

Editorial

Liebe Leserinnen und Leser,

vor Ihnen liegt nunmehr die bereits achtzehnte Ausgabe des E-Journals **Anwendungen und Konzepte in der Wirtschaftsinformatik (AKWI)**.

Wie schon in den bisherigen Ausgaben werden viele charakteristische Bereiche der Wirtschaftsinformatik behandelt: Geschäftsprozesse im weiteren Sinne, Systeme wie das SAP ERP System, der Betrieb von Systemen und neuere Ansätze aus dem Bereich der künstlichen Intelligenz, gerade ChatGPT. Dies erfolgt in verschiedenen Anwendungen wie logistische Fragestellungen und im Rahmen von Forschungsarbeiten sowie Projektarbeiten mit Industrieunternehmen.

Mehrheitlich wurde an Themen mit einem direkten Bezug zur künstlichen Intelligenz gearbeitet. Im Einzelnen finden Sie Arbeiten zur GPT-3.5-unterstützten Entwicklung von Prozessmodellen, der Spezifikation einer KI-Funktion am Beispiel des Fußgängerschutzes, Analyse der Chancen und Risiken von KI-basierten Large Language Models am Beispiel von ChatGPT in der innerbetrieblichen IT, Prompt Engineering mit ChatGPT 3.5 für Midjourney am Beispiel eines Use Cases für Wildlife Photography, ein auf maschinellem Lernen basierender Ansatz zur Vorhersage der Mitarbeiterfluktuation mit Schwerpunkt auf der Vorhersage der Austrittsgründe, Ansätze zur Entwicklung und Anwendung intelligenter Handelssysteme mittels Machine Learning und Deep Learning und damit methodisch verwandt Automatisierungspotenziale von Büro- und Verwaltungsarbeiten anhand von Large Language Models am Beispiel von ChatGPT.

Die weiteren Arbeiten adressieren die seit längerem existierenden und im Grunde weiterhin jungen Themen nämlich die Entwicklung eines Unterweisungsmoduls für Microsoft Dynamics NAV 2016, der Entwurf eines Anwendungsfalls für die Resilienz der Lieferkette auf der Grundlage von Process Mining, der Einfluss eines Business Intelligence Systems auf die Analyse am Beispiel der Stadtverwaltung Flensburg in Deutschland, der Vergleich und die Evaluierung verschiedener Clustering Algorithmen und Methoden zur Anwendung auf Wetterdaten zum Definieren von Wetterereignisprofilen und deren Charakteristiken, die Anwendung von Object Centric Process Mining am Beispiel einer Fallstudie, die Organisatorische Integration von User Experience-Mitarbeitern in der agilen Softwareentwicklung, die Kundensegmentierung im Vollmachtsgeschäft mittels Fuzzy Clustering, die Entwicklung einer plattformunabhängigen Chatbot-Frontend-Anwendung, Ausgestaltung eines digitalen Zwillings im Rahmen von Manufacturing-X und schließlich die Entwicklung eines Referenzmodells zur Strukturierung des Einsatzes von Self-Service Business Intelligence in Unternehmen.

Über Ihr Interesse an der Zeitschrift freuen wir uns und wünschen Ihnen Freude bei der Lektüre.

Regensburg, Fulda, Luzern und Wildau, im Dezember 2023.

Frank Herrmann, Norbert Ketterer, Konrad Marfurt und Christian Müller



Christian Müller



Konrad Marfurt



Norbert Ketterer



Frank Herrmann

Experimente zur GPT-3.5-unterstützten Entwicklung von Prozessmodellen

Carlo Simon, Stefan Haag und Lara Zakfeld
Fachbereich Informatik
Hochschule Worms
Erenburgerstr. 19, 67549 Worms
Email: {simon,haag,zakfeld}@hs-worms.de

ZUSAMMENFASSUNG

Noch können Prozessmodelle nicht automatisch aus textuellen Beschreibungen generiert werden. Die Vorteile einer solchen Technologie lägen in kürzeren Entwicklungszeiten und einer kohärenteren, weil algorithmischen Interpretation des Inputs. Dazu müssten Methoden der formalen Modellierung und der Linguistik kombiniert werden. Die aktuellen GPT-Modelle eröffnen gerade hinsichtlich der linguistischen Analyse neue Möglichkeiten.

Dieser Beitrag untersucht, ob die Fähigkeiten von GPT-3.5, Texte zusammenzufassen, kausale Zusammenhänge zu erkennen oder Begriffe zu klassifizieren, auch ausreichen, um formale Prozessmodelle zu erstellen. Hierzu wurden systematisch Experimente zur Nutzung von GPT-3.5 durchgeführt. Dabei konnte beobachtet werden, dass die Resultate abhängig von den gewählten Transformer-Optionen erheblich variieren. Die Protokolle der Experimente werden in diesem Beitrag zusammengefasst und aus den Ergebnissen dann mögliche, künftige Forschungsaktivitäten abgeleitet.

SCHLÜSSELWÖRTER

GPT-3.5, Prozessmodellierung, Künstliche Intelligenz

DISCLAIMER

Dieser Beitrag ist eine übersetzte und überarbeitete Version des Papers *Experiments on GPT-3 Assisted Process Model Development* in den *Proceedings of the 37th International ECMS Conference on Modelling and Simulation* (Simon et al., 2023).

DIE NÄCHSTE DISRUPTION?

ChatGPT verwendet eine Chat-ähnliche Oberfläche, um mit Benutzern in natürlicher Sprache zu kommunizieren. Das von OpenAI entwickelte ChatGPT ist damit eine Umgebung, die maschinelles Lernen und künstliche Intelligenz einem breiten Publikum zugänglich macht. Viele Autoren schreiben ihr das Potenzial zu, bestehende Technologien zu verändern oder abzulösen. Eine von vielen Veröffentlichungen zum Thema ist <https://www.theguardian.com/commentisfree/2023/jan/07/chatgpt-bot-excel-ai-chatbot-tech>.

ChatGPT kann Fragen in natürlicher Sprache besser beantworten als bisherige Systeme. Hierzu wurde das zugrundeliegende Sprachmodell GPT-3.5, ein Large Language Model (LLM), mit über 235 Millionen Internet-Textdokumenten in Englisch und über 10

Millionen Textdokumenten in anderen Sprachen trainiert (Kublik und Saboo, 2022, S. 6). Zudem wurden Sourcecode und Dokumentationen von GitHub sowie Stack Overflow-Diskussionen genutzt. Hierdurch kann ChatGPT auch Fragen zur Softwareentwicklung und zur Erstellung formaler Spezifikationen beantworten.

Aus der prinzipiellen Fähigkeit, mit formalen Sprachen umzugehen, ergibt sich die Frage, welche Auswirkungen diese Technologie auf die Themenfelder Modellierung und Simulation haben könnte. Dieser Beitrag untersucht, ob und wie GPT-3.5 Menschen beim Entwickeln von simulierbaren Modellen assistieren kann.

In der vorliegenden Arbeit werden daher Prozesse, ihre natürlichsprachliche Beschreibung sowie deren formale Modellierung betrachtet. Das praktische Potential von GPT-3.5 wird anhand einer Reihe systematischer Experimente untersucht. Dabei haben sich die drei folgenden Herausforderungen und zugehörigen Lösungsansätze herauskristallisiert:

1. Trotz der riesigen Textmenge zum Trainieren von GPT-3.5 stehen nur vergleichsweise wenige natürlichsprachliche Prozessbeschreibungen zusammen mit ihrer formalen Spezifikation frei im Internet zur Verfügung. Die Autoren vermuten, dass insbesondere die grafische Repräsentation der Prozessmodelle den Zugriff erschwert. Sie haben daher ein Training formaler Methoden in den Dialog integriert.
2. Prozessmodelle von Organisationen basieren auf aktuellen Gegebenheiten. Das Trainingsmaterial für GPT-3.5 ist aber allgemeiner Text aus dem Internet, der bis 2021 erstellt wurde. Die Autoren haben daher versucht, aktuelle Fakten über in den Dialog integrierte Trainings in das System einzuspielen.
3. Menschen entwickeln Prozessbeschreibungen systematisch. Die Dialoge sollten daher so gestaltet werden, dass Nutzer ein „natürliches“ Modellierungserlebnis haben. Die Autoren haben hierfür die Fähigkeit von GPT genutzt, aufeinander folgende Anfragen als Dialog zu interpretieren, um das System Prozessmodelle in einer ähnlichen Reihenfolge erzeugen zu lassen, wie dies menschliche Modellierer üblicherweise tun. In diesen Dialogen werden die Modelle aufgebaut.

Die Idee des vorgestellten Ansatzes besteht darin, GPT-3.5 in einem fragengestützten Dialog schrittweise passende formale Spezifikationen aus einer gegebenen Prozessbeschreibung entwickeln zu lassen. Die Forschungsfrage lautet:

Welche Schritte im Prozess der (Prozess-)Modellentwicklung kann GPT-3.5 unterstützen?

Die vorliegende Arbeit unterliegt Einschränkungen: Da OpenAI seine Modelle permanent aktualisiert, ist es praktisch unmöglich, beim Wiederholen der Experimente auch die erzielten Ergebnisse zu reproduzieren. Dieser Beitrag ist wie folgt aufgebaut: Zunächst bettet der Abschnitt VERWANDTE ARBEITEN das Paper in einen größeren Rahmen ein und bietet grundlegende Informationen zu LLMs. Der Abschnitt EIN EXPERIMENTALLABOR beschreibt die verwendeten Systeme und gibt einen Überblick über die experimentellen Phasen. Die Experimente und ihre Ergebnisse werden in den drei folgenden Abschnitten beschrieben. Der Artikel endet mit ZUSAMMENFASSUNG UND AUSBLICK.

VERWANDTE ARBEITEN

Komplexe Tätigkeiten wie das Ableiten formaler Prozessspezifikationen aus narrativen Prozessbeschreibungen werden üblicherweise in Teilschritte gegliedert. Abbildung 1 zeigt, wie menschliche Modellierer typischerweise eine gegebene Prozessbeschreibung in ein Prozessmodell überführen. Zunächst werden die Prozesselemente identifiziert (Rollen und Akteure, Ereignisse und Auslöser, Aktivitäten, Geschäfts- und Informationsobjekte). Darauf aufbauend wird der Prozessfluss gegliedert (in Sequenzen, Alternativen, Iterationen und Nebenläufigkeiten). Schließlich wird das Modell um quantitative Informationen ergänzt. Hierfür muss ein Mensch den zu modellierenden Prozess verstehen.

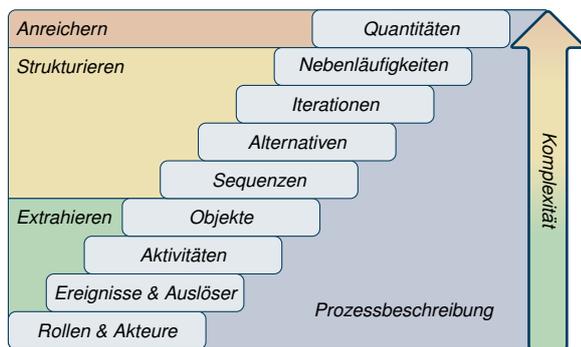


Abbildung 1: Ein Vorgehensmodell für die formale (Prozess-)Modellierung

GPT extrahiert hingegen aus einer Prozessbeschreibung zuvor gelernte Muster inklusive deren Transformation in Modelle, um diese zu übersetzen. Diese werden dann zu einem Ganzen verbunden. Benutzer können bei Bedarf zwischen alternativen Mustern wechseln, und so unterschiedliche Ergebnisse erzielen.

(Kecht et al., 2023) untersuchen anhand von Kundengesprächen auf Twitter die Fähigkeit von Chatbots, Geschäftsprozesse aus natürlicher Sprache zu extrahieren. Sie stoßen aber aufgrund der begrenzten Fähigkeiten von Chatbots, Antworten zu generieren, an die Grenzen solcher Systeme.

Da das Transformer-Modell von GPT-3.5 leistungsfähiger und weniger deterministisch als Chatbots ist, ist der in diesem Beitrag beschriebene Ansatz ein völlig anderer. Um dies zu verdeutlichen, werden zunächst die Begriffe Sprachbasierter Agent, Transformer-

Sprachmodell und insbesondere GPT-3.5 erläutert. Schließlich wird der hier vorgestellte Ansatz in einen größeren Forschungskontext eingebettet.

Zum Zeitpunkt der ursprünglichen Einreichung waren weder GPT-4 noch ein endgültiger Zeitplan für dessen Veröffentlichung verfügbar. In der Zwischenzeit durchgeführte initiale Tests mit ChatGPT Plus (das einen etwas eingeschränkten Zugang zu GPT-4 ermöglicht) zeigen jedoch, dass die hier gesammelten Erfahrungen ihre allgemeine Gültigkeit behalten.

Sprachbasierte Agenten

Sprachbasierte Agenten (Conversational Agents) simulieren menschenähnlich Unterhaltungen mit Benutzern. Diese Unterhaltungen können dabei als Dialoge ablaufen, die frühere Gesprächsverläufe mit berücksichtigen (Adamopoulou und Moussiades, 2022). Diese Systeme lassen sich irgendwo im folgenden Spektrum verorten:

- Regelbasierte Chatbots reagieren auf bestimmte Stichworte (vgl. Weizenbaum (1966)).
- Sprachassistenten nutzen maschinelles Lernen, um Eingaben in natürlicher Sprache zu „verstehen“. So vermeiden sie, nur auf bestimmte Stichworte anzusprechen. Sie reagieren somit menschenähnlicher, sind aber immer noch regelbasiert (vgl. Lopez et al. (2017)).

Transformer-Sprachmodelle

Transformer-Sprachmodelle verwenden Deep Learning, um Eingabesequenzen in Ausgabesequenzen umzuwandeln. Ein solches Modell wird probabilistisch trainiert, um Wörter und ihre Kontexte mit anderen Wörtern und deren Kontexten in Beziehung zu setzen. Das Modell kann dann entscheiden, welche dieser Kontexte mit höherer Wahrscheinlichkeit mit einem bestimmten Thema verbunden sind als andere. Bei einer Folge von Eingabewörtern berücksichtigt ein Transformer zudem den Kontext, um so das wahrscheinlichste Wort zu finden, das als nächstes folgt (Vaswani et al., 2017).

Jede Benutzereingabe und Systemausgabe wird in numerischer Form in einem Kontextvektor gespeichert. Dieser Vektor stellt das aktuelle Gedächtnis des Modells dar und hält es während eines Dialogs im Kontext. Durch Benutzereingaben kann er Informationen aufnehmen, die in den ursprünglichen Trainingsdaten nicht enthalten waren (Vaswani et al., 2017).

Die probabilistische Natur dieser Sprachmodelle macht sie zu natürlichen Feinden deterministischer Prozessbeschreibungen. Selbst vom Benutzer eingegebene Prozessfakten können von anderen zuvor gelernten Mustern überlagert werden. Das Problem, Faktenwissen verbindlich zu integrieren, ist noch ungelöst.

GPT-3.5

GPT-3.5 ist ein hochmoderner Transformer der Non-Profit-Organisation OpenAI. GPT-3.5 wurde mit einer großen Menge von Textdaten trainiert und ist in der Lage, natürlich anmutende, flüssige und kohärente Texte zu erzeugen. Aufgrund des großen Trainingskorpus kann ein spezifisches Feintuning durch „learning by doing“ zu überraschend guten Ergebnissen führen.

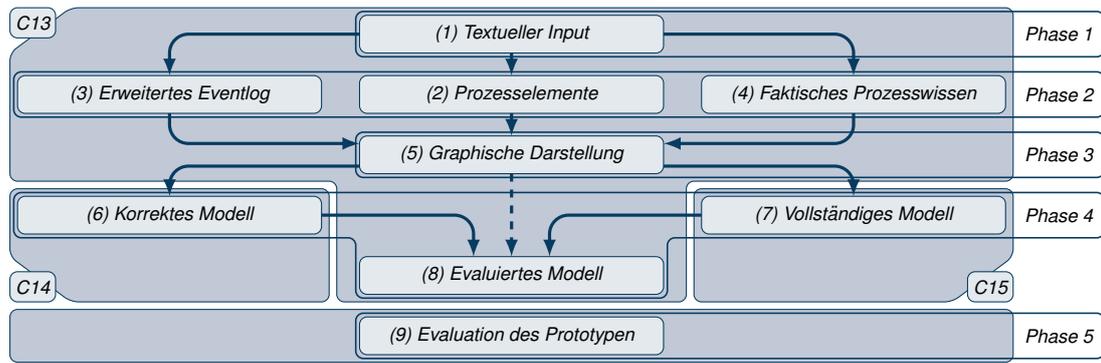


Abbildung 2: Phasen und Artefakte eines Forschungsplans zur Extraktion von Modellen aus natürlichsprachlichen Texten, siehe (Haag und Simon, 2022)

Eine der größten Stärken von GPT-3.5 ist seine Vielseitigkeit in einem breiten Spektrum von Anwendungen. Es kann für Aufgaben wie Übersetzung, Textvervollständigung, Schreiben von Zusammenfassungen, aber auch dem Erstellen von Chatbots und virtuellen Assistenten eingesetzt werden. GPT-3.5 zeichnet sich auch durch die Fähigkeit aus, kontextabhängige Texte zu generieren, die sich an die Präferenzen und den Sprachstil des Benutzers anpassen.

Transformation von Sprache in ein Modell

(Mendling et al., 2014) benennen 25 Herausforderungen der semantischen Prozessmodellierung, unter anderem als 13. Herausforderung diejenige, natürlichsprachliche Textbeschreibungen in Modelle zu transformieren. Abbildung 2 stellt diese Herausforderung in einen Kontext und nennt die verschiedenen Artefakte, die zur Lösung erstellt werden müssen. Zudem wird gezeigt, wie die Ergebnisse mit zwei der anderen Herausforderungen zusammenhängen, nämlich die Korrektheit und die Vollständigkeit von Modellen zu verifizieren (Haag und Simon, 2022).

Die meisten Beiträge, die sich mit solchen Transformationen beschäftigen, verwenden vorverarbeitete Daten aus Texten (z.B. Fliedl et al. (2005); Nolte (2020)). Aktuelle NLP-Technologien wie GPT-3.5 oder andere Transformer wurden bisher noch nicht untersucht.

Die Aufgabenstellung in ihrer allgemeinen Form ist aus zwei Gründen noch ungelöst:

1. Regelbasierte Automaten handhaben die Komplexität formaler Modellierung bislang unzuverlässig.
2. Methoden, die auf maschinellem Lernen basieren, benötigen in der Trainingsphase große Mengen an korrekt annotierten Beispielen. Diese sind jedoch schlicht nicht in ausreichender Menge verfügbar.

GPT-3.5 scheint diese Einschränkungen (zumindest in Ansätzen) zu überwinden.

EIN EXPERIMENTALLABOR

Die Experimente wurden mit den von OpenAI als „Labor“ zur Verfügung gestellten Werkzeugen durchgeführt. Jedes Experiment wurde in einem Laborjournal dokumentiert, das auf der Website der Arbeitsgruppe veröffentlicht ist. Es enthält alle Voreinstellungen, beschreibt die Experimente und umfasst die vollständigen Gesprächsverläufe (Haag, 2023).

GPT-3.5 und ChatGPT

Die umgangssprachlich als ChatGPT bezeichnete Web-Oberfläche (<https://chat.openai.com/>) ist genau genommen eine natürlichsprachliche Schnittstelle zu den Sprachmodellen GPT-3.5 und GPT-4. Auf deren Basis erstellt die Applikation kontextrelevante Antworten, die denen eines Menschen ähnlich sind (OpenAI, 2022). ChatGPT ist in der Lage Dialoge zu führen, um Antworten aufgrund bereits durchgeführter Interaktionen zu verbessern.

Ein an das System gesendeter Text, der als „Programmiersprache in normalem Englisch“ (oder einer anderen unterstützten Sprache) fungiert, sowie ein von ihm generierter Text wird als Prompt bezeichnet. Drei Arten von Input-Prompts werden unterschieden (Tingiris, 2021, S. 6-10):

Zero-shot prompts geben eine Aufgabenstellung oder einen Text vor, auf denen weitere Prompts aufgebaut werden sollen.

One-shot prompts sind Beispiele, anhand derer sich ableiten lässt, wie eine Aufgabe zu lösen ist.

Few-shot prompts liefern mit Hilfe mehrerer Beispiele Muster, denen gefolgt werden soll.

OpenAI Playground

Auch der OpenAI Playground (<https://platform.openai.com/playground>) ist eine webbasierte Schnittstelle, die benutzt werden kann, ohne eine Zeile Code zu schreiben (Tingiris, 2021, S. 20). Er bietet Zugang zu verschiedenen GPT-basierten Sprachmodellen, wobei *text-davinci-003* das zum Zeitpunkt der ursprünglichen Einreichung aktuelle Standardmodell war. Mehrere Parameter beeinflussen die Ausgabe des Systems (Brown et al., 2020), unter anderen die folgenden:

Temperatur steuert die Determiniertheit der Ausgabe. Je niedriger der Wert im Intervall [0,1] liegt, desto deterministischer und repetitiver verhält sich das System. *Maximale Länge* begrenzt die Länge des Prompts, der in Token - einer numerischen Darstellung von Wortteilen - unterteilt wird. Für die meisten Sprachmodelle liegt die maximale Länge bei 2.048 Token, was etwa 1.500 Wörtern entspricht (Tingiris, 2021, S. 12). Bei Erreichen dieser Grenze stoppt die Ausgabe, d.h. es wird weder gekürzt noch umgeschrieben.

Token Highlighting färbt Token gemäß der Wahrscheinlichkeit, mit die sie generiert wurden.

Bei den Experimenten wurde die Werte für diese Parameter variiert:

- Teilweise wurde die Temperatur auf 0 statt der üblichen 0,7 gesetzt, um aufgrund der erzielten Determiniertheit Referenzprompts zu bilden.
- Die maximale Länge wurde erhöht. Dies beeinflusst die tatsächlich gegebene Antwort nicht, ermöglicht aber längere Gesprächsverläufe.
- Volles Token Highlighting wurde erlaubt, um Token mit vergleichsweise geringer Wahrscheinlichkeit zu erkennen.

Alle anderen Parameter wurden auf ihren Standard-einstellungen belassen.

Phasen der Experimente

Die Experimente wurden in drei Phasen durchgeführt:

1. *Erweitern einer Beschreibung*: GPT-3.5 wurde eingesetzt, um eine gegebene Prozessbeschreibung mit weiteren Details anzureichern.

2. *Informationsextraktion*: Anschließend wurde GPT-3.5 aufgefordert, prozessrelevante Informationen aus dieser erweiterten Beschreibung zu extrahieren.

3. *Formale Transformation*: Schließlich sollte GPT-3.5 diese Antworten in formalisierter Syntax darstellen.

Nach dieser kurzen Einführung in die technische Umgebung werden in den folgenden Abschnitten der experimentelle Aufbau und die beobachteten Ergebnisse beschrieben.

ERWEITERN EINER BESCHREIBUNG

In einem vorgelagerten Schritt wurde ein einfacher Prozess aus der Literatur übernommen (Simon et al., 2022a). Seine textuelle Beschreibung und eine Visualisierung sind in Abbildung 3 dargestellt.

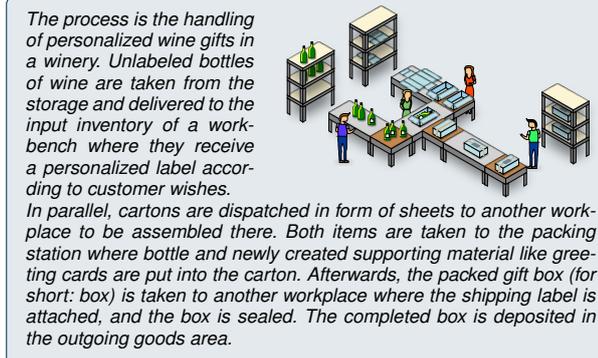


Abbildung 3: Der Beispielprozess und seine Visualisierung

Aus diesem Text sollte GPT-3.5 Aktivitäten, Ereignisse und Prozessstrukturen extrahieren. Aufgrund der vielversprechenden Ergebnisse war das erste Ziel, die Komplexität mit Hilfe von GPT-3.5 zu erhöhen.

Konfiguration des Experiments

Dem OpenAI Playground wurde die ursprüngliche Prozessbeschreibung und die Liste der in Abbildung 4 dar-

gestellten Modifikationen vorgelegt. Verbunden damit war die Anweisung, hieraus eine erweiterte Prozessbeschreibung zu erstellen. Es wurden zehn Durchläufe mit einer Temperatur von 0 und weitere zehn Durchläufe mit einer Temperatur von 0,7 durchgeführt.

1. *The process starts with a customer order*
2. *From this order, amount and type of wine bottles are used to determine the bottles to be packed*
3. *The process not only handles gifts, but also standard orders. If an order is a gift, the personalized label for the bottle needs to be printed at an own workplace. If it is not a gift, this step isn't necessary as a stock label is used*
4. *The wine bottles are not necessarily packaged in size one, but there are different package sizes: 1, 3, 6, 12 bottles per package*
5. *Before sealing the box, a quality check is conducted for breakage, leakages, package contents, and overall looks*
6. *If during quality check an issue is discovered, the process should continue with the corresponding process step. The possible issues are: damaged bottle, damaged label on bottle, missing or damaged gift card, insufficient stuffing material, damaged cartons, missing or damaged shipping label*
7. *The process doesn't start immediately after receiving an order, but all orders are processed once a day*

Abbildung 4: Modifikationen zur Erweiterung der Beschreibung

Die so entwickelte Beschreibung ist in Abbildung 5 angegeben. Es handelt sich dabei um eine leicht bearbeitete, konsolidierte Version der Antworten von GPT-3.5. Die Gründe für die Bearbeitung werden nachfolgend erläutert. Trotz der Bearbeitung werden die beeindruckenden Potenziale großer Sprachmodelle deutlich.

Ergebnisse der ersten Phase

Ziel der ersten Phase war das Erstellen einer Prozessbeschreibung als Basis für die folgenden Phasen.

Die linguistischen Fähigkeiten ermöglichen es GPT-3.5, der Beschreibung Details hinzuzufügen und Prozesselemente zu modifizieren. Auf die Frage nach der Güte der generierten Prozessbeschreibungen bemängelte das Modell jeweils zu Recht das Fehlen integraler Konzepte des Geschäftsprozessmanagements wie etwa Leistungsmessung, Prozessziele oder die Anbindung an einen größeren organisatorischen Kontext.

The daily process of handling customer orders starts with the determination of the amount and type of wine bottles required for each order. Customers can order packages of one, three, six or twelve bottles each. The bottles are taken from storage to the input inventory at the workbench. If the order is a gift, the corresponding bottles receive a personalized label according to customer specifications that is printed at a small workstation. If it is a standard order, the bottles receive a stock label instead. In parallel, cartons are dispatched to the assembly workplace, where they are assembled. At the packing station, the bottles and materials are put into the carton. This includes printing of these materials, such as accompanying documents and, for gift orders, greeting cards. Next, a quality check is conducted to ensure that there are no breakages, leakages, or discrepancies in the package contents and, if it is a gift package, to verify that the overall appearance is suitable. If any issues are found during the quality check, the process continues with the corresponding step to rectify the issue. The shipping label is attached to the box at another workplace, where it is also sealed. Lastly, the package is deposited in the outgoing goods area.

Abbildung 5: (Teil-)automatisch erweiterte Beschreibung

Tabelle 1: Berechnete Wahrscheinlichkeiten für das Auftreten verschiedener Token in 10 OpenAI Playground Prompts bei Temperatur 0

Token	Lauf 1	Lauf 2	Lauf 3	Lauf 4	Lauf 5	Lauf 6	Lauf 7	Lauf 8	Lauf 9	Lauf 10
<i>corresponding</i>	27,41	28,46	28,18	27,07	27,83	28,15	28,06	27,89	27,96	28,73
<i>personalized</i>	23,31	22,18	22,67	22,68	22,80	22,43	22,69	22,96	23,19	21,90
<i>appropriate</i>	19,55	19,74	19,47	19,56	19,54	19,67	19,57	19,44	19,36	19,70
<i>label</i>	5,62	5,62	5,79	5,72	5,71	5,81	5,73	5,78	5,74	5,64
<i>necessary</i>	5,34	5,22	5,25	5,25	5,30	5,21	5,22	5,18		5,26
<i>respective</i>									5,20	
Summe	81,23	81,22	81,36	81,28	81,18	81,27	81,27	81,25	81,45	81,23

Alle zehn Durchläufe bei einer Temperatur von 0 führten zur genau gleichen Formulierung. Die zehn Durchläufe bei einer Temperatur von 0,7 hingegen lieferten jeweils unterschiedliche Ergebnisse hinsichtlich Formulierung und Detaillierungsgrad. Für eine verlässlichere Wiederholbarkeit sollte daher stets eine Temperatur von 0 verwendet werden. Eine Wiederholbarkeit ist aber selbst dann nicht garantiert, wie das Beispiel des Begriffs „corresponding“ im vierten Satz der in Abbildung 5 gezeigten Ausgabe zeigt:

Tabelle 1 listet die von GPT-3.5 berechneten Wahrscheinlichkeiten für das Auftreten dieses Tokens in zehn Durchläufen. Obwohl die Temperatur auf 0 gesetzt wurde, variieren die Werte zwischen den verschiedenen Läufen um etwa 4,6%. Die alternativen Terme besitzen immer niedrigere Wahrscheinlichkeiten und in allen Durchläufen sind die 3 folgenden Terme immer die gleichen. Der geringste Abstand zwischen den beiden obersten Wahrscheinlichkeiten beträgt jedoch vergleichsweise geringe 4,1%. Im neunten Durchlauf wechseln die fünftplatzierten Token, wenn dies auch keinen Einfluss auf die tatsächliche Ausgabe hat. Aufgrund dieser Beobachtungen kann aber nicht ausgeschlossen werden, dass ein solcher Wechsel auch auf der ersten Position vorkommen könnte. Die daraus folgende Änderung des Kontextvektors könnte Auswirkungen auf den weiteren Verlauf der Ausgabe haben.

Die Prozessbeschreibung in Abbildung 5 ist nicht die Kopie eines einzelnen Prompts, sondern eine leicht bearbeitete Kombination aus mehreren. Keine einzelne Version war aus menschlicher Sicht vollkommen überzeugend. So war beabsichtigt, die zuletzt hinzugefügte Information (Nr. 7: „The process doesn’t start immediately after receiving an order, but all orders are processed once a day“) als ersten Satz in die Beschreibung aufzunehmen. Stattdessen wurde diese Aussage nahezu wörtlich immer am Ende platziert. Mit Ausnahme des ersten stammen alle anderen Sätze der finalen Version aus den verschiedenen Prompts. Dies sollte die Auswirkungen des Eingriffs möglichst gering halten. Jedoch fehlen in der Beschreibung noch wichtige Informationen. So werden keine Akteure, Rollen oder Geschäfts- und Informationsobjekte definiert. Auch wird der Kontrollfluss ohne die explizite Nennung von Kontrollstrukturen beschrieben. Dies motivierte die Entscheidung, GPT-3.5 auf die Fähigkeit zu untersuchen, implizite Informationen aufzudecken und fehlende Informationen abzuleiten.

INFORMATIONSEXTRAKTION

In der zweiten Phase des Experiments wurde untersucht, inwiefern GPT-3.5 prozessrelevante Informationen aus einer Prozessbeschreibung extrahieren kann. Hierfür wurden 13 Fragen formuliert, die dem System gestellt wurden.

Konfiguration des Experiments

Dem System wurde der Prozess aus Abbildung 5 und die Aufforderungen aus Abbildung 6 bei einer maximalen Länge von 2.048 Token eingegeben. Es wurden vier verschiedene Settings betrachtet.

1. Provide an ordered list of the workplaces that are used in this process.
2. Provide an ordered list of business roles or actors in this process.
3. Provide an ordered list of events that occur in this process.
4. Provide an ordered list of activities that need to be conducted in this process.
5. Provide an ordered list of business objects in this process.
6. Provide an ordered list of information objects in this process.
7. Is there a sequence in this process? If so, which?
8. Is there a decision in this process? If so, which?
9. Is there an alternative in this process? If so, which?
10. Is there an exclusivity in this process? If so, which?
11. Is there a concurrency in this process? If so, which?
12. Is there an iteration in this process? If so, which?
13. Is there a loop in this process? If so, which?

Abbildung 6: Aufforderungen zur Analyse der Prozessbeschreibung

1. Zunächst wurden alle Aufrufe bei einer Temperatur von 0 durchgeführt. Der Prozess wurde zusammen mit jeder Frage in einem separaten Prompt an das System gesendet. Dadurch konnte das System keinen fragenübergreifenden Kontextvektor aufbauen.
2. Die Temperatur für den zweiten Versuch wurde ebenfalls auf 0 gesetzt. Allerdings wurden die Fragen im selben Prompt nach der jeweils vorhergehenden Antwort an das System gesendet, so dass das Modell dieses Mal einen Kontextvektor erstellen konnte.
3. Für den dritten Versuch wurde die Temperatur auf 0,7 gesetzt. Die Fragen wurden wie im zweiten Versuch an das System gesendet. Zur Prüfung der Wiederholbarkeit wurde dieser Versuch fünfmal wiederholt.

4. Schließlich wurde die Reihenfolge der Fragen randomisiert, die Temperatur jedoch auf 0,7 belassen. Diese Prozedur wurde zwanzigmal wiederholt.

Eine vollständige Untersuchung aller 13! möglichen Permutationen scheitert an der Anzahl der notwendigen Auswertungen.

Die Antworten von GPT-3.5 wurden, wie nachstehend erläutert, aus einer menschlichen Perspektive auf Vollständigkeit und Korrektheit hin bewertet.

Ergebnisse der zweiten Phase

Folgende weitere Beobachtungen wurden gemacht:

- Explizite Informationen über Arbeitsplätze, Tätigkeiten und einige Informationen über den Kontrollfluss wurden erfolgreich extrahiert.
- Implizite Informationen wurden geraten:
 - Die Antworten zu Rollen und Akteuren reichten von insgesamt zwei sehr generischen Rollen (Kunden und Arbeiter) bis hin zu zehn Rollen (ein dedizierter Mitarbeiter je identifizierter Aktivität). Allerdings wurden in drei Durchläufen überhaupt keine Rollen identifiziert, sondern stattdessen Arbeitsplätze ausgegeben.
 - Geschäfts- und Informationsobjekte wurden ebenfalls geraten und beinhalteten fälschlicherweise u. a. Materialien, Arbeitsplätze oder Kontrollstrukturen.
- Die Aktivitäten wurden nahezu korrekt erkannt.
- Die Extraktion von Ereignissen schlug faktisch fehl. Die Versuche lieferten eher Listen von Aktivitäten, die teilweise nicht einmal der Beschreibung entsprachen.
- Die Kontrollstrukturen zeigen ein unklares Bild:
 - Sequenzen wurden in allen Läufen identifiziert. Sie sind jedoch in der Reihenfolge angeordnet, in der die Aktivitäten in der Beschreibung erscheinen, was nicht der beabsichtigten Reihenfolge innerhalb des Prozesses entspricht. Auch wurden einige Aktivitäten, die nebenläufig stattfinden sollten, in eine Sequenz gegliedert.
 - Die Frage nach der Existenz von Nebenläufigkeiten wurde in 22 der 27 Läufe bejaht, auch wenn zuvor nur eine einzige, durchgehende Sequenz identifiziert wurde.
 - In 23 Läufen wurde die Alternative korrekt erkannt.
 - In 15 Läufen wurde die beabsichtigte Iteration nach der Qualitätskontrolle identifiziert. In 3 Läufen wurden Angaben zu Flaschenmengen, Kartonmontage oder Etikettendruck als Iterationen fehlinterpretiert. In 9 Durchläufen wurde keinerlei Iteration erkannt.

Aus Sicht der Autoren ist dies ein bemerkenswertes Ergebnis, wenn man bedenkt, dass es sich um erste Versuche mit dieser neuen Technologie handelt. Weitere Verbesserungen der Ergebnisse scheinen möglich, wenn der Kontextvektor systematisch durch passend formulierte Fragen in der richtigen Reihenfolge aufgebaut wird.

FORMALE TRANSFORMATION

Die bisher besprochenen Transformationen haben Reintext erzeugt. In diesem Beitrag geht es jedoch um die GPT-3.5-unterstützte Entwicklung von Prozessmodellen. Solche Modelle müssen in einer formalen Spezifikationsprache beschrieben werden, die speziellen syntaktischen Regeln folgt, wie z.B. Fluss- oder BPMN-Diagramme. Anspruchsvoller wäre eine Sprache mit

Ausführungssemantik wie Petri-Netze. Die extrahierten Ergebnisse müssen dann zusammengeführt werden, um eine Darstellung des Prozesses zu erhalten. Der Gesamtprozess besteht aus Aktivitäten, Ereignissen, Prozessstrukturen und anderen typischen Prozessmodellkomponenten. Diese Elemente müssen schließlich in einer Form abgebildet werden, die von einer Prozessmodellierungsumgebung verarbeitet werden kann.

Konfiguration des Experiments

Die Überführung natürlichsprachlicher Aussagen durch die Aufforderungen aus Abbildung 6 in eine formale Darstellung erfordert eine Modellierungsumgebung mit einer offenen API oder einem Klartext-Dateiformat.

Die Autoren haben sich für die Spezifikationsprache des Process-Simulation.Center (P-S.C) entschieden. Das P-S.C ist ein integriertes Managementsystem, das Modellierung, Simulation und Dokumentation von Prozessen mit Hilfe von Petri-Netzen ermöglicht. Das Tool bietet zudem Spezifikationsprachen für Organigramme und Swimlanes und unterstützt die Verknüpfung von Prozessen mittels Prozesslandkarten (Simon et al., 2022b).

Diese Versuche wurden in zwei Abschnitten durchgeführt. Der erste Abschnitt war wie folgt strukturiert:

1. Aufruf der Prozessbeschreibung bei Temperatur 0.
2. Abfrage der Arbeitsplätze, da aufgrund der vorherigen Versuche eine bessere Extraktion der anderen Prozesselemente erwartet wurde.
3. Extrahieren der Aktivitäten in einer Verb-Substantiv-Form und Formatieren der Ausgabe für das P-S.C.
4. Extrahieren der Ereignisse und formatieren. Da das Wort „event“ nicht verlässlich funktionierte, wurde stattdessen das Wort „trigger“ verwendet.
5. Extrahieren der Strukturen „sequence“, „branching“, „merging“, „iteration“ und „concurrency“ in der angegebenen Reihenfolge und formatieren der Ausgabe entsprechend der jeweiligen Struktur.

Dieses Experiment wurde einmal bei einer Temperatur von 0 und fünfmal in ChatGPT mit unterschiedlicher Reihenfolge der Aufforderungen bzgl. der Kontrollstruktur wiederholt.

Im zweiten Abschnitt wurden die an das System gestellten Aufforderungen umformuliert und mit Beispielen angereichert. Die ersten Fragen wurden in der selben Reihenfolge gestellt, jedoch wurden diejenigen nach den Kontrollstrukturen in „branching“, „merging“, „concurrency“, „iteration“ und „sequence“ geändert. Diese Anweisungen wurden einmal bei Temperatur 0 und fünfmal bei Temperatur 0,7 durchgeführt.

Ergebnisse der dritten Phase

Der erste Abschnitt lieferte passable Ergebnisse, wenn GPT-3.5 mit genauen Anweisungen und Beispielen für die korrekte Formatierung versorgt wurde. Im weiteren Verlauf der Konversationen entfernte sich das System jedoch immer mehr von den bereits (mutmaßlich) identifizierten Prozesselementen und -strukturen.

- Eine korrekte Spezifikation verlangt eine definierte Syntax und Schlüsselwörter. Daneben erhalten die einzelnen Elemente (sprechende) Namen. Bei einer Temperatur von 0,7 nummerierte ChatGPT die Elemente in mehreren Läufen jedoch durch, statt sich an den vorgegebenen Beispielen zu orientieren. Syntaktisch wichtige Zeichen wie Semikolons wurden nur in Ausnahmefällen gesetzt. Bei einer Temperatur von 0 arbeitete das System insgesamt besser, aber immer noch mit den grundlegend gleichen Fehlern.
- Das Extrahieren von Ereignissen (bzw. Triggern) funktionierte schlechter als das von Aktivitäten. Eine Umformulierung der Aufforderung und die Angabe von weiteren Beispielen verbesserten die Ergebnisse.
- Kontrollstrukturen wurden nicht erfolgreich erkannt. Die Resultate ließen sich mit einer veränderten Reihenfolge der Abfragen leicht verbessern. Trotzdem übersah das System weiterhin beabsichtigte Strukturen oder integrierte nicht vorhandene. Die Aktivitäten wurden so serialisiert, wie sie in der Beschreibung erscheinen, und explizite Korrekturen wurden größtenteils ignoriert. GPT-3.5 kann rudimentäre Prozessmodelle für Aktivitäten und Ereignisse erstellen. Allerdings sind die Resultate bzgl. der Kontrollstrukturen faktisch unbrauchbar. Unter diesem Gesichtspunkt können die Probleme von GPT-3.5, eine gegebene formale Syntax korrekt umzusetzen, vorerst zurückgestellt werden.

Zusammenfassung und Ausblick

Die in diesem Beitrag beschriebenen Experimente geben eine erste Antwort auf die Frage:

Welche Schritte im Prozess der (Prozess-)Modellentwicklung kann GPT-3.5 unterstützen?

Die Experimente wurden in die Phasen *Erweitern einer Beschreibung*, *Informationsextraktion* und *Formale Transformation* unterteilt, die unterschiedliche linguistische Fähigkeiten von GPT-3.5 adressieren: Generieren von Text, Zusammenfassen und Übersetzen (in eine formale Sprache).

Bedenkt man, wie neu diese Technologie ist, sind die erzielten Ergebnisse beeindruckend. Im Vergleich zu früheren Arbeiten an Chatsystemen im Bereich Prozessmodellierung stellt sie eine vielversprechende Entwicklung dar. Durch die riesige Menge an Trainingsdaten eröffnen Transformer neue Horizonte. Die derzeitigen Einschränkungen verhindern ein unbeaufsichtigtes Erstellen von Prozessmodellen. Eine assistierte Durchführung ist möglich, mehr aber noch nicht.

Die Gründe dafür sind in jeder Phase der Experimente zu beobachten:

Erweitern einer Beschreibung: Auch wenn die GPT-Sprachmodelle auf einem sehr großen Datensatz trainiert werden, verfügen sie nicht über Allgemeinwissen oder Menschenverstand. Sie sind nicht in der Lage, Wissen abzuleiten, sondern sind beschränkt auf das, was sie während ihres Trainings gelernt haben. Dies behindert eine sinnvolle Erweiterung bestehender Beschreibungen. Ein Feintuning kann schwierig sein, da spezifisches Domänenwissen nicht im gleichen Umfang zur Verfügung steht wie die allgemeinen Trainingsdaten.

Eine weitere Einschränkung ist von Bedeutung: Auch wenn es den Anschein hat, ist GPT-3.5 nicht in der Lage logisch zu denken und zu rechnen. Was wie logisches Denken aussieht, ist lediglich die Ausgabe der wahrscheinlichsten Wort- und Zahlenkombinationen unter Berücksichtigung des Kontextvektors.

Informationsextraktion: Die Beschreibungen müssen so präzise und klar wie möglich sein, um genaue Ergebnisse zu erhalten. Der Aufwand, der hierzu betrieben werden muss, könnte aber bereits in das Erstellen eines Modells investiert werden.

Eine Wissensextraktion ist möglich, wenn dem System die richtigen Fragen in der richtigen Formulierung und in der richtigen Reihenfolge gestellt werden. In den hier beschriebenen Experimenten konnten jedoch nicht immer alle relevanten Informationen erfasst werden.

Formale Transformation: GPT-3.5 kann auch bei der Erstellung formaler Modelle helfen. Prozesselemente und (zumindest teilweise) Strukturen lassen sich extrahieren und serialisieren. Das Resultat muss jedoch sorgfältig aufbereitet werden, um es auch für Modellierungssoftware verwenden zu können.

In Zukunft sind Arbeiten in verschiedenen Bereichen notwendig, um LLMs zur Geschäftsprozessmodellierung in der Praxis einsetzen zu können. Es existieren weitere Transformer-Sprachmodelle unterschiedlicher Implementierung ebenso wie andere, ältere Ansätze, die noch nicht detailliert untersucht wurden. Dies erfordert eine große Anzahl gut dokumentierter und evaluierter Experimente. Dazu zählt auch das Bereitstellen von geeigneten Trainingsdaten, in denen textuelle und formale Prozessbeschreibungen miteinander verknüpft sind.

Prompt Design und Engineering, das Erstellen „guter“ Anweisungen an die Modelle, könnte sich als weites, sich stetig veränderndes Forschungsfeld erweisen. Die systematische Entwicklung von Prompts wird großen Einfluss auf die Qualität der Prozessmodelle haben.

Die wichtigste Aufgabe besteht nach Ansicht der Autoren jedoch darin, die inhärent probabilistische Wissensbasis um Faktenwissen zu erweitern.

Auch die zwischenzeitliche Veröffentlichung von GPT-4 mitsamt der Integration in ChatGPT Plus ändert nur wenig an diesen Arbeitsfeldern.

Einerseits ist der Zugang zu GPT-4 kostenpflichtig und teilweise beschränkt und somit der allgemeinen Öffentlichkeit nicht ohne weiteres zugänglich. Aus diesen Gründen wurden durch die Autoren bisher nur vorläufige Tests mit ChatGPT Plus durchgeführt. Hierbei fällt negativ ins Gewicht, dass ChatGPT Plus keine Möglichkeiten zur Konfiguration des Systemverhaltens oder gar zur Anpassung des Modells besitzt.

Andererseits zeigt ChatGPT Plus im beschriebenen Anwendungsfall kein nennenswert verbessertes Verhalten. Dies ist unerwartet, da GPT-4 seinen Vorgängern gerade linguistisch deutlich überlegen ist. Jedoch konnten keine Änderungen bei der Extraktion von Ereignissen und Kontrollstrukturen oder der Formalisierung des Outputs festgestellt werden. Zu den Verbesserungen von GPT-4 gegenüber GPT-3 und den unterschiedlichen Fähigkeiten siehe (Bubeck et al., 2023).

DANKSAGUNG

Diese Forschung wird gefördert durch „ProFIL - Programm zur Förderung von Forschungspersonal, Infrastruktur und forschendem Lernen der Hochschule Worms“.

LITERATUR

- Adamopoulou, E. und L. Moussiades (2022): An Overview of Chatbot Technology. In: *Artificial Intelligence Applications and Innovations*, Hg. I. Maglogiannis; L. Iliadis und E. Pimenidis. Springer, Cham, S. 373–383.
- Brown, T.; B. Mann; N. Ryder; M. Subbiah; J. D. Kaplan; P. Dhariwal; A. Neelakantan; P. Shyam; G. Sastry; A. Askell; S. Agarwal; A. Herbert-Voss; G. Krueger; T. Henighan; R. Child; A. Ramesh; D. Ziegler; J. Wu; C. Winter; C. Hesse; M. Chen; E. Sigler; M. Litwin; S. Gray; B. Chess; J. Clark; C. Berner; S. McCandlish; A. Radford; I. Sutskever und D. Amodei (2020): Language Models are Few-Shot Learners. In: *Advances in Neural Information Processing Systems*, Hg. H. Larochelle; M. Ranzato; R. Hadsell; M.F. Balcan und H. Lin. Curran Associates, Inc., Bd. 33, S. 1877–1901.
- Bubeck, S.; V. Chandrasekaran; R. Eldan; J. Gehrke; E. Horvitz; E. Kamar; P. Lee; Y. T. Lee; Y. Li; S. Lundberg; H. Nori; H. Palangi; M. T. Ribeiro und Y. Zhang (2023): Sparks of Artificial General Intelligence: Early experiments with GPT-4. <https://arxiv.org/abs/2303.12712/> (last accessed: 10.04.2023).
- Fliedl, G.; C. Kop und H. C. Mayr (2005): From textual scenarios to a conceptual schema. *Data and Knowledge Engineering*, 55(1):20–37.
- Haag, S. (2023): *Protocol of Experiments on GPT-3 for Process Model Generation*. Techn. Ber., HS Worms.
- Haag, S. und C. Simon (2022): Extracting Models from Spoken Text - A Research Proposal. In: *EMISA 2022*. S. in Veröffentlichung.
- Kecht, C.; A. Egger; W. Kratsch und M. Röglinger (2023): Quantifying chatbots' ability to learn business processes. *Information Systems*, 113.
- Kublik, S. und S. Saboo (2022): *GPT-3 - Building Innovative NLP Products Using Large Language Models*. O'Reilly, Sebastopol, CA.
- Lopez, G.; L. Quesada und L. A. Guerrero (2017): Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. In: *Advances in Human Factors and Systems Interaction*, Hg. I. L. Nunes. Springer, Cham, S. 241–250.
- Mendling, J.; H. Leopold und F. Pittke (2014): 25 Challenges of Semantic Process Modeling. *International Journal of Information Systems and Software Engineering for Big Companies*, 1(1):78–94.
- Nolte, F. R. (2020): *Text to Process Model: Automating Process Model Creation from Text*. Dissertation, WWU Münster.
- OpenAI (2022): ChatGPT: Optimizing Language Models for Dialogue. <https://openai.com/blog/chatgpt/> (last accessed: 10.04.2023).
- Simon, C.; S. Haag und L. Zakfeld (2022a): Stratification of Timed Petri Nets at the Example of a Production Process. In: *ECMS 2022: 36th International ECMS Conference on Modelling and Simulation*. S. 128–134.
- Simon, C.; S. Haag und L. Zakfeld (2022b): The Process-Simulation.Center. In: *SIM-SC: Special Track at SIMUL 2022: The Fourteenth International Conference on Advances in System Simulation*, Hg. F. Herrmann. Lissabon (Portugal), S. 74–77.
- Simon, C.; S. Haag und L. Zakfeld (2023): Experiments on GPT-3 Assisted Process Model Development. In: *ECMS 2023: 37th International ECMS Conference on Modelling and Simulation*. S. 270–276.
- Tingiris, S. (2021): *GPT-3 - An unofficial first look at the general-purpose language processing API from OpenAI*. Packt, Birmingham, UK.
- Vaswani, A.; N. Shazeer; N. Parmar; J. Uszkoreit; L. Jones; A. N. Gomez; L. Kaiser und I. Polosukhin (2017): Attention Is All You Need. In: *31st Conference on Neural Information Processing System*. Curran Associates.
- Weizenbaum, J. (1966): ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Communications of the ACM*, 9(1):36–45.

Ausgestaltung eines digitalen Zwillings im Rahmen von Manufacturing-X

Yohannas Alemayehu

Hochschule Pforzheim
Tiefenbronnerstr. 65
75175 Pforzheim
alemayeh@hs-pforzheim.de

Christian Jakobi

Hochschule Pforzheim
Tiefenbronnerstr. 65
75175 Pforzheim
jakobich@hs-pforzheim.de

Max-Arthur Klink

Hochschule Pforzheim
Tiefenbronnerstr. 65
75175 Pforzheim
klinkmax@hs-pforzheim.de

Julius Weber

Hochschule Pforzheim
Tiefenbronnerstr. 65
75175 Pforzheim
weberjuu@hs-pforzheim.de

Frank Morelli

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
frank.morelli@hs-pforzheim.de

ABSTRACT

Die Ausgestaltung von digitalen Zwillingen im Rahmen von Manufacturing-X, einer Initiative zur Förderung eines vernetzten Datenraums in der Ausrüsterindustrie, eröffnet sowohl für Unternehmen in der Ausrüsterindustrie als auch für die Betreiber der Plattform erhebliche Chancenpotenziale. Für einen effektiven und effizienten Einsatz gilt es, zugehörige Rahmenbedingungen zu kennen und notwendige Schritte auf beiden Seiten für eine performante Umsetzung einzuleiten. Hierfür werden Lösungsmöglichkeiten, ausgewählte Geschäftsmodelle und ein eigenes entwickeltes Vorgehensmodell zur Implementierung vorgestellt. Zur Ausgestaltung des Vorgehensmodells erfolgt eine detaillierte Behandlung der technischen Realisierung. Zur Schaffung von Anreizen werden Vorschläge aus anderen Bereichen zur effizienten Nutzung von Manufacturing-X aufgegriffen.

SCHLÜSSELWÖRTER –

Manufacturing-X; digitaler Zwilling; Geschäftsmodell; Vorgehensmodell; Implementierung

EINLEITUNG

Als Gründe für die zunehmende Bedeutung von Digitalisierung lassen sich veränderte geopolitische Rahmenbedingungen, Klimawandel und globalen Krisen anführen. Um die Transformation erfolgreich zu bewältigen, benötigen Unternehmen eine transparente Darstellung ihrer Prozesse und Ressourcen über die gesamten Wertschöpfungskette hinweg. Die notwendige Grundlage hierfür bilden digitale Ökosysteme, welche das Teilen von Daten zwischen Unternehmen und weiteren Akteuren der Industrie ermöglichen. Zwei wichtige Bausteine auf dem Weg zu einem digitalen Ökosystem wurden bereits durch GAIA-X und digitale Zwillinge im europäischen Umfeld gelegt:

- Hauptziele von GAIA-X sind die Schaffung einer sicheren Datenaustausch-Infrastruktur in Europa, die Stärkung der digitalen Souveränität und die Reduzierung von Abhängigkeiten von ausländischen Technologien und Dienstleistungen. GAIA-X verfolgt ein föderalistisches, offenes Konzept, das Interaktionen zahlreicher Dienste und Systeme fördert (Otto et al., 2022; Weiss, 2022).
- Catena-X zielt auf den Aufbau eines kollaborativen Datenökosystems ab, das auf Kooperation und ge-

meinsamer Datenverwaltung basiert. Catena-X fokussiert sich speziell auf die gesamte Lieferkette der Automobilindustrie. GAIA-X dient dabei als Framework bzw. Grundlage für Catena-X, was die Zusammenarbeit und Integration der beiden Initiativen erleichtert (BMW Group, 2023; Catena-X, o. D.).

Während die Automobilindustrie mit Catena-X bereits Pionierarbeit geleistet hat und einen GAIA-X Use Case darstellt, soll nun auch die Ausrüsterindustrie dem Beispiel der Automobilindustrie, durch den Aufbau ihres Datenraums Manufacturing-X, folgen (Plattform Industrie 4.0, 2022). Da in diesem Kontext digitale Zwillinge als elementare Bausteine fungieren, wird im Rahmen dieser Arbeit der Anwendungsfall digitale Zwillinge und dessen Treiber sowie die Einbettung im Kontext von Manufacturing-X untersucht. Dabei sind auch rechtliche, organisatorische und technischen Rahmenbedingungen zu beachten bzw. zu gestalten, um die Kollaboration in diesem föderativem Datenökosystem für die Akteure langfristig zu gewährleisten. Der vorliegende Artikel untersucht die Thematik aus folgender Forschungsperspektive: *Welche Vorgehensweise bietet sich für die Ausgestaltung eines digitalen Zwillings im Rahmen von Manufacturing-X an?*

Manufacturing-X

Manufacturing-X als gemeinsame Initiative von Akteuren aus Wirtschaft, Politik und Wissenschaft, verfolgt das

Ziel, einen Datenraum im Sinne der Industrie 4.0 für eine digital vernetzten Ausrüsterindustrie zu schaffen. Das „X“ steht für „Exchange“ und zielt auf neue Technologien und Innovationen wie zum Beispiel Data Sharing ab (Bundesministerium für Wirtschaft und Klimaschutz, 2023b, 2023c).

Ein Ziel besteht in der Erhöhung bzw. Verbesserung von Resilienz. Ferner soll die datenbasierte Abhängigkeit deutscher und europäischer Unternehmen von dominierenden Digitalplattformen reduziert werden. Neben der Erhöhung der Resilienz ist auch die Stärkung der Wettbewerbsfähigkeit ein wichtiges Ziel der Initiative. Zusätzlich strebt die Manufacturing-X-Initiative auch die Förderung von Nachhaltigkeit an, indem es einen Beitrag zur Klimaneutralität und zum Ökologischen Fußabdruck leisten will (Plattform Industrie 4.0, 2022).

Als Akteure sind sowohl große Konzerne als auch kleine und mittelständische Unternehmen, ggf. auch Startups, im Blickfeld. Als besonders relevant erweisen sich Akteure aus der Industrie 4.0 Sektor. Daneben spielen auch Cloud Provider oder Software-Entwickler eine wichtige Rolle. Aktive Player in Catena-X oder GAIA-X sowie Verbände sind ebenfalls angesprochen, um möglichst viele relevante Akteure zu involvieren (Bitkom e.V., 2023).

Digitale Zwillinge

Ein digitaler Zwilling steht für die virtuelle Repräsentation eines physischen Objekts oder Systems. Die Kopplung der virtuellen mit der realen Welt ermöglicht es, Daten zu analysieren und Probleme schon vor dem Auftreten zu erkennen. Das hieraus erwachsende bessere Verständnis komplexer Systeme soll zur Optimierung der abgebildeten Systeme und Objekte führen. Der Einsatz von digitalen Zwillingen kann in der Praxis von der Abbildung einzelner Komponenten (Komponentenzwillinge) bis hin zur Abbildung ganzer Prozesse (Prozess-Zwillinge) reichen (Grösser, 2018; IBM, 2023).

Digitale Zwillinge benötigen ein abzubildendes Objekt, beispielsweise eine Maschine oder ein Produkt aus der realen Welt, den zugehörigen Zwilling im digitalen Raum sowie Informationen, welche beides miteinander verbindet. Die Daten werden den digitalen Zwillingen mithilfe von Sensoren, IoT-Systemen und zahlreichen anderen Quellen, die mit dem Objekt in der realen Welt in Verbindung stehen, zur Verfügung gestellt. Anschließend setzt man Algorithmen und KI ein, um die gesammelten Daten zu analysieren und daraus Modelle zu entwickeln (Esri, 2021; Grösser, 2018).

Der Datenfluss bei digitalen Zwillingen findet in beide Richtungen automatisch statt (vgl. Abb. 1). Die hieraus resultierenden Ergebnisse beeinflussen die Realität direkt. Im Unterschied dazu basiert ein Digital Shadow zwar auch auf einem automatischen Datenfluss bei der Erstellung des Modells und reagiert auf Änderungen der Realität. Aus den Ergebnissen abgeleitete Änderungen müssen jedoch manuell angestoßen werden. Besteht darüber hinaus kein Online-Datenfluss von Realität zum Modell, spricht man lediglich von einem Digital Model (van der Aalst et al., 2021).

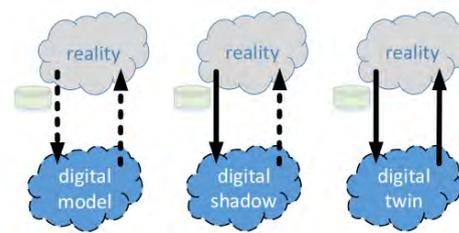


Abbildung 1: Digital Twin, Digital Shadow und Digital Model
Quelle: (van der Aalst et al., 2021)

Die Anwendungsbereiche für digitale Zwillinge reichen von der Optimierung von Bau- und Infrastrukturprojekten über die personalisierte Medizin und den Automobilbau bis hin zur Luft- und Raumfahrtindustrie. Auch in der Logistik können digitale Zwillinge zur Prozessoptimierung, Effizienzsteigerung und Qualitätsverbesserung eingesetzt werden. Grundsätzlich gibt es drei Arten von digitalen Zwillingen, die unterstützend in den unterschiedlichsten Branchen eingesetzt werden können: digitale Zwillinge für Produkte, digitale Zwillinge für Produktionsprozesse und digitale Zwillinge für Leistungsfähigkeit (IBM, 2023; Siemens, 2019).

Zu den Herausforderungen für die Implementierung von digitalen Zwillingen gehören Datensicherheit, Datenschutz, technische Fehler, Implementierungskosten sowie Schulungsaufwände für die Mitarbeiter im Umgang mit der neuen Technologie. Auch die Komplexität der Systeme spielt in diesem Zusammenhang eine zentrale Rolle: Unternehmen müssen Vorkehrungen treffen, um die Sicherheit und Integrität ihrer Daten zu gewährleisten und die bestmögliche digitale Abbildung zu schaffen (DHL Trend Research, 2019; PwC, 2018).

RAHMENBEDINGUNGEN

Unternehmen sind bestrebt, ihre Kernkompetenzen und geistiges Eigentum zu schützen. Beispielsweise wollen Produkthersteller i.d.R. ihr Produktdesign sowie sonstige technische Informationen nicht teilen (Uhlenkamp et al., 2020). Eine Studie von PwC zeigt, wie groß die Sorge in Bezug auf die Datensicherheit ist: Dies erweist sich als zentraler Aspekt für viele deutsche Unternehmen, weshalb sie keinen Datenaustausch mit anderen Akteuren eingehen wollen (Pauer et al., 2018). Darüber hinaus befürchten entsprechend der PwC-Studie mehr als die Hälfte der befragten Führungskräfte, durch digitale Kooperation die Kontrolle über die eigenen Daten zu verlieren (Pauer et al., 2018). Entsprechend erweist sich Datensouveränität als weitere Herausforderung und wichtige Rahmenbedingung für Manufacturing-X (VDMA, 2022): Unter Datensouveränität versteht man die „Hoheit eines Dateneigners über die Nutzung der eigenen Daten souverän und exklusiv zu entscheiden“ (Bergs et al., 2020). Dieser Sachverhalt spricht gegen eine Realisierung von Manufacturing-X im Sinne einer monolithischen zentralen Plattform. Ein föderativer Ansatz soll stattdessen sicherstellen, dass Unternehmen die Kontrolle über ihre eigenen Daten behalten und gleichzeitig von den Vorteilen eines gemeinsamen Datenökosystems zu profitieren (VDMA, 2022).

Um einen sicheren und vertrauensvollen Datenaustausch zu gewährleisten, müssen die europäischen rechtlichen Verordnungen und Gesetze eingehalten werden. Hierzu zählen z.B. der EU Data Act, das Vertragsrecht, das Lieferkettengesetz, der Data Governance Act, der Digital Markets Act und das Kartellrecht. Allerdings sind viele rechtlichen Aspekte zum Datenaustausch noch ungeklärt. Daher erweist es sich als relevant, den rechtlichen Entwicklungen hinreichend Aufmerksamkeit zu schenken und sich adäquat an den rechtlichen Rahmen anzupassen (Plattform Industrie 4.0, 2022).

Die Ausrüsterindustrie in der EU umfasst ca. 80.000 Unternehmen, welche eine Vielfalt an heterogenen Produkten produzieren und eine komplexe offene Wertschöpfungsnetzwerk aufweisen (VDMA, 2022). Um Manufacturing-X erfolgreich umzusetzen ist es daher eine besondere Herausforderung, für diese starke Heterogenität einen umfassenden organisatorischen Rahmen zu schaffen und diesen zu orchestrieren (Plattform Industrie 4.0, 2022; VDMA, 2022).

Die Organisationsstruktur von Manufacturing-X muss geprägt sein von Transparenz und Offenheit, sowohl für nationale als auch internationale gegenwärtige und zukünftige Beziehungen. Gerade auch deshalb wird angestrebt, eine Zusammenarbeit mit Catena-X zu ermöglichen. Des Weiteren hat die offene und transparente Organisationsstruktur von Manufacturing-X zum Ziel, die aktive Beteiligung und Mitgestaltung der beteiligten und interessierten Unternehmen, politischen Institutionen, Verbände und wissenschaftlichen Akteure bei der Entwicklung der Initiative zu fördern (Plattform Industrie 4.0, 2022).

In Bezug auf die Zusammenarbeit sind Monopolbildungen zu verhindern, Wettbewerb zu fördern und Chancengleichheit zu gewährleisten (VDMA, 2022). Ein Austausch von hochsensiblen Informationen basiert auf Vertrauen und sollte bei souveränen Partnern ohne äußere Zwänge erfolgen (VDMA, 2022).

Analog zu Catena-X müssen die Stakeholder von Manufacturing-X eine organisatorische Basis schaffen. Entsprechende Strukturen wie bei Catena-X, beispielsweise in Form von Onboarding-Partnern für Coachings, Schulungen etc., existieren bei Manufacturing X jedoch nicht. Aufgrund der Heterogenität der beteiligten Unternehmen und der hohen Anzahl erweist es sich als unsicher, ob eine solche Konstellation sich etablieren wird (Catena-X Automotive Network Konsortialprojekts, 2023).

Neben rechtlichen und organisatorischen Rahmenbedingungen gibt es bei der Implementierung von digitalen Zwillingen auch technische Voraussetzungen, die erfüllt werden müssen. Diese lassen sich hierarchisch in Maschinenebene, Datenebene und Analyseebene einteilen. Zusätzlich gibt es noch technische Rahmenbedingungen allgemeiner Art, die keiner einzelnen Ebene zugeordnet werden können (Neugebauer et al., 2023).

Auf Maschinenebene ist vor allem die Verwendung von IoT-Geräten bzw. das Ausstatten vorhandener Geräte und Produkte mit Sensorik und Sendeelementen von großer Bedeutung. Die Grundlage digitaler Zwillinge bilden

Daten, die direkt an den physischen Objekten erfasst werden müssen. Mithilfe der Sendeelemente können die Daten kabelgebunden oder kabellos an die Dateninfrastruktur gesendet werden (Fischer, 2020; Human, 2020; Neugebauer et al., 2023).

Die Verwaltung der erfassten Daten findet auf der Datenebene statt. Die benötigte Dateninfrastruktur kann sich hierbei aus einem Cloud- und einem On-Premise-Anteil zusammensetzen (Neugebauer et al., 2023). Zusätzlich sollten bereits bestehende System, wie ERP, PDM, PLM und CAD in den Datenfluss eingebunden werden, um die Daten der physischen Objekte mit Umgebungs- und Prozessdaten zu verbinden (exone.de, 2023). Ein Datenmanagement-Tool sowie ein Registry-Dienst helfen dabei, die Daten für die Analyse vorzubereiten und den Überblick über die Daten zu behalten (Schnicke und Kuhn, 2020). Ein Registry-Dienst ist notwendig, um die dynamischen Systeme zu verwalten. Der Dienst dient als zentrales Verzeichnis, in dem Informationen über Verwaltungsschalen und Teilmodelle, wie beispielsweise Metadaten, Zugangspunkte und Schnittstellenbeschreibungen gespeichert sind. Es ermöglicht den Zugriff und die Verwaltung sowie das Durchsuchen digitaler Zwillinge anhand bestimmter Kriterien, beispielsweise Tags, Eigenschaften oder Zeiträume (Catena-X, 2022; Schnicke und Kuhn, 2020).

Der Nutzen aus der großen Datenmenge soll sich auf der Analyseebene vollziehen. Hier werden die Daten auf Benutzeroberflächen, wie beispielsweise Tablets, Smartphones und AR/VR-Brillen für den Menschen visualisiert und mithilfe von Software aufbereitet. Optimierungs-, Simulations-, Prognose-, Reporting- und Überwachungsprogramme verwenden die Daten, um spezifische Analysen zum Verhalten der Maschinen und Produkte zu erstellen und daraus im weiteren Verlauf einen Nutzen zu ziehen (Fischer, 2020; Neugebauer et al., 2023).

Betrachtet man die allgemeinen technischen Rahmenbedingungen, ist es wichtig, Schnittstellen zwischen den Ebenen und auch zwischen den Unternehmen zu schaffen, über die die Daten in einem einheitlichen Datenformat übertragen werden können. Die horizontale und vertikale Kommunikation zwischen Maschinen kann auf Basis der „Produktionssprache“ OPC UA stattfinden. Die standardisierte Darstellung digitaler Zwillinge kann mithilfe sogenannter Verwaltungsschalen erfolgen (vgl. Abb. 2). Diese fungieren als „genormter Datenstecker“ zwischen Wertschöpfungspartnern und bieten eine harmonisierte Schnittstelle zum gesamten Produktionssystem, wodurch der Datenaustausch erleichtert werden soll (Plattform Industrie 4.0, 2022).

Verwaltungsschalen umschließen ein Objekt und ermöglichen das Referenzieren von Teilmodellen dieses Objekts, welche unterschiedliche Sichten des digitalen Zwillinges darstellen. Hierdurch wird die Implementierung verschiedener Versionen eines digitalen Zwillinges ermöglicht. Die Verwaltungsschale verknüpft Eigenschaften und Funktionen mit semantischen Informationen, um die Bedeutung von Daten zu definieren. Dies führt zu einer einheitlichen Informationsgrundlage, auch

wenn verschiedene Hersteller unterschiedliche Feldnamen für Komponenten verwenden. Die Nutzer selbst sind für den Inhalt der Teilmodelle verantwortlich und können verschiedene, ihren Bedürfnissen entsprechende, Arten von Teilmodellen definieren (Schnicke und Kuhn, 2020).



Abbildung 2: Verwaltungsschale
Quelle: (Schnicke und Kuhn, 2020)

AUSGESTALTUNG EINES DIGITALEN ZWILLINGS FÜR MANUFACTURING-X

Einsatz von digitalen Zwillingen

Manufacturing-X bietet die Möglichkeit, digitale Zwillinge zu verwalten und anderen teilnehmenden Unternehmen zu Verfügung zu stellen. Im Folgenden werden Möglichkeiten für die Anwendung digitaler Zwillinge im Kontext von Manufacturing-X aufgezeigt.

Nach erfolgter Implementierung kann damit begonnen werden, Daten zu sammeln und daraus einen digitalen Zwilling zu erstellen. Beim Speichern der Daten sollte man einen hybriden Ansatz wählen. Entsprechend sind statische Daten in der Cloud und dynamische Daten initial On-Premise speichern, um dauerhaft große Datenströme zu vermeiden und das System so nicht zu überlasten. Allerdings ergibt es Sinn, dynamische Daten in regelmäßigen Abständen in der Cloud zu speichern, um Datenverluste bei Ausfällen zu vermeiden.

Die intern gespeicherten Daten der digitalen Zwillinge sollten in festgelegten Abständen in dem für andere Unternehmen einseharen Datenraum registriert werden. Hierbei kann jedes Unternehmen einen eigenen souveränen Datenraum betreiben und in diesem seine Daten den teilnehmenden Unternehmen zu Verfügung stellen. Damit dieses Datenökosystem funktioniert, ist es notwendig, eine einfache und durchgängige Datenvernetzung sowie die Bereitschaft zum Teilen von Daten sicherzustellen (Plattform Industrie 4.0, 2023). Darüber hinaus bietet das Einbinden von Suchmaschinenanbietern für Manufacturing-X Services die Möglichkeit, das Datenökosystem auf Basis von Stichwörtern nach digitalen Zwillingen zu durchsuchen (VDMA, 2022).

Grundsätzlich steigt mit der Menge der verwendeten Daten auch die Qualität der digitalen Zwillinge und somit die Genauigkeit der Analysen, die mithilfe dieser erstellt werden können. Eine Notwendigkeit für den Erfolg von Manufacturing-X und dem damit verbundenen Datenaustausch zwischen Unternehmen ist die Verwendung einheitlicher Standards beim Speichern und Verarbeiten der

Daten. Die für diesen Anwendungsfall beste Möglichkeit dies sicherzustellen, ist die Verwendung von Verwaltungsschalen. Diese speichern alle Teilmodelle eines Assets ab. Beispiele für mögliche Teilmodelle sind in Abbildung 3 zu sehen.



Abbildung 3: Mögliche Teilmodelle einer Verwaltungsschale
Quelle: (Zentralverband Elektrotechnik- und Elektronikindustrie, 2016)

Die Teilmodelle sind dann als Tabellen aufgebaut und speichern alle wichtigen Daten, die im Anschluss als einheitliches Datenformat mit den anderen Unternehmen geteilt werden können. Das Teilmodell „Energieeffizienz“ ist beispielhaft in Abbildung 4 beschrieben.

		Merkmalsdefinition		Merkmalsausprägung								
Hierarchy	ID	(preferred) Name	Definition	Unit of measure	Data type	Value list	Value	Expression semantic	Expression logic	Views	API/ FTA/	Contents
←	AAB010	Electrical energy	This is a group of properties concerning about electrical energy consumption.	-	-	-	-	-	-	Performance	-	-
	AAB011	Electrical consumption actual	Current, actual electrical consumption.	W	REAL	D..*	93.6 [W]	Measurement	Equal	Performance	-	-
-	AAB012	Electrical consumption cumulative energy	Integrated electrical consumption over time.	Wh	REAL	D..*	138.86 [Wh]	Measurement	Equal	Performance	-	-
-	AAB013	Electrical consumption cumulative start date	Date and time the integration of electrical consumption was started.	-	UTC Date & Time	n/a	2002-05-30T09:30:10Z	Measurement	Equal	Performance	A	For XML UTC time format see: http://www.w3schools.com/xml/schema_dtypes_date.asp
←	AAB020	Pneumatic energy	This is a group of properties concerning about pneumatic energy consumption.	-	-	-	-	-	-	Performance	-	-
-	AAB021	Actual supply pressure	Supply pressure of the asset sensed at the inlet of the asset.	bar	REAL	D..*	8 [bar]	Measurement	Equal	Performance	-	-
-	AAB022	Pneumatic consumption actual	Current, actual pneumatic consumption.	l/h	REAL	D..*	212 [l/h]	Measurement	Equal	Performance	-	-
-	AAB023	Pneumatic consumption cumulative energy	Integrated pneumatic consumption over time.	l	REAL	D..*	3424 [l]	Measurement	Equal	Performance	-	-
-	AAB024	Pneumatic consumption cumulative start date	Date and time the integration of pneumatic consumption was started.	-	UTC Date & Time	n/a	2002-05-30T09:30:10Z	Measurement	Equal	Performance	-	-

Abbildung 4: Teilmodell "Energieeffizienz"
Quelle: (Zentralverband Elektrotechnik- und Elektronikindustrie, 2016)

Um sicherzustellen, dass den Unternehmen auch beim Teilen von Teilmodellen ihrer Assets die volle Kontrolle über ihre Daten erhalten bleibt, sollten Unternehmen wählen können, welche Teilmodelle eines Assets preisgegeben werden. Innerhalb eines Teilmodells können zudem Felder, die nicht als „mandatory“ gekennzeichnet sind, außen vor gelassen werden, um bestimmte Daten zu schützen. Als Konsequenz dürfen diese Unternehmen allerdings auch keinen Zugriff auf die Teilmodelle und Felder innerhalb der Teilmodelle haben, die sie selbst nicht im Datenökosystem geteilt haben. Dies führt dazu, dass Unternehmen sehr genau abwägen müssen, welche Daten

sie teilen und welche Daten sie für sich behalten wollen. Gleichzeitig motiviert es Teilnehmer viele Daten preiszugeben, um selbst auch den größtmöglichen Nutzen aus der Zusammenarbeit in Manufacturing-X zu ziehen.

Ausgewählte Geschäftsmodelle

Nutzen durch digitale Zwillinge entstehen insbesondere dann, wenn die erzeugten Daten analysiert und für Simulationen und Vorhersagen verwendet werden. Der konkrete Mehrwert variiert stark zwischen den Unternehmen und hängt von verschiedenen Faktoren ab. Nachfolgend werden potenzielle Geschäftsmodelle vorgestellt, die sich i.d.R. mithilfe von digitalen Zwillingen erschließen lassen:

grund der gesammelten Echtzeitdaten und deren Abgleich Vorhersagen über drohende Störungen treffen. Dies ermöglicht proaktive Maßnahmen wie z.B. den rechtzeitigen Austausch von Verschleißteilen. Wartungsintervalle lassen sich mithilfe der Daten des Condition Monitorings, welche mit dem digitalen Zwilling verknüpft sind, am tatsächlichen Bedarf orientieren. Collaborative Condition Maintenance ermöglichen einen koordinierten Einsatz von Instandhaltungsmaßnahmen über mehrere Partner hinweg. Der Fokus liegt hierbei auf dem Austausch von Wissen, Ressourcen und Fachkenntnissen, um die Instandhaltungsprozesse zu optimieren (Bundesministerium für Bildung und Forschung, 2020; Bundes-

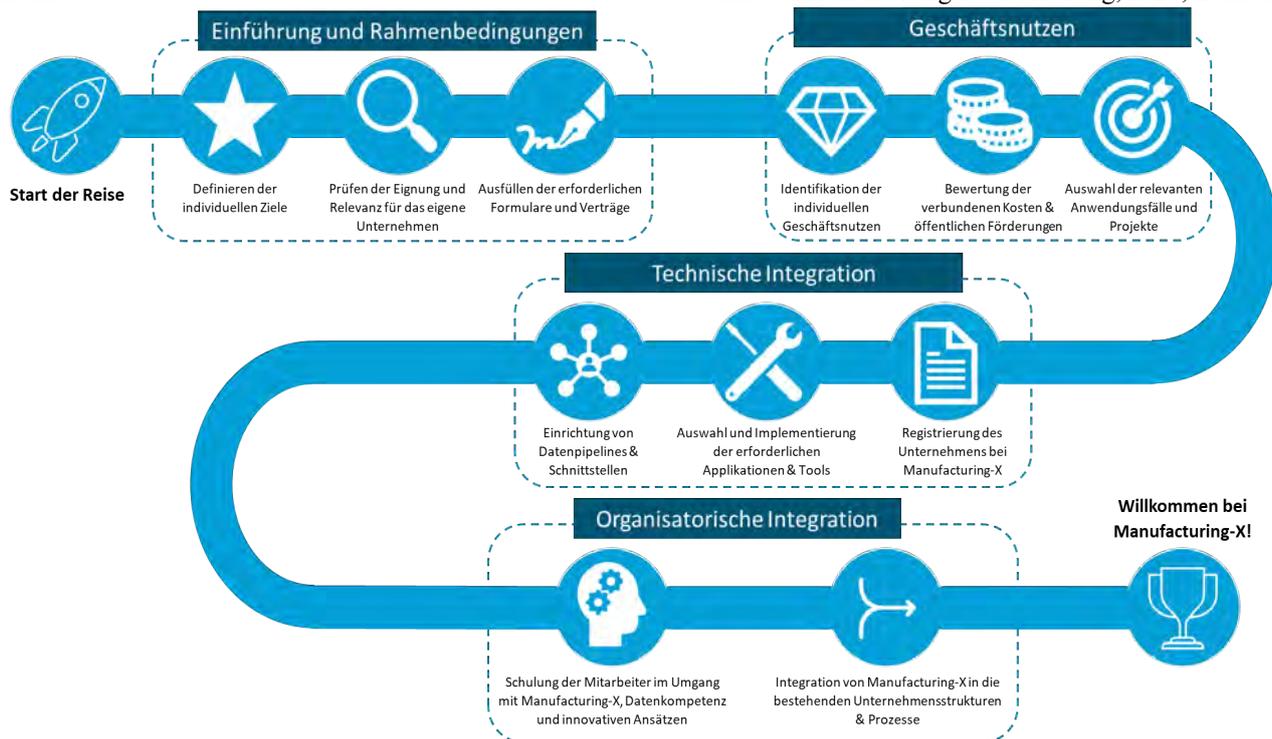


Abbildung 5: Manufacturing-X Phasenplan
Quelle: Eigene Darstellung

- Virtueller Marktplatz für den Handel mit digitalen Zwillingen: Ein wesentlicher Aspekt besteht darin, dass die Eigentümer der digitalen Zwillinge die volle Kontrolle über ihre Daten behalten und eigens darüber entscheiden können, wie ihre Daten genutzt werden (VDMA, 2022).
- CO₂- und ESG-Monitoring mithilfe des digitalen Produktpasses: Der wachsende Umfang an Berichtspflichten für Hersteller von Produkten, wie sie beispielsweise im Lieferkettengesetz oder dem Product Carbon Footprint (PCF) festgelegt sind, erstreckt sich zunehmend über die gesamte Wertschöpfungskette. In diesem Zusammenhang kann die Einführung eines digitalen Produktpasses hilfreich sein, um relevante digitale Produktdaten zu generieren und zu managen.
- Predictive Maintenance und Collaborative Condition Monitoring: Digitale Zwillinge können auf-

- grund der gesammelten Echtzeitdaten und deren Abgleich Vorhersagen über drohende Störungen treffen. Dies ermöglicht proaktive Maßnahmen wie z.B. den rechtzeitigen Austausch von Verschleißteilen. Wartungsintervalle lassen sich mithilfe der Daten des Condition Monitorings, welche mit dem digitalen Zwilling verknüpft sind, am tatsächlichen Bedarf orientieren. Collaborative Condition Maintenance ermöglichen einen koordinierten Einsatz von Instandhaltungsmaßnahmen über mehrere Partner hinweg. Der Fokus liegt hierbei auf dem Austausch von Wissen, Ressourcen und Fachkenntnissen, um die Instandhaltungsprozesse zu optimieren (Bundesministerium für Bildung und Forschung, 2020; Bundes-
- ministerium für Wirtschaft und Klimaschutz, o. D.).
- Manufacturing-as-a-Service: Hierdurch können Unternehmen ungenutzte Kapazität ihre Fertigungssysteme als Service an Dritte anbieten (VDMA, 2022). Ein digitaler Zwilling ermöglicht es hierbei potenziellen Kunden, die verfügbaren Ressourcen und Fähigkeiten des Fertigungssystems virtuell zu erkunden und ihre Produktionsanforderungen entsprechend anzupassen.
- Equipment-as-a-Service: In diesem Geschäftsmodell erhalten Betreiber von Produktionsmitteln eine Lizenz zur Nutzung, anstatt diese zu erwerben (Pay-per-Use-Geschäftsmodelle). Dies vermeidet hohe Anfangsinvestitionen und schafft höhere Flexibilität. Gleichzeitig führt dieses Modell zu einer kontinuierlichen Einkommensquelle für den Hersteller oder Anbieter des Services. Um eine unterbrechungsfreie Nutzung zu gewährleisten und Wartungen vorausschauend zu planen und durchzuführen,

ist es erforderlich, dass die Eigentümer der Produktionsmittel Daten über den aktuellen Zustand erfassen und auswerten können (VDMA, 2022). Dies lässt sich durch einen digitalen Zwilling bewerkstelligen.

Vorgehensmodell zur Implementierung

Der nachfolgend vorgestellte Phasenplan (vgl. Abb. 5) hat zum Ziel Interessenten einen Leitfaden für die Implementierung von Manufacturing-X an die Hand zu geben. Das Konzept orientiert sich an Catena-X und berücksichtigt spezifische Rahmenbedingungen in der Manufacturing-X Domäne wie z.B. die Anzahl und Heterogenität der Teilnehmer (Bundesministerium für Wirtschaft und Klimaschutz, 2023a).

1. Einführung der Rahmenbedingungen:

In der ersten Phase definieren die Nutzer ihre strategischen Ziele und begutachten die Rahmenbedingungen für die Teilnahme an Manufacturing-X. Zusätzlich prüfen sie die Eignung und Relevanz von Manufacturing-X für das eigene Unternehmen. Darüber unterzeichnen sie die zugehörigen Verträge.

Den Betreibern der Manufacturing-X Plattform obliegt die Aufgabe, umfassende Eingangsinformationen bereitzustellen. Es steht ein zentrales Portal mit Unterstützungsmaterialien und Ressourcen zur Verfügung, um den Einstieg in Manufacturing-X zu erleichtern. Darüber hinaus liegt ein allgemeiner Standardvertrag vor, der den aktuellen rechtlichen Rahmenbedingungen entspricht und der die Bedürfnisse und Anforderungen der Teilnehmer berücksichtigt.

2. Geschäftsnutzen:

In der zweiten Phase identifizieren die Nutzer den individuellen Geschäftsnutzen und die Mehrwerte für ihr Unternehmen durch die Teilnahme an Manufacturing-X. Weiterhin analysieren sie die damit verbundenen Kosten und mögliche öffentliche Förderungen. Sie wählen relevante Anwendungsfälle und Projekte aus, die im Rahmen von Manufacturing-X umgesetzt werden sollen.

Die Manufacturing-X Betreiber stellen umfassende Informationen über mögliche Anwendungsfälle und Potenziale von Manufacturing-X bereit. Es werden auch Informationen und Kriterien zur Bewertung von Kosten und aktuellen öffentlichen Förderungsmöglichkeiten bereitgestellt. Zusätzlich sollten Richtlinien und Empfehlungen zur Identifizierung und Umsetzung der Anwendungsfälle bereitgestellt werden. Dabei werden etablierte erfolgreiche Einsätze als Best Practices ausgewiesen.

3. Technische Integration:

Zusätzlich zur Registrierung bei Manufacturing-X wählen die Nutzer die erforderlichen Applikationen und Tools aus. Diese werden implementiert und dienen als Grundlage für die Umsetzung der Anwendungsfälle. Darüber hinaus erfolgt die Einrichtung von Datenpipelines und Schnittstellen zur sicheren und effizienten Datenübertragung.

Die Manufacturing-X Betreiber halten einfache und benutzerfreundliche Tools und Formulare zur Registrierung und Integration neuer Teilnehmer bereit. Der Onboarding-Prozess ist geprägt von Automatisierung und Standardisierung. Es steht eine sichere und skalierbare Dateninfrastruktur für den Datenaustausch und die Zusammenarbeit zur Verfügung. Zusätzlich werden Schulungen in Form von Videos und Anleitungen zur reibungslosen Integration angeboten.

4. Organisatorische Integration:

Die Nutzer etablieren eine Kultur des Datenaustauschs und der Zusammenarbeit innerhalb ihres Unternehmens. Zusätzlich stellen sie die Integration von Manufacturing-X in die bestehenden Unternehmensstrukturen und Prozesse sicher. Außerdem sorgen sie für die erforderlichen Skills bei ihren Mitarbeitern für den Umgang mit Manufacturing-X, Datenkompetenz und innovativen Ansätzen. In dieser vierten Phase stellen die Manufacturing-X Betreiber Leitlinien und Best Practices für die organisatorische Integration von Manufacturing-X in die Unternehmensstrukturen der Teilnehmer bereit. Darüber bieten sie Schulungen in Form von Videos oder Onlineworkshops an, um eine erfolgreiche Integration zu unterstützen.

Schwerpunkt technische Integration

Im Folgenden wird noch detaillierter auf die Phase der technischen Integration eingegangen. In diesem Zusammenhang werden auch Technologien erörtert, die von Nutzern zur Implementierung von Manufacturing X in Betracht gezogen werden können. Aufgrund des großen Anbieterspektrums, insbesondere im Cloud-Bereich, handelt es sich um eine exemplarische Auswahl, von der im Falle von spezifischen Anforderungen eines Unternehmens abgewichen werden kann.

Austausch von Echtzeitdaten mittels OPC UA

OPC UA („Open Platform Communications Unified Architecture“) eignet sich als Standard für den Transport von Echtzeitdaten, da es sich innerhalb der Fertigungsindustrie bereits als Hersteller- und plattformunabhängiges Kommunikationsprotokoll etabliert hat. Die Herstellerunabhängigkeit von OPC UA ermöglicht eine nahtlose Integration bzw. Verknüpfung von Maschinen unterschiedlicher Hersteller.

In der folgenden Architekturbetrachtung wird davon ausgegangen, dass es sich um moderne Maschinen handelt, die bereits über eine OPC UA-Schnittstelle verfügen. Folglich entfällt die Notwendigkeit, zusätzliche Adapter oder Gateways für die Implementierung einer OPC UA-basierten Kommunikation zu installieren. Durch die Herstellerunabhängigkeit von OPC UA wird eine Abhängigkeit zu großen Softwareunternehmen vermieden, was ansonsten häufig der Fall ist, insbesondere bei Kommunikationssystemen, die auf COM-Technologien basieren. Diese Unabhängigkeit erhöht die Flexibilität und Autonomie innerhalb der Systemarchitektur. OPC UA bietet ein breites Spektrum an Einsatzmöglichkeiten, was die

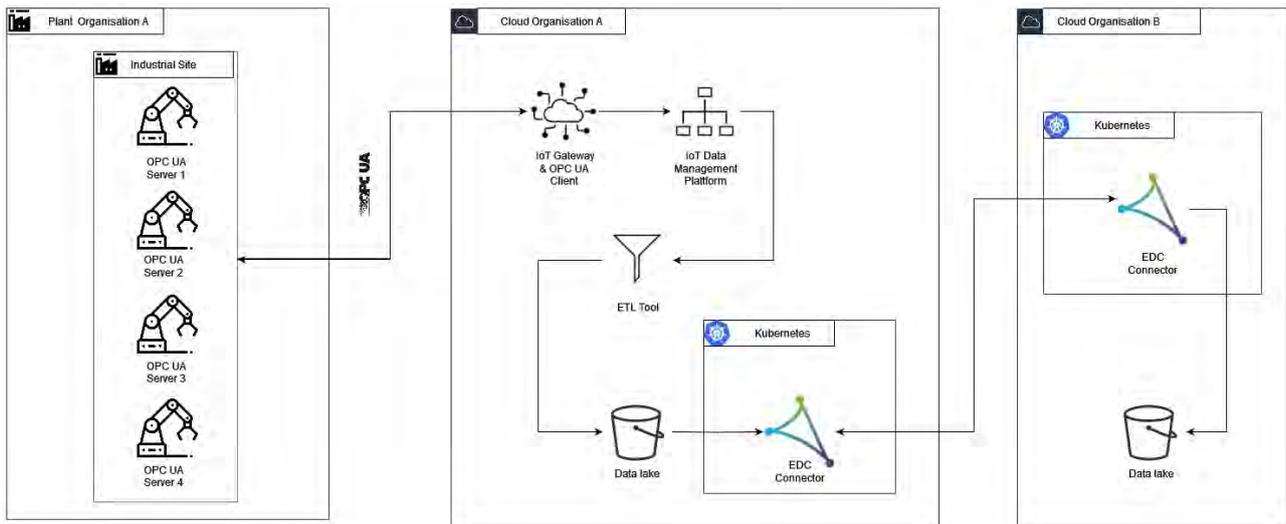


Abbildung 6: Beispiel Architektur
Quelle: Eigene Darstellung

Interoperabilität fördert. Zusätzlich zeichnet sich OPC UA durch eine hohe Skalierbarkeit aus (IPC2U, o. D.; OPC Router, o. D.).

Darüber hinaus gewährleistet OPC UA in industriellen Umgebungen wie der Fertigungsindustrie eine sichere Datenübertragung durch die Implementierung eines umfassenden Sicherheitskonzepts. Dieses Konzept umfasst Authentifizierung, Autorisierung, Verschlüsselung und die Sicherstellung der Datenintegrität durch den Einsatz von Zertifikaten (ascalab, o. D.; OPC Unified Architecture, 2018). Das Bundesamt für Sicherheit in der Informationstechnik (BSI) hat in den Jahren 2015 und 2021 Sicherheitsanalysen für OPC UA veröffentlicht, die zwei Sicherheitsmodi aufzeigen: Sign und Sign&Encrypt. Der Hauptunterschied besteht darin, dass Sign&Encrypt eine erweiterte Sicherheit bietet, da Nachrichten nicht nur signiert, sondern auch verschlüsselt werden. Die Schlussfolgerung aus diesen Sicherheitsanalysen lautet, dass OPC UA in Verbindung mit Sign&Encrypt eine effektive Abwehr gegen die meisten Bedrohungen darstellt, mit Ausnahme von Angriffen wie 'Denial of Service' und 'Server Profiling', gegen die nur ein begrenzter Schutz geboten werden kann (vom Dorp et al., 2022).

OPC UA unterstützt zwei Kommunikations-mechanismen: Client/Server und Publisher/Subscriber. Die Client/Server Kommunikation stellt die „traditionelle“ Kommunikation zwischen zwei Parteien dar. Dabei senden einer oder mehrere OPC-Clients eine Anfrage an den OPC UA-Server und erhalten von diesem eine Antwort. Die Publisher/Subscriber-Kommunikation hingegen basiert auf einer indirekten Interaktion über einen Message Queue Broker. Ein Publisher stellt dabei Informationen in einem Nachrichtenkanal zur Verfügung, den die Subscriber „abonnieren“ können, um die Daten zu erhalten, sobald der Publisher sie veröffentlicht.

Das nachfolgende Architekturmodell basiert auf der Entscheidung, ein OPC UA Client/Server-Modell zu implementieren. Es beruht auf der Annahme, dass lediglich ein Dienst die Daten von den Maschinen abfragt. Die Unternehmensarchitektur kann aus einem oder mehreren OPC

UA-Servern bestehen. Das Gegenstück hierzu ist der OPC UA Client. OPC UA-Server sind die Einheiten, die Daten generieren und bereitstellen. In der vorliegenden Modellierung sind dies die Maschinen, die Sensordaten erzeugen und bereitstellen. Der OPC UA-Client hingegen ist die Software oder das Dienstprogramm, das die Sensordaten vom OPC UA-Server abrufen und weiterverarbeitet, z.B. ein Gateway. Bezüglich der ausgetauschten Daten bietet OPC UA die Möglichkeit, Daten semantisch zu beschreiben. Das bedeutet, dass neben „rohen“ Datenwerten zusätzliche Informationen wie Herkunft, Format und ähnliches bereitgestellt werden können. Dies kann zu einer verbesserten Datenanalyse und -nutzung führen, da die Daten mehr Informationen beinhalten und verständlicher sind, was das Risiko von Fehlinterpretationen reduziert (IPC2U, o. D.; OPC Router, o. D.; Phoenix Contact, o. D.; OPC Unified Architecture, 2018; Siemens, o. D.). Für die potenzielle Implementierung von Manufacturing X ist eine Edge-to-Cloud-Architektur zu empfehlen. Wie diese im Detail aussieht, hängt zukünftig von IPCEI-CIS ab, einem Förderprogramm, das für die Entwicklung fortschrittlicher Technologien für ein Multi-Provider Cloud-Edge-Continuum konzipiert ist (Plattform Industrie 4.0, 2022). Zusätzlich sind die spezifischen Bedürfnisse und Anforderungen eines Unternehmens für die Architektur ausschlaggebend. Unabhängig von diesen Faktoren, bietet sich die folgende generische Edge-to-Cloud-Architektur an (vgl. Abb. 6).

Generische Manufacturing X Architektur

In einem Unternehmen fungieren Maschinen, die Sensordaten erzeugen, als OPC UA-Server. In dem Beispielmodell existieren vier Maschinen und somit vier individuelle OPC UA-Server. Diese Server kommunizieren mit einem in der Cloud betriebenen IoT-Gateway, das gleichzeitig als OPC UA-Client dient. Abhängig von den spezifischen Anforderungen kann entweder ein Client/Server- oder ein Publisher/Subscriber-Kommunikationsmodell ausgewählt werden. In dem Modell ist ein Client/Server-Modell modelliert, da das IoT-Gateway die

einzigste Komponente ist, die Daten abrufen. Das IoT-Gateway fungiert als Schnittstelle, die die Daten von den Edge-Geräten abrufen und an ein Datenmanagement-Tool weiterleitet. Dieses in der Cloud betriebene Tool sammelt, verarbeitet und speichert die Daten und kann für eine erste Datenbereinigung und Analyse genutzt werden. Das ETL-Tool (Extraction, Transformation, Load) ist für das Extrahieren, Transformieren und Weiterleiten der Daten an den Data Lake verantwortlich. Über dieses können die Daten bei Bedarf in das erforderliche Format umgewandelt werden. Der Data Lake dient als zentrale Speicherstelle für alle Daten und kann flexibel für Analysen oder maschinelles Lernen genutzt werden. Er ist mit einem Eclipse Dataspace Connector des Eclipse Dataspace Components Frameworks verbunden, welcher in einem Kubernetes-Cluster betrieben wird, um einen sicheren und kontrollierten Datenaustausch mit anderen Unternehmen zu ermöglichen (AWS Workshop Studio, o. D.; Eclipse Foundation, o. D.; European Commission, 2023; Kolodziej & Vazquez, 2023; Sodabathina et al., 2022).

Einheitliches Datenformat

OPC UA unterstützt speziell drei Encoding-Formate für das Codieren und Decodieren von Daten: OPC UA Binary, XML und JSON. Diese Formate werden von OPC UA zur Datenübertragung und -speicherung bereitgestellt. Es ist somit möglich, Daten entweder binär oder als Text in Form von XML oder JSON zu übertragen. Alle drei Formate bieten sich als ein einheitliches Datenformat an. Insbesondere XML und JSON sind weit verbreitete Formate und werden von vielen Anwendungen akzeptiert. Das binäre Format von OPC UA ist ein spezielles, für die maximale Nutzung von OPC UA konzipiertes Format. Es ist vor allem auf schnelles Codieren und Decodieren ausgelegt. Unabhängig von der gewählten Option sollten keine anderen Datenformate als die genannten drei in Betracht gezogen werden, um mögliche Komplikationen oder Fehler bei Transformationen zu vermeiden (OPC Foundation, o. D.a, o. D.b, o. D.c, o. D.d).

Tools

Bei der Wahl der Cloud-Umgebung, des Datenmanagement-Tools, des ETL-Tools und des Data Lakes besitzt jedes Unternehmen individuelle Präferenzen und Anforderungen, die es zu berücksichtigen gilt. Zudem muss die Entscheidung auf der Grundlage der aktuellen IT-Architektur getroffen werden, um die bestmögliche Kompatibilität und Integration zu gewährleisten. Grundsätzlich sollten alle gewählten Tools Eigenschaften wie Skalierbarkeit, Flexibilität, Sicherheit, Verfügbarkeit und eine breite Unterstützung für Cloud-Plattformen aufweisen.

Für Unternehmen, die von Grund auf neu starten, können Cloud-Umgebungen wie Amazon Web Services (AWS) oder Microsoft Azure in Betracht gezogen werden. AWS bietet Dienste wie beispielsweise AWS Glue für Datenmanagement und ETL-Aufgaben sowie AWS S3 als Data Lake. Microsoft Azure bietet Dienste wie Azure Data Factory für Datenmanagement und ETL und Azure Data

Lake Storage als Speicherlösung (Amazon Web Services, Inc., o. D.a, o. D.b; Microsoft, o. D.a, o. D.b).

Eclipse Data Space Connector

Der in der Architektur verwendete Eclipse Dataspace Connector ist ein Bestandteil des Eclipse Dataspace Component Frameworks, das von der Eclipse Foundation entwickelt wird. Es handelt sich um ein Open-Source Framework, das einen sicheren und kontrollierten Datenaustausch zwischen unterschiedlichen Parteien ermöglicht. Dabei wird das Konzept des International Data Spaces (IDS) implementiert, welches regelt, wie ein sicherer Datenaustausch zwischen mehreren Parteien innerhalb eines Datenraums realisiert werden kann. Ein wesentlicher Vorzug des Eclipse Dataspace Connectors ist seine Fähigkeit, automatische Vertragsverhandlungen durchzuführen. Es ist erwähnenswert, dass der Eclipse Dataspace Connector auch Teil der Projekte GAIA-X und Catena-X ist (Catena-X Automotive Network Konsortialprojekts, 2022; Eclipse Foundation, o. D.; European Commission, 2023; Fraunhofer-Institut für Software- und Systemtechnik ISST; Kanal, 2022; Kolodziej & Vazquez, 2023).

Schwerpunkt organisatorische Integration

Wie bereits erwähnt sind Unternehmen sehr stark daran interessiert ihre Kernkompetenzen und geistiges Eigentum zu schützen (Uhlenkamp et al., 2020). Daher besteht eine große Herausforderung für Manufacturing-X darin, Unternehmen dazu zu motivieren, aktiv ihre Daten zu teilen, insbesondere wenn es sich um vertrauliche Informationen handelt. Um dieser Herausforderung zu begegnen, wird nachfolgend ein Anreizsystem vorgestellt, das Unternehmen motivieren soll, ihre Daten aktiv mit dem Datenökosystem von Manufacturing-X zu teilen. Bei der Entwicklung dieses Anreizsystems ist auf Erfahrungen aus anderen Bereichen, wie zum Beispiel Gaming oder Dating Plattformen, Bezug genommen worden.

Eine häufige Grundlage für digitale Anreizsysteme bildet die Einführung eines Krypto Token. Diesen können die Teilnehmer von Manufacturing-X nicht käuflich erwerben, sondern nur als eine Art Belohnung erhalten, wobei deren Wert durch die Menge und Qualität der geteilten Daten bestimmt wird. Die Tokens lassen sich über eine Blockchain-Plattform wie Polygon automatisch an Unternehmen ausgeben. Unternehmen können die erhaltenen Tokens dann dazu nutzen, um detailliertere Daten zu kaufen und freizuschalten oder andere Vorteile zu erhalten (bspw. Steuererleichterungen, Fördergelder, etc.). Dabei muss jedoch erwähnt werden, dass anfallende GAS-Fees sich hierbei als problematisch erweisen können.

Auf der Grundlage der Krypto-Token lässt sich eine Rangliste erstellen, ähnlich wie bei vielen Videospiele. Die Unternehmen können auf Grundlage der Anzahl vorhandener Krypto-Token in ihren Wallets in verschiedene Ränge eingestuft werden, wobei die Rangliste z.B. vierteljährlich zurückgesetzt wird. Als Konsequenz haben Unternehmen nur Zugriff auf die Daten von anderen Teil-

nehmern, welche sich im gleichen Rang oder darunter befinden. Eine Voraussetzung für den Einsatz besteht darin, dass die Krypto Wallets der Unternehmen bei Manufacturing-X registriert sein müssen, um die Gesamtanzahl Ihrer Krypto Tokens einsehen zu können. Diese Registrierung ließe sich als Voraussetzung für die Teilnahme an Manufacturing-X festlegen. Eine mögliche Herausforderung besteht hierbei darin, den Unternehmen einen Anreiz zu bieten, eine hohe Platzierung in der Rangliste anzustreben. In diesem Zusammenhang lässt sich argumentieren, dass eine hohe Platzierung die Sichtbarkeit eines Unternehmens erhöht und somit potenziell neue Geschäftsmöglichkeiten eröffnet.

In Hinblick auf die Datensouveränität, die eine zentrale Rolle in Manufacturing-X spielt, haben Unternehmen die Möglichkeit, ihre eigenen Daten zu klassifizieren und festzulegen, die als vertraulich oder nicht vertraulich gelten. Um Unternehmen zu motivieren, auch vertrauliche Daten zu teilen, ließe sich ein Ansatz aus der Dating-App Lovoo adaptieren: Dort können Nutzer erst die Bilder anderer Teilnehmer sehen, wenn sie selbst ausreichend eigene Bilder hochgeladen haben. Entsprechend bestünde der Ansatz in Manufacturing-X darin, dass nur Unternehmen, die ihre eigenen vertraulichen Daten geteilt haben, Zugang zu den vertraulichen Daten anderer Unternehmen erhalten. Eine potenzielle Herausforderung besteht darin, dass Unternehmen versucht sein könnten, absichtlich ihre Daten vertraulicher zu klassifizieren, um Zugang zu den vertraulichen Daten anderer Teilnehmer zu erhalten. Um diesem Problem entgegenzuwirken, könnte eine Abstimmung anderer Unternehmen über die Richtigkeit der Klassifizierung eingeführt werden.

Manufacturing-X bietet die Möglichkeit, einen virtuellen Marktplatz für den Handel mit Daten zu etablieren (VDMA, 2022). Auf diesem Marktplatz können Unternehmen ihre Daten verkaufen und dadurch einen direkten monetären Anreiz erhalten, ihre Daten zu teilen. Diese Möglichkeit schafft eine Win-Win-Situation, da Unternehmen durch den Verkauf ihrer Daten zusätzlich Umsatz generieren können, während andere Unternehmen Zugang zu wertvollen Daten erhalten, um diese für ihre eigenen Zwecke zu nutzen.

Weiterhin besteht die Möglichkeit, dass die Manufacturing-X Betreiber den Nutzern über eine benutzerfreundliche Website oder Software hochwertige Analysefunktionen zur Verfügung stellen. Insbesondere für kleine und mittelständische Unternehmen (KMU), die möglicherweise nicht über die erforderlichen Ressourcen und Fachkenntnisse für die Datenanalyse verfügen, kann dies von hoher Attraktivität sein. Angesichts dessen, dass viele potenzielle Teilnehmer von Manufacturing-X spezialisierte KMU sind, repräsentieren Analysefunktionen einen potenziellen Mehrwert. Optionale könnte die Freischaltung der Analysefunktionen durch die Krypto Token erfolgen. Dadurch würde nicht nur ein Anreiz für KMU geschaffen, sondern auch ein indirekter Anreiz zum Teilen von Daten, da der Erhalt von Krypto-Token an das Teilen von Daten gekoppelt ist.

FAZIT

Bei einem digitalen Zwilling handelt es sich um eine virtuelle Repräsentation eines physischen Objekts, Systems oder Prozesses. Er lässt sich für Simulationen, Analysen und Optimierungen von realen Objekten oder Systemen verwenden. Im Rahmen von Manufacturing-X vermag es ein digitaler Zwilling, Effizienz, Produktivität und Nachhaltigkeit von Fertigungsprozessen zu verbessern. Manufacturing-X baut auf Erfahrungen von Initiativen wie Catena-X auf, besitzt jedoch Spezifika wie eine große Anzahl und Heterogenität an potenziellen Teilnehmern. Entsprechend erweist sich eine Analyse der Rahmenbedingungen als notwendig für eine nachhaltige Nutzung. Hierzu zählen zunächst die Gewährleistung von Datensicherheit, Datensouveränität, vertrauensvollen Datenaustausch und rechtliche Rahmenbedingungen. Darüber hinaus sind jedoch auch technische und organisatorische Herausforderungen zu meistern. Es zeigt sich, dass sich eine föderative Struktur mit souveränen Datenräumen als ein wichtiger Bestandteil des Datenökosystems von Manufacturing-X erweist. Für die teilnehmenden Unternehmen ist es erforderlich, vorab ein geeignetes Geschäftsmodell zum Einsatz eines digitalen Zwillings zu definieren. Hierzu werden ausgewählte Beispiele wie beispielsweise Predictive Maintenance und Manufacturing-as-a-Service gegeben.

Zur Bewältigung dieser komplexen Aufgabe ist ein Phasenmodell erarbeitet worden, das Einführung und Rahmenbedingungen, Geschäftsnutzen, technische sowie organisatorische Integration systematisch würdigt. Dabei werden relevante Aspekte sowohl für die Unternehmen als Nutzer als auch für die Manufacturing-X Betreiber thematisiert. Die technische Implementierung erfährt eine detaillierte Behandlung. Hierfür erfolgt eine Würdigung der OPC UA Standards als adäquate Lösungsmöglichkeit und die Vorstellung einer generische Manufacturing-X Architektur. Für die organisatorische Integration werden Vorschläge gemacht, die einerseits auf Erfahrungen aus dem Gaming-Bereich und von Dating Apps basieren und etablierte Konzepte auf Marktplätzen aufgreifen.

LITERATUR

Amazon Web Services, Inc. (o. D.a) *Data Lakes auf AWS* [Online]. Verfügbar unter <https://aws.amazon.com/de/big-data/datalakes-and-analytics/datalakes/> (Abgerufen am 5 Juli 2023).

