

System Center öffnet sich

Mit „System Center“ bietet Microsoft ein umfassendes Portfolio für das IT-Management. Der darin enthaltene Virtual Machine Manager unterstützt demnächst auch VMware.

SEITE 12

ERP-Einführung

Lesen Sie im vierten Teil der CW-Serie „ERP im Mittelstand“, wie sich Fachabteilungen in ein ERP-Projekt einbringen und zu einem gemeinsamen Prozessverständnis kommen können.

SEITE 14

COMPUTERWOCHE.de

CRM-Matchmaker

Suchen Sie Produkte und Dienstleister zur Verwaltung Ihrer Kundenbeziehungen? Der CRM-Matchmaker ermittelt für Sie aus rund 180 Leistungsprofilen die geeignete Lösung oder den passenden Partner.

www.computerwoche/crm-matchmaker.de

COMPUTERWOCHE.de

PRODUKTE & TECHNOLOGIEN

10

COMPUTERWOCHE 22/2008

Keine Angst vor Data Mining

Am Beispiel der Kundenkategorisierung zeigen sich die Stärken und Schwierigkeiten von Data-Mining-Verfahren. Erfolgsfaktoren sind ein solides fachliches Verständnis, saubere Daten, Kontrollen und gute Tools.

VON HILMAR BUCHTA*

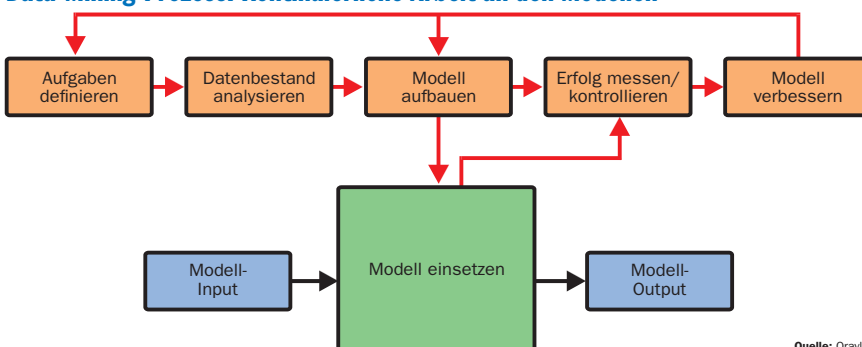
Mit Verfahren des Data Mining lassen sich geschäftsrelevante Muster, Regeln und Auffälligkeiten in Massendaten entdecken. Viele Anwender scheuen aber vor diesem Thema zurück, weil sie es für schwer beherrschbar und komplex halten. Doch die wachsende Verbreitung sollte Mut machen. So ist Data Mining aus den intelligenten Vorschlagslisten bei Web-Shops nicht mehr wegzudenken, und auch im Marketing und Vertrieb lässt sich ein echter Mehrwert erzielen. Ein typisches Beispiel ist die Bewertung des Umsatzpotenzials von Kunden und Interessenten und deren Zuordnung in Segmente oder Kategorien.

Erfolgsfaktoren für Data Mining

- Prozesse und betriebswirtschaftlichen Hintergrund des Problems verstehen;
- Anforderungen der Anwender und Ziel des Modells erfassen;
- Daten sorgfältig auswählen und vorbereiten;
- Qualität unterschiedlicher Mining-Algorithmen vergleichen;
- Verarbeitung automatisieren;
- Modell kontinuierlich kontrollieren und verbessern.

Das Marketing kann dadurch Adressaten für Kampagnen gezielter auswählen und Änderungen im Kaufverhalten schneller erkennen. Für nachfolgende Vertriebsaktionen könnte dann ein Ziel lauten, im nächsten Jahr 20 Prozent der Kunden von der Kategorie „kauft nur sporadisch“ in die bessere Kategorie „Top-Käufer“ zu bringen. Werden Kategorien mit Bezug auf Artikelgruppen gebildet, lassen sich sehr leicht Chancen für Cross-Selling erkennen und beispielsweise durch geschickte Gutschein-Maßnahmen unterstützen. Bei automatisierten Bestellsystemen, wie sie häufig im Internet zu finden sind, könnte umgekehrt ein deut-

Data-Mining-Prozess: Kontinuierliche Arbeit an den Modellen



Die Erstellung von Mining-Modellen (orange) und ihre laufende Verbesserung (blau) erfolgen getrennt.

Quelle: Oraylis

liches Abweichen der aktuellen Bestellung von der Kategoriestufe des Kunden auf einen Betrugsversuch von dritter Seite hindeuten. Diese Bestellung könnte dann manuell geprüft werden.

