

Predictive Analytics – mit Prognosen in eine gesicherte Zukunft

Detlef E. Schröder
Oracle Deutschland B.V. & Co. KG
Düsseldorf

Schlüsselworte

Vorhersagen, Entscheidungen, Vorgehensmodell, Entscheidungsgrundlage, Problemstellung, Data Mining Modelle, Datenqualität, Redesign, Entscheidungsunterstützung, Prognose

Einleitung

Die meisten Entscheidungen in Unternehmen werden heute, trotz besserer Informationslage, nicht auf Basis von Prognosen und Vorhersagen getroffen, sondern aus dem Bauch heraus. Hier gilt es noch viele Wettbewerbsvorteile zu heben.

Ist Situation

An vielen Stellen in unserem Alltag leben wir mit Vorhersagen und richten uns danach. Das Wetter zum Beispiel. Wenn in der Wettervorhersage Regen angekündigt ist, nehmen wir einen Schirm mit. Und natürlich auch kennen wir hier die Fehlerhaftigkeit der Vorhersagen und der damit verbundenen Probleme. Wenn es regnet obwohl Sonnenschein angekündigt war, fehlt der Regenschirm natürlich. Die Komplexität der Wettervorhersage ist uns bewusst und doch schauen wir hin.

Wenn es dann um Entscheidungen in den Geschäftsprozessen geht, worauf vertrauen wir dann. Was dient als Entscheidungsgrundlage und mit welchen Informationen sind diese untermauert. Nicht selten ist das Bauchgefühl der sicherste Begleiter und Helfer. Die BI-Systeme und Datenquellen liefern alles Mögliche an Kennzahlen und Berichten. Aber in wie weit ist dies auf die Möglichkeit der Vorhersage hin abgeklopft und mit welchen Qualitäts- und Sicherheitszahlen arbeiten wir?

Noch diffiziler wird es, wenn die Entscheidung in einem Meeting nur von einer Person herbeigeführt wird, nämlich der, die am höchsten in der Hierarchie steht. Dann ist das wie in Abb. 1, das Hippo.

Highest
Paid
Person's
Opinion

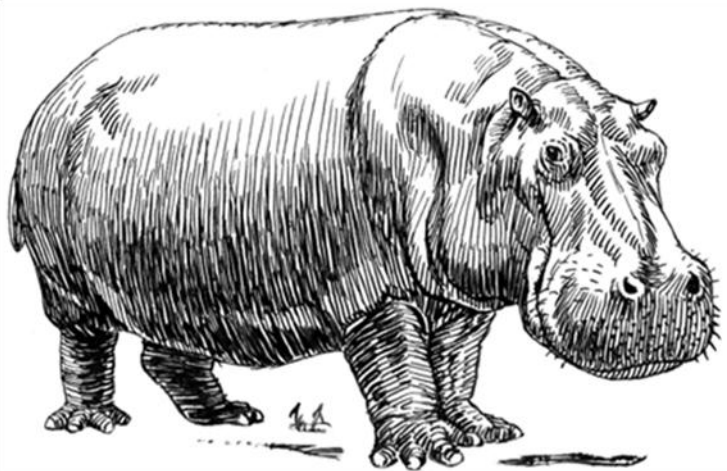


Abb. 1: Das Bauchgefühl des Entscheiders

Vorgehensmodell

Oft liegen viele Daten in den Unternehmen vor und es werden quasi täglich neue Quellen erschlossen und weitere Daten hinzugefügt. Der Ansatz muss allerdings von der Problemstellung oder Fragestellung her aufgerollt werden. Dann ergibt sich folgendes Vorgehensmodell.

Zuerst muss die Fragestellung sauber umrissen und mit den Implikationen auf alle Randbereiche her beleuchtet werden. Dies kann nur von dem Business her geschehen. Was könnte alles der Entscheidung dienen und was würde es bedeuten. Die Problemdefinition sollte so konkret wie möglich sein und noch nicht von den nachfolgenden Phasen limitiert sein.

Er ist dann kann ich mich auf die Suche nach einem oder mehreren Modellen machen, die das Problem erklären und mathematisch berechnen können. An dieser Stelle sollte möglichst der Analyst und das Business zusammen arbeiten, um eine genaue Auswahl treffen zu können und nicht unnötig viele Zyklen zu durchlaufen.

Erst nach dieser Phase kann ich die notwendigen und hinreichenden Daten ermitteln, die für die Modelle zur Verfügung gestellt werden müssen. Hier ist es notwendig, dass der Analyst mit der IT zusammen die Definitionen und Notwendigkeiten abstimmt und die Quellen definiert.

In der Regel werden diese Daten nun noch bereinigt und vervollständigt und dann können die Modelle erstellt werden. Wo und von wem diese Aufgaben erledigt werden, hängt auch von der Herkunft der Daten ab und sollte möglichst in der Datenbank stattfinden, dann können die Aufgaben gebündelt und Datenzentrisch abgearbeitet werden.

Der letzte Schritt ist dann die praktische Umsetzung und Evaluierung des Prozesses und die Überprüfung, ob wirklich die Antwort zur Frage passt und interpretiert werden kann. Dann kann man wieder in den Prozess der Redefinition und Korrektur einsteigen.

Die Abbildung 1 verdeutlicht diesen Prozess.

