

## **Stornoanalyse in einem Unfallversicherungsbestand**

Andreas Reuß  
Universität Ulm  
Sektion Aktuarwissenschaften  
89069 Ulm  
reuss@mathematik.uni-ulm.de

Hans-Joachim Zwiesler  
Universität Ulm  
Sektion Aktuarwissenschaften  
89069 Ulm  
zwiesler@mathematik.uni-ulm.de

### **Zusammenfassung**

Bei der Unfallversicherung handelt es sich um eine für deutsche Versicherungsunternehmen profitable Sparte, in der der Wettbewerbsdruck in den letzten Jahren zugenommen hat. Beim betrachteten Versicherungsunternehmen führten die bisherigen Anstrengungen zur Steigerung des Neugeschäfts nicht zur erwünschten Erhöhung des Marktanteils, da im vorhandenen Bestand an Versicherungsverträgen ein hoher Anteil an Stornierungen zu verzeichnen war. Um zukünftig gezielte Maßnahmen zur Senkung der Stornoquote ergreifen zu können, ist es deshalb notwendig, das Stornoverhalten im Bestand detailliert zu analysieren. Hierbei sind die Besonderheiten der Unfallversicherung zu beachten wie beispielsweise die häufig vereinbarte mehrjährige Mindestlaufzeit von Verträgen.

Im Rahmen einer Fallstudie wurden Storno- und Bestandsdaten aus der operativen Datenbank eines Unfallversicherers mit Hilfe des SAS Enterprise Miner analysiert. Dabei ist es trotz der begrenzten Datengrundlage gelungen, Risikofaktoren für Storno sowie Bestandssegmente mit hoher Stornoquote zu identifizieren. Außerdem wurde mit Hilfe eines Entscheidungsbaums ein Modell zur Prognose des Stornoverhaltens entwickelt, das als Grundlage für gezielte Marketingaktionen dienen kann. Aus den Analysen ergaben sich ferner Vorschläge für eine Verbesserung der Datenhaltung. Die Vorgehensweise orientierte sich am Prozessmodell CRISP-DM (Cross-Industrial Standard Process Modell for Data Mining).

**Keywords:** Unfallversicherung, Storno, CART, Enterprise Miner, CRISP-DM

## **1 Einleitung**

Daten, Daten, Daten: diese Charakterisierung trifft für Unternehmen in allen Branchen zu, insbesondere auch für Versicherungsunternehmen. Denn in Versicherungsunternehmen werden tagtäglich Daten in großer Menge erzeugt, beispielsweise im Bestandsführungssystem, im Inkassosystem, beim Außendienst usw. Die moderne Informationstechnologie ermöglicht die effiziente Erfassung und Speicherung der erzeugten Daten.

Diese riesigen Datenmengen bergen eine Unmenge an Informationen und Zusammenhängen, an die bei ihrem Erstellen noch gar nicht gedacht wurde. Ferner werden die Daten oft ab einem gewissen Zeitpunkt weder weiterverwendet noch systematisch ausgewertet.

Ziel dieses Artikels ist es, anhand eines Fallbeispiels aufzuzeigen, wie aus solch riesigen Datenmengen nützliche Informationen extrahiert und schließlich unternehmerisch genutzt werden können. Dabei können bereits in einem ersten Schritt mit begrenztem Aufwand äußerst interessante Erkenntnisse gewonnen werden. Erfahrungen aus anderen Branchen wie z.B. dem Versandhandel zeigen, dass solche Informationen ein wichtiger Wettbewerbsfaktor sein können.

