

Analytisches CRM

Seminararbeit

nach einem Thema von Prof. Dr.-Ing. Gerhard Weikum
Lehrstuhl für Datenbanken und Informationssysteme
Fachrichtung 6.2 – Informatik
Universität des Saarlandes, Saarbrücken, 2003



Christian Ludt
Matr.-Nr. 2018979

Michael Bernward Schmidt
Matr.-Nr. 2029855

Vorwort

Das Thema ‘Analytisches CRM’ war Bestandteil des Seminars ‘CRM und SRM: Customer und Supplier Relationship Management’ im WS 2002/2003, das in Zusammenarbeit mit der SAP Retail Solutions angeboten wurde.

Wir danken Ilonka Schmidt für das Ausleihen zahlreicher Bücher und das Zusenden Ihrer Diplomarbeit. Weiter danken wir unserer Ansprechpartnerin bei SAP Elisabeth Thieser, die bei Fragen stets schnell und kompetent geantwortet hat.

Saarbrücken, Februar 2003

Inhaltsverzeichnis

Vorwort	i
1 Einführung	1
2 OLAP	5
2.1 Definition	5
2.2 FASMI-Anforderungen	6
2.3 Operationen des OLAP	7
2.4 RFM Analyse	9
3 Data Mining	13
3.1 Data Mining als Teilaspekt von analytischem CRM	13
3.1.1 Definition DataMining	13
3.2 Data Mining Methoden	14
3.2.1 Assoziationsanalyse	15
3.2.2 Entscheidungsbäume	18
3.2.3 Clustering	21
4 Kundenanalyse	23
4.1 Kundenverhaltensanalyse	23
4.1.1 Kaufverhaltensanalyse	23
4.1.2 Abwanderungsverhaltensanalyse	24
4.1.3 Zufriedenheitsanalyse	24
4.2 Kundenwertanalyse	24

5	Customer-Lifetime-Value	27
5.1	Kundenwert	27
5.2	Kundenlebenszyklus	27
5.3	Berechnung CLTV	28
5.3.1	Beispielrechnung CLTV	29
6	Fazit	33
A	Abkürzungsverzeichnis	35
B	Beispielrechnung Informationsgewinnverhältnis	37
	Abbildungsverzeichnis	43
	Tabellenverzeichnis	43
	Literaturverzeichnis	45

Kapitel 1

Einführung

Die Leistungsfähigkeit eines Customer Relationship Management (CRM) Systems wird nicht zuletzt durch die Leistungsfähigkeit der Analysemethoden bestimmt. Grundsätzlich geht es dabei darum, auf vorhandenen (historischen) Daten verschiedene Analysen durchzuführen. Diese Analysefunktionen werden unter dem Begriff analytisches CRM (aCRM) zusammengefasst. Die Größe der Ausprägung dieses Systems innerhalb des CRM Systems ist von Anbieter zu Anbieter unterschiedlich. Dennoch gibt es durchaus Gemeinsamkeiten. Eine dieser Gemeinsamkeiten ist die Notwendigkeit einer großen und umfangreichen Datenbasis. Dies wird in der Regel durch die Verwendung eines Data Warehouse (DW) sichergestellt. Mit diesen Daten kann dann das aCRM arbeiten. Aus diesem Grund ist das aCRM sowohl logisch als auch physisch an das Data Warehouse gekoppelt, d.h. die Funktionen des aCRM sind im Data Warehouse implementiert. Häufig wird ein so ausgestattetes Data Warehouse auch als Business Warehouse (BW) bezeichnet. Die Ergebnisse der Analysefunktionen können dann von den verschiedenen CRM Modulen (z. B. Marketing, Sales, Service usw.) verwendet werden.

Die wesentlichen Ziele eines analytischen CRMs stellen sich wie folgt dar:

- Gewinnung neuer betriebswirtschaftlich relevanter Informationen
- Unterstützung im operativen und strategischen Bereich
- Analysieren vorhandener Kundenbeziehungen
- Analysieren und prognostizieren von Kundenverhalten

- Stabile Kundenbeziehungen

Beim ersten Punkt sind vor allem die Aspekte *neu* und *relevant* von Bedeutung. Eine Analyse an deren Ende ein triviales Ergebnis steht ist sinnlos. Als Beispiel sei hier die Warenkorbanalyse eines Elektronikmarktes genannt. Ist das Ergebnis einer solchen Analyse, daß Kunden, die einen Drucker gekauft haben auch ein Druckerkabel kaufen, keine neue Information, da der Zusammenhang zwischen Drucker und Druckerkabel offensichtlich ist. Dem entgegen ist ein Zusammenhang zwischen den Verkäufen von CD-ROM Laufwerken und Scanner nicht so offensichtlich (ein solcher Zusammenhang sei an dieser Stelle angenommen). Diese Information kann durchaus von betriebswirtschaftlicher Bedeutung sein. So könnte ein Verkäufer einem Kunden, der ein CD-ROM Laufwerk gekauft hat auch einen Scanner empfehlen. Eine solche Analyse kann einem Unternehmen also bei der Erledigung seiner operativen Aufgaben helfen. Aber auch bei der Lösung strategischer Probleme kann eine Analyse aus dem Bereich aCRM hilfreich sein. Eine Analyse des Kaufverhaltens der Kunden kann z. B. nützlich sein, wenn es darum geht, neue Märkte zu erschließen. Somit kann die Kundenanalyse zur Lösung eines strategischen Problems eingesetzt werden. Generell lässt sich sagen, daß all diese Verfahren dazu dienen, die vorhandenen Kundenbeziehungen zu stabilisieren. Dies ist von entscheidender Bedeutung. Eine Studie von Reichheld/Sasser [8] ergab, daß bei einer Vermeidung der Kundenabwanderungsrate um 5% eine Gewinnsteigerung zwischen 25% und 85% erzielt werden kann.



Abbildung 1.1: Closed-Loop Szenario[1]

Wie die Abbildung 1.1 zeigt, werden die Ergebnisse einer Analyse dazu verwendet, neue Aktionen zu planen. Nachdem diese Aktionen durchgeführt wurden,

gilt es, das Kundenfeedback aufzuzeichnen und in dem Datenbestand abzubilden. Nach diesem Schritt stehen die neuen Informationen für weitere Analysen bereit, welche dann wiederum zur Planung neuer Aktionen verwendet werden können. Es entsteht somit ein Closed-Loop Szenario.

Wie eingangs bereits erwähnt, wird für die erfolgreiche Verwendung der aCRM Methoden eine umfangreiche Datenbasis benötigt. Aus diesem Grund muss auch im Bereich des aCRM großen Wert auf die Realisierung eines Closed-Loop Szenarios gelegt werden.

Das aCRM kommt in allen Teilbereichen des CRMs zum Einsatz. Im Marketing ist es vor allem die RFM Analyse, die heute für die Planung von Kampagnen eingesetzt wird. Diese RFM Methode zählt nicht zu den präzisesten Analyseverfahren, zeigt sich aber im Bereich des Marketing als ausreichend. Im Vertrieb (engl.: Sales) werden aCRM Analysen eingesetzt, um die Kennzahlen (Umsatz, Gewinn, usw.) eines Kunden bestimmen zu können. Diese können sehr hilfreich beim täglichen Umgang mit dem Kunden sein. So können z. B. individuelle Angebote für einen Kunden erstellt werden, mit deren Hilfe man die Umsatzzahlen eines Kunden steigern möchte. Ein solches individuelles Angebot wird natürlich auf Basis einer Kundenanalyse erstellt. Für das Eingangsbeispiel könnte das bedeuten, daß einem Kunden, der eine CD-ROM ordert auch gleich ein Angebot für einen Scanner unterbreitet wird. Im Bereich Service können Analysen über die Kundenzufriedenheit und der Produktqualität durchgeführt werden. Die Ergebnisse können dann zur Verbesserung der Kundenzufriedenheit und zur Steigerung der Produktqualität genutzt werden. Im Bereich des aCRM ist natürlich die Analyse des Kunden der Haupteinsatzbereich. Die hier gewonnen Erkenntnisse können in den anderen Bereichen, wie oben erwähnt, eingesetzt werden. Bei der Analyse des Kunden geht es vor allem darum, sein Kaufverhalten zu analysieren. Mit den so gewonnen Ergebnissen kann man dann auch bedingt Aussagen über das zukünftige Kaufverhalten machen. Als letztes Teilgebiet sei hier die Analyse der Kundenkanäle (Internet, telefon, Brief, persönliche Kontakte, usw.) genannt. Diese ermöglicht es beispielsweise festzustellen, welche Waren über welchen Kanal besonders häufig verkauft wurden. Auf Grund dieser Ergebnisse können dann neue Waren nur an bestimmte Kunden oder über bestimmte Kanäle angeboten werden.

Der Einsatzbereich des aCRM ist also sehr vielseitig, wobei der Schwerpunkt natürlich eindeutig auf der Analyse des Kunden liegt.

