

Berne, Switzerland, November 2008

Methodological Training in Statistical Data Mining

Three-day training course

Monday, March 16 till Wednesday, March 18, 2009

or

Monday, September 7 till Wednesday, September 9, 2009

at Villa Boveri in Baden, Switzerland

given by Dr. Diego Kuonen, CStat, Statoo Consulting

Dear Madam or Sir,

data mining technology and methodology have been applied to understand and to optimise various processes within business and industry, academia, engineering and government. It is widely believed that data mining will have a 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 potentials and limitations of data mining and with 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/.

We look forward seeing you in Baden.

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

Yours sincerely



Dr. Diego Kuonen, CStat
CEO, Statoo Consulting

Methodological Training in Statistical Data Mining

Three-day training course

at Villa Boveri in Baden, Switzerland

given by Dr. Diego Kuonen, CStat, Statoo Consulting

Description

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

Course goals

The naïve and blind “black-box” use of 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 methods 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 the 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 practical 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 learnt 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, Barry Callebaut, Bayer Consumer Care, Bühler, CSS, Daimler Chrysler, Decathlon, Helsana, John Deere, Manor, MAN Turbo, Mobiliar, Novelis, Phonak, PostFinance, Procter & Gamble Manufacturing, Roche Diagnostics, Saudi Arabian Oil Company, 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 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
 - 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)
 - Neural networks
 - Tree-based methods (e.g. CART, C4.5 and C5.0, CHAID)
 - Ensemble learning (e.g. bagging, subbagging, random forests, 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
 - Recent lessons – what has been learnt?
 - Comparison of chosen supervised learning methods
- Criteria for potential data mining success
- Conclusion
- References and resources
- Appendix
 - Data mining myths *versus* realities

About the speaker

Diego Kuonen, PhD in Statistics and CStat, is founder and CEO of Statoo Consulting, Switzerland (www.statoo.com). He has extensive experience in applying data mining within large and small companies in Switzerland and throughout Europe. Statoo Consulting is a software-vendor independent Swiss consulting firm specialised in statistical consulting and training, data analysis and data mining 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 only.

Date and hour

Monday, March 16 till Wednesday, March 18, 2009 (**course one**) or Monday, September 7 till Wednesday, September 9, 2009 (**course two**). Courses start at 09.00 and end at 17.30.

Place and accommodation

Villa Boveri, Haus für Kultur, Ländliweg 5, 5400 Baden, Switzerland; see www.villa-boveri.ch. For hotels in Baden please contact us. Our recommendation is the Hotel Blume in Baden; see www.blume-baden.ch.

Course fee and discounts

Public course fee	CHF 3'300.–
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 organization 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 February 3, 2009 (course one) or July 28, 2009 (course two) . No other discounts apply.

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

Registration

See separate registration form or 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.

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

* Required Information

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

- Monday, March 16 till Wednesday, March 18, 2009 (course one)
- Monday, September 7 till Wednesday, September 9, 2009 (course two)
- Public course fee of CHF 3'300.–**
- 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 **February 3, 2009 (course one)** or **July 28, 2009 (course two)**. 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 (not applicable). The number of participants is limited to 20 with a minimum of 5 people. Statoo Consulting reserves the right to cancel a course up to 14 days prior to the course due to insufficient enrolment. 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. Statoo Consulting reserves the right to cancel a course for any reasons beyond its control. Statoo Consulting is not liable for any participants' expenses incurred from cancelled courses.

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.