

Auszug aus:

Neuronale Netze im Marketing-Management

Praxisorientierte Einführung in modernes Data-Mining.

Wiedmann, Klaus-Peter / Buckler, Frank (Hrsg.) Gabler Verlag 2001. 322 S. Mit 94
Abb. u. 31 Tab.-322 S.. - Br.

ISBN: 3-409-11673-7

Preis: 88,00 DM - 642,00 öS - 77,00 sfr

Zu beziehen unter: www.gabler-online/bookshop oder im Fachhandel

Klaus-Peter Wiedmann/ Frank Buckler/ Holger Buxel

Data Mining - ein einführender Einblick

1. Einleitung	17
2. Begriff des Data Mining	19
3. Prozess des Data Mining	22
4. Anwendungsfelder des Data Mining.....	25
5. Methoden des Data Mining	27
6. Umsetzungsaspekte des Data Mining	29
7. Fazit.....	32

1. Einleitung

Der steigende Wettbewerbsdruck sowie die wachsende Umfelddynamik, denen Unternehmen zunehmend im Rahmen des absatzmarktgerichteten Handelns ausgesetzt sind, zwingen diese dazu, ihr internes wie externes Umfeld systematisch zu analysieren und die im Rahmen der Analyse anfallenden Daten als wichtige Ressource des Unternehmens zu begreifen. Der enorme Preisverfall und die zunehmend ansteigende Leistungsfähigkeit elektronischer Datenverarbeitungs- und Speichermedien ermöglichen den Unternehmen, Daten in großer Menge zu verarbeiten. Zeitgleich führt die weiter voranschreitende informationstechnologische Vernetzung dabei zu einem Anstieg an Daten im Unternehmen, die zur Beantwortung einer Vielzahl an Frage- und Aufgabenstellungen auf operativer wie strategischer Ebene im Unternehmen genutzt werden können.

Neben Scannerkassen beim Verkauf, der EDV-gestützten Buchhaltung, dem automatisierten Fertigungsprozess oder der Protokollierung der Kundenkontakte, bei denen digitale Informationen als Ergebnis von Transaktionsprozessen i.w.S. anfallen, erlangt in diesem Zusammenhang zunehmend auch das Internet besondere Bedeutung. Im Rahmen der Abwicklung von E-Commerce-Aktivitäten können wöchentlich bis zu mehreren hundert Megabyte Daten in Logfiles und anderen Datenformen entstehen, die weitgehende Aufschlüsse über das Nutzungsverhalten der Nachfrager liefern und zur Optimierung des Sortiments, der Webseite oder der Erstellung von Customer Profiles genutzt werden können.

Vor dem Hintergrund der zunehmenden Bedeutung der Information als Wettbewerbsvorteil sowie der zu beobachtenden Technologiedynamik entwickelte sich als Reaktion auf die Problematik sehr großer Datenbestände, die mit klassischen Analysemethoden nicht mehr auswertbar sind, Ende der 80er Jahre die interdisziplinäre Forschungsrichtung „Knowledge Discovery in Databases“ (KDD), die mittlerweile überwiegend als Data Mining bezeichnet wird. Die Triebkräfte dieser integrativen Forschungsrichtung bilden die Forschungsrichtungen Statistik, Datenbanksysteme, Visualisierung, Fuzzyssetheorie, Künstliche Intelligenz mit den Teilbereichen Genetische Algorithmen, Neuronale Netze und Maschinelles Lernen, sowie Anwendungsgebiete aus der Medizin, den Natur- und Wirtschaftswissenschaften (vgl. Multhaupt, 1999, S. 8).

KDD und Data Mining sind im betriebswirtschaftlichen Kontext Synonyme.