Kapitel 2

OLAP

2.1 Definition

Der On-Line Analytical Process (OLAP) ist ein wesentlicher Teil des aCRM. Seine Hauptaufgabe ist es, Informationen aus dem Data Warehouse in verständlicher Form darzustellen. Der OLAP definiert sich dabei wie folgt (vgl.[26]):

Online Analytical Processing (OLAP) is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user

Diese Definition beschreibt sehr genau, welches die markanten Eigenschaften des OLAP sind. Zunächst wird definiert, daß es sich um eine Software handelt, die von Analysten, Managern und Mitarbeitern genutzt werden soll. Dies bedeutet wiederum, daß die Bedienung und Verwendung des OLAP so transparent gestaltet werden muß, daß die oben erwähnte Zielgruppe dieses problemlos einsetzen können. Weiterhin fordert die obige Definition, daß ein OLAP schnell, konsistent und interaktiv sein soll. Hierbei soll es möglich sein, verschiedene Sichten auf die vorhandenen Daten zu erhalten. Dies ist recht verständlich, da die verschiedenen Personenkreise der Zielgruppe unterschiedliche Sichten der Daten benötigen (Manager vs. Marketing Mitarbeiter). Diese Sichten werden aus Daten gewonnen, die

über verschiedene Kanäle in das DW gelangt sind, und dort von rohen Daten zu konsistenten Informationen verarbeitet wurden. Schließlich fordert die obige Definition, daß die Daten in ihrer realen Dimensionalität dargestellt werden können. Dies ist das entscheidende Merkmal eines OLAP.

2.2 FASMI-Anforderungen

Die obige Definition beschreibt sehr gut die wesentlichen Eigenschaften des OLAP. Leider ist dies nur eine der möglichen Definitionen. Bis heute konnte man sich nicht auf eine einheitliche Definition verständigen. Dies führt nun wieder dazu, daß es einige OLAP Produkte gibt, die der obigen Definition nicht genügen. Bereits 1993 wurde ein ‘white paper’ zum Thema OLAP veröffentlicht, welches jedoch von vielen Fachleuten abgelehnt wurde, da es im Auftrag einer Firma erstellt wurde. Aus diesem Grund wurde versucht, eine Anforderungsdefinition zu entwickeln, die nur die wesentlichsten Eigenschaften von OLAP erfüllt. Es entstand der FASMI Test. FASMI ist hier eine Abkürzung, wobei jeder Buchstabe für eine Anforderung steht (vgl. [23]).

- (F)ast
- (A)nalysis
- (S)hared
- (M)ultidimensional
- (I)nformation

Fast bedeute in diesem Zusammenhang, daß die Ergebnisse des OLAP innerhalb weniger Sekunden zur Verfügung stehen. Dabei geht man davon aus, daß die Antwortzeit unter 30 Sekunden liegen sollte. *Analysis* bedeutet, daß es in irgendeiner Art und Weise möglich sein muß, eine (mehrere) Analysen mit den Daten durchzuführen. Auf Grund der Tatsache, daß hier die Analysemethoden nicht weiter eingeschränkt wurden, ergibt sich hier natürlich ein sehr weiches Kriterium. Eine Analyse kann sowohl die einfache Berechnung eines Durchschnitts, als auch die Berechnung einer umfangreichen Clusteranalyse sein. Einfacher ist

es wiederum, bei dem Begriff der *Shared* Anforderung. Diese Anforderung wird erfüllt, wenn die Software es ermöglicht mit mehreren Anwendern die Software gleichzeitig zu verwenden. Für den einzelnen Nutzer soll dies natürlich nicht sichtbar sein. *Multidimensional* bedeutet, daß es dem Benutzer möglich ist, multidimensionale Analysen durchzuführen. Hierzu folgendes Beispiel: Ein Manager möchte den Umsatz an Elektrogeräten im ersten Quartal für den Bereich Saarbrücken wissen. Bei dieser Abfrage haben wir zwei Dimensionen. Zum einen die Zeit (erstes Quartal: Januar-März) und zum anderen die Elektrogeräte (Computer, Fernseher, usw.). Der OLAP bleibt natürlich nicht nur auf zwei Dimensionen beschränkt, sondern kann n-dimensional sein. Das Limit legt hier die einzelne Software fest. Der letzte Punkt ist trivial. *Information* bedeutet schlicht und einfach, daß der OLAP mit einer großen Menge an Daten umgehen können muß. Diese Bedingung ist in der Regel erfüllt und wurde hier nur der Vollständigkeit halber aufgeführt.

2.3 Operationen des OLAP

Die Hauptaufgabe des OLAP ist die multidimensionale Darstellung von Informationen. Um diese Darstellung besser zu steuern, stehen vier Operationen zur Verfügung (vgl.[4]). Dies sind:

- Roll-up
- Drill-down
- Dice/Slice
- Pivot

Um die obigen Funktionen zu verdeutlichen, soll folgendes Beispiel verwendet werden.

Die Abbildung 2.1 zeigt einen Würfel der die drei Dimensionen Time (Zeit), Location (Ort) und Item (Ware) widerspiegelt. Hier kann man nun sehr gut die multidimensionale Darstellung der Daten erkennen. Zudem sieht man auch,

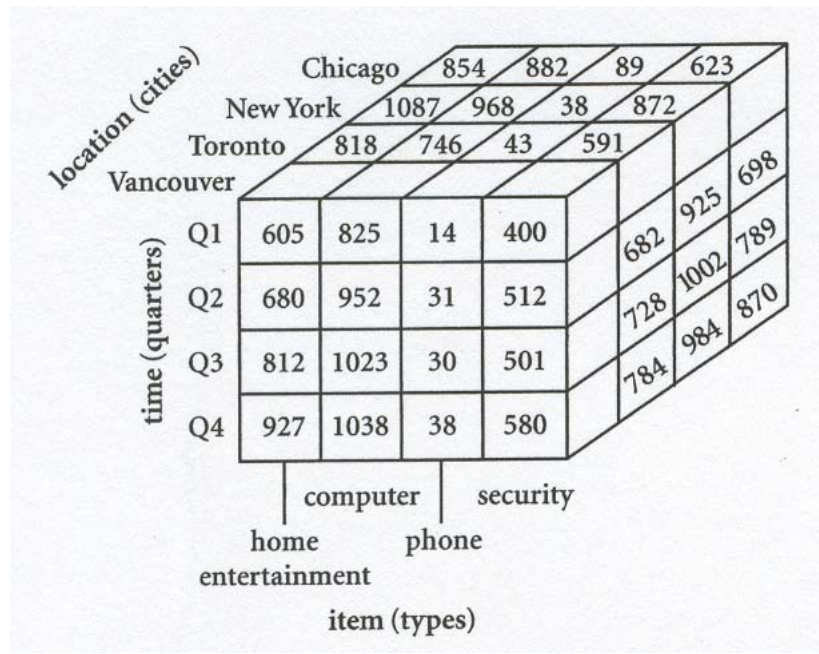


Abbildung 2.1: Beispiel OLAP

daß das Gewinnen von relevanten Informationen aus einem solchen Würfel nicht trivial ist. Aus diesem Grund gibt es die oben aufgezählten Operationen.

Beim Roll-up wird eine oder mehrere Dimensionen zusammen „gerollt“. Für die Dimension Location könnte das bedeuten, daß nun nicht mehr die einzelnen Städte betrachtet werden, sondern die Länder. New York und Chicago würden dann zu der Location USA und die Städte Vancouver und Toronto zur Location Kanada zusammengefasst. Dies kann z. B. wichtig sein, wenn man betrachten will, welches Land den größten Umsatz verursacht. Durch die Roll-up Operation wird also eine Abstraktion erreicht. Dem entgegen steht die Operation Drill-down. Bei dieser werden die Informationen im Bezug auf eine (mehrere) Größe(n) genauer betrachtet. Im obigen Beispiel könnte man auf die Abstraktion von Quartalen verzichten und an Stelle dessen die einzelnen Monate darstellen. Eine solche detaillierte Aufschlüsselung kann z.B. bei der Bestimmung der Kosten, die von den Kunden verursacht werden, von Bedeutung sein. Die nächste Operation ist Dice und Slice. Bei diesen beiden Operationen werden nur noch Teilmengen des Würfels betrachtet. Ein Beispiel zeigt folgende Abbildung 2.2.

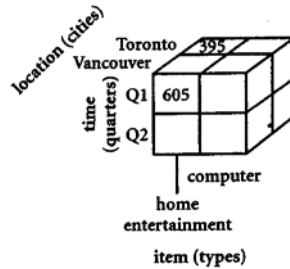


Abbildung 2.2: Beispiel Dice

Wie man erkennen kann, wurde dieser kleine Würfel aus dem großen Würfel extrahiert. In diesem Fall werden nur noch die Städte in Kanada in den ersten zwei Quartalen für die Waren home entertainment und computer betrachtet. Ein Slice wird gebildet, indem man den Würfel um eine Dimension reduziert. In dem obigen Fall könnte die Dimension Location auf die Stadt Vancouver beschränkt werden, wodurch sich die Dimension des Würfels auf zwei reduziert und so aus dem Würfel eine Tabelle wird. Die letzte Operation ist die Pivot Operation. Diese Operation dient im wesentlichen der Darstellung der Informationen. Extrahiert man den oben genannten Slice erhält man eine Tabelle in deren Zeilen die Quartale stehen und in deren Spalten die Waren zu finden sind. Die Pivot Operation dreht die Tabelle, so daß die Waren jetzt als Zeilen und die Quartale als Spalten dargestellt werden.

