

## Einführung in Methoden für statistisches Data Mining

### Zweitägiger Ausbildungskurs

14.+15. März 2006 in Hamburg

Dozent:

**Dr. Diego Kuonen**

CEO, Stattoo 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 Stattoo Consulting, Lausanne, Schweiz. Stattoo 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 Vize-Präsident der Schweizerischen Gesellschaft für Statistik und Präsident der Sektion Statistik in Business und Industrie.

#### Kurs-Outline

siehe Rückseite

#### Vorkenntnisse

Der Kurs setzt elementare Kenntnisse in Statistik und Erfahrungen mit (multipler linearer) Regression voraus.

#### Teilnahmegebühr

1400 EUR zzgl. MwSt. (1200 zzgl. MwSt. für jeden weiteren Teilnehmer aus Ihrem Unternehmen).

#### Beurteilungen von Teilnehmern dieses Kurses

Kursinhalt gesamt: sehr gut/gut: 100%  
 Wurden Ihre Erwartungen erfüllt: ja: 100%

#### Anmeldung

unter [www.statsoft.de/akt\\_veranstaltung.html](http://www.statsoft.de/akt_veranstaltung.html)

## Outline

- Introduction
- Applicability of data mining
- 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 (*e.g.* data integration, data cleaning, data transformation, data reduction, data discretisation)
- 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 (*e.g.* 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 (*e.g.* inductive bias and model complexity, score functions, internal validation, external validation)
  - 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, generalised additive models)
  - Neural networks
  - Tree-based methods (*e.g.* CART, C4.5 and C5.0, CHAID)
  - Ensemble learning (*e.g.* bagging, subagging, random forests, boosting)
  - The curse of dimensionality (*e.g.* feature extraction, feature selection: filters, wrappers, embedded methods)
  - Evaluating and comparing classifiers
  - A look forward
- Data mining myths versus realities
- Conclusion
- References and resources

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

### Teilnahmebedingungen:

1. Die Anmeldung ist verbindlich.
2. Die Teilnehmer erhalten unmittelbar nach der Anmeldung eine Bestätigung per E-mail zugesandt. Drei Wochen vor dem Seminar teilen wir den genauen Veranstaltungsort mit, senden eine Anfahrtsskizze, Hotelempfehlungen und eine Rechnung.
3. Die Teilnahme am Seminar kann bis zu 3 Wochen vor Kursbeginn abgesagt werden, ohne dass Stornogebühren anfallen. Bei einer späteren Absage oder bei Nichterscheinen fallen die vollen Kursgebühren an.
4. StatSoft behält sich vor, den Kurs nicht durchzuführen, wenn die erforderliche Mindestteilnehmerzahl nicht erreicht wird oder einzelnen Teilnehmern abzusagen, wenn eine Teilnehmerhöchstzahl überschritten wird. StatSoft setzt die Teilnehmer mindestens 10 Tage vor Kursbeginn von einer Absage in Kenntnis. Muss eine Veranstaltung aufgrund höherer Gewalt, Ausfall des Dozenten oder aufgrund eines sonstigen unvorhersehbaren Ereignisses abgesagt werden, so kann dieses auch kurzfristig erfolgen. StatSoft wird in diesen Fällen einen Ersatztermin anbieten oder bereits gezahlte Kursgebühren erstatten.
5. Die Kursgebühr ist sofort nach Rechnungserhalt fällig.
6. Die Kursunterlagen werden den Teilnehmern nur zur persönlichen Verwendung ausgehändigt. Die Vervielfältigung, auch nur von Teilen, ist untersagt.