

Datenanalyse, Data Science und Machine Learning mit R

Datenanalysen (data.table), Visualisierung (ggplot2) und Umsetzung von Machine Learning-Algorithmen mit R

Kurzbeschreibung

Die Programmiersprache R ist mit über zwei Millionen Nutzern weltweit eine der meist-verwendeten Programmiersprachen im Bereich Statistik und Data Science. Verglichen mit anderen Programmiersprachen (Java, C, C#,...) ist R beim PYPL Index auf Platz 7 der weltweit meist genutzten Programmiersprachen. Viele Softwareanbieter bieten Schnittstellen zu R, als eine der „de-facto“ Data Science-Programmiersprachen an. R wird oft mit RStudio als Entwicklungsumgebung (IDE) genutzt. Die Sprache bietet verschiedene Möglichkeiten Daten einzulesen, zu visualisieren und anhand von Machine Learning Algorithmen zu analysieren.

Was lernen Sie in diesem Seminar?

Das zweitägige Seminar bietet eine Erweiterung Ihrer Programmierkenntnisse in R im Bereich data science, um Datenanalysen mit data.table durchzuführen, eigene Grafiken mit ggplot2 zu erstellen und Machine Learning Algorithmen in R umzusetzen. Konkrete Inhalte sind:

Inhalt

- Einführung in das Data Science Paket data.table
- Weiterführende Anwendungen eines data.tables (lapply im data.table und der := Operator)
- Grafiken erstellen mit dem Paket ggplot2
- Eigene Funktionen schreiben und Control Flows verwenden
- Einführung in Machine Learning
- Detaillierte Beschreibung von Entscheidungsbaum und Logistischer Regression und Umsetzung mit R
- Überblick über weitere Machine Learning Algorithmen: SVM, Random Forest, k-means und Umsetzung mit R
- Overfitting, Train-Test-Split und Cross Validation in Theorie und Umsetzung mit R

An wen richtet sich das Seminar?

Das Seminar richtet sich an Datenanalysten, data scientists und interessierte R-Benutzer, die ihre Kenntnisse in den oben genannten Bereichen erweitern wollen. Erste Erfahrungen mit R sind notwendig, um mit den Grundlagen von R sicher umgehen zu können. Teilnehmer sollten über folgende Erfahrung verfügen: Pakete installieren und laden, Funktionen aufrufen, einen Wert einer Variable zuweisen, erste einfache Plots erstellen, Statistiken wie min / max / mean berechnen, einen Vektor mit c() erzeugen.

Grundlegende Vorkenntnisse im Bereich der Statistik (Begriffsdefinitionen wie bspw. Mittelwert, Median, Standardabweichung, Quantil, Dichtefunktion, Normalverteilung) sind sehr hilfreich für das Verständnis. Kenntnisse grundlegender mathematischer Symbole und Begriff (Summenzeichen, Integral, Funktion, Ableitung, Menge der natürlichen und reellen Zahlen, Vektor, Matrix) sind hilfreich für das Verständnis. Kenntnis der booleschen Algebra mit den logischen Operatoren (UND, ODER, NICHT) sind von Vorteil.

Didaktischer Aufbau des Seminars?

Dieses Seminar ist sehr praxisorientiert und das Erlernen und Coden der Inhalte in R steht im Vordergrund. Die Teilnehmer arbeiten direkt und selbstständig mit der Programmiersprache R und der Entwicklungsumgebung RStudio, so dass das Erlernete direkt geübt und vertieft werden kann. Der Trainer moderiert dabei verschiedene Aufgaben und begleitet die Teilnehmer durch die einzelnen Lerneinheiten.

Was ist an Technik mitzubringen?