Allerdings kann eine Kundenkategorisierung, die nur auf tatsächlich getätigten Umsätzen basiert, nur für Bestandskunden festgelegt werden, zum Beispiel anhand von Schwellenwerten. Verlässliche Daten über Neukunden und Interessenten fehlen hingegen. Kunden mit sehr hohem Potential werden so eventuell nicht rechtzeitig erkannt und daher nicht optimal betreut. Im schlimmsten Fall wechseln sie zur Konkurrenz. Um dies zu verhindern, benötigt der Anbieter eine realistische Einschätzung des Potenzials des Kunden. Bei Neukunden kann Data Mining hier die entscheidende Prognose liefern und somit eine frühzeitige Einschätzung ermöglichen.

Mit bekannten Daten trainieren, unbekannte prognostizieren

Voraussetzung hierfür ist natürlich, möglichst viele Daten über Neukunden und Interessenten zur Verfügung zu haben und in die Auswertung einzubringen. Zwar lässt eine solche Datenbasis keine hundertprozentigen Vorhersagen zu, aber selbst eine Wahrscheinlichkeit von 70 Prozent würde dem Marketing schon helfen, sein Budget gezielter einzusetzen.

Im Data Mining gibt es diverse Algorithmen, die sich für die beschriebene Aufgabe eignen und eine oder mehrere Eigenschaften auf der Basis anderer Eigenschaften vorhersagen. Häufige Verfahren sind Entscheidungsbäume, Clustering, Naive Bayes oder eventuell neuronale Netze. Diese Verfahren werden mit vorhandenen und bekannten Daten trainiert, das heißt, sie erkennen selbst die für die Prognose relevantesten Eigenschaften und Muster und leiten dann Regeln für die Vorhersage der Zielgrößen, in diesem Beispiel die Kun-

denkategorie, ab. Nach der Trainingsphase sind die Verfahren in der Lage, auch neue und bislang unbekannte Daten mit diesen Erkenntnissen einer Kundenkategorie zuzuordnen (siehe Tabelle „Training und Vorhersage“). Diese Intelligenz hat Data Mining in der Vergangenheit häufig in die Nähe von Computermagie gerückt. Tatsächlich handelt es sich aber um mathematische und statistische Verfahren wie Häufigkeitsanalysen, Korrelationen und Wahrscheinlichkeitsrechnungen, die dank der modernen Werkzeuge im Bereich Business Intelligence nicht nur für Experten und Visionäre geeignet sind,

Nach sechs Monaten lässt sich der Kunde einordnen

Für das Beispiel Kundenanalyse sei angenommen, dass sich die Kategorie nach etwa sechs Monaten aufgrund der Verkaufszahlen stabil zuordnen lässt. Bei Überschreitung eines bestimmten Umsatzes gilt der Kunde zum Beispiel als „Top-Käufer“. Bei Neukunden und Interessenten muss diese Kategorie prognost-

verwendet beispielsweise diskrete Daten, die eventuell aus den vorhandenen Daten zu transformieren sind. Ferner müssen relationale Daten und Zeitreihen oft noch in ein einfaches Zeilenschema überführt werden, um einen Zusammenhang zwischen Eingabe- und Vorhersagewerten herstellen zu können. Ist anfangs nicht klar, welche Algorithmen sich am besten für das Problem eignen, sollten Anwender verschiedene Algorithmen testen und die Ergebnisse vergleichen.

Das Beispiel setzt voraus, dass die zu prognostizierende Eigenschaft, hier die Kundenkategorie, sich direkt ableiten lässt, für den Fall, dass genügend Daten vorhanden sind. Neben dieser Vorhersage können Data-Mining-Verfahren aber auch mit dem Ziel eingesetzt werden, die Kundenkategorien dann anschließend sprechende Namen wie „Doppeltes Haushaltseinkommen“ oder „Alleinerziehend“ zuzuordnen. In diesem Fall bezieht sich die Prognose darauf, in welches Cluster der Kunde wahrscheinlich fällt.

Wann ändern Kunden ihr Kaufverhalten?