Triebkräfte des Data Mining

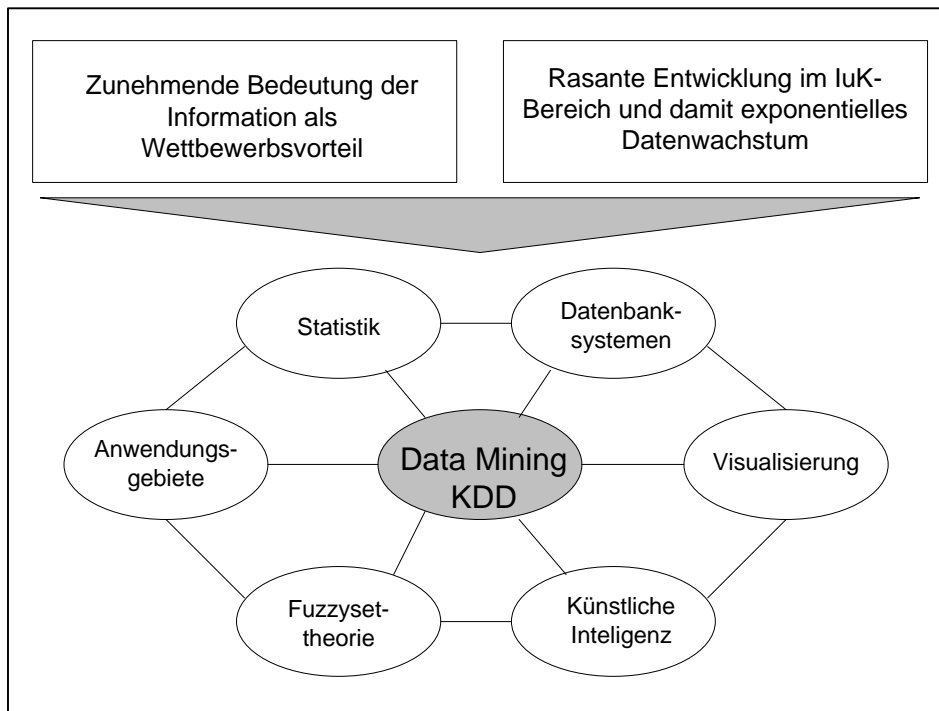


Abbildung 1: Triebkräfte des Data Mining

Die Forschungsrichtung Data Mining hat in den letzten Jahren eine Reihe von Tools hervorgebracht, die Unternehmen dabei helfen können, „in ihre riesigen Datenberge hinabzusteigen, um nach wertvollen Erkenntnissen zu graben“, die oftmals bares Geld wert sind. Dabei lohnt es sich, in all den Bereichen eines Unternehmens „einen Stollen“ zu schlagen, wo systematisch Daten gesammelt werden. Der Lohn reicht von immensen Einsparungen bei Mailings oder in der Fertigung über den Schutz vor finanziellen Risiken bis hin zu Umsatzsteigerungen durch Erkenntnisse bei Verbundkäufen. Sie können dabei als innovative Methoden zur Suche und Filterung sowie Analyse von Informationen aus großen Datenbeständen aufgefasst werden.

2. Begriff des Data Mining

Wie oft bei jungen Forschungsrichtungen, ist das Begriffsverständnis zu Beginn recht uneinheitlich. Während sich in der englischsprachigen Literatur bspw. eine deutliche Abgrenzung der Begriffe KDD und Data Mining findet, werden beide aufgrund einer inhaltlichen Deckungsgleichheit im Deutschen fast ausschließlich synonym gebraucht (vgl. Mulhaupt, 1999, S. 8). Insbesondere im betriebswirtschaftlichen Kontext und der Anwendung im Marketing hat sich mittlerweile zunehmend der Data Mining-Begriff eingebürgert. Trotz des einheitlichen Begriffsgebrauchs wird der Data Mining-Begriff selbst jedoch auf inhaltlicher Ebene z.T. unterschiedlich gebraucht. Tabelle 2 zeigt eine Vielzahl unterschiedlichster Definitionen des Data Mining.

Data Mining Definition	Autor ¹
Data Mining finds novel, valid, potentially useful and ultimately understandable patterns in mountains of data.	K. Burn-Thornton Professor, University Plymouth
Data Mining is finding new and useful knowledge in data.	Gregory Piatetsky-Shapiro Editor, KDnuggets.com
Data Mining is the application of statistical decision theory to huge, messy data sets to maximize profits.	Warren S. Sarle SAS Institute Inc.
Fill in the blanks. "If I only knew _____ then I could do _____." Data Mining is providing the answer in the first blank.	Ed Freeman Accrue Software Inc.
"KDD beschreibt automatisierte Verfahren, mit denen Regelmäßigkeiten in Mengen von Datensätzen und in eine für Nutzende verständliche Form gebracht werden." (Der zugrunde liegende Data Mining-Begriff ist lt. Author äquivalent mit "Knowledge Discovery in Databases" (KDD)).	Reginald Ferber GMD-IPSI und TU Darmstadt (Ferber, R., 1999)

¹ Autoren ohne Quelle werden aus einer persönlichen Kommunikation im März 2000 mit den Autoren in der Data Mining-Mailing-Liste auf www.nautilus-sys.com zitiert.

„Data Mining is the process of discovering meaningful new correlations, patterns and trends by „mining“ large amount of stored data using pattern recognition technologies, as well as statistical and mathematical technics“	Ashby/ Simms (1998)
“Data Mining ist the exploration and analysis, by automatic and semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules”	Berry/ Linoff (1997)
Data Mining is ... “a problem solving methodology that finds logical or mathematical descriptions, eventually of a complex nature, of patterns and regularities in aset of data”	Decker/ Focardi (1995)
“Data Mining is the process of discovering advantageous patterns in data”	John (1997)
“Data Mining finds relationships and can help anticipate the future based on past data”	Newquist (1996)
“Data Mining is a decision support process where we look in large data bases for unknown and unexpected patterns of information”	Parsaye (1996)

Tabelle 1: Data Mining-Definitionen

Auch in anderen Funktions- und Fachbereichen, wie der Mathematik und den Ingenieurwissenschaften sind Verfahren, die ebenfalls „unknown pattern“ und „meaningful relationships“ zwischen Variablen entdecken, entwickelt worden. Nur werden für denselben Vorgang Begriffe wie „Adaption“, „Parameterschätzung“ oder „Funktionsapproximation“ verwendet (siehe Tabelle 2). Besieht man sich alle Data Mining-Definitionen, so ist allen gemein, dass sie sich, inhaltlich betrachtet, auf einen Vorgang beziehen, den man umgangssprachlich auch als „Lernen“ bezeichnen könnte. Nur dass Data Mining das Lernen einer Maschine (d.h. eines Computers) und erst im zweiten Schritt das des bedienenden Menschen meint.

