NUTZERBINDUNG DURCH VERHALTENSBASIERTE BIG-DATA-ANALYSE

Dr. Nora Vollmers
Andreas Berghammer
Luca Trautmann
PROCON IT Aktiengesellschaft
Parkring
57-59,
85748 Garching bei München
E-Mail: nora.vollmers@procon-it.de
E-Mail: andreas.begrhammer@procon-it.de

Abstract

Kundenbeziehungsmanagement ist ein gewinnbeeinflussender Faktor für Unternehmen und hat in den letzten Jahren stark an Bedeutung gewonnen. In diesem Artikel wird vorgestellt, wie durch verhaltensbasierte Big-Data-Analyse Nutzerbindungen aufgebaut werden können. Die drei grundlegenden Fragen hierbei sind: Welche Dienste werden genutzt? Wer sind meine aktivsten Nutzer? Gibt es Gemeinsamkeiten oder Synergien zwischen den Nutzerverhalten? Die Herausforderung besteht in der effizienten Aufbereitung und Auswertung von großen Datenmengen.

Schlüsselwörter

Big Data, Data Lake, Kundenbeziehungsmanagement

Einleitung

Standen früher ausschließlich einzelne Verkaufsabschlüsse im Vordergrund, liegt heutzutage der Fokus auf einer langfristigen Kunden- und Geschäftsbeziehung. Geschätzt ist es mindestens fünfmal teurer einen Neukunden zu gewinnen als die Zufriedenheit eines Stammkunden konstant aufrecht zu erhalten oder gar zu erhöhen. Dieser Entwicklungsprozess ist in Abb. 1 dargestellt. Kundenzufriedenheit, -bindung und -wertermittlung sind wichtige gewinnbeeinflussende Faktoren. Zum Erlangen dieser Kundeninformationen ist eine Datenerhebung notwendig. Diese Kenntnisse ermöglichen es den Unternehmen Kundenbindungseffekte zu schaffen. (Töpfer and Mann 2008)

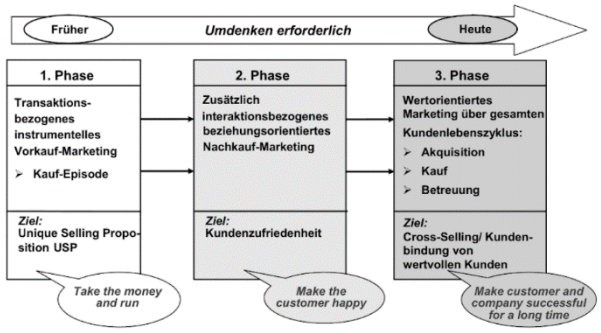

Abb. 1: Vom Beeinflussungsmarketing zum wertorientierten Beziehungsmarketing (Töpfer and Mann 2008)

Unternehmen grenzen sich schon lange nicht mehr nur über das Endprodukt voneinander ab, sondern ebenfalls über digitale Zusatzdienste.
Um die Wirksamkeit bzw. Rentabilität der Zusatzdienste, sowie die dadurch entstehenden Kosten nachvollziehbar zu machen, ist es von Bedeutung alle daraus resultierenden Daten zusammenzuführen und geeignet zu aggregieren. Ein geeigneter zentraler Speicherort kann mittels eines Data Lakes geschaffen werden.

Problem der Datenverantwortlichkeiten

Ein häufiges Problem bei der Analyse der Daten ist, dass diese meist nicht zentral an einem Ort vorliegen oder ihre Zugänglichkeit beschränkt ist. Häufig betreiben verschiedene Abteilungen unterschiedliche Dienste oder sind für einzelne Applikationen verantwortlich. Im Idealfall monitoren die Abteilungen bereits Nutzerzahlen oder Zugriffsstatistiken ihrer eigenen Anwendungen. Häufig werden allerdings nur kleine Teile der Daten berücksichtigt. Dies kann verschiedene Gründe haben.
Selbst kleine Applikationen können große Datenmengen produzieren, deswegen ist die Verarbeitung dieser Datenmasse mit konventionellen Methoden oft nicht möglich und verlangt eine entsprechende Infrastruktur. Hinzu kommt, dass diese Daten, gerade Applikationslogs oder von Sensoren generierte Daten, oftmals in verschiedenen semistrukturierten Formen, wie z.B. JSON, XML, Text oder CSV vorliegen. Das Erstellen eines Gesamtprofils eines Nutzers ist durch die Verteilung der Verantwortlichkeiten, nicht existenter Schnittstellen und komplexen Prozessen meist nicht möglich.

