

stefan PAPP
wolfgang WEIDINGER
mario MEIR-HUBER
bernhard ORTNER
georg LANGS
rania WAZIR



HANDBUCH DATA SCIENCE

Mit **Datenanalyse**
und **Machine Learning**
Wert aus Daten generieren

HANSER

Kunden zum Unternehmen. Konkret soll hier ersichtlich werden, in welchem Kontext der Kunde das Unternehmen aktuell kontaktiert hat.

Der letzte Schritt ist das Vorhersagen von möglichen künftigen Interaktionen sowie Upsell- und Cross-Sell-Möglichkeiten.

Als oberste Maxime gilt dabei das Ziel, den vorhandenen Kundenstamm zu halten, d. h. die Customer Retention zu maximieren, und zugleich in neue Segmente vorzustoßen. Die Grundlage hierfür können interne (CRM, ERP . . .) sowie externe Daten (Social Media) sein, die von unterschiedlichen Datensilos auf einer zentralen Plattform gespeichert und mit Analysen ausgewertet werden.

Dabei kann Predictive Analytics helfen, entsprechende Muster zu erkennen und diese auszuwerten. Beispielsweise kann man feststellen, wann ein Kunde den Verkaufsprozess verlässt, und eventuell lässt sich der Grund hierfür ermitteln. Diese Erkenntnis ist die Grundlage, um entsprechende Marketingmaßnahmen zu optimieren und auszubauen.

1.1.4 Zusammenfassung

Der Erfolg dieser Strategien lässt sich an einem erhöhten Return on Invest und einer reduzierten Time to Market ablesen. Da die Umsetzung von Visionen immer einen erhöhten Bedarf an Ressourcen hat, ist ein Top-down-Ansatz, bei dem die Strategie von einem Mitglied des Vorstandes angestoßen und in den Abteilungen umgesetzt wird, besser geeignet als ein Bottom-up-Ansatz, bei dem die Vision in einer Abteilung entsteht und dann auf weitere Abteilungen ausgedehnt wird.

Beim Top-down-Ansatz werden etwaige Organisationsprobleme in den einzelnen Abteilungen minimiert, da die Abteilungen von Anfang an involviert sind. Somit steht von Anfang an die Gesamtlösung im Fokus. Wir gehen in [Kapitel 8](#) genauer darauf ein.

Erstes Ziel der Gesamtlösung kann z. B. sein, die Anzahl der Datensilos zu verringern, und so einen konzernweiten Data Lake aufzubauen. Der nächste Schritt wäre dann, innerhalb des Data Lake die Daten intelligent zu verknüpfen, um einen Mehrwert zu generieren, da die Daten nun harmonisiert, d. h. in einem zentralen Punkt gesammelt, aufbereitet werden.

Parallel dazu können einzelne Fachabteilungen beginnen aus den Daten einen Mehrwert zu generieren, indem sie ihre Probleme anhand dieser Daten lösen. Das führt zu diversen Prototypen oder Initiativen, die in ein „Digitales Investment“ umgewandelt werden können. Um den maximalen Nutzen aus den Daten zu generieren, ist auch eine

entsprechende Governance-Strategie notwendig, um die Datenverarbeitungsschritte und somit die Wertschöpfungskette zu dokumentieren und die angewandten Schritte nachvollziehbar zu machen. Zusätzlich erhöht die Governance den Grad der Vertrauenswürdigkeit der Plattform, da nun ersichtlich ist, woher die Daten kommen und wohin sie gespeichert werden.

Im Kern der Entwicklung des digitalen Geschäftsmodells steht also eine Transformation vom verarbeitenden Gewerbe hin zu einem Dienstleistungsgewerbe, für die ein kontinuierliches Abonnement nötig ist oder die pro Volumen bezahlt wird. Das heißt, „dass physische Produkte nur noch Hilfsmittel sind, um ein Ziel zu erreichen oder ein Kundenproblem zu adressieren“ ¹ und so einen kontinuierlichen Cashflow zu generieren, anstatt einen einmaligen.

Gegenüber der traditionellen Software-Industrie ergeben sich in der Daten-Industrie folgende Unterschiede:

- Es gibt eine zentralisierte Datenplattform.
- Kunden sind Miterzeuger der Lösungen oder Produkte.
- Neue Lösungen und Produkte können laufend optimiert und skaliert werden.
- Es bestehen erhöhte Anforderungen an die Automatisierung sowie die Qualitätssicherung aufgrund der Menge an Daten, die verarbeitet werden.
- Die Organisationsform der Personen, die Big Data anwenden, ist agil und passt sich der Problemstellung kontinuierlich an. Sie erlaubt auch entsprechende Fehler zu machen, folgt aber den vorher definierten Anwendungsfällen bzw. der Vision.