Begriff:	Verwendung im Fachgebiet:
Lernen	Biologie, Psychologie, Künstliche Intelligenz, Maschinelles Lernen
Funktionsapproximation	Mathematik
Parameterschätzung	Klassische Statistik
Adaption	Ingenieurwissenschaften
Data Mining	Betriebswirtschaft insb. Marketingforschung

Tabelle 2: Begriffe für Verfahren zur Wissensgewinnung

Die Begriffe in Tabelle 2 sind nicht vollständig gleichbedeutend. Dies resultiert vor allem aber darin, dass wie der Begriff des Lernens oder des Data Mining selbst, die meisten anderen Begriffe je nach Zielsetzung unterschiedlich definiert werden.

Nach unserer Auffassung ist es aufgrund der Begriffsvielfalt sinnvoll, den Begriff Data Mining für die Analyse von Daten bei einer betriebswirtschaftlichen Fragestellung zu verwenden. Einige oft genannte Eigenschaften des Data Mining wie „Verständlichkeit“, „Analyse großer Datenmengen“, „teilweise chaotische Daten“ oder „regelmäßige Anwendung“ sind u.E. nicht definitorisch relevant, da Analyseergebnisse auch ohne diese Eigenschaften betriebswirtschaftlichen Nutzen erzeugen können. Vielmehr scheint insbesondere für das Data Mining konstitutiv, dass Hypothesen über Gesetzmäßigkeiten der relevanten Welt aus Erfahrungswerten selbstständig erzeugt werden.

Data Mining ist das computergestützte Lernen aus Daten bei einer betriebswirtschaftlichen Fragestellung.

Lernen ist das Aneignen von Hypothesen über Gesetzmäßigkeiten der Welt anhand von Erfahrungen.

3. Prozess des Data Mining

In der wissenschaftlichen Literatur zum Thema Data Mining dominiert mittlerweile die Sichtweise, dass Data Mining als Prozess zu begreifen ist, der durch die Anwendung von automatischen und semiautomatischen Methoden auf einen großen Datenbestand gültige, neue, möglicherweise nützliche und verständliche Muster identifiziert.

Der Data Mining-Prozess wird typischer Weise im Detail durch fünf oder mehr Prozessschritte beschrieben. Vereinfachend dargestellt wird bei Untersuchungen prinzipiell auf folgenden Weise vorgegangen:

Als erstes muss das zu untersuchende Objekt festgelegt und anhand von Eigenschaften beschrieben werden können. So ist das Objekt bei der Standortwahl der Standort, und die Eigenschaften so etwas wie Lage, Wetter, Arbeitslosenquote, Steuersatz etc.. Bei einer Kundenanalyse ist das Objekt der Kunde selbst mit den bekannten Eigenschaften, wie Alter, Wohnort, gekaufte Produkte, Jahresumsatz u.Ä.m.

Im zweiten Schritt wird das zu untersuchende Objekt anhand seiner Eigenschaften analysiert. Dies kann mittels zweier grundsätzlicher Herangehensweisen erfolgen:

Die eine Gruppe von Methoden (Interdependenzanalysen) gruppiert die Objekte anhand ihrer Eigenschaften und erzeugt daraus wenige, für alle repräsentative, Objekte. So bilden alle Warenkörbe mit ähnlichem Inhalt genauso eine Gruppe wie die anhand von demographischen Eigenschaften gebildeten Kundensegmente.

Die zweite Gruppe von Methoden (Dependenzanalysen) versucht den Zusammenhang von Eigenschaften des Untersuchungsobjektes zu einer anderen Eigenschaft herauszufinden. Dazu gehört beispielsweise der Zusammenhang von Personenmerkmalen zu ihrer Antwort- oder Kaufwahrscheinlichkeit genauso wie der Zusammenhang zwischen vergangenen Absatzwerten zu zukünftigen.

Vor der Analyse muss die Problemstellung samt aller relevanten Größen definiert werden.

Im Kern können Dependenz- und Interdependenzanalysen durchgeführt werden

Die enge Sichtweise eines fünfstufigen Data Mining-Prozesses beschreibt indes nicht alle relevanten Problembereiche, die mit der Gestaltung eines geeigneten Data Mining-Prozesses verbunden sein können. Daher finden sich in der Literatur auch detailliertere Prozess-Schemata (siehe Tabelle 3), die sich an den Stufen Problemdefinition, Auswahl der Datenbasis, Datenaufbereitung, Datenreduktion, Festlegung der Methode, Data

Mining i.e.S., Interpretation der Muster und Ergebnispräsentation und -anwendung orientieren. (vgl. Poddig/Huber, 1999, S. 665f.).

Stufe	Kennzeichen
1 Problemdefinitor	<p>Neben der Definition eines geeigneten Ziels muss geprüft werden, ob die formulierten Ziele mit Hilfe der zur Verfügung stehenden Daten in einem gegebenen zeitlichen und finanziellen Rahmen erreicht werden können. So gibt es in der Praxis grundsätzlich zwei Arten von Zielstellungen:</p> <p>Zum einen kann man vor allem daran interessiert sein, verständliche Erklärungen für Zusammenhänge zu erhalten.</p> <p>Auf der anderen Seite kann das vorrangige Ziel darin bestehen, Voraussagen treffen zu können.</p> <p>Im Idealfall können beide Zielstellungen zugleich erfüllt werden.</p>
2 Auswahl der Datenbasis	Auswahl einer Datenbasis, mit der das Ziel am besten erreicht werden kann.
3 Datenaufbereitung	Häufig liegen die Daten in einer Form vor, die unvollständig ist oder Ausreißer enthält. Die Daten sind mit einer geeigneten Vorgehensweise bei derartigen Problemen zu behandeln.
4 Datenreduktion	Mit Hilfe geeigneter Verfahren der Datenreduktion können hochdimensionale Datenbasen auf niedrigdimensionalere ohne (oder nur mit einem geringen) Informationsverlust abgebildet werden.

5 Festlegung der Methoden	<p>Insbesondere die Wahl der Auswertungsmethode hängt dabei stark von dem Untersuchungsziel ab, das mit dem Data Mining verfolgt wird. Bei der Analyse des zu untersuchenden Objektes, anhand seiner Eigenschaften können zwei grundsätzliche Herangehensweisen zum Einsatz kommen: Die eine Gruppe von Methoden (Interdependenzanalysen) gruppiert die Objekte anhand ihrer Eigenschaften und erzeugt daraus wenige, für alle repräsentative, Objekte.</p> <p>Die zweite Gruppe von Methoden (Dependenzanalysen) versucht den Zusammenhang von Eigenschaften des Untersuchungsobjektes zu einer anderen Eigenschaft herauszufinden. Dazu gehört beispielsweise der Zusammenhang von Personenmerkmalen zu ihrer Antwort- oder Kaufwahrscheinlichkeit genauso wie der Zusammenhang zwischen vergangenen Absatzwerten zu zukünftigen.</p>
6 Data Mining i.e.S.	Nach der Auswahl einer geeigneten Methode muss diese angewandt werden, um die gesuchten Muster in den Daten zu identifizieren.
7 Interpretation der Muster	Die Ergebnisse des Data Mining-Prozesses müssen mit den gesetzten Zielvorstellungen verglichen und vor diesem Hintergrund interpretiert werden. Aus den resultierenden Daten können sich neue Fragestellungen und Ziele ergeben, die durch die Wiederholung aller oder einzelner Schritte von 1 bis 6 analysiert werden müssen.
8 Ergebnispräsentation & -anwendung	Die Ergebnisse werden im abschließenden Schritt meist visuell aufbereitet, um eine einfache Interpretation zu gewährleisten und somit die zentralen Aspekte auch den nicht an der Analyse beteiligten Unternehmensmitgliedern zu verdeutlichen.

Tabelle 3: Prozess des Data Mining

4. Anwendungsfelder des Data Mining

Data Mining kann in allen Bereichen Anwendung finden, in denen Daten systematisch erhoben werden. Die Herangehensweise dabei basiert auf einigen grundlegenden Prinzipien, die - einmal verstanden - auf alle Anwendungsfelder übertragen werden können.

Data Mining kann unabhängig von der Bedeutung der Daten prinzipiell in allen Problemfeldern gleichermaßen Anwendung finden

Ein beliebtes Anwendungsfeld des Data Mining ist beispielsweise auf der Basis von Kundendatenbanken die Gründe dafür herauszufinden, weshalb durch Briefe angeschriebene Personen antworten und andere nicht. Kann die sogenannte „Responsewahrscheinlichkeit“ einer Person gut geschätzt werden, so kann man dazu übergehen, nur noch Personen mit ausreichender Aussicht auf Erfolg anzuschreiben. So spart beispielsweise die Firma Microsoft einige Millionen Dollar Portokosten pro Jahr, indem sie mit Hilfe von Neuronalen Netzen die Antwortquote von zwei auf acht Prozent steigerte.

Ein anderes neuartiges Anwendungsfeld tritt bei der Bestückung von Web-Seiten mit Werbebannern auf. Jeder Internetsurfer überträgt auch Informationen über seine Herkunft, seine Surfhistorie und vieles mehr an die angesteuerte Web-Seite. Insofern man die Ursachen herausfinden kann, die dazu führen, dass ein Werbebanner angeklickt wird, können jedem Surfer genau die Banner präsentiert werden, die ihn interessieren. Einer Website mit 5 Millionen Seitenaufrufen pro Monat in den USA ist es beispielsweise gelungen, durch zielgerichtete Bannerzuordnung die Klickraten von einem auf zwei Prozent zu erhöhen. Damit erhöhte sich auch der pro Banner erzielte Erlös. Effektiv bedeutete dies eine Umsatzsteigerung von 1,2 Millionen Dollar.