Hat sich der Anwender für ein Modell entschieden, kann es unbetreut und automatisiert weiterverwendet werden. Für die Trainingsphase benutzen wir in diesem Beispiel vorhandene Daten aus einem ERP-System. Für alle Kunden, die bereits mehr als sechs Monate Kunde sind, wird die Kategorisierung zu diesem Zeitpunkt gelesen. Dies setzt aber voraus, dass die Kategorisierung immer wieder protokolliert wird, was schon allein wegen des Nutzens für die Erkennung von Änderungen im Käuferverhalten sinnvoll ist. Nachdem das Modell mit diesen Daten trainiert wurde, kommen die Daten von Neukunden hinzu; dann kann das Mining-Modell zur Vorhersage der

Vorhersagen über Kundenkategorien sind nie hundertprozentig sicher.

ziert werden. Zunächst wird hier ein Mining-Modell definiert. Das Modell legt unter anderem fest, welche Daten als Eingabe verwendet werden können und welche Eigenschaften vorhergesagt werden sollen. Das setzt Kenntnisse der betriebswirtschaftlichen Zusammenhänge des jeweiligen Problems voraus. Die Datenvorbereitung für Data Mining darf dann auch nicht unterschätzt werden, da von ihr letztlich die Möglichkeiten der Mustererkennung und Vorhersage abhängen. So sind bei der Wahl der Datenquellen die Besonderheiten des jeweiligen Mining-Algorithmus zu beachten. Viele Algorithmen

Kundenkategorie genutzt werden.

Eine Möglichkeit, Data-Mining-Verfahren in die IT-Landschaft einzubinden, ist die Extraktion, Transformation und Laden (ETL) von Daten in ein Data Warehouse. Der Ladeprozess kann das Training und die Abfragen in einem Job kombinieren und die Daten vollautomatisch (zum Beispiel nachts) verarbeiten. Die gewünschten Daten stehen danach für das Marketing im Data Warehouse bereit. Auf diesem Weg finden Änderungen im Kaufverhalten schnell einen Niederschlag in den Daten, etwa wenn der Algorithmus festgestellt hat, dass inzwischen andere Eigenschaften für die Vorhersage herangezogen werden müssen. Diese automatischen Modellanpassungen können direkt beim Training des Modells stattfinden. Die vorhergesagte Kategorie schreiben wir zur Unterscheidung von der berechneten Kategorie in ein eigenes Feld, um sie von den „sicheren“ Kategorien abzusetzen.

Ergänzende Kundendaten verbessern die Prognosen

Trotz dieses systematischen Vorgehens sind wirklich gesicherte Prognosen bezüglich des Kaufverhaltens schwierig, solange keine ausreichenden Daten vorliegen. Unternehmen müssen daher alle Möglichkeiten zum Sammeln von Daten ausnutzen. Betreiben sie beispielsweise einen Web-Shop, können sie Daten der Kunden erfassen lassen oder sie durch Gewinnspiele oder Kundenkarten zur Herausgabe weiterer Angaben bewegen. Eine andere Quelle sind Agenturen, die kundenrelevante Informationen verkaufen. Einige für die Prognose benötigte Daten werden sich trotzdem nur schätzen lassen. Dafür gibt es recht gut funktionierende Verfahren, um beispielsweise das Alter aus dem Vornamen abzuleiten oder das Jahresgehalt aus Wohnort und Straße (zum Beispiel über den Mietspiegel). Solche abgeleiteten Kenntnisse sind aber unbedingt in getrennten Feldern zu speichern und müssen vom Mining-Algorithmus analysiert werden, um eventuell noch eine andere Gewichtung vorzunehmen. Und selbst dann werden Unternehmen nie über vollständige Kundendaten und Interessendaten verfügen. Moderne Mining-Algorithmen können aber auch mit Null-Werten umgehen und interpretieren diese korrekterweise so, dass die betreffenden Informationen nicht bekannt sind.

Probleme und Aufwände im Data Mining

Selbst wenn alle Daten vorhanden wären, bleibt die Frage, ob eine Prognose überhaupt möglich ist. Selbstverständlich ist dies nicht: Eventuell gibt es keinerlei Zusammenhang zwischen den erhobenen, den vorliegenden Daten und der zu prognostizierenden Kundenkategorie. Glück-

licherweise geben die Mining-Algorithmen hierzu zumindest eine Rückmeldung. Beim Training wird typischerweise nur ein Teil der vorhandenen Daten genutzt und das Ergebnis sofort auf den übrigen Teil angewendet und überprüft. Zugleich wird diese Kreuzvalidierung gleich mehrfach vorgenommen, um eine Aussage zur Qualität der Vorhersage treffen zu können. Besondere Darstellungsformen wie das Lift-Diagramm können dann die Prognosegenauigkeit eines Algorithmus darstellen.

