

Ein generisches Kreislaufmodell zur Einbettung von Data-Mining-Analysen in die Geschäftsprozesse von Unternehmen – mit einem Fallbeispiel aus der Versicherung

Andreas Reuß und Hans-Joachim Zwiesler
Universität Ulm

Zusammenfassung:¹

Die intelligente Informationsgewinnung wird ein immer wichtigeres strategisches Erfolgskriterium für Unternehmen, da Daten in immer größerem Umfang vorhanden sind und sich die Software und Hardware zu ihrer Verarbeitung drastisch verbessert hat. Die unter Schlagworten wie "Data-Mining" oder "Knowledge Discovery in Databases" zusammengefassten Verfahren liefern jedoch nur dann wirklich unternehmerischen Nutzen, wenn sie voll in die betrieblichen Prozesse eingegliedert werden. Hierfür entwickeln wir in der vorliegenden Arbeit ein umfassendes, in sich geschlossenes Prozessmodell. Dieses Kreislaufmodell betont insbesondere den zyklischen Charakter des zugrunde liegenden Prozesses. Das Modell verbindet dabei die bisher in der Literatur diskutierten Aspekte der Datenanalyse mit einer detaillierten Beschreibung der Einbindung derartiger Prozesse in das betriebliche Umfeld unter konsequenter Nutzung von Kosten-Nutzen-Analysen. Unser Modell verwendet mehrere Abstraktionsebenen und eignet sich als generisches Referenzmodell für Unternehmen im Allgemeinen. Darüber hinaus zeigen wir an einem spezifischen Beispiel aus der Versicherung, wie das Modell an eine konkrete Situation angepasst und in der Praxis angewendet werden kann.

Schlagworte:

Data-Mining, Knowledge Discovery in Databases, Stornoprophylaxe, Unfallversicherung

Abstract:

The intelligent extraction of information from data is an increasingly important strategic success criterion for businesses, since a steadily increasing amount of data is available and both software and hardware for the processing of data have improved dramatically. However, the methods summarized by terms like "data mining" or "knowledge discovery in databases" only provide benefits in a business context if they are completely integrated into the business processes. In this paper, we are developing a comprehensive,

¹Working Paper, Mai 2005

self-contained process model for this purpose. It is a cyclical model that emphasizes the cyclical character of the underlying process. The model combines the aspects of data analysis discussed in the literature with a detailed description of the integration of these processes in the business environment under consistent usage of cost-benefit-analyses. Our model uses different levels of abstraction and is suited as a generic reference model for businesses in general. In addition, we are illustrating how the model can be adapted to a specific situation and applied in practice by referring to a specific example from the insurance sector.

Keynotes:

Data Mining, Knowledge Discovery in Databases, Cancellation Prophylaxis, Accident Insurance

Inhaltsübersicht

1. Einleitung
2. Einordnung des Data-Mining-Begriffs
3. Übersicht über Prozessmodelle für Data-Mining
 - 3.1 Prozessmodelle für KDD
 - 3.1.1 Prozessmodell von Chapman et al. – CRISP-DM (2000)
 - 3.1.2 Fazit
 - 3.2 Einbindung des KDD-Prozesses in das betriebliche Umfeld
 - 3.2.1 Ansatz von Berry und Linoff (1997, 2000)
 - 3.2.2 Ansatz von Knobloch (2001, 2004)
 - 3.2.3 Fazit
4. Generisches Kreislaufmodell zur Einbettung von Data-Mining-Analysen in die Geschäftsprozesse von Unternehmen
 - 4.1 Das generische Kreislaufmodell
 - 4.2 Die einzelnen Stufen des Kreislaufs im Detail
 - 4.2.1 Spezifikation der Fragestellung
 - 4.2.2 Durchführung der Data-Mining-Analysen
 - 4.2.3 Umsetzung der Erkenntnisse in Maßnahmen
 - 4.2.4 Messung und Bewertung der Ergebnisse
 - 4.3 Die Anwendung des Kreislaufmodells
5. Anwendung des Modells auf die Bestandssicherung in der Unfallversicherung
 - 5.1 Spezifikation der Fragestellung
 - 5.2 Durchführung der Data-Mining-Analysen
 - 5.3 Umsetzung der Erkenntnisse in Maßnahmen
 - 5.4 Messung und Bewertung der Ergebnisse
 - 5.5 Weitere Vorgehensweise
6. Zusammenfassung und Ausblick

Literatur

1. Einleitung

Versicherungsunternehmen besitzen eine Fülle von Daten, die vielfältige Informationen über das Unternehmen und seine Kunden enthalten. Damit werden diese Daten zu einer wichtigen Quelle zur Verbesserung der Produkte und Prozesse im Unternehmen und somit zur Generierung von Wettbewerbsvorteilen. Dies setzt allerdings voraus, dass die Unternehmen die strategische Bedeutung der in diesen Daten verborgenen Informationen erkennen und Prozesse der intelligenten Informationsgewinnung aus Daten initiieren sowie geeignet strukturieren. Gerade an letzterem hapert es in der Praxis jedoch oft. Und interessanterweise zeigt sich, dass auch die in der wissenschaftlichen Literatur diskutierten Prozessmodelle nicht alle aus betriebswirtschaftlicher Sicht relevanten Aspekte der Datenanalyse im betrieblichen Umfeld umfassen, insbesondere im Hinblick auf die Bewertung und Nutzung der mithilfe der Datenanalyse gewonnenen Erkenntnisse. Diese Lücke versuchen wir mit der vorliegenden Arbeit zu schließen, indem wir ein in sich geschlossenes Gesamtmodell vorstellen, mit dem Data-Mining-Analysen in die Geschäftsprozesse insbesondere auch bei Versicherungsunternehmen eingebettet werden können.

Der Begriff Data-Mining bezieht sich dabei nur auf einen Teilschritt dieses Prozesses, nämlich auf die Anwendung moderner datengetriebener Analyseverfahren zur Erkennung von Mustern in Daten. Bereits seit langem ist in der wissenschaftlichen Literatur diskutiert worden, dass dies nur ein, wenn auch ganz wesentlicher, Schritt eines umfassenderen Prozesses ist. So gehören z. B. die Interpretation der Ergebnisse und die Aspekte der Datenaufbereitung als unverzichtbare Bestandteile dazu. Für diese erweiterte Verfahrensklasse hat sich der Begriff Knowledge Discovery in Databases (KDD) entwickelt. Für KDD existieren eine Reihe von Modellen in der Literatur, für die mit CRISP-DM² ein Obermodell entwickelt wurde, das alle wesentlichen, bis dato in der wissenschaftlichen Literatur diskutierten Aspekte umfasst. Allerdings fokussiert dieses Modell eindeutig auf den Aspekt der Datenanalyse und beschäftigt sich nur rudimentär mit der Umsetzung der Ergebnisse im betrieblichen Umfeld. Insbesondere wird in diesem Modell kaum aufgezeigt, welche Phasen und Schritte notwendig sind, um die Ergebnisse in der betrieblichen Praxis einzusetzen. Zu dieser Einbindung des KDD-Prozesses in das betriebliche Umfeld gibt es in der Literatur den Ansatz von Berry und Linoff³, welcher von Knobloch⁴ erweitert wur-

²Cross-Industry Standard Process for Data Mining, vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000).

³Vgl. *Berry/Linoff* (1997) und *Berry/Linoff* (2000).

⁴Vgl. *Knobloch* (2001) und *Knobloch* (2004).

de. Allerdings sind diese Ansätze nur sehr wenig konkret und enthalten keine detaillierte Beschreibung der Aufgaben, die den einzelnen Schritten des Prozesses zugeordnet sind.

In der vorliegenden Arbeit stellen wir ein Modell vor, das all diese Ansätze umfasst und eine vollständige, in sich geschlossene Beschreibung des Prozesses liefert, wie Data-Mining sinnvoll im betrieblichen Umfeld eingesetzt werden sollte. Es handelt sich dabei um ein allgemeines, branchenübergreifendes Modell, welches jedoch insbesondere für den Einsatz in Versicherungsunternehmen geeignet ist.

Entsprechend diesem Ziel der vorliegenden Arbeit geben wir im zweiten Abschnitt zunächst einen Überblick über die Einordnung der Begriffe Data-Mining und Knowledge Discovery in Databases, ehe wir im dritten Abschnitt die wissenschaftliche Literatur zu Prozessmodellen für KDD und deren Einbindung in das betriebliche Umfeld zusammenfassen. Im Hauptteil der vorliegenden Arbeit, dem vierten Abschnitt, stellen wir das von uns entwickelte Modell vor und beschreiben detailliert die einzelnen Stufen des Kreislaufs. Auf Basis der Analysen des vorhergehenden Abschnitts zeigen wir dabei, dass es die bisherigen in der Literatur diskutierten Ansätze umfasst und gleichzeitig deren wesentliche Defizite behebt. Zur Veranschaulichung unseres Modells und zum Nachweis seiner Praxistauglichkeit beschreiben wir anschließend im fünften Abschnitt eine konkrete Anwendung unseres Modells auf die Bestandssicherung in der Unfallversicherung.

Mit dem von uns entwickelten Modell stellen wir einen Ansatz vor, der wesentliche Defizite bisheriger Modelle beseitigt und damit eine theoretisch fundierte und praktisch handhabbare Basis liefert, um Data-Mining systematisch als strategisches Instrument zur Verbesserung der Wettbewerbsfähigkeit von (Versicherungs-)Unternehmen einzusetzen.

2. Einordnung des Data-Mining-Begriffs

Zunächst wollen wir den Begriff Data-Mining näher erläutern und einordnen.⁵ Hierzu ist eine Unterscheidung zwischen zwei Grundtypen von Datenanalyseproblemen hilfreich. Als Unterscheidungsmerkmal dient dabei der Grad der Hypothesenfreiheit, also das Ausmaß, in dem Hypothesen des Anwenders eine Rolle spielen.

Bei der Analyse des Stornoverhaltens in einem Unfallversicherungsbestand können sich beispielsweise folgende Fragestellungen ergeben:

- "Nimmt die Stornowahrscheinlichkeit mit zunehmender Vertragslaufzeit ab?"

⁵Die Ausführungen dieses Abschnitts orientieren sich an *Knobloch* (2001), S. 67–72, sowie den dort angegebenen Quellen.

- "Welche Kombinationen von Vertragsmerkmalen führen zu einer hohen Stornowahrscheinlichkeit?"

Bei der ersten Fragestellung handelt es sich um eine typische hypothesengetriebene Fragestellung. Dabei gibt der Anwender eine Hypothese vor, die anschließend mit Hilfe der vorhandenen Daten überprüft wird, sodass der Anwender schließlich die Ergebnisse interpretieren kann. Hier steht also die Hypothesenverifikation im Vordergrund und man spricht in diesem Zusammenhang auch von einem Top-Down-Problem.

Bei der zweiten Fragestellung hingegen ist keine zu überprüfende Hypothese vorgegeben; es handelt sich also um eine hypothesenfreie Fragestellung. In diesem Fall werden vielmehr aus dem Datenbestand selbständig Hypothesen generiert, welche anschließend vom Anwender interpretiert werden. Deshalb spricht man in diesem Zusammenhang auch von einem Bottom-Up-Problem, also von der datengetriebenen Aufdeckung bisher unbekannter Muster und Beziehungen.⁶

Für die Beantwortung der ersten Fragestellung kommen grundsätzlich die traditionellen Verfahren der Statistik zur Anwendung. Die zweite Fragestellung erfordert hingegen eine neue Art von Analyseverfahren, welche meist unter dem Schlagwort Data-Mining zusammengefasst werden. Ein zentrales Merkmal des Data-Mining ist also die Hypothesenfreiheit und die damit verbundene datengetriebene Analyse mit dem Ziel der Datenmustererkennung.⁷

Im Idealfall würde also ein Data-Mining-System vollkommen autonom aus einem beliebigen Datenbestand die für den Anwender interessantesten Auffälligkeiten herausfiltern. Diese häufig mit dem Schlagwort Data-Mining verbundenen Erwartungen können zumindest mit den heute bekannten Methoden nicht erfüllt werden.⁸ Vielmehr müssen vor Beginn der Analyse zumindest der Anwendungsbereich und die Art der erwarteten Ergebnisse spezifiziert werden. Knobloch bemerkt hierzu außerdem, dass "eine völlig freie Suche nach 'irgendwelchen' Auffälligkeiten in den Daten im Allgemeinen als unsinnig erachtet" wird.⁹ Diese Feststellungen gelten u.E. insbesondere im betrieblichen Umfeld. Entscheidend ist also die Tatsache, dass keine Hypothesen vom Anwender vorgegeben werden, sondern vielmehr Hypothesen datengetrieben generiert werden.¹⁰

⁶Vgl. *Bigus* (1996), S. 9.