Im Folgenden werden nun die Bereiche vorgestellt, die eine generelle Digitalisierungsstrategie für die Entwicklung eines Geschäftsfeldes in der Regel umfasst.

1.2 Umsetzung einer Digitalisierungsstrategie

Die Digitalisierungsstrategie leitet sich von der Strategie des Unternehmens ab und unterstützt diese mit Big-Data-Werkzeugen. In der Digitalisierungsstrategie wird der operative Fokus von Big-Data-Initiativen festgelegt und es wird definiert, wie Daten und die darauf aufbauende Modellierung geschäftsseitig interpretiert werden. Sie legt auch

fest, wie aus Daten Informationen gewonnen werden und welche Fragen des Unternehmens beantwortet werden.

1.2.1 Daten

Daten liegen in Unternehmen aktuell horizontal, d. h. in verschiedenen Abteilungen, oder vertikal, d. h. gekoppelt an verschiedene Funktionen, fragmentiert und in Silos vor. Zusätzlich wächst der Anteil der kritischen Informationen, die außerhalb der üblichen Prozesse generiert werden. Das Ziel der Datenstrategie ist also einen Prozess zu erstellen, der verschiedenste Datenformate verarbeiten und in ein strukturiertes und orchestrierbares Format überführen kann. Dabei lassen sich die Daten in vier verschiedenen Eigenschaften beschreiben.

- **Volumen:** Beschreibt die Menge der Daten, die durch tägliche Geschäftsprozesse erhoben werden. Hierbei handelt es sich um eine Größenordnung – wie etwa Gigabyte, Terabyte oder Petabyte.
- **Velocity:** Beschreibt die Geschwindigkeit der Daten, die während einer Session oder Transaktion generiert werden. Sensordaten sind typischerweise Daten mit einer sehr hohen Geschwindigkeit, da sie oftmals unmittelbar verarbeitet werden müssen. Erkennt man beispielsweise mit Sensoren Probleme in der Produktion, so will man innerhalb weniger Minuten reagieren können.
- **Veracity:** Dieser Wert beschreibt die Vertrauenswürdigkeit bzw. die Genauigkeit der Daten. Als vertrauens erhöhende Maßnahmen können beispielsweise Data Lineage, also das Nachzeichnen der einzelnen Datenverarbeitungsschritte und Datenströme, sowie entsprechende Signaturmechanismen verwendet werden. Zusätzlich kann ein Wasserzeichen mitgeführt werden, das den letzten Zeitpunkt der Verwendung dokumentiert.
- **Variety:** Beschreibt die Varietät der Daten. So muss eine gesamte Plattform unterschiedliche Daten, wie Sprachdaten oder Textdaten, konsumieren können, um effektiv und effizient betrieben werden zu können. Zusätzlich muss sie auch die gängigen Konnektoren zu den einzelnen im Unternehmen verwendeten Schnittstellen haben, um so die benötigten Daten effizient zur Verfügung zu stellen.

1.2.2 Modellierung und Analyse

Daten selbst erzeugen noch keinen Wert für das Unternehmen. Um aus den Daten einen Wert zu generieren, werden analytische Modelle oder Algorithmen benötigt. Typische Anwendungsfälle sind Optimierungsaufgaben, Voraussagen für die nächste Rechnungsperiode oder Risikoklassifizierungen. Es muss also ein Plan erstellt werden, wie die analytischen Ressourcen, wie Personal oder Hardware, am effizientesten und effektivsten eingesetzt werden können und somit das meiste Potenzial haben, einen Wert für das Unternehmen zu generieren. Dabei müssen auch das automatisierte Training und die Anpassung an neue, dem System noch unbekannte Werte berücksichtigt werden. Eine mögliche Lösung wäre beispielsweise etwaige Probleme bereits während des System-Designs zu berücksichtigen oder während des Betriebs darauf angemessen zu reagieren.

Hierbei ist es wichtig, den gesamten Ende-zu-Ende-Prozess im Auge zu behalten, einschließlich der Anwender, der Komplexität der Lösung sowie des Managements der Modelle und der Daten selbst.