Amazon Web Services, Inc. (o. D.b) *Was ist ETL? – Extract Transform Load erklärt – AWS* [Online]. Verfügbar unter <https://aws.amazon.com/de/what-is/etl/> (Abgerufen am 5 Juli 2023).

ascolab (o. D.) *OPC UA Security* [Online]. Verfügbar unter <https://www.ascolab.com/en/technology-unified->

- architecture/technology-security.html (Abgerufen am 30 Juni 2023).
- AWS Workshop Studio (o. D.) *Set up OPC UA Server* [Online]. Verfügbar unter <https://catalog.us-east-1.prod.workshops.aws/workshops/73235981-96cd-4da4-afdf-c2a2c2e9fe88/en-US/onboarding/setup-opc-ua-server> (Abgerufen am 30 Juni 2023).
- Bergs, T, Brecher, C, Schmitt, R & Schuh, G (Hg.) (2020) *Internet of Production - Turning Data into Value: Statusberichte aus der Produktionstechnik 2020*, Aachen, Fraunhofer-Institut für Produktionstechnologie IPT.
- Bitkom e.V. (2023) *Manufacturing-X: Mögliche Anwendungen, Anforderungen und Akteure aus Sicht der Digitalwirtschaft* [Online]. Verfügbar unter <https://www.bitkom.org/sites/main/files/2023-04/230417BitkomManufacturing-XDigitalwirtschaft.pdf>.
- BMW Group (2023) *Catena X: Innovation durch Kooperation* [Online]. Verfügbar unter <https://www.bmwgroup.com/de/news/allgemein/2023/catenax.html> (Abgerufen am 30 April 2023).
- Bundesministerium für Bildung und Forschung (2020) *Collaborative Condition Monitoring* [Online]. Verfügbar unter https://www.plattform-i40.de/IP/Redaktion/DE/Kurzmeldungen/2020/2020-02-18_Collaborative-Condition-Monitoring.html (Abgerufen am 6 Juli 2023).
- Bundesministerium für Wirtschaft und Klimaschutz (o. D.) *Collaborative Condition Monitoring* [Online]. Verfügbar unter <https://www.bmwk.de/Redaktion/EN/Artikel/Digital-World/GAIA-X-Use-Cases/collaborative-condition-monitoring.html> (Abgerufen am 6 Juli 2023).
- Bundesministerium für Wirtschaft und Klimaschutz (2023a) *Weg zur aktiven Teilnahme - Informationen für Datenanbieter und -nutzer* [Online]. Verfügbar unter <https://catena-x.net/de/catena-x-einfuehren-umsetzen/einfuehrung-von-catena-x> (Abgerufen am 6 Juli 2023).
- Bundesministerium für Wirtschaft und Klimaschutz (2023b) *Manufacturing-X: Initiative zur Digitalisierung der Lieferketten in der Industrie* [Online]. Verfügbar unter <https://www.plattform-i40.de/IP/Navigation/DE/Manufacturing-X/Initiative/initiative-manufacturing-x.html> (Abgerufen am 7 Juli 2023).
- Bundesministerium für Wirtschaft und Klimaschutz (2023c) *Manufacturing-X – FAQ* [Online]. Verfügbar unter https://www.plattform-i40.de/IP/Redaktion/DE/Standardartikel/ManufacturingX_FAQ.html (Abgerufen am 7 Juli 2023).
- Catena-X (o. D.) *Die Vision von Catena-X* [Online]. Verfügbar unter <https://catena-x.net/de/vision-ziele> (Abgerufen am 30 April 2023).
- Catena-X (2022) *Semantic Layer/Digital Twins* [Online]. Verfügbar unter <https://catena-x.net/de/angebotestandards/digitaler-zwilling> (Abgerufen am 30 Juni 2023).
- Catena-X Automotive Network Konsortialprojekts (2022) *Eclipse Dataspace Connector (EDC): zentrale Komponente von Catena-X | Catena-X* [Online]. Verfügbar unter <https://catena-x.net/de/angebotestandards/edc-die-zentrale-komponente-fuer-die-navigation> (Abgerufen am 5 Juli 2023).
- Catena-X Automotive Network Konsortialprojekts (2023) *Weg zur aktiven Teilnahme* [Online]. Verfügbar unter <https://catena-x.net/de/catena-x-einfuehren-umsetzen/einfuehrung-von-catena-x> (Abgerufen am 6 Juli 2023).
- DHL Trend Research (2019) „Digital Twins in Logistics: A DHL perspective on the impact of digital twins on the logistics industry“ [Online]. Verfügbar unter <https://www.dhl.com/content/dam/dhl/global/core/documents/pdf/glo-core-digital-twins-in-logistics.pdf> (Abgerufen am 30 Juni 2023).
- Eclipse Foundation (Hg.) (o. D.) *Eclipse Dataspace Components* [Online]. Verfügbar unter <https://projects.eclipse.org/projects/technology.edc> (Abgerufen am 30 Juni 2023).

- Esri (2021) *Technologie für digitale Zwillinge und GIS | Was ist ein digitaler Zwilling?* [Online]. Verfügbar unter <https://www.esri.com/de-de/digital-twin/overview> (Abgerufen am 30 Juni 2023).
- European Commission (2023) *Eclipse Dataspace Components* [Online]. Verfügbar unter <https://eclipse-edc.github.io/docs/#/README> (Abgerufen am 3 Juli 2023).
- exone.de (2023) *Digital Twins: Die Zukunft für Unternehmen?* [Online]. Verfügbar unter <https://www.exone.de/ratgeber/digital-twins/> (Abgerufen am 30 Juni 2023).
- Fischer, C. (2020) „Warum dem Digital Twin die Zukunft gehört“, *Deutsche Telekom MMS GmbH*, 15. April [Online]. Verfügbar unter <https://blog.telekom-mms.com/digital-stories/warum-dem-digital-twin-die-zukunft-gehört> (Abgerufen am 30 Juni 2023).
- Schnicke, F. und Kuhn, T. (2020) „Digitaler Zwilling: Aufbau der Industrie 4.0-IT-Infrastruktur“, *Fraunhofer IESE*, 15. Oktober [Online]. Verfügbar unter <https://www.iese.fraunhofer.de/blog/digitale-zwillinge-it-infrastruktur/> (Abgerufen am 30 Juni 2023).
- Fraunhofer-Institut für Software- und Systemtechnik ISST *Eclipse Dataspace Components (EDC) im Gesundheitswesen* [Online]. Verfügbar unter <https://www.isst.fraunhofer.de/de/geschaeftsfelder/gesundheitswesen/technologien/Eclipse-Dataspace-Components.html> (Abgerufen am 5 Juli 2023).
- Grösser, S. Prof. Dr. (2018) „Definition: Digitaler Zwilling“, *Springer Fachmedien Wiesbaden GmbH*, 19. Februar [Online]. Verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/digitaler-zwilling-54371> (Abgerufen am 30 Juni 2023).
- Human, S. (2020) „Das Beste zweier Welten: der Digital Twin“, *Industry of Things*, 3. Februar [Online]. Verfügbar unter <https://www.industry-of-things.de/das-beste-zweier-welten-der-digital-twin-a-900237/> (Abgerufen am 30 Juni 2023).
- IBM (2023) *Was ist ein digitaler Zwilling?* | IBM [Online]. Verfügbar unter <https://www.ibm.com/de-de/topics/what-is-a-digital-twin> (Abgerufen am 30 Juni 2023).
- IPC2U (o. D.) *Die Standards OPC DA und OPC UA ganz einfach erklärt* [Online]. Verfügbar unter <https://ipc2u.de/artikel/wissenswertes/die-standards-opc-da-und-opc-ua-ganz-einfach-erkl-rt/> (Abgerufen am 30 Juni 2023).
- Kanal, M. (2022) *Der Eclipse Dataspace Connector (EDC) – Architektur und Nutzen des Frameworks* [Online]. Verfügbar unter <https://blog.doubleslash.de/der-eclipse-dataspace-connector-edc-architektur-und-nutzen-des-frameworks> (Abgerufen am 5 Juli 2023).
- Kolodziej, M. & Vazquez, P. (2023) *Enabling data sharing through data spaces and AWS* [Online]. Verfügbar unter <https://aws.amazon.com/de/blogs/publicsector/enabling-data-sharing-through-data-spaces-aws/> (Abgerufen am 30 Juni 2023).
- Microsoft (o. D.a) *Azure Data Factory – Datenintegrationsdienst | Microsoft Azure* [Online]. Verfügbar unter <https://azure.microsoft.com/de-de/products/data-factory> (Abgerufen am 5 Juli 2023).
- Microsoft (o. D.b) *Azure Data Lake Storage Gen2 – Preise | Microsoft Azure* [Online]. Verfügbar unter <https://azure.microsoft.com/de-de/pricing/details/storage/data-lake/> (Abgerufen am 5 Juli 2023).
- Neugebauer, J., Heilig, L. & Voß, S. (2023) „Anforderungsanalyse zur Umsetzung eines digitalen Zwillings im Containerterminal“, *HMD Praxis der Wirtschaftsinformatik*, Vol. 60, No. 1, S. 110–131 [Online]. DOI: 10.1365/s40702-022-00941-1.
- OPC Foundation (o. D.a) *Design goals* [Online]. Verfügbar unter <https://reference.opcfoundation.org/Core/Part1/v105/docs/5.3> (Abgerufen am 30 Juni 2023).
- OPC Foundation (o. D.b) *OPC UA Binary* [Online]. Verfügbar unter <https://reference.opcfoundation.org/Core/Part6/v105/docs/5.2> (Abgerufen am 30 Juni 2023).

- OPC Foundation (o. D.c) *OPC UA JSON* [Online]. Verfügbar unter <https://reference.opcfoundation.org/Core/Part6/v105/docs/5.4> (Abgerufen am 30 Juni 2023).
- OPC Foundation (o. D.d) *OPC UA XML* [Online]. Verfügbar unter <https://reference.opcfoundation.org/Core/Part6/v105/docs/5.3> (Abgerufen am 30 Juni 2023).
- OPC Router (o. D.) *Was ist OPC UA? Die wichtigsten Begriffe im Überblick* [Online]. Verfügbar unter <https://www.opc-router.de/was-ist-opc-ua/> (Abgerufen am 30 Juni 2023).
- (2018) *OPC Unified Architecture: Interoperabilität für Industrie 4.0 und das Internet der Dinge* [Online], 8. Aufl., OPC Foundation. Verfügbar unter <https://opcfoundation.org/wp-content/uploads/2017/11/OPC-UA-Interoperability-For-Industrie4-and-IoT-DE.pdf> (Abgerufen am 30 Juni 2023).
- Otto, B., Hompel, M. ten & Wrobel, S. (2022) *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*, Cham, Springer International Publishing; Imprint Springer.
- Pauer, A., Nagel, L., Fedkenhauser, T., Fritzsche-Sterr, Y. & Resetko, A. (2018) „Data exchange as a first step towards data economy“ [Online]. Verfügbar unter <https://www.pwc.de/en/digitale-transformation/data-exchange-as-a-first-step-towards-data-economy.pdf> (Abgerufen am 6 Juli 2023).
- Phoenix Contact (o. D.) *OPC UA* [Online]. Verfügbar unter <https://www.phoenixcontact.com/de-de/technologien/kommunikationstechnologien/opc-ua> (Abgerufen am 30 Juni 2023).
- Plattform Industrie 4.0 (2022) „Whitepaper „Manufacturing-X“: Eckpunkte für die Umsetzung von „Manufacturing-X“ im produzierenden Gewerbe zur Sicherung des Wettbewerbsstandortes Deutschland“ [Online]. Verfügbar unter https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Manufacturing-X_lang.pdf?__blob=publication-File&v=1 (Abgerufen am 30 Juni 2023).
- Plattform Industrie 4.0 (2023) *Manufacturing-X: Initiative zur Digitalisierung der Lieferketten in der Industrie* [Online]. Verfügbar unter <https://www.plattform-i40.de/IP/Navigation/DE/Manufacturing-X/Initiative/initiative-manufacturing-x.html> (Abgerufen am 30 Juni 2023).
- PwC (2018) „Der digitale Zwilling: Erwartungen und Einschätzungen der deutschen Bevölkerung mit besonderem Fokus auf Diabeteserkrankungen“ [Online]. Verfügbar unter <https://www.pwc.de/de/gesundheitswesen-und-pharma/pwc-studie-der-digitale-zwilling.pdf> (Abgerufen am 30 Juni 2023).
- Siemens (o. D.) *OPC UA – strukturierte Daten bis in die Cloud* [Online]. Verfügbar unter <https://www.siemens.com/de/de/produkte/automatisierung/industrielle-kommunikation/opc-ua.html> (Abgerufen am 30 Juni 2023).
- Siemens (2019) *Digital Twin | Siemens* [Online]. Verfügbar unter <https://www.plm.automation.siemens.com/global/en/our-story/glossary/digital-twin/24465> (Abgerufen am 30 Juni 2023).
- Sodabathina, R., Shan, J. & Ulloa, M. (2022) *Building event-driven architectures with IoT sensor data* [Online], AWS Architecture Blog. Verfügbar unter <https://aws.amazon.com/de/blogs/architecture/building-event-driven-architectures-with-iot-sensor-data/> (Abgerufen am 30 Juni 2023).
- Uhlenkamp, J.-F., Hribernik, K. & Thoben, K.-D. (2020) „Wie Digitale Zwillinge Unternehmensgrenzen überwinden“, *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, Vol. 115, s1, S. 84–89.
- van der Aalst, W. M. P., Hinz, O. & Weinhardt, C. (2021) „Resilient Digital Twins“, *Business & Information Systems Engineering*, Vol. 63, No. 6, S. 615–619 [Online]. DOI: 10.1007/s12599-021-00721-z.
- VDMA (2022) „Manufacturing-X: Denkansätze zum Aufbau und zur Etablierung eines deutschen und europäischen Datenökosystems für das produzierende Gewerbe mit der Ausrüsterindustrie als Nukleus und Multiplikator“ [Online]. Verfügbar unter <https://>

vdma.org/documents/34570/55087429/VDMA-Whitepaper%20Manufacturing-X.pdf/7e799522-d86d-5004-32de-4388ee891a8c (Abgerufen am 30 Juni 2023).

vom Dorp, J., Merschjohanna, S., Meier, D., Patzer, F., Karch, M. & Haas, C. (2022) *Sicherheitsanalyse OPC UA* [Online], Bundesamt für Sicherheit in der Informationstechnik. Verfügbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Studien/OPCUA/OPCUA_2022.pdf?__blob=publicationFile&v=10.

Weiss, A. (2022) „Gaia-X – Grundlagen für den Aufbau föderierter, digitaler Ökosysteme nach europäischen Regeln“, *Datenschutz und Datensicherheit - DuD*, Vol. 46, No. 4, S. 227–232 [Online]. DOI: 10.1007/s11623-022-1593-8.

Zentralverband Elektrotechnik- und Elektronikindustrie (2016) „Beispiele-Verwaltungsschale-Industrie-40-Komponente-Basisteil: Fortentwicklung des Referenzmodells für die Industrie 4.0-Komponente“ [Online]. Verfügbar unter https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2016/November/Beispiele_zur_Verwaltungsschale_der_Industrie_4.0-Komponente_-_Basisteil/Beispiele-Verwaltungsschale-Industrie-40-Komponente-White-Paper-Final.pdf (Abgerufen am 5 Juli 2023).

Entwicklung eines Referenzmodells zur Strukturierung des Einsatzes von Self-Service Business Intelligence in Unternehmen

Julian Beck

Kai Holzweißig

Wirtschaftsinformatik

Wirtschaftsinformatik

Duale Hochschule
Baden-Württemberg Stuttgart
Rotebühlstr. 133
70197 Stuttgart

E-Mail: julian-beck@outlook.de

Duale Hochschule
Baden-Württemberg Stuttgart
Rotebühlstr. 133
70197 Stuttgart

E-Mail: kai.holzweissig@dhbw-
stuttgart.de

ABSTRACT

Mit „Self-Service Business Intelligence (SSBI)“ wird gegenwärtig ein Ansatz diskutiert, der Mitarbeiter*innen dazu befähigen soll, anhand einfach zu bedienender BI-Anwendungen, benutzerdefinierte Datenanalysen weitgehend selbstständig durchzuführen. Eine Kernidee hierbei ist, den Wissensaustausch zwischen den verschiedenen Anwendergruppen zu fördern, um ihre Data-Science-Kompetenzen weiterzuentwickeln. Stand heute ist allerdings nicht ausreichend beschrieben, inwiefern eine Zusammenarbeit bei diesen Aufgaben sinnvoll stattfinden kann. Darüber hinaus fehlen bislang Erkenntnisse über die zeitliche Abfolge der Aufgaben und die dabei involvierten Anwender*innen. Im vorliegenden Artikel wird deshalb anhand einer Literaturanalyse sowie qualitativer empirischer Forschungsbefunde ein Referenzmodell für den Einsatz von SSBI in multidisziplinären Teams vorgeschlagen, das die komplexen Interdependenzen zwischen den Anwenderrollen offenlegt und Handlungsempfehlungen zur Strukturierung des Analyseprozesses gibt. Die Ergebnisse zeigen, dass etablierte Modelle wie „Cross-Industry Standard Process for Data Mining (CRISP-DM)“ keine vollständige Unterstützung für SSBI bieten. Anhand der Forschungsbefunde können vier Trajektorien identifiziert werden, die den Verlauf von verschiedenen Analyseaufgaben aufzeigen und dabei die Zusammenarbeit der Rollen hervorheben. Dem übergeordnet wird eine Reihe von Aktivitäten dargelegt, die eine Ausgangsbasis für die Ableitung weiterer Trajektorien schafft. Insgesamt unterstützt das Referenzmodell Unternehmen darin, BI-relevante Abläufe schneller anzupassen und Prozessschwächen leichter identifizieren zu können.

SCHLÜSSELWÖRTER

Self-Service Business Intelligence, Referenzmodellierung, Datenanalyse, Kooperation

EINLEITUNG: SSBI IN UNTERNEHMEN NUTZEN

Die Erfassung und Analyse von Big Data eröffnet Unternehmen neue Möglichkeiten, ihre Prozesse zu optimie-

ren, um damit ihre Produktivität steigern und ihren Wettbewerbsvorteil ausbauen zu können (Daradkeh 2021). Hierbei stehen Unternehmen u.a. vor der Herausforderung, eine datengetriebene Unternehmenskultur zu schaffen, in der Handlungen durch Daten und Analysen gestützt werden (Berndtsson et al. 2020). Um eine solche Kultur in Unternehmen erfolgreich zu etablieren, bedarf es auch geeigneter Analyseplattformen, die einfach zu bedienen sind (Berndtsson et al. 2018). Solche Plattformen ermöglichen es Unternehmen, Fachwissen wirksam einzusetzen und den Mangel an Analyseexpert*innen zu kompensieren (Lismont et al. 2019; Ploder et al. 2020). Vor diesem Hintergrund wird immer häufiger der Begriff „Self-Service Business Intelligence (SSBI)“ als Lösungsdomäne vorgefunden, die „Gelegenheitsnutzer“ in die Lage versetzen soll, benutzerdefinierte Analysen selbstständig durchzuführen (Alpar und Schulz 2016; Passlick et al. 2020; Lennerholt et al. 2020). Bei SSBI-Plattformen handelt es sich um softwaregestützte Werkzeuge, die das übergeordnete Ziel verfolgen, den Wissensaustausch zwischen den diversen SSBI-Anwenderrollen zu fördern (Passlick et al. 2017). Weder in der Forschung noch in der Praxis ist bislang ausreichend beschrieben, inwiefern eine Zusammenarbeit und ein Informationsaustausch der Rollen bei ihren verschiedenen Aufgaben stattfindet (Michalczyk et al. 2020). Die Untersuchung der zeitlichen Abfolge der Aufgaben innerhalb von SSBI stellt auch aus Unternehmenssicht einen vielversprechenden Forschungsansatz dar. Hierdurch können Schwachstellen im Analyseprozess leichter identifiziert und ein strukturierter Arbeitsablauf etabliert werden, durch den unerfahrene Anwender*innen potenziell bessere Leistungen erzielen können (Lismont et al. 2019). Insgesamt existieren bislang keine methodischen Vorgehensweisen, welche den Einsatz von SSBI in Unternehmen strukturierend beschreiben. Andere etablierte Methoden wie „Cross-Industry Standard Process for Data Mining (CRISP-DM)“ bieten hier keine vollständige Unterstützung, da sie zu abstrakt bzw. begrenzt sind. Aufgrund der in Forschung und Praxis bislang fehlenden Formalisierung von SSBI-Tätigkeiten ist es das Ziel des vorliegenden Beitrags, generische Muster von SSBI-Tätigkeiten zu identifizieren und in einem Referenzmodell abzubilden. Dabei sollen die komplexen Interdependenzen zwischen den Anwenderrollen offengelegt und

Erkenntnisse über die verschiedenen Anwenderrollen und deren Aufgaben für die Modellkonstruktion gewonnen werden. Anhand des resultierenden Referenzmodells sollen Unternehmen, insbesondere solche, die heute noch kein SSBI nutzen, in die Lage versetzt werden, ihre SSBI-Tätigkeiten aufzubauen, zu strukturieren und zu steuern.

AKTUELLER STAND DER FORSCHUNG UND PRAXIS

Zur Entwicklung des Referenzmodells wird im Folgenden die konzeptionelle Basis anhand des Stands der Forschung und Praxis gelegt. Als Ausgangspunkt wird hierbei die umfassende Literaturrecherche von Michalczyk et al. zum Thema „SSBI State-Of-The-Art“ herangezogen (Michalczyk et al. 2020). Eine zusätzlich durchgeführte explorative Literaturrecherche einschlägiger Datenbanken ergänzt den Stand mit weiterem relevanten Quellenmaterial. Die entsprechend angepasste bzw. erweiterte Konzeptmatrix ist in Tabelle 1 im Anhang dargestellt. Anhand der Konzeptmatrix ist erkennbar, dass die SSBI-Domäne in mehrere Dimensionen gegliedert werden kann. Von besonderer Bedeutung im Rahmen dieses Beitrags sind die Dimensionen „Anwenderrollen“ („User-Centric“) und „Governance-Herausforderungen“ („Governance-Centric“), da sie wesentliche fachliche Aspekte, aber keine technischen Umsetzungsdetails betreffen. Gleichzeitig bilden diese Dimensionen den Schwerpunkt für die weiteren Betrachtungen in Bezug auf die verschiedenen Anwendergruppen sowie deren Tätigkeiten. Die Dimension „User-Centric“ beschäftigt sich mit dem Ziel, die Nutzerakzeptanz von SSBI-Systemen zu erhöhen, damit Anwender unabhängig ihrer Qualifikationen und ihrem Tätigkeitsbereich mit großen und vielfältigen Datensätzen umgehen können. Dementgegen thematisiert die Dimension „Governance Centric“ die Verwaltung der Daten und inwiefern durch eine fehlende Datenqualität die Ergebnisse der Analysen beeinträchtigt werden können. Im weiteren Verlauf des Beitrags wird mit dem Data-Science-Trajectories-Modell ein Ansatz zur Konzeptualisierung des Gesamtzusammenhangs von SSBI-Tätigkeiten betrachtet.

Beschreibung von SSBI-Anwenderrollen

Grundsätzlich können bei SSBI mehrere Anwenderrollen unterschieden werden. Eine in der Literatur gängige Unterscheidung erfolgt in die drei Kategorien *Casual User* (z.B. Manager, Business User), *Power User* (z.B. Data Scientist, Controller) und *IT User* (z.B. Entwickler) (Alpar und Schulz 2016).

Casual User verfügen über verhältnismäßig geringes Wissen zur Durchführung von Analyseaufgaben (Spahn et al. 2008). Sie erhalten typischerweise Zugriff auf Informationen (Berichte), die bereits erstellt worden sind oder bei denen sie vor der Verarbeitung nur noch Parameter setzen müssen (Michalczyk et al. 2020). Weiterhin haben sie die Möglichkeit, auf Berichte und Dashboards zuzugreifen, welche eine „Drill anywhere“-Möglichkeit

bieten. *Power User* können nicht nur auf Informationen zugreifen, sondern auch neue Informationen oder Informationsquellen erstellen (Michalczyk et al. 2020). Typische Aufgaben konzentrieren sich auf die Datenintegrität oder die Analyse von aufbereiteten Daten. Damit haben diese Anwender erhöhte Anforderungen bezüglich der Funktionsbreite und -tiefe. *IT User* sind für die Bereitstellung der notwendigen technischen Infrastruktur zuständig und fungieren damit als Wegbereiter für die Nutzung der SSBI-Systeme (De Mauro et al. 2018).

Charakterisierung von SSBI-Tätigkeiten

Die genannten Anwenderrollen führen verschiedene SSBI-Tätigkeiten durch, die in diesem Abschnitt genauer diskutiert werden. Eine Klassifizierung von SSBI-Tätigkeiten kann je nach vorausgesetzter Selbstständigkeit der Anwender und der dabei benötigten Systemunterstützung nach dem Self-Service-Konzept von Alpar und Schulz in drei Ebenen erfolgen (Alpar und Schulz 2016). Die Ebenen reichen, wie in Abbildung 1 dargestellt, von der Informationsnutzung (Level 1) über die Informationserstellung (Level 2) bis hin zur Erstellung von Informationsressourcen (Level 3). Die systemtechnische Unterstützung kompensiert bei dem Einsatz von SSBI die fehlenden technischen und/oder analytischen Fähigkeiten der Anwender*innen (Michalczyk et al. 2020). Mit zunehmendem Self-Service Level benötigen die Anwender*innen in Abhängigkeit von ihren Kompetenzen und Fähigkeiten dabei vermehrt Unterstützung (Alpar und Schulz 2016). Durch diesen Ansatz können technisch versierte Anwender*innen wie Data Scientists sowie IT-Abteilungen von den Analyseanfragen der übrigen Anwender*innen entlastet werden (Michalczyk et al. 2020).

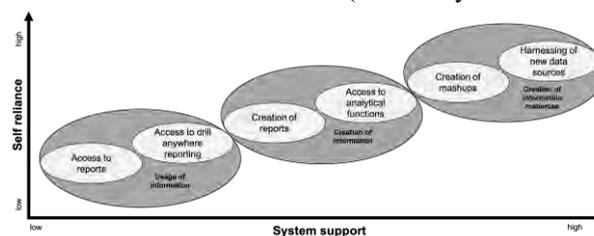


Abbildung 1: Klassifizierung von SSBI-Tätigkeiten (Alpar und Schulz 2016)

Informationsnutzung

Der Zugriff auf bereits erstellte Berichte erfordert von der Rolle *Casual User* den geringsten Grad an Selbstständigkeit und Systemunterstützung (Michalczyk et al. 2020). Im Gegensatz zu traditionellen Berichtssystemen werden die Analysemöglichkeiten im SSBI-Umfeld nicht auf eine bestimmte Anzahl an Standardberichten beschränkt (Alpar und Schulz 2016). Vor allem *Power User* erstellen bei Bedarf zusätzliche Berichte als Basis für den zugrundeliegenden Entscheidungsprozess der *Casual User* (Michalczyk et al. 2020). Damit kann den Anwender*innen Zugang zu allen Berichten verschafft werden, die für sie potenziell relevant sind (Alpar und Schulz 2016). Eine Verbesserung diesbezüglich stellt auf dieser Ebene die Bereitstellung von Berichten und Dashboards mit einer

„Drill anywhere“-Möglichkeit dar, welche durch eine Such-, Filter- und Navigationsfunktion subsumiert werden kann (Smuts et al. 2015).

Erstellung von Informationen

Bei der Berichterstellung werden zunächst die Unternehmensanforderungen der *Casual User* aus fachlicher Sicht erfasst und an die *Power User* übergeben (Michalczyk et al. 2020). Dies ist notwendig, da es aufgrund sich ändernder Anforderungen nicht möglich ist, mit den Standardberichten den gesamten Informationsbedarf zu decken (Alpar und Schulz 2016). Neben der Erstellung von Berichten und Diagrammen kann den Anwendern auch die Möglichkeit gegeben werden, Funktionen des Bereichs „Advanced Analytics“ anzuwenden (Alpar und Schulz 2016).

Erstellung von Informationsressourcen

Eine weitere Möglichkeit für die Rolle *Power User* besteht darin, verschiedene Funktionen mit wiederverwendbaren Komponenten zu kombinieren, die von der Rolle *IT User* bereitgestellt werden (Alpar und Schulz 2016). Hierbei wird es den Anwender*innen ermöglicht, Berichte, Dashboards und andere BI-Ansichten zu personalisieren, indem visuelle Elemente, Daten und Sichten (Views) in eine anpassbare Arbeitsumgebung gezogen werden (Kobieltus 2009). Im Allgemeinen führt die Nutzung neuer Datenquellen (z.B. die Kombination von Daten aus verschiedenen Abteilungen) zu umfassenderen Erkenntnissen für die Entscheidungsfindung (Michalczyk et al. 2020). Daher kann es Anwender*innen ermöglicht werden, neue Datenquellen, die nicht von den IT Usern bereitgestellt werden, selbstständig für die Analyse einzubinden und mit Unternehmensdaten zu kombinieren (Bani-Hani et al. 2018).

Insgesamt gibt das oben behandelte Self-Service-Konzept von Alpar und Schulz zwar Aufschluss über die diversen SSBI-Tätigkeiten der Anwender*innen, allerdings bleibt dabei unklar, inwiefern Interdependenzen zwischen diesen Tätigkeiten bestehen (Michalczyk et al. 2020). So ist beispielsweise die Erstellung eines Berichts eine notwendige Voraussetzung für dessen Zugriff (Michalczyk et al. 2020). Diese Abhängigkeiten gilt es, in dem zu entwickelndem Referenzmodell zu berücksichtigen.

Analyse des Data-Science-Trajectories-Modell

In der SSBI-Forschung gibt es Stand heute kaum Erkenntnisse über die zeitliche Abfolge oder Beteiligung von Anwenderrollen im Rahmen von SSBI (Michalczyk et al. 2020). Etablierte Methoden wie CRISP-DM bieten keine vollständige Unterstützung für den Einsatz von SSBI innerhalb von Unternehmen, da sie zu abstrakt bzw. zu begrenzt sind (Michalczyk et al. 2020). Dennoch gilt unter den bestehenden Vorgehensmodellen CRISP-DM nach wie vor als Industriestandard (Fernandes et al. 2017). Die zunehmende Menge vielfältigster Daten sowie der experimentelle und explorative Charakter von Data-Science-Projekten erfordern allerdings flexiblere Vorgehensmodelle als die bisherigen (Martinez-Plumed et al. 2021).

Vor diesem Hintergrund wird mit „Data Science Trajectories (DST)“ ein Modell vorgeschlagen, das eine Überarbeitung von CRISP-DM (Chapman et al. 2001) darstellt, ohne jedoch dessen Kernaktivitäten zu verwerfen (Martinez-Plumed et al. 2021). Das DST-Modell (vgl. Abbildung 2) enthält einen äußeren Kreis mit Data Science (oder explorativen) Aktivitäten, einen inneren Kreis mit CRISP-DM (oder zielgerichteten) Aktivitäten und im Kern Datenmanagement-Aktivitäten (Martinez-Plumed et al. 2021). In Anbetracht der verschiedenen Szenarien, in denen Daten verwendet werden können, sind die folgenden Datenmanagement-Aktivitäten zu unterscheiden: (Martinez-Plumed et al. 2021)

Data Acquisition: Einbindung oder Erstellung relevanter Daten, z.B. durch die Installation von Sensoren

Data Simulation: Simulation komplexer Systeme, um nützliche Daten zu erstellen und kausale Fragestellungen zu formulieren

Data Architecting: Gestaltung des logischen und physischen Datenlayouts und Integration verschiedener Datenquellen

Data Release: Bereitstellung der Daten durch Datenbanken, Schnittstellen und Visualisierungen

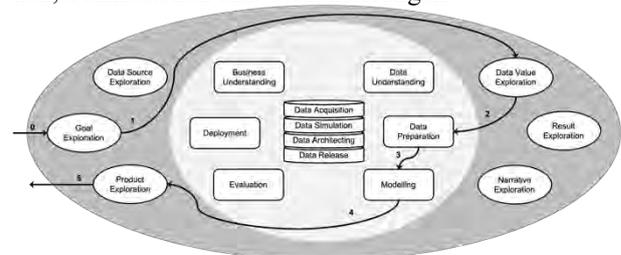


Abbildung 2: Data-Science-Trajectories-Modell mit beispielhafter Trajektorie (Martinez-Plumed et al. 2021)

Im Gegensatz zu CRISP-DM müssen im DST-Modell die Aktivitäten nicht in einer vorgegebenen Reihenfolge durchgeführt werden (Martinez-Plumed et al. 2021). Stattdessen werden sogenannte Trajektorien (Verlaufs-pfade) durch das Modell gelegt, welche Vorgehensweisen für diverse Arten von Data-Science-Projekten berücksichtigen (Martinez-Plumed et al. 2021). Die Trajektorien sollen als Vorlage für Data Scientists bei der Planung ihrer Projekte dienen, indem sie Aktivitäten aufzeigen, die abhängig vom Kontext zu ihren Arbeitsabläufen hinzugefügt oder aus ihnen entfernt werden können (Martinez-Plumed et al. 2021). Folgende Gründe sind ausschlaggebend für die Verwendung des DST-Modells als Grundlage für das angestrebte Referenzmodell:

Flexibilität: Im SSBI-Umfeld werden verschiedenste Tätigkeiten durchgeführt, die durch Trajektorien dargestellt werden können.

Verallgemeinerbarkeit: Die Trajektorien sind als Vorschlag für die Durchführung eines Projektes anzusehen und können auf das jeweilige Unternehmensumfeld angepasst werden.

Nachverfolgbarkeit: Durch den definierten Prozess kann der Status eines Projektes verfolgt werden, womit Optimierungspotenziale leichter identifizierbar sind.

Zur Adaption des Modells auf den SSBI-Kontext gilt es bei der anschließenden Modellkonstruktion Anpassungen an den äußeren Kreisen vorzunehmen, um die Anwenderrollen mit ihren Tätigkeiten abbilden zu können. Außerdem sind für die Darstellung der Trajektorien weitere Befunde aus der Unternehmenspraxis zu erheben.

ENTWICKLUNG EINES SSBI-REFERENZMODELLS

Forschungsmethodisches Vorgehen

Das angewandte Untersuchungsdesign orientiert sich an dem Prozessmodell „elaborated Action Design Research“ (Mullarkey und Hevner 2019). Dabei bildet der methodische Ansatz von „Design Science Research“ (Peffers et al. 2007) die theoretische Grundlage, die um Konzepte der Aktionsforschung ergänzt wird (Sein et al. 2011). Um empirische Daten zu den Abläufen eines SSBI-Analyseprozesses zu erheben, wurden neun leitfadengestützte Experteninterviews in einem international tätigen Großunternehmen geführt und die Ergebnisse davon wörtlich transkribiert (Kruse 2015). Bei den Experten handelte es sich um Mitarbeiter des Controllings in einem Großunternehmen, die in ihrem Arbeitsumfeld Zugriff auf diverse SSBI-Tools haben und unterschiedliche Schwerpunkte bei der Analyse setzen. Aufgrund der überschaubaren Fallauswahl ist aber darauf hinzuweisen, dass die Erkenntnisse dieser Arbeit keinen Anspruch auf Repräsentativität erheben. Die Module des verwendeten Interviewleitfadens orientieren sich an den Definitionen der Phasen von CRISP-DM, um zu überprüfen, inwiefern die Prozessschritte von SSBI und Data Mining bei der Analyse übereinstimmen. Die Fragen innerhalb dieser Module setzen einen Fokus auf Handlungsabläufe und Meinungen bzw. Einschätzungen bezüglich des Analyseprozesses. Daher sind die Fragen möglichst offen formuliert, um den Befragten Raum zu geben, ihre Relevanzsysteme darzustellen und bislang unbekannte Aspekte einzubringen (Kruse 2015). Die Synthese der Befunde erfolgte mittels einer inhaltlich strukturierenden Inhaltsanalyse (Kuckartz 2018). Nach der Empfehlung von Kuckartz und Rädiker wurde die konkrete Umsetzung mit der Software MAXQDA vorgenommen (Kuckartz et al. 2020). Die anschließende Artefaktkonstruktion wurde anhand der erzielten theoretischen und praktischen Erkenntnisse in Form einer Referenzmodellierung umgesetzt (Schütte 1998). Insbesondere die Auswertung des Interviewmaterials anhand einer „Konfiguration von Kategorien“ lieferte hierfür relevante Informationen (Kuckartz 2018). Durch diese Methodik konnten Übereinstimmungen in den Analyseaufgaben der Experten mittels der Häufigkeit von bestimmten Kombinationen von Codes festgestellt werden (z.B. bei wie vielen Experten die Phase „Business Understanding“ durchgeführt wurde).

Ableitung eines Master-Referenzmodells für Analyseaufgaben

Basierend auf den Ergebnissen der Literaturanalyse sowie der Analyse der Befragung wurde das in Abbildung 3 dargestellte Modell abgeleitet. Aufgrund des Umstandes, dass basierend auf dem Referenzmodell, ebenso wie bei DST, Trajektorien abzuleiten sind, kann das zu entwickelnde Modell als Metamodell oder Master-Referenzmodell verstanden werden (Schütte 1998). Im inneren Kreis werden die Datenmanagement- und CRISP-DM-Aktivitäten dargestellt. Obwohl die Rolle *IT User* nur bei den Aktivitäten *Data Understanding*, *Data Preparation* und *Data Archiving* beteiligt ist, wird sie trotzdem dem inneren Kreis repräsentativ zugeordnet, um ihre Funktion als Wegbereiter für die Nutzung von SSBI-Systemen zu verdeutlichen. In den äußeren Kreisen werden die Anwenderrollen *Casual User* und *Power User* mit ihren korrespondierenden SSBI-Aktivitäten dargestellt. Im Vergleich zum DST-Modell fehlt die Datenmanagement-Aktivität *Data Acquisition*, da diese in ihrer Bedeutung mit der Aktivität *Harnessing of new data sources* übereinstimmt. Außerdem konnten durch die Befragungsergebnisse noch keine Trajektorie für die Aktivitäten *Access to analytical functions* und *Data Simulation* abgeleitet werden.

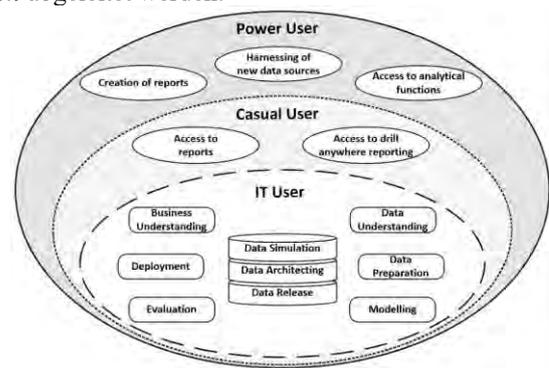


Abbildung 3: Entwickeltes Master-Referenzmodell für SSBI-Analyseprozesse

Ableitung von Trajektorien für Analyseaufgaben

Neben der Entwicklung des Master-Referenzmodells zielte das Forschungsdesign darauf ab, Trajektorien, also Muster für Verlaufspfade von Data-Science-Aktivitäten, zu identifizieren. Anhand der Befragungsergebnisse waren vier Trajektorien ableitbar, die im Folgenden diskutiert werden. Eine wesentliche Bedeutung in den folgenden Abbildungen haben sowohl die Linientypen, welche unterschiedliche Anwenderrollen repräsentieren, als auch die in Klammern gesetzten Nummerierungen, die optionale Aktivitäten oder Schleifen kennzeichnen.

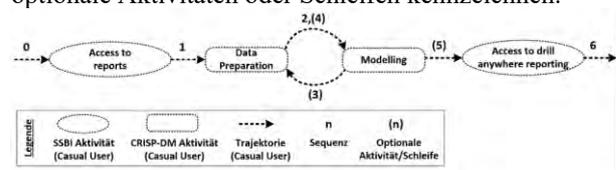


Abbildung 4: SSBI-Analyseaufgabe eines Casual User

Die Trajektorie in Abbildung 4 zeigt das Vorgehen eines *Casual User* bei der Analyse im Kontext der Informationsnutzung. Eine solche Aufgabe ist vom prinzipiellen Vorgehen her wenig komplex, weshalb dieses Modell zwar hinsichtlich der Vollständigkeit einen sinnvollen Beitrag für die Theorie darstellt, bei dem in Bezug auf die Praxis allerdings fragwürdig ist, ob es einen größeren Nutzen für die Implementierung von Prozessen bietet. Grundsätzlich werden bei dieser Form der Analyse zunächst ein Bericht aufgerufen und bestimmte Parameter gesetzt, um die Datenbankabfrage einzugrenzen. In der Phase *Modelling* kann der Bericht anschließend nach den jeweiligen Anforderungen angepasst werden, indem der Anwender bzw. die Anwenderin auf verschiedenen Ebenen *drillt* oder Attribute filtert. Hinsichtlich des Referenzmodells wird unter *Modelling* der Einsatz verschiedener Visualisierungen zur Aufbereitung eines Berichts, basierend auf einer definierten Datengrundlage verstanden. Außerdem besteht eine Rückkopplungsschleife zur Phase *Data Preparation*, falls durch Änderung der Parameter andere Daten betrachtet werden sollen. Da bei dieser Aufgabe lediglich Informationen konsumiert werden, beläuft sich die Dauer auf wenige Minuten.

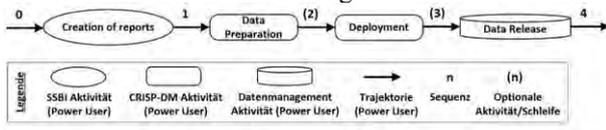


Abbildung 5: SSBI-Analyseaufgabe eines Power User mit Fokus auf der Datenvorbereitung

Die Trajektorie in Abbildung 5 stellt ebenso eine verhältnismäßig kurzzeitige Aufgabe dar, deren Dauer auf 30 Minuten geschätzt wird. Dabei ist ein bereits bestehender Bericht von einem Power User erneut zu generieren, indem im Rahmen der *Data Preparation* die Datengrundlage an die entsprechenden Anforderungen angepasst wird. Dazu werden die Attribute der Datenbankabfrage eingeschränkt und die daraus resultierende Ergebnismenge zusätzlich durch spezifische Parameter gefiltert, damit die Abfrage nur die relevanten Daten für den Bericht liefert. Für das abschließende *Deployment* wird sowohl die Zielgruppe auf den Bericht selbst, als auch die zugehörige Datengrundlage berechtigt, weshalb als *Datenmanagement-Aktivität* das *Data Release* folgt. Diese Aktivitäten können entfallen, wenn die Zielgruppe den Bericht als einfache Visualisierung ohne SSBI-Funktionalität erhält.

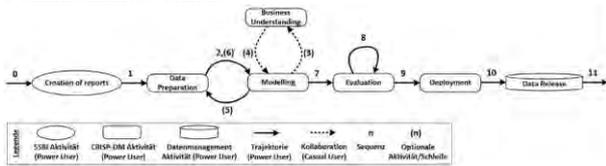


Abbildung 6: SSBI-Analyseaufgabe eines Power User ohne Einarbeitungsbedarf

Bei der Analyse in Abbildung 6 handelt es sich um eine aufwändigere Aufgabe, die einen zeitlichen Horizont von sechs Wochen vorsieht, was in der Beteiligung mehrerer Personen begründet liegt. Eine tiefgehende Einarbeitung

wird bei dieser Aufgabe nicht benötigt, vorausgesetzt die grundlegende Zielsetzung ist verständlich und es besteht eine klar definierte Vorgabe zur Datengrundlage. Abhängig von der Aufgabe muss bei der *Data Preparation* entweder eine Datenquelle erstellt oder eine bestehende Datenquelle als Grundlage genutzt werden. Zu dieser Phase bestehen Rückkopplungsschleifen, falls während der Entwicklung Änderungen an den Daten erforderlich sind. Bei der anschließenden Phase *Modelling* können ebenso Rückkopplungsschleifen auftreten, die aus den kontinuierlichen Abstimmungsterminen mit der Rolle *Casual User* resultieren. Das ist jedoch als optional zu betrachten, da der Bericht auch erst nach Fertigstellung der Entwicklung der Zielgruppe vorgestellt werden kann. Die anschließende Evaluation findet in Zusammenarbeit der *Power User* untereinander statt, wodurch mehrere Zyklen existieren. Prinzipiell sind auch hier Rückkopplungsschleifen vorhanden, die allerdings nicht dargestellt werden, da in Abhängigkeit des Änderungsgrunds unterschiedliche Phasen zu wiederholen sind. Die Vorgehensweise der danach folgenden Aktivitäten (*Deployment* und *Data Release*) stimmen mit den Beschreibungen zu Abbildung 5 überein.

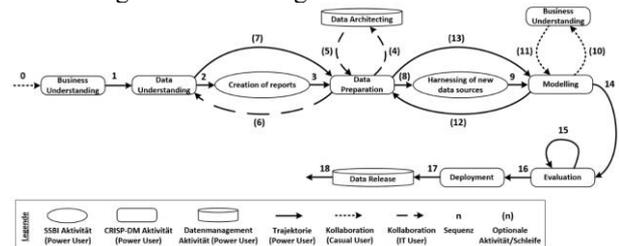


Abbildung 7: SSBI-Analyseaufgabe eines Power User mit hoher Komplexität

Die Trajektorie in Abbildung 7 repräsentiert die Vorgehensweise bei der Analyse der meisten Experten, wobei die Dauer dieser Analysen von einigen Wochen bis hin zu mehreren Monaten variiert. Das Modell weist viele optionale Aktivitäten sowie Schleifen auf, weshalb grundsätzlich weitere Trajektorien daraus abgeleitet werden könnten. Allerdings wird dies nicht als sinnvoll erachtet, da erst im Verlauf des Analyseprozesses deutlich wird, ob die jeweiligen Aktivitäten notwendig sind oder entfallen können. Bezüglich der Ausführungen zu dieser Analyseaufgabe werden lediglich Aspekte erläutert, die nicht bereits im Zusammenhang mit Abbildung 6 genannt worden sind. Es handelt sich bei dieser Analyse um einen komplexen Aufgabentyp, der eine Einarbeitung verantwortlicher *Power User* erfordert. In der Phase des *Business Understanding* werden in Zusammenarbeit mit einem *Casual User* die spezifischen Anforderungen und Zielsetzungen im Dialog erarbeitet. Bei dem darauffolgenden *Data Understanding* wird sich mit den wesentlichen Daten für den Bericht auseinandergesetzt. Zu dieser Phase bestehen Rückkopplungsschleifen, falls während der Entwicklung Rückfragen hinsichtlich der Daten entstehen und daher ein *IT User* konsultiert werden muss. Bei der *Data Preparation* kann ebenso die Unterstützung der Rolle *IT User* benötigt werden, falls bestimmte Daten fehlen, die daher entsprechend anzufordern sind. Dabei

wird von dem *IT User* die Datenmanagement-Aktivität *Data Architecting* durchgeführt, die in der Gestaltung eines logischen und physischen Datenlayouts besteht. Falls bei der *Data Preparation* die Anbindung einer neuen Datenquelle erforderlich ist, wird die Aktivität *Harnessing of new data sources* von dem *Power User* ausgeführt, um die bestehenden Unternehmensdaten mit weiteren Daten anzureichern. Die danach folgenden Aktivitäten entsprechen den Erläuterungen zu Abbildung 6.

Insgesamt wird durch die Trajektorien deutlich, wieso CRISP-DM keine vollständige Unterstützung für den Einsatz von SSBI innerhalb von Unternehmen bietet, da die SSBI-Analyseaufgaben von sehr unterschiedlichem Umfang und Inhalt sind. In Bezug auf das DST-Modell kann außerdem festgehalten werden, dass die Trajektorien aufgrund der vielen Rückkopplungsschleifen eine höhere Dynamik aufweisen.

KRITISCHE REFLEXION UND BEDEUTUNG FÜR DIE PRAXIS

Das Ziel des vorliegenden Beitrags bestand in der Entwicklung eines Referenzmodells zur Strukturierung und Steuerung von SSBI-Tätigkeiten in Unternehmen. Es konnte gezeigt werden, dass der SSBI-Analyseprozess grundlegend nach den Phasen von CRISP-DM strukturiert werden kann. Allerdings bietet das CRISP-DM-Vorgehensmodell keine vollständige Unterstützung für den SSBI-Bereich, da dort die Analyseaufgaben in ihrem Umfang und Inhalt deutlich variieren. Das entwickelte Referenzmodell definiert eine Reihe von Aktivitäten, um dem SSBI-Analyseprozess vollständig Rechnung tragen zu können. Im Wesentlichen beinhaltet das Modell Datenmanagement Aktivitäten, CRISP-DM-Aktivitäten und SSBI-spezifische Aktivitäten, die unterschiedlichen Anwenderrollen zugeordnet sind. Anhand der Befunde der durchgeführten Befragung konnten vier Trajektorien abgeleitet werden, die exemplarisch den Verlauf von verschiedenen SSBI-Analyseaufgaben aufzeigen und dabei die Zusammenarbeit der Rollen hervorheben. Diese Trajektorien bieten eine Referenz für den Einsatz von SSBI in Unternehmen und können an die unternehmensspezifischen Gegebenheiten angepasst werden. Ferner bietet das Referenzmodell eine Ausgangsbasis, um weitere Analyseaufgaben abbilden zu können. Insgesamt können anhand des entwickelten Modells Abläufe in Unternehmen schneller adaptiert und Prozessschwächen leichter identifiziert werden.

In Bezug auf die verwendete Forschungsmethodik ist kritisch anzumerken, dass eine Generalisierbarkeit der Ergebnisse der Befragung in Anbetracht der Stichprobencharakteristik mit Vorsicht betrachtet werden muss. Weiterhin bedürfen das erstellte Master-Referenzmodell und die dazugehörigen Trajektorien hinsichtlich ihres unternehmenspraktischen Nutzens einer weiteren Evaluation. Als abschließendes Fazit kann festgehalten werden, dass sich die Diskussion des SSBI-Analyseprozesses in Forschung und Praxis noch in einem frühen Stadium befindet. Derzeitige Forschungsprojekte thematisieren bislang kaum das eigentliche Vorgehen bei der Analyse,

weshalb es weiterer wissenschaftlicher Arbeiten auf diesem Gebiet bedarf, um Unternehmen zukünftig optimal bei dem Einsatz von SSBI unterstützen zu können.

LITERATUR

- Alpar P, Schulz M (2016) Self-Service Business Intelligence, *Business & Information Systems Engineering*, 58:51–155
- Bani-Hani I, Pareigis J, Tona O, Carlsson S (2018) A holistic view of value generation process in a SSBI environment: a service dominant logic perspective, *Journal of Decision Systems*, 27:46–55
- Berndtsson M, Forsberg D, Stein D, Svahn T (2018) BECOMING A DATA-DRIVEN ORGANISATION, *Research-in-Progress Papers*, S 1–9
- Berndtsson M, Lennerholt C, Svahn T, Larsson P (2020) 13 Organizations' Attempts to Become Data-Driven, *International Journal of Business Intelligence Research*, 11:1–21
- Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2001) CRISP-DM 1.0 Step-by-step data mining guide, SPSS Inc., USA
- Daradkeh MK (2021) Enterprise Data Lake Management in Business Intelligence and Analytics, In: Tavana et al. (Hrsg) *Integration Challenges for Analytics, Business Intelligence, and Data Mining*, IGI Global, S 92–113
- Fernandes E, Carvalho R, Holanda M, van Erven G (2017) Educational Data Mining: Discovery Standards of Academic Performance by Students in Public High Schools in the Federal District of Brazil, In Rocha et al. (Hrsg) *Recent Advances in Information Systems and Technologies*, Springer International Publishing, Cham, S 287–296
- Kobielus J (2009) *Mighty Mashups: Do-It-Yourself Business Intelligence For The New Economy*, Forrester Research, Cambridge
- Kruse J (2015) *Qualitative Interviewforschung, Ein integrativer Ansatz, 2., überarbeitete und ergänzte Auflage*, Beltz Juventa, Weinheim, Basel
- Kuckartz U (2018) *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*, 4. Auflage, Beltz Juventa, Weinheim, Basel
- Kuckartz U, Rädiker S (2020) *Fokussierte Interviewanalyse mit MAXQDA. Schritt für Schritt*, 1. Auflage, Springer Fachmedien Wiesbaden, Wiesbaden
- Lennerholt C, van Laera J, Söderström E (2020) User related Challenges of Self-Service Business Intelligence, *Information Systems Management*, 38:309–323
- Lismont J, van Calster T, Óskarsdóttir M, vanden Broucke, S, Baesens B, Lemahieu W, Vanthienen J (2019) Closing the Gap Between Experts and Novices Using Analytics-as-a-Service: An Experimental Study, *Business & Information Systems Engineering*, 61:679–693
- Martinez-Plumed F, Contreras-Ochando L, Ferri C, Hernandez-Orallo J, Kull M, Lachiche N, Ramirez Quintana MJ, Flach PA (2021) CRISP-DM Twenty Years

- Later: From Data Mining Processes to Data Science Trajectories, *IEEE Transactions on Knowledge and Data Engineering*, 33:3048–3061
- De Mauro A, Greco M, Grimaldi M, Ritala P (2018) Human resources for Big Data professions: A systematic classification of job roles and required skill sets, *Information Processing & Management*, 54:807–817
- Michalczyk S, Nadj M, Azarfar D, Mädche A, Gröger C (2020) A State-of-the-Art Overview and Future Research Avenues of Self-Service Business Intelligence and Analytics, In: *Proceedings of the 28th European Conference on Information Systems (ECIS 2020)*, An Online AIS Conference, June 15-17
- Mullarkey MT, Hevner AR (2019) An elaborated action design research process model, *European Journal of Information Systems*, 28:6–20
- Passlick J, Lebek B, Breitner M (Hrsg) (2017) A Self-Service Supporting Business Intelligence and Big Data Analytics Architecture, In: *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*. St. Gallen, S 1126–1140
- Passlick J, Guhr N, Lebek B, Breitner MH (2020) Encouraging the use of self-service business intelligence – an examination of employee-related influencing factors, *Journal of Decision Systems*, 29:1–26
- Peppers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A Design Science Research Methodology for Information Systems Research, *Journal of Management Information Systems*, 24:45–77
- Ploder C, Bernsteiner R, Dilger T (2020) Improving Decision Quality for Business Users Based on Cloud-based Self-Service Business Intelligence Tools, *Cloud Computing and Data Science*, 1:1–11
- Schütte R (1998) Grundsätze ordnungsmäßiger Referenzmodellierung, Konstruktion konfigurations- und anpassungsorientierter Modelle, Gabler, Wiesbaden
- Sein MK, Henfridsson O, Purao S, Rossi M, Lindgren (2011) Action Design Research, *MIS Quarterly*, 35:37–56
- Smuts M, Scholtz B, Calitz A (2015) Design Guidelines for Business Intelligence Tools for Novice Users, In: *Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists - SAICSIT '15*. ACM Press, New York, S 1–15
- Spahn M, Kleb J, Grimm S, Scheidl S (2008) Supporting Business Intelligence by Providing Ontology-Based End-User Information Self-Service, In: *OBI '08: Proceedings of the first international workshop on Ontology-supported business intelligence*, Article No.: 10, S 1-12

KONTAKT

Julian Beck

Wirtschaftsinformatik
Duale Hochschule Baden-Württemberg
Stuttgart
Rotebühlstr. 133
70197 Stuttgart
E-Mail: julian-beck@outlook.de

Kai Holzweißig

Wirtschaftsinformatik
Duale Hochschule Baden-Württemberg
Stuttgart
Rotebühlstr. 133
70197 Stuttgart
E-Mail: kai.holzweissig@dhbw-stuttgart.de

ANHANG

	Perspective		User Role			Levels of Self-Service							Total
	User-Centric	Governance-Centric	Casual User	Power User	IT User	Usage of Information		Creation of information		Creation of information resources			
						Access to reports	Access to drill anywhere reporting	Creation of reports	Access to analytical functions	Creation of mashups	Harnessing of new data sources	Deployment of analytical functions	
Alpar & Schulz 2016	X	X	X	X		X	X	X		X	X		9
Bani-Hani et al. 2018	X		X	X	X	X		X	X	X	X		9
Berndtsson et al. 2018	X	X	X	X		X		X	X				7
Berndtsson et al. 2020		X	X	X									3
Daradkeh 2021		X											1
De Mauro et al. 2018			X	X	X								3
Kobielus 2009	X		X	X						X			4
Lennerholt et al. 2020	X	X	X	X	X			X			X		7
Lismont et al. 2019	X		X	X	X				X				5
Michalczyk et al. 2020	X	X	X	X	X	X	X	X	X	X	X	X	12
Passlick et al. 2017	X	X	X	X		X	X	X	X	X	X	X	11
Passlick et al. 2020	X	X	X										3
Ploder et al. 2020	X		X	X									3
Smuts et al. 2015	X		X			X	X	X		X	X		7
Spahn et al. 2008	X		X			X		X		X	X	X	7
Total	12	8	14	11	5	7	4	8	5	7	7	3	91

Tabelle 2: Konzeptmatrix zur Literaturanalyse

A machine learning based approach on employee attrition prediction with an emphasize on predicting leaving reasons

Fabian Engl
Vitesco Technologies GmbH
People Analytics and Technology;
Ostbayerische Technische
Hochschule Regensburg
Siemensstraße 10-12, 93055 Regensburg

Email: fabian.engl@oth-regensburg.de

Frank Herrmann
Ostbayerische Technische
Hochschule Regensburg
Labor Wirtschaftsinformatik, SAP
und Produktionslogistik
Galgenbergstraße 32, 93053 Regensburg

Email: frank.herrmann@oth-regensburg.de

Abstract—Using Vitesco Technologies as an example, this article examines whether machine learning models are suitable for detecting employee attrition at an early stage, with the aim of uncovering underlying reasons for leaving. Nine different machine learning algorithms were examined: K-nearest-neighbors, Naive Bayes, logistic regression, a support vector machine, a neural network, a random forest, adaptive boosting, and two gradient boosting models. A three-way-holdout validation method was implemented to assess the quality of the results and measure both the f-score and the degree of model generalization. Initially, it was found that tree-based methods are best suited for classifying employees. A multiclass classification approach showed that under certain conditions it is even possible to predict the underlying leaving reasons.

I. INTRODUCTION

According to a 2018 study conducted by Bund Verlag, 71 percent of German employees reported a lack of joy in their daily work, with one in seven actively considering the prospect of quitting their jobs (Bund-Verlag 2018). However, these circumstances are not limited to just one country, but rather depict a global issue that notably impacts the younger generation (Kelly 2023). This development, combined with a labor market in favor of employees, leads to an increased number of employee-sided resignations. In business, this problem is referred to as employee attrition.

This poses a range of challenges for companies such as Vitesco Technologies, with disengaged employees resulting in reduced productivity, a higher number of workplace accidents, and in the case of resignations even burnout among the remaining colleagues as they have to leverage the additional workload (Wallace 2023). Adding to that, resignations come with substantial additional costs, as replacing a position can often cost three to four times the position's annual salary (Navarra 2022). Therefore, it should be of interest for HR management to identify and address existing dissatisfaction on the side of employees and to mitigate these issues. Yet, HR decisions often rely on subjective judgments, which cannot

fully encompass the complexity of these conflicts. As a result, companies are increasingly implementing data-driven approaches like machine learning (Chugani 2023).

II. PROBLEM DESCRIPTION

Prior to this publication, Vitesco Technologies' did not use machine learning models in an HR context. The goal of this research is to examine whether such models can identify employees at risk of leaving and uncover associated attrition factors. This paper analyses a data set with a total of 1,500 employees from the Chinese locations Tianjin, Wuhu, Changchun and Shanghai and is structured into two parts:

First, we review machine learning algorithms that are commonly used for employee attrition prediction and identify the optimal model for our data set. Here, the emphasis is placed on relevant pre-processing steps and metrics that ensure qualitative machine learning predictions. Building on these findings, the second part of this study delves deeper into distinguishing between reasons for employee attrition. The goal is to answer whether any correlations exist within the data that can provide insights into the employees motivations for leaving.

III. DEFINITION EMPLOYEE ATTRITION

There exist many definitions for employee attrition, which all share a common understanding that it refers to an employee leaving without any action or influence from the company (Arqawi et al. 2022). On one hand, possible reasons for such resignations include a fundamental dissatisfaction (Srivastava and Eachempati 2021). On the other hand, retirement and premature death are also considered employee attrition (Jain and Nayyar 2018). A few publications also list internal job changes and promotions, presuppose the elimination of the position, or even account for layoffs due to poor performance or internal restructuring (Raza et al. 2022; Gim and Im 2023; Alduayj and Rajpoot 2018; Alao and Adeyemo 2013). The

subsequent prediction of employee attrition aims to identify these departures and uncover the accompanying influencing factors.

This study only considers employees who voluntarily left the company. Reasons for leaving include professional and educational development, higher salaries, a relocation (for personal reasons), problems with managers, leadership or company values, and a lack of work-life balance.

IV. LITERATURE REVIEW

The specific focus of employee attrition prediction studies varies depending on the author's research background. Mainly, two perspectives can be distinguished: a business-oriented perspective and an IT perspective. Business-driven publications first develop theses about the reasons for employee attrition and represent them in the form of statistically evaluable features. Here machine learning methods serve only to verify their assumptions (Srivastava and Eachempati 2021). Moreover, they usually include a high proportion of descriptive statistics from which they derive their hypotheses. Some studies also present solutions completely without the use of machine learning (Guerranti and Dimitri 2023; Pawar, Saraf and Pradhan 2023). In the IT-oriented literature, the focus lies on the implementation of machine learning models. In these articles, there is often no formulation of such hypotheses. Since this article examines employee attrition prediction from a computer science perspective, mainly economically inspired publications are not further considered in the following literature review.

Despite numerous publications, there is hardly any practice-related literature on employee attrition prediction. Due to the lack of real industry data, the vast majority of authors resort to the "IBM HR Analytics Employee Attrition & Performance" data set (Cf. et al Fallucchi et al. 2020; Alduayj and Rajpoot 2018; Najafi-Zangeneh et al. 2021; Jain and Nayyar 2018; Bhatta et al. 2022). It contains a partial extract of IBM's employee data. At the time of this article, to our knowledge, only four publications use real company data. Of these, Alao and Adeyemo, for example, analyzed attrition in Southwest Nigerian government institutions, covering a period of over 30 years. However, the data set includes only 309 resignations among 4326 entries; in contrast to roughly 750 in our data set (Alao and Adeyemo 2013). Therefore, the results of Alao and Adeyemo cannot be applied to our use case as large companies such as Vitesco Technologies tend to have a much higher employee turnover rate.

Two other publications by Sikaroudi et al. and Srivastava and Eachempati examine employee attrition within the automotive supplier Arak and a "mid-sized fast-moving consumer goods (FMCG) company" (Sikaroudi, Ghousi and Sikaroudi 2015; Srivastava and Eachempati 2021). Both review multiple machine learning models.

However, the article on Arak only lists the accuracy of the models, which alone is not a sufficient machine learning metric. Other authors supplement it with additional metrics such as precision, recall, or the f-score (Najafi-Zangeneh et al. 2021). Adding to that the accuracy of just under 90 percent can

hardly be put into comparison as no information about the size and structure of the data is given. Basic characteristics used in our approach, such as gender or salary, are also missing. Sikaroudi et al. recommend the use of Naive Bayes, which requires statistically independent characteristics (Ertel 2016).

Srivastava and Eachempati devote a considerable amount of their paper to the optimization of the algorithms hyperparameter and show how they can significantly increase the quality of predictions. The authors advise using a deep neural network because in their case it was able to correctly classify 92 percent of employees. However, they only covers seven features, including the number of projects assigned and the amount of time invested in them. It can be concluded from this that the analyzed employees work predominantly in a project-oriented manner. This way of working and the selected characteristic also do not match our present use case.

The last study based on real company data examines employee attrition of insurance agents (Valle and Ruz 2015). Yet, their intention contradicts our research goal as firstly, the authors do not consider financial or sociological employee-related factors and make their predictions purely on the basis of performance indicators. Secondly, the results also serve to identify employees who should be terminated due to poor performance, which the researchers justify with the "practical interest to the call center" (Valle and Ruz 2015). These two assumptions severely limit the transfer of the findings to other, less competitive business sectors, especially because other publications have already demonstrated clearly that correlations exist between personal well-being and termination behavior (cf. et al. Fallucchi et al. 2020; Jain and Nayyar 2018; Alduayj and Rajpoot 2018; Aggarwal et al. 2022).

Three other research groups used information from publicly available databases as well as professional social networks (Dahan et al. 2020; Alaskar, Crane and Alduailij 2019; Kisaog 2014). Here, however, quantitative as well as qualitative shortcomings severely limited the explanatory power of the models (Dahan et al. 2020; Kisaog 2014).

The remaining publications based on the IBM data differ in their structure, in some cases considerably. While researchers from real companies in particular pay little attention to data pre-processing steps such as feature engineering, they play a central role in many of the IBM publications (Cf. et al. Alduayj and Rajpoot 2018; Alao and Adeyemo 2013; Gim and Im 2023; Pawar, Saraf and Pradhan 2023; Raza et al. 2022). This lack of focus is also criticized by Najafi-Zangeneh et al., who emphasize that especially the feature selection process must fit the characteristics of the data set (Najafi-Zangeneh et al. 2021). Several of the authors conclude that careless use of features increases model complexity, complicates evaluations, and may even negatively affect validity (Alduayj and Rajpoot 2018; Alao and Adeyemo 2013). Various proposed solutions are given for feature reduction and selection approaches. These include statistical analysis or machine-learning-based approaches (Gim and Im 2023; Gopinath and Subhashini 2020; Gim and Im 2023). In addition to feature engineering, many articles initially analyze the data using visual representations

to strengthen the general understanding of the use case and to detect inconsistencies (Cf. Raza et al. 2022; Gopinath and Subhashini 2020; Najafi-Zangeneh et al. 2021).

Some authors also criticize weaknesses of the IBM data set, which has a strong imbalance between the two employee classes (Soner et al. 2022; Raza et al. 2022). This can lead to biased machine learning models (Gim and Im 2023). They solve this problem by using synthetic data to scale up to the minority class (Raza et al. 2022; Soner et al. 2022). However, they ignore the risks of this synthetic data and only discuss its benefits. According to them, for example, a large training data set has a positive influence on the quality of the results (Raza et al. 2022).

Despite numerous publications on the use of machine learning models for employee attrition prediction, research gaps remain. Most notably, there is a lack of representative contributions from companies supplying real world data. This circumstance makes it difficult to transfer existing findings to companies like Vitesco Technologies. This is mostly due to the fact that company researchers do not provide information about the knowledge base used as well as its structure. In addition, many publications neglect the relevance of a clean data basis and detailed feature engineering. Recommendations about suited machine learning models also vary so widely, that even the authors using IBM data are in disagreement. This shows that there is no universally applicable model but rather that it is essential to implement and compare multiple algorithms to identify the one most suited for the individual use case.

V. MODEL VALIDATION

Before comparing the machine learning models, it must first be clarified how their suitability and the quality of their predictions can be measured. Since no reference projects exist within the HR department of Vitesco Technologies, this article first explores common validation methods and evaluates them in the context of the present use case. The f-score, which is regarded in the literature as a decisive indicator for the quality of a model, serves as the central indicator (cf. et al. Alduayj and Rajpoot 2018; Soner et al. 2022; Kisaog 2014; Bhatta et al. 2022; Raza et al. 2022). However, especially for small data sets with many features, overfitting may occur despite a high f-score (Vabalas et al. 2019). A reliance on key figures alone does not guarantee generalized model predictions. Generalization ensures that trained models can apply learned patterns and relationships to unseen data entries. To assess both generalization and the f-score, a validation process tailored to Vitesco Technologies' human resources management was developed:

Usually, holdout or cross-validation is used to validate machine learning models. Holdout validation splits the data into a training and a validation data set (Jung 2022). The model receives the former to learn the data patterns and can use the latter to verify that it has interpreted the relationships correctly. The holdout validation is well suited for measuring the degree

of generalization, as the holdout data set remains unseen by the model until verification. However, the subdivision greatly reduces the size of the training data (Raschka 2020). Adding to that, it only allows for one validation cycle, because reusing the data can result in the model merely memorize the labels and thus acting in a biased manner (Raschka 2020). This lack of multilevel validation as well as the reduced training base, can lead to large variations in the results which a reliable model evaluation.

Cross-validation, on the other hand, allows for multiple training cycles using the same data set. To do this, it first divides the data set into k subsets, with each one once serving as the validation data set for the remaining training data. In the literature, a typical value for k lies between 5 and 10, since too low values also lead to the problems mentioned before (Raschka 2020). In this way, cross-validation trains k different models, which are then merged. The averaging counteracts strong fluctuations in key figures like the f-score. This is particularly beneficial for small data sets, as more entries remain for model training (Vabalas et al. 2019). In practice, cross-validation is often complemented by stratification, which ensures that each partial data set represents all existing classes equally (Berrar 2018). Models trained using cross-validation usually exhibit less bias compared to holdout validation (Raschka 2020). However, multiple uses of the data can lead to overfitting, especially when class entries are very similar (Varma and Simon 2006).

None of the presented validation approaches are able to simultaneously ensure generalization and prevent overfitting when the data sample size is small. A combination of both approaches, on the other hand, minimizes their drawbacks: A Three-Way-Validation-Method makes more optimal use of the training data while ensuring at the same time measuring the model generalization (Raschka 2020). It is based on an initial (stratified) data split followed by a cross-validation of the training data. The entire validation process is depicted in figure 1.

VI. DATA PRE-PROCESSING

The quality of machine learning predictions also strongly depends on the available information and its quality (Batista and Monard 2002). Accurate predictions require a clean as well as conclusive data set. This severely limits the use of unprocessed system extracts or raw data, which in practice usually has errors, gaps or simply insufficient expressiveness (Keim et al. 2006). Given this context, it is crucial to conduct an initial quality assessment to identify any content deficiencies or gaps before proceeding with the implementation.

A. Visual Data Exploration

Visual data exploration (short VDE) is a graphical analysis process that provides deeper insights into the data structure and allows for initial conclusions about structural discrepancies (Keim 2001). It is particularly suited as an initial tool to reveal existing correlations and provide a general understanding of

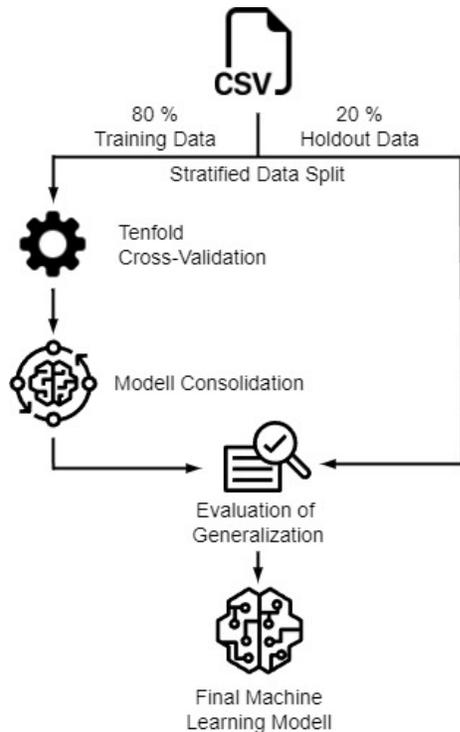


Fig. 1. Three-Way Holdout Validation Process

the use case (Cox 2017). The VDE can uncover irregularities and errors in the database, in particular, contradictory statistical distributions and outliers (Gabler 2008; Datta and Davim 2022). Highlighting and correcting these is important because undetected erroneous entries can critically affect the final machine learning predictions (Cox 2017).

In the case of Vitesco Technologies, it primarily uncovered errors of the manual data preparation such as duplicate identifiers, incorrect categories, and unrealistic entries such as a contract end date in the year 2121. Such irregularities are common as real-world industry data, to some degree, usually contains erroneous or out-of-date data (Keim et al. 2006). Nevertheless, the VDE alone is not sufficient enough to eliminate all structural deficiencies, as especially gaps in the data remain. Such blanks and NULL-values were filled using a KNN imputation. This approach is widely regarded as the most suitable and versatile method for data imputation (Emmanuel et al. 2021; Ismail, Abidin and Maen 2022). While other algorithms, particularly regression methods, may yield slightly more precise results, they often demonstrate only marginal improvements in direct comparison (Makaba and Dogo 2019). Hence, we decided to forgo a comparison of multiple approaches.

B. Feature Engineering

Classification models try to divide data entries into different classes based on their features. For structured, relational data such as in the present use case, each column – except for the class label – represents a feature. For effective model training, these must be both easily machine-processable and

-interpretable (Jung 2022). Feature engineering significantly contributes to increasing the quality of information by reducing the (raw) data to only the most relevant information. For this reason, many authors refer to it as an essential part of data pre-processing (Verdonck et al. 2021; Duboue 2020; Najafi-Zangeneh et al. 2021). The process of feature engineering consists of two parts: First, it requires the examining and transforming of the data into meaningful features to best represent the use case (Fallucchi et al. 2020). Secondly, each of these columns is analyzed to determine its relevance and is removed in case it's not relevant for the classification (Alao and Adeyemo 2013). In summary, feature engineering translates a use case from the real world into a knowledge base that can be interpreted by machines and thus acts as a link between the business view and the data-driven machine learning models (Duboue 2020).

In HR especially, decisions are based on subjective assessments of employees (Fallucchi et al. 2020). This introduces bias that can harm the correct prediction of employee attrition. Here subjective opinions may cause existing patterns to be ignored because they contradict human expectations. For example, Jain and Nayyar uncovered that employees in their use case were more likely to quit the closer they lived to their work location, and Arqawi et al. noted that in their data, there was no correlation between employee performance and resignations (Jain and Nayyar 2018; Arqawi et al. 2022). Both results are contrary to conventional leaving reasons. Feature engineering, on the other hand, ensures an objective approach and combines both perspectives by first representing business assumptions in the form of data and then having a machine learning algorithm identifying existing patterns. A clear data set containing representative features also helps HR employees to better interpret the final results of machine learning models (Duboue 2020). Interpretability plays a significant role as employee attrition can only be counteracted if dissatisfaction factors are identified at an early stage and appropriate measures are derived from them (Fallucchi et al. 2020). For this reason, internal HR experts and their know-how were integrated into our feature engineering process to better map reasons for termination to available system information and to elaborate new features (For questions about the data set and the developed features, please contact Fabian Engl using the contact details provided).

After the feature engineering and creation phase, the relevance of the features must be evaluated (Alao and Adeyemo 2013). In some circumstances, the presence of features actually harms the performance of the models. This is particularly relevant in the case of Vitesco Technologies as several models are going to be implemented and the optimal underlying data bases differ depending on the algorithm (Duboue 2020). This evaluation can be done in three different ways: using filters, wrappers, or embedded in the training phase of the machine learning models (Gim and Im 2023). In the case of Vitesco Technologies, a backwards feature selection approach was used followed by an ordinal data encoding.

VII. BINARY CLASSIFICATION

The machine learning models were implemented using the Python libraries scikit-learn and XGBoost, with random states guaranteeing a constant evaluation, with the latter ensuring deterministic results and thus enabling evidence-based comparison (scikit learn 2023).

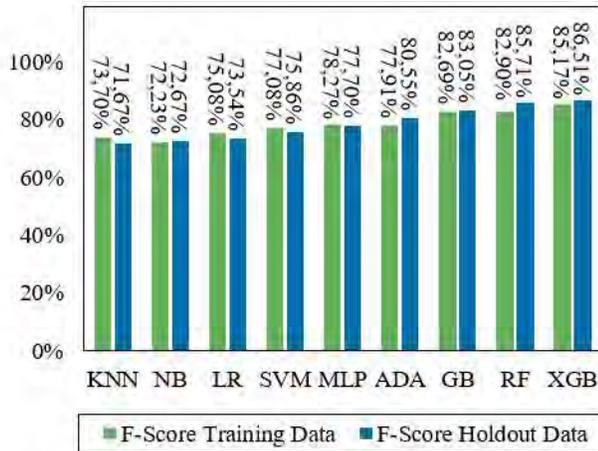


Fig. 2. F-scores of the binary machine learning models after feature engineering

A first comparison using un-optimized model shows that all achieve similar results in both cross- and holdout validation, which indicates that the holdout data represents the training data well. All machine learning models detected patterns in the HR data and were able to classify unseen data sets using the derived patterns. Overall, the tree-based algorithms – the Random Forest (RF), adaptive boosting (ADA) and both gradient boosting methods (GB and XGB) – lead the comparison. All four models achieve f-scores above 80 percent on the holdout data set (see Figure 3). However, the degree of generalization differs.

The K-nearest-neighbors (KNN) algorithm, the Naive Bayes (NB), and logistic regression (LR) record below-average results in direct comparison. Their f-scores mostly lie below 75 percent. In the case of the KNN approach, these scores result from the spatial distribution of the employee data within the feature space. Using a principle component analysis to display the feature space in a two-dimensional scatter plot, large-scale overlaps of both employee classes become apparent (see Figure 3). The entries are so close to each other that an evaluation based on neighbors does not allow for meaningful conclusions. Since the KNN algorithm, bases its predictions solely on this spatial distance of the features (Ertel 2016), it is unsuitable for the current use case.

The Naive Bayes model also underperforms in a direct comparison, which was to be expected as the model presupposes the independence of all features (Ertel 2016). Especially in the context of employee data, this is rarely the case. For example, in Vitesco Technologies' HR systems, certain characteristics such as compensation are calculated from a combination of other information such as the number of service years. In

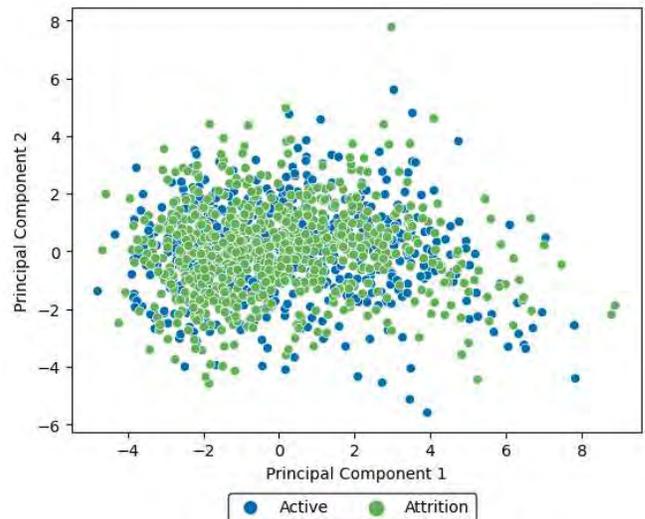


Fig. 3. Two-dimensional visualization of the KNN feature space using a principle component analysis

theory, the model is generally not suitable as a tool for employee turnover forecasting. Nevertheless, some researchers achieved accurate forecasts using Naive Bayes (Valle and Ruz 2015). In our use case, the underwhelming results and doubts about the model's suitability are enough to rule it out as unsuitable.

The logistic regression also only achieved f-scores of 75.08 and 73.54 percent. This is because the data set has many binary-categorical columns, which make it difficult to calculate clear boundaries within features (Kemalbay and Korkmazoğlu 2014). Similar to the KNN algorithm, the logistic regression seems incompatible with the employee data studied.

The support vector machine (SVM) achieves f-scores of 77.08 and 75.86 percent but shows a rather high difference between training and holdout data. However, algorithms – especially the SVM – have hyperparameters that allow for optimization of the predictions (see chapter Hyperparameter Optimization). In case of the SVM these may allow the construction of more representative hyper-planes possibly increasing the f-scores further.

A similar trend can be observed for the neural network (MLP), which comes close to matching the performance of tree-based models achieving f-scores of 78.27 and 77.70 percent, respectively. With a difference of 0.57 percent, it also records a high degree of generalization. However, there is still a significant performance gap compared to the leading machine learning model.

Since both AdaBoost and Random Forest achieve significantly higher f-scores in the holdout data set, there is reason to believe that the holdout data contains more clearly distinguishable employees and therefore slightly positively biases the results. The extreme gradient boosting procedure, which also achieved better holdout values, supports this hypothesis. Nevertheless, the comparison clearly shows that without hyperparameter tuning the tree-based methods achieve the best

results in the case of Vitesco Technologies and stand out in particular due to their high precision (For a more detailed list of all key figures as well as questions regarding the implementation, please contact Fabian Engl using the contact details provided).

A. Hyperparameter Optimization

Hyperparameter optimization allows to further improve the performance of machine learning methods by tuning the model to the available data and its structure (Duboue 2020). Hyperparameters are adjustment screws that make minor changes in how a model operates, thus enhancing performance (Bhatta et al. 2022). They are set before model training and do not dynamically adjust during the validation process (Nokeri 2021). The number of parameters depends on the algorithm.

The two most common methods for determining hyperparameters include a grid search and a random search (Agrawal 2021). The grid search uses a list of values for each parameter, checks all possible combinations and returns the best combination (Bhatta et al. 2022). While this approach guarantees optimal parameter matching within the given values, it requires a lot of computational resources to do so. If the data set contains many features or if a model has many parameters, this process can take several days or even weeks (Ertel 2016). The runtime of the random search, on the other hand, is unaffected by both and always remains constant (Agrawal 2021). Instead of fixed parameter values, it is given plausible value ranges for each parameter and a fixed number of iterations. It chooses random hyperparameters within these ranges and repeats this process as many times as specified. This way, it usually finds a near best set saving a huge amount of time (Agrawal 2021).

In our case limited computational capacity restricted a large-scale grid search, so both methods were combined into a two-step search: A random search first narrowed down the value ranges of the parameters as best as possible, ensuring that the resulting combination is as close to the actual optimum as possible. Afterwards, a grid search examined the immediately adjacent range to fine-tune the parameters. This approach allows for a significantly larger initial search space to be covered while reducing the drawbacks of both search methodologies. The optimized results show that some methods benefit more from hyperparameter optimization than others (see Figure 4). While AdaBoost and the normal gradient boosting model achieved a significant performance increase, most of the other models only saw minor improvements. For the neural network, the random forest and the extreme gradient boosting method, it even lead to slightly lower holdout results. However, this development does not necessarily mean that the overall performance worsened, as the holdout data primarily serves to evaluate the generalization of the machine learning model in our validation process. Consequently, a qualitative machine learning model must both achieve a high f-score in the cross-validation as well as comparable results with the holdout data. Considering this, all models recorded better overall results, which are reflected either in an improvement of the holdout f-scores or in the degree of generalization.

The Random Forest was able to reduce the difference between the cross-validation and the holdout evaluation by 1.57 percent while both scores themselves showed only marginal improvements. The adaptive and the normal gradient boosting models showed the biggest improvements with an average improvement of 5.77 and 2.83 percent respectively. In the case of adaptive boosting a modification of tree-depth resulted in performance on par with the other leading tree-based machine learning models. This result is based on a maximum depth of three. This modification of the depth changes the basic approach of the algorithm, as it is originally based on weak learners – so-called decision stumps (Freund and Schapire 1997). Nevertheless, this adjustment leads to the highest holdout f-score and the most generalized prediction among all models. The neural network and the SVM, on the other hand, fail to catch up to the tree-based models even after hyperparameter optimization and are therefore neglected in the following chapters.

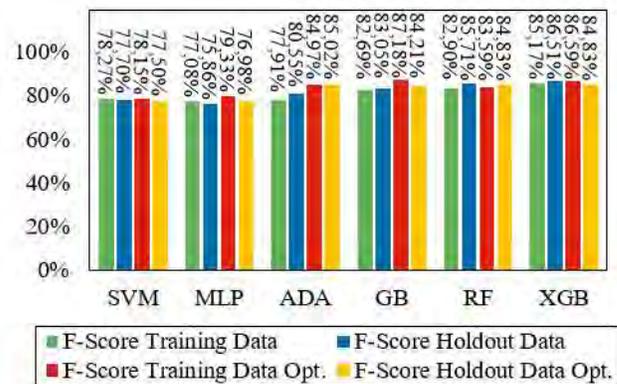


Fig. 4. F-scores of the machine learning models after the hyperparameter optimization

B. Conclusion on binary employee attrition prediction

After hyperparameter optimization, the tree-based machine learning models stand out as the most accurate models. It can be clearly stated that the KNN method, the Naive Bayes, the logistic regression, the neural network and the support vector machine are either incompatible or under-performing in our use case. While none of the other algorithms achieved reliable f-scores above 80 percent using our data, the results of the tree-based models stagnated at around 85 percent. Yet, due to the neglectable differences between them, no clear recommendation can be given as all seem equally suited to predict employee attrition prediction at Vitesco Technologies.

Although in our case the tree-based methods are generally suitable for employee attrition prediction, they differ in their data pre-processing and model optimizing requirements. Especially in the context of HR the choice of a machine learning model depends on more factors than just the f-score or degree of generalization. Internal resources, IT know-how and available technical resources contribute significantly to the long-term success of data-driven employee attrition detection

strategies. In this study, boosting-based models required less extensive feature engineering, which could prove beneficial if resources are limited. In particular, pre-built frameworks such as XGBoost achieve accurate predictions even with little hyperparameter optimization.

VIII. MULTICLASS CLASSIFICATION

The first binary employee attrition classification showed that machine learning models were able to correctly classify employees at risk of leaving with an 85 percent certainty. Yet this simple differentiation does not give any insights into the underlying motives for leaving which need to be revealed and understood in order to introduce HR guidelines to counteract employee attrition. For this reason this chapter wants to answer whether data patterns exist that also allow for a multi-class classification of the leaving reasons. This requires a modification to the validation process as the f-score in its used form is only suitable for binary classification problems (Hossing and Sulaiman 2015). This chapter is uses the macro-f-score which calculates individual f-scores at class level and then calculates the average (Grandini, Bagli and Visani 2020).

The available leaving reasons are based on voluntary information provided by employees. The extend to what they reflect the reality can not be confirmed. Vitesco Technologies distinguishes between the following leaving reasons: a higher salary (SAL), better career opportunities (CAR), further education (FUE), a relocation (REL), a lack of work-life balance (WLB) and problems with the company, its culture or the leadership (CCL). The class of active employees (ACT) remains unchanged. This detailed split of former employees leads to a strong class imbalance problem as these six new classes were previously combined into one. Such strong imbalances can have a negative impact on machine learning results (Nguyen, Cooper and Kamei 2011).

A. Imbalanced-Class-Problem

An imbalanced-class problem exists when one or more classes contain significantly more entries than the rest (Tharwat 2020). In the case of Vitesco Technologies, this affects the class of active employees. This condition can distort machine learning metrics and hurt predictions (Najafi-Zangeneh et al. 2021). The imbalance can be counteracted by under- or oversampling the data set. Undersampling involves reducing the size of the majority class(es) to the number of entries in the smallest class, while oversampling instead increases the size of all the minority class(es) using synthetically generated data entries (Chawla et al. 2002).

B. Initial Evaluation of patterns

A random oversampling algorithm was used for the initial validation to analyze whether leaving-reason-specific correlations exist within the data set. Random oversampling uses existing data points to generates new entries with similar features (IBM imbalanced-learn 2023).

This initial analysis clearly shows that all models fail to differentiate all classes correctly. The f-scores regarding the

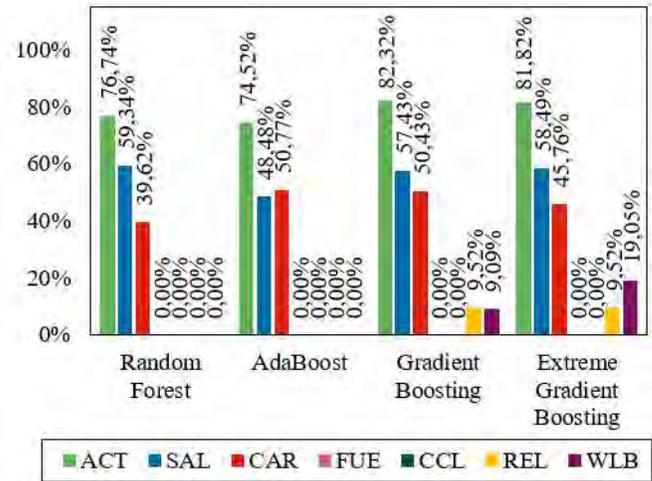


Fig. 5. Holdout f-scores of the machine learning models after applying random oversampling

class of the active employees stays mostly unchanged. While resignations due to better salary or a career advancement can still partially be identified, the remaining four reasons were not detect at all or only very unreliably. Resignations due to relocation or as a result of a lacking work-life balance were detected by the gradient boosting models, but only so inconsistently that their predictions provide little to no value for HR. With maximum f-scores of 9.52 and 19.09 percent, respectively, the results are basically on par with or just slightly better than random guessing.

The fact that only the first two leaving reasons are recognized might be due to the purely company-sided representation of the employees in the data set. All of the last four classes represent leaving reasons that are driven by intrinsic motivation. Factors such as the concordance of personal beliefs with Vitesco Technologie’s company values or the experienced leadership style are based on individual perception and can vary between employees. The company-sided data set does not contain any features that represent individual well being and therefore, does not provide sufficient information to clearly identify leaving reasons affected by it. Adding to that the last four classes make up for only roughly 13 percent of all resignations. Such small classes could hurt performance even further as too few entries exist for the machine learning models, to derive clear patterns. Since both facts hinder a clear differentiation of all classes, the following multi-class classification focuses on first three classes only: active employees and former employees who left due to a higher salary or better career advancement opportunities.

C. Comparison of under- and oversampling techniques

Based on the selected classes, several under- and oversampling methods were compared. In direct comparison, the undersampling algorithms achieve significantly lower macro-f-scores and show higher fluctuations between the machine learning

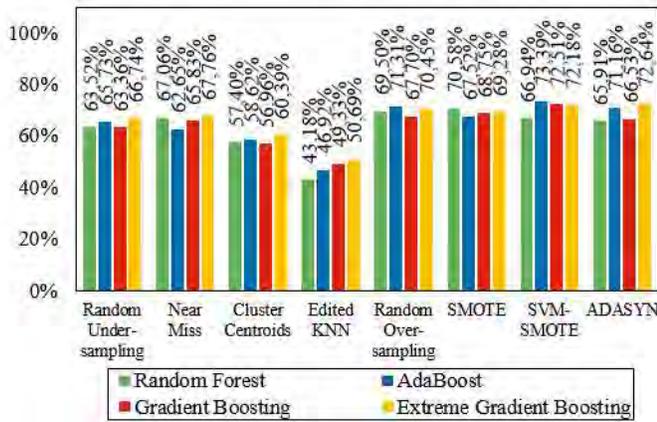


Fig. 6. Holdout f-scores of the machine learning models after applying different under- and oversampling techniques

models (see figure 6). Of all oversampling algorithms, the SVM-SMOTE achieves the most stable macro-f-scores across all models. For this reason it is used to counteract the class imbalance for the following comparison.

D. Evaluation of the multi-class classification

Just as in binary classification, all four machine learning models underwent both a feature engineering phase and a hyperparameter optimization. Similar to the previous chapter, the tree-based machine learning models exhibit only slight variations, with adaptive boosting achieving the highest level of generalization. The extreme gradient boosting model in particular achieves the best f-scores across all classes achieving 86.29, 74.73 and 66.67 percent. However, all models show a noticeable decline in generalization ranging from 6.21 to 14.22 percent. This probably results from the use of synthetic data.

Looking only at the active employees the class f-scores match the the binary classification results. This means they were identified with a certainty of roughly 85 to 86 percent again. This was to be expected as the class was neither modified nor affected by the excluded employee entries. Employees who resigned due to a higher salary or better career opportunities could only be detected in roughly 71 to 75 and 65 to 67 percent respectively. Although these f-scores may appear relatively low in comparison, it's crucial to consider them within the broader context of employee attrition prediction. Given that active employees are accurately classified in over 85 percent of cases, this indicates that the two primary classes of active and former employees are quite distinct. Considering this, the machine learning models can generally differentiate employees who have left Vitesco Technologies and even correctly predict the corresponding leaving reasons in more than two thirds of cases.

E. Conclusion on multi-class employee attrition prediction

In summary, machine learning models are not only capable of identifying employees at risk of leaving but also offer some insight into their leaving motives. In the case of Vitesco

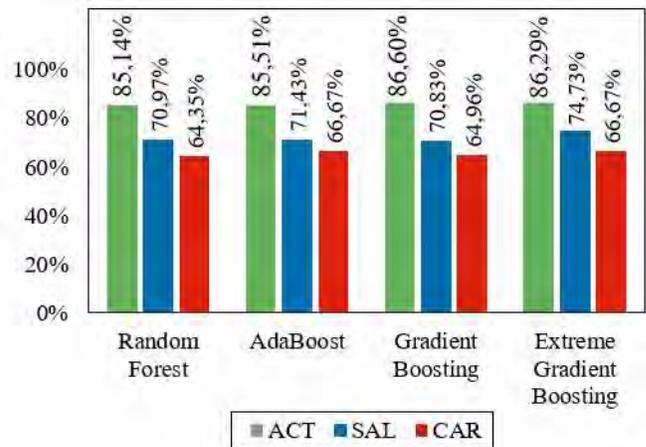


Fig. 7. Holdout f-scores of the machine learning models on class level

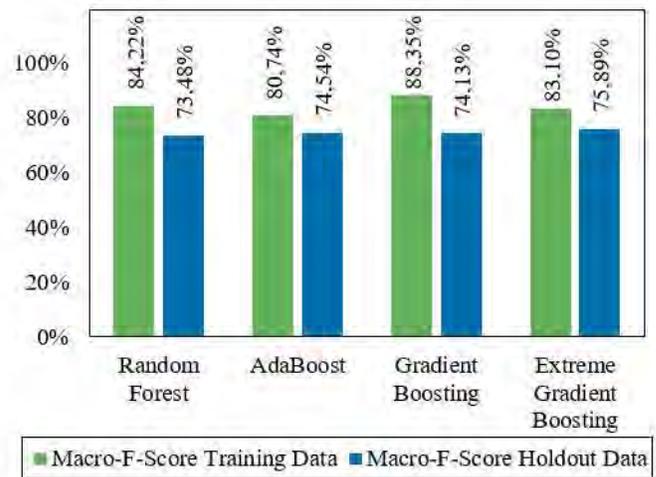


Fig. 8. Macro-f-scores of the machine learning models

Technologies, the models consistently identified employees who left due to higher salary or better career opportunities. However, despite these reliable (partial) findings, all machine learning models struggle to predict all leaving reasons. One potential explanation lies in the choice of features, which may not adequately capture the employee's perspective. Many of the collected features were constructed based on a company's perspective on employee attrition and do not encompass subjective factors like the well-being of employees. Expanding the data set to incorporate additional attributes reflecting employee well-being could solve this problem. A possible solution could be asking employees for individual feedback and experiences in their exit survey.

IX. FINAL CONCLUSION

In summary, the results of this paper show that machine learning models can detect distinct patterns in Vitesco Technologies' workforce data and accurately identify employees at risk of leaving. A combination of cross-validation and holdout validation is used to assess the quality of the evidence. The

(macro-)f-score is the primary metric of comparison because it most reliably reflects the predictive and differentiating capabilities of the algorithms.

Due to structural errors in the system extraction, extensive data pre-processing and preparation was necessary prior to implementation of the machine learning algorithms. Inconsistencies were first detected and corrected using a visual data exploration technique. A KNN-imputation subsequently completed structural gaps. Adding to that, the different machine learning models required varying degrees of feature engineering and hyperparameter optimization.

The comparison clearly showed that tree-based machine learning approaches are best suited for the classification of Vitesco Technologies' employee data. The random forest, the AdaBoost algorithm and both gradient boosting models achieved f-scores of around 85 percent. Therefore, all are equally suitable for employee attrition prediction. However, the extend of necessary data pre-processing steps varied significantly. In this study boosting-based approaches required less detailed feature engineering. The findings showed that the XGBoost framework achieved both a high f-score and a pronounced degree of generalization even without intensive feature engineering or hyperparameter optimizations.

In addition to identifying employees at risk of leaving, the multi-class classification approach also allowed partial conclusions to be drawn about employees motives. However, characteristics that depicted the individual well-being of the employees were missing here. As a result, it was not possible to adequately represent all reasons for leaving. Only departures due to a higher salary or better promotion prospects were reflected in the data. However, the machine learning models were able to identify both with certainty of around 75 and 67 percent respectively.

The findings of the machine learning algorithms allow for an evidence-based alignment of Vitesco Technologies' Chinese HR policy with the employee needs. However, before the findings can be transferred to other countries and locations of Vitesco Technologies, an evaluation of the cultural influences on the results of the machine learning models is required. Furthermore, there are additional challenges to overcome: both the limited computing resources and the small data set restricted the evaluations in many places. Therefore, before integrating the proposed algorithms into the HR landscape the research results should first be validated during an initial test phase verifying the findings using real employee data and adjusting the models in case of deviations. In order to reduce manual data preparation in the future and ensure qualitative predictions, an automated data pipeline needs to be implemented.

X. SUMMARY

Using Vitesco Technologies as an example, this article examines whether machine learning models are suitable for detecting dissatisfaction on the part of employees at an early stage and thus being able to preventively counteract terminations. Tree-based methods were found to be the most

suitable for classifying employee data. These include a random forest, the AdaBoost algorithm and two different gradient boosting models. Finally, it was shown that a prediction of the associated reasons for termination is also possible under certain conditions.

REFERENCES

- Aggarwal, S. et al. (2022). "Employee Attrition Prediction Using Machine Learning Comparative Study". In: *Intelligent Manufacturing and Energy Sustainability*. Singapore: Springer, pp. 453–466.
- Agrawal, T. (2021). "Hyperparameter Optimization Using Scikit-Learn". In: *Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient*. Berkeley, USA: Apress, pp. 31–51.
- Alao, D. and A. B. Adeyemo (Mar. 2013). "Analyzing Employee Attrition using Decision Tree Algorithms". In: *Computing, Information Systems, Development Informatics and Allied Research Journal* 4.1, pp. 17–28.
- Alaskar, L., M. Crane, and M. Alduailij (2019). "Employee Turnover Prediction Using Machine Learning". In: *Advances in Data Science, Cyber Security and IT Applications*. Cham, Schweiz: Springer International Publishing, pp. 301–316.
- Alduayj, S. S. and K. Rajpoot (Nov. 2018). "Predicting Employee Attrition using Machine Learning". In: *2018 International Conference on Innovations in Information Technology*, pp. 93–98.
- Arqawi, S. M. et al. (Nov. 2022). "Predicting Employee Attrition and Performance using Deep Learning". In: *Journal of Theoretical and Applied Information Technology* 100.21, pp. 6526–6536.
- Batista, G. and M. C. Monard (Jan. 2002). "A Study of K-Nearest Neighbour as an Imputation Method." In: 30, pp. 2–11.
- Berrar, D. (Jan. 2018). "Cross-Validation". In: *Encyclopedia of Bioinformatics and Computational Biology* 1, pp. 542–545.
- Bhatta, S. et al. (2022). "Machine Learning Approach to Predicting Attrition Among Employees at Work". In: *Artificial Intelligence Trends in Systems*. Cham, Schweiz: Springer International Publishing, pp. 285–294.
- Bund-Verlag (Sept. 2018). *Innere Kündigung - Fünf Millionen Arbeitnehmer haben innerlich gekündigt*. URL: <https://www.bund-verlag.de/aktuelles~F%C3%BCnf-Millionen-Arbeitnehmer-haben-innerlich-gek%C3%BCndigt~.html> (visited on 07/24/2023).
- Chawla, N. V. et al. (June 2002). "SMOTE: Synthetic Minority Over-sampling Technique". In: *Journal of Artificial Intelligence Research* 16, pp. 321–357.
- Chugani, J. (July 2023). *How HR Professionals Can Co-exist With AI Tools*. URL: <https://www.forbes.com/sites/forbeshumanresourcescouncil/2023/07/12/how-hr-professionals-can-coexist-with-ai-tools/> (visited on 07/24/2023).
- Cox, V. (2017). *Translating Statistics to Make Decisions*. 1st ed. Salisbury, UK: Springer Link, pp. 41–100.

- Dahan, S. et al. (May 2020). “Predicting Employment Notice Period with Machine Learning: Promises and Limitations”. In: *McGill Law Journal* 65, pp. 711–753.
- Datta, S. and J. P. Davim (2022). *Machine Learning in Industry*. 1st ed. Cham, Schweiz: Springer, pp. 5–6.
- Duboue, P. (2020). *The Art of Feature Engineering: Essentials for Machine Learning*. 1st ed. Cambridge, UK: Cambridge University Press, pp. 21–43.
- Emmanuel, T. et al. (Oct. 2021). “A survey on missing data in machine learning”. In: *Journal of Big Data* 8, pp. 1–37.
- Ertel, W. (2016). *Grundkurs Künstliche Intelligenz*. 4th ed. Wiesbaden: Springer Fachmedien, pp. 207–209, 237–239, 298, 304.
- Fallucchi, F. et al. (Dec. 2020). “Predicting Employee Attrition Using Machine Learning Techniques”. In: *Computers* 9.4, pp. 86, 1–17.
- Freund, Y. and R. E. Schapire (Aug. 1997). “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”. In: *Journal of Computer and System Sciences* 55, pp. 119–139.
- Gabler, T. (2008). “Auffälligkeiten in Häufigkeitsverteilungen”. In: *Wirtschaftsstatistik im Bachelor - Grundlagen und Datenanalyse*. 1st ed. Wiesbaden: Springer Link, pp. 117–148.
- Gim, S. and E. T. Im (2023). “A Study on Predicting Employee Attrition Using Machine Learning”. In: *Big Data, Cloud Computing, and Data Science Engineering*. Cham, Schweiz: Springer International Publishing, pp. 55–69.
- Gopinath, R. and M. Subhashini (Dec. 2020). “Employee Attrition Prediction in Industry using Machine Learning Techniques”. In: *International Journal of Advanced Research in Engineering & Technology* 11, pp. 3329–3341.
- Grandini, M., E. Bagli, and G. Visani (Aug. 13, 2020). *Metrics for Multi-Class Classification: an Overview*. arXiv: 2008.05756[cs,stat].
- Guerranti, F. and G. M. Dimitri (Jan. 2023). “A Comparison of Machine Learning Approaches for Predicting Employee Attrition”. In: *Applied Sciences* 13.1, pp. 267, 1–8.
- Hossing, M. and M. N. Sulaiman (Mar. 2015). “A Review on Evaluation Metrics for Data Classification Evaluations”. In: *International Journal of Data Mining & Knowledge Management Process* 5.2, pp. 1–11.
- IBM imbalanced-learn (2023). *Naive random over-sampling*. URL: https://imbalanced-learn.org/stable/over_sampling.html (visited on 07/24/2023).
- Ismail, A. R., N. Z. Abidin, and M. K. Maen (Feb. 2022). “Systematic Review on Missing Data Imputation Techniques with Machine Learning Algorithms for Healthcare”. In: *Journal of Robotics and Control* 3.2, pp. 143–152.
- Jain, R. and A. Nayyar (Nov. 2018). “Predicting Employee Attrition using XGBoost Machine Learning Approach”. In: *2018 International Conference on System Modeling & Advancement in Research Trends*, pp. 113–120.
- Jung, A. (2022). *Machine Learning: The Basics*. 1st ed. Singapore: Springer Nature, pp. 3–22, 117–142, 172.
- Keim, D. A. (Aug. 2001). “Visual exploration of large data sets”. In: *Communications of the ACM* 44.8, pp. 38–44.
- Keim, D. A. et al. (July 2006). “Challenges in Visual Data Analysis”. In: *Tenth International Conference on Information Visualisation*, pp. 9–16.
- Kelly, J. (Feb. 2023). *Workers In China Have Their Own Version Of Quiet Quitting And Acting Your Wage: ‘Huminerals’ Are Extracted, Exploited And Disposed Of*. URL: <https://www.forbes.com/sites/jackkelly/2023/02/23/china-workers-have-their-version-of-quiet-quitting-and-acting-your-wage-huminerals-are-extracted-exploited-and-disposed-of/> (visited on 07/24/2023).
- Kemalbay, G. and Ö. B. Korkmazoğlu (Jan. 2014). “Categorical Principal Component Logistic Regression: A Case Study for Housing Loan Approval”. In: *Procedia - Social and Behavioral Sciences* 109, pp. 730–736.
- Kisaog, Z. Ö. (Sept. 2014). “Employee Turnover Prediction using Machine Learning based Methods”. Masterarbeit. Ankara, Türkei: Middle East Technical University. URL: <https://etd.lib.metu.edu.tr/upload/12617686/index.pdf>.
- Makaba, T. and E. Dogo (Nov. 2019). “A Comparison of Strategies for Missing Values in Data on Machine Learning Classification Algorithms”. In: *2019 International Multidisciplinary Information Technology and Engineering Conference*, pp. 1–7.
- Najafi-Zangeneh, S. et al. (Jan. 2021). “An Improved Machine Learning-Based Employees Attrition Prediction Framework with Emphasis on Feature Selection”. In: *Mathematics* 9.11, pp. 1226, 1–14.
- Navarra, K. (Apr. 2022). *The Real Costs of Recruitment*. URL: <https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/the-real-costs-of-recruitment.aspx> (visited on 07/24/2023).
- Nguyen, H. M., E. W. Cooper, and K. Kamei (Apr. 2011). “Borderline over-sampling for imbalanced data classification”. In: *International Journal of Knowledge Engineering and Soft Data Paradigms* 3.1, pp. 24–29.
- Nokeri, T. C. (2021). *Data Science Revealed: With Feature Engineering, Data Visualization, Pipeline Development, and Hyperparameter Tuning*. 1st ed. Berkeley, USA: Apress, p. 22.
- Pawar, N., N. Saraf, and T. Pradhan (Jan. 2023). “Role of Analytics in HR-Attrition for effective Decision Making”. In: *International Research Journal of Modernization in Engineering Technology and Science*, pp. 185–188.
- Raschka, S. (Nov. 10, 2020). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. arXiv: 1811.12808[cs,stat]. URL: <http://arxiv.org/abs/1811.12808>.
- Raza, A. et al. (Jan. 2022). “Predicting Employee Attrition Using Machine Learning Approaches”. In: *Applied Sciences* 12.13, pp. 6424, 1–17.
- scikit learn (2023). *Glossary of Common Terms and API Elements*. URL: https://scikit-learn.org/stable/glossary.html#term-random_state (visited on 07/24/2023).
- Sikaroudi, A., R. Ghousi, and A. Sikaroudi (Oct. 2015). “A data mining approach to employee turnover prediction (case

- study: Arak automotive parts manufacturing)”. In: *Journal of Industrial and Systems Engineering* 8.4, pp. 106–121.
- Soner, S. et al. (Dec. 20, 2022). *Predictive Deep Learning approach of employee attrition for imbalance datasets using SVM SMOTE algorithm with Bias Initializer*. preprint. In *Review*, pp. 1–19.
- Srivastava, P. and P. Eachempati (Nov. 2021). “Intelligent Employee Retention System for Attrition Rate Analysis and Churn Prediction: An Ensemble Machine Learning and Multi-Criteria Decision-Making Approach”. In: *Journal of Global Information Management* 29.6, pp. 1–29.
- Tharwat, A. (Jan. 2020). “Classification assessment methods”. In: *Applied Computing and Informatics* 17.1. Publisher: Emerald Publishing Limited, pp. 168–192.
- Vabalas, A. et al. (Nov. 2019). “Machine learning algorithm validation with a limited sample size”. In: *PLOS ONE* 14.11, pp. 1–20.
- Valle, M. A. and G. A. Ruz (Oct. 2015). “Turnover Prediction in a Call Center: Behavioral Evidence of Loss Aversion using Random Forest and Naïve Bayes Algorithms”. In: *Applied Artificial Intelligence* 29, pp. 923–942.
- Varma, S. and R. Simon (Feb. 2006). “Bias in error estimation when using cross-validation for model selection”. In: *BMC Bioinformatics* 7, pp. 91, 1–8.
- Verdonck, T. et al. (Aug. 2021). “Special issue on feature engineering editorial”. In: *Machine Learning*, pp. 1–12.
- Wallace, L. (Mai 2023). *Five Hidden Costs Of Employee Attrition*. URL: <https://www.forbes.com/sites/forbeseq/2023/03/21/five-hidden-costs-of-employee-attrition/> (visited on 07/24/2023).

