

# Analytic Intelligence heute

## Data Mining wird mitunter unterschiedlich interpretiert. Wie lautet Ihre Interpretation?

Data Mining ist keine für sich allein stehende Analysetechnik, sondern ein Prozess, in dessen Verlauf durch Auswahl, Untersuchung und Modellierung grosser Datenmengen vorher unbekannte Zusammenhänge entdeckt werden können. Ein solcher Data-Mining-

**GABRIELE DOBENECKER\***

Prozess ist aber keine vollautomatische «Frage-Antwort-Lösung», sondern eine Kombination verschiedener Arbeitsschritte. Im Zentrum des Prozesses steht die Modellbildung: Nach entsprechender Selektion und Aufbereitung von Daten werden Abhängigkeiten der verwendeten Merkmale in einem statistischen Modell abgebildet. Der Blick geht dabei sowohl in die Vergangenheit (erklärende Modelle) als auch in die Zukunft (vorhersagende Modelle).

Der intrinsische Unterschied zu gewöhnlichen Business-Intelligence-Tools besteht darin, dass nicht nur quantitative Aspekte wie «Was? Wie viel? Wie oft?» berücksichtigt werden, sondern auch qualitative Fragestellungen wie «Warum ist etwas so?» oder «Was wird passieren?» behandelt werden können. Deshalb ermöglicht Data Mining nicht nur die Analyse der Vergangenheit, sondern ermöglicht auch einen Blick in die Zukunft. Typische Fragestellungen, die mit Data-Mining-Technologien beantwortet werden können, sind z.B.: «Welcher Kunde wird mit hoher Wahrscheinlichkeit seinen Vertrag kündigen?» und «Wie kann ich mein Kreditrisiko minimieren?»

## Was erfassen bzw. umfassen Ihre Data-Mining-Produkte? Welche Methoden nutzen sie?

SAS Institute bietet mit dem SAS Enterprise Miner eine grafische, intuitive «Point-and-click»-Arbeitsumgebung für die Entwicklung und Durchführung von Data-Mining-Projekten. Die von SAS Institute entwickelte Data-Mining-Methodologie, der SEMMA-Prozess (Stichprobenziehung – Exploration – Modifikation – Modellierung – Auswertung), wird im Enterprise

miner als grafisches Prozessflussdiagramm abgebildet. Die grafische Darstellung ermöglicht ein einfaches Navigieren innerhalb des Prozesses und ein Modifizieren einzelner Arbeitsschritte, wobei die jeweiligen Zwischenergebnisse und Metainformationen automatisch an die nachfolgenden Schritte weitergegeben werden.

Der SAS Enterprise Miner bietet eine grosse Bandbreite an integrierten Modellierungsalgorithmen – u.a. mehrere Varianten von Entscheidungsbaumverfahren, neuronale Netzwerke, lineare und logistische Regressionsanalyse für normal- und nicht normalverteilte Daten, Memory Based Reasoning, Kohonen-Netzwerke und SOMs (Self-organizing Maps), Bagging und Boosting, einfache Implementation zweistufiger Modelle, Clusterverfahren, Zeitreihenanalyse und Assoziationsanalyse.

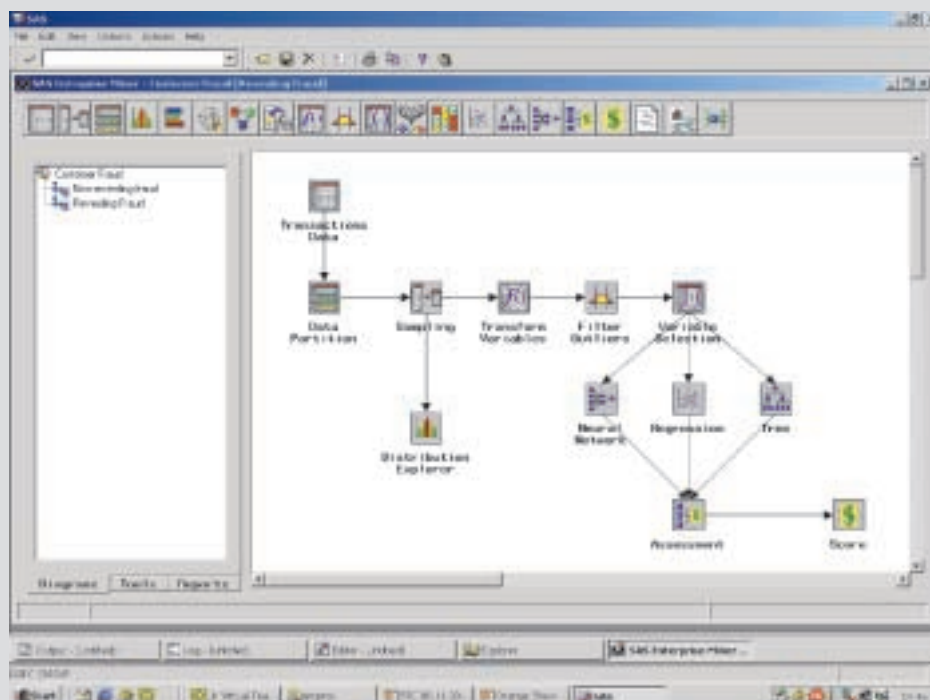
Allen methodischen Ansätzen ist gemeinsam, dass sie simultan eine grosse Anzahl von Einzelmerkmalen (z.B. Alter, Geschlecht, Ausbildung, Stellung, Kaufverhalten) zur Bewertung einer Zielgrösse (z.B. Kaufwahrscheinlichkeit, Stornowahrscheinlichkeit, Betrugswahrscheinlichkeit) verwenden.