⁷Vgl. *Knobloch* (2001), S. 69.

⁸Vgl. *Küppers* (1999), S. 92.

⁹*Knobloch* (2001), S. 70.

¹⁰Vgl. *Küppers* (1999), S. 51.

Man spricht im Zusammenhang mit Data-Mining heute deshalb von tendenziell hypothesenfreien Fragestellungen. Dies zeigt sich auch im obigen Beispiel: dort wird vorgegeben, dass die Daten im Hinblick auf Unterschiede in der Stornowahrscheinlichkeit zu untersuchen sind. Als Ergebnis werden Kombinationen von Vertragsmerkmalen erwartet, die zu einer besonders hohen Stornowahrscheinlichkeit führen.

Auf die Vielzahl der im Zusammenhang mit Data-Mining in der wissenschaftlichen Literatur verwendeten Begriffe sowie deren Entwicklung gehen wir an dieser Stelle nicht näher ein, sondern verweisen auf die entsprechende Literatur.¹¹ Wir möchten jedoch festhalten, dass sich aus der Vielzahl von Begriffen zwei Begriffe durchgesetzt haben: "Data-Mining" und "Knowledge Discovery in Databases" (KDD).¹² Eine grundlegende Definition der Begriffe wird deshalb an dieser Stelle vorgestellt und erläutert:¹³

- "*Knowledge discovery in databases* (KDD) is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data."
- "*Data mining* is a step in the KDD process consisting of particular data mining algorithms that, under some acceptable computational efficiency limitations, produces a particular enumeration of patterns."

Data-Mining bezeichnet also hier die Anwendung von Algorithmen zur Extraktion von Mustern aus Rohdaten. Fayyad et al. weisen jedoch darauf hin, dass die blinde Anwendung von Data-Mining-Algorithmen gefährlich sein kann.¹⁴ KDD hingegen umfasst zusätzliche Schritte, die notwendig sind, um sicherzustellen, dass tatsächlich gültige, neuartige, nützliche und verständliche Informationen generiert werden. Hierzu gehören die Einbindung von vorhandenem Fachwissen sowie die Interpretation der Ergebnisse, aber auch die gesamte Vor- und Nachbereitung der Daten. Diese Definition von KDD als komplexer Prozess bestehend aus einer Reihe von Aktivitäten neben der eigentlichen Datenanalyse wird im Folgenden verwendet und durch die Vorstellung zugehöriger Prozessmodelle konkretisiert.

Auf eine Beschreibung der vorhandenen Data-Mining-Verfahren und eine Systematisierung der Verfahren verzichten wir mit Verweis auf die vorhandene Literatur zu diesem Aspekt.¹⁵ Das im Folgenden zu entwickelnde Prozessmodell soll jedenfalls so flexibel und allgemeingültig ausgestaltet sein, dass die Berücksichtigung unterschiedlicher vorhandener oder künftig entwickelter Data-Mining-Verfahren möglich ist.

¹¹Vgl. Küppers (1999), S. 17–28.

¹²Knobloch (2001), S. 63–64.

¹³Vgl. Fayyad/Piatetsky-Shapiro/Uthurusamy (1996), S. 6–9.

¹⁴Vgl. Fayyad/Piatetsky-Shapiro/Uthurusamy (1996), S. 4.

¹⁵Vgl. Küppers (1999), S. 51–80, und Knobloch (2001), S. 80–85.

Bei der Anwendung von Data-Mining-Methoden im betrieblichen Umfeld (beispielsweise in Versicherungsunternehmen) ist die Datenmustererkennung alleine nicht ausreichend; reine Analysetätigkeiten generieren nämlich keinen betriebswirtschaftlichen Nutzen.¹⁶ Um den Aufwand für die Analysen aus betriebswirtschaftlicher Sicht zu rechtfertigen, ist es vielmehr notwendig, wie von Berry und Linoff gefordert, Daten in Informationen, Informationen in Handlungen und Handlungen in Erträge umzusetzen. Berry und Linoff verwenden für diesen Ansatz den Begriff des "virtuous cycle of data mining".¹⁷ Diesen Ansatz diskutieren wir in Abschnitt 3.2.1 genauer.

In jüngster Zeit hat sich in diesem Umfeld ein weiterer Begriff entwickelt, der Begriff des "Organizational data mining" (ODM), welcher wie folgt definiert ist:¹⁸

- "Organizational data mining (ODM) is defined as leveraging data mining tools and technologies to enhance the decision making process by transforming data into valuable and actionable knowledge to gain a competitive advantage."

Hier steht also der Einsatz von Data-Mining-Methoden zur Entscheidungsunterstützung im Vordergrund. Zentral ist dabei, wie schon bei Berry und Linoff, die Generierung von Wissen, welches in konkrete Maßnahmen umgesetzt werden kann ("actionable knowledge").

Aus den hier aufgeführten Definitionen von Data-Mining, KDD und ODM ergeben sich unmittelbar folgende Fragen:

- Wie kann der KDD-Prozess geeignet strukturiert werden, um mit Hilfe von Data-Mining tatsächlich umsetzbares Wissen zu generieren?
- Wie kann der KDD-Prozess in andere Geschäftsprozesse eingebettet werden, um eine wirkungsvolle Umsetzung des generierten Wissens in konkrete Maßnahmen im Sinne des ODM zu ermöglichen?

Die in der wissenschaftliche Literatur diskutierten Ansätze zur Beantwortung dieser Fragen werden im folgenden Abschnitt vorgestellt und gehen anschließend in die Entwicklung eines allgemeinen Modells zur Einbettung von Data-Mining-Analysen in Geschäftsprozesse ein.

¹⁶Vgl. Knobloch (2001), S. 103.

¹⁷Vgl. Berry/Linoff (1997), S. 17f.

¹⁸Vgl. Nemati/Barko (2004), S. 2.

3. Übersicht über Prozessmodelle für Data-Mining

In der wissenschaftlichen Literatur zu Data-Mining und KDD finden sich zahlreiche Vorschläge für Modelle zur Strukturierung des KDD-Prozesses. Säußerlich gibt eine Übersicht über fünf Prozessmodelle, die allesamt in Jahren 1996-1998 publiziert wurden, und fasst die Hauptschritte in einem eigenen, verallgemeinerten Ansatz zusammen.¹⁹ In diesem Zeitraum wurden noch einige weitere Modelle publiziert, die interessante Aspekte des KDD-Prozesses beleuchten.²⁰ Als Ansatz zur Vereinheitlichung und Standardisierung der unterschiedlichen Ansätze unter Berücksichtigung praktischer Erfahrungen mit Data-Mining-Projekten ist schließlich das CRISP-DM Modell (Cross-Industry Standard Process for Data-Mining) zu sehen, welches im Jahr 2000 publiziert wurde.²¹ Dieses wird, wie Umfragen zeigen, auch häufig in der Praxis angewendet.²²

Im Folgenden erläutern wir zunächst die wichtigsten Aspekte dieser Prozessmodelle für KDD. Anschließend gehen wir auf die Einbettung des KDD-Prozesses in die Geschäftsprozesse eines Unternehmens ein.

3.1 Prozessmodelle für KDD

Grundsätzlich ist festzuhalten, dass sich der KDD-Prozess in mehrere Phasen unterteilen lässt, die sukzessive zu durchlaufen sind. Der Output einer Phase dient dabei als Input für die darauf folgenden Phasen. Dabei werden die Phasen in der Regel nicht einmalig linear durchlaufen und führen sofort zum gewünschten Ergebnis. Vielmehr ist häufig ein Zurückspringen auf eine frühere Phase notwendig. Es handelt sich also insgesamt um einen komplexen iterativen Prozess.²³

Eine weitere wichtige Feststellung ist die Tatsache, dass trotz des technologischen Fortschritts vollständig autonome KDD-Systeme (noch) nicht existieren.²⁴ Vielmehr handelt es sich bei KDD um eine Aufgabe, die aus komplexen Interaktionen zwischen einem Menschen und einer Datenbank besteht, wobei der Mensch sich geeigneter Methoden bedient.

¹⁹Vgl. *Säußerlich* (2000), S. 22–33.

²⁰Vgl. *Adriaans/Zantinge* (1996), S. 37–78, und *Cabena/Hadjinian/Stadler/Verhees/Zanasi* (1998), S. 41–59.

²¹Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000).

²²Vgl. *Piatetsky-Shapiro* (2002).

²³Vgl. beispielsweise *Adriaans/Zantinge* (1996), S. 37f., *Berry/Linoff* (1997), S. 22f., *Brachman/Anand* (1996), S. 52, *Cabena/Hadjinian/Stadler/Verhees/Zanasi* (1998), S. 42, *Fayyad* (1996), S. 23, *John* (1997), S. 5, und *Knobloch* (2001), S. 87f.

²⁴Vgl. Abschnitt 2.

Dies hat zur Folge, dass den am Prozess beteiligten Personen eine entscheidende Rolle zukommt und dass ein KDD-Prozess mit erheblichem Arbeitsaufwand verbunden ist, welcher insbesondere für die betriebswirtschaftliche Bewertung zu berücksichtigen ist.²⁵

Besondere Beachtung verdienen somit die in den Prozess involvierten Personen sowie deren evtl. unterschiedliche Interessen und Vorkenntnisse. Grundsätzlich kann zwischen zwei beteiligten Personengruppen unterschieden werden, die im Rahmen des KDD-Prozesses eng zusammenarbeiten müssen: zum einen sind das Experten für das jeweilige Anwendungsgebiet, zum anderen Experten für Data-Mining-Analysen. Gelegentlich werden auch Spezialisten für Datenmanagement als dritte Gruppe genannt.²⁶

In den KDD-Prozess fließt somit sowohl Fachwissen über den Anwendungsbereich also auch Wissen über KDD- und Data-Mining-Methoden ein. Dem Hintergrundwissen aus dem konkreten Anwendungsgebiet kommt dabei eine wichtige Rolle zu.²⁷

Adriaans und Zantinge weisen außerdem darauf hin, dass Data-Mining sinnlos ist, wenn nicht auf Basis der Ergebnisse der Analysen auch geeignete Maßnahmen ergriffen werden.²⁸ Allerdings gehen sie nicht näher auf die Vorgehensweise bei der Nutzung der Ergebnisse im betrieblichen Umfeld ein. Außerdem sehen Adriaans und Zantinge Data-Mining idealerweise als fortlaufenden Prozess: Organisationen sollten ständig mit den Daten arbeiten, neue Informationsbedürfnisse identifizieren, und die Daten zu verbessern versuchen, um die gesetzten Ziele besser zu erreichen.²⁹

Cabena et al. betonen, dass Ausgangs- und Endpunkt des KDD-Prozesses die unternehmerischen Ziele sein müssen. Diese Ziele sollten außerdem in allen Phasen des Prozesses als Orientierung dienen. Der Prozess selbst ist iterativ, wobei Rückkopplungen zwischen den einzelnen Phasen explizit vorgesehen sind. Trotz technologischem Fortschritt ist der Prozess nicht autonom, sondern erfordert erheblichen Arbeitsaufwand.³⁰

Auf Details der einzelnen Modelle gehen wir mit Ausnahme von CRISP-DM hier nicht näher ein. Wir verweisen vielmehr auf die jeweils angegebene Literatur.

²⁵Vgl. *Brachman/Anand* (1996), S. 38, und *Cabena/Hadjinian/Stadler/Verhees/Zanasi* (1998), S. 42.

²⁶Vgl. *John* (1997), S. 5f., und *Cabena/Hadjinian/Stadler/Verhees/Zanasi* (1998), S. 44f.

²⁷Vgl. *Brachman/Anand* (1996), S. 47 und S. 51f.

²⁸Vgl. *Adriaans/Zantinge* (1996), S. 81.

²⁹Vgl. *Adriaans/Zantinge* (1996), S. 38.

³⁰Vgl. *Cabena/Hadjinian/Stadler/Verhees/Zanasi* (1998), S. 42.