ANWENDUNG VON OBJECT CENTRIC PROCESS MINING AM BEISPIEL EINES USE CASES

Florian Heck
HS Pforzheim
Tiefenbronnerstr. 65,
75175 Pforzheim
heckflor@hs-pforzheim.de

Frank Morelli
HS Pforzheim
Tiefenbronnerstr. 65,
75175 Pforzheim
frank.morelli@hs-pforzheim.de

Janna Meyer
MEHRWERK GmbH
Karlsruher Straße 88,
76139 Karlsruhe
janna.meyer@mehrwerk.net

SCHLÜSSELWÖRTER

Process Mining, Object Centric Process Mining, Process Analyzer, Divergenz, Konvergenz

ABSTRACT

Die Anwendung von Object Centric Process Mining (OCPM) erweist sich als hilfreich, Geschäftsprozesse besser zu verstehen und sie nachhaltig zu optimieren. Dabei löst OCPM einige Probleme des traditionellen Process-Mining-Einsatzes und kann als digitaler Prozess-Zwilling die Realität besser abbilden. Der vorliegende Artikel erläutert zum einen die Problematiken von Single-Event Logs im Sinne von Divergenz und Konvergenz. Darüber hinaus werden die theoretischen Grundlagen zu den OCPM-Event-Logs vorgestellt und im weiteren Verlauf in einer praktischen Anwendung veranschaulicht. Die praktische Anwendung der OCPM-Lösung erfolgt mit der Process-Mining-Software der Firma MEHRWERK.

EINLEITUNG

Das Geschäftsprozessmanagement ist für die Wettbewerbsfähigkeit und für den Erfolg eines Unternehmens von entscheidender Bedeutung (Laue et al. 2021). Aufgrund der fortschreitenden Digitalisierung verfügen die meisten Unternehmen zusätzlich über umfangreiche Geschäftsprozessdaten, die in verschiedenen Informationssystemen gespeichert sind. Die manuelle Analyse dieser Daten ist oft zeitaufwändig, fehleranfällig und nicht immer vollständig. Um die Analyse und die damit einhergehende Geschäftsprozessoptimierung zu erleichtern, können Process-Mining-Techniken eingesetzt werden, die automatisiert und intersubjektiv überprüfbar Daten aus verschiedenen Quellen kombinieren und visualisieren. Object Centric Process Mining (OCPM) stellt in diesem Zusammenhang eine Erweiterung der konventionellen Process-Mining-Techniken dar, um noch bessere und genauere Analyseergebnisse erzielen zu können. Dies beinhaltet beispielsweise die Koordination mehrerer Geschäftsprozesse.

Die Motivation dieser Arbeit beruht auf dem Potenzial von OCPM und konventionellem Process Mining, Geschäftsprozesse effektiver und effizienter zu gestalten. Wenn Unternehmen besser verstehen, wie ihre Prozesse funktionieren, können sie diese gezielt optimieren und so Kosten senken, Durchlaufzeiten verringern, die Qualität verbessern und die Kundenzufriedenheit erhöhen. Process Mining bietet daher ein großes Chancenpotenzial, die Wettbewerbsfähigkeit von Unternehmen zu steigern.

Das Ziel dieser Arbeit ist es, den aktuellen Stand des Wissens zum Thema OCPM zu untersuchen und anzuwenden. Dazu werden zuerst die theoretischen Grundlagen von konventionellem sowie von objekt-zentriertem Process Mining dargelegt. Bislang gibt es kaum Veröffentlichungen über praxisnahe Anwendung von Object Centric Process Mining. Entsprechend ist es ein weiteres Ziel dieses Artikels, den Nutzen von OCPM anhand eines generischen Fallbeispiels darzustellen. Hierzu wird ein Purchase-to-Pay (P2P)-Prozess ausgewählt und anschließend analysiert, um zu zeigen, wie sich OCPM für die Geschäftsprozessoptimierung einsetzen lässt. Der Fokus liegt hierbei auf der eigentlichen Analyse und nicht auf der zugehörigen Datenextraktion und Datenaufbereitung. Darüber hinaus erfolgt ein Vergleich von konventionellem Process Mining mit OCPM und eine Auswertung, welche Vor- und Nachteile OCPM im Anwendungsfall mit sich bringt. Daran schließt sich eine Generalisierung der gewonnenen Erfahrungen an. Die wichtigsten Erkenntnisse werden in einem Fazit zusammengefasst und ein Ausblick für OCPM gegeben.

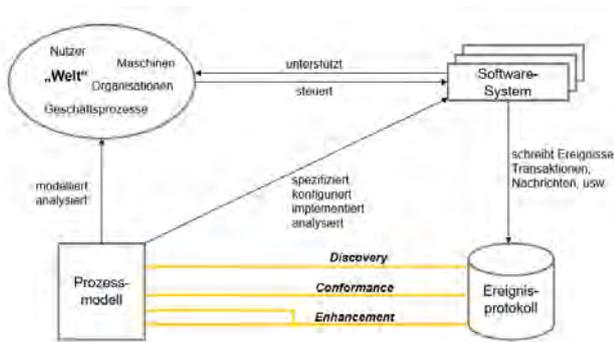
Process Mining

Bei Process Mining handelt es sich um eine datenbasierte Methode zur Analyse und Verbesserung von Geschäftsprozessen. Es gilt als junges interdisziplinäres Forschungsgebiet, welches einerseits maschinelles Lernen und Data Mining, sowie Prozessmodellierung und Prozessanalyse verbindet (van der Aalst 2011). Laue et al. definieren Process Mining als „systematische Analyse von Prozessen auf Basis von Ereignisprotokollen“ (Laue et al. 2021). Dabei wird ein Geschäftsprozess immer im Ganzen von seinem Startpunkt bis zum Endpunkt analysiert, inklusive aller zugehörigen Varianten.

Aufgrund der Unterstützung von Informationssystemen bei der Prozessausführung lassen sich zunehmend mehr Daten über Ereignisse, Transaktionen usw. sammeln. Diese kann man in ein Event-Log (dt. Ereignisprotokoll) überführen (vgl. Abb.1). Ereignisprotokolle sind der Ausgangspunkt für die Anwendung von Process Mining. Nach van der Aalst gibt es drei Hauptanwendungen von Process Mining: Discovery (dt. Erkennung), Conformance (dt. Übereinstimmung) und Enhancement (dt. Verbesserung) (van der Aalst 2012). Die Discovery-Funktion wird eingesetzt, um ein Prozessmodell basierend auf dem Event-Log durch einen Process Mining Algorithmus automatisch zu erstellen und bildet damit die Grundlage für weitere Aktivitäten. Dieses Ist-Prozessmodell beschreibt die tatsächliche „Welt“, durch die in den Informationssystemen Daten generiert werden.

Die automatisch abgeleiteten Prozessmodelle lassen sich dazu verwenden, die Leistung von Prozessen zu messen, Prozessprobleme zu identifizieren und Optimierungspotentiale aufzudecken. Auf Grundlage dieser Erkenntnisse können dann datenbasiert situationsgerechte Entscheidungen getroffen werden, um Verbesserungen in den Abläufen zu implementieren.

Abbildung 1: Übersicht über Process Mining und den damit verbundenen Möglichkeiten



(Quelle: Eigene Darstellung auf Basis von (van der Aalst 2011))

Abb.1 fasst die zuvor beschriebenen Zusammenhänge zwischen der Realität, den Informationssystemen, dem Ereignisprotokoll und dem Prozessmodell zusammen. Die gelben Pfeile stellen zudem die Wechselbeziehungen zwischen den drei Hauptanwendungen und dem Prozessmodell sowie dem Ereignisprotokoll dar.

Object Centric Process Mining

Das Ziel von OCPM ist es, Prozesse in ihrer Gesamtheit zu analysieren und somit die Grenzen des konventionellen Process Minings zu überwinden. Dafür muss in der neuen Domäne von Process Mining auf einen eindeutigen Case Key verzichtet werden: Zugehörige Geschäftsprozesse durchlaufen in der Praxis viele Stationen (z.B.

Abteilungen), die an unterschiedlichen Objekten arbeiten: Beispielsweise erhält die Finanzabteilung Rechnungen, der Einkauf bestellt Materialien und die Versandabteilung versendet Pakete. Da diese Bereiche keinen gemeinsamen Case Identifier haben, entstehen im traditionellen Process Mining mehrere Event-Logs für die Abläufe, bzw. ein Event-Log für jedes Objekt. Die reale Prozesswelt lässt sich damit jedoch nicht abbilden, da diese auf mehrere Objekte referenziert. OCPM vereinfacht somit die Prozessanalyse und erlaubt einen tieferen Einblick in die Prozesslandschaft (Celonis 2023).

Zudem kann bei OCPM eine Aktivität auf mehr als exakt einen Fall verweisen. Es wird angenommen, „dass es mehrere Case Notions (sogenannte Objekttypen) gibt und dass ein Fall sich auf eine beliebige Anzahl von Objekten beziehen kann, die verschiedenen Objekttypen entsprechen“ (van der Aalst 2019). Für diese neue Art des Process Minings wird ein anderes Datenmodell für Event Logs benötigt. Der Vorgang des sog. Glättens transformiert ein objektzentriertes Event-Log in ein konventionelles Event-Log, indem ein einziger Objekttyp ausgewählt wird. Dies erfordert einen adäquaten Umgang mit auftretenden Problemen wie Defizit, Konvergenz und Divergenz und welche im folgenden Kapitel erläutert werden (Berti und van der Aalst 2022)

Lösungsansätze durch Object Centric Process Mining

Abb. 2 zeigt einen Beispielbestellprozess, anhand dessen sich die Probleme erläutern lassen, mit vier Objekttypen (Bestellung, Artikel, Paket und Route) und den zugehörigen Multiplizitäten.

Abbildung 2: Einfacher Bestellprozess



(Quelle: eigene Darstellung auf Basis von (van der Aalst 2020b))

Eine Bestellung setzt sich aus einem oder aus mehreren Artikeln zusammen. Ein Artikel kann zusätzlich aus mehreren Einheiten bestehen. Im Versand erfolgt die Verpackung der Artikel. Es besteht die Möglichkeit, mehrere Artikel in einem Paket zu versenden. Sind beispielsweise nicht alle Artikel sofort verfügbar, können Artikel aus einer Bestellung in mehreren Paketen versandt werden. Zudem müssen Artikel in einem Paket nicht aus der gleichen Bestellung kommen. Die Lieferung findet dann auf Basis einer vorgeplanten Route statt, kann aber, falls der Empfänger oder die Empfängerin nicht sofort anzutreffen ist, auch auf einer anderen Route ausgeliefert werden. Typischerweise erfolgt auf einer Route die Auslieferung mehrerer Pakete.

Selbst für diesen einfachen Beispielprozess erweist es sich als schwierig, eine einzige Case Notion bzw. einen

einzig Objekttypen auszuwählen: Es muss davon ausgegangen werden, dass Defizite, Konvergenz- und/ oder Divergenzprobleme auftreten.

Defizit im Sinne von Process Mining bedeutet, dass aufgrund der Wahl einer Case Notion Events unabsichtlich verloren gehen, da dieses sich nicht auf die gewählte Case Notion bezieht. Ein Beispiel dafür ist in Abb. 3 zu sehen: Für das Event mit der Aktivität „Bestellung aufgeben“ und „Bestellung bestätigen“ existiert keine Paketnummer. Wenn beim Glättungsprozess der Daten das Objekt „Paket“ als Case Notion gewählt wird, gehen die zwei anderen Events verloren, da sie keinem Paket zugehörig sind.

Abbildung 3: Event-Log für einen Bestellprozess, um das Defizit zu zeigen

Timestamp	Aktivität	Artikel	Bestellung	Paket
...
20.05.2019	Bestellung aufgeben	[880005, 880006, 880007]	[990002]	
...
20.05.2019	Bestellung bestätigen	[880009, 880010, 880011]	[990003]	
20.05.2019	Artikel nicht auf Lager	[880004]	[990001]	[66001]
...

Timestamp	Aktivität	Artikel	Bestellung	Paket
...
20.05.2019	Artikel nicht auf Lager	[880004]	[990001]	[66001]
...

(Quelle: Eigene Darstellung auf Basis von (van der Aalst 2020b))

Wählt man beispielsweise „Artikel“ als Case Notion in der Tabelle aus Abb. 4 aus, wiederholt sich das Event im Event-Log, da nun die Aktivität „Bestellung aufgeben“ für Artikel 880005, Artikel 880006 und Artikel 88007 aufgeteilt wird. Diese unbeabsichtigte Vervielfältigung charakterisiert das Konvergenzproblem: Es erschwert, im weiteren Verlauf genaue Aussagen über Kosten, Durchlaufzeiten usw. zu treffen.

Abbildung 4: Event-Log für einen Bestellprozess, um Konvergenz zu zeigen

Timestamp	Aktivität	Artikel	Bestellung
...
20.05.2019	Bestellung aufgeben	[880005, 880006, 880007]	[990002]
...
20.05.2019	Bestellung bestätigen	[880009, 880010, 880011]	[990003]
20.05.2019	Artikel nicht auf Lager	[880004]	[990001]
...

Timestamp	Aktivität	Artikel	Bestellung
...
20.05.2019	Bestellung aufgeben	[880005]	[990002]
20.05.2019	Bestellung aufgeben	[880006]	[990002]
20.05.2019	Bestellung aufgeben	[880007]	[990002]
...
20.05.2019	Bestellung bestätigen	[880009]	[990003]
20.05.2019	Bestellung bestätigen	[880010]	[990003]
20.05.2019	Bestellung bestätigen	[880011]	[990003]
20.05.2019	Artikel nicht auf Lager	[880004]	[990001]
...

(Quelle: Eigene Darstellung auf Basis von (van der Aalst 2020b))

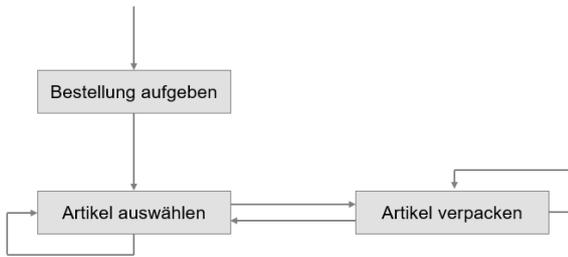
Divergenz im Sinne von Process Mining bringt verschiedene Events in einen möglichen kausalen Zusammenhang, welcher in Realität nicht existiert. Abb. 5 zeigt ein beispielhaftes Event-Log, in dem es für jeden Artikel die Aktivitäten „Bestellung aufgeben“, „Artikel auswählen“ und „Artikel verpacken“ gibt. Wählt man das Objekt „Artikel“ als Case Notion, lässt sich eine klare und logische Abfolge der drei genannten Aktivitäten erkennen. Geht man demgegenüber von „Bestellung“ als Case Notion aus, geht die Kausalität der „Artikel“ verloren: Wie Abb. 6 zeigt, reihen sich die Aktivitäten in einer nicht der Realität entsprechenden Abfolge (van der Aalst 2020b).

Abbildung 5: Event-Log für einen Bestellprozess, um Divergenz zu zeigen

Timestamp	Aktivität	Artikel	Bestellung
...
20.05.2022	Bestellung aufgeben	[880005, 880006, 880007]	[990002]
22.05.2022	Artikel auswählen	[880005]	[990002]
22.05.2022	Artikel auswählen	[880006]	[990002]
23.05.2022	Artikel verpacken	[880006]	[990002]
23.05.2022	Artikel verpacken	[880005]	[990002]
22.05.2022	Artikel auswählen	[880007]	[990002]
23.05.2022	Artikel verpacken	[880007]	[990002]
...

(Quelle: Eigene Darstellung auf Basis von (van der Aalst 2020b))

Abbildung 6: Ablauf für Bestellprozess durch Divergenz



(Quelle: eigene Darstellung auf Basis von (van der Aalst 2020b))

OCPM hingegen erlaubt es, eine Aktivität auf mehrere Objekte zu beziehen. In Bezug auf das Beispiel kann die Aktivität „Artikel verpacken“ sich sowohl auf eine Bestellung als auch auf mehrere Artikel beziehen und zielt somit darauf ab, die Probleme von Defiziten, Konvergenz und Divergenz zu lösen (van der Aalst 2019).

Object Centric Event-Logs

Der XES-Standard bezieht sich auf Event-Logs, in denen eine einzelne Case Notion gewählt werden muss. Bei Object Centric Event-Logs (OCEL) handelt es sich um eine neue Art von Event-Logs, die es ermöglichen, die Realität ohne die vorher beschriebene Konvergenz- bzw. Divergenzproblematik abzubilden. Dazu wurde von der PADS der RWTH Aachen der OCEL-Standard entwickelt, welcher als Standarddatenformat für OCPM gilt. Dessen Zweck ist es „einen allgemeinen Standard für den Austausch objektbezogener Ereignisdaten mit mehreren Case Notions bereitzustellen“ (Ghahfarokhi et al. 2021). Weiterhin soll der Standard durch die JavaScript Object Notation (JSON)/XML Serialisierungen von OCEL mehrere Sprachen und Systeme unterstützen und damit Interoperabilität ermöglichen. Zudem ist der Standard auch erweiterbar. Abschließend soll dieser Standard in individuellen Anwendungen verwendbar sein und unterstützt werden.

Ein Ausschnitt eines Order Management Beispiellogs ist in Abb. 7 zu sehen. In diesem besteht das erste Event aus der Aktivität „Bestellung bestätigen“ und bezieht sich auf die Objekte „Artikel“, „Produkt“ und „Bestellung“. Zudem gibt es auch Attribute, welche in der Darstellung „Gewicht“ und „Preis“ sind. Weitere typische Beispiele wären Purchase-to-Pay (P2P) oder Order-to-Cash (O2C) Prozesse.

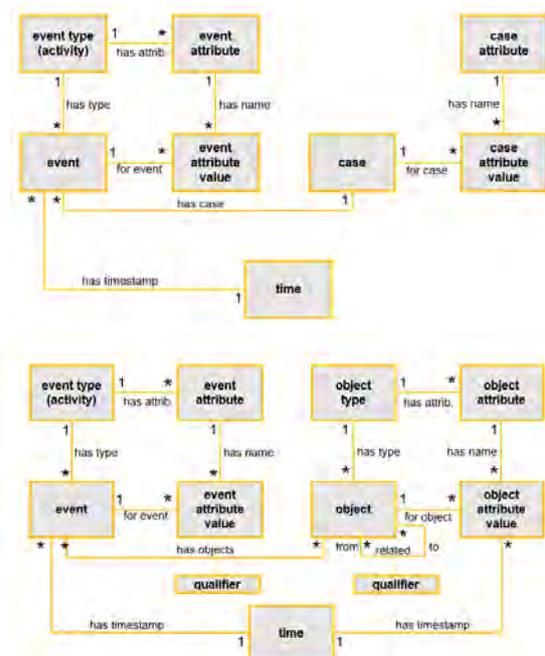
Abbildung 7: Ausschnitt aus einem Order Management Log

Event ID	Timestamp	Aktivität	Gewicht	Preis	Artikel	Produkt	Bestellung	Paket
1350	23.05.2019 12:55	Bestellung bestätigen	2432	98398	[8800096', '8800094', '8800096']	[iPhone X', 'Echo Plus', 'Echo Show 8]	[9900024]	
1360	23.05.2019 13:04	Bestellung bestätigen	348	453498	[8800081', '8800080', '880079]	[Echo Show 8', 'MacBook Air]	[9900020]	[9600001]
1370	23.05.2019 13:32	Artikel auswählen	98	12999	[8800090]	[Echo Show 8]	[9900020]	
1380	23.05.2019 13:39	Artikel auswählen	28	8999	[8800039]	[Fire Stick 4K]	[9900111]	[9600001]
1390	23.05.2019 13:57	Bestellung aufgeben	1263	18498	[880110', '880111]	[Echo', 'Kindle]	[9900027]	

(Quelle: eigene Darstellung auf Basis von (PADS - RWTH Aachen University 2022))

Der konzeptionelle Unterschied in den Datenmodellen zwischen klassischem Process Mining und OCPM zeigen die Metamodelle in Abb. 8: Im klassischen Process Mining existiert ein „Case“ (mit einer eindeutigen „Case ID“). Diese Klasse wird bei OCPM durch „Object“ ersetzt, die ihrerseits mit dem „Object Type“ in Form einer N:1 Kardinalität verbunden ist. Jedes Objekt hat genau einen Objekt Identifikator und einen Objekttypen. Dem „Object Type“ lassen sich wiederum mehrere Objektattribute in Form einer 1:N Kardinalität zuordnen. Bei Object Centric Event Data (OCED) sind „Event“ und „Object“ mit einer N:M Beziehung miteinander verbunden. Somit ist es möglich, dass mehrere Objekte in einem Event enthalten sind oder ein Objekt in mehreren Events. Mit Hilfe eines „Qualifier“ wird die Zuordnung von Events zu einem Objekt im Eventlog realisiert. Auch zwischen Objekten können N:M Beziehungen bestehen. Entsprechend sorgt auch hier ein „Qualifier“ für die entsprechende Zuordnung (van der Aalst 2023).

Abbildung 8: Metamodelle zur Konzeptualisierung von traditionellen Eventlogs (oben) und zu objektzentrierten Eventdaten (unten)



(Quelle: eigene Darstellung auf Basis von (van der Aalst 2023))

ANWENDUNG VON OCPM

Der nachfolgend vorgestellte Purchase-to-Pay-Prozess stellt einen generischen Einkaufsprozess im Unternehmen dar: Er beginnt mit der Beschaffung und endet mit der Bezahlung der Ware (vgl. Abb. 9).

Abbildung 9: Vereinfachter P2P-Prozess

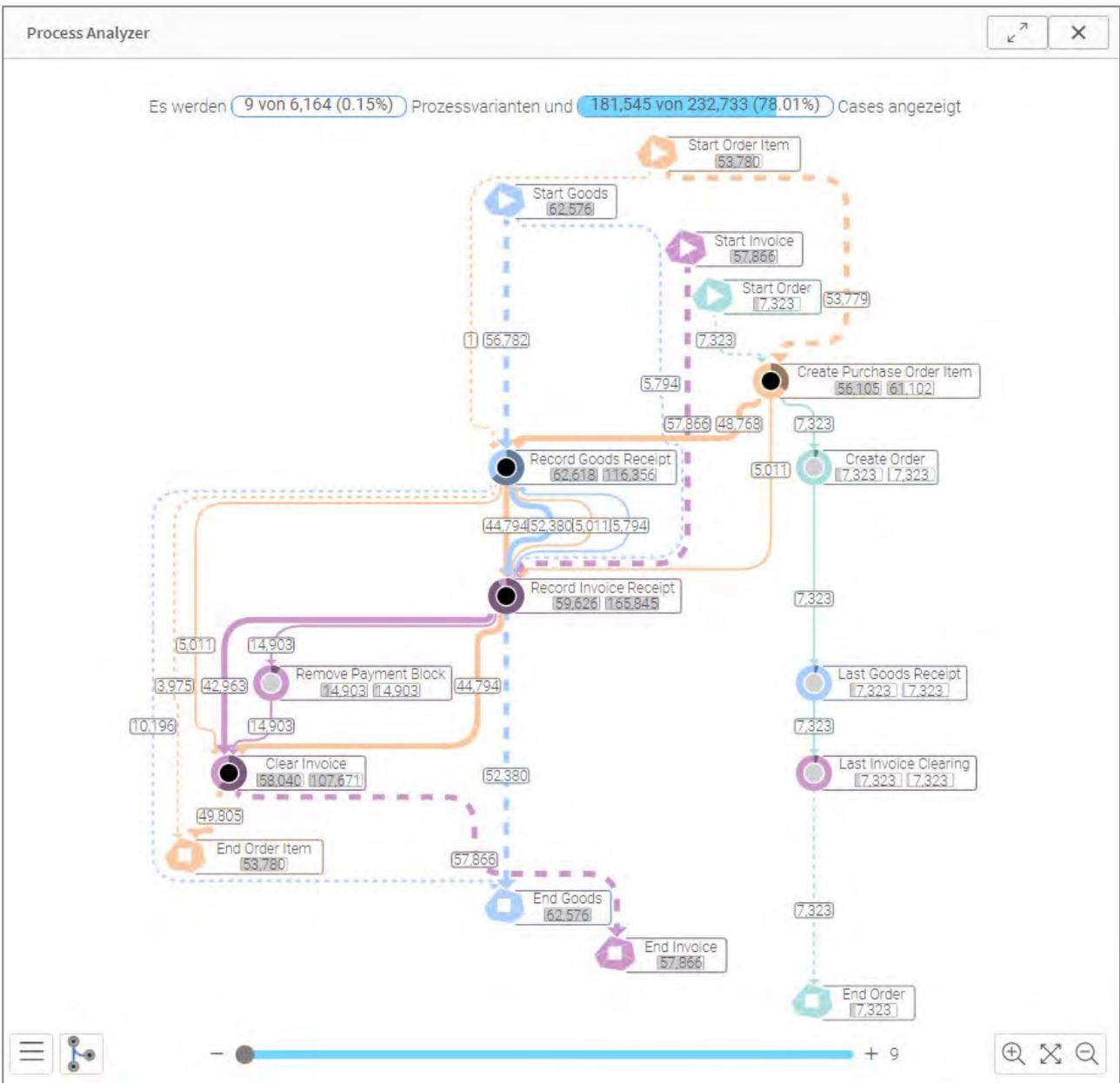


(Quelle: Eigene Darstellung auf Basis von (Hartel et al. 2021))

Nach dem Prozessstart erfolgt eine Bestellung von Materialien und im Anschluss die Übertragung an den Lieferanten. Anschließend werden die Waren vom Lieferanten an das Unternehmen geliefert und der Wareneingang dokumentiert. Darauf folgt der Erhalt der Rechnung und die abschließende Zahlung (Hartel et al. 2021).

Für die OCPM-Analyse wird ein Datensatz verwendet, welcher von MEHRWERK zur Verfügung gestellt wurde. Grundsätzlich stellt die PADS der RWTH Aachen öffentlich zugängliche Datensätze für OCPM zur Verfügung. Diese sind aber deutlich kleiner als der verwendete Datensatz und beinhalten daher lediglich eingeschränkte Analysemöglichkeiten. Weiterhin fehlen Case Dimensionen, d.h. Attribute wie beispielsweise Lieferant, Kostenbereich, in diesen öffentlichen Datensätzen.

Abbildung 10: Process Analyzer



(Quelle: Auszug MPM OCPM Anwendung)

Process Analyzer

Um den zu analysierenden Prozess zu visualisieren, wird der Process Analyzer (dt. Prozessanalysator) genutzt. Dieser bietet eine typische Process Mining Funktionalität, die Analysten dabei hilft, Prozessvarianten zu verstehen und Varianten zu entdecken, weshalb er auch als Variant Explorer bezeichnet wird. Um den Prozess als Graphen darzustellen, werden Knoten und Kanten verwendet. Knoten sind als Kreise dargestellt und repräsentieren die Aktivitäten, welche der Prozess durchläuft. Für die Aktivitäten werden außerdem ein bis zwei weitere Prozessleistungsindikatoren angezeigt (Anzahl an Events und Anzahl an Interaktionen). Kanten stellen die Verbindungslinien zwischen den Knoten her und ihre Häufigkeit lässt sich anhand der beschrifteten Zahl und der Linienbreite erkennen. Eine breite Linie stellt eine hohe Häufigkeit dar und vice versa. Verbindungslinien, die zum Start- oder Endevent führen, werden als gestrichelte Linie angezeigt, da diese keine Prozessschritte sind. Über dem Prozessverlauf wird zusätzlich der relative Anteil der ausgewählten Varianten an allen Prozessvarianten angezeigt. Ferner erfolgt der Ausweis, welche Anzahl an Cases durch die ausgewählten Varianten nominal und relativ abgebildet werden. Ein Nutzer kann die Anzahl an angezeigten Varianten über einen Slider auswählen. Die Reihenfolge der Varianten richtet sich nach der Häufigkeit, somit ist die häufigste Prozessvariante immer die erste Variante. Es besteht ebenfalls die Möglichkeit, die durchschnittliche Wartezeit oder den Anteil der Cases direkt an den Kanten anzuzeigen, um insbesondere Engpässe zu erkennen. Ziel des Process Analyzer ist es, die Komplexität des Prozesses zu reduzieren, um das Prozessverständnis zu vergrößern. Abb. 10 zeigt den Process Analyzer und stellt 9 von 6.194 Prozessvarianten dar, womit bereits 78,01 % der Cases abgedeckt sind. Bereits mit neun Varianten ist der Prozess komplex und nicht

intuitiv nachvollziehbar.

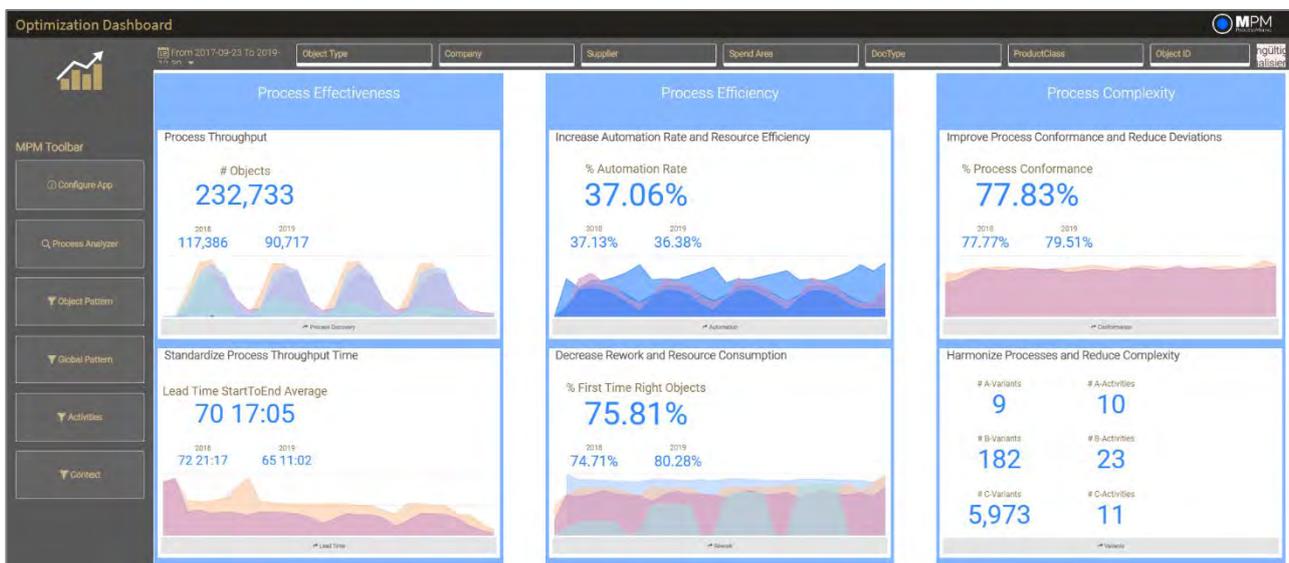
Der Nutzer kann sich entsprechend weniger Prozessvarianten anzeigen lassen oder spezifisch nach einzelnen Aktivitäten filtern. Dafür können NutzerInnen eine Aktivität auswählen und entscheiden, ob ausschließlich Prozessverläufe mit dieser Aktivität oder ohne diese Aktivität dargestellt werden sollen. Zudem kann man sich gezielt Prozessverläufe anzeigen lassen, die mit dieser Aktivität beginnen oder enden.

Arbeitsblätter

Das Optimization Dashboard in Abb. 11 bildet einen Ausgangspunkt für eine Analyse, da es verschiedene Informationen in aggregierter Form anzeigt. Dafür werden Process Performance Indicators (PPIs) (dt. Prozessleistungsindikatoren) zu allen Objekten dargestellt. Prozessverantwortlichen steht damit ein Werkzeug zur Verfügung, mit dem sich überprüfen lässt, welche Initiativen zur Prozessverbesserung den größten Erfolg haben. Hierfür werden drei Bereiche der Optimierung abgedeckt: Effektivität, Effizienz und Komplexität.

Der linke Block stellt die Prozesseffektivität anhand des Prozessdurchsatzes und der Standarddurchlaufzeit dar. Zusätzlich werden die Kennzahlen im Zeitverlauf pro Jahr angezeigt. Der P2P-Prozess hat insgesamt einen Durchsatz von 232.733 Objekten. Im dazugehörigen Liniendiagramm ist für jeden Objekttyp eine wellenartige Linie zu sehen. Dieser Verlauf könnte auf einen saisonalen Einkaufsprozess hinweisen, jedoch wurden im konkreten Fall die Daten vervielfältigt, um den Datensatz zu vergrößern und hat damit keinen realitätsnahen Hintergrund. Die durchschnittliche Durchlaufzeit von Anfang bis Ende beträgt 70 Tage, 17 Stunden und 5 Minuten,

Abbildung 11: Optimization Dashboard



(Quelle: Auszug MPM OCPM Anwendung)

(siehe Abb. 18). 2018 waren es ca. 72 Tage und 2019 hat sich die Standarddurchlaufzeit bereits auf ca. 65 Tage verkürzt. Das entspricht einer Reduktion um 9,7 % und folglich waren die Optimierungsmaßnahmen erfolgreich.

Der mittlere Block zeigt die Prozesseffizienz an und teilt sich in zwei Abschnitte auf. Diese Abschnitte dienen jeweils dem Ziel, die Prozesseffizienz zu erhöhen. Inhaltlich geht es dabei einerseits um die Erhöhung der Automatisierungsrate und Ressourceneffizienz sowie andererseits um die Senkung der Nacharbeit und des Ressourcenverbrauchs. Im P2P-Prozess hat sich über die letzten zwei Jahre die Automatisierungsrate von 37,13 % auf 36,38 % verringert und repräsentiert somit einen zentralen Handlungsbedarf, da ein strategisches Unternehmensziel die Steigerung der Automatisierungsrate ist. Die Nacharbeit ist von 2018 auf 2019 gesunken, wie aus der Steigerung der fehlerfreien Bearbeitung von 74,71 % auf 80,28 % („First Time Right Objects“) hervorgeht. Es ist ersichtlich, dass die verschiedenen Objekttypen im Prozess große Unterschiede aufweisen.

Der Objekttyp „Goods“, welcher in Blau im Liniendiagramm dargestellt wird, zeigt eine nahezu waagrechte Linie und der Objekttyp „Order“ weist eine wellenartige Linie auf. Das sind Anhaltspunkte, um eine Ursache-Wirkungsanalyse durchzuführen.

Der rechte Block im Optimierungsdashboard bildet die Prozesskomplexität ab und teilt sich in zwei Abschnitte. Die zugehörigen Zielsetzungen sind zum einen die Verbesserung der Prozesskonformität und Reduzierung der Abweichungen sowie zum anderen die Harmonisierung der Prozesse und die Komplexitätsreduktion. Im Bereich der Prozesskonformität ist im Zeitverlauf von 2018 auf 2019 eine kleine Steigerung von 2,3 % ersichtlich und das Liniendiagramm zeigt eine nahezu gerade Linie und somit keine Auffälligkeiten. Für den Teil der Prozessharmonisierung werden die Varianten und Aktivitäten innerhalb des P2P-Prozesses anhand einer ABC-Analyse aufgegliedert. Der A-Teil beschreibt ca. 80 %, der B-Teil 15 % und der C-Teil ca. 5 %.

Alle sechs Visualisierungen haben unter dem Liniendiagramm einen grauen Balken, welcher den Nutzer direkt zum entsprechenden Arbeitsblatt weiterleitet, um die Analyse zu vertiefen.

Weiterhin gibt es eine Vielzahl an Arbeitsblättern, auf die hier nicht im Detail eingegangen wird. Dazu gehört das Start-to-End-Lead-Time-Arbeitsblatt, welches sich auf die Prozessoptimierung in der zeitlichen Dimension fokussiert. Das Automatisierungsarbeitsblatt dient dazu, Bereiche und Aktivitäten zu identifizieren, die durch Automatisierung optimiert werden können. Zusätzlich dient das Monitoring Arbeitsblatt zur Echtzeitüberwachung von unvollendeten Cases und zeigt damit das der Trend der Echtzeitüberwachung bereits in der Praxis angewandt

wird. Abschließend beinhaltet die OCPM-Lösung noch spezifische Arbeitsblätter für Detailanalysen, Konformitätsprüfungen und diverse Übersichts- und Benchmarkingarbeitsblätter.

DISKUSSION (VERGLEICH TPM & OCPM)

Um die im Theorieteil erläuterten Unterschiede von OCPM und konventionellem Process Mining aufzuzeigen, wird im Folgenden der identische P2P-Prozess mit dem konventionellen Ansatz und dem objektzentrierten Ansatz analysiert. Dafür wurde von MEHRWERK für beide Ansätze eine Anwendung bereitgestellt, in der die Daten bereits importiert wurden. Die Analyse des P2P-Prozess wurde bereits in Kap. 3 mit OCPM vollzogen. Für die Anwendung von konventionellem Process Mining wird „Order Item“ als Case Notion gewählt.

Ein großer Unterschied ergibt sich bei Extraktion und Aufbereitung der Daten: Für konventionelles Process Mining müssen diese aufwändig aufbereitet und geglättet werden, da die Daten in objektzentrierter Form in Informationssystemen vorliegen. OCPM kann diese objektzentrierte Form fast ohne Transformationsschritte übernehmen und reduziert im Anwendungsfall den Aufwand erheblich. Das Zusammenspiel der Objekte beim OCPM muss von einem Data Analysten festgelegt werden: Dieser definiert, welche Interaktionen oder Übergabepunkte, sogenannte Touchpoint-Aktivitäten, zwischen den Objekten genutzt werden sollen. Die Experten von MEHRWERK schätzen den Aufwand der Extraktion und des Imports insgesamt für OCPM 20–50 % geringer ein als im konventionellen Mining.

Die zwingende Auswahl einer Case Notion erlaubt nur eine Sicht auf den Prozess. Um die Case Notion im klassischen Process Mining zu ändern, müssen die Daten erneut geglättet und importiert werden, was einen weiteren erheblicher Mehraufwand bedeutet. Das führt in der Praxis zu Entscheidungsproblemen, da alle Stakeholder einer Process Mining Analyse sich auf eine Case Notion einigen müssen. Die Finanzabteilung wünscht sich dabei i.d.R. eine Analyse, welche sich auf die Rechnungen fokussiert, die Logistikabteilung demgegenüber eine Analyse mit Fokus auf den Lieferscheinen. Zwangsläufig können mit dem konventionellen Ansatz im Gegensatz zu OCPM nicht beide Wünsche parallel befriedigt werden.

Zusätzlich sind die Probleme von Konvergenz und Divergenz auch in der praktischen Anwendung zu beobachten. Abb. 12 und Abb. 13 zeigen einen Ausschnitt aus den jeweiligen Optimization Dashboards, welche die vier wichtigsten PPIs visualisiert. Abb. 12 zeigt dabei die PPIs der konventionellen Process Mining Analyse und Abb. 13 die PPIs der OCPM-Analyse.

Die identische Datenbasis wird durch die Anzahl der Cases bzw. der Anzahl der Objekte evident. Die Anzahl der Events unterscheidet sich zwischen den zwei Ausprägungen jedoch erheblich: Der konventionelle Ansatz zählt 849.401 Events, wobei die OCPM-Analyse nur 558,882 Events zählt. Die Differenz von 290.519 Events ist dabei auf Konvergenz, also der unbeabsichtigten Vielfältigkeit von Events, zurückzuführen.

Abbildung 12: PPIs des P2P-Prozess nach konventioneller Process Mining Analyse



(Quelle: Auszug MPM OCPM Anwendung)

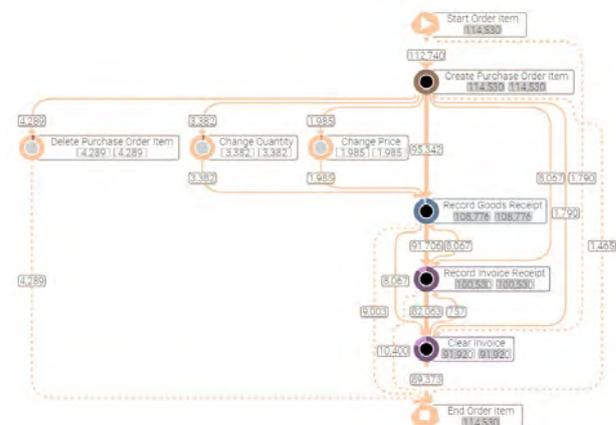
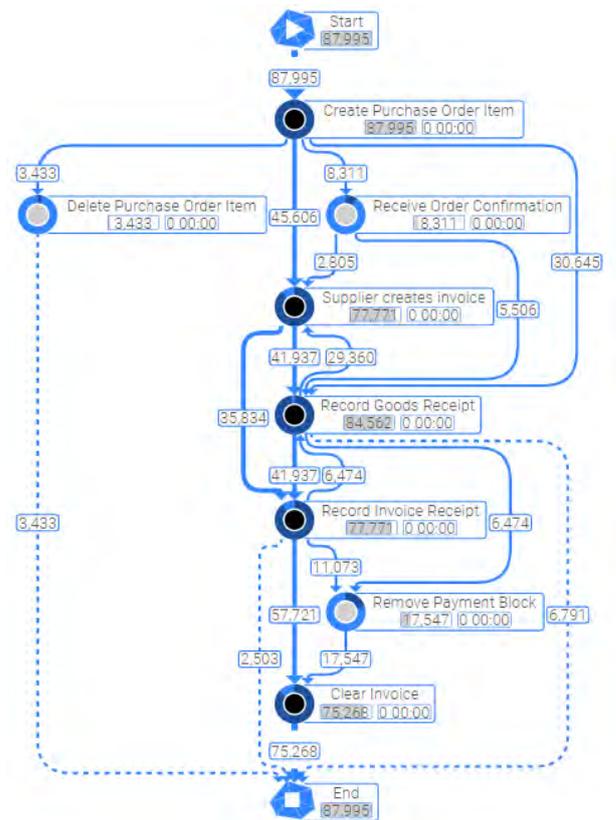
Abbildung 13: PPIs des P2P-Prozess nach OCPM Analyse



(Quelle: Auszug MPM OCPM Anwendung)

Neben der Differenz in der Anzahl der Events existiert ein weiterer Unterschied: In der konventionellen Process Mining Analyse hat der Prozess 8.218 Varianten, jedoch zeigt die objektzentrierte Analyse nur 2.662 Varianten an. Dies ist hauptsächlich auf die Divergenzproblematik zurückzuführen, da diese dazu führt, dass Events in einen kausalen Zusammenhang gebracht werden, der in der Realität nicht existiert. Entsprechend fälschlicherweise vermutete Zusammenhänge führen zu Rückschlüssen auf Prozessverläufen, die in dieser Form nie in der Praxis ausgeführt wurden. Abb. 14 zeigt die Process Analyzer mit den zehn häufigsten Prozessvarianten aus dem konventionellen Process Mining (oben) und den Process Analyzer aus OCPM (unten). Die OCPM-Version ist nach dem Order Item sortiert, um die Prozessverläufe besser vergleichen zu können. Es zeigt sich, dass sich die Aktivitäten bereits in den ersten zehn häufigsten Varianten unterscheiden: In der OCPM-Analyse werden unter anderem die Aktivitäten „Change Price“ und „Change Quantity“ angezeigt, welche im konventionellen Ansatz nicht enthalten sind.

Abbildung 14: Process Analyzer von OCPM (unten) und klassischem Process Mining (oben)



(Quelle: Auszug MPM OCPM Anwendung)

Dafür sind im Process Analyzer des konventionellen Ansatzes Aktivitäten wie „Receive Order Confirmation“ enthalten, die der OCPM Process Analyzer nicht ausweist. Diese Unterschiede resultieren aus der jeweiligen Anzahl der Prozessvarianten und der Events. Als Konsequenz von Informationsdefiziten, Divergenz und Konvergenz lassen sich durch das klassische Process Mining keine aussagekräftigen Prozessmodelle erstellen. Dem durch konventionellem Process Mining erstellten Prozessverlauf fehlen zudem die Interaktionen zwischen den verschiedenen Objekten.

Die Modellierung der Interaktionen für OCPM erweist sich gegenwärtig generell als problematisch: In der praktischen Anwendung sind die Beziehungen zwischen den einzelnen Objekten nicht immer transparent. Somit hängt die Qualität der OCPM-Datenmodelle von den Fähigkeiten eines Data Engineers ab.

ZUSAMMENFASSUNG & AUSBLICK

Um den aktuellen Wissensstand von OCPM zu zeigen, wurden die theoretischen Grundlagen erläutert und aufgezeigt, dass die zwingende Auswahl einer einzigen „Case Notion“ im konventionellen Process Mining zu Problemen im Hinblick auf Informationsdefizite, Konvergenz und Divergenz führt. Für die Bewertung des aktuellen Entwicklungsstandes von OCPM wurde mit Hilfe eines P2P-Prozesses ein Anwendungsfall generiert. Sowohl der zugehörige Datensatz als auch die Process-Mining-Software wurden von der Firma MEHRWERK gestellt. Hierdurch konnten die Analysemöglichkeiten von OCPM realitätsnah in einzelnen Arbeitsblättern dargestellt werden. Es zeigt sich dabei auch, dass die Datenbereitstellung bei OCPM wesentlich weniger aufwändig und gleichzeitig flexibler für den Einsatz in der Praxis ist.

Als Fazit lässt sich ziehen, dass sich durch OCPM Optimierungspotenziale besser als durch klassisches Process Mining ableiten lassen: Aufwändige Aufbereitungen und Glättungen der Daten können reduziert werden. Ferner sind Konvergenz- und Divergenzprobleme durch OCPM wesentlich besser lösbar. Aufgrund dieser Vorteile ist davon auszugehen, dass OCPM aufgrund der zuvor aufgezeigten Vorteile mittelfristig traditionelles Process Mining komplett ersetzen wird. Hieraus lassen sich Impulse für die integrierte Analyse verschiedener Geschäftsprozesse und zugehörigen Verbesserungen ableiten.

Weiterer Forschungsbedarf ergibt sich aus der Analyse von stetigen Daten (van der Aalst 2020a): Gegenwärtig geht man davon aus, dass Ereignisdaten diskret sind. Durch das wachsende Anwendungsgebiet von Process Mining können jedoch in Zukunft auch stetige Daten zum Einsatz kommen. Wenn diese jedoch in diskreter Form dargestellt werden müssen, kann dies zu mehr falsch-positiven oder falsch-negativen Werten führen. Hieraus erwächst die Gefahr, dass nicht stattgefundenen Aktivitäten dem Ereignisprotokoll hinzugefügt werden, was wiederum zu falsch abgeleiteten Schlussfolgerungen und Entscheidungen führen kann (van der Aalst 2020a).

LITERATUR

Berti, Alessandro; van der Aalst, Wil M. P. (2022): OC-PM: analyzing object-centric event logs and process models. In:

International Journal on Software Tools for Technology Transfer, S. 1–17.

Ghahfarokhi, Anahita Farhang; Park, Gyunam; Berti, Alessandro; van der Aalst, Wil M. P. (2021): OCEL: A Standard for Object-Centric Event Logs. In: European Conference on Advances in Databases and Information Systems: Springer, Cham, S. 169–175.

Hartel, Dirk H.; Javorovic, Filip; Norberg, Leonard (2021): Anwendungsmöglichkeiten von Process Mining am Beispiel des Purchase-to-Pay-Prozesses für Raw Materials bei Bosch Power Tools — ein Theorie-Praxis-Transfer. In: Supply Management Research: Springer Gabler, Wiesbaden, S. 161–187.

Laue, Ralf; Koschmider, Agnes; Fahland, Dirk (2021): Prozessmanagement und Process-Mining. Grundlagen. München, Berlin: De Gruyter Oldenbourg; Walter de Gruyter GmbH (De Gruyter Studium).

van der Aalst, Wil M. P. (2011): Process mining: Discovery, Conformance and Enhancement of business processes. Springer, Berlin.

van der Aalst, Wil M. P. (2012): Process Mining: Over-view and Opportunities. In: ACM Transactions on Management Information Systems 3 (2), S. 1–17.

van der Aalst, Wil M. P. (2019): Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data. In: International Conference on Software Engineering and Formal Methods: Springer, Cham, S. 3–25.

van der Aalst, Wil M. P. (2020a): Academic View: Development of the Process Mining Discipline. In: Process Mining in Action: Springer, Cham, S. 181–196.

van der Aalst, Wil M. P. (2020b): Genuine Process Improvements Using Object-Centric Process Mining. Virtual Lectures Series on Business Process Management. Universität Würzburg, 30.11.2020.

van der Aalst, Wil. M. P. (2023). Object-Centric Process Mining: Unraveling the Fabric of Real Processes. Mathematics, 11(12), 2691.

Weblinks

Celonis (2023): What is object-centric process mining? Online verfügbar unter <https://www.celonis.com/blog/what-is-object-centric-process-mining-ocpm/>, zuletzt aktualisiert am 07.02.2023, zuletzt geprüft am 07.02.2023.

PADS - RWTH Aachen University (2022): OCEL Standard. Online verfügbar unter <https://ocel-standard.org/>, zuletzt aktualisiert am 05.12.2022, zuletzt geprüft am 02.02.2023.

Entwicklung eines Unterweisungsmoduls für Microsoft Dynamics NAV 2016

Patrick Hilgart

Pfleiderer Teisnach GmbH & Co. KG
Adolf-Pfleiderer-Straße 19, 94244 Teisnach
E-Mail: p.hilgart@pf-t.de

Professor Dr. Frank Herrmann

Ostbayerische Technische Hochschule Regensburg
Innovationszentrum für Produktionslogistik und Fabrikplanung
Galgenbergstraße 32, 93053 Regensburg
E-Mail: frank.herrmann@oth-regensburg.de

SCHLÜSSELWÖRTER

ERP-System, Microsoft Dynamics NAV 2016,
Unterweisung

ABSTRAKT

Sicherheit am Arbeitsplatz ist in der heutigen Zeit unerlässlich. Die psychische und physische Gesundheit aller Beschäftigten eines Unternehmens dürfen niemals vernachlässigt werden. Um die Mitarbeiter zu sensibilisieren, sieht der Gesetzgeber regelmäßige Unterweisungen vor. Diese bedeuten für den Unternehmer nicht nur Zeitaufwand während den Unterweisungsterminen, sondern auch bei der Organisation. Die fortschreitende Digitalisierung der Geschäftsprozesse kann diese Aufgaben erleichtern, menschliche Fehler minimieren und Arbeitszeit einsparen. Die zentralen ERP-Systeme der Unternehmen eignen sich auch in diesem Bereich als Plattform. Für die im Unternehmen Pfleiderer Teisnach GmbH & Co. KG eingesetzte ERP-Software Microsoft Dynamics NAV 2016 soll ein Unterweisungsmodul programmiert werden, dass auch diese Geschäftsprozesse digitalisiert und zum Großteil automatisiert.

1 EINFÜHRUNG

In jedem Unternehmen herrschen Gefahren am Arbeitsplatz. Viele sind offensichtlich, andere wiederum kaum zu erkennen. Insbesondere durch Routinearbeiten und der daraus folgenden Unachtsamkeit der Mitarbeiter, erhöht sich das Unfallrisiko erheblich. Sicherheitsunterweisungen finden daher nicht nur zu Beginn eines Arbeitsverhältnisses statt, sondern müssen regelmäßig wiederholt werden. Dies wird durch den Gesetzgeber in Deutschland nach §12 des Arbeitsschutzgesetzes vorgeschrieben [1]. Die Unterweisungen sollen den Mitarbeiter an die Gefahren des eigenen Arbeitsplatzes und der betrieblichen Umgebung erinnern. Sie sind ein wichtiges Mittel, um Arbeitsunfälle zu vermeiden und die Gesundheit der Mitarbeiter zu schützen.

Die Unterweisungen bedeuten für den Arbeitgeber nicht nur Zeitinvestitionen während den konkreten Unterweisungsterminen, sondern auch durch den erheblichen Verwaltungsaufwand. Dieser wird meist an die einzelnen Abteilungsleiter abgegeben. Sie übernehmen die Verantwortung zur Unterweisung der eigenen Mitarbeiter. Zum

einen müssen die benötigten Unterweisungsunterlagen ständig anhand aktueller Gesetze und Normen überarbeitet werden, zum anderen gilt es sicherzustellen, dass alle erforderlichen Mitarbeiter an den Unterweisungsterminen teilnehmen. Für die leitenden Angestellten bedeutet dies zusätzlichen Arbeitsaufwand.

Diese dezentrale Struktur erschwert den Überblick für den Arbeitgeber, da alle Dokumente, Termine und Teilnehmerlisten zunächst von den jeweiligen Abteilungsleitern anzufordern sind. Bei Dokumentenanfragen durch staatliche Behörden stellt dies ein Problem dar. Die Daten müssen nach einem Arbeitsunfall schnellstmöglich zur Verfügung stehen.

Die fortschreitende Digitalisierung der Unternehmen kann auch auf die Geschäftsprozesse der Unterweisungen ausgeweitet werden. Dazu bieten sich die bereits eingesetzten ERP-Systeme an. Diese dienen als zentrale Verwaltungssoftware aller relevanten Betriebsabläufe.

Die im Unternehmen Pfleiderer Teisnach GmbH & Co. KG eingesetzte ERP-Software soll durch ein Modul erweitert werden, um die Geschäftsprozesse der Unterweisungen zu zentralisieren und zu automatisieren.

2 PROBLEMBESCHREIBUNG

Das Unterweisungsmodul muss den Anforderungen des Arbeitgebers, der Abteilungsleiter und des Gesetzgebers entsprechen. Das Modul muss die gewünschten Funktionalitäten bereitstellen und in einer Softwarekomponente zusammenfassen. Sie muss sich dabei nahtlos in das ERP-System integrieren. Dabei darf die Updatefähigkeit der Software nicht beeinträchtigt werden. Die Aktualisierung auf eine neue Programmversion muss auch zukünftig ohne Modulanpassungen möglich sein.

2.1 Funktionsumfang Unterweisungsmodul

Das Unterweisungsmodul muss den Arbeitgeber und die Abteilungsleiter bei ihrer Arbeit unterstützen und diese erleichtern. Trotz der dezentralen Struktur, soll die Verwaltung der Unterweisungen an einer Stelle erfolgen. So können allgemeingültige Unterweisungsvorlagen schnell

an neue Gesetze und Regelungen angepasst werden. Diese beinhalten auch alle Arbeitsanweisungen und andere durch den Unternehmer freigegebene Dokumente. Sie bilden dabei die Grundlage für spezifische Unterweisungsaufträge der einzelnen Abteilungen. Die Generierung soll schnell und einfach durch wenige Klicks erfolgen.

Der Gesetzgeber schreibt den Unternehmen vor, die Unterweisungen der einzelnen Mitarbeiter regelmäßig durchzuführen [1]. Um keine Fälligkeitstermine zu verpassen, müssen die Abteilungsleiter diese ständig im Blick behalten. Die zusätzliche Arbeit soll durch das Unterweisungsmodul abgenommen werden. Es soll die verantwortlichen Personen über die anstehenden Termine per E-Mail automatisch informieren. Dabei müssen auch die erforderlichen Teilnehmer aufgelistet werden. Der Benachrichtigungszeitpunkt soll monatlich frei wählbar sein.

Für den Arbeitgeber wiederum ist eine Gesamtübersicht der Unterweisungstermine erforderlich. Diese dient sowohl zur Kontrolle der erfolgreich durchgeführten, als auch der anstehenden Unterweisungen einzelner Personen. Die Darstellung muss am Computer erfolgen und auch auf Papier ausgedruckt werden können. Die Zusammenfassungen helfen weiterhin auch zur Terminplanung und Zusammenlegung zukünftiger Unterweisungen.

Die Dokumentation der Unterlagen ist ein weiterer großer Aspekt des Unterweisungsmoduls. Der Gesetzgeber sieht vor, dass die Unterweisungsunterlagen, wie Arbeitsanweisungen und Teilnehmerlisten archiviert werden müssen. Die Aufbewahrungsdauer variiert dabei je nach Unterweisungsthema, jedoch empfiehlt die DGUV zwei Jahre [2]. Das Unternehmen sieht vor, die Daten wie eine Personalakte bis zu zehn Jahre lang zu speichern. Das Unterweisungsmodul soll nicht nur bereits durchgeführte Unterweisungsaufträge ablegen, sondern auch die dazugehörigen Schulungsunterlagen und Teilnehmerlisten. Diese sollen importiert und abgespeichert werden, um alle Dokumente digital vorzuhalten. Eine automatisch angelegte Unterweisungshistorie ordnet die Daten den entsprechenden Mitarbeitern zu. Vergangene Unterweisungen finden sich damit in einer Mitarbeiteransicht wieder, in der alle detaillierten Informationen aufrufbar sind. Anfragen von staatlichen Behörden können in kurzer Zeit bearbeitet werden.

2.2 ERP-Systemintegration

In mittelständischen Unternehmen finden zumeist ERP-Systeme ihren Einsatz, um die betrieblichen Abläufe zu digitalisieren. Die wichtigsten Geschäftsprozesse werden bereits durch die Standardfunktionen der ERP-Software abgebildet. Sie wird damit zur zentralen Rolle in der Digitalisierung der alltäglichen Betriebsabläufe. Es ist somit naheliegend, dieses System durch ein Modul zu erweitern, dass die Unterweisungstätigkeiten übernimmt.

Die im Unternehmen Pfeleiderer Teisnach GmbH & Co. KG eingesetzte ERP-Software Microsoft Dynamics NAV 2016 soll um ein Unterweisungsmodul erweitert werden. Die dazu erforderliche Entwicklerlizenz ist be-

reits vorhanden. Die nahtlose Integration in das Basissystem ist dabei unumgänglich. Als wichtigstes Kriterium ist dabei die Updatefähigkeit der Software zu nennen. Die Aktualisierung auf neue Programmversionen soll ohne Anpassungen des Moduls oder des neuen Standardsystems möglich sein. Die von Microsoft bereitgestellten Objekte im Nummernbereich 1 bis 50.000 [3] werden während einer Aktualisierung wieder auf die Originalversion zurückgesetzt. Es wird daher dringend abgeraten, diese Objekte zu verändern. Da die durch das Unterweisungsmodul benötigten Informationen nicht vollumfänglich in den Standardtabellen gespeichert sind, müssen diese in zusätzlichen Tabellen ausgelagert und verknüpft werden. Die Datenkonsistenz muss trotz der aufgeteilten Datenspeicherung stets erhalten bleiben. Das Löschen von Daten aus den Standardtabellen muss auch die Bereinigung der Erweiterungstabelle zur Folge haben. Das Unterweisungsmodul muss auf bestimmte Ereignisse innerhalb der Standardprogrammierung reagieren können.

2.3 Datenübertragung Sage HR Suite

Sämtliche Mitarbeiterdaten sind in der Personaldatenbank Sage HR Suite geführt. Die kontinuierliche Datenpflege aus der Personalabteilung gewährleistet stets den aktuellen Stand der Daten. Einige Informationen benötigt auch das ERP-System in den Standardfunktionen. Bestimmte Buchungsvorgänge erfolgen mit Bezug auf einen Mitarbeiter. Beispielsweise sind in den Serviceaufträgen die entsprechenden Beschäftigten hinterlegt. Insbesondere die an das Unterweisungsmodul gestellten Anforderungen, können nur auf Basis aktueller Mitarbeiterdaten erfüllt werden. Die ständige Aktualisierung der Datenbank ist somit unerlässlich.

Die Übertragung der Daten soll jedoch nicht manuell erfolgen, da dies zusätzlichen Arbeitsaufwand für die Personalabteilung bedeuten würde. Weiterhin können menschliche Fehler verhindert werden, indem der Datenabgleich automatisch erfolgt. Die betroffenen Abteilungen sollen über veränderte Datensätze informiert werden, damit diese wissen, wann die aktualisierten Daten zur Bearbeitung bereitstehen. Um beispielsweise die erforderlichen Unterweisungen zuzuordnen oder um mitarbeiterbezogene Buchungen durchzuführen.

Die Personaldatenbank ist die maßgebende Informationsquelle, welche das ERP-System übernehmen muss.

3 LÖSUNG UNTERWEISUNGSMODUL

Das Unterweisungsmodul wird aus mehreren Programmabschnitten zusammengesetzt, welche die einzelnen Anforderungen umsetzen. Das Softwaredesign und die Bedienbarkeit folgen dem Aufbau der Standardfunktionalitäten des ERP-Systems. Damit soll selbst neuen Benutzern der einfache Einstieg in das Modul ermöglicht werden.

3.1 Einrichtung

Einige Einstellungen des Unterweisungsmoduls sollen von den Benutzern vorgenommen werden können. Jeder Abteilungsleiter wünscht sich andere Zeitpunkte zur automatischen E-Mail Benachrichtigung, um eine abgestufte Erinnerungshierarchie zu gestalten. So kann die Assistenz schon vor dem Abteilungsleiter eine Erinnerung erhalten, um den Termin vorzubereiten. Die Konfiguration benötigt dazu drei Einträge durch den Anwender. Die betreffende Abteilung, die E-Mail-Adresse und den Benachrichtigungszeitpunkt. Dieser bestimmt, wie viele Monate im Voraus die automatische E-Mail abgesendet wird.

Darüber hinaus kann der Abgleich der Mitarbeiterdaten mit der Personalsoftware aktiviert und deaktiviert werden. In seltenen Fällen kann es vorkommen, dass ein Mitarbeiter in die ERP-Datenbank eingetragen werden muss, ohne dass dieser in der Personalsoftware vorhanden ist. Auch wenn dieser Zeitraum nur wenige Tage beträgt, entfernt der Datenabgleich den Beschäftigten immer wieder aus der ERP-Datenbank. Es empfiehlt sich daher, die Synchronisation in diesem Zeitraum auszusetzen. Für diese Konfigurationen steht ein eigenes Einrichtungsfenster zur Verfügung.

3.2 Unterweisungsvorlagen

Alle im Unternehmen durchzuführenden Unterweisungen basieren auf Vorlagen, die durch die Vorgesetzten und dem Arbeitgeber gepflegt werden. Diese ermöglichen die Standardisierung und Aktualisierung der Unterweisungsdokumente an zentraler Stelle, bei minimalem Arbeitsaufwand. Gesetzesänderungen oder neue durch den Arbeitgeber freigegebene Dokumente, können für alle Abteilungen schnell eingearbeitet werden. Alle Unterweisungen erfolgen somit anhand aktueller Gesetze und Regelungen.

Die Vorlagen beinhalten nicht nur die benötigten Unterweisungsvorlagen und die dazugehörigen Beschreibungen, sondern speichern auch das erforderliche Unterweisungsintervall. Dort wird festgelegt, in welchen Zeitabständen die Unterweisung durchgeführt werden muss. Darauf basieren die späteren Berechnungen der Fälligkeitstermine.

Eine Zuordnungstabelle ermöglicht die Verknüpfung zwischen der Mitarbeiterliste und den einzelnen Unterweisungsvorlagen. Damit erfolgt die Festlegung, welche Mitarbeiter an den jeweiligen Unterweisungen teilnehmen müssen. Ein Assistent unterstützt den Abteilungsleiter bei dieser Konfiguration, indem er die Mitarbeiterliste zur Verfügung stellt und Filterungen nach Abteilungen

und Schichten ermöglicht. Die Auswahl und Zuordnung ganzer Mitarbeitergruppen kann in kurzer Zeit erfolgen.

3.3 Unterweisungsaufträge

Das ERP-System von Microsoft wickelt viele Aufgaben mithilfe von Arbeitsaufträgen ab. Auch das Unterweisungsmodul arbeitet nach diesem bekannten Vorgehen. Zu jedem fälligen Unterweisungstermin wird durch die jeweilige Fachabteilung ein Auftrag erstellt und abgearbeitet.

Die Datenablage erfolgt in zwei Abschnitten. Der sogenannte „Auftragskopf“ speichert alle relevanten und allgemeingültigen Daten einer Unterweisung. Die Beschreibungen, die Dokumente, den Termin und den Status. Die „Auftragszeilen“ wiederum beinhalten die spezifischen Teilnehmerinformationen, wie Personalnummer, den Namen und die Teilnahmebestätigung. Die Zeilen sind dabei an die Auftragsnummer des Kopfes gekoppelt. Durch das Löschen eines Auftragskopfes werden auch die dazugehörigen Zeilen mithilfe einer Löschweitergabe entfernt, um die Datenkonsistenz zu gewährleisten.

Die Ablage aller Unterweisungsdaten erfolgt als Kopie in der SQL-Datenbank. Dies ist erforderlich um nachträgliche Änderungen zu unterbinden. Veränderungen der Mitarbeiterliste dürfen die Unterweisungsdaten nicht beeinflussen. Beispielsweise kann sich der Nachname eines Mitarbeiters während der Beschäftigungsdauer ändern.

Im Unternehmen befinden sich immer wieder Personen, die in der Mitarbeiterliste nicht geführt werden, aber dennoch an Unterweisungen teilnehmen müssen. Darunter fallen Praktikanten oder Ferienhelfer, die an einer Erstunterweisung am Arbeitsplatz teilnehmen. Der Kursleiter kann diese Personen manuell im Unterweisungsauftrag eintragen. Durch Ändern des Mitarbeiternamens wird die Personalnummer gelöscht. Zusätzlich wird der Name überschrieben, wenn eine Personalnummer ausgewählt wird. So können falsche Zuordnungen im Unterweisungsauftrag zwischen der Personalnummer und dem Mitarbeiternamen verhindert werden, obwohl die Ergänzung manuell geschieht.

Jeder Auftrag im ERP-System verfügt über ein Statusfeld. Die Unterscheidung erfolgt zwischen „Offen“, „Geplant“ und „Beendet“. Neu angelegte Aufträge erhalten den Status „Offen“ um anzuzeigen, dass dieser noch bearbeitet wird. Sobald der Unterweisungstermin festgelegt wurde, wird der Status auf „Geplant“ umgestellt.

Der Abschluss eines Auftrages kann nur erfolgen, wenn der Kursleiter die Anwesenheit aller zugeordneten Mitarbeiter kontrolliert. Alle Teilnahmebestätigungen müssen als Dokument im Unterweisungsauftrag abgespeichert und die Person in der Liste abgehakt sein. Diese Hilfestellung soll den Überblick gewährleisten und fehlerhafte Teilnahmebuchungen verhindern. Dies hätte falsche Fälligkeitberechnungen zu Folge.

3.4 Unterweisungsassistent

Die Erstellung neuer Unterweisungsaufträge kann schnell und einfach durch einen bereitgestellten Unterweisungsassistenten erfolgen. Dieser kann auf Basis einer konkreten Unterweisungsvorlage gestartet werden und führt den Anwender durch die Neuanlage.

Die Schritte erfolgen nach Themengebieten sortiert nacheinander. Der erste Teil beschäftigt sich mit den allgemeinen Unterweisungsdaten. Die Vorlagennummer und Beschreibungsfelder sind bereits mit den Daten aus der Unterweisungsvorlage gefüllt. Zusätzlich muss der Kursleiter eingetragen werden, da dieser sich bei jedem Auftrag ändern kann. Im zweiten Schritt erfolgt der Eintrag des vorgesehenen Unterweisungstermins. Zuletzt wird die Mitarbeiterliste angezeigt. Durch die Auswahl der Filter nach Abteilungen und Schichten können auch mehrere Mitarbeiter schnell gefunden werden. Die Filterung der nach der Unterweisungsvorlage zugeordneten Beschäftigten, kann durch einen Mausklick erfolgen. Abschließend wird der neue Unterweisungsauftrag automatisch generiert, gespeichert und dem Benutzer vorgelegt. Der benutzerfreundliche Assistent kann neue Unterweisungsaufträge in nur wenigen Schritten befüllen und erzeugen.

3.5 Unterweisungsmatrix

Der Arbeitgeber und die Vorgesetzten müssen ständig den Überblick über die fälligen Unterweisungen behalten, damit die gesetzlichen Anforderungen erfüllt werden. Das Unterweisungsmodul muss die benötigten Informationen aufbereiten und bereitstellen.

Die Unterweisungsmatrix ist eine interaktive Ansicht aller fälligen Unterweisungstermine. Die Ansicht erfolgt in tabellarischer Form, wobei die Zeilen aus den einzelnen Unterweisungsvorlagen und die Spalten aus den einzelnen Monaten eines Jahres bestehen. Das angezeigte Zeitfenster kann durch den Benutzer spezifiziert werden, indem die Auswahl nach dem gewünschten Monat in der dritten Spalte erfolgt. Die Matrix zeigt die zwei Vormonate, das ausgewählte Monat und die darauf folgenden neun Monate an. Die Tabelleneinträge sind entweder leer oder mit Zahlen gefüllt. Die Zahl gibt Aufschluss darüber, wie viele Mitarbeiter an der jeweiligen Unterweisung in diesem Monat teilnehmen müssen. Durch Auswahl der gewünschten Teilnehmerzahl wird die Liste der erforderlichen Mitarbeiter aufgerufen.

Gewünschte Filter passen die Berechnungen an. Der Benutzer kann die einbezogenen Mitarbeiter auf deren Abteilungen und Schichten einschränken. Es können entweder nur fällige, oder auch erfolgreich durchgeführte Unterweisungstermine eingeblendet werden.

Eine eigenständige Liste informiert über alle angeforderten Erstunterweisungen, da die Berechnung der Fälligkeitstermine nicht möglich ist. Die Auflistung innerhalb des aktuellen Kalendermonates würde das Matrixergebnis erheblich verfälschen, weswegen diese unabhängig von den Folgeunterweisungen betrachtet werden.

3.6 Bericht ausstehender Unterweisungen

Alle fälligen Unterweisungen können auch durch einen Bericht dargestellt werden. Diese sind im ERP-System üblich, um Daten aufzubereiten und die Ergebnisse für den Benutzer auszugeben. Die Darstellung erfolgt auf Basis des DIN A4 Papierformats, wodurch der einfache Ausdruck der Informationen ermöglicht wird.

Berichte unterstützen auch die Verwendung von benutzerdefinierten Filtern, um die berechneten Daten entsprechend einzugrenzen. Der Unterweisungsbericht übernimmt Beschränkungen nach Abteilungen und Schichten. Die angegebenen Filter werden auch auf dem erstellten Papier angedruckt, um auf die Datengrundlage hinzuweisen.

Wie bei der Unterweisungsmatrix erfolgt die Ergebnisaufbereitung durch zwei eigenständige Tabellen. Diese können durch den Benutzer jeweils aktiviert oder deaktiviert werden. Die erste Tabelle kategorisiert die fälligen Folgeunterweisungen entsprechend den Vorlagen und zeigt dazu die fälligen Mitarbeiter mit dem dazugehörigen Unterweisungstermin an. Die zweite Tabelle listet die benötigten Erstunterweisungen mit den erforderlichen Teilnehmern auf.

3.7 E-Mail-Benachrichtigung

Das Unterweisungsmodul fungiert nicht nur als passive Informationsquelle, sondern ist eine aktive Komponente zur Unterstützung des Arbeitgebers und der verantwortlichen Abteilungsleiter. Sie erhalten automatische Benachrichtigungen, wenn fällige Unterweisungen anstehen. Der Zeitpunkt ist dabei in monatlichen Abstufungen frei wählbar. Dies ermöglicht den Aufbau einer Benachrichtigungshierarchie. Assistenzen können bereits vor den Abteilungsleitern benachrichtigt werden, um die Termine entsprechend zu planen und vorzubereiten. Mehrere Datensätze mit verschiedenen Zeitpunkten der Benachrichtigung ermöglichen diese Abstufung. Die Zusammenfassung der fälligen Unterweisungen erfolgt nach zugeordneter Abteilung des Mitarbeiters.

Microsoft Dynamics NAV 2016 stellt für automatische Prozesse einen sogenannten „NAV Application Server“ bereit. Dieser führt gewünschte Aufgaben zu einem festgelegten Zeitplan selbständig aus. Eine Benutzerinteraktion ist damit nicht erforderlich [4]. Zu jedem ersten Tag im Monat wird die Programmierung der E-Mail-Benachrichtigungen ausgeführt und diese damit bei Bedarf zugestellt.

Der pro Person fällige Unterweisungstermin ergibt sich anhand des letzten Auftrages und des Zeitintervalls aus der dazugehörigen Unterweisungsvorlage. Von diesem berechneten Fälligkeitsdatum muss der gewählte Benachrichtigungszeitpunkt abgezogen werden, um den Absendetermin der Benachrichtigung festzustellen. Liegt dieses Datum im aktuellen Monat, folgt die Ablage der fälligen Unterweisung mit den dazugehörigen Mitarbeiterdaten in einer temporären Tabelle. Die E-Mails werden nach der erfolgreichen Berechnung automatisch generiert. Die Kategorisierung und Zusammenstellung erfolgt nach den einzelnen Unterweisungsvorlagen, die in

der temporären Tabelle abgelegt worden sind. Die Formatierung mit HTML5 Steuerelementen verbessert die Lesbarkeit der Benachrichtigung. Jede Unterweisung erhält eine eigene Überschrift, mit anschließender Auflistung der jeweiligen Personalnummern und Namen der Mitarbeiter. Dazu wird die temporäre Tabelle nach der aktuellen Unterweisung gefiltert und die dadurch gefundenen Mitarbeiter Zeile für Zeile entnommen und der E-Mail hinzugefügt.

Die Abteilungsleiter erhalten die Informationen über die erforderlichen Folgeunterweisungen immer rechtzeitig, um diese nach Gesetzesvorgabe durchzuführen.

3.8 Mitarbeiteransicht

Alle relevanten Mitarbeiterinformationen können durch eine spezielle Ansicht abgerufen werden, die auf das Thema Unterweisungen zugeschnitten ist.

Alle betrieblichen Ressourcen sind im ERP-System zentral in einer Tabelle hinterlegt. Dort erfolgt die Speicherung der ausführlichen Mitarbeiterdaten. Doch nicht alle Informationen sind für die Unterweisungen relevant. Aufgrund des Datenschutzes bleiben zusätzliche Informationen ausgeblendet. Zeitgleich sind die mitarbeiter-spezifischen Unterweisungsdaten gesammelt darzustellen.

Aus diesen Gründen wird eine Mitarbeiteransicht entworfen, die nur die Anzeige der nötigsten Informationen ermöglicht. Dazu zählen die Personalnummer, der Name, die Abteilung und die Schicht. Alle anderen Daten aus der sogenannten Ressourcentabelle können nicht angezeigt werden.

Eine Liste informiert den Benutzer über alle erfolgreich abgeschlossenen Unterweisungen des jeweiligen Mitarbeiters. Eine chronologische Auflistung ermöglicht den schnellen Überblick und hilft dabei, die entsprechenden Unterweisungsaufträge zu finden. Dort hinterlegt sind alle benötigten Dokumente zur Unterweisung. Die Ausgabe der kompletten Historie an staatliche Behörden kann somit in kurzer Bearbeitungszeit erfolgen.

Eine weitere Tabelle informiert über die dem jeweiligen Mitarbeiter zugewiesenen Unterweisungsvorlagen. Dort erfolgt die Festlegung, welche Unterweisungen für den Beschäftigten vorgeschrieben sind. Sowohl die Fälligkeitsberechnungen aus den E-Mail-Benachrichtigungen, der Unterweisungsmatrix als auch des Berichts ausstehender Unterweisungen basieren auf diesen Zuordnungen.

Diese Ansicht sammelt alle relevanten Informationen eines Beschäftigten und bereitet sie zur Anzeige und zur Konfiguration auf.

3.9 Auftragshistorie

Die Historie abgeschlossener Unterweisungen kann auf mehrere Arten angezeigt werden.

Eine eigenständig aufrufbare Liste zeigt alle Unterweisungsaufträge mit dem Status „Beendet“. Diese Liste unterstützt die Eingrenzung nach gewünschten Datumsbereichen, Unterweisungen oder Beschreibungen. Dennoch

ist es schwierig die Teilnahmen eines bestimmten Mitarbeiters zu finden.

Aus diesem Grund zeigt die Mitarbeiteransicht alle abgeschlossenen Unterweisungen einer einzelnen Person an (siehe Mitarbeiteransicht).

Jedoch werden nicht alle Personen in der Ressourcentabelle geführt. Dort fehlen sowohl Ferienarbeiter als auch Praktikanten, wodurch diese Ansicht nicht aufrufbar ist. Aus diesem Grund existiert eine Tabelle, in der alle Teilnahmebestätigungen mit den dazugehörigen Unterweisungen einzeln verknüpft und aufgeführt sind. Diese Liste verfügt über sehr viele Einträge und soll daher als Suchmaske fungieren. Durch Tabellenfilter auf Namen und andere Informationen der Unterweisungen können damit alle Teilnehmer schnell gefunden werden.

3.10 Subscriber

Microsoft Dynamics NAV 2016 stellt sogenannte Publisher und Subscriber zur Verfügung. Diese dienen dem Informationsaustausch innerhalb der ERP-Software.

Die Publisher werden in der Programmierung aufgerufen und versenden dabei eine Nachricht an einen oder mehrere Subscriber.

Diese horchen auf eine spezifische Nachricht und führen daraufhin den eigenen Programmcode aus [5]. Microsoft entkoppelt damit die einzelnen Programmbausteine voneinander. Besonders hervorzuheben ist die Trennung der Standardprogrammierung zu weiteren Softwaremodulen von fremden Softwareherstellern oder Eigenentwicklungen.

Diese Funktionalität wird zur Wahrung der Datenbankkonsistenz verwendet. Das Unterweisungsmodul speichert die benötigten Daten in eigenen Tabellen und erweitert damit unter anderem die Ressourcentabelle aus dem ERP-System Standard. Sobald Mitarbeiterdaten aus dieser gelöscht werden, sind die verknüpften Daten auch aus den Modultabellen zu entfernen.

Mehrere Subscriber horchen dazu auf Löschvorgänge in den Standardtabellen. Durch deren Auslösung verwirft das Unterweisungsmodul die zugehörigen Daten automatisch. Eine weitere Benutzeraktion ist nicht erforderlich. Dies verhindert verwaiste Datensätze in den Tabellen des Unterweisungsmoduls.

Die benötigten Publisher sind seitens Microsoft bereits in der Standardprogrammierung hinterlegt. Somit entfällt der Einbau eigener Publisher.

Durch dieses Vorgehen sind keine Änderungen an der Standardprogrammierung erforderlich. Dadurch kann die ERP-Software jederzeit auf eine neuere Version migriert werden, ohne dass das Unterweisungsmodul oder die Standardprogrammierung erneut anzupassen sind. Auch andere Unternehmen können die Softwareerweiterung ohne Probleme mit einbinden.

4 LÖSUNG DATENABGLEICH

Die Personalsoftware Sage HR Suite bietet keine definierte Schnittstelle, um die Mitarbeiterdaten an eine andere Software zu übertragen. Daher muss der Zugriff auf die Daten direkt auf der SQL-Datenbank erfolgen. Durch Analyse dieser wird die Tabellenstruktur untersucht, um die benötigten Informationen zu finden. Abschließend erfolgt die Datenübernahme durch das ERP-System.

4.1 Sage HR Suite Datenbank

Die Personaldatenbank wird auf einem Microsoft SQL Server 2016 ausgeführt. Die dazu verfügbare kostenlose Software SQL Server Management Studio, ermöglicht den schnellen und einfachen Überblick in die Struktur der Datenbank. Voraussetzung dafür ist die Verwendung eines Microsoft SQL Servers ab Version 2008 bis Version 2019 [6].

Tabellennamen und Abkürzungen in der Datenbank sind in deutscher Sprache verfasst. Wichtige Bestandteile sind durch die selbsterklärenden Bezeichnungen schnell auffindbar. Die Übertragung beschränkt sich auf die für das Unterweisungsmodul notwendigen Informationen.

Für die Extrahierung der benötigten Daten sind vier verschiedene Tabellen von Bedeutung. Die Mitarbeiterinformationen liegen in einer Tabelle mit der Bezeichnung „Arbeitnehmer“. Das Modul benötigt daraus nur die Personalnummer und den Namen des Mitarbeiters. Filter auf das abgespeicherte Austrittsdatum verwerfen bereits ausgeschiedene Personen. Ein leeres oder ein zukünftiges Datum deutet auf die aktuelle Beschäftigung hin. Aussortiert werden auch Datensätze, die nicht explizit dem Mandanten des Unternehmens zugeordnet sind. Damit entfallen Einträge aus Teststellungen und Mitarbeiter-schulungen. Die dazu erforderliche Firmenbezeichnung befindet sich in einer Tabelle mit dem Namen „Mandant_OV“. Eine Join-Operation dieser ermöglicht die Filterung nach dem Namen des Unternehmens.

Die Abbildung der Organisationsstruktur erfolgt in der Tabelle „Orga_Stamm“. Dort abgelegt sind die einzelnen Abteilungen, Schichten, sowie die Hierarchie des Unternehmens. Zur Übertragung vorgesehen ist nur die Organisationsnummer. Diese enthält sowohl die Abteilungs-, als auch die Schichtbezeichnung als Kürzel.

Eine Zuordnungstabelle mit dem Namen „Orga_AN_Zuord“ verknüpft die Mitarbeiterdaten zum Organisationsstamm. Dort enthalten sind mehrere Einträge zu jeder einzelnen Person, um die komplette Strukturzuordnung abzubilden. Die einzelnen Datensätze sind mit einem Zeitfenster versehen, welches vom SQL-Server mit dem aktuellen Datum abgeglichen wird. Veraltete Einträge werden somit bei der Abfrage ausgeschlossen.

Der abgesetzte SQL-Befehl extrahiert alle derzeit aktiven Mitarbeiter, verknüpft mit dem dazugehörigen Organisationsstamm. Die weitere Verarbeitung erfolgt während der Synchronisierung der beiden Datenbanken.

4.2 Microsoft Dynamics NAV 2016 mit externer Datenbank verbinden

Ohne Erweiterung bietet das ERP-System keine Möglichkeit, sich mit einer externen Datenbank zu verbinden. Jedoch werden die Einbindung und die Verwendung von .NET-Framework Funktionen unterstützt. Diese können als Umgebungsvariablen in der Programmierung angelegt und aufgerufen werden. Das Unterweisungsmodul verwendet diese .NET Funktionen, um den externen Datenbankzugriff zu ermöglichen.

Die Ausführung der ERP-Software erfolgt durch ein eigenes Dienstkonto auf dem Serversystem. Dadurch authentifiziert sich die Software mit diesem Konto gegenüber anderen Computern und Diensten im Netzwerk. Der Server der Personaldatenbank muss Lesevorgänge durch dieses Dienstkonto gestatten. Der Serveradministrator muss das Leserecht auf die vier benötigten Tabellen aus dem vorherigen Kapitel vergeben.

Nachdem die Berechtigungen korrekt gesetzt sind, kann der SQL-Befehl durch die .NET-Framework Funktionen vorbereitet und abgesetzt werden. Die Datenbankantwort erfolgt allerdings in Datentypen aus dem .NET-Framework. Sie müssen in die Datentypen des ERP-Systems konvertiert werden [7]. Der Funktionsaufruf erfolgt für jede Spalte der Personaldatenbank einzeln. Abschließend wird der extrahierte Wert Variablen des ERP-Systems zugewiesen.

Die Daten stehen somit zur weiteren Verarbeitung durch das ERP-System bereit. Dies wird im nächsten Kapitel betrachtet.

4.3 Personaldatenabgleich

Die durch die Personaldatenbank bereitgestellten Informationen müssen mit der Ressourcentabelle der ERP-Datenbank synchronisiert werden. Der Vergleich beruht auf Detailinformationen, da die Änderungen für die Benachrichtigungen an die Abteilungsleiter relevant sind.

Der Datenabgleich erfolgt Zeile für Zeile aus der Antwort der Personaldatenbank. In den übertragenen Daten befinden sich jedoch mehrere Datensätze pro Mitarbeiter, da der abgesetzte SQL-Befehl alle einzelnen Zugehörigkeiten des Organisationsstammes ausgibt. Zwei Übersetzungstabellen helfen der Programmierung dabei, die Organisationsnummern aus der Datenbankantwort den einzelnen Abteilungen und Schichten der ERP-Software zuzuordnen. Findet sich die Organisationsnummer innerhalb der beiden Tabellen, handelt es sich um die unterste Schicht der Firmenorganisation und damit um den jeweiligen Arbeitsplatz des Mitarbeiters. Existiert die Organisationsnummer nicht in den Übersetzungstabellen, handelt es sich um eine übergeordnete Organisationsschicht des Unternehmens. Der Datensatz wird im ERP-System nicht benötigt und kann ohne genauere Betrachtung verworfen werden.

Der Datenabgleich erfolgt in mehreren Einzelschritten, um die exakten Unterschiede der Datensätze festzustellen.

Im ersten Schritt wird geprüft, ob die aktuell zu bearbeitende Personalnummer in der ERP-Datenbank gefunden

werden kann. Schlägt die Suche fehl, handelt es sich um einen neuen Mitarbeiter, der in die Ressourcentabelle eingetragen werden muss. Anschließend wird eine E-Mail an den Vorgesetzten gesendet, wodurch er über den neuen Eintrag informiert wird. Die weitere Prozedur bricht in diesem Fall ab.

Erzielt die Suche jedoch einen Treffer, erfolgt der Vergleich der Abteilung und der Schichtzugehörigkeit. Unterschiede haben die Aktualisierung der ERP-Datenbank und die Benachrichtigung per E-Mail an den entsprechenden Abteilungsleiter zur Folge.

Das ERP-System unterstützt temporäre Markierungen der einzelnen Datensätze während der Ausführung eines Programmabschnittes. Die Speicherung der Markierungen ist jedoch nicht möglich. Während dem Datenabgleich werden alle betrachteten Datensätze der ERP-Datenbank markiert. Nach erfolgreicher Abarbeitung aller Zeilen aus der Personaldatenbank, verfügen ausschließlich aktive Mitarbeiter über eine Markierung. Alle anderen Datensätze sind veraltet und werden abschließend aus der ERP-Datenbank gelöscht.

Durch die Synchronisierung der Datenbestände muss die manuelle Datenpflege lediglich in der Personaldatenbank erfolgen. Das ERP-System übernimmt die Informationen automatisch und ohne Benutzerinteraktion.

Nicht nur das Unterweisungsmodul benötigt die aktuellen Daten um die Anforderungen korrekt zu erfüllen, sondern auch andere Module sind darauf angewiesen. Buchungsvorgänge der Serviceaufträge oder Produktionsaufträge hinterlegen ebenfalls den verantwortlichen Mitarbeiter. Die Datensynchronisierung vereinfacht und beschleunigt somit auch andere Geschäftsprozesse im Unternehmen.

5 EVALUATION

Die Evaluierung des Unterweisungsmoduls wird in mehrere Kriterien aufgeteilt, welche im folgenden Abschnitt im Detail erläutert werden. Das Kernthema bildet die Einsparung von Arbeitszeit sowohl bei den Abteilungsleitern, als auch bei der Geschäftsführung. Andere Aspekte, wie die Reduzierung menschlicher Fehler können jedoch nur qualitativ betrachtet werden, da die Fehlerraten nicht einfach zu ermitteln sind. Das Unterweisungsmodul wird noch nicht produktiv ausgeführt, weswegen die Daten durch ein Testsystem erfasst wurden. Dieses wird parallel zu den alten Geschäftsprozessen betrieben, um die Funktionalität zu testen und die Mitarbeiter an den Umgang zu gewöhnen.

5.1 Wiederkehrende Aufgaben

Es folgt die Betrachtung von Aufgaben, die regelmäßig durch das Personal wiederholt werden müssen. Diese erfordern die meiste Arbeitszeit und bieten daher das höchste Einsparungspotential.

5.1.1 Einarbeiten neuer Arbeitsanweisungen und Gesetze in die Unterweisungsvorlagen

Die Einarbeitung neuer Gesetze und Arbeitsanweisungen in bestehende Unterweisungsvorlagen ist immer wieder

in unregelmäßigen Zeitabständen durchzuführen. Der Arbeitsaufwand hängt dabei stark vom Umfang der Änderungen ab. Diese sind bei jedem Änderungsauftrag individuell und können nicht explizit aufgelistet werden. Bei der alten Arbeitsweise musste jede Abteilung die vorgenommenen Änderungen in die eigenen Vorlagen übernehmen.

Die Überarbeitung der Unterweisungsvorlagen kann in erster Linie nicht beschleunigt oder vereinfacht werden, da das Unterweisungsmodul an diesem Prozess keine Veränderung bringt.

Jedoch wird jetzt die Möglichkeit geschaffen, Gesetzesänderungen an einer zentralen Unterweisungsvorlage für alle betroffenen Abteilungen und Vorgesetzten einzuarbeiten. Diese Arbeit muss somit nicht durch mehrere Personen einzeln vorgenommen werden.

Als Beispiel wird die Abteilung Produktion herangezogen. Diese ist in 4 Schichten aufgeteilt, wobei jeder Schichtführer seine eigenen Mitarbeiter unterweisen muss. Die Schichtführer sind dafür verantwortlich, die Unterweisungen termingerecht und nach den neuesten Regelungen durchzuführen. Jedoch unterscheiden sich im alten System die Unterweisungsvorlagen von jedem Vorgesetzten, da diese jeweils einem anderen Rhythmus folgen. Damit ist jeder Schichtführer gezwungen, die eigenen Vorlagen bei einem anstehenden Termin zu überarbeiten.

Durch die Zentralisierung des Unterweisungsmoduls entfällt nun die mehrfache Durchführung dieser Tätigkeit. Nur noch der Abteilungsleiter muss die Vorlage anpassen, welche automatisch in alle neuen Unterweisungsaufträge übernommen wird. Im Fall der Produktion kann diese Aufgabe damit um den Faktor 4 beschleunigt werden. Ähnliche Beschleunigungsfaktoren finden sich auch in anderen Bereichen des Unternehmens wieder.

Feststellen fälliger Unterweisungen

Jeden Monat müssen die einzelnen Abteilungsleiter überprüfen, welche Unterweisungen im aktuellen Monat durchzuführen sind. Dabei sind auch die erforderlichen Mitarbeiter festzustellen.

Diese Tätigkeit findet im alten System in der Regel zu jedem Monatsbeginn statt und erfordert je nach Abteilungsgröße etwa eine Arbeitsstunde.

Das Unterweisungsmodul als aktive Komponente übernimmt diese Aufgabe vollständig autonom, indem es den Abteilungsleiter per E-Mail-Benachrichtigung informiert. Alternativ können die Informationen über die Unterweisungsmatrix oder dem Bericht ausstehender Unterweisungen binnen weniger Minuten abgerufen werden. Damit entfällt die manuelle Aufgabe des Abteilungsleiters nahezu vollständig. Pro Abteilung des Unternehmens werden so 12 Arbeitsstunden im Jahr eingespart.

5.2 Dokumentation durchgeführter Unterweisungen

Die Gesetzesgrundlage fordert die lückenlose Dokumentation durchgeführter Unterweisungen. Auch in diesem Bereich ist das Unterweisungsmodul anzutreffen und unterstützt die Tätigkeiten der verantwortlichen Mitarbeiter.

Zuvor wurden die Unterweisungsdokumente und Teilnahmebestätigungen auf einem Netzwerklaufwerk des Unternehmens gespeichert. Zusätzlich wünschen sich die verantwortlichen Mitarbeiter die Ablage in Papierformat, um gegen Datenverlust abgesichert zu sein.

Das Unterweisungsmodul übernimmt die Aufgabe der digitalen Dokumentenablage innerhalb des ERP-Systems und löst damit das Netzwerklaufwerk als Speicherort ab. Die physikalische Archivierung wird dennoch weiterhin gewünscht.

Daher ändert sich nur der digitale Ablageort der Unterweisungsdokumente, jedoch nicht der dazu erforderliche Arbeitsaufwand, um die Dokumente zu digitalisieren. Somit kann der Vorgang nicht maßgeblich im Bereich der Zeiteinsparung verbessert werden.

Jedoch können bei dieser Tätigkeit menschliche Fehler verringert werden. Durch den Abschluss eines Unterweisungsauftrages kontrolliert das Modul, ob die Teilnahme aller Mitarbeiter bestätigt wurde. Dies verhindert effektiv, dass abwesende Personen übersehen werden und die Unterweisung nicht erhalten. Somit wird zusätzlich sichergestellt, dass die gesetzlichen Vorgaben eingehalten werden.

Die Benutzer des Testsystems bestätigen die Verringerung der menschlichen Bearbeitungsfehler und der damit einhergehenden Arbeitserleichterung.

5.3 Dokumentenanfragen

Die Dokumentenanfragen von staatlichen Behörden sind ein wichtiges Thema für das Unterweisungsmodul, welche in kürzester Zeit bearbeitet werden müssen. Nach einem Arbeitsunfall sind alle Unterweisungsunterlagen dieser Person auszuhändigen. Dieser Fall wird für die Evaluierung betrachtet.

Das Auffinden und Zusammenstellen einer Mitarbeitermappe durch das alte Ablagesystem dauert gemäß Erfahrungswerten zwischen 30 und 60 Minuten. Dies ist dem geschuldet, da die Dokumente je nach Unterweisungsthema an verschiedenen Orten aufbewahrt wurden. In manchen Fällen erfolgte die Ablage ausschließlich in Papierform, weswegen die Aktenschränke bei jeder Anfrage händisch durchsucht wurden.

Die zentralisierte Mitarbeiteransicht des Unterweisungsmoduls kann diese Aufgabe in kurzer Zeit erledigen. Durch die Auflistung aller teilgenommenen Unterweisungen eines Mitarbeiters wird die Übersicht direkt zur Verfügung gestellt. Die Suche nach einem Mitarbeiter im Modul wird mit etwa 2 Minuten bemessen. Das Ausdrucken der Unterweisungsdokumente kann anschließend mit einem Zeitaufwand von etwa 5-10 Minuten erfolgen. Damit ergibt sich im schlechtesten Fall ein Zeitbedarf von 12 Minuten.

Als Vergleichswerte werden der bestmögliche Fall des alten Ablagesystems und der schlechtmögliche Fall des Unterweisungsmoduls herangezogen. Selbst bei Betrachtung dieses negativen Extremfalls, kann die Bearbeitungszeit um 18 Minuten verkürzt werden. Dies ergibt eine Zeiteinsparung von 60%.

Die Dokumente werden jedoch nur zu bestimmten Ereignissen durch Behörden oder der Geschäftsführung angefordert, weswegen keine aussagekräftigen Langzeit-Metriken dargestellt werden können.

Das Unterweisungsmodul bietet für den Arbeitgeber große Einsparungspotentiale bei der Arbeitszeit und verhindert effektiv menschliche Fehler im Umgang mit den Unterweisungen.

Die oben genannten Aspekte zeigen, dass sich die Zeit- und Geldinvestition in die Entwicklung des Unterweisungsmoduls für den Unternehmer lohnt.

6 ZUSAMMENFASSUNG

Unterweisungen sind in jedem Unternehmen ein wichtiges Thema, um die Mitarbeiter über Gefahren im Betrieb aufzuklären. Die Sicherheit der Mitarbeiter hat oberste Priorität und darf keinesfalls vernachlässigt werden. Dies wird durch gesetzliche Regelungen bestätigt, die den Arbeitgeber zu regelmäßigen Unterweisungen verpflichten. Es ist darauf zu achten, dass jeder Beschäftigte an den erforderlichen Schulungen teilnimmt. Für die Unternehmen bedeutet dies nicht nur Zeitaufwand während der Unterweisungstermine, sondern auch hohen organisatorischen Verwaltungsaufwand. Damit einhergehen die Personalkosten für die aufgewendete Arbeitszeit der Beschäftigten, insbesondere der verantwortlichen Leitungsebene.

Bereits in mittelständischen Unternehmen kann es daher von Vorteil sein, die Arbeit mit den Unterweisungen zu digitalisieren und zu automatisieren. Dabei bietet sich häufig das bereits verwendete ERP-System an, welches durch Softwarekomponenten erweitert werden kann.

Das entwickelte Unterweisungsmodul ist in der Lage, die Geschäftsprozesse weitestgehend zu automatisieren. Darunter fällt sowohl die Dokumentation vergangener Unterweisungen, als auch die Historie einzelner Mitarbeiter. Die hinterlegten Unterweisungsvorlagen stehen an zentraler Stelle zur Verwaltung und Bearbeitung bereit. Die Einarbeitung von Gesetzesänderungen und neuen Normen in die Unterweisungsdokumente kann in kurzer Zeit erfolgen. Neue Unterweisungsaufträge erhalten diese Änderungen automatisch. Alle neuen Schulungen erfolgen so auf Basis der aktuellsten Dokumente und Vorschriften. Als wichtigster Bestandteil ist die Informationsgewinnung durch das Unterweisungsmodul zu benennen. Der Arbeitgeber kann in wenigen Schritten sowohl vergangene Unterweisungsdaten abrufen, als auch alle demnächst fälligen Folgeunterweisungen ausgeben. Automatische Benachrichtigungen an die verantwortlichen Personen sparen zusätzliche Arbeitszeit ein, während darüber hinaus menschliche Bearbeitungsfehler auf ein Minimum reduziert werden. Die durch staatliche Behörden angeforderten Dokumente sind schnell abrufbar und zur Übergabe bereitgestellt. Sowohl der interne, als auch externe Kommunikationsaufwand wird deutlich reduziert.

Die Verwendung eines Unterweisungsmoduls im ERP-System ist ein weiterer Schritt in die Digitalisierung des Unternehmens. Dies bedeutet für den Arbeitgeber weitere vereinfachte Geschäftsprozesse, die nicht nur

menschliche Fehler reduzieren, sondern auch Arbeitszeit des Personals einsparen. Bereits für mittelständische Unternehmen ist die Zeit- und Kosteninvestition in ein Unterweisungsmodul durchaus lohnenswert.

LITERATUR

- [1] Bundesrepublik Deutschland: Gesetz über die Durchführung von Maßnahmen des Arbeitsschutzes zur Verbesserung der Sicherheit und des Gesundheitsschutzes der Beschäftigten bei der Arbeit (Arbeitsschutzgesetz – ArbSchG) URL: http://www.gesetze-im-internet.de/arbSchG/_12.html Abruf am 12. Juli 2022
- [2] Deutsche Gesetzliche Unfallversicherung e.V. (2012): Unterweisung – Bestandteil des betrieblichen Arbeitsschutzes.
- [3] Gayer Michaela (2008): Einführung in die Programmierung von Microsoft Dynamics NAV 5.0. MBS-Training, S. 7.
- [4] Microsoft Corporation (2022): Configuring NAS Services. URL: <https://docs.microsoft.com/en-us/dynamics-nav/configuring-nas-services> Abruf am 28. Juli 2022
- [5] Microsoft Corporation (2022): Subscribing to Events. URL: <https://docs.microsoft.com/en-us/dynamics-nav/subscribing-to-events> Abruf am 12. August 2022
- [6] Microsoft Corporation (2022): Herunterladen von SQL Server Management Studio (SSMS). URL: <https://docs.microsoft.com/de-de/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver16> Abruf am 15. Juli 2022
- [7] Microsoft Corporation (2022): Mapping Between .NET Framework and C/AL Types. URL: <https://docs.microsoft.com/en-us/dynamics-nav/mapping-between-.net-framework-and-c-al-types> Abruf am 16. August 2022

AUTOREN

Patrick Hilgart absolvierte zunächst eine Ausbildung als Elektroniker für Betriebstechnik im Unternehmen Pfeiderer Teisnach GmbH & Co. KG. Nach erfolgreichem Berufsabschluss folgte von 2018 bis 2022 ein duales Studium im Fachbereich Informatik an der OTH-Regensburg. Seither ist er im Unternehmen in den Bereichen ERP-Systementwicklung und Rechenzentrum tätig.

Professor Dr. Frank Herrmann wurde in Münster geboren und studierte Informatik an der Rheinisch-Westfälischen Technischen Hochschule in Aachen. Nach seinem Diplom 1989 arbeitete er bei dem Fraunhofer Institut IITB in Karlsruhe. Während dieser Zeit promovierte er 1996 über Ressourcenbelegungsplanungsprobleme. Von 1996 bis 2003 arbeitete er für die SAP AG in verschiedenen Funktionen, zuletzt als Direktor. Im Jahr 2003 wurde er Professor für Produktionslogistik an der Ostbayerischen Technischen Hochschule in Regensburg. Seine Forschungsthemen sind Planungsalgorithmen, Optimierung und Simulation für die operative Produktionsplanung und -steuerung. Er ist Leiter des Innovations- und Kompetenzzentrums für Produktionslogistik und Fabrikplanung (IPF).

HINREICHENDE SPEZIFIKATION EINER KI-FUNKTION AM BEISPIEL FUßGÄNGERSCHUTZ

Bernhard Knie

Fachbereich Industrial Engineer-
ing

**Università degli studi di Roma
Tor Vergata**
Via Cracovia n.50
00133 Roma
E-Mail: b.knie@aol.com

Prof. Stefan Kubica

Fachbereich Wirtschaftsinforma-
tik

**Technische Hochschule
Wildau**
Hochschulring 1
15745 Wildau
E-Mail: skubica@th-wildau.de

ABSTRAKT

Transparenz und Erklärbarkeit sind essenzielle Faktoren bei der Entwicklung sicherer KI-Modelle für den Fußgängerschutz. Dies ermöglicht nicht nur eine verbesserte Akzeptanz, sondern fördert auch das Vertrauen der Nutzer in die präventiven Fähigkeiten dieser Systeme.

Die Integration von Sensoren, maschinellem Lernen und fortschrittlichen Algorithmen ermöglicht eine präzise Analyse von Fußgängerverhalten und Umgebungsfaktoren. Dabei spielt die Echtzeitverarbeitung eine entscheidende Rolle, um schnell auf sich ändernde Situationen zu reagieren.

Zusätzlich zur Technologieentwicklung betont diese Analyse die Notwendigkeit strikter Sicherheitsmaßnahmen, um potenzielle Risiken und Sicherheitslücken zu minimieren. Dazu gehören Sicherheitsprotokolle für die Datenspeicherung, den Modellschutz und den Zugriff auf KI-Systeme.

Die Einhaltung gesetzlicher und ethischer Richtlinien ist ein weiterer wichtiger Aspekt. Regulatorische Vorgaben und ethische Standards sollen sicherstellen, dass KI-Technologien im Automobilsektor nicht nur sicher, sondern auch im Einklang mit gesellschaftlichen Werten und Normen eingesetzt werden.

Die Bedeutung von sicherer KI im Kontext des Fußgängerschutzes im Automobilsektor wird weiter wachsen. Durch die richtige Balance zwischen Technologieentwicklung, Sicherheitsmaßnahmen und regulatorischer Compliance kann Künstliche Intelligenz dazu beitragen, die Verkehrssicherheit zu verbessern und Unfälle zu minimieren.

SCHLÜSSELWÖRTER

KI-Modelle, Automotive, Fußgängerschutz
1 HINTERGRUND UND MOTIVATION

Die Entwicklung von KI-Funktionen, abgebildet durch tiefe neuronale Netze für den Fußgängerschutz ist eine Reaktion auf die zunehmende Komplexität und Vielfalt von Verkehrssituationen, insbesondere in städtischen Gebieten. Traditionelle Bilderkennungsalgorithmen haben Schwierigkeiten, die Vielfalt menschlicher Bewegungen und die komplexe Umgebung angemessen zu erfassen. Tiefe neuronale Netze, insbesondere Convolutional Neural Networks (CNNs), haben sich als leistungsfähige Werkzeuge erwiesen, um die Herausforderungen im Fußgängerschutz zu bewältigen. Tiefe neuronale Netze können hierarchisch abstrahierte Merkmale aus Bildern extrahieren. In Bezug auf den Fußgängerschutz ermöglicht dies eine automatische Lernfähigkeit, um komplexe Muster und Merkmale, die für die Erkennung von Fußgängern relevant sind, zu identifizieren. Zudem ermöglichen sie eine bessere Bewältigung von Varianzen. Fußgänger können in unterschiedlichen Umgebungen, bei verschiedenen Lichtverhältnissen und in unterschiedlichen Positionen auftreten. Tiefe neuronale Netze können lernen, mit dieser Varianz umzugehen und robuste Modelle zu entwickeln, die in verschiedenen Szenarien gut funktionieren. [1]

Ein weiterer Aspekt ist das End-to-End-Lernen, bei dem das Modell direkt aus den Rohdaten lernt, ohne manuell definierte Merkmale. Dies ist besonders vorteilhaft, wenn komplexe Beziehungen zwischen den Eingabeinformationen und der Fußgängererkennung schwer vorhersehbar sind. Schlüssel zum Erfolg sind jedoch vor allem große, zusammenhängende Datensätze. Tiefe neuronale Netze profitieren von großen Trainingsdatensätzen. Mit der Zunahme verfügbarer Daten können tiefe Modelle besser

generalisieren und Muster erkennen, was zu einer verbesserten Leistung in der Fußgängererkennung führt. Die rasante Entwicklung im Bereich der tiefen neuronalen Netze wird durch kontinuierliche Forschung vorangetrieben. Neue Architekturen, Optimierungstechniken und Trainingsstrategien ermöglichen eine ständige Verbesserung der Leistung von Modellen im Fußgängerschutz.

2 PROBLEMSTELLUNG

Der Einsatz von Künstlicher Intelligenz im Bereich des Fußgängerschutzes bringt zwar vielversprechende Fortschritte hinsichtlich der Verkehrssicherheit, wirft jedoch auch einige Herausforderungen und Problemstellungen auf, die sorgfältig berücksichtigt werden müssen:

Unvorhersehbare menschliche Interaktionen:

Fußgänger verhalten sich oft weniger vorhersehbar als andere Verkehrsteilnehmer. Ihre Bewegungen können durch verschiedene Faktoren wie Ablenkung, unerwartete Handlungen oder das Überqueren außerhalb von Fußgängerüberwegen geprägt sein. KI-Funktionen müssen in der Lage sein, diese unvorhersehbaren Verhaltensweisen zu erkennen und darauf angemessen zu reagieren.

Vielfalt der Umgebungsfaktoren:

Fußgänger bewegen sich in unterschiedlichsten Umgebungen, sei es in städtischen Gebieten, ländlichen Regionen oder auf stark frequentierten Straßen. Die Vielfalt der Umgebungsfaktoren, wie unterschiedliche Lichtverhältnisse, Wetterbedingungen und Straßentypen, stellt eine Herausforderung für die KI dar, die Modelle müssen robust gegenüber solchen Variablen sein.

Datenschutz und Ethik:

Die Verarbeitung von Informationen über Fußgänger, insbesondere durch KI-Funktionen in vernetzten Fahrzeugen, wirft Datenschutz- und Ethikfragen auf. Wie werden die Daten gesammelt, gespeichert und genutzt? Wie wird die Privatsphäre der Fußgänger geschützt? Es ist wichtig, einen ausgewogenen Ansatz zu finden, der sowohl die Sicherheit als auch die Privatsphäre respektiert.

Fehlende Normen und Standards:

Im Bereich der KI für den Fußgängerschutz fehlen oft einheitliche Normen und Standards. Die Entwicklung von Richtlinien und Normen ist entscheidend, um eine konsistente und sichere Implementierung von KI-Systemen zu gewährleisten und das Vertrauen der Öffentlichkeit zu stärken.

Adversarial Attacks:

KI-Modelle im Fußgängerschutz könnten anfällig für Adversarial Attacks sein, bei denen absichtlich gestaltete Eingaben die KI dazu verleiten, falsche Entscheidungen zu treffen. Der Schutz vor solchen Angriffen erfordert fortschrittliche Sicherheitsmaßnahmen.

Haftungsfragen und rechtliche Aspekte:

Im Falle von Unfällen oder Fehlfunktionen von KI-gesteuerten Fahrzeugen stellt sich die Frage der Haftung. Klare rechtliche Rahmenbedingungen und Verantwortlichkeiten müssen definiert werden, um Rechtsunsicherheiten zu minimieren.

Die Lösung dieser Problemstellungen erfordert eine multidisziplinäre Herangehensweise, die technologische, ethische, rechtliche und soziale Aspekte gleichermaßen berücksichtigt. Es ist entscheidend, dass Entwickler, Forscher, Gesetzgeber und die Gesellschaft gemeinsam daran arbeiten, sichere und verantwortungsbewusste KI-Systeme im Bereich des Fußgängerschutzes zu gewährleisten.

3 ZIELSETZUNG

Das übergeordnete Ziel besteht darin, eine robuste und verlässliche KI-Funktion zu entwickeln, die den Fußgängerschutz in Verkehrsszenarien optimiert. Die Zielsetzung umfasst mehrere spezifische Aspekte. Elementar ist zunächst eine präzise Fußgängererkennung, bei der die KI-Funktion in der Lage ist, Fußgänger unabhängig von deren Position, Bewegung oder Umgebungsfaktoren zu erkennen. Dies gewährleistet eine zuverlässige Grundlage für weitere Sicherheitsmaßnahmen. Desweiteren muss eine verlässliche Verhaltensvorhersage sichergestellt werden. Die KI soll in der Lage sein, das Verhalten von Fußgängern akkurat zu modellieren und vorherzusagen. Dies ermöglicht es, potenziell gefährliche Situationen frühzeitig zu identifizieren und angemessen darauf zu reagieren. Ein weiterer Aspekt ist die Echtzeit-Reaktionsfähigkeit.