Ein wichtiges Merkmal des SAS Enterprise Miner ist sein direkter Zu-

griff auf den Datenbestand im Data Warehouse. Damit profitiert die Software von den Vorteilen der SAS Software: Plattformunabhängigkeit, Flexibilität, leistungsfähige Anwendungen, Skalierbarkeit und durchgehendes Metadatenkonzept. Die Bewertungsmodule sämtlicher Data-Mining-Analysen können durch einen automatisch erzeugten Code direkt im gesamten Data Warehouse angewendet und umgesetzt werden. Das Data-Mining-Konzept von SAS versteht sich damit auch als Teil und Erweiterung von Data Warehousing. Ein Data Warehouse als Zusammenführung themenorientierter, konsolidierter und gereinigter Daten stellt dabei die optimale Basis für Data Mining dar. Denn Data Mining ohne ein Data Warehouse erfordert einen erheblichen Aufwand für das Zusammentragen und Vorbereiten der Daten.

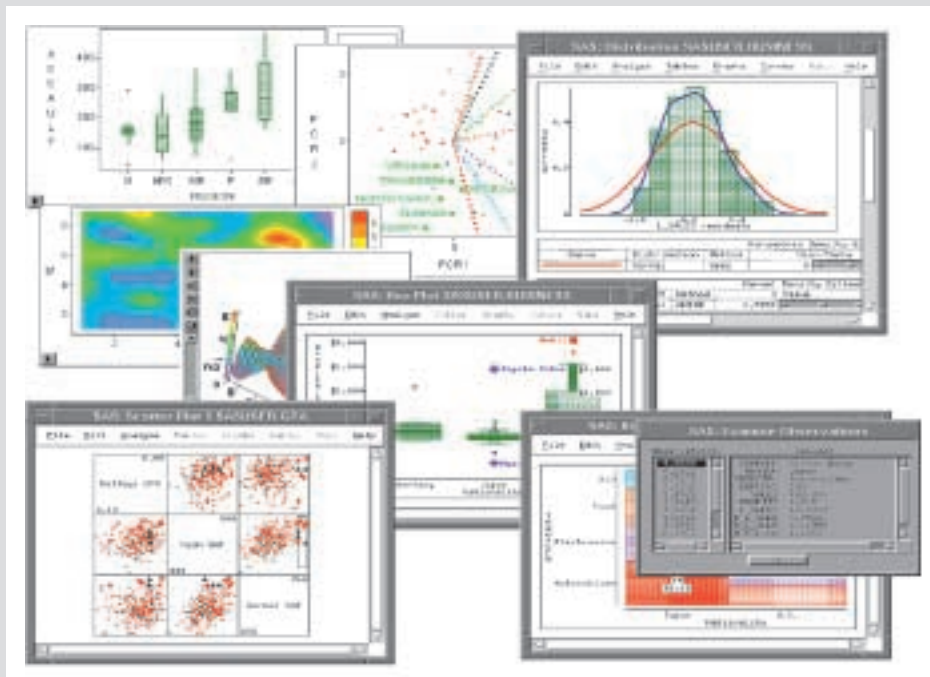
## Welche der Data-Mining-Methoden favorisieren Sie, wenn Sie eine überhaupt bevorzugen – und wenn, warum diese?

Es gibt beim Data Mining keine «beste» Methode. Die Eignung von Methoden ist von der Aufgabenstellung und der Art der verfügbaren Daten abhängig. Geeignete Data-Mining-



Grafische Benutzeroberfläche im SAS Enterprise Miner.

\* Marketing Director, SAS Schweiz.



Dynamische, interaktive, grafische Analyse mit Visualisierungswerkzeugen von SAS.