Big Data: Datenaufbereitung

Die Einführung eines zentralen Speicherorts für alle Formate - in Form eines sogenannten Data Lake - ermöglicht es, alle Daten der entsprechenden Abteilungen und Zuständigkeitsbereiche zu vereinen und fachübergreifend zugänglich zu machen. Technologisch hat sich Hadoop bei den meisten Branchen als Data Lake etabliert. Es handelt sich hierbei um ein Open-Source-Projekt, dass die Speicherung von Daten über eine große Anzahl von Rechnern, sogenannten Knoten unterstützt. Konzerne setzen hierbei in der Regel auf orchestrierte und vor allem supportete Lösungen von Firmen wie Cloudera, Hortonworks oder MapR. Zusätzlich werden in Bundles Komponenten (in der Regel auch aus dem Open-Source-Bereich) zur Verarbeitung der Daten oder Steuerung der Zugriffe angeboten, um ein ganzheitliches Big-Data-Konzept zu ermöglichen. (Mathis 2017)
So können z.B. Daten verschiedenen Typs und Herkunft mit Hilfe von Tools wie Kafka, Sqoop oder Flume in den Data Lake übertragen werden. Man nennt diesen Prozess „ingest“ (Pasupuleti and Purra 2015).
Datenquellen können hierbei Filestreams, Folder, Datenbanken aber auch Bildquellen, Audiofiles oder Videos sein. Die Daten werden in dem sogenannten Source Layer in ihrer Ursprungsform übertragen. So werden hier unter anderem die relevanten JSON-, CSV-, XML-, Text-, DB- und andere Dateien aus den Quellsystemen abgelegt. Im Unterschied zu Data-Warehouse-Systemen, bei denen in der Regel nur die aggregierten Daten gespeichert werden, bleiben im Data Lake die Rohdaten erhalten.
Zur Überführung der Datenformate in ein einheitliches Format wird ein Tool benötigt, das komplexe Transformationen durchführt. Durch diese Transformationen werden die Daten abfragbar. Ein beliebtes Tool hierbei ist Apache Spark. Durch das Nutzen vieler Rechner im Verbund werden riesige Datenmengen „in memory“ prozessiert, parallel verarbeitet und gespeichert. Es gibt eine Vielzahl an Bibliotheken um z.B. mit Hilfe von Scala, Python oder Java Daten verschiedenster Herkunft zu extrahieren und abfragbar in Tabellen auf dem sogenannten Prepared Layer zu speichern. Beliebte Tools auf dem Data Lake sind hierbei HIVE für Daten mit definiertem Schema oder HBase als NoSQL Komponente. (Mathis 2017)

Big Data: Datenauswertung

Sind die Daten im Prepared Layer abfragbar gespeichert, können Datascientists mit Ad-Hoc-Analysen beginnen. Die Ergebnisse können mittels direkter Ausführung von SQL-Querys berechnet werden. Als Query-Editor kann ein einfacher Datenbankmanager verwendet werden, der sich z.B. via JDBC Treiber mit HIVE verbindet. Die Parallelisierung der Abfragen mit Hilfe von Big-Data-Engins reduziert die Rechenzeit. Durch die Speicherung aller relevanten Daten im Data Lake sind der Kreativität des Datascientisten keine Grenzen gesetzt. Einfache ABC-Analysen, um verschiedene Benutzerklassen zu identifizieren, können genauso relevant sein, wie das Finden von Synergien, selten genutzten Diensten, unbekannten Mustern, Trends, usw.
Stellt man zum Beispiel durch Daten fest, dass Benutzer relativ komplizierte Wege durch die Menüführung auf sich nehmen, um eine häufig benutzte App zu finden, so könnte diese durch ein Update zukünftig auf dem Startbildschirm erscheinen.
Die kombinierte Nutzung von Applikationen kann bei der Datenauswertung berücksichtigt werden. Das Kombinationsmuster kann verwendet werden, um Ähnlichkeitsstrukturen zu analysieren. Dadurch können dem Kunden weitere Applikationen empfohlen werden, die möglicherweise in seinem Interessengebiet liegen. Lizenzen für Dienste die kaum oder gar nicht benutzt werden können reduziert oder ganz eingespart werden. Nutzer, die innerhalb einer Applikation bisher keine Beachtung fanden, können durch das Zusammenführen der Daten als neue Zielgruppe identifiziert werden.
Dadurch können in der Automobilbranche Fragestellung wie z.B. „Ist der Alltagsfahrer, der die Applikation selten nutzt, vielleicht gar nicht so interessant wie der Gelegenheitsfahrer, der meine Applikation ständig im Einsatz hat?“ beantwortet werden.
Des Weiteren bietet der Data Lake verschiedene Schnittstellen und Tools für Machine-Learning-Komponenten. Neben manuellen Abfragen können auch intelligente Modelle entwickelt werden, welche auf großen Datenmengen trainiert werden. Dadurch ist beispielsweise eine individuelle digitale Unterstützung möglich.

