

Berne, Switzerland, May 2007

## Methodological Training in Statistical Data Mining

Three-day training course

**Tuesday, September 4 till Thursday, September 6, 2007**

at Villa Boveri in Baden, Switzerland

given by Dr. Diego Kuonen, Statoo Consulting

Dear Madam, dear Sir,

Data mining technology and methodology have been applied to optimize various processes within business and industry, academia, engineering and government. It is widely believed that data mining will have profound impact on our society and that data mining can bring real value. But how can data mining contribute to achieving operational excellence? Is data mining worth the trouble or is it "statistical *déjà vu*"?

This **three-day training course** will provide you with an overview of the potential and limitations of data mining and a thorough methodological, practical and, most importantly, software-vendor independent coverage of state-of-art data mining techniques. It highlights its applicability to accumulated data, and it will enable you to apply the presented methodology and its underlying philosophy to your own or benchmark data.

You can find a detailed description of the training and a registration form on the attached pages or on our homepage at [www.statoo.com/dm/](http://www.statoo.com/dm/).

We look forward seeing you in Baden.

Please do not hesitate to contact us if you have any questions.

Yours sincerely,



Dr. Diego Kuonen  
CEO, Statoo Consulting

## Methodological Training in Statistical Data Mining

Three-day training course

at Villa Boveri in Baden, Switzerland

given by Dr. Diego Kuonen, CEO, Statoo Consulting

### Description

This training course for professionals will provide you with a thorough methodological and practical coverage of state-of-art data mining techniques that identify expected and unexpected trends in data. This course will provide practical experience with data mining and throughout the course illustrations of the concepts and methods will be given. Moreover, you will be able to apply what you have learned within a state-of-art data-mining workbench using your own or benchmark data.

### Course goals

The naïve and blind use of “black-box” data mining software packages has its obvious pitfalls and can, and probably often does, lead to practically worthless results and misleading conclusions. Data mining is easy to do badly. It is therefore important to understand enough of the characteristics of the underlying data mining methodologies (both their advantages and their pitfalls) to be able to make an informed choice about which data mining method to use and also to be able to critically appraise their own results and those of others. In this course we will apply a “white-box” methodology, which emphasises an understanding of the algorithmic and statistical model structures underlying the “black-box” software.

### Training

Instruction proceeds from tangible examples to theory – from the big picture, or “whole”, to details, or “parts” – and from a conceptual understanding to ability to perform specific statistical data mining tasks.

Consequently, the course begins with a brief discussion of the role and applicability of data mining to empower companies to extract previously unrealised information from their data repositories. Next, a general overview of data mining, the art and science of learning from data, will be given. Only then we do see individual tools in detail and note how they fit into the big picture. As such, in the main part of this training a software-vendor independent overview of the statistical data mining terminology and methods, resources and business issues will be given. For all techniques considered the basic methodology will be explained and illustrated with examples. Finally, the course will enable you to apply the presented methodology and its underlying philosophy to your own or benchmark data.

In summary, this three-day course divides class time between lectures covering, in a software-vendor independent way, the methodological aspects and practical applications of statistical data mining, and between hands-on practice, where you will have a chance to try on your own the methods learned in the course within a state-of-art data mining workbench using your own or benchmark data.

### References

All former participants from companies like **ABB, Alstom, Bayer Consumer Care, Bühler, Daimler Chrysler, Decathlon, John Deere, MAN Turbo, Phonak, PostFinance, Procter & Gamble Manufacturing, Roche Diagnostics, Siemens** or **Total** would recommend this course to others. Based on their feedback we extended the training with representative applications and examples.

## Outline data mining methodology

- 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

## About the speaker

Diego Kuonen, PhD in Statistics, is founder and CEO of Statoo Consulting, Switzerland ([www.statoo.com](http://www.statoo.com)). Statoo Consulting is a software-vendor independent Swiss consulting firm specialized in statistical consulting and training, data analysis, data mining, analytical CRM and bioinformatics services. Currently, Dr. Diego Kuonen is also vice president of the Swiss Statistical Society and president of its section Statistics in Business and Industry.

## Prerequisites

Participants should be familiar with basic statistics, including multiple linear regression.

A laptop with pre-installed *STATISTICA Data Miner* course-license which runs 30 days. StatSoft will provide this license before the course begins.

## Presentation

The lecture will be given, depending on the participants, in English, French or German. During the course questions may be asked in English, French or German. Training documents will be all in English. All participants will receive a printed version of the documentation for personal use.

## Date and hour

Tuesday, September 4 till Thursday, September 6, 2007. Courses start at 08.30 and end at 17.00.

## Place & Accommodation

Villa Boveri, Haus für Kultur, Ländliweg 5, 5400 Baden, Switzerland; see [www.abb-wfs.ch/villa.html](http://www.abb-wfs.ch/villa.html) or [www.abb-wfs.ch/VillaBoveri\\_V3.pdf](http://www.abb-wfs.ch/VillaBoveri_V3.pdf). For hotels in Baden please have a look at [www.baden.ch/hotel/](http://www.baden.ch/hotel/).

## Course fee & discounts

Public course fee	<b>SFR 2'950.00</b> (~ € 1'950.00)
Academic discount	30% off public course fee. No other discounts apply.
Group discount	Group discounts are available if two or more individuals from the same organisation register together and at the same time. Please contact us for further information. No other discounts apply.
Early bird discount	10% off public course fee if you register till <b>June 30, 2007</b> . No other discounts apply.

Prices include printed documentation for personal use and *STATISTICA Data Miner* course-license, which runs 30 days, coffee breaks and lunch but not VAT. All participants will receive an attendance certificate.

## Registration

See separate registration form or [www.statoo.com/dm/](http://www.statoo.com/dm/).

## Contact Information

For further information about the training please contact Cosimo Caforio, phone +41 (0) 56 631 31 10 and fax +41 (0) 56 631 31 12 or email [dm@statoo.com](mailto:dm@statoo.com).

## Registration form for three-day training course Methodological Training in Statistical Data Mining given by Dr. Diego Kuonen, Statoo Consulting

To register please fill out this form completely and fax it to **+41 (0) 56 631 31 12** or register online at [www.statoo.com/dm/](http://www.statoo.com/dm/).

\* Required Information

First Name*	
Last Name*	
Company*	
Department/Function*	
Address*	
ZIP/Postal Code*	
City*	
Country*	
Phone*	
Fax	
Email*	
Date and Signature*	
Comments	

- Tuesday, September 4 till Thursday, September 6, 2007
- Public course fee of SFR 2'950.00** (~ €1'950.00)
- Academic course fee      30% off public course fee. Please attach a copy of your certification. No other discounts apply.
- Early bird discount      10% off public course fee if you register till **June 30, 2007**. No other discounts apply.

### Terms and Conditions

Prices include printed documentation for personal use only, *STATISTICA Data Miner* course license, which runs for 30 days, coffee breaks and lunch but not VAT. Payment of the course registration fee is required prior to the start of the course. Cancellations received in writing more than 30 days before the start of the course will be refunded 100% of the course fee. Cancellations received between 30 and 14 days prior to the course will be refunded 50% of the course fee. We regret that no refunds are allowed for cancellations received within 14 days of the course start date.

### Contact Information

For further information please contact Cosimo Caforio, phone +41 (0) 56 631 31 10 and fax +41 (0) 56 631 31 12 or email [dm@statoo.com](mailto:dm@statoo.com).