Mit diesen vier Operationen ist es nun möglich alle gewünschten Informationen mit der gewünschten Präzision darzustellen. Im folgenden soll die RFM Analyse als ein Beispiel für die Verwendung von OLAP dargestellt werden.

2.4 RFM Analyse

Im Marketing stellt sich immer wieder die Frage, wie eine Kampagne geplant werden soll, damit diese möglichst erfolgreich ist. Früher wurden die Marketingkampagnen überwiegend nach dem Gießkannenprinzip gestaltet, dies bedeutet, daß die Werbemittel an alle Kunden gleichermaßen verteilt werden. Dies erwies sich jedoch als wenig erfolgreich. Um eine Steigerung in diesem Bereich zu er-

reichen, ging man dazu über, die Werbemittel gezielt an bestimmte Kunden zu verteilen. Es stellte sich heraus, daß der zweite Ansatz viel erfolgreicher war. Das Problem ist nun die Bestimmung des Kundenkreises, an die sich die Kampagne wendet. Hierzu wird heute häufig die RFM Methode eingesetzt. RFM ist hierbei eine Abkürzung und steht für:

- Recency
- Frequency
- Monetary

Diese drei Begriffe stehen für drei Eigenschaften, die jeder Kunde besitzt. Die Recency beschreibt dabei, wie lange der letzte Kauf eines Kunden zurückliegt. Die Frequency beschreibt, wie häufig ein Kunde gekauft hat und schließlich wird durch das Monetary der finanzielle Umfang der Transaktionen wiedergespiegelt. Die Kundenbeziehung wird also auf diese drei Werte reduziert. Das gesamte Modell beruht dabei auf drei betriebswirtschaftlichen Annahmen:

1. Die Wahrscheinlichkeit, daß ein Kunde einen erneuten Kauf durchführt ist um so größer, je kürzer der Zeitabstand zum letzten Kauf ist.
2. Die Wahrscheinlichkeit, daß ein Kunde einen erneuten Kauf durchführt ist um so größer, je häufiger ein Kunde kauft.
3. Die Wahrscheinlichkeit, daß ein Kunde einen erneuten Kauf durchführt ist um so größer, je mehr Geld der Kunde pro Kauf ausgibt.

Unter der Voraussetzung das die obigen Annahmen gelten, läßt sich nun der RFM Wert eines Kunden bestimmen. Hierzu werden für jede Eigenschaft (Recency, Frequency, Monetary) fünf Klassen eingeführt. Die Kunden werden nun zunächst in die fünf Klassen entsprechend der Recency eingeteilt, d. h. in jeder Klasse befinden sich 20% der Kunden. Alle fünf Klassen der Recency werden nun in jeweils 5 Klassen entsprechend der Frequency unterteilt. Hierdurch ergeben sich nun 25 Klassen (1-1),(1-2),(1-3),(1-4),(1-5),(2-1)...(5-5). Anschließend werden die nun 25 Klassen wiederum entsprechend des Monetary eingeteilt. Hierdurch ergibt sich folgendes Bild.

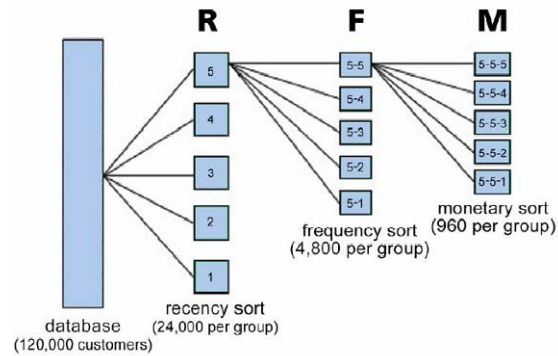


Abbildung 2.3: RFM

Geht man von einem Kundenstamm von 120 000 Kunden aus, kann die Marketingkampagne auf 960 Kunden beschränkt werden. In diesem Fall werden allerdings nur die Topkunden (5-5-5) in die Marketingkampagne einbezogen.

Die RFM Analyse besitzt jedoch auch einige Nachteile. Zum einen ist es eine recht einfache Analyse, die einen Kunden lediglich auf die drei Eigenschaften RFM reduziert. Das Abbilden von qualitativen Größen, wie z.B. des Weiterempfehlungspotentials, ist nicht möglich. Desweiteren ist nicht explizit festgelegt, wie die drei Werte bestimmt werden. Soll für die Recency nur die letzte Woche, der letzte Monat oder das letzte Jahr betrachtet werden? Ebenso kann für den Monetary-Wert der Gesamtumsatz eines Kunden, der Umsatz eines bestimmten Produkts oder jede andere Größe, die mit dem Umsatz des Kunden in Verbindung steht, verwendet werden. Auch die Klassifizierung der Kunden ist frei wählbar. Auf Grund dieser sehr variablen Gestaltung kann natürlich keine absolut präzise Aussage über den Wert eines Kunden getroffen werden. Dies ist aber im Fall des Marketing auch nicht zwingend erforderlich. Im Vergleich schneidet eine Kampagne, die mit dem Gießkannenprinzip gestaltet wurde immer noch wesentlich schlechter ab, als eine Kampagne, die mit der RFM Analyse geplant wurde. Die oben beschriebenen Nachteile sind für den Bereich des Marketing durchaus akzeptabel.

Kapitel 3

Data Mining

3.1 Data Mining als Teilaspekt von analytischem CRM

Der Begriff Data Mining wurde bereits in den 60er Jahren in der Statistik für die hypothesenfreie Suche nach Datenmustern verwendet [6]. Der Grundstein für die konzeptionelle Entwicklung des Data Mining wurde 1989 mit dem interdisziplinären Workshop „Knowledge Discover in Databases“ im Rahmen der 11th International Joint Conference of Artificial Intelligence in Detroit gelegt.

Im folgenden wird der Begriff des *Data Mining* kurz erklärt, bevor als ausgewählte Data Mining Verfahren die Assoziationsanalyse, das Entscheidungsbaumverfahren und das Clustering erläutert werden.

3.1.1 Definition DataMining

Der Begriff des *Data Mining* wird in der Literatur nicht einheitlich verwendet. Häufig werden z. B. die Begriffe Knowledge Discovery in Databases (KDD) und Data Mining synonym benutzt. Andere Autoren sehen in KDD den Oberbegriff, d.h. KDD beschreibt den

„... process of using data mining methods (algorithms) to extract

(identify) what is deemed knowledge ...“

während Data Mining als

„...step in the KDD process consisting of particular data mining algorithms ...“

zu verstehen ist [3, Seite 9].

Neben der oben erwähnten Definition sollen im folgenden exemplarisch weitere Definitionen von Data Mining aufgeführt werden.

Definition nach Herb Edelstein [2, Seite 1]:

„Data Mining is a process which uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid and accurate predictions.“

Definition nach Jesus Mena [5]:

„Data Mining ist ein Prozess zur Aufdeckung nutzbringender und aussagekräftiger Muster, Profile und Trends.“

Trotz unterschiedlicher Definitionen, ist bei alle Autoren eine Gemeinsamkeit festzustellen: das Aufdecken von Muster. Und um diese Muster aufzudecken, werden entsprechende Data Mining Methoden benötigt.

3.2 Data Mining Methoden

In der Literatur findet man häufig die Unterscheidung in beschreibende (Englisch: descriptive) und prognostizierende (Englisch: predictive/prescriptive) Funktionen des Data Mining [4].

Deskriptives Data Mining verfolgt das Ziel, die Eigenschaften der Daten einer Datenbank allgemein zu beschreiben und somit ein besseres Verständnis für die Daten zu erwirken. Prognostizierendes Data Mining analysiert die vorhandenen Daten und leitet daraus Möglichkeiten ab, Vorhersagen zu treffen. Aus diesem

Grunde wird überwiegend prognostizierendes Data Mining im analytischen CRM eingesetzt.

Autoren nennen als häufigste Aufgaben des Data Mining die Abweichungserkennung, die Abhängigkeitsentdeckung, die Klassifikation oder Vorhersage und die Segmentierung. Im folgenden wird jeweils ein Verfahren aus den drei zuletzt genannten Aufgaben näher erläutert.

3.2.1 Assoziationsanalyse

Die Assoziationsanalyse ist eine Methode der Abhängigkeitsentdeckung, bei der es darum geht, unbekannte Zusammenhänge zwischen Variablen zu entdecken und zu quantifizieren. Solche Produktassoziationsanalysen haben das Ziel, Cross-Selling-Potenziale zu realisieren. D.h. man ermittelt, welche der Kunden, die Produkt A gekauft haben, auch für den Kauf von Produkt B in Frage kommen. Somit können diese Kunden gezielt angesprochen werden. Als eine Ergänzung zu der oben dargestellten Cross-Selling-Analyse können Warenkorbanalysen durchgeführt werden. Unter Verwendung von Assoziationsanalysen wird dann z. B. untersucht, welche Produktkombinationen überdurchschnittlich häufig zusammen erworben werden. Diese Ergebnisse könne dann als erste Ausgangspunkte herangezogen werde, um z. B. die Sortimentgestaltung oder die Layoutplanung von Supermärkten oder Katalogen zu verbessern. Als ein weiteres Beispiel hierfür sei der Online-Shop Amazon genannt. Auf seiner Webseite (<http://www.amazon.de>) spricht Amazon seinen Kunden Kaufempfehlungen auf Grund von vergangenen Produktkäufen anderer Käufer des im Moment in den Warenkorb gelegten Produktes aus.