Ende-zu-Ende-Prozess der Datenverarbeitung

Die Resultate der technischen Überlegungen, die während der Analyse-Phase getroffen worden sind, beispielsweise welche Daten wie analysiert werden, müssen für die Entscheidungsträger und Stakeholder aufbereitet werden. Der Mehrwert der Arbeit ergibt sich erst, wenn die Endbenutzer, beispielsweise Manager oder Angestellte, die Analyse-Ergebnisse direkt verwenden können.

Ob sie verwendet werden, hängt meist von einigen nichttechnischen Parametern ab, wie Nachvollziehbarkeit und Bedienbarkeit der Benutzeroberfläche. Zusätzlich werden organisatorische Ressourcen benötigt, beispielsweise jemand, der die Strategie aktiv weiter voranbringt sowie Leute mit den entsprechenden analytischen und technischen Fähigkeiten, um den Ende-zu-Ende-Prozess umzusetzen.

Wird nun ein Produkt, das physisch vorliegt, in eine Dienstleistung transformiert, dann muss das neue Produkt ein Bedürfnis oder ein Problem besser lösen, als dies bisher der Fall war, und so die Customer Experience weiter steigern, um sich durchzusetzen.

Daher ist es unerlässlich, die strategischen Ziele mit den längerfristigen operativen Zielen des Business in Einklang zu bringen. Eine solche Strategie kann unterschiedlich angewandt werden und benötigt mitunter einige Zeit, insbesondere wenn Investitionen in Hardware-Ressourcen notwendig werden. Alternativ zu den eigenen Investitionen kann auch ein Cloud-Ansatz gewählt werden, in dem die ganze Infrastruktur in der Cloud, etwa AWS oder Azure, angeboten wird. Das hat den Vorteil, dass schnell

begonnen werden kann, jedoch steigen die Kosten mit der Zeit. Zusätzlich entsteht ein Vendor-Lock-in, da nach der Wahl eines Cloud-Providers ein Wechsel nur mit erheblichem Mehraufwand möglich ist.

Im Gegensatz dazu gibt es den On-Premise-Ansatz, der die physische Hardware in der Firma aufbaut und somit erhebliche Investitionen zu Beginn des Projekts erfordert. Im On-Premise-Ansatz entstehen wesentlich höhere Kosten, da eine dynamische Ressourcenzuteilung nicht möglich ist. Man muss die gesamte Kapazität vorhalten und kann auf geänderte Anforderungen nur mit erheblichem Einsatz neuer Ressourcen reagieren.

Die Umsetzung der Digitalisierungsstrategie umfasst folgende Überlegungen:

- Die erhobenen Daten des angebotenen Service bilden die Grundlage für weitere Aktionen. Diese Daten können beispielsweise verwendet werden, um den Service selbst zu optimieren und etwaige Schwachstellen zu beheben.
- Über einen möglichst vernetzten digitalen Kanal kann mit dem Endbenutzer, dem Service oder anderen Partnern kommuniziert werden. Zusätzlich lässt ein digitaler Kanal die Anwendung von analytischen Modellen zu, da diese meist erheblich mehr Ressourcen benötigen, als bei der Datenerhebung verfügbar wären.
- Durch eine zentrale Plattform werden alle Prozesse und Services automatisiert verwaltet und stehen so schnell und übersichtlich dem entsprechenden Endbenutzer oder der Business Unit zur Verfügung.

Datenprodukte

Daten und Informationen können beispielsweise über Sensoren am physischen Produkt (IoT) oder in Social-Media-Anwendungen sehr nahe am Entstehungsort gesammelt und ausgewertet werden. Dabei wird letztendlich Wissen, wie ein Service oder Produkt benutzt wird, generiert. Das Produkt wird im ersten Schritt in ein intelligentes Produkt verwandelt, das heißt das Produkt erhält die Fähigkeit seinen physikalischen Zustand zu beschreiben, indem es mit Sensoren ausgestattet wird.

Ein Wesensmerkmal von intelligenten Produkten aus Kundensicht ist, dass sie einen Zusatznutzen zum herkömmlichen Produktnutzen liefern, der durch Datensammlung zustande kommt. Ein Beispiel hierfür wäre ein Gabelstapler, in dem Sensoren verbaut sind. Neben der eigentlichen Funktion als Gabelstapler, kann nun der Hersteller z. B. ein Logistikprogramm anbieten, sofern er entsprechende Sensoren verarbeitet hat, um den