Die KI-Funktion muss in Echtzeit auf sich ändernde Verkehrssituationen zu reagieren, insbesondere wenn Fußgänger unvorhersehbares Verhalten zeigen. Eine kurze Reaktionszeit und hohe Präzision ist entscheidend, um Kollisionen zu vermeiden. Zudem muss die Bilderkennung Fehlerresistenz und Robustheit aufweisen. Insbesondere die Objekterkennung soll robust gegenüber Fehlern und Unsicherheiten sein. Dies beinhaltet die Fähigkeit, auch in herausfordernden Umgebungen wie schlechtem Wetter oder ungünstigen Lichtverhältnissen zuverlässig zu funktionieren. Erreicht werden kann dies nur mithilfe einer Sensorfusion für umfassende Wahrnehmung. Daher müssen verschiedene Sensorinformationen, wie Kamera, Radar und Lidar, effektiv fusionieren, um eine umfassende Wahrnehmung der Verkehrsumgebung und der Fußgänger zu gewährleisten.

Schlussendlich muss die KI-Funktion automatisch und präzise Notbremsungen oder andere Kollisionsvermeidungsmaßnahmen einleiten, wenn eine potenzielle Gefahr für Fußgänger erkannt wird. Ein weiterer Aspekt ist die Anpassungsfähigkeit an verschiedene Verkehrsszenarien. Die entwickelte KI-Funktion soll sich nahtlos an verschiedene Verkehrsumgebungen anpassen können, darunter städtische Gebiete, ländliche Straßen oder komplexe Kreuzungen unter Einhaltung ethischer und rechtlicher Standards.

Die Zielsetzung umfasst auch die Entwicklung der KI-Funktion unter Berücksichtigung ethischer Prinzipien und geltender rechtlicher Standards. Datenschutz und Privatsphäre sollen dabei ebenso berücksichtigt werden wie die Einhaltung gesetzlicher Vorgaben. Durch die Verfolgung dieser Zielsetzung streben wir die Schaffung einer innovativen und sicheren KI-Funktion im Fußgängerschutz an, die die Verkehrssicherheit erhöht und gleichzeitig ethische und rechtliche Anforderungen erfüllt.

4 EXEMPLARISCHE UMSETZUNG DER FUNKTION

Die Fußgängererkennung durch Künstliche Intelligenz ist ein komplexer Prozess, der Bildverarbeitung, maschinelles Lernen und neuronale Netze integriert. Zunächst muss die Funktion mit der Erfassung von Bilddaten beginnen. Dies kann von verschiedenen Sensoren wie Kameras aufgenommen werden, die am Fahrzeug angebracht sind. Die aufgenommenen Bilddaten durchlaufen anschließend eine Datenpräverarbeitug. Hierzu gehören Schritte wie Rauschunterdrückung, Kontrastanpassung und möglicherweise die Normalisierung der Helligkeit. Ein Objekterkennungsalgorithmus wird auf die vorverarbeiteten Bilddaten angewendet. Dieser Algorithmus identifiziert potenzielle Objekte im Bild, einschließlich Fußgänger. Basierend auf den erkannten Objekten wird eine Region-of-Interest (ROI) um jeden potenziellen Fußgänger definiert. Dies hilft, den Fokus auf relevante Bereiche des Bildes zu legen und die Verarbeitung zu optimieren. Die Merkmale der erkannten Objekte, insbesondere der Fußgänger, werden extrahiert. Dies kann das Aussehen, die Größe, die Form und andere relevante visuelle Merkmale umfassen. Darauf folgend werden die extrahierten Merkmale einem trainierten neuronalen Netzwerk zugeführt.

Ein tiefes neuronales Netzwerk (Deep Neural Network, DNN) ist hierbei typisch. Das Netzwerk wurde zuvor mit großen Datensätzen von Fußgänger- und Nicht-Fußgängerbildern trainiert. Das neuronale Netzwerk verwendet Aktivierungsfunktionen und gewichtete Verbindungen zwischen Neuronen, um Entscheidungen darüber zu treffen, ob das erkannte Objekt tatsächlich ein Fußgänger ist. Entscheidend ist hier, dass die Ausgabe des neuronalen Netzwerks auf einen Schwellenwert angewendet wird, um zu bestimmen, ob ein erkanntes Objekt als Fußgänger klassifiziert wird. Dies hilft, Fehlklassifikationen zu minimieren. In einigen Anwendungen kann die Fußgängererkennung auch mit Daten anderer Sensoren kombiniert werden, um die Position der Fußgänger im dreidimensionalen Raum zu bestimmen. Dies ist insbesondere für autonome Fahrzeuge wichtig. Basierend auf den Ergebnissen der Fußgängererkennung trifft das Fahrzeug Entscheidungen und reagiert entsprechend. Dies kann das Auslösen eines Bremsvorgangs, eine Anpassung der Geschwindigkeit oder das Einleiten von Ausweichmanö-

vern umfassen. Abbildung 1 skizziert den Funktionsablauf zusammenfassend und gibt einen Überblick über die grundlegenden Schritte in der Anwendung einer KI-Fußgängererkennungsfunktion.

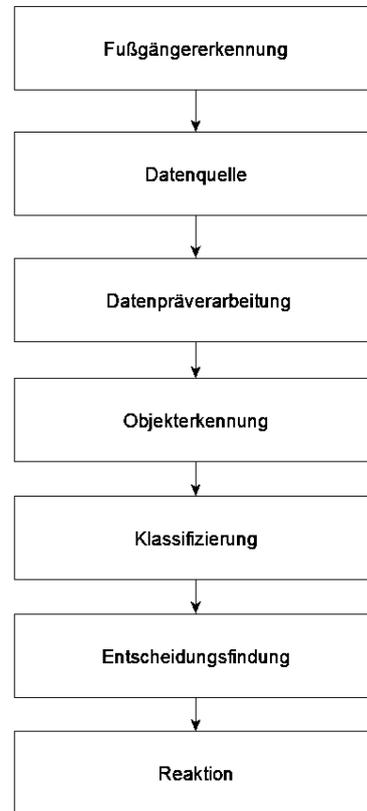


Abbildung 1: Hierarchischer Ablauf der KI-Funktion zur Fußgängererkennung.

Die spezifische Implementierung kann je nach den Anforderungen der Anwendung und den verfügbaren Sensordaten variieren.

5 GOAL STRUCTURING NOTATION

Die Goal Structuring Notation (GSN) ist eine Methode, die in der Entwicklung sicherheitskritischer Systeme verwendet wird, um die rationale Struktur von Argumentationen für sicherheitsrelevante Ziele und Entscheidungen zu modellieren. Die GSN bietet eine klare und strukturierte Darstellung von Sicherheitsargumentationen. Durch die Verwendung von Diagrammen und Notationen wird die Komplexität reduziert, was zu einer leichteren Verständlichkeit für alle Beteiligten führt. Dabei hilft sie den Argumentationsprozess transparent zu machen. Durch die sichtbare Darstellung von Argumenten und den Beziehungen zwischen verschiedenen Sicherheitszielen und -entscheidungen wird es einfacher, die Sicherheitsannahmen und -überlegungen nachzuvollziehen. Hierbei ermöglicht sie eine klare Verbindung zwischen den Sicherheitszielen und den Argumenten, die diese Ziele unterstützen.

Dies erleichtert die Überprüfung der Konsistenz und Relevanz der Argumentation im Hinblick auf die gesetzten Sicherheitsziele. Darüber hinaus erlaubt die Notation die Wiederverwendung von Argumentationsstrukturen. In sicherheitskritischen Systemen gibt es oft ähnliche Sicherheitsziele oder -entscheidungen, wodurch es sinnvoll ist, bewährte Argumentationsstrukturen zu speichern und in verschiedenen Kontexten wiederzuverwenden. Ein weiterer Faktor ist die Darstellung von klaren Strukturen und Beziehungen. So können Sicherheitsexperten und Prüfer die Argumentation leicht überprüfen. Dies trägt zur Qualitätssicherung und zur Identifizierung von Schwachstellen im Sicherheitskonzept bei. Die GSN ist somit auch ein effektives Mittel zur Dokumentation von Sicherheitsargumentationen. Diese Dokumentation ist nicht nur für Entwickler und Ingenieure, sondern auch für Aufsichtsbehörden, Prüfer und andere Interessengruppen von großer Bedeutung. Sie wird in vielen sicherheitskritischen Branchen, wie der Luft- und Raumfahrt, der Medizintechnik und der Automobilindustrie, eingesetzt. Sie unterstützt die Einhaltung von Sicherheitsstandards, indem sie eine strukturierte Darstellung der Sicherheitsargumentation ermöglicht. Insgesamt trägt die Goal Structuring Notation dazu bei, die Entwicklung sicherheitskritischer Systeme durch eine klare, transparente und überprüfbare Argumentationsstruktur zu unterstützen. Dies ist besonders wichtig in Bereichen, in denen die Einhaltung von Sicherheitsstandards und die nachvollziehbare Dokumentation von Sicherheitsüberlegungen von höchster Bedeutung sind.

6 ASSURANCE CASE

Ein Assurance Case ist eine formale Argumentationsstruktur, die verwendet wird, um die Sicherheit, Verlässlichkeit oder Konformität eines Systems zu dokumentieren und zu demonstrieren. Diese Argumentationsstruktur besteht aus einer Hierarchie von Claims (Behauptungen), die mit Argumenten und Evidenzen gestützt werden. Der Zweck eines Assurance Cases besteht darin, das Vertrauen in die Fähigkeiten eines Systems zu stärken und die Überzeugung zu vermitteln, dass das System sicher und zuverlässig ist.

Die Behauptung ist die Hauptaussage, die im Assurance Case vertreten wird. Sie könnte eine Sicherheitsaussage, eine Leistungsaussage oder eine Konformitätsaussage sein. Zum Beispiel: "Das autonome Fahrzeug gewährleistet die Sicherheit von Fußgängern". Anschließend werden Ziele und falls erforderlich Unterziele definiert. Die Annahme ist die rationale Randbedingung, die die Behauptung unterstützt. Sie kann aus mehreren Subannahmen oder Argumenten bestehen, die aufeinander aufbauen, um die Hauptbehauptung zu stützen. Fälle die sich innerhalb und außerhalb der Operational Design Domain (ODD) befinden, werden zuverlässig erkannt. Die Evidenzen sind konkrete Daten, Fakten oder Ergebnisse, die dazu dienen, die Argumente zu unterstützen. Das können Forschungsergebnisse, Testergebnisse, Simulationen

oder andere nachprüfbare Informationen sein. Beispielsweise könnten Evidenzen für eine präzise Fußgängererkennung Testergebnisse von umfangreichen Datensätzen von Fußgängererkennungsalgorithmen sein. Hierbei verdeutlicht Abbildung 2 die Struktur mit der letztendlich die Argumentation unter Einbindung der GSN realisiert werden kann.

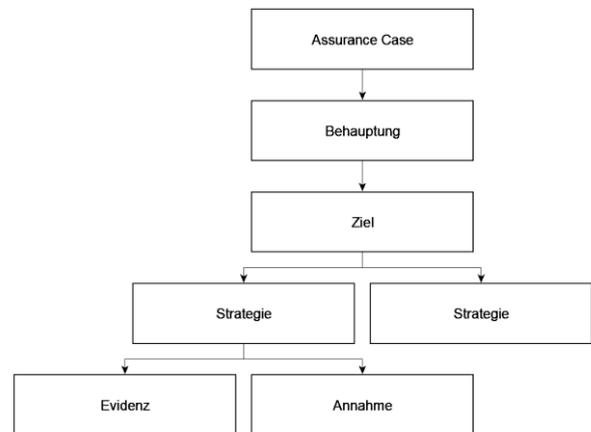


Abbildung 2: Generelle Struktur eines Assurance Case.

Darüber hinaus müssen Gegenmaßnahmen definiert werden, die ergriffen wurden, um Risiken zu mitigieren und die Sicherheit zu gewährleisten. Wenn beispielsweise identifizierte Risiken eine potenzielle Fehlfunktion der Fußgängererkennung beinhalten, könnten Gegenmaßnahmen redundante Sensoren oder kontinuierliche Überwachungssysteme sein. Eine Risikobewertung bewertet potenzielle Risiken und deren Auswirkungen auf das System. Dies hilft bei der Identifizierung kritischer Bereiche, die besondere Aufmerksamkeit erfordern. Auch hier muss Bezug auf einschlägige Standards und Normen genommen werden, um sicherzustellen, dass das System den allgemein anerkannten Anforderungen entspricht.

7 KOMBINATION BEIDER KONZEPTE

Die Goal Structuring Notation und der Assurance Case sind eng miteinander verbundene Konzepte und können effektiv miteinander kombiniert werden, um komplexe Sicherheitsargumentationen und -strukturen zu modellieren. Grundgerüst ist eine Gesamtstruktur in der GSN, die die zentrale Sicherheitsbehauptung (Claim) und die unterstützenden Ziele (Goals), Strategien (Strategies), Beweise (Evidence) und Annahmen (Assumptions) enthält. Jedes GSN-Element (Claim, Goal, Strategy, Evidence, Assumption) ist als eine spezifische Sicherheitsbehauptung im Assurance Case zu betrachten. Jedes Element trägt zur Gesamtargumentation bei. Anschließend erfolgt die Goal Decomposition in der GSN, um übergeordnete Sicherheitsziele in untergeordnete Ziele zu unterteilen. Jedes untergeordnete Ziel wird zu einer spezifischen Behauptung im Assurance Case. Strategien in der GSN mit konkreten Maßnahmen oder Nachweisen im Assurance

Case werden verbunden. Jede Strategie sollte durch geeignete Beweise gestützt werden, um ihre Wirksamkeit zu zeigen.

Diese Evidenz-Elemente in der GSN sind notwendig um die Sammlung von Beweisen für jede Sicherheitsbehauptung darzustellen. Dies können Daten aus Tests, Simulationen, Analysen oder anderen Quellen sein. Alle Annahmen in der GSN können grundsätzlich als Risikofaktoren betrachtet werden, die im Assurance Case transparent gemacht und gemanagt werden müssen. Diese Annahmen sollten klar dokumentiert und überwacht werden.

Jederzeit sind klare Verbindungen und Beziehungen zwischen den GSN-Elementen im Assurance Case notwendig. Zum Beispiel sollte jedes Goal direkt zur Unterstützung eines übergeordneten Claims beitragen. Sollten in der GSN besonders sicherheitskritische Punkte oder Aspekte vorliegen, müssen sie in einem Assurance Case besonders gründlich überprüft und dokumentiert werden. Standardisierte Diagramme und Grafiken existieren in beiden Notationsformen, um die Argumentationsstrukturen visuell darzustellen. Dies kann die Verständlichkeit und Klarheit der Sicherheitsargumentation verbessern. Da sowohl GSN als auch Assurance Case lebende Dokumente sind, sollten sie kontinuierlich aktualisiert und angepasst werden, um Änderungen im System, in den Anforderungen oder im Sicherheitskontext widerzuspiegeln.

Die Kombination von GSN und Assurance Case bietet eine strukturierte und systematische Möglichkeit, Sicherheitsargumentationen zu modellieren und zu dokumentieren, was insbesondere in sicherheitskritischen Anwendungen wie autonomen Fahrzeugen von entscheidender Bedeutung ist. [6]

Abbildung 3 zeigt einen exemplarischen GSN-Baum mit Fokus auf die Softwarequalität einer KI-Funktion. Zu Beginn ist das übergeordnete Ziel G1 Wartbarkeit (Maintainability) definiert. Die Strategie S1 verweist auf die entsprechenden Unterziele G1.1 bis G1.5, sowie auf einen Lösungsansatz E.1. Entscheidend ist der stringente Aufbau des Baumes entlang der Argumentationskette. Die Qualitätskriterien Wiederverwendbarkeit, Testbarkeit, Modifizierbarkeit und Analysierbarkeit bilden hier die Unterziele des Hauptziels Wartbarkeit. Der Lösungsansatz ist in dem Beispiel ein Verweis auf einschlägige Standards zur Softwarequalität wie z.B. die ISO 25010 Norm für Qualitätskriterien von Software, IT-Systemen und Software-Engineering.

Die Norm bietet ein Modell zur Evaluierung und Spezifikation der Qualitätsmerkmale für Softwareprodukte. Die ISO/IEC 25010 definiert acht Hauptqualitätsmerkmale, die weiter in Untermerkmale unterteilt sind. Wartbarkeit ist eines der Untermerkmale und ermöglicht Entwicklern und Qualitätsmanagern sicherzustellen, dass Softwareprodukte die erforderlichen Qualitätsstandards erfüllen.

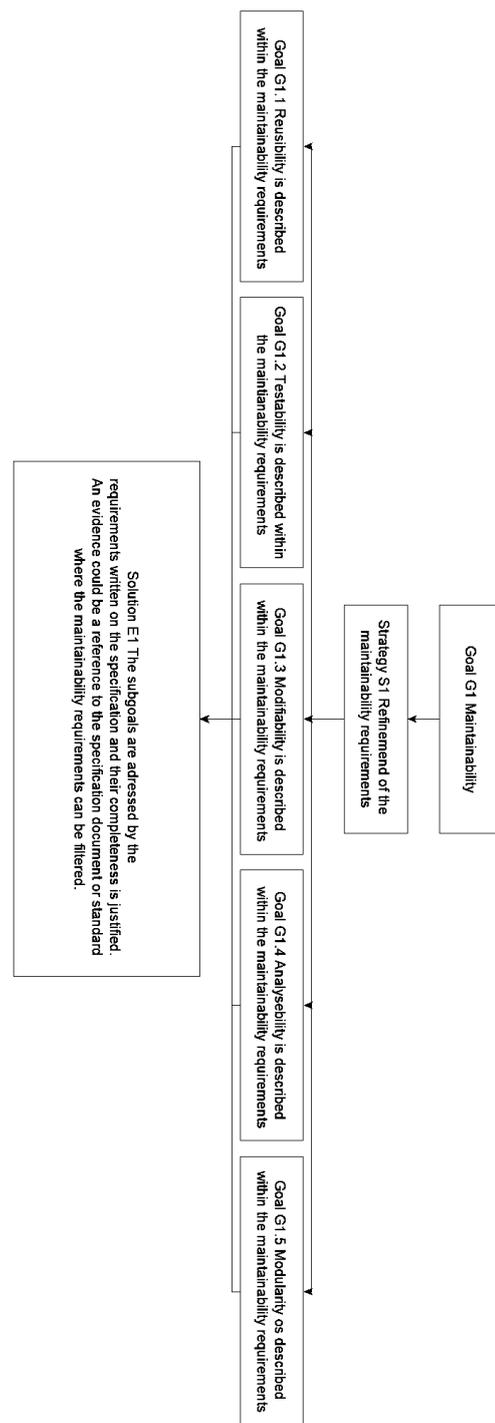


Abbildung 3: Beispiel eines GSN-Baums zur Fußgängererkennung mit Fokus auf die Softwarequalität des Algorithmus.

8 ZUSAMMENFASSUNG

Abschließend kann mit den zuvor genannten Methoden und Werkzeugen eine hinreichende Spezifikation einer KI-Funktion im Fußgängerschutz realisiert werden. Entscheidend ist hierbei vor allem die genaue Überprüfung der Randbedingungen unter Einbindung aller Aspekte der Entwicklung, Implementierung und Bewertung sowie die Hinzunahme von Standards und Normen. Die größte Herausforderung bleibt, dass viele KI-Modelle

weiterhin als "Black Box" betrachtet werden, da es schwierig sein kann, genau nachzuvollziehen, wie sie zu bestimmten Entscheidungen kommen. Die Undurchsichtigkeit von KI-Systemen erschwert die Identifizierung von Fehlern, unerwartetem Verhalten oder Diskriminierung. Es ist wichtig zu betonen, dass die Sicherheit von KI ein laufender Prozess ist, der sich mit der Entwicklung neuer Technologien und Angriffsmethoden weiterentwickeln wird. Die oben genannten Maßnahmen dienen als Ausgangspunkt.

LITERATUR

- [1] Harald Rueß; Simon Burton (2022): Safe AI - How is this Possible? In: CoRR abs/2201.10436.
- [2] Gauerhof, Lydia; Munk, Peter; Burton, Simon (2018): Structuring Validation Targets of a Machine Learning Function Applied to Automated Driving. In: Computer Safety, Reliability, and Security: Springer International Publishing, S. 45–58.
- [3] Assion, Felix; Schlicht, Peter; Greßner, Florens; Günther, Wiebke; Hüger, Fabian; Schmidt, Nico; Rasheed, Umair (2019): The Attack Generator: A Systematic Approach towards Constructing Adversarial Attacks. In: Proc. 2019 IEEE Conf. Comput. Vision and Pattern Recognition Workshops.
- [4] Burton, Simon; Gauerhof, Lydia; Sethy, Bibhuti Bhusan; Habli, Ibrahim; Hawkins, Richard (2019): Confidence Arguments for Evidence of Performance in Machine Learning for Highly Automated Driving Functions. In: Computer Safety, Reliability, and Security, Bd. 11699: Springer International Publishing, S. 365–377.
- [5] Geirhos, Robert; Rubisch, Patricia; Michaelis, Claudio; Bethge, Matthias; Wichmann, Felix A.; Brendel, Wieland (2018): ImageNet-Trained CNNs Are Biased towards Texture; Increasing Shape Bias Improves Accuracy and Robustness. In: Proc. 7th Int. Conf. Learning Representations.
- [6] Schwalbe, Gesina; Knie, Bernhard; Sämann, Timo; Dobberphul, Timo; Gauerhof, Lydia; Raafatnia, Shervin; Rocco, Vittorio (2020): Structuring the Safety Argumentation for Deep Neural Network Based Perception in Automotive Applications. In: António Casimiro, Frank Ortmeier, Erwin Schoitsch, Friedemann Bitsch und Pedro Ferreira (Hg.): Computer Safety, Reliability, and Security. SAFECOMP 2020 Workshops. Cham: Springer International Publishing, S. 383–394.
- [7] ISO (2017): ISO/IEC/IEEE 12207:2017: Systems and Software Engineering Software Life Cycle Processes. 1. Aufl.
- [8] ISO (2019): ISO/PAS 21448:2019(En): Road Vehicles Safety of the Intended Functionality.
- [9] 16th ACM/IEEE Int. Conf. Formal Methods and Models for System Design (2018): IEEE.
- [10] Burton, Simon; Herd, Benjamin (2023): Addressing uncertainty in the safety assurance of machine-learning. In: Frontiers in Computer Science 5. DOI: 10.3389/fcomp.2023.1132580.
- [11] Carlini, Nicholas; Wagner, David (2017): Adversarial Examples Are Not Easily Detected: Bypassing Ten Detection Methods. In: Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security: Association for Computing Machinery (AISeC '17), S. 3–14.
- [12] Schleiss, P., Carella, F., & Kurzidem, I. (2022, November). Towards continuous safety assurance for autonomous systems. In 2022 6th International Conference on System Reliability and Safety (ICSRS) (pp. 457–462). IEEE.
- [13] Kindermans, Pieter-Jan; Schütt, Kristof T.; Alber, Maximilian; Müller, Klaus-Robert; Erhan, Dumitru; Kim, Been; Dähne, Sven (2018): Learning How to Explain Neural Networks: PatternNet and PatternAttribution. In: Proc. 6th Int. Conf. on Learning Representations.
- [14] Bagschik, G.; Menzel, T.; Maurer, M. (2018): Ontology Based Scene Creation for the Development of Automated Vehicles. In: Proc. 2018 IEEE Intelligent Vehicles Symp: IEEE, S. 1813–1820.
- [15] Computer Safety, Reliability, and Security (2018): Springer International Publishing.
- [16] Bojarski, Mariusz, et al. "End to end learning for self-driving cars." arXiv preprint arXiv:1604.07316 (2016).
- [17] Proc. 10th European Congress Embedded Real Time Systems (2020).
- [18] Proc. 2018 IEEE Intelligent Vehicles Symp (2018): IEEE.
- [19] Bernhard, Julian, et al. "Risk-based safety envelopes for autonomous vehicles under perception uncertainty." 2022 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2022.
- [20] Leveson, Nancy (2012): Engineering a Safer World: Systems Thinking Applied to Safety: MIT Press (Engineering Systems).

Organizational Integration of User Experience Contributors in Agile Software Development

Dominik Oelfin

Pforzheim University
Tiefenbronner Straße 65
75175 Pforzheim
E-Mail: oelfindo@hs-pforzheim.de

Frank Morelli

Pforzheim University
Tiefenbronner Straße 65
75175 Pforzheim
E-Mail: frank.morelli@hs-pforzheim.de

ABSTRACT

In today's software industry, agile software development methods have largely replaced traditional processes. At the same time, organizations are increasingly recognizing that user experience plays an important role in the quality and success of software. While agile methods are widely used in practice and therefore influence organizational considerations, the integration of user experience contributors in such organizations poses several challenges. The presented research shows that the involvement of user experience in software development is considered valuable in theory, but its prioritization lags behind in practice. Therefore, we investigate factors that need to be considered to improve the organizational integration of user experience contributors in agile software development. These factors can be divided into structural aspects and aspects of processes and practices. Organizational design, roles involved, and further structural aspects such as organizational autonomy serve as a basis to allow for effective and efficient involvement of user experience contributors in product development processes. Organizational practices shape the way of how the processes of user experience and agile software development can be combined. In particular, the interconnectedness of processes, mutual prioritization, and close collaboration can be identified as key factors. Finally, we consolidate these factors into a framework and enrich them with practical implications. This framework can serve as a starting point for researchers and practitioners to further explore and improve the organizational integration of user experience contributors in agile software development.

KEYWORDS

Agile Software Development, Agile Methodology, Human Centered Design, User Centered Design, User Experience

INTRODUCTION

User experience (UX) has a significant impact on the perceived quality of software, encompassing the “emotions, beliefs, preferences, perceptions, comfort, behaviors, and accomplishments that occur before, during and after use” of software (ISO 9241-210:2019 2019). Roles that are explicitly working on UX are therefore an integral part of organizations for the successful development of software (Bruun et al. 2018, 1–3). As such, the tasks of UX contributors include conducting user research, creating product wireframes and prototypes, assessing usability and accessibility of products and analyzing product interactions (Bruun et al. 2018, 4). Responsibilities often overlap among roles (e.g. “user researcher”, “visual designer”, “interaction designer” and “information architect”), with some teams combining these distinct responsibilities into a unified UX designer position. The specific naming, responsibilities, and tasks of these roles can vary depending on the organization, product, and context.

The responsibility of UX managers is to plan, oversee, and measure the overall UX process and practice (Bruun et al. 2018, 4). Furthermore, another key objective of UX management is to raise awareness about the importance

of UX and the value it creates within the organization (Szóstek 2012). To integrate the UX management strategy, a UX mission should be derived directly from the business goals of the organization (Rosenberg 2019, 30).

The present state of UX influence in an organization can be referred to as UX maturity (Chapman and Plewes 2014, 14). UX maturity has two direct effects on the organization: The more mature the integration of UX into development, the more influence UX has on product decisions. In addition, a higher level of maturity leads to a higher awareness of the importance of UX and thus to a better allocation of resources for the UX team.

In recent years, agile methods have significantly changed the existing organization of software development in many companies (Gerster, Dremel, and Prashant 2018, 1). Teams are becoming more interdisciplinary, development cycles are becoming shorter, and products are brought to market faster (Dima and Maassen 2018, 316). These changes have implications for the integration of UX management into the development process. Where development cycles are more tightly timed, the need for time-consuming UX practices is questioned – the influence of UX contributors seems to fade (McInerney and Maurer 2005, 19). Nevertheless, especially in the context of innovative agile development, UX is a crucial need to develop competitive user-centered products. Interdisciplinary collaboration is becoming increasingly im-

portant. The quality of developed software is strongly dependent on the quality of the collaboration of the teams involved and thus significantly determined by organizational aspects.

Studies have empirically demonstrated the influence of organizational structures of software development teams on software quality (Nagappan, Murphy, and Basili 2008). However, their conclusions do not explicitly refer to agile software development and the involvement of UX contributors. Furthermore, various studies on the integration of UX practices and principles in development organizations can be found (Kashfi, Feldt, and Nilsson 2019; Gray, Toombs, and Gross 2015). For the most part, existing publications examine new practices to bring agile methods to UX. However, the necessary organizational approaches to combine UX contributors and development in one organization are only marginally considered yet (Ferreira, Sharp, and Robinson 2012).

Based on this, the following research objectives are covered within this paper:

- What is state of the art for integrating UX in agile software development?
- Which organizational factors influence the integration of UX contributors in agile software development?
- Which implications for the integration of UX contributors in agile software development can be derived from these factors?

We base our research on the Design Science Research method (Hevner et al. 2004; Hevner and Chatterjee 2010). In this paper we present the outcomes of a review of existing literature to provide an overview of the persisting knowledge base. We focus on the organizational design of agile software development teams as well as specifics of structures for UX contributors. Furthermore, we reflect upon existing processes and practices integrating UX practices into agile software development. Hereby we aim to identify influential factors that affect the organizational integration of UX contributors in agile software development and their respective impact on the integration. To add practical insights, enriching the theories gained from literature research, we present a case study that we conducted, analyzing a real-world example of an integration of UX contributors into a specific agile software development organization. Based on the findings from the literature review and the case study we introduce a combined framework containing influential factors and their implications for the organizational integration of UX contributors in agile software development.

ORGANIZATIONAL DESIGN OF AGILE SOFTWARE DEVELOPMENT AND UX

A central concept of agile methods is that their processes benefit from the unique skillsets of each team member and each team (Cockburn and Highsmith 2001, 132).

Consequently, based on their strengths, the team members are assigned to specific roles within the agile software development team. Gerster et al. (2020) identify three groups of roles within a product development team: key roles, utility roles and contributor roles. Based on the Scrum methodology, the key roles include product owner, technical product owner and Scrum master (Gerster et al. 2020, 88). Utility roles are fulfilled by the regular team members. Based on their individual skills they are assigned to design, develop, test, integrate, maintain, or operate the product. In addition to this core team, there is an extended team consisting of specialists for specific tasks, brought in as required. These can participate within the team either for short terms but full-time or flexibly as part-time support. Although they enhance the ability of the team it is not intended that they carry out all the tasks related to their field of expertise, but rather transfer their skills to the core team members and support them in their decisions (Gerster et al. 2020, 88).

Gerster et al. (2020) furthermore describe a generic model of an organizational structure for agile software development – the “fully agile unit”. This model is based on an organizational structure used at Spotify, firstly described by Kniberg and Ivarsson (2012) and extended by Gonçalves and Lopes (2014) (Gonçalves and Lopes 2014). The model of a fully agile unit is characterized by the definition of four distinct groupings: tribe, squad, chapter, and guild. The whole of a unit is called a tribe and reflects a product area within a software development organization. Organizations can consist of multiple tribes. Tribes (product areas) consist of multiple products, each developed by one or multiple teams, referred to as squads. Squads are structured like Scrum teams and contain the roles previously described (Kniberg and Ivarsson 2012, 1–5). Several team members of different squads can come together to form a chapter. A chapter typically combines members with similar skills and tasks. Like a matrix structure, each chapter typically has a chapter lead, responsible for all chapter members, who is also part of a specific squad. Alternatively, several team members of multiple squads can form a guild. These are less formal and usually do not have a strict leading role, making them a “community of interest” (Kniberg and Ivarsson 2012, 10). Chapters and guilds are beneficial to keep the separate squads engaged with each other, resulting in economies of scale through reduced work-redundancy, while keeping them as autonomous as possible. Gerster et al. (2020) furthermore add shared teams to the existing concepts.

Involvement of UX

Gerster et al. (2020) position UX contributors as part of the extended team, being assigned full-time while actively participating for short terms. Apart from this generalized positioning, Gerster et al. (2020) do not elaborate further on the integration of UX contributors in the fully agile unit. Kniberg and Ivarsson (2012) do not explicitly mention UX. However, a screenshot in their article shows

that UX could be considered as a separate squad. We assume that the concept of shared teams is relevant with regards to the integration of UX contributors into the organizational model. Furthermore, positioning UX contributors as part of the core team can be envisioned as well. To further analyze the positioning of UX, we now focus on the organizational structures that UX contributors might find themselves in.

Organizational Structures for UX Contributors

Centralized Structure

A centralized UX structure is characterized by the existence of a single UX team with a company-wide focus, serving as a supporting unit for all product divisions. The organizational structure of this UX team can be further divided into smaller functional teams focusing on specific topics and areas within the field of UX (Kaplan and Pernice 2019). UX contributors are flexibly assigned to projects for the individual products, meaning that a single contributor can work for various products over the course of time or at the same time. The assignment of the UX contributors is determined by their expertise in a specific field of UX rather than profound product knowledge (Kaplan and Pernice 2019).

Software development can benefit from a centralized UX structure in multiple ways: From product perspective, the UX consistency of all products can be ensured, as all UX effort is coming from a team reporting in a single organizational structure. Based on the comprehensive skill set with regards to all aspects of UX provided by the different UX contributors, the centralized team can always provide the most suitable contribution to the project. While this is a benefit of the centralized approach, it is also a challenge that it faces. UX contributors are lacking in-depth product knowledge and additional effort and time is necessary to provide a qualitative assessment of the user requirements for the product (Kaplan and Pernice 2019). Bringing UX to the table of product decisions can be hard in a centralized system, as the UX team is seen as a separate unit. Product divisions need to proactively request the consultation by the UX team which presupposes that the need for UX consultation and its value is recognized by the product development team first (Rosenbaum et al. 2000, 340; Kaplan and Pernice 2019).

Overall, Rohn (2007) rates a centralized UX organization to be most effective for most companies (Rohn 2007, 4). Oppositional, Rosenberg (2019) assesses centralized UX to be a model of the past, that was seen as the optimal mature structure “for the first two decades of corporate UX practice” (Rosenberg 2019, 30).

Distributed Structure

Distributed structures embody a contradictory approach to the organizational integration of UX. UX contributors are distributed across the company and the product divisions. In a distributed model, the UX work force is located much closer to the development work force of the products (Rosenberg 2019, 30). Often UX contributors are structurally directly integrated into the development teams of specific products. As a result, individual UX contributors collaborate consistently with the same developers and product managers within their respective teams. For most product development teams, there are only a few UX contributors working on the same product – sometimes even only a single contributor. Consequently, the UX contributors within a distributed model need to be more versatile in terms of their UX skills. However, they develop stronger product specific domain knowledge over time (Kaplan and Pernice 2019).

The main advantage of distributed UX contributors is that they work directly at the product development level. Thus, it is easier for them to influence product decisions by having a fixed place in regular team processes. Additionally, the value of UX can be better demonstrated by product specific improvements, sparking long-lasting trust by the direct product management. With these benefits, the distributed model seems suitable for implementation in an agile setting. However, the strict timing of agile methods can limit the possible effort spent on UX topics. Disadvantages of a distributed model are also evident looking at alignment issues. Ensuring a consistent UX over multiple products and preventing redundant efforts with separated UX contributors demands sophisticated communication (Kaplan and Pernice 2019). Szóstek (2012) suggests to “provide them with a forum, place and time to meet and discuss their projects as a way to support and learn from each other” (Szóstek 2012, 13).

	Centralized	Distributed	Matrix	Federated
Number of UX contributors per team	++	o	+	+
Variety of UX capabilities	++	o	+	+
UX consistency	++	o	+	++
Visibility of UX value for product team	o	+	+	+
Product knowledge of UX contributors	o	++	++	++
UX presence in product team	o	++	++	++
UX influence on product decisions	o	++	++	++

Table 1: Effects of organizational structures for UX contributors (++ = high, + = medium, o = low)

Matrix Structure and Federated Structure

Matrix structures and federated structures are intended to combine the benefits of both the centralized and the distributed system, while compensating their disadvantages. Within a matrix structure, UX contributors are organizationally connected to two different managers, a product manager and a UX manager. To ensure a consistent organizational structure, one reporting line to one manager (either UX or product) is formally prioritized, while the other is rather informal (Kaplan and Pernice 2019). The main advantage of UX contributors in a matrix setting is their direct product involvement, combined with the strengths of an overarching central UX organization (Rohn 2007, 4).

Rosenberg (2019) introduces another term for an organizational structure that claims to combine the benefits of centralized and distributed UX – the federated structure. In contrast to the matrix structure, the federated model envisions multiple dedicated UX teams embedded within the product divisions, not only single contributors. These product specific UX teams are backed by a centralized UX team, responsible for company-wide visibility of UX, consistency standards, and specifically skilled UX support (Rosenberg 2019, 31).

Overall, matrix and federated settings improve the presence of UX within the product development teams, leading to higher value put on the work of UX contributors as well as better product knowledge of the UX contributors. Hence, the influence of UX input on product decisions is higher. Nonetheless, both matrix and federated models have their disadvantages. Within the matrix setting, UX contributors “may feel pulled between their two different managers” (Kaplan and Pernice 2019). Additionally, the added organizational complexity through the matrix structure can slow down decision making and demands additional alignment. Within a federated setting it can be challenging to claim the authority of the central UX team in providing standards as the product level UX teams might question their decisions based on product knowledge. UX consistency is hence more difficult to ensure (Rosenberg 2019, 34).

Applying the Fully Agile Unit Model to UX

Although there is no clear positioning of UX within the fully agile unit, based on the research on organizational structures for UX contributors, the concepts can be mapped to each other: Centralized UX teams use the concept of shared teams, supporting multiple product development teams (squads) or even multiple product areas (tribes). Distributed UX contributors on the other hand are directly incorporated into the product development teams and can hence be seen as either part of the core or extended team, depending on the UX focus of the organization. Providing them a forum (Szóstek 2012, 13) can be compared with the guild concept of the fully agile unit. Furthermore, organizing UX contributors as chapters is comparable to the matrix UX structure, having UX contributors incorporated within the product development

teams but still collectively organized by a semi-central UX team. Federated UX teams, as described by Rosenberg (2019) can be implemented as separate squads within the respective product area (tribe).

Autonomy

To provide faster decisions, agile software development features the concept of self-managing individuals and self-organizing teams (Beck et al. 2001b, eleventh principle). A key feature of self-organizing teams is autonomy. Moe et al. (2008) introduce three levels of autonomy: external, internal, and individual autonomy.

External autonomy is determined by the influence of team-external parties, such as management and other individuals. The higher the influence of these, the lower the external autonomy of a team. Internal autonomy is high when the decision power within the team is shared between all team members. Decisions are not taken by e.g., a team lead, but by either the entire team or on-demand authorized subgroups. Individual autonomy refers to high individual control over the process of working on assigned tasks – characterized by few rules and constraints. Individual autonomy, although providing freedom to the individual team members, can lead to challenges with regards to internal autonomy. With self-structured work of the individuals, the frequency of interaction with other team members decreases, making it more difficult to achieve cohesive team decisions (Moe, Dingsøyr, and Dybå 2008, 78).

Moe et al. (2008) conclude that traditional software development teams are characterized by high individual autonomy but low internal and external autonomy. Agile teams on the other hand demand for a balanced amount of all three levels of autonomy. To achieve this, first, external management influence must be limited to increase the authority of the team in project decisions. Thus, team members can better identify with the project and hence foster interaction and shared decision-making. Second, there is a need for redundancy in the roles of team members to ensure that contributors can assist or substitute each other. Consequently, Moe et al. (2008) recommend that the team members acquire multiple skills to achieve functional redundancy. Although redundancy is typically considered counterproductive and uneconomical, within self-organizing teams it is crucial to achieve flexibility and support rapid decision-making (Moe, Dingsøyr, and Dybå 2008, 82–83).

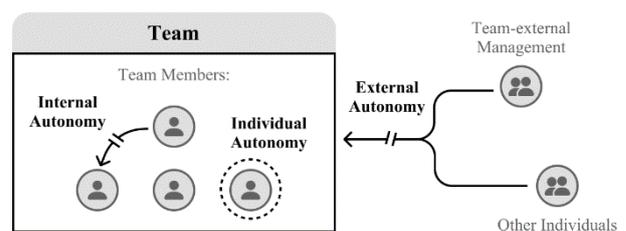


Figure 1: External, internal, and individual autonomy

Impact of Organizational Structures on Software Quality and UX

Structural organization influences the quality of software via multiple factors. Damaševičius (2010) mentions that there is a direct relation between the quality of software code produced and the team structures that produced the code. Especially when it comes to interfacing software components, they can only be of high quality if the communication between the involved contributors was working well (Damaševičius 2010, 12). This is referring to Conway's law, which in detail states that "[...] organizations which design systems [...] are constrained to produce designs which are copies of the communication structures of these organizations." (Conway 1968, 31) Consequently, the determined organizational structure crucially influences the expected outcome and should hence be carefully planned and observed.

In literature, there is no fundamental evidence that organizational structure directly influences the UX perceived by the users. Nonetheless, indirect relations can be identified. Software quality directly determines the utility and usability of a software as low quality of software results in usability issues for the user. Hence, good software quality resulting amongst others from optimized organizational structures influences the perception of the software by the users and therefore determines the UX of the software. Furthermore, assuming UX contributors as part of the product development teams, Conway's law can also be applied to them. It is hence crucial that UX contributors closely collaborate both with the development team as well as with potential other UX contributors working on connected features and software components.

PROCESS DESIGN OF AGILE SOFTWARE DEVELOPMENT AND UX

Challenges

Agile software development methodologies have gained popularity due to their iterative and customer-centric approach. The user-centric and iterative approach of UX implies that agile software development and UX fit well together. However, integrating UX practices into agile projects presents several challenges. One of the main concerns highlighted by Larusdottir et al. (2012) is the lack of the big picture of UX in agile projects. Because agile software development focuses on individual features rather than holistic UX, both the UX work and its outcome can be fragmented in an agile environment (Larusdottir, Cajander, and Gulliksen 2012, 2).

Another challenge identified by Chamberlain et al. (2006) is the power struggle between UX contributors and developers (Chamberlain, Sharp, and Maiden 2006, 152). UX may be perceived as optional or peripheral to development work, leading to a lack of collaboration between UX contributors and developers (Jurca, Hellmann, and Maurer 2014, 28). Silva da Silva et al. (2012) find that UX contributors are often not perceived as full mem-

bers of the product development team and work on multiple projects simultaneously. This approach to UX hinders close collaboration between UX contributors and developers (Silva da Silva et al. 2012, 749–50).

Kuusinen and Väänänen-Vainio-Mattila (2012) point out that UX work is often not included in development processes, and decisions regarding its inclusion are not made by UX specialists. As a result, UX issues are considered too late in the process, leading to inefficiencies in the use of UX resources (Kuusinen and Väänänen-Vainio-Mattila 2012, 145–47). In addition, Chamberlain et al. (2006) point out differences in time frames, communication issues, and a reluctance to understand each other's needs within the collaboration between UX and development (Chamberlain, Sharp, and Maiden 2006, 152).

To overcome the challenges mentioned above, the integration of agile software development and UX is dependent on processes that promote collaboration, coordination, and communication. The goal is to integrate the benefits of agile software development and UX without limiting either in its practices or methods.

Practices

Big Picture

Consistency throughout a product is essential for a positively perceived UX. To achieve consistency, UX contributors require a vision of the entire product in advance of creating the product design (Kuusinen 2014, 266). Agile software development, especially Scrum, on the other hand is feature-oriented. Hence, developers focus primarily on the short-term scope of the current iteration rather than the big picture of the product (Larusdottir, Cajander, and Gulliksen 2012, 6). Larusdottir et al. (2012) mention that it is more convenient for developers to work on a small fraction of the product at a time, while they are not responsible for keeping track of the big picture of the product – for which the product owner is responsible (Larusdottir, Cajander, and Gulliksen 2012, 4). However, to achieve a positively perceived UX, a collectively known big picture is crucial. Consequently, it is suggested that each role within the product development team should be responsible for keeping the product vision in mind. Kuusinen and Väänänen-Vainio-Mattila (2012) summarize this practice as “all together from early on” (Kuusinen and Väänänen-Vainio-Mattila 2012, 145).

Big or Little Design Upfront

A common perception of UX practices is that they involve “big design upfront” (Larusdottir, Cajander, and Gulliksen 2012, 6), a demand for fixed requirements and a strict handoff to development at the end of the design work, making it difficult to include in regular iterations. In contrast, Beyer (2010) argues that UX processes are in fact iterative as well (Beyer 2010, 11). Thus, the literature refers to “little design upfront” as an appropriate approach for UX within agile software development – acknowledging that certain design tasks need to be done upfront while affirming that the main UX activities can be

done iteratively in parallel with the development iterations (Silva da Silva et al. 2012, 744).

Parallel and Iterative Tracks

Although UX activities can and should be performed iteratively throughout the development sprints, it is essential to allow sufficient time for UX contributors to focus on the design prior to the implementation of the respective artifact (Jurca, Hellmann, and Maurer 2014, 29). Consequently, the literature suggests shifting the main UX activities to a parallel track, one sprint ahead of the implementation sprint of the respective feature. With this method, UX contributors work on the upcoming feature one sprint ahead, while overseeing and supporting the implementation during the current sprint (Sy 2007, 118).

With this “one sprint ahead” practice in place, UX and development work is done in parallel synchronized tracks (Kieffer, Ghouti, and Macq 2017, 578). While the members of both tracks work independently on their current sprint items, regular collaboration is required to support the current implementation. Several research papers recommend the practice of parallel tracks (Larusdottir, Cajander, and Gulliksen 2012; Sy 2007; Miller 2005; Salah, Petrie, and Paige 2009). It serves to combine the benefits of regular interaction between UX and development, while allowing them the necessary autonomy to organize and pursue their activities individually.

To fulfill the purpose of having UX and development working closely together, it is essential that these parallel tracks are both synchronized and interwoven, as mentioned in the third principle of user-centered agile software development by Brhel et al. (2015) (Brhel et al. 2015, 18–19). The separation into different tracks should not lead to less interaction and collaboration between UX and development.

To support the previously introduced parallel tracks, it is necessary that UX contributors and developers work on the same sets of features. While it is common in agile software development to break large parts of the big picture into smaller pieces to work on, it is less common for UX activities. Therefore, to integrate UX into agile software development, it is expected that UX contributors will work on smaller feature sets rather than designing the entire product end to end. These small feature sets, tied to the current scope of the sprint, incrementally add up to a growing prototype of the product (Kieffer, Ghouti, and Macq 2017, 578).

In addition, the iterative nature of both agile software development and user-centered design should be encouraged, as described in the second principle of Brhel et al. (2015) (Brhel et al. 2015, 17–18). Although UX contributors create initial prototypes of the features one sprint ahead of development, these prototypes should not remain unchanged. The design should be iteratively revised both within the preceding UX sprint and within the implementation sprint. Iterative changes should be based on

the regular interaction of the UX contributors with other UX contributors, developers, and other involved stakeholders such as management, customers, and users (Brhel et al. 2015, 17).

Handoff and Collaboration

Traditional software development methods rely heavily on formal handoff processes between the various contributors and development stages involved. Agile software development, on the other hand, explicitly refrains from incorporating fixed handoff processes to promote the iterative and lightweight collaboration between the different contributors and teams involved. Larusdottir et al. (2012) find that most of the collaboration between UX contributors and developers takes place through informal practices – further highlighting the need for close personal connections between the collaborating roles (Larusdottir, Cajander, and Gulliksen 2012, 6). Additionally, close connections foster early and frequent communication. Jones and Thoma (2019) advocate for UX contributors and developers to build these relationships to enable faster problem identification, problem solving, ideation, and decision making. Moreover, direct connections between UX contributors, product owners, and developers improve the understanding of each other’s role, fostering mutual understanding and the sharing of skills and knowledge (Jones and Thoma 2019, 35–36).

The different roles of a product development team need effective artifacts to discuss and communicate the current state and upcoming plans (Brhel et al. 2015, 22). These artifacts should be adaptive, as opposed to fixed documentation, in line with the second value of the Agile Manifesto (“Working software over comprehensive documentation” (Beck et al. 2001a)). Bruun et al. (2018) argue that wireframes, mockups, and prototypes created by UX contributors can serve this purpose as effective and up-to-date artifacts for communication both within the team and with external stakeholders, such as customers and users (Bruun et al. 2018, 7). In addition to communicating and informing, these artifacts can also be used to actively involve developers and product management in the design of the solution, enabling iterative co-creation rather than strictly isolated delivery steps (Jones and Thoma 2019, 27).

Although agile software development focuses on the rapid creation of functional prototypes, the previously mentioned artifacts have gained popularity within agile methodology and can therefore facilitate the integration with other UX-based processes and tools (Brhel et al. 2015, 22).

CASE STUDY

Goal

To evaluate and demonstrate the relevance of the integration of UX contributors in agile software development, we conducted a case study, adding to both the relevance

cycle and the design cycle of our Design Science Research approach. This case study examines how a large international software company integrates product development teams and UX contributors for a specific product area. The company has multiple product areas, and it embraces agile innovation for its products.

In terms of organizational structure, the study describes the structure of the case organization and analyzes how the UX department fits into the company's overall structure and product development teams. It compares these findings to established organizational design principles and generic agile unit and UX team structures. The study also explores organizational goals and team members' experiences related to the structure.

On the process side, the study evaluates the application of defined practices within the organization and how they are supported by managerial activities. It focuses on the collaboration between UX contributors, product owners, and developers, while also assessing the priority of UX within the organization and its impact on UX contributor integration.

Method

Referring to the Design Science Research method, this case study serves as a practical context to further develop, evaluate, and refine the findings and concepts from the literature presented earlier.

The case study is primarily qualitative, gathering insights from experts and practitioners in user-centered agile software development through five interviews. These interviews spanned various roles, including UX management (two UX managers), UX practice (a UX lead), product management (a product owner), and development (a developer). Given the global distribution of interviewees, the interviews were conducted online and were semi-structured, with tailored key questions.

For UX managers in Germany and the USA, the interviews focused on organizational design and expectations for UX, development, and product management collaboration. The interview with a USA-based UX lead centered on practical application of processes and their impact on role coordination. A Canadian product owner's interview explored product management's UX integration expectations. The interview with a German front-end developer delved into working with UX contributors and how organizational structures affect role collaboration. All interviewees were questioned about their learning experiences from working with other roles.

Additionally, a workshop and group discussion involved eight UX contributors from various locations (Germany, Canada, Poland, China). It dealt with preferred organizational structures and experiences with integrating UX and agile development practices. Furthermore, the study ex-

amined relevant documents outlining processes and practices and was complemented by observations of departmental activities.

Findings

Structural Organization

Product owners as well as development teams of the specific products within the product area are organizationally cumulated under the individual heads of product. Each head of product directly reports to the chief product officer of the product area. Structurally, UX is not part of the product development teams. All UX contributors for the different products are combined in a separate product area specific UX department under the head of UX who directly reports to the chief product officer of the product area.

Within the UX department, the UX contributors are grouped by their different locations (Northern America, Europe, Asia), each having a respective UX manager. In addition, individual contributors act as UX leads for the different products. These directly report to the head of UX. UX contributors are structurally reporting to their location managers but work in teams for the different products, led by the respective UX leads. Product area specific UX departments are common across the entire company. Together with a central UX department, they form a federated UX structure. The central UX department acts as a supporting unit.

The structural integration of the UX contributors in the considered case does not follow a specific structure of the fully agile model but applies a combination of its characteristics. Within the UX department, UX contributors form distinct groups, similar to squads, led by their location-specific UX managers. However, their main activities are performed within chapters that focus on individual products and are led by the UX leads. Additional guilds for specialized topics can be found within the UX department. In the context of the separated product development teams, the UX department can be classified as a supporting shared team.

In terms of autonomy, the interviewed UX managers emphasize that a certain amount of external autonomy is beneficial. Considering the different locations of the UX managers and teams, external and internal autonomy allows them to manage their teams in a way that best fits the team's culture and way of working. In addition, internal autonomy allows product-specific UX teams to collaborate on solutions without being dependent on the decision of individual managers. As a result, the UX lead and UX managers see their responsibility in enabling their teams to make decisions together, rather than driving the decisions themselves. Increased external autonomy, on the other hand, invokes the feeling of siloed departments, as criticized by the UX contributors. To achieve organizational goals and product goals, the focus should be on the collaboration and connectivity of the teams and contributors involved. According to one of the

UX managers, the balance lies in having autonomy for the team while also incorporating a connected approach that brings together the right expertise and perspectives.

According to the UX lead, having a dedicated UX department for the product area improves the influence of UX and helps align UX efforts with business goals. Because UX workloads vary across products, the flexible matrix structure of the UX department allows UX managers to quickly adjust the assignments of UX contributors. Being part of a larger UX department enables the UX contributors to collaborate on solutions. Additionally, having the smaller UX teams for each product promotes quick decision-making and increases organizational agility. While these factors highlight the benefits of having a separate UX department within the product area, the interviewees also express a desire for closer proximity between UX contributors and development teams. The product owner believes that integrating UX effectively into the development lifecycle requires the UX contributors to be as close to the development teams as possible. In the current structure, UX contributors report that they do not feel part of the product development teams. The consulted developer mentions that having UX contributors on the same team would significantly improve feedback, allowing for easier clarification and better agreement on realistic requirements.

Processes and Practices

Big Picture: The interviewed UX lead mentions that focusing on the big picture in advance helps to streamline the development process. In practice, the UX contributors experience contradicting realities. They report that some product development teams involve UX early in the process while others start to consult UX at a very late stage. While the former leads to a better understanding of the product and helps to identify gaps of the product vision early, the latter results in a lack of the big picture and hence in conflicting UX. The UX contributors further argue that missing a collective product vision leads to ever-changing requirements due to misalignment within the product development teams. As a result, in some cases, UX contributors report to work on features without having a clear definition of the use case.

According to one of the interviewed UX managers, UX involvement in the early envisioning phases of new products or projects is often limited and needs to be improved. This impression coincides with the experience of the UX lead who mentions that in early product envisioning phases, UX is excluded from most strategy planning. This disconnect consequently leads to problems when product owners ask for a comprehensive look at the product appearance without involving UX from the beginning, as the UX lead reports. The other interviewed UX manager also confirms that involvement in the envisioning phase should be cross-functional. Previous projects that were either entirely driven by product management or entirely driven by UX in their envisioning phases lacked either user-centricity or technical feasibility. The

product owner emphasizes the benefits of involving UX early in the process, not only for improving the usability of the product, but also for refining requirements. However, the product owner further mentions that “[...] it is currently only a matter of personal preference if [product owners] want to work with UX or not.”

One Sprint Ahead: During the interview, the UX lead recalls former processes where UX work was integrated directly into the development sprints, working on the same features as development during the same sprint. As this approach caused stress and dissatisfaction among the UX contributors, the developers as well as the product owners, the lead time of UX activities ahead of development was increased. Both the product owner and the UX lead concurringly report that currently UX activities take place not only one but multiple sprints ahead of development – usually four to six weeks in advance of the implementation, allowing for sufficient refinement and iterations. For the product owner, this added time allows for better capacity planning and it ensures that the developers are equipped with sophisticated UX artifacts when the implementation begins.

Parallel and Iterative Tracks: According to the interviews conducted, the UX and development work does not have a specific synchronization with respect to the development sprints. Nevertheless, a product area specific product design and development process features several touch points to support the synchronicity of the UX and development tracks. In practice, UX contributors experience alignment issues. For instance, features are released without a final UX sign-off performed. One interviewee states that sometimes small improvements and compromises get lost, cumulatively impacting the UX of the product. The product owner and UX contributors identify limited personnel and time resources as a reason leading to these process deviations. Overall, iterative handoffs between UX and development are common. Even during the implementation sprint, adjustments are made to the previously created designs. While regular change is characteristic for agile software development, both UX contributors as well as developers sometimes report this as unexpected rework. The product owner mentions: “Developers prefer the path of least resistance. They don't want to refactor what they have done before. That's normal. [...] So, you have to be prepared to explain the reasons for the change.” As a result of this, UX improvements that do not directly impact the functionality may be rejected because of the effort required to implement them. The UX lead concurringly mentions that UX improvements are difficult to achieve unless they were included in the initial design phase. This leads to a perceived lack of UX investment in certain products.

Collaboration: Two main styles of collaboration are evident during the case study: formal and informal collaboration. The formal collaboration mainly entails fixed regular meetings. Additionally, a significant part of the col-

laboration is supported by informal means of communication. These include ad-hoc calls between UX contributors and developers as well as conversations via chat messages for quick discussions. This proves to be particularly valuable when facing roadblocks during development, as mentioned by the developer. Informal collaboration also resonates well with the product owner, who mentions quick refinement of requirements in close collaboration with UX contributors. All consulted roles agree that personal connections and relationships are key to successful collaboration. Especially the UX lead and the UX managers highlight the importance of a deep connection and close relationships between UX contributors, product owners, and developers. Without these strong relationships UX is often rather seen as a blocker than a benefit, as the UX lead has experienced. Therefore, the UX lead encourages UX contributors to build up strong partnerships with their respective developers.

Besides building strong relationships, another enabler for effective collaboration mentioned by the interviewees is gaining general knowledge about the respective other roles that they are working with. The UX contributors report that collaboration especially works well when the responsible product owner is well-versed around UX and consequently values UX practices. The interviewed product owner expresses the desire to have a role within the development teams to “champion” UX topics to convey the importance of UX among the developers. The UX lead also emphasizes the importance of developers understanding the value of good UX and the need to work hand in hand to achieve it. Since developers are initially not trained regarding the practices of UX contributors and vice versa, it takes time for the roles involved to adapt and to effectively collaborate, as stated by the interviewed developer.

UX Priority: According to both UX managers, UX is considered very important in the organization, but there are challenges in fully integrating it into the product development. The main challenge identified is the partly low prioritization of UX within the product development processes. The priority of UX related work is mainly influenced by the respective product owners. The interviewed product owner as well as the UX contributors and the UX lead note that the influence of UX within the different product development teams varies significantly due to the different preferences of the individual product owners. The UX lead summarizes: “UX is often considered important in theory but then tends to be deprioritized in practice.”

The interviewed product owner considers UX a top priority but acknowledges that prioritizing UX too high can also result in challenges of deprioritizing non-UX related tasks. The UX contributors furthermore mention that UX related processes such as the product area internal product design and development process are mainly enforced within the UX team but not fully adopted by the product development teams. Hence, UX contributors wish for a

more stable representation within the processes of the product development teams.

FRAMEWORK OF INFLUENTIAL FACTORS AND IMPLICATIONS

To design a framework, we aimed to combine the insights from our literature research (knowledge base) and case study (environment). We therefore clustered the insights to visualize connections and contradictions between literature and practice and to identify mutual relations between the researched aspects. Based on these clusters we consolidated the insights into actionable factors, iteratively evaluated by the experience from the case study (design cycle of the Design Science Research methodology).

As a result of the consolidation, we identify the following as the main factors of the organizational integration of UX contributors in agile software development:

- (1) Organizational structure of agile product development teams and UX contributors.
- (2) Degree of autonomy of the teams involved.
- (3) Priority given to UX in agile software development processes.
- (4) Agility of UX processes and practices.
- (5) Mutual integration of agile software development and UX processes.
- (6) Mutual knowledge about the roles involved.
- (7) Influence of collaboration and relationships between the roles involved.

It can be said that the identified factors are either structure-related or process-related. Furthermore, they are either characteristics of UX, of agile software development, or of the interface between the two. The identified factors contribute to the success of the integration by improving the effectivity and efficiency of software development.

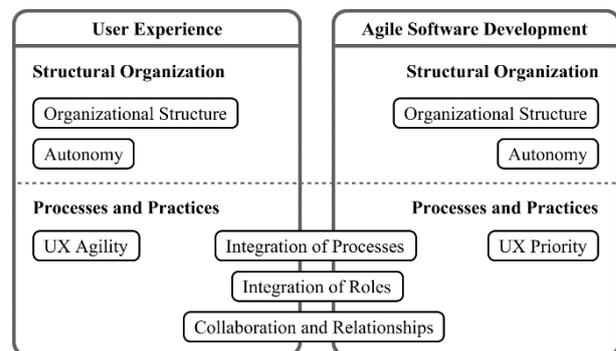


Figure 2: Influential factors for the organizational integration of UX contributors in agile software development

Structural Organization

(1) *Organizational Structure* – An organizational design that allows for the addition of collaborative structures

serves as a foundation for the effective integration of UX contributors in agile software development:

The fully agile unit, consisting of tribes and squads, serves as a systematic model to represent the generic structure of agile software development teams (Goncalves and Lopes 2014). With the concepts of chapters, guilds and shared teams, the fully agile unit provides multiple ways to integrate the core product development teams with additional structures.

The structural integration of UX contributors requires a balance between the comprehensiveness of shared UX teams and the sense of direct involvement of UX contributors in product development teams:

UX contributors are not specifically represented within the fully agile unit model. However, a set of typical organizational structures with regards to UX is introduced within this paper. These include centralized, distributed, matrix, and federated approaches (Kaplan and Pernice 2019; Rohn 2007; Szóstek 2012; Rosenberg 2019). The evaluation of these structures shows advantages and disadvantages of each of them when integrated into agile software development. The centralized approach facilitates consistency and broad UX skills but limits in-depth product knowledge and agility. While distributed UX contributors across different product development teams can compensate for these disadvantages by being deeply involved in the product, their ability for consistency and breadth of UX activities is limited.

Matrix structures and federated approaches aim to combine the advantages of both by applying centralized as well as distributed aspects to the organizational structure. Especially in terms of agility, they benefit from having UX contributors or smaller UX teams directly integrated in product development teams or product areas. Their ability to quickly adapt to changing product requirements is enhanced by deep product knowledge. At the same time, the individual UX contributors benefit from the support of a larger group of UX contributors around them. However, using a matrix or federated structure increases the complexity of the organizational structure when viewed in the context of the product development team structure. The case study shows that the organizational design of product development teams and UX teams can differ substantially. Furthermore, the case study shows that a federated structure consisting of product area specific UX teams can limit the actual integration of UX contributors into the product development teams by reducing the sense of belonging.

Finally, both the theoretical research and the conducted case study show that the different organizational structures regarding UX all have advantages and disadvantages. Based on the contrasting characteristics of the centralized and distributed approaches, it is advisable to adopt a structure that combines both approaches. However, it is then crucial to maintain a balance between deep integration and supportive central backing.

(2) Autonomy – Balancing autonomy and connectedness of the involved teams enables effective and efficient collaboration:

Moe et al. (2008) introduce autonomy as a characteristic of self-organizing teams, common to agile software development. They recommend a balanced amount of external, internal, and individual autonomy for agile software development teams (Moe, Dingsøyr, and Dybå 2008, 82–83). However, based on the findings of the case study, autonomy should always be balanced with connectedness to foster collaboration and prevent the emergence of silos. Regarding the integration of UX and agile software development, the case study shows that external autonomy is beneficial in terms of enabling the different working styles of the teams involved. However, as the integration of UX contributors with the product development teams is critical, autonomy should not limit the connectivity of the roles involved. Since external autonomy can amplify the feeling of disconnection, it should be balanced with shared processes. Internal and individual autonomy, on the other hand, support collaborative decision-making within the teams involved and should therefore be encouraged.

Processes and Practices

(3) UX Priority – The prioritization of UX by product management enables the effective involvement of UX contributors in the software development:

Since none of the major agile software development methodologies explicitly incorporate UX contributors, pure agile software development processes do not reserve time and resources for involving UX practices (Kane 2003, 1). To integrate UX, it is therefore necessary to prioritize UX tasks within the entirety of tasks performed during the development process. The case study shows that this UX priority is mainly influenced by the product owners, who determine the criticality of the tasks to be performed. UX should hereby not fall behind other development tasks. While organizations with a high level of UX maturity naturally set a high priority to UX, organizations with a lower level of UX maturity tend to de-prioritize UX due to time and resource constraints. Without a high UX priority, UX contributors are limited in their ability to influence the product. In addition to product management, UX management can also influence the prioritization of UX by raising awareness of the importance of UX.

(4) UX Agility – Incorporating agile considerations such as iterative and incremental design into the UX processes enables UX contributors to efficiently contribute to an agile software development process:

While improvements in UX prioritization help bring agile software development closer to UX considerations, UX contributors should also incorporate agile considerations into the way they work. As Kieffer et al. (2017) point out, UX contributors are used to creating a design for the entire product. Working on tightly constrained features instead, as is common in agile software develop-

ment, requires process changes in the way UX contributors work (Kieffer, Ghouti, and Macq 2017, 578). While the big picture of the product vision should still be top of mind, iterations should only focus on the current solution to a specific identified problem. Furthermore, for agile processes it is crucial to acknowledge that requirements can change frequently and even at a late point in time during development, as mentioned in the second principle of the Agile Manifesto (Beck et al. 2001b). The case study highlights that this frequency of change can be unfamiliar to UX contributors, who may feel overwhelmed by changing requirements. However, the iterative nature of UX processes supports the increasing agility of UX activities (Beyer 2010, 11).

(5) Integration of Processes – Combined practices enable UX, product management and development to efficiently co-create the product while maintaining process agility:

The theoretical research of this paper shows that specific processes and practices should be considered for the integration of UX in agile software development.

Emphasizing a common big picture of the product to be developed improves the outcome of the collaboration between UX contributors, product management, and development. While agile methods focus primarily on short-term feature development (Larusdottir, Cajander, and Gulliksen 2012, 6), an overall product vision is necessary to achieve a consistent and thus positively perceived UX of the product (Kuusinen 2014, 266). This not only enables the UX contributors to create consistent designs, but also supports the product management in defining coherent features and precise requirements as the case study shows. The vision should be developed jointly by all involved roles, keeping them engaged from early on. This also encourages all roles to feel equally involved and to identify with the product – increasing their engagement and consequently the quality of the outcome (Chamberlain, Sharp, and Maiden 2006, 152).

To improve the direct collaboration between UX contributors and development, their individual processes must be interwoven (Brhel et al. 2015, 17–19). Working in parallel tracks on the same feature sets helps achieve this (Larusdottir, Cajander, and Gulliksen 2012; Sy 2007; Miller 2005; Salah, Petrie, and Paige 2009). To ensure effective transfer of UX considerations into implementation, the research recommends giving UX contributors a lead time to prepare designs, ideally one sprint ahead of the implementation (Sy 2007, 118). In practice, as the case study shows, these lead times for UX can also be extended to provide additional time for thorough research and ultimately improve the quality of the designed solution. However, excessively extending the time difference between UX work and development can hinder the synchronicity between the two tracks and consequently reduce the agility of the process, as retrieved from the case study.

During their parallel work, UX contributors and developers focus on the same feature sets derived from the product vision. Agile principles are applied through iterative and incremental approaches (Kieffer, Ghouti, and Macq 2017, 578). However, the case study indicates that iterative changes can sometimes be perceived negatively by both developers and UX contributors. In particular, unsubstantiated requests for change or frequently changing requirements frustrate the UX contributors. Requests for design changes after the implementation furthermore frustrate developers. To mitigate frustration, it is important for UX contributors and developers to be open to change and for product owners to thoroughly evaluate and explain changing requirements.

(6) Integration of Roles – Improving the mutual knowledge about collaborative roles within product development teams improves the effectiveness of the collaboration:

Both agile software development and UX employ specifically defined roles, introduced in this paper. When it comes to integrating the roles of UX contributors with the agile software development environment, these predefined roles must work jointly. According to Jones and Thoma (2019), a thorough understanding of each other's role significantly improves the ability to collaborate (Jones and Thoma 2019, 35–36). This is further supported by the findings of the case study, which emphasize that UX contributors, product owners, and developers must first understand each other's roles in order to develop effective ways to support each other and utilize the benefits of each other's work.

(7) Collaboration and Relationships – Informal organization in the form of strong individual relationships between the collaborating roles increases the efficiency and effectiveness of the collaboration and compensates for possible compromises in the formal organization:

Theory emphasizes that traditional software development relies on formal handoff processes, while agile development favors iterative and lightweight collaboration between contributors and teams. Larusdottir et al. (2012) find that in agile environments, most collaboration between UX contributors and developers occurs through informal practices, emphasizing the need for close personal connections among team members (Larusdottir, Cajander, and Gulliksen 2012, 6). Furthermore, the fourth principle of the Agile Manifesto states that “business people and developers must work together daily throughout the project” (Beck et al. 2001b). Based on the insights from this paper, this principle should also be applied to additional roles such as UX contributors, working with both business and development. Establishing these connections leads to early and frequent communication and enables faster problem identification, problem solving, idea generation, and decision making between UX contributors and developers.

In practice, the case study confirms two main styles of collaboration: formal and informal. Formal collaboration

involves regular meetings where product owners, developers, and UX contributors come together to exchange on the current state of their work. The case study highlights that informal communication plays a significant role in supporting collaboration. Ad-hoc meetings and chat messages allow for quick efficient discussions, particularly beneficial when fast decisions are required during iterations. Furthermore, close connections allow for rapid refinements of requirements through informal collaboration between UX contributors and product owners.

It can be concluded that personal connections and relationships are essential for successful collaboration across different roles, especially when located in structurally separated teams. Strong relationships between UX contributors, product owners, and developers are crucial for overcoming the perception of UX as an obstacle rather than a valuable asset.

Overall, the collaboration should not be overly dictated by strict processes. Strong interpersonal relationships should take precedence over strict adherence to rigid organizational structures. Especially in a federated organizational structure, effective communication plays a key role in successful collaboration across multiple teams involved. Prioritizing close collaboration is further supported by the first value of the Agile Manifesto, which values “individuals and interactions over processes and tools” (Beck et al. 2001a). More precisely, the sixth principle of the Agile Manifesto states that “the most efficient and effective method of conveying information to and within a development team is face-to-face conversation” (Beck et al. 2001b). The research shows that this principle also applies to the collaboration between UX contributors, product owners, and developers.

Levels of Organizational Influence

The identified factors can be influenced either on an individual, team-internal, or team-external level. While the organizational structure of both the agile product development teams as well as the UX contributors is mainly predetermined by the structure of the wider product area (external organization), the specific integration of processes and roles can be influenced by the management of the individual teams (team organization). Furthermore, the priority given to UX, the agility of UX and the autonomy of the involved teams are heavily influenced by the directly involved team management (team organization). Lastly, collaboration and building relationships among the involved contributors from product management, development and UX directly depends on the individual involved team members and can hence be classified as an aspect of self-organization.

This classification shows that the integration is not only determined by external organization but can be substantially influenced by the direct management (UX and product management) and further by the individuals working in the teams. As the research shows, especially self-organization, focusing on building relationships and

fostering direct collaboration, crucially improves the integration of UX contributors and agile product development teams.

Application as Critical Success Factors

To evaluate the usefulness of this framework with regards to the practical management of software development organizations, we draw an interrelation to the concept of critical success factors. According to Rockart (1979), critical success factors are “the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization” (Rockart 1979). Hence, these factors must be focused and continually measured by management (Rockart 1979).

Our process of deriving and evaluating the influential factors for the successful integration of UX contributors in agile software development can be compared to the process of deriving critical success factors (described by Cooper 2008, 4). The process consists of an introductory workshop, interviews, and a focusing workshop for consolidation. Based on our research, the influential factors that we outline are in fact crucial for the success of an integration. Necessary management focus is especially expressed in factors such as UX priority, UX agility and integration of processes and roles. Continuous measurement of these factors seems beneficial for the overall success. Therefore, we can infer that our identified influential factors are actionable factors that support the successful integration of UX contributors into agile software development. However, to completely fulfill the requirements of critical success factors, according to Rockart (1979), a proof of extensiveness of the identified factors needs to be researched.

CONCLUSIONS

Synthesizing our research with regards to the objectives set at the beginning, we can conclude the following:

What is state of the art for integrating UX in agile software development?

Agile methodologies are widely used in the field of software development. With the increasing awareness of the importance of UX in this field, literature research shows that organizations now face challenges in effectively enhancing agile software development with the benefits of UX considerations. Our research confirms that the principles of agile software development and UX are indeed compatible. However, to effectively incorporate UX considerations into agile software development methodologies, organizational aspects of integrating UX contributors and agile product development teams need to be considered. Since prominent agile methodologies do not include UX contributors in their models, new models for the integration need to be defined. To start the development of such models, this paper has investigated the factors that need to be considered for the integration.

Which organizational factors influence the integration of UX contributors in agile software development?

Organizational theory indicates two perspectives on organization: the structural perspective and the process perspective. Throughout this paper, this distinction proved to be appropriate to explore the factors of the organizational integration in a structured manner. In terms of structural factors, our research has shown that organizational design has a direct impact on the integration of UX contributors in agile software development. In particular, the organizational structures concerning the product development teams and UX contributors, the roles involved, and the intended autonomy of these were identified as the main aspects of the integration in terms of structural organization. From a process perspective, our research identified several areas of practices that need to be considered to support an effective integration. These include the product vision (jointly developing a big picture, aiming for little design upfront), synchronicity of UX and development work (parallel and iterative tracks), and process agility. Collaboration was identified as a critical factor, with formal and informal approaches emphasized in both the literature research and the case study.

Which implications for the integration of UX contributors in agile software development can be derived from these factors?

The aspects identified before were used to create a framework of influential factors for the integration of UX contributors in agile software development. Based on the findings from the literature review and the case study, the factors were enhanced with specific implications. Recognizing that the application of organizational structures always balances benefits and drawbacks, our research shows that integrated processes and practices can effectively mediate these drawbacks and compensate for compromises that must be made in the rigid organizational structure. The main areas that these processes should cover were derived from principles set in literature. The case study mainly confirmed the usage of these practices in practice. Finally, the research showed that organizational structure does not only cover factors that can be influenced by management or are externally provided, but there is a significant aspect of self-organization involved. Both, the theoretical research and the case study showed that individually established informal relationships between the roles involved significantly improves the effectiveness of the integration of UX contributors and agile product development teams.

Limitations and Future Research

To ensure the broader applicability of the framework, the factors and implications found in this paper should be validated through additional studies. The case study conducted was limited to one particular integration of UX contributors in a specific organizational setting. Thus, aspects of future research should entail to conduct similar studies in additional organizations, covering different organizational conditions. Furthermore, the amount and diversity of the consulted participants in such studies

should be increased to ensure statistic generalizability. With the factors and implications validated and further enhanced through additional studies, the herein developed framework can then comprehensively support practitioners to improve the organizational integration of UX contributors in agile software development.

LITERATURE

- Beck, Kent, Mike Beedle, Arie van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning, et al. 2001a. "Manifesto for Agile Software Development." 2001. <http://agilemanifesto.org/>.
- . 2001b. "Principles behind the Agile Manifesto." 2001. <http://agilemanifesto.org/principles.html>.
- Beyer, Hugh. 2010. *User-Centered Agile Methods*. Penn State: Morgan & Claypool Publishers.
- Brhel, Manuel, Hendrik Meth, Alexander Maedche, and Karl Werder. 2015. "Exploring Principles of User-Centered Agile Software Development: A Literature Review." *Information and Software Technology* 61 (May). <https://doi.org/10.1016/j.infsof.2015.01.004>.
- Bruun, Anders, Marta Kristin Larusdottir, Lene Nielsen, Peter Axel Nielsen, and John Stouby Persson. 2018. "The Role of UX Professionals in Agile Development: A Case Study from Industry." In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, 352–63. Oslo Norway: ACM. <https://doi.org/10.1145/3240167.3240213>.
- Chamberlain, Stephanie, Helen Sharp, and Neil Maiden. 2006. "Towards a Framework for Integrating Agile Development and User-Centred Design." In *Extreme Programming and Agile Processes in Software Engineering*, edited by Pekka Abrahamsson, Michele Marchesi, and Giancarlo Succi, 4044:143–53. Berlin, Heidelberg: Springer. https://doi.org/10.1007/11774129_15.
- Chapman, Lorraine, and Scott Plewes. 2014. "A UX Maturity Model: Effective Introduction of UX into Organizations." In *Design, User Experience, and Usability. User Experience Design Practice*, edited by Aaron Marcus, 12–22. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-07638-6_2.
- Cockburn, Alistair, and Jim Highsmith. 2001. "Agile Software Development, the People Factor." *Computer* 34 (11): 131–33. <https://doi.org/10.1109/2.963450>.
- Conway, Melvin E. 1968. "How Do Committees Invent?" *Datamation Magazine* 14 (4): 28–31.
- Cooper, Vanessa. 2008. "The Critical Success Factor Method: A Review and Practical Example." In .
- Damaševičius, Robertas. 2010. "On The Human, Organizational, and Technical Aspects of Software Development and Analysis." In *Information Systems Development: Towards a Service Provision Society*, edited by George Angelos Papadopoulos, Wita Wojtkowski, Gregory Wojtkowski, Stanislaw

- Wrycza, and Jože Zupancic, 11–19. Boston, MA: Springer US. https://doi.org/10.1007/b137171_2.
- Dima, Alina Mihaela, and Maria Alexandra Maassen. 2018. “From Waterfall to Agile Software: Development Models in the IT Sector, 2006 to 2018. Impacts on Company Management.” *Journal of International Studies* 11 (2): 315–26. <https://doi.org/10.14254/2071-8330.2018/11-2/21>.
- Ferreira, Jennifer, Helen Sharp, and Hugh Robinson. 2012. “Agile Development and User Experience Design Integration as an Ongoing Achievement in Practice.” In *2012 Agile Conference*, 11–20. Dallas, TX: IEEE. <https://doi.org/10.1109/Agile.2012.33>.
- Gerster, Daniel, Christian Dremel, Walter Brenner, and Prashant Kelker. 2020. “How Enterprises Adopt Agile Forms of Organizational Design: A Multiple-Case Study.” *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* 51 (January): 84–103. <https://doi.org/10.1145/3380799.3380807>.
- Gerster, Daniel, Christian Dremel, and Kelker Prashant. 2018. “Scaling Agility: How Enterprises Adopt Agile Forms of Organizational Design.” In . San Francisco, CA.
- Gonçalves, Ester, and Eurico Lopes. 2014. “Implementing Scrum as an IT Project Management Agile Methodology in a Large Scale Institution.” In , 461. Academic Conferences International Limited.
- Gray, Colin M., Austin L. Toombs, and Shad Gross. 2015. “Flow of Competence in UX Design Practice.” In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3285–94. CHI '15. New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/2702123.2702579>.
- Hevner, Alan, and Samir Chatterjee. 2010. “Design Science Research in Information Systems.” In *Design Research in Information Systems*, by Alan Hevner and Samir Chatterjee, 22:9–22. Integrated Series in Information Systems. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-5653-8_2.
- Hevner, Alan, Salvatore March, Jinsoo Park, and Sudha Ram. 2004. “Design Science in Information Systems Research.” *MIS Quarterly* 28 (1): 75–105.
- ISO 9241-210:2019. 2019. “Ergonomics of Human-System Interaction – Part 210: Human-Centred Design for Interactive Systems.” International Organization for Standardization.
- Jones, Alexander, and Volker Thoma. 2019. “Determinants for Successful Agile Collaboration between UX Designers and Software Developers in a Complex Organisation.” *International Journal of Human-Computer Interaction* 35 (20): 1914–35. <https://doi.org/10.1080/10447318.2019.1587856>.
- Jurca, Gabriela, Theodore D. Hellmann, and Frank Maurer. 2014. “Integrating Agile and User-Centered Design: A Systematic Mapping and Review of Evaluation and Validation Studies of Agile-UX.” In *2014 Agile Conference*, 24–32. Kissimmee, FL: IEEE. <https://doi.org/10.1109/AGILE.2014.17>.
- Kane, D. 2003. “Finding a Place for Discount Usability Engineering in Agile Development: Throwing down the Gauntlet.” In *Proceedings of the Agile Development Conference, 2003. ADC 2003*, 40–46. Salt Lake City, UT: IEEE. <https://doi.org/10.1109/ADC.2003.1231451>.
- Kaplan, Kate, and Kara Pernice. 2019. “Where Should UX Report? 3 Common Models for UX Teams and How to Choose Among Them.” Nielsen Norman Group. 2019. <https://www.nngroup.com/articles/ux-team-models/>.
- Kashfi, Pariya, Robert Feldt, and Agneta Nilsson. 2019. “Integrating UX Principles and Practices into Software Development Organizations: A Case Study of Influencing Events.” *Journal of Systems and Software* 154 (August): 37–58. <https://doi.org/10.1016/j.jss.2019.03.066>.
- Kieffer, Suzanne, Aissa Ghouti, and Benoit Macq. 2017. “The Agile UX Development Lifecycle: Combining Formative Usability and Agile Methods.” *Hawaii International Conference on System Sciences 2017 (HICSS-50)*, January. <https://aisel.aisnet.org/hicss-50/cl/hci/7>.
- Kniberg, Henrik, and Anders Ivarsson. 2012. “Scaling Agile @ Spotify with Tribes, Squads, Chapters & Guilds.” <http://www.agileleanhouse.com/lib/lib/People/HenrikKniberg/SpotifyScaling.pdf>.
- Kuusinen, Kati. 2014. “Improving UX Work in Scrum Development: A Three-Year Follow-Up Study in a Company.” In *Human-Centered Software Engineering*, edited by Stefan Sauer, Cristian Bogdan, Peter Forbrig, Regina Bernhaupt, and Marco Winckler, 259–66. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-44811-3_17.
- Kuusinen, Kati, and Kaisa Väänänen-Vainio-Mattila. 2012. “How to Make Agile UX Work More Efficient: Management and Sales Perspectives.” In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, 139–48. NordiCHI '12. New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/2399016.2399037>.
- Larusdottir, Marta, Åsa Cajander, and Jan Gulliksen. 2012. “The Big Picture of UX Is Missing in Scrum Projects.” *CEUR Workshop Proceedings 922* (January): 42–48.
- McInerney, Paul, and Frank Maurer. 2005. “UCD in Agile Projects: Dream Team or Odd Couple?” *Interactions* 12 (6): 19–23. <https://doi.org/10.1145/1096554.1096556>.
- Miller, L. 2005. “Case Study of Customer Input for a Successful Product.” In *Agile Development Conference (ADC'05)*, 225–34. Denver, CO: IEEE Comput. Soc. <https://doi.org/10.1109/ADC.2005.16>.
- Moe, Nils Brede, Torgeir Dingsøy, and Tore Dybå. 2008. “Understanding Self-Organizing Teams in

- Agile Software Development.” In *19th Australian Conference on Software Engineering (Aswec 2008)*, 76–85. Perth.
<https://doi.org/10.1109/ASWEC.2008.4483195>.
- Nagappan, Nachiappan, Brendan Murphy, and Victor Basili. 2008. “The Influence of Organizational Structure on Software Quality: An Empirical Case Study.” In *Proceedings of the 13th International Conference on Software Engineering - ICSE '08*, 521. Leipzig: ACM Press.
<https://doi.org/10.1145/1368088.1368160>.
- Rockart, John F. 1979. “Chief Executives Define Their Own Data Needs.” *Harvard Business Review* 57 (2): 81–93.
- Rohn, Janice Anne. 2007. “How to Organizationally Embed UX in Your Company.” *Interactions* 14 (3): 25–28. <https://doi.org/10.1145/1242421.1242440>.
- Rosenbaum, Stephanie, President Janice, Anne Rohn, and Director Humburg. 2000. “A Toolkit for Strategic Usability: Results from Workshops, Panels, and Surveys.” *CHI Letters* 2 (1): 337–44.
<https://doi.org/10.1145/332040.332454>.
- Rosenberg, Daniel. 2019. “The Business of UX Management.” *Interactions* 26 (3): 28–35.
<https://doi.org/10.1145/3318131>.
- Salah, Dina, Helen Petrie, and Richard F Paige. 2009. “Towards a Framework for Integrating User Centred Design and Agile Software Development Processes.” In *Extreme Programming and Agile Processes in Software Engineering*, edited by Abrahamsson, P., Marchesi, M., Succi, G. Vol. 4044. Berlin, Heidelberg: Springer.
- Silva da Silva, Tiago, Milene Selbach Silveira, Frank Maurer, and Theodore Hellmann. 2012. “User Experience Design and Agile Development: From Theory to Practice.” *Journal of Software Engineering and Applications* 05 (10): 743–51.
<https://doi.org/10.4236/jsea.2012.510087>.
- Sy, Desirée. 2007. “Adapting Usability Investigations for Agile User-Centered Design.” *Journal of Usability Studies* 2 (3): 112–32.
- Szóstek, Agnieszka. 2012. “A Look into Some Practices behind Microsoft UX Management.” In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. Vancouver, BC: ACM.
https://dl.acm.org/doi/abs/10.1145/2212776.2212833?casa_token=SO8cRAduSLIAAAAA:g7gWih0qVID9IB5GHvZkgdaOOXUHtuzjTF71VvBT1Iz0Qx_BanL_KUPWB-CPJ8cfk2N713qXNgk.

On the Impact of a Business Intelligence System on Analysis Effort: A Case Study of Flensburg Municipality in Germany

Kai Petersen
Hochschule Flensburg
Fachbereich Wirtschaft
Kanzleistraße 91-93
24943 Flensburg
kai.petersen@hs-
flensburg.de

Amelie Wasse
Stadt Flensburg
Zentrale Dienste
Rathausplatz 1
24937 Flensburg
Wasse.Amelie@
Flensburg.de

Tim Cruse
Hochschule Flensburg
Fachbereich Wirtschaft
Kanzleistraße 91-93
24943 Flensburg
timcruse1@gmail.com

Jacob Sietas
Hochschule Flensburg
Fachbereich Wirtschaft
Kanzleistraße 91-93
24943 Flensburg
mail@jacobsietas.de

Jan M. Gerken
Hochschule Flensburg
Fachbereich Wirtschaft
Kanzleistraße 91-93
24943 Flensburg
jan.gerken@hs-
flensburg.de

Keywords

business intelligence, resource efficiency, process analysis, impact

ABSTRACT

Background: Business intelligence provides decision support to decision-makers in public administrations. So far, researchers have studied the use of business intelligence systems concerning, for example, efficiency gains of the organization. However, based on our literature study, researchers have yet to focus on the impact of business intelligence systems on the efficiency of analytical processes in the public administration context.

Objective: This study's objective is to investigate the impact of introducing a business intelligence system in public administration.

Method: A case study was conducted to investigate the impact. We studied four departments at Flensburg Municipality in Germany. We used interviews and analyzed documentation to describe the analytical processes and to estimate efficiency gains.

Results: The BI system led to efficiency gains across departments. The gains varied from 1.77% to 47%. Losses were observed regarding media changes.

Conclusions: The impact on efficiency varies with the attributes of the analytical process (e.g., repetitions in the process). Hence, further investigations are needed to build theories regarding the relationship between efficiency and

the use of BI systems in public administrations.

INTRODUCTION

Even though business intelligence systems are now used in various areas in public administration, this area is still seen as a relevant field of research McDonald III et al. (2022). The relevance can be emphasized by the fact that public administrations make decisions with far-reaching consequences that affect not only their own organizations, but entire administrative districts - be it states, federal states, or municipalities. The effect on entire administrative districts emphasizes the need for complete and comprehensive data for decision-making processes in public administration Abai et al. (2015).

The information silos caused by clear departmental boundaries in public administrations make it challenging to share information Teixeira et al. (2014). Business intelligence systems can make a contribution by providing relevant information to all stakeholders through a single point. This can reduce the process-related effort of information procurement and evaluation as well as the number of system breaks and thus contribute to process efficiency in public administration. However, it still needs to be determined to what extent BI systems can make a positive contribution to improving the efficiency of decision-making processes in the public administration.

Most attention in the past has been drawn to the level of organizations. Less attention has been paid to the level of individuals Trieu (2017). In this study, we examined the efficiency of BI systems at the individual level by having process participants estimate execution times.

Using a case study approach, the efficiency of BI systems in four different departments of the Flensburg city administration, focusing on the estimated and perceived impact on

time efficiency. In each department, we examined two versions of the same process. First, process experts estimated the execution time without the use of an implemented BI system. Afterward the execution time for the process with the use of an implemented BI system was estimated.

The results show that implementing a BI system in public administration can lead to efficiency gains. However, challenges with data integration and report generation have also become apparent. In both areas, there is further potential for automation and standardization.

The remainder of the paper is structured as follows: Section presents background information and related work. Section describes the methodology followed for studying the impact of the BI system. Section presents the results of our study. Section discusses practical and research implications. Section concludes the paper.

RELATED WORK

Theoretical Framework

The theoretical framework of this literature emerges from four perspectives, namely (i) the domain perspective, (ii) the impact perspective, (iii) the output perspective and (iv) the information system type perspective.

(i) Domain perspective: Competitive pressure is of secondary importance for the success of BI systems Zaied et al. (2018). This is especially true in public administration. Public administrations differ from private companies in their mission. Public administrations are not evaluated based on profits and losses Abai et al. (2015). Moreover, they perform sovereign tasks, so they are not in a competitive situation either. However, administrations are also subject to financial restrictions. Internal process efficiency is, therefore, of particular importance for public administration.

(ii) Impact perspective: The impact of business intelligence systems has been investigated for private sector companies in various studies (for a detailed literature review see Trieu 2017).

BI systems add value at both the operational and strategic level Yogev et al. (2013). Al-Eisawi et al. (2020) found a positive impact of BI system efficiency on organizational efficiency. The organizational efficiency of an organization can be defined as its ability to use resources optimally for their intended purpose Al-Eisawi et al. (2020).

Elbashir et al. (2011) identified four different types of BI systems' benefits, including (i) organizational benefits, (ii) business supplier / partner relation benefits, (iii) internal process efficiency benefits and (iv) customer intelligence benefits. While the first factor relates to benefits at the organizational level, the remaining factors relate to benefits at the business process level. Benefits related to internal process efficiency include the following items: a) improved efficiency of internal processes, increase in employee productivity, c) reduction of the cost of effective decision-making, and d) reduced operational costs. The authors found a positive relationship between business process performance and organizational performance.

These results are partly consistent with the findings of Trieu (2017). He cites (i) minimising the mistargeting of customers, (ii) transforming business processes, (iii) enriching organizational intelligence, and (iv) developing new or improving products or services as the possible positive impacts of BI systems. With regard to the tasks of public administration, it seems obvious that especially the dimension of process efficiency is of particular importance for public administration. In the context of this paper, we focus on efficiency to distinguish it from effectiveness and possibly other impacts.

(iii) Output perspective: In the past, numerous papers have been written on success factors of business intelligence systems (for comprehensive literature reviews see Gaardboe & Jonasen 2018, Magaireah, HidayahSulaiman & Ali 2019, Zaied et al. 2018). Many studies cite top management support and data quality as critical success factors for BI systems in general Zaied et al. (2018), Magaireah, HidayahSulaiman & Ali (2019). In contrast, it was also shown that data quality is of little importance for public administration Magaireah, HidayahSulaiman & Ali (2019). However, information quality has proven to be an important dimension for the success of BI systems in public administration, with the elements of security (information quality) and availability (information quality) being of utmost importance, while the elements of accuracy, comprehensibility and completeness are only of medium importance Zaied (2012). There are also divergent views on the importance of top management support, which assess its role as less important Zaied (2012).

Although the findings may be relevant for discussing our results, they are irrelevant to the literature review. Critical success factors do not consider the impact of BI systems in general nor resource efficiency in particular. They represent input factors that are not relevant for our literature research. Therefore, publications related to input factors are considered irrelevant. However, the divergent results for public administration show that BI systems in public administration must be subjected to a separate analysis.

(iv) Information system type perspective: A BI system can be considered a type of information system. This is manifested by the fact that research on business intelligence system adoption, utilization, and success mostly uses framework models for the evaluation of information systems in general, as Ain et al. (2019) has shown in a systematic literature review. The authors Ain et al. (2019) include a total of 123 relevant essays in their analysis. They identify 16 papers that use the DeLone & McLean IS success model. A further 15 papers use the technology acceptance model. Both models are suitable for evaluating and analyzing information systems in general. This result, especially the high importance of the TAM, was confirmed in another publication Chi & Mahmud (2020).

This is also confirmed by Zakaria et al. (2011). They identified the efficiency and effectiveness of work processes as a Key Performance Indicator to deliver services for the National Registration Department (NRD), Pulau Pinang.

Even if a broad perspective is conceivable, we focus on BI systems and exclude other types of information systems.

Literature Review

The literature search comprises two steps and consists of the actual literature search and the screening of the publications found. The first step restricts the search results to the domain of public administration, the information system type Business Intelligence and resource efficiency perspective in a broad sense. In the second, the general relevance of the researched articles for the present work is examined. A distinction is made between the input-oriented and output-oriented evaluation of BI systems. In addition, we examine whether process efficiency is addressed.

Ain et al. (2019) conducted a systematic literature review on BI system adoption, utilization, and success. For this purpose, they developed a search strategy that we adapted in this analysis to search for publications on BI systems in public administration. The terms related to BI systems were adopted. We have not included the terms that contain other search terms as sub-terms. For example, we have omitted the search term "business intelligence system" because the term "business intelligence" is included as a search term. We also restricted our search on terms related to the system level and excluded terms that are only related to certain aspects of a business intelligence system (such as OLAP, data warehouse, and dashboard). We restricted the search to the domain of public administration. For this purpose, search terms related to public administration were added.

Our search strategy (see for details Table 1) aims at a high precision. Therefore, the search was limited to searching titles, abstracts, and metadata. The exact design of the search strategy depended on the possibilities and the available fields of the search interfaces. If we have not received any search results, we have dispensed with the impact related terms. This ensures that finding no results was not caused by a character limitation of the search mask. For semantic scholar, we limited our search to one term per perspective and relied on the semantic search options (public sector, business intelligence, efficiency). Even though the reproducibility is limited, it adds another perspective to the literature search with standardised Boolean queries. In the case of SemanticScholar and Google Scholar, the screening was limited to the first 50 search results, whereby we relied on the relevance ranking.

The research did not lead to any evaluable results on the narrower topic of increased resource efficiency through business intelligence systems within public administration (see Table 2). Thus, no search results were obtained that systematically show how business intelligence systems affect the efficiency of work processes in terms of time saved in daily work within public administration. However, many results were obtained which highlight the input perspective (e.g. critical success factors) Magaireah et al. (2019), Hartley & Seymour (2015), Magaireah et al. (2017) or the output perspective (higher decision quality) Hočevar & Jaklič (2010), Boselli et al. (2011), Abai et al. (2015). Furthermore, we found support for the relevance of our research questions, as Fischer et al. Fischer et al. (2021) found in a systematic literature review that so far only a few studies deal with the impact of digitalisation in public administration.

METHODOLOGY

The case study was conducted as part of a research project with the city of Flensburg. The three-year research project focused on the introduction of a business intelligence system in public administration. We describe the case study following the guidelines by Runeson and Höst Runeson & Höst (2009), which were specifically designed for the software context (e.g., describing the case, units of analysis, information sources, etc.). The case study focused on the Flensburg municipality in Germany. Our units of analysis were processes involving multiple departments. The research aim was to look at the differences in process steps conducted before the introduction of the BI system and after its introduction.

Research goal and questions

The goal of the research is to determine effort savings through the introduction of a business intelligence system in municipalities. Thus, our main research question is: *What is the estimated impact on effort through the introduction of a business intelligence system in a municipality?* We provide data from a specific case and a set of analysis processes to achieve the goal. We also studied different departments. We analyzed the data from different perspectives, and hence formulated sub-questions:

1. What is the impact of the business intelligence system on the department level?
2. What is the impact of the business intelligence system on different activity categories?
3. What is the impact on individual activities required in the analysis process?

The departments, the business analysis system, and the analysis processes are described in the following.

Case and Context

We first describe the case (the city Flensburg municipality) and thereafter we briefly introduce the units of analysis (analysis process) and the departments participating in the process.

The case - Flensburg Municipality

Flensburg comprises a population of approximately 98000 people. The municipality employs 1.600 people. The municipality is organized into 12 different departments (e.g. finance, central services, education and sports office, etc.). As mentioned earlier, data collection was conducted in a total of four departments, which were involved in the analysis processes (units of analysis).

Departments

The four departments are briefly described below.

Department 1 (D1) - Traffic planning and regulation: The traffic planning and regulation department is responsible for the general planning of mobility and traffic guidance in the city of Flensburg. Traffic data and routing are a central, municipal management instrument. Traffic planning is of particular importance with regard to the desired change of the type of transport in the city.

Table 1: Search Strategy

Category	Content
Business Intelligence related terms	business intelligence, BI system
Public administration related terms	public administration, public sector, public management, municipality
Impact related terms	efficiency, time saving
Sources	IEEE Xplore, ScienceDirect, EmeraldInsight, Google Scholar, SemanticScholar, Scopus
Years	2000 - 11/2022
Search Fields	Title, Abstract, Metadata (if applicable)
Language	English

Table 2: Search Results

Source	Search Query	Results
Scopus	TITLE-ABS-KEY (("business intelligence" OR "bi system") AND ("public administration" OR "public sector" OR "public management" OR municipality) AND (efficiency OR "time saving"))	16
IEEE Xplore	((("All Metadata":"Business Intelligence" OR ("All Metadata":"BI system"))) AND (("All Metadata":"public administration") OR ("All Metadata":"public sector") OR ("All Metadata":"public management") OR ("All Metadata":"municipality"))) AND (("All Metadata":"efficiency") OR ("All Metadata":"time saving")))	9
ScienceDirect	Title, abstract or author-specified keywords: ("Business Intelligence"OR "BI system") AND ("public administration"OR "public sector"OR "public management"OR "municipality") AND ("efficiency"OR "time saving")	2
EmeraldInsight	((title:"Business Intelligence") OR (title:"BI System") OR (abstract:"Business Intelligence") OR (abstract:"BI System")) AND ((title:"public management") OR (title:"public sector") OR (title:"public administration") OR (abstract:"public management") OR (abstract:"public sector") OR (abstract:"public administration") OR (abstract:"municipality") OR (title:"municipality"))	35
Google Scholar	("Business Intelligence"OR "BI system") AND ("public administration"OR "public sector"OR "public management"OR "municipality") AND ("efficiency"OR "time saving")	approx. 16.800
SemanticScholar	"business intelligence"public sector"efficiency"	12.900

Department 2 (D2) - Statistics office: The statistics office of the city of Flensburg is responsible for processing and providing various statistical surveys and key figures. It acts as an internal and external point of contact for information on statistical data of the city. The business intelligence project is intended to automate and digitize the statistical data report, which the statistics office previously created.

Department 3 (D3) - Youth service planning: The youth service planning department is responsible for planning early intervention and youth services activities. In addition, an important part of the tasks is the assessment of the city's social situation. The social data prepared and presented by the department represent a strong management instrument for the city of Flensburg.

Department 4 (D4) - Education and sports office: The education and sports office is responsible for planning

and coordinating schools in the city area. This includes the schools on the one hand, but also the preceding sections, such as the daycare centers, and the succeeding sections, such as the transition to work, on the other hand. The data used by the education and sports office can be used to plan the future demand for teachers, the required number of classrooms to be provided, and childcare spots, for example.

Business Intelligence System

The Business Intelligence software QlikSense was used to implement the project requirements. QlikSense offers the possibility to create interactive dashboards and thus provides a self-service business intelligence system. This includes self-service visualizations that can be examined using search, select and drill down. At each step, an immediate update of all relevant tables and charts to the current context occurs. In addition, QlikSense allows drag and drop of tables, charts

and data fields in the front end. The Business Intelligence software caters to the need for a full visual front-end, permitting the creation of basic visualizations by non-professional users without the necessity of extensive training. In the project, QlikSense is being used as a mostly off-the-shelf product with few add-ons. While the company provides a cloud solution, QlikSense is deployed through an internal server on premise due to data safety and protection regulations in the public administration of Flensburg.

In public administration, mandatory reports must be created each year. These processes require a significant amount of data preparation. The created reports are then only available in a rigid form, so any dynamic work with the contained data is not possible. The Business Intelligence System allows easy web-based access to the data, and updating the data can be done with significantly less effort.

Within the traffic planning and regulation department, the possibilities offered by the Business Intelligence System include, for example, a digitized display of all traffic monitoring stations as well as a dynamic detailed view of each station. This includes the number of vehicles, speed, type of vehicle, time and date or the direction. For the statistics office, the Business Intelligence system provides insights into the general population data of the city. For the youth service planning department, the advantage results from the fact that the preparation and presentation of the data can take place with significantly less effort. Through the Business Intelligence system it is possible to do an automated update of tables and charts each year. For the education and sports office, the Business Intelligence system offers the advantage of free exploration of the number of students over a fairly large time span. The system provides a digital location of the schools in the city area and shows in which districts pupils live, for example.

Units of analysis

The units of analysis are different analytical processes. We have selected four processes that originate from different departments and are supported by the implemented BI system within the BI-F2022 project. By choosing processes from different departments, a cross-departmental comparison becomes possible. Choosing processes that are supported by a recently implemented BI system ensures that the assessment of impact is based on initial experience. Similarly, there is still a familiarity with the previous process flows. This allows for direct comparison by the experts. We looked at the processes *Analyze traffic data* (traffic planning and regulation), *Create population statistics* (statistics office), *Create social atlas* (youth services planning) and *Create school development planning report* (education and sports office). Processes were partially interdepartmental. For these processes, only the process steps of the respective department were considered.

Process 1 (P1) in D1 - Analyze traffic data: The *analyze traffic data* aims to collect, process, and analyse traffic data. As a result, processed traffic data is provided. These include, for example, time-related traffic counts and speed measurements. The process comprises five sub-processes. Four of these sub-processes describe different types of data collection in the course of traffic measurement. Traffic data

is taken from a variety of sources. Manual counts by pupils, speed displays, manual counts by employees of the administration, and traffic measurement devices are used. These sub-processes comprise a total of 34 activities, which serve, for example, the preparatory organisation, data collection, data evaluation, and data communication. For the *pupil count* sub-process, these include, for example, planning with the schools, preparing the traffic count documents, hiring the necessary vehicles, briefing the pupils, sending to and receiving Excel spreadsheets from the schools, and checking the data for plausibility. The fifth sub-process contains only one activity. This serves to answer questions from colleagues and planning offices.

Process 2 (P2) in D2 - Create population statistics:

The population statistics provides aggregated information on the demographic structure of the city of Flensburg. The population statistics contains, for example, information on population figures per city district, per gender and per religion. The process create population statistics is carried out to prepare, analyse and publish the corresponding data. It comprises a total of 12 activities. Current population figures and relocations are analysed. Both types of data are queried and processed for further analysis, checked for plausibility and analyzed in relation to certain characteristics. Once the data aggregation is complete, the actual report is produced.

Process 3 (P3) in D3 - Create social atlas:

The social atlas provides social data for Flensburg. It covers the topics of population, labour market and employment, housing, social security, and assistance with education. The data is presented for small-scale geographical areas. The social atlas is published on the website of the municipality of Flensburg¹ and used for further planning. The process *create a social atlas* is used to map population and social data to small-scale geographical areas. The process includes a total of 72 activities, with only 26 activities directly related to the department. Therefore, only these activities were considered in the further course. Due to the numerous interfaces to the other departments, the full organisational potential may not be represented here. However, the study focuses on resource efficiency in relation to individual departments. The process includes activities such as requesting data by email, receiving relevant data by email, integrating and preparing this data, creating figures and tables, creating the document and integrating feedback from different stakeholders (e.g. responsible officer and head of department). Finally, the social atlas is compiled and published.

Process 4 (P4) in D4 - Create school development planning report:

The school development report serves to prepare a needs-based school development plan in coordination with the neighbouring offices and districts in accordance with the legal framework. The process consists of 32 activities. The process includes the integration of different data sources (PRIMUS, Statistical Office North), the preparation of the data, the creation of figures and tables, the creation and publication of the report as well as the coordination with internal and external stakeholders (e.g. school headmasters and neighbouring municipal administrations).

¹siehe <https://www.flensburg.de/Leben-Soziales/Familie-Soziales/Sozialatlas/>

Data collection

The data collection was conducted after an initial introductory phase of the BI system so respondents had already gained some familiarity with the system. The strength of case studies is the use of data triangulation, namely collecting data of different types (e.g., interviews and documentation). We used multiple types of data for the purposes of modeling the analysis processes and collecting data about effort. First, a process analysis was carried out and the workflows were documented. In the second step, execution times were determined.

Capturing analysis processes

Unstructured interviews were conducted for data collection. Experts were interviewed about the process flows and the process times. Experts were selected who, on the one hand, have departmental and process-related expertise and, on the other hand, are familiar with the BI system that has been introduced.

For the BI-F2022 project, so-called multipliers were established in the departments. The multipliers form the interface between the BI team and the departments. The study participants work within the departments. They can therefore be seen as experts at the interfaces between the departments and the BI system. This enables a robust assessment of the impact of the BI system on process performance. Experts were interviewed in several sessions, depending on time availability and process complexity.

As part of the process modeling, a prior BI and a BI-enabled process were modeled for each of the processes under consideration. The processes were modeled collaboratively in 14 meetings. Process documentation was used as input to the meetings. A step-by-step walk-through was conducted during the meetings. The interviews lasted one to two hours.

Modeling was performed using the PICTURE method, which was developed specifically for business process management in public administration. It is based on the analysis of more than 600 process models from public administration (see here and in the following Becker et al. (2007, 2015)). The PICTURE method comprises domain-specific process building blocks that are used to map activities within an administration. Process building blocks are defined sets of reoccurring activities within a specific domain. In the following, we will refer to activities as process building blocks as part of the analytical processes. Process building block types group activities (i.e. process building blocks) belonging together. The groups are: (i) information gathering and coordination, (ii) information flows, (iii) content work, (iv) media change, and (v) documentation. Figure 1 provides an overview of the building block types and the process building blocks. We only provide examples, as a high number of process building blocks were estimated by the study participants. The figure also makes explicit what the building block type categories are about. For example, media change is about transferring data to different types, or transferring digital information to a physical representation (printing).

Effort data

After describing the processes, the process experts were asked to estimate the execution times. They were asked to esti-

Table 3: Number of process building blocks estimated per building block type

Process building block type	prior Bi	BI-enabled	Total
Information gathering	5	5	10
Information flows	33	27	60
Content work	37	30	67
Media change	3	9	12
Documentation	26	13	39
Total	150	99	249

mate the duration of process execution with and without the help of the BI system. Only one person per process made the estimates. This is to be considered a weakness. However, technical familiarity and familiarity with the BI system were necessary for the estimation. Competencies in both areas and at their interface are of high importance and limited the number of possible experts. Due to the heterogeneity of the processes, the figures collected nevertheless allow us to make cross-comparisons and identify patterns.

The activities were broken down to a low level of granularity, so that estimations became easier as the activities were less complex. At the same time, a high number of activities needed to be estimated. Table 3 shows the number of activities for each process building block type. In order to make assessments on a higher level of abstraction, we used the grouping of activities on various levels of abstraction (see e.g. Figure 1).

The estimates were extrapolated to the annual effort for each process to account for the fact that each process was performed multiple times and to ensure comparability. The annual times were summed and compared.

Data analysis

The data were analyzed from different perspectives, mapping to the research questions. First, analyses were carried out in relation to the different departments. Then analyses were made in relation to the building block type categories (see Figure 1). Finally, a detailed evaluation is carried out for the process building blocks, whereby we also look at individual activities. The measurement unit to capture the impact on time is captured in minutes. We present the data in tabular form and calculate the difference between the analysis processes without (referred to as prior BI process) and with the BI system (referred to as BI-enabled).

Threats to validity

In this section, we discuss different threat types, namely descriptive validity (factual accuracy), theoretical validity, generalizability (external and internal), and interpretive validity (objectivity of the researcher).

Descriptive validity (factual accuracy): Factual accuracy is concerned with our ability to accurately describe the truth. In this study, actual measurements of efficiency gains would have been preferred. Due to the lack of available historical data and ongoing measurements of the efficiency of the analytical process, we needed to rely on estimates. To get accurate estimates, we modeled the process on a low granularity level so that the estimated activities were not

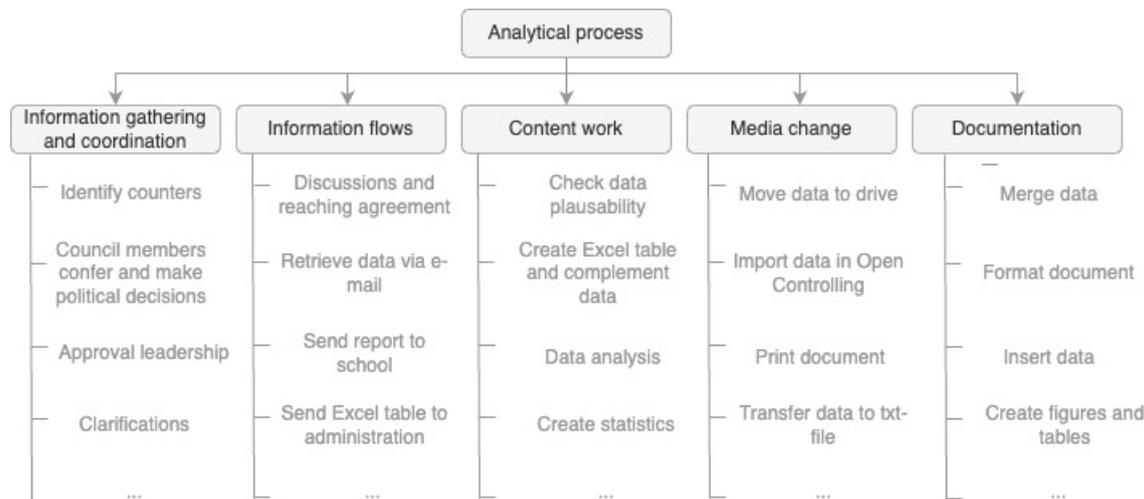


Figure 1: Overview of process building block types and process building blocks

too complex. Later, we grouped/aggregated activities to understand where efficiency gains and losses occur.

Theoretical validity: Theoretical validity is concerned with factors that influence the outcome (impact on efficiency due to the introduction of a BI system) without the researcher being able to control them. The study was conducted with the multipliers working on the project. However, the employees hired or released for the project often do not yet have a great deal of experience in public administration and also in the processes they assessed. This can lead to inaccuracies in the estimation of the implementation times. Likewise, the implementation frequencies of the traffic analysis are based on empirical values, so the actual frequencies may deviate from the frequencies assumed here.

Generalizability: Internal generalizability is concerned with whether findings are transferable to other units of analysis within the same organization. We studied multiple departments to reduce the threat. We found that, for the organization studied, the efficiency gains and losses varied between departments concerning building block type categories. We also found that all departments achieved efficiency gains, even though the magnitude varied. Given that we studied a single organization, we cannot claim generalizability concerning other organizations. As it is more likely that similar contexts yield similar findings (cf. Ghaisas et al. (2013)), we reported contextual information regarding processes and the public administration organization.

Interpretive validity (objective researcher): Threats to interpretive validity lead to limitations in the conclusions drawn by the researchers (e.g. due to biases). The threat to interpretive validity is low as the researchers did not have a personal stake as they were external evaluators. In addition, the data was interpreted by multiple researchers looking at the data (observer triangulation).

Results

The results of the analysis of resource efficiency are presented below. This is first considered in relation to the de-

partments and the processes mapped there. This is followed by analyses with regard to the phases as well as the process building blocks.

Research Question 1: Impact on effort on department level

Table 4 shows a comparison of the time saved by the department processes. It can be seen that the impact of the BI system varies. However, it should be noted in this context that this can also be due to the selection of processes. This will become clear in the further course of the detailed examination. The BI system achieves a particularly high impact in creating the social atlas (47.4% effort saving). Likewise, the effects of the analysis of traffic data and the creation of the population statistics can be rated as positive (15.3% and 16.6%, respectively). The increase in efficiency creating the report on school development planning is rather negligible (1.8% effort saving).

In the following, we discuss the processes and impact on effort in further detail for each department.

Traffic planning and regulation

Manual counts by pupils generate a large amount of data because counts are conducted by students on one day a year. The data collection is not affected by the BI system. After introducing the BI system, the data is no longer imported and processed by the clerical staff. This is expected to save 4 hours and 45 minutes of work time per year.

The manual counts by employees of the administration are only performed in special cases and therefore take place irregularly, estimated ten times a year. The sub-process is similar to the manual counts by pupils and changes with the BI system only after the data is received. After the traffic count data is received, the data can be sent to the statistical office. Thus, Excel evaluations no longer take place. This results in a time saving of 20 minutes per run, i.e. 3 hours and 20 minutes per year.

Table 4: Estimated and aggregated total process times in minutes in the departments for prior BI and BI-enabled processes.

Department	Analytical processes	Prior BI Process Estimation	BI-enabled Process Estimation	Difference	%
<i>Traffic</i>					
Planning and regulation	Analyze traffic data	5900	5000	900	15.25
Statistics Office	Create population statistics	7211	6011	1200	16.64
Youthe Services Planning	Create social atlas	3162	1662	1500	47.44
Education and sports office	Create school development planning report	13578	13338	240	1.77
Sum	-	298518	26011	3840	8.1

Speed displays change locations every two to three weeks. The data is sent to the clerk's office by the Technisches Betriebszentrum Anstalt öffentlichen Rechts (TBZ), a service provider for the city of Flensburg, an estimated seven times a year. The data must then be uploaded to the speed display manufacturer's website. Through the BI system, the data can be downloaded directly from the manufacturer's site and transferred to the statistics office. Separate evaluation of the data, as was necessary before the BI system, is no longer necessary. This allows a time saving of 25 minutes per run, which amounts to an annual time saving of 2 hours and 55 minutes.

Traffic measurement devices change location weekly from March to October. Once a month, the last four locations are evaluated. The data can be found on the manufacturer's website. After the data is downloaded from the website, it is integrated into the BI system. The data can then be analyzed and further processed by the BI system. A time saving of 20 minutes per run is expected. The annual time saving is estimated at 4 hours.

In summary, this results in an annual time saving of 15 hours for the clerical staff in the process of traffic data analysis. In addition, it should be noted that the data is located in the BI system and can therefore be viewed by colleagues or planning offices at any time. Therefore, weekly queries will be reduced. Since not all parties involved have access to the BI system yet, the total time savings due to the omitted queries cannot yet be estimated. Therefore, this activity was excluded from the analysis.

Statistics Office

The population statistics is created annually in the statistics office. The population statistics are published on the website of the city of Flensburg. Working in different software applications such as EwoPEaK or Migrapro remains unchanged. The individual activities change in the last steps of the process. As soon as the data is derived by the clerks for plausibility and according to various criteria, the manual creation of the numerical summary takes place, which is taken over by the BI system in the TARGET process. An annual time saving of 20 hours is expected.

Early Intervention and Youth Assistance and Youth Services Planning

The social atlas is compiled in the youth services planning department. For the compilation of the social atlas, comprehensive data must be systematised and analysed. The first step in the process is to request data from seven different departments. The requesting, waiting, and merging of different data formats is greatly reduced in the new BI-enabled process. Almost all data is available in the BI system. Currently, the data is loaded manually into the BI system by the BI team, which is expected to be converted in the near future. After the data has been received in the BI system, the process can be continued in the department of youth services planning.

During the project phase, a typesetting system for professional document generation (LATEX) was used for report generation. The BI system generates needed graphics and figures, and text generation in Latex reduces formatting work. These points, together with the existing data base (quality assured and up-to-date) in the BI system, provide an estimated annual time savings of 25 hours. At this point in time, it is only possible to speak of an estimated time saving, as the social atlas is an annual product and is, therefore, only being created for the second time using the BI system.

Education and Sports Office

The report for the school development planning is created annually. In this process, the BI system handles the calculations and the generation of graphics and tables. It should be noted that prior to the BI system, the graphics were generated by PRIMUS. According to the staff, the graphics can now be generated more in line with requirements, since the scope for adaptation by PRIMUS was very small.

In addition, as with the social atlas, the typesetting system LATEX was used for report generation. The time saved by the calculations and generated figures/graphs, as well as in the reduced formatting work, amounts to 4 hours annually. The data fed into the BI system can also be used for further inquiries, resulting in synergy effects for the clerical staff.

Research Question 2: Impact on effort on building block type categories

Five building block type categories were distinguished in the process models. The result of the process comparison for the building block categories can be seen in Table 5. Negative time differences indicate that more time must be spent on these phases when using the BI system. Positive values indicate a gain in efficiency.

The BI system provides support for the processes under consideration, particularly in terms of content-related administrative work and in terms of writing and documentation. Information flows are also made more efficient. No increase in efficiency can be seen in information procurement and coordination.

The efficiency gains are offset by efficiency losses in the phase *media change*. Over the entire year, efficiency gains are estimated to be 3840 minutes.

A comparison of processes and phases (see Table 6) with regard to time savings shows a differentiated picture for the individual processes. While the process population statistics benefits in particular in the content-related administrative work, in the process of creating the social atlas an increase in efficiency can be seen in particular in the area of writing / documentation. Furthermore, challenges can be seen in the phase media change. This applies in particular to the creation and publication of the social atlas and the analysis of traffic data.

Research Question 3: Impact on effort on process building block level

The evaluation in terms of process building blocks (Figure 7) shows from a further perspective the areas in which efficiency gains can be realized by introducing and making the BI system permanent. Note that, due to the high number of activities captured, the process building blocks in Table 7 were grouped. However, there are also areas where efficiency losses have to be accepted.

The highest efficiency gains can be achieved in the execution of specific activities. Overall, a time saving of 1185 minutes can be realized. However, a detailed analysis of the processes shows that in particular the preparation of the population statistics can be shortened by 1200 minutes. However, the execution of special activities in the creation and publication of the social atlas is extended by 15 minutes. In this process, the times for the creation of figures and tables are reduced by 45 minutes. However, exporting figures and tables from the BI system will require 60 minutes in the future.

For the process building block *Edit document/information*, a total time saving of 910 minutes is identified. In the following, we highlight how the savings were achieved by looking at the individual activities associated with *Edit document/information*. In particular, for the creation of the social atlas, efficiency gains are realized due to the shortening or elimination of the activities *inserting data into Excel spreadsheet* (300 minutes), *creating figures and tables/checking references and tables* (180 minutes), *formatting document* (90 minutes), *insert data into table* (30 minutes), *insert data* (15 minutes), and *copy Excel table from previous year, rename and adjust* (10 minutes) will be achieved. There are also 285 minutes of efficiency gained in the traffic data analysis

process. These result from the elimination of *saving and analyzing Excel tables* (250 minutes) and the elimination of *saving PDF files* (35 minutes).

The third highest efficiency gains are realized for the process building block *create document/information* (860 minutes). Especially for the Education and Sports Office and Youth Services Planning department, this activity is positively influenced. For the activity *insert figures, create tables and charts* in the municipal education planning process (Education and Sports Office), a time saving of 240 minutes is assumed. A comparable activity (*document creation (texts and tables)*) allows for a time saving of 1575 minutes in the creation of the social atlas (Department of Youth Services Planning). This is contrasted with the new activity *create / generate social atlas / write texts (LaTeX)*, which causes a time expenditure of 1080 minutes. Thus, taking into account the new effort, a time saving of 495 minutes is realized. In the process of traffic analysis, the annual effort in document creation is reduced by 120 minutes.

A time saving of 385 minutes is expected for the process building block *Check content/Make decisions*. Efficiency gains are expected in the processes *Create and publish social atlas* (325 minutes) and *Analyze traffic data* (300 minutes). However, increased effort is also expected for the school development planning report (an additional 240 minutes). The comparison of the areas indicates different expectations: While in the processes for analyzing traffic data and creating the social atlas, it is assumed that there will be an efficiency gain in the plausibility of the data, in the course of the report for education planning, it is assumed that there will be an increase in the time required to verify the data in the dashboard. For traffic data analysis, it is estimated that the activities *check data for plausibility* and *look through excel spreadsheet and correct if necessary* will no longer be necessary in the future. On the other hand, the activity *check data for plausibility and correct if necessary* will be required, but it can be done 300 minutes faster.

The added value of a central data pool becomes particularly clear for the process building block *Send document/information*: Requests to the statistics office for special evaluations in the course of compiling the social atlas can reduce the time required by 60 minutes. In total, a time saving of 105 minutes is assumed.

An additional effort is identified for the activity *to adopt data in EDP*. The additional effort, as well as the challenges of integrating the data, is again evident here. The additional effort is estimated at 450 minutes.

DISCUSSION

Summary of key findings

Comparison of departments (overall): Time effort savings from 1.77 to 47.44 % of the total effort were achieved across the departments. That is, we observe that the introduction of the BI system had varying effects on the efficiency of the analytical processes.

To understand the effects, we looked at different activity categories (referred to as Building Block Type Categories in the PICTURE method). The largest savings can be seen for

Table 5: Estimated total process times in minutes by application phase for prior BI and BI-enabled processes (negative values mean a negative effect)

Building Block Type Category	Prior BI Process Estimation	BI-enabled Process Estimation	Difference	%
Information gathering and coordination	1980	1980	0	0.00
Information flows	1286	1171	115	8.94
Content work	13370	11050	2320	17.35
Media change	185	635	-450	-243.24
Documentation	13030	11175	1855	14.24
Sum	29851	26011	3840	12.86

Table 6: Comparison of the time difference in minutes between prior BI process and BI-enabled processes in relation to the application phase (negative values mean a negative effect)

Building Block Type Category	Create school development report	Create social atlas	Analyze traffic data	Create population statistics
Information gathering and coordination	0	0	0	0
Information flows	0	115	0	0
Content work	0	310	810	1200
Media change	-5	-45	-400	0
Documentation	245	1120	490	0

content work (17.35% savings) and documentation (14.24% savings). A negative impact was mainly observed for media changes. Media changes occur repeatedly when information is transformed into different formats or transferred to different locations.

We found that information gathering and coordination were not affected for any of the studied departments. For the remaining building block type categories, it was evident that some departments were not affected by the introduction of the BI system (see Table 6). In addition, the degree by which the departments were affected varied substantially. Overall, the study shows that depending on the context, the impact of introducing a BI system can vary greatly.

To understand where time savings occur, we analyzed the process building blocks and looked at the individual activities and sub-activities. Based on the time saved in the statistics office, the education and sports office, and the department of youth services planning, it is clear that the BI system supports reporting tasks in particular. However, processes were selected that have a strong focus on reporting. This is evident from the process models and also the analysis in terms of activities. In the prior BI processes, reporting-related activities take up much time. The process building blocks *Edit Document / Information* (7305 minutes) and *Create Document / Information* (5640 minutes) take up 43,37% of the total time of the prior BI processes.

In the prior BI process, more time is only spent on *special activities* (8825 minutes), whereby clear references to the preparation of reports can also be seen here (e.g. the elements *publication (digital, print) of school development plan* and *preparation of population statistics* in the processes *Create school development planning report* and *create population statistics*).

We found a perceived positive impact on tasks that have a comprehensive reporting function.

Practical implications

We found that the gains and losses concerning efficiency varied between departments. This shows that the gains highly depend on the process being studied. The processes of the departments were heterogeneous, and may also be affected by the classification of the activities done during the modeling of the processes in the organization. Also, how many times certain activities are repeated, played a role. Thus, when wanting to assess the potential of efficiency gains due to BI systems, we still have to conduct individual assessments.

The estimation of impacts was possible with relatively little effort from the perspective of data collection (14 interviews). When having fine-granular activities of low complexities, we suggest assessing potential efficiency gains as part of introducing a BI system using the PICTURE method.

Research implications

The study of the related work showed that we did not find studies on the effect of business intelligence systems in the public administration context. Hence, to the best of our knowledge this is the first investigation with the main focus on BI system effects on efficiency in the public administration context. Therefore, further studies in different contexts are needed. We studied public administration in Germany, and suggest studying organizations in different countries. Also, as the study focused on a medium-sized municipality, administrations of larger and smaller cities should be studied.

We also would like to highlight the need for measurement programs that allow us to easily record efficiency data as, in

Table 7: Overview of resource efficiency by process building block in minutes (negative values mean a negative effect).

Process building block - grouped	Estimated prior BI duration	Estimated BI-enabled duration	Difference
Discussion and reaching agreements	600	600	0
Change work place for data collection	780	780	0
Consult	300	300	0
Calculate	750	0	750
Transfer data to IT	125	575	-450
Print document	60	60	0
View document	120	120	0
Edit document/information	7305	6395	910
Receive document/information	384	374	10
Create document/information	5640	4780	860
Send document/information	902	797	105
Formal check	20	20	0
Content check/plausibility check	2875	2490	385
Clarification questions	1080	1080	0
Execution of specific administrative tasks	8825	7640	1185
File process outcome	85	0	85

our case, no historical data was available. Thus, we needed to rely on estimates.

CONCLUSIONS

In this study, we investigated the impact of introducing a BI system in public administration. We investigated the case of Flensburg municipality. Our research question was: *What is the estimated impact on effort through the introduction of a business intelligence system in a municipality?*

Overall, a perceived positive effect of the introduction of the BI system on resource efficiency can be observed. These positive effects can be increased in the future by mapping further processes and taking other departments into account. Especially in areas where reporting is of high importance, the BI system is expected to add value to the daily work. However, challenges in report generation were also identified. For example, a new time effort of 1,080 minutes is expected in the creation of the social atlas using LaTeX, the tool chosen for documentation as part of the project.

This challenge was addressed early in this project, and work began on developing a reporting tool that was used prototypically in the creation of the social atlas. However, the reporting tool was not yet considered at the time of the time estimates. It should be noted here that in the process *create social atlas* the previous time required for document creation was estimated at 1,575 minutes. Thus, a time saving of 495 minutes is expected even without the reporting tool.

Further potential can be expected in relation to the processes described in the other departments involved. For example, the statistics office, the immigration office, and the economic youth welfare office are among the departments involved in the creation of the social atlas. Efficiency gains can be expected here as well. Data integration is a challenge. Further automation potential should be examined here, and data integration should be further simplified.

For this study, four processes were selected that are primarily designed to prepare reports. These processes do not cover the complexity and diversity of the operational and strategic decision-making processes of a public administration. Rather, they only give a first impression of the efficiency gains that can be realised through a BI system. This gives rise to the need for research to analyse further decision-making processes in public administration. Large-scale quantitative studies are challenging due to the diversity of processes and their concrete implementations.

The efficiency gains were not compared with the costs within the scope of this study. However, this is necessary for an economic decision. Since the BI system has only been introduced in four departments so far, an ex-post economic analysis is not possible at this time. This results in a need for further research that includes the business cost perspective. In this context, it should also be noted that this study only measured efficiency, as the name suggests, but not the effectiveness of the BI system. When cost factors are taken into account, they should be contrasted not only with the efficiency of the decision-making processes, but also with the quality of the decisions.

Given the lack of studies focusing on the impact of BI systems on efficiency in public administration, we suggest replicating this study in different contexts.

Acknowledgements

This study was prepared as part of the BI-F2022 project². The BI-F2022 project is funded by the Federal Ministry of Digital and Transport as part of the mFund innovation initiative³. We acknowledge the financial support from our funding agency and the collaborative efforts of our project partners in this research.

References

- ²see <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfund-projekte/bif2022.html>
³see <https://www.bmvi.de/DE/Themen/Digitales/mFund/Ueberblick/ueberblick.html>

- Abai, N. H. Z., Yahaya, J. H. & Deraman, A. (2015), Incorporating business intelligence and analytics into performance management for the public sector issues and challenges, in '2015 International Conference on Electrical Engineering and Informatics (ICEEI)', pp. 484–489.
URL: <https://doi.org/10.1109/ICEEI.2015.7352549>
- Ain, N., Vaia, G., DeLone, W. H. & Waheed, M. (2019), 'Two decades of research on business intelligence system adoption, utilization and success—a systematic literature review', *Decision Support Systems* **125**, 113113.
URL: <https://doi.org/10.1016/j.dss.2019.113113>
- Al-Eisawi, D., Serrano, A. & Koulouri, T. (2020), 'The effect of organisational absorptive capacity on business intelligence systems efficiency and organisational efficiency', *Industrial Management & Data Systems* .
URL: <https://doi.org/10.1108/IMDS-02-2020-0120>
- Becker, J., Algermissen, L., Pfeiffer, D. & Räckers, M. (2007), 'Bausteinbasierte modellierung von prozesslandschaften mit der picture-methode am beispiel der universitätsverwaltung münster', *Wirtschaftsinformatik* **49**, 267–279.
URL: <https://doi.org/10.1007/s11576-007-0063-0>
- Becker, J., Pfeiffer, D., Räckers, M., Falk, T. & Czerwonka, M. (2015), Semantic business process modelling and analysis, in 'Handbook on Business Process Management 1: Introduction, Methods, and Information Systems', Springer, pp. 187–217.
- Boselli, R., Cesarini, M. & Mezzanzanica, M. (2011), Public service intelligence: evaluating how the public sector can exploit decision support systems, in 'XXI International RESER Conference "Productivity of Services NextGen-Beyond Output/Input"', Fraunhofer Verlag.
- Chi, T. W. & Mahmud, I. (2020), 'Business intelligence system adoption: A systematic review of two decades', *International Journal of Industrial Management* **6**, 1–8.
URL: <https://doi.org/10.15282/ijim.6.0.2020.5624>
- Elbashir, M. Z., Collier, P. A. & Sutton, S. G. (2011), 'The role of organizational absorptive capacity in strategic use of business intelligence to support integrated management control systems', *The Accounting Review* **86**(1), 155–184.
URL: <https://www.jstor.org/stable/29780228>
- Fischer, C., Heuberger, M. & Heine, M. (2021), 'The impact of digitalization in the public sector: a systematic literature review', *dms - der moderne staat - Zeitschrift für Public Policy, Recht und Management* .
URL: <https://doi.org/10.3224/dms.v14i1.13>
- Gaardboe, R. & Jonassen, T. S. (2018), 'Business intelligence success factors: a literature review', *Journal of Information Technology Management* **29**(1), 1–15.
- Ghaisas, S., Rose, P., Daneva, M., Sikkil, K. & Wieringa, R. J. (2013), Generalizing by similarity: lessons learnt from industrial case studies, in M. M. Burnett, H. Giese, T. Nguyen & Y. Brun, eds, 'Proceedings of the 1st International Workshop on Conducting Empirical Studies in Industry, CESI 2013, San Francisco, California, USA, May 20, 2013', IEEE Computer Society, pp. 37–42.
URL: <https://doi.org/10.1109/CESI.2013.6618468>
- Hartley, M. & Seymour, L. (2015), Key success factors for implementing business intelligence in south african public sector organisations, in '9th IDIA Conference, IDIA2015', pp. 249–265.
- Hočvar, B. & Jaklič, J. (2010), 'Assessing benefits of business intelligence systems—a case study', *Management: journal of contemporary management issues* **15**(1), 87–119.
- Magaireah, A. I., HidayahSulaiman & Ali, N. (2019), 'Identifying the Most Critical Factors to Business Intelligence Implementation Success in the Public Sector Organizations', *The Journal of Social Sciences Research* **5**(2), 450–462.
URL: <https://ideas.repec.org/a/arp/tjssrr/2019p450-462.html>
- Magaireah, A. I., Sulaiman, H. & Ali, N. (2017), Theoretical framework of critical success factors (csfs) for business intelligence (bi) system, in '2017 8th International Conference on Information Technology (ICIT)', IEEE, pp. 455–463.
URL: <https://doi.org/10.1109/ICITECH.2017.8080042>
- Magaireah, A. I. et al. (2019), 'Identifying the most critical factors to business intelligence implementation success in the public sector organizations', *The Journal of Social Sciences Research* **5**(2), 450–462.
URL: <https://ideas.repec.org/a/arp/tjssrr/2019p450-462.html>
- McDonald III, B. D., Hall, J. L., O'Flynn, J. & van Thiel, S. (2022), 'The future of public administration research: An editor's perspective', *Public Administration* **100**(1), 59–71.
URL: <https://doi.org/10.1111/padm.12829>
- Runeson, P. & Höst, M. (2009), 'Guidelines for conducting and reporting case study research in software engineering', *Empir. Softw. Eng.* **14**(2), 131–164.
URL: <https://doi.org/10.1007/s10664-008-9102-8>
- Teixeira, R., Afonso, F., Oliveira, B., Portela, F. & Filipe Santos, M. (2014), Business intelligence to improve the quality of local government services - case-study in a local government town hall, in 'Proceedings of the International Conference on Knowledge Management and Information Sharing - KMIS, (IC3K 2014)', INSTICC, SciTePress, pp. 153–160.
- Trieu, V.-H. (2017), 'Getting value from business intelligence systems: A review and research agenda', *Decision Support Systems* **93**, 111–124.
URL: <https://doi.org/10.1016/j.dss.2016.09.019>
- Yogev, N., Even, A. & Fink, L. (2013), 'How business intelligence creates value: An empirical investigation', *International Journal of Business Intelligence Research (IJBIR)* **4**(3), 16–31.
URL: <http://doi.org/10.4018/ijbir.2013070102>
- Zaied, A. N. H. (2012), 'An integrated success model for evaluating information system in public sectors', *Journal of Emerging Trends in Computing and Information Sciences* **3**(6), 814–825.

Zaied, A. N. H., Grida, M. O. & Hussein, G. S. (2018), 'Evaluation of critical success factors for business intelligence systems using fuzzy ahp', *Journal of Theoretical and Applied Information Technology* **96**(19), 6406–6422.

Zakaria, Z., Yaacob, M., Yaacob, Z., Noordin, N., Mohamed Sawal, M. Z. H. & Zakaria, Z. (2011), 'Key performance indicators (kpis) in the public sector: a study in malaysia', *Asian Social Science* **7**(7).

URL: <https://doi.org/10.5539/ass.v7n7p102>

Contact

KAI PETERSEN is a professor of software engineering at University of Applied Sciences Flensburg and Blekinge Institute of Technology (BTH), Sweden, Germany. He received his Ph.D. from BTH in 2010. His research interests are Agile Software Development, Software Testing, Evidence-Based Software Engineering, and Software Measurement. His research has been conducted in close collaboration with companies and with an empirical focus.

AMELIE WASSE (M. Sc.) works for the municipality of Flensburg. Her main focus is on the digitalization of internal processes and the examination of the local health office.

JACOB SIETAS is a business intelligence analyst in the food industry. Previously, he worked as a research assistant in the BI-F2022 project. Business intelligence and data science are two of his obsessions. He has acquired his knowledge across industries in the insurance sector and in the publishing industry.

TIM CRUSE received his Master of Arts degree in Business Management/Business Informatics from the University of Applied Sciences Flensburg, which he obtained from September 2019 to May 2022. Prior to that, he completed his Bachelor of Science in Marketing/Technical Business Administration at the University of Applied Sciences Hamburg from October 2015 to July 2019.

JAN M. GERKEN is Professor of Data Science at Flensburg University of Applied Sciences. His research and teaching activities are particularly focus on Text Analytics, Natural Language Processing and Data Science. Previously, he was co-founder of a technology company specializing in semantic analysis and visualization of document landscapes.

Customer Segmentation in the Power of Attorney Business by Means of Fuzzy Clustering

Peter Rausch
Nuremberg Institute of Technology Georg Simon
Ohm
Keßlerplatz 12
90489 Nuremberg, Germany
peter.rausch@th-nuernberg.de

Michael Stumpf
Nuremberg Institute of Technology Georg Simon
Ohm
Keßlerplatz 12
90489 Nuremberg, Germany
michael.stumpf@th-nuernberg.de

Keywords

Customer Segmentation, Fuzzy Clustering, Modeling, Machine Learning, Data Mining, Targeting

ABSTRACT

Over the past years, a large market of service providers for powers of attorney has emerged in Germany. The number of officially registered general powers of attorney raised from almost 326,000 in 2005 to more than 5.3 million in 2021, according to the German Federal Chamber of Notaries. Service offers include consulting services on powers of attorney, text blocks and templates, annual update services and emergency call services. In general, the market contains all segments of the population, irrespective of their demographic characteristics. In this difficult and competitive environment, marketing and sales play an important role. In particular, not much research in the field of customer segmentation and marketing campaign preparation on this industry can be found. Thus, this research had the following goals which could all be achieved: The first goal was to compare the actual regional demographic structures of the German population with the regional structures of the customer base of a service provider in order to identify underrepresented population groups. The data on underrepresented groups was used to identify customer segments which can be used for targeting when future marketing campaigns are prepared. To solve this issue popular deterministic and fuzzy cluster approaches were successfully transferred to the use case and suitable configurations were explored as a second goal. As a third goal a recommendation, which of the analyzed approaches is most suitable in the present case, was derived. The approaches were evaluated on a real-world example of a service provider with more than 80,000 customers. It is important to note that despite of the fact that the regarded industry partner addresses a special market, the promising findings can be transferred to many other areas.

1. INTRODUCTION

Any person can get into a situation of being unable to articulate the own will by an accident, an illness or even slowly by old age. In some cases decisions which need immediate action, for instance, concerning medical treatments, important personal economic matters or urgent business matters, can not be made or communicated anymore. The number of people recorded in Germany who mainly needed care for the reasons listed above has risen from about 625,000 since 1995 [12] to currently approximately 1.3 million people [15]. Thus, it is not surprising that general, durable and medical powers of attorney are attracting increasing attention due to personal experiences among family and friends, but also because of media reports. According to the Federal Chamber of Notaries in Germany, the number of officially registered general powers of attorney raised from 325,637 in 2005 to 5,366,795 in 2021 in Germany [5]. The market for related products includes all parts of the population, independent from their demographic characteristics such as age, gender or place of residence. In this competitive environment, marketing and sales play an important role, since the above-mentioned products are not self-explanatory and, similarly to insurance products, the need has to be aroused. This is complicated by the fact that different demographic groups have to be targeted individually via campaigns, in order to achieve the highest possible impact. For instance, social media campaigns for younger and older people should be different. In order to support customer segmentation as a basis for preparing industry specific campaigns, very few well-founded documented findings can be found in science and reports from the field.

By means of this research this gap should become smaller and multiple goals have to be achieved. The first aim is to compare the actual regional demographic structures of the German population with the regional structures of the customer base in order to identify underrepresented population groups of the regarded service provider's customers. Knowledge about underrepresented groups is important for targeting when future marketing campaigns, for instance, via social media channels, are prepared. In case of the regarded service provider with many thousands of customers it is impossible to generate this knowledge without automated approaches. In particular, we suggest a cluster approach to further process the results of the comparison of the regional actual and customer structures. It is intended to find underrepresented customer segments in the customer base. The demographic characteristics of the related clusters play

an important role in social media campaigns. The results of an appropriate cluster approach are important, because different customer segments should be differently addressed by means of marketing campaigns on social media channels. In our case, it has to be taken into account that we have to deal with mixed categorical and continuous data. We will see that cases exist in which the customers do not have a homogeneous structure. This means, that individual instances do not really fit to a certain cluster but also match to another cluster based on their attributes by a slightly lesser degree. To plan effective marketing campaigns in our case the issue of identifying and handling a huge number of intermediate objects has to be solved. For the purpose, we will transfer a Fuzzy C-Means clustering algorithm to the use case and search for suitable configurations as a second goal. Finally, as a third goal a recommendation which approach is most appropriate in the presented case will be given. The approaches will be evaluated on a real-world example of a service provider with more than 80,000 customers.

At first, in Section 2 we will introduce the related use case and provide some background information to get a better understanding of the issue we intend to solve. Afterwards, in Section 3, we provide an overview of the current state of the art of approaches in this field. Based on the results of the industry partner's data, which is described in Section 4, different approaches and configurations are analyzed in Section 5. For this purpose, we start with a brief overview of deterministic clustering and the transfer to our use case. Then, fuzzy approaches are tested in different configurations to get improved results. Subsequently, the benefits and open issues of the approaches are discussed and compared in Section 6. Finally, in Section 7 the results are summarized and further enhancements will be discussed. It is important to note that despite the regarded industry partner addresses a special market, the analyzed problems can be found in many other areas and the proposed solutions can be transferred.

2. CASE DESCRIPTION

Before we go into details some background information on the industry, the industry partner and the related use case is needed to understand the issue and to evaluate the proposed solution later on in Section 6. As already mentioned, the number of officially registered general powers of attorney strongly increased. A main reason for this is the demographic change in Germany which will intensify according to forecasts ([25]). In general, everybody can quickly get into a situation in which a power of attorney is needed even at a young age. So this issue basically affects everyone. In order to set up valid powers of attorney correctly, a number of legal requirements must be regarded (see for instance, §1829 especially sub-paragraph 5 German BGB ([4]). Furthermore, experts advise updating powers of attorney at least every three to five years ([27]). Accordingly, a market has emerged with a variety of service providers. Services offered include consulting services on powers of attorney, text blocks and templates, annual update services, and emergency call services. Providers range from notaries, publishers, insurers or financial services brokers, automobile clubs as intermediaries, to platforms supporting the easy creation, distribution and sale of legal documents. The products and services of the regarded company include services to create general, durable as well as medical powers of attorney and a

hotline service for emergency cases. Since these services are useful for almost all parts of the population, the structure of the customers in terms of their demographic features should correspond to the structure of the population. By comparing the current customer structure with the demographic structure, mismatches can be used to identify potential starting points for future marketing campaigns. For instance, if younger people are underrepresented in the customer structure a targeted campaign addressing this customer group can be launched. If we consider clusters of customers, the number of clusters is unknown, in advance. Furthermore, it is likely that many customers do not fit perfectly to a certain cluster. Since, we may have to deal with many intermediate customer instances a dichotomic assignment to clusters would distort the results and would not be a good representation of the reality. Based on such a segmentation positive impacts of campaigns could not be fully exploited, or in borderline cases a decision maker might choose the wrong option when faced with several alternative courses of action. For this reason we will consider fuzzy approaches in this research. Before more information on the data set is given, we will provide a detailed review on the subject and approaches applied in the field of customer segmentation.

3. RELATED WORK

In literature, the term market segmentation describes the idea of grouping similar consumers in terms of specific characteristics [23]. Based on this segmentation, attractive segments can be identified and targeted, and marketing actions for each segment can be customized. According to [7], different categories of segmentation, for instance, geographic, demographic, psychological, psychographic, sociocultural, user-related, user-situation and benefit segmentation, are distinguished. Also mixtures of segmentation approaches, so-called hybrid segmentation methods, are possible [7]. In our paper, based on the regarded input data, see Section 4, demographic and geographic segmentation is addressed. Thus, we will focus on a hybrid segmentation approach. Since the data set is unlabeled, see Section 4, unsupervised learning approaches have to be considered. In particular, many successful applications of cluster approaches can be found. In some cases, segmentation is used to address the more or less volatile behavior of customers, see for instance, [6]. The idea is to develop relationships with customers based on accurate information which is optimized for the customers' taste and preferences [6]. Other research focuses on customer segmentation for an online seller of automobile accessories and fittings based on a k-means approach. The idea is to recommend methods to increase customer influx and give suggestions for better performance and sales [21]. A disadvantage of this approach is the a priori selection of the number of clusters which has an impact on the quality of the algorithm's outcome [13]. Additionally, this approach tends to have issues in cases when clusters are of different sizes, densities or when their shapes are not globular [8]. Also, outliers can cause issues because the approach is based on the arithmetic mean of points in an n-dimensional space [13]. Besides, in our case categorical data has to be processed which is not supported by the k-means approach [8]. In [11] the case of demographic customer segmentation of banking users is discussed, but as well as [21] they do not address the aspect of fuzzy memberships to clusters.

[28] applies fuzzy clustering to customers' segmentation in securities industry. The idea is to identify the customers with similar characteristics and value by means of a Fuzzy C-Means (FCM) approach which will be explained in Section 5.3. Many other successful applications of customer segmentation by means of fuzzy clustering can be found. For instance, fuzzy clustering is applied in the fields of energy consumers [22], telecom [1], life insurance [14] and B2B for food and beverages merchants [16]. [19] analyze the case of implementing marketing strategies by observing dynamic changes in the customer segments over time. Like in the other cases, they use a FCM approach. These many successful practical applications indicate that fuzzy clustering may also lead to promising results in the present case. Nevertheless, it must be analyzed whether the FCM approaches' results are of such a good quality that a recommendation can be made in terms of goal 3, see Section 1. Besides the fact that we consider a different line of business, the FCM algorithm which is used in most of the mentioned use cases has some limitations. The approach has issues with outlier data and is sensitive to initialization. Hence, a general optima can not be guaranteed [19]. This problem needs to be addressed. Before we go into details, the data set and the preprocessing step will be explained.

4. DATA SET AND PREPROCESSING

As already mentioned the basic idea is to cluster customers in terms of their demographic and geographic features and use the results for the preparation of marketing campaigns. The data set provided by the industry partner is unlabeled and contains mixed data types. The feature "year of birth" was converted to a 0-1-normalized attribute. For reporting purposes it can be reconverted or transformed to the customers' age, depending on the purpose of the analysis. In addition to this attribute, the features "gender" and "place of residence" were included. These features are used, because social media campaigns can be targeted by these attributes and this data could be provided by the industry partner. The feature "gender" is categorical data, but we used it as binary variable because the data from the officially registered German population based on the last census was just available for two genders, see [24]. When the data from the 2022 census currently conducted in Germany will be published, certainly data on all genders will be available. For our case, we just considered the genders "male" and "female". However, it is recommended to add other genders as soon as the new related official statistics is available.

The "place of residence" feature is present in multiple different granularity levels. For our study, we had to restrict our comparisons of the customer base to the demographic structure of the population on the level federal states. Although the customer data was available in the form of postal codes, data for population statistics was available only in a separate geographical region structure which can not be matched exactly to postal codes. In general, a more fine-grained analysis would be possible, if more detailed data from the government would be accessible. Because of this, the lowest common denominator where regions between the industry data set and the population statistics match is the federal state level. So, in the industry partner's input data set as well as in the data set from the official German population statistics preprocessing was executed to unify "place of resi-

dence" to the federal state level. The information about the postal code for each instance of the industry partner's data set is still available. This can be used to display instances on a geographic map with more detail than the federal state level. Additionally, the feature "marital status" (yes/no) was considered as a binary value for the first set of experiments.

The above mentioned features can be used to target customers, for instance, in the preparation process of social media campaigns on popular platforms, see for instance [18]. To choose an appropriate approach for the segmentation of the customers a pre-analysis was executed. For this purpose, the data was analyzed, and accumulations of characteristics could be identified. For instance, in Germany, about 41.8% of the total population are married [26]. The data set of the regarded company with about 80,000 instances included about 69.4% with this marital status. It is already apparent at this point that there are major discrepancies between the customer structure and the population structure. Of course, the results of this pre-analysis will be refined, elaborated and enhanced in the next section.

5. APPROACH

5.1 OVERVIEW

As already outlined in Section 1 the basic idea of this research is to identify underrepresented population groups in the customer base in order to use this information for targeting when future marketing campaigns are prepared. To compare the actual demographic structure of the German population with the structure of the customer base the data which was described in Section 4 will be used. Due to a mix of demographic and geographic characteristics, a hybrid segmentation approach, see Section 3, is chosen.

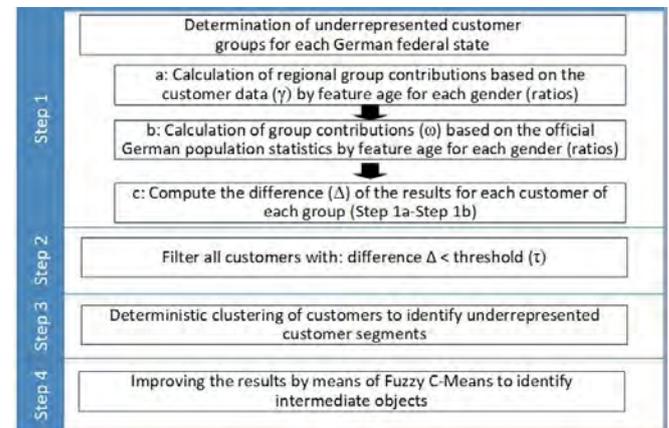


Figure 1: Approach: overview

As shown in Figure 1, we determine underrepresented customer groups for each German federal state, in Steps 1a-c. In Step 1a, the group contributions γ_ϕ by the features age a and gender g based on the customer data for each federal state ϕ are calculated as the intersection of the related feature set divided by the total number of customers Π in federal state ϕ :

$$\gamma_{(a \cap g, \phi)} = \frac{|\pi_{a, \phi} \cap \pi_{g, \phi}|}{\Pi} \quad \forall \phi = 1, \dots, \Phi \quad (1)$$

$\pi_{a, \phi}$ and $\pi_{g, \phi}$ denote the set of customers who have a certain characteristic profile in the customer base.

Analogously, in Step 1b the group contributions of the officially registered population ω by features age a and gender g based on the official German population statistics are determined for each federal state ϕ divided by the total number of citizens ϱ in federal state ϕ :

$$\omega_{(a \cap g, \phi)} = \frac{|\rho_{a, \phi} \cap \rho_{g, \phi}|}{\varrho} \quad \forall \phi = 1, \dots, \Phi \quad (2)$$

$\rho_{a, \phi}$ and $\rho_{g, \phi}$ denote the set of officially registered citizens who have a certain characteristic profile in the official German population statistics. These ratios will be used as references and will be compared to the results of Step 1a in the following Step 1c

This means that each individual group of Step 1a is compared to its corresponding group of Step 1b, like shown in Equation 3:

$$\Delta_{(a \cap g, \phi)} = \gamma_{(a \cap g, \phi)} - \omega_{(a \cap g, \phi)} \quad \forall \phi = 1, \dots, \Phi \quad (3)$$

By computing the differences $\Delta_{(a \cap g, \phi)}$, the contributions of each age and gender group in the dataset is compared to its corresponding average group contribution in the German population for each state. So, instances with positive $\Delta_{(a \cap g, \phi)}$ denote that this instance is part of an overrepresented group in the customer data set. Whereas negative values of $\Delta_{(a \cap g, \phi)}$ indicate that the instance is part of an underrepresented group.

If a difference is smaller than a user defined threshold, all objects of the related group are filtered in Step 2. Afterwards, in Step 3 all filtered objects are clustered to identify the underrepresented customer segments. Finally, in Step 4, the results are improved by means of a Fuzzy C-Means approach.

5.2 DETERMINISTIC CLUSTERING

5.2.1 APPROACH AND CONFIGURATION

According to Figure 1 (Steps 3 and 4), the filtered objects which include all underrepresented customers are clustered. Since, cluster analysis has proven its efficiency in the area of customer segmentation, see Section 3, we apply this method to our case. Deterministic cluster approaches are based on the assumption that an object o (of all O filtered object instances I) is always member of only one cluster c . Thus, the following condition has to be satisfied:

$$\mu_{oc} \in \{0, 1\}; \forall o = 1, \dots, O \quad (4)$$

In our case, we choose a hierarchical approach for Step 3, see Figure 1. This category of clustering allows to visualize the results as dendrograms. By means of a dendrogram, the analyst can determine a threshold for the proximity measure and fix it in a way that the clusters have a desired homogeneity. Thus, in advance, there is no knowledge about the number of clusters necessary. Within the hierarchical approaches, we have selected an agglomerative complete linkage clustering. It tends to form small compact groups and avoids the chaining effect which might be an issue related to other hierarchical approaches [9]. In detail, the steps shown in Figure 2 are executed.

Like other agglomerative approaches, each object initially forms a cluster. Subsequently, this initial partition is modified by successively merging the clusters into larger aggregates. For this reason, the distances or the similarities between the clusters have to be calculated. In our case, we

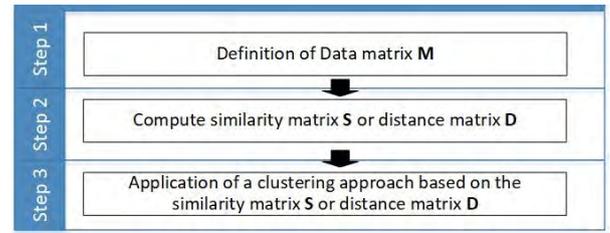


Figure 2: Hierarchical-agglomerative clustering methods

use the Gower distance. In general, it is used to compute the distances between objects having numerical and categorical features [10] as in our case. As Equation 5 shows, the Gower distance allows to weight the impact of each of the N features individually [10] - which is another requirement in our case.

$$d(o, c) = \frac{\sum_{k=1}^N \delta_{oc}^{(k)} d_{oc}^{(k)}}{\sum_{k=1}^N \delta_{oc}^{(k)}} \quad (5)$$

For binary features, the distances are computed as shown in Equation 6, and metric features are included according to Equation 7.

$$d_{oc}^{(k)} = \begin{cases} 1 & \text{if } x_{ok} \neq x_{ck} \\ 0 & \text{if } x_{ok} = x_{ck} \end{cases} \quad (6)$$

$$d_{oc}^{(k)} = \frac{|x_{ok} - x_{ck}|}{R(k)} \quad \text{with } R(k) = \max_o x_{ok} - \min_o x_{ok} \quad (7)$$

In our case, we use one binary variable for each federal region of Germany. Thus, for Germany we need 16 binary variables to indicate if a customer belongs to a federal region or not. By means of the weight $\delta_{oc}^{(k)}$ the influence of each geographic binary variable is reduced to $\frac{1}{16}$ to avoid an inappropriate large influence of the "place of residence" feature on the distance. Due to the reasons mentioned in Section 4 the "gender" feature was also regarded as a binary variable with a weight of one. With the same weight the age was considered as a numerical value. Furthermore, the feature "marital status" was initially included in the Gower distance. However, as the analyses showed, this feature had no significant influence on the cluster structures and is therefore no longer considered in the further explanations. To compute the distances or the similarities between the clusters, the maximum distance or the minimum similarity of a cluster pairs' elements are determined, see Equation 8.

$$d_{lm} = \text{Max}\{d_{oc}\}_{(I_o, I_c) \in P_l \times P_m} \quad \text{or} \quad s_{lm} = \text{Min}\{s_{oc}\}_{(I_o, I_c) \in P_l \times P_m} \quad (8)$$

When the distances, like in our case, or the similarities between all clusters have been determined, the clusters which have the minimum distance or the maximum similarity to each other are merged.

Afterwards, the proximities between the new agglomerate P_q resulting from the fusion of the clusters P_l and P_m and the remaining classes have to be calculated. This procedure is repeated until a certain number of clusters is achieved or all objects are in one cluster. In the latter case, the cluster structure can be visualized by means of a dendrogram. The

dendrogram helps to define a threshold for the maximum tolerated distance, and the cluster can be split. Finally, a desired number of more or less compact clusters is resulted. A disadvantage of this approach are the high computational costs for analyses [17], which our experiments confirm. Even with the filtered subset of customers and a powerful computer, $\frac{O(O-1)}{2}$ comparisons must be made to perform the algorithm. In our case, it was possible to test several configurations of the algorithm within a day. But with the complete number of customers, the performance of the available hardware would have been exceeded with many million comparisons. Another problem with this cluster procedure is that the selection of the clusters to be merged is determined only by two extremely positioned objects of the cluster pairs. In addition, this method does not detect outliers. The extent to which these problems are relevant in our case is explained below in combination with the presentation of the results of the deterministic part of the experiments.

5.2.2 DETERMINISTIC RESULTS

As Figure 3 shows, 10 clusters (C_{HI1}, \dots, C_{HI10}), which instances are denoted by their numbers, are the result of the deterministic cluster approach. Due to the hierarchical approach the clusters varied widely in size. Besides 3 large clusters also 3 medium and 4 small clusters were identified.

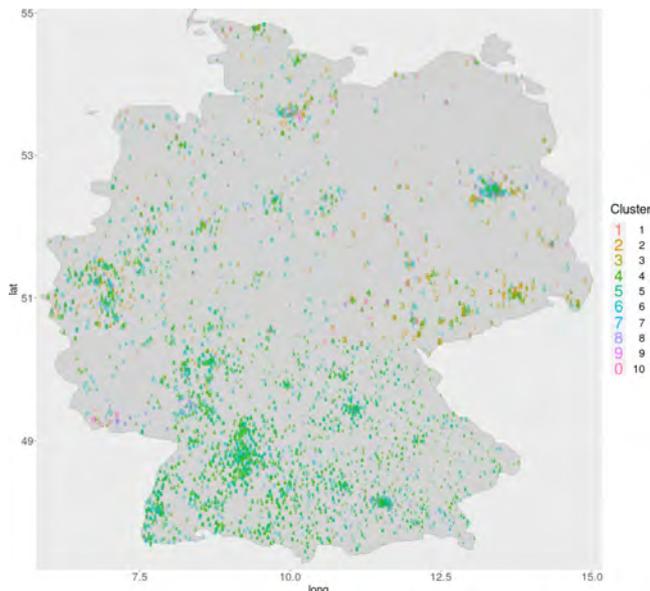
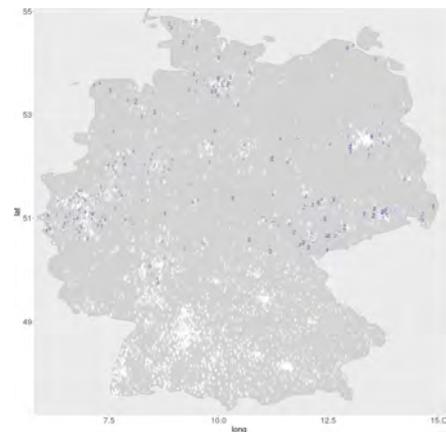


Figure 3: Visualization of the deterministic cluster results (The larger the symbol size the higher the under-representation.)

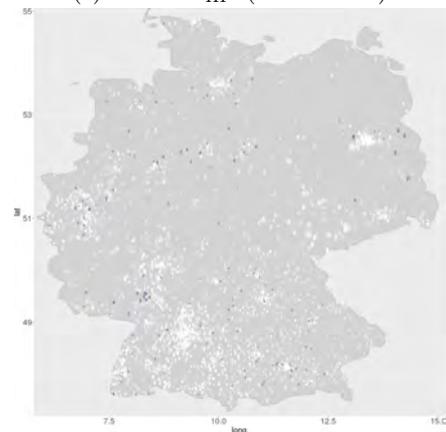
The large clusters (C_{HI4} , C_{HI5} , and C_{HI6}) are not geographically grouped but are relatively compact in the age dimension. If the range between first and third quartile around the median is considered the age dimension of cluster C_{HI6} has a range of ± 2 and cluster C_{HI4} and C_{HI5} are almost as compact with a range of ± 3 , see Table 1.

For smaller clusters this is not always the case, but for instance, cluster C_{HI9} , which represented roughly 1% of the filtered data set, had a range of ± 3 around the birth year of 1976 and its instances were all male and geographically

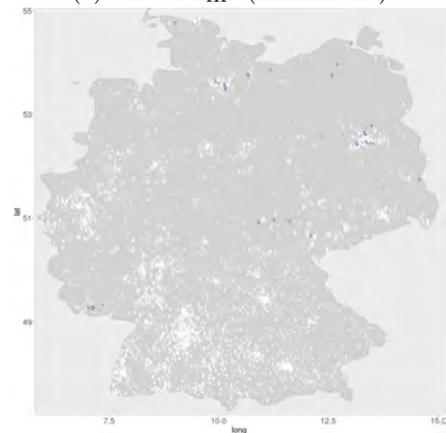
located in different compact areas in the northern half of Germany, see Fig. 4c. Yet, there are also medium clusters with a larger scatter overlapping others, like, for instance, clusters C_{HI2} and C_{HI8} , see Fig. 4a and 4b.



(a) Cluster C_{HI2} (medium size)



(b) Cluster C_{HI8} (medium size)



(c) Cluster C_{HI9} (small size)

Figure 4: Visualization of selected individual clusters in hierarchical results (blue symbols represent the corresponding cluster membership)

Both of these clusters consist again of males, around a certain age (C_{HI2} : 1934 ± 4 , C_{HI8} : 1940 ± 3) but overlap geographically.

Cluster-No.	Percentage	Gender	Birth Year (19xx)				
			Min	1.Q	Median	3.Q	Max
C_{HI2}	6.0 %	male	35	40	43	45	51
C_{HI4}	38.0 %	male	80	87	90	93	96
C_{HI5}	15.3 %	female	21	29	32	35	40
C_{HI6}	27.8 %	female	73	90	92	94	96
C_{HI8}	5.1 %	male	27	31	34	37	43
C_{HI9}	0.9 %	male	64	73	76	79	85

Table 1: Selected statistical data for hierarchical cluster results

However, they are not only intersecting by location, but if minimum and maximum of the birth year is considered they also overlap in the age dimension. The minimum and maximum values of year of birth are for cluster C_{HI2} in the range 1935-1951 and for cluster C_{HI8} in the range 1927-1943, see Table 1. So, some instances of cluster C_{HI8} are well within the first quartile (1940) and the median (1943) year of birth of cluster C_{HI2} . It is to be doubted that the instances possessing features of different clusters should exclusively have the degree of belonging 0 or 1 to exactly one cluster. In addition, a non-dichotomous mapping would be helpful to better estimate the potential of marketing campaigns because there might be cases where additional customers can be reached by a single campaign. Therefore, in the following will be investigated to what extent the results can be improved by fuzzy approaches, see Figure 1, Step 4.

5.3 FUZZY PART

In contrast to deterministic cluster approaches, condition 4 is relaxed in fuzzy cluster analyses, so that an object can be member of multiple clusters:

$$\mu_{oc} \in [0, 1] \quad (9)$$

This relaxed condition ensures that the above mentioned problems will be solved. In order to be able to evaluate the quality of the results of the fuzzy algorithm, the approach should be examined more closely beforehand.

5.3.1 APPROACH AND CONFIGURATION

As mentioned in Section 3, the Fuzzy C-Means (FCM) approach, which is one of the most known and most commonly used variant of fuzzy cluster analysis algorithms [20], is used in many cases in the field of customer segmentation. It is based on the classical ISODATA algorithm and minimizes the generalized variance criterion σ_m [2]:

$$\sigma_m = \sum_{c=1}^C \sum_{o=1}^O \mu_{oc}^m \cdot |\mathbf{f}_o - \bar{\mathbf{u}}_c|^2 \quad (10)$$

$$\text{with } \bar{\mathbf{u}}_c = \frac{\sum_{o=1}^O \mu_{oc}^m \mathbf{f}_o}{\sum_{o=1}^O \mu_{oc}^m}, \quad \forall c = 1, \dots, C$$

\mathbf{f}_o represents the feature vector of object o and $\bar{\mathbf{u}}_c$ the centroid of a cluster c . The power $m \in [1, \infty[$ is to be determined by the user. The use of the objective function 10 identifies relatively homogeneous partitions. Besides, the elements of

Experiment	1.Q - Median		3.Q - Median		Max - Min	
	Median	Min	Median	Max	Median	Max
Hierarchical	-3.0	-3.0	+3.0	+4.0	17.5	23.0
FCM	-2.5	-4.0	+2.0	+6.0	12.5	27.0
FCM with Start U	-2.0	-4.0	+2.0	+6.0	12.5	27.0

Table 2: Selected statistical properties of age dimension over all clusters between experiments

the membership matrix U must satisfy the following conditions:

$$\mu_{oc} \geq 0 \quad \forall c = 1, \dots, C; \quad \forall o = 1, \dots, O \quad (11)$$

$$\sum_{c=1}^C \mu_{oc} = 1 \quad \forall o = 1, \dots, O \quad (12)$$

It should be noted that the fulfillment of condition 12 can distort the results in case of outliers which would be best represented by very small membership values to all clusters. Considering this aspect, it has to be analyzed whether this approach can improve the results of the deterministic method in Section 5.2.2. Since we need a start partition to solve the problem described above, it appeared reasonable to use the results of Section 5.2.2. Alternative experiments with random start partitions and different seeds confirmed that this approach is superior and achieves the best results for our use case. Details will be given later in Section 5.3.2. The use of the deterministic start partition obviously seems to help solving the problem that FCM is sensitive to the start partition. To determine the FCM results, first, the class centroids of each cluster $\bar{\mathbf{u}}_c$ must be determined for all clusters $c = 1, \dots, C$. In the next step, the new membership values of the objects are calculated by Equation 13:

$$\mu_{oc}^{new} = \left(\sum_{h=1}^C \left(\frac{|\mathbf{m}_o - \bar{\mathbf{u}}_c|}{|\mathbf{m}_o - \bar{\mathbf{u}}_h|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (13)$$

$$\forall c = 1, \dots, C; \quad \forall o = 1, \dots, O$$

According to (13), the membership value of an object o to the cluster c is determined by the distance between the object and the cluster centroids. In case the features of object O_o are identical to those of a centroid, the membership value is set to 1. For all other μ_{ih}^{new} -values with $h \neq c$, zero is set [3]. The approach ends after the q -th iteration if

$$\max_{((O_o, C_h) \in E \times C)} |\mu_{oh}^{(q)} - \mu_{oh}^{(q-1)}| \leq \xi, \quad (14)$$

where ξ is specified by the user. If $m \rightarrow 1$, the result tends to the result of a deterministic ISODATA-method. For $m \rightarrow \infty$, the membership values tend to the reciprocal of the number of classes $\frac{1}{C}$.

5.3.2 FUZZY RESULTS

At first, an FCM configuration was used to determine clusters without any previous knowledge about the cluster structure. The fuzzy clustering yields, as configured, 10 clusters (C_{F01}, \dots, C_{F010}). The instances of the filtered data set are grouped into 4 large, 5 medium-sized and 1 small cluster. As it can be seen in Table 3 the amount of small clusters is reduced.

Experiment	Cluster		
	Small < 3%	Medium < 9%	Large ≥ 9%
Hierarchical	4	3	3
FCM	1	5	4
FCM with Start U	0	6	4

Table 3: Cluster sizes of different experiments

Another finding about the statistical properties of the corresponding cluster instances concerning the dimension “age” is shown in Table 2. The year differences between first quartile and median are slightly reduced over all clusters when considering the median. However, the maximum difference over all clusters is slightly increased. The same effect occurs with the difference between the median and third quartile of the year difference. It is also slightly reduced, but the maximum difference is also slightly increased. Furthermore, the year range between minimum and maximum over all clusters is also decreased over all clusters. However, as in the previous two considerations, the maximum value is also increased. This means that most of the clusters have a slightly smaller age range between first quartile and median, as well as between median and third quartile. Considering the maximum and minimum values over all clusters the median of the age range is reduced. So, the majority of the clusters is more compact concerning the age dimension. However, there are also clusters which are broader than in the hierarchical case.

As it can be seen in Figure 5, the geographic location had more influence on cluster formation.

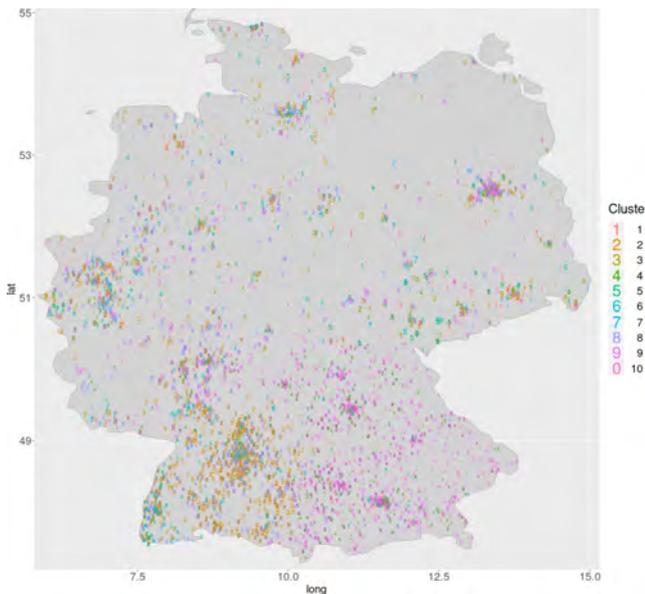


Figure 5: Visualization of the FCM cluster results (without knowledge)

In the second set of experiments, the hierarchical clusters C_{HI1}, \dots, C_{HI10} were used as starting partition for the cluster membership values U . The computation yielded 10 clusters C_{FH1}, \dots, C_{FH10} . As shown in Table 3 no small clusters were formed.

Considering the age dimension, the results were marginally improved by a minimal decrease in the age range of first

quartile to median, see Table 2). However, the results in terms of the geographic properties of clusters improved significantly, as it can be seen in Figure 6, in comparison to the hierarchical cluster result in Figure 3.

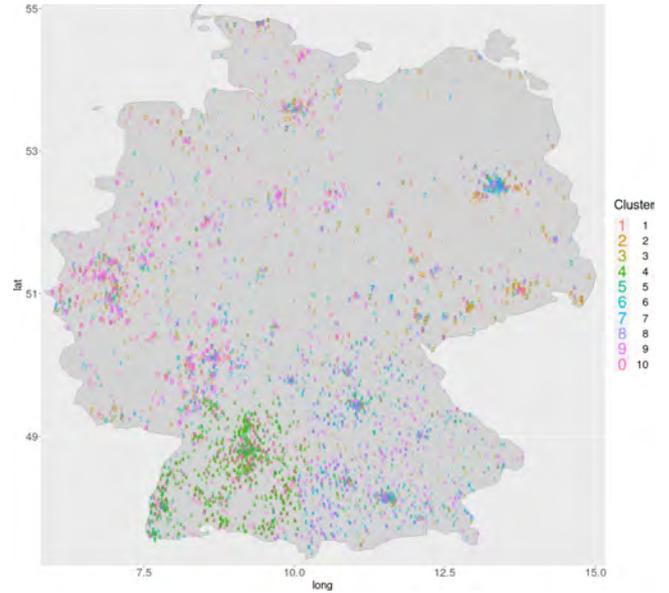


Figure 6: Visualization of the FCM cluster results with a hierarchical start partition U

More specifically, for instance, clusters C_{FH4} and C_{FH9} have roughly the same properties as C_{HI4} , C_{F02} and C_{F09} , see Table 4 - being male and roughly around year of birth 1990. It seems that the fuzzy approaches found a reasonable geographic split for C_{HI4} into two or more separate clusters C_{F02} and C_{F09} for the first iteration without knowledge, whereby the combined sum of the cluster sizes of 39.0% is slightly larger than the original cluster with 38.0% of all objects. In the second iteration with a start partition generated by a hierarchical approach the results are improved even more, since C_{FH4} and C_{FH9} combined represent only 30.2% instead of 38.0%. So, the remaining instances were allocated to other, more suitable clusters.

Taking up on the geographic distinction the clusters C_{F02} and C_{FH4} have a rather dominant region BW (short for Baden-Württemberg) and clusters C_{F09} and C_{FH9} have a rather dominant region BY (short for Bavaria). The fainter the blue in Figure 7, the lower the membership degree to the assigned cluster. So, the fuzzy approaches managed better to associate clusters to a geographic region.

Besides the size of the clusters between first and second iteration, also the age dimension is more compact in the second iteration. Comparing the minimum and maximum values in the age dimension, the clusters found by the second iteration of the fuzzy approach are much more compact, see Table 4.

Cluster-No.	Percentage	Gender	Birth Year (19xx)					Region
			Min	1.Q	Median	3.Q	Max	
C_{HI4}	38.0 %	male	80	87	90	93	96	all
C_{F02}	18.9 %	male	86	89	91	94	96	BW+
C_{F09}	20.1 %	male	69	85	89	91	96	BY+
C_{FH4}	14.4 %	male	86	88	91	93	96	BW+
C_{FH9}	15.8 %	male	88	90	91	93	96	BY+

Table 4: Selected statistical data of clusters in all three experiments (Hierarchical, FCM, FCM with hierarchical start partition U)

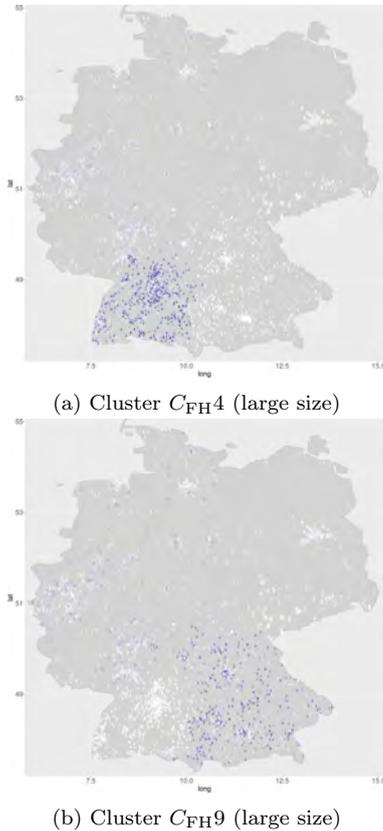


Figure 7: Visualization of selected individual clusters in FCM with hierarchical start partition (The fainter the blue symbol, the lower the membership for the corresponding cluster. The larger the symbol size the higher the under-representation.)

6. EVALUATION

After testing many configurations of deterministic cluster algorithms, a configuration was found which delivered a useful segmentation of the underrepresented customer groups. Yet, to prepare a concrete marketing campaign, it was an issue that the deterministic approach does not consider that some clusters overlapped more or less with others. Thus, a crisp assignment of the respective objects is not an ideal representation. It causes problems when social media campaigns are planned, since transparency is not given. The presented fuzzy approach solves these issues. The tendency towards larger clusters and smaller “residual clusters” in the results of the fuzzy approach indicates that the potential of marketing campaigns for the bigger clusters is much larger than

it would be expected regarding the corresponding results of the deterministic clusters. On the other hand, for the smaller clusters which were detected by the deterministic approach the effect of targeted campaigns would have been overestimated. It is obvious that in the first case many more underrepresented customers could be targeted by means of a single campaign. In the second case, a waste of resources for campaigns which might have a lower impact due to an overestimation of the positive impact can be avoided. Thus, the fuzzy approach provides a better quality of information and a more realistic representation of the underestimated customer segments is achieved. This information is also important for the evaluation of campaigns.

As outlined in Section 5.3.2, the better quality of the fuzzy results is especially evident for the dimensions “age” and “place of residence”. Since the filtered data set consisted roughly of 49 % female and 51 % male instances all clustering attempts resulted in a strict distinction between genders. This could have been mitigated by reducing the weight of the gender attribute. However, since gender is a possible configuration parameter for marketing campaigns we decided not to do so.

Additionally, it should be mentioned that the first set of the fuzzy experiments (without a starting partition of a hierarchical cluster approach) could be executed more swiftly with less computational resources. It generated reasonable results and outperformed the deterministic outcome. In this study, computational effort was not be a big issue, but when analyzing customer groups in other domains this aspect should be considered.

7. CONCLUSION AND FUTURE WORK

Due to many reasons which were mentioned at the beginning, the power of attorney business is a growing market. Since potential customers can basically come from all population groups, marketing campaigns that leverage the greatest potential on a target group-specific basis are of great importance. In order to identify major underrepresented target groups customer segmentation plays an important role. Unfortunately, this issue has been relatively unexplored in this business. By means of this research, we narrowed the gap by analyzing approaches to support targeting for marketing campaigns by means of cluster analyses and developed a course of action. All research goals could successfully be achieved.

In detail, we compared the actual regional demographic structures of the German population with the regional structures of the regarded service provider’s customer base in order to

identify underrepresented population groups (goals 1 and 2). In the analyzed case, mixed categorical and continuous data had to be considered. The clusters were identified by means of deterministic and fuzzy clustering approaches combined with a Gower distance to address the issue of mixed data types. As we compared the results, it became obvious that the FCM approach comes along with several benefits compared to the hierarchical cluster approach and its results were superior to those of the deterministic variants used in the present case. The FCM approach identified more or less overlapping clusters in a way that the quality of the customer segmentation was improved. The fuzzy analyses tended towards larger clusters and smaller “residual clusters”. Thus, valuable information for targeting was generated and can be used for the preparation of future marketing campaigns to focus on the bigger groups and to avoid an unnecessary waste of resources for inefficient campaigns. By means of more realistic results, the post evaluation of marketing campaigns can also be carried out at a better level of information.

In any case, we recommend the FCM approach for the analyzed use case (goal 3). Even for the worst FCM result in a setting when no start partition of a deterministic approach is available, the fuzzy results outperformed the deterministic results. Obviously, the issues of FCM mentioned in Section 3 and Section 5.3.1 are not so severe in the present use case as to lead to worse results. Hence, in cases with larger data volumes and increasingly higher efforts to generate a deterministic start partition, we still recommend the FCM method.

Furthermore, it is recommended to repeat the analysis which was presented here from time to time because the structures of the underrepresented population groups in the customer base may change over time.

Nevertheless, there are still ideas for further improvements. It might be possible to differentiate the identified cluster structures by adding additional features, such as “education”. The question is, however, whether this data could be obtained from governmental sources with the appropriate mapping to other demographic and geographic characteristics. Apart from that, the geographic feature had to be mapped to the level of regions due to the reasons explained in Section 4. Possibly, finer granular data could be available after the 2022 census currently conducted in Germany. Moreover, the geographic data would also be more up-to-date and other genders as already mentioned in Section 4 could be considered.

Yet, an important step has been taken toward a solution in the field of customer segmentation for the power of attorney business. With achieving all objectives which were set, marketing campaigns can now be carried out more effectively by means of the results obtained.

Acknowledgements

This research is supported by JURA DIREKT GmbH, Nuremberg/Germany. We thank Mr. Domenico Anic to make this research possible. Furthermore, we are grateful to him and his team for their time and their great support.

REFERENCES

- [1] G. Asokan and S. Mohanavalli. Fuzzy Clustering for Effective Customer Relationship Management in Telecom Industry. In D. Nagamalai, E. Renault, and M. Dhanuskodi, editors, *Trends in Computer Science, Engineering and Information Technology*, pages 571–580, Berlin, Heidelberg, 2011. Springer.
- [2] J. C. Bezdek. A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms. *IEEE Trans. on PAMI*, 2(1):1–8, 1980.
- [3] J. C. Bezdek and J. C. Dunn. Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions. *IEEE Transactions on Computers*, 24(8):835–838, 1975.
- [4] Bundesministerium der Justiz. Bürgerliches Gesetzbuch (BGB): § 1829 Genehmigung des Betreuungsgerichts bei ärztlichen Maßnahmen, 2023. https://www.gesetze-im-internet.de/bgb/_1829.html (Online; Last accessed: 2023-05-15).
- [5] Bundesnotarkammer. Jahresbericht und Statistik: Das Zentrale Vorsorgeregister in Zahlen, 2023. <https://www.vorsorgeregister.de/footer/jahresbericht-und-statistik> (Online; Last accessed: 2023-05-15).
- [6] S. Das and J. Nayak. Customer Segmentation via Data Mining Techniques: State-of-the-Art Review. In J. Nayak, H. Behera, B. Naik, S. Vimal, and D. Pelusi, editors, *Computational Intelligence in Data Mining*, pages 489–507. Springer Nature Singapore, 2022.
- [7] N. de Vries and P. Moscato. Consumer Behaviour and Marketing Fundamentals for Business Data Analytics. In P. Moscato and N. de Vries, editors, *Business and Consumer Analytics: New Ideas*, pages 119–164. Springer Nature Switzerland, 2019.
- [8] N. de Vries, L. Olech, and P. Moscato. Introducing Clustering with a Focus in Marketing and Consumer Analysis. In P. Moscato and N. de Vries, editors, *Business and Consumer Analytics: New Ideas*, pages 165–212. Springer Nature Switzerland, 2019.
- [9] T. Eckes and H. Roßbach. *Clusteranalysen*. Kohlhammer, Stuttgart et al., 1980.
- [10] J. C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 27(4):857–872, 1971.
- [11] R. Gupta, H. Kumar, T. Jain, A. Shrotriya, and A. Sinha. Demographic customer segmentation of banking users based on k-prototype methodology. *12th International Conference on Cloud Computing, Data Science & Engineering*, pages 578–584, 2022.
- [12] P. M. Hoffmann, U. Hütter, M. T. Korte, and C. von Ferber. Die Lebenslage älterer Menschen mit rechtlicher Betreuung: Abschlussbericht zum Forschungs- und Praxisprojekt, 2005. <https://www.bmfsfj.de/resource/blob/78932/459d4a01148316eba579d64cae9e1604/abschlussbericht-rechtliche-betreuung-data.pdf> (Online; Last accessed: 2023-05-15).
- [13] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR)*, pages 651–666, 2010.

- [14] G. Jandaghi, H. Moazzez, and Z. Moradpour. Life insurance customers segmentation using fuzzy clustering. *World Scientific News*, pages 24–35, 2015.
- [15] Kester-Haessler-Forschungsinstitut für Betreuungsrecht. Entwicklung des Betreuungsrechts / Betreuungszahlen, 2022. <http://www.betreuungsrecht.de/betreuung/entwicklung-des-betreuungsrechts-betreuungszahlen/> (Online; Last accessed: 2023-05-15).
- [16] N. R. Maulina, I. Surjandari, and A. M. M. Rus. Data Mining Approach for Customer Segmentation in B2B Settings using Centroid-Based Clustering. In *16th International Conference on Service Systems and Service Management (ICSSSM), 3-15 July 2019*. IEEE, 2019.
- [17] S. Mehrotra and S. Kohli. Data Clustering and Various Clustering Approaches. In S. Bhattacharyya, S. De, I. Pan, and P. Dutta, editors, *Intelligent Multidimensional Data Clustering and Analysis*, pages 90–108. IGI Global, 2016.
- [18] Meta Platforms. Ad targeting: Help your ads find the people who will love your business., 2023. <https://www.facebook.com/business/ads/ad-targeting>, (Online; Last accessed: 2023-05-15).
- [19] S. Munusamy and P. Murugesan. Modified dynamic fuzzy c-means clustering algorithm - application in dynamic customer segmentation. *Applied Intelligence*, 50:1922–1942, 2020.
- [20] J. Nayak, B. Naik, and H. Behera. Fuzzy C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014. *Computational Intelligence in Data Mining*, 2:133–149, 2015.
- [21] R. Punhani, V. Arora, A. Sai Sabitha, and V. Shukla. Segmenting e-Commerce Customer through Data Mining Techniques. *Journal of Physics: Conference Series*, 1714(1):012026, 2021.
- [22] H. Schäfer, J. L. Viegas, M. C. Ferreira, S. M. Vieira, and J. M. C. Sousa. Analysing the segmentation of energy consumers using mixed fuzzy clustering. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–7, 2015.
- [23] T. Schlager and M. Christen. Market Segmentation. In C. Homburg, M. Klarmann, and A. Vomberg, editors, *Handbook of Market Research*, pages 939–967. Springer Nature Switzerland, 2022.
- [24] Statistische Ämter des Bundes und der Länder. Regionaldatenbank Deutschland, 2023. <https://www.regionalstatistik.de/genesis/online?operation=table&code=12111-04-01-4-B&bypass=true&levelindex=0&levelid=1659423506709#abreadcrumb> (Online; Last accessed: 2023-05-15).
- [25] Statistisches Bundesamt. Bevölkerung: Mitten im demografischen Wandel, 2023. <https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/demografie-mitten-im-wandel.html> (Online; Last accessed: 2023-05-15).
- [26] Statistisches Bundesamt. Bevölkerungsstand: Bevölkerung nach Familienstand, 2023. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/familienstand-jahre-5.html> (Online; Last accessed: 2023-05-15).
- [27] Stiftung Warentest. So funktioniert die Patientenverfügung, 2022. <https://www.test.de/Vorsorgevollmacht-und-Patientenverfuegung-Wie-Sie-rechtzeitig-Klarheit-schaffen-4641470-5384641/> (Online; Last accessed: 2023-05-15).
- [28] D. Zheng. Application of Silence Customer Segmentation in Securities Industry Based on Fuzzy Cluster Algorithm. *Journal of Information & Computational Science*, (10):4337–4347, 9 2013.

Designing a Use Case for Supply Chain Resilience Based on Process Mining

Angela
Sekulovska

University of
Ljubljana
Kardeljeva
ploščad 17
1000 Ljubljana
as0052@
student.uni-lj.si

Frank
Morelli

Pforzheim
University
Tiefenbronner
Straße 65
75175 Pforzheim
frank.morelli@
hs-pforzheim.de

Artur
Siurdyban

Process Mining
Capability Lead
Europe
Wrocław, Poland
artur.siurdyban@
accenture.com

Anton
Manfreda

University of
Ljubljana
Kardeljeva
ploščad 17
1000 Ljubljana
anton.manfreda@
ef.uni-lj.si

Frank
Schätter

Pforzheim
University
Tiefenbronner
Straße 65
75175 Pforzheim
frank.schaetter@
hs-pforzheim.de

KEYWORDS

Supply Chain Resilience Management, Process Mining, SCOR Model, Key Resilience Areas, Design Science Research

ABSTRACT

Amidst global disruptions, the significance of Supply Chain Resilience (SCR) has surged in scholarly and practical discourse. This paper endeavors to craft an industry-neutral conceptual dashboard, tailored for in-house consultants, to measure and oversee SCR. Drawing from the SCOR model's resilience metrics, Key Resilience Areas (KRAs), and Accenture's SCR application, this study presents a practical use case for deploying such a dashboard. Expert evaluation by process mining and supply chain specialists highlighted potential enhancements for the dashboard.

INTRODUCTION

In recent times, a series of events took place that have captivated economies worldwide and triggered major global repercussions. These events, which range from political conflicts to natural disasters, have caused substantial disruptions in global supply chains (SCs), affecting everything from raw materials, movement of people to finished goods and SC operations (Sheng et al., 2020). Such events prove relentlessly that “we live and operate in an ever-changing and turbulent environment” (Neubauer, 2018).

As a consequence of the growing frequency and size of SC disruptions, companies are facing significant challenges due to continuous changes in consumer behaviors, markets, and SCs, leading to an imbalance in their operations. Consequently, the concept of Supply Chain Resilience Management (SCRM) has garnered considerable attention in recent times. Global SCs have been made public by events like the COVID-19 pandemic, emphasizing the urgent need for enterprises to proactively manage SC risks (Bret et al, 2021). These destabilizing effects have highlighted how crucial it is to establish robust supply networks that can take shocks, quickly adapt to

changing conditions, and recover effectively. The speed and magnitude of these changes necessitate swift responses from leaders who must embrace agile methodologies and accelerate the transformation of their chains. Enhancing the resilience of processes involves applying new insights and abilities while adopting alternative measurements for process-focused performance (Schätter et al, 2022). Additionally, robust data and analytics capabilities are crucial for comprehending complications, anticipating potential disruptions, and rapidly formulating effective strategies in response (Accenture, 2023). Resilience is both defined as a capacity that is developed in response to a shock or disruption and as being more pro-active in helping the company get ready for a disruption (Melnyk et al, 2014). The capability to learn from past disruptions to be able to predict and overcome future disruptions, by reducing the level of risk and being aware of the SC's vulnerabilities (Aman & Seuring, 2021). Therefore, being aware of SC vulnerabilities and possible risks can increase the effectiveness of the response to future disruption, mitigate its damage and restore the state of the SC.

On the other hand, over the past two decades process mining (PM) has emerged as a particularly promising contender and as a relatively new field of study that falls somewhere between process modeling and analysis, computational intelligence, and data mining (van der Aalst et al., 2012). PM aims to identify, monitor, and improve business processes by extracting data from the event logs that are easily accessible in today's information systems. Event data has drastically increased in the past decades, and subsequently many PM techniques have also advanced significantly. Thus, PM has become highly important in management trends that are concerned with process improvement (van der Aalst, 2012; van der Aalst, 2016).

The foundation of PM is an event log, which contains events that relate to a specific process instance aka *a case* and refers to a clearly defined step in a process aka *an activity*. In other words, a case is a specific identifier,

similar to an item from a purchase order, an invoice, or an order number, while an activity is a description of what took place, such as when a purchase order was created or when goods were received. It is necessary to chronologically order and consider all of the events that were connected to one case as "one run" of a process. Nevertheless, event logs can hold further information related to the events, such as a *timestamp* that indicates when an event has occurred, details about the resource i.e., device or person that is initiating or performing the activity, or *data elements* recorded along with the event (van der Aalst, 2012).

The information gathered from event logs is usually presented on dashboards, so that the person responsible can quickly identify issues and take corrective actions to help the organization perform better (Rasmussen et al., 2009, p.3). Pauwels et al., (2009) defines a dashboard as a reasonably compact group of linked key performance indicators and fundamental performance drivers that represents both immediate and long-term objectives and can be viewed by all members of the organization. A dashboard is a visual representation of the most crucial data required to accomplish one or more goals, collected, and organized on a single screen for easy monitoring (Few, 2006, p.26). Managers frequently use dashboards to monitor a company's performance, and since the COVID-19 pandemic, their use has increased (Reibstein et al., 2023). Dashboards can be categorized based on their level of detail and timeframe into strategic, tactical, and operational tiers (Few, 2006, p.33).

This paper aims to design a conceptual industry-neutral SCRM dashboard for the Purchase-to-Pay (P2P) process based on the event log used in PM. Tailored for in-house consultants at the mid-level management, the mockup dashboard aims to address challenges in ensuring uninterrupted supplies during unforeseen disruptions from the side of inbound logistics. Consequently, the objective of the paper is to answer the following research questions:

1. *How can insights into the resilience of the supply chain through the P2P process be obtained using an SCRM dashboard conception, considering a specific use case?*
2. *Which opportunities and limitations does the dashboard offer for resilience management of the P2P process from a tactical or strategic point of view?*

This paper stands out because it consolidates various resilience-focused theoretical concepts, forming a robust foundation for crafting a mockup dashboard for SCRM. Notably, it stands as a pioneering work by relying on the newest version of the SCOR model, introducing a redefined resilience performance category. This category further breaks down into three distinct sections, encompassing a broader range of metrics across diverse supply chain processes.

Interestingly, there's a noticeable gap in existing literature where no prior papers have delved into the design of such a dashboard, let alone the creation of a mockup that could serve as a visual aid for practitioners. The objective was to illustrate how an SCRM dashboard could be tailored to cater to users at both strategic and tactical levels, potentially leveraging the capabilities of process mining. While the dashboard includes certain design principles and heuristics, its primary aim was to offer a visual summary, essentially acting as an initial framework for an IT artifact. Furthermore, expert evaluation was conducted to gain deeper insights into the essential components that such SCR dashboards should encompass.

Throughout this entire process, our endeavor remained guided by the seven principles of Design Science Research (DSR), ensuring a systematic and methodical exploration of creating and evaluating this innovative IT artifact for SCRM (Hevner et al. 2004 ; Peffers et al., 2007). However, the structure of this paper doesn't strictly adhere to the predefined process activities of DSR. Instead, it incorporates and considers these principles when designing the mockup dashboard for SCRM, aligning with the advice of Hevner et al. (2004) who encourage flexible incorporation rather than strict adherence to predefined steps. Elaborating extensively on DSR principles might potentially divert attention from our main objective, which is to showcase the development and evaluation of this innovative IT artifact for SCRM.

PRACTICAL MEASURES OF SCR

In the ever-evolving landscape of global commerce, ensuring the resilience of a SC has become a paramount concern for organizations across industries. To address this challenge, a multitude of frameworks, solutions, and metrics have been developed and proposed in the academic and business literature. When designing a mockup dashboard for supply chain resilience management, it is essential to select the most effective and globally recognized tools. In this endeavor, we turned to the Supply Chain Operations Reference (SCOR) model, a framework widely adopted worldwide for SCM (ASCM, 2022). Additionally, we incorporated the Key Resilience Areas (KRAs) concept, a set of practical measures devised by Schätter et al. (2022), to help organizations pinpoint vulnerabilities within their SCs and evaluate their current state. This combined approach provides a comprehensive foundation for enhancing SCR and serves as the backbone for the subsequent sections that delve into the specifics of SCOR model resilience metrics and the utilization of KRAs in this context.

SCOR Resilience Metrics

The SCOR model is a comprehensive framework that provides organizations with a standardized approach for managing and improving their SC processes. Developed in 1996 by the Supply Chain Council, the SCOR model offers a common language and set of metrics to define, measure, and evaluate SC activities (Selitto et al., 2015). Its newest version, SCOR 14.0, outlines seven key SCM

management processes: Orchestrate, Plan, Source, Transform, Return, Fulfill, and Order. These processes cover the end-to-end SC, from strategic planning and sourcing of materials to manufacturing, logistics, and customer service (ASCM, 2022). The SCOR model enables organizations to map and analyze their SC operations, identify areas for improvement, and optimize performance. By aligning with the SCOR model, businesses can enhance collaboration, increase efficiency, and achieve greater SC visibility and control. It serves as a valuable tool for organizations across industries seeking to optimize their SC operations and drive competitive advantage (Phadi & Das, 2021).

Within the SCOR model, the Performance section places emphasis on measuring and evaluating the results achieved through the execution of SC processes. This approach enables a thorough examination, evaluation, and diagnosis of SC performance. By considering these elements, organizations can successfully measure the effectiveness and efficiency of their SC operations and identify areas for improvement. Performance measurement is done by performance attributes which include various metrics. In fact, each attribute has three levels of metrics that function as diagnostic for the previous level. Therefore, Level-3 metrics act as diagnostic indicators for Level-2 metrics and so on. This provides valuable insights into the overall performance of the SC and enables organizations to identify areas for improvement or address performance gaps at various levels (ASCM, 2022).

SCOR 14.0 has redefined their performance attributes by categorizing them under Resilience, Sustainability, and Economic performance. The focus will be on resilience-only performances, for the purpose of this paper. The model provides three sub-attributes under “Resilience”, this are Reliability (RL), Responsiveness (RS), and Agility (AG). Below you can find the description of each resilience attribute:

Reliability (RL): Deals with the ability to execute tasks in accordance with established expectations. It concerns ensuring the predictability of process outcomes. Key metrics associated with Reliability encompass the timely delivery of products or services, meeting the prescribed quantity, and providing accurate documentation.

Responsiveness (RS): Focuses on how quickly tasks related to customer orders are completed. It looks at how fast businesses can respond and satisfy customer needs.

Agility (AG): Describes the ability to respond to unplanned external influences, disruptions and/or events.

Key Resilience Areas (KRAs)

The paper by Schätter et al (2022) “Supply Chain Resilience Management Using Process Mining” aims to identify relevant data needed for strategic monitoring of resilience status and provide a base for using PM in SCRM. The result is eight distinguished KRAs, which help to

have an initial understanding of the state of the SC, and is easier to apply than more complicated analytical methods. Below you can find a short description for each of the KRAs:

- **KRA1 - Geographic Distribution:** Provides visibility into the locations and distribution of entities, for example assessing the risk of disruptions due to supplier concentration in one specific area or country.
- **KRA2 - Sourcing Strategy:** Evaluates the impact of supply disruptions by analyzing material sourcing strategies, with a focus on single-sourcing vs multi-sourcing approaches.
- **KRA3 - Warehouse Materials:** Enhances resilience through buffer stocks of critical materials in warehouses, ensuring SC continuity.
- **KRA4 - Average Storage Time:** Measures the average time materials are stored in warehouses to identify critical materials and assess buffer effectiveness.
- **KRA5 - Transport Delays:** Identifies critical transport relations by analyzing delivery durations and assessing delay-prone materials.
- **KRA6 - Consolidation of Deliveries:** Promotes cost-effective and resilient SC practices by consolidating (combining) material deliveries.
- **KRA7 - Transport Distance:** Analyzes transport distances and regional deliveries to reduce the risk of large-scale disruptions.
- **KRA8 - Intra-Logistics Processes:** Offers transparency into internal supply chain processes, assessing dependencies on critical materials and revealing potential weaknesses.

PM DASHBOARD FOR SCRM

Integration of resilience concepts

The theoretical foundation of the research is composed of three fundamental concepts that were discussed in the previous chapters. These concepts include SCR, the SCOR model along with its resilience attributes, and the KRAs. Notably, Accenture’s SCR application significantly contributed to the conceptual development of the SCRM mockup dashboard. This application was developed in collaboration with Celonis, a leading process mining vendor, and aims to assist companies in enhancing their SCR by leveraging real-time data on disruptive events, conducting SC operational risk, assessment, and providing intelligent solutions to mitigate disruptions (Celonis, n.d.).

Throughout the course of the research, collaborative discussions were held with the Process Mining Capability Lead for Europe, who played an important role in clarifying the concept of SCR. Mr. Being directly involved in the development of Accenture’s application, he provided valuable insights, and expertise that greatly influenced the design of the mockup dashboard.

The SCRM mockup dashboard drew inspiration from Accenture's application and skilfully incorporated theoretical insights from the SCOR model, KRAs, and various other elements such as OLAP tables and visualizations that are often included in PM solutions. Although we were granted access to Accenture's application, its functionality and elements served as an inspiring reference rather than a direct replication.

Due to the inherent complexity of covering the SC from end-to-end, it became necessary to narrow the scope of the research. Consequently, it was decided that the mockup dashboard would be industry neutral, offering greater flexibility and standardization, particularly in integrating the resilience metrics from the SCOR model. During the process of considering Accenture's application as a reference for the dashboard design, a thorough analysis of its functionalities revealed a primary focus on tracking external risks associated with suppliers, materials, and supplier's plants, predominantly covering the inbound logistics aspect of the SC. As a result, while narrowing the scope, the decision was made to concentrate on the inbound logistics, requiring the selection of a specific process within this domain. Additionally, Accenture's application was primarily designed to cater to users at the operational level, facilitating immediate action to prevent disruptions. However, for this conceptual dashboard the focus shifted to mid-level management, envisioning the end-users as in-house consultants. In this role, they would be responsible for investigating any anomalies and disruptions, and subsequently reporting their findings directly to the executives.

The final step in delimiting the scope was to select an appropriate process within the inbound logistics domain that would serve as the foundation for the envisioned dashboard's construction. This selection would determine the specific resilience metrics and elements incorporated into the dashboard. The procedure of selecting the process is discussed in more detail within the following subchapters.

Selection of P2P process for the Use Case

The decision to focus on the P2P process for the use case and the dashboard design was carefully considered, driven by practicality and strategic intent. The P2P process, also known as procure-to-pay, is a fundamental and widely adopted component in SCs across various industries. It refers to the procurement activities of companies, and with the increasing digitalization, businesses have developed systems to track the P2P process from end to end, covering everything from requisitioning goods and services to payment to suppliers. These systems offer numerous benefits, including high-cost savings, increased efficiency, and enhanced procurement and financial visibility. In addition, their aim is to optimize the entire purchasing process and streamline operations (Barton, 2023).

Among the various solutions, Celonis offers to its clients, one focuses on procurement, specifically for the P2P process, and has accumulated numerous success stories. As part of its offerings, Celonis Academy provides a P2P demo, which effectively showcases how PM can be applied to enhance the P2P process. Within this demo, a simplified specified "to-be" P2P process model is included. Since the research did not center around a specific company, this process model served as an example for visualizing the activities within the P2P process. Additionally, when selecting resilience-related metrics from the SCOR model, this process model was taken into consideration. Taking into account Accenture's SCR application and the domain of inbound logistics, the procure-to-pay (P2P) process seemed like a reasonable choice for the use case and the design of the mockup dashboard.

P2P and the SCOR Model

The SCOR model categorizes six major processes at Level - 0 and 1, namely Orchestrate, Plan, Order, Source, Transform, Fulfill and Return. Among these processes, the Source process emerged as the most suitable in the context of the research, specifically within the inbound logistics domain. The Source process encompasses a series of activities associated with procurement, order placement, order scheduling, handling deliveries, receipt of goods and services, and their subsequent transfer within the SC. It is further broken down into four sub-processes: S1 Strategic Procure, S2 Direct Procure, S3 Indirect Procure, and S4 Source Return (ASCM, 2022).

Nevertheless, the SCOR model does not directly and explicitly cover the P2P process, since the process itself falls within the activities listed under the Source process. In the attempt to find a process that closely resembled or effectively covered the P2P process, after careful consideration, the S2 Direct Procure was selected, which by the definition is the procurement process for acquiring products or services that are part of the final product. The S2 Direct Procure includes multiple activities organized in a hierarchy as follows: S2.1 Establish Order Signal, S2.2 Schedule Product Delivery, S2.3 Manage Inbound Transport, S2.4 Receive Product, S2.5 Inspect and Verify, S2.6 Transfer Product, and S2.7 Authorize Supplier Payment (ASCM, 2022). Subsequently, from the listed activities only the first five were considered as they matched the nature of the activities from the P2P process. Subsequently, S2.6. Transfer Product and S2.7 Authorize Supplier Payment were not considered when it came to the selection of the resilience metrics. Therefore, with the selection of the S2 Direct Procure process as a surrogate for the P2P process it was possible to identify resilience-related metrics.

After choosing the process S2 Direct Procure and the subsequent sub-processes that applied to the P2P process, the metrics related to the resilience performance attributes were selected. Later, these metrics were incorporated into the design of the mockup dashboard. Given the abundance of resilience metrics available, it was essential

to narrow down the choices while ensuring an adequate representation of metrics falling under the resilience performance attributes. To achieve this, each metric's description and formula provided by the SCOR model was considered, ensuring that the chosen metrics capture the essence of reliability, responsiveness, and agility within the process (Table 1).

Reliability	
Metric	Description
RL.3.13 Delivery Item Accuracy from the Supplier	The percentage of orders in which all items ordered from the supplier are the items actually provided and no extra items are provided.
RL.3.14 Delivery Quantity Accuracy from Supplier	The percentage of orders in which all quantities delivered by the supplier match the ordered quantities, within mutually agreed tolerances.
RL.3.15 Supplier Achievement to Original Organization Commit Date	The percentage of orders that are received by the organization's commit date.
RL.3.16 Delivery Organization Location Accuracy	The percentage of orders that are delivered to the correct organization location.
RL.3.46 Fill Rate	Fill rate, from the supplier's perspective, is a measure that quantifies the level of order fulfillment achieved within a specific timeframe, typically expressed as a percentage.
Responsiveness	
RS.3.37 Average Days per Schedule Change	Number of days each schedule change impacts the delivery date/Total number of changes.
RS.3.38 Average Release Cycle of Changes	Cycle time for implementing change notices/Total number of changes.
Agility	
AG.3.6 Additional Source Volumes Obtained in 30 days	Additional source volumes determined.
AG.3.12 Current Source Volume	Amount of each item purchased.

Table 1: Selected metrics from the SCOR model. Adapted from ASCM (2022)

In addition to the metrics selected from the S2 Direct Source, the mockup dashboard also incorporates higher-level metrics that enhance its comprehensiveness. Among these metrics is the Level 1 metric for reliability, denoted as RL.1.2 Perfect Supplier Order Fulfillment. This metric gauges the extent to which supplier orders fulfill delivery expectations, encompassing accurate documentation, complete deliveries, and absence of damage. In the context of the mockup dashboard, this metric portrays the collective effectiveness of all suppliers in meeting order fulfillment requirements. Furthermore, the dashboard integrated the Level 2 metric for reliability, RS.2.2 Source Cycle Time. This metric, which represents the average duration for completing Source processes, has been strategically adapted in the dashboard as the "Order Fulfillment Cycle." Through this translation, it signifies the average time suppliers take to fulfill orders (ASCM, 2022). The inclusion of these additional metrics underscores the dashboard's aim to provide a holistic view of supplier performance and order fulfillment dynamics.

P2P and KRAs

The last theoretical concept that was considered within the creation of the mockup dashboard were the KRAs, which were developed as a conceptual framework for evaluating the current state of SCR. The KRAs have been

designed to identify and assess any vulnerabilities within the SC that could impact its operations during a disruptive event. Based on the nature of the P2P process, three specific KRAs were carefully selected and considered in the design phase of the mockup dashboard. Below you can find the selected KRAs for the P2P process along with the reasoning behind the choice:

KRA1 - Geographic Distribution of Entities is focused on gaining visibility into the locations and distribution of entities within the SC, such as factories, warehouses, suppliers, and customers. Understanding the geographic spread of these entities provides an initial indication of the SCR. For instance, identifying large concentrations of suppliers in specific 53 areas can highlight potential risks of significant disruptions if a risk affects an entire region (Schätter et al., 2022). Furthermore, KRA1 in the context of the P2P process, proves essential as it offers a comprehensive overview of all suppliers and their respective locations, including their density. For instance, if a company heavily relies on suppliers concentrated in a specific area or country, a disruptive event like an earthquake or a flood could pose significant risks. Such an event may damage not only the suppliers' warehouses but also disrupt the overall infrastructure of the affected country, consequently impacting the flow of materials within the SC. Additionally, it could present challenges in promptly finding alternative suppliers to ensure uninterrupted delivery of goods.

KRA2 - Sourcing Strategy of Materials is focused on assessing the impact of the sourcing strategy for materials within the SC. While a single-sourcing approach may offer some cost benefits, it also poses significant risks in managing supply disruptions. In such cases, there might be no redundancies for these materials, leaving the SC vulnerable to shortages in case of supplier failures (Schätter et al., 2022). Based on the description of KRA2, it is evident that it holds significant relevance for the assessment of the P2P process, given its direct involvement with sourcing strategies of materials. This aspect proves valuable in evaluating the resilience state, particularly when a manufacturer heavily relies on a single supplier for a critical material. If the sole supplier faces external or internal challenges and becomes unable to fulfill its obligations, it could potentially lead to severe disruptions for the manufacturer. To mitigate this risk, it becomes imperative to proactively minimize dependency on a single supplier by diversifying the sourcing strategy and identifying alternative suppliers in advance.

KRA5 - Transport Delay is focused on evaluating transport delays within the SC. If target data about delivery durations are compared with actual delivery durations, the delays within the SC can become evident. Nonetheless, based on this critical transport relations can then be identified (Schätter et al., 2022). Receiving purchase orders (POs) on time is of crucial importance for ensuring the seamless flow of operations for businesses. Consequently, KRA5 proposes a critical aspect that must

be carefully considered within the P2P process, which involves analyzing the promised and actual delivery dates of suppliers. By relying on such data points, companies can make informed decisions about the reliability of suppliers in fulfilling their committed delivery dates. Such analysis empowers businesses to take appropriate actions, whether it involves continuing to work with dependable suppliers or reevaluating their partnerships with those who consistently miss their delivery commitments.

Mockup Dashboard

The mockup dashboard was designed to comprehensively address the three distinct aspects of resilience outlined in the SCOR model: reliability, responsiveness, and agility. Consequently, the dashboard structure comprises four distinct tabs, each dedicated to a specific facet of resilience. These tabs are identified as follows: Resilience Overview P2P, Reliability Deep Dive, Responsiveness Deep Dive, and Agility Deep Dive. Within the context of each tab, a prominent illustrative banner is positioned in the upper-left corner. This banner acts as a visual bridge, connecting cases derived from the event log to their corresponding appearances in the Celonis EMS dashboard analysis. Adjacent to this banner, an essential component facilitates the sharing and exporting of the content specific to each tab (as depicted in Figures 2 through 5).

Moreover, beneath these elements, another banner is situated. This banner is prominently displayed in each tab along with the name of the tab and with global indicators. It is important for items on a dashboard to have a consistent visual appearance; items that serve the same purpose or convey the same meaning should look the same. To prevent erroneous interpretations and preserve communication clarity, consistency in visual appearance and element selection is crucial (Few, 2006, p.143). Such repetitive patterns can work to simplify visual complexity through familiarity and predictability. The global indicators encompass critical metrics, including total count of POs, the total value of all POs, the average value of POs, and the count of POs exposed. The term "exposed POs" specifically refers to purchase orders that have been exposed to potential risks within the operational context. In the bottom-left corner, the dashboard features a conspicuous tab navigation bar. This navigational bar empowers users to effortlessly transition between different sections. This tab bar provides convenient access to the tabs, facilitating a userfriendly navigation experience (Figure 1 to 4).

The initial tab, labeled "Resilience Overview P2P," is divided into three distinct sections, each corresponding to a fundamental aspect of resilience. Here the gestalt principle of enclosure is incorporated, where fill colors and borders are established to group different informations together and set them apart from each other (Few, 2006, p. 76). Specifically, the "Reliability" section is structured with two distinct components, mirroring the arrangement found in the other two sections. The initial component is closely linked with the "Perfect Supplier Order" metric,

as depicted within the SCOR model. Simultaneously, the second component is exclusively centered around the "Supplier at Risk" metric, which was taken as an idea from Accenture's SCR application. Within this section, the dashboard provides a comprehensive presentation of essential information for executives. Strategic dashboards are meant to give the decision maker a quick overview of key metrics and often include simple display mechanisms, and they simply present what is going on (Few, 2006, p.32). The previously mentioned section encompasses details such as the percentage of suppliers categorized as "at risk," the total count of suppliers in jeopardy, and a numerical breakdown of suppliers falling under the "at risk" classification. Collectively, this integrated section aims to underscore the overall reliability of suppliers.

The segments devoted to "Responsiveness" and "Agility" follow a similar structure. However, the "Responsiveness" section pertains to materials and the speed of their delivery, incorporating the SCOR metric known as the "Order Fulfillment Cycle Time." The corresponding element within this section mirrors the components featured in the "Suppliers at Risk" section but is tailored to materials. Lastly, the "Agility" section revolves around countries impacted by risks.

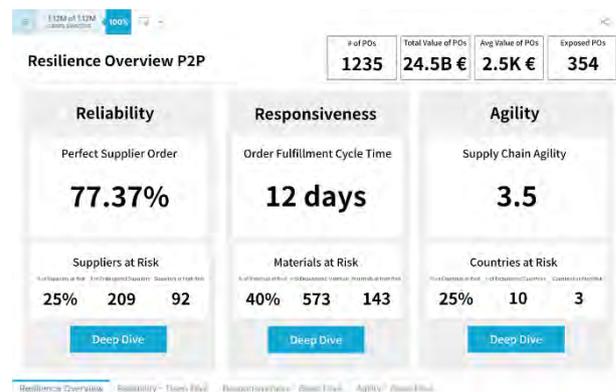


Figure 1: Resilience Overview P2P

Consequently, the KPI for "Supply Chain Agility" represents a conceptual measure designed to evaluate how effectively the SC can adapt to external influences, particularly within the context of inbound logistics. Additionally, this section provides information about countries that are at risk or have been affected by risks, akin to the content found in the "Suppliers at Risk" section. Each of these sections includes a "deep dive" button intended to guide the user to the corresponding resilience aspects within the tabs for in-depth exploration (see Figure 1). Dashboards designed for strategic purposes are not designed to be interactive and yield further analysis, because often strategic managers are not directly responsible (Few, 2006, p. 32). Nevertheless, the "deep dive" buttons main purpose is to take the user to a dashboard with analytical purposes to further explore each aspect of the SC resilience.

Next, the second tab labeled "Reliability Deep Dive" (Fig 2), filters can be found beneath the banner featuring global indicators, which in this tab also includes the total number of suppliers. These filters enable users to narrow down the analysis by supplier, materials, country, risk, and date range (selectable by months). Below this section, you'll encounter the same metrics displayed in the initial dashboard tab. These metrics include the total number of suppliers at risk, the percentage of suppliers at risk, the number of endangered suppliers, and the total value of goods at risk. Directly beneath, an OLAP table can be found which includes the preselected SCOR Reliability metrics, along with the option that would allow users to search for specific suppliers. Tactical dashboards serve as powerful tools primarily intended for conducting in-depth analysis by consolidating and presenting comprehensive data. These dashboards empower users to gain deeper insights into performance metrics over time and thereby foster impactful, data-driven decision-making across all tiers of the business (Rahman et al., 2017).

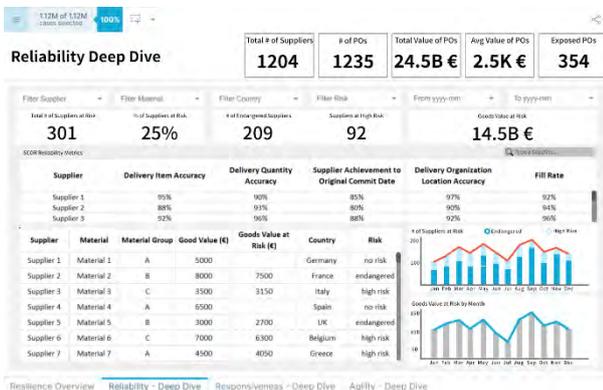


Figure 2: Reliability Deep Dive

Meanwhile, on the right-hand side of the dashboard, two bar charts come into view. The first chart depicts the number of suppliers at risk, divided into two categories: "endangered" and "high risk", plotted across months. The second visualization illustrates the value of goods at risk over months. In the left hand-side another OLAP table can be found that includes the following dimensions: Supplier, Material, Material Group (associated with the ABC analysis), Goods Value in Euros, Goods Value at Risk, Country, and Risk Type (no risk, endangered, and high risk). It's worth noting that the filters are meant to exclusively impact the second OLAP table and the accompanying visualizations, they do not extend their influence on the SCOR metrics presented in the initial OLAP table (see Figure 2). OLAP tables when accompanied with visualizations and through combination of interactive and navigational functions can reduce the recognition load and refine views on any particular area of interest in the data, while enhancing the user's understanding of the exploration (Techapichetvanich & Datta, 2005).

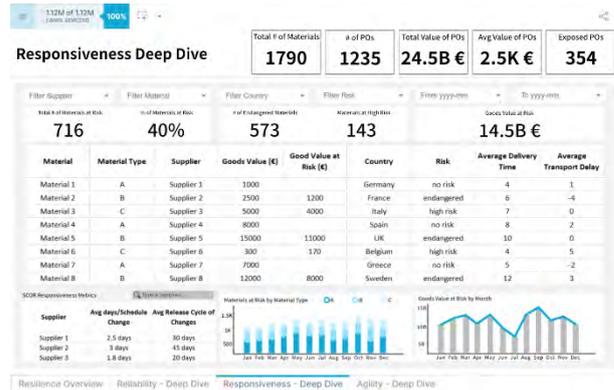


Figure 3: Responsiveness Deep Dive

In the corresponding "Responsiveness Deep Dive" tab, an additional numerical depiction of the total number of materials is provided within the top banner. This tab also features the same filters that were included in the previous section. Similarly, the OLAP table structure is maintained, incorporating several dimensions to facilitate comprehensive analysis. Furthermore, the OLAP table in this tab includes KPIs for both the average delivery time and the average transport delay. Adjacent to this, on the left-hand side of the dashboard, another OLAP table is positioned, encompassing the preselected SCOR responsiveness metrics. On the right-hand side of the dashboard, two visualizations are displayed. One of these visualizations was already introduced in the previous tab, while the second one is a bar chart that specifically illustrates the number of materials at risk categorized by materials type (ABC) across different months (Figure 3).

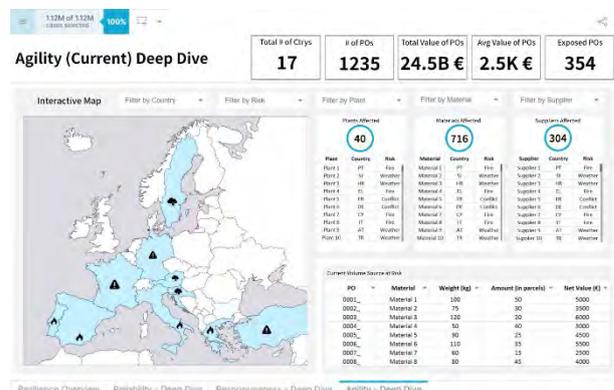


Figure 4: Agility Deep Dive

The final tab of the dashboard is labeled "Agility (Current) Deep Dive" (Figure 4). This tab is designed to be powered by real-time data, which justifies its name "current." Within this tab, users can access a range of filters including country, risk, plant, material, and supplier. The integration of real-time data is intended to provide up-to-the-minute insights into the company's agility. Prominently displayed within this tab is an interactive map that dynamically showcases countries affected by various risks. Geographic Information Systems capabilities, encompassing visualization, spatial analysis, optimization, and site research, significantly contribute to SCM decision-making. They provide crucial insights for supplier

selection, facility location, supply chain network configuration, and asset management, facilitating more efficient SCM and logistics activities while offering valuable guidance for strategic decisions (Tsakiridi, 2021). The inspiration for the risk categories was drawn from Accenture's SCR application, resulting in three risk types: "fire" (indicated by a fire symbol) to denote instances of fires occurring in countries, "conflict" (depicted by a danger symbol) to highlight strikes, wars, or other conflicts at a national level, and "weather" (represented by a weather symbol) which encompasses floods, storms, and other supernatural weather-related events. Positioned on the left-hand side are three tables, each presenting data on affected plants, materials, and suppliers. The tables furnish essential numerical indicators along with pertinent information, including the plant's name, the material and supplier impacted, the associated country, and the specific risk type (fire, weather, or conflict). Initially, the interactive map illustrates all countries affected by risks, with corresponding symbols indicating the risk type.

