R Einführung für Data Scientists

Seminarbroschüre

INFORMATIONSFABRIK
DATEN VERSTEHEN, ENTSCHEIDUNGEN TREFFEN

Bereich: Data Science

Seminar: Einführung R für Data Scientists

Dauer: 3 Tage

Beschreibung des Seminarinhalts

R ist die am häufigsten verwendete Programmiersprache, wenn es um Statistik und Data Science geht. R wird häufig mit dem R Studio als Entwicklungsumgebung (IDE) genutzt. Die Sprache bietet zahlreiche Möglichkeiten Daten zu visualisieren und zu analysieren.

In dieser Schulung lernen Sie das Potenzial der Sprache und der IDE zu nutzen.

Zunächst vermitteln wir Ihnen die grundlegenden Funktionen von R. Im zweiten Teil wird dieses Wissen dann anhand einer konkreten Data Science Aufgabe gefestigt und weiter ausgebaut. Am Ende halten sie mit dem erstellten R Skripten Ihren ersten Schlüssel zu einer erfolgreichen Datenanalyse in den Händen.

Inhalt

- Allgemeine Grundlagen der Skriptsprache R
  - Datentypen,
  - Objekte,
  - Funktionale Programmierung
- Umsetzung eines exemplarischen Data Science Prozesses in R
  - Visualisierung + Vorverarbeitung der Daten
  - Modellierung mittels Machine Learning Algorithmen
  - Nutzung der Modelle für Predictions

Zielgruppe

Angehende Data Scientists, Business Analysten, BI Experten

Voraussetzungen

Programmiererfahrung in einer anderen Sprache

Ziele des Seminars

Nach dem Seminar

- sind Sie in der Lage R als Programmiersprache und R Studio für Ihre Projekte zu nutzen
- kennen Sie die grundlegenden Konzepte, Funktionen und Strukturen von R
- wissen Sie, wie Sie Data Science Aufgaben mit R lösen können

Inhalt

Tag 1

R Einführung

- Datentypen und Objekte in R
- Subsetting
- Kontroll-Strukturen
- Funktionen
- Scoping Rules/ Umgebung (Environment)
- Debugging Tools
- Style Guide

Tag 2 und 3

Data Science mit R

- Data Science Prozess in R
- Daten Selektion
  - Einlesen von Flatfiles
  - Datenbankanbindung
- Daten Vorverarbeitung
  - Explorative Datenanalyse
    - Plots (Säulendiagramme, Boxplots, Histogramme)
    - Exportieren in PDFs
    - Duplikate, Fehlende Werte und statistische Aufbereitung der Daten
  - Datenbereinigung
- Modellierungsprozess
  - Machine Learning Algorithmen
  - Evaluierungskriterien + Kreuzvalidierung
- Modellnutzung
  - Exportieren, Laden und Scoring von Daten
- Ausblick und weitere Möglichkeiten in R
  - Weitere Kriterien zur Messung der Modellgüte: ROC, AUC, Lift
  - Entscheidungsbaumvisualisierung
  - Feature Importance