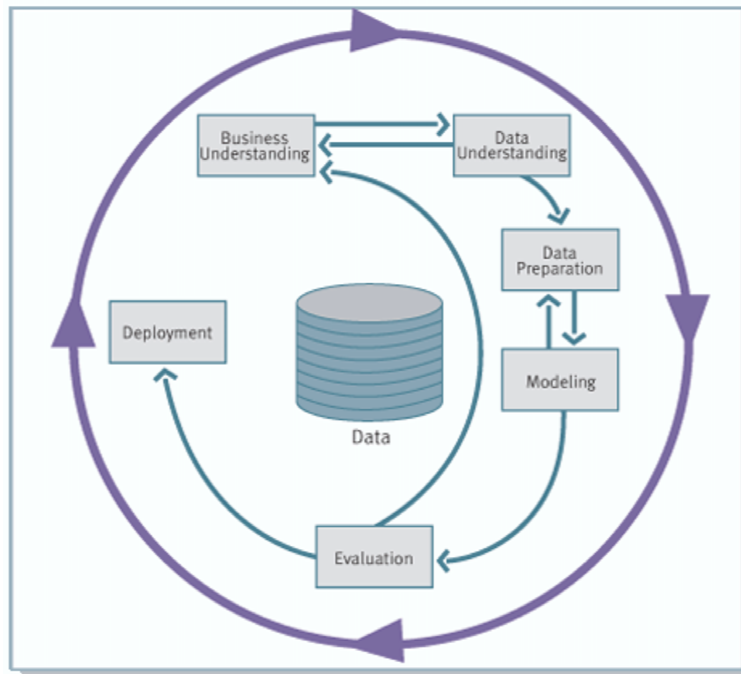


Abbildung 1: Das CRISP-DM Modell von Chapman et al. (Quelle: *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 13)

3.1.1 Prozessmodell von Chapman et al. – CRISP-DM (2000)

CRISP-DM unterteilt den KDD-Prozess in sechs Phasen³¹: "business understanding", "data understanding", "data preparation", "modeling", "evaluation" und "deployment". Diese werden wiederum in generische Aufgaben unterteilt. Im Gegensatz zu den oben genannten Modellen wird bei CRISP-DM eine sehr detaillierte Beschreibung der in den jeweiligen Phasen auszuführenden Tätigkeiten gegeben. Diese Beschreibung erfolgt auf generischer Ebene, wodurch eine Anpassung auf die jeweilige Analysesituation möglich ist.

Abbildung 1 zeigt den Lebenszyklus eines KDD-Projekts nach CRISP-DM. Die Darstellung der Phasen und Aufgaben in einer festen Reihenfolge ist dabei eine idealisierte Darstellung des KDD-Prozesses. In der Praxis lassen sich die einzelnen Phasen nicht strikt voneinander trennen. Zudem handelt es sich auch hier um einen iterativen Prozess, bei dem immer wieder Rückkopplungen zwischen den einzelnen Phasen vorgesehen sind. Die wichtigsten Rückkopplungen sind in Abbildung 1 ebenfalls aufgeführt.³²

³¹Bei CRISP-DM wird Data-Mining als Synonym für KDD verwendet. Im Folgenden soll dennoch stets der Begriff KDD im Sinne der Definition aus Abschnitt 2 verwendet werden.

³²Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 13.

”Data” in der Mitte der Abbildung verdeutlicht, dass die Daten im Mittelpunkt des Prozesses stehen. Der äußere Kreis veranschaulicht die zyklische Natur des KDD. Ein KDD-Projekt ist nicht zu Ende, wenn einmal Daten ausgewählt, analysiert und die Ergebnisse angewendet 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.³³

3.1.2 Fazit

Die in der Literatur diskutierten Prozessmodelle unterteilen den KDD-Prozess jeweils in mehrere Phasen. Die Aufteilung sowie die Benennung der Phasen unterscheidet sich dabei von Modell zu Modell. Dennoch lässt sich bei genauerem Vergleich feststellen, dass die inhaltlichen Unterschiede recht gering sind. Darüber hinaus zeigt sich, dass das als Referenzmodell konzipierte CRISP-DM tatsächlich den überwiegenden Teil der von den anderen Autoren identifizierten Aspekte des KDD-Prozesses abbildet.

Darüber hinaus liegt CRISP-DM eine wegweisende und praxistaugliche Methodik zugrunde, die vier Abstraktionsebenen vorsieht. Vom Allgemeinen zum Speziellen sind dies: Phasen, generische Aufgaben, spezialisierte Aufgaben und Prozessinstanzen.³⁴

Auf der obersten Ebenen wird der Prozess in mehrere Phasen unterteilt. Jede dieser Phasen wiederum besteht aus einer Reihe generischer Aufgaben. Hierbei soll es sich um Aufgaben handeln, die allgemein genug sein sind, um möglichst alle denkbaren Analysesituationen abzudecken. Die ersten beiden Abstraktionsebenen beschreiben also das grundlegende Prozessmodell.

Die dritte Ebene beschreibt, wie die in der zweiten Ebene spezifizierten Aufgaben in bestimmten Situationen ausgeführt werden. Die vierte Ebene, die Prozessinstanz, stellt eine Abfolge von Handlungen, Entscheidungen und Ergebnissen eines konkreten KDD-Prozesses dar. Die dritte und vierte Abstraktionsebene beschreiben also einen konkreten KDD-Prozess.

Diese Methodik erlaubt also die systematische Abbildung des generischen Modells auf konkrete Analysesituationen. Dies ist u.E. für die praktische Umsetzung im durch vielfälti-

³³Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 13.

³⁴Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 9-11.

ge denkbare Analysesituation gekennzeichneten betrieblichen Umfeld sehr hilfreich und erlaubt ein zielgerichtetes und auf die konkrete Analysesituation angepasstes Durchlaufen des KDD-Prozesses. Eine solche Methodik fehlt bei den Beschreibungen der anderen genannten Prozessmodelle. Diese beschränken sich vielmehr auf eine grobe Beschreibung der Tätigkeiten in den einzelnen Phasen bzw. auf konkrete Anwendungsbeispiele.

Insgesamt ist CRISP-DM als Referenzmodell für den KDD-Prozess gut geeignet und dient deshalb als Ausgangspunkt für die weiteren Analysen in Abschnitt 4. Soweit weitere zusätzliche Aspekte anderer Autoren relevant sind, werden sie dort entsprechend gekennzeichnet.

Abschließend möchten wir festzuhalten, dass einige der bisher genannten Prozessmodelle die Wichtigkeit der Umsetzung der Ergebnisse der Analysen in konkrete Maßnahmen durchaus betonen. Jedoch wird bei keinem der bisher vorgestellten Prozessmodelle näher auf die hierzu notwendigen Prozess-Schritte eingegangen. Diesen aus betriebswirtschaftlicher Sicht entscheidenden Schritt der Einbettung des KDD-Prozesses in die Geschäftsprozesse des betrachteten Unternehmens beleuchten wir im nächsten Abschnitt näher.

3.2 Einbindung des KDD-Prozesses in das betriebliche Umfeld

Als erster konkreter Ansatz zur Einbettung des KDD-Prozesses in die Geschäftsprozesse eines Unternehmens kann der von Berry und Linoff entwickelte "virtuous cycle of data mining" genannt werden.³⁵ Dieser Ansatz wird sowohl von Thuraisingham als auch von Knobloch aufgegriffen und konkretisiert.³⁶

3.2.1 Ansatz von Berry und Linoff (1997, 2000)

Dem Modell von Berry und Linoff liegt die Erkenntnis zugrunde, dass die Extrahierung interessanter Informationen aus Daten allein nicht genügt. Vielmehr müssen aus den Informationen geeignete Maßnahmen abgeleitet werden, die dem Unternehmen Ertrag bringen. Um dies zu erreichen, muss der Data-Mining-Prozess in andere Geschäftsprozesse eingebettet werden. Das Hauptaugenmerk liegt hier somit nicht auf der Datenanalyse mit Data-Mining-Methoden sondern vielmehr auf den Maßnahmen, die auf Basis der mit Hilfe von Data-Mining-Methoden gewonnenen Erkenntnisse durchgeführt werden.³⁷

³⁵Vgl. *Berry/Linoff* (1997), S. 22–33, und *Berry/Linoff* (2000), S. 39–64.

³⁶Vgl. *Thuraisingham* (1999), S. 93–103, *Knobloch* (2001), S. 103–108, und *Knobloch* (2004), S. 340–342.

³⁷Vgl. *Berry/Linoff* (1997), S. 18.

Als Rahmen für die Einbettung von Data-Mining-Analysen in Geschäftsprozesse dient der so genannte "virtuous cycle of data mining". Dabei handelt es sich um ein Kreislauf-Modell bestehend aus vier Stufen. Die einzelnen Stufen hängen insofern voneinander ab, dass der Output einer Stufe als Input für die folgende Stufe dient.³⁸

In der ersten Stufe werden zunächst Fragestellungen identifiziert, bei denen Datenanalyse mit Data-Mining-Methoden von Nutzen sein kann. Anschließend werden Data-Mining-Methoden eingesetzt, um die Daten in unternehmerisch nutzbare Informationen zu transformieren. Die Informationen sollten dabei so ausgeprägt sein, dass sich geeignete Maßnahmen ableiten lassen ("actionable information"). In der dritten Stufe werden aus den extrahierten Informationen geeignete Maßnahmen abgeleitet und durchgeführt. Hierbei ist sicherzustellen, dass eine Bewertung der Maßnahmen möglich ist. Schließlich werden in der vierten Stufe die Ergebnisse der Maßnahmen gemessen, um herauszufinden, ob und ggf. wie die vorhandenen Daten optimal ausgeschöpft werden können. Dies ist eine wichtige Voraussetzung für eine schrittweise Verbesserung der Ergebnisse, wird jedoch in der Praxis häufig vernachlässigt. Idealerweise handelt es sich insgesamt um einen fortlaufenden Prozess im Sinne eines Kreislaufs.³⁹

Der von Berry und Linoff beschriebene Ansatz wird u. a. von Thuraisingham aufgegriffen. Die einzelnen Schritte werden in der Beschreibung jedoch nur wenig konkretisiert.⁴⁰

3.2.2 Ansatz von Knobloch (2001, 2004)

Knobloch beschreibt auf Basis des "virtuous cycle of data mining" ein Handlungsschema für betriebswirtschaftliche Datenanalyseprozesse, welches genau die von Berry und Linoff vier Stufen umfasst: "Spezifikation der Fragestellung", "Durchführung der Untersuchung", "Umsetzung der Untersuchungsergebnisse" und "Evaluierung der Untersuchungssituation". Die einzelnen Stufen werden dabei etwas konkretisiert.⁴¹

Knobloch weist außerdem auf einige wichtige Aspekte hin: so ist beispielsweise in der ersten Stufe insbesondere auch eine Beurteilung der zu erwartenden Wirtschaftlichkeit des Vorhabens dringend geboten. Darüber hinaus ist zu analysieren, inwiefern das Vorhaben in die Unternehmensorganisation einzubetten ist und ob die datenschutzrechtlichen

³⁸Vgl. *Berry/Linoff* (1997), S. 23.

³⁹Vgl. *Berry/Linoff* (1997), S. 23–30.

⁴⁰Vgl. *Thuraisingham* (1999), S. 98–100.

⁴¹Vgl. *Knobloch* (2001), S. 103–108, und *Knobloch* (2004), S. 340–342.

Rahmenbedingungen die Durchführung der Analysen erlauben.⁴²

3.2.3 Fazit

Die beiden vorgestellten Ansätze identifizieren und beschreiben die grundlegenden Phasen eines Prozessmodells, das die Einbettung von Data-Mining-Analysen in Geschäftsprozesse ermöglicht. Dabei werden auch einige in den einzelnen Phasen durchzuführenden Tätigkeiten erläutert. Die Beschreibung der Phasen ist dabei teilweise sehr grob bzw. exemplarisch und somit nicht im Sinne eines Referenzmodells verwendbar, da eine systematische Zusammenstellung aller Aufgaben einer Phase auf generischer Ebene nur teilweise gegeben ist.

Ein weiterer wichtiger und u.E. unbedingt zu berücksichtigender Aspekt ist die durchgehende Ausrichtung des Prozesses auf die zu Grunde liegenden betriebswirtschaftlichen Fragestellungen. Data-Mining ist dabei nur als Mittel zum Zweck zu sehen und darf deshalb nicht der Hauptfokus des Prozessmodells sein. Darüber hinaus ist die Integration einer ökonomischen Kosten-Nutzen-Analyse in die einzelnen Phasen eines solchen Prozesses aus betriebswirtschaftlicher Sicht dringend geboten.

Im Folgenden werden nun das CRISP-DM als Referenzmodell für den KDD-Prozess einerseits und die Ansätze von Berry und Linoff bzw. Knobloch zur Einbettung des KDD-Prozesses in die Geschäftsprozesse andererseits kombiniert und als Grundlage für ein umfassendes Prozessmodell verwendet. Anschließend wird das Modell um bisher nicht oder nur unzureichend berücksichtigte Aspekte erweitert und entsprechend weiterentwickelt. Ferner werden einige bisher nicht ausreichend präzierte Teilprozesse detaillierter ausgearbeitet. Ziel ist es dabei, ein Prozessmodell zu beschreiben, welches alle in der Literatur als wichtig erachteten Aspekte berücksichtigt und in einem konsistenten Modellansatz vereinigt. Dies umfasst sowohl die in Abschnitt 3.1 aufgeführten Prozessmodelle für KDD als auch die in Abschnitt 3.2 diskutierten Ansätze zur Einbettung des KDD-Prozesses in das betriebliche Umfeld. Dabei werden sowohl die einzelnen Phasen des Prozesses als auch die generischen Aufgaben präziert und vollständig beschrieben.

