

## Einführung in Methoden für statistisches Data Mining

### Zweitägige Ausbildungskurse in Hamburg

24.+25. April 2007 in Hamburg

+

27.+28. November 2007

Dozent:

**Dr. Diego Kuonen**

CEO, Statoo Consulting, 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 Statoo Consulting, Schweiz. Statoo 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 mehreren Schweizer Technischen Hochschulen und 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 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%
Fachliche Kompetenz des Dozenten:	sehr gut: 100%
Wurden Ihre Erwartungen erfüllt:	ja: 100%

#### Anmeldung

unter

[www.statsoft.de/anmelden.html](http://www.statsoft.de/anmelden.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, random forests, boosting)
  - The curse of dimensionality (*e.g.* feature extraction, feature selection: filters, wrappers, embedded methods)
  - Evaluating and comparing classifiers
  - Comparing regression models
  - A look forward
  - Recent lessons – what has been learnt?
- Data mining myths *versus* realities
- Criteria for potential data mining success
- Conclusion
- References and resources

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

### Teilnahmebedingungen

1. Die Anmeldung ist verbindlich. Nach Eingang der Kursanmeldung erhält der Teilnehmer eine Anmeldebestätigung.
2. Die Teilnahme kann bis zu 4 Wochen vor Kursbeginn schriftlich abgesagt werden, ohne dass Stornengebühren anfallen. Bei einer Absage bis 2 Wochen vor Schulungsbeginn fallen 50% des Kurspreises als Stornengebühren an. Bei einer späteren Absage oder bei Nichterscheinen fallen die vollen Kursgebühren an.
3. StatSoft behält sich vor, die jeweiligen Kurse nicht durchzuführen, wenn eine 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.
4. Die Kursgebühr ist sofort nach Rechnungserhalt ohne Abzug fällig.
5. Die Auswahl des Kurses liegt in der Verantwortung des Teilnehmers. Die Schulungen werden von qualifiziertem Personal so gestaltet, dass ein durchschnittlicher aufmerksamer Teilnehmer das Schulungsziel erreichen kann. Ein Erfolg wird nicht geschuldet.
6. Kursunterlagen, soweit von StatSoft gestellt, werden den Teilnehmern nur zur persönlichen Verwendung ausgehändigt. Die Urheberrechte liegen bei StatSoft, oder StatSoft steht ein eigenes Nutzungsrecht zur Verwendung und Weitergabe im Rahmen der Schulung zu. Die Vervielfältigung, auch nur von Teilen, ist untersagt.
7. Sofern den Teilnehmern zu Lern- und Übungszwecken Computer zur Verfügung gestellt werden, ist es den Teilnehmern untersagt, Datenträger, welche nicht von StatSoft für die Schulung freigegeben wurden, auf den Computern zu verwenden. Von denen von StatSoft zur Verfügung gestellten Computern dürfen Daten nicht heruntergeladen, kopiert oder auf anderen Datenträgern gespeichert werden. Die Teilnehmer und der Kunde haften gegenüber StatSoft für alle durch eine Zuwiderhandlung entstehenden Schäden und Folgeschäden, auch bezüglich von Ansprüchen die von Dritten gegenüber StatSoft geltend gemacht werden, in unbegrenzter Höhe.