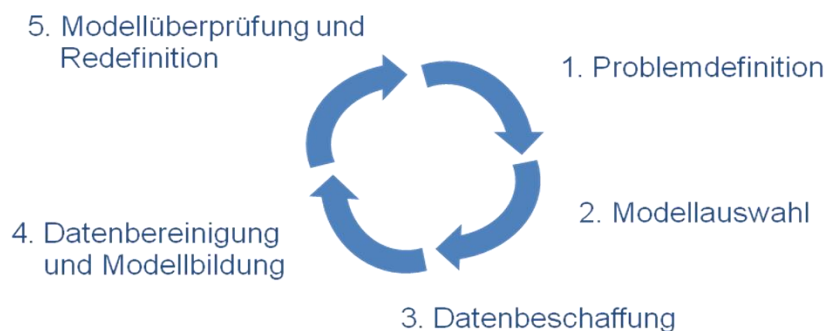


Abb.2: Prognose - Kreislauf

Beispiel – Fälle

Zwei Beispiele sollen dies verdeutlichen

Zuerst der größte Mobiltelefonanbieter in der Türkei. Hier wird oft mit Prepaid Guthaben telefoniert und die Bindung in Verträge ist nicht so, wie zum Beispiel in Deutschland. Das Prepaid Guthaben wird dabei auch als Ersatz für Bargeld verwendet. Es existiert ein reger Handel mit dem Prepaid Vermögen, das im Umlauf ist.

Die Daten speichert Turkcell in einem Datawarehouse mit über einer Petabyte an Rohdaten, verteilt auf sechs Exadata, die das Volumen auf 100TB komprimieren. Das Problem von Turkcell ist der Missbrauch mit Prepaid Guthaben und dies führt zu nichtbezahlten Dienstleistungen. Die überwiegende Anzahl an Transaktionen laufen legal ab und stellen kein Problem da. Die betrügerischen Aktivitäten allerdings belaufen sich auf mehrere Millionen pro Jahr.

Die Muster einer betrügerischen Aktion zu identifizieren und dann im aktuellen Aufkommen zu identifizieren war die Herausforderung, gemessen an der Datenmenge. Turkcell konnte mit den InDatabase Methoden der Oracle Datenbank die Identifikation einer betrügerischen Handlung um Stunden reduzieren und damit viel an potentiellern Verlust verhindern.

Bei der Erledigung einer solchen Aufgabe, geht es nicht nur darum ein wenig Software zu kaufen, nicht einmal nur Hardware, sondern es spielen viel Faktoren zusammen, die zu so einem Erfolg geführt haben, Soft und Hardware war ein wesentlicher davon.

Der zweite Kunde ist aus dem medizinischen Bereich. Die Wirkung von Medikamenten ist von vielen Faktoren abhängig und ist ein komplexes Zusammenspiel auch unterschiedlicher verabreichter Medikamente. Die medizinische Universität von Pittsburgh hat auf über 200 Quellen zurückgegriffen um eine möglichst Breite Datenbasis zu haben, auf der dieses Zusammenspiel umfassend betrachtet werden kann, von der reinen Genetik bis hin zu Proteinen und der Medikamentenhistorie der Patienten.

Mit Hilfe all dieser Daten sollte die Wirksamkeit von Medikamenten untersucht und den Ärzten als Hilfestellung an die Hand gegeben werden. Wenn ein Arzt aus seiner Erfahrung vielleicht auf Hunderte von ähnlichen Fällen zurückgreifen kann, dann kann ein solches System tausende sicherstellen. Jeder Mensch ist anders, aber vielleicht doch an manchen Stellen vergleichbar. Der Verlust, der einer Krankenkasse oder Gesellschaft durch Fehlmedikationen entsteht lässt sich nicht beziffern, aber der potentielle Nutzen ist enorm, der hier mit Datamining - Mitteln erreicht wird.

Software und Integration

Als Basis für die hier angeführten Beispiele diente jeweils die Oracle Datenbank und die darin eingebauten Mining Mechanismen, und diese im Zusammenspiel mit der Mining Software R. Die In Database Mining Umsetzung ermöglicht es nicht nur representative Datenmengen zu untersuchen, sondern auf alle Daten zu greifen, um die Modellbildung und – Anwendung durch zu führen, und das performant.

Des weiteren hilft es die nachfolgenden Systeme mit den Ergebnissen, oder Teilschritten , zu versorgen, das nicht anderes als ein Zugriff auf eine Datenbank notwendig ist, wie ihn die meisten Reporting und Analysesysteme so wie so schin durchführen. De sweiterne wird auch die Datenbewegung minimiert. Es werden keine Daten in eine extra Miningengine kopiert, und damit nicht nur Schnittstellen und Zeit auf sich genommen, sondern auch ein Sicherheitsproblem eliminiert.

Die erläuterte Systemlandschaft wird in der Abbildung 3 noch einmal zusammengefasst.



Abb.3: Analyse - Systemlandschaft

Zusammenfassung

Durch den Einsatz von Prognosen in der Entscheidungsunterstützung können Sie entscheidende Wettbewerbsvorteile erzielen und den Entscheidungsprozess transparenter gestalten.

Oracle bietet dazu die notwendige Plattform an, um die Daten zu sammeln, zu speichern, zu analysieren und auch wieder zur Verfügung zu stellen.

Alles beginnt mit der Identifizierung und Beschreibung des Problems.

Kontaktadresse:

Detlef E. Schröder
Oracle Deutschland B. V. & Co KG
Riesstr. 25
D-80992 München

Telefon: +49 (0) 40-89091 423
E-Mail: detlef.e.schroeder@oracle.com
Internet: www.oracledwh.de