Verwendung

Die SAP-Bibliothek gibt folgende Verwendung für die Data Mining Methode *Assoziationsanalyse* an [11]:

Die Assoziationsanalyse dient dazu, Regelmäßigkeiten vor allem bei geschäftlichen Vorgängen zu finden und entsprechende Regeln zu formulieren, in der Art „Wenn ein Kunde Produkt A kauft, kauft er

auch Produkt B und C“. [...]

Die Assoziationsanalyse hilft Ihnen also zum Beispiel dabei, Cross-Selling-Chancen aufzudecken. Sie können die Regeln dazu nutzen, assoziierte Produkte in einem Katalog, im Supermarkt oder im Webshop zusammen anzuordnen, oder mit einer Marketingkampagne für Produkt C gezielt Kunden anzusprechen, die bereits Produkt A gekauft haben. Bei der Assoziationsanalyse werden die Regeln beim Training des Modells auf historischen Daten ermittelt. Sie können die ermittelten Assoziationsregeln anzeigen und exportieren.

Einsatzmöglichkeiten

Einsatzmöglichkeiten werden in der SAP-Bibliothek wie folgt angegeben [12]:

Mit Data Mining können Sie automatisch interessante Muster und schwer aufzuspürende Zusammenhänge in großen Datenmengen ermitteln. Data Mining liefert Ihnen Erkenntnisse und Zusammenhänge, die bisher verborgen blieben oder außer acht gelassen wurden, weil sie für nicht analysierbar gehalten wurden.

Da jedes Unternehmen unterschiedliche Anforderungen an Data Mining hat, ist es nicht möglich, feste Modelle zur Vorhersage auszuliefern. Mit den im SAP BW zur Verfügung stehenden Data-Mining-Methoden können Sie jedoch Modelle entsprechend Ihren Anforderungen selbst anlegen und damit aus den Daten in Ihrem SAP BW entscheidungsrelevante Informationen ermitteln. Sie können zum Beispiel Verhaltensweisen von Kunden analysieren und Trends vorher-sagen, indem Sie Muster im Kundenverhalten erkennen und nutzen. Dadurch können Sie Fragen wie die folgenden beantworten:

- Welchen Kunden sollte wann welches Angebot unterbreitet werden?
- Welche Kunden sind abwanderungsgefährdet?
- Wie hoch ist das Cross-Selling-Potential für ein neues Produkt?

Regel-Syntax

Assoziationsregeln haben immer folgende Form:

If A und B und ... und X gekauft *Then* Y gekauft.

Die Anzahl der Elemente in der Bedingung (*If*-Part) ist abhängig vom jeweiligen Unternehmen bzw. von dessen Produkten.

Bei hochwertigen Produkten, wie z.B. einem PC, würde im *If*-Part nur ein Produkt (hier der PC) auftauchen, und in dem *Then*-Part dann z.B. ein Monitor oder Drucker. In Supermärkten hingegen würden im *If*-Part mehrere Produkte auftauchen.

Unsicherheitsgrad der Assoziationsregeln [9]

Zusätzlich zu dem *If*- und *Then*-Part gibt es zwei weitere Zahlen, die den Unsicherheitsgrad einer Regel ausdrücken.

Die erste Zahl wird *Support* für die Regel genannt. Der Support ist die Anzahl der Transaktionen, die alle Produkte sowohl der Bedingung (*If*-) als auch der Konsequenz (*Then*-Part) einschließen. (Manchmal wird der Support auch als prozentualer Anteil an den gesamten Transaktionen ausgedrückt.)

Die andere Zahl ist bekannt als *Confidence* der Regel. Confidence ist der Quotient aus der Anzahl der Transaktionen die im *If*- und *Then*-Part enthalten sind (den Support) und der Anzahl der Transaktionen aus der *If*-Bedingung.

Die Begriffe können am einfachsten anhand eines Beispiels verdeutlicht werden:

Wir nehmen an, ein Supermarkt habe 100.000 Transaktionen an den Kassen, wobei 2.000 davon die Produkte A und B enthielten, darunter 800, die Produkt C enthielten. Die dazugehörige Assoziationsregel lautet wie folgt:

If A und B gekauft *Then* C gekauft.

Die Regel hat einen Support von 800 Transaktionen (bzw. $0,8\% = 800/100.000$) und einen Confidence von $800/2.000 = 40\%$.

Lift ist ein weiterer interessanter Parameter in der Assoziationsanalyse. Lift ist nichts anderes als das Verhältnis von Confidence zu Expected Confidence. Nehmen wir weiter an, das Produkt C tauche in insgesamt 5.000 Transaktionen auf. Deshalb ist der Expected Confidence $5.000/100.000 = 5\%$ und der Lift $= 40\%/5\% = 8$.

3.2.2 Entscheidungsbäume

Entscheidungsbaumverfahren

Entscheidungsbäume werden seit den 60er Jahren zur Klassifikation von Objekten eingesetzt. Klassifikationsprobleme haben zur Zielsetzung, die Merkmale herauszufinden, die die Zugehörigkeit zu einer Gruppe ausmachen. Mit Hilfe dieser Merkmale können dann neue Objekte bereits existierenden Klassen zugeordnet werden. Als Anwendungsgebiet kommen somit im Rahmen von analytischem CRM diverse Vorhersagemodelle in bezug auf Kundenverhalten in Frage. Am häufigsten werden Entscheidungsbäume eingesetzt, um zu bestimmen, welches Verhaltensprofil besonders oft zum Verlust eines Kunden geführt hat. Dadurch lassen sich rechtzeitig entsprechende Gegenmaßnahmen einleiten, um dem Absprung des Kunden vorzubeugen.

In der Praxis kommen unterschiedliche Verfahren zum Einsatz wie z. B. CART¹, ChAID² sowie ID3 und dessen Nachfolger C4.5 [7, Seite 34].

Bei Entscheidungsbaumverfahren handelt es sich um Verfahren des überwachten Lernens, d. h. das Modell wird anhand einer Trainingsdatenmenge aufgebaut, für die die Klassifikation der einzelnen Sätze bereits bekannt ist. Zu Beginn ist die gesamte Trainingsdatenmenge der Wurzel des Baumes zugeordnet. Anhand bestimmter Tests wird die Gesamtdatenmenge unterteilt, und die so entstehenden

¹CART steht für Classification an Regression Trees

²ChAID steht für Chi-Squared Automated Interaction Detection

Teildatenmengen werden den Knoten des Baumes zugeordnet. Dabei wird das Ziel verfolgt, die ‘Reinheit’ in bezug auf die Klassenzugehörigkeit im jeweiligen Knoten zu optimieren. Ist eine weitere Unterteilung eines betrachteten Knotens nicht möglich, so wird er zum Blatt- oder Endknoten. Dabei wird ihm diejenige Klassenzugehörigkeit zugewiesen, die in der Menge der Daten dieses Knotens am häufigsten vorkommt.

Unterteilungskriterien

Als Unterteilungskriterium werden häufig die Ausprägungen eines Merkmals verwendet. Theoretisch gibt es dann so viele Teilmengen, wie es Ausprägungen für ein Attribut gibt. Da die Zahl der möglichen Ausprägungen jedoch sehr hoch sein kann, wird die Anzahl der betrachteten Ausprägungswerte z. T. begrenzt. In der Praxis werden häufig nur Unterteilungen betrachtet, die die Datenmenge in zwei Teilmengen unterteilen. So entstehen sog. Binärbäume, d.h. jeder Knoten hat genau zwei Söhne.

Prinzipiell geht es darum, die Merkmale zu finden, anhand derer die Daten am besten diskriminiert werden können, d.h. die Klassenzugehörigkeit bestimmt werden kann. Hierfür werden sog. Auswahlmaße berechnet.

Auswahlmaße

Das Ziel bei der Unterteilung der Datenmenge liegt darin, die Unreinheit in den Knoten zu minimieren bzw. die Güte der Unterteilung zu maximieren.

Eines der bekanntesten Auswahlmaße ist der sog. Informationsgewinn. Er basiert auf dem Informationswert, der auch als Entropie einer Wahrscheinlichkeitsverteilung bezeichnet wird. Danach berechnet sich der Informationswert einer Datenmenge S gemäß folgender Formel, wobei $\text{freq}(C_j, S)$ die Anzahl der Sätze in S bezeichnet, die der Klasse C_j angehören, $|S|$ beziffert die Gesamtzahl der Sätze in S :

$$\text{info}(S) = - \sum_{j=1}^k \left(\frac{\text{freq}(C_j, S)}{|S|} \times \log_2 \frac{\text{freq}(C_j, S)}{|S|} \right)$$

Bezogen auf einen Entscheidungsbaum betrachten wir die Trainingsdatensmenge T , die aufgrund eines konkreten Tests in die Teilmengen T_1, T_2, \dots, T_n

unterteilt wird.