Big Data: Datenschutz

Big Data bringt auch kritische Aspekte mit sich. Die Analyse von Kundendaten impliziert einen gewissenhaften Umgang mit den Daten. Die Einhaltung des Datenschutzes ist hierbei ein wichtiger Aspekt. Die personenbezogenen Kundendaten können durch Anonymisierung, ein entsprechendes Rechte-Rollen-Konzept und weiteren Compliance-Vorkehrungen geschützt werden. Im Vordergrund sollten im Bereich von Big Data immer Gruppen oder Cluster stehen. Die Analyse von Einzelpersonen ist in der Regel nicht notwendig und sollte im Idealfall erst gar nicht möglich sein, um Zuwiderhandlung zu unterbinden und das Vertrauen der Kunden zu stärken.

Big Data: Vorteile für Unternehmen

Die Analyse von Massendaten ist für ein Unternehmen vorteilhaft. Pro Monat und angebotenem Dienst, wie z.B. Websites oder Apps, mehrere Millionen Einträge durch Kundenaktionen in Logfiles produziert. Diese treten je nach Implementierung in unterschiedlichsten strukturierten, semistrukturierten oder unstrukturierten Formaten auf.
Die Möglichkeit diese enormen, inhomogenen Datenmengen performant zu verarbeiten, zu speichern und zu analysieren bieten zum Beispiel die bereits erwähnten

moderne Open Source Technologien Hadoop, HIVE und Spark.
Diese neuen Technologien ermöglichen die Parallelisierung der Rechenvorgänge in Memory und somit sind komplexe Transformationen auf Datensatzgrößen mehrerer Milliarden Zeilen in kurzer Zeit durchführbar. Ein Beispiel aus der Praxis für die schnelle Verarbeitung ist das Prozessieren von Logfiles, die in den Source Layer von mehreren Webapplikationen ingestiert werden. So werden in der Sekunde ca. 10.000 Einträge durch Benutzerinteraktionen generiert. Dadurch entsteht eine stetig wachsende Datenmenge von ca. 1.5TB pro Monat. Mit Hilfe der in memory Verarbeitung kann via Batchprozessing die Aufbereitung eines kompletten Monats in weniger als 30 Minuten durchgeführt werden. Es können bestimmte Textfelder mit Hilfe von Regular Expressions extrahiert, Ergebnisse mit diversen Mappings verbunden und Kennzahlen berechnet werden. Möglich wäre auch ein Streaming-Szenario, das alle Daten sofort bei deren Erscheinen in Echtzeit in den Prepared Layer überträgt.
Der digitale Datenspeicher eines Produktes ermöglicht das automatische Lesen von Daten. Infolgedessen wird eine vollkommen dezentrale Produktionsumgebung geschaffen und die Produkte selbst werden zu wichtigen Informationsträgern. (Herrmann 2018) Diese Informationen können sich die Unternehmen zunutze machen, um das Produkt individuell auf den Kunden anzupassen und somit eine Nutzerbindung aufzubauen.
Die gemessenen Kundeninteraktionen können nun im Data Lake gespeichert werden. Somit können anschließend durch einfache SQL-Querys Analysen durchgeführt werden, um z.B. in einer Art ABC-Analyse Kennzahlen über Nutzergruppen zu ermitteln.
Bei diesen Gruppen geht es beispielsweise darum, herauszufinden, welche Nutzer neu, durchgehend oder nach längerer Pause wieder im System aktiv sind. Zusätzlich kann analysiert werden, welche Anwendungen bzw. welche Kombinationen dieser die Kunden verwenden.
Diese Erkenntnisse können verwendet werden, um zu errechnen, wie beliebt angebotene Dienste sind, ob Synergie- oder Ausschlusseffekte existieren oder wie sich das Kundenverhalten über die Zeit auf den Systemen verändert.
Durch intelligente Algorithmen können auf die Kunden abgestimmte Marketing-Aktionen gestaltet und automatisiert verbessert werden. Die Kundenbindung wird dadurch messbar erhöht. Dies geschieht einerseits mit traditionellen Methoden, indem Nutzer, welche eine längere Zeit nicht aktiv waren, spezielle Angebote erhalten und andererseits mit Verfahren aus dem Bereich Machine Learning. Ein Beispiel sind Empfehlungen auf Basis von Clustering-Algorithmen, bei denen bereits verwendete Dienste und möglicherweise interessante Systeme abgeglichen werden.

