

# Data-Mining Grundlagenworkshop für Six Sigma Belts und Methodenexperten

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

Termine 2021:

16.06.–18.06.2021 in der Nähe von Koblenz *(mit Option auf einen Covid-19 bedingten Wechsel auf ein Online-Training)*

22.11.– 24.11.2021 in der Nähe von Bad Nauheim *(mit Option auf einen Covid-19 bedingten Wechsel auf ein Online-Training)*

Referenten:

Prof. Dr. Maik Kschischo, Hochschule Koblenz -University of applied sciences

Björn Noreik, BNB-Qualitätsstatistik und Training

**Motivation:** Unternehmen treffen regelmäßig datenbasierte Entscheidungen. Bei diesen Daten handelt es sich häufig um Prozess-, Produkt- oder Felddaten. Durch die Digitalisierung ergeben sich hier vollkommen neue Möglichkeiten, um Zusammenhänge zu erkennen, Abläufe zu verstehen und zu steuern und um gezielt Nutzen zu generieren. Dabei werden Experten jedoch immer häufiger mit komplexeren Datenstrukturen konfrontiert, bei denen klassische Vorgehensweisen zur Datenaufbereitung und statistische Standardverfahren häufig ineffizient sind oder sogar versagen. Um es ganz klar zu sagen: Tabellenkalkulationsprogramme sind keine Option.

**Zielgruppe:** Das Angebot richtet sich speziell an Six-Sigma Belts und Methodenexperten zum Beispiel aus dem Qualitätsmanagement, die sich einen Überblick verschaffen möchten über die Begriffe, Softwarepakete, Möglichkeiten sowie Grenzen rund um die Themen Big Data und Data-Mining. Während dieses dreitägigen Workshops werden methodische- und strukturelle Ansätze für die Analyse komplexerer Datenstrukturen vorgestellt, verglichen und diskutiert. Teilnehmerinnen und Teilnehmer sollten in grafischer und statistischer Datenanalyse geübt sein, um den vollen Nutzen aus diesem Workshop zu ziehen.

**Ziel/Nutzen:** Dieser Grundlagen-Workshop bietet die ideale Gelegenheit, gemeinsam in die Themen des Data-Minings einzutauchen und ausgewählte Methoden direkt mit den Software-Paketen Minitab®, SalfordPredictiveModeler(SPM)®, KNIME® und der Programmiersprache R zu erleben, vergleichen und anzuwenden. Der Workshop vermittelt grundlegende Begriffe sowie Methoden zur strukturierten Analyse komplexerer Daten. Als Ergebnis können Sie eigene Analysen komplexer Daten planen und durchführen.

**Hinweis:** Im Vordergrund steht der durchgängige Prozess von der Datenerfassung, über verschiedenste statistische Ansätze zur Lösungsfindung bis hin zur Empfehlung und Absicherung von Maßnahmen. Die Erkenntnisse aus dem Workshop sind genereller Struktur, sodass diese sich anschließend mit verschiedenen Werkzeugen/Softwarepaketen umsetzen lassen. Für diesen Workshop empfehlen wir die Installation der Statistiksoftwarepakete Minitab®, KNIME® sowie R-Studio®.

## **Inhalte** (08:30 Uhr bis ca. 17:00 Uhr)

- Erfahrungsaustausch, Erwartungen und Ziele
- Explorative Datenanalyse und Datenvorbereitung
- Modelle entwickeln mit Minitab (Korrelation, Regression und ANOVA) - Auffrischung *(Minitab)*
- Bewerten der Modell- und Prognosequalität mit Vertrauensbereichen *(Minitab)*
- 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
- Vorstellung der Analysesoftware KNIME®
- Entwicklung unterschiedlicher Analyseprozesse in KNIME®

## **Inhalte** (08:30 Uhr bis ca. 17:00 Uhr)

- Vorstellung typischer Data-Mining Ansätze
- Gerichtete und ungerichtete Analysen sowie Feature Selection(*KNIME*)
- Entwickeln und bewerten von Trainings-und Testdatensatz
- Entwickeln von Modellen und Bewerten der Prognosequalität
  
- Verifizierung und Validierung von Ergebnissen aus dem Data-Mining (*Data-Mining als Prozess mit KNIME und Minitab*)
- Vorstellung der Möglichkeiten mit dem Salford PredictiveModeler(SPM)<sup>®</sup> von Minitab
  
- Gruppenübung
- Präsentation der Gruppenergebnisse -Vom Problem zur Empfehlung
- Besprechung offener Punkte, Erfahrungsaustausch und nächste Schritte

## **Inhalte** *(08:30 Uhr bis ca. 17:00 Uhr)*

- Vorstellung von R und seinen Grundkonzepten
- Warum ist R eine der wichtigsten Sprachen für Data Science?
- Was ist anders in R? Einige Beispiele für die Möglichkeiten und der prinzipielle Ablauf einer Datenanalyse in R
- Grundkonzepte der Programmierung und wesentliche Grundkonzepte der Sprache R zum Mitmachen
- Einstieg in einfache Analysen mit R

## Anmeldung und weitere Informationen

- Über [essc@sixsigmaclub.de](mailto:essc@sixsigmaclub.de)
- Verbindliche Anmeldungen werden erbeten bis 6 Wochen vor dem jeweiligen Workshoptermin
  
- *Teilnahmegebühr, inklusive elektronischer Unterlagen(\*), Erfrischungen und Mittagessen(\*\*)*
- *Für alle drei Tage: 1950.- € (für ESSC-D Mitglieder: 1550.- €).*
- *Werden Sie am besten gleich Vereinsmitglied und profitieren Sie von dem günstigeren Angebot*  
*(\*) Wir arbeiten ressourcenschonend. Gegen einen Aufpreis von 80€ stellen wir Ihnen die Unterlagen*  
*zusätzlich auch gerne gedruckt und als Ordner zur Verfügung.*  
*(\*\*) Im Falle eine Präsenzveranstaltung*

Wir empfehlen folgende Softwarepakete für den Workshop:

- Office®
- Statistiksoftware Minitab®
- Salford PredictiveModeler(SPM)® - *(optional)*
- KNIME®
- R-Studio® inkl. bestimmter Pakete, die wir Ihnen vor dem Workshop noch mitteilen werden

Bitte installieren Sie diese Pakete vorab als lizenzierte- oder Testversionen.

**Hinweis:** Die Installation der oben genannten Pakete ist optional. Teilnehmerinnen und Teilnehmer sind auch eingeladen, bei den Analysen zuzuschauen und aktiv mitzudenken. Wer auch etwas Erfahrung in der Anwendung der unterschiedlichen Softwarepakete sammeln möchte, sollte diese vor dem Workshop jedoch installieren.