Zuerst wird der Informationswert $\text{info}(T)$ für den gesamten Baum berechnet. Der so errechnete Wert verdeutlicht die Informationsmenge, die benötigt wird, um die Klassifikation eines Trainingssatzes zu bestimmen. Anschließend kann ein ähnliches Maß berechnet werden, nun aber nachdem die Gesamtmenge T in Abhängigkeit der n Ergebnisse eines Tests X aufgeteilt wurde. Und zwar berechnet sich die erwartete Informationsmenge in diesem Fall als gewichtete Summe über die Teilmengen des Knotens:

$$\text{info}_X(T) = \sum_{i=1}^n \frac{|T_i|}{|T|} \times \text{info}(T_i)$$

Der Informationsgewinn mißt nun die erwartete Entropieminderung, die durch die Aufteilung der Daten mittels eines bestimmten Attributs erzielt wird. Die Formel zur Berechnung des Informationsgewinns lautet:

$$\text{gain}(X) = \text{info}(T) - \text{info}_X(T)$$

Je größer die Entropieminderung ausfällt, desto reiner wird der durch den Split entstehende Teildatenbestand in bezug auf die Klassenstruktur. Der Informationsgewinn stellt also ein Maß für die Güte eines gefundenen Splits dar. Es wird diejenige Unterteilung gewählt, die dieses Maß maximiert.

Der Informationsgewinn hat den Nachteil, daß er Attribute mit einer großen Anzahl möglicher Ausprägungen bevorzugt. Um dies zu verhindern, hat Quinlan, der Entwickler des ID3- und C4.5-Algorithmus die Normierung des Informationsgewinns mittels der sog. Splitinformation vorgeschlagen, wodurch sich das Informationsgewinnverhältnis ergibt. Die Splitinformation berechnet sich wie folgt:

$$\text{split info}(X) = \sum_{i=1}^n \frac{|T_i|}{|T|} \times \log_2\left(\frac{|T_i|}{|T|}\right)$$

Damit läßt sich dann das Informationsgewinnverhältnis nach folgender Formel berechnen:

$$\text{gain ratio}(X) = \text{gain}(X)/\text{split info}(X)$$

Das Informationsgewinnverhältnis berücksichtigt somit die Anzahl der Ausprägungen eines Attributes.

Mittels Auswahlmaße erfolgt also die Unterteilung des Baumes, bis alle Knoten rein sind oder eine weitere Unterteilung nicht möglich ist. Dies führt jedoch zu einer Überanpassung des Baumes an die Trainingsdaten, d. h. der Baum funktioniert gut mit den Trainingsdaten, lässt sich aber nicht auf unabhängige Testdaten verallgemeinern. Dieses Phänomen wird als „Overfitting“ bezeichnet. Deshalb werden zur Bestimmung von Endknoten Abbruchkriterien definiert. Man spricht in diesem Zusammenhang von Pruning-Techniken, die jedoch an dieser Stelle nicht weiter erläutert werden sollen.

Eine Beispielrechnung findet sich in Anhang B.

3.2.3 Clustering

Clustering wird zur Kundensegmentierung verwendet. Segmentierung allgemein hat zum Ziel, eine Datenmenge, die noch nicht klassifiziert ist, in Gruppen aufzuteilen, Dabei werden die Objekte so aufgeteilt, daß die Ähnlichkeit innerhalb der Gruppe möglichst hoch und zwischen den Gruppen möglichst gering ist. Die Merkmale, die zur Beschreibung bzw. Bildung der verschiedenen Cluster herangezogen werden, sollten gute Diskriminatoren sein.

Bei der Kundensegmentierung werden nun sämtliche potentielle und aktuelle Kunden in Untergruppen (Kundensegmente), die in sich bzgl. ihrer Marktreaktion homogen, untereinander jedoch heterogen sind, aufgeteilt. In den meisten Fällen werden die Kunden anhand Demographie (Durchschnittsalter, Geschlecht, Familienstand), Kaufverhalten, etc. in homogene Gruppen zusammengefaßt.

Kapitel 4

Kundenanalyse

4.1 Kundenverhaltensanalyse

Durch Kundenverhaltensanalysen gewinnen Unternehmen weiteres Wissen darüber, wer ihre Kunden sind. Somit können Kundenprofile erstellt und Verhaltensmuster erkannt werden. Vorrangig werden folgende Aspekte analysiert

- Kaufverhalten
- Abwanderungsverhalten
- Zufriedenheit

Das Ziel einer Kundenverhaltensanalyse ist immer das Entdecken von unbekanntem Mustern im Kundenverhalten basierend auf historischen Daten. Um diese Muster zu entdecken, bedient man sich den im vorangegangenen Kapitel beschriebenen Data Mining Methoden.

4.1.1 Kaufverhaltensanalyse

Bei der Kaufverhaltensanalyse soll u.a. ermittelt werden, welche Kunden welche Produkte zusammen bzw. nicht zusammen gekauft werden. Hierzu dienen die in Kapitel 3.2.1 beschriebenen Produktassoziationsregeln.

4.1.2 Abwanderungsverhaltensanalyse

Das Abwanderungsverhalten eines Kunden spielt eine sehr große Rolle für ein Unternehmen. Liegt die prognostizierte Abwanderung eines potentiellen Kunden in einer frühen Phase des Customer-Lifetimes (vgl. Kapitel 5), so lohnt es sich meist für das Unternehmen nicht, überhaupt eine Geschäftsbeziehung mit diesem Kunden zu beginnen. Liegt die vorhergesagte Abwanderung aber in einer späten Phase, so kann der Abwanderung rechtzeitig entgegengewirkt werden.

4.1.3 Zufriedenheitsanalyse

Die Zufriedenheit eines Kunden kann nur schwer analysiert werden. Meist beruhen die Einschätzungen auf dem subjektiven Empfinden von Mitarbeitern oder auf von den Kunden ausgefüllten Fragebögen.

4.2 Kundenwertanalyse

Über die Kundenwertanalyse läßt sich ermitteln, welchen Wert einzelne Kunden oder Kundensegmente für ein Unternehmen haben. Auf Basis dieser Informationen können dann vom Management Investitionsentscheidungen getroffen werden. Im einzelnen gehören folgende Analysen zu den Kundenwertanalysen:

- **Kundenprofitabilitätsanalyse**
Am einfachsten läßt sich die Kundenprofitabilität als Differenz zwischen Erlös und Kosten pro Kunde errechnen. Aufschlußreicher ist jedoch eine detaillierte Kundendeckungsbeitragsanalyse, die unter Einbeziehung verschiedener Erlösarten, Produkt- und Vertriebskosten ein in sich stimmiges Bild der Kundenprofitabilität liefert.
- **ABC-Analyse**
Die ABC-Klassifikation dient dazu, Kunden auf Basis von Profitabilitätsdaten in A-, B- oder C-Kunden einzuteilen. Dadurch läßt sich beispielsweise ermitteln, mit wieviel Prozent der Kunden wieviel Umsatz gemacht wird. Auf Basis dieser Informationen können beispielsweise Marketing-, Vertriebs- und Kundenservice-Ressourcen zugeordnet werden.

Absatzmenge	30 ST
Bruttoumsatz	500
- Erlösschmälerungen	20
Nettoerlös	480
- Produktkosten	250
Deckungsbeitrag I	230
- Vertriebseinzelkosten	20
- Kampagnen- und Promotionskosten	10
- Kundenbezogene Auftragskosten	10
- Kundenbezogene Frachtkosten	40
Deckungsbeitrag II	150
- Kundenbesuche	30
- Kundenunterstützung	10
- Kundenpflege	50
Deckungsbeitrag III	60

Tabelle 4.1: Kundenprofitabilität

- Customer-Lifetime-Value-Analyse

Die Customer-Lifetime-Value-Analyse (CLTV-Analyse) betrachtet den Kundenwert von Kunden eines Segments über die gesamte Kundenbeziehung hinweg. Customer-Lifetime-Value ist eine Kennzahl, die den Vermögenswert eines typischen Kunden eines Kundensegments für einzelne Zeitabschnitte einer Kundenbeziehung (Lifetime-Perioden) angibt. Die CLTV-Analyse dient zur Einschätzung, ob sich die Investition in die Akquisition weiterer Kunden eines Kundensegments lohnt. Auf Basis dieser Analysen können die Kunden in Kundensegmente eingeteilt werden. Es kann entschieden werden, wie mit welchem Kundensegment umgegangen werden soll. Beispielsweise, ob vielleicht bestimmte Marketingkampagnen durchgeführt werden sollen oder ob in ein bestimmtes Kundensegment investiert werden soll.

Generell lässt sich sagen, daß der Kundenwert aus zwei Hauptkomponenten besteht. Die erste Komponente sind die quantitativen Eigenschaften eines Kun-

den wie Umsatz, Einzelkosten, Akquisitionskosten usw. Diese lassen sich recht einfach aus dem Datenbestand extrahieren. Die zweite Komponente sind die qualitativen Eigenschaften eines Kunden, wie z. B. Weiterempfehlungs-Potential oder Up/Cross-Selling Potentiale. Diese sind nun im Gegensatz zu den quantitativen Eigenschaften nicht mehr so leicht zu ermitteln. Für die Bestimmung der qualitativen Merkmale werden überwiegend subjektive Einschätzungen der Mitarbeiter verwendet. Zudem müssen diese Eigenschaften noch auf einen Zahlenwert transferiert werden, damit später zusammen mit den quantitativen Eigenschaften ein kommulierter Kundenwert entsteht.

Im nächsten Kapitel wird nun Customer-Lifetime-Value-Analyse näher erläutert.

Kapitel 5

Customer-Lifetime-Value

Eine der am häufigsten durchgeführten Analysen im Bereich aCRM ist sicher die Customer-Lifetime-Analyse. Bei dieser Analyse wird der Kundenwert über den gesamten Kundenlebenszyklus betrachtet. Der auf diese Weise berechnete Wert wird als Customer-Lifetime-Value (CLTV) bezeichnet. Dieser CLTV eines Kunden kann dann dazu verwendet werden die Beziehung zum Unternehmen zu bewerten. Die Customer-Lifetime-Analyse ist dabei wesentlich präziser, als die RFM Analyse. Zudem kann die Customer-Lifetime-Analyse auch dazu verwendet werden, Vorhersagen über das zukünftige Kundenverhalten zu treffen.

5.1 Kundenwert

Zur Berechnung des CLTV wird zunächst eine Bestimmung des Kundenwertes festgelegt. Die Art und Weise der Bestimmung des Kundewertes ist dabei nicht festgelegt. Im Bereich des Lead-Management werden Kunden entsprechend ihres Leads nach hot, warm oder cold eingestuft. Um diese Einteilung nun für die Bestimmung des CLTV zu verwenden, muss hier eine Abbildung in den reellen Zahlenraum erfolgen (z.B. hot=1, warm=2, cold=3).

5.2 Kundenlebenszyklus

Nachdem ein Kundenwert berechnet wurde, wird dieser dann bezüglich des gesamten Kundenlebenszyklus betrachtet. Dieser Zyklus kann in sechs Phasen ein-

geteilt werden.

1. Kennenlernphase
2. Startphase
3. Penetrationsphase
4. Reifephase
5. Krisenphase
6. Trennungsphase

Jede dieser Phasen hat bestimmte Eigenschaften, die für die Phase charakteristisch sind. Die Kennenlernphase zeichnet sich dadurch aus, daß es erste Kontakte zwischen dem Kunden und dem Unternehmen gibt. An diese Phase schließt sich die Startphase an. In dieser werden die ersten Geschäfte zwischen Kunden und Unternehmen abgeschlossen. Es beginnt sich ein Vertrauen zwischen dem Kunden und dem Unternehmen aufzubauen. In der Penetrationsphase beginnt sich eine Routine in der Abwicklung der Geschäfte einzustellen. Das Kontaktvolumen zwischen Kunden und Unternehmen nimmt ab, auf Grund der Tatsache, daß Geschäfte mehr und mehr „automatisiert“ ablaufen. Der Penetrationsphase schließt sich dann die Reifephase an. Dies ist die Phase, in der der Kunde den höchsten Umsatz erzeugt. Die Abwicklung der Geschäfte läuft nun sehr routiniert ab. Das Kontaktvolumen ist minimal und die Kundenzufriedenheit hat ihren höchsten Wert erreicht. Die Reifephase geht dann über in die Krisenphase. In dieser nimmt die Kundenzufriedenheit ab, und die Umsätze sinken. Schließlich endet der Kundenlebenszyklus in der Trennungsphase. In dieser geht der Umsatz eines Kunden stetig zurück bis er schließlich bei 0 angekommen ist.

5.3 Berechnung CLTV

Der CLTV betrachtet nun den Kundenwert entlang dieses Lebenszyklus. Für die Berechnung des CLTV wird folgende Formel verwendet (vgl.[27]).

$$\text{CLTV} = \sum_{n=1}^T S(n)V(n)D(n)$$

- $S(n)$ = Survival Funktion
- $V(n)$ = Kundenwert Funktion
- $D(n)$ = Discounting Funktion

Die Survival Funktion dient dazu, darzustellen, wie groß die Wahrscheinlichkeit ist, daß ein Kunde den nächsten Lebenszyklus erreicht. Als Beispiel sei hier der Verkauf von Autos genannt. Eine große Menge von Personen wird sich wahrscheinlich zu einem Auto informieren, jedoch wird ein deutlich geringerer Teil auch tatsächlich das Auto kaufen. Dieses Verhalten soll mit der Survival Funktion simuliert werden. Der nächste Faktor ist der Kundenwert über dessen Bestimmung weiter oben bereits eingegangen wurde. Schließlich bezieht der CLTV noch eine Discounting Funktion mit ein. Diese wird dazu verwendet, um die Inflation mit in das Modell einzubeziehen. Die Customer-Lifetime-Analyse basiert, wie alle anderen Analysemethoden auch, auf den momentan verfügbaren (historischen) Daten. Mit diesen wird dann versucht, eine Prognose für die Zukunft zu erstellen. Aus diesem Grund muss hier die Inflationsrate mit in die Betrachtungen eingehen. Die Discounting Funktion kann jedoch noch erweitert werden, so daß beispielsweise Steuererhöhungen mit in das Modell einbezogen werden können.

5.3.1 Beispielrechnung CLTV

Das folgende Beispiel zeigt die vereinfachte Berechnung des CLTV für eine Gaststätte.

Der Kundenwert wird in diesem Beispiel durch den Deckungsbeitrag dargestellt. Die Survival Funktion entspricht der jährlichen Fluktuation. Auf die Discounting Funktion wurde hier verzichtet, um das Beispiel einfach und verständlich zu gestalten. Zunächst lässt sich feststellen, daß die CLTVs der einzelnen Kunden deutliche Differenzen aufzeigen. Ein Kunde der Klasse „lebendiges Inventar“ erwirtschaftet einen um den Faktor 81 größeren Deckungsbeitrag. Betrachtet man

Gasttyp:	"lebendiges Inventar"	Stamm Gast	Student
Umsatz:	2.548,00 €	509,60 €	39,20 €
Nettoumsatz:	2.196,55 €	439,31 €	33,79 €
Kosten:	50%	50%	50%
Deckungsbeitrag:	1.098,28 €	219,66 €	16,90 €
Jährliche Fluktuation:	12%	12%	20%
1. Jahr:	1.098,28 €	219,66 €	16,90 €
2. Jahr:	966,48 €	193,30 €	13,52 €
3. Jahr:	850,50 €	170,10 €	10,81 €
4. Jahr:	748,44 €	149,69 €	6,92 €
5. Jahr:	658,63 €	131,73 €	5,53 €
Summe (CLTV)	4.322,34 €	864,47 €	53,68 €

Abbildung 5.1: Beispiel CLTV für eine Gastätte

nur das Ergebnis des ersten Jahres, ist der Unterschied der beiden Klassen geringer (Faktor 68). Eine einfache Hochrechnung des ersten Jahres würde also zu einem verfälschten Ergebnis führen. Der Grund hierfür ist, daß die beiden Klassen unterschiedliche Survival Funktionen haben. Mit den in der Grafik ermittelten Ergebnissen kann der Betreiber der Gaststätte nun feststellen, daß es von existentieller Bedeutung ist, die Kunden, die der Klasse „lebendiges Inventar“ angehören zu halten.

Die Customer-Lifetime-Analyse besitzt jedoch auch einige Nachteile. Das größte Problem im Bereich des CLTV ist die Berechnung des Kundenwertes. Hier gibt es keine genauen Methoden, die festschreiben, wie dieser Wert zu bestimmen ist. Deshalb kann hier praktisch jedes Verhalten modelliert werden. Wie bereits mehrfach erwähnt, wird auf Grund der Ergebnisse der Vergangenheit eine Aussage über die Zukunft getroffen. Diese kann natürlich immer nur als bedingt korrekt betrachtet werden, da sich das Verhalten eines Kunden in der Zukunft durchaus verändern kann. Dennoch ist die Customer-Lifetime-Analyse eine der beliebtesten Analysen und findet unter anderem auch in der SAP Software ihren Einsatz, wie die Abbildung zeigt:

