

mehr als die Hälfte der Männer in dieser Altersgruppe zu.

Das Wissen um die Einführung einer Digitalen Rentenübersicht ist gering. Bislang hat nur ein Fünftel der Befragten von der Digitalen Rentenübersicht gehört.

Die Ergebnisse zum Kenntnisstand über die zu erwartenden Rentenleistungen korrelieren mit Erkenntnissen einer forsa-Umfrage im Auftrag des GDV⁵. Danach haben sich rund zwei Drittel der Deutschen im Alter zwischen 18 und 65 Jahren schon intensiver mit dem Thema Altersvorsorge beschäftigt. Für ein Drittel der Befragten war dies bis dato nicht der Fall. Dabei sticht gerade die Gruppe der 18 bis 29jährigen hervor, da sich von ihnen deutlich über 50% noch nicht mit dem Thema Altersvorsorge beschäftigt haben und über 40% von ihnen angeben, nicht für das Alter vorzusorgen.

Umso wichtiger ist es, jedem einzelnen mit der Digitalen Rentenübersicht in klarer und verständlicher Weise aufzuzeigen, wie sich unter den gegebenen Voraussetzungen die eigenen Rentenansprüche weiter entwickeln werden. Insbesondere die Berufseinsteiger haben noch großen Informations- und Beratungsbedarf. Die Digitale Rentenübersicht kann damit ein zusätzlicher und wichtiger Baustein für eine umfassende Beratung sein, die alle bislang ergriffenen Vorsorgemaßnahmen beinhaltet und bestehende Vorsorgelücken aufzeigt.

¹ Begründung zum Gesetzesentwurf, A. Allgemeiner Teil, Abschnitt 2, 21.10.2020

² Stellungnahme des GDV zum Referentenentwurf.

³ § 6 RentÜG, Abs. 3: „Die erste Betriebsphase soll 21 Monate nach Inkrafttreten dieses Gesetzes beginnen und nach zwölf Monaten enden. [...]“.

⁴ Umfrage vom 12.3.2021, durchgeführt im Auftrag der Aeiforia GmbH

⁵ forsa-Umfrage „Altersvorsorge“, veröffentlicht am 21.02.2019.

Dr. Annika Bergbauer / Alexander Horn

Datenprojekte scheitern viel zu oft: Entscheider müssen die Risiken kennen und managen

Datengetriebene Geschäftsmodelle und Produkte sind die Verheißung der Stunde. Daten werden häufig als das neue Öl bezeichnet. Sie sind ein Rohstoff, dessen Vorkommen sich permanent und in atemberaubender Geschwindigkeit vergrößert: Die International Data Corporation (IDC) schätzt, dass sich die weltweite Datenmenge von rund 33 Zettabyte im Jahr 2018 auf 175 Zettabyte im Jahr 2025 mehr als verfünffachen wird¹.

Klar ist: Die wirtschaftliche Nutzung der enormen Datenmengen hat bereits heute eine hohe praktische Relevanz. Eine Bitkom-Umfrage aus dem Jahr 2020 zeigt, dass für 85% der deutschen Unternehmen mit mehr als 50 Beschäftigten die intensive Datennutzung heute und in Zukunft eine sehr große Rolle spielt². Das Problem: Laut Schätzungen scheitern 60 bis 80% aller Datenprojekte³.

Auch Versicherungsunternehmen stehen vor dieser Herausforderung und müssen die Risiken von Datenprojekten zielgerichtet managen. Voraussetzung dafür ist unter anderem, die typischen Fehler zu kennen und zu vermeiden. Dabei mag es vielleicht überraschen, dass Datenprojekte in der Regel nicht aus (daten-)technischen, sondern aus menschlichen Gründen fehlschlagen. Die folgenden Risiken sind die fünf häufigsten Gründe für das Scheitern von Datenprojekten.