Im Einzelhandel sind beispielsweise Warenkorbanalysen Aufgabe des Data Mining. Hier können moderne Methoden herausfinden, welche Produkte häufig im Verbund gekauft werden. Diese Erkenntnisse können dazu genutzt werden, um im Verbund gekaufte Produkte nebeneinander im Regal zu positionieren und so die Abverkaufswahrscheinlichkeit zu erhöhen. Sind zudem über ein Kundenkarten-System noch die Warenkörbe einer bestimmten Person zuzuordnen, so entstehen weitere Möglichkeiten. Zum einen können Eigenschaften von Kunden herausgefunden werden, die in gewissem Sinne besonders lukrativ sind. Diese Kundensegmente können dann dementsprechend mit Kundenbindungsmaßnahmen wie Geburtstagsgeschenken bearbeitet werden.

Abverkaufsprognosen werden beispielsweise im Einzelhandel standardmäßig zum Zwecke einer optimalen Warenbestellung durchgeführt. Durch den Einsatz moderner Data Mining-Methoden kann versucht werden, Verlaufsmuster über die Umsatzentwicklungen der Vergangenheit zu bestimmen, welche auf eine positive oder negative Umsatz-

entwicklung in der Zukunft hindeuten. In einer Studie der Universität Osnabrück, die in einem Supermarkt durchgeführt wurde, konnte die Fehlerquote bei der Einschätzung zukünftiger Abverkäufe im Vergleich zum herkömmlich eingesetzten Verfahren der „Gleitenden Durchschnitte“ bis zu 50% gesenkt werden. (Thiesing, F. M./ Vornberger, T., 1998, S.115)

Die genannten, vor allem aus dem Absatz stammenden Beispiele, sollen nicht darüber hinwegtäuschen, dass die Einsatzgebiete des Data Mining letztlich kaum begrenzt sind. Einige weitere Beispiele zeigt Tabelle 4.

Funktionsbereiche	Anwendungsgebiet des Data Mining
Marketing	Absatz- und Marktanteilsprognosen
	Segmentierung von Kunden in Verhaltensklassen
	Cross Selling: gut verkaufbare Produktkombinationen
	Gewinnmaximales Mailing (Industrie oder Versandhandel)
	Kostenminimale Routenplanung im Außendienst
	Missbrauchserkennung von Versicherungsnehmern
	WWW: Click-maximierendes Zuordnen von Werbebannern
Finanzierung	Devisenkursprognosen zum Hedging
	Lieferanten-Qualifikation
	Bonitätsprognosen von Kunden
Controlling	Target Costing
	Früherkennung von Unternehmenskrisen
Personal	Personalselektion bzw. Matchinganalysen
	Prognose von Fehlleistungen
Strategisches Management	Simulation alternativer Zukünfte, insbesondere i.V.m. Preis-Absatz-Funktionen
	Optimale Standortwahl
Fertigung	Fehlererkennung in der Fertigung (Qualitätssicherung)
	Optimale Reihenfolgeplanung in der Produktion

Tabelle 4: Anwendungsbeispiele des Data Mining geordnet nach Funktionsbereichen

5. Methoden des Data Mining

Der Werkzeugkasten einer Data Mining-Software besteht aus Methoden verschiedener Wissenschaftsbereiche. Dies sind zum einen die Methoden der klassischen Statistik. Weiterhin sind Methoden des Maschinellen Lernens und der Künstlichen Intelligenz, Neuronale Netze und Entscheidungsbäume sehr verbreitet und erfolgreich. Eine Übersicht über einige wichtige Data Mining-Methoden vor dem Hintergrund ihrer Zuordnung zur Interdependenz- bzw. Dependenzanalyse gibt Tabelle 5:

Fachgebiet	Interdependenzanalyse	Dependenzanalyse
Künstliche Intelligenz, Maschinelles Lernen	Association Discovery, Sequenzanalyse ...	Decision-Trees/ Regression-Trees ...
Neuronale Netze	Neuro-Clustering (SOM, ART, etc.)	Feed-Forward-Netze (MLP, RBF, etc.)
Klassische multivariante Statistik	Clusteranalyse, Faktoranalyse, etc.	Regressionsanalyse, Diskriminanzanalyse, K-Nächste-Nachbarn, etc.

Tabelle 5: Übersicht wichtiger Data Mining-Methoden

Alle der aufgeführten Verfahren unterscheiden sich indes hinsichtlich ihres Anwendungsspektrums:

Methode	Kennzeichen
Assoziationsregeln	Verfahrensklasse, mit der sich Verbundbeziehungen zwischen gemeinsam auftretenden Objekten ermitteln lassen. Liefern Vorhersagen über das Auftreten eines Objektes in Abhängigkeit anderer Objekte.
Entscheidungsbaumverfahren	Verfahrensklasse zur Klassifikation, die die zu klassifizierenden Objekte unter Berücksichtigung ihrer Merkmalsausprägungen sukzessiv in Klassen aufteilt, unter Rückgriff auf die Annahme, dass für jeden Klassenwert bestimmte Attribute oder Attributkombinationen verantwortlich sind.