However, once filters are applied (as depicted in Figure 5), such as selecting a specific risk type like "fire," both the map and the corresponding tables undergo dynamic changes to reflect the filtered criteria. The culminating feature on this tab is an OLAP table, meticulously curated to include the chosen agility metrics derived from the SCOR model, specifically centered around "Current Volume Source." This table showcases comprehensive information, encompassing each PO at risk, the corresponding material ordered, its weight, quantity, and individual value.

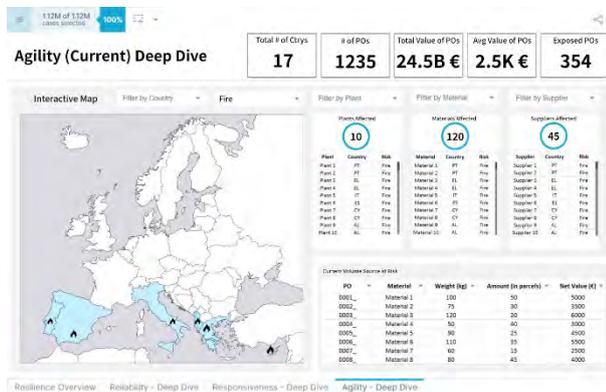


Figure 5: Agility Deep Dive with the Fire filter applied

P2P Use Case

The Use Case is essentially a demonstration walkthrough of the intended dashboard functionality, offering a comprehensive understanding of how the dashboard achieves its core goals. To enhance reader comprehension, additional figures (Figures 6 to 8) have been included. These figures offer visual cues that correspond to the steps a potential inhouse consultant would take while utilizing the dashboard.

As a hypothetical scenario, the in-house consultant initiates access to the dashboard within the Celonis EMS.

Upon entry, their initial view is directed to the "Resilience Overview P2P" tab. Subsequently, they are provided the option to investigate a specific aspect, such as "Reliability," which encompasses two distinct sections. Within the first section, a vital metric—termed "Perfect Supplier Order"—is showcased. In this instance, the metric registers at 77.37%. This metric serves as a comprehensive indicator of collective supplier performance in fulfilling their orders. Notably, the maximum attainable value is 100%, signifying optimal efficiency in meeting the company's order delivery requirements. In the present case, the metric reveals certain suppliers demonstrating inefficiency in fulfilling orders. Transitioning to the second focal point, as depicted in Figure 7, the in-house consultant encounters three pivotal information metrics. These metrics unveil that a quarter (25%) of the suppliers are in a precarious state of risk. Of this subset, 209 suppliers are categorized as "endangered," while 92 are flagged as "high risk." The identification of such risk-laden suppliers raises a preliminary flag, indicating potential concerns warranting a more profound exploration. Consequently, the in-house consultant possesses the option to delve deeper by engaging the "Deep Dive" button—an action depicted as numeral 3 in Figure 6.

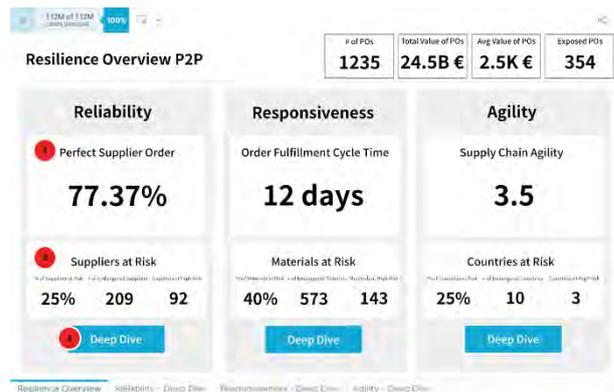


Figure 6: Resilience Overview P2P Demonstration

If the consultant opts to conduct a more in-depth exploration of the emerging issues, they can initiate this process by clicking the "Deep Dive" button, which leads them to the dedicated "Reliability Deep Dive" tab (Figure 7). Here, the in-house consultant gains the opportunity to observe the comprehensive list of suppliers, irrespective of their risk classification—indicated as point number 1 in Figure 7. To proceed, the consultant might choose to refine their focus by utilizing the filtering option to isolate suppliers categorized as "high risk." This action, illustrated as point number 2 in Figure 7, results in the display of exclusively those suppliers that have been identified as being exposed to elevated risk levels. Subsequently, the in-house consultant can elect to delve deeper into the performance of these specific high-risk suppliers based on the SCOR Reliability metrics. This entails leveraging the search functionality to scrutinize the historical performance of these suppliers. Should the consultant's concerns extend to comprehending the external factors influencing the heightened risk for these suppliers, they can navigate to the "Agility Deep Dive" section.

This option is accessible via the tab menu situated at the lower part of the page, as denoted by point number 4 in Figure 7.

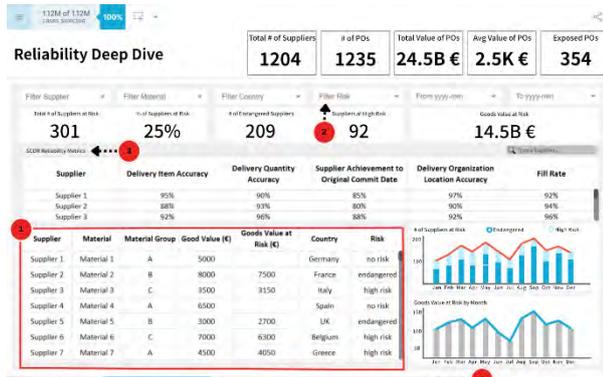


Figure 7: Reliability Deep Dive Demonstration

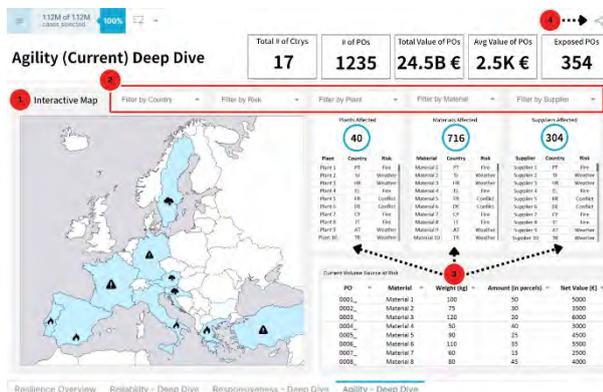


Figure 8: Agility Deep Dive Demonstration

Upon accessing the "Agility Deep Dive" tab, the in-house consultant is presented with an interactive map showcasing countries influenced by risks. This interactive map encompasses multiple functionalities, although in this mockup representation, it primarily serves as a visual depiction of countries impacted by risks—indicated as the primary function (point number 1 in Figure 8). To refine the analysis's scope, the consultant possesses the capability to implement various filters. These filters hold the potential to induce dynamic changes both within the interactive map and across the three associated tables (point number 3 in Figure 8). These tables correspond to plants, materials, and suppliers, and their content adapts according to the selected filters. Upon achieving a satisfactory level of insight, should the inhouse consultant consolidate their findings and share them with organizational executives, an option for data extraction is available. This functionality is readily accessible through a button situated in the top-right corner of the interface, which can be found also in all the tabs (point number 4 in Figure 8).

Moreover, the design of the mockup dashboard facilitates an exploration of the KRAs to evaluate SCR. Specifically, the previously mentioned KRAs—KRA1: Geographic Distribution of Entities, KRA2: Sourcing Strategy of Materials, and KRA5: Transport Delay—can be assessed using various components within the mockup

dashboard. For instance, KRA1 can be hypothetically evaluated through interactive map functionality or by employing filtering options within the "Reliability Deep Dive" tab (Figure 2). These features would provide insights into the number of suppliers originating from specific countries. Likewise, the assessment of KRA2 can be accomplished using filters within the "Responsiveness Deep Dive" tab (Figure 3), shedding light on sourcing strategies for materials. Addressing KRA5, the mockup dashboard incorporates dedicated metrics to offer an overview of transport delay metrics, encompassing the time taken to procure particular materials and any associated delays (Figure 3).

EXPERT EVALUATION

The evaluation phase was a crucial step aiming to validate the design artifacts created during the development process. In this chapter, we present the results of the evaluation activity for the SCR mockup dashboard, which was conducted with two experts, referred to as Expert 1 and Expert 2, to maintain their anonymity. The evaluation process involved separate online meetings with each expert, during which the dashboard's design, functionality, and potential improvements were discussed. The meetings were conducted by presenting the use case from the previous chapter and the discussion afterwards was conducted openly. Subsequently, the discussions with both experts were recorded and transcribed. From these transcripts, themes within the discussions were identified for coding the expert's comments and organizing them into categories/themes of discussion.

Expert Profiles

- Expert 1 holds a background in finance and has transitioned into business analysis, with a primary focus on PM projects. This professional trajectory has equipped Expert 1 with a unique perspective on development of data-driven and process optimization solutions. In addition, they originally authored Accenture's SCR application and they were directly involved in its development.
- Expert 2 serves as a Senior Manager at Accenture within the field of SC and operations strategy. With over 15 years of consulting experience, Expert 2 specializes in guiding clients through strategic transformation endeavors, encompassing a holistic view of processes. They were not familiar with the SCR application that Accenture had developed.

Evaluation Process

The evaluation process commenced by introducing the use case for the SCR mockup dashboard to each expert separately. A presentation outlined the dashboard's functionalities, the use case it addresses, and its intended impact on SCR. Following the presentation, in-depth discussions ensued, delving into various aspects of the dashboard's design.

Themes of Discussion

I - Risk Assessment and Analytical Depth

Both experts provided valuable perspectives on risk assessment methodologies. Expert 1 discussed the SCOR model as an alternative approach to assess resilience, diverging from Accenture's SCR application, which followed an asset-driven approach. Expert 1 elaborated that the asset-driven method considers all company assets exposed to risks, such as materials, suppliers, and even supplier's plants. Despite supplier plants not being owned by the company, they house materials that may eventually belong to the company. In case of risks like fires, these plants directly affect material flow. Thus, Expert 1 suggested incorporating SCOR resilience aspects and asset-driven aspects in separate tabs on the dashboard. This insight underscores the importance of evaluating SC risks from various angles

Expert 2 delved into risk type classification, proposing a dynamic approach that reflects supplier behavior and commodity criticality. For instance, a consistently "at risk" supplier, producing a critical commodity, holds more significance for manufacturers than an "at risk" supplier dealing with a non-critical material. Expert 2 also highlighted the need to analyze historical risk trends by regions, countries, and the historical performance of suppliers impacted by external factors in more detail.

The feedback converged on the theme of analytical depth. Expert 1 found the initial "Resilience Overview P2P" tab suitable for executives but noted a lack of histograms for quick tracking of historical performance trends. Expert 1 also observed that deep-dive tabs for in-house consultants were overly simplistic and needed more metrics, visualizations, and elements. Moreover, Expert 2 agreed that the initial tab serves executives well, suggesting that with the ability to further drill down into specific metrics and elements, other tabs could provide an ideal level of detail for in-house consultants. Expert 2's proposal to explore specific event investigations addresses the growing demand for deeper insights into the dynamic SC nature. By integrating risk event data with PO information, the dashboard could offer a comprehensive view of SC performance, crucial for effective decision-making.

II - Key Performance Indicators and User Experience

The experts' insights converge on the significance of well-selected KPIs that align with user roles and requirements. Expert 1's endorsement of the SCOR reliability metrics as interesting and effective demonstrates the potential to capture critical supplier performance aspects. Expert 2's emphasis on categorizing order fulfillment cycles and understanding diverse lead times for different materials adds depth to the understanding of SC dynamics. For instance, he emphasized that certain commodities have an order fulfillment cycle of a few days, as opposed to packaging that has an order fulfillment cycle of several weeks. These observations collectively emphasize the

need for more specific KPIs that cater to the diverse needs of different company settings.

The discussion with both experts also highlighted the central role of user experience in the dashboard's design. Overall, both experts suggested that incorporating colors into the dashboard's design is crucial. Expert 1 recommended refining the color and icon representation for risks in the interactive map, addressing it as an essential element of the user experience. On the other hand, Expert 2 introduced the concept of color-coding to visually indicate different risk levels, a technique that could improve the dashboard's usability and provide clear insights to the user. These recommendations highlight the potential of innovative visual representation techniques in effectively communicating complex data.

III - Process Mining

In terms of PM-specific elements, the experts provided insights from different angles. Expert 1 highlighted the logical inclusion of PM elements, sparking a discussion about developing a "Supply Chain Centric Data Model." However, Expert 1 acknowledged that developing such a model would be inherently intricate, demanding substantial effort from the designated team. The crux of the complexity, as pointed out, lies in sourcing the right data for the dashboard and experimenting with diverse services to monitor external risks like environmental disasters and weather data.

Likewise, Expert 2 expressed the feasibility of integrating PM elements, particularly if specific risk events are linked to specific POs. For a comprehensive deep dive, he suggested that the dashboard should incorporate facets commonly seen in regular PM solution dashboards for monitoring the P2P process. Examples include metrics such as on-time delivery, full delivery, and automation rate. If these elements were to be incorporated, they could enhance the dashboard's utility.

IV - Potential Improvements and Data Quality

Both experts provided valuable suggestions for enhancing the dashboard's effectiveness. Expert 1's primary recommendations for potential improvements revolved around incorporating a time perspective. This could involve integrating histograms to illustrate trends over time, implementing clearer scales for metrics, and enhancing visual clarity through the use of color-coded indicators. These enhancements would ensure that users have a comprehensive understanding of the data they are viewing and its significance. Nevertheless, Expert 2 pointed out potential extensions to the analysis that could further enrich the dashboard's insights. One such extension involves taking into account quality risks linked to materials. By considering these additional dimensions, the dashboard could offer a more comprehensive and nuanced view of the SC's performance and potential vulnerabilities.

In addition, Expert 2 strongly emphasized that the overall concept of the dashboard appears promising. However,

he expressed a significant concern regarding data quality, underlining that many of his clients continue to grapple with this issue even today. According to him, the most substantial challenge would involve sourcing and upholding precise, dependable data inputs for the dashboard.

Both experts significantly contributed to the evaluation of the mockup dashboard, providing valuable insights regarding its usability and potential for conceptual enhancement. The feedback received from these experts has illuminated a spectrum of opportunities as well as areas necessitating improvement for the SCR mockup dashboard. Their observations centered around several prominent themes, including risk assessment, analytical depth, user experience, and resilience-related metrics.

DISCUSSION

With the groundwork established in the preceding chapters, the focus now turns to the core of this study: the investigation into the resilience of the SC through the innovative lens of a conceptual SCR dashboard. This discussion chapter engages with the research questions that have guided the exploration. Firstly, the focus is the demonstration of the intended functionality of the mockup that elucidate the synergy between the developed concept and the first research question at hand: *"How can insights into the resilience of the supply chain through the P2P process be obtained using an SCR dashboard conception, considering a specific use case?"*. The walkthrough vividly illustrated the dashboard's ability to empower an in-house consultant to delve into the intricacies of SCR. While it is important to acknowledge that alternate use cases could have been explored, the example chosen for this research encapsulates the envisioned objectives the mockup dashboard aims to achieve. Specifically, it enables in-depth exploration and provides dual layers of analysis for assessing resilience. The "Resilience Overview P2P" section caters to the executive level, offering a consolidated perspective, while also allowing for further investigation for mid-level management by furnishing more detailed analyses across three distinct dimensions of resilience: reliability, responsiveness, and agility. Nevertheless, the conceptual dashboard included filtering options that would allow for drilling down and focusing on specific areas of interest, as well as various metric and elements that are of crucial importance for the SC.

In the assessment of the second research question *"Which opportunities and limitations does the dashboard offer for resilience management of the P2P process from a tactical or strategic point of view?"*, key insights emerged from the expert evaluation. These insights shed light on the potential advantages and constraints associated with the dashboard's effectiveness for resilience management. Notably, the experts acknowledged several positive aspects of the presented dashboard concept. The experts found the proposed dashboard concept intriguing, with

one of them remarking that it introduced a fresh perspective on SCRM. Expert 2 observed that the initial tab included in the mockup provided an optimal level of analysis for executives. Conversely, the more detailed tabs, such as the deep dive sections, were deemed valuable for managers operating at the tactical level within organizations. In addition, Expert 2 underscored the concept's potential for further development, suggesting the incorporation of additional elements such as a quality perspective and diverse drill-down categorizations of risks and the entities they impact. Turning to the limitations, both experts stressed the significance of color usage within the dashboard, highlighting it as a notable drawback. Few (2006, p. 62) highlighted the importance of colors, warning the overusing colors can cause confusion, but that also the misusing colors can lead users to wrong conclusions when reading information from a dashboard. The mockup dashboard's absence of vibrant colors resulted in a lack of emphasis on crucial metrics or data, that would make it challenging for users to identify potential disruptions within the SC. Nevertheless, it is more effective to display alerts using color-coded indicators that signify various levels of threats. These threats have predefined performance target thresholds and so-called zones e.g. a green zone representing the target area, a yellow zone indicating a warning level, and a red zone signifying a critical area of concern (Karami & Safdari, 2016).

Moreover, a significant constraint identified was the relatively limited availability of visualizations within the mockup, which hindered the capacity to compare current data with historical trends. Such a comparative analysis should be deemed essential for both tactical and strategic decision-making perspectives. However, an overflow of information can result not only in ignoring details but also in making inaccurate decisions (Yigitbasioglu & Velcu, 2012). Iselin (1988) discovered a relationship where the accuracy of decision-making and the amount of provided information follow an inverted U-shape. This relationship suggests that there exists an optimal level of accounting information within reports that leads to the most precise decisions. However, the experts' criticism suggests a deficiency in visualizations, particularly in time-related graphs. This highlights the potential need for increased visualization within SCRM, considering its inclusion of various perspectives and entities within a SC. Therefore, colors and color coding, as well as the time-related visualizations for tracking risks and entities across time were deemed important considering the complex data representation that is required to have an overview on the SCR.

While the DSR principles influenced the design and development of the mockup dashboard, it's essential to recognize the iterative nature of this methodology. The paper didn't explicitly delve into these principles, but it's crucial to note that DSR inherently involves multiple refinement cycles before reaching a "final product." Ideally, the feedback gathered from experts could have been

utilized to improve and enhance the conceptual dashboard's design. Nevertheless, the input gathered from the expert evaluation not only contributes to forthcoming refinements of the dashboard but also provides a clearer path for design focus points and future development.

Considering a broader study, like a survey, might clarify which elements are crucial for inclusion in an SCRM mockup dashboard. This approach could distinguish between essential components, potential value additions, and those that might not be necessary. Additionally, previous research highlighted a comprehensive set of criteria applicable across various dashboard types. These criteria, covering aspects such as user customization, knowledge discovery, security, information delivery, alerting, visual design, and integration with system connectivity, offer a valuable framework for assessing the effectiveness of such dashboards (Karami et al., 2017).

The evaluation of the mockup dashboard, however, raised additional questions regarding the practical development process and the challenges associated with integrating the necessary data to operationalize the dashboard. Of particular interest was Expert 1's suggestion to create a "Supply Chain Centric Data Model," a concept not covered within this paper but holding potential for future exploration. Delving into this concept could unlock the true potential of PM in driving process optimization within companies, even for elements outside their immediate control. This could empower companies to take both reactive and proactive actions when addressing external influences on their SCs. Moreover, Expert 2 highlighted the persistent data-related issues as a fundamental challenge that would likely arise in bringing such a concept to fruition. This challenge, as well could serve as a valuable avenue for future research.

In addition, a thorough examination of the SCOR model and its specific breakdown of resilience facets in SCs opens the door to carefully selecting metrics when tailoring solutions for individual enterprises. Given that the SCOR model serves as a versatile framework applicable across diverse industries and company types, it offers a plethora of available metrics falling under the "Resilience" performance category (ACSM, 2022). Additionally, it accommodates various process types. Practitioners can leverage mockup dashboards as exemplars to streamline data organization, ensuring its alignment with both executive and tactical levels within an organization. Nonetheless, it is crucial to recognize that these discrete sub-dimensions of resilience play a pivotal role in conceptualizing and comprehending SCR within varying contexts and across diverse types of companies. This understanding becomes particularly paramount in the context of inventing solutions aimed at mitigating challenges posed by unforeseen external factors that frequently impact SCs.

The development of the SCRM mockup dashboard stands as a significant milestone, offering a visual representation that provides stakeholders with a sneak peek into the envisioned dashboard's interface, data visualizations, and interaction designs tailored for enhancing and overseeing SCRs. Creating this preview and subjecting it to expert evaluation has revealed a profound understanding of the complex challenges inherent in SCR. While established design principles, color schemes, and visualization techniques have been acknowledged in the development of other types of dashboards, identifying the most crucial elements specifically for SCRM requires further exploration, as pertinent literature on this subject remains scarce. Conducting a broader study focused on gathering insights and data from experts across various SC processes in this field would be imperative for identifying these essential elements and metrics, and creating a checklist that could serve dashboard developers as a guide when deciding how and what they should consider when designing such innovative solutions for SCRM.

Despite its static nature the mockup significantly contributes to theoretical advancement, thus deepening the understanding of how information based on the SCOR model resilience aspects could empower resilient decision-making. Furthermore, the mockup provides a conduit for knowledge dissemination. This paper could serve as a means of communication for the academic community, simultaneously directing future research endeavors and enhancement of the SCRM conceptual dashboard, and potentially developing a functional prototype.

CONCLUSION

The primary goal of this paper was to create an industry-neutral conceptual dashboard for the P2P process. This dashboard was designed based on the event log of PM and intended to cater to in-house consultants at the mid-management level. Furthermore, the paper was crafted with the intention of addressing the research questions that steer the course of the research.

The initial research question, "How can a conceptual SCR dashboard provide insights into supply chain resilience within the P2P process, focusing on a specific use case?" was addressed through the creation and presentation of a mockup dashboard. This mockup effectively illustrated how the dashboard could potentially offer insights into SCR. It showcased the dashboard's intended functionality, highlighting its capability to assist consultants in investigating issues that impact the P2P process, with a focus on various important entities and aspects.

Answering the second research question, "What opportunities and limitations does the dashboard present for enhancing resilience management of the P2P process from tactical and strategic standpoints?" involved soliciting insights and feedback from experts in the fields of SCM and PM. The feedback emphasized the well-considered integration of metrics from the SCOR model and elements from Accenture's SCR application. However, the

experts identified the absence of time-related visualizations and color-coded information as crucial additions that would aid in easier tracking of disruptions and issues within the dashboard. These additions were seen as particularly valuable for mid-level managers and executives.

Nevertheless, the conceptual dashboard proposed in this paper could serve as a steppingstone for further research that could potentially lead to the development of a prototype using the Celonis EMS. In addition, the proposed dashboard could offer a visual foundation for understanding the three distinct resilience dimensions, along with organizing key metrics and elements that are crucial for addressing disruptions that impact the SCR within the context of the P2P process.

REFERENCES

- Accenture. (2023). Supply chain disruption: Supply chain networks of the future must have resilience and sustainability at their heart. <https://www.accenture.com/ae-en/insights/consulting/supply-chain-disruption>
- Aman, S., & Seuring, S. (2021). Analysing developing countries approaches of supply chain resilience to COVID-19. *The International Journal of Logistics Management*, 34(4), 909–934. <https://doi.org/10.1108/ijlm-07-2021-0362>
- ASCM. (2022). *ASCM Supply Chain Operations Reference Model SCOR Digital Standard*. <https://scor.ascm.org/performance/introduction>
- Banton, C. (2023). *Purchase-to-Pay (P2P): Definition, Process, Steps, and Benefits*. Investopedia. <https://www.investopedia.com/terms/p/purchasetopay.asp#:~:text=Purchase%2Dto%2Dpay%20is%20a,saves%20costs%2C%20and%20reduces%20risk>
- Bret, L., Dussud, M., Metral, L., Ladier, A. L., & Trilling, L. (2021). Towards a model assessing supply chain resilience strategies. *Procedia CIRP*, 103, 14–19. <https://doi.org/10.1016/j.procir.2021.10.001>
- Celonis. (n.d.). *Accenture - Supply Chain Resilience*. <https://www.celonis.com/ems/ems-store/supply-chain-resilience/>
- Few, S. (2006). *Information Dashboard Design*. http://books.google.ie/books?id=7-FrPwAACAAJ&dq=Information+Dashboard+Design.&hl=&cd=1&source=gbs_api
- Hevner, Alan & R, Alan & March, Salvatore & T, Salvatore & Park, & Park, Jinsoo & Ram, & Sudha. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*. 28. 75-.
- Iselin, E. R. (1988). The effects of information load and information diversity on decision quality in a structured decision task. *Accounting, Organizations and Society*, 13(2), 147–164. [https://doi.org/10.1016/0361-3682\(88\)90041-4](https://doi.org/10.1016/0361-3682(88)90041-4)
- Karami, M., & Safdari, R. (2016). From Information Management to Information Visualization. *Applied Clinical Informatics*, 07(02), 308–329. <https://doi.org/10.4338/aci-2015-08-ra-0104>
- Karami, M., Langarizadeh, M., & Fatehi, M. (2017). Evaluation of Effective Dashboards: Key Concepts and Criteria. *The Open Medical Informatics Journal*, 11(1), 52–57. <https://doi.org/10.2174/1874431101711010052>
- Melnyk, Steven & Closs, D.J. & Griffis, Stanley & Zobel, Christopher & Macdonald, John. (2014). Understanding supply chain resilience. *Supply Chain Management Review*. 18. 34-41.
- Neubauer, M. (2018). Supply Chain Resilience Support in S-BPM. *Proceedings of the 10th International Conference on Subject-Oriented Business Process Management*. <https://doi.org/10.1145/3178248.3178263>
- Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., Wierenga, B., & Wiesel, T. (2009). Dashboards as a Service. *Journal of Service Research*, 12(2), 175–189. <https://doi.org/10.1177/1094670509344213>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/mis0742-1222240302>
- Phadi, N. P., & Das, S. (2021). The Rise and Fall of the SCOR Model: What After the Pandemic? *Computational Management*, 253–273. https://doi.org/10.1007/978-3-030-72929-5_12
- Rahman, A. A., Adamu, Y. B., & Harun, P. (2017). Review on dashboard application from managerial perspective. *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*. <https://doi.org/10.1109/icriis.2017.8002461>
- Rasmussen, N. H., Bansal, M., & Chen, C. Y. (2009). *Business Dashboards*. John Wiley & Sons. http://books.google.ie/books?id=23A5oPOQ7pYC&printsec=frontcover&dq=978-0-470-41347-0&hl=&cd=1&source=gbs_api
- Reibstein, D., Hoyne, N., & Pauwels, K. (2023). Dashboards: From Performance Art to Decision Support.

NIM Marketing Intelligence Review, 15(1), 60–63.
<https://doi.org/10.2478/nimmir-2023-0009>

Schätter, F., Morelli, F., & Haas, F. (2022). SUPPLY CHAIN RESILIENCE MANAGEMENT USING PROCESS MINING. In 36th International ECMS Conference on Modelling and Simulation, ECMS 2022 (pp. 121-127).

Sellitto, M. A., Pereira, G. M., Borchardt, M., da Silva, R. I., & Viegas, C. V. (2015). A SCOR-based model for supply chain performance measurement: application in the footwear industry. *International Journal of Production Research*, 53(16), 4917–4926.
<https://doi.org/10.1080/00207543.2015.1005251>

Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2020). COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions. *British Journal of Management*, 32(4), 1164–1183.
<https://doi.org/10.1111/1467-8551.12441>

Techapichetvanich, K., & Datta, A. (2005). Interactive Visualization for OLAP. *Computational Science and Its Applications – ICCSA 2005*, 206–214.
https://doi.org/10.1007/11424857_23

Tsakiridi, Anastasia. (2021). Applications of Geographic Information Systems (GIS) in Supply Chain Management: Systematic Literature Review.

van der Aalst, W. (2012, July). Process Mining. *ACM Transactions on Management Information Systems*, 3(2), 1–17. <https://doi.org/10.1145/2229156.2229157>

Van Der Aalst, W. M. P. (2016). *Process Mining*. Springer. http://books.google.ie/books?id=hUEG-DAAAQBAJ&pg=PR6&dq=978-3-662-49850-7&hl=&cd=2&source=gbs_api

van der Aalst, W., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., Bose, J. C., van den Brand, P., Brandtjen, R., Buijs, J., Burattin, A., Carmona, J., Castellanos, M., Claes, J., Cook, J., Costantini, N., Curbera, F., Damiani, E., de Leoni, M., . . . Wynn, M. (2012). Process Mining Manifesto. *Business Process Management Workshops*, 169–194.
https://doi.org/10.1007/978-3-642-28108-2_19

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, 13(1), 41–59.
<https://doi.org/10.1016/j.accinf.2011.08.002>

Automatisierungspotenziale von Büro- und Verwaltungsarbeiten anhand von Large Language Models am Beispiel von ChatGPT

Sina Stumpp

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
E-Mail: stumppsi@hs-pforzheim.de

Frank Morelli

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
E-Mail: frank.morelli@hs-pforzheim.de

ABSTRACT

Dieser Beitrag untersucht das Potenzial von Large Language Models (LLMs) am Beispiel von ChatGPT für die Automatisierung von Büroarbeiten in Beratungsunternehmen. Auf der Basis von Experteninterviews werden zwei Anwendungsfälle identifiziert und Kriterien für die Anwendung des Tools definiert. Es handelt sich dabei um "Erstellung von PowerPoint-Präsentationen" und "Deeskalation beim Verfassen von E-Mails". Die Studie unterstreicht die Notwendigkeit einer robusten KI-Strategie und intensiver Schulungen, um Risiken zu minimieren und Mitarbeiter zu schulen.

SCHLÜSSELWÖRTER –

Large Language Model, ChatGPT, Office Automation, Consulting, Transformer

EINLEITUNG

Natural Language Processing (NLP) als Teilgebiet der Künstlichen Intelligenz (KI) konzentriert sich auf sprachliche Disziplinen. Diese Disziplin hat zum Ziel, natürliche menschliche Sprache zu verstehen und zu generieren, um eine generische Konversation zwischen Mensch und Maschine zu ermöglichen und gleichzeitig sprachliche Herausforderungen wie grammatische Regeln, Ausnahmen, Zweideutigkeiten, Dialekte, Ironie und Sprachwandel zu bewältigen (Ghavami 2020; Moloi/Marwala 2021; Taulli 2022; Wang u. a. 2023).

Der Launch der Webanwendung ChatGPT gab im November 2022 den Startschuss für eine neue Welle disruptiver Sprachtechnologien (Chui u. a. 2016; Haleem u. a. 2022). Dieses Large Language Modell (LLM), mit der Fähigkeit natürliche Konversationen zu führen und permanent Wissen zur Verfügung zu stellen, hat seither Menschen weltweit begeistert. ChatGPT basiert auf dem NLP-Lernprozess: Während der Vorverarbeitung kommen dabei Tokenization, Part of Speech Tagging, Named Entity Recognition, Stemming und Stop Word Identification als NLP-Techniken innerhalb der „input embedding“-Schicht zum Einsatz (Onslow 2023; Ghavami 2020; Kuiler 2022).

Neben textintensiven Sektoren, wie der Bildungs- und Forschungsbranche, beschäftigen sich auch Beratungsunternehmen mit der Technologie (ZBW - Leibniz Information Centre for Economics 2023). Zum einen zeichnet es sich ab, dass die Implementierung von digitalen Technologien zu Produktivitätssteigerung in der internationalen Arbeitswelt führt (Dwivedi u. a. 2023). Zum anderen ist davon auszugehen, dass der Umgang mit KI-Techno-

logien und -Tools wie ChatGPT für BeraterInnen in Zukunft wichtig sein wird. Die Suche nach potenziellen Automatisierungsfällen innerhalb eines Büro- und Verwaltungsrahmens erweist sich im Kontext von LLMs als wesentlicher Bestandteil für eine zukunftsweisende Strategie. Sie hat das Potenzial, Arbeit effizienter zu gestalten. Die praktische Implementierung solcher Tools stellt Unternehmen jedoch vor Herausforderungen, da bisher die Einordnung und Konkretisierung von benutzerfreundlichen Einsatzszenarien fehlen: Zum gegenwärtigen Zeitpunkt besteht kein vollumfängliches Verständnis für die realistische Leistungsfähigkeit der Technologie. Die Integration von LLMs in Unternehmensumgebungen und Verwaltungsprozesse wirft zudem noch unbeantwortete Fragen der Datensicherheit, Ethik und Verantwortlichkeit auf. Diese sind aber entscheidend für eine zugehörige Implementierung in die Unternehmenslandschaft.

Ziel der Arbeit ist es, die Identifikation von geeigneten Anwendungsfällen der Technologie ChatGPT, basierend auf dem kostenlos zugänglichen Modell GPT 3.5, für den regelmäßigen Einsatz innerhalb von Beratungsunternehmen. Dabei sollen Potenziale zur Einsparung von Arbeitszeit und Ressourcen identifiziert sowie passende Anleitungen zur Automatisierung der Aufgaben mit ChatGPT erstellt werden. Als Rahmenbedingung ist die aktuelle Gesetzgebung (z.B. KI-Verordnung und DSGVO) zu berücksichtigen.

Die gewonnenen Erkenntnisse basieren zum einen auf einer qualitativen Datenerhebung mit Hilfe von fünf Experteninterviews. Es wurden sowohl Personen aus leitenden Positionen im Bereich KI-Consulting und klassischer Unternehmensberatung befragt, sowie Personen, die durch ihren Beruf als Programmierer mit ChatGPT konfrontiert waren. Zum anderen erfolgt die Analyse zweier Use Cases im Büro- und Verwaltungskontext, die sich im Rahmen der Experteninterviews als relevante Anwendungsfälle herauskristallisiert haben. Im einzelnen handelt es sich dabei um „PowerPoint Präsentation erstellen“

und „Erstellung einer E-Mail zur Deeskalation in der Kundenkommunikation“.

Large Language Models (LLMs)

Aufgrund ihrer Leistungsfähigkeit haben sich Large Language Models (LLMs), zu denen auch ChatGPT gehört, als neue Subkategorie von Sprachmodelle etabliert. Ein essenzieller Indikator für den Wissensumfang eines Sprachmodells ist die Parameteranzahl, ein Synonym für die Anzahl an Gewichten, die bestimmen wie das Netzwerk auf Eingaben reagiert und somit auch die Ausgaben angibt. Als „large“ gilt ein LLM, wenn es eine hohe Anzahl an Parametern besitzt und das Training mit Big Data, in Form von Textdaten, durchgeführt wurde. Studien belegen, dass eine wachsende Größe des LLMs die Leistung des Modells in erster Linie positiv beeinflusst, sodass derzeit der Trend dahin geht, immer größere LLMs zu entwickeln (Bender 2021; Goetze/Abramson 2021). Grundsätzlich wird zum einen darauf abgezielt, korrekte natürliche Sprache zu erzeugen, die von Menschen nicht als maschinell generiert einzuordnen ist. Zum anderen geht es darum, dass Modelle selbst ein Sprachverständnis generieren. Darauf basierend können LLMs komplexe Textaufgaben wie Fragenbeantwortung, Klassifizierung etc. lösen (Zhao u. a. 2023). Mithilfe der Berechnung von Wahrscheinlichkeiten für Wörter und deren Reihenfolge werden dafür autoregressiv Wortsequenzen erzeugt (Bender u. a. 2021; Ghavami 2020).

Die Eingabe erfolgt über Prompts, d.h. schriftliche Befehle der Nutzer, die tokenisiert in die Encoderschicht übermittelt werden. LLMs sollen dabei angepasst an Nutzererwartungen handeln. Dafür werden meist entsprechende Beispiele in den Trainingsdatensatz integriert, damit das Modell anhand derer sprachliche Konzepte hinzulernen kann. In diesem Zusammenhang müssen Wahrheit und Unvoreingenommenheit der Ausgaben gesichert sein sowie Nutzungen mit falschen, illegalen oder gefährlichen Intentionen unterbunden werden (Ouyang u. a. 2022). Mittlerweile haben einige LLMs die Fähigkeit sich einen „moralischen Kompass“ anzueignen, indem sie schädliche Konzepte wie bspw. Diskriminierung oder Stereotypen als Teil ihres Trainings erkennen oder speziell von Entwicklern dazu instruiert werden. Anschließend ist es ihnen möglich, sich selbst und andere moralisch zu korrigieren oder Befehle abzuweisen (Ganguli u. a. 2023).

Neben der Möglichkeit zur Spezialisierung auf bestimmte Anwendungsbereiche, besitzen LLMs die Kompetenz zur Generalisierung. Dies ermöglicht die Bearbeitung von unbekanntem Befehlen bzw. beschreibt die Fähigkeit, aus dem Training gezogenes Wissen auf neue Situationen anzuwenden und kontextspezifische Antworten liefern können (Ferrara 2023).

Durch intensives Training mit Textdokumenten erlangen LLMs ein tiefgehendes Verständnis für menschliche Sprache, ermöglichen die Generierung von natürlichen Texten und können zudem multimodale Daten wie Bilder und Videos verarbeiten (Lund u. a. 2023; Yin u. a. 2023). Die Anwendungsbereiche sind vielseitig und umfassen

Textgenerierung, Übersetzung, Klassifizierung, Informationsextraktion, Fragenbeantwortung und Sentiment-Analyse. LLMs bieten auch Formulierungshilfen, unterstützen kreative Inhalte, verstehen Emotionen in Texten, können auf wenigen Beispielen basierend generalisieren (Chohan 2023; Chui u. a. 2022; Zhang u. a. 2023; Zhao u. a. 2023). Ihr Verständnis für Kontext und Zusammenhänge ermöglicht situationsabhängige Reaktionen (Dwivedi u. a. 2023; Haleem u. a. 2022).

Sind LLMs auf Mensch-Computer-Dialoge ausgerichtet sind, ermöglicht dies natürliche Konversationen (Bozkurt u. a. 2023; Haleem u. a. 2022). Zudem agieren sie dabei als effektive Assistenten oder Co-Autoren in Zusammenarbeit mit Menschen, ohne diese zu ersetzen (Polonsky/Rotman 2023). Darüber hinaus erleichtern LLMs die Automatisierung von wiederholenden, zeitaufwändigen Textaufgaben. Auf Basis von maßgeschneiderten Trainingsdaten können LLMs flexibel an individuelle Szenarien angepasst werden, was eine vielseitige Anwendung ermöglicht (Ferrara, 2023). LLMs bewältigen mühelos große Mengen an unstrukturierten Daten, d.h. sie benötigen keine vorherige Datenstrukturierung und können Informationen aus weniger organisierten Quellen verarbeiten (Bender u. a. 2021; Haleem u. a. 2022).

Zudem lassen sie sich nahtlos in bestehende Systeme und Workflows integrieren, um zusätzliche Funktionalitäten und Verbesserungen bereitzustellen (OpenAI 2023a). Durch die Implementierung von LLMs in Geschäftsprozesse können weiterführend Routineaufgaben automatisiert werden und Mitarbeiter besitzen die Möglichkeit sich auf anspruchsvollere Aufgaben zu fokussieren (ZBW – Leibniz Information Centre for Economics 2023). Somit optimieren sie bei einer effektiven Nutzung Prozesse, sparen Zeit, steigern die Unternehmensproduktivität und tragen zur Kosteneffizienz bei, sodass Unternehmen sich dementsprechend einen Wettbewerbsvorteil sichern und das Wirtschaftswachstum fördern können (Dwivedi u. a. 2023; Goldstein u. a. 2023; Lund u. a. 2023).

In der Betrachtung der Risiken von LLMs offenbart sich eine Vielzahl von Herausforderungen, die sowohl ethische als auch praktische Implikationen mit sich bringen. Der undurchsichtige "Black Box"-Aufbau von LLMs führt zu einer mangelnden Nachvollziehbarkeit ihrer Ausgaben, da der komplexe interne Mechanismus nicht immer logisch ersichtlich ist (Aiyappa u. a. 2023). Diese Intransparenz birgt das Risiko von Falschinformationen, die durch Halluzinationen seitens der Modelle erzeugt werden (Wang u. a. 2023). Datenschutz wird ebenfalls gefährdet, da Eingabedaten in Trainingsdaten integriert werden, was nicht nur unerwünschte Datenschutzrisiken schafft, sondern auch LLMs zu potenziellen Zielen für Cyberangriffe macht (Kuiler 2022). Zusätzlich besteht die Gefahr, dass sich trotz Anonymisierung von Eingabedaten Personen oder Unternehmen durch Kontexthinweise identifizieren lassen (Patsakis/Lykousas 2023).

Die Fehlbarkeit des Trainings durch menschlichen Einfluss sowie die Unzureichendheit von repräsentativen

Trainingsdaten verstärken Verzerrungen in den generierten Ausgaben und verursachen eine erhöhte Wahrscheinlichkeit für toxische und von Stereotypen durchzogene Inhalte (Bender u.a. 2021; Goldstein u.a. 2023; Ouyang u. a. 2022).

Diese Risiken erstrecken sich über verschiedene Bereiche, von der individuellen Nutzung bis hin zu potenziellen Auswirkungen auf Arbeitsmärkte und die Integrität wissenschaftlicher Forschung (Chohan 2023; Polonsky/Rotman 2023). Eine weitere Herausforderung besteht darin, dass die Anwendung von LLMs eine fundierte Kenntnis der korrekten Formulierung von Prompts erfordert, um effektiv genutzt werden zu können (Goldstein u. a. 2023). Insgesamt verdeutlichen die genannten Risiken die Notwendigkeit eines umfassenden Verständnisses und einer kritischen Reflexion im Umgang mit LLMs in diversen Anwendungsbereichen, um einem unberechtigt hohen Vertrauen in die KI vorzubeugen (Kasneci u. a. 2023).

ChatGPT

ChatGPT ist ein von OpenAI entwickeltes LLM, das seit seiner Veröffentlichung im November 2022 über die Grenzen der Technologiebranche hinaus große Wellen geschlagen und aufgrund seiner disruptiven Art einen neuen Standard in linguistischer Intelligenz gesetzt hat (Lund u. a. 2023; Wang u. a. 2023). Mit über 100 Millionen Nutzern weltweit innerhalb der ersten zwei Monate gilt das, in der Basisversion kostenlos zugängliche, LLM als erstes Sprachmodell mit Massenanzug (Aiyappa u. a. 2023). Ein Grund dafür ist die Entwicklung des Modells als Chatbot, was die Zugänglichkeit für technisch-unaffine Menschen erleichtert (Dwivedi u. a. 2023).

Über ein Texteingabefeld können Nutzer Prompts formulieren und im Anschluss beantwortet ChatGPT die Befehle durch das Generieren von Text. In die Kategorie der Chatbots fällt ChatGPT durch die Fähigkeit, Bezug auf vorherige Prompts und Antworten zu nehmen, und somit einen menschlichen Dialog mit dem Nutzer zu führen (Lund u. a. 2023). ChatGPT kann Rückfragen stellen, unangemessene Anfragen verweigern, Fehler eingestehen sowie korrigieren, sodass es sehr intuitiv wirkt und den Anschein erwecken kann, es würde eine Art Gewissen besitzen (Ganguli u. a. 2023).

Zur Entwicklung von GPT-3 wurde das transformerbasierte neuronale Netzwerk zunächst mit einem sehr großen Datensatz, bestehend aus einer Vielzahl von ungekennzeichneten Textkorpora, trainiert, sodass dieses ein umfassendes Verständnis von Sprache entwickelt (Kasneci u.a. 2023; Narasimhan u.a. 2018). In diesem Schritt lernt ein Modell, wie Sprache gebildet wird, wie Kontext Aussagen beeinflusst und welche Bedeutung hinter Wörtern und Kombinationen dieser stecken, indem es Muster und Beziehungen in den Trainingsdaten enthaltenen Texten erkennt und seine Parameter entsprechend anpasst (Ferrara 2023; Zhao u.a. 2023). Beim Training von ChatGPT wurde dafür eine gefilterte Version von Common-Crawl, einer Bündelung von Websites und Dokumenten, die über die letzten acht Jahre durch Webcrawling ge-

sammelt wurden, der Datensatz WebText, der Redditeinträge, inklusive angegebener Quellen, beinhaltet, sowie zwei Buchkorpora und alle englischen Wikipediaeinträge verwendet (Patsakis/Lykousas 2023). Auch wenn ChatGPT in über 50 Sprachen, wie bspw. Englisch, Spanisch, Französisch, Deutsch, Arabisch und Japanisch, nutzbar ist, bestanden mindestens 96 % der Trainingsdaten aus englischen Daten (Ouyang u. a. 2022). Um wie viele Dokumente und Wörter es sich genau handelt, wurde von OpenAI nicht veröffentlicht.

(1) *Supervised Fine-Tuning:*

Beim Vortraining wird ein Grundverständnis von Sprache vermittelt. Hierdurch besitzt das Sprachmodell zwar breitgefächerte Kompetenzen. Diese reichen jedoch i.d.R. nicht aus, um qualitative hochwertige Antworten in ausgewählten Anwendungsbereichen hervorzubringen. Entsprechend gilt es, diese mit mit Supervised Finetuning zu ermöglichen (Ouyang u. a. 2022). Hierbei wird das Sprachmodell mit einem kleineren, gelabelten und speziell auf das gewünschte Nutzungsverhalten ausgelegten Datensatz gefüttert (Wang u. a. 2023). Da ChatGPT in erster Linie einen einfachen Kommunikationsaustausch ermöglichen soll, wird das Modell in diesem Schritt für zwischenmenschliche Konversationen im Chatumfeld zugeschnitten. Eine ausgewählte Gruppe an Menschen, die als „Labeler“ (Kennzeichner) agieren, verfasst dafür eine Reihe an Prompts sowie die, nach ihrem Ermessen, gewünschten Ziel-Ausgaben auf diese, die das Modell, nach Verarbeitung der Daten, als Handlungsorientierung einsetzt (Ouyang u. a. 2022). Supervised Finetuning gilt als beliebter Ansatz, um den Aufwand, Daten manuell zu kennzeichnen, zu minimieren. Neben Domänenwissen lassen sich mit Hilfe von Finetuning auch bestimmte Richtlinien oder Einschränkungen implementieren, um klassische Risiken wie Bias oder toxische Ausgaben zu vermeiden und die Ausgaben an menschliche Wertevorstellungen anzupassen (Goldstein u. a. 2023; Ouyang u. a. 2022). Anschließend ist es dem Sprachmodell möglich sich selbst moralisch zu korrigieren.

(2) *Reinforcement Learning from Human Feedback:*

Das Sprachmodell ist zu diesem Zeitpunkt nicht dazu fähig, die Qualität von Antworten einzuschätzen, d.h. zu wissen, welche Antwort auf einen Prompt die beste ist, wenn es mehrere davon produziert. Deshalb findet die Anwendung eines Reward Models (Belohnungsmodell) Gebrauch (Gao u. a. 2023; Ouyang u. a. 2022). Zur Erstellung des Reward Models werden dem Sprachmodell verschiedene Prompts mehrmals gestellt, womit am Ende zu jedem Prompt eine Auswahl an Antwortmöglichkeiten existiert. Danach wird diese Sammlung Labelern vorgelegt, die die Ausgaben von „am besten“ zu „am schlechtesten“ ranken (Ziegler u. a. 2019). Wenn genug Rankingbeispiele gesammelt worden sind, erfolgt auf Basis der Feedbackdaten ein separates Reward Model Training, das vorhersagt, welche Ausgabe ein Mensch für eine gegebene Eingabe bevorzugt, indem es für jede Antwortmöglichkeit eine „Skalarbelohnung“ ausgibt (Ferrara 2023; Ouyang u. a. 2022). Anhand dessen wird eine Policy (Strategie) erlernt, die mit dem Proximal Policy

Optimization (PPO) Algorithmus, einer Gruppe von Policy Gradient Methoden von OpenAI, Skalarbelohnungen mittels Gradient Descent maximiert (Ouyang u. a. 2022; Schulman u. a. 2017). Das Sprachmodell gibt nach Implementierung der Policy bei Promptanfragen auf Nutzervorlieben zugeschnittene Texte aus und beendet den Trainingsvorgang, sodass ChatGPT von dort an einsatzbereit ist. Seit dem Launch von ChatGPT wurden mehrere geupdatete Versionen veröffentlicht, die laut OpenAI noch besser an die Nutzerwünsche angepasst wurden. Grund dafür ist vor allem der Einsatz von vergangenen Nutzeranfragen und Nutzerbewertungen zu den generierten Ausgaben als neue Trainingsdaten. ChatGPT wird somit aktiv weiterentwickelt und gewinnt mit mehr Nutzereingabedaten der auch an mehr Sprachverständnis (Aiyappa u. a. 2023).

Promptdesign

Die Formulierung der Prompts, innerhalb der Use Cases, ist an den gängigsten Promptdesign-Techniken sowie OpenAIs „GPT Best Practices“ orientiert (OpenAI 2023b). Grundsätzlich gibt es drei Ansätze für das Design von Prompts:

- **Zero-Shot:** Enthält ein Prompt eine Aufgabenstellung ohne passende Lösungsbeispiele, spricht man von einem zero-shot Prompt. Die Beantwortung des Befehls beruht stattdessen auf der Generalisierung von Trainingsdaten, indem mit allgemeinem Wissen und mit Hilfe von erkannten Mustern eine möglichst angemessene Lösung generiert wird.
- **One-Shot:** Liefert man dem Modell nur ein einziges Beispiel, in Form von einer Eingabe mit dazugehöriger Ausgabe, so kann ChatGPT, basierend darauf, eine eigene Ausgabe formulieren. Dafür lernt es aus diesem einzelnen Beispiel, um anschließend ähnliche Aufgaben ausführen können, ohne weitere Trainingsdaten zu benötigen.
- **Few-Shot:** Im Gegensatz zu einem one-shot Prompt, umfasst ein Few-Shot Prompt normalerweise mehrere Beispiele, mit denen das Modell die Aufgabe bewältigen soll.

Es kann hilfreich sein, ChatGPT mit one- oder few-shot Prompts in eine gewünschte Richtung zu lenken, wenn das Modell Schwierigkeiten hat einen Befehl zu verstehen (Brown u. a. 2020).

USE CASES

Die Auswahl der Use Cases ist auf Basis von durch Experteninterviews ermittelten Anwendungsbeispielen erfolgt, ersichtlich in Tabelle 1.

Insgesamt wurden sechs Kategorien mit 29 Anwendungsbeispielen identifiziert. Aufgrund ihrer täglichen Relevanz in der Consultingbranche wurden davon die Use Cases „PowerPoint Präsentation erstellen (17)“ und „Deeskalierende E-Mail-Kommunikation (22)“ für eine detaillierte Umsetzungsanalyse ausgewählt.

Tabelle 1: Durch Experteninterviews identifizierte Use Cases

Nebenkategorie	Nr	Use Case	
Programmierung	1	Codegeneration	
	Wissensmanagement	2	Wissensrecherche
		3	Marktanalyse
		4	Extrahierung von Informationen, Zusammenfassungen/ Executive Summaries
		5	Strukturierung von Inhalten
		6	Extrahierung von Informationen & Hinzufügen von Informationen in Exceltabellen (über API)
Schreibassistenz	7	Formulierungshilfe	
	8	Paraphrasieren	
	9	E-Mail (-Kommunikation)	
	10	Dokumentenerstellung (z.B. Proposals)	
	11	Websitextgeneration, Storeintraggeneration für Apps	
	12	Übersetzung	
	13	Kreative Texte Schreiben & Inspiration	
	14	Reden schreiben	
	15	Redetext für Pitches & Präsentationen schreiben	
	16	Corporate Language (vorheriges Finetuning notwendig)	
Außenwirkung	17	Powerpoints erstellen	
	18	Kommunikations- Coaching/ Training mit ChatGPT	
Beratende & prüfende Tätigkeiten	19	Risikoerkennung	
	20	Prüfung von Dokumenten	
	21	Bewertungssystem	
	22	Beschwerde Management/ Sentiment Analyse in (E-Mail) -kommunikation bspw. Zur Deeskalation	
Freizeitgestaltung	23	Urlaub planen	
	24	Rezepte	
	25	Ernährungsplan	
	26	Workoutplan	
	27	Anlageberater (über API)	

„PowerPoint Präsentation erstellen“

In Consultingfirmen ist die Erstellung von Präsentation täglicher Bestandteil der Arbeit. Bis zur fertigen Präsentation müssen mehrere Schritte durchlaufen werden, die bei manueller Durchführung die folgenden Punkte umfassen:

1. Wissensrecherche: Sammlung von passenden Informationen über das Internet, Studien, Bücher o.ä.
2. Strukturierung der Informationen: Zusammenstellung und Sortierung der Informationen für ein Präsentationsformat. Dazu gehört die Bestimmung einer Präsentationsgliederung, die Formulierung von Titeln und Texten und einer Storyline mit passendem Fazit.
3. Präsentationserstellung: Umsetzung der Struktur in PowerPoint durch manuelles Hinzufügen von Folien, Titeln, Texten, Bildern, Grafiken o.ä.
4. Überarbeitung und Verbesserung: Anpassung von Inhalten und Design, aufgrund von Abänderungswünschen oder Unternehmensrichtlinien.

Da jeder Schritt einen Zeitaufwand von mehreren Stunden bedeuten kann, würde die Möglichkeit, diese Arbeit zu automatisieren, eine erhebliche Zeiteinsparung ermöglichen. Mit dem Use Case soll überprüft werden, ob ChatGPT als Automatisierungstechnologie in Frage kommt. Eine Visualisierung des Automatisierungsprozesses zeigt Abb. 1.

Anforderungsanalyse

Tabelle 2: Anforderungsanalyse Use Case 1

Ziel	Der Nutzer erhält eine qualitativ hochwertige Präsentation mit einer faktisch korrekten inhaltlichen Übersicht zu Generative AI in Deutschland, die für Menschen ohne Vorkenntnisse zum Thema verständlich und nachvollziehbar ist.
Umfang	<ul style="list-style-type: none"> - Neu erstellte PowerPoint - 5 Folien (1 Deckblatt & 4 Inhaltsfolien) - Titel & Folieninhalte
Zustand bei Erfolg	Es existiert eine neu erstellte PowerPoint Präsentation, die mit einer logischen Struktur eine Übersicht über das Thema Generative AI in Deutschland bietet. Die Präsentation hat leicht verständliche Textinhalte und ein ansprechendes Design.
Zustand bei Scheitern	<ul style="list-style-type: none"> - Stufe 1: es existiert keine Präsentation - Stufe 2: die PPT hat keine logische Struktur - Stufe 3: textliche Inhalte sind faktisch inkorrekt oder nicht verständlich - Stufe 4: die PPT hat kein ansprechendes Design
Untersuchtes System	ChatGPT
Primär-Akteure	Nutzer, ChatGPT
Sekundär-Akteure	PowerPoint
Erfolgsszenario	Ersichtlich in Abbildung 1
Erweiterungen	Drei Varianten
Potentielle direkte Probleme	<ul style="list-style-type: none"> - ChatGPT ist nicht verfügbar, da Server ausgelastet sind o.ä. - ChatGPT braucht mehrere Prompts, um einen einzelnen Schritt zufriedenstellend auszuführen - ChatGPT führt einzelne Schritte unabh.ngig von der Promptanzahl nicht richtig aus - Makro ist fehlerhaft oder lässt sich nicht ausführen

Stakeholder

Tabelle 3: Stakeholder Use Case 1

Primärstakeholder	Kunden, Beratungsunternehmen, Berater/ Angestellter, Teammitglieder
Sekundärstakeholder	Endnutzer, Konkurrenz

User Story

Laura, Mitarbeiterin bei "TechPro Consulting", nutzt ChatGPT-Makros, um kurzfristig eine umfangreiche PowerPoint-Präsentation über Generative AI für ein wichtiges Kundengespräch zu erstellen. Nach Abklärung der inhaltlichen Punkte erstellt ChatGPT automatisch eine vorläufige Präsentation als Makro. Laura führt das Makro im Visual Basic Editor aus, passt die Präsentation an die Markenrichtlinien an und ergänzt diese mit ihrer Expertise. Die Nutzung von ChatGPT-Makros steigert ihre Effizienz und reduziert i.d.R. Stress, sodass sie zuversichtlich für das Kundengespräch ist und weitere Projekte erfolgreich bewältigen kann.

Rahmenbedingungen

Um die Einhaltung der Unternehmens-Policy zu gewährleisten, darf der Unternehmensname nicht als Kontextorientierung in den Prompts genannt werden. Dementsprechend ist eine gebrandete Präsentation im Corporate Design mit ChatGPT 3.5 nicht umsetzbar, sondern muss manuell für diesen Zweck angepasst werden.

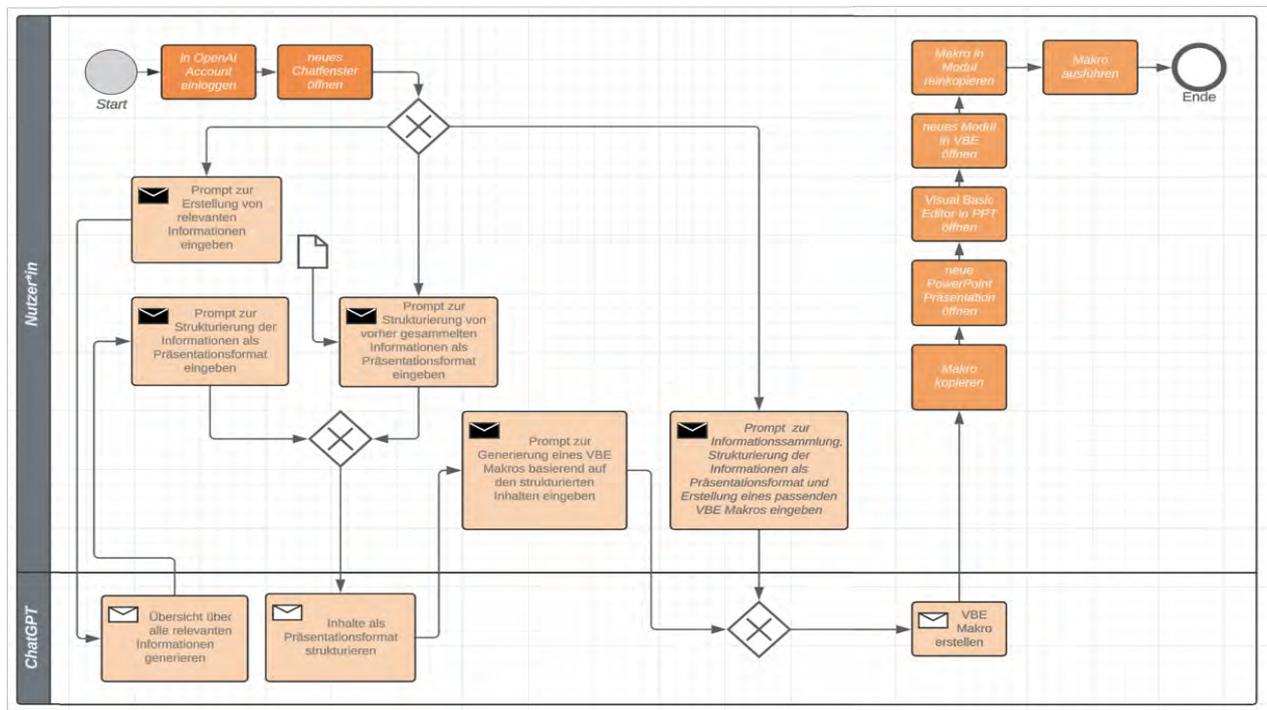


Abb. 1: Prozessablauf (Quelle: Eigene Darstellung)



Abb. 2: Teilergebnis Variante 1 zum Thema Generative KI (Quelle: Eigene Darstellung)



Abb. 3: Teilergebnis Variante 2 zum Thema Generative KI (Quelle: Eigene Darstellung)



Abb. 4: Teilergebnis Variante 3 zum Thema Generative KI in der Automobilbranche (Quelle: Eigene Darstellung)

Erkenntnisse

Haben Nutzer noch keine Erfahrung mit dem Einsatz von ChatGPT oder in der Erstellung von VBA-Makros, dauert die komplette Durchführung des Use Cases deutlich länger als bei erfahrenen Nutzern. Zudem kann Unwissen über die korrekte Formulierung von Prompts zu schlechteren Ergebnissen führen. Des Weiteren besteht die Möglichkeit, dass die Use Case Voraussetzungen noch nicht erfüllt sind und sich Nutzer beispielsweise über die Unternehmens-Policy informieren oder einen neuen Account erstellen müssen, was die Nutzung von ChatGPT aufwändiger als die manuelle Durchführung gestaltet. Grundlegende Kenntnisse sind daher erforderlich, damit der Use Case mit Hilfe von ChatGPT effizient ausgeführt wird. Zudem ist zu beachten, dass während der Durchführung der Use Cases mehrmals von ChatGPT generierte VBA-Makros bei Ausführung Fehler verursacht haben und so nicht funktioniert haben. Erst nach diversen Neugenerierungen konnten bestimmte Makros eingesetzt werden. Es hat sich jedoch als zeitsparender erwiesen, den Prompt neu zu stellen als nach einzelnen Fehlern zu fragen. Als technisch-unaffine Person können Codefehler nichtsdestotrotz verunsichernd wirken und zur Ablehnung von ChatGPT in diesem Einsatzszenario führen. Weiterhin ist zu erwähnen, dass ChatGPT 3.5 ein rein textbasiertes Tool ist und auch VBA-Makros nicht auf

die Erstellung von visuellen Inhalten ausgelegt ist. Zwar besteht die Möglichkeit Bilder oder Grafiken per Makro einzufügen, dafür muss jedoch der passende Dateipfad zum Speicherort auf dem Desktop im Makro hinterlegt werden. Dies würde eine manuelle Speicherung von Bildern und Grafiken bedeuten und zusätzliche Zeit kosten. Die eingebaute Designerfunktion von PowerPoint, die sich als erweiterndes Tool zur visuellen Gestaltung nutzen lässt, kann ggf. im Unternehmenskontext aus Sicherheitsgründen gesperrt sein. Außerdem ist darauf zu achten, dass, aufgrund von dem begrenzten Kontextfelds ChatGPTs, sich keine großen Foliendecks kreieren lassen. Stattdessen bricht der VBA-Code frühzeitig ab und weist darauf hin die restlichen Folien in gleicher Weise manuell hinzuzufügen.

Neben den aus der einschlägigen Literatur abgeleiteten Risiken finden auch von Experten angesprochene Risiken Anwendung: Den Ausgaben von ChatGPT ist zu entnehmen, dass viele Formulierungen ähnlich klingen und persönliche Nuancen in Texten deutlich weniger werden. Hierdurch entsteht die Gefahr sich nicht von Konkurrenten, die ähnlich vorgehen, zu differenzieren. Dies kann zu negativen Auswirkungen auf das Image des Beratungsunternehmens haben und zum anderen bei regelmäßiger Wiederholung zur negativen Beeinflussung von Soft Skills der Mitarbeiter führen, da sie verlernen Texte zu schreiben und strukturieren und potenziell dazu neigen, weniger achtsam in der Umsetzung von sprachlichen Aufgaben zu werden. Dies ist vor allem problematisch innerhalb der Wissensrecherche, da ChatGPT 3.5 kein Zugriff auf aktuelle Daten hat. Hinzukommend birgt die Wissensrecherche durch ChatGPT ein erhöhtes Risiko Falschinformationen zu verbreiten. Dies kann für Unternehmensberatungen projektspezifisch und / oder imagebezogene negative Auswirkungen zur Folge haben. Neben den Risiken durch ChatGPT existiert auch das Risiko mit Makros Malware-Infektionen oder unbefugten Zugriff auf sensible Daten zu ermöglichen, da diese Schadcode enthalten können, der beim Öffnen des Dokuments ausgeführt wird. Zurzeit ist dieses Risiko jedoch als gering einzustufen.

Trotz der Risiken sind die Use Case Anforderungen mehrheitlich erfüllt: So konnte eine PowerPoint mit logischer Struktur erstellt werden, die eine grundlegende Übersicht über generative KI mit Bezug zur Automobilbranche bietet (siehe hierzu Abb. 2 bis 4). Die Textinhalte sind überwiegend leicht verständlich und nachvollziehbar.

Als großes Defizit hat sich jedoch das Präsentationsdesign erwiesen. Es ist i.d.R. im Anschluss notwendig Folien manuell anzupassen, zu gestalten und visuelle Elemente hinzuzufügen. Folglich ist ChatGPT für die Vollautomatisierung des Prozesses ungeeignet, als unterstützendes Tool aber durchaus sinnvoll.

Unter Berücksichtigung der Grenzen und Risiken kann die Basis von Präsentation in kurzer Zeit erstellt und dafür mehr Aufwand in die Anpassung an Zielgruppen und ins Storytelling durch Design gesteckt werden. Auch

die von den Experten angegebene Maximalanzahl von sechs aufzuwendenden Prompts wird nicht überschritten. Für den Use Case existieren diverse Möglichkeiten, um an das gleiche Ziel zu kommen. Im vorliegenden Fall sind drei Varianten näher untersucht worden:

- Variante 1 (ein Prompt): In Fall von Zeitmangel liefert Variante 1 am schnellsten ein Ergebnis. Trotz kürzeren Inhalten ist die Wissensessenz vorhanden und kann als Grundlage dienen.
- Variante 2 (zwei Prompts): Möchte der Nutzer Kontrolle über die exakten Folieninhalte und Quellen erhalten, erscheint Variante 2 empfehlenswert. Da der Zeitaufwand für manuelle Recherche hoch ist, muss der Nutzer jedoch genug Zeit einplanen.
- Variante 3 (drei Prompts): Variante 3 ist für alle Nutzer interessant, die längere und genauere Texte erwarten, aber bei denen ChatGPT die Ausführung aller Schritte in Eigenleistung erbringen soll. Dazu ist es für Nutzer, die Einzelschritte nach ihren Vorstellungen personalisieren wollen, die beste Variante im direkten Vergleich.

In allen Fällen ist es essenziell, alle generierten Inhalte auf faktische Wahrheit zu überprüfen und toxische oder von Bias beeinflusste Texte zu entfernen.

„Deeskalierende E-Mail-Kommunikation“

Auch wenn Anwendungen wie Microsoft Teams oder Zoom sich zunehmend an Beliebtheit erfreuen, ist die E-Mail in der Consultingbranche nach wie vor das Hauptkommunikationstool mit Kunden und zwischen Mitarbeitern. Wenn es zu einem externen oder internen Konflikt kommt, erfolgt eine Beschwerde oftmals über dieses Medium. Dies erschwert häufig den Deeskalationsprozess im Vergleich zu persönlichen Gesprächen (Jong/van Eekelen 1999). Eine manuelle Bearbeitung einer Beschwerdemail umfasst verschiedene Schritte:

1. Leseverständnis: Lesen und Verstehen der E-Mail
2. Stimmungseinschätzung: Inhaltliche Analyse und Abschätzung zur Stimmung des Absenders
3. Lösungsfindung: Entwicklung geeigneter Maßnahmen zur Konfliktlösung
4. Kommunikation: Formulierung einer deeskalierenden Antwort, inklusive Lösungsmöglichkeiten

Anforderungsanalyse

Tabelle 4: Anforderungsanalyse Use Case 2

Ziel	Der Nutzer erhält eine professionelle, empathische und deeskalierende E-Mail, die er als Antwort auf die Kunden-E-Mail einsetzen kann.
Umfang	- 1 Antwort-E-Mail - Sentiment Analyse der Kunden-E-Mail
Zustand bei Erfolg	Es existiert eine neu erstellte E-Mail, die einen logischen Aufbau hat und auf die Inhalte der Kunden-E-Mail eingeht. Die generierte E-Mail muss die Bedenken vom Kunden konkret ansprechen und Lösungsansätze aufzeigen. Sie kann ohne große Abänderungen eingesetzt werden.
Zustand bei Scheitern	- Stufe 1: es existiert keine E-Mail - Stufe 2: die E-Mail hat keine logische Struktur - Stufe 3: die E-Mail kann nicht deeskalierend eingesetzt werden
Untersuchtes System	ChatGPT
Primär-Akteure	Nutzer, ChatGPT
Sekundär-Akteure	Kunde
Erfolgsszenario	Ersichtlich in Abbildung 5
Erweiterungen	-
Potentielle direkte Probleme	- ChatGPT ist nicht verfügbar, da Server ausgelastet sind o.ä. - ChatGPT braucht mehrere Prompts, um einen einzelnen Schritt zufriedenstellend auszuführen - ChatGPT führt einzelne Schritte unabhängig von der Promptanzahl nicht richtig aus

Die richtige Formulierung einer E-Mail in einer Konfliktsituation kann essenziell für die weitere Beziehung der Parteien und den Verlauf von Projekten sein. Eine Automatisierung des Prozesses anhand von ChatGPT (vgl. Abb. 2) vermag ggf. eine sachorientierte Perspektive auf die Situation fördern und eine optimierte Antwort, mit Berücksichtigung der Bedürfnisse der Involvierten, generieren.

Stakeholder

Tabelle 5: Stakeholder Use Case 2

Primärstakeholder	Kunden, Beratungsunternehmen, Berater/ Angestellter
Sekundärstakeholder	Andere Mitarbeiter, Andere Kunden, Konkurrenz

User Story

Als Consultant bei einer renommierten Beratungsfirma ist Lennart täglich mit der Kommunikation mit Kunden und Projektbeteiligten konfrontiert. Ein Kunde, Herr Müller, hat eine E-Mail gesendet, in der er emotional gereizt die Unzufriedenheit über das Nichteinhalten einiger Projektvorgaben ausdrückt. Lennart muss schnell und professionell reagieren, um die Situation zu deeskalieren und das Vertrauen des Kunden zu wahren. Er beschließt, das ChatGPT-System zu nutzen, um eine Sentiment-Analyse der E-Mail zu generieren und anschließend eine deeskalierende E-Mail als Antwort auf die Beschwerde von Herrn Müller zu erstellen.

Rahmenbegrenzungen

Um die Einhaltung der Unternehmens Policy zu gewährleisten, darf der Unternehmensname sowie der des betroffenen Kunde nicht als Kontextorientierung in den Prompts genannt werden. Weiterführend sollen keine projektinternen Informationen, sondern lediglich eine allgemeine Umschreibung der Gesamtsituation genutzt

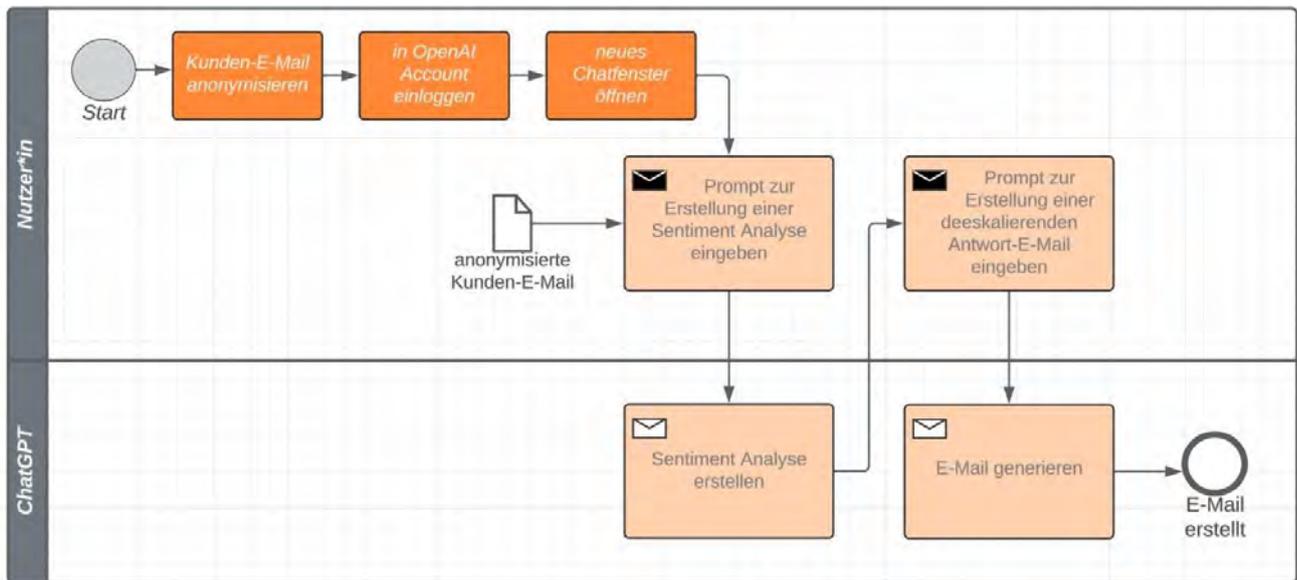


Abb. 5: Prozessablauf (Quelle: Eigene Darstellung)

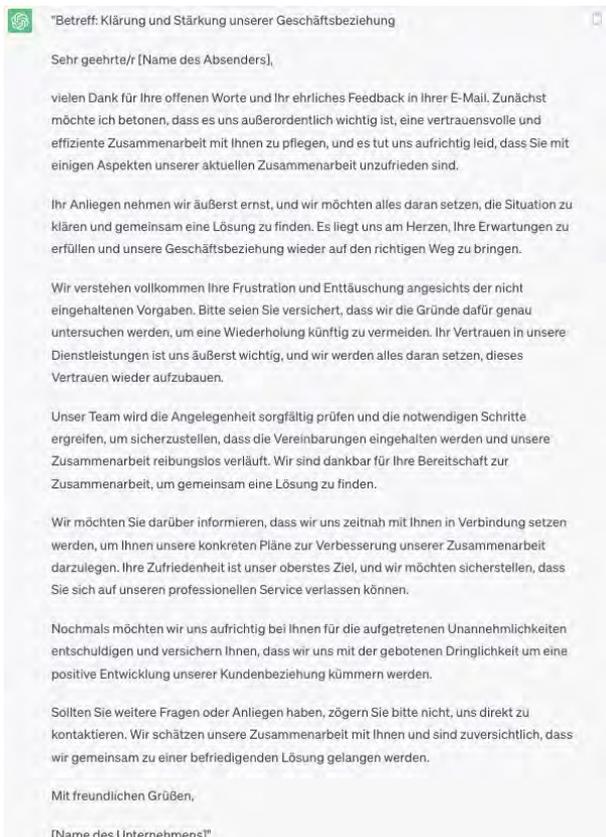


Abb. 6: Ergebnis deeskalierende E-Mail

Erkenntnisse

Zusätzlich zu den Einschränkungen aus dem vorherigen Use Case besteht hierbei die Möglichkeit, dass Nutzer den Vorgang einer Sentiment Analyse nicht kennen und so möglicherweise Schwierigkeiten haben, ihr Anliegen in Worte zu fassen. Des Weiteren ist zu erwähnen, dass Konflikte meist nicht plötzlich auftreten, sondern i.d.R. eine Entstehungsgeschichte haben. Es ist folglich fragwürdig, ob eine einzige E-Mail als Orientierung den

Konflikt vollumfänglich abbilden kann, vor allem wenn Teile der Kommunikation auch in persönlichem, mündlichem Kontakt stattgefunden haben. Ferner ist es möglich, dass ChatGPT aufgrund von Datenschutzrichtlinien keine umfangreiche Analyse durchführen kann. Da bestimmte Konfliktbereiche eine eigene Interpretation erfordern, ist dies für ChatGPT folglich nicht möglich. Zusätzlich ist die Sentiment Analyse auf das Spektrum negativ bis positiv begrenzt. Trotz einer umfangreichen Anzahl von nuancierten Emotionen sind diese nicht angegeben. Die emotionalen Hintergründe des Absenders sind entsprechend vereinfacht dargestellt und bilden eventuell nicht die Wirklichkeit ab.

Neben den aus wissenschaftlichen Erkenntnissen abgeleiteten Risiken sehen die befragten Experten weitere Risiken: Die Fähigkeit von ChatGPT, Gefühle aus Sätzen zu erkennen, basiert auf mathematischen Algorithmen und nicht auf einem wirklichen Verständnis von Emotionen. Entsprechend ist davon auszugehen, dass die Sentiment Analyse einer Maschine eine andere ist als die eines Menschen. Die Frage, ob ChatGPT die relevanten Emotionsnuancen verstehen kann, ist daher unklar. Des Weiteren ist vor allem in einem interkulturellen Kontext die Beratung von ChatGPT kritisch zu sehen, da der Datensatz und die Labeler, die ChatGPT trainieren, hauptsächlich aus der westlichen Welt stammt und Stimmungen in verschiedenen Kulturen sprachlich unterschiedlich ausgedrückt werden (Dwivedi u. a. 2023). Dementsprechend kann es zu Missinterpretationen kommen. Ferner entstehen durch die Weiterentwicklung von Sprache neue Wörter und Wortbedeutungen, die, aufgrund von ChatGPTs veraltetem Datenstand, sich nicht mehr angemessen interpretieren lassen. Entsprechend kann der Einsatz von ChatGPT im Unternehmenskontext einen negativen Einfluss auf das Image des Unternehmens und dessen Arbeit haben.

Die Erprobung des Use Cases zeigt dennoch ein positives Ergebnis, da die für das Erfolgsszenario erforderlichen Kriterien erfüllt werden konnten (vgl. Abb. 6). So ist eine

professionell formulierte E-Mail generiert worden, die auf die Bedenken des Kunden eingeht und eine Lösungsfindung anbietet, ohne die maximale Anzahl an Prompts von sechs zu überschreiten. Vor dem Versenden müsste in diesem Einsatzszenario die E-Mail nur noch geringfügig individualisiert werden. Daraus folgt, dass die Nutzung von ChatGPT zur Deeskalation in der E-Mail-Kommunikation für Unternehmenberatungen ein sinnvolles Hilfsmittel sein kann, wenn die geltenden Vorschriften und Rahmenbedingungen dabei berücksichtigt werden. Nichtsdestotrotz ist zwischen verschiedenen Konfliktsituationen zu differenzieren. Besteht die Möglichkeit, die Situation in einem persönlichen Gespräch aufzuklären, sollte man diese Maßnahme wählen, um Missverständnisse leichter aus der Welt zu schaffen und besser auf individuelle Bedürfnisse der Konfliktparteien eingehen zu können. Ist dies nicht umsetzbar, kann ChatGPT, vor allem bei schriftlichen Konversationen eine Mediatorrolle einnehmen. Gerade aufgrund der Tatsache, dass Menschen oftmals mit negativen Emotionen auf Infragestellungen der eigenen Leistung reagieren, können Consultingfirmen von ChatGPT profitieren und dieses Risiko einschränken. Dennoch erweist es sich als empfehlenswert, dieses Konzept hauptsächlich bei einfachen Konflikten einzusetzen, die keine längere Konfliktvergangenheit haben. Abschließend ist es unerlässlich, die Ausgaben nochmals zu überprüfen und für die individuellen Konversationspartner anzupassen.

DISKUSSION

Generelle Anwendungsempfehlung

Eine große Stärke von ChatGPT liegt in Analyse, Strukturierung und Schreiben von Texten in menschlicher Sprache. Folglich ist vor allem die Eingabe von eigenen Referenztexten, auf dessen Basis ChatGPT Befehlen nachgeht, eine geeignete Lösung, um das Risiko, Falschinformationen zu erhalten, zu reduzieren. Eine weitere Kompetenz des LLMs ist die Erstellung von Inhalten nach Vorgaben und Referenzeingaben des Nutzers, die als erster Entwurf anschließend weiterverarbeitet und angepasst werden kann. Dabei vermag es sich um Newsletter, Blogs, E-Mails, Reden, Präsentationsinhalte oder ähnliches halten, die mit Hilfe von ChatGPT weiterführend an eine vom Nutzer definierte Zielgruppe angepasst oder in eine oder mehrere, vom Tool beherrschte Sprache übersetzt werden. Weiterhin besteht die Möglichkeit, Referenztexte mit bestimmten Kriterien abzugleichen, die im Vorfeld definiert wurden, um den Text zu kategorisieren. Eine weitere Stärke von ChatGPT besteht in der Möglichkeit zur Einübung von verschiedenen rollen- bzw. Persona-basierten Dialogen. Im direkten Kundenkontakt ist jedoch tendenziell vom Einsatz abzuraten, da im Kundensupport i.d.R. nicht genügend Kontrollmechanismen vorhanden sind. Entsprechend sollte ChatGPT als unternehmensinternes Tool angesehen werden. Unabhängig vom jeweiligen Anwendungsfall sind die Ausgaben jedoch stets noch von Menschen zu überprüfen, bevor man sie final verwendet.

Limitationen

In der vorliegenden Arbeit setzen sich die Limitationen aus drei beeinflussenden Faktoren zusammen: Erstens beschränkt die Wahl von Experteninterviews als Teilmethodik den Umfang der Daten, sodass die Ergebnisse nicht auf die Allgemeinheit übertragbar sind. Zudem ist die geringe Zahl der Befragten nicht repräsentativ für die Gesamtheit der Beratungsindustrie, zumal die Experten auch angegeben haben, sich privat mit neuen Technologien zu beschäftigen. Dementsprechend ist zu vermuten, dass bei ihnen das Interesse, LLMs in ihren Arbeitsalltag zu integrieren, höher als im Durchschnitt ist. Die Untersuchung der Use Cases wird des Weiteren durch das Nutzungswissen der Forschenden eingeschränkt, sodass folglich nur so viel Wissen und Fähigkeiten von ChatGPT nutzbar waren, wie die Forschenden durch die eigenen Promptingkompetenzen extrahieren können. Weiterhin gibt es, aufgrund der Aktualität des Themas, permanente Weiterentwicklungen und Neueröffnungen. So lagen zum Zeitpunkt der Untersuchung noch keine Informationen über den Einsatz von GPT-4 vor. Es ist davon auszugehen, dass die GPT 3.5 Version für Unternehmen zunehmend unattraktiv wird.

SCHLUSSFOLGERUNG UND AUSBLICK

Diese Arbeit untersucht das Potenzial von LLMs am Beispiel von ChatGPT 3.5 für die Automatisierung von Büro- und Verwaltungsarbeiten innerhalb von Consultingunternehmen. Anhand von Experteninterviews sind verschiedene Anwendungsfälle identifiziert worden, in denen man ChatGPT sinnvoll einsetzen kann, um Arbeitszeit und Ressourcen einzusparen. Als Voraussetzung sind passende Anforderungskriterien an das Tool zu definieren.

Die Ergebnisse zeigen, dass die Integration von ChatGPT in die Unternehmensumgebung vielversprechende Möglichkeiten bietet, bestimmte Aufgaben (teilweise) zu automatisieren und Prozesse zu optimieren. Die praxisorientierten Use Cases "Powerpoint Präsentation erstellen" und „Erstellung einer E-Mail zur Deeskalation in der Kundenkommunikation“ verdeutlichen, wie man ChatGPT in der Realität anwenden kann, um die Effizienz und Genauigkeit von Bürotätigkeiten zu verbessern. Aus den Interviews lässt sich eine allgemeine Bereitschaft zur Nutzung von ChatGPT im Arbeitskontext feststellen. Allerdings weisen die Ergebnisse der Befragung auch auf Risikopotenziale hin, insbesondere in Bezug auf Datensicherheit, Ethik und Verantwortlichkeit.

Die Integration von LLMs muss zudem von einer passenden KI-Strategie des Unternehmens begleitet und umfangreiche Aufklärungsarbeit geleistet werden, um die Wahrscheinlichkeit des Eintretens von Risiken zu minimieren und die Mitarbeiterkompetenzen an den zeitgemäßen Standard anzupassen. Die Arbeit hat gezeigt, dass der aktuelle Forschungsstand im Bereich der Automatisierung und LLMs bereits zahlreiche Einsatzmöglichkeiten in verschiedenen Branchen aufzeigt, aber speziell für den Einsatz von LLMs in Büro- und Verwaltungsarbeiten noch weiterführende Forschung benötigt wird.

In den kommenden Jahren ist mit einem deutlichen Anstieg der Größe und Leistungsfähigkeit von LLMs sowie mit der Entwicklung anwendungsspezifischer LLMs zu rechnen. Das bedeutet, dass es auf der einen Seite die Basis-LLMs geben wird, die eine generelle Sprachkompetenz besitzen und eine Vielzahl an Textaufgaben befriedigend erledigen können und auf der anderen Seite werden LLMs zukünftig auf konkrete Umgebungen und Bedürfnisse, wie Unternehmen oder Codingplattformen, angepasst werden.

LITERATUR

- Aiyappa, Rachith u. a. (2023): Can we trust the evaluation on ChatGPT?, arXiv preprint arXiv:2303.12767v1.
- Bender, Emily M. u. a. (2021): On the Dangers of Stochastic Parrots, in: Association for Computing Machinery (Hrsg.): FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York, S. 610–623.
- Bozkurt, Aras u. a. (2023): Speculative Futures on ChatGPT and Generative Artificial Intelligence (AI). A Collective Reflection from the Educational Landscape, in: *Asian Journal of Distance Education*, 18. Jg., S. 53–130.
- Brown, Tom B. u. a. (2020): Language Models are Few-Shot Learners, in: Neural Information Processing Systems Foundation, Inc. (NeurIPS) (Hrsg.): Advances in Neural Information Processing Systems 33, Red Hook, NY, S. 1877–1901.
- Chohan, Usman W. (2023): Generative AI, ChatGPT & The Future of Jobs, Notes on the 21st Century, Critical Blockchain Initiative.
- Chui, Michael/Manyika, James/Miremadi, Mehdi (2016): Where machines could replace humans - and where they can't (yet). The technical potential for automation differs dramatically across sectors and activities, in: McKinsey Quarterly, S. 1–12.
- Chui, Michael/Roberts, Roger/Yee, Lareina (2022): Generative AI is here. How tools like ChatGPT could change your business, Quantum Black AI by McKinsey.
- Ferrara, Emilio (2023): Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models, arXiv preprint arXiv:2304.03738.
- Ganguli, Deep u. a. (2023): The Capacity for Moral Self-Correction in Large Language Models, arXiv preprint arXiv:2302.07459v2.
- Gao, Leo/Schulman, John/Hilton, Jacob (2023): Scaling Laws for Reward Model Overoptimization, in: Krause, A. u. a. (Hrsg.): Proceedings of the 40th International Conference on Machine Learning, PMLR 202, S. 265–279. 131
- Ghavami, Peter (2020): Big Data Analytics Methods. Analytics techniques in data mining, deep learning and natural language processing, De Gruyter, Boston/Berlin.
- Goetze, Trystan S./Abramson, Darren (2021): Bigger Isn't Better. The Ethical and Scientific Vices of Extra-Large Datasets in Language Models, in: Association for Computing Machinery (Hrsg.): WebSci '21. Proceedings of the 13th ACM Web Science Conference 2021, S. 69–75.
- Goldstein, Josh A. u. a. (2023): Generative Language Models and Automated Influence Operations. Emerging Threats and Potential Mitigations, arXiv preprint arXiv:2301.04246.
- Dwivedi, Yogesh K. u. a. (2023): So what if ChatGPT wrote it?. Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy, in: *International Journal of Information Management*, 71. Jg., S. 102642.
- Haleem, Abid/Javaid, Mohd/Singh, Ravi Pratap (2022): An era of ChatGPT as a significant futuristic support tool. A study on features, abilities, and challenges, in: *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2. Jg., Nr. 4, S. 1–8.
- Jong, Jan A. de/van Eekelen, Ilse M. (1999): Management consultants: what do they do?, in: *Leadership & Organization Development Journal*, 20. Jg., Nr. 4, S. 181–188.
- Kasneci, Enkelejda u. a. (2023): ChatGPT for good? On opportunities and challenges of large language models for education, in: *Learning and Individual Differences*, 103. Jg., Nr. 102274, S. 1–9.
- Kuiler, Erik W. (2022): Ethics, in: Schintler, Laurie A./McNeely, Connie L. (Hrsg.): Encyclopedia of Big Data, Springer International Publishing, Cham, S. 452–454.
- Lund, Brady D. u. a. (2023): ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing, in: *Journal of the Association for Information Science and Technology*, 74. Jg., Nr. 5, S. 570–581.
- Moloi, Tankiso/Marwala, Tshilidzi (2021): Artificial Intelligence and the Changing Nature of Corporations. How Technologies Shape Strategy and Operations, Springer International Publishing, Cham.
- Narasimhan, Karthik u. a. (2018): Improving Language Understanding by Generative Pre-Training, OpenAI.134 Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz. Wie KI die Gesellschaft und Marketing beeinflusst, Springer Fachmedien, Wiesbaden.
- Onslow, Mike (2023): Unraveling the Enigma of Tokens in ChatGPT. Your Pathway to Proficiency, URL: https://medium.com/@mike_onslow/unraveling-the-enigma-of-tokens-in-chatgpt-your-pathway-to-proficiency-4ee200e18cc3, Stand: 23. August 2023.
- OpenAI (2023a): ChatGPT plugins, URL: <https://openai.com/blog/chatgptplugins>, Stand: 19. August 2023.
- OpenAI (2023b): GPT Best Practices, URL: <https://platform.openai.com/docs/guides/gpt-best-practices>, Stand: 20. August 2023.

- Ouyang, Long u. a. (2022): Training language models to follow instructions with human feedback, arXiv preprint arXiv:2203.02155.
- Patsakis, Constantinos/Lykousas, Nikolaos (2023): Man vs the machine. The Struggle for Effective Text Anonymisation in the Age of Large Language Models, arXiv preprint arXiv:2303.12429v1.
- Polonsky, Michael Jay/Rotman, Jeffrey D. (2023): Should Artificial Intelligent Agents be Your Co-author? Arguments in Favour, Informed by ChatGPT, in: *Australasian Marketing Journal*, 31. Jg., Nr. 2, S. 91–96.
- Schulman, John u. a. (2017): Proximal Policy Optimization Algorithms, arXiv preprint arXiv:1707.06347, S. 1–12.
- Taulli, Tom (2022): Grundlagen der Künstlichen Intelligenz. Eine nichttechnische Einführung, Springer, Berlin/Heidelberg.
- Wang, Fei-Yue u. a. (2023): What Does ChatGPT Say. The DAO from Algorithmic Intelligence to Linguistic Intelligence, in: *IEEE/CAA Journal of Automatica Sinica*, 10. Jg., Nr. 3, S. 575–579.
- Yin, Shukang u. a. (2023): A Survey on Multimodal Large Language Models, arXiv preprint arXiv:2306.13549v1.
- ZBW - Leibniz Information Centre for Economics (2023): Economics of ChatGPT. A Labor Market View on the Occupational Impact of Artificial Intelligence, Kiel.
- Zhang, Wenxuan u. a. (2023): Sentiment Analysis in the Era of Large Language Models. A Reality Check, arXiv preprint arXiv:2305.15005v1.
- Zhao, Wayne Xin u. a. (2023): A Survey of Large Language Models, arXiv preprint arXiv:2303.18223.
- Ziegler, Daniel M. u. a. (2019): Fine-Tuning Language Models from Human Preferences, arXiv preprint arXiv:1909.08593v2.

Anwendung von Machine Learning und Deep Learning im automatisierten Finanzhandel

Ansätze zur Entwicklung und Anwendung intelligenter Handelssysteme

Dogus Tansel (M. Sc.)

Prof. Dr. Harald Ritz

Prof. Dr. Oliver Hein

Technische Hochschule
Mittelhessen

Technische Hochschule
Mittelhessen

Technische Hochschule
Mittelhessen

Fachbereich MND
Wilhelm-Leuschner-Straße 13
61169 Friedberg
E-Mail: tansel.dogus@gmail.com

Fachbereich MNI
Wiesentraße 14
35390 Gießen
E-Mail: harald.ritz@mni.thm.de

Fachbereich MND
Wilhelm-Leuschner-Straße 13
61169 Friedberg
E-Mail: oliver.hein@mnd.thm.de

Abstract

Im Zuge der steigenden Verfügbarkeit von Daten (bis zum Jahr 2025 weltweit über 180 Zettabyte (Redgate 2021)) und durch die rasante Entwicklung im Rahmen künstlicher Intelligenz kristallisieren sich zunehmend zukunftsorientierte Ansätze zur Automatisierung des finanziellen Handels mit Blick auf Machine Learning (ML)- und Deep Learning (DL)-Methoden heraus. Insbesondere DL-Modelle haben in den letzten Jahren im Finanzsektor an Bedeutung gewonnen, da sie in der Lage sind, komplexe Zusammenhänge und Muster innerhalb der Datenmengen zu erkennen, die für menschliche Händler schwer zugänglich sind (Magwa 2023).

Der vorliegende Artikel stellt ganzheitliche Ansätze zur Entwicklung intelligenter Handelssysteme und die damit einhergehende Nutzbarkeit vor. Hierbei ist das Ziel, zu untersuchen, wie der Einsatz von DL-Modellen innerhalb eines Handelssystems realisiert werden kann. Im Rahmen dieser Realisierung steht dabei die Entwicklung und Anwendung eines algorithmischen Handelssystems für das Treffen automatisierter Handelsentscheidungen im Fokus. Als Ergebnis wurde deutlich, dass der Einsatz von DL-Modellen eine Unterstützung für das aktive Handeln mit Preis- oder Trendvorhersagen sein kann. Teilnehmer auf den Finanzmärkten, die algorithmische Handelssysteme einsetzen, stehen vor der spannenden Aufgabe, verschiedene Ansätze zu berücksichtigen. Ob sie sich für den Einsatz von klassischen ML- oder erweiterten DL-Modellen entscheiden, um Muster in umfangreichen Datenmengen zu erkennen, oder ob Techniken zur Preis- und Trendvorhersage eingesetzt werden, hängt von den individuellen Zielen und Präferenzen ab.

Schlüsselwörter

Algorithmic Trading, Deep Learning, Machine Learning, Tradingbot, Stable Baselines3

15,77 Mrd. US\$ im Jahr 2023 wird auf 23,74 Mrd. US\$ für das Jahr 2028 geschätzt (Mordor Intelligence o. D.).

Angesichts dieser auffallenden Zahlen und des anhaltenden Wachstumspotenzials der aufkommenden Technologieenerungen ist es fundamental, den Markt für algorithmischen Handel weiter zu erforschen und zu verstehen. Aufgrund der divergierenden Menge an Daten und der damit einhergehenden Analyse auf dem Finanzmarkt wird es für konventionelle Händler, auch als Trader bezeichnet, immer herausfordernder die richtige Handelsentscheidung auf der Grundlage der Marktsituation zu treffen.

I. AUSGANGSSITUATION

Die dynamische Welt der Finanzmärkte hat sich in den letzten Jahrzehnten radikal verändert, wobei technologische Fortschritte und die Verfügbarkeit großer Datenmengen neue Möglichkeiten für den Handel eröffnet haben. In diesem Zeitalter des digitalen Wandels hat sich der algorithmische Handel zu einem Schlüsselfaktor entwickelt, der den Finanzsektor sowie die Art und Weise, wie Handelsentscheidungen getroffen werden, revolutioniert. So erzielte laut SelectUSA allein der automatisierte Handel im Jahr 2018 bereits um die 60 – 73% des gesamten Handelsvolumens auf dem US-Aktienmarkt (Shah 2019). Dabei wird die erwartete weltweite Steigerung der Marktgröße des algorithmischen Handels mit derzeit

Die einst konventionelle Analyse des Marktes und das Eröffnen von Handelspositionen profitiert durch den Einsatz intelligenter Handelssysteme und optimiert die Entscheidungsfindung mit Hilfe von großen Datenmengen in Echtzeit. ML, insbesondere DL, hat sich als leistungsstarke Methode erwiesen, um Muster in riesigen Datensätzen zu erkennen und präzise Handelsentscheidungen zu

treffen, die für Trader schwer zugänglich sind. Dieser Ansatz verspricht, die herkömmlichen Methoden des algorithmischen Handels zu erweitern und in Echtzeit auf sich ändernde Marktbedingungen zu reagieren.

II. VERWANDTE ARBEITEN

In der Forschung gibt es inzwischen vielfältige Arbeiten, die sich mit der Entwicklung von Handelssystemen für die Teilbereiche des Finanzmarktes beschäftigen. Einige dieser Forschungsarbeiten haben ihren Schwerpunkt auf den Einsatz von LSTM-Modellen zur Prognose des Schlusspreises von Finanzanlagen gelegt (Huang, Huang und Chen 2022), (Singh, Thulasiram und Thavaneswaran 2022). Andere Forschungsarbeiten fokussieren sich auf die Anwendung und Kombination von Long Short-Term Memory (LSTM) und Deep Reinforcement Learning (DRL)-Modellen, um zeitliche Muster in Finanzdaten präziser zu analysieren und darauf aufbauend Handelsstrategien zu entwickeln (Ansari, et al. 2022). Insgesamt haben diese Arbeiten gezeigt, dass DL-Modelle komplexe Zusammenhänge in den Marktbewegungen erkennen sowie die Entscheidungsfindung in Echtzeit verbessern.

III. HERAUSFORDERUNGEN

Das Ziel des algorithmischen Handelssystems ist es, einen ML-basierten Handelsbot für die Generierung von Kauf- und Verkaufsaufträgen (Ordererzeugung), Platzierung von Kauf- und Verkaufsaufträgen (Orderausführung), das Managen von bereits bestehenden Positionen (Schließen, Halten) sowie die Einhaltung von Risikomanagementmaßnahmen (z. B. Stop Loss, Take Profit) inkl. entsprechender Zuverlässig- und Geschwindigkeit zu entwickeln. Um dieses Ziel zu erreichen, sind insbesondere folgende Herausforderungen mit der Verwendung verbunden und sicherzustellen (Quantified Strategies 2023):

Die **Datenqualität** spielt eine entscheidende Rolle für die Genauigkeit und Effektivität der Handelsstrategien, die durch ML-Algorithmen entwickelt werden. Schlechte Daten können zu ungenauen Vorhersagen und fehlerhaften Handelsentscheidungen führen, weswegen eine Sicherstellung der Qualität essenziell für den Erfolg eines Handelssystems ist. Die Qualität und Vollständigkeit der Trainingsdaten sind entscheidend für die Leistung der Modelle bzw. des Systems.

Verzerrte oder unvollständige Daten können zu fehlerhaften Handelsstrategien führen, weshalb auf eine Reduzierung der **Datenverzerrung (Bias)** Rücksicht genommen werden muss.

Overfitting (zu dt. Überanpassung) beschreibt den Zustand eines Modells, sobald dieses zu stark auf historische Daten trainiert worden ist und dadurch den Datensatz memorisiert. Infolgedessen entstehen weniger effektive Ergebnisse im Einsatz bei der Anwendung auf neue und unbekannte Daten.

Underfitting (zu dt. Unteranpassung) tritt auf, wenn ein Modell während des Trainings zu wenig auf die Merkmale des Trainingsdatensatzes eingeht und dadurch nicht in der Lage ist, selbst die vorhandenen Muster angemessen zu erfassen. Im Gegensatz zum Overfitting deutet Underfitting darauf hin, dass das Modell zu einfach bzw. zu allgemein gehalten ist, um die zugrunde liegenden Strukturen innerhalb der Daten zu verstehen.

Die **Interpretierbarkeit** der Modelle durch den Anwender muss bei jeder Handelsentscheidung nachvollziehbar sein. **Menschliches Eingreifen** ist trotz der fortschrittlichen Technologie weiterhin notwendig und sollte mit der Zeit immer mehr reduziert werden. Sicherstellung bei der **Einhaltung gesetzlicher Vorschriften** für einen risikofreien Handel.

Sobald bei der Implementierung eines ML-basierten algorithmischen Handelssystems die aufgeführten Herausforderungen berücksichtigt werden, können Trader von den zahlreichen Vorteilen profitieren, die ML in Bezug auf eine präzisere Marktanalyse, automatisierte Handelsentscheidungen und verbesserte Risikomanagementstrategien mit sich bringt. Durch die effektive Nutzung von ML kann der algorithmische Handel effizienter und profitabler gestaltet werden, während menschliches Fachwissen und Überwachung gewährleisten, dass die Strategien den regulatorischen Standards entsprechen. Somit eröffnet der Einsatz von ML im algorithmischen Handel neue Perspektiven für Trader und trägt zur kontinuierlichen Weiterentwicklung und Innovation auf den Finanzmärkten bei. Die Einhaltung wird dazu beitragen, dass das Handelssystem effektiv und nützlich für den operativen Einsatz durch Trader ist sowie die Effizienz und Genauigkeit der Handelsprognosen zu verbessern.

IV. THEORETISCHER RAHMEN

In diesem Abschnitt wird zunächst auf den theoretischen Hintergrund und dann im darauffolgenden Kapitel auf die Implementierung eingegangen.

Candlesticks

Candlesticks, auch als Kerzenhalter bezeichnet, stellen eine nicht-konventionelle, aber bedeutende Darstellungsmethode für die Analyse bzw. den Handel auf Märkten dar. Bei der Analyse einzelner Candlestick-Diagramme liefern diese Informationen über die Eröffnungs-, Höchst-, Tiefst- und Schlusskurse des Marktes. In der englischen Sprache wird dies oft mit der Abkürzung OHLC (Open, High, Low, Close) abgekürzt (Udagawa 2018). Mit diesen Kursdaten bzw. dieser Datenstruktur wird im weiteren Verlauf das Training und die Anwendung des Handelssystems auf dem Markt durchgeführt.

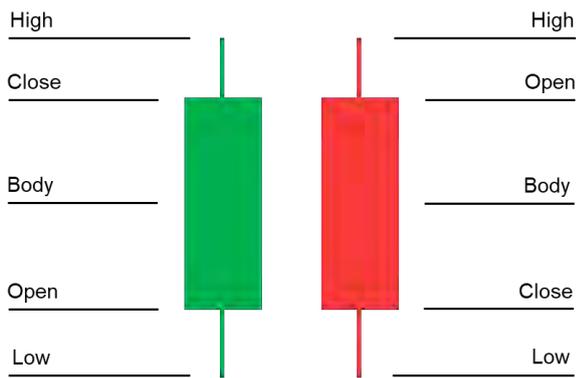


Abbildung 1 Candlesticks (Bullish links, Bearish rechts)

Algorithmische Handelssysteme

Konventionelle algorithmische Handelssysteme bauen auf der Grundlage markttechnischer Bedingungen auf. Dabei werden bewährte technische Indikatoren für das Implementieren von Handelssignalen verwendet. Aufgrund dessen existieren unterschiedliche Arten von markttechnischen Handelssystemen:

Trendfolgesysteme versuchen, von bestehenden Trends in den Finanzmärkten zu profitieren. Sie identifizieren Aufwärts- oder Abwärtstrends und setzen auf steigende oder fallende Kurse (Hofmann 2014). Im Gegensatz zu trendfolgenden Systemen versuchen *Kontra-Trend-Systeme* Trendumkehrungen zu erkennen und davon zu profitieren. Sie setzen darauf, dass der Markt nach einem starken Trend in die entgegengesetzte Richtung korrigiert. (Economy-Pedia o. D.) *Channel-Breakout-Systeme* signalisieren mögliche Handelssignale beim Ausbruch von Kursen aus einem etablierten, durch Trendlinien hergeleiteten Preiskanal (Kuepper, Stapleton und Schmitt 2021).

Machine Learning

ML bildet einen Teilbereich der KI ab, bei dem Algorithmen entwickelt werden, die auf der Grundlage von aufbereiteten Daten lernen können. Hierbei ist das automatische Erkennen von Patterns innerhalb der Datensätze essenziell, um Vorhersagen mit einer hohen Genauigkeit treffen zu können. ML greift dabei auf verschiedene Instrumente der statistischen Analyseverfahren wie Regression, Klassifikation und Clustering zurück (Alpaydm 2004, 1-14).

ML-Standardprozess



Abbildung 2 ML-Prozess (vgl. (Wuttke o. D.))

Ganz gleich, um welche Art von ML-Modell (Supervised, Unsupervised und Reinforcement Learning) es sich handelt, das grundlegende Prinzip der Ausführung bleibt unabhängig vom Anwendungsbereich dasselbe. Im ersten Schritt wird der zu bearbeitende (Trainings-)Datensatz definiert und in das gewünschte System eingespeist. Anschließend erfolgt der Aufbau des Modells unter Verwendung von Algorithmen und Methoden sowie notwendigen Datenaufbereitungsverfahren. Dabei werden Zusammenhänge, Patterns, Korrelationen und Strukturen aus den Daten abgeleitet und dem System als Ausgabe zur Verfügung gestellt (trainiertes Modell). Zum Abschluss können auf Grundlage des trainierten Modells neue und unbekannte Test- sowie Validierungsdaten eingespielt werden, um bspw. Vorhersagen für einen Anwendungsfall zu erstellen (Wuttke o. D.).

ML im algorithmischen Handel

ML bietet aufgrund seiner Fähigkeit, Patterns in großen Datenmengen zu erkennen, vielfältige Möglichkeiten für den Handel. Diese fortschrittliche Technologie kann auf abwechslungsreiche Weise Einfluss nehmen und den Handelsprozess optimieren. Nachfolgend sind einige der potenziellen Einsatzbereiche und Anwendungsmöglichkeiten von ML im Handel aufgeführt (Chopra 2022).



Abbildung 3 Anwendung ML im algorithmischen Handel (vgl. (Chopra 2022))

Sentiment-Analyse: ML kann mithilfe von Daten aus verschiedenen Quellen wie Nachrichten, sozialen Medien und anderen unstrukturierten Daten Tradern ermöglichen die Marktsituation oder das Anlegerverhalten zu erkennen. Mit der Hilfe von Natural Language Processing (NLP) wird der Kontext der Daten analysiert, ausgewertet sowie verstanden, um die sogenannte Marktstimmung zu bestimmen.

Mustererkennung: ML entlastet die Trader von zeitaufwändigen Aufgaben der Datensammlung sowie -verarbeitung und trägt zur Reduzierung von manuellen Arbeiten durch Automatisierung der Analyse bei.

Echtzeit-Datenprognose: Maschinelle Lernalgorithmen können kontinuierlich Echtzeitdaten verarbeiten und lernen, um die Genauigkeit ihrer Vorhersagen zu verbessern. Sie können wichtige Aspekte wie globale Wetterbedingungen, politische Unruhen und den Klimawandel berücksichtigen, die direkte Auswirkungen auf die Handelsbranche haben.

HFT-Maschine: Mit der Hilfe einer Hochfrequenzhandelsmaschine wird Tradern das Eröffnen und Schließen

von tausenden Transaktionen pro Tag unter der Zuhilfenahme maschineller Algorithmen ermöglicht.

Mit einem Blick auf die Anwendungsbereiche kristallisiert der Einsatz von ML in der Erweiterung algorithmischer Handelssysteme neben der vereinfachten Handelsentscheidung deutliche bzw. bemerkbare Vorteile für einzelnen Teilnehmer des Finanzmarktes heraus (Quantified Strategies 2023):

- Verbesserte Genauigkeit von Handelssignalen
- Erhöhte Geschwindigkeit in der Auftragsausführung
- Verbessertes Risikomanagement durch die Reduzierung menschlicher Fehler und Einflüsse
- Zeitersparnis durch Automatisierung von wiederkehrenden Eingaben
- Situationsbedingte Anpassungsfähigkeit von Handelsstrategien

Reinforcement Learning



Abbildung 4 Reinforcement Learning (vgl. (Bhatt 2018))

Reinforcement Learning (verstärkendes Lernen) unterscheidet sich von anderen Formen des maschinellen Lernens, da es darauf abzielt, eine optimale Strategie zu entwickeln, um Probleme zu lösen, ohne vorab bereitgestellte Trainingsdaten. Ein Agent folgt vordefinierten Regeln und einem Belohnungssignal, um durch Trial-and-Error eine erfolgreiche Strategie zu erlernen. Der Fokus liegt auf der Gesamttaktik oder Abfolge von Aktionen, die zu einem gewünschten Zustand führen, anstatt auf einzelnen Aktionen. Der Algorithmus bewertet und lernt aus positiven Aktionen, um eine effektive Strategie zu entwickeln. Ein Beispiel hierfür ist ein Brettspiel, bei dem die Reihenfolge der richtigen Züge entscheidend für den Erfolg ist (Mockenhaupt 2021, 141-142), (Alpaydm 2019, 14).