Risiko 1: Zu hohe Erwartungen

Daten und Algorithmen sind nicht die magische Lösung aller Probleme. Denn Datenanalysen können nur die Probleme lösen, für die sie designt wurden, und das auch nur so gut, wie die vorhandenen Daten sind. Beispielsweise ermöglichen es Datenanalysen, Muster im Kaufverhalten zu erkennen, dem Kunden daraufhin Angebote mit einer höheren Erfolgswahrscheinlichkeit zu machen (Next Best Offer) und so den Absatz von Policen.

Die beste Datenanalyse nutzt jedoch wenig, wenn die dafür genutzten Daten auf-

grund der geringen Kontaktfrequenz der Kunden mit der Versicherung hoffnungslos veraltet sind. Des Weiteren wird die Erwartungshaltung häufig durch Marketing erhöht. Denn Aussagen wie „Fortune 1000 Unternehmen konnten ihren jährlichen Gewinn durch Datenanalyse steigern“ oder „Return on Equity wuchs durch verbesserten Datenzugang“ sind bestenfalls Korrelationen und keine Kausalitäten.

Risiko 2: Überschätzung der Umsetzungsgeschwindigkeit

Datenverständnis und -vorbereitung nehmen viel mehr Zeit in Anspruch, als Entscheider glauben. Umfragen unter Datenexperten ergaben, dass sie bis zu 80% ihrer Zeit mit der Datenvorbereitung statt mit der Datenanalyse verbringen⁴. Häufig sind mehrere Abstimmungsschleifen zwischen Fachbereich und Datenexperten notwendig, um die Daten zu verstehen und das Ziel zu konkretisieren.

Risiko 3: Der Ergebnisbeitrag ist unklar

Wird nicht klar definiert, welchen Beitrag das Datenprodukt leisten soll, stiftet es oft nur geringen Nutzen oder wirkt sogar kontraproduktiv. Ein Beispiel: Entscheider möchten das trendy Gadget der Konkurrenz nutzen – einen Next-Best-Offer-Algorithmus. Schneidet das Unternehmen im Verkauf allerdings bereits gut ab, die Risikoselektion ist jedoch schlecht, kann das Datenprodukt dieses Problem nicht lösen und es senkt die Akzeptanz für Datenprojekte, da sie nutzlos erscheinen. Das heißt, ohne eine fachliche Analyse und Definition der operativen Problemstellungen kann kein nützliches Datenprodukt erzeugt werden.

Dr. Annika Bergbauer

Consultant bei der Business- und IT-Beratung Q_PERIOR.

Alexander Horn

Associate Partner bei Q_PERIOR und Leiter des Bereichs Process Intelligence & Architecture.

Risiko 4: Mangelnde Akzeptanz bei den betroffenen Einheiten

Dass viele Jobs durch Algorithmen ersetzt werden ist eine weitverbreitete, jedoch in der Regel unberechtigte Angst. Darum sollten Entscheider in Datenprojekten von Anfang an alle Stakeholder einbinden, um diese Ängste zu nehmen und die Mitarbeiter zu begeistern. Denn mangelnde Akzeptanz führt zu inneren Verweigerungshaltungen und schlimmstenfalls zu einem bewussten falschen Training von Algorithmen.

Risiko 5: Die Expertise in den Unternehmen ist noch zu gering

Datenexperten sind Mangelware. Auch die Assekuranz spürt diesen Fachkräftemangel, da sie häufig noch nicht als attraktiver IT-Arbeitgeber wahrgenommen wird. Das liegt meist daran, dass Versicherer überwiegend noch keine klare Datenstrategie haben und folglich noch nicht für Datenexperten attraktiv sind, die gerne mit Gleichgesinnten an spannenden Projekten arbeiten möchten.

Wie können die Risiken für das Scheitern von Datenprojekten verringert werden?

Beginnen wir direkt mit letztgenanntem Risiko. Mangelnde Datenexpertise im Unternehmen kann durch zwei Ansätze behoben werden: Durch die Zusammenarbeit mit externen Partnern (zum Beispiel mit spezialisierten Beratungen) oder das Einstellen beziehungsweise Ausbilden eigener Datenexperten. Mit externen Partnern kann schnell, aber vergleichsweise kostspielig, Expertise an Bord geholt werden. Interne Expertise aufzubauen dauert zwar länger, ist aber tendenziell günstiger, zumal die kurzfristige Lösung über externe Experten schon aus Kostengründen nicht den Aufbau eigener Expertise ersetzt.

