Data mining: Den Kunden kennen lernen.

Erschienen in: acquisa, Ausgabe 06/2000

Wissen über die Kunden ist in Unternehmen reichlich vorhanden. Dennoch wird es deshalb noch lange nicht genutzt: Denn nur wer die richtigen Fragen stellt, findet mit Software für Datamining und Database-Marketing kluge Antworten.

An Zahlen mangelt es den wenigsten Unternehmen. Gerade im Vertriebsbereich fällt so einiges an: Verkaufszahlen der einzelnen Gebiete, Umsatzprovisionen und alle möglichen Informationen über den Kunden. Dennoch werden selbst recht einfache Fragestellungen nicht mit Hilfe der EDV beantwortet, sondern bleiben dem Gespür des einzelnen Verkäufers überlassen: »Welchen Kunden sollte wann welches Angebot unterbreitet werden?« oder »Bei welchem Kundenprofil lohnt sich ein Außendienstbesuch?« sind Fragen, die sich heute mit Softwareunterstützung beantworten lassen.

Während aber das Sammeln und Archivieren von Transaktionsdaten in vielen Unternehmen schon seit Jahren selbstverständlich ist, beginnen die Firmen erst nach und nach diese Daten auch zu nutzen, um ihre Kunden besser kennen zu lernen. Hauptprobleme sind dabei in der Vergangenheit die Archivierungsart und -weise und die Komplexität und Bedienerunfreundlichkeit vieler Analyseinstrumente.

Ungenutztes Cross-Selling-Potenzial

So stellte es kein Problem dar, die Anzahl oder das Durchschnittsalter neuer Kunden oder Interessenten auszugeben. In den wenigsten Fällen wurden jedoch die Daten zur Beantwortung entscheidungsrelevanter Fragen genutzt. Schon gar nicht, wenn die Fragen etwas komplexer werden: »Wie hoch ist das Cross-Selling-Potenzial für ein neues Produkt?« oder »Wie lassen sich Top-Interessenten mit hohen Lifetime-Values gewinnen?«

Betrachtet man den Charakter solcher Fragestellungen, wird schnell klar, dass sich die Antwort nicht aus einem einzelnen Datenfeld oder Kundenmerkmal ergibt, sondern in der richtigen Kombination unterschiedlicher Kundeninformationen entsteht. So kann beispielsweise die Angebotsaffinität eines Kunden von einer Vielzahl von Merkmalen wie Alter, Geschlecht, Familienstand, demographischen Typologien, bisher gekauften Produkten, gezeigtem Produktinteresse, Zahlungsmoral und einer Reihe weiterer Eigenschaften abhängen.

Mit Datamining lassen sich solche Informationen systematisch verknüpfen. Denn unter Datamining versteht man eine Reihe von Analysen, mit deren Hilfe neue, häufig unerwartete Zusammenhänge festgestellt werden. Der Übergang zwischen klassischen statistischen Methoden und Datamining ist fließend. Ein deutlicher Unterschied liegt allerdings darin, dass Datamining im Gegensatz zur Statistik nicht so viel Wert auf Annahmen beziehungsweise Verteilungsvoraussetzungen legt.

Was ist Datamining wirklich?

Neu am Datamining sind allerdings nicht die Methoden. Neu ist, dass Unternehmen heute über Rechnerleistungen verfügen können, um auf großen Datenbeständen diese Algorithmen in akzeptablen Zeiten berechnen zu lassen. Dabei wird der Begriff derzeit recht großzügig verwendet. So neigen gerade die großen Softwarehersteller dazu, schon die Berechnung eines Mittelwertes als Datamining zu verkaufen.

Beim Datamining unterscheidet man zwei grundsätzliche Vorgehensweisen:
- Überprüfung von Hypothesen über Muster innerhalb der Daten
- Entdecken von bisher unbekannten Mustern oder Regeln in den Daten

Während für die erste Vorgehensweise vorab eine entsprechende Hypothese entwickelt werden muss, versucht man im zweiten Fall neue Muster in den Daten zu finden.

Ein Beispiel für die Validierung von Hypothesen: Ein Hersteller von Automobilzubehör hat folgende Vermutung:
»Junge Autofahrerinnen mit billigen Autos kaufen deutlich weniger Autozubehör als männliche gleichaltrige Autofahrer mit gleichwertigen Autos.« In diesem Fall würden die Umsatz- und Absatzmengenunterschiede in der Produktgruppe Autozubehör zwischen den beiden Gruppen analysiert und die oben genannte Hypothese kann entweder bestätigt oder widerlegt werden.