Verfahren für Kundensegmentierung sind z.B. Clusteranalysen, SOMs und Regressionsverfahren. Mit Hilfe solcher Verfahren können bestehende Kunden in immer neue Gruppen aufgeteilt werden. Demgegenüber eignen sich Assoziationsanalysen insbesondere für so genannte Warenkorb- und Kassenbonanalysen im Handel. Mit Hilfe dieses Verfahrens lassen sich Kaufkorrelationen feststellen, also mit welcher Häufigkeit Kunden bestimmte Produkte oder Produktkombinationen kaufen. Für die Bildung prediktiver Modelle, d.h. Modelle, die es ermöglichen einen Blick, in die Zukunft zu werfen, eignen sich vor allem Entscheidungsbaumverfahren und neuronale Netze.

**Was bietet Ihr Produkt bezüglich Visualisierung – oder falls sie darauf verzichten: Warum tun sie das?**

Mit Hilfe zwei- und dreidimensionaler Grafiken lassen sich Merkmale visualisieren und erste Erkenntnisse über interessante Zusammenhänge in den Daten gewinnen. Zusätzlich ist ein leistungsfähiges Werkzeug für explorative Visualisierung und Analyse im Enterprise Miner integriert. Dieses Werkzeug umfasst Verteilungsanalysen mit Hilfe von Histogrammen, die Visualisierung von Zusammenhängen in Form von Scatter-Plot-Matrizen, Linien-Grafiken sowie drehbare, dreidimensionale Darstellungen u.a.m. Alle diese Werkzeuge sind so miteinander verknüpft, dass Selektionen aus

einem Werkzeug in allen anderen Auswertungen entsprechend sichtbar gemacht werden.

Die Ergebnisse der Modellierungsalgorithmen stehen in tabellarischer und grafischer Form (Pie Chart, Distance Plot, 3D Cluster Profile Chart, Baumdiagramme, Ringdiagramme) für die Interpretation zur Verfügung. Die Qualität von Modellen, die mit unterschiedlichen Modellierungsverfahren und -varianten gewonnen wurden, können Sie im SAS Enterprise Miner mit Hilfe von Lift Charts, Profit Charts, ROI Charts u.v.a.m. vergleichen.

**In welcher Art und Weise nehmen Sie Einfluss auf die User Ihrer Produkte hinsichtlich «Gesetzestreue» – wenn überhaupt?**

In der Regel werden mit SAS Software riesige, von den Unternehmen gesammelte Datenmengen analysiert. Der Grossteil dieser Daten ist jedoch nicht persönlich. Wir nehmen aus diesem Grund keinen Einfluss auf unsere Kunden. In Bezug auf Datenschutz empfehlen wir unseren Kunden folgende Prinzipien: Data Mining ist sowohl für Unternehmen als auch Verbraucher von Nutzen. Der Missbrauch persönlicher Daten ist eine äusserst schlechte Geschäftspraktik, der Verbraucher wird sich von Unternehmen, die seine Privatsphäre nicht achten, schnell abwenden.

Die Erstellung eines Data Marts und das darauf aufsetzende Data Mining bedeuten die Herstellung eines

Verhaltensprofils des Kunden aus singulären Informationen (unternehmensintern wie -extern). Damit ist die Frage nach der Zulässigkeit im Sinne der geltenden Datenschutzrichtlinien zwingend. Generell wird in Unternehmen die Überlassung von Informationen durch die Kunden in den allgemeinen Geschäftsbedingungen geregelt. Die Überlassung erfolgt in der Regel zur ordnungsgemässen Abwicklung der die Kundschaft begründenden Transaktionen. Die Vernetzung singulärer Daten und die Herstellung von Verhaltensprofilen geht klar darüber hinaus – datenschutzrechtlich zulässig ist das besprochene Vorgehen nur dann, wenn den Kunden der Verwendungszweck eindeutig mitgeteilt worden ist.

SAS ist mit 36 Prozent Marktanteil weiterhin die Nummer eins beim Data Mining und bleibt eines der führenden Unternehmen im Gesamtmarkt für Business Intelligence (BI). Dies geht aus dem 2003 erschienenen IDC-Report «Worldwide Business Intelligence Forecast and Analysis, 2003–2007» hervor, der die BI-Anbieter entsprechend ihrer weltweiten Umsätze einordnet.

«Durch die massiven Investitionen in Forschung und Entwicklung kann SAS weiterhin eines der breitgefächertsten BI-Portfolios anbieten und zusammen mit seinen Data Warehousing-Tools und analytischen Anwendungen verkaufen», so Dan Vesset, Research Manager bei IDC für die Bereiche Analytik und Data Warehousing. «Mit seiner langjährigen führenden Position im wachsenden Markt für Data Warehousing ist SAS optimal für die zunehmende Bedeutung der «predictive analysis», also der vorhersagenden Analyse, bei Business Intelligence positioniert.»



SAS Institute AG  
Ruchstückstrasse 6  
8306 Brüttsellen  
Tel: 01 805 74 74  
Fax: 01 805 74 99  
E-Mail: info@sch.sas.com  
www.sas.com/switzerland