Deep Learning

DL bzw. tiefes Lernen bildet wiederum einen spezifischen Teilbereich des MLs ab, der sich auf den Einsatz künstlicher NN konzentriert. Diese Netze bestehen aus mehreren Schichten von Neuronen, die hintereinandergeschaltet sind, um komplexe Funktionen zu approximieren. Durch das Lernen aus Daten sind NN in der Lage, anspruchsvolle Aufgaben wie Bild- oder Spracherkennung durchzuführen (Bhattacharyya, et al. 2020, 8-9).

Deep Reinforcement Learning

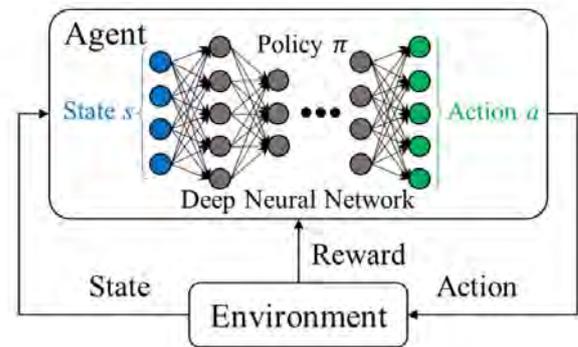


Abbildung 5 DRL (Huang, et al. 2019, 82195)

DRL kombiniert Reinforcement Learning mit neuronalen Netzwerken (NN), um komplexe Entscheidungsprobleme zu bewältigen. DRL-Agenten können Wissen aus Rohdaten, wie visuellen Informationen, ohne vordefinierte Merkmale erlernen. DRL baut auf früheren Fortschritten im Reinforcement Learning auf, indem es NN verwendet, um hochdimensionale Probleme zu bewältigen. CNNs sind Teil dieser DL-Algorithmen und ermöglichen das direkte Lernen aus rohen und hochdimensionalen Daten. Die Spezialisierung des DRL konzentriert sich darauf, optimale Wertfunktionen Q zu finden, um die beste Entscheidung anzunähern (Vijayan PV 2020).

Deep Q-Learning

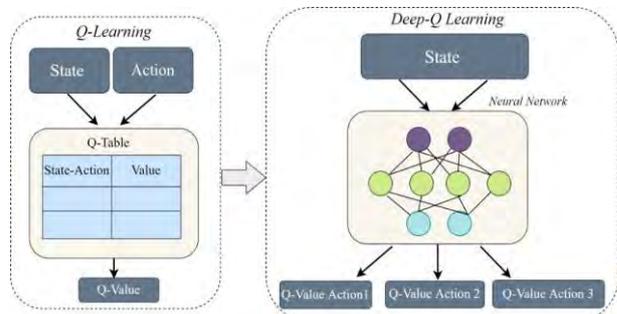


Abbildung 6 Deep Q-Learning (Ansari, et al. 2022, 127473)

Das Q-Learning basiert auf der Q-Funktion, die die erwarteten Belohnungen für Aktionen in einem Zustand berechnet. Die optimale Q-Funktion, Q^* , repräsentiert die maximale Belohnung von einem Zustand ausgehend, wenn die beste Aktion gemäß einer Richtlinie π gewählt wird. Mit diesen Erkenntnissen ordnet das Q-Learning jedem Zustands-Aktionspaar einen entsprechenden Q-Wert innerhalb einer Tabelle zu, wobei der höchste Wert in dieser Tabelle die optimale Q-Funktion darstellt, also die maximale Belohnung. Das Deep Q-Learning (DQL) kombiniert diese Idee mit DRL, indem es ein NN verwendet, um Eingabezustände auf Paare (Q-Wert, Aktion) abzubilden und dadurch die bestmögliche Entscheidung mit der höchsten erwarteten Belohnung zu erhalten (TensorFlow o. D.), (Wang 2020).

Long Short-Term Memory

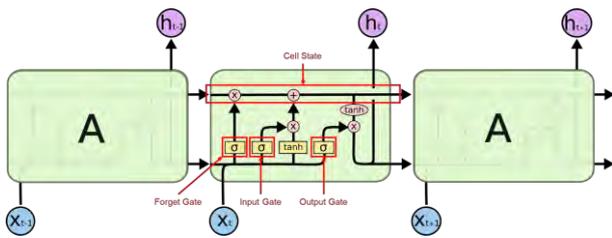


Abbildung 7 LSTM (Olah 2015) (Anpassung, vgl. (Lang 2022))

Recurrent Neural Networks (RNN) sind spezialisierte neuronale Netzwerke zur Verarbeitung von Sequenzen, da sie Informationen aus früheren Schritten speichern und für zukünftige Schritte verwenden können. Dies ist besonders nützlich bei aufeinanderfolgenden Daten, z. B. in der Sprachverarbeitung oder Zeitreihenanalysen.

RNNs haben jedoch das Problem des „Vergessens“, wenn die Trainingsdauer zunimmt und viele Rückkopplungsschleifen in den Hidden-Layern auftreten. Um dieses Problem zu lösen, wurden Long Short-Term Memory (LSTM) entwickelt. LSTMs basieren auf RNNs, verfügen jedoch über Mechanismen, die es ihnen ermöglichen, langfristige Abhängigkeiten besser zu erfassen, indem sie entscheiden, welche Informationen behalten und welche vergessen werden sollen. Dies geschieht durch die Verwendung von Toren wie dem Forget Gate, dem Input Gate und dem Output Gate, um das Langzeitgedächtnis effizient zu aktualisieren und relevante Informationen zu erhalten. Dies ermöglicht LSTMs, langfristige Abhängigkeiten in den Daten effektiver zu berücksichtigen und ist besonders in der Verarbeitung sequenzieller Informationen von Vorteil (Kempkes o. D.), (Luber und Litzel 2018 II), (Lang 2022).

V. BASELINE-HANDELSSYSTEM

Nachdem der theoretische Rahmen mit seinen Konzepten rund um das algorithmische Handeln sowie einem Einblick in die Themen des MLs vermittelt wurde, wird im folgenden Kapitel der Aufbau eines Baseline-Handelssystems präsentiert. Dieses Handelssystem dient als Ausgangspunkt und Basis für die weiteren Entwicklungen, Experimente sowie der Integration von DL-Modellen.

Programmiersprache und Frameworks

Das Handelssystem wird mit Hilfe der Programmiersprache Python aufgebaut, da Python eine weit verbreitete und beliebte Sprache für ML (Gülen 2022) sowie und algorithmisches Handeln (Quantified Strategies 2023) ist. Es bietet eine breite Palette von Bibliotheken und Frameworks, die sich besonders gut für den Aufbau eines Handelssystems eignen.

Einige sinnvolle Frameworks für den Aufbau des Handelssystems mit Python sind (Thakar 2023):

- NumPy: Bibliothek für numerische Berechnungen, dass eine effiziente Handhabung von großen Datenmengen ermöglicht.
- Pandas: Framework für Datenmanipulation und Analyse, dass besonders gut geeignet ist, um Finanzdaten zu verarbeiten.
- TensorFlow & Keras: Framework für ML und der Implementierung von DL-Modellen bzw. dem Aufbau von NN.
- Scikit-learn: Bibliothek für ML mit vorgefertigten Algorithmen wie Regression und Klassifikation für die Modellvalidierung

Architektur und Design



Abbildung 8 Wechselseitige Interaktion (Freepik Company S.L. o. D.), (One Financial Markets o. D.), (Python Software Foundation o. D.)

Die drei Hauptkomponenten der Implementierung bestehend aus dem Finanzmarkt, der Handelsplattform und der Python-Applikation bilden die grundlegenden Bausteine der geplanten Umsetzung des Handelssystems. Der Finanzmarkt stellt die Daten und Informationen bereit, auf deren Grundlage Handelsentscheidungen getroffen werden. Die Python-Applikation übernimmt die Analyse der Daten, die Implementierung von Handelsstrategien und die Ausführung bzw. Signalisierung von Handelsaufträgen. Die Handelsplattform MetaTrader 5 (MT5) fungiert als Schnittstelle zwischen der Python-Applikation und dem Finanzmarkt. Sie ermöglicht die Interaktion mit dem Finanzmarkt, indem sie Positionen platziert und verwaltet sowie notwendige Kontoinformationen abrufen. Eine effiziente Zusammenarbeit dieser drei Hauptkomponenten gewährleistet den reibungslosen Betrieb des algorithmischen Handelssystems.

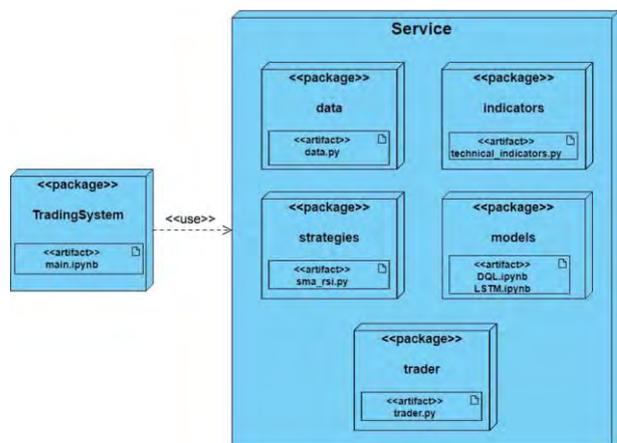


Abbildung 9 Verteilungsdiagramm

Zu Beginn der Implementierung eines Handelssystems müssen geeignete Komponenten festgelegt werden.

Diese Komponenten sind im Verteilungsdiagramm des Baseline-Handelssystems dargestellt (siehe Abbildung 9). Die Abbildung zeigt die geplante Ordnerstruktur bzw. den modularen Aufbau des Systems mit dem Namen „AlgoBot“. Die Pakete in der Abbildung repräsentieren die verschiedenen Module und Funktionen, die im Handelssystem enthalten sein werden.

- data-Modul: Diese Komponente ist verantwortlich für das Herunterladen und Vorverarbeiten von Marktdaten aus verschiedenen Quellen, z. B. Finanzmärkten, APIs oder anderen Datenanbietern.
- indicators-Modul: Zuständig für die Berechnung technischer Indikatoren auf der Basis von Bibliotheken und Frameworks.
- strategies-Modul: Diese Komponente enthält die Logik, um Handelsentscheidungen basierend auf den vorverarbeiteten Daten zu treffen. Dies kann eine Vielzahl von Strategien umfassen, z. B. einen Indikator-Crossover-Ansatz. Ein Indikator-Crossover-Ansatz bezieht sich darauf, Handelsentscheidungen auf Basis von Überkreuzungen (Crossovers) von technischen Indikatoren zu treffen. Hierbei werden typischerweise zwei oder mehrere Indikatoren verwendet. Es wird ein Handelssignal generiert, wenn diese beiden Indikatoren sich über- oder untereinander kreuzen.
- models-Modul: Diese Komponente beinhaltet die implementierten DL-Ansätze für den Handel.
- trader-Modul: Diese Komponente gewährleistet die Verbindung zu den Finanzmärkten und der damit einhergehenden Hauptfunktionalität für das Öffnen und Schließen von Handelspositionen. Ihre Aufgabe besteht darin, die Handelsstrategien und -entscheidungen an die MT5-Plattform zu senden
- TradingSystem-Modul: Dieses Modul fungiert als der zentrale Einstiegspunkt oder das Hauptprogramm des Handelssystems. In den unterschiedlichen Funktionen werden verschiedene Komponenten des Handelssystems aufgerufen und miteinander verknüpft, um den Handelsprozess zu steuern als auch zu koordinieren.

Funktionalitäten

```
class Data:
    def __init__(self):...

    def fetch_data(self, symbol, timeframe, start_pos, num_bars):...

    def download(self, symbol):...
```

Abbildung 10 Data-Modul

Durch das Data-Modul wird dem Handelssystem ermöglicht, historische Marktdaten im OHLC-Format (siehe Abschnitt Candlesticks) zu sammeln, zu verarbeiten und für die technische Analyse zu nutzen. Es ermöglicht dem Handelssystem, Kursdaten von verschiedenen Finanzmärkten über die Handelsplattform abzurufen, wie z.B. Forex-Währungspaare oder Aktienkurse und diese Daten für die Generierung von Handelssignalen und die Ausführung von Handelsstrategien zu verwenden. Die „fetch_data“-Methode empfängt dabei ein gewünschtes

Symbol über einen ausgewählten Zeitraum aus der Handelsplattform MT5. Dem Nutzer steht es dabei frei, den Zeitstempel der Daten zu wählen und die Anzahl der zurückgegebenen Candlesticks anzugeben. Wohingegen die „download“-Methode für das Laden der Daten aus der YahooFinance-API zuständig ist.

Date	Open	High	Low	Close	Adj Close	Volume
2003-12-01	189.330002	189.660004	187.740005	187.630005	187.630005	0
2003-12-02	187.669998	188.809998	187.559998	188.009995	188.009995	0
2003-12-03	188.020004	188.240005	186.490005	187.089996	187.089996	0
2003-12-04	187.029999	187.029999	185.899994	186.220001	186.220001	0
2003-12-05	186.190002	186.740005	185.830002	185.880005	185.880005	0
...
2023-09-18	183.307999	183.332001	182.740005	183.307999	183.307999	0
2023-09-19	182.770004	183.468994	182.731995	182.770004	182.770004	0
2023-09-20	183.089005	183.341003	182.492996	183.089005	183.089005	0
2023-09-21	182.834000	182.919006	181.087997	182.834000	182.834000	0
2023-09-22	181.389008	182.304001	181.199005	181.628998	181.628998	0

5157 rows × 6 columns

Abbildung 11 Datenausgabe

Das Trader-Modul enthält neben den Klassenmethoden auch weitere Methoden für den operativen Einsatz. Bei der Erstellung eines Objekts der Klasse Trader wird eine Accountbezeichnung und ein Passwort angegeben, die zur Initialisierung des Kontos und zur Anmeldung bei der Handelsplattform erforderlich sind.

```
class Trader:
    def __init__(self, account, password):...

    def __del__(self):...

    def buy(self, symbol, quantity):...

    def sell(self, symbol, quantity):...

    def close(self, symbol, quantity):...
```

Abbildung 12 Trader-Modul

Die Methoden „buy“ und „sell“ ermöglichen den Kauf und Verkauf von Handelsobjekten mit einer angegebenen Menge zu einem aktuellen Preis auf dem Markt. Um eine offene Position zu schließen, wird die Methode „close“ verwendet.

```
class TechnicalIndicators:
    def __init__(self):...

    def get_SMA(self, data):...

    def get_SMA100(self, data):...

    def get_RSI(self, data):...

    def indicators_stack(self, data):...
```

Abbildung 13 Indicators-Modul

Das Indicators-Modul stellt dem Handelssystem gängige technische Indikatoren für die Analyse der Finanzmärkte auf Basis geladener Kursdaten zur Verfügung. Die Klasse „TechnicalIndicators“ nutzt das Bibliothekspaket „ta-lib“ (Technical Analysis Library) zur Berechnung

und Generierung verschiedener technischer Indikatoren. Diese Indikatoren können zur Identifizierung von Trends, Volatilität, Momentum und anderen Aspekten der Marktdynamik verwendet werden.

```
def sma_rsi(df):
    close = df["close"]
    rsi = talib.RSI(close, 20)
    sma1 = talib.SMA(close, 20)
    sma2 = talib.SMA(close, 50)

    signal = ""

    if crossover(sma2.iloc[-1], sma1.iloc[-1]) or (rsi.iloc[-1] > 70):
        signal = "sell"
    elif crossover(sma1.iloc[-1], sma2.iloc[-1]) or (rsi.iloc[-1] < 30):
        signal = "buy"

    return signal
```

Abbildung 14 Strategies-Modul: SMA-RSI-Crossover

Das Strategies-Modul ist für die zur Verfügungstellung unterschiedlicher Handelsstrategien verantwortlich. Es beinhaltet verschiedene Funktionen, die jeweils eine spezifische Handelsstrategie implementieren. Diese Strategien können je Anforderung auf die Marktsituation ausgewählt und angewendet werden. Die implementierte Handelsstrategie innerhalb des Moduls ist die SMA-RSI-Crossover-Strategie.

Der modulare Aufbau des Baseline-Handelssystems ermöglicht durch die Trennung der einzelnen Funktionalitäten eine bessere Gesamtübersicht. Innerhalb der „main.ipynb“-Datei im TradingSystem-Modul wird das gesamte Handelssystem orchestriert und ausgeführt. Hier werden alle Komponenten des Systems zusammengeführt, einschließlich des Traders, der technischen Indikatoren und Datenverarbeitung. Die „main.ipynb“-Datei steuert den Ablauf des Handelssystems und ruft die entsprechenden Funktionen auf, um Daten abzurufen, Handelssignale zu generieren, Handelspositionen zu platzieren sowie Informationen über das Handelskonto zu erhalten. Außerdem können in dieser Datei auch die verschiedenen Strategien mit den einhergehenden Parametern des Handelssystems konfiguriert werden.

Das folgende Sequenzdiagramm veranschaulicht die Zusammenarbeit der Komponenten unter Berücksichtigung der unterschiedlichen Module im Rahmen des Baseline-Handelssystems während der Initialisierung einer Handelsposition mit der SMA-RSI-Crossover-Strategie.

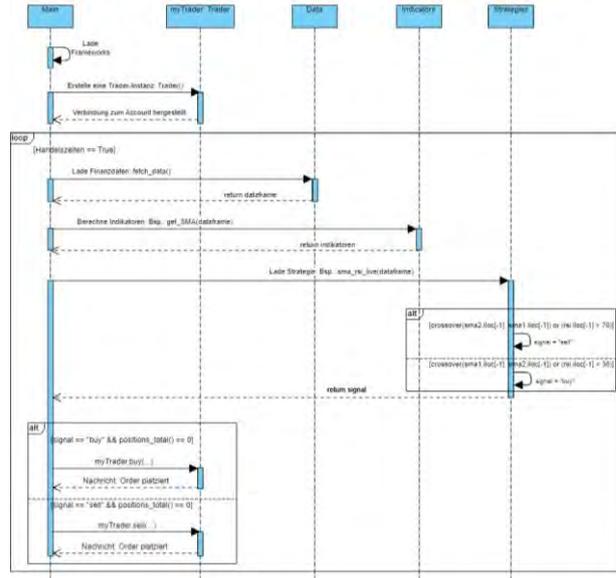


Abbildung 15 Handelssystem-Sequenzdiagramm

VI. DL-HANDELSAGENTEN

In diesem Kapitel werden verschiedene Ansätze der Implementierung von DL-Modellen für den Handel beleuchtet, um diese als Handelsagenten in das bestehende algorithmische Handelssystem zu integrieren.

Feature Engineering

Um eine konsistente und markttechnische Grundlage für die Analyse zu schaffen, ist es von entscheidender Bedeutung einen allgemeinen Datensatz zu erstellen, der für die Anwendung verschiedener Ansätze geeignet ist. Dadurch ist es möglich, unterschiedliche Modelle und Algorithmen auf derselben Datengrundlage zu trainieren. Die Konstruktion eines solchen universellen Datensatzes trägt i. d. R. dazu bei, die Vergleich- und Wiederholbarkeit der Modelle zwischen unterschiedlichen Algorithmen bzw. Handelsansätzen sicherzustellen.

Dafür werden in der Anwendung jedes Handelsansatzes folgende Schritte vorausgesetzt:

- Daten laden,
- Datensatz mit Indikatoren anreichern und
- Daten ggf. normalisieren (MinMaxScalers()) → Werte zw. 0 und 1 erzeugen)

LSTM Handelsansatz

Die Entscheidung zur Verwendung eines LSTM-Ansatzes basiert auf der bereits dargelegten Tatsache einer effektiven Leistungsfähigkeit im Umgang mit Zeitreihendaten. Ähnlich zu anderen Ansätzen, wie sie in den Arbeiten von Huang und Singh dargelegt wurden, liegt der Fokus dieses Ansatzes darauf den Schlusspreis zu bestimmen.

```
df = Data().download(symbol="GBPJPY=X")
df = TechnicalIndicators().indicators_stack(df)
df = df.drop(["Volume"], axis=1)
df.dropna(inplace=True)
df
```

Abbildung 16 LSTM: Daten laden

Zu Beginn der Implementierung eines LSTM-Modells muss die zu vorhersagende Zielvariable definiert werden. In diesem Ansatz setzt sich die Zielvariable aus dem Schlusspreis bzw. Wert einer Kerze des nächsten Tages zusammen. Für das DataFrame wird eine neue Spalte mit der Bezeichnung „target“ erzeugt. Das Setzen des Schlusspreises des nächsten Tages wird durch eine Iteration über alle verfügbaren Handelstage abgerufen und in der Spalte gespeichert. Im Anschluss werden ergebnisverzerrende Spalten entfernt und der Datensatz normalisiert.

Die Grundlage für das erfolgreiche Vorhersagen des LSTM-Modells bildet das festgesetzte bzw. vergangene Zeitfenster, dass das Modell für die Berechnungen der Zielvariable miteinbezieht. Die Auswahl des Zeitfensters innerhalb der Funktion „transformLookBack()“ wird mit der folgenden Abbildung näher erläutert:

```
def transformLookBack(df, look_back):
    # Anzahl der Merkmale (Spalten) im DataFrame
    num_features = df.shape[1] - 1

    # Zweidimensionale Liste von Werten, die aus Zeitschritten der Vergangenheit bestehen. Die Größe der Liste richtet sich dem übergebenen Zeitfenster und der Länge des Datensatzes.
    dataX = [
        [df[i - look_back:i, j] for i in range(look_back, df.shape[0])]
        for j in range(num_features)
    ]

    # Transponieren der dataX-Matrix, um die gewünschte Form zu erhalten.
    dataX = np.moveaxis(dataX, 0, 2)

    # Extrahieren der Zielvariable aus der letzten Spalte des DataFrames.
    dataY = df[look_back:, -1].reshape(-1, 1)

    return dataX, dataY

# Fenstergröße
look_back = 25

X, y = transformLookBack(df, look_back)
```

Abbildung 17 LSTM: Lookback

Die Funktion erwartet neben einem DataFrame einen Parameter „look_back“, der die Anzahl der vergangenen Zeitschritte festlegt, die zur Vorhersage eines zukünftigen Werts verwendet werden. Dies entspricht dem blauen Fenster innerhalb der Abbildung 18.

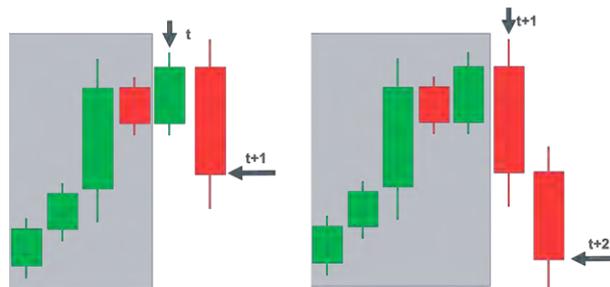


Abbildung 18 LSTM-Lookback-Logik

Anschließend erstellt die Funktion eine Liste von Eingangsdaten „dataX“, indem sie durch die Werte des DataFrames iteriert und ab dem Zeitpunkt t bis t+n-1 Sequenzen von vergangenen Werten jeder Kerze innerhalb des Fensters erstellt werden. Die Zielvariable bzw. der Schlusswert „dataY“ für den Zeitpunkt t+1 wird demnach in Abhängigkeit von der zuletzt betrachteten Kerze zum Zeitpunkt t festgelegt. Dieser Wert wird aus der letzten Spalte des DataFrames zum Zeitpunkt t+1 extrahiert, um die zugehörigen Zielwerte für den betrachteten Zeitraum zu liefern. In der darauffolgenden Iteration wird demnach versucht vom Zeitpunkt t+1 den Schlusswert der nächsten Kerze t+2 zu berechnen.

Nach dem Aufteilen in Trainings- und Testdaten erfolgt im nächsten Schritt der Aufbau sowie das Trainieren des LSTM-Netzwerks, welches wie folgt konstruiert wird:

Layer	Typ	Input Dim.	Output Dim.	Aktivierungsfunktion
1	LSTM	20	150	tanh
2	Dropout	150	150	-
3	LSTM	150	300	tanh
4	Dropout	300	300	-
5	Dense	300	1	linear

Tabelle 1 Aufbau des NN-Layer

Nach dem Training kann auf Basis der Testdaten folgender Graph erzeugt werden.

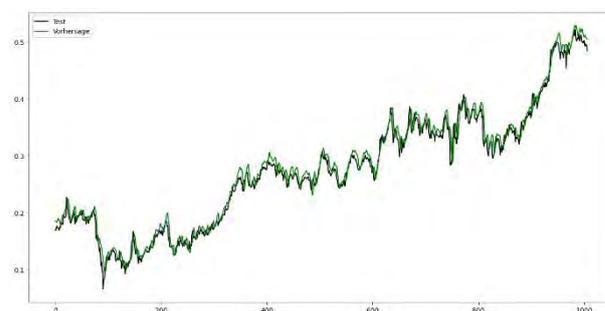


Abbildung 19 LSTM-Vorhersagenplot

Die schwarze Linie bildet den tatsächlichen Werteverlauf (Validierungsdaten) des Graphen ab, während der grüne Graph auf der Grundlage des LSTM-Modells berechnet wurde. Hierbei wird erkenntlich, dass das Modell durchaus in der Lage ist, Zeitreihendaten möglichst nahe an den tatsächlichen Werten zu berechnen. An bestimmten

Momenten erfolgt ein identischer Verlauf, wobei wiederum an anderen Stellen eine Abweichung bzw. eine versetzte Darstellung zu einem späteren Zeitpunkt der tatsächlichen Werte auffällt.

Mit diesen Ergebnissen können unter Einhaltung bestimmter Regeln und durch die Erweiterung des trainierten Modells wichtige Einblicke in den automatisierten Handel gewonnen werden. Dadurch können Trader informiert handeln, sobald bspw. erwartet wird, dass der Schlusswert des nächsten Tages gegenüber dem Vortag sinkt.

Deep RL Handelsansatz

In diesem Abschnitt werden die Konzepte angelehnt an die Implementierung eines DRL-Ansatzes für das Spielen von Atari (Mnih, et al. 2013) mit Blick auf die Handelsebene vertieft, indem die Implementierung von DRL mithilfe NN betrachtet wird.

Zum Start der Implementierung werden die benötigten Bibliotheken in das Programm aufgenommen. Neben den bereits bekannten Frameworks wird das OpenAI-Gym-Framework Anytrading (Haghpanah o. D.) sowie für den Einsatz von RL-Techniken Stable Baselines3 (SB3) (Raffin o. D.) geladen.

Anschließend erfolgt das Deklarieren eines DataFrames mit dem erforderlichen Datensatz aus der YahooFinance-Umgebung und dem Anreichern zusätzlicher Features bzw. Spalten aus dem Indicators-Modul

Wie aus dem klassischen RL bekannt, wird eine Umgebung bzw. Environment für das Trainieren eines RL-Modells benötigt. Da in diesem Ansatz Forex-Währungspaare gehandelt werden sollen, bezieht sich die Variable an dieser Stelle auf die „ForexEnv“-Umgebung der Anytrading-Bibliothek. Das Environment bezieht sich auf die Umgebung oder das System, in dem ein RL-Agent agiert und lernt sowie den damit einhergehenden Spielregeln für das Entwickeln der bestmöglichen Aktionen.

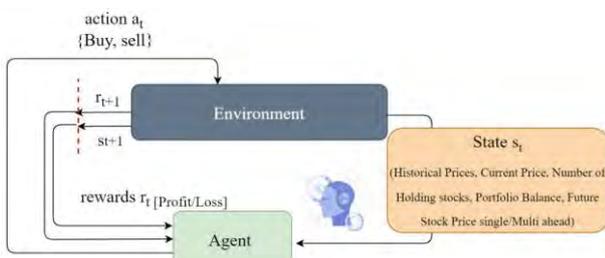


Abbildung 20 RL im Handel (Anpassung, vgl. (Ansari, et al. 2022, 127471))

In diesem Ansatz wird lediglich das Kaufen und Verkaufen von Währungspaaren anvisiert, weswegen der Agent dem Environment „env“ zwei durchführbare Aktion $a_i = [0, 1]$ übermittelt. Dabei steht 0 für das Verkaufen und 1 für das Kaufen. Nach jeder durchlaufenen Iteration regis-

triert das Environment die übermittelte Aktion und berechnet daraufhin die Belohnung bzw. den Reward (Profit/Verlust) als auch den nächsten Schritt in Iteration $t+1$. Um optimale Entscheidungen auf der Basis von Aktionen a_i in einem Zustand s_i zu ermöglichen, wird mithilfe von SB3 ein Deep Q-Network (DQN)-Modell aufgespannt bzw. der Agent initialisiert. Dafür wird eine Instanz durch die Variable „env_maker“ erstellt und anschließend in eine vektorisierte Umgebung „DummyVecEnv“ gewickelt. Diese ist erforderlich, um RL-Modelle in der SB3-Bibliothek zu verwenden. Die „DummyVecEnv“ ermöglicht es, die Umgebung parallel zu betreiben, was für das Training von RL-Modellen von Vorteil ist. Eine Möglichkeit den Agenten zu initialisieren, ist der Aufbau eines bekannten NN aus Dense-Layern sowie Aktivierungsfunktionen.

```
# Erstelle eine Dummy-Umgebung, um das wiederholende Training
# innerhalb des Netzwerks zu ermöglichen
env_maker = lambda: ForexEnv(df=df, frame_bound=(1, len(df)),
window_size=1)
env = DummyVecEnv([env_maker])

# Initialisierung
model = DQN("MlpPolicy", env, verbose=1)

# Training des DQN-Modells
model.learn(total_timesteps=1000000)
```

Abbildung 21 Initiierung einer DQN-Environment

Als Richtlinie π verwendet das Modell die „MlpPolicy“, die auf einem Multi-Layer Perceptron (MLP) basiert. Die Umgebung, in der das Modell trainiert wird, wird durch die Variable „env“ repräsentiert, die zuvor erstellt wurde. Mit diesem Konstrukt und dem Befolgen der optimalen Richtlinie wird die bestmögliche Entscheidung aus den Tupeln [Q-Wert, Aktion] mit der höchsten erwarteten Belohnung getroffen.

Mit dem Ausführen der Lernfunktion werden 1 Mio. Lernschritte innerhalb der erzeugten Environments durchgeführt und folgendes Ergebnis ersichtlich:



Abbildung 22 Ergebnis DQL

Das DQL-Modell hat über einen bestimmten Zeitraum einen positiven Reward-Wert und eine Erhöhung des Portfolios erzielen können. Das „Total Profit“ von 1,055003 bzw. 105,50% vermittelt einen Gewinn von ca. 5,5% des zu Beginn bestehenden Gesamtkapitals.

VII. ANWENDUNG UND ERGEBNISSE

In diesem Abschnitt werden die Implementierungs- bzw. Anwendungsmöglichkeiten auf den realen Markt untersucht. Hierzu werden die zuvor präsentierten Modelle mitsamt ihrer spezifischen Konfiguration in eigenständigen Anwendungsszenarien integriert. Für den praktischen Einsatz des LSTM- und DQN-Modells auf dem echten Markt ist es notwendig, spezielle Hilfsfunktionen zu entwickeln, die die aktuellen Marktdaten verarbeiten. Hierfür wird die Grundstruktur des Modells verwendet und in eine neue, verschachtelte Funktion eingebettet.

Für beide Ansätze wird eine Funktion „get_prediction(model)“ in der „main.ipynb“ des TradingSystem-Moduls implementiert, die je nach übergebenen Modell einen prognostizierten Schlusswert oder ein Kauf- oder Verkaufssignal ausgibt. Innerhalb dieser Funktion werden aktuelle Daten geladen, verarbeitet und normalisiert. Schließlich erfolgt das Übergeben der „Live-Daten“ an das trainierte Modell, welches im Anschluss eine Prognose „prediction“ zur Verfügung stellt. Anhand dieser Prognose wird in den vordefinierten Handelszeiten eine Handelsentscheidung bzw. Handelsposition platziert.

LSTM

```
if prediction[0] > price:
    if mt5.positions_total() == 0:
        print(prediction[0])
        print(mt5.positions_total())
        trader.buy(symbol, 0.5)
        print("Handelsposition platziert")
else:
    if mt5.positions_total() == 0:
        print(prediction[0])
        print(mt5.positions_total())
        trader.sell(symbol, 0.5)
        print("Handelsposition platziert")
    time.sleep(5)
```

Abbildung 24 LSTM-Handelsbedingung

Im ersten Schritt wird überprüft, ob die Vorhersage des LSTM-Modells „prediction“ über dem aktuellen Angebotspreis liegt. Wenn dies der Fall ist und keine offenen Positionen vorhanden sind (überprüft mit „mt5.positions_total() == 0“), wird über das Objekt der Klasse Trader ein Kaufauftrag für 0,5 Einheiten des Währungspaares „GBPJPY“ platziert. Wenn die Vorhersage niedriger ausfällt als der aktuelle Angebotspreis und keine offenen Positionen bestehen, erfolgt die Platzierung eines Verkaufsauftrags für 0,5 Einheiten von „GBPJPY“ mithilfe der Funktion „trader.sell(symbol, 0.5)“.

In diesem Fall wurde ein Anstieg des Kurses prognostiziert, weshalb eine Kaufposition eröffnet wurde. Diese Position erwies sich über mehrere Stunden hinweg als profitabel (siehe Abbildung 23).

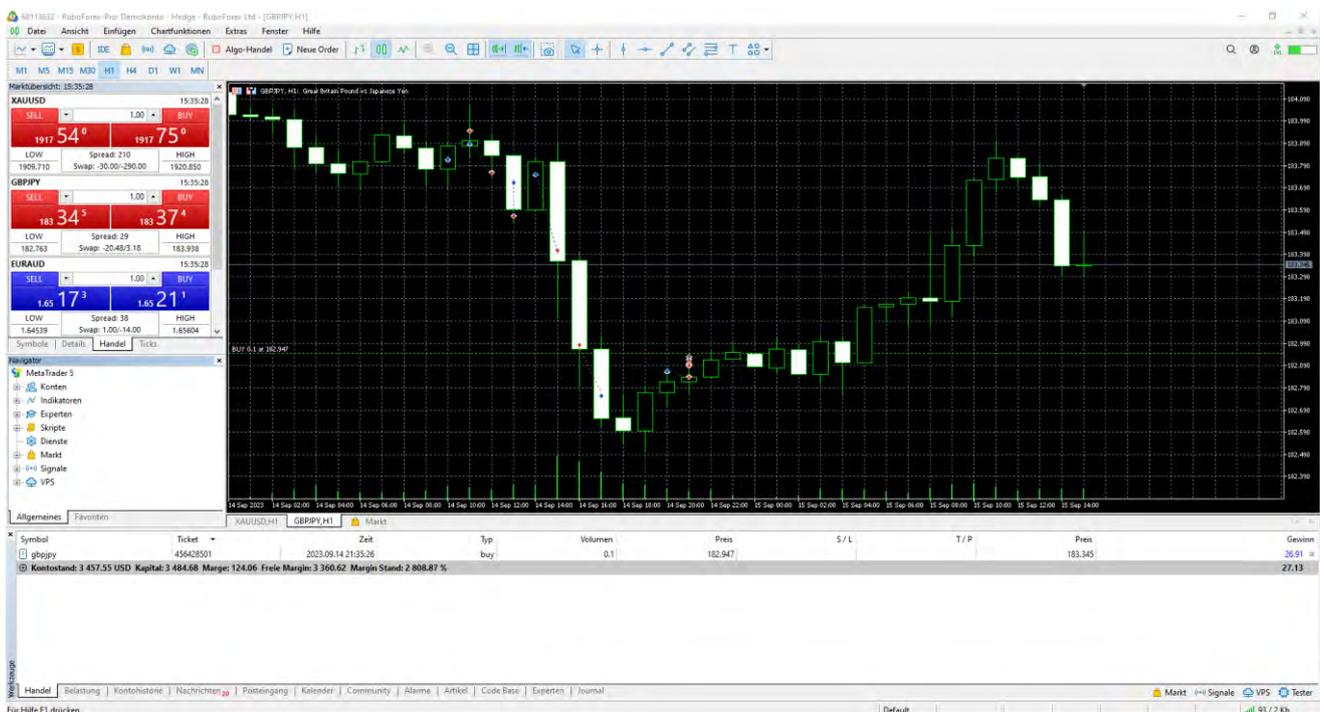


Abbildung 23 LSTM-Einsatz im MT5

DQL

```
if prediction == 1:
    if mt5.positions_total() == 0:
        print(prediction)
        print(mt5.positions_total())
        trader.buy(symbol, 0.5)
        print("Handelsposition platziert")
else:
    if mt5.positions_total() == 0:
        print(prediction)
        print(mt5.positions_total())
        trader.sell(symbol, 0.5)
        print("Handelsposition platziert")
    time.sleep(5)
```

Abbildung 25 DQL-Handelsbedingung

Zur Vorhersage eines Kauf- (1) oder Verkaufssignals (0) muss ein korrekter Datensatz das zuvor trainierte Modell übergeben werden.



Abbildung 26 DQL-Einsatz im MT5

Im Beispiel in Abbildung 26 ist ersichtlich, dass die beiden eröffneten Verkaufspositionen in den ersten Stunden mit einem Verlust konfrontiert wurden, während sich der Markt in einer bullischen Bewegung befand, bevor sie schließlich der vorhergesagten Variable „prediction“ mit dem Wert 0 gefolgt sind. Dies führte zu einem Gewinn von über 150 € auf dem Demokonto.

VIII. FAZIT UND AUSBLICK

Im Rahmen dieser Arbeit wurde unter Zuhilfenahme von gängigen ML-Bibliotheken wie Keras, Tensorflow, Gym und Stable Baselines3 eine Python-Applikation zum Treffen automatisierter Handelsentscheidungen konzipiert und entwickelt. Hierbei wurde ein zuvor implementiertes Baseline-Handelssystem, das Handelsentscheidungen auf der Grundlage technischer Indikatoren getroffen hat, mit DL-Methoden erweitert und trainiert. Dabei wurden verschiedene Architekturen, darunter DRL und LSTM-Ansätze zur Vorhersage von Preisen sowie einer Handelsrichtung (Trend) im Devisenmarkt (Forex) eingesetzt. Während sich der DRL-Ansatz unter Zuhilfenahme des DQN-Konzepts aus bekannten Anwendungen wie selbstlernenden Agenten zum Spielen von Atari durch das Abschätzen von Aktionen und Belohnungen entwickelte, konzentriert sich der LSTM-Ansatz auf die

Vorhersage zukünftiger Marktentwicklungen basierend auf Zeitreihendaten.

Die entwickelte Python-Applikation bietet nicht nur die Möglichkeit, historische Daten zu analysieren und Handelsentscheidungen zu treffen, sondern ermöglicht auch den Handel in Echtzeit auf verschiedenen Märkten durch die Integration der Handelsplattform MetaTrader 5. Dabei werden die DL-Modelle kontinuierlich aktualisiert, um die jüngsten Marktentwicklungen zu berücksichtigen sowie die Genauigkeit der berechneten Vorhersagen zu optimieren.

In jedem Fall zeigt die Entwicklung im Bereich des algorithmischen Handels, dass fortgeschrittene Technologien und automatisierte Datenanalysen eine immer wichtigere Rolle in der Handelsbranche spielen und den Weg für präzisere, effizientere und profitablere Handelsstrategien ebnen. Ein wichtiger Aspekt ist die Notwendigkeit einer kontinuierlichen Weiterentwicklung und Anpassung von Handelsstrategien. Hierbei bildet der Einsatz von DL eine weitreichende Unterstützung der bereits bestehenden Systeme, weshalb diese weder ignoriert noch ersetzt werden dürfen. Die Finanzmärkte unterliegen ständigen Veränderungen, sei es durch geopolitische Ereignisse, wirtschaftliche Entwicklungen oder technologische Fortschritte. Daher ist es von entscheidender Bedeutung, dass Handelssysteme flexibel und anpassungsfähig sind, um auf neue Gegebenheiten zu reagieren.

Neben den erweiterten Ansätzen des DL innerhalb algorithmischer Handelssysteme gibt es das sogenannte Quantitative Trading (Quant Trading). Im Gegensatz zum konventionellen Algorithmic Trading, das auf festgelegten Algorithmen basiert, zeichnet sich Quantitative Trading durch einen erweiterten Ansatz aus. Hier liegt der Fokus auf der Anwendung fortgeschrittener statistischer und mathematischer Modelle zur Entwicklung von Handelsstrategien. Dabei wird gezielt auf den Einsatz von Wahrscheinlichkeitstheorie und Zeitreihenanalyse gesetzt. Dies ermöglicht eine präzisere und leistungsstärkere Herangehensweise an den Finanzmärkten (Sharma, Anderson und Schmitt 2021).

Die Möglichkeit menschlichen Eingreifens sollte während der Einführung algorithmischer Systeme beibehalten werden, da die Verwendung von DL-basierten Vorhersagen weiterhin als Empfehlung und Unterstützung angesehen werden muss, anstatt sie als Ersatz für menschliche Entscheidungen zu verwenden. Mit den entwickelten und angewandten Ansätze kann festgehalten werden, dass die Integration von ML und DL in den algorithmischen Handel ein vielversprechendes und mächtiges Werkzeug darstellt. Mit dieser Integration ist ein entscheidender Schritt in Richtung einer effizienteren, präziseren und automatisierten Handelspraxis auf den Finanzmärkten untersucht worden.

LITERATUR

- Alpaydm, Ethem . 2004. *Introduction to Machine Learning*. Cambridge, Massachusetts: The MIT Press.
- Alpaydm, Ethem. 2019. *Maschinelles Lernen*. Berlin, Boston: De Gruyter Oldenbourg.
- Ansari, Yasmeen, Sadaf Yasmin, Sheneela Naz, Hira Zaffar, Zeeshan Ali, Jihoon Moon, and Seungmin Rho. 2022. "A Deep Reinforcement Learning-Based Decision Support System for Automated Stock Market Trading." *IEEE Access*.
- Bhatt, Shweta. 2018. *Reinforcement Learning 101*. 19 März. Accessed Juli 26, 2023. <https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292>.
- Bhattacharyya, Siddhartha, Vaclav Snasel, Aboul Ella Hassanien, Satadal Saha, and B. K. Tripathy. 2020. *Deep Learning: Research and Applications*. Berlin: De Gruyter.
- Chopra, Yashica. 2022. *Machine Learning for Trading – Can It Predict the Trend?* 16 Juni. Accessed Mai 02, 2023. <https://www.datatobiz.com/blog/machine-learning-for-trading/>.
- Economy-Pedia. o. D. *Arten von Handelssystemen*. Accessed Juli 21, 2023. <https://de.economy-pedia.com/11029980-types-of-trading-systems>.
- Freepik Company S.L. o. D. *Stock Icons*. Accessed August 14, 2023. https://www.flaticon.com/free-icon/stock-market_2910311?term=stock&page=1&position=85&origin=tag&related_id=2910311.
- Gülen, Kerem. 2022. *How to choose a programming language for your machine learning project?* 17 November. Accessed Juli 31, 2023. <https://dataconomy.com/2022/11/17/best-language-for-machine-learning/>.
- Haghpanah, Mohammad Amin. o. D. *gym-anytrading*. Accessed September 01, 2023. <https://github.com/AminHP/gym-anytrading>.
- Hofmann, Jonathan. 2014. *Markttechnische Handelssysteme für ausgewählte Wechselkurse*. Wiesbaden: Springer Gabler Wiesbaden.
- Huang, Peisen, Guanhai Huang, and Hongyu Chen. 2022. "Comprehensive scoring trading model based on LSTM prediction." *International Conference on Electronics and Devices, Computational Science (ICEDCS)*. Marseille, Frankreich. 303-307.
- Huang, Xuefei, Seung Ho Hong, Mengmeng Yu, Yuemin Ding, and Junhui Jiang. 2019. "Demand Response Management for Industrial Facilities: A Deep Reinforcement Learning Approach." *IEEE Access*.
- Kempkes, Christoph. o. D. *Neuronale Netze und der MNIST Datensatz*. 14 April. Accessed Juli 29, 2023. <https://medium.com/@wahlschwabe/neuronale-netze-und-der-mnist-datensatz-4cf57bc4b22>.
- Kuepper, Justin, Chip Stapleton, and Kirsten Rohrs Schmitt. 2021. *Channeling: Charting a Path to Success*. 21 Oktober. Accessed Juli 21, 2023. <https://www.investopedia.com/trading/channeling-charting-path-to-success/>.
- Lang, Niklas. 2022. *Long Short-Term Memory Networks (LSTM) – einfach erklärt!* 4 Juni. Accessed Juli 29, 2023. <https://databasecamp.de/ki/lstm>.
- Luber, Stefan, and Nico Litzel. 2018 II. *Was ist ein Long Short-Term Memory?* 12 November. Accessed Juli 29, 2023. <https://www.bigdata-insider.de/was-ist-ein-long-short-term-memory-a-774848/>.
- Magwa, Lusanele. 2023. <https://www.robeco.com/de/de/einblicke/2023/03/machine-learning-modelle-koennen-interessante-zusammenhaenge-identifizieren>. 01 März. Accessed Mai 02, 2023. <https://www.robeco.com/de/de/einblicke/2023/03/machine-learning-modelle-koennen-interessante-zusammenhaenge-identifizieren>.
- Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. *Playing Atari with Deep Reinforcement Learning*. DeepMind Technologies: arXiv.
- Mockenhaupt, Andreas. 2021. *Digitalisierung und Künstliche Intelligenz in der Produktion*. Wiesbaden: Springer Vieweg .
- Mordor Intelligence. o. D. *ALGORITHMIC TRADING MARKET SIZE & SHARE ANALYSIS - GROWTH TRENDS & FORECASTS (2023 - 2028)*. Accessed Juli 03, 2023. <https://www.mordorintelligence.com/industry-reports/algorithmic-trading-market>.
- Olah, Christopher. 2015. *Understanding LSTM Networks*. 27 August. Accessed Juli 29, 2023. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- One Financial Markets. o. D. *ONE | MT5*. Accessed August 14, 2023. <http://www.vimarkets.me/one-mt5>.
- Python Software Foundation. o. D. *The Python Logo*. Accessed August 14, 2023. <https://www.python.org/community/logos/>.
- Quantified Strategies. 2023. *Best Programming Language for Algorithmic Trading Strategies and Systems?* 23 Juni. Accessed Juli 31, 2023. <https://www.quantifiedstrategies.com/best-programming-language-for-algorithmic-trading-strategies/>.
- . 2023. *Trading Strategies with AI and Machine learning*. 13 Juni. Accessed Juli 24, 2023. <https://www.quantifiedstrategies.com/trading-strategies-with-ai-and-machine-learning/>.
- Raffin, Antonin. o. D. *Stable-Baselines3 Docs - Reliable Reinforcement Learning Implementations*.

- Accessed September 01, 2023. <https://stable-baselines3.readthedocs.io/en/master/>.
- Redgate. 2021. *Volumen der jährlich generierten/replizierten digitalen Datenmenge weltweit in den Jahren 2012 und 2020 und Prognose für 2025 (in Zettabyte)*. 08 September. Accessed Juni 05, 2023. <https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/>.
- Shah, Raj. 2019. *Global Algorithmic Trading Market to Surpass US\$ 21,685.53 Million by 2026*. 05 Februar. Accessed Juni 03, 2023. <https://www.businesswire.com/news/home/20190205005634/en/Global-Algorithmic-Trading-Market-to-Surpass-US-21685.53-Million-by-2026>.
- Sharma, Rakesh, Somer Anderson, and Kirsten Rohrs Schmitt. 2021. *What Is Quantitative Trading? Definition, Examples, and Profit*. 13 Mai. Accessed September 18, 2023. <https://www.investopedia.com/terms/q/quantitative-trading.asp>.
- Singh, Japjeet, Ruppa Thulasiram, and Aerambamoorthy Thavaneswaran. 2022. "LSTM based Algorithmic Trading model for Bitcoin." *IEEE Symposium Series on Computational Intelligence (SSCI)*. Singapur, Singapur. 344-351.
- TensorFlow. o. D. *Introduction to RL and Deep Q Networks*. Accessed Juli 30, 2023. https://www.tensorflow.org/agents/tutorials/0_intro_rl.
- Thakar, Chainika. 2023. *Popular Python Libraries for Algorithmic Trading*. 08 Februar. Accessed Juli 31, 2023. <https://blog.quantinsti.com/python-trading-library/>.
- Udagawa, Yoshihisa. 2018. "Predicting Stock Price Trend Using Candlestick Chart Blending Technique." *2018 IEEE International Conference on Big Data (Big Data)*. doi:10.1109/BigData.2018.8622402.
- Vijayan PV, Vishnu. 2020. *Deep Reinforcement Learning: Artificial Intelligence, Machine Learning and Deep Learning — Introduction, Overview and Contrast for Beginners*. 03 August. Accessed Juli 30, 2023. <https://medium.com/@vishnudevijayanpv/deep-reinforcement-learning-artificial-intelligence-machine-learning-and-deep-learning-e52cb5974420>.
- Wang, Mike. 2020. *Deep Q-Learning Tutorial: minDQN*. 18 November. Accessed Juli 30, 2023. <https://towardsdatascience.com/deep-q-learning-tutorial-mindqn-2a4c855abffc>.
- Wuttke, Laurenz. o. D. *Machine Learning: Definition, Algorithmen, Methoden und Beispiele*. Accessed Juli 26, 2023. <https://datasolut.com/was-ist-machine-learning/>.

Analyse der Chancen und Risiken von KI-basierten Large Language Models am Beispiel von ChatGPT in der innerbetrieblichen IT

Cem Sidar Bulut

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
bulutcem@hs-pforzheim.de

Frank Morelli

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
frank.morelli@hs-pforzheim.de

SCHLÜSSELWÖRTER

Large Language Models, ChatGPT, Natural Language Processing

EINLEITUNG

Die digitale Transformation ermöglicht es zum einen, verschiedene KI-Technologien einzusetzen. Zum anderen erwachsen hieraus jedoch auch Risiken und Probleme. Bei einer der betroffenen Anwendungen handelt es sich um Chatbots unter Verwendung von künstlicher Intelligenz (KI). Chatbots sind vielseitig in unterschiedlichen Bereichen des Unternehmens einsetzbar. Ursprünglich waren diese primär dafür vorgesehen, einen menschlichen Dialog in Schriftform nachzustellen, um Kunden eine virtuelle und sofortige Unterstützung zu bieten. Der heutige Stand der Technik reicht weit darüber hinaus, wodurch sich neue Use Cases schaffen lassen.

Im Rahmen dieser Bachelorarbeit werden Möglichkeiten eines ChatGPT-Einsatzes in der innerbetrieblichen Informationstechnologie eines Unternehmens untersucht und diskutiert. Neben den Chancenpotenzialen fließen auch Risikoaspekte wie die Gefährdung der IT-Sicherheit oder der Umgang mit sensiblen Daten bei der Einbettung von ChatGPT ein. Aus dieser Betrachtungsweise heraus werden Handlungsempfehlungen abgeleitet.

ZIELSETZUNG

Der KI-basierte Chatbot-Markt lässt sich durch rapides Wachstum charakterisieren und erfreut sich starker Aufmerksamkeit in der Öffentlichkeit. Die dem Artikel zugrunde liegende Bachelorthesis verfolgt das generelle Ziel, für KI-basierten Chatbots, durch das Sammeln von möglichen Use Cases im Unternehmen, Transparenz für zugehörige EntscheiderInnen auf Basis des State-of-the-Art zu schaffen. Ferner wird durch die Analyse von Chancen und Risiken im speziellen Fall von ChatGPT angestrebt, Handlungsempfehlungen für den innerbetrieblichen Einsatz im Unternehmen zu geben. Verantwortlichen in einer IT-Abteilung soll es darüber hinaus erleichtert werden, generische Einsatzmöglichkeiten von KI-basierten Chatbots auf das individuelle Unternehmensumfeld zu übertragen.

EINSATZMÖGLICHKEITEN

KI-basierte Algorithmen besitzen die Fähigkeit, eigenständig Schlussfolgerungen zu ziehen und aus dem Feedback neue Dinge zu erlernen. Konversationale KI beschreibt die Kommunikation mit Menschen mithilfe von natürlicher Sprache. Darunter fallen virtuelle Assistenten, konversationale Agenten und Chatbots. Im Rahmen von Natural Language Processing und insbesondere mit Large Language Models (LLMs) wird angestrebt, natürliche Sprache korrekt zu verarbeiten und in einem schlüssigen Kontext wieder auszugeben. Bei einer generativen KI ist es u.a. möglich, Antworten auf nicht im Vorfeld trainierten Fragen zu geben.

In diesem Kontext findet ChatGPT, ein Produkt der OpenAI, mit seinem Sprachmodell Davinci als generative NLP-getriebene Plattform besondere Aufmerksamkeit, sowohl generell in der Öffentlichkeit als auch speziell in der Unternehmenswelt. Es ist davon auszugehen, dass ChatGPT einen erheblichen Einfluss allgemein auf die zukünftige Arbeitsweise im Unternehmen und im Speziellen im IT-Bereich haben wird.

Bereits zum jetzigen State-of-the-Art gibt es viele Möglichkeiten, beim Einsatz von Routinearbeiten MitarbeiterInnen effizient zu unterstützen (vgl. Abb. 1).

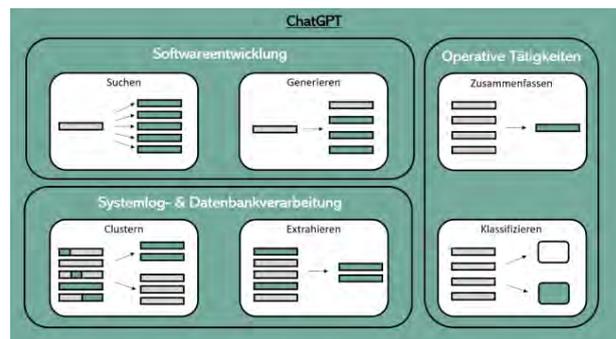


Abbildung 1: Darstellung der möglichen Einsatzpotentiale

Die Bewältigung von Aufgaben durch ChatGPT erfordert i.d.R. einen Impuls durch menschlich generierte Prompts. Entsprechend lässt sich der Einsatzbereich im Sinne einer virtuellen Assistenz charakterisieren.

RISIKEN BEIM EINSATZ

Eine Gefahr besteht darin, dass der Einsatz dieser Technologie kritisches Denken bei MitarbeiterInnen hemmt und die Bereitschaft, eigene innovative Lösungswege zu finden, abnimmt. Umgekehrt kann es durch die zur Verfügung gestellten Daten mit entsprechendem Prompting zu unbefugten Zugriffen und damit zu einer Verletzung der Datensicherheit kommen.

Bei ChatGPT wird seitens OpenAI ein Self-Hosting nicht angeboten. Laut der offiziellen Privacy Policy von OpenAI werden persönliche Daten gesammelt. Entsprechend Artikel 6 der DSGVO erfordert dies i.d.R. die Einwilligung von allen Betroffenen. Generell ist im Sinne des Datenschutzes zu sagen, dass hierbei nur geringe Transparenz auf Anwendungsseite herrscht.

Bei ChatGPT ist eine selbstständige Arbeitsweise von ChatGPT zum heutigen Stand nicht möglich bzw. nicht ratsam, da falsche Schlussfolgerungen bzw. auf Annahmen basierende Outputs generiert werden. Im Sinne von Fairness ist darauf zu achten, dass es nicht zu einer Diskriminierung von Personengruppen kommt.

Mit dem Einsatz von ChatGPT ergeben sich auch juristische Problempunkte: Beispielsweise steigt mit dem ChatGPT-Einsatz prinzipiell das Risiko, dass durch externe Quellen, beispielsweise in Form von Phishing-Kampagnen, Schaden im Unternehmen entsteht. Entsprechend stellt sich die Frage, wer in einem solchen Fall dafür zur Rechenschaft zu ziehen ist. Darüber hinaus erweisen sich urheberrechtliche Aspekte bei generativen KI-Modellen als ungeklärt.

HANDLUNGSEMPFEHLUNGEN

Generell müssen strikte Vorgaben gesetzt und eingehalten werden, um das Sprachmodell nachhaltig im Arbeitsalltag implementieren zu können. Ein überwachter Lernansatz ermöglicht es, eine Feinabstimmung für das LLM vorzunehmen und damit das Modell auf bestimmte Anwendungsszenarien gezielt vorzubereiten. Bei ChatGPT stehen für die Optimierung des Sprachmodells drei Methoden zur Verfügung, Supervised Fine Tuning (SFT), Reward Model (RM)-Training und Reinforcement Learning via Proximal Policy Optimization (PPO). SFT eignet sich vor allem gut, wenn dem System hierfür bereits Trainingsdaten zur Verfügung stehen. Entsprechende Erkenntnisse wurden aus der Simulation im OpenAI Playground gewonnen: SFT trägt sowohl zur Begrenzung der mangelnden Informationssicherheit bei als auch zur Optimierung der Suche nach unternehmensinternen Informationen.

Eine weitere Handlungsempfehlung, um das Datenschutzrisiko zu verringern, ist der Aufbau von ChatGPT auf einem standardisierten Service wie z.B. dem von Azure OpenAI. Dies ermöglicht eine eigenständige Verwaltung des Sprachmodells. Weiterhin müssen flankierende Maßnahmen zur Sensibilisierung der MitarbeiterInnen im Hinblick auf Datensicherheit und Datenschutz getroffen werden.

Darüber hinaus kann man seit dem 25.04.2023 die Verarbeitung der Daten seitens OpenAI unterbinden. Eine Unternehmenspolicy sollte die zugehörige Einstellung zur Deaktivierung entsprechender Möglichkeiten standardmäßig vorschreiben.

Generell erweist sich eine regelmäßige Überwachung und Bewertung des Einsatzes von ChatGPT im Unternehmen als sinnvoll. Durch Kennzahlen ist es möglich, weitere Rückschlüsse auf Effektivität und Effizienz im laufenden Betrieb wie zum Beispiel dem Einsatz des Sprachmodells zu ziehen.

LITERATUR

- Khowaja, Sunder Ali; Khuwaja, Parus; Dev, Kapal** (2023): ChatGPT Needs SPADE (Sustainability, Privacy, Digital divide, and Ethics) Evaluation: A Review. Online verfügbar unter <http://arxiv.org/pdf/2305.03123v1>.
- Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N. et al.** (2017): Attention Is All You Need. Online verfügbar unter <http://arxiv.org/pdf/1706.03762v5>.
- Ouyang, Long; Wu, Jeff; Jiang, Xu; Almeida, Diogo; Wainwright, Carroll L.; Mishkin, Pamela et al.** (2022): Training language models to follow instructions with human feedback. Online verfügbar unter <http://arxiv.org/pdf/2203.02155v1>.
- Dong, Yihong; Jiang, Xue; Jin, Zhi; Li, Ge** (2023): Self-collaboration Code Generation via ChatGPT. Online verfügbar unter <http://arxiv.org/pdf/2304.07590v2>.
- Olmo, Alberto; Sreedharan, Sarath; Kambhampati, Subbarao** (2021): GPT3-to-plan: Extracting plans from text using GPT-3. Online verfügbar unter <http://arxiv.org/pdf/2106.07131v1>.
- Lim, Weng Marc; Gunasekara, Asanka; Pallant, Jessica Leigh; Pallant, Jason Ian; Pechenkina, Ekaterina** (2023): Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. In: The International Journal of Management Education 21 (2), S. 100790. DOI: 10.1016/j.ijme.2023.100790.
- Fuchs, Simon; Drieschner, Clemens; Wittges, Holger** (2022): Improving Support Ticket Systems Using Machine Learning: A Literature Review. Online verfügbar unter <https://scholarspace.manoa.hawaii.edu/handle/10125/79570>.
- Hartmann, Ernst A.** (2022): Digitalisierung souverän gestalten II. Unter Mitarbeit von Ernst A. Hartmann. Berlin, Heidelberg: Springer Berlin Heidelberg. Online verfügbar unter <https://library.open.org/handle/20.500.12657/51926>
- Rahaman, Md. Saidur; Ahsan, M. M. Tahmid; Anjum, Nishath; Rahman, Md. Mizanur; Rahman, Md Nafizur** (2023): The AI Race is on! Google's Bard and Openai's Chatgpt Head to Head: An Opinion Article. In: SSRN Journal. DOI: 10.2139/ssrn.4351785.

VERGLEICH UND EVALUIERUNG VERSCHIEDENER CLUSTERING ALGORITHMEN UND METHODEN ZUR ANWENDUNG AUF WETTERDATEN ZUM DEFINIEREN VON WETTEREREIGNISPROFILIEN UND DEREN CHARAKTERISTIKEN

Julian Lukas Ruben Erath
Duale Hochschule
Baden-Württemberg, Stuttgart
Wirtschaftsinformatik
Paulinenstraße 50
70178 Stuttgart
E-Mail: julian.erath@ibm.com

ABSTRACT

In der vorgelegten Arbeit werden Clusteranalysen an Wetterdaten aus Ontario, Kanada durchgeführt, um Profile für verschiedene Wetterereignisse zu erstellen. Dabei werden verschiedene Clustering-Algorithmen, darunter KMeans, hierarchisches agglomeratives Clustering, Gauß'sche Mischmodelle und DBSCAN, mit Hilfe der Forschungsmethode Design Science Research (DSR) untersucht und angewandt. Das Ziel ist die Identifizierung der am besten geeigneten Methode für die Definition von Wetterereignisprofilen anhand deren Parameter sowie die Optimierung der Analysen und Methoden, um genaueste Ergebnisse zu erzielen. Die Qualität der Ergebnisse wird gemessen anhand der Genauigkeit, Leistungsfähigkeit und den vorgestellten Evaluationsmetriken. Die Wetterereignisprofile (WEP) könnten in der Wettervorhersage zur Bestimmung von verschiedenen Arten von Wetterereignissen und zur Anzeige relevanter Informationen in Dashboards verwendet werden. Ebenso können Modelle für WEP zur Anomaliedetektion entwickelt werden. Die Daten stammen aus sieben Jahren historischer Wetterdaten von IBM Deutschland GmbH und haben das Potenzial für zukünftige Forschung, insbesondere im Bereich der Identifizierung hochkorrelierter Wettermerkmale und der Auswahl relevanter Merkmale.

In the work presented clustering analyses are performed on weather data from Ontario, Canada to generate profiles for various weather events. Different clustering algorithms, including KMeans, hierarchical agglomerative clustering, Gaussian mixed models, and DBSCAN, are investigated and applied using the Design Science Research (DSR) research method. The goal is to identify the most appropriate method for defining weather event profiles based on their parameters and to optimize the analyses and methods to obtain the most accurate results. The quality of the results is measured by the accuracy, performance and evaluation metrics presented. Weather event

profiles (WEP) could be used in weather forecasting to determine different types of weather events and to display relevant information in dashboards. Likewise, models for WEP can be developed for anomaly detection. The data comes from seven years of historical weather data from IBM Deutschland GmbH and has the potential for future research, especially in the area of identifying highly correlated weather features and selecting relevant features.

SCHLÜSSELWÖRTER

Clusteranalyse, Clustermodelle, Wetterdaten, Meteorologie, Wetterdaten Gruppierung, Wetterereignisprofile

1. EINFÜHRUNG

1.1. Motivation

Das Wetter nimmt in vielen Aspekten des menschlichen Lebens eine zentrale Rolle ein. Seit dem Beginn zivilisierten Lebens streben Menschen danach, die Abläufe und Muster von Wetterveränderungen zu verstehen und vorherzusagen. Fortgeschrittene und intelligente Analysetechniken helfen den Menschen dabei, die Auswirkungen des Wetters typisieren und vorhersagen zu können. Im Zeitalter der Informationstechnologien entsteht nun ein Trend, Machine Learning (ML) Techniken und automatisierte Big Data Analysen auf Wetterdaten anzuwenden, um das Wetter noch effektiver kategorisieren, analysieren und vorhersagen zu können (Grabbe et al. 2014). In der Entdeckung von Mustern und Abläufen im Wetter wurden bereits große Fortschritte erzielt, wie in Fathi u. a. 2021 systematisch erörtert wurde. Es gibt jedoch eine Vielzahl von Wetterbedingungen, welche die genaue Analyse, Kategorisierung von Wetterereignissen und deren Vorhersage erschweren. Daher ist es von Relevanz, den Einsatz von ML-Techniken im Bereich der Meteorologie genauer zu untersuchen.

1.2. Problemstellung

Die bisher zumeist eingesetzten Methoden zur Wetteranalyse beruhen auf physikalischen Gleichungen, wie den numerischen Wettervorhersagemodellen. Der Einsatz dieser Modelle ist lediglich für eine genaue Vorhersage des Wetters von bis zu zwei Wochen bewährt. Über diesen Zeitraum hinaus werden physikalische Modelle aber sehr ungenau. Zudem ist deren Berechnung sehr komplex und erfordert viel Rechenleistung (Holmstrom et al. 2016). Darüber hinaus können sie aufgrund ihrer begrenzten räumlichen und zeitlichen Auflösung bestimmte Wetterereignisse wie lokale Gewitter, Stürme oder andere Wetterkatastrophen nicht genau kategorisieren und vorhersagen (de Lima et al. 2013; Ferstl et al. 2017). Das enorme Volumen an meteorologischen Daten macht auch eine manuelle, vollumfängliche Analyse und Kategorisierung durch Meteorologen unmöglich und rentiert sich aus betriebswirtschaftlicher Sicht nicht (de Lima et al. 2013). Daher setzen Wissenschaftler und Data Scientisten vermehrt ML und Big Data Analysetechniken zur Analyse von Wetterdaten ein. Diese haben den Vorteil, dass sie teil- oder vollautomatische Analysen von Wetterdaten ermöglichen und das Klassifizieren und Auswerten von Wetterdaten, übernehmen können (Wei Fang et al. 2014). Für viele Anwendungsfälle werden klassifizierte Wetterdaten benötigt. Daher wird untersucht, wie die Kategorisierung von Wetterdaten mittels unüberwachten ML-Ansätzen teil- oder auch vollautomatisiert werden kann. Hier muss zunächst genauer untersucht werden, aus welchen Parametern sich solche WEP zusammensetzen und wie diese genau definiert werden können. Der Einsatz von ML-Ansätzen hat auch den Vorteil, dass durch diese Zusammenhänge in den Wetterdaten gefunden werden können, die einem Menschen nicht auffallen würden. Dies kann sowohl für die Kategorisierung der Wetterdaten als auch für eine weitere Auswertung genutzt werden (Pooja et al. 2020). Das Thema dieser Thesis ist vor allem im Hinblick auf die momentane Entwicklung des Klimawandels interessant. Mithilfe von Big Data und ML-Methoden analysierte Klimadaten können insofern effektiver ausgewertet werden, da sie im Sinne der Forschung bezüglich des Klimawandels dokumentiert und ihr zeitlicher Trend analysiert werden kann. Interessant ist insbesondere die Betrachtung zur Entwicklung der definierten WEP und deren Merkmale, gerade hinsichtlich extremer Wetterereignisse.

Bisher existiert in der Literatur keine klare Definition von WEP durch Clustergruppen. Mit dieser Problemstellung befasst sich die vorgelegte Arbeit. Im Praxisteil dieser Arbeit sollen Wetterdaten geclustert und ähnliche Wetterereignisse in Gruppen organisiert werden.

Bei der Literaturrecherche dieser Arbeit wurde deutlich, dass die Forschung in relevanten Features bei der Clusteranalyse auf Wetterdaten bereits sehr vorangeschritten ist. Die Ergebnisse bisheriger Forschung können als Basis der weitergehenden Forschung in dieser Arbeit genutzt werden. Bisherige Forschung beschäftigt sich viel mit der Auswertung der Ergebnisse von Clusteranalysen,

im Spezifischen auch Clusteranalysen bezogen auf Wetterdaten, bspw. in Form von Performanz, Evaluationsmetriken und auch der Qualität der Ergebnisse. Diese Forschung wird ebenfalls als Basis für den Praxisteil dieser Arbeit genutzt. Des Weiteren ging aus der Literaturanalyse hervor, dass einige Forschende bereits erfolgreich Clusteranalysen eingesetzt haben, um Wettercluster zu generieren und auszuwerten. Die identifizierten Wettercluster werden zumeist für eine direkte weitere Analyse, bspw. zur Evaluation effektiver Maßnahmen bei bestimmten Wetterereignissen genutzt. Einige Forschende beschreiben zum Teil auch, wie sich die identifizierten Wetterereignisse zusammensetzen und beschreiben relevante Parameter, jedoch werden keine klaren WEP in größerer Menge und über eine größere geographische Fläche definiert. Auch wurde in der Forschung bereits eine Vielzahl von Clusteringalgorithmen und Methoden angewandt. Jedoch wurden bisher nicht mehrere spezifische Algorithmen ausgewählt und unter spezifischen Gesichtspunkten verglichen, um die effizientesten Algorithmen und Methoden zum Clustering von WEP zu identifizieren.

1.3. Zielsetzung

Die Zielsetzung der vorgelegten Bachelorarbeit ist die Beantwortung der folgenden Forschungsfrage: *„Welche Algorithmen und Methoden der Clusteranalyse eignen sich am besten zum Definieren von WEP, gemessen anhand der identifizierten Kriterien und welche Möglichkeiten gibt es, diese weiter zu optimieren?“* Zur Einlösung der Zielsetzung und somit der Lösung der erkannten Problemstellungen bei der Analyse von Wetterdaten, soll der ML-Ansatz des Clusterings verwendet werden, um Wetterdaten zu gruppieren und so WEP zu erstellen. Dieser Ansatz soll dazu beitragen, die räumliche und zeitliche Variabilität des Wetters unter Einfluss einer Vielzahl von Wetterparametern zu berücksichtigen und genauere Vorhersagen zu treffen. Durch die Erstellung von WEP sollen so auch die Charakteristiken verschiedener Wetterereignisse analysiert werden können. Dazu sollen verschiedene Clustering-Algorithmen angewandt und verglichen werden. Die Arbeit wird sich auf die Identifizierung von charakteristischen Mustern und Merkmalen von WEP konzentrieren, sowie auf den Vergleich relevanter Clustering Algorithmen. Diese sollen verglichen werden, um anhand adäquater Kriterien die leistungsstärksten, effizientesten und genauesten Algorithmen zum Definieren von WEP zu evaluieren. Als Ergebnis dieser Arbeit sollen klare WEP für konkrete geographische Regionen definiert werden. Auch sollen mehrere spezifische Algorithmen ausgewählt und unter spezifischen Gesichtspunkten verglichen werden, um die effizientesten Algorithmen und Methoden zum Clustering von WEP zu identifizieren.

Die Ergebnisse dieser Arbeit tragen dazu bei, die bestehenden und zukünftigen Big Data und ML-Techniken im Bereich der Meteorologie zu verbessern und damit das tägliche Leben der Menschen effektiv zu erleichtern.

Der Beitrag der vorgelegten Arbeit zur Wissenschaft ist dabei, wie in Gregor und Hevner 2013 beschrieben, ein

Verbesserungsbeitrag, indem die genaue Anwendung der Clusteringmethoden auf Wetterdaten analysiert und evaluiert wird und die ML-Methodik in Form von positiver Steigerung der Effizienz, Produktivität und Qualität optimiert wird (Gregor und Hevner 2013).

1.4. Vorgehen und kritische Auswahl der Forschungsmethodik

Im Theorieteil der Arbeit wird eine Literaturrecherche mit Konzeptmatrix nach Webster und Watson 2002 durchgeführt. Im Praxisteil der vorgelegten Arbeit wird Design Science Research (DSR) nach Gregor und Hevner 2013 als Forschungsmethodik zur Entwicklung eines Prototyps angewandt (siehe auch: Hevner und Chatterjee 2010; Hevner und March 2004).

Die Methode des iterativen Prototypings nach Wilde und Hess 2007 soll im Rahmen des DSR erweiternd eingesetzt werden. Sie beinhaltet die schrittweise Entwicklung und Verfeinerung eines Prototyps oder Artefakts, um die Forschungsfragen zu beantworten und das Designproblem zu lösen. Es wird in dem iterativen Prozess ein Prototyp entwickelt, getestet und evaluiert, bis das Artefakt die Anforderungen erfüllt. Dieser Ansatz ermöglicht es, Designentscheidungen zu verfeinern, Fehler zu korrigieren und sicherzustellen, dass das endgültige Artefakt den praktischen Anforderungen entspricht.

Zur Evaluierung wird dabei die Kreuzvalidierung nach Shao 1993 verwendet. Diese ist eine Technik in der Statistik und ML, um die Leistung eines Modells zu bewerten und sicherzustellen, dass es generalisierbare Ergebnisse liefert. Dabei wird der Datensatz in verschiedene Teile aufgeteilt, wobei Teile für das Training und andere für die Validierung des Modells verwendet werden. Dies hilft, Overfitting zu verhindern und die Zuverlässigkeit der Modellvorhersagen zu erhöhen.

In dem Forschungsrahmen für DSR werden diese Methoden effektiv zusammen verwendet, indem das iterative Prototyping eingesetzt wird, um das technische Artefakt zu entwickeln und kontinuierlich zu verbessern. Während dieses Prozesses werden die Daten aus den Prototyp-Tests gesammelt und zur Evaluierung und Optimierung genutzt. Die Kreuzvalidierung wird dann verwendet, um die Leistung des entwickelten Artefakts zu bewerten, indem die Genauigkeit und Vorhersagefähigkeit des entwickelten Artefakts überprüft wird. Dies trägt dazu bei, sicherzustellen, dass das Artefakt nicht nur während der Entwicklung, sondern auch in der Validierungsphase angemessen funktioniert und generalisierbare Ergebnisse liefert. Die Kombination dieser Methoden ermöglicht es, DSR in der Entwicklung und Evaluierung von technischen Artefakten fundiert und methodisch durchzuführen, wodurch die Qualität und Nützlichkeit dieser Artefakte verbessert wird.

Schlussendlich wird so ein Prototyp zum Clustering der Wetterdaten zum Definieren von WEP implementiert werden und so zur Verprobung und Evaluierung der Methodik, Parameter und Algorithmen genutzt werden. Die Forschungsrelevanz des Projektes ist dadurch gegeben, dass die vorgelegte Arbeit mit ihrer Implementierung des

Artefakts und dessen Ergebnissen zur Wissensgrundlage für weitergehende Forschung und Nutzung in der Praxis beiträgt.

2. DISKUSSION DES AKTUELLEN STANDS DER FORSCHUNG

2.1. Meteorologie

In der Literatur werden verschiedene Datenquellen (wie ERA5) diskutiert. Eine Beschreibung der verschiedenen Datenquellen, relevante Institutionen und wichtige Instrumente sowie deren Vor- und Nachteile werden im Weiteren diskutiert.

Eine der wichtigsten Institutionen bei der Forschungsarbeit mit Wetterdaten ist die European Centre for Medium-Range Weather Forecasts (ECMWF). Die ECMWF ist eine unabhängige regierungsübergreifende Organisation, welche international unterstützt wird. Ihr Ziel ist es, Wettervorhersagen und Klimaprognosen von hoher Qualität zu erstellen. Hierfür definieren sie internationale Standards für Big Data Themen im Bereich der Meteorologie. Das ECMWF arbeitet auch mit nicht europäischen Nutzern und Staaten zusammen (Pelosi et al. 2020). Von besonderer Relevanz ist ERA5 („the fifth generation of atmospheric reanalysis of the global climate to be produced“) (Hersbach et al. 2020), welches die aktuelle Version des globalen Klima- und Wetter-Reanalyse-Datensatzes des ECMWF ist. ERA5 ist ein umfassendes und standardisiertes globales Klima- und Wetterdatenarchiv mit hoher zeitlicher und räumlicher Auflösung (ECMWF 2023a). ERA5 besteht aus einer Vielzahl von stündlich gemessenen atmosphärischen, land- und ozeanbezogenen Klimaparametern. Die Daten decken die Erde in einem Gitter mit einem Abstand von 30km ab und teilen die Atmosphäre in 137 Ebenen, beginnend mit der Erdoberfläche, bis zu einer Höhe von 85km ein. Qualitätsgesicherte monatliche Aktualisierungen von ERA5 (vom Jahr 1959 bis heute) werden innerhalb von 3 Monaten nach der Echtzeit vom ECMWF veröffentlicht. Vorläufige tägliche Aktualisierungen des Datensatzes stehen den Nutzern innerhalb von 5 Tagen nach der Echtzeit zur Verfügung (ECMWF 2023b). Eine genaue Auflistung der enthaltenen Parameter, detaillierte Erklärungen zu diesen und weitere Standardisierungen sind in der ECMWF Dokumentation zu ERA5 zu entnehmen (ECMWF 2023c). Aus Tabelle 2 sind die relevanten meteorologischen Parameter (zusammengetragen aus der Literatur), welche die Wetterereignisse definieren, zu entnehmen.

2.2 Einsatz von Clustering in der Meteorologie für die Definition von WEP

In diesem Kapitel wird der Stand der Forschung zum Einsatz von Clustering in der Meteorologie für die Definition von WEP diskutiert. Es wird darauf eingegangen, welche Fortschritte Forschende in diesem Bereich schon gemacht haben, welche Algorithmen, Methoden, Themen und Aspekte dazu relevant sind, derzeitige Problemstel-

lungen im Bereich identifiziert und welche Aspekte dieses Themas derzeit und zukünftig noch zu bearbeiten sind.

Die Notwendigkeit ML-Techniken in der Meteorologie zu verwenden, wird auch durch Liljequist und Cehak 1984 deutlich. Sie erklären, dass in der Meteorologie die Schwierigkeit besteht, dass meteorologische Phänomene - kontrastär zu bspw. der Experimentalphysik - nicht simuliert und befreit von störenden Einflüssen untersucht werden können. So müsse „der Meteorologe seine Beobachtungen und Messungen oft durchführen und sein Beobachtungsmaterial nachher statistisch bearbeiten“, um genug Material für seine Forschung zu erlangen. Zudem ist viel Material aufgrund von störenden Einflüssen unbrauchbar (Liljequist und Cehak 1984). Zusätzlich müssen auch genug Messungen vorhanden sein, damit die Resultate nicht von Zufälligkeiten verfälscht sind. Um diese Probleme zu lösen, kann Maschinelles Lernen, bspw. in Form einer Clusteranalyse verwendet werden, um anfallende meteorologische Daten dauerhaft und automatisiert zu speichern und zu analysieren. Diese Daten dann durch vordefinierte WEP zu labeln, bietet für Meteorologen und Data Scientisten einen enormen Mehrwert, was die Bedeutung dieser Forschung weiter verdeutlicht.

Grabbe et al. 2014 haben den Einsatz von Clustering Techniken auf Wetterdaten bereits untersucht, um den Einfluss des Wetters auf geplante Ankunftszeiten von Flugzeugen an Flughäfen und deren Verspätung zu analysieren und um zu evaluieren, welche Maßnahmen bei bestimmten Wetterereignissen effektiv sind, um das Verkehrsflussmanagement zu optimieren. Sie haben herausgefunden, dass es möglich ist, historische Wetterdaten und Flugverkehrsarchivdaten zu clustern, um Auskunft darüber geben zu können, welche Art von Verkehrsmanagement als Reaktion auf verschiedene Wetterereignisse effektiv ist. Genutzt wurde hierfür die Clustering Technik des Expectation Maximization (EM) Algorithmus als Erweiterung des KMeans Algorithmus, um für jede Stunde ein bestimmtes Wetterereignis identifizieren zu können. Hierzu wurden Wetterdaten des Localized Aviation MOS Program (Ghirardelli 2005) genutzt. Insbesondere die Wetterfaktoren Windrichtung, Windgeschwindigkeit, Niederschlag, Niederschlagswahrscheinlichkeit, Gewitterwahrscheinlichkeit, Wahrscheinlichkeit von gefrorenem Niederschlag, Wahrscheinlichkeit für Schnee, Wolkenbildung und Sichtbarkeit wurden als die relevantesten Wetterparameter identifiziert, welche den größten Einfluss auf die Verspätung von Flugzeugen und am Flughafen und den Einsatz von Verkehrsmanagementinitiativen haben. Grabbe et al. 2014 konnten am Newark Liberty International Airport fünf Wettercluster für 2012 identifizieren. Auffällig war, dass ein einzelner Cluster (s. Tabelle 1, Cluster 2) ca. 57% aller Datenpunkte enthielt, welcher grundsätzlich unauffälliges Wetter und normale Abläufe am Flughafen enthält. Zwei Wettercluster wurden identifiziert, welche Unwetter enthielten, welche mit Verspätungen am Flughafen und den Einsatz von Verkehrsmanagementinitiativen korrelieren.

Einer dieser Cluster ist charakterisiert durch erhöhte Niederschlagswahrscheinlichkeit und einer niedrigen Wolkendecke (s. Tabelle 1, Cluster 0), der andere Cluster ist charakterisiert durch schlechtes Wetter, starke Winde und viel Niederschlag. Diese Cluster machen jeweils ca. 7,5% und 3,4% aller Datenpunkte aus. Die Cluster-Ergebnisse der Arbeit sind in Tabelle 1 dargestellt. Die Autoren haben jedoch nur teilweise spezifische WEP definiert. Auch wurden nicht mehrere spezifische Clustering Algorithmen unter Variation ihrer Parameter verglichen und auf ihre Performanz und die Qualität ihrer Ergebnisse überprüft.

Tabelle 1 Clusterergebnisse von Grabbe et al. 2014: Beschreibung der Wettercluster am Chicago O'Hare International Airport 2012

Cluster Index	Number of Members	Description
0	660	Reduced ceilings and elevated probability of precipitation
1	1833	Fair daytime weather hours
2	5045	Early morning and nighttime operations
3	923	Moderate daytime weather hours
4	299	Bad weather – reduced ceilings and visibility, increased winds, increased probability of precipitation

Pooja et al. 2020 stellen eine Selektionsmethode von Features für das Clustering von Wetterdaten vor, bei der mittels Tanimoto Correlation Coefficient Ähnlichkeiten zwischen Features gefunden werden. Damit können relevante Features mit höherer Selektionsgenauigkeit für die Clusteranalyse gewählt werden. Die Korrelation zwischen Features wird berechnet, um größere Ähnlichkeiten zwischen Wetterfeatures zu finden und die Genauigkeit der Clusteranalyse zu erhöhen. Features mit kleiner Ähnlichkeit werden entfernt. Zur genauen Zusammensetzung Berechnung des Tanimoto Correlation Coefficient wird auf Pooja et al. 2020 verwiesen. Mittels des Tanimoto Correlation Coefficient ermittelte Features, welche über eine hohe Ähnlichkeit verfügen, werden dann für die Clusteranalyse ausgewählt, um eine möglichst genaue Wetteranalyse zu gewährleisten. Nach der Selektion relevanter Features wird eine Clusteranalyse ausgeführt, um Wetterdaten in Clustern zu gruppieren. Pooja et al. 2020 verwenden hierfür den EM-Algorithmus, wobei bei der Clusteranalyse die erwartete Likelihood-Wahrscheinlichkeit zwischen den Clusterzentren und den Datenpunkten berechnet wird, um mittels der Maximum a posteriori function die erwartete Wahrscheinlichkeit zum möglichst exakten Zuweisen der Datenpunkte in die richtigen Cluster zu erreichen. Die Clusterergebnisse werden anschließend durch linear program boosting classification kategorisiert.

Tabelle 2 Relevante meteorologische Parameter

Faktor	Symbol	Einheit	Messung	Berechnung	Beschreibung
Luftfeuchtigkeit / Wasserdampfkonze ntration	/	g/m ³	Hygrometer	$f = \frac{h_w}{V}$	Anteil des Wasserdampfis am Gasgemisch der Luft
Temperatur	°	°C oder °K	Thermometer	°K = 273,15 + °C	Wärmegrad der Luft
Luftdruck	Pa oder bar	Pa oder mbar	Barometer / Manometer / Anemometer	1 Pa = 1 Nm ² 1 mbar = 100Nm ² = 1 hPa	Die auf eine Fläche wirkende Kraft
Luftdichte	ρ	kg/m ³	Barometer / Manometer	$\rho = \frac{p \cdot M}{R \cdot T}$	Masse Luft, die in einem bestimmten Volumen enthalten ist
Luftströme (Vertikal) Stärke	v ↑	km/h	Anemometer	l = km/h	Gibt an, wie schnell sich die Luft an einem spezifischen Punkt in vertikaler Richtung vorbewegt
Luftströme (Horizontal) Stärke	v ↔	km/h	Anemometer	l = km/h	Gibt an, wie schnell sich die Luft an einem spezifischen Punkt in horizontaler Richtung vorbewegt
Luftströme (Vertikal) Richtung			Anemometer		Gibt die genaue Richtung an, in die sich die Luft auf vertikaler Achse an einem spezifischen Punkt vorbei bewegt
Luftströme (Horizontal) Richtung	Winkel in ° oder Himmelsrichtung	° oder Himmelsrichtung	Anemometer	0 – 360 ° / sechzehn 22,5°-Schritte (Nord (N), Nordnordost (NNE), Nordost (NE), Ostnordost (ENE), usw.)	Gibt die genaue Richtung an, in die sich die Luft auf horizontaler Achse an einem spezifischen Punkt vorbei bewegt
Böen	v ↔	km/h	Anemometer	l = km/h	Starke Luftbewegung von kurzer Dauer
Niederschlag	mm	mm	Ombrometer / Hyetometer	$1 = \frac{l}{m^2} = 1 \frac{dm^3}{m^2} = \frac{(0,1m)^3}{m^2} = 0,001 \frac{m^3}{m^2} = 0,001 m = 1 mm$	Gibt die Menge an Wasser an, das sich aus Wolken, Nebel oder Dunst oder Luftfeuchtigkeit in einer definierten Zeitspanne auf einer 1m ² großen Fläche sammelt
Sonneneinstrahlung / Sonnenintensität	E	W/m ²	Pyranometer		Bestrahlungsstärke der Sonne auf die Erdoberfläche
Schneedichte	-	kg/m ³	Bestimmte Menge Schnee wird aus der Schneedecke ausgestochen und gewogen		Masse Schnee pro Quadratmeter in der Schneeschicht
Schneefläche / Schneeealbedo	-	0 - 1	Satellitenbild Auswertung	0 - 1	Misst die Reflektion des schneebedeckten Teils einer Fläche der Erde als Fraktion der Solarstrahlung (Kurzwellen), die vom Schnee aus dem solaren Spektrum reflektiert werden
Taupunkt	°	°C oder °K	Taupunktspiegel-hygrometer	°K = 273,15 + °C	Kondensationspunkt von Wasser in der Luft
Bedeckungsgrad / Bevölkerungsgrad		ISO/CIE-Standard ISO 15469:2004(E) / CIE S 011/E:2003	ISO/CIE-Standard ISO 15469:2004(E) / CIE S 011/E:2003		Ansammlung von kondensiertem Wasser in der Erdatmosphäre

Zusammengetragen aus
Liljequist und Cehak 1984;
Grabbe et al 2014;
Fathi et al. 202;
Hasan et al. 2016;
ECMWF 2023

Zur Evaluation der Ergebnisse wird auf vier Parameter zurückgegriffen. Die Genauigkeit der Feature Selektion, die Genauigkeit der Cluster, die False Positive Rate und die Durchlaufzeit. Obwohl Pooja et al. 2020 eine Zwei-Schritt-Methodik einführen, vergleichen und evaluieren sie nicht verschiedene Clustering Algorithmen unter Variation ihrer Parameter. Allerdings wurden effektive Messkriterien für die Performanz und die Qualität der Ergebnisse eingeführt und effektive Methodiken zur Selektion relevanter Features. Auch die identifizierten Wettercluster mit relevanten Kriterien und Charakteristiken werden nicht als WEP vorgestellt.

Das und Sun 2014 nutzen Assoziationsregeln, um den Zusammenhang zwischen bestimmten Wetterereignissen und Unfallstatistiken zu analysieren und evaluieren die Effektivität verschiedener Maßnahmen. Mit diesen konnten sie einfache Regelsätze einführen, welche Assoziationen zwischen Wetterparametern, Wetterereignissen und geeigneten Maßnahmen zum Vorbeugen von Unfällen erzeugen. Außerdem konnten versteckte Muster in den Daten analysiert werden, um ein besseres Verständnis über die Wetterdaten zu erlangen. Obwohl Das und Sun 2014 keine Clusteralgorithmen einsetzen, identifizieren sie Wetterereignisse über Assoziationsregeln und stellen u.a. fest, dass bei Regen eine erhöhte Unfallgefahr herrscht. Es ist denkbar, diese Assoziationsregeln auch zum konkreten Identifizieren von WEP oder die Methodik erweiternd mit der Clusteranalyse einzusetzen.

Xu et al. 2015 stellen eine kaskadierte Clusteringmethode bestehend aus KMeans Algorithmen vor, mit welchem sechs Wettercluster generiert werden. Hierfür wird ein kaskadiertes KMeans Clustermodell auf zehn ausgewählte Wetterparameter angewandt. Die maximale Euklidische Distanz zwischen den Datenpunkten eines Clusters und dem Clusterzentrum wird als Kante eines Clusters zum Berechnen der optimalen Zuweisung der Datenpunkte in die Cluster genutzt. Xu et al. 2015 analysieren viele Parameter ihrer vorgestellten kaskadierten KMeans Clustermethode. Allerdings werden keine genauen WEP definiert und auch keine anderen Clusteralgorithmen miteinander verglichen.

Ferstl et al. 2017 nutzen agglomeratives hierarchisches Clustering (HAC) für die Analyse der zeitlichen Veränderung in Ensembles von Wettervorhersagen. Sie vergleichen keine verschiedenen Clusteringmethoden und definieren auch keine WEP, aber nutzen HAC erfolgreich, um Wetterensembles anhand verschiedener Parameter zeithierarchisch für Wettervorhersagen zu clustern. De Lima et al. 2013 stellen zwei Clusteringansätze vor, die durch Klassifikation ergänzt werden. Es wird HAC zum Clustern der Wetterdaten verwendet und anschließend eine Klassifikation auf die erstellten Testdaten angewandt. Zur Auswertung der Ergebnisse wird eine neu vorgestellte Similarity Metric verwendet. Der zweite Ansatz kombiniert eine Klassifikation mit einer dichte-basierten Clusteranalyse, bei welcher drei Cluster identifiziert werden. Ausgewertet werden die Ergebnisse mit einer Kreuzvalidierung und Klassifikations-Performanz-

Indexes. Erfolgreich konnten extreme Wetterereignisse identifiziert werden. Die genauen Charakteristiken der Extremwetterereignisse wurden nicht vorgestellt. Die Details zu den Methodiken sind sehr ausführlich beschrieben.

Nach eingehender Literaturanalyse geht hervor, dass sich zukünftige Arbeiten im Gebiet des Clusterings von Wetterdaten sowohl mit der Evaluierung und dem Vergleich von verschiedenen Clusteralgorithmen und der Variation ihrer Parameter beschäftigen muss als auch mit der genauen Beschreibung der identifizierten WEP.

2.3 Machine Learning

2.3.1 Normalisierung

Die Daten-Normalisierung (oft auch als Daten-Standardisierung bezeichnet) ist eine Methode zur Aufbereitung von Analysedaten. Sie bezeichnet den Prozess, bei dem Parameterwerte verschiedener Intervalle mit verschiedenen physikalischen Einheiten auf einen einheitlichen Intervall (bspw. -1 bis 1 oder 0 bis 1) skaliert werden (Zhou 2021). Dieser Schritt ist insbesondere bei der Anwendung von Algorithmen von Relevanz, die auf der Skaleninvarianz basieren, so wie die meisten Clusteralgorithmen, welche die Distanz der Datenpunkte zueinander berechnen (Jo 2019). Zur Berechnung der Distanz wird zumeist die Euklidische Distanz, in manchen Fällen aber auch die Hausdorff Distanz, Manhattan Distanz oder Cosine Distanz verwendet. Für weitere Details sei auf Liu und Deng 2020 verwiesen.

Jo, J.M. 2019 diskutiert verschiedene Normalisierungsmethoden wie simple feature scaling, min-max scaling, maximum-absolute scaling, Z-score scaling und robust scaling. Diese basieren auf verschiedenen mathematischen Vorgehensweisen, um die Datenparameter auf eine gemeinsame Skala zu bringen, indem die Standardnormalverteilung aller Parameter so umgerechnet werden, dass sie einen Mittelwert von 0 und eine Standardabweichung von 1 haben. Für die detaillierten mathematischen Hintergründe sei auf Zhou 2021 sowie Jo, J.M. 2019 verwiesen.

2.3.2 Auswahl der Anzahl an Cluster

Ein Großteil, der in der Literatur in der Arbeit mit Wetterdaten verwendeten Clustermethoden nutzen Algorithmen, welche nicht automatisch die optimale Anzahl der zu generierenden Cluster berechnen. Sie benötigen eine Methode zum Finden der optimalen Clusteranzahl k . Im Sinne des Praxisteils der Arbeit ist damit die Anzahl der zu identifizierenden WEP zu verstehen. Es ist davon auszugehen, dass für verschiedene Regionen der Erde unterschiedliche Anzahlen von WEP identifiziert werden. In diesem Kapitel werden Methoden diskutiert, mit welchen die optimale Anzahl an Clustern in einem Datensatz - unter Einsatz einer bestimmten Clustering Methode - identifiziert werden können.

In Liu und Deng 2020 wird zu diesem Zweck die ‚Ellenbogen-Methode‘ (engl. ‚Elbow Method‘) vorgestellt. Diese stellt ein Verfahren zur Bestimmung der optimalen

Anzahl an Clustern bei einer Clusteranalyse dar. Eine ausgewählte Clustering Methode wird iterativ durchgeführt, wobei die Anzahl der zu generierenden Cluster k mit der ersten Iteration 1 beträgt und für jede Iteration i um 1 erhöht wird, bis jeder Datenpunkt x ein eigenes Cluster darstellt, also $k = x_{max}$. Bei jeder Iteration wird die Summe der quadrierten Distanzen aller Datenpunkte jedes Clusters zum Clusterzentrum berechnet. Der Durchschnitt aller quadratischen Distanzen der Clusterzentren zu allen Datenpunkten der Cluster, kann als Wert einer Kostenfunktion verstanden werden, welcher zu minimieren ist. Üblicherweise wird für diese Berechnung die Euklidische Distanz verwendet (Syakur et al. 2018). Für diese Berechnung kann die Kostenfunktion $L = \sum_{i=1}^k \sum_{x \in C_i} |x - C_i|^2$ verwendet werden, wobei L die Kostenfunktion ist, mit x ein Element des Clusters C_i und k der Anzahl an Cluster $|C_i|$ (Liu und Deng 2020). Mit steigender Anzahl der Cluster k wird der Wert L der Kostenfunktion kleiner. Vor der Stelle der optimalen Clusteranzahl k ist die Verminderung des Wertes L der Kostenfunktion groß und nach der Stelle der optimalen Clusteranzahl k ist dieser stark verringert. In einem sogenannten Elbow-Plot wird der Wert der Kostenfunktion in Abhängigkeit zu k dargestellt. Der Elbow-Plot beschreibt typischerweise einen Bogen. An der Stelle der optimalen Clusteranzahl k ist eine starke Beugung des Graphen festzustellen, woher die visuelle Elbow-Methode ihren Namen nimmt. An dieser Stelle, an der der Winkel des Plots minimal ist, erreicht k also seinen optimalen Wert. In Yuan und Yang 2019 wird außerdem die Silhouette-Methode vorgestellt, welche als Evaluationsmetrik für Clusteranalysen verwendet wird. Die Details zu Evaluationsmetriken für Clusteranalysen und zu Silhouette-Methode werden in Kapitel 2.3.6. vorgestellt. Allerdings sei vorweggegriffen, dass Evaluationsmetriken auch zur Bestimmung der optimalen Clusteranzahl k in einem Clusteringmodell verwendet werden können. Ähnlich wie bei der Elbow Methode wird eine ausgewählte Clustering Methode iterativ durchgeführt, indem für $i=1$ auch $k=1$ und für jede Iteration i um 1 erhöht wird, bis $k = x_{max}$. Bei jeder Iteration wird der durchschnittliche Silhouette Koeffizient über alle Datenpunkte berechnet und jenes Clustermodell mit Clusteranzahl k gewählt, bei dessen Iteration der durchschnittliche Silhouette Koeffizient maximal ist.

2.3.3. Möglichkeiten zum Messen und Evaluieren der Performanz der ML-Algorithmen

Die diskutierten ML-Algorithmen werden im Praxisteil der Arbeit auf die Wetterdatensätze angewandt werden, um Datenpunkte in WEP zu clustern. Für jeden Algorithmus wird dessen Effizienz evaluiert. Es sind drei Messwerte von Relevanz: Die Performanz des Algorithmus, der Wert des Evaluationsindex und die Qualität der Ergebnisse in Form von der Genauigkeit zu identifizierten WEP (Pooja et al. 2020; De Lima et al. 2013). Ein Evaluationsindex gibt an, wie gut ein Algorithmus gleiche Datenpunkte in gleiche Cluster und ungleiche Datenpunkte in ungleiche Cluster sortiert. Gesucht wird also

eine hohe Intra-Cluster-Ähnlichkeit und eine geringe Inter-Cluster-Ähnlichkeit (Zhou 2021).

Zu diesem Zweck soll der bereits in Kapitel 2.3.2. vorgestellte Silhouette Koeffizient verwendet werden. Neben diesem werden in Zhou 2021 auch noch weitere Validitätsindizes vorgestellt (Jaccard Coefficient, Fowlkes und Mallows Index, Rand Index, Davies–Bouldin Index, Dunn Index). Aufgrund der Prominenz des Silhouette Koeffizienten in der Forschungsliteratur wird in dieser Arbeit auf diesen zurückgegriffen.

Die Silhouette-Methode wird in Yuan und Yang 2019 vorgestellt. Bei dieser wird berechnet, wie gut ein Datenpunkt in das ihm zugewiesene Cluster passt. Hierzu werden die Faktoren Kohäsion und der Silhouette Koeffizient verwendet. Kohäsion beschreibt die Ähnlichkeit, zwischen einem Datenpunkt und dessen Cluster. Wird der Datenpunkt mit einem fremden Cluster verglichen, wird dieser Wert Separation genannt. Dieser Vergleich wird über den Silhouette Koeffizient berechnet, der Werte zwischen -1 und 1 annehmen kann. Nimmt der Silhouette Koeffizient den Wert 1 an, so ist ein Datenpunkt dem zugewiesenen Cluster sehr ähnlich, bei -1 sind sich Datenpunkt und Cluster sehr unähnlich. Der Silhouette Wert kann auch genutzt werden, um zu evaluieren, ob das genutzte Clustermodell akzeptabel und passend ist. Der Silhouette Koeffizient wird wie folgt berechnet:

$$s(x) = \frac{b(x) - d(x)}{\max\{d(x), b(x)\}} = \begin{cases} 1 - \frac{d(x)}{b(x)}, & d(x) < b(x) \\ 0, & d(x) = b(x) \\ \frac{b(x)}{d(x)} - 1, & d(x) > b(x) \end{cases}$$

Formel 1 Algorithmus zum Berechnen des Silhouette Koeffizienten

Kalkuliert wird die durchschnittliche Distanz $d(x)$ eines Datenpunktes x zu den anderen Datenpunkten desselben Clusters. Je kleiner $d(x)$, desto mehr passt x zu dem jeweiligen Cluster, es besteht also eine hohe Kohäsion. $d(x)$ entspricht der Intra-Cluster-Unähnlichkeit des Datenpunktes x , welche minimiert werden soll. Der Durchschnitt von $d(x)$ über alle Datenpunkte eines Clusters C wird die Cluster-Unähnlichkeit von C genannt. In einem zweiten Schritt wird die durchschnittliche Distanz aller Datenpunkte $b(x)$ eines fremden Clusters zu x berechnet. Diese wird als Inter-Cluster-Unähnlichkeit bezeichnet, welche maximiert werden soll. Je größer $b(x)$, desto weniger passt x zu dem jeweiligen Cluster. Es besteht also eine große Separation.

Ein weiterer wichtiger Faktor zum Messen der Performanz eines Algorithmus ist dessen Komplexität O . Diese beschreibt die Menge an Ressourcen, die bei Anwendung eines Algorithmus benötigt werden. Es wird zwischen Zeitkomplexität und Speicherkomplexität unterschieden. Die Zeitkomplexität gibt an, wie viel Zeit ein Algorithmus bei der Ausführung benötigt, während die Speicherkomplexität angibt, wie viel Speicher- und Rechenressourcen für die Ausführung benötigt werden (Firdaus und Uddin 2015).

3 EXPERIMENTELLES DESIGN

Im folgenden Kapitel der Arbeit wird auf das Design und die Implementierung des erstellten Artefakts eingegangen. Zuerst werden die akquirierten und verwendeten Daten genauer erklärt. Anschließend werden die im Theorieteil der Arbeit identifizierten Methoden und Algorithmen in einem Python Programm mittels Jupyter Notebooks unter Verwendung der Python Libraries Pandas und Sklearn implementiert. Wie im vorangegangenen Kapitel diskutiert, wird das zu entwickelnde Artefakt iterativ verbessert und evaluiert.

3.1 Population, Datenquellen und Datensätze

Die für die Erstellung des Designartefakts / Prototypen der Arbeit verwendeten Wetterdaten werden von der IBM Deutschland GmbH und ihren Partnern im Bereich der Meteorologie zur Verfügung gestellt. Die zur Verfügung gestellten Datensätze enthalten aggregierte und abgeleitete Wettermerkmale aus sieben Jahren (Juli 2015 bis Dezember 2022) historischer ERA5-Wetterdaten, aufgeteilt auf die Region Ontario im Süden von Zentralkanada. Für die Analyse wurde die kanadische Provinz Ontario in 55 Regionen aufgeteilt, welche in Abb. 1 dargestellt sind und für jede Region wurde mittels der historischen aggregierten Wetterdaten vom Jahr 2015 bis 2022 die Clusteranalyse unter Anwendung des implementierten Modells durchgeführt, um für jede Region bedeutende WEP zu definieren. Die Aggregationen der Wetterparameter sind räumlich nach den Regionen und zeitlich nach den Messstunden erzeugt. Die Wetterdaten werden in stündlicher Granularität gemessen.



Abbildung 1 Karte der Regionen der genutzten aggregierten historischen Wetterdaten in Ontario, Kanada

Die für die Nutzung in der Analyse des Praxisteils dieser Arbeit ausgewählten ERA5 Parameter, aufgeteilt nach Regionen, sind die folgenden Wetterparameter:

- maximale Böenstärke (`max_windgust`), welcher die höchste Windböenintensität der betrachteten Stunde des Datenpunktes in m/s angibt
- durchschnittliche Temperatur (`avg_temp`) in Kelvin
- maximale Schneedichte (`max_snow_density_6`) in kg/m^3
- maximaler kumulierter Niederschlag (`max_cumulative_precip`) in mm
- maximale kumulierte Eisschicht (`max_cumulative_ice`) in mm

- maximaler kumulierter Schnee (`max_cumulative_snow`) in mm
- maximale durchschnittliche Windgeschwindigkeit (`avg_windspeed`) in m/s
- durchschnittliche Änderung des Luftdrucks (`avg_pressure_change`) in Pa
- durchschnittliche Windrichtung (`avg_winddir`) in Kreisgrad (für die Analyse wurde der Sinus (`avg_winddir_sin`) und Kosinus (`avg_winddir_cos`) des Gradwertes genutzt)

Der ursprüngliche Datensatz enthielt weitere Parameter, die für die eigentliche Analyse redundant waren, bspw. Bodenfeuchte, der Taupunkt oder die erweiterte Schneedichte. Die Parameter Taupunkt, Feuchtkugelttemperatur und Temperaturveränderung sind redundant, da sie stark mit dem Parameter Lufttemperatur korrelieren (s. Abb. 2). Bei Verwendung mehrerer stark korrelierender Temperaturparameter wird das Ergebnis der Clusteranalyse verfälscht. Zudem sind im ursprünglichen Datensatz für jeden Parameter der minimale, der maximale und der Durchschnittswert für die betrachtete Stunde enthalten. Es wird jeweils nur einer dieser Werte benutzt, der bei der PCA als der relevanteste identifiziert wurde. Die Relevanz der Parameter wird durch die Höhe der entsprechenden Parameterwerte in den Eigenvektoren der Hauptkomponenten wiedergegeben. Mittels der Literaturanalyse und einer PCA wurden die relevantesten Parameter bestimmt und in das Datenobjekt zur Anwendung im Clustermodell aufgenommen. Zudem wurden weitere benötigte Parameter integriert. Die durchschnittliche Temperaturänderung wurde aus bestehenden Daten berechnet und der kategorische Wert der Windrichtung mittels Berechnung von Sinus und Cosinus der Einheit Grad in zwei numerische Werte für die Windrichtung umgerechnet und in zwei neue Spalten des Datenobjektes aufgenommen. Bei der Durchführung der Analyse werden nur die in diesem Kapitel erläuterten Wetterparameter genutzt, während bei der Visualisierung der Ergebnisse alle in Abb. 2 dargestellten Parameter genutzt werden, um weitere Zusammenhänge zwischen den Wetterparametern auswerten zu können, ohne die eigentliche Analyse durch redundante Daten zu verfälschen. Durch Darstellung der Datenparameter über den gesamten Zeitverlauf in Zeitreihendiagrammen und durch Erkundung des Datenobjektes können weitere wichtige Erkenntnisse über die Daten und die Analyse gewonnen werden. Durch eine solche Auswertung wurde der Input des Modells iterativ optimiert. Mittels der ‚StandardScaler‘ Klasse der scikit-learn Bibliothek werden diese normalisiert (Scikit Learn 2023b), um trotz unterschiedlicher physikalischer Einheiten und variierender Skalen der Einheitswerte eine effektive Analyse zu ermöglichen. So wurden die Daten durch Normalisierung auf eine vergleichbare Skala gebracht.

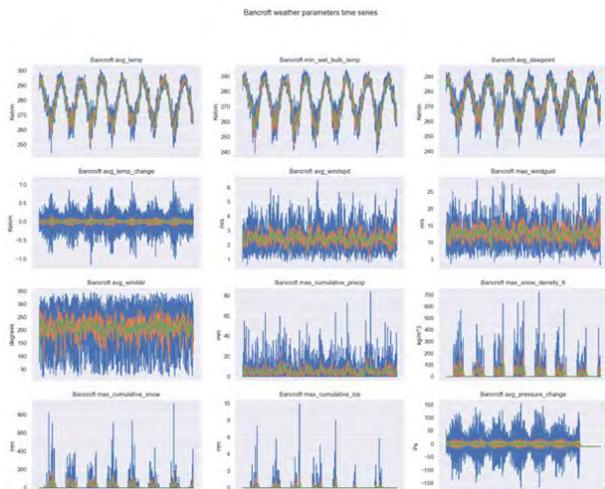


Abbildung 2 Wetterparameter der Region Bancroft von den Jahren 2015 bis 2022 dargestellt in Zeitreihendiagrammen

3.2 Aufbau des Modells und Durchführung der Analyse

Nach Durchführung der Datenakquise und der Datenvorverarbeitung, inklusive Feature Engineering zum Entwickeln neuer Parameter und Optimieren der bestehenden Parameter, beginnt der Aufbau des Clustering Modells. An dieser Stelle variiert die Implementierung der verschiedenen Clusteringmethoden. Zur Vorbereitung des KMeans Clustering Modells, des HAC-Modells und der Gauß'schen Mischmodelle muss zuerst die optimale Anzahl der zu generierenden Cluster für den geladenen Wetterdatensatz gefunden werden. Hierfür wird die in Kapitel 2.3.3 eingeführte Ellenbogen-Methode, unterstützt durch den Silhouette Score, eingesetzt. Die Ergebnisse dieser Methoden sind für den KMeans Algorithmus, angewandt auf die Region Bancroft, beispielhaft in Abb. 3 dargestellt.