Darüber hinaus kann der Aufbau interner Datenexpertise auch über die Schulung sogenannter Data Citizens⁵ erfolgen. Diese Fachkollegen können dann mit low- oder no-code-Software, wie zum Beispiel der SAP Analytics Cloud oder Microsoft Power BI, ohne Programmierkenntnisse erste Datenvisualisierungen durchführen. Jedoch sind die Möglichkeiten einer solchen Software für die Erstellung tiefgreifender Datenanalysen begrenzt, da sie durch ihre kurze Entwicklungshistorie ihr Potenzial noch nicht völlig ausgeschöpft haben und statisti-

sches Fachwissen letztlich nicht völlig ersetzt werden kann. Beispielsweise können no-code-Analysetools selbstständig Korrelationen aber keine Kausalitäten bestimmen.

Das CRISP-DM Modell zur Datenverarbeitung

Zentrales Hilfsmittel bei der Durchführung von Datenprojekten sind etablierte Standards, durch welche die Risiken für das Scheitern verringert werden können. Das Datenverarbeitungskonzept CRISP-DM (Cross-Industry Standard Process for Data Mining) gibt unter anderem einen Rahmen für die Lösung der oben genannten Risiken 1 bis 4. Insgesamt gliedert sich das Modell in 6 Phasen.

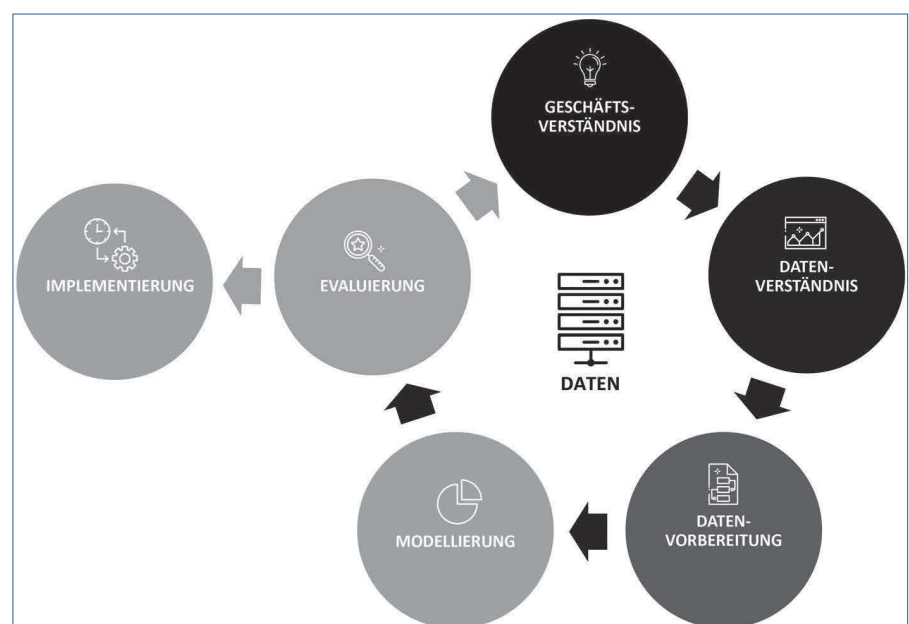
In *Phase 1* verstehen Datenexperten das Geschäft, um die Zusammenhänge in den Daten zu erkennen. Aus dem Geschäftsverständnis, dem identifizierten operativen Problem und einem klar umrissenen Ergebnisbeitrag wird dann das konkrete Ziel definiert. Damit steuert Phase 1 des CRISP-DM dem dritten Risiko für Datenprojekte entgegen, wenn der Ergebnisbeitrag und der Nutzen des Datenprodukts unklar sind.