Ein anderes Problem bei der Verwendung des Data-Mining-Ansatzes zur Vorhersage von Eigenschaften besteht darin, dass es sich um eine sich selbst erfüllende Prophezei handeln könnte. Die Wahrscheinlichkeit einer starken Kundenbindung ist nämlich sicher nicht unabhängig von dem aufgewendeten Marketing-Budget für einen Kunden. Dieser Vorwurf lässt sich zwar nicht komplett entkräften, jedoch kann man das Verfahren gefahrlos zum Beispiel sechs Monate laufen lassen und sich dann anschauen, inwieweit die Vorhersagen tatsächlich eingetreten sind. Wenn sich nach dieser Erprobungsphase eine gute Übereinstimmung offenbart, sind die Ergebnisse künftig für Marketing-

Training und Vorhersage

Alter	Bruttogehalt	Kinder	...	Kundenkategorie	Bekannte Daten (Trainings-Dataset)
36	20-30	Ja	...	A	
39	30-40	Ja	...	B	
...	
20	20-30	Nein	...	??	Neue Daten (Vorhersage)
30	30-40	Nein	...	??	
45	30-40	Ja	...	??	

Für die Bildung von Kategorien wird das Modell zunächst mit bekannten Daten trainiert. Noch unbekannt Kategorien lassen sich später ungefähr vorhersagen.

Zwecke brauchbar. Natürlich entstehen Aufwände bei der Erstellung des Modells und vor allem später bei ihrer weiteren Pflege. Wie dargestellt, lässt sich das Modell beispielsweise in einem ETL-Prozess automatisiert und ohne Benutzereingriffe verwenden.

Regelmäßige Kontrollen und Änderungen des Modells

Trotzdem ist das Modell von Zeit zu Zeit zu überprüfen und an neue Gegebenheiten anzupassen. So könnten sich beispielsweise die Voraussetzungen, die zum Modell geführt haben, geändert haben, oder die Vorhersagequalität ist nicht mehr ausreichend hoch. Vielleicht werden

auch andere Daten für die Vorhersage gebraucht. Anwender sollten zudem regelmäßig prüfen, in welchem Umfang die Vorhersagen tatsächlich wie prognostiziert eingetreten sind, also, ob der Kunde tatsächlich nach den genannten sechs Monaten in der vorhergesagten Kategorie lag. In den meisten Fällen wird jedoch der Nutzen, der unter anderem durch einen gezielteren Einsatz von Marketing-Budgets und durch eine frühe Kundenbindung entsteht, den Aufwand mehr als kompensieren und zu einem erfreulichen Gesamtergebnis führen.

Neben guten und möglichst vollständigen Daten sind leistungsfähige Tools ein Schlüssel

zum Erfolg von Data-Mining-Projekten. Sie sollten sich intuitiv bedienen lassen und erschwinglich sein. Funktional müssen sie mehrere Mining-Verfahren parallel auf ein Problem anwenden können, die Ergebnisse grafisch darstellen und bewerten helfen sowie Auswertungen automatisieren. Häufig ist Data Mining als Zusatzprodukt zu Business-Intelligence-Lösungen erhältlich. Beispiele sind der „DB2 Intelligent Miner“ von IBM oder die integrierten Mining-Funktionen in Microsofts „SQL Server“.

Ausgereiftes Angebot an Werkzeugen für Data Mining

Mit der Funktionsbibliothek „Weka“ ist sogar eine Open-Source-Lösung für Data Mining und seine Automatisierung erhältlich. Die Bedienung ist etwas umständlicher als bei den kommerziellen Werkzeugen. Insgesamt haben die Werkzeuge heute aber einen Stand erreicht, der den Einsatz dieser Technologien beherrschbar und kontrollierbar macht und einen echten Mehrwert besonders für Marketing- und Vertriebsprozesse bietet. (as) ♦



*HILMAR BUCHTA ist Principal Consultant bei der Oraylis GmbH in Düsseldorf.

KLEINE HELFER

Kostenfreie oder preiswerte Soft- und Hardware, die IT-Profis das Leben leichter macht.

Bildschirminhalte einfach aufzeichnen

Sie wollen Ihren Kunden die Vorzüge eines Produkts näherbringen oder Mitarbeitern die Bedienung einer Software erläutern, ohne trockene Texte erstellen zu müssen? Dann greifen Sie zu einer Software wie „Camtasia Studio 5“: Damit lassen sich interaktive Demos und Präsentationsvideos erstellen, bearbeiten und verteilen.

