



European  
Six Sigma Club  
Deutschland e. V.

# Modellierung, Visualisierung und Interpretation komplexer Daten

Einführung in Anwendungen moderner Data-Mining-Methoden und -Werkzeuge

Termin: 24. – 26. Juni 2019

Referenten:

Prof. Dr. Maik Kschischo, Hochschule Koblenz - University of Applied Sciences

Björn Noreik, BNB-Qualitätsstatistik und Training

Die Menge an Daten steigt täglich. Nur statistische Methoden erlauben noch, dieser Flut an Information wertvolle Einsichten zu entlocken. Klassische Methoden stoßen dabei schnell an Grenzen. Dieser Workshop richtet sich an alle, die Erfahrung in der statistischen Datenanalyse mitbringen und Interesse an modernen Analyseverfahren haben. Teilnehmerinnen und Teilnehmer dieses Workshops lernen moderne Data-Mining und „machine learning“ Ansätze kennen und können diese Erkenntnisse auf eigene Fragestellungen übertragen.

Teilnahmegebühr: 1950 € (für ESSC-D Mitglieder 1750 €)

Anmeldung per Email: [essc@sixsigmaclub.de](mailto:essc@sixsigmaclub.de) Der Veranstaltungsort ist noch offen.

Analysesoftwarepakete:

Statistiksoftware Minitab®, Salford Predictive Modeler (SPM)®, KNIME® und R-Studio®

(Die vorgestellten und verwendeten Verfahren können auch in anderen Applikationen Verwendung finden)

- Vorstellung der Teilnehmerinnen und Teilnehmer
- Erfahrungsaustausch, Erwartungen und Ziele
  
- Modelle entwickeln mit Minitab (Korrelation, Regression und ANOVA) (*Auffrischung*)
- Bewerten der Modell- und Prognosequalität
- Absichern von Modellen mit Hilfe der Kreuzvalidierung (*Minitab*)
- Einblick Reduktion der Datenkomplexität anhand Multivariater Analysen (*Minitab*)
  
- Industrie 4.0 in Produktionsprozessen – Transformation von Daten in Nutzen
- Typische Datenstrukturen und die sich daraus ergebenden Herausforderungen
- Typische Data-Mining Verfahren, Ansätze und Begriffe
- Clusteranalyse und –prognose (*SPM*)
- Reduktion der Datenkomplexität mit Data-Mining Ansätzen (*SPM*)

- Analyse komplexer Daten mit Hilfe typischer Data-Mining Ansätze (*SPM o. KNIME*)
- Bewerten der Modell- und Prognosequalität
- Absichern von Modellen mit Hilfe der Kreuzvalidierung

Eine Fragestellung 3 Softwarepakete –Analyse eines Datensatzes mit

- SPM (*vorbereitetes Beispiel*)
- KNIME (*vorbereitetes Beispiel*)
- R-Studio (*vorbereitetes Beispiel*)
  
- Gruppenübung zu des Themen des Vortags
- Präsentation der Gruppenergebnisse - Vom Problem zur Empfehlung
  
- Verifizierung und Validierung von Ergebnissen aus dem Data-Mining (*Data-Mining als Prozess*)
- Übung, Besprechung offener Punkte, Erfahrungsaustausch und nächste Schritte

## Einführung in R

- Grundkonzepte
- Datenstrukturen
- Pakete installieren
  
- Visuelle Datenanalyse mit R
- Analyse eines bekannten Datensatzes mit R
  
- Gruppenübung, eigenständige Analyse mit Hilfe
  
- Diskussion
- Optional: Reporting in R und R-Markdown

## Teilnehmerinnen und Teilnehmer

- Regelmäßige Durchführung von statistischer und grafischer Datenanalyse
- Gutes Verständnis für Varianz- und Regressionsanalyse
- Six Sigma Belts und/oder Methodenexperten

## Software: (Vorinstallierte Lizenzen oder Testversionen)

- Microsoft Office®
- Statistiksoftware Minitab®
- Salford Predictive Modeler (SPM)®
- KNIME®
- R-Studio® mit Packages (ggf. vor Ort)