



# Data Science mit Python

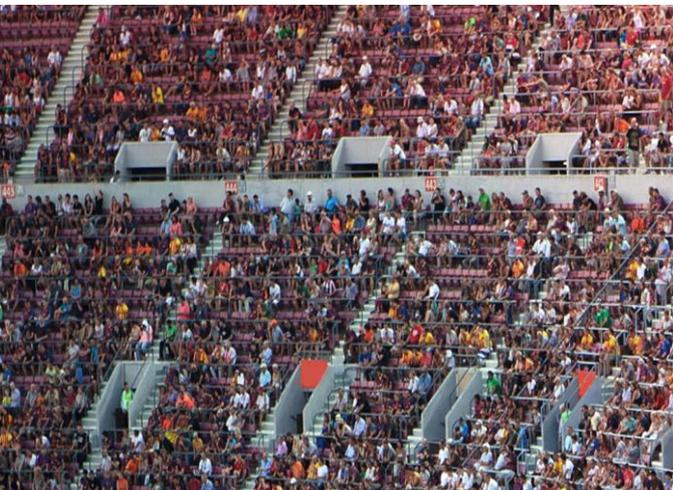
Datenanalyse, Visualisierung und erste Machine  
Learning Algorithmen

# Data Science mit Python

Datenanalyse, Visualisierung und erste Machine Learning Algorithmen

## Kurzbeschreibung

Python ist eine Programmiersprache, die auch im Bereich Data Science etabliert ist. Aufgrund ihrer klaren Syntax und leicht verständlichen Struktur eignet sie sich insbesondere für Einsteiger in Data Science, die bereits mit den Grundlagen des Programmierens vertraut sind. Bei richtiger Anwendung macht Python das Analysieren großer Datenmengen einfach, effizient und kann Unternehmen rasch neue Erkenntnisse aus dem eigenen Datenbestand verschaffen. Darüber hinaus ist Python plattformunabhängig – es läuft unter Linux, Mac und Windows ebenso wie unter Unix und vielen Unix-Systemen.



## Inhalt

- Grundlagen von Python
- Einführung in das Data Science Paket pandas und dem DataFrame
- Daten aus flat files (csv, txt, excel) einlesen und schreiben
- Daten anpassen, konvertieren, modellieren
- Visualisierung von Daten mit seaborn
- Erste Supervised Algorithmen aus dem Machine Learning
- Clustering Algorithmen (K-Means, DBSCAN)

## Was lernen Sie in diesem Lehrgang?

Das dreitägige Seminar schafft einen kompakten und umfangreichen Einstieg in die Data Science- und Programmiersprache Python. Die Teilnehmenden sind anschließend in der Lage, Python selbstständig für Datenanalysen im Unternehmen zu nutzen. Das Seminar umfasst hierbei eine Einführung in die Datenstruktur DataFrame vom Paket pandas (effizient Daten anpassen und verändern), Daten aus flat files (csv) einlesen bzw. schreiben, Visualisierung mit matplotlib und seaborn, erste eigene Funktionen schreiben und enthält einen Einstieg in Machine Learning mit den Algorithmen Entscheidungsbaum, K-Nearest Neighbor, einem einfachen Neuronales Netz und den Cluster Verfahren K-Means und DBSCAN.

# Data Science mit Python

Datenanalyse, Visualisierung und erste Machine Learning Algorithmen

## **An wen richtet sich dieser Lehrgang?**

Das Seminar richtet sich an angehende Data Scientists, BI-Analysten und Datenanalysten und an der Programmierung in Python interessierte Fachkräfte bzw. Projektleiter, welche schon erste Programmiererfahrung in Python haben und die Grundlagen der Programmiersprache Python für Data Science-Projekte erlernen möchten, um damit eigenständig an data mining Projekten mitzuwirken oder Python Code besser verstehen zu können. Die Teilnehmer benötigen keine Erfahrung mit Python, sollten aber bereits Erfahrung mit einer Programmiersprache (z.B. VBA, Java, C, R, etc.) gemacht haben, um das Grundkonzept einer Programmiersprache zu verstehen (Variable in einer Programmiersprache, Zuweisung von Werten zu einer Variablen, Aufrufen von Funktionen, Parameter einer Funktion).

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

## **Didaktischer Aufbau des Lehrgangs?**

Dieses Seminar ist sehr praxisorientiert. Die Teilnehmer arbeiten direkt und selbstständig mit der Programmiersprache Python in der Entwicklungsumgebung Jupyter Lab, so dass das Erlernte direkt geübt und vertieft werden kann. Der Trainer moderiert dabei verschiedene Aufgaben und begleitet die Teilnehmer durch die einzelnen Lehreinheiten.

## **Was ist an Technik zu organisieren?**

- Die Übungen finden auf Laptops statt. Es wird empfohlen, dass Teilnehmer einen eigenen Laptop mit der vorab installierten Software bereitstellen. Die Bitkom Akademie kann für die Präsenzs Schulungen jedoch auf Anfrage auch eigene Schulungslaptops bereitstellen.
- Die Teilnehmer sollten prüfen, ob firmeneigene Laptops Zugangsbeschränkungen beim Herunterladen von Dateien und Programmen haben, die der Referent bereitstellt. Die digitalen Unterlagen (Skript, Code, Dateien) werden über eine Cloud zur Verfügung gestellt. Die Einwahl in fremde WLAN-Netze sollte daher möglich sein.
- Die Beispiele werden mit dem Windows Betriebssystem umgesetzt. Wird ein anderes Betriebssystem verwendet (z.B. MacOS), sollte der Umgang vertraut sein.