Digitale Zusatzdienste: Das smarte Fahrzeug

Längst schon nicht mehr kauft man ein Fahrzeug ausschließlich aus Gründen der Fortbewegung. Beim Erwerb eines neuen Autos erwartet der Kunde: „Freude am Fahren“, „Vorsprung durch Technik“, „das Beste oder nichts“. Die Hersteller setzen auf Zusatzdienste, die den Fahrkomfort maximieren, um sich von der Konkurrenz abzugrenzen. So besitzen besonders moderne Fahrzeuge gehobener Preisklasse neben einer Reihe an Sensoren, die intelligente Assistenzsysteme ermöglichen, auch eine hohe Anzahl an zusätzlichen multimedialen Diensten. Sprachassistenten verschiedener Hersteller, wie Apple, Google oder Amazon können frei gewählt werden. Diese ermöglichen es E-Mails während der Fahrt mit der integrierten Microsoft-Outlook-App zu diktieren, Kalendereinträge im Google Kalender zu erstellen oder die Lieblingsplaylist in Spotify abzuspielen. Um jeder Kundengruppe gerecht zu werden, reicht es nicht aus, z.B. nur Apple-Dienste exklusiv anzubieten. Somit befindet sich eine Vielzahl an Diensten im Fahrzeug, die über ähnliche Services verfügen. Abspielen von Musik ist beispielsweise neben Spotify mit Amazon Music, Google Music oder Apple Music möglich. Darüber hinaus kann das parkende Fahrzeug via Handyapp aufgefunden oder mit dieser von der Ferne aus abgesperrt werden. Zusätzlich existiert eine Reihe von Service-Websites um z.B. Schäden zu melden, Werkstatttermine zu vereinbaren usw. Das Angebot all dieser Dienste ist eine große Investition für den Automobilhersteller.
So ist es nicht verwunderlich, dass ein großes Interesse besteht herauszufinden, welche Dienste besonders häufig und welche eher sporadisch genutzt werden. Des Weiteren liegt das Interesse darin, die Kundenzufriedenheit durch das schnellere Finden von beliebten Diensten zu erhöhen.

Fazit

Zusammenfassend lässt sich sagen, dass Firmen, die ihre immer anspruchsvoller werdenden Kunden verstehen und ihre Dienste gezielt auf die Bedürfnisse anpassen können, zukünftig in der Lage sein werden sich von der Konkurrenz abzusetzen und ihre Marktstellung auszubauen.

Literatur

Herrmann, F. 2018. “The Smart Factory and Its Risks.” *Systems* 6, no. 4: 38.

Mathis, C. 2017. “Data Lakes”. *Datenbank-Spektrum*, 17(3), 289-293.

Pasupuleti, P. and B. S. Purra. 2015. “Data Lake Development with Big Data”. Packt Publishing Ltd.

Töpfer A. and A. Mann. 2008. “Kundenzufriedenheit als Basis für Unternehmenserfolg”. In *Handbuch Kundenmanagement: Anforderungen, Prozesse, Zufriedenheit, Bindung und Wert von Kunden 2008,* A. Töpfer. Springer-Verlag, 37-79.

Kontakt

Dr. Nora Vollmers wurde in Mönchengladbach geboren und studierte theoretische Physik an der Universität Paderborn, wo sie 2016 in diesem Fachbereich ebenfalls

promovierte. Danach arbeitete sie als quantitative Analystin bei einer japanischen Investmentbank in Hong Kong. Bei PROCON IT ist sie Big Data Engineer und vorrangig für das Big Data Backend in diversen Projekten verantwortlich. Ihre E-Mail-Adresse lautet Nora.Vollmers@procon-it.de.

Andreas Berghammer wurde in Dachau geboren und absolvierte eine Ausbildung zum Fachinformatiker bevor er 2016 seinen Bachelor in Wirtschaftsinformatik berufsbegleitend an der Technischen Hochschule Deggendorf abschloss. Als Projektleiter für Big-Data-Projekte bringt er umfangreiche Erfahrung in den Bereichen SAP BW, BI-Infrastruktur sowie der Entwicklung diverser Data-Pipelines im Big-Data-Umfeld mit. Seine E-Mail-Adresse lautet Andreas.Berghammer@procon-it.de.

Luca Trautmann wurde in München geboren und schloss ihren Bachelor in Statistik an der Ludwig-Maximilians-Universität München ab. Weiterführend begann sie 2017 mit dem Studium zum Master der Statistik mit sozial- und wirtschaftswissenschaftlicher Ausrichtung. Bei PROCON IT ist sie im Zuge ihrer Masterarbeit als Werkstudentin tätig. Ihre E-Mail-Adresse lautet Luca.Trautmann@procon-it.de.