Genetische Algorithmen	Universelles Optimierungsverfahren, das nach dem Prinzip der „Natürlichen Auslese“ aus einer Menge an Lösungen diejenige bestimmt, welche eine möglichst gute i.S. einer „erlesenen“ Lösung darstellt. Eignet sich auch für nicht-metrische Optimierungsprozesse.
Clusteranalyse	Ziel der Clusteranalyse ist die Zerlegung von Mengen an Objekten bei gleichzeitiger Betrachtung aller relevanten Merkmale so in Teilmengen, dass die Ähnlichkeit zwischen den Objekten eines Clusters möglichst groß, die zwischen den Gruppen jedoch möglichst gering ist.
Faktoranalyse	Die Faktorenanalyse untersucht multivariate Dateien, bei denen die Annahme sinnvoll ist, dass die beobachteten Variablen von Einflussgrößen abhängig sind, die selbst nicht unmittelbar gemessen werden können oder wurden und als Faktoren bezeichnet werden.
Regressionsanalyse	Die Regressionsanalyse dient der Analyse von linearen Beziehungen zwischen einer metrisch abhängigen und einer oder mehreren unabhängigen metrischen Variablen.
Diskriminanzanalyse	Mit Hilfe der Diskriminanzanalyse kann untersucht werden, ob und wie gut sich Gruppen von Elementen, die durch eine Reihe an Variablen beschrieben werden, voneinander unterscheiden.
K-Nächste-Nachbarn	Klassifikationsverfahren (gleiche Anwendung wie Diskriminanzanalyse), welche die Klasse eines Objektes gemäß der unter K ähnlichen Objekten am häufigsten vorkommende Klasse bestimmt.
Neuronale Netze	Universelle Approximatoren, „Baukastensystem“, aus einfachen Rechenoperatoren, durch deren Kombination logische Zusammenhänge und Regeln darstellbar sind.

Tab.6: Wichtige Verfahren des Data Mining

Bedeutung Neuronaler Netze

Neuronale Netze stellen eine Verfahrensklasse dar, die fast alle Aufgabengebiete von der Dependenz- bis zur Interdependenzanalyse abdeckt. Dabei besteht ihr großer Vorteil darin, dass sie als Lernverfahren universell eingesetzt werden kann. Neuronale Netze können ebenso einfache lineare wie auch hoch komplexe nichtlineare Zusammenhänge aufdecken.

Grenzen der Einsatzmöglichkeiten Neuronaler Netze bestehen in Einzelfällen da, wo es darum geht, Zusammenhänge verständlich darzustellen. Als Konsequenz daraus folgt jedoch nicht der Verzicht auf Neuronale Netze, sondern meist eine sinnvolle Ergänzung

mit Methoden zur verständlichen Darstellung. Somit ist es möglich, die hohe Erklärungsgüte Neuronaler Netze mit der Verständlichkeit der Ergebnisse anderer Methoden zu kombinieren (mehr dazu im nächsten Beitrag dieses Bandes)

Neuronale Netze zählen zu den bedeutendsten Data Mining-Methoden. Daher scheint es lohnenswert, sich im Rahmen dieses Buches auf sie und ihre Anwendungsmöglichkeiten zu konzentrieren.

6. Umsetzungsaspekte des Data Mining

Inwieweit die dem Data Mining innewohnenden Chancenpotenziale umgesetzt werden können, hängt neben der methodisch-konzeptionellen Vorgehensweise auch maßgeblich davon ab, inwieweit die für unternehmerische Entscheidungen notwendigen Datenbestände im Unternehmen in einer brauchbaren Form vorgehalten werden und inwieweit eine organisationale Implementierung des Data Mining gelingt.

Hinsichtlich der Datenbasis kann konstatiert werden, dass sich die im Unternehmen vorhandenen Daten in einem weit höheren Maße als strategische Waffe nutzen lassen, wenn es gelingt, die marktgerichteten Aktivitäten des Unternehmens bereichsübergreifend und damit ganzheitlich zu betrachten, zu analysieren und zu optimieren, indem die an unterschiedlichen Stellen des Unternehmens anfallenden Daten in einem entscheidungsorientierten Datenmodell zusammengeführt werden (vgl. Wiedmann/ Jung, 1998, S.324). Der vermehrt zu beobachtende Aufbau von Data Warehouses, bei der eine Vernetzung mit anderen abteilungsspezifischen Datenbeständen im Rahmen eines integrierten Ansatzes zur Bildung eines ganzheitlichen Managementunterstützungssystems stattfindet, kann wohl vor allem darauf zurückgeführt werden, dass sich vermehrt die Erkenntnis durchsetzt, dass

- durch die Integration verteilter Datenmassen in einem Gesamtsystem ein Mehrwert aus den vorhandenen Informationen durch die Nutzarmachung von Verbundbeziehungen zwischen den verteilten Datenbeständen erzielt werden kann, da
- die meisten unternehmerischen Entscheidungen in der Regel einen funktionsbereichsübergreifenden Charakter aufweisen.

Gefordert sind heute daher vielmehr integrierte Ansätze, die das enge abteilungsspezifische „Kästchendenken“ überwinden und das Wissen aus unterschiedlichen Bereichen in Form von Daten zusammenführen und vertiefen.