⁴²Vgl. *Knobloch* (2001), S. 105.

4. Generisches Kreislaufmodell zur Einbettung von Data-Mining-Analysen in die Geschäftsprozesse von Unternehmen

Hinsichtlich der Beschreibung des Prozessmodells orientieren wir uns an der bei CRISP-DM verwendeten Methodik.⁴³ Im Folgenden beschreiben wir zunächst die ersten beiden Abstraktionsebenen (Phasen und generische Aufgaben). Auf die dritte und vierte Ebene (spezialisierte Aufgaben und Prozessinstanzen) gehen wir im Rahmen des Fallbeispiels in Abschnitt 5 ein.

Als Ausgangspunkt für die Modellentwicklung wählen wir das CRISP-DM Modell. Dieses Modell erweitern wir nun, um eine Einbettung der Datenanalyse in die Geschäftsprozesse eines Unternehmens zu ermöglichen. Als Anhaltspunkt hierzu verwenden wir das von Berry und Linoff entwickelte Konzept des "virtuous cycle of data mining" und das darauf aufbauende Handlungsschema für betriebswirtschaftliche Datenanalyseprozesse von Knobloch.

4.1 Das generische Kreislaufmodell

Das Modell sieht grundsätzlich acht Phasen vor, wobei sich die einzelnen Phasen insgesamt vier Stufen analog zum "virtuous cycle of data mining" zuordnen lassen (vgl. Abbildung 2). Die Phasen orientieren sich an den Phasen des CRISP-DM Modells. Im Einzelnen besteht das Modell aus den folgenden Phasen bzw. Stufen:

- I. Spezifikation der Fragestellung
 - Spezifikation der Ziele und Rahmenbedingungen
 - Analyse der Datengrundlage
- II. Durchführung der Data-Mining-Analysen
 - Vorverarbeitung der Daten
 - Analyse der Daten (Data-Mining)
 - Evaluierung der Ergebnisse
- III. Umsetzung der Erkenntnisse in Maßnahmen
 - Ableitung geeigneter Maßnahmen
 - Implementierung der Maßnahmen
- IV. Messung und Bewertung der Ergebnisse

⁴³Vgl. Abschnitt 3.1.2.

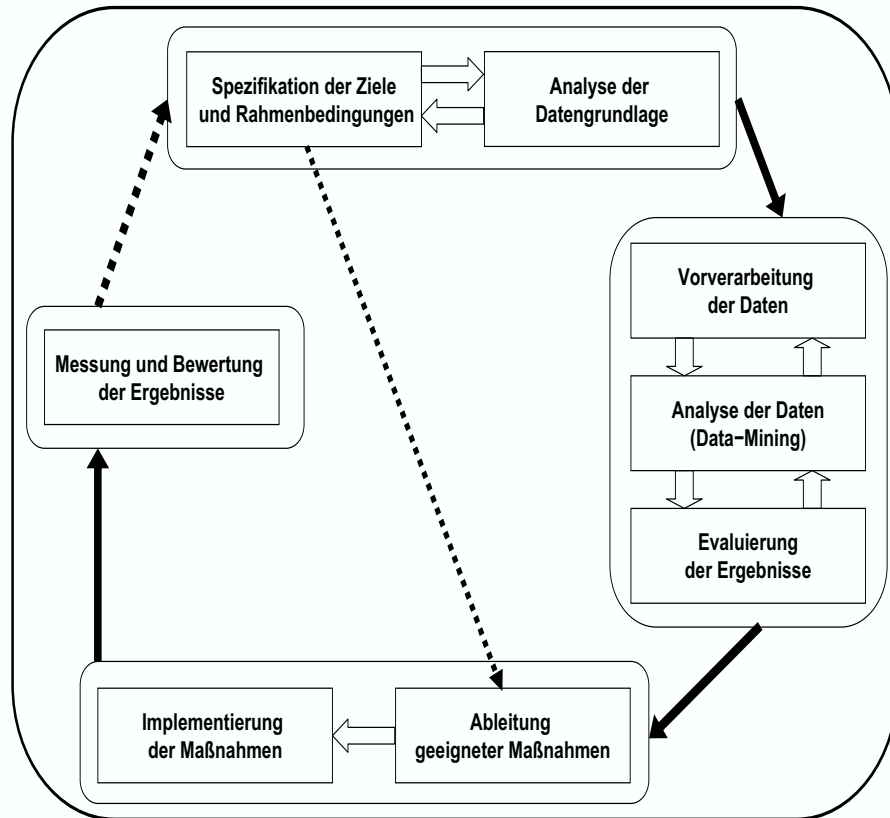


Abbildung 2: Generisches Kreislaufmodell zur Einbettung von Data-Mining-Analysen in die Geschäftsprozesse (Quelle: eigene Darstellung)

Die einzelnen Stufen bauen insofern aufeinander auf, dass der Output einer Stufe als Input für die folgende(n) Stufe(n) dient. Die Phasen werden dabei in der Regel nicht linear durchlaufen, vielmehr sind vielfältige Rückkopplungen zwischen den Phasen vorgesehen. Dies gilt insbesondere für die einzelnen Phasen innerhalb einer Prozess-Stufe und wird in Abbildung 2 durch Pfeile zwischen den Phasen einer Prozess-Stufe angedeutet. Zusätzlich wird in der Abbildung die enge Verknüpfung zwischen der Spezifikation der Ziele und Rahmenbedingungen in Stufe I und der Ableitung geeigneter Maßnahmen in Stufe III durch einen gestrichelten Pfeil angedeutet. Hierdurch soll verdeutlicht werden, dass zum einen bei der Ableitung der Maßnahmen die vorher ausgearbeiteten Ziele und Rahmenbedingungen zu berücksichtigen sind und zum anderen die gesamte Datenanalyse in Stufe II letztlich Mittel zum Zweck ist, um möglichst gut geeignete Maßnahmen zur Erreichung der betriebswirtschaftlichen Ziele zu konzipieren.

Das vorgeschlagene Modell geht dabei insbesondere in den folgenden Aspekten über die bisher in der wissenschaftlichen Literatur diskutierten Ansätze hinaus:

- Es ist als ein Kreislauf konzipiert, der vollständig alle Phasen des Prozesses zur systematischen Analyse von Daten und zur Einbindung derartiger Analysen in Geschäftsprozesse beschreibt. Der Prozess ist dabei grundlegend auf die Frage ausgerichtet, welchen Nutzen das Unternehmen aus derartigen Analysen zieht. Die Modellierung als Kreislauf ist dabei von grundsätzlicher Bedeutung, da die Nutzung intelligenter Datenanalysen erst dadurch zu einem brauchbaren strategischen Werkzeug wird, dass sie regelmäßig durchgeführt wird. Dies ermöglicht es, die entwickelten Modelle und angewendeten Maßnahmen auf ihre Wirksamkeit hin zu überprüfen, und gleichzeitig ergeben sich erhebliche Einsparungspotenziale durch die wiederholte Nutzung des vorher gewonnenen Know-hows und der im Rahmen des Prozesses entwickelten Modelle und Maßnahmen.
- Unser Modell legt konsequent Wert auf den Einsatz geeigneter Kosten-Nutzen-Analysen zur Überprüfung der Wirksamkeit von Modellen und Maßnahmen. Hierzu sehen wir den Einsatz monetärer Kosten-Nutzen-Analysen in verschiedenen Prozessphasen vor – ein Aspekt, der in allen bisherigen Modellen nur unzureichend berücksichtigt wird.
- Unser Modell enthält neben einer generischen Beschreibung der einzelnen Modellphasen zusätzlich deren konkrete Ausformulierung in Form der einzelnen im Rahmen der jeweiligen Phase durchzuführenden Aufgaben. Erst hierdurch wird das Modell konkret einsetzbar. Im Rahmen eines Fallbeispiels in Abschnitt 5 werden wir zeigen, wie das Modell auf versicherungsspezifische Fragestellungen angewendet werden kann.

Die genannten Aspekte werden im folgenden Abschnitt im Rahmen einer detaillierten Beschreibung der einzelnen Phasen verdeutlicht.

Voraussetzung für die Initiierung eines solchen Prozesses ist die Identifikation von Geschäftsprozessen, bei denen mithilfe der zur Verfügung stehenden Daten das Potenzial besteht, durch intelligente datengetriebene Datenanalyse (Data-Mining) Wissen zu generieren, welches in konkrete Maßnahmen zur besseren Steuerung eines bestehenden Geschäftsprozesses bzw. zum Aufsetzen eines neuen Geschäftsprozesses umgesetzt werden kann. Ziel ist es hierbei, insgesamt betriebswirtschaftlichen Nutzen zu generieren und zur besseren Zielerreichung des Unternehmens beizutragen. Wurde ein solches Potenzial grundsätzlich erkannt, so kann die erste Stufe des Prozesses initiiert werden.

Mögliche Ansatzpunkte zur Identifikation des Potenzials von Data-Mining-Analysen sind die Betrachtung von Geschäftsprozessen, die sehr stark auf Daten beruhen, das Sam-

meln von Fragen und Beobachtungen, die im Unternehmen gemacht werden, und die Befragung von Entscheidungsträgern im Unternehmen.⁴⁴

Eine solche Potenzialanalyse ist mit dem Kreislauf insofern verknüpft, dass zum einen ein konkreter Kreislauf durch eine derartige Potenzialanalyse initiiert wird und zum anderen die Ergebnisse eines konkreten Kreislaufs in zukünftige Potenzialanalysen einfließen sollten.

4.2 Die einzelnen Stufen des Kreislaufs im Detail

Im folgenden Abschnitt beschreiben wir die wichtigsten Aspekte der einzelnen Stufen sowie die Interaktionen zwischen den Stufen.

4.2.1 Spezifikation der Fragestellung

Die in der Literatur diskutierten Prozessmodelle sehen als erste Stufe und grundlegende Voraussetzung für den Gesamtprozess meist die eindeutige Spezifikation der Fragestellung vor.

Vor Beginn der eigentlichen Datenanalyse müssen insbesondere folgende Fragen beantwortet werden: Was wollen wir wissen und was wollen wir mit dem Wissen anfangen? Notwendig ist also die Formulierung eines spezifischen Informationsbedürfnisses bezogen auf eine spezifische Maßnahme. Dies unterstreicht die Tatsache, dass Data-Mining sinnlos ist, wenn die Ergebnisse der Analysen nicht in geeignete Maßnahmen umgesetzt werden.⁴⁵

Eine ausführliche Diskussion der Fragestellungen und Ziele ist insbesondere deshalb notwendig, weil sichergestellt werden muss, dass nicht unnötig viel Zeit darauf verwendet wird, die falschen Fragen zu beantworten. Komplementär hierzu ist allerdings auch eine grundlegende Analyse der Datengrundlage hinsichtlich Struktur, Umfang und Qualität erforderlich. Hierdurch wird sichergestellt, dass vorhandenen Daten grundsätzlich zur Beantwortung der Fragestellungen geeignet sind.⁴⁶

Diese erste Stufe des Prozesses erfordert eine enge Zusammenarbeit zwischen den Experten aus dem Anwendungsgebiet, welche über das notwendige Hintergrundwissen verfügen,

⁴⁴Vgl. *Berry/Linoff* (1997), S. 23–25.

⁴⁵Vgl. *Adriaans/Zantinge* (1996), S. 81.

