

# Methoden für statistisches Data Mining

**Zweitägiger Ausbildungskurs in Hamburg, Deutschland**

3. + 4. Dezember 2013

**Dozent:**  
**Dr. Diego Kuonen, CStat PStat CSci**  
 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 nicht erwarteter Muster oder Strukturen oder Modelle oder 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 beachtetes, wertvolles Wissen 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, CStat PStat CSci, ist Gründer, CEO und CAO von Statoo Consulting, Schweiz. Er besitzt mehrjährige praktische Erfahrungen mit Data Mining Anwendungen für namhafte Schweizer und internationale Unternehmen. Statoo Consulting ist ein softwareunabhängiges Schweizer Beratungsunternehmen mit Schwerpunkt auf statistischer Beratung und Schulung, Datenanalyse und Data Mining Dienstleistungen. Dr. Diego Kuonen, CStat PStat CSci, besitzt mehrjährige Schulungserfahrungen und ist Dozent an der Universität Genf, Schweiz. Derzeit ist er ebenfalls Präsident der Schweizerischen Gesellschaft für Statistik.

## **Kurs-Outline**

siehe Rückseite

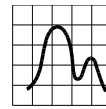
## **Vorkenntnisse**

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

## **Weitere Details**

unter

[www.statsoft.de/mining](http://www.statsoft.de/mining) und/oder [www.statsoft.de/schulungen](http://www.statsoft.de/schulungen)



## Outline

- *Introduction*
- *Applicability of data mining*
- *What is data mining?*
  - *Is data mining "statistical déjà vu"?*
  - *What distinguishes data mining from statistics?*
- *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*
  - *Correspondence analysis*
  - *Cluster analysis (e.g. hierarchical algorithms, partitioning algorithms, using clustering in practise)*
  - *Kohonen's self-organising 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, multivariate adaptive regression splines)*
  - *Neural networks*
  - *Tree-based methods (e.g. CART, C4.5 and C5.0, CHAID)*
  - *Ensemble learning (e.g. bagging, subbagging, random forests, arcing, boosting, stochastic gradient tree boosting)*
  - *The curse of dimensionality (e.g. feature extraction, feature subset selection: filters, wrappers, embedded methods)*
  - *Evaluating and comparing classifiers*
  - *Comparing regression models*
  - *A look forward*
  - *Comparison of chosen supervised learning methods*
  - *Recent lessons – what has been learnt?*
- *Criteria for potential data mining success*
- *Conclusion*
- *References and resources*

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

## Weitere Details

unter

[www.statsoft.de/mining](http://www.statsoft.de/mining) und/oder [www.statsoft.de/schulungen](http://www.statsoft.de/schulungen)