Um nützliche Informationen zu extrahieren, ist eine systematische Analyse der Kundendaten notwendig. In diesem Zusammenhang wird häufig das Schlagwort Data Mining genannt; die Umsetzung der damit verbundenen Ideen ist jedoch oft unklar. Neben der korrekten Anwendung der zum „Datenschürfen“ zur Verfügung stehenden Techniken ist zur Zielerreichung nämlich eine Vielzahl weiterer Schritte notwendig, von der sorgfältigen Zieldefinition bis hin zur Überprüfung der Ergebnisse. Insbesondere ist ein strukturiertes Vorgehen zwingend erforderlich. Dies wird im folgenden Abschnitt beispielhaft für die Stornoanalyse in einem Unfallversicherungsbestand beschrieben. Die Vorgehensweise orientierte sich dabei am Prozessmodell CRISP-DM (Cross-Industrial Standard Process for Data Mining), welches im dritten Abschnitt allgemein diskutiert wird. Zur Durchführung der Analysen wurde der SAS Enterprise Miner verwendet.

## **2 Fallbeispiel**

### **2.1 Analyse der unternehmerischen Fragestellungen**

Wir betrachten in einem Fallbeispiel einen deutschen Unfallversicherer und zwar insbesondere den Bereich der Privaten Risikounfallversicherung<sup>60</sup>. Diese Sparte zeichnet sich sowohl durch hohen Wettbewerbsdruck als auch durch hohe Rentabilität aus. In den letzten 25 Jahren lag die Schadenquote in der Privaten Unfallversicherung stets zwischen 50% und 60%, in 2002 erreichte sie branchenweit 52,4%. Das Beitragsvolumen betrug 5,6 Mrd. € und es gab 153 Anbieter von Privaten Unfallversicherungen<sup>61</sup>.

Beim betrachteten Unternehmen hat die Anzahl der Kunden, die ihren Vertrag zum Ende der vereinbarten Laufzeit kündigen, in den letzten 5 Jahren leicht zugenommen. Dies hat dazu geführt, dass der Marktanteil des betrachteten Versicherers trotz eines beachtlichen Neugeschäfts ungefähr konstant geblieben ist. Deshalb stellt sich die Frage, wie die Stornoquote durch geeignete Maßnahmen gesenkt, der Bestand gesichert und damit die Wettbewerbschancen in dieser profitablen Sparte erhöht werden können.

Als erster Schritt soll hierzu das Stornoverhalten in der Vergangenheit detailliert untersucht werden. Ziel ist es dabei, Bestandssegmente mit besonders hoher Stornoquote – so genannte Stornonester - zu identifizieren, zu beschreiben und zu interpretieren. Auf Basis dieser Analysen soll das Stornoverhalten in der Zukunft prognostiziert werden, d.h. es sollen Verträge mit hoher Stornogefährdung identifiziert werden. Durch gezielte, auf die stornogefährdeten Segmente ausgerichtete Marketingaktionen soll dann versucht werden, die Stornoquote signifikant zu reduzieren. Der Erfolg des Projektes lässt sich dabei durch die erreichte Senkung der Stornoquote messen. Die Frage der Rentabilität der einzelnen Verträge wird hierbei zunächst nicht betrachtet.

Aus diesen Unternehmenszielen lassen sich entsprechende Ziele für die Datenanalyse ableiten: Auf Basis von Vergangenheitsdaten soll ein Modell entwickelt werden, welches besonders stornogefährdete Bestandssegmente identifiziert. Dabei soll eine Beschreibung und Interpretation der identifizierten Bestandssegmente möglich sein. Ferner soll das Modell eine möglichst genaue Prognose des zukünftigen Stornoverhaltens erlauben. Als

---

<sup>60</sup> Eine genaue Beschreibung des Produkts findet man bei Riebesell (1991), S.533ff.

<sup>61</sup> Vgl. GDV (2003), S.47.

Beurteilungskriterium dient hierbei der Lift, also das Verhältnis zwischen der Stornoquote in den identifizierten Segmenten und der Stornoquote im Gesamtbestand<sup>62</sup>.

## 2.2 Analyse der Datengrundlage

Im Rahmen des Fallbeispiels standen Daten aus der Data-Warehouse-Datenbank des betrachteten Versicherers zur Verfügung. Diese konnten ohne größere Vorbereitungen zu einem Stichtag aus dem System extrahiert werden. Weitere mögliche Datenquellen wurden zunächst nicht herangezogen. Anhand dieses Fallbeispiels lässt sich also zeigen, dass auch mit begrenztem Aufwand im Hinblick auf die Datenbereitstellung erste interessante Data Mining Analysen möglich sind.