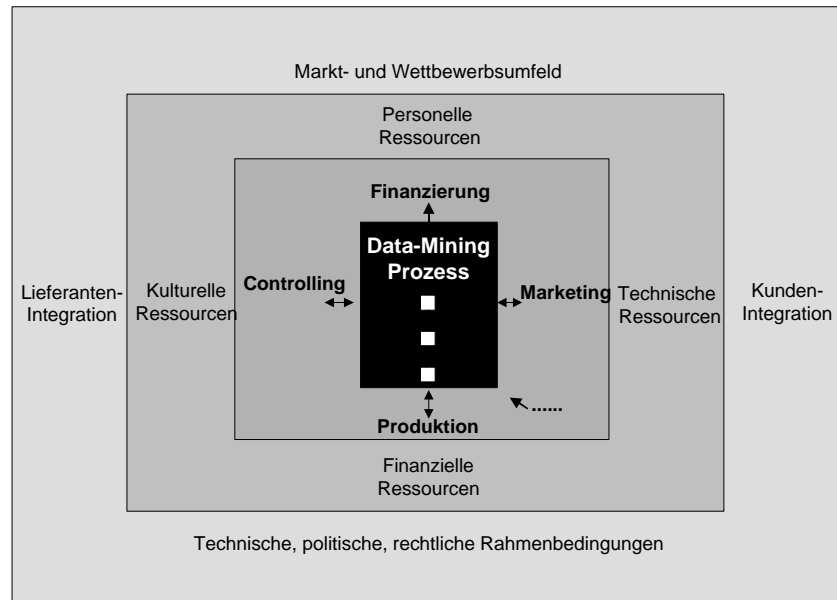


Abbildung 2: Bezugsebenen bei der Umsetzung des Data Mining, Quelle: Laslopp, 2000

Neben Fragestellungen der Datenorganisation gilt es, im Rahmen der organisationalen Implementierung, die Entscheidung zur Einrichtung eines Data Mining-Systems im Gesamtkontext einer Unternehmung situationsspezifisch zu bewerten (Lasslop, 2000):

Die situativen Rahmenbedingungen beeinflussen maßgeblich die Potenziale, die mit der Einführung des Data Mining verbunden sind. So beeinflusst der Grad der informatischen Kunden- und Lieferantenintegration (z. B. über Internet, EDI) erheblich die Menge und Güte der zur Verfügung stehenden Daten. Dies gilt ebenso für das Umfeld der Unternehmung: Je turbulenter und instabiler die technologische und die Konsumententwicklung am Markt sind, desto höher ist tendenziell der Wert von Informationssystemen, die eine Unternehmung in die Lage versetzen, flexibel auf Veränderungen zu reagieren. Die Umsetzung des Data Mining kann in diesem Fall sowohl zur direkten Generierung von Kundennutzen als auch im Sinne von Frühwarnsystemen eingesetzt werden.

Des weiteren sind die Art und der Umfang der Ressourcenbelastung im Vorfeld einer Implementierung von Data Mining möglichst genau abzuschätzen. Neben finanziellen Ressourcen werden zur zielgerichteten Implementierung des Data Mining vor allem auch technische, personelle und kulturelle Ressourcen benötigt, um eine weitreichende Integration der Datenbestände und Ankopplung an die Funktionsbereiche wie Marketing, Produktion und Controlling im Sinne eines weitreichenden Data Warehouse-Ansatzes sicherstellen zu können.

Die Gestaltung von Data Mining-Prozessen umfasst in der Regel die Integration verschiedener Abteilungen und Aufgabenträger. Zur erfolgreichen Umsetzung des Data Mining im Rahmen von Marketinganwendungen sind daher letztlich auf personeller Ebene funktionsübergreifende Teams zu bilden, bei der die personelle Besetzung bestimmter Rollen entscheidend für den Erfolg von Data Mining-Projekten ist (vgl. Berson/Smith/Thearling, 2000):

- Die Person des *Data Miners* sollte die Kompetenz hinsichtlich aktueller Data Mining-Verfahren mit einem Verständnis für die damit zu bearbeitende Marketing-Fragestellung bündeln.
- Der *Data Designer* ist verantwortlich für den Entwurf eines Datenmodells, welches das Data Mining optimal unterstützt. Dazu bedarf es neben fundierten Kenntnissen über fortschrittliche Datenbankarchitekturen auch der ständigen Überwachung der eingeführten Systeme.
- Der *Projekt-Manager* koordiniert die Aktivitäten der einzelnen Teammitglieder und plant und überprüft den Projektfortgang. Zudem ist er allgemeiner Ansprechpartner für die beteiligten Fachabteilungen.
- Ein Mitglied sollte die speziellen Anforderungen der Marketing-Abteilungen innerhalb des Projektteams vertreten (*Marketing-Schnittstelle*).
- Ein weiteres Mitglied sollte die Verantwortung für die Datenbeschaffung aus den unternehmensinternen Datenbanken sowie die Integration der darin enthaltenen Daten (Vollständigkeit, Richtigkeit, Kompatibilität, Redundanzen, ...) tragen (*Datenbank-Schnittstelle*).
- Schließlich sollte eine Person die Entwicklung der *Anwender-Schnittstelle* koordinieren. Sie ist gemäß den Kenntnissen und Anforderungen der unternehmensinternen Systembenutzer auszugestalten, was eine enge Zusammenarbeit mit diesen bereits in einem frühen Planungsstadium nahe legt.

