

PPC und Data Mining

Seminar aus Informatik

LV-911.039

Michael Brugger

Fachbereich der Angewandten Informatik
Universität Salzburg

28. Mai 2010

Inhalt

- 1 Literatur
- 2 Data Mining
- 3 Predictive Process Control
- 4 STATISTICA QC Data Miner

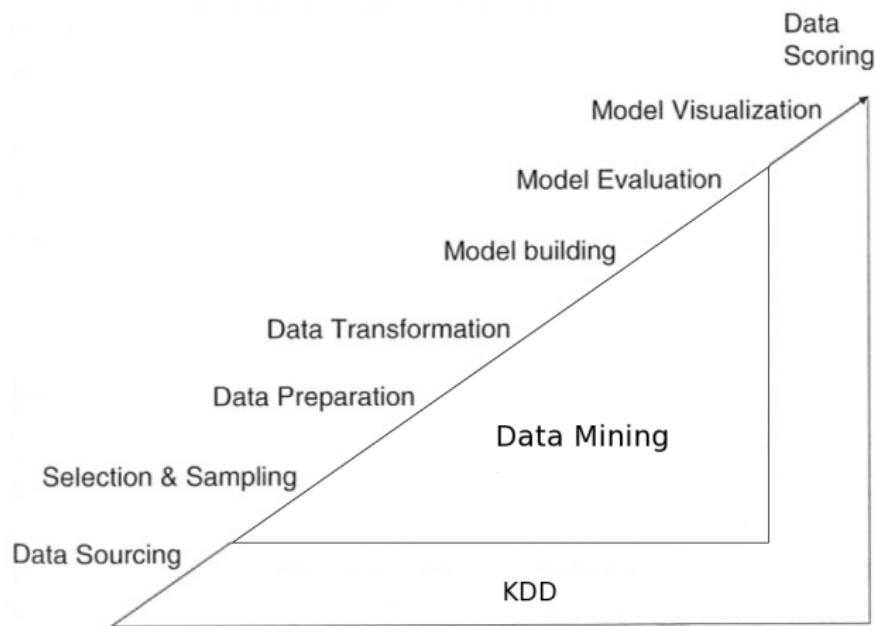
Handbook of Statistical Analysis and Data Mining

... The Handbook of Statistical Analysis and Data Mining Applications is a comprehensive professional reference book that guides business analysts, scientists, engineers and researchers (both academic and industrial) through all stages of data analysis, model building and implementation ...

Begriffsdefinition

- stammt ursprünglich aus dem Bereich der Statistik und kennzeichnet dort die selektive Methodenanwendung zur Bestätigung vorformulierter Hypothesen
- Erforschung und Analyse großer Datenmengen mit automatischen oder halbautomatischen Werkzeugen, um bedeutungsvolle Muster und Regeln aufzufinden (Berry und Linoff)
- Methodik zur Problemlösung um logische oder mathematische, zum Teil komplexe Beschreibungen von Mustern und Regelmäßigkeiten in Datensätzen zu entdecken (Decker und Focardi)

KDD und Data Mining



PPC

... ist ein Ansatz um Variationen und Möglichkeiten zu finden, steuerbare Parameter in einer Produktionsumgebung zu stabilisieren um Qualitätsmerkmale zu halten bzw. verbessern zu können ...

Ziel

- Vorhersagen von Disfunktionalitäten innerhalb eines Prozesses, einer Fertigungsstufe oder einer Produktionsniederlassung
- Zeit- und Geldersparnis durch Verbessern der Produktqualität
- Verringerung der Standzeiten von Produktionseinheiten und dadurch höhere Maschinenauslastung

Problem Definition

- sehr hohe Komplexität
- Entwicklung als langwieriger und experimenteller Prozess
- Punkt, an welchem der Prozess sich zu verschlechtern beginnt
- Unterteilung in beeinflussbare und nicht beeinflussbare Faktoren
- Auffinden von möglichen Variationen und Intervallen in beeinflussbaren Parametersettings und Sollwertvorgaben

Parameter

Unterteilung der Parameter in die Kategorien:

- Prädiktoren, *unabhängige Variablen*
- Zielvariablen, *abhängige Variablen*
- Erzeugte abhängige Variablen

Methoden

Statistische Analysen

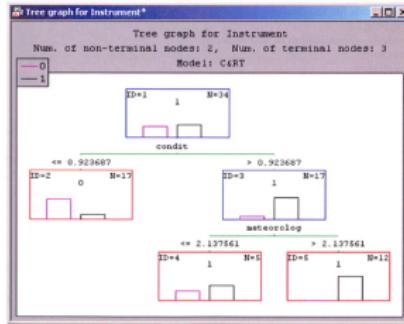
z.B. eine Regressionsanalyse mit dem Ziel Beziehungen zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen festzustellen

$$y = f(x_1, x_2, \dots, x_n) + e$$

Methoden

Statische Analysen

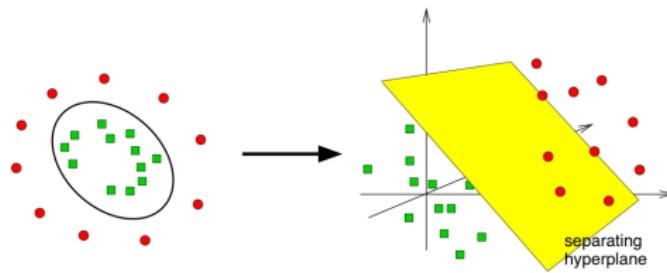
z.B. *Chi-square Automatic Interaction Detector*, eingesetzt bei der Konstruktion von Entscheidungsbäumen. CHAIDs kommen zur Anwendung, wenn eine Aussage über die Abhängigkeit von Variablen gemacht werden muss, dazu wird der Chi-Quadrat-Abstand berechnet



Methoden

Dynamische Analyse

z.B. Support Vector Machine, mathematisches Verfahren der Mustererkennung, eine Klassifizierungsmethode zur Unterteilung von Datenpunkten in 2 Klassen. Dabei wird versucht Datenpunkte durch eine Hyperebene zu trennen



STATISTICA

- Leistungsfähige Plattform, entwickelt für die Analyse von Industriellen Produktionsumgebungen
- Analytische Werkzeuge und Techniken um ein Optimum an Qualität zu erreichen



Vielen Dank!