Betrachten wir dies am Beispiel Next Best Offer. Nehmen wir an, es existiert ein Defizit im Vertrieb von Versicherungsprodukten bei Bestandskunden (zu geringe Bestandsdurchdringung). Als Ursache für das Vertriebsdefizit kommt zum einen

„Versicherer haben überwiegend noch keine klare Datenstrategie und sind folglich noch nicht für Datenexperten attraktiv, die gerne mit Gleichgesinnten an spannenden Projekten arbeiten möchten“

mangelndes Verkaufsgeschick in Frage. Zum anderen ist es aber auch möglich, dass der Versicherer seine Bestandskundendaten noch nicht in Erfolg versprechende Vertriebs-Leads übersetzen kann. Next Best Offer kann helfen, die Erfolgswahrscheinlichkeit der Kundenkontakte zu steigern, da die Leads zutreffender

Abbildung: CRISP-DM: In jeder Phase kann eine Rückkopplung auf die vorangehenden Phasen erfolgen.



sind. Mangelndes Verkaufsgeschick wird dadurch nicht gelöst.

Betrachtet man diese zwei Ursachen, wird deutlich, dass das Datenprodukt auf die Ursache des operativen Problems abgestimmt sein muss. Daher sollten Datenexperten und Entscheider offen besprechen, was sie leisten können und was nicht. Denn ein Next-Best-Offer-Algorithmus kann zwar Vorschläge mit hoher Abschlusswahrscheinlichkeit generieren, kann aber nicht garantieren, dass die Versicherungsprodukte dann auch wirklich von den Kunden gekauft werden (Risiko 1).

Die mangelnde Umwandlung des vom Algorithmus vorgeschlagenen Produkts in einen Kauf kann unter anderem daran liegen, dass die Kundenberater dem Algorithmus misstrauen und sich nicht an die Empfehlung des Programms halten. Um mangelnde Akzeptanz (Risiko 4) zu verhindern, sollten alle Beteiligten des Datenprojektes bereits in Phase 1 planvoll angesprochen und auf die damit einhergehenden Veränderungen vorbereitet werden. Dies sollte mittels einer Stakeholderanalyse und einer darauf aufbauenden Kommunikationsstrategie erfolgen. Das Einbeziehen von Betroffenen – also ein aktives Teilen von Wissen und die Möglichkeit zur Mitwirkung an oder zumindest die Erläuterung von Entscheidungen – ist unverzichtbar für die Akzeptanz.

Die Struktur der Daten wird in *Phase 2* verstanden. Es werden erste deskriptive Statistiken erzeugt, also Mittelwerte und Gruppen (zum Beispiel Geschlechter- oder Altersverteilung) betrachtet. Auf Basis dieser Betrachtung kann gegebenenfalls schon ersichtlich werden, dass die vorhandenen Daten nicht zum gewünschten Ziel passen. Sind zum Beispiel nur Daten aus einer Versicherungssparte vorhanden, sind die Möglichkeiten für eine Ableitung guter Cross-Selling-Potenziale begrenzt.

Stimmen die vorhandenen Daten noch nicht mit der Zielsetzung überein, muss eines oder beides angepasst werden. Hier ist der Dialog zwischen Datenexperten und Entscheidern notwendig.

Auf Basis des Datenverständnisses werden die Daten in *Phase 3* vorbereitet. Daten mit unterschiedlichen Formaten und Maßeinheiten werden vereinheitlicht, anonymisiert, fehlende Datenpunkte ersetzt,

Ausreißer und offensichtliche Fehler entfernt. Diese Datenmanipulation sollte nie ohne das Wissen und die Expertise der Fachkollegen und Entscheider passieren, da sie das Analyseergebnis stark beeinflussen kann. Fachkollegen können erkennen, ob auffällige Datenpunkte wirklich Fehler sind oder es gegebenenfalls ein historisches Ereignis gab (wie eine Server-Downtime oder eine Finanzkrise), das diese scheinbare Dateninkonsistenz erklärt.