⁴⁶Vgl. *Brachman/Anand* (1996), S. 48f.

und den Datenanalysten, welche die Ziele in konkrete Anforderungen an die Data-Mining-Analysen übersetzen müssen. Evtl. vorhandene Missverständnisse zwischen den Experten aus dem Anwendungsgebiet und den Datenanalysten müssen möglichst frühzeitig beseitigt werden.⁴⁷

Die Fragestellung muss dann (evtl. durch Zerlegung in Teilprobleme) präzise genug formuliert werden, damit sie zu einem lösbar ist und zum anderen die Ergebnisse messbar sind. Darüber hinaus sind auch realistische Erwartungen an das Projekt zu formulieren. Außerdem ist eine erste Kosten-Nutzen-Analyse von essenzieller Bedeutung.⁴⁸

Auf Basis der ausgearbeiteten betriebswirtschaftlichen Fragestellung muss dann die Fragestellung für die Datenanalyse präzise formuliert werden, um sicherzustellen, dass sich die Ergebnisse anschließend auch implementieren lassen. Außerdem muss die Vorgehensweise bei der Implementierung der auf Basis der Ergebnisse der Datenanalyse abgeleiteten Maßnahmen ausgearbeitet werden.⁴⁹

Aufbauend auf diesen Erkenntnissen schlagen wir analog zu CRISP-DM eine Aufteilung der ersten Stufe in zwei Phasen vor.⁵⁰ Die erste Phase umfasst die Spezifikation der Ziele und Rahmenbedingungen, wohingegen die zweite Phase eine grundlegende Analyse der Datengrundlage zum Inhalt hat. Die beiden Phasen sind eng miteinander verknüpft.⁵¹

In der ersten Phase sind folgende generische Aufgaben vorgesehen:

- Sammlung von Hintergrundinformationen über die Ausgangslage des Unternehmens
- Ausarbeitung der betriebswirtschaftlichen Ziele
- Ausarbeitung der möglichen Maßnahmen zur Erreichung der Ziele
- Ausarbeitung der Vorgehensweise bei der Implementierung der Maßnahmen
- Festlegung von Erfolgskriterien aus betriebswirtschaftlicher Sicht
- Analyse der Rahmenbedingungen
- Analyse der zu erwartenden Wirtschaftlichkeit
- Ableitung der Ziele für die Data-Mining-Analysen
- Festlegung von Erfolgskriterien für die Data-Mining-Analysen

⁴⁷Vgl. *John* (1997), S. 6f.

⁴⁸Vgl. *Cabena/Hadjinian/Stadler/Verhees/Zanasi* (1998), S. 45–47.

⁴⁹Vgl. *Pyle* (1999), S. 19–22.

⁵⁰Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 16–22 (Phasen "business understanding" und "data understanding").

⁵¹Diese Verknüpfung wird in Abbildung 2 durch Pfeile zwischen den beiden Phasen angedeutet.

- Erstellung eines Projektplans

Die Aufgaben entsprechen im Wesentlichen den Aufgaben in der Phase "business understanding" im CRISP-DM.⁵² Die einzelnen Aufgaben hängen voneinander ab und müssen deshalb evtl. mehrmals iterativ durchlaufen werden.

Im betrieblichen Umfeld sind dabei zwei Aspekte besonders hervorzuheben. Zum einen betrifft dies die Spezifikation möglicher Maßnahmen, die auf Basis des generierten Wissens ergriffen werden könnten. Hierzu ist ein Katalog möglicher Maßnahmen zu erarbeiten. Außerdem ist zur Vorbereitung von Phase 7 die Implementierung der Maßnahmen zu diskutieren. In Anlehnung an Pyle können hier fünf der sechs W's als Orientierung dienen: wer, wie, was, wann, wo.⁵³

Zum zweiten kommt im betrieblichen Umfeld der Wirtschaftlichkeitsanalyse besondere Bedeutung zu. Diese umfasst eine erste Abschätzung von potenziellem Nutzen sowie von entstehenden Kosten und soll die Frage beantworten, ob bei den gegebenen Daten, Rahmenbedingungen und möglichen Maßnahmen eine sinnvolle Zielerreichung möglich erscheint. Hierzu sind zum einen die Kosten für die Datenbeschaffung und die Durchführung der Analysen sowie die Kosten für die Ableitung und Durchführung der Maßnahmen zu prognostizieren. Diesen wird dann eine monetäre Bewertung des Nutzens durch die Umsetzung der Maßnahmen gegenübergestellt. Aus dieser Kosten-Nutzen-Analyse ergeben sich Anforderungen sowohl an die Genauigkeit der Ergebnisse der Datenanalyse als auch an die Wirksamkeit der Maßnahmen. Diese können beispielsweise in Form von Benchmarks festgehalten werden. Verschiedene Ansätze zur Wirtschaftlichkeitsbetrachtung werden von Küppers vorgestellt⁵⁴. Die Wirtschaftlichkeitsanalyse wird in Phase 7 aufgegriffen und verfeinert sowie schließlich in Phase 8 auf Basis der tatsächlichen Kosten und des tatsächlichen Nutzens aktualisiert.

Die zweite Phase ("Analyse der Datengrundlage") umfasst folgende generische Aufgaben:

- Beschaffung der vorhandenen Daten
- Anreicherung mit externen Daten
- Beschreibung der vorhandenen Daten
- Exploration der vorhandenen Daten

⁵²Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 16–19.

⁵³Vgl. *Pyle* (1999), S. 22. Das sechste w (warum) wurde bereits bei der Ausarbeitung der betriebswirtschaftlichen Ziele behandelt.

⁵⁴Vgl. *Küppers* (1999), S. 117–121.

- Überprüfung der Datenqualität

Diese Aufgaben entsprechen im Wesentlichen den Aufgaben in der Phase "data understanding" im CRISP-DM.⁵⁵ Lediglich die Beschaffung externer Daten wird hier als separate Aufgabe aufgeführt. Dies geschieht in Anlehnung an Adriaans und Zantinge, die der Anreicherung der vorhandenen Daten mit externen Daten besondere Bedeutung beimessen.⁵⁶ Der Aufwand für diese Phase darf im betrieblichen Umfeld nicht unterschätzt werden. Dies gilt insbesondere dann, wenn die Daten ursprünglich nicht zum Zweck der Datenanalyse erfasst wurden und bisher nicht weiter aufbereitet wurden.

Zwischen diesen beiden Phasen bestehen vielfältige Abhängigkeiten, die ein streng lineares Durchlaufen der Phasen nicht möglich machen. Ergebnis der ersten beiden Phasen und somit der ersten Stufe des Prozesses ist eine auf der Spezifikation der betriebswirtschaftlichen Ziele und Rahmenbedingungen sowie auf der Analyse der Datengrundlage basierende, klar definierte und auf die Datengrundlage abgestimmte Aufgabenstellung für die Datenanalyse. Außerdem wird bereits in der ersten Stufe die Ableitung und Durchführung von Maßnahmen, sowie die Messung der Ergebnisse inkl. der Wirtschaftlichkeitsanalyse vorbereitet.

4.2.2 Durchführung der Data-Mining-Analysen

Grundlage für die Durchführung der Data-Mining-Analysen sind die in der ersten Stufe klar definierte Aufgabenstellung und Zielsetzung sowie die darauf abgestimmte Datengrundlage.

Die Durchführung der Data-Mining-Analysen unterteilen wir hier in drei Phasen, wobei sich die Phasen am CRISP-DM-Modell orientieren. Den Mittelpunkt bildet dabei die eigentliche Data-Mining-Analyse. Sie wird umrahmt von der notwendigen Vorverarbeitung der Daten sowie von der Evaluierung der Ergebnisse. Wie schon in der ersten Stufe sind vielfältige Iterationen zwischen diesen drei Phasen notwendig.⁵⁷

Im Folgenden stellen wir die in der jeweiligen Phase durchzuführenden generischen Aufgaben dar. Details zu den einzelnen Aufgaben finden sich bei CRISP-DM sowie bei Knobloch.⁵⁸ Außerdem wird im Rahmen des Fallbeispiels in Abschnitt 5 auf einige wichtige

⁵⁵Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 20–19.

⁵⁶Vgl. *Adriaans/Zantinge* (1996), S. 38.

⁵⁷Diese Verknüpfungen werden in Abbildung 2 durch Pfeile zwischen den drei Phasen angedeutet.

⁵⁸Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 48–59, und *Knobloch* (2001), S. 87–103.

Aspekte näher eingegangen.

Phase 3: Vorverarbeitung der Daten

- Selektion (Datensätze, Felder)
- Bereinigung (Datenqualität)
- Manipulation (abgeleitete Attribute, abgeleitete Datensätze)
- Integration (Zusammenführung verschiedener Datenquellen, ggf. Aggregation)
- Formatierung (Anforderung des Software-Tools)

Phase 4: Analyse der Daten (Data-Mining)

- Auswahl der Modellierungsmethode(n) (Festhalten der zugrunde liegenden Annahmen)
- Festlegung des Test-Designs (Validierung und Test des Modells)
- Modellkalibrierung (Parametereinstellungen, Beschreibung der jeweiligen Modelle)
- Technische Evaluierung (Genauigkeit und Generalisierbarkeit, Parameteroptimierung, dabei Berücksichtigung der Data-Mining-Ziele)

Phase 5: Evaluierung der Ergebnisse

- Bewertung der Ergebnisse (Abgleich mit Data-Mining-Zielen)
- Review des Analyseprozesses
- Bestimmung der nächsten Schritte (Umsetzung der Ergebnisse in Maßnahmen oder Versuch der Verbesserung der Ergebnisse)

Besonders hervorzuheben ist die Tatsache, dass hierbei ein Abgleich mit den in der ersten Stufe spezifizierten Zielen und Erfolgskriterien stattfinden muss. Ergebnis der zweiten Stufe sind Ergebnisse, die anhand der in der ersten Phase spezifizierten Erfolgskriterien bewertet wurden und zur Ableitung geeigneter Maßnahmen grundsätzlich geeignet sind.

4.2.3 Umsetzung der Erkenntnisse in Maßnahmen

Bei der nun folgenden dritten Stufe handelt es sich um eine entscheidende und keinesfalls triviale Stufe im Prozess: die mit Hilfe der Data-Mining-Analysen gewonnenen Informationen müssen nun in den betrachteten Geschäftsprozess eingebracht werden. Denn nur durch die Umsetzung der Erkenntnisse in konkrete Maßnahmen können durch die Datenanalyse ökonomische Vorteile für das Unternehmen erreicht werden. Außerdem darf bei

der Planung hierfür geeigneter Maßnahmen die Frage der Messung der Ergebnisse nicht vernachlässigt werden.⁵⁹

Hieraus ergeben sich zwei zentrale Fragen für diese Phase:

- Welche Maßnahmen sollen basierend auf den Ergebnissen der Analysen eingeleitet werden, um die Zielerreichung zu gewährleisten?
- Wie müssen die Maßnahmen ausgestaltet werden, damit eine Überprüfung der Wirksamkeit der Maßnahmen möglich ist?

Die hierfür notwendigen grundlegenden Überlegungen wurden bereits in ersten Stufe erarbeitet. Zu beachten ist jedoch, dass es aufgrund der Vielfältigkeit der mit Hilfe von Data-Mining extrahierbaren Informationen viele mögliche, auf diesen Informationen basierende Maßnahmen gibt. Darüber hinaus spielen die Ziele und Rahmenbedingungen des Unternehmens eine entscheidende Rolle bei der Ableitung geeigneter Maßnahmen. In dieser Phase sind deshalb die Experten aus dem Anwendungsgebiet sehr stark involviert.

Darüber hinaus ist festzuhalten, dass das Vorhandensein eines für die Fragestellung geeigneten Modells nicht notwendigerweise bedeutet, dass auf Basis der Erkenntnisse des Modells sofort geeignete Maßnahmen durchgeführt werden können.⁶⁰ Im Folgenden werden deshalb alle für die Ableitung geeigneter Maßnahmen notwendigen grundlegenden Tätigkeiten dargestellt.

Grundsätzlich sollten die Ergebnisse der Datenanalyse und ihr Bezug zu den betriebswirtschaftlichen Fragestellungen und Zielen zunächst in Form eines Reports zusammengefasst werden. Wichtig ist hierbei eine betriebswirtschaftlich orientierte Darstellung. Darauf aufbauend sollten Wege aufgezeigt werden, wie die gewonnenen Erkenntnisse bestmöglich betriebswirtschaftlich genutzt werden können.⁶¹ Hierzu sind insbesondere folgende mögliche Maßnahmen zu nennen: die Durchführung einmaliger Maßnahmen (z. B. gezielte Marketingmaßnahmen), die Modifikation operativer Geschäftsprozesse (z. B. Einteilung neuer Kunden in Kundenklassen), die Einrichtung eines auf den Erkenntnissen beruhenden Frühwarnsystems sowie Verbesserungen der Datenhaltung.⁶²

Die Tätigkeiten dieser Stufe sind teilweise vergleichbar mit den Tätigkeiten in der Phase "deployment" in CRISP-DM.⁶³ Im betrieblichen Umfeld ist jedoch eine Konkretisierung notwendig. Hierzu werden die Tätigkeiten zwei Phasen zugeordnet: zunächst gilt

⁵⁹Vgl. *Berry/Linoff* (1997), S. 27f., und *Knobloch* (2001), S. 106.

⁶⁰Vgl. *John* (1997), S. 22f.

⁶¹Vgl. *Cabena/Hadjinian/Stadler/Verhees/Zanasi* (1998), S. 59.

⁶²Vgl. *Brachman/Anand* (1996), S. 47f., und *Knobloch* (2001), S. 106.

⁶³Vgl. *Chapman/Clinton/Kerber/Khabaza/Reinartz/Shearer/Wirth* (2000), S. 60–62.

es, geeignete Maßnahmen abzuleiten. Anschließend müssen diese gewählten Maßnahmen implementiert werden.