- Die Übungen finden auf Laptops statt. Es wird empfohlen, dass Teilnehmer einen eigenen Laptop mit der vorab installierten Software (R und RStudio) mitbringen. Die Bitkom Akademie kann auf Anfrage auch Schulungslaptops bereitstellen. Die Teilnehmer müssen prüfen, ob firmeneigene Laptops Zugangsbeschränkungen beim Herunterladen bzw. Installieren von zusätzlichen Paketen aus dem Internet haben.
- Die digitalen Unterlagen (Skript, Code, Dateien) werden über eine Cloud zur Verfügung gestellt. Die Einwahl in fremde WLAN-Netze sollte daher möglich sein.
- Im Idealfall ist der USB Port der Laptops freigeschaltet, so dass als Backup Pakete, verwendete Daten oder sonstige Unterlagen per USB-Stick übertragen werden können. Bitte stellen Sie ggfalls sicher, dass Ihr Gerät über USB Ports oder entsprechende Adapter verfügt.
- Bitte stellen Sie sicher, dass Sie mit dem Betriebssystem (Windows, Mac OS, Linux), mit dem Sie arbeiten, vertraut sind.

Buchung des Seminars in Kombination mit „Data Science mit Python“ (optional)

Die Bitkom Akademie bietet auch ein Seminar zur Programmiersprache **Python** (dreitägig) an, welches zusätzlich zu den in R behandelten Themen eine Einführung in die Programmiersprache Python gibt und daher auch für Teilnehmer geeignet ist, welche keine Erfahrung mit Python, aber generell Erfahrung mit einer Programmiersprache haben. Weiter Informationen finden Sie [hier](#).

Kontaktieren Sie uns – wir beraten Sie gern.

Bitkom Akademie | Albrechtstraße 10 | 10117 Berlin
T 030 27576-540 | info@bitkom-akademie.de
Weitere Seminare finden Sie unter www.bitkom-akademie.de

Seminarprogramm Tag 1 – Einführung in R

Uhrzeit	Agenda
09.00-09.30	Akkreditierung und Ausgabe der Seminarunterlagen
09.30-10.00	Begrüßung durch den Seminarleiter <ul style="list-style-type: none">▪ Vorstellungsrunde & Erwartungshaltung der Teilnehmer
10.30-11.30	HANDS-ON: Data.table Paket – Einführung <ul style="list-style-type: none">▪ Einführung in die Besonderheit vom data.table Paket und Ähnlichkeit zu SQL-Abfragen▪ Erzeugen eines data.table▪ Möglichkeiten, Zeilen und Spalten zu extrahieren▪ Berechnungen direkt auf Spalten durchführen▪ Berechnungen nach Variablen gruppieren
11.30-11.45	Kaffeepause mit Gelegenheit zum Erfahrungsaustausch und Networking
11.45-13.15	HANDS-ON: Datenmanipulation <ul style="list-style-type: none">▪ Manipulationen auf einem data.table bei Zeilen und Spalten▪ Zusammenfassen und Filtern von Daten▪ Variablen erstellen, löschen, ändern▪ Fehlende Werte ersetzen▪ lapply() und die Anwendung in einem data.table (mit .SD und .SDcols)
13.15-14.15	Mittagspause
14.15-15.45	HANDS-ON: Datenvisualisierung mit ggplot2 <ul style="list-style-type: none">▪ Das Konzept hinter Grammar of Graphics▪ Die Layer von ggplot2 zur Erstellung erster Plots und zum Plotten von Statistiken▪ Darstellungen (Größe der Datenpunkte, Farbe, Gruppierung) mit einer Variable variieren oder festsetzen▪ Mehrere Subplots erstellen, Plots anpassen und abspeichern▪ Erstellen von Histogrammen, Boxplots, Scatterplots, Dichtefunktion
15.45-16.00	Kaffeepause mit Gelegenheit zum Erfahrungsaustausch und Networking
16.00-17.30	HANDS-ON: Control Flows <ul style="list-style-type: none">▪ Eine eigene Funktion schreiben▪ Default Parameter in einer Funktion setzen▪ For Schleifen, If-Else Bedingungen und While Schleife
17.30	Ende des ersten Seminartags

Kontaktieren Sie uns – wir beraten Sie gern.