6. Fazit

Data Mining-Tools sind Hilfsmittel, die durch Nachbilden bestehender Zusammenhänge oder durch geeignetes Zusammenfassen von Informationen selbstständig aus Datenbeständen Erklärungen und Prognosen erzeugen können.

Wie dargestellt, kann die automatische Gewinnung von Wissen, wirkliche „Schätze“ bergen. Ob und inwieweit sich eine Expedition in das eigene „Datenbergwerk“ lohnt, wird man immer erst im Nachhinein wissen.

Unternehmer, die sich dem sicherlich lohnenden und auch zunehmend notwendigen Abenteuer „Data Mining“ stellen wollen, sollten drei wichtige Punkte beachten:

- *Datenbasis aufbauen*: Die Grundlage des Data Mining ist eine umfassende Datenbasis. Daher sollte nach Möglichkeit ein unternehmensübergreifendes Data Warehouse eingerichtet werden.
- *Geeignete Data Mining-Software beschaffen*: Es sind derzeit eine Reihe von Softwarepaketen erhältlich, die jedoch alle Vor- und Nachteile besitzen. Daher sollte idealerweise ein Fachmann die Software auf die spezifischen Anforderungen hin auswählen.
- *Data Mining-Know-How erwerben*: Ein erfolgreicher Data Mining-Prozess, inklusive der sinnvollen Interpretation der Ergebnisse, erfordert eine gewisse Expertise und Erfahrung. Die beste (jedoch kurzfristig kostspieligere Variante) ist die Beauftragung eines Data Mining-Beratungsteams, welches von der Implementierung bis zur Projektdurchführung den gesamten Miningprozess professionell zu leiten imstande ist.

Voraussetzung für all dies ist, dass bei betrieblichen Entscheidern ein Grundverständnis des Data Mining besteht. Denn nur damit ist es möglich, die immensen Potentiale, die notwendigen Randbedingungen, sowie die Limitationen des Data Mining zu erkennen.

Ein Grundverständnis des Data Mining ist demnach eine Voraussetzung, um ein „Pro vs. Kontra“ bezüglich der Implementierung eines Data Mining-Systems entscheiden zu können. Ein solches bildhaftes Verständnis zu erzeugen, ist Ziel dieses Buches.

Literatur

- ASHBY, C./SIMMS, S.: Data Mining – Research Brief. URL <http://theweb.badm.sc.edu/798gstud/simsj/Datamininggrb.html>, 1998
- BERRY, M.J.A./LINOFF, G.: Data Mining Techniques for Marketing, Sales and Customer Support, New York, 1997
- BERSON, A./SMITH, S./THEARLING, K.: Building Data Mining Applications for CRM, New York u.a. 2000
- DECKER, K.M./FOCARDI, S.: Technology Overview: A Report on Data Mining, CSCS-ETH, Swiss Scientific Computer Center, 1995
- JOHN, G.: Enhancements to the Data Mining Process, Doctorial Dissertation, Department of Computer Science, University Stanford, 1997
- LASSLOP, I.: Data Mining als Grundlage des Relationship-Marketing – Einsatzfelder und Aussagewert, Arbeitspapier Nr. 142 der Wissenschaftlichen Gesellschaft für Marketing und Unternehmensführung, Münster 2000
- MULTHAUPT, M.: Data Mining: Grundlagen und Methoden, Arbeitspapier des Instituts für Wirtschaftswissenschaften der Technischen Universität Braunschweig, Braunschweig 1999
- NEWQUIST, H.P.: Data Mining: The AI Methamorphosis. Database Programming & Design, www.dbpd.com/newquist.htm, 1996
- PARSAY, K.: New Realms of Analysis. Database Programming & Design, 1996
- PODDING, T./HUBER, C.: Data Mining und Knowledge Discovery in Databases, in: WiSt, Heft 12/1999, S. 663-666
- SÄUBERLICH, F.: KDD und Data Mining als Hilfsmittel zur Entscheidungsunterstützung, Frankfurt a.M. u.a. 2000
- THIESING, F.M./VORNBERGER, O.: Abverkaufsprognose im Supermarkt mit Neuronalen Netzen, in: BIETHAHN, J./HÖNERLOH, A. /KUHL, J./LEISEWITZ, M.C./NISSEN, V./TIETZE, M. (Hrsg.): Betriebswirtschaftliche Anwendungen des Soft Computing, Braunschweig/Wiesbaden, 1998
- WIEDMANN, K.P./JUNG, H.H.: CORIM[®]: Ein neuer Ansatz im Feld des integrierten Informationsmanagements, in: HIPPER, H./MEYER, M./WILDE, K.D. (Hrsg.): Computer Based Marketing, Braunschweig u.a., 1998, S. 317-326