Im Einzelnen umfasst Phase 6 ("Ableitung geeigneter Maßnahmen") folgende Tätigkeiten:

- Zusammenstellung möglicher Maßnahmen:

Grundlage sind die Ergebnisse der ersten Phase (betriebswirtschaftliche Ziele, Katalog möglicher Maßnahmen, Erfolgskriterien) sowie die evaluierten Ergebnisse der Datenanalyse aus der zweiten Stufe (insbesondere hinsichtlich der Genauigkeit und Validität der gewonnenen Erkenntnisse).

Zunächst wird in Form eines Reports der Bezug zwischen den Ergebnissen der Datenanalyse und den betriebswirtschaftlichen Fragestellungen hergestellt. Dabei sind insbesondere alle grundsätzlich in Maßnahmen umsetzbare Erkenntnisse zu ermitteln und im Hinblick auf Rahmenbedingungen und Restriktionen des Unternehmens zu bewerten. Auch auf mögliche Schwierigkeiten bei der Umsetzung ist einzugehen.

- Wirtschaftlichkeitsanalyse:

Anschließend werden alle denkbaren Maßnahmen hinsichtlich der zu erwartenden Wirtschaftlichkeit bewertet. Hierzu wird die bereits in der ersten Stufe durchgeführte monetäre Kosten-Nutzen-Analyse auf Basis der gewonnenen Erkenntnisse aktualisiert und ggf. verfeinert.

- Auswahl der Maßnahmen:

Auf Basis der Wirtschaftlichkeitsanalyse möglicher Maßnahmen werden dann unter Berücksichtigung aller anderen relevanten Restriktionen die durchzuführenden Maßnahmen ausgewählt.

- Strukturierung der Maßnahmen:

Die Maßnahmen müssen anschließend geeignet strukturiert werden, um eine hinreichend detaillierte Auswertung der Ergebnisse zu ermöglichen. Dabei soll zum einen eine Überprüfung der mit den Data-Mining-Analysen gewonnenen Informationen möglich sein. Zum anderen wird auch eine (monetäre) Bewertung des Erfolgs der Maßnahmen sichergestellt und vorbereitet.

Die sich anschließende Phase 7 ("Implementierung der Maßnahmen") umfasst folgende Tätigkeiten:

- Planung der Durchführung und Überwachung der Maßnahmen:

Bei der Planung der Durchführung der Maßnahmen ist besonderes Augenmerk darauf zu legen, dass alle für die spätere Bewertung notwendigen Daten erfasst werden.

Außerdem muss sichergestellt werden, dass bei der operativen Umsetzung die Maßnahmen tatsächlich wie geplant durchgeführt werden und die Ergebnisse erfasst werden.

Ein weiterer wichtiger Punkt ist die Analyse der dynamischen Aspekte bei der Umsetzung, also die Frage inwiefern sich Rahmenbedingungen ändern können, sodass entweder die Ergebnisse der Data-Mining-Analysen nicht länger gültig sind oder die gewählten Maßnahmen nicht weiter sinnvoll erscheinen. Hierzu sind geeignete Vorgehensweisen zur fortlaufenden Überwachung der Maßnahmen zu definieren.

- Durchführung und Überwachung der Maßnahmen:

Hierzu sind zunächst alle für die Maßnahmen notwendigen Vorarbeiten abzuschließen. Anschließend kann dann die planmäßige Durchführung und Überwachung der Maßnahmen sowie die Erfassung der Ergebnisse beginnen.

4.2.4 Messung und Bewertung der Ergebnisse

Die Bewertung sowohl der Qualität der gewonnenen Erkenntnisse als auch des Erfolgs der daraus abgeleiteten Maßnahmen stellt ein wichtiges Feedback dar, welches Grundvoraussetzung für eine schrittweise Verbesserung der Zielerreichung ist. Die Bewertung bezieht sich dabei auf die in der ersten Phase erarbeiteten Ziele und Erfolgskriterien. Erst durch diese Stufe wird also der Kreislauf geschlossen. Die detaillierte Analyse der Ergebnisse kann dabei sowohl Ansatzpunkte für zukünftige Verbesserungen liefern als auch weitere interessante Fragen aufwerfen. Hierzu sollten die Ergebnisse auch im Unternehmen kommuniziert werden, um die Potenziale von Data-Mining-Analysen zu verdeutlichen.⁶⁴

Trotz der essenziellen Bedeutung dieser Stufe wird der Messung und Bewertung der Ergebnisse oft nicht ausreichend Aufmerksamkeit geschenkt. Dies zeigt sich z. B. darin, dass dieser Schritt bei den in Abschnitt 3 aufgeführten Prozessmodellen entweder gar nicht explizit berücksichtigt wird oder nur sehr grob erläutert wird.

Im Einzelnen sind in dieser Stufe folgende Tätigkeiten vorgesehen:

- Bewertung der Erkenntnisse der Data-Mining-Analysen:

Die mithilfe der eingesetzten Analyseverfahren abgeleiteten Erkenntnisse werden anhand realer Daten validiert. Hierdurch lässt sich feststellen, ob die eingesetzten Modelle angemessen sind. Dabei erfolgt nochmals ein Abgleich mit den in der ersten Phase formulierten Zielen der Datenanalyse.

⁶⁴Vgl. *Berry/Linoff* (1997), S. 28–30, und *Knobloch* (2001), S. 106–108.

Außerdem wird durch die Validierung überwacht, ob Modelle und Erkenntnisse veraltet sind und aktualisiert werden müssen.

- **Bewertung der Maßnahmen:**
Hierbei wird analysiert, welche Rückschlüsse sich aus den Ergebnissen der Maßnahmen ziehen lassen. Wesentlicher Aspekt ist dabei eine monetäre Kosten-Nutzen-Analyse, d. h. eine Quantifizierung des betriebswirtschaftlichen Nutzens der Maßnahmen und eine Gegenüberstellung des Aufwands. Wiederum erfolgt ein Abgleich mit den in der ersten Phase formulierten betriebswirtschaftlichen Zielen.
- **Gesamtbewertung des Kreislaufs in einem abschließenden Report:**
Abschließend sind die einzelnen Stufen des Kreislaufs zu bewerten. Dazu gehört die Frage, ob die Zielsetzung sinnvoll und erreichbar war, ob die Datenbasis geeignet war, ob die Erkenntnisse aus der Datenanalyse zufriedenstellend waren und ob die ergriffenen Maßnahmen erfolgreich waren.

Dabei ist insbesondere zu analysieren, inwiefern die Ergebnisse der Datenanalyse und die Maßnahmen aufeinander abgestimmt waren und zueinander gepasst haben. Es sind durchaus Situationen denkbar, in denen die Ergebnisse der Datenanalyse äußerst zufriedenstellend sind, die anschließend ergriffenen Maßnahmen jedoch nicht den gewünschten Erfolg bringen. Diesen Interaktionen zwischen Modellen und Maßnahmen ist besondere Aufmerksamkeit zu schenken.

Die so gewonnenen Erkenntnisse und erkannten Verbesserungsmöglichkeiten werden in einem abschließenden Report zusammengefasst. Alle Erfahrungen im Prozessverlauf sollten ebenfalls erfasst werden. Darauf basierend kann dann ein weiterer Zyklus initiiert werden.

4.3 Die Anwendung des Kreislaufmodells

Das im vorherigen Abschnitt dargestellte Modell sieht die Einbettung von Data-Mining-Analysen in Geschäftsprozesse nicht als einmalige Tätigkeit an. Vielmehr sollte es sich um einen kontinuierlichen Prozess handeln, bei dem der Kreislauf immer wieder durchlaufen wird. Jeder Zyklus liefert dabei Rückschlüsse auf die Wirksamkeit der Modelle und Maßnahmen und damit zum Potenzial von Data-Mining-Analysen. Diese Erfahrungen sind bei einem erneuten Durchlauf unbedingt zu berücksichtigen. In Abbildung 2 wird dies durch den gestrichelten Pfeil zwischen Stufe IV und Stufe I angedeutet.

Ein erneuter Durchlauf des Kreislaufs ist dabei i. d. R. viel einfacher und kostengünstiger als der erste Durchlauf, da auf den (bei korrekter Anwendung des Modells) gut dokumentierten Erfahrungen der bisherigen Durchläufe aufgebaut werden kann. Insbesondere bleiben die betriebswirtschaftlichen Fragestellungen oft über einen längeren Zeitraum im Wesentlichen unverändert und gleichermaßen relevant. Gerade dann eignet sich das Kreislaufmodell besonders, da hierdurch sowohl die Effizienz der Analysen als auch die Qualität der Ergebnisse schrittweise gesteigert werden kann. Dabei sollte eine weitgehende Standardisierung der Vorgehensweise angestrebt werden, insbesondere im Hinblick auf die oftmals besonders zeitaufwändige Datenaufbereitung. Aus den Ergebnissen können sich aber auch ganze neue Fragestellungen ergeben, die wiederum zur Initiierung eines komplett neuen Kreislaufs führen können.

5. Anwendung des Modells auf die Bestandssicherung in der Unfallversicherung

Das im vorherigen Abschnitt auf generischer Ebene beschriebene Modell zur Einbindung von Data-Mining-Analysen in Geschäftsprozesse wird nun anhand eines praxisorientierten Fallbeispiels näher erläutert. Dabei gehen wir auf die konkrete Ausgestaltung der im Modell generisch beschriebenen Aufgaben und deren konkrete Abfolge in einer Prozessinstanz ein. Hierzu werden im Folgenden für jede Phase des in Abschnitt 4 entwickelten Modells einige Anmerkungen zu konkreter Realisierung im Kontext der Bestandssicherung in der Unfallversicherung gemacht.⁶⁵ Die Darstellung ist dabei nicht vollständig, sondern dient vornehmlich dazu, die besonders wichtigen Aspekte hervorzuheben.

Beim betrachteten Versicherungsunternehmen, einem großem deutschen Unfallversicherer, gab es schon seit längerer Zeit Überlegungen und auch konkrete Maßnahmen zur Bestandssicherung in der Unfallversicherung. Hierzu wurden ausgewählte Kunden, deren Unfallversicherungsvertrag ablief, mit speziellen Direktmarketingmaßnahmen angesprochen, um die Kunden zu einer Verlängerung ihres Vertrags zu bewegen. Die Auswahl der anzusprechenden Kunden basierte dabei im Wesentlichen auf allgemeinen Restriktionen von Direktmarketingmaßnahmen sowie auf Erfahrungswerten der beteiligten Fachbereiche.

Da diese Direktmarketingmaßnahmen in der Vergangenheit oft nicht den gewünschten Erfolg gebracht hatten, entwickelte sich die Idee, die Maßnahmen gezielter auf diejenigen

⁶⁵Vgl. hierzu auch die Ausführungen in *Reuss/Zwiesler* (2004).

Kunden auszurichten, welche besonders stornogefährdet sind. Da umfangreiche Daten über in der Vergangenheit stornierte und verlängerte Verträge vorlagen, bot sich hierzu die Anwendung von Data-Mining-Verfahren an, um geeignete Modelle zur Prognose des Stornoverhaltens zu entwickeln.

Hier wurde also mit der Bestandssicherung in der Unfallversicherung durch Direktmarketingmaßnahmen ein Geschäftsprozess identifiziert, bei dem die Einbettung von Data-Mining-Analysen deutliches Verbesserungspotenzial versprach. Auf Basis dieser grundlegenden Ideen wurde die erste Stufe des Prozesses initiiert.

5.1 Spezifikation der Fragestellung

Zunächst galt es, die Ausgangslage des Unternehmens zu analysieren und die betriebswirtschaftlichen Ziele auszuarbeiten.

Bei der Unfallversicherung handelt es sich um eine Sparte, die sowohl durch hohe Rentabilität als auch durch hohen Wettbewerbsdruck gekennzeichnet ist. Dies zeigt sich einerseits durch branchenweit niedrige Schadenquoten, andererseits jedoch auch durch eine hohe Anzahl von Verträgen, die nach Ablauf der vereinbarten Mindestlaufzeit storniert werden. Die hohe Stornoquote führte beim betrachteten Unternehmen dazu, dass trotz erfolgreichen Neugeschäfts der Marktanteil in den letzten Jahren nicht gesteigert werden konnte.