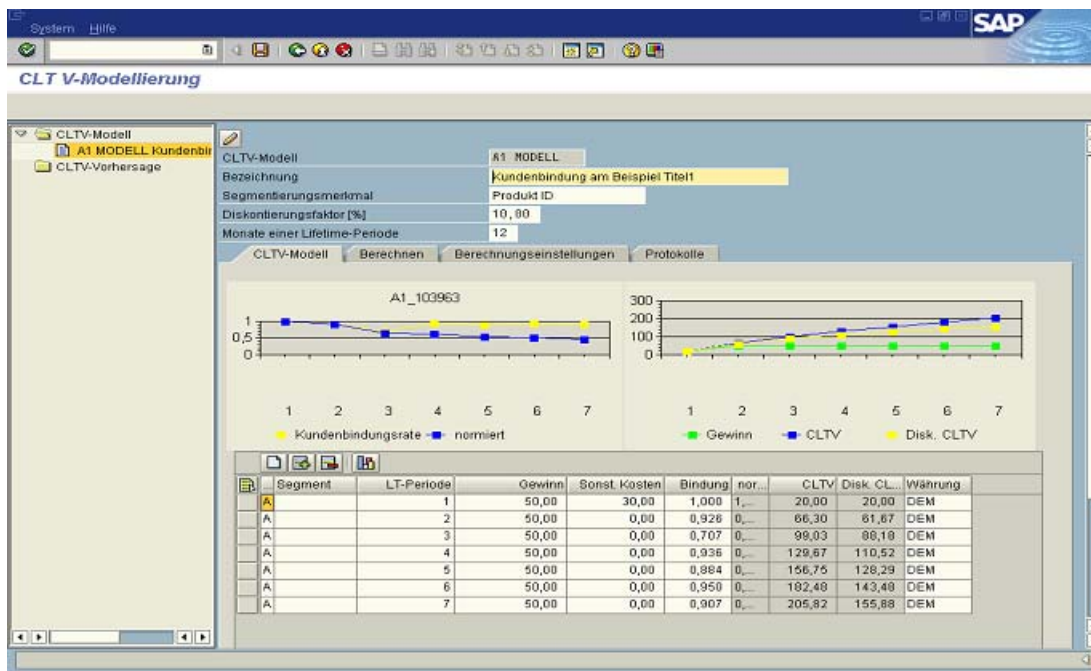


Abbildung 5.2: CLTV im SAP Modul

Kapitel 6

Fazit

Das aCRM leistet einen erheblichen Beitrag zur Gewinnung neuer relevanter Informationen. Die Problemstellung, die nun verbleibt ist, wie diese Informationen weiterverarbeitet werden. Aus einer Information lassen sich durchaus zwei völlig unterschiedliche Handlungen definieren. Es ist also unbedingt erforderlich, hier zwischen Information und Handlungsanweisung zu trennen. aCRM kann „nur“ Informationen liefern, keine Handlungsanweisungen. Weiterhin ist für eine gute Analyse eine breite Datenbasis nötig. Sind die Daten nur lückenhaft vorhanden und schlichtweg falsch, wird auch die Analyse des aCRM falsche Ergebnisse zurückliefern. Diese Bedingung einer umfangreichen und korrekten Datenbasis ist nicht so einfach zu erfüllen, wie es sich zunächst anhört. Für ein Unternehmen bedeute dies, daß es den Mitarbeitern Zeit geben muss, Daten zu pflegen bzw. einzufügen, und diese Zeit kostet wiederum Geld. Der Erfolg des aCRM ist auch wiederum nur schwer messbar, da es keinen kausalen Schluss zwischen den Ergebnissen und den durchgeführten Aktionen gibt. Eine Bewertung kann immer nur für die ausgeführte Aktion durchgeführt werden. Dennoch ist das aCRM eine bedeutende Stütze fast aller CRM Systeme.

Anhang A

Abkürzungsverzeichnis

In das Abkürzungsverzeichnis wurden diejenigen Kürzel aufgenommen, die entweder an mehreren Stellen in der Seminararbeit verwendet wurden, oder die eine feste Bedeutung im Analytischen Customer Relationship Managementbereich gewonnen haben. Nicht aufgenommen wurden einmalig gebrauchte Abkürzungen sowie Abkürzungen ohne expliziten CRM-Bezug (etwa ISBN).

aCRM	Analytisches CRM
BI	Business Intelligence
BW	Business Warehouse
CLTV	Customer Lifetime Value
CRM	Customer Relationship Management
DW	Data Warehouse
OLAP	Online Analytical Process
SRM	Supplier Relationship Management

Anhang B

Beispielrechnung zur Bestimmung des Informationsgewinnverhältnisses

Diese Beispielrechnung stammt aus [7, Anhang S. XVII ff].

Die folgende Tabelle B.1 zeigt die Trainingsdatensätze für das in der Literatur häufig zitierte Wetterbeispiel [10, S. 18ff] [13, S. 97ff]. Und zwar ist abhängig von den Wetterbedingungen vorherzusagen, ob ein Golfspiel stattfindet oder nicht. Die zu prognostizierende Variable ist also „Play“ mit den möglichen Ausprägungen „Yes“ oder „No“.

Als mögliche Attribute, anhand derer die Trainingsdatenmenge aufgeteilt werden kann, stehen „Outlook“, „Temperature“, „Humidity“ und „Windy“ zur Verfügung. Die folgende Abbildung B.1 verdeutlicht, wie die Aufteilung der Trainingsdaten aussähe, je nachdem welches Attribut herangezogen wird. Der obere Knoten ist als Wurzel zu interpretieren; er beinhaltet das Attribut, nachdem gesplittet wird, sowie die Anzahl der Datensätze der Kategorie „Yes“ bzw. „No“. Die untergeordneten Knoten beinhalten die Ausprägung des Merkmals und ebenfalls die Anzahl der Datensätze der Kategorie „Yes“ und „No“.

Die Berechnung des Informationsgewinnverhältnisses wird nachfolgend für die einzelnen Baumvarianten demonstriert.

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Tabelle B.1: Wetterdaten, Quelle: [13, S. 44]

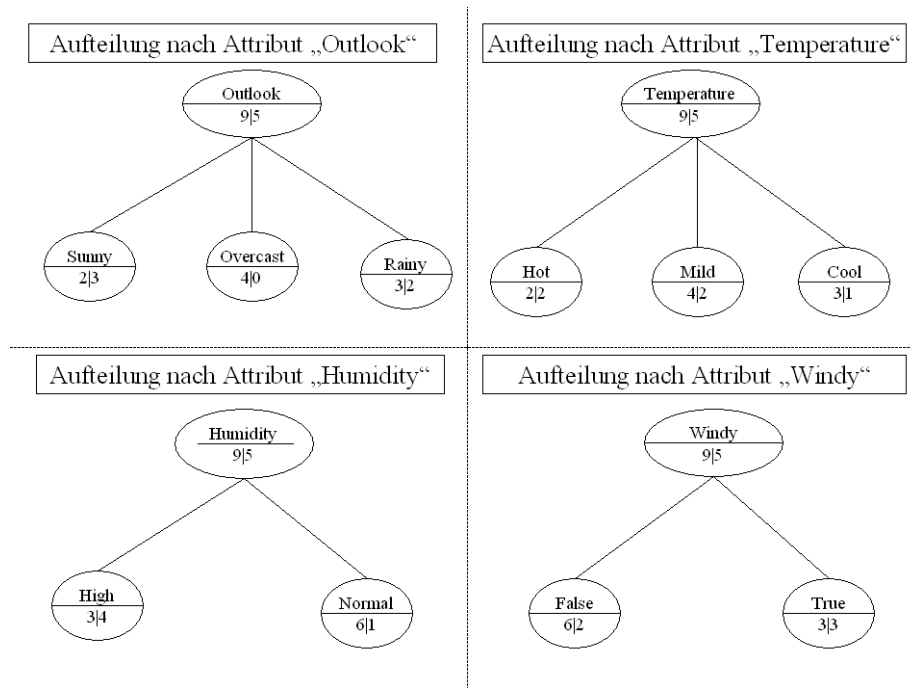


Abbildung B.1: Mögliche Aufteilung des Entscheidungsbaumes für die Wetterdaten, Quelle: [7, Anhang, S. XVIII]

Schritt 1: Berechnung des Informationswertes $\text{info}(T)$ der gesamten Trainingsdatensmenge

$$\text{Info}(T) = \text{info}([9|5]) = - \left(\frac{9}{14} \times \log_2 \left(\frac{9}{14} \right) + \frac{5}{14} \times \log_2 \left(\frac{5}{14} \right) \right) = 0,940$$

Schritt 2: Berechnung des Informationswertes im Anschluß an eine Aufteilung nach dem Attribut

(a) Outlook mit den Teilmengen $T_1 = 5$, $T_2 = 4$, $T_3 = 5$

$$\begin{aligned} \text{Info}_{\text{outlook}}(5|4|5) &= \frac{5}{14} \times \left[- \left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5} \right) \right] + \\ &\quad \frac{4}{14} \times \left[- \left(\frac{4}{4} \log_2 \frac{4}{4} \right) \right] + \\ &\quad \frac{5}{14} \times \left[- \left(\frac{3}{5} \log_2 \frac{3}{5} + \frac{2}{5} \log_2 \frac{2}{5} \right) \right] \\ &= 0,693 \end{aligned}$$

(b) Temperature mit den Teilmengen $T_1 = 4$, $T_2=6$, $T_3 = 4$

$$\begin{aligned} \text{Info}_{\text{temperature}}(6|4|6) &= \frac{4}{14} \times \left[- \left(\frac{2}{4} \log_2 \frac{2}{4} + \frac{2}{4} \log_2 \frac{2}{4} \right) \right] + \\ &\quad \frac{6}{14} \times \left[- \left(\frac{4}{6} \log_2 \frac{4}{6} + \frac{2}{6} \log_2 \frac{2}{6} \right) \right] + \\ &\quad \frac{4}{14} \times \left[- \left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) \right] \\ &= 0,91 \end{aligned}$$