# Data Science mit Python

Datenanalyse, Visualisierung und erste Machine Learning Algorithmen



## Zusatzinformationen

- Das Seminar findet in einer kleinen Gruppe mit mind. 5 und max. 12 Teilnehmern statt. Unser Referent kann dadurch auf individuelle Fragestellungen besser eingehen.
- Bitte stellen Sie sicher, dass Sie mit dem Betriebssystem (Windows, Mac OS, Linux), mit dem Sie arbeiten, vertraut sind.
- Die Bitkom Akademie ist [anerkannter Bildungsträger in Baden-Württemberg](#) und [Nordrhein-Westfalen](#). Teilnehmer haben im Rahmen des Bildungszeitgesetzes die Möglichkeit, Bildungsurlaub bzw. eine Bildungsfreistellung zu beantragen. Auf Anfrage erstellen wir auch Anträge auf Anerkennung unserer Veranstaltungen in anderen Bundesländern.
- Lunch und Getränke sind im Seminarpreis enthalten.
- Tipp: Nutzen Sie für Ihre Anreise zu unseren Akademie-Seminaren die [Sonderkonditionen unserer Partner](#).

# Seminarprogramm

## Data Science mit Python

### Begrüßung durch den Seminarleiter

- Vorstellungsrunde & Erwartungshaltung der Teilnehmer

### Hintergrund von Python

- Anwendungsmöglichkeiten von Python
- Installieren von Paketen
- Die Entwicklungsumgebung Jupyter Lab
- Aspekte von Python, welche von anderen Programmiersprachen abweichen

### Einführung in das Data Science Paket pandas

- Elemente eines pandas data.frame
- Eine Zeile oder Spalte auswählen, hinzufügen und verändern
- Boolean indexing
- Ein leeres data.frame erstellen

### Mittagspause

### Grundlegende Statistiken mit pandas

- Statistiken mit einem data.frame berechnen (Anzahl an Beobachtungen, Summe, Mittelwert, Median, Minimum, Maximum, Varianz, ...)
- Methoden im data.frame, um einen Überblick der Daten zu erhalten
- Kreuztabelle (Kontingenztafel)
- Auswertungen nach einer Variable gruppieren
- Fehlende Werte löschen oder ergänzen

### Daten einlesen

- Das Arbeitsverzeichnis setzen
- Eine CSV Datei einlesen/schreiben
- Überblick über nützliche Parameter beim Lesen und Schreiben
- Große Dateien mit dem Paket pandas einlesen
- Das pickle modul zum Speichern von Python Objekten

### Ende des ersten Seminartages

Tag  
1

# Seminarprogramm

## Data Science mit Python

### Begrüßung durch den Seminarleiter und Rückblick auf Tag 1

#### Control Flows

- Die range() Funktion
- Eine eigene Funktion schreiben und default Parameter setzen
- For Schleifen, If-Else Bedingungen
- Logische Vergleichsoperatoren

#### Visualisierung mit matplotlib und pandas

- Die Hauptelemente beim Plotten
- Einen Plot anpassen (x- und y-Achse ändern, Beschriftungen, Legende und Titel) und speichern
- Auswahl von Farben
- Erstellen von Scatterplot, Linienplot, Barplot, (gruppiertes) Histogram, Boxplot

#### Mittagspause

#### String Manipulation mit pandas; Einführung numpy

- Information aus Text in einem DataFrame extrahieren
- Text verändern und anpassen
- Grundlagen von numpy, das die Basis von pandas ist

#### Machine Learning

- Begriffserklärungen (Machine Learning, Deep Learning, Künstliche Intelligenz)
- Einteilung von Machine Learning Algorithmen (Supervised vs. Unsupervised Learning)
- Train-Test Split der Daten
- Schritte beim Trainieren eines Machine Learning Algorithmus
- Overfitting
- One-Hot Encoding

#### Ende des zweiten Seminartages

Tag  
2

# Seminarprogramm

## Data Science mit Python

### Begrüßung durch den Seminarleiter und Rückblick auf Tag 2

#### Entscheidungsbaum

- Einführung in den Algorithmus
- Einen Entscheidungsbaum in Python umsetzen
- Erste Ergebnisse validieren

#### K-Nearest Neighbor und Neuronale Netze

- K-Nearest Neighbor in Python trainieren
- Ein einfaches Neuronales Netz in scikit-learn umsetzen
- Ergebnisse validieren

#### Mittagspause

#### K-Means Clustering

- Einführung in den Algorithmus
- Den K-Means Algorithmus in Python umsetzen
- Güte von Cluster Ergebnissen einschätzen

#### DBSCAN

- Einführung in den Cluster Algorithmus
- DBSCAN mit scikit-learn in Python umsetzen
- Ergebnisse validieren

#### Ende des Seminars

Tag  
3

# Ihr Referent



## Jan Köhler

**Founder, Data Scientist, Deep Learning Scientist  
Enable AI und Bosch Center for Artificial Intelligence,  
Robert Bosch GmbH**

Jan Köhler ist Gründer von EnableAI und ist dort als Dozent und Berater im Bereich Künstlicher Intelligenz (Machine Learning, Deep Learning, Data Science) tätig. