

Data Mining - Zukunftsthema im BI Umfeld

Wohin geht die Reise im Bereich Business Intelligence? Die Antwort darauf ist nicht eindeutig. Immer vielschichtiger sind die Angebote, die sich unter dem Schlagwort BI tummeln. Ein Trend ist allerdings absehbar: Immer mehr Business Logik soll von Computersystemen übernommen werden. Dadurch können Entscheidungen des Managements besser vorbereitet werden.

Data Mining kann hier eine entscheidende Rolle spielen. Data Mining (engl. „to mine“ = graben, schürfen) bietet die Möglichkeit, Daten zu klassifizieren und Muster bzw. statistischen Auffälligkeiten zu identifizieren.

Dies geschieht durch die Anwendung von statistischen Verfahren, künstlichen neuronalen Netzen, Fuzzy-Clustering-Verfahren oder auch genetischer Algorithmen.

Das klassische Beispiel einer Data Mining Anwendung ist die Warenkorbanalyse. Hier wird versucht, Abhängigkeiten zwischen gekauften Produkten und ihres Standortes im Supermarkt zu erkennen und Preis bzw. Standort darauf auszurichten. Jedem geläufig dürfte auch die Amazon Kaufempfehlung sein. Hier geht der Umfang der berücksichtigten Daten noch weiter, indem auch Käufe anderer Kunden mit ähnlichem Kaufverhalten einbezogen werden.

Die Einführung einer Data Mining Logik für einen Geschäftsbereich entspricht einem mehrstufigen Prozess. Zunächst werden historische Daten in zwei Bereiche unterteilt, einem kleineren für die Modellbildung und einem größeren für die Verifizierung bzw. dem Training dieses Modells.

Auf der Grundlage des Modelldatenpools wird ein geeignetes Data Mining Modell entwickelt. Dabei spielt die Problemstellung der zu beantworteten Frage eine entscheidende Rolle. Nicht alle Data Mining Modelle beantworten alle Fragestellungen gleichermaßen qualitativ.

So eignet sich das Assoziationsverfahren für eine Warenkorbanalyse, um eine Kundensegmentierung im Marketing vorzunehmen ist ein Clustering die geeignete Methode.

Nach der Entscheidung für ein Verfahren wird das Data Mining Modell erstellt. Hierzu werden vom System automatisch Regeln aus dem Modelldatenpool erstellt.

Anschließend müssen diese Regeln überprüft und verfeinert werden. Dazu verwendet man den Trainingsdatenpool.

Wenn das Modell verifiziert ist, lässt es sich auf neu anfallende Daten anwenden und kann diese entsprechend der Regeln klassifizieren.

Die Aussagen, die eine Data Mining Anwendung trifft, sind immer unscharf. Man erhält je nach Qualität des Modells bzw. des gewählten Verfahrens ein Ergebnis der Form:

„Mit einer Wahrscheinlichkeit von 91,2 % vertreibt dieser Kunde unser Produkt XY“.

Die Modellverwendung lässt sich deutlich weiter fassen, wenn auch Einflussfaktoren außerhalb der historischen Daten Beachtung finden.

Ein Beispiel dafür ist die Berücksichtigung des Geschäftsklimaindex bei der Vertriebsplanung.

Data Mining bietet sehr viele Chancen, zukünftig anfallende Daten in verschiedensten Geschäftsbereichen von Unternehmen auszuwerten und Entscheidungsgrundlagen daraus zu formulieren. Die große Herausforderung besteht darin, geeignete Fragestellungen und Modelle zu entwickeln und deren Ergebnissen bis zu einem gewissen Grad zu vertrauen – letztendlich ist Data Mining schließlich „nur“ Mathematik.

Weitere Info: www.antares-is.de