(c) Humidity mit den Teilmengen $T_1 = 7$, $T_2=7$

$$\text{Info}_{\text{humidity}}(7|7) = 0,788^1$$

(d) Windy mit den Teilmengen $T_1 = 8$, $T_2 = 6$

$$\text{Info}_{\text{windy}}(8|6) = 0,892$$

Schritt 3: Berechnung des Informationsgewinns für

(a) Outlook

$$\text{Gain}(\text{outlook}) = 0,940 - 0,693 = 0,247$$

(b) Temperature

$$\text{Gain}(\text{temperature}) = 0,940 - 0,91 = 0,029$$

(c) Humidity

$$\text{Gain}(\text{humidity}) = 0,940 - 0,788 = 0,152$$

(d) Windy

$$\text{Gain}(\text{windy}) = 0,940 - 0,892 = 0,048$$

Schritt 4: Berechnung der Splitinfo für

(a) Outlook

$$\begin{aligned} \text{splitinfo}_{\text{outlook}}(5|4|5) &= \frac{5}{14} \times \log_2 \frac{5}{14} + \frac{4}{14} \times \log_2 \frac{4}{14} + \frac{5}{14} \times \log_2 \frac{5}{14} \\ &= 1,577 \end{aligned}$$

(b) Temperature

$$\text{splitinfo}_{\text{temperature}}(4|6|4) = 1,362$$

(c) Humidity

$$\text{splitinfo}_{\text{humidity}}(7|7) = 1,000$$

(d) Windy

$$\text{splitinfo}_{\text{windy}}(8|6) = 0,985$$

Schritt 5: Berechnung der Gain ratio für

(a) Outlook

$$\text{Gain ratio}(\text{outlook}) = 0,247 / 1,577 = 0,156$$

(b) Temperature

$$\text{Gain ratio}(\text{temperature}) = 0,029 / 1,362 = 0,021$$

(c) Humidity

$$\text{Gain ratio}(\text{humidity}) = 0,152 / 1 = 0,152$$

(d) Windy

$$\text{Gain ratio}(\text{windy}) = 0,048 / 0,985 = 0,049$$

Die nachfolgende Tabelle soll die Ergebnisse noch einmal übersichtlich zusammenfassen.

Wie aus der Tabelle ersichtlich wird, nimmt das Informationsgewinnverhältnis für das Merkmal „Outlook“ den höchsten Wert an. Somit wird dieses Attribut für eine erste Unterteilung der Trainingsdatenmenge herangezogen.

Outlook		Temperature	
Info:	0,693	Info:	0,911
Gain:	0,247	Gain:	0,029
Split Info:	1,577	Split Info	1,362
Gain ratio:	0,156	Gain ratio:	0,021
Humidity		Windy	
Info:	0,788	Info:	0,892
Gain:	0,152	Gain:	0,048
Split Info:	1,000	Split Info	0,985
Gain ratio:	0,152	Gain ratio:	0,049

Tabelle B.2: Gewinnverhältnissberechnung für die Bäume in Abbildung B.1

Abbildungsverzeichnis

1.1	Closed-Loop Szenario[1]	2
2.1	Beispiel OLAP	8
2.2	Beispiel Dice	9
2.3	RFM	11
5.1	Beispiel CLTV für eine Gastätte	30
5.2	CLTV im SAP Modul	31
B.1	Mögliche Aufteilung des Entscheidungsbaumes für die Wetterdaten	39

Tabellenverzeichnis

4.1	Kundenprofitabilität	25
B.1	Wetterdaten, Quelle: [13, S. 44]	38
B.2	Gewinnverhältnsberechnung für die Bäume in Abbildung B.1 . .	42

Literaturverzeichnis

- [1] A. Duffner; H. Henn: *CRM verstehen nutzen anwenden!*. Max Schimmel Verlag, Würzburg 2001, ISBN 3-920834-92-5
- [2] H. Edelstein: *Building Profitable Customer Relationships With Data Mining*; in PDF-Dokument: <http://www.twocrows.com/crm-dm.pdf>, letzter Abruf: 09.02.2003
- [3] U. M. Fayyad; Gregory Piatetsky-Shapiro; P. Smyth; R. Uthurusamy (1996): *From Data Mining to Knowledge Discovery: An overview*; in: *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Menlo Park, California 1996, S. 1-34
- [4] J. Han; M. Kamber (2001): *Data Mining, Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco u.a. 2001
- [5] J. Mena (2000): *Data Mining und E-Commerce, Wie Sie Ihre Online-Kunden besser kennen lernen und gezielter ansprechen*. Symposion Publishing GmbH, Düsseldorf 2000.
- [6] G. Piatetsky-Shapiro (1993): *Database MiningTM is term trademarked by HNC*; in txt-Dokument, <http://www.kdnuggets.com/news/95/n3.txt>, letzter Abruf: 09.02.2003
- [7] I. Schmidt (2002): *Praktischer Einsatz von Data Mining im Rahmen von analytischem CRM am Beispiel der SAP-Lösung mySAP CRM*. Universität Mannheim, 2002
- [8] F.F. Reichheld; W.E. Sasser, jr. (2000): *Zero Migration: Dienstleister im Sog der Qualitätsrevolution*; in M. Bruhn; Christian Homburg: *Handbuch*

- Kundenbindungsmanagement: Grundlagen, Konzepte, Erfahrungen*. 3. Aufl., Gabler Verlag, Wiesbaden 2000
- [9] Resampling Stats, Inc.: *Association Rules*; in html-Dokument http://www.resample.com/xlminer/help/Assocrules/associationrules_intro.htm, letzter Abruf: 10.02.2003
- [10] J.R. Quinlan (1997): *C4.5: Programs For Machine Learning*. Morgan Kaufmann Publishers, San Mateo, 1997
- [11] SAP-Bibliothek: *Assoziationsanalyse (SAP-Bibliothek - Customer Relationship Management)*; in html-Dokument <http://help.sap-ag.de/sapdocu/crossind/biw/bw30b/helpdata/DE/da/ec293b31de281de10000000a114084/content.htm>, letzter Abruf: 09.02.2003
- [12] SAP-Bibliothek: *Assoziationsanalyse (SAP-Bibliothek - Data Mining)*; in html-Dokument <http://help.sap-ag.de/sapdocu/crossind/biw/bw30a/helpdata/DE/da/ec293b31de281de10000000a114084/content.htm>, letzter Abruf: 09.02.2003
- [13] I.H. Witten; E. Frank (2000): *Data Mining, Praktische Werkzeuge und Techniken für das maschinelle Lernen*. Carl Hanser verlag, München u.a. 2000
- [14] S. Helmke, M. Uebel, W. Dangelmaier: *Effektives Customer Relationship Management*. Gabler, Wiesbaden 2002, ISBN 3-409-21767-3
- [15] *Analyseinstrumente im Database Marketing*; in html-Dokument <http://www.database-marketing.de/dbmanalyse.htm>, letzter Abruf: 09.02.2003
- [16] *Mit den richtigen Kundendaten zum gewünschten Erfolg*; in html-Dokument <http://www.ecin.de/strategie/crmkundendaten/>, letzter Abruf: 09.02.2003
- [17] *Umgekehrtes E-CRM*; in html-Dokument <http://www.symposion.de/ecrm-hb/e-crm-12.htm>, letzter Abruf: 09.02.2003
- [18] *Advanced Customer Data Analysis Services*; in html-Dokument. http://thinklab.telecomitalia.com/data_e.htm, letzter Abruf: 09.02.2003

- [19] *Customer Life Time Value*; in html-Dokument <http://www.abseits.de/cltv.htm>, letzter Abruf: 09.02.2003
- [20] *Database Marketing Institute*; in html-Dokument <http://www.dbmarketing.com/articles/Art104a.htm>, letzter Abruf: 09.02.2003
- [21] *Die analytische CRM-Komponente*; in html-Dokument <http://www.orbis.de/pdf/iconrol.pdf>, letzter Abruf: 09.02.2003
- [22] *Data Warehousing*; in html-Dokument <http://www3.informatik.tu-muenchen.de/lehre/WS2000/DBS-kossmann/folien6.pdf>, letzter Abruf: 09.02.2003
- [23] *OLAP Report*; in html-Dokument <http://www.olapreport.com>, letzter Abruf: 09.02.2003
- [24] *OLAP Council*; in html-Dokument <http://www.olapcouncil.org/>, letzter Abruf: 09.02.2003
- [25] *German OLAP and Data Warehouse Forum*; in html-Dokument <http://olap.winf.ruhr-uni-bochum.de/>, letzter Abruf: 09.02.2003
- [26] *OLAP Knowledge Bases*; in html-Dokument <http://www.compinfo-center.com/entsys/olap.htm>, letzter Abruf: 09.02.2003
- [27] *Customer Lifetime Modeling and its use for Customer Retention Planning*; in PDF-Dokument <http://www-stat.stanford.edu/~saharon/papers/lrv.pdf>, letzter Abruf: 09.02.2003