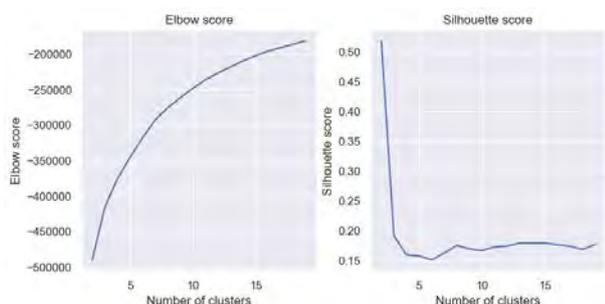


Abbildung 3 Beispielhafte Ergebnisse der Ellenbogen- und Silhouette-Methode zum Identifizieren der optimalen Anzahl Cluster eines Modells (KMeans, reguläre Methode)

In Abb. 3 wird ersichtlich, dass die mathematisch optimale Anzahl an Clustern zwei beträgt, an dieser Stelle beträgt der Silhouette Score 0,52. In der Literaturanalyse wurde erörtert, dass für eine bestimmte Region kleiner als 30km² eine Wetterclusteranzahl von ca. 5 bis 12 zu erwarten ist. Auch nach iterativer manueller und visueller Evaluation der Ergebnisse, weist die Genauigkeit der Ergebnisse für eine Clusteranzahl zwischen 5

und 12 eine sehr hohe Qualität auf. Daher wurde der weitere Aufbau der Analyse in drei Methoden unterteilt: Reguläres Clustering, kaskadiertes Clustering und eine Clusteranalyse nach durchgeführter PCA zur Dimensionsreduktion. Die Details zu den jeweiligen Methoden werden in den folgenden Kapiteln diskutiert.

3.2.1 Reguläres Clustering Model

Sollte für die Ergebnisse der Ellenbogen- und Silhouette-Methode eine optimale Clusteranzahl von zwei herauskommen, so wird bei der Anwendung des regulären Modells für die verwendete Anzahl der Cluster das zweithöchste lokale Maximum des Silhouette Score genutzt. Ist für das Ergebnis dieser Methoden die optimale Clusteranzahl höher als zwei, wird das globale Maximum genutzt. Mit diesem Wert für die Anzahl der Cluster wird das jeweilige Modell anschließend auf den Wetterdaten der einzelnen Regionen trainiert. Die Ausführungszeit, die beanspruchten Rechenressourcen und der Silhouette Score wird gemessen. Für jeden Parameter wird der Durchschnittswert, der maximale Wert, der minimale Wert sowie der Median für jedes Cluster berechnet und in Radardiagrammen dargestellt. Die so dargestellten Cluster verfügen über normalisierte Werte, da für die Visualisierung in Radardiagrammen eine vergleichbare Skala nötig ist. Aus den Visualisierungen geht zudem hervor, aus wie vielen Datenpunkte jeder Cluster besteht. Die in Radardiagrammen visualisierten Daten sind beispielhaft für die Region Bancroft unter Verwendung des KMeans Algorithmus in Abb. 4 dargestellt.

Die Radardiagramme können visuell evaluiert werden, um grobe Informationen über die Wettercluster zu erlangen wie bspw. die generelle Form, die Anzahl der enthaltenen Datenpunkte und ausschlaggebende Parameter einzelner Cluster (bspw. kumulatives Eis, kumulativer Schnee und Schneedichte für Cluster 2 und 4). Für eine detailliertere Evaluierung der Wettercluster werden Distogramme zu jedem Parameter aller Cluster erzeugt. In diesen können die Details der einzelnen Parameter betrachtet und ausgewertet werden. Insbesondere fallen Korrelationen zwischen bestimmten Clustern und Parametern auf, bspw. den verschiedenen Temperaturparametern. Auch können durch Visualisierung der Ergebnisse in Form von Distogrammen signifikante, typische und charakteristische Parameter für einzelne Cluster identifiziert werden. Beispielhaft sei hier der Cluster 2 in Abb. 5 genannt, welcher durch hohes kumulatives Eis, eine hohe Schneedichte und eine nordöstliche bis östliche Windrichtung gekennzeichnet ist. Zuletzt werden für die reguläre Clustering Methode auch die Clusterstatistiken zur Detailanalyse (s. Tabelle 3) sowie Durchlaufzeit, CPU-Ausführzeit und der genutzte physikalische Speicher (RAM) des Modells angegeben, welche in Tabelle 6 zusammengeführt sind.

Tabelle 3 Cluster-Statistiken der Wetterdaten (für die Parameter Windböenstärke und Lufttemperatur. KMeans, reguläre Methode)

Cluster	max_windstf	max_windstg	max_windstb	max_windstm	max_windstl	max_windstc	max_windstn	max_windste	max_windstf	max_windstg	max_windstb	max_windstm	max_windstl	max_windstc	max_windstn	max_windste	max_windstf	max_windstg	max_windstb	max_windstm	max_windstl	max_windstc	max_windstn	max_windste
0	15298	16.0238268	2.6292362	11.76448	14.75548	16.17051	17.96031	23.95741	15298	276.789207	6.2003982	246.86762	272.58764	276.256279	281.988887	308.844006	15298	276.789207	6.2003982	246.86762	272.58764	276.256279	281.988887	308.844006
1	1762	28.841789	3.1027069	3.10854	3.24101	3.40066	3.54899	21.97772	2762	327.625494	4.6889124	261.89319	262.249489	266.351616	272.603787	278.302001	1762	327.625494	4.6889124	261.89319	262.249489	266.351616	272.603787	278.302001
2	498	16.3888462	4.84882178	9.54551	14.81512	16.70488	17.96248	23.95741	498	276.789214	4.6504788	268.489211	268.788802	271.658702	273.238141	276.208919	498	276.789214	4.6504788	268.489211	268.788802	271.658702	273.238141	276.208919
3	208	24.8888882	4.88888148	3.88884	3.88889	3.88891	3.88893	28.88888	208	276.789202	5.1544272	248.788812	267.788887	271.664418	273.788816	286.888811	208	276.789202	5.1544272	248.788812	267.788887	271.664418	273.788816	286.888811
4	2298	16.3888462	4.84882178	9.54551	14.81512	16.70488	17.96248	23.95741	2298	286.288208	4.6488222	271.711288	268.921038	286.921098	283.788131	300.84414	2298	286.288208	4.6488222	271.711288	268.921038	286.921098	283.788131	300.84414
5	492	34.882222	3.8888264	3.88882	3.88882	3.88881	3.88881	27.88884	492	286.20882	4.6488222	271.888818	268.941778	286.921098	283.788131	300.84414								

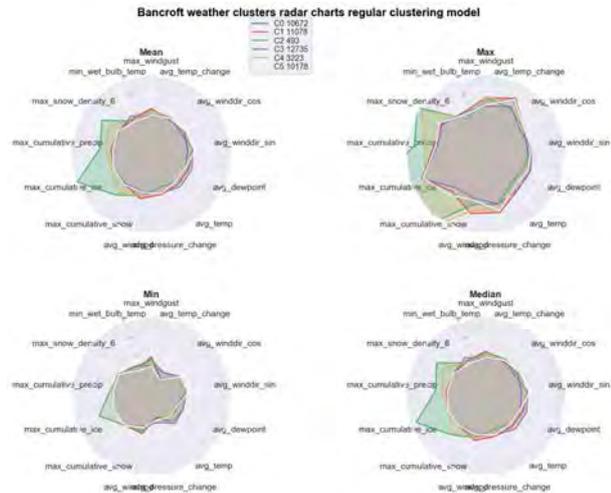


Abbildung 4 Beispielhafte Visualisierung der Cluster der Wetterdaten (KMeans, reguläre Methode) in Radardiagrammen

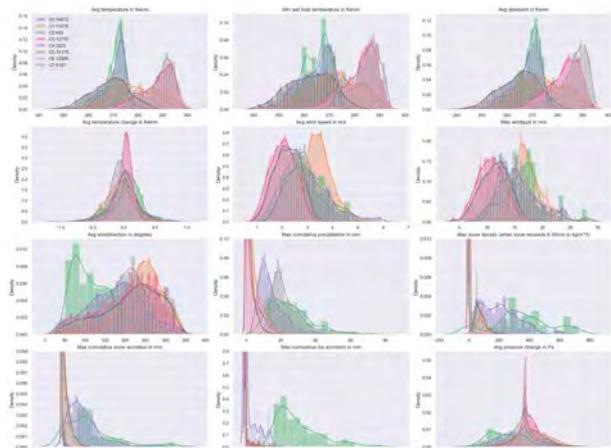


Abbildung 5 Beispielhafte Visualisierung der Cluster der Wetterdaten (KMeans, reguläre Methode) in Distogrammen

3.2.2 Kaskadiertes Clustering Model

Ergibt sich für die Ellenbogen- und Silhouette-Methode eine optimale Clusteranzahl von zwei, so wird zusätzlich zu dem regulären Clustermodell das kaskadierte Modell eingesetzt. Bei diesem werden die Wetterdaten einer Region zuerst in zwei Cluster gruppiert. Durch iterative Implementierung und Evaluation konnte herausgefunden werden, dass ein Cluster zumeist ein Wettercluster ist, der Extremwetterereignisse enthält, wenn dieser 5% bis 20% der Datenpunkte der Gesamtmasse enthält. Die Ergebnisse dieses ersten Schrittes des kaskadierten Clusterings sind in Abb. 6 und 7 zu sehen.

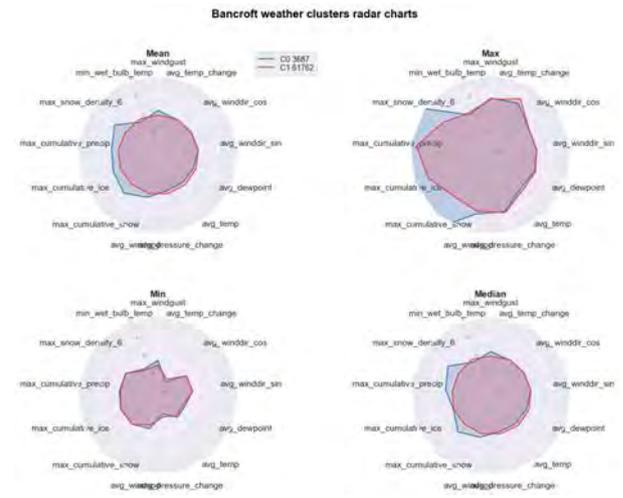


Abbildung 6 Beispielhafte Visualisierung der Cluster der Wetterdaten (KMeans, erster Schritt der kaskadierten Methode) in Radardiagrammen

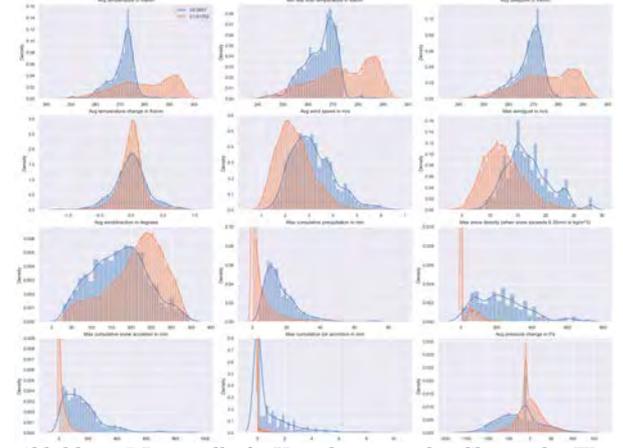


Abbildung 7 Beispielhafte Visualisierung der Cluster der Wetterdaten (KMeans, 1. Schritt der kaskadierten Methode) in Distogrammen

Tritt wie in den Abb. 6 und 7 der Fall ein, dass ein Wettercluster 5% bis 20% der Datenpunkte enthält, wird dieser als ein Wettercluster betrachtet, der Extremwetterereignisse enthält. Auf diesen wird in einem zweiten Schritt die Ellenbogen- und Silhouette-Methode angewandt und mit der hierbei identifizierten optimalen Anzahl an Clustern, das Extremwettercluster in einem zweiten Schritt regulär geclustert. In Abb. 8 sind für die Region Bancroft unter Verwendung des KMeans Algorithmus beispielhaft die Ellenbogen- und Silhouette-Methode für den zweiten Schritt des kaskadierten Modells zu sehen. Die endgültigen Ergebnisse der kaskadierten Methode werden ebenfalls in Radar- und Distogrammen dargestellt (s. Abb. 9 und 10). Es fällt auf, dass die Wetterereignisse viel genauer und klarer definiert sind und der endgültige Silhouette Score viel höher ist, die Cluster sind mathematisch also optimaler getrennt. Allerdings ist zu bedenken, dass in der kaskadierten Methode nur die Extremwetterereignisse und deren Charakteristiken eindeutig definiert und analysiert werden können.

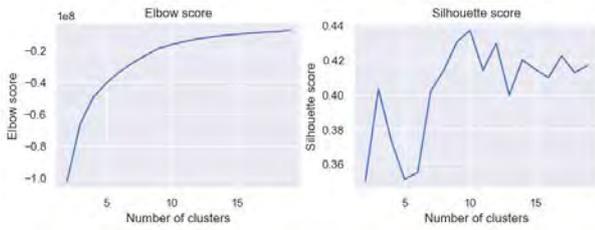


Abbildung 8 Beispielhafte Ergebnisse der Ellenbogen und Silhouette-Methode (KMeans, 2. Schritt kaskadiertes Modell)

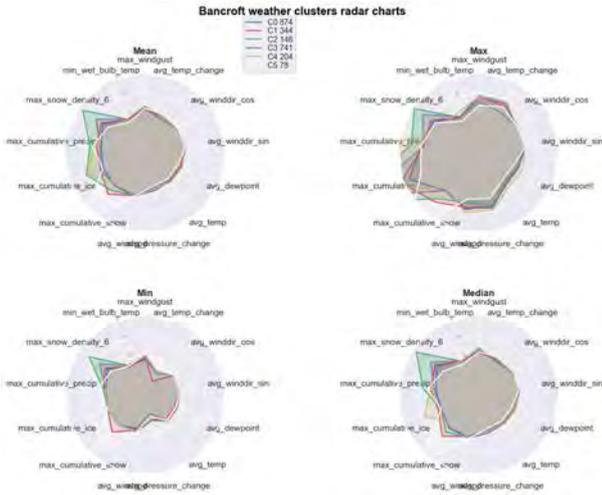


Abbildung 9 Beispielhafte Visualisierung der Cluster der Wetterdaten (KMeans, 2. Schritt der kaskadierten Methode) in Radardiagrammen

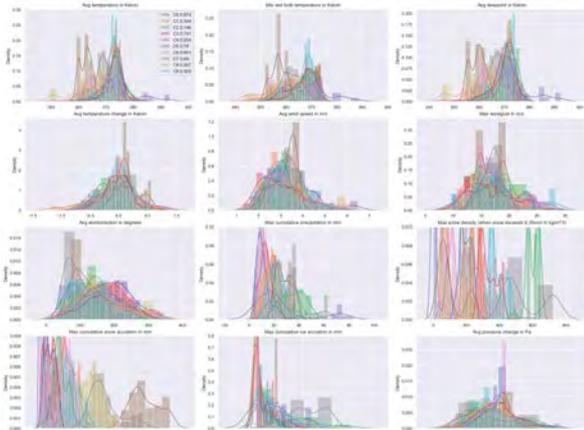


Abbildung 10 Beispielhafte Visualisierung der Cluster der (2. Schritt der kaskadierten Methode) in Distogrammen

3.2.3 PCA Clustering Model

Die in Kapitel 2.3.1 eingeführte PCA nach Zhou 2021 oder Abdi und Williams 2010 wird genutzt, um die hohe Dimensionalität der Wetterdatensätze zu reduzieren, indem die große Anzahl der Wetterparameter in ein kleineres Set von Hauptkomponenten transformiert wird. Der größte Teil des Informationsgehalts der Parameter bleibt erhalten, der allerdings nur durch Aufschlüsselung der Hauptkomponenten für Menschen interpretierbar wird (für Details sei auf Zhou 2021 oder Abdi und Williams 2010 verwiesen). Für die PCA wird eine Anzahl von drei Hauptkomponenten gewählt, damit das Clusterergebnis neben den bisher aufgeführten Visualisierungen auch in

einem dreidimensionalen Raum dargestellt und visuell ausgewertet werden kann. Zudem kann die Zusammensetzung der Hauptkomponenten untersucht werden, um die Relevanz der ursprünglichen Wetterparameter für die einzelnen Hauptkomponenten selbst zu untersuchen und so schlussendlich auch die Relevanz einzelner Wetterparameter im Allgemeinen für die Analyse evaluieren zu können. Dieses Wissen wurde iterativ zur Verbesserung der Modelle genutzt. Die Ergebnisse der PCA-Methode werden zusätzlich zu den Visualisierungen in Radardiagrammen (Abb. 11) und Distogrammen (Abb. 12) zur Bewertung der Clusterergebnisse auch in einem dreidimensionalen Graphen visualisiert. Zusätzlich werden die Varianzpfleile der Eigenvektoren der Parameter im 3D-Plot angegeben (Abb. 13 und 14). Diese Auswertung unterstützt iterativ dabei, die Daten und Ergebnisse sowie die Zusammenhänge der Parameter und deren Relevanz besser zu verstehen und die Clusteranalyse generell zu optimieren. Die PCA-Methode ermöglicht das Betrachten neuer Aspekte bei der Clusteranalyse, wie bspw. die Verteilung im dreidimensionalen Raum und insbesondere das Verständnis der Wetterextreme (die orangenen Datenpunkte / Cluster 3 in Abb. 14). Allerdings müssen die Hauptkomponenten zur Ergebnisbetrachtung und zur Definition der WEP in die ursprünglichen Wetterparameter zurückgeschlüsselt werden.

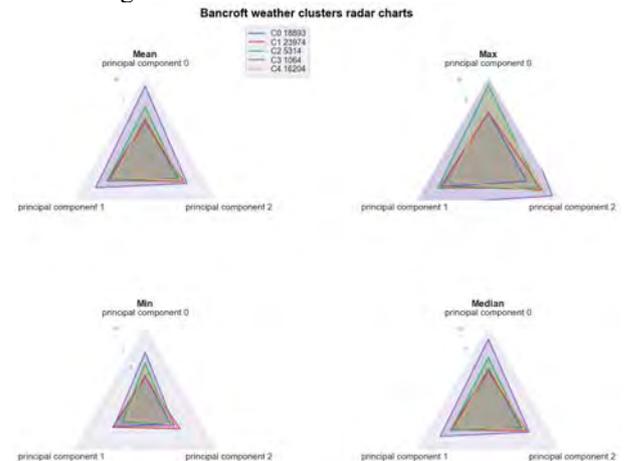


Abbildung 11 Beispielhafte Visualisierung der Cluster der Wetterdaten (KMeans, reguläre PCA-Methode) in Radardiagrammen

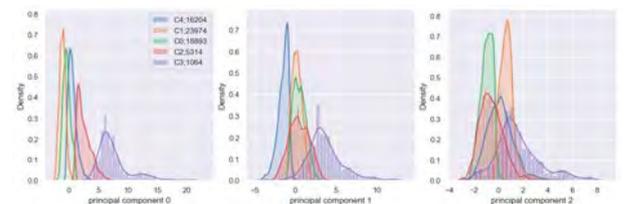


Abbildung 12 Beispielhafte Visualisierung der Cluster der Wetterdaten (KMeans, reguläre PCA-Methode) in Distogrammen



Abbildung 13 Varianzpfleile der regulären PCA-Methode im dreidimensionalen Raum

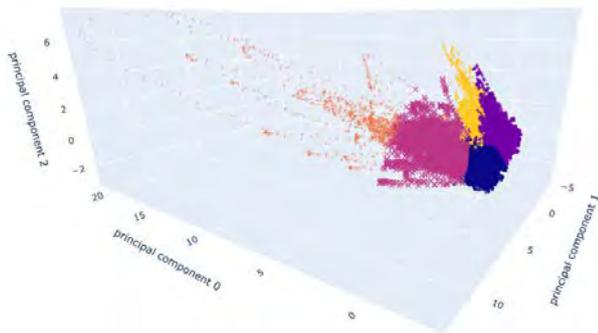


Abbildung 14 Beispielhafte Visualisierung der Cluster der Wetterdaten (KMeans, reguläre PCA-Methode) im dreidimensionalen Raum

3.2.4 Definition der Wetterereignisprofile

Zur Definition der WEP werden alle Datenpunkte unter Betrachtung des genauen Datums und der genauen Uhrzeit, zu der sie eingetreten sind, in Verbindung mit dem zugewiesenen Clusterlabel betrachtet. Zeitlich benachbarte Datenpunkte, die dem gleichen Wettercluster angehören, werden dem gleichen Wetterereignis zugeordnet. Die Nachbarschaft hat eine Toleranz von drei Nachbarn / Stunden. Unter Betrachtung der zeitlichen Dauer und der Ergebnis- und Parametervisualisierung der Wetterereignisse können diese aus meteorologischer Sicht qualitativ beschrieben, definiert und benannt werden. Zur qualitativen Messung der Genauigkeit werden die generierten WEP mit konkreten historischen Wetterereignissen verglichen, die von der kanadischen Regierung für die jeweiligen Regionen erfasst und veröffentlicht wurden (Government of Canada 2022). Insbesondere ist relevant, ob die genannten Wetterparameter und ihre Intensität, die geographische Einordnung, sowie die zeitliche Abgrenzung übereinstimmen. Die Genauigkeit der WEP werden darin gemessen, wie viel Prozent der verglichenen Wetterereignisse übereinstimmen. Die Qualität der WEP gibt an, wie hoch der Detailgrad der WEP ist und wie gut die Modelle die Wetterparameter ein- und voneinander abgrenzen können, um das WEP zu beschreiben. Dieser Schritt wird für alle durchgeführten Methoden, unter Implementierung aller genannten Clustering Algorithmen durchgeführt. Das iterativ erarbeitete und evaluierte Artefakt, das in diesem Kapitel beschrieben ist, wird auf alle in Abb. 1 dargestellten Regionen unter sukzessiver Implementierung der vorgestellten Clustering Algorithmen angewandt. Diese ergeben die identifizierten WEP als Kerneergebnisse dieser Arbeit (Tabelle 4), welcher nach Aggregation zeitlicher benachbarter Datenpunkte die Wetterereignisse in Tabelle 5 (beispielhaft für die GMM auf der Region Bancroft) ergeben. Beispielhaft wird an dieser Stelle das WEP des Clusters 2 des kaskadierten KMeans Clustermodells erläutert, um eine beispielhafte Ausführung der durch die vorgestellten Modelle generierten Ergebnisse als Kerneergebnisse dieser Arbeit zu geben. Dieses stellt Schneestürme mit viel Niederschlag dar, welches mit 270 Datenpunkten 0,41% der Gesamtdatenmenge ausmacht. Dieses WEP ist geprägt durch sehr niedrige Temperaturen unter dem Gefrierpunkt, starke Südostwinde bis zu 6 m/s, starke Windböen mit bis zu 23 m/s, hohen Niederschlag mit bis zu 40

mm, niedrige Schneedichte mit 0 bis 200 kg/m³ und einer sehr hohen kumulative Schneedecke mit bis zu 500 mm. Eine genauere Analyse und die Ergebnisauswertung erfolgen im folgenden Kapitel.

4 ERGEBNISAUSWERTUNG UND DISKUSSION

4.1 Untersuchung und Evaluation der Definition von WEP

Über die Zusammensetzung und Ausprägungen der Parameter eines Clusters lassen sich Charakteristiken für einzelne WEP beschreiben. Mittels der Clusteranalyse konnte somit über die WEP und deren Charakteristiken ein konkretes Wetterereignis sowie dessen Zeitraum und Dauer bestimmt werden. In diesem Kapitel wird beispielhaft der Fokus auf die Region Bancroft gelegt. Diese befindet sich geographisch mittig in Ontario, mit einigem Abstand zu den sehr kalten Gebieten im Norden und mit genügend Abstand zu den Gebieten im Süden, wo der sogenannte ‚Lake effect‘ das Klima maßgeblich beeinflusst (Hjelmfelt 1990). Für die Region Bancroft wurden unter Anwendung der beschriebenen Modelle und Methoden die in Tabelle 4 identifizierten WEP für die Region Bancroft hier beispielhaft die GMM tabellarisch dargestellten WEP identifiziert. Dabei wurden die erzeugten Radardiagramme, Distogramme und Clusterstatistiken der jeweiligen Algorithmen und Methoden manuell qualitativ ausgewertet und mithilfe der Literatur und Ergebnissen der Literaturanalyse aus Kapitel 2 qualitativ benannt und charakterisiert. Es fällt auf, dass sich gleiche WEP, identifiziert durch unterschiedliche Algorithmen, in der Anzahl der enthaltenen Datenpunkte, der Dauer und der Ausprägung der Parameter ähneln. Zudem wurden durch die regulären Clustering Methoden allgemeinere Wetterereignisse identifiziert, wie generell Schneefall, warme Tage oder windige Wetterereignisse. Mittels der kaskadierten Methode konnten detaillierte Wetterextreme gefunden werden. Die PCA-Methode wurde zur Verdeutlichung der Verteilung der Datenpunkte in einem dreidimensionalen Raum genutzt, eine Aufschlüsselung der Hauptkomponenten zur Detailanalyse ist aus Zeitgründen nicht erfolgt. Das größere Cluster, nach dem ersten Clustering Schritt der kaskadierten Methoden, wird als gewöhnliches Wetterevent mit allgemein moderat ausgeprägten Parametern betrachtet. Es fällt auf, dass die verschiedenen Algorithmen grundsätzlich ähnliche bis gleiche WEP erkennen. Das in Kapitel 4.3.4 erläuterte WEP ‚Schneesturm mit viel Niederschlag‘ wird mit ähnlichen Charakteristiken durch das kaskadierte KMeans Modell und durch das kaskadierte GMM-Modell erkannt, ebenso wie die WEP ‚Moderater Schneefall‘, ‚Starker Schneefall mit viel liegendem Schnee‘, ‚Frontdurchlauf / Langanhaltender gefrierender Regen‘ und ‚Schneesturm mit sehr niedrigen Temperaturen‘. Die WEP ‚Sturm mit gefrierendem Regen‘ und ‚Milder Schneefall‘ wurden von allen kaskadierten Modellen erkannt. Durch die regulären Clustermodelle wurden ebenfalls ähnliche WEP identifiziert.

Tabelle 4 Tabellarische Auflistung der identifizierten WEP für die Region Bancroft für die GMM

GMM Reguläres Clustering	Ereignis Name	Ereignis Beschreibung	GMM Kaskadiertes Clustering	Ereignis Name	Ereignis Beschreibung
Cluster 0 20633 Datenpunkte 31,53% der Gesamt-menge	Windiger und warmer Sturm	<ul style="list-style-type: none"> Hohe Temperaturen bis zu 298°K Moderater bis starker Wind (2 bis 4 m/s) Moderate bis starke Windböen (12 bis 22 m/s) Moderater Niederschlag (um die 20 mm) Kein Schnee oder Eis 	Cluster 0 2920 Datenpunkte 4,46% der Gesamt-menge	Moderater Schneefall	<ul style="list-style-type: none"> Niedrige Temperaturen nur wenig höher als 0°C Hohe Schneedichte (um die 200 kg/m³) Moderate kumulative Schneedecke (200 bis 400 mm) Wenig Niederschlag (10 bis 20 mm) Mittelstarke Südwestwinde (um 2,5 m/s)
Cluster 1 10133 Datenpunkte 15,48% der Gesamt-menge	Gewöhnliches Ereignis	<ul style="list-style-type: none"> Moderate Ausprägung der meisten Parameter Kein Schnee oder Eis 	Cluster 1 1040 Datenpunkte 1,59% der Gesamt-menge	Starker Schneefall mit viel liegendem Schnee	<ul style="list-style-type: none"> Niedrige Temperaturen nur wenig höher als 0°C Moderater bis hoher Niederschlag (mit 20 bis 40 mm) Moderate Winde um die die 2,5 m/s) Hohe Schneedichte (um die 350 bis 500 kg/m³) Hohe kumulative Schneedecke (200 bis 400 mm) Sehr hohe kumulative Eisdecke (bis zu 8 mm)
Cluster 2 10551 Datenpunkte 16,12% der Gesamt-menge	Gewöhnliches Ereignis	<ul style="list-style-type: none"> Moderate Ausprägung der meisten Parameter Kein Schnee oder Eis 	Cluster 2 430 Datenpunkte 0,66% der Gesamt-menge	Sturm mit gefrierendem Regen / Schwere Schnee- und Eissturm	<ul style="list-style-type: none"> Niedrige Temperaturen nur wenig um die 0°C Starke Ostwinde (mit bis zu 4,5 m/s) Starke Windböen (mit bis zu 20 m/s) Hoher Niederschlag (mit bis zu 60 mm) Hohe Schneedichte (bis 700 kg/m³) Hohe kumulative Schneedecke (bis 700 mm) Hohe kumulative Eisdecke (bis zu 8 mm) Sehr hohe negative Änderungen des Luftdrucks (bis zu 100 Pa)
Cluster 3 9578 Datenpunkte 14,63% der Gesamt-menge	Gewöhnliches Ereignis	<ul style="list-style-type: none"> Moderate Ausprägung der meisten Parameter Kein Schnee oder Eis 	Cluster 3 2630 Datenpunkte 4,02% der Gesamt-menge	Moderater Regen mit liegendem Schnee	<ul style="list-style-type: none"> Niedrige Temperaturen nur wenig höher als 0°C Keine Schneedichte Niedrige kumulative Schneedecke (bis 50 mm) Wenig Niederschlag (mit bis zu 15 mm) Schwache Westwinde (um die 3 m/s) Niedrige kumulative Schneedecke (bis zu 100 mm) Moderate kumulative Eisdecke (bis zu 2 mm)
Cluster 4 3900 Datenpunkte 5,96% der Gesamt-menge	Moderater Schneesturm	<ul style="list-style-type: none"> Niedrige Temperaturen wenig unter 0°C Moderate Winde bis zu 4 m/s Hohe Windböenstärke bis zu 23,5 m/s Niedriger bis moderater Niederschlag (10 bis 22 mm) Hohe Schneedichte um die 300 bis 450 kg/m³ Moderate kumulative Schneedecke (200 bis 400 mm) 	Cluster 4 261 Datenpunkte 4,02% der Gesamt-menge	Schneesturm mit viel Niederschlag	<ul style="list-style-type: none"> Sehr niedrige Temperaturen unter 0°C Starke Südostwinde (bis zu 4 m/s) Starke Windböen (mit bis zu 25 m/s) Hoher Niederschlag (mit bis zu 40 mm) Niedrige Schneedichte (0 bis 200 kg/m³) Sehr hohe kumulative Schneedecke (bis zu 900 mm) Moderate kumulative Eisdecke (bis zu 2 mm)
Cluster 5 7716 Datenpunkte 11,79% der Gesamt-menge	Kälte mit niedrigem Niederschlag und Wind unter dem Gefrierpunkt	<ul style="list-style-type: none"> Sehr niedrige Temperaturen unter 0°C Schwache Westwinde (mit 2 bis 4 m/s) Schwache Windböen (um die 15 m/s) Niedriger Niederschlag (5 bis 10 mm) Niedrige Schneedichte (bis zu 200 kg/m³) Niedrige kumulative Schneedecke (bis zu 100 mm) 	Cluster 5 4497 Datenpunkte 6,87% der Gesamt-menge	Milder Schneefall	<ul style="list-style-type: none"> Niedrige Temperaturen um die 0°C Mittelhoh bis niedrige Schneedichte (um die 100 kg/m³) Mittelstarke Westwinde (2 bis 4 m/s) Mittelstarke Windböen (mit bis zu 22 m/s) Wenig Niederschlag (5 bis 10 mm) Moderate kumulative Schneedecke (bis zu 180 mm)
Cluster 6 1528 Datenpunkte 2,33% der Gesamt-menge	Sturm mit gefrierendem Regen / Schwere Schnee- und Eissturm	<ul style="list-style-type: none"> Sehr niedrige Temperaturen unter 0°C Starke Südostwinde (mit bis zu 4,5 m/s) Starke Windböen (mit bis zu 25 m/s) Hoher Niederschlag (mit bis zu 45 mm) Hohe Schneedichte (500 bis 700 kg/m³) Hohe kumulative Schneedecke (400 bis 700 mm) Hohe kumulative Eisdecke (bis zu 8 mm) Sehr hohe Änderungen des Luftdrucks (bis zu 100 Pa) 	Cluster 6 934 Datenpunkte 1,43% der Gesamt-menge	Frontdurchlauf / Langanhaltender gefrierender Regen	<ul style="list-style-type: none"> Mittlere Temperaturen (von 270°K bis 280°K) Starke Ostwinde (mit bis zu 6 m/s) Starke Windböen (mit bis zu 28 m/s) Hoher Niederschlag (bis zu 35 mm) Moderate kumulative Schneedecke (bis zu 400 mm) Hohe kumulative Eisdecke (bis zu 4 mm) Lange Dauer (bis zu mehreren Tagen)
Cluster 7 1306 Datenpunkte 2,00% der Gesamt-menge	Gefrierender Regen ohne Wind und viel Eis und Schnee	<ul style="list-style-type: none"> Niedrige Temperaturen um die 0°C Starke Südostwinde (mit bis zu 4,5 m/s) Starke Windböen (mit bis zu 25 m/s) Moderater Niederschlag (mit bis zu 20 mm) Moderate Schneedichte (bis zu 300 kg/m³) Niedrige kumulative Schneedecke (bis zu 100 mm) Moderate kumulative Eisdecke (bis zu 2,5 mm) Sehr hohe Änderungen des Luftdrucks (bis zu 100 Pa) 			

Tabelle 4 Auswertung und Vergleich der vorgestellten Clusteralgorithmen und Methoden anhand des Silhouette Scores, der Performanz, der Genauigkeit und der Qualität der generierten WEP

	Silhouette Score	Performanz in Durchlaufzeit, CPU-Ausführungszeit und genutzte Rechenressourcen			Anzahl Cluster	Genauigkeit	Qualität / Detail
		Execution time	CPU-execution time	Genutzte Rechenressourcen (physikalischer Speicher)			
KMeans Regular method	0,195426366	1.923,3 sec	2.996,6 sec	743.153,66 MB	5	7/10	Moderat
KMeans Cascaded method	0,676872128	2.207,19 sec	3.447,26 sec	875.134,98 MB	10	6/10	Hoch
KMeans PCA method	0,339207292	1.971,96 sec	3.031,43 sec	442.576,896 MB	-	-	-
KMeans total	0,403835262	6.102,45 sec	9.475,28 sec	2.060.865,54 MB	5 to 10	6.5/10	Hoch
HAC Regular method	0,158407099	70.046,84 sec	28.367,62 sec	2.486.848	4	5/10	Moderat
HAC Cascaded method	0,579056567	91.060,9 sec	39.714,67 sec	2.932.902,40	3	2/10	Niedrig
HAC PCA method	0,336365936	84.056,21 sec	40.565,7 sec	2.784.217,60	-	-	-
HAC total	0,357943201	245.163,95 sec	108.647,99 sec	5.203.968,749 MB	3 to 4	3.5/10	Niedrig bis moderat
GMM Regular method	0,057888212	3.184,08 sec	10.144,34 sec	890.085,38	8	7/10	Moderat
GMM Cascaded method	0,312381864	3.728,01 sec	12.232,32 sec	921.899,01	7	7/10	Sehr hoch
GMM PCA method	0,16188301	1.606,51 sec	4.070,52 sec	498.655,23	-	-	-
GMM total	0,177384362	8.518,60 sec	26.447,18 sec	2.310.639,62 MB	7 to 8	7/10	Sehr hoch
Regular method	0,49252297	4.185,9 sec	7.106,79 sec	965.820,42	2	7/10	Low
PCA method	0,68362696	2.816,01 sec	4.596,45 sec	725.385,22	-	-	-
DBSCAN	0,58807496	7.120,20 sec	11.816,37 sec	1.691.205,63 MB	2	7/10	Low

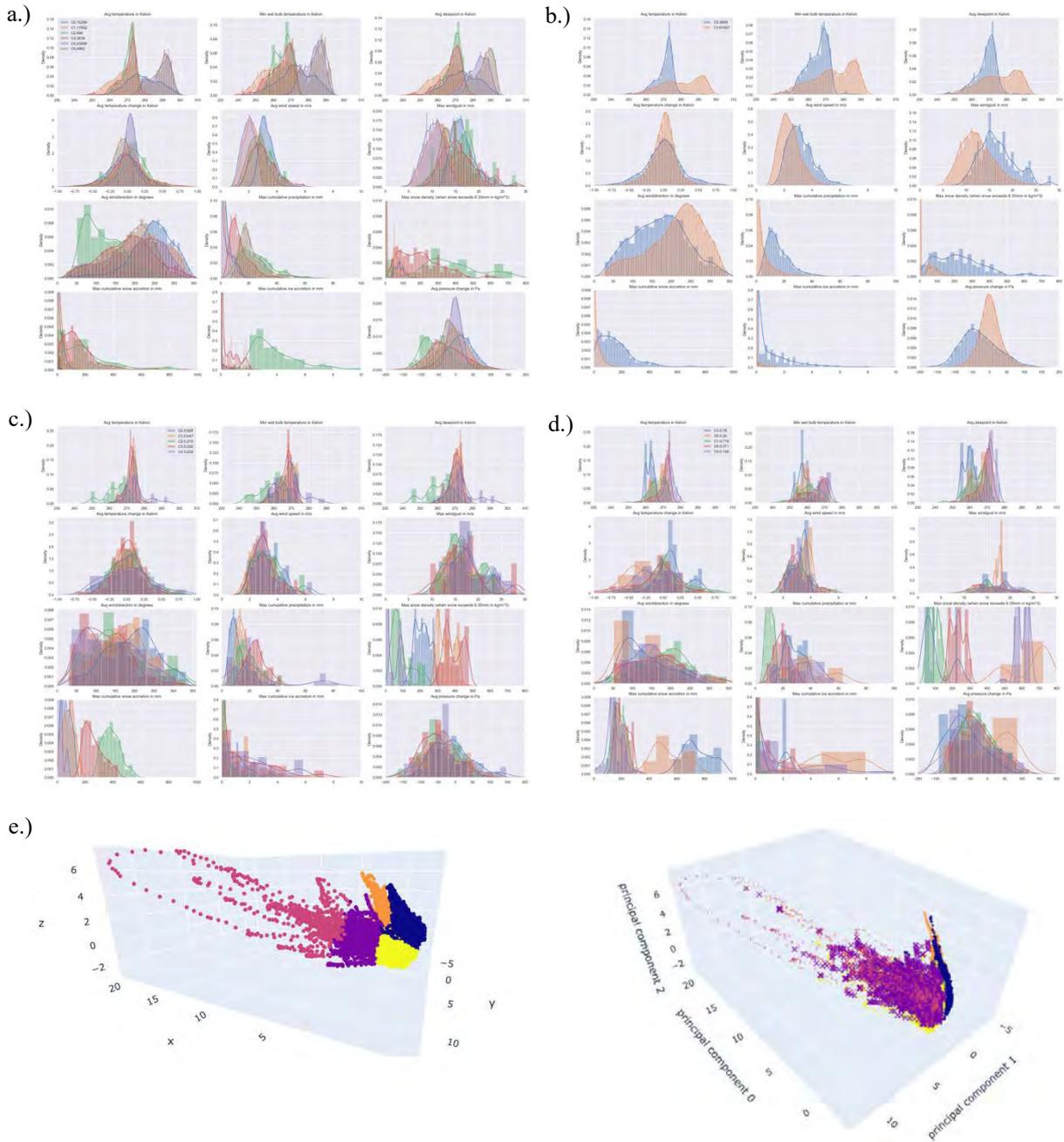


Abbildung 15 Ausprägungen der Wetterparameter der WEP für KMeans für die Region Bancroft
 a.) reguläre Methode b.) erster Schritt der kaskadierten Methode c.) kaskadierten Methode (Cluster 0 bis 4)
 d.) kaskadierten Methode (Cluster 5 bis 9) e.) PCA-Methode

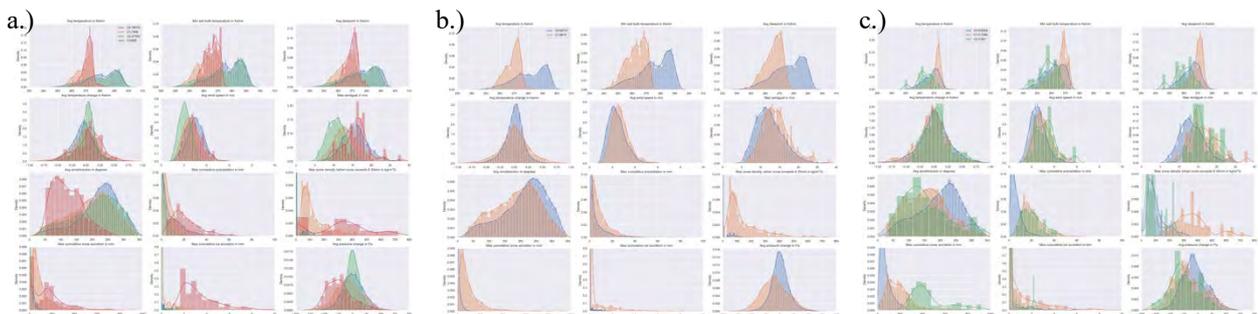


Abbildung 16 Ausprägungen der Wetterparameter der WEP für HAC für die Region Bancroft
 a.) reguläre Methode b.) erster Schritt der kaskadierten Methode c.) kaskadierten Methode

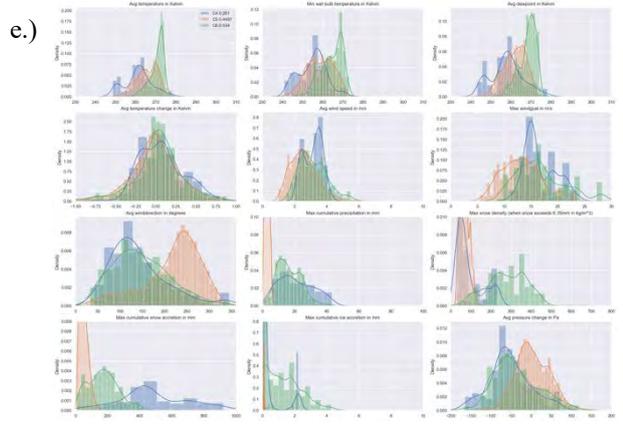
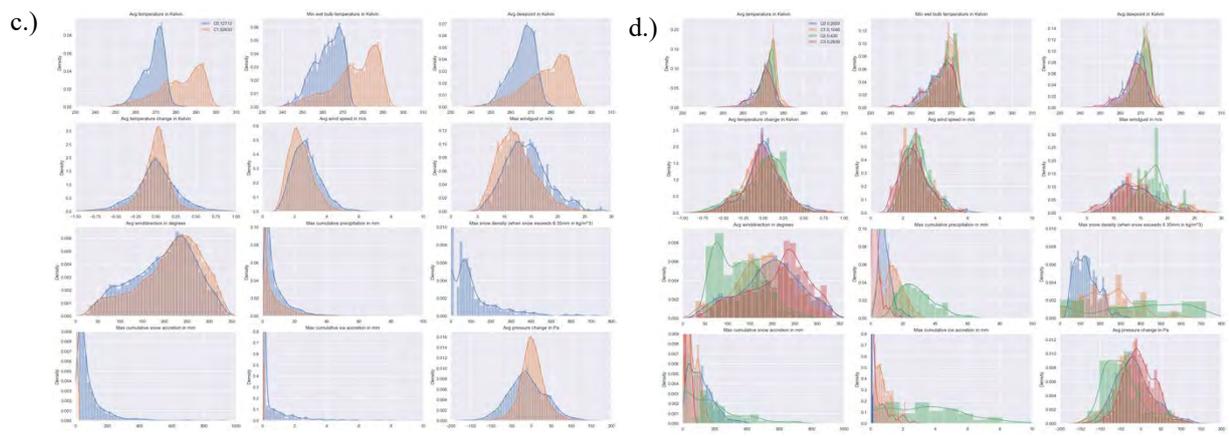
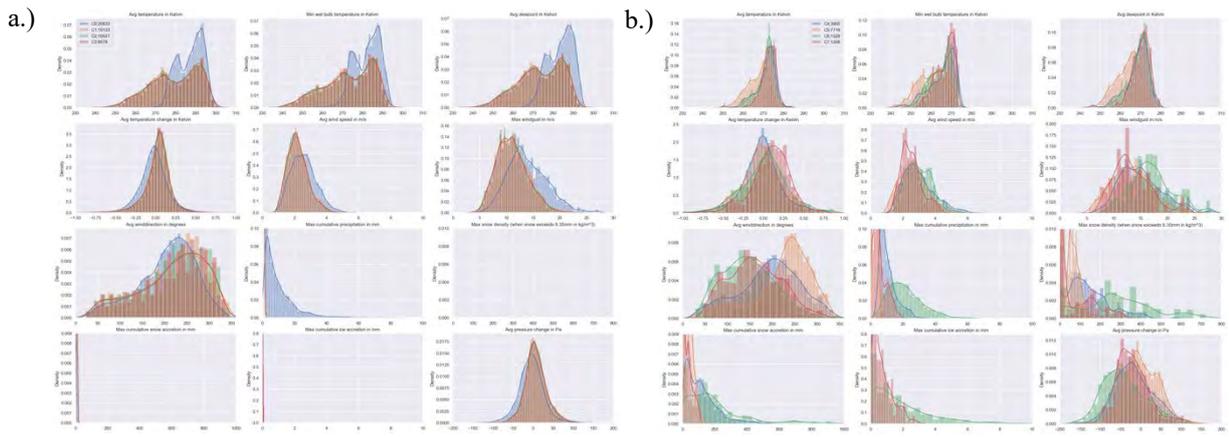
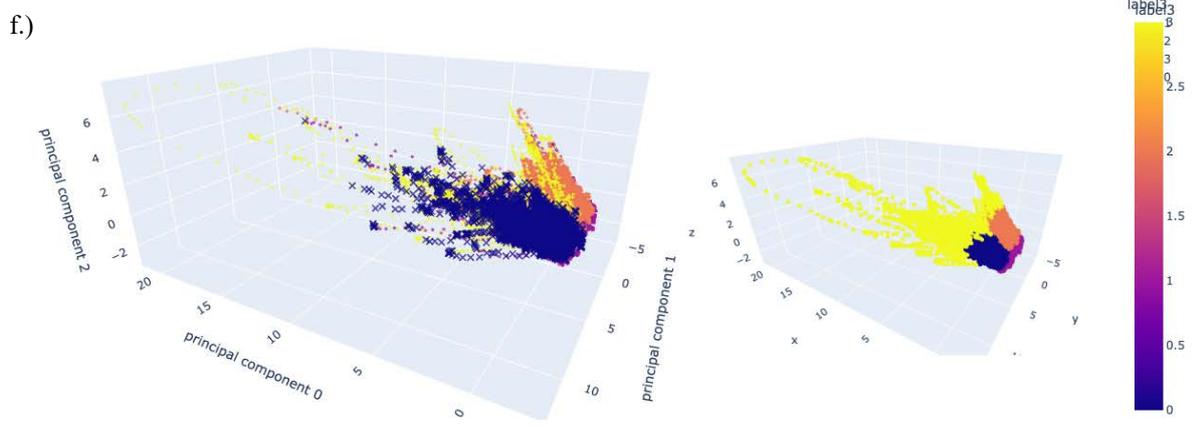


Abbildung 17 Ausprägungen der Wetterparameter der WEP für GMM für die Region Bancroft
 a.) reguläre Methode (Cluster 0 bis 3)
 b.) reguläre Methode (Cluster 4 bis 7)
 c.) erster Schritt der kaskadierten Methode
 d.) kaskadierte Methode (Cluster 0 bis 3)
 e.) kaskadierte Methode (Cluster 4 bis 6)
 f.) PCA-Methode



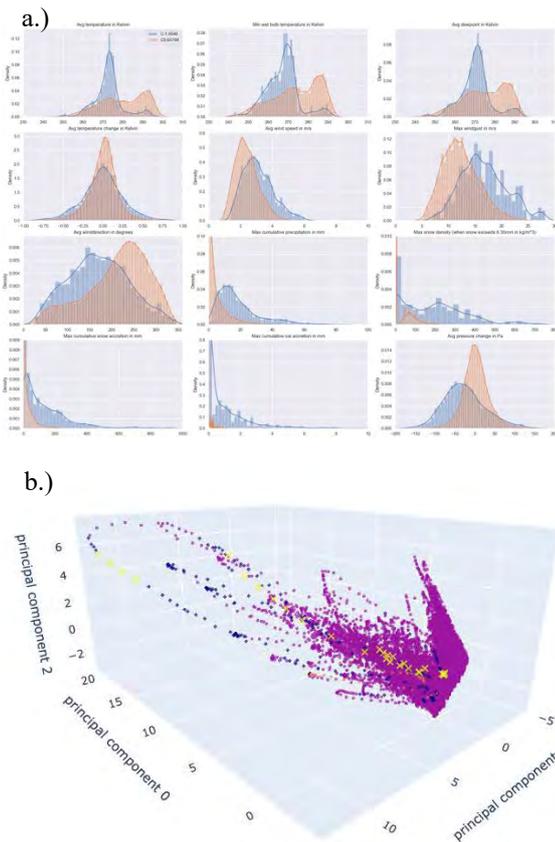


Abbildung 18 Ausprägungen der Wetterparameter der WEP für DBSCAN für die Region Bancroft
 a.) reguläre Methode b.) PCA-Methode

Generell ist der Silhouette Score für die kaskadierten Methoden sehr viel höher als für die regulären und für die PCA-Methoden. Die niedrigsten Durchlaufzeiten und damit die höchste Performanz wurden durch die KMeans und DBSCAN Modelle erreicht. Diese besitzen, wie im Theorieteil bereits deutlich wurde, auch die niedrigste Komplexität. Zudem wurden durch diese Modelle auch am wenigsten Rechenressourcen gebraucht. Unter der Genauigkeit der Clusterergebnisse wird verstanden, wie richtig die Clusterergebnisse verglichen mit reellen historisch dokumentierten Wetterereignissen sind. Mit der Qualität der Ergebnisse ist gemeint, wie detailliert ein WEP durch ein Modell beschrieben werden kann. Für beide Kriterien erzielen KMeans und GMM die besten Ergebnisse. Zudem ist zu erkennen, dass GMM trotz sehr niedrigem Silhouette Score qualitativ sehr hochwertige Clusterergebnisse erzielt. HAC schneidet im Vergleich am schlechtesten ab. Nicht nur sind die Clusterergebnisse qualitativ nicht hochwertig, auch die verbrauchten Rechenressourcen und die Durchlaufzeiten sind unverhältnismäßig hoch.

4.3 Bewertung der Ergebnisse

Die vorgestellten Algorithmen und Methoden wurden erfolgreich eingesetzt, um WEP zu finden und zu definieren. Die WEP konnten erfolgreich charakterisiert und anhand relevanter Parameter und deren Ausprägungen, der Dauer der Wetterereignisse eines WEP und der Anzahl der zugehörigen Datenpunkte charakterisiert werden. Anschließend wurden über Aggregation zeitlicher Nachbarn mit gleichem Label erfolgreich konkrete Wetterereignisse identifiziert. Insbesondere WEP, welche im Frühling und Winter stattfinden und über starke Ausprägungen der Wind- und Niederschlagsparameter charakterisiert sind, konnten erfolgreich identifiziert werden. Einige Wetterereignisse wie Fluten oder Dürren konnten nicht erfasst werden, obwohl diese laut Literatur stattgefunden haben. Auch wurden wenige bis keine Extremwetterereignisse im Sommer identifiziert. Dies kann daran liegen, dass im kanadischen Sommer generell gemäßigteres Wetter stattfindet. Da aber in der Literatur von Stürmen, Fluten, Dürren und Hitzewellen im Sommer berichtet wurde, ist auch möglich, dass diese Wetterereignisse vor allem durch Parameter identifiziert werden, welche nicht in den genutzten Datensätzen vorhanden sind. In der Literatur wird diesbezüglich insbesondere die Luft- und Bodenfeuchte diskutiert. In den Ergebnissen zu den WEP und Wetterereignissen ist zu erkennen, dass einige Wetterereignisse häufig miteinander korrelieren und gemeinsam nacheinander auftreten. Es ist zu analysieren, inwiefern diese Ereignisse zusammenhängen. Denkbar ist, dass einige WEP verschiedene Stufen des gleichen Ereignisses darstellen, bspw. drei Phasen eines Sturmes. Insbesondere ist die Dauer der Ereignisse relevant, welche durch die implementierten Modelle gut analysiert werden kann. Zum Identifizieren von extremen WEP wurde erkannt, dass der erste Schritt der kaskadierten Clustermethode den Datensatz in zwei Cluster teilen sollte, wobei hingegen der größere Cluster ca. 90% der Datenpunkte besitzt. Insgesamt wurden KMeans und GMM als die Algorithmen mit den qualitativ hochwertigsten Ergebnissen zum Definieren von WEP erkannt, wobei KMeans eine geringfügig bessere Performanz und einen höheren Silhouette Score aufweist. HAC ist für den diskutierten Anwendungsfall aufgrund der qualitativ minderwertigen Ergebnisse und der sehr schlechten Performanz irrelevant. DBSCAN ist grundsätzlich zum Identifizieren spezifischer WEP ungeeignet, ist aber hervorragend dazu in der Lage, Wetterextreme von gewöhnlichen Wetterereignissen zu separieren.

4.4 Beantwortung der Forschungsfragen

Im Weiteren wird die in Kapitel 1.2 gestellte Forschungsfrage beantwortet und die Antworten diskutiert. *Welche ML-Algorithmen eignen sich für das Clustering von Wetterdaten zum Definieren von WEP? Und welche Algorithmen und Methoden der Clusteranalyse eignen sich am besten zum Definieren von WEP, gemessen anhand der identifizierten Kriterien und welche Möglichkeiten gibt es, diese weiter zu optimieren?*

Für das Clustering von Wetterdaten eignen sich insbesondere der KMeans Algorithmus und die Gauß'schen Mischmodelle. Zum Identifizieren grober Wetterereignisse ist die vorgestellte reguläre Methode geeignet, zum Identifizieren von spezifischen Wetterereignisse oder Wetterextremen hingegen die kaskadierte Methode. Die Möglichkeiten zur weiteren Optimierung der vorgestellten und diskutierten Modelle werden in Kapitel 6 und 7 diskutiert. Als Evaluationskriterien werden die folgenden genutzt. Die Performanz, gemessen über Durchlaufzeiten, CPU-Ausführungszeiten und genutzter RAM, ist geeignet zum Bewerten der eingesetzten Algorithmen. Auch die qualitative Evaluierung über Qualität und Genauigkeit der Ergebnisse ist von hoher Relevanz und geeignet zum Bewerten der Modelle. Hierzu sollten eine erweiterte Referenztable oder manuell gelabelte Datensätze erstellt werden, um eine qualitativ hochwertige Auswertung zu ermöglichen. Der Silhouette Score als Evaluationsmatrix zum Bewerten der Aufteilung der Datenpunkte in die Cluster ist nicht optimal geeignet. Es sollten weitere Evaluationsmetriken (wie der Jaccard Coefficient, Fowlkes Index, Mallows Index, Rand Index, Davies-Bouldin Index oder Dunn Index) verprobt und evaluiert werden, um zu analysieren, welche Evaluationsmatrix für diesen Anwendungsfall am besten geeignet ist. Diese Anforderung ist insbesondere deswegen notwendig, da die qualitativen Clusterergebnisse für die GMM-Modelle eine sehr hohe Qualität aufweisen, der Silhouette Score für diese Modelle aber sehr niedrig ist.

4.5 Rückbezug zur Theorie

Der in Grabbe et al. 2014 und Pooja et al. 2020 genutzte EM-Algorithmus konnte im Rahmen dieser Praxisarbeit in Form der Gauß'schen Mischmodelle erfolgreich eingesetzt werden, um Wetterdaten zu clustern und WEP zu definieren. Der Algorithmus konnte erfolgreich auf die diskutierten Methoden (kaskadiert und PCA) angewandt werden. Korrelationsanalysen der Wettercluster wie sie von Grabbe et al. 2014 angemerkt und durchgeführt wurden, werden im Zukunftsausblick dieser Arbeit diskutiert und sind grundsätzlich möglich. Wie auch bereits Grabbe et al. 2014 demonstrieren, werden durch Clustering von Wetterdaten zumeist ein bis wenige Cluster generiert, die einen Großteil der Datenpunkte enthalten und welche unauffälliges und gewöhnliches Wetter enthalten und mehrere Cluster, die deutlich weniger Datenpunkte enthalten, dafür aber auffälliges Wetter oder Extreme aufweisen. Pooja et al. 2020 nutzen zur Evaluierung der Clusterergebnisse eine ähnliche Methode wie diese Arbeit (Genauigkeit der Feature Selektion, die Genauigkeit der Cluster, die False Positive Rate und die Durchlaufzeit), es sind aber in zukünftiger Forschung weitere und optimalere Evaluationsmethoden zu untersuchen, wie im vorherigen Kapitel diskutiert wurde. Die Erweiterung der Clustermethodik durch bspw. Klassifikation wie in Pooja et al. 2020 oder De Lima et al. 2013 ist in weitergehender Forschung und zur Anwendung in der Praxis sinnvoll. Auch sollten kaskadierte Zwei-Schritt Methodiken untersucht werden, die mehrere Modelle (wie DBSCAN und GMM oder Clustering und Klassifikation) kombinieren.

In dieser Arbeit wurde diesbezüglich die bereits von Xu et al. 2015 vorgestellte kaskadierte Methode untersucht, welche insbesondere erfolgreich zum Definieren von Wetterextremen angewandt wurde. Die in Ferstl et al. 2017 und De Lima et al. 2013 genutzten HAC-Modelle erwiesen sich für den konkreten Anwendungsfall der vorgelegten Arbeit als nicht effizient.

De Lima et al. 2013 kombinieren in ihrer Arbeit zudem Dichte-basierte Modelle (wie DBSCAN) mit Klassifikation. In weiterer Forschung sollte der kombinierte Ansatz über DBSCAN auch weiter untersucht werden. Die in Kapitel 2.1.4 diskutierten Wetterkategorien wie sie von Liljequist und Cehak 1984 sowie von Jahn 2015 beschrieben werden, stellten sich im Praxisteil der Arbeit als zu oberflächlich heraus. Es wurden insbesondere detailliertere Wetterereignisse erkannt, die zudem speziell auf die analysierte Region zugeschnitten sind. Mittels der vorgestellten Modelle zum Definieren von WEP können Listen von Wetterkategorien vereinfacht und detailliert für weitere spezifische Regionen der Erde generiert und untersucht werden. Diese können dann den im Theorie-teil dieser Arbeit diskutierten Wetterkategorien untergeordnet werden.

4.6 Herausforderungen und Limitationen

Die erzielten Ergebnisse dieser Arbeit werden durch einige Herausforderungen und Limitationen begrenzt. Diese werden in diesem Kapitel diskutiert. Im Rahmen der vorgelegten Arbeit war es aufgrund zeitlicher Einschränkungen und dem begrenzten Umfang nicht möglich, alle Daten der in Kapitel 4.1 vorgestellten Regionen Kanadas zu analysieren. Diese könnten jedoch wertvolle Erkenntnisse zum Thema liefern und sollten in zukünftiger Forschung ausgewertet werden. Zudem wurden in dieser Arbeit nur Wetterdaten analysiert, welche typisch für das Klima in Ontario und durch Kälte, Schnee, Eis und starke Winde gekennzeichnet sind. Eine erweiterte Anwendung und Auswertung der Modelle auf weitere Klimaregionen der Erde ist erforderlich, um die Modelle weiter zu evaluieren und zu optimieren. In weitergehenden Analysen sollte zudem ein vollständiges Spektrum an Wetterparametern genutzt werden, wie es in Kapitel 2.1.3 vorgestellt wurde. Insbesondere die im Praxisteil dieser Arbeit fehlenden Parameter Luftfeuchte, Bodenfeuchte und Sonneneinstrahlung könnten große Auswirkungen auf die Analyse haben und besonders relevant zum Identifizieren von Wetterereignissen wie Fluten, Dürren oder Hitzewellen sein. Auch die manuelle, qualitative Auswertung der Analyseergebnisse im Praxisteil dieser Arbeit erwies sich aufgrund fehlenden meteorologischen Fachwissens als Herausforderung. Die generierten Ergebnisse in Form von WEP sollten im Weiteren von meteorologischen Experten ausgewertet und charakterisiert werden. Zuletzt war im Umfang der Arbeit auch eine vollumfängliche Auswertung der PCA-Methode angedacht. Aufgrund der zeitlichen Beschränkung der Arbeit konnte diese allerdings nicht fertig umgesetzt werden.

4.7 Kritische Analyse der erzielten Ergebnisse

Vorteile der vorgestellten Modelle sind, dass diese insbesondere Wetterextreme im Detail charakterisieren und identifizieren können. Die vorgestellten Clusteranalysen können auf Grundlage der vorherrschenden Wetterparameter zu anderen Ergebnissen für verschiedene Klimaregionen der Erde führen. Obwohl in dieser Arbeit ein Schwerpunkt auf die Ergebnisse der kaskadierten Methode gelegt wurde, ist auch eine detailliertere Betrachtung der Clusterergebnisse der regulären Clusteringmethode interessant. So ist denkbar, diese nach Jahreszeiten zu sortieren, um noch aufschlussreichere Aussagen über diese Wetterereignisse treffen zu können. Bspw. wäre es denkbar, dass die größeren Cluster der regulären Methoden in gewöhnliche milde Ereignisse verschiedener Jahreszeiten gliederbar sind. Eine Analyse, aufgegliedert nach Jahreszeiten, ist in weitergehender Forschung notwendig. In der Literatur wird eine Vielzahl weiterer Cluster-Algorithmen genannt, welche aufgrund der geringeren Relevanz nicht weiter betrachtet wurden. Eine Anwendung und Evaluierung dieser Algorithmen auf den spezifischen Anwendungsfall ist in weiterer Forschung notwendig. Hier sind insbesondere Affinity Propagation, Mean-Shift, Ward Hierarchical Clustering, OPTICS und BIRCH (Scikit Learn 2023a) interessant. Bei der Betrachtung der Ergebnisse fällt zudem auf, dass die GMM, trotz ihres geringen Silhouette Scores, qualitativ sehr hochwertige und genaue Ergebnisse liefern. Es liegt daher nahe, dass der Silhouette Score nicht die optimale Evaluationsmetrik zur Bewertung der Clusterergebnisse von Wetterdaten ist. In weitergehender Forschung ist die Anwendung weiterer Evaluationsmetriken zu analysieren, um jene zu identifizieren, die für den Anwendungsfall des Clusters von Wetterdaten am geeignetsten sind. Interessant sind hierbei insbesondere der Jaccard Coefficient, der Fowlkes Index, der Mallows Index, der Rand Index, der Davies–Bouldin Index, der Dunn Index (Zhou 2021) und die von De Lima et al. 2013 vorgestellte Similarity Metric. Auch sollte in zukünftiger Forschung die Skalierbarkeit der Modelle in Bezug auf die geographische Größe der zu analysierenden Regionen untersucht werden, um festzustellen, welche räumliche Granularität am geeignetsten zur Definition von WEP ist. Dies ist umso mehr der Fall, da ähnliche WEP für verschiedene Regionen Ontarios identifiziert wurden. Zu-dem besteht die Notwendigkeit der Auswertung der Analyseergebnisse durch meteorologische Fachexperten, um komplexe Zusammenhänge der Wetterparameter und auch geographische Ursachen wie den nordamerikanischen Lake Effect adäquat auswerten zu können. Auch sollte der Ansatz des Clusterings mit weiteren Methoden zum Definieren von WEP detaillierter verglichen werden. Insbesondere die Klassifikation und Assoziationsregeln sind hierfür relevant. Nach der erfolgreichen Anwendung der kaskadierten Methode in der vorgelegten Arbeit, ist zudem zu untersuchen, welche weiteren Zwei-Schritt-Methoden zur Optimierung der Analyse genutzt werden können. Als Beispiel sei hier die bereits diskutierte Verbindung von Clustering und Klassifikation oder eine Verbindung

verschiedener Clusteringalgorithmen genannt. Die qualitativ hochwertigen Ergebnisse der regulären DBSCAN Methode könnten bspw. als erster Schritt des kaskadierten Modells genutzt werden, um die Stärken des Dichtebasierten Modells zur Identifikation von Ausreißern auszunutzen. In einem zweiten Schritt könnte dann KMeans oder GMM genutzt werden, welche besonders eine Stärke darin zeigen, qualitative und genaue Ergebnisse zur Identifizierung von Wetterextremen zu erzielen. Zuletzt muss auch der Einsatz der PCA-Methode vollumfänglich, auch unter Einsatz weiterer Algorithmen und Methoden, verprobt und evaluiert werden.

5 ZUSAMMENFASSUNG UND AUSBLICK

In der vorgelegten Arbeit wurden auf Grundlage der durchgeführten Literaturrecherche zur Clusteranalyse von Wetterdaten zum Definieren von WEP relevante Clusteringalgorithmen und Methoden, unter Betrachtung und Optimierung der in der Literatur identifizierten relevanten Faktoren, auf verschiedene Regionen Kanadas angewandt. Anschließend wurde mittels maschinellen Lernens von Clusteranalysen Profile für verschiedene Wetterereignisse zu identifiziert. Zur Einlösung der Zielsetzung wurde die Forschungsmethodik Design Science Research ergänzt durch iteratives Prototyping und die Kreuzvalidierung verwendet. Durch iterative Implementierung von regulären, kaskadierten und Hauptkomponentenanalyse-Modellen wurden erfolgreich WEP identifiziert. Dies geschah unter Anwendung der Clustering Algorithmen KMeans, hierarchisches agglomeratives Clustering, Gauß'sche Mischmodelle und DBSCAN. Geographisch eingeschränkt wurde die Analyse auf die Regionen Ontarios unter Betrachtung relevanter identifizierter Parameter und der zeitlichen Granularität der Wetterereignisse. Die verschiedenen eingesetzten Algorithmen und Methodiken wurden unter den Aspekten der Performanz, anhand einer Evaluationsmatrix sowie der Qualität und Genauigkeit der Ergebnisse verglichen. Gemessen wurden diese Kriterien in Form der Durchlaufzeiten, der CPU-Ausführungszeiten und der genutzten Rechenressourcen, des Silhouette Scores und einer manuellen, qualitativen Auswertung der Qualität und Genauigkeit der Modelle. So konnten Stärken und Schwächen der jeweiligen Modelle eruiert werden und das kaskadierte KMeans und das kaskadierte GMM-Modell als die geeignetsten für die Definition von extremen WEP identifiziert werden. Mit den regulären KMeans und den regulären GMM-Modellen konnten insbesondere gemäßigte und gewöhnliche Wetterereignisse identifiziert werden.

In weitergehender Arbeit mit Wetterdaten können die identifizierten WEP und die implementierten Modelle für eine Reihe von Anwendungsfällen in der Forschung und Praxis genutzt werden. Die WEP mit ihren Charakteristiken können bspw. in Anwendung auf Wettervorhersagen untersucht werden, um detaillierte Informationen (wie ausschlaggebende und gefährliche Ausprägungen bestimmter Parameter oder Handlungsempfehlungen bei

Wetterextremen) zu vorhergesagtem Wetter zu bekommen und diese Informationen in einem Dashboard als Warnung anzuzeigen. Weiterhin können bestehende WEP über eine Korrelationsanalyse mit weiteren Faktoren untersucht werden, um bspw. effiziente Maßnahmen gegen die Auswirkungen von Katastrophen und Schäden bei bestimmten Wetterereignissen zu untersuchen. Hierzu können Modelle für diverse WEP gebaut werden. Denkbar sind Modelle zur Vorhersage und zur Vorbeugung von Stromausfällen bei bestimmten Wetterereignissen (Eskandarpour und Khodaei 2016). Weiterhin kann in zukünftiger Forschung die Erkennung und Vorhersage der definierten WEP mittels Wetterdaten untersucht werden und die Effektivität von ML-Modellen zur Wettervorhersage mit physikalischen Methoden verglichen werden. Es besteht also eine Vielzahl von Anwendungsfällen, in denen die implementierten Artefakte dieser Arbeit, sowie die identifizierten WEP in zukünftiger Forschung genutzt werden können.

LITERATUR

- Abdi, H. / Williams, L.J. (2010): Principal component analysis, in: Wiley interdisciplinary re-views: computational statistics, 2(4), pp.433-459.
- Ackerman, S. / Knox, J. (2011): Meteorology. Jones & Bartlett Publishers.
- Akande, A. / Costa, A.C. / Mateu, J. / Henriques, R. (2017): Geospatial analysis of extreme weather events in Nigeria (1985–2015) using self-organizing maps, in: Advances in Meteorology, 2017.
- Ban, Z. / Liu, J. / Cao, L. (2018): Superpixel segmentation using Gaussian mixture model, in: IEEE Transactions on Image Processing, 27(8), pp.4105-4117.
- Bellman, R. (1957): Dynamic programming. Princeton University, NJ, Princeton University Press, New Jersey.
- Chandola, V. / Banerjee, A. / Kumar, V. (2009): Anomaly detection: A survey, in: ACM computing surveys (CSUR). Jul 30;41(3):1-58.
- Chu, X. / Ilyas, I.F. / Krishnan, S. / Wang, J. (2016): Data cleaning: Overview and emerging challenges, in: Proceedings of the 2016 international conference on management of data (pp. 2201-2206).
- Cui, M. (2020): Introduction to the k-means clustering algorithm based on the elbow method. Accounting, Auditing and Finance, 1(1), pp.5-8.
- Das, S. / Sun, X. (2014): Investigating the pattern of traffic crashes under rainy weather by association rules in data mining, in: Transportation Research Board 93rd Annual Meeting (No. 14-1540). Transportation Research Board Washington DC.
- de Lima, Glauston, R.T. / Stephan, S. (2013): A new classification approach for detecting severe weather patterns, in: Computers & geosciences 57 (2013): 158-165.
- ECMWF (2023a): ERA5: data documentation. URL: <https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation>, Abruf: 01.03.2023, 13:36 Uhr
- ECMWF (2023b): ERA5: reanalysis datasets for forecasts. URL: <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>, Abruf: 01.03.2023, 13:44 Uhr
- ECMWF (2023c): ERA5: data documentation parameter listings. URL: <https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#ERA5:datadocumentation-Parameterlistings>, Abruf: 01.03.2023, 13:59 Uhr
- Epstein, E.S. (1969): A scoring system for probability forecasts of ranked categories, in: Journal of Applied Meteorology (1962-1982), 8(6), pp.985-987.
- Eskandarpour, R. / Khodaei, A. (2016): Machine learning based power grid outage prediction in response to extreme events, in: IEEE Transactions on Power Systems, 32(4), pp.3315-3316.
- Fathi, M. / Haghi Kashani, M. / Jameii, S. M. / Mahdipour, E. (2022): Big Data Analytics in Weather Forecasting: A Systematic Review, in: Archives of Computational Methods in Engineering 29.2 (2022, Springer): 1247–1275
- Ferstl, F. / Kanzler, M. / Rautenhaus, M. / Westermann, R. (2017): Time-hierarchical clustering and visualization of weather forecast ensembles, in: IEEE transactions on visualization and computer graphics, 23(1), pp.831-840.
- Firdaus, S. / Uddin, M.A. (2015): A survey on clustering algorithms and complexity analysis, in: International Journal of Computer Science Issues (IJCSI), 12(2), p.62.
- Géron, A. (2019): Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc.
- Ghirardelli, J.E. (2005): An Overview of the Redeveloped Localized Aviation Mos Program (Lamp) For Short-Range Forecasting.
- Giordani, P. / Ferraro, M.B. / Martella, F. (2020): Introduction to Clustering. Springer Singapore.
- Grabbe, S.R. / Sridhar, B. / Mukherjee, A. (2014): Clustering days with similar airport weather conditions, in: 14th AIAA Aviation Technology, Integration, and Operations Conference (p. 2712).
- Gregor, S. / Hevner, A.R. (2013): Positioning and Presenting Design Science Research for Maximum Impact, in: MIS Quarterly, Jg. 37, Nr. 2, S. 337-355
- Government of Canada (2022): Canada's top 10 weather stories of 2022. URL: <https://www.canada.ca/en/environment-climate-change/services/top-ten-weather-stories/2022.html>, Abruf: 20.04.2023, 14:05 Uhr
- Hasan, N. / Uddin, M.T. / Chowdhury, N.K. (2016): Automated weather event analysis with machine learning, in: International Conference on Innovations in Science 2016, Engineering and Technology (ICISSET) (pp. 1-5). IEEE.
- Hegland, M. (2003): Algorithms for Association Rules, in: Mendelson, S., Smola, A.J. (eds) Advanced Lectures on Machine Learning. Lecture Notes in Computer Science(), vol 2600. Springer, Berlin, Heidelberg.

- Hersbach, H. / Bell, B. / Berrisford, P. / Hirahara, S. / Horányi, A. / Muñoz-Sabater, J. / Nicolas, J. / Peubey, C. / Radu, R. / Schepers, D. / Simmons, A. (2020): The ERA5 global reanalysis, in: *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp.1999-2049.
- Hevner, A. / Chatterjee, S. (2010): *Design Research in Information Systems, Theory and Practice*. Hrsg. von R. Sharda/S. Voß. Bd. 22. *Integrated Series in Information Systems*. New York, NY, USA: Springer New York, NY.
- Hevner, A. / March, S.T. / Park, J. / Ram, S. (2004): *Design Science in Information Systems Research*, in: *MIS Quarterly* 28.1, S. 75–105.
- Hjelmfelt, M.R. (1990): Numerical study of the influence of environmental conditions on lake-effect snowstorms over Lake Michigan, in: *Monthly Weather Review*, 118(1), pp.138-150.
- Holmstrom, M. / Liu, D. / Vo, C. (2016): Machine learning applied to weather forecasting. *Meteorology. Appl. Dec*; 10: 1-5.
- Horenko, I. / Dolaptchiev, S.I. / Eliseev, A.V. / Mokhov, I.I. / Klein, R. (2008): Metastable decomposition of high-dimensional meteorological data with gaps, in: *Journal of the atmospheric sciences*, 65(11), pp.3479-3496.
- Hupfer, P. / Kuttler, W. (2005): *Witterung und Klima. Eine Einführung in die Meteorologie und Klimatologie*, 11. Auflage
- Jahn, M. (2015): Economics of extreme weather events: Terminology and regional impact models. *Weather and Climate Extremes*, 10, pp.29-39.
- Jo, J.M. (2019): Effectiveness of normalization preprocessing of big data to the machine learning performance, in: *The Journal of the Korea institute of electronic communication sciences*, 14(3), pp.547-552.
- Kassambara, A. (2017): *Practical guide to cluster analysis in R: Unsupervised machine learning*, 1. Auflage, Sthda.
- Kotsiantis, S. / Kanellopoulos, D. (2006): Association rules mining: A recent overview, in: *GESTS International Transactions on Computer Science and Engineering*. 2006 Jan;32(1): 71-82.
- Liljequist, G.H. / Cihak, K. (1984): *Allgemeine Meteorologie*. 3. Auflage, Springer-Verlag.
- Liu, F. / Deng, Y. (2020): Determine the number of unknown targets in open world based on elbow method, in: *IEEE Transactions on Fuzzy Systems*, 29(5), pp.986-995.
- Liu, F. / Ting, K.M. / Zhou, Z.H. (2012): Isolation-based anomaly detection, in: *ACM Trans Knowl. Discov. Data* 6(1): Article 3
- Mitchell, T. (1997): *Machine learning*. McGraw Hill, New York
- Moon, T.K. (1996): The expectation-maximization algorithm, in: *IEEE Signal processing magazine*, 13(6), pp.47-60.
- Pelosi, A. / Terribile, F. / D'Urso, G. / Chirico, G.B. (2020): Comparison of ERA5-Land and UERRA-MESCAN-SURFEX reanalysis data with spatially interpolated weather observations for the regional assessment of reference evapotranspiration. *Water*, 12(6), p.1669.
- Pooja, S.B. / Balan, R.S. / Anisha, M. / Muthukumar, M.S. / Jothikumar, R. (2020): Techniques Tanimoto correlated feature selection system and hybridization of clustering and boosting ensemble classification of remote sensed big data for weather forecasting. *Computer Communications*, 151, pp.266-274.
- Poteraş, C.M. / Mihăescu, M.C. / Mocanu, M. (2014): An optimized version of the K-Means clustering algorithm, in *2014 Federated Conference on Computer Science and Information Systems* (pp. 695-699). IEEE.
- Ray, P. (ed) (2015): *Mesoscale meteorology and forecasting*. Springer.
- Runkler, T.A. (1999): *Probabilistische und Fuzzy Methoden für die Clusteranalyse*, in: Seising, R. (eds) *Fuzzy Theorie und Stochastik*. Computational Intelligence. Vieweg+Teubner Verlag, Wiesbaden.
- Scikit Learn (2023a): Clustering. URL: <https://scikit-learn.org/stable/modules/clustering.html>, Abruf: 07.03.2023, 14:33 Uhr
- Scikit Learn (2023b): Preprocessing. URL: <https://scikit-learn.org/stable/modules/preprocessing.html>, Abruf: 19.04.2022, 16:37 Uhr
- Sagiroglu, S. / Sinanc, D. (2013): Big data: A review, in: *International conference on collaboration technologies and systems (CTS) 2013 May 20* (pp. 42-47). IEEE.
- Savaresi, S.M. / Boley, D.L. / Bittanti, S. / Gazzaniga, G. (2002): Cluster selection in divisive clustering algorithms, in: *Proceedings of the 2002 SIAM International Conference on Data Mining* (pp. 299-314). Society for Industrial and Applied Mathematics.
- Spektrum Akademischer Verlag, Heidelberg, (2000): *Lexikon Der Geowissenschaften: Atmosphäre*. URL: <https://www.spektrum.de/lexikon/geowissenschaften/atmosphaere/1060>, Abruf: 23.02.2023, 13:46 Uhr
- Syakur, M. A. / Khotimah, B. K. / Rochman, E. M. S. / Satoto, B. D. (2018): Integration k-means clustering method and elbow method for identification of the best customer profile cluster, in: *IOP conference series: materials science and engineering* (Vol. 336, p. 012017). IOP Publishing.
- The Weather Network (2022): The Weather Network. URL: <https://www.theweathernetwork.com/en/news/weather/>, Abruf: 24.04.2023, 15:06 Uhr
- Thudumu, S. / Branch, P. / Jin, J. / Singh, J. (2020): A comprehensive survey of anomaly detection techniques for high dimensional big data, in: *Journal of Big Data*. Dec;7: 1-30.
- Fang, W. / Sheng, V.S. / Wen, X. / Pan, W. (2014): Meteorological data analysis using mapreduce, in: *The Scientific World Journal*, 2014.
- Webster, J. / Watson, R.T. (2002): Analyzing the past to prepare for the future: Writing a literature review, in: *MIS quarterly*. Jun 1: xiii-xiii.

Xu, Q. / He, D. / Zhang, N. / Kang, C. / Xia, Q. / Bai, J. / Huang, J. (2015): A short-term wind power forecasting approach with adjustment of numerical weather prediction in-put by data mining, in: IEEE Transactions on sustainable energy, 6(4), pp.1283-1291.

Yuan, C. / Yang, H. (2019): Research on K-value selection method of K-means clustering algorithm. J, 2(2), pp.226-235.

Zhou, Z.H. (2021): Machine learning. Springer Nature.

KONTAKT

JULIAN L.R. ERATH, geboren am 21.01.2001 in Sindelfingen, Deutschland, absolvierte 2023 seinen Bachelor of Science in Wirtschaftsinformatik Data Science an der Dualen Hochschule Baden-Württemberg in Stuttgart. Seit 2020 arbeitet er bei der IBM Deutschland GmbH, mit welcher er sein duales Studium abschloss. Seit Abschluss seines Bachelorstudiums arbeitet er als Data Consultant bei der IBM Deutschland GmbH und absolviert berufsbegleitend seinen Master of Science in Data Science an der Hochschule der Medien in Stuttgart. Julian.Erath@ibm.com

Prompt Engineering mit ChatGPT 3.5 für Midjourney am Beispiel eines Use Cases für Wildlife Photography

Jonathan Kreiss

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim

E-Mail: kreissjo@hs-pforzheim.de

Frank Morelli

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim

E-Mail: frank.morelli@hs-pforzheim.de

ABSTRACT

Das vorliegende Paper beleuchtet das Phänomen des Prompt Engineerings anhand der Anwendung von ChatGPT. In praktischen Anwendungen wird analysiert, was einen effektiven Prompt ausmacht und wie ChatGPT gezielt genutzt werden kann, um präzise Bildbeschreibungen für Midjourney zu generieren. Auf Grundlage dieser Erkenntnisse werden zudem Variationen für verschiedene Use Cases abgeleitet.

SCHLÜSSELWÖRTER

Prompt Engineering, Midjourney, ChatGPT

1. Theoretische Grundlagen Prompt Engineering Problemstellung und Zielsetzung

Das Konzept des Prompt Engineerings gewinnt mit der zunehmenden Verbreitung von Large Language Models (LLMs), wie beispielsweise OpenAIs ChatGPT, immer mehr an Bedeutung. Prompt Engineering zielt darauf ab, präzise Anweisungen, sogenannte Prompts, für LLMs zu entwickeln, um LLMs gezielt zu steuern und dadurch die Qualität der generierten Antworten zu verbessern. Die Forschungslage bezüglich der Merkmale eines präzise formulierten Prompts ist bis dato noch begrenzt. In diesem Zusammenhang hat der folgende Artikel das Ziel, einen Überblick über die Methoden zur Entwicklung effektiver Prompts zu liefern.

Die Bedeutung effektiver Prompts

Ein präziser formulierter Prompt steigert die Wahrscheinlichkeit nützlicher Antworten von Large Language Models (LLMs) zu erhalten. Um Prompt Engineering auf unterschiedliche Use Cases anzuwenden zu können, erfordert es Expertise im Prompt Engineering. Diese Fähigkeit ermöglicht es Nutzern, maßgeschneiderte Anweisungen für unterschiedliche Szenarien zu erstellen und so die Leistung von LLMs zu maximieren. Prompt Engineering ist hierbei entscheidend, um das volle Potenzial von LLMs auszuschöpfen. Nachfolgend werden die hierzu gewonnenen Erkenntnisse aus einer Bachelor-Thesis zusammengefasst.

Open AI Best Practices

OpenAI, das für die Entwicklung von ChatGPT verantwortliche Unternehmen, hat einige Best Practices zum Thema Prompt Engineering mit ChatGPT veröffentlicht, die im nachfolgenden kurz vorgestellt werden:

1. Trennung von Anweisung und Kontext

Zu empfehlen ist es, in einem Prompt zu Beginn präzise Anweisungen und anschließend den Kontext zu platzieren. Zur klaren Trennung zwischen Anweisung und Kontext können Trennzeichen wie "###" oder "====" eingesetzt werden.

2. Detaillierte und spezifische Anweisungen

Klare Anweisungen formulieren, etwa in Bezug auf:

Länge: Begrenzen der gewünschten Länge des Texts, beispielsweise auf maximal 3 Sätze, 300 Wörter oder 1000 Zeichen.

Format: Spezifizieren des gewünschten Formats, sei es in Form von Stichpunkten, einer Tabelle, Absätzen oder mit Zwischenüberschriften.

Schreibstil: Präzisieren des gewünschten Schreibstils, ob wissenschaftlich, geschäftlich, kreativ, humorvoll, formal oder in einem anderen Stil.

3. Ausgabeformat angeben

Einige Beispiele für mögliche Ausgabeformate sind:

Tabellenform: Die gewünschte Antwort wird in tabellarischer Form präsentiert, um Informationen übersichtlich und strukturiert darzustellen.

Stichworte: Die Antwort wird in Form von Stichworten oder Schlagworten verfasst, um Schlüsselinformationen hervorzuheben.

Sätze mit Zwischenüberschriften: Die Antwort erfolgt in gegliederten Sätzen, die durch Zwischenüberschriften strukturiert sind, um eine klare Struktur in der Antwort zu schaffen.

Pro/Kontra Liste: Die Antwort wird in Form einer Liste von Argumenten aufgeteilt, um Vor- und Nachteile zu verdeutlichen.

Schritt-für-Schritt Erklärung: Die Antwort wird schrittweise erläutert, um komplexe Konzepte oder Abläufe verständlich darzustellen.

4. Show don't tell

Statt ausführliche Instruktionen zu schreiben, ist es oft zielführender, ein oder mehrere Beispiele direkt in den Prompt zu integrieren. Die verschiedenen Arten des Promptings werden im nächsten Abschnitt ausführlicher beleuchtet.

Promptingarten

In der Forschung und Anwendung des Promptings haben sich drei Hauptvarianten herausgebildet: Zero-Shot, One-Shot und Few-Shot Prompting. Diese Varianten unterscheiden sich in der Anzahl der Beispiele, die in den Prompt integriert werden:

- **Zero-Shot Prompting:** In dieser Variante sind keinerlei Beispiele im Prompt enthalten. Der Prompt besteht lediglich aus einer reinen Aufforderung ohne vorherige Beispiele. Eignet sich besonders für einfache Instruktionen wie zum Beispiel Fragen.
- **One-Shot Prompting:** Bei dieser Methode wird präzise ein einziges Beispiel in den Prompt integriert, welches als Grundlage für die gewünschte Antwort dient. Eignet sich für komplexere Instruktionen.
- **Few-Shot Prompting:** Beim Few-Shot Prompting hingegen werden mindestens zwei, oft aber mehrere Beispiele in den Prompt aufgenommen. Diese Beispiele dienen dazu, dem Modell zusätzlichen Kontext und Informationen zur Verfügung zu stellen, um die Qualität und Relevanz der generierten Antwort zu steigern. Eignet sich vor allem für sehr komplexe Instruktionen.

Chain-of-Thought Prompting

Chain-of-Thought (CoT)-Prompting ist besonders geeignet für komplexe und neuartige Aufgabenstellungen. Im Vergleich zum herkömmlichen Prompting werden

bei CoT-Prompting gezielt Zwischenschritte in die Anweisung miteingebaut. CoT-Prompting ermöglicht es, komplizierte Sachverhalte in mehrere einfachen Schritte aufzuteilen.

Promptingmatrix

Die folgende Matrix soll eine Hilfestellung geben, bei welcher Aufgabenstellung, welche Promptingart verwendet werden kann (siehe Abb. 1).

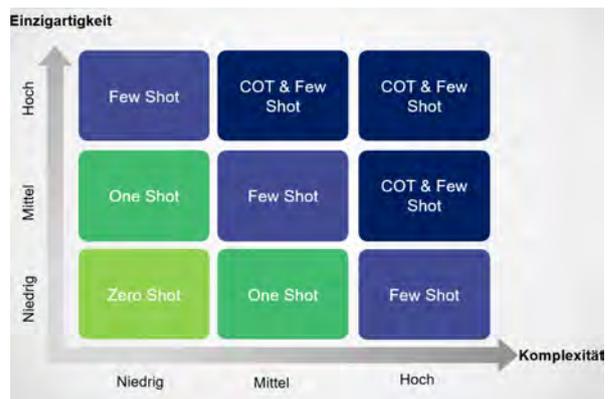


Abb. 1: Promptingmatrix

- Y-Achse: "Einzigartigkeit" – Diese Achse gibt das Ausmaß der Einzigartigkeit einer gegebenen Anweisung an. Eine höhere Position auf dieser Achse deutet darauf hin, dass die Anweisung für Aufgaben verwendet wird, die neuartig sind und wahrscheinlich nicht im Trainingsdatensatz des Large Language Models (LLM) enthalten sind.
- X-Achse: "Komplexität" – Die X-Achse stellt die Komplexität einer Anweisung dar. Anweisungen mit geringer Komplexität umfassen klar formulierte Fragen oder einfache Aufgabenstellungen, die direkt auf Informationen abzielen. Eine höhere Position auf dieser Achse weist auf komplexere Anweisungen hin die eine mehrstufige Verarbeitung erfordern können.

Prompt Engineering – ein Prozess

In der Praxis beginnt der Prozess des Prompt Engineering mit einer grundlegenden und simplen Anweisung. Im Verlauf wird der Prompt schrittweise verfeinert und jede erzeugte Antwort sorgfältig überprüft. Dieser iterative Prozess setzt sich fort, bis der Prompt eine derart präzise Form erreicht, dass die generierte Antwort den Erwartungen entspricht und einen konkreten praktischen Nutzen bietet.

Geringfügige Änderungen, wie beispielsweise das Austauschen von Wörtern, können erhebliche Auswirkungen auf die erzeugte Antwort haben. Die Unterscheidung zwischen effektiven und weniger effektiven Anweisungen erfordert oft experimentelles Ausprobieren und Er-

proben unterschiedlicher Formulierungen. Dieser kontinuierliche Anpassungsprozess ist entscheidend, um die Qualität und Relevanz der generierten Antworten zu optimieren und den angestrebten Mehrwert zu gewährleisten.

2. Prompting mit ChatGPT 3.5 für Midjourney

Ausgangslage

Zur besseren Vergleichbarkeit zwischen den generierten Bildern und den dazugehörigen Prompts wird ein Tiger als Motiv für alle Generierungen ausgewählt. Zudem wurden bei jeder Bilderstellung dieselben Parameterwerte festgelegt, um die Vergleichbarkeit zwischen den Bildern zu gewährleisten

Vorgehensweise

Die Generierung geeigneter Prompts für Midjourney mithilfe von ChatGPT 3.5 steht im Mittelpunkt der Untersuchung. Die Methodik dieses Vorhabens ist in mehrere aufeinanderfolgende Prozessschritte unterteilt:

1. *Prompt-Formulierung mit ChatGPT:* Entwicklung präziser Prompts, um qualitativ hochwertige textuelle Bildbeschreibungen zu generieren.
2. *Bildgenerierung in Midjourney:* Die in Schritt 1 erstellten Prompts werden in Midjourney getestet
3. *Bewertung der generierten Bilder:* Ein KI-Tool wird zur Bewertung und Analyse der erzeugten Bilder eingesetzt.
4. *Analyse und Optimierung der Prompts:* Die genutzten Prompts werden analysiert und verbessert, um die Bildqualität zu verbessern.

Der Prozess wird nach Abschluss des vierten Schrittes iterativ wiederholt, da Prompt Engineering keine lineare Methode darstellt, sondern fortlaufende Anpassungen und Verfeinerungen erfordert, um optimale Ergebnisse zu erzielen.

3. Praktische Umsetzung (Ergebnisse)

Die KI-basierte Bildbewertungs Website <https://aesthetics.everypixel.com/> wird genutzt, um die generierten Bilder zu bewerten und anschließend miteinander zu vergleichen. Die Einschätzung des Bildes erfolgt in Prozent, wobei das Maximum bei 100% liegt. Je näher die Bewertung an 100% heranreicht, desto ästhetischer wurde das Bild bewertet.

Es ist jedoch wichtig anzumerken, dass die ästhetische Bewertung durch die KI nicht gänzlich objektiv ist und stark von den Merkmalen des Algorithmus beeinflusst wird. Es bleibt unklar, nach welchen genauen Kriterien der Algorithmus Bilder als ästhetisch oder nicht ästhetisch einstuft. Dennoch ermöglicht der Einsatz des Algorithmus eine Bewertung und somit auch ein Vergleichbarkeit zwischen den generierten Bildern.

3.1 Tiger

Struktur des finalen Prompts

Der folgende Screenshot, zeigt die Struktur des finalen Prompts. Dieser Prompt lässt sich vielfältig nutzen, um verschiedene Variationen zu generieren. Dabei können alle Parameter verändert werden, um neue Bildbeschreibungen zu generieren. Im nächsten Kapitel (3.2) werden zudem Abwandlungen für andere Tierarten vorgestellt.

Rank	Bild
1.	
99,40%	Wildlife Photography of a Tiger. Full Body Shot. Background Jungle. Captured by Sony Alpha a9 II with Sony FE 400mm f/2.8 GM OSS lens. Low-angle wide shot. Dappled sunlight. Intense.
2.	
99,20 %	Wildlife Photography of a Tiger. Full Body Shot. Background Jungle. Captured by Sony Alpha a9 II with Sony FE 400mm f/2.8 GM OSS lens. Low-angle wide shot.
3.	
99,00 %	Wildlife Photography of a Tiger. Full Body Shot. Background Jungle. Captured by Canon Sony Alpha a9 II.

Tabelle 1

Certainly, here's the modified breakdown with "Photography Type" in place of "Wildlife Photography of a Tiger":

Aspect	Description
Photography Type	Wildlife Photography of a Tiger
Subject	Tiger
Type of Shot	Full Body Shot
Background	Jungle
Camera Used	Sony Alpha a9 II
Lens Used	Sony FE 400mm f/2.8 GM OSS lens
Shot Angle	Low-angle wide shot
Lighting	Dappled sunlight
Pose	Crouching and ready to pounce

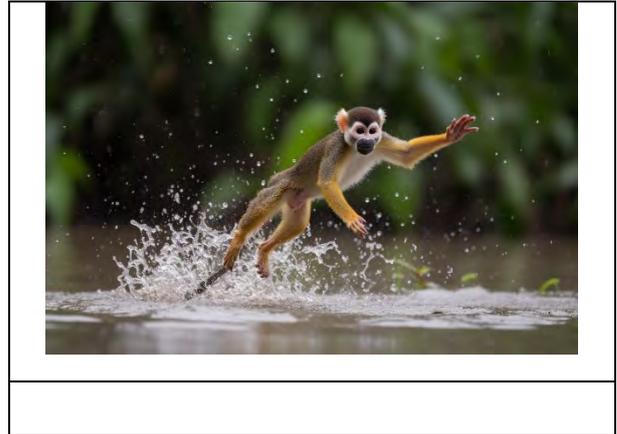
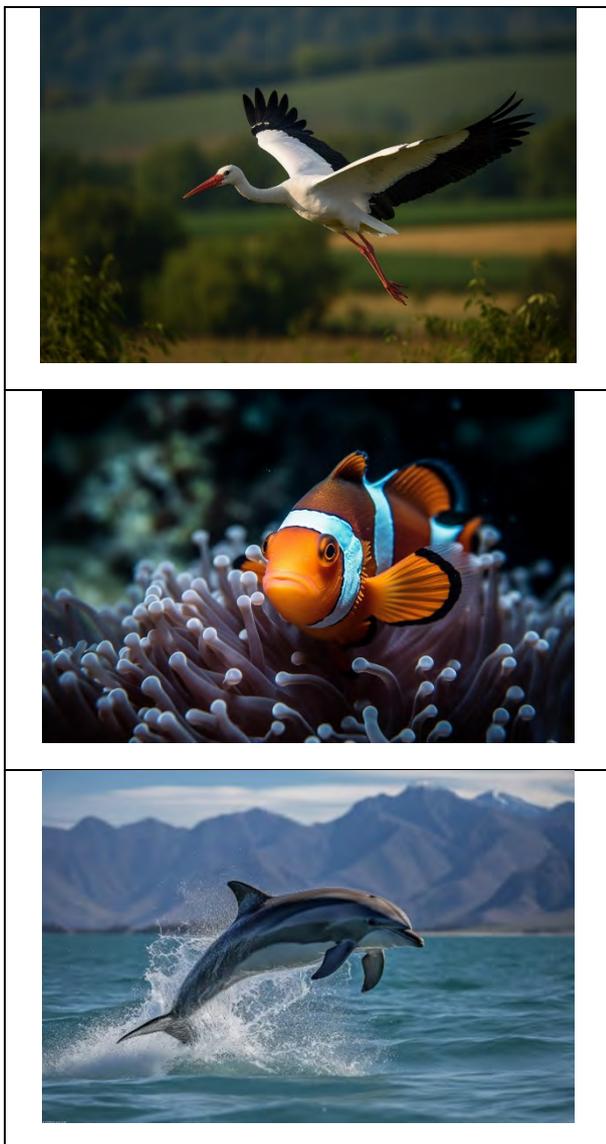


Tabelle 2

3.2 Abwandlungen andere Tiere



Fazit

Das Prompt Engineering gewinnt mit der zunehmenden Verbreitung von LLMs immer mehr an Bedeutung. Durch präzise formulierte Prompts lässt sich die Antwort von LLMs gezielt steuern, was zu verbesserten Ergebnissen führt. Ein Beispiel dafür ist das gezielte Trainieren von ChatGPT für die Generierung von Bildbeschreibungen in Midjourney. Die drei Hauptvarianten des Promptings: Zero-Shot-, One-Shot- & Few-Shot-Prompting bieten differenzierte Ansätze, die je nach Aufgabenkomplexität variieren.

Für komplexere Anweisungen wie die Generierung einer Bildbeschreibung eignen sich One-Shot bzw. Few-Shot Anweisungen mit einem oder mehreren Beispielen.

LITERATUR

Brown et. al (2020) Language Models are Few-Shot Learners, OpenAI

<https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api>

Brown, T, et al. (2020). Language Models are Few-Shot Learners. arXiv (Cornell University). <https://arxiv.org/pdf/2005.14165.pdf>

Entwicklung einer plattformunabhängigen Chatbot-Frontend-Anwendung

Simon Ludwig

Technische Hochschule Mittelhessen
Fachbereich MND
Wilhelm-Leuschner-Str. 13
61169 Friedberg
E-Mail: simon.ludwig@mnd.thm.de

Prof. Dr. Harald Ritz

Technische Hochschule Mittelhessen
Fachbereich MNI
Wiesenstraße 14
35390 Gießen
E-Mail: harald.ritz@mni.thm.de

Kategorie

Masterarbeit

Schlüsselwörter

App-Entwicklung, Chatbot, Continuous Integration (CI), Continuous Delivery (CD), Frontend-Entwicklung, KI, Plattformunabhängigkeit, Prototyping, Testautomatisierung

Zusammenfassung

Für die Technische Hochschule Mittelhessen, bei der das Leitbild auf der Bewältigung technischer Herausforderungen unter Verwendung neuester Erkenntnisse ausgerichtet ist, griff der Prüfungsausschuss-Vorsitzende der Studiengänge B.Sc. und M.Sc. Wirtschaftsinformatik im Jahre 2021 den Trend der Chatbots auf. Gemeinsam mit Master-Studierenden der Wirtschaftsinformatik leitete er die Entwicklung von „Winfy“, einem Chatbot, der Fragen von Studierenden bezüglich Prüfungsangelegenheiten einfach und schnell beantworten sollte, ein.

Ogleich ebenjener Chatbot Fragen der Nutzer ausgesprochen präzise in seiner vorliegenden Version beantwortet, ist dessen Nutzung zum Zeitpunkt des Verfassens der Abschlussarbeit nur auf Endgeräten mit großen Bildschirmen bzw. einer hohen Auflösung möglich. Das fehlende responsive Verhalten sorgt dafür, dass die Anwendung auf mobilen Endgeräten bisher mit Einschränkungen bedient werden kann. Die seit Jahrzehnten wachsende Nutzung von mobilen Endgeräten wie Smartphones und Tablets zwingt Entwickler und Betreiber von Anwendungen zum Umdenken. Denn es ist davon auszugehen, dass sich dieser Trend fortsetzen, gar intensivieren wird. Durch neue Technologien wie der Virtual-Reality-Brille oder Smartwatches wird der sowieso schon sehr diverse Endgeräte-Markt noch vielfältiger werden.

Ziel der Masterarbeit war es daher, eine plattformunabhängige Anwendung zu entwickeln, die auf allen gängigen Endgeräten lauffähig sein sollte – also sowohl auf Smartphones und Tablets als auch Desktop-Geräten und Notebooks. Neben der Plattformunabhängigkeit umfassten die Ziele weitere Anforderungen, die sich aus der vorherigen Anwendung, dem ISO 25010 und aktuellen Gesetzestexten ableiten ließen.

So sollte neben einer barrierefreien Anwendung, eine mehrsprachige, responsive Oberfläche entstehen. Zudem wurde sich auf die langfristige Wartbarkeit, die Wirtschaftlichkeit der Entwicklung, den Betrieb und die Benutzerfreundlichkeit fokussiert.

Mit den gesetzten Zielen und Rahmenbedingungen wurden mögliche Entwicklungsansätze ermittelt, mit deren Hilfe die Anwendung entwickelt werden konnte. Die möglichen Entwicklungsansätze umfassten dabei den nativen, Cross-Plattform, Hybriden und Web-Entwicklungsansatz. Nach einer umfangreichen Nutzwertanalyse, die alle Interessengruppen, Ziele und Anforderungen berücksichtigten, erwies sich die Entwicklung einer Webanwendung als am sinnvollsten.

Der native Entwicklungsansatz stellte sich als ungeeignet heraus, da für die Entwicklung separate Quellcodes für jedes zu unterstützende Betriebssystem notwendig gewesen wären. Nicht nur die initiale Entwicklung, sondern auch die Wartung mehrerer Projekte hätte einen hohen zeitlichen Aufwand zur Folge gehabt.

Da der Web-Entwicklungsansatz im Vergleich zu der Cross-Plattform- und hybriden Entwicklung einen geringeren Wartungsaufwand, eine schnellere Implementierung, geringere Einstiegshürden und den Aufbau einer CI-/CD-Pipeline ermöglicht, wurde dieser als am sinnvollsten betrachtet. Nachdem ein adäquater Entwicklungsansatz eruiert wurde, sollte die Entwicklung der Anwendung nutzerorientiert erfolgen. Dies mit dem Ziel, den Nutzer bei allen Entscheidungen in den Mittelpunkt zu stellen.

Auf Grundlage der vorherigen Version des Chatbots und der Anforderungen aus dem Kontext der Prüfungsangelegenheiten wurden die Anwendungsfälle ausgearbeitet. In einem durch das Prototyping angepassten Wasserfallmodell wurden auf dieser Grundlage Prototypen der Anwendung entworfen. Durch dieses Vorgehen konnte bereits früh Feedback zur Oberfläche eingeholt und in die Entwicklung einbezogen werden.

Die Oberfläche wurde dabei unter Berücksichtigung aktueller Studien und Untersuchungen aufgebaut. So war es ein Ziel, die kognitive Belastung bei der Bedienung durch den Nutzer zu minimieren. Dabei wurden u.a.

Konzepte wie die Daumen-Zone und die Daumen-Impact-Area aufgegriffen. Weiterhin wurde sichergestellt, dass eine leichtgewichtige Anwendung entsteht, die sowohl mit einer schnellen Lade- als auch Bediengeschwindigkeit aufwartet. Mehrere Studien und Untersuchungen zeigen die Wichtigkeit schnell reagierender Anwendungen auf. Eines der daraus abgeleiteten Ziele ist es daher, die Schwellwerte einzuhalten. Mit einer initialen Ladezeit von etwa 700 Millisekunden und einer Ladezeit beim wiederholten Besuch von 200 Millisekunden, wurden die gesetzten Ziele weit übertroffen. Damit die Zufriedenheit der Interaktion mit dem Chatbot gemessen werden kann, wurde die Oberfläche um das Bewerten der Antworten erweitert.

Ein mit der Benutzerfreundlichkeit eng verwobenes Konzept stellte die Qualitätssicherung der Anwendung dar. Dieses spielte bei der Entwicklung eine entscheidende Rolle. Eine minderwertige Anwendung wirkt sich so nicht nur negativ auf das Benutzererlebnis, sondern auch auf die Wartbarkeit, die Wirtschaftlichkeit und die Plattformunabhängigkeit einer Anwendung aus. Eine Ausfallzeit kann beispielsweise nicht nur imageschädigende Auswirkungen zur Folge haben, sondern beeinträchtigt auch das Benutzererlebnis. Auch Laufzeitfehler im Quellcode würden dazu führen, dass die Anwendung nachträglich korrigiert werden müsste. Es galt daher sowohl während, als auch nach der Entwicklung qualitätssichernde Maßnahmen zu ergreifen, um Fehler zu vermeiden und diese proaktiv behandeln zu können.

Für diesen Zweck wurden neben automatisierten Tests und statischen Quellcodeanalysen auch Akzeptanztests entwickelt. Mithilfe der statischen Quellcodeanalyse, der Unit- und End-To-End-Tests sollten Fehler bereits vor der Inbetriebnahme einer neuen Version gefunden werden. Zur Reduzierung der Auswirkungen eines Ausfalls des Systems wurde zusätzlich ein Monitoring-System implementiert, das sowohl die Frontend- als auch die Backend-Anwendung kontinuierlich überwacht. Laufzeitfehler der Anwendung sollen so schnell erkannt und entsprechende Maßnahmen eingeleitet werden. Im Fall eines Fehlers sollte der Betreiber durch ein Ticket benachrichtigt werden.

Damit Änderungen schnell und unkompliziert eingepflegt und in Betrieb genommen werden können, bestand ein weiterer Aspekt in der Implementierung einer CI-/CD-Pipeline. In Verbindung mit einem dazu passenden Branching-System und automatisierten Testfällen wurde so ein Prozess entwickelt, der sicherstellt, dass ausschließlich Änderungen in Betrieb genommen werden, die vorher getestet wurden und den Qualitätsansprüchen genügen.

Die entwickelte Pipeline wurde so konfiguriert, dass sicherheitsrelevante Schlüssel und Passwörter innerhalb des Quellcodeverwaltungssystems, nicht aber innerhalb des Projektes, gespeichert werden.

Fazit

Die Entscheidung, eine Webanwendung zu entwickeln, ermöglichte es, die Anwendung plattformunabhängig zu entwickeln. Endgeräte mit einem Browser ermöglichen den barrierefreien Zugang zum Chatbot. Sowohl Smartphones, Tablets, Laptops als auch Desktop-PCs unterstützen die Anwendung. In über 20 Tests wurde bewiesen, dass die Anwendung auf modernen Browser funktionsfähig ist. Die Anwendung wurde auf den o.g. Geräten mehreren Tests unterzogen.

Die multilinguale, responsive Oberfläche, die Einhaltung von gängigen Barrierefreiheitsregeln und das Implementieren einer Text-To-Speech- und Dark-Mode-Funktionalität ermöglichen es, dass die Anwendung auch von Menschen mit körperlichen Einschränkungen genutzt werden kann. Durch die entwickelte CI-/CD-Pipeline wird sichergestellt, dass die Anwendung leicht zu warten ist, die Software den Qualitätsrichtlinien entspricht und neue Versionen einfach und schnell in Betrieb genommen werden können. Durch eine entsprechende Dokumentation und Installationsanweisungen wurde zudem sichergestellt, dass die Anwendung von zukünftigen Entwicklern unkompliziert gewartet und verwaltet werden kann. Das nutzerzentrierte Vorgehen stellte die Nutzer des Chatbots bei allen Entscheidungen in den Mittelpunkt. Das Ergebnis dieses Vorgehens ist eine Anwendung, die sich durch eine einfach zu bedienende Oberfläche und eine sehr schnelle Lade- und Bediengeschwindigkeit auszeichnet.

Mit der Entwicklung einer modernen, plattformunabhängigen Webanwendung wurden Hürden in der Bedienung des Chatbots abgebaut und die Anwendung einem noch breiteren Publikum zugänglich gemacht. Die Anwendung kann sich nun zu einem zentralen Anlaufpunkt bei Fragen rund um das Studium entwickeln.

Literatur

Baxter, Kathy; Catherine Courage; Kelly Caine: Introduction to User Experience. in: Understanding your Users. Waltham USA. Elsevier Science (2015)

Boehm, Barry; Basili, Victor: Software Defect Reduction Top 10 List. in: IEEE Computer 34. Washington, D.C USA. IEEE Computer Society Press (2001).

França, Joyce; Soares, Michel S.: SOAQM: Quality Model for SOA Applications based on ISO 25010. Barcelona Spain. SciTePress (2015)

Ritz, Harald; Tansel, Dogus: Entwicklung eines KI-basierten FAQ-Chatbots für die Hochschule im Bereich Prüfungsangelegenheiten, in: AKWI, 17 / 2023, S.81-92, DOI: <https://doi.org/10.26034/lu.akwi.2023.n17>