Aus betriebswirtschaftlicher Sicht war deshalb eine Senkung der Stornoquote anzustreben, um den Bestand und die damit verbundenen Erträge zu sichern. Hierzu sollten geeignete Direktmarketingmaßnahmen durchgeführt werden. Um Streuverluste zu vermeiden, sollten die Maßnahmen auf die besonders stornogefährdeten Verträge ausgerichtet werden. Hierzu war grundsätzlich vorgesehen, den Kunden, deren Verträge zur Verlängerung anstanden, per Brief oder Telefon verschiedene Angebote zu unterbreiten. Dabei standen Angebote zur Umstellung auf neue Versicherungsbedingungen (mit entsprechend günstigeren Prämien) bzw. Angebote zum Einschluss zusätzlicher Leistungen sowie allgemeine Informationen zur Unfallversicherung zur Auswahl. Bei der Implementierung der Maßnahmen war die Bildung geeigneter Kontrollgruppen zu beachten, um eine Messung des Erfolgs der Maßnahmen zu ermöglichen.

Entscheidend war an dieser Stelle, dass eine Verbindung zwischen den betriebswirtschaftlichen Zielen einerseits (hier: Reduktion der Stornoquote) und den möglichen Maßnahmen zur Erreichung der Ziele andererseits (hier: Direktmarketingmaßnahmen zur Sen-

kung der Stornoquote) hergestellt wurde. Darauf aufbauend wurden die Ziele für die Datenanalyse erarbeitet.

Hierbei war es wichtig, zunächst den Begriff "Storno" im Hinblick auf die geplanten Maßnahmen exakt zu definieren: unter "Storno" sollte hier die Kündigung durch den Versicherungsnehmer zum Ende der vereinbarten Mindestlaufzeit verstanden werden. Kündigt der Versicherungsnehmer nicht bis spätestens 3 Monate vor Ablauf, so verlängert sich der Vertrag automatisch um ein weiteres Jahr. Andere Stornogründe, wie z. B. die Kündigung durch den Versicherer im Schadenfall oder die Kündigung wegen Tod des Versicherungsnehmers haben nur untergeordnete Bedeutung und wurden bei den folgenden Analysen nicht betrachtet.

Dies hatte erhebliche Auswirkungen auf die Spezifikation der Aufgabenstellung für die Datenanalyse. Hierzu sollten zunächst alle innerhalb eines gewissen Zeitraums (1 Jahr) ablaufenden Verträge betrachtet werden. Da dieser Zeitraum in der Vergangenheit lag, konnte für jeden dieser Verträge festgestellt werden, ob er verlängert oder durch den Versicherungsnehmer storniert wurde. Durch eine Gegenüberstellung von verlängerten und stornierten Verträge sollten Kombinationen von Vertragsmerkmalen identifiziert werden, die Verträge mit besonders hoher Stornogefährdung kennzeichnen. Aus fachlicher Sicht war dabei eine Beschreibung bzw. Interpretation der so identifizierten Bestandssegmente wünschenswert.

Solche Analysen sind aus betriebswirtschaftlicher Sicht nur dann sinnvoll, wenn dem Aufwand für die Analysen und die daraus abgeleiteten Maßnahmen auch entsprechende Erträge gegenüber stehen. Bereits in dieser ersten Phase soll deshalb eine grobe Gegenüberstellung der zu erwartenden Aufwände und Erträge durchgeführt werden. Hieraus lassen sich dann Erfolgskriterien in Form von Benchmarks sowohl für die Datenanalyse als auch für die Maßnahmen ableiten.

Um den Erfolg der geplanten Direktmarketingmaßnahmen messen zu können, war es notwendig, den monetären Wert eines verlängerten Vertrags zu bestimmen. Hierzu musste also ein geeignetes Kundenwert- bzw. Vertragswertmodell vorhanden sein. Dann konnte analysiert werden, wie stark die Stornoquote durch die Maßnahmen gesenkt werden müsste, damit die Erträge aus den zusätzlich verlängerten Verträgen die Aufwände für die Maßnahmen übersteigen. Für diese Analyse musste ferner abgeschätzt werden, wie gut die Maßnahmen dabei auf besonders stornogefährdete Segmente ausgerichtet werden können. Da die Auswahl der Segmente auf den Modellen beruhen sollte, war dabei also die Güte der Modelle relevant. Basierend auf dieser Analyse konnten geeignete (realis-

tische) Benchmarks für Datenanalyse und Maßnahmen festgelegt werden. Die Güte der mit Hilfe von Data-Mining-Verfahren entwickelten Modelle wurde dabei durch Lift-Werte beschrieben.⁶⁶

In der zweiten Phase (”Analyse der Datengrundlage”) wurde analysiert, ob die oben formulierten Ziele für die Datenanalyse mit Hilfe der vorhandenen Daten erreichbar scheinen. Dabei stellte sich beispielsweise die Frage, ob ausreichend historisierte Datenbestände vorliegen, anhand derer festgestellt werden kann, welche Verträge innerhalb eines gewissen Zeitraums in der Vergangenheit zur Verlängerung anstanden und ob diese verlängert oder durch den Versicherungsnehmer storniert wurden. Außerdem war zu prüfen, ob für jeden dieser Verträge ausreichend viele Vertragsmerkmale vorlagen, wobei diese ebenfalls möglichst historisiert sein sollten, um den Status des Vertrags zum Zeitpunkt des Ablaufs rekonstruieren zu können. Ein weiterer wichtiger Aspekt war die Qualität der Daten, also die Frage inwiefern fehlerhafte oder fehlende Werte enthalten sind. Erste explorative Analysen gaben dabei Aufschluss darüber, ob die vorhandenen Daten diesen Anforderungen genügen.

Die beiden ersten Phasen wurden mehrmals iterativ durchlaufen, um eine Übereinstimmung der Ziele, der geplanten Maßnahmen und der vorhandenen Daten sicherzustellen.

5.2 Durchführung der Data-Mining-Analysen

Zur Bestimmung der stornogefährdeten Bestandssegmente wurden so genannte Entscheidungsbäume eingesetzt, da diese eine Beschreibung und Interpretation der Bestandssegmente ermöglichen.⁶⁷ Die Auswahl der Modellierungsmethode orientierte sich also an den betriebswirtschaftlichen Zielen, und nicht umgekehrt.

Zunächst wurden alle innerhalb eines Beobachtungszeitraums von 12 Monaten ablaufenden Verträge selektiert. Die in diesem Zeitraum vom Versicherungsnehmer zum Ablauf stornierten Verträge wurden dann 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 fehler-

⁶⁶Unter dem Lift versteht man dabei das Verhältnis zwischen der Stornoquote in den durch das Modell als stornogefährdet identifizierten Segmenten und der durchschnittlichen Stornoquote in allen Segmenten zusammengenommen.

⁶⁷Auf eine detaillierte Beschreibung der eingesetzten Verfahren sowie der zugehörigen Terminologie muss hier aus Platzgründen verzichtet werden. Einen grundlegenden Überblick über diese Themen enthält *Breiman/Friedman/Olshen/Stone* (1984); außerdem möchten wir auf die weitere, einschlägige Literatur zu Entscheidungsbäumen, beispielsweise auf *Berry/Linoff* (1997), verweisen.

hafter Werte nicht zur Analyse herangezogen werden konnten.⁶⁸ Zeitabhängige Merkmale wie Geburtsdatum und Datum des Vertragsbeginns wurden dabei geeignet transformiert.

Aus den Rohdaten wurden außerdem weitere aussagekräftige Vertragsmerkmale abgeleitet: beispielsweise konnte aus der Tariffkennziffer, 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 waren detaillierte Kenntnisse der Funktionsweise sowie der Tariffkalkulation 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. Dies zeigt, wie wichtig die Einbeziehung des vorhandenen Fachwissens ist.

Für die Data-Mining-Analyse wurde im konkreten Fall aus der Vielzahl der zur Verfügung stehenden Entscheidungsbaum-Algorithmen der für die Fragestellung geeignete CART-Algorithmus (Classification and Regression Trees) ausgewählt.⁶⁹ Zur Beurteilung der Güte der damit erzeugten Modelle wurde der Datensatz in einen Trainings- und einen Validierungs-Datensatz aufgeteilt. Mit Hilfe des Trainings-Datensatzes wurde ein Entscheidungsbaum aufgebaut, der anschließend mit Hilfe des Validierungs-Datensatzes bewertet wurde. Diese Vorgehensweise stellte sicher, dass das Modell sich nicht zu gut an die Trainingsdaten anpasste (so genanntes "Overfitting") und somit auch bei Anwendung auf neue Daten zu optimalen Ergebnissen führte.

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 Know-how beim Umgang mit Data-Mining-Verfahren erfordert.

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 besteht⁷⁰ und einen Lift von etwa 3,4 aufweist, d. h. dass bei gezielter Anspra-

⁶⁸Die Analyse erlaubt damit auch Rückschlüsse auf mögliche sinnvolle Veränderungen der Datenhaltung.

⁶⁹Vgl. *Breiman/Friedman/Olshen/Stone* (1984).

⁷⁰Dies bedeutet, dass der Bestand an ablaufenden Verträgen in 28 Segmente mit unterschiedlich hoher Stornoquote aufgeteilt wird.

che von 5% der Kunden mit ablaufenden Verträgen ca. 3,4 Mal mehr stornogefährdete Verträge erreicht werden als bei zufälliger Auswahl. Falls nun mit einer Marketingmaßnahme eine gewisse Anzahl stornogefährdeter Kunden erreicht werden soll, so kann durch Anwendung des Modells die Anzahl der hierfür anzusprechenden Kunden gegenüber einer zufälligen Auswahl um ca. 70% reduziert werden.

Das durch Modifikation der Parameter sowie der verwendeten Vertragsmerkmale bestimmte "optimale" Modell erlaubte die Identifikation der wichtigsten Vertragsmerkmale für die Prognose des Stornoverhaltens. Dies waren insbesondere die bisherige Vertragslaufzeit, die Tarifgruppe, die versicherte Invaliditätssumme und das Alter des Versicherungsnehmers.

Ferner erlaubte das Modell eine Beschreibung der stornogefährdeten Bestandssegmente. Zum Beispiel ergab sich ein Segment, welches ca. 0,15% des Bestands umfasste und in dem ca. 72% der Verträge storniert wurden. Dieses Segment war 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 war das Zusammenspiel der verschiedenen Vertragsmerkmale von besonderem Interesse.

Ein Vergleich der erzielten Ergebnisse mit den in der ersten Phase definierten Data-Mining-Zielen zeigte, dass das Modell insgesamt gesehen als fachlich plausibel und mit einem Lift von 3,4 bezogen auf 5% des Bestands als geeignete Grundlage für eine zielgerichtete Marketingaktion anzusehen war.

5.3 Umsetzung der Erkenntnisse in Maßnahmen

Auf Basis der gewonnenen Erkenntnisse wurde nun eine zielgerichtete Direktmarketingmaßnahme konzipiert mit dem Ziel, stornogefährdete Verträge von einer Stornierung abzuhalten. Die hierfür in Phase 6 ("Ableitung geeigneter Maßnahmen") notwendigen Tätigkeiten bauten auf den Ergebnissen von Phase 1 auf.

Zur Selektion der anzusprechenden Kunden wurde das in der Data-Mining-Stufe entwickelte Modell auf alle innerhalb eines gewissen Zeitraums ablaufenden Verträge angewendet. Der Zeitraum war so gewählt, dass ausreichend Zeit für die Durchführung der Maßnahmen blieb. Hierbei war insbesondere die Kündigungsfrist von 3 Monate zu berücksichtigen. Abhängig von der Anzahl der so selektierten Verträge ließ sich die Güte der Selektion mit Hilfe von Lift-Werten abschätzen.

Aus betriebswirtschaftlicher Sicht waren eine ganze Reihe möglicher Maßnahmen zur Erreichung des Ziels der Senkung der Stornoquote denkbar. Diese Maßnahmen wurden nun zunächst mit den Beschreibungen der selektierten Bestandssegmente abgeglichen, wobei analysiert wurde, ob die geplante Maßnahme für das jeweilige Segment geeignet erscheint. Dies erforderte Fachwissen sowohl aus dem Fachbereich Unfall als auch aus dem Direktmarketing und zusätzlich auch ein gewisses Maß an Kreativität.

Anschließend wurde die zu erwartende Wirtschaftlichkeit der möglichen Maßnahmen analysiert und verglichen. Dabei spielte sowohl die Güte der durch das Modell gegebenen Selektion als auch die zu erwartende Wirksamkeit der Maßnahme eine wichtige Rolle. Die hier unterstellte zu erwartende Wirksamkeit der Maßnahmen stellte eine Benchmark für die mit der Durchführung der Maßnahmen betrauten Personen dar.

