

Bern, Schweiz, im Dezember 2007

## ***Methodological Training in Statistical Data Mining***

**Montag, 19. Mai bis Mittwoch, 21. Mai 2008**

oder

**Montag, 15. September bis Mittwoch, 17. September 2008**

in der Villa Boveri in Baden, Schweiz

Dozent: Dr. Diego Kuonen, Statoo Consulting

Sehr geehrte Damen und Herren,

Data Mining Techniken und Methoden werden heute in den verschiedensten Unternehmen angewendet, um die unterschiedlichsten Prozesse zu verstehen und zu optimieren. Bekannte Anwendungsgebiete sind

- Forschung & Entwicklung, z. B. neue Möglichkeiten frühzeitig erkennen;
- Produktion, z. B. störende Faktoren in komplexen Fertigungsprozessen aufspüren;
- Analytisches Marketing, z. B. Kundenverhalten voraussagen, Tendenzen im Markt qualifizieren.

Heute gehen führende Analysten davon aus, dass angewandtes Data Mining eine noch stärkere Auswirkung auf unsere Gesellschaft haben wird. Aber wie kann man durch Data Mining einen wesentlichen Beitrag zum Unternehmenserfolg erzielen? Ist es die Mühe wert, oder einfach ein statistisches „*déjà vu*“?

An diesem **dreitägigen Kurs** gehen wir genau auf diese Fragen ein. Konkret werden die Möglichkeiten und Limitationen von Data Mining vermittelt, indem wir auf den Data Mining Prozess und die darin enthaltenen statistischen Data Mining Methoden eingehen. Sie erhalten so eine **softwareunabhängige Übersicht** der zugrunde liegenden Philosophie und Methoden, wie man aus Daten lernen kann.

Eine detaillierte Kursbeschreibung sowie ein Anmeldeformular finden Sie im Anhang oder auf unserer Homepage unter [www.statoo.com/dm/](http://www.statoo.com/dm/).

Über Ihre Teilnahme würden wir uns freuen.

Gerne stehen wir Ihnen bei Fragen und für weitere Informationen zur Verfügung.

Freundliche Grüsse,



Dr. Diego Kuonen

CEO, Statoo Consulting

## ***Methodological Training in Statistical Data Mining***

Dreitägiger Kurs in der Villa Boveri in Baden, Schweiz

Dozent: Dr. Diego Kuonen, Statoo Consulting

### **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 Modelle oder Trends in Daten eingesetzt. Der Kurs vermittelt auch praktische Erfahrungen mit Data Mining und veranschaulicht Konzepte und Methoden.

### **Kursziele**

Der naive und „*black-box*“ Gebrauch von Data Mining Software hat seine offensichtlichen Gefahren und führt häufig zu praktisch wertlosen Resultaten und irreführenden Schlussfolgerungen. Data Mining falsch anzuwenden ist so einfach. Es ist folglich wichtig, genug von den Eigenschaften der zugrunde liegenden Methoden sowie ihre Vorteile und ihre Gefahren zu verstehen, um in der Lage zu sein, eine richtige Wahl bezüglich der angewandten Methoden zu treffen. Mit diesem Kurs vermitteln wir eine Transparenz („*white-box*“) in Sachen Methoden und deren mögliche Anwendungen und somit die Basis für qualitativ hochstehende Auswertungen.

### **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 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. Gleichzeitig wird Ihnen ermöglicht, die vorgestellten Techniken praktisch an Beispieldaten und, im Anschluss zum Kurs, an eigenen Daten anzuwenden. Dieses Konzept mit Fokus auf die Methodik und angereichert durch praktische Beispiele hat sich in früheren Kursen bestens bewährt.

### **Referenzen**

Alle Teilnehmer aus früheren Kursen von Firmen wie **ABB, Alstom, Barry Callebaut, Bayer Consumer Care, Daimler Chrysler, Decathlon, John Deere, MAN Turbo, Mobiliar, Novelis, Phonak, PostFinance, Procter & Gamble Manufacturing, Roche Diagnostics, Saudi Arabian Oil Company, Siemens** oder **Total** empfehlen diesen Kurs weiter. Basierend auf deren Feedback haben wir den Kurs mit weiteren Anwendungen und Beispielen ergänzt.

## Kursinhalt des Methodikteils

- *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-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, 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*

## Über den Dozenten

Der promovierte Statistiker Dr. Diego Kuonen ist Gründer und CEO von Statoo Consulting, Schweiz ([www.statoo.com](http://www.statoo.com)). 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 statistische Beratung und Schulung, Datenanalyse und Data Mining Dienstleistungen. Derzeit ist Dr. Diego Kuonen Vize-Präsident der Schweizerischen Gesellschaft für Statistik und Präsident der Sektion Statistik in Business und Industrie.

## Voraussetzungen

Der Kurs setzt elementare Kenntnisse in Statistik und Erfahrungen mit multipler Regression voraus. Die Teilnehmer müssen einen Laptop mit vorinstallierter *STATISTICA Data Miner* Lizenz mitbringen. Diese Lizenz wird rechtzeitig vor Kursbeginn durch StatSoft zur Verfügung gestellt.