Bitkom Akademie | Albrechtstraße 10 | 10117 Berlin
T 030 27576-540 | info@bitkom-akademie.de
Weitere Seminare finden Sie unter www.bitkom-akademie.de

bitkom
akademie

Seminarprogramm Tag 2 – Anwendung von R

Uhrzeit	Agenda
08.30-09.00	Begrüßung durch den Seminarleiter und Rückblick auf Tag 1
09.00-10.30	Überblick über Machine Learning <ul style="list-style-type: none">▪ Einführung in Machine Learning▪ Unterschied Supervised – Unsupervised Learning (überwachtes – unüberwachtes Lernen)▪ Overfitting, Train-Test-Split und cross-validation (Kreuzvalidierung)
10.30-10.45	Kaffeepause mit Gelegenheit zum Erfahrungsaustausch und Networking
10.45-12.15	HANDS-ON: Entscheidungsbaum in R <ul style="list-style-type: none">▪ Big Data und die vier Vs von Big data▪ Aufteilung der Daten in Test-Train, ein Model erstellen und validieren▪ Grundlagen des Algorithmus▪ Umsetzung in R▪ Validieren der Ergebnisse (u.a. confusion matrix, sensitivity, accuracy)▪ Anpassen von Hyperparametern im Training▪ cross-validation (Kreuzvalidierung)
12.15-13.15	Mittagspause
13.15-14.45	HANDS-ON: Logistische Regression in R <ul style="list-style-type: none">▪ Aufteilung der Daten in Test-Train, ein Model erstellen und validieren▪ Grundlagen des Algorithmus▪ Umsetzung in R▪ Validieren der Ergebnisse (u.a. AIC; BIC; confusion matrix, sensitivity, accuracy)▪ ROC curve und AUC▪ Interpretation der Ergebnisse▪ cross-validation (Kreuzvalidierung)
14.45-15.00	Kaffeepause mit Gelegenheit zum Erfahrungsaustausch und Networking
15.00-16.30	HANDS-ON: Machine Learning Algorithmen in R <ul style="list-style-type: none">▪ Überblick der Algorithmen Support Vector Machine (SVM), Random Forest und K-means▪ Umsetzung der Algorithmen in R▪ Validieren der Ergebnisse
16.30	Ende des Seminars

Kontaktieren Sie uns – wir beraten Sie gern.

Bitkom Akademie | Albrechtstraße 10 | 10117 Berlin
T 030 27576-540 | info@bitkom-akademie.de
Weitere Seminare finden Sie unter www.bitkom-akademie.de

bitkom
akademie

Ihr Referent



Jan Köhler

**Technischer Projektleiter, deep learning
Bosch Center for Artificial Intelligence
Robert Bosch GmbH**

Jan Köhler arbeitet und forscht derzeit an den neuesten Technologien im Bereich Data Science, Machine Learning und Deep Learning im Bosch Center for Artificial Intelligence in Stuttgart. Zuvor unterstützte er über viele Jahre als Data Scientist unterschiedliche Geschäftsbereiche bei den verschiedenen Stufen eines Data Mining Projekts von der Beratung bis zur Implementierung. Die technische Umsetzung erfolgte mit R oder Python. Seit 2012 ist er als Trainer im Bereich Data Science tätig und schult verschiedene Ebenen, vom Projektmitarbeiter bis zum Vorstand.

Preise

1.190 €* Regulär

990 €* für Bitkom-Mitglieder

**Die angegebenen Preise sind in Netto-Beträgen ausgewiesen.*

Termine und Veranstaltungsorte

Die Termine entnehmen Sie bitte der Website der Bitkom Akademie. [hier](#) ↗

Kontaktieren Sie uns – wir beraten Sie gern.

Bitkom Akademie | Albrechtstraße 10 | 10117 Berlin
T 030 27576-540 | info@bitkom-akademie.de
Weitere Seminare finden Sie unter www.bitkom-akademie.de

bitkom
akademie