Zur Verfügung standen Daten zu allen in den letzten 5 Jahren stornierten Verträgen, sowie zu allen aktuell aktiven Verträgen. Zu jedem Vertrag wurden dabei allgemeine Vertragsmerkmale, Daten zum Versicherungsnehmer und zu den versicherten Personen, sowie die Details der vereinbarten Versicherungsleistungen gespeichert. Insgesamt handelte es sich um deutlich mehr als 100.000 Verträge mit jeweils mindestens 65 Vertragsmerkmalen.

Bei der Analyse sind zwei Besonderheiten der Risikounfallversicherung zu beachten. Zum einen werden meist Mehrjahresverträge abgeschlossen, wobei die Mindestlaufzeit heute in der Regel 5 Jahre beträgt. Falls der Versicherungsnehmer nach Ablauf dieser Mindestlaufzeit nicht kündigt (mit einer Frist von 3 Monaten), wird der Vertrag automatisch jeweils um ein weiteres Jahr verlängert. Zum anderen ist es möglich, dass im Rahmen eines Vertrages mehrere Personen versichert sind. Insbesondere muss der Versicherungsnehmer nicht notwendigerweise auch versicherte Person sein.

Um einen ersten Überblick über die Struktur der Daten zu erhalten wurden mit Hilfe von Methoden der explorativen Datenanalyse die Verteilungen der wichtigsten Vertragsmerkmale sowie Abhängigkeiten zwischen den Vertragsmerkmalen analysiert.<sup>63</sup> Die Untersuchung der Datenqualität hat gezeigt, dass die fehlende Historisierung wichtiger Vertragsmerkmale ein erhebliches Problem darstellt. Auch traten bei einigen Feldern viele fehlerhafte oder fehlende Werte auf. Dennoch waren die zur Verfügung stehenden Daten zur Analyse der Fragestellung grundsätzlich geeignet.

---

<sup>62</sup> Vgl. Berry und Linoff (2000), S.186ff.

<sup>63</sup> Hierzu wurden insbesondere die Funktionalitäten von SAS/BASE und SAS/STAT herangezogen.

### **2.3 Aufbereitung der Daten**

Im Fallbeispiel wurden alle innerhalb eines Beobachtungszeitraums von 12 Monaten ablaufenden Verträge analysiert. Die in diesem Zeitraum vom Versicherungsnehmer zum Ablauf stornierten Verträge wurden dabei mit den nicht stornierten Verträgen verglichen.

Der Vergleich basierte auf den zur Verfügung stehenden Vertragsmerkmalen, wobei einige Merkmale aufgrund der fehlenden Historisierung bzw. wegen fehlender und fehlerhafter Werte nicht zur Analyse herangezogen werden konnten. Zeitabhängige Merkmale wie Geburtsdatum und Datum des Vertragsbeginns wurden zunächst geeignet transformiert.

Aus den Rohdaten wurden außerdem weitere aussagekräftige Vertragsmerkmale abgeleitet: beispielsweise konnte aus der Tarifikennziffer, die mehr als 1.000 mögliche Ausprägungen hat, eine sinnvolle Einteilung der Verträge in Tarifgruppen abgeleitet werden. Auch die Anzahl der versicherten Erwachsenen bzw. Kinder sowie die gesamte versicherte Invaliditätssumme ergaben sich durch Kombination verschiedener Merkmale. Zur Ableitung geeigneter Vertragsmerkmale sind detaillierte Kenntnisse der Funktionsweise sowie der Tarifikalkulation der Privaten Unfallversicherung erforderlich. Werden die den Daten zu Grunde liegenden versicherungstechnischen Zusammenhänge nicht beachtet, so kann dies erhebliche Auswirkungen auf die Qualität und Akzeptanz der Ergebnisse haben.