Erfahrungsgemäß nimmt die Datenaufbereitung in Phase 3 von allen Phasen am meisten Zeit in Anspruch (Risiko 2), da sie kaum Standardentscheidungen beinhaltet. Klein erscheinende Änderungswünsche erzeugen viel Arbeit im Code, denn es können sich Verschiebungen im gesamten Datensatz ergeben. Allein das erneute Erzeugen der Analyse kann einige Zeit beanspruchen, da große Datensätze lange Rechenzeiten haben. So gibt es ein agiles vor und zurück zwischen den ersten drei Phasen. Entscheider sollten unbedingt ausreichend Zeit und Budget für die Datenvorbereitung bereitstellen.

Sind die Daten vorbereitet, kann in *Phase 4* in Abhängigkeit von Ziel und Daten ein Analysemodell gewählt werden. Stehen zum Beispiel mehrere Millionen Datenpunkte zur Verfügung und es sollen Korrelationen gefunden werden, bietet sich maschinelles Lernen an. Soll hingegen ein kausaler Zusammenhang gefunden werden – soll also herausgefunden werden, welcher Faktor der Treiber für andere Faktoren ist – bieten sich Regressionen aus der klassischen Statistik an.

In *Phase 5* werden die Ergebnisse plausibilisiert. Bleiben wir beim Beispiel Next Best Offer und nehmen an, dass sehr hohe und aus der praktischen Erfahrung unplausible Abschlusswahrscheinlichkeiten berechnet werden. Aufgrund dieser unplausiblen Ergebnisse könnte festgestellt werden, dass der zugrunde liegende Datensatz auch solche Individuen mit einbezieht, die das vorgeschlagene Produkt bereits besitzen. In diesem Fall muss der Datensatz angepasst werden. Selten erhält man bei der ersten Vorlage der Ergebnisse ein abschließendes Resultat, wenn die Fachexperten auf Basis ihrer Erfahrung ein anderes Ergebnis erwarten und Datenexperten daraufhin weitere Evidenz erzeugen müssen. Oder bei der Besprechung tauchen weitere Datenquellen auf, an die vorher nicht gedacht wurde.

Hat die Evaluierung ein plausibles Ergebnis gezeigt, wird das Datenprodukt in der letzten Phase (*Phase 6*) implementiert. In dieser Phase wird die Anwendung in die vorhandene Systemlandschaft integriert. Beispielsweise soll der Next-Best-Offer-Algorithmus als Webapplikation präsentiert werden. Hierzu entwickeln Anwendungsdesigner eine Oberfläche und binden den Algorithmus im Backend an.

Fazit

Der CRISP-DM-Ansatz bietet die Möglichkeit, die mit Datenprojekten einhergehenden Risiken aktiv zu managen. Die vielleicht wichtigste Best Practice ist die ehrliche und vertrauensvolle Kommunikation mit mindestens den wichtigsten Stakeholdern während aller Phasen des Datenprojektes. Seien Sie offen für ein vor und zurück zwischen den Phasen der Datenverarbeitung. Starten Sie ihr Datenprojekt mit einem klar erfassten operativen Problem, beschreiben Sie es genau und erörtern Sie mit Datenexperten die Möglichkeiten und Unmöglichkeiten einer Lösung durch ein Datenprodukt. Denn trotz der Risiken sind Datenprojekte die unverzichtbare Basis für innovative datengetriebene Geschäftsmodelle.

¹ Studie des US-Festplattenherstellers Seagate und des IT-Marktbeobachters IDC: <https://www.seagate.com/de/de/our-story/data-age-2025/>

² <https://www.bitkom.org/Presse/Presseinformation/Ohne-Daten-laeuft-in-der-deutschen-Wirtschaft-kaum-noch-etwas>.

³ <https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/>.

⁴ <https://rapidminer.com/blog/data-prep-time-consuming-tedious/>; <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>

⁵ *Data citizens haben ein konzeptionelles Verständnis von analytischen Methoden, coden aber nicht selbst, sondern arbeiten mit low- oder no-code Software.*