

Einführung in Methoden für statistisches Data Mining

Zweitägiger Ausbildungskurs

5.+6. Oktober 2005 in Hamburg

Dozent:

Dr. Diego Kuonen

CEO, Statio Consulting, Lausanne, Schweiz

Beschreibung

Dieser Kurs richtet sich an Mitarbeiter in Unternehmen, die gründliche methodische und praktische Kenntnisse über moderne Data-Mining-Techniken erwerben möchten. Diese werden zur Identifikation erwarteter und nicht erwarteter Muster und Trends in Daten eingesetzt. Der Kurs vermittelt auch praktische Erfahrungen mit Data Mining und veranschaulicht Konzepte und Methoden.

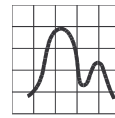
Training

Die Unterweisung reicht von praktischen Beispielen bis zur Theorie, erklärt Gesamtkonzepte, vermittelt aber auch Fähigkeiten zur Bearbeitung spezifischer Aufgaben. So beginnt der Kurs mit einer Diskussion der Rolle und der Eignung von Data Mining, Unternehmen in die Lage zu versetzen, bislang nicht beachtete, wertvolle Informationen aus ihren Datenbeständen zu ziehen. Es folgt ein kurzer Überblick zum Data Mining, der Kunst und Wissenschaft, aus Daten zu lernen. Danach werden einzelne Werkzeuge genauer betrachtet, und es wird erklärt, wie sie sich in das Gesamtbild einfügen. Schwerpunkt des Kurses bildet der softwareunabhängige Überblick der Methoden und der Terminologie statistischen Data Minings. Für alle vorgestellten Techniken wird die grundlegende Methodik erklärt und anhand von Beispielen veranschaulicht.

Über den Dozenten

Der promovierte Statistiker Dr. Diego Kuonen ist Gründer und CEO von Statio Consulting, Lausanne, Schweiz. Statio Consulting ist ein softwareunabhängiges Schweizer Beratungsunternehmen mit Schwerpunkt auf statistische Beratung und Schulung, Datenanalyse, Data Mining, analytisches CRM und Dienstleistungen für Bioinformatik. Dr. Diego Kuonen besitzt mehrjährige Schulungserfahrungen und ist Dozent an den Swiss Federal Institutes of Technology Lausanne und Zürich sowie an anderen Schweizer Universitäten. Derzeit ist er ebenfalls Präsident des Organisationskomitees der Schweizer Statistiktage, Vize-Präsident der Schweizerischen Gesellschaft für Statistik und Präsident der Sektion Statistik in Business und Industrie.

Der Kurs wird in deutscher Sprache abgehalten, alle Präsentationsmaterialien sind in Englisch.



Outline

- Introduction
- What is data mining?
 - Is data mining `statistical déjà vu'?
 - But, why is statistics needed?
 - What distinguishes data mining from statistical analysis?
 - Two analytical approaches to data mining
 - The knowledge discovery in databases process
 - Data mining and business intelligence
- A process model for data mining
- Data and data preprocessing
 - Data sources
 - Why data preprocessing?
 - Major tasks in data preprocessing
- Data mining techniques and tasks
- Description and visualisation
- Characterising multivariate data
- Dissimilarity and distance measures
- Unsupervised methods (class discovery)
 - Principal component analysis
 - Multidimensional scaling
 - Cluster analysis (hierarchical algorithms, partitioning algorithms, using clustering in practise)
 - Kohonen's self-organizing maps
 - Affinity grouping or association rules
 - A look forward
- Supervised methods (class prediction)
 - Introduction
 - Classification modelling (e.g. discriminant analysis, support vector machines, nearest neighbour classification, naïve Bayes classifier)
 - Regression modelling (e.g. multiple linear models, generalised linear models, nonparametric regression models)
 - Neural networks
 - Tree-based methods (e.g. CART, C4.5 and C5.0, CHAID)
 - Ensemble learning (e.g. bagging, arcing, random forests, Bayesian model averaging, boosting)
 - The curse of dimensionality (feature extraction, feature selection, wrappers, embedded methods)
 - Evaluating and comparing classifiers
 - A look forward
- Data mining myths versus realities
- Conclusion
- References and resources

Teilnahmegebühr

1.400 EUR zzgl. MwSt. (1.200 zzgl. MwSt. für jeden weiteren Teilnehmer aus Ihrem Unternehmen).

Anmeldung

unter http://www.statsoft.de/akt_veranstaltung.html