Ein Beispiel für die Entdeckung von neuen, unbekannten Zusammenhängen kann das Entwickeln von neuen logischen Regeln sein. Als ein Ergebnis könnte sich zum Beispiel folgende Regel aus der Analyse herauskristallisieren:
Nach der Analyse des Datenbestandes (Bondaten eines Supermarktes) ergab sich, dass alle Kunden, die Stangensellerie, Quark, Joghurt und frische Kräuter gekauft haben, auch teuren Rotwein kaufen und insgesamt einen überdurchschnittlich hohen Kassenbon haben. Werden solche Ergebnisse auf das eigene Unternehmen im Marketing, Produktplatzierung, Produktlistung, Einkauf angewandt, nutzt das Unternehmen Database-Marketing.

Im oben genannten Beispiel mit dem Autozubehör könnte man bei Bestätigung der Hypothese sich verstärkt auf die Zielgruppe junger Männer mit billigen Autos konzentrieren, um Streuverluste zu verringern. Das zweite Beispiel hat starken Einfluss auf die strategische Entscheidung, ob in diesem Supermarkt auch weiterhin eine - isoliert betrachtet - defizitäre Gemüsetheke bestehen bleibt. Würde man diese Gemüsetheke schließen, bestände die Gefahr, dass wesentlich mehr Umsatz wegfällt als nur der Umsatz mit Gemüse. Die Analyse hilft in diesem Fall, das Risiko und die Anzahl der betroffenen Kunden/Einkäufe besser abzuschätzen. Eine andere Anwendungsmöglichkeit des Ergebnisses wäre auch, durch gezielte Werbung in dieser Käufergruppe erstens den Rotweinabsatz zu erhöhen und zweitens Kunden mit überdurchschnittlich hohem Kassenbon in den Laden zu locken.

Unverzichtbar zum Datamining dazu gehört das Databasemarketing. Häufig werden die Begriffe auch synonym gebraucht. So war es besonders in der Anfangszeit des Database-Marketing (80ziger und frühe 90ziger Jahre) üblich, nur von Database-Marketing zu sprechen und die Analysen, wie Sie heute unter OLAP und Datamining fallen, als Teil des Database-Marketings zu betrachten.

Heute wird Datamining als ein eigener Prozess betrachtet, der seinen Platz hat innerhalb des Regelkreises: Daten sammeln - Daten analysieren - Daten nutzen.

Dieser Datamining-Prozess lässt sich in sechs Phasen unterteilen.

- Definition der Hypothesen beziehungsweise Wissen um die Geschäftsprozesse, die den zu analysierenden Daten zu Grunde liegen sollen
- Datenextraktion und wenn notwendig Aggregation und Transformation der benötigten Daten des Datawarehouse
- Analyse der Daten, Modellerstellung und –optimierung
- Evaluierung der Analyseergebnisse
- Validierung der Analyseergebnisse
- Anwendung und Umsetzung der Ergebnisse => Database-Marketing

Die Phasen des Dataminings

Während sich die Phasen 2, 3, und 5 heute bequem und schnell mit Hilfe verschiedener Datamining-Tools beziehungsweise statistischer Software-Pakete realisieren lassen, setzen die Phasen 1, 4 und 6 das entsprechende Branchenwissen und Erfahrung voraus. Da die Kombination Analytiker mit entsprechendem Branchenwissen und Erfahrung nicht überall gegeben ist, ist es wichtig, im Team zu arbeiten. Phase 1 ist sehr wichtig, denn auch wenn nach noch unbekannten Mustern in den Daten gesucht wird, muss man dem Datamining-Tool ja sagen, wo und wie gesucht werden soll. Besonders in Phase 4 und 6 profitieren alle am Datamining-Prozess Beteiligten stark von einer grafischen Aufbereitung der Ergebnisse, denn so fällt es leichter, die häufig sehr komplexen Zusammenhänge zu verstehen und daraus die richtigen Strategien und Umsetzungen zu entwickeln.

Heute geht die Entwicklung dahin, dass man Database-Marketing mehr als Anwendung beziehungsweise Strategie auf Basis der durch die Analysen gefundenen Ergebnisse für die jeweiligen Zielgruppen betrachtet, während sich OLAP und Datamining zu eigenständigen Disziplinen entwickelt haben. Diese Entwicklung wurde nicht zuletzt dadurch erleichtert, dass mit den heute in großer Zahl vorhandenen Softwarepaketen zu OLAP und Datamining, die früher in den Firmen mit Database-Marketing üblichen, selbst geschriebenen Anwendungen/Analyseprogramme nach und nach ersetzt werden. Ein zweiter positiver Aspekt der entwickelten Softwarepakete ist, dass die Analytiker und Anwender heute mehr Zeit haben, sich mit den Analysen und ihren Ergebnisse auseinander zu setzen, damit dieses Wissen zur Optimierung der Geschäftsprozesse genutzt werden kann.