Der zu analysierende Datensatz bestand schließlich aus mehr als 100.000 Verträgen mit jeweils 75 Merkmalen und einer binären Zielvariable. Insgesamt wurden etwa 9% der betrachteten Verträge durch den Versicherungsnehmer storniert. Bei der Auswahl der Daten mussten geeignete Näherungen verwendet werden, da wichtige Vertragsmerkmale nicht historisiert werden.

### **2.4 Modellierung**

Grundsätzlich handelt es sich bei der Fragestellung um ein typisches Klassifikationsproblem, für das eine riesige Bandbreite möglicher Modellierungsansätze denkbar ist. Im Fallbeispiel wurden Entscheidungsbäume als Modellierungstechnik gewählt. Diese sind gut für Klassifikationsprobleme und zur Analyse großer Datenmengen geeignet und erlauben insbesondere auch eine Beschreibung und Interpretation des Prognosemodells. Vor allem berücksichtigen diese modernen Algorithmen automatisch die vorhandenen Abhängigkeiten zwischen den Vertragsmerkmalen und weisen deshalb strukturelle Vorteile gegenüber klassischen (logistischen) Regressionsmo-

dellen auf. Aus der Vielzahl der zur Verfügung stehenden Entscheidungsbaum-Algorithmen wurde der CART-Algorithmus (Classification and Regression Trees) ausgewählt<sup>64</sup>. Für die Durchführung der Analysen wurde der SAS Enterprise Miner verwendet. Insbesondere die Visualisierung des Modells mit Hilfe des Tree Rings hat sich dabei als sehr hilfreich erwiesen.

Zur Beurteilung der Güte der mit Hilfe von CART erzeugten Modelle wurde eine Aufteilung in einen Trainings- und einen Validierungs-Datensatz vorgenommen. Mit Hilfe des Trainings-Datensatzes wurde ein Entscheidungsbaum aufgebaut, der anschließend mit Hilfe des Validierungs-Datensatzes bewertet wurde. Dadurch wird sichergestellt, dass das Modell sich nicht zu gut an die Trainingsdaten anpasst (so genanntes „Overfitting“)<sup>65</sup>.

Aufgrund der Tatsache, dass das zu analysierende Merkmal Storno nur in 9% der Verträge auftrat, hat es sich als sinnvoll erwiesen, die entsprechenden Verträge höher zu gewichten als die nicht stornierten Verträge (so genanntes „Oversampling“)<sup>66</sup>.

Um ein möglichst gutes Modell zu finden, wurden die Parametereinstellungen geeignet variiert. Außerdem wurde versucht, durch Modifikationen des Datensatzes (z.B. durch Hinzunahme weiterer abgeleiteter Merkmale) eine Verbesserung des Modells zu erzielen. Die Parametrisierung ist auch bei dieser Art von Modellen eine komplexe Aufgabe, die ein grundlegendes Verständnis der Funktionsweise der zugrunde liegenden Algorithmen und somit erhebliches statistisches Know-how erfordert. Das Oversampling führt dazu, dass Liftwerte sowie die Größe von Segmenten geeignet umgerechnet werden müssen.

Die Modelle wurden auf Basis des Lifts in einem Segment verglichen, das 5% des Bestands entspricht. Diese iterative Vorgehensweise führte letztlich zu einem Baum, der aus 28 Blättern besteht und einen Lift von etwa 3,4 aufweist, d.h. dass bei gezielter Ansprache von 5% der Kunden mit ablaufenden Verträgen ca. 3,4 Mal mehr stornogefährdete Verträge erreicht werden als bei zufälliger Auswahl. Dies bedeutet, dass sich durch gezielte Auswahl stornogefährdeter Kunden auf Basis des Modells der gleiche Effekt mit 70% geringeren Kosten im Vergleich zur zufälligen Auswahl der Kunden erzielen lässt.

---

<sup>64</sup> Vgl. Breiman et al. (1984).

<sup>65</sup> Vgl. Berry und Linoff (2000), S.183ff.

<sup>66</sup> Vgl. Ebenda.

## **2.5 Bewertung der Ergebnisse**

Im Fallbeispiel ermöglicht das „optimale“ Modell die Identifikation der wichtigsten Vertragsmerkmale für die Prognose des Stornoverhaltens. Dies sind insbesondere die bisherige Vertragslaufzeit, die Tarifgruppe, die versicherte Invaliditätssumme und das Alter des Versicherungsnehmers.

Ferner erlaubt das Modell eine Beschreibung der stornogefährdeten Bestandssegmente. Zum Beispiel ergibt sich ein Segment, welches ca. 0,15% des Bestands umfasst und in dem ca. 72% der Verträge storniert wurden. Dieses Segment ist durch bestimmte Tarifgruppen, eine bisherige Laufzeit von genau 10 Jahren und eine versicherte Invaliditätssumme im Bereich von 82.000€ bis 86.000€ gekennzeichnet. Darüber hinaus ergaben sich weitere Segmente mit überdurchschnittlich hoher Stornoquote. Bei der Analyse der Segmente ist insbesondere das Zusammenspiel der verschiedenen Vertragsmerkmale von besonderem Interesse.

Insgesamt gesehen ist das Modell als fachlich plausibel und mit einem Lift von 3,4 bezogen auf 5% des Bestands als geeignete Grundlage für eine erste zielgerichtete Marketingaktion anzusehen.

## **2.6 Anwendung der Ergebnisse**

Das im Rahmen des Fallbeispiels entwickelte Modell zur Prognose des Stornoverhaltens kann zum Design einer gezielten Marketingaktion zur Stornoreduktion verwendet werden. Dazu wird für zukünftig ablaufende Verträge das Stornoverhalten mit Hilfe des Modells prognostiziert. Hierdurch ergeben sich Bestandssegmente, in denen eine hohe Stornoquote zu erwarten ist. Mit Hilfe geeigneter Marketinginstrumente kann das Unternehmen nun versuchen, die entsprechenden Kunden bereits im Vorfeld gezielt anzusprechen, um eine Stornierung zu verhindern.

Bei einer solchen Marketingaktion werden stets auch Kontrollgruppen gebildet, anhand derer das Modell überprüft werden kann. Dies ist Teil einer systematischen Auswertung und monetären Bewertung solcher Marketingmaßnahmen. Auch hierbei ist die Anwendung geeigneter statistischer Verfahren erforderlich.

Die durchgeführten Analysen erlauben darüber hinaus Hinweise für eine Verbesserung der Datenhaltung. Insbesondere sollte die Historisierung einiger wichtiger Vertragsmerkmale in Erwägung gezogen werden.

### 3 CRISP-DM als Prozessmodell

Das Fallbeispiel zeigt, dass eine systematische Datenanalyse als Prozess aufgefasst werden sollte, der sinnvoll strukturiert und zielgerichtet durchgeführt werden muss. Dieser Prozess wird häufig als „Knowledge Discovery in Databases“ (KDD) bezeichnet. Darunter versteht man den „nicht-trivialen Prozess der Identifizierung von gültigen, neuen, potenziell nützlichen und letztlich verständlichen Mustern in Daten“<sup>67</sup>. Data Mining ist dabei nur ein Teilabschnitt des Prozesses, nämlich „der Schritt im KDD-Prozess, in dem mit Hilfe bestimmter Algorithmen nach interessanten Mustern in Daten gesucht wird“<sup>68</sup>. Oft werden beide Begriffe jedoch (fälschlich) synonym verwendet.

Der Ablauf eines solchen Prozesses kann in mehrere Abschnitte aufgebrochen werden. Hierzu wurde eine Vielzahl von Modellen entwickelt, wobei sich das Vorgehen im Fallbeispiel am Prozessmodell CRISP-DM orientierte. Hierbei handelt es sich um einen Standard für Data Mining Projekte, der sich auf verschiedenste Fragestellungen in den unterschiedlichsten Branchen übertragen lässt, und darüber hinaus unabhängig von der verwendeten Software ist.<sup>69</sup>

CRISP-DM unterteilt ein Data Mining Projekt in sechs Phasen: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation und Deployment. Diese werden wiederum in generische Aufgaben unterteilt. Abbildung 1 zeigt den Lebenszyklus eines Data Mining Projekts nach CRISP-DM<sup>70</sup>.

Die Darstellung der Phasen und Aufgaben in einer festen Reihenfolge ist eine idealisierte Darstellung des Data Mining Prozesses. In der Praxis lassen sich die einzelnen Phasen nicht strikt voneinander trennen. Vielmehr handelt es sich um einen iterativen Prozess, bei dem immer wieder Rückkopplungen zwischen den einzelnen Phasen vorgesehen sind.

„Daten“ in der Mitte der Abbildung verdeutlicht, dass die Daten im Mittelpunkt des Prozesses stehen. Der äußere Kreis veranschaulicht die zyklische Natur des Data Mining. Ein Data Mining Projekt ist nicht zu Ende, wenn

---

<sup>67</sup> Vgl. Fayyad et al. (1996).

<sup>68</sup> Vgl. Ebenda.

<sup>69</sup> Eine Diskussion verschiedener möglicher Prozessmodelle findet man beispielsweise in Säuberlich (2000).

<sup>70</sup> Vgl. Chapman et al. (2000), S.13.



einmal Daten ausgewählt, analysiert und die Ergebnisse angewandt wurden. Gerade bei mehrfacher Anwendung ergibt sich eine (kosten-)effiziente Nutzung der dabei gewonnenen Kenntnisse für das Unternehmen. Gleichzeitig erlaubt dies, die Modelle in regelmäßigen Abständen anzupassen und stetig zu verbessern. Außerdem ergeben sich aus einem konkreten Projekt oft weitere Fragestellungen. Erfahrungen aus früheren Projekten können und sollten dabei genutzt werden.

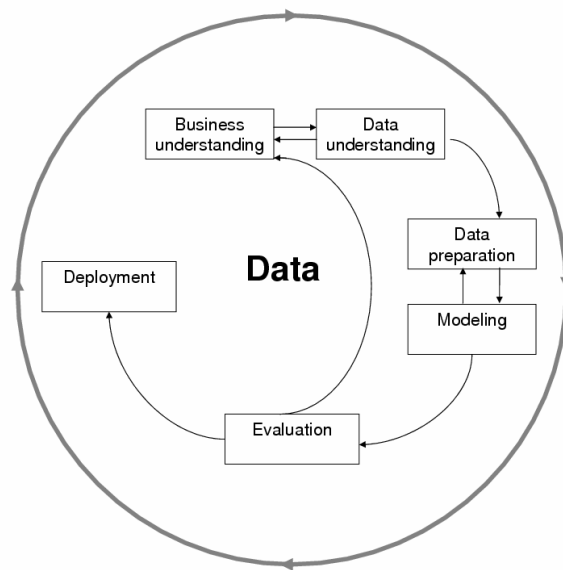


Abbildung 1: Phasen des CRISP-DM Prozessmodells

Im Folgenden sollen die einzelnen Phasen etwas näher beschrieben werden<sup>71</sup>. Die konkreten Tätigkeiten in der jeweiligen Phase wurden bereits im vorhergehenden Abschnitt diskutiert.

- a) Business Understanding: In der ersten Phase eines Projekts werden die unternehmerischen Fragestellungen und Ziele sorgfältig herausgearbeitet, um daraus anschließend die Aufgabenstellung für das Data Mining abzuleiten. Dieser Schritt ist besonders wichtig, um sicherzustellen, dass Data

---

<sup>71</sup> Vgl. Chapman et al. (2000), S.16ff.

Mining nicht die richtigen Antworten auf die „falschen“ Fragen liefert. Ferner müssen bereits in dieser frühen Phase Erfolgsfaktoren festgelegt werden, auf Basis derer der Erfolg des Projekts gemessen wird. Außerdem erfordert ein solches Projekt die genaue Planung aller Schritte bis hin zur Umsetzung der Ergebnisse, sowie eine Beurteilung der Ressourcen, Restriktionen und Voraussetzungen.

- b) **Data Understanding:** In der zweiten Phase soll ein grundlegendes Verständnis über die für die Analyse zur Verfügung stehenden Daten gewonnen werden. Ziel ist eine genaue Analyse der Datenquellen, des Datenumfangs, der Datenstruktur und der Datenqualität. Es ist wichtig, in dieser Phase auch zu analysieren, ob die in der ersten Phase formulierten Ziele mit Hilfe der zur Verfügung stehenden Daten überhaupt erreicht werden können. Falls dies nicht der Fall ist, müssen entweder weitere Datenquellen hinzugezogen oder aber die Fragestellung modifiziert werden.
- c) **Data Preparation:** In der dritten Phase wird eine fundierte Datengrundlage für die eigentliche Datenanalyse geschaffen. Hierzu müssen die aus verschiedenen Quellen stammenden Daten zusammengeführt, bereinigt und geeignet formatiert werden. Aus den vorhandenen Variablen werden für die Analyse besonders hilfreiche Variable abgeleitet. Schließlich wird eine zur Analyse geeignete Teilmenge ausgewählt.
- d) **Modeling:** Diese Phase bildet den Kern des Projekts. Hier erfolgen zunächst die Auswahl und anschließend der Einsatz der für die Fragestellungen geeigneten Modellierungstechniken. Die resultierenden Modelle werden aus technischer Sicht bewertet und verglichen, insbesondere in Hinblick auf Genauigkeit und Generalisierbarkeit. Über iterative Schritte wird schließlich das endgültige Modell gesucht, welches aus Sicht der Datenanalyse nur noch marginale Verbesserung zulässt.
- e) **Evaluation:** Nach der technischen Bewertung in der vorangegangenen Phase steht nun die fachliche Bewertung des Modells im Vordergrund. Hierbei wird untersucht, ob das Modell zur Beantwortung der unternehmerischen Fragestellungen geeignet und fachlich plausibel ist.
- f) **Deployment:** In der letzten Phase werden die gewonnenen Erkenntnisse letztlich unternehmerisch eingesetzt. Hierbei ist eine genaue Planung, Überwachung und Auswertung notwendig.

Der Zeitaufwand für die einzelnen Phasen ist in der Regel unterschiedlich hoch: oft beanspruchen die ersten drei Phasen allerdings mehr als die Hälfte des Zeitaufwands, insbesondere bei erstmaliger Durchführung solcher Analysen.

## **4 Fazit und Ausblick**

Ziel des Fallbeispiels war es aufzuzeigen, wie das Schlagwort Data Mining umgesetzt wird, so dass es zählbare Erfolge bringt. Anhand des Fallbeispiels wird klar, dass sich mit ersten Data Mining Analysen auf Basis einer begrenzten Datengrundlage Ergebnisse erzielen lassen, die bei systematischer Umsetzung zu einer erheblich besseren Fokussierung von Marketingaktionen führen können. Im Fallbeispiel ermöglichen raffinierte statistische Verfahren eine Steigerung der Treffergenauigkeit um den Faktor 3,4. Dieses gibt dem Unternehmen die Möglichkeit, abwanderungswillige Kunden frühzeitig zu erkennen und entsprechende Gegenmaßnahmen zu ergreifen.

Entscheidend für den Erfolg derartiger Analysen sind die richtige Herangehensweise sowie die Auswahl der geeigneten statistischen Verfahren. Nur eine strukturierte Vorgehensweise stellt sicher, dass die für das Unternehmen wichtigen Fragen beantwortet werden. Dazu ist ein sequentielles Vorgehen, Durchhaltevermögen sowie eine konsequente Auswertung der Ergebnisse notwendig.

Erst durch regelmäßige Durchführung und schrittweise Verbesserung kommen die positiven Effekte in aller Breite zum Tragen. Durch systematische Auswertung von Marketingaktionen, sukzessive Anpassung der Datenhaltung, sowie ständige Verbesserung der Prognosemodelle kann unter Einbeziehung der betroffenen Unternehmensbereiche schrittweise ein Frühwarnsystem zur Stornoprophylaxe entwickelt werden. Dieses gibt dem Unternehmen die Möglichkeit, abwanderungswillige Kunden frühzeitig zu erkennen und entsprechende Gegenmaßnahmen zu ergreifen. Ein solches Frühwarnsystem kann dann eine erste Komponente eines Kundenwertmodells sein.

Ferner erfordert eine zielgerichtete Analyse erhebliches Hintergrundwissen über die zugrunde liegenden versicherungstechnischen Zusammenhänge sowie statistisches Know-how bei der Auswahl und dem Einsatz geeigneter Modellierungstechniken. Die iterative Suche nach einem bestmöglichen Modell kann dabei nicht automatisiert werden. Entscheidungsbäume bieten in diesem Zusammenhang den Vorteil, dass die Modelle keine „Black Box“, sondern vielmehr einfach interpretierbar sind. Durch die Einschränkung auf ein Verfahren hielt sich der Modellierungsaufwand in Grenzen.<sup>72</sup>

---

<sup>72</sup> Alternative Modellierungsansätze werden beispielsweise in Berry und Linoff (1997) diskutiert.

Das Vorhandensein eines komplexen unternehmensübergreifenden Data Warehouse ist in einem solchen ersten Schritt nicht unbedingt erforderlich. Im Rahmen der Analysen ergeben sich jedoch oft Hinweise auf eine Verbesserung der Datenhaltung, die für die Konzeption und Weiterentwicklung eines Data Warehouse sehr hilfreich sein können. Mit Hilfe der dargestellten Methodik lassen sich beliebige Fragestellungen in den unterschiedlichsten Sparten der Versicherungswirtschaft analysieren. Dabei müssen stets die versicherungstechnischen Zusammenhänge beachtet werden. Eine Analyse des Stornoverhaltens analog zum Fallbeispiel erscheint dabei insbesondere für die Kfz-Versicherung Erfolg versprechend. Die Analyse und Nutzung der in den Daten verborgenen Informationen stellt aber auch für Fragen wie Cross-Selling und Kundenwertanalyse ein erhebliches Potenzial dar.

## Literatur

- [1] Berry, M.J.A. und Linoff, G.S (1997). *Data Mining Techniques: For Marketing, Sales and Customer Support*. New York: Wiley.
- [2] Berry, M.J.A. und Linoff, G.S (2000). *Mastering Data Mining*. New York: Wiley.
- [3] Breiman, L., Friedman, J.H., Olshen, R.A. und Stone, C.J. (1984). *Classification and Regression Trees*. Belmont, California: Wadsworth.
- [4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. und Wirth, R. (2000). *CRISP-DM 1.0 – Step-by-Step Data Mining Guide*. Download unter <http://www.crisp-dm.org/CRISPWP-0800.pdf> am 23.02.2004.
- [5] Fayyad, U.M., Piatetsky-Shapiro, G. und Smyth, P (1996). *From Data Mining to Knowledge Discovery: An Overview*. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. und Uthurusamy, R. (Eds.). *Advances in Knowledge Discovery and Data Mining*. Menlo Park, California: AAAI Press.
- [6] Gesamtverband der Deutschen Versicherungswirtschaft e.V. [GDV] (2003). *Statistisches Taschenbuch der Versicherungswirtschaft 2003*. Karlsruhe: Verlag Versicherungswirtschaft.
- [7] Riebesell, H. (1991). *Unfallversicherung*. In: Grosse, W. (Hrsg.). *Versicherungsenzyklopädie, Band 4*. Wiesbaden: Gabler.
- [8] Säuberlich, F. (2000). *KDD und Data Mining als Hilfsmittel zur Entscheidungsunterstützung*. Frankfurt: Lang.