Unter Berücksichtigung sonstiger Rahmenbedingungen (u. a. des Gesamtbudgets der Marketingmaßnahme) wurden dann die Maßnahmen ausgewählt. Hinsichtlich der Strukturierung der Maßnahmen ist anzumerken, dass jeweils ausreichend große Kontrollgruppen gebildet werden müssen. Diese Kontrollgruppen weisen grundsätzlich die gleiche Struktur wie die mit Marketingmaßnahmen kontaktierten Segmente auf, werden jedoch nicht angesprochen. Durch Bildung ausreichend großer Kontrollgruppen konnte später analysiert werden, inwiefern die Maßnahmen zu einer (statistisch) signifikanten Reduktion der Stornoquote geführt haben. Außerdem konnte durch einen Vergleich der selektierten Segmente mit dem restlichen Bestand die Güte der verwendeten Modelle anhand realer Daten analysiert werden. Dies erlaubte ferner eine Abschätzung darüber, ob das Modell veraltet ist und deshalb auf Basis neuerer Daten aktualisiert werden sollte.

Die so geplante Direktmarketingmaßnahme wurde in der folgenden Phase ("Implementierung der Maßnahmen") operativ umgesetzt. Dabei wurde sichergestellt, dass für alle innerhalb des gewählten Zeitraums ablaufenden Verträge alle relevanten Daten erfasst werden. Dies umfasste beispielsweise alle Vertragsmerkmale zu Beginn und Ende des Zeitraums sowie Angaben darüber, ob und ggf. aus welchem Grund der Vertrag nicht verlängert wurde. Außerdem wurde festgehalten, ob der Kunde überhaupt erreicht wurde.

Hinsichtlich der operativen Umsetzung sind geeignete Kontrollmechanismen notwendig, um sicherzustellen, dass die Maßnahmen wie geplant durchgeführt werden. Dazu gehörte beispielsweise die laufende Überwachung des Stornoverhaltens in den betrachteten Segmenten, welches Hinweise auf evtl. Fehlsteuerungen liefern kann. Vor der Durchführung waren die Mitarbeiter ggf. geeignet zu schulen.

5.4 Messung und Bewertung der Ergebnisse

Die Messung und Bewertung der Ergebnisse lässt sich in drei Abschnitte untergliedern.

Zunächst wurde durch Betrachtung aller nicht kontaktierten Verträge eine Bewertung des entwickelten Prognosemodells vorgenommen. Dies geschah durch eine Gegenüberstellung des gemäß Modell zu erwartenden Lift-Werts mit dem real beobachteten Lift-Wert in den jeweiligen Segmenten und mit einer Analyse der statistischen Signifikanz der beobachteten Unterschiede. An dieser Stelle wurde also endgültig festgestellt, ob die Data-Mining-Analysen eine Fokussierung auf die tatsächlich stornogefährdeten Segmente ermöglicht haben und ob bzw. zu welchem Grad somit die in Phase 1 formulierten Data-Mining-Ziele erreicht wurden.

Aus diesen Analysen ergaben sich zusammen mit den Erkenntnissen der Data-Mining-Analysen selbst (Stufe 2 des Kreislaufs) wichtige Anhaltspunkte für eine Verbesserung der Modelle. Beispielsweise kann sich die Hinzunahme weiterer Daten, die genauere Betrachtung gewisser Vertragsmerkmale oder auch die Anwendung alternativer Data-Mining-Verfahren als sinnvoll erweisen.

Im zweiten Abschnitt erfolgte eine detaillierte Bewertung der Maßnahmen. Hierbei wurde analysiert, in welchem Segment welche Maßnahmen zu einer signifikanten Veränderung der Stornoquote geführt haben. Hierdurch ließ sich feststellen, ob die Maßnahmen insgesamt bzw. in den einzelnen Segmenten erfolgreich waren.

Eine umfassende Bewertung der Maßnahmen erfordert jedoch zusätzlich eine monetäre Kosten-Nutzen-Analyse. Die in Phase 6 verfeinerte Kosten-Nutzen-Analyse wurde deshalb für jedes Segment und jede Maßnahme aktualisiert. Diese Analyse lieferte wichtige Anhaltspunkte für Verbesserungen der Marketingmaßnahmen.

Im dritten Abschnitt wurden schließlich die im Prozessverlauf gewonnenen Erkenntnisse systematisch zusammengefasst. Als Fazit ließ sich hierbei festhalten, dass mit den durchgeführten Data-Mining-Analysen und den daraus abgeleiteten Maßnahmen der Einstieg in einen Kreislauf mit dem Ziel der mittelfristigen Bestandssicherung gelungen ist. Für zukünftige Maßnahmen ist dabei eine erhebliche Effizienzsteigerung durch eine Standardisierung der Vorgehensweise anzustreben. Außerdem soll das erkannte Verbesserungspotenzial systematisch genutzt werden.

5.5 Weitere Vorgehensweise

Um eine schrittweise Verbesserung des Prozesses der Bestandssicherung zu erreichen, wurden nach einem ersten Durchlauf des Kreislaufs folgende Maßnahmen diskutiert.

Auf Basis der Erkenntnisse kann die Vorgehensweise bei der Data-Mining-Analyse verfeinert werden. Dies betrifft sowohl die Datengrundlage als auch die angewendeten Verfahren. Um eine genauere Bewertung des Erfolgs der Maßnahmen zu ermöglichen, wird außerdem eine Verfeinerung des Modells zur monetären Bewertung derartiger Maßnahmen angestrebt. Dies umfasst insbesondere die Einbeziehung des spartenübergreifend ermittelten Kundenwerts. Auf Basis der so verbesserten Modelle können auch die Direktmarketingmaßnahmen geeignet modifiziert werden. Insbesondere soll in Zukunft genauer untersucht werden, welche Kunden auf welche Maßnahme besonders gut reagieren.

All diese Ideen können durch die Initiierung eines weiteren Kreislaufs umgesetzt werden.

6. Zusammenfassung und Ausblick

Das in den vorangegangenen Kapiteln von uns vorgestellte Modell liefert eine theoretisch fundierte und praktikable Basis für die systematische Anwendung von Data-Mining-Analysen im betrieblichen Umfeld. Gerade vor dem Hintergrund der bis dato in der Literatur analysierten Modelle sind uns folgende Aspekte besonders wichtig:

- Das Hauptaugenmerk liegt stets auf den Maßnahmen, nicht auf der Datenanalyse; Data-Mining ist lediglich Mittel zum Zweck.
- Ein weiterer Schwerpunkt ist die Messung der erzielten Ergebnisse; dadurch wird die Grundlage für eine schrittweise Verbesserung der Maßnahmen sowie der Modelle gelegt.
- Eine durchgehende monetäre Kosten-Nutzen-Betrachtung ist notwendig; hierzu müssen geeignete Modelle vorhanden sein. Die Kosten-Nutzen-Analyse erfolgt dabei an drei Stellen: vor Beginn der Analysen, vor der Ableitung der Maßnahmen und schließlich bei der Bewertung der Ergebnisse.
- Für den Erfolg sind sowohl Fachwissen und Datenwissen aus dem Anwendungsbereich als auch Data-Mining Know-how erforderlich. Nicht zuletzt ist jedoch auch ein gewisses Maß an Kreativität notwendig.

- Es handelt sich um ein Kreislaufmodell, bei dem durch einen kontinuierlichen Prozess eine schrittweise Verbesserung angestrebt wird; Data-Mining ist somit als regelmäßiger Prozess anzusehen und nicht als einmalige Aktivität.

Zur konkreten Anwendung dieses Modells in einem Versicherungsunternehmen bedarf es jeweils einer systematischen Potenzialanalyse. Dabei wird untersucht, bei welchen Geschäftsprozessen der Einsatz von Data-Mining-Analysen besonders Erfolg versprechend ist. Auf Basis einer derartigen Potenzialanalyse kann das Versicherungsunternehmen dann eine entsprechende Prioritätenliste aufstellen, für welche Prozesse derartige Analysen initiiert und evtl. anschließend regelmäßig durchgeführt werden sollen und wie diese systematisch und überprüfbar in erfolgssteigernde Maßnahmen umgesetzt werden. Hierfür liefert das vorgestellte Modell eine umfassende strukturelle Basis.

Literatur

- Adriaans, Z. / Zantinge, D.* (1996):
Data mining.
Addison-Wesley, Harlow
- Berry, M. / Linoff, G.* (1997):
Data mining techniques: for marketing, sales, and customer support.
John Wiley & Sons, New York
- Berry, M. / Linoff, G.* (2000):
Mastering data mining: the art and science of customer relationship management.
John Wiley & Sons, New York
- Beyer, D. / Orsteifen, C.* (Hrsg.) (2004):
SAS in Hochschule und Wirtschaft: Proceedings der 8. Konferenz der SAS-Anwender
in Forschung und Entwicklung (KSFE).
Shaker, Aachen
- Bigus, J.* (1996):
Data mining with neural networks: solving business problems - from application development to decision support.
McGraw-Hill, New York
- Breiman, L. / Friedman, J. / Olshen, R. / Stone* (1984):
Classification and regression trees.
Wadsworth, Belmont
- Cabena, P. / Hadjinian, P. / Stadler, R. / Verhees, J. / Zanasi, A.* (1998):
Discovering data mining: from concept to implementation.
Prentice Hall, Upper Saddle River
- Brachman, R. / Anand, T.* (1996):
The process of knowledge discovery in databases.
in: *Fayyad, U. / Piatetsky-Shapiro, G. / Smyth, P. / Uthurusamy, R.* (Hrsg.): Advances
in knowledge discovery and data mining, AAAI Press, Menlo Park, S. 37–57
- Chapman, P. / Clinton, J. / Kerber, R. / Khabaza, T. / Reinartz, T. / Shearer, C. / Wirth, R.* (2000):
CRISP-DM 1.0 - Step-by-step data mining guide.
<http://www.crisp-dm.org/CRISPWP-0800.pdf> [Stand: 16.12.2004]
- Fayyad, U. / Piatetsky-Shapiro, G. / Uthurusamy, R.* (1996):
From data mining to knowledge discovery: an overview.
in: *Fayyad, U. / Piatetsky-Shapiro, G. / Smyth, P. / Uthurusamy, R.* (Hrsg.): Advances
in knowledge discovery and data mining, AAAI Press, Menlo Park, S. 1–34
- Fayyad, U. / Piatetsky-Shapiro, G. / Smyth, P. / Uthurusamy, R.* (Hrsg.) (1996):
Advances in knowledge discovery and data mining.
AAAI Press, Menlo Park

Fayyad, U. (1996):

Data mining and knowledge discovery: making sense out of data.
in: IEEE Expert, 11.Jahrgang, Heft 5, S. 20–25

John, G. (1997):

Enhancements to the data mining process.
Dissertation, University of Stanford

Knobloch, B. (2001):

Der Data-Mining-Ansatz zur Analyse betriebswirtschaftlicher Daten.
in: Informationssystem-Architekturen, 8. Jahrgang, Heft 1, S. 59–115

Knobloch, B. (2004):

A framework for organizational data analysis and organizational data mining.
in: *Nemati, H. / Barko, C.* (Hrsg.): Organizational data mining: leveraging enterprise data resources for optimal performance, Idea Group Publishing, Hershey, S. 333–356

Küppers, B. (1999):

Data mining in der Praxis: ein Ansatz zur Nutzung der Potentiale von Data mining im betrieblichen Umfeld.
Lang, Frankfurt/M

Nemati, H. / Barko, C. (2004):

Organizational data mining: an introduction.
in: *Nemati, H. / Barko, C.* (Hrsg.): Organizational data mining: leveraging enterprise data resources for optimal performance, Idea Group Publishing, Hershey, S. 1–8

Nemati, H. / Barko, C. (Hrsg.) (2004):

Organizational data mining: leveraging enterprise data resources for optimal performance.
Idea Group Publishing, Hershey

Piatetsky-Shapiro, G. (2002):

KDnuggets: Polls: What main methodology are you using for data mining?
<http://www.kdnuggets.com/polls/2002/methodology.htm> [Stand: 16.12.2004]

Pyle, D. (1999):

Data preparation for data mining.
Morgan Kaufmann, San Francisco

Reuß, A. / Zwiesler, H.-J. (2004):

Stornoanalyse in einem Unfallversicherungsbestand.
in: *Beyer, D. / Orsteifen, C.* (Hrsg.): SAS in Hochschule und Wirtschaft: Proceedings der 8. Konferenz der SAS-Anwender in Forschung und Entwicklung (KSFE), Shaker, Aachen, S. 265–276.

Säuberlich, F. (2000):

KDD und Data Mining als Hilfsmittel zur Entscheidungsunterstützung.
Lang, Frankfurt/M

Thuraisingham, B. (1999):

Data mining: technologies, techniques, tools, and trends.
CRC Press, Boca Raton

Any correspondence concerning this working paper should be directed to

Andreas Reuß
Sektion Aktuarwissenschaften
Ulm University
D-89069 Ulm
Germany

phone: +49-(0)731-50-31251
fax: +49-(0)731-50-31252
email: a.reuss@ifa-ulm.de