## Präsentation

Das Training wird je nach Teilnehmern in Deutsch, Englisch oder Französisch abgehalten. Während dem Kurs können Fragen in Deutsch, Englisch oder Französisch gestellt werden. Alle **Kursunterlagen sind in Englisch**. Jeder Teilnehmer erhält eine Version der gedruckten Kursunterlagen zur persönlichen Verwendung.

## Datum und Zeit

Montag, 19. Mai bis Mittwoch, 21. Mai 2008 (**Frühjahrskurs**) oder Montag, 15. September bis Mittwoch 17. September 2008 (**Herbstkurs**). Der Kurs beginnt jeweils um 09.00 Uhr und dauert bis 17.30 Uhr.

## Kursort und Unterkunft

Villa Boveri, Haus für Kultur, Ländliweg 5, 5400 Baden, Schweiz; siehe auch [www.villa-boveri.ch](http://www.villa-boveri.ch). Für Übernachtungen in Baden empfehlen wir das Hotel Blume; siehe [www.blume-baden.ch](http://www.blume-baden.ch).

## Teilnahmegebühr

Für Teilnehmer aus Unternehmen	<b>CHF 3'300.–</b> Bitte kontaktieren Sie uns, sollten sich mehrere Teilnehmer aus Ihrem Unternehmen gleichzeitig anmelden wollen.
Hochschulrabatt	<b>30%</b> auf den Normalpreis.
Frühbucherrabatt	<b>10%</b> für Teilnehmer aus Unternehmen, wenn die Anmeldung vor dem <b>24. März 2008 (Frühjahrskurs)</b> oder <b>21. Juli 2008 (Herbstkurs)</b> erfolgt.

Preise beinhalten eine Version der gedruckten Kursunterlagen zur persönlichen Verwendung sowie eine *STATISTICA Data Miner* Kurslizenz, die ab dem Kursbeginn für 30 Tage läuft, sowie alle Pausengetränke und Mittagessen. Preise sind exklusive MwSt. (nicht anwendbar). Jeder Teilnehmer erhält ein Kurszertifikat.

## Anmeldung

Siehe Formular im Anhang oder auf unser Homepage unter [www.statoo.com/dm/](http://www.statoo.com/dm/).

## Kontaktinformationen

Bei Fragen und für weitere Informationen über den Kurs steht Ihnen Cosimo Caforio zur Verfügung: Telefon +41 (0) 56 631 31 10, Fax +41 (0) 56 631 31 12 oder Email [dm@statoo.com](mailto:dm@statoo.com).

## Anmeldeformular für den dreitägigen Kurs

### *Methodological Training in Statistical Data Mining*

Dozent: Dr. Diego Kuonen, Statoo Consulting

Bitte füllen Sie für Ihre Anmeldung folgende Felder vollständig aus und faxen Sie dieses Formular anschliessend an **+41 (0) 56 631 31 12** oder melden Sie sich unter [www.statoo.com/dm/](http://www.statoo.com/dm/) an.

\* Zwingende Angaben

Vorname*	
Nachname*	
Unternehmen*	
Abteilung*	
Adresse*	
PLZ und Ort*	
Land*	
Telefon direkt*	
Telefax	
Email*	
Datum und Unterschrift*	
Bemerkungen	

- Montag, 19. Mai bis Mittwoch, 21. Mai 2008 (Frühjahrskurs)
- Montag, 15. September bis Mittwoch, 17. September 2008 (Herbstkurs)
- Normalpreis für Teilnehmer aus Unternehmen von CHF 3'300.–**
- Hochschulrabatt                      30% auf den Normalpreis. Bitte schicken Sie uns eine Bestätigung.
- Frühbucherrabatt                      10% für Teilnehmer aus Unternehmen, wenn die Anmeldung vor dem **24. März 2008 (Frühjahrskurs)** oder **21. Juli 2008 (Herbstkurs)** erfolgt.

#### Bedingungen

Die Kursgebühr beinhaltet eine Version der gedruckten Kursunterlagen zur persönlichen Verwendung, eine *STATISTICA Data Miner* Kurslizenz, die ab dem Kursbeginn für 30 Tage läuft, Pausenverpflegung und Mittagessen. Preise sind exklusive MwSt (nicht anwendbar). Die Kursgebühr ist vor Kursbeginn zu entrichten. Bei einer schriftlichen Annullierung, die mehr als 30 Tage vor Kursbeginn bei uns eingeht, wird die Kursgebühr vollumfänglich zurückerstattet. Bei einer schriftlichen Annullierung zwischen 30 und 14 Tagen vor Kursbeginn erhalten Sie 50% der Kursgebühren zurück. Wir bedauern, dass wir bei einer kurzfristigen Annullierung (weniger als 14 Tage) keine Kursgebühren zurückerstatten können.

#### Kontaktinformationen

Bei Fragen und für weitere Informationen über den Kurs steht Ihnen Cosimo Caforio zur Verfügung: Telefon +41 (0) 56 631 31 10, Fax +41 (0) 56 631 31 12 oder Email [dm@statoo.com](mailto:dm@statoo.com).