Camtasia wurde entwickelt von dem amerikanischen Anbieter TechSmith, von dem auch das Screen-Grabbing-Tool „Snagit“ stammt. Das Programm erlaubt es dem Anwender, beliebige auf dem Bildschirm angezeigte Inhalte (Videos, aber auch Arbeitsschritte beim Bedienen eines Programms) direkt aufzuzeichnen und nachzubearbeiten. Dabei lässt sich wahlweise der komplette Bildschirminhalt erfassen, ein Anwendungsfenster auswählen oder aber ein individueller Bereich definieren, der von der Software aufgenommen werden soll. Die Bedienung des Tools ist intuitiv und bietet viele

Voreinstellungen, die sich aber anpassen lassen. So lässt sich die Bildrate verändern und die Aufzeichnung des Tons deaktivieren beziehungsweise auf bestimmte Tonquellen beschränken. Außerdem kann der Benutzer auswählen, ob Mauszeigeraktivitäten mit erfasst werden sollen oder nicht. Erwähnenswert sind auch die Möglichkeiten der Nachbearbeitung. So können Teile der Aufzeichnung gelöscht oder Zwischentitel eingefügt werden. Camtasia ermöglicht es dem Anwender auch, in bestimmte Anzeigebereiche hineinzuzoomen oder Hinweise einzubauen, um Aspekte hervorzuheben. Ist das Projekt fertig gestellt, können unterschiedliche Formate gewählt werden, in denen die Präsentation ausgegeben werden kann. Neben dem Camtasia-eigenen Dateiformat lassen sich avi-Files, Quicktime-Filme, Shockwave-Dateien und vieles mehr erzeugen.

Für Testzwecke gibt es eine voll funktionsfähige Demoversion. Die Vollversion von Camtasia kostet 260 Euro.

(Martin Seiler)

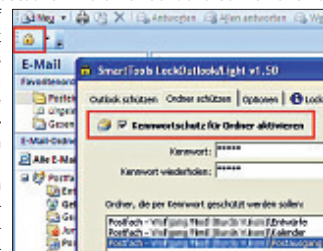
CW-FAZIT: Camtasia ist einfach zu bedienen und bietet umfangreiche Gestaltungsmöglichkeiten für interaktive Präsentationen. Das rechtfertigt den hohen Preis.

Einbruchssicheres Outlook

Ob im Vorstandszimmer, im Planungsbüro oder im Einwohnermeldeamt - Outlook zeigt an vielen Büroarbeitsplätzen von morgens bis abends Präsenz. Die darin gespeicherten Inhalte wie Korrespondenz, Aufgaben oder Kontaktdaten sind nicht selten vertraulicher Natur, die man weder neugierigen Kollegen noch Datenspielen leichtfertig zugänglich machen sollte. Outlook bietet zwar einen Basisschutz, doch der beschränkt sich auf eine optionale Passwortabfrage beim Start. Wer nach einer alltagsstauglichen Sperre sucht, die sich auch für den kurzen Abstecher in die Kaffeeküche schnell aktivieren lässt, sollte einen Blick auf „LockOutlook 1.5“ von Smart Tools Publishing werfen. Das Tool bietet eine Reihe von Schlössern, mit denen sich je nach Bedarf die verschiedensten Zugänge und Winkel von Outlook verriegeln lassen.

LockOutlook kann über ein Schlosssymbol in der Symbolleiste jederzeit aktiviert werden. In den Programmeinstellungen steht eine Reihe von Einstellmöglichkeiten zur Verfügung. Zunächst vergibt der Benutzer ein Kennwort für die Komplettsperre, darüber hinaus können auch gezielt einzelne Ordner gegen unautorisierten Zugriff geschützt werden. Möglich ist auch eine automatische Sperre, die sich nach einer definierten Inaktivitätsphase einschaltet.

Noch granularere Möglichkeiten stellt die Pro-Version bereit: Hier lassen sich wahlweise Unterordner oder sogar einzelne Elemente wie Mails oder Kontakte schützen. Teilen sich mehrere Benutzer eine Outlook-Instanz, können damit außerdem individuelle Beschränkungen je Benutzer vergeben werden. Und sogar Programmbefehle lassen sich hiermit kontrollieren, so dass eine Sekretärin beispielsweise nur Mails weiterleiten darf. LockOutlook 1.5 ist kompatibel mit Outlook 2000 bis 2007 und kann kostenlos unter <http://www.add-in-world.com/katalog/ol-lockoutlook/> heruntergeladen werden. Die Pro-Version ist auf Anfrage erhältlich. (Wolfgang Miedl)



CW-FAZIT: Outlook beherbergt oft sensible Daten. Einen umfassenden Schutz gegen neugierige Kollegen und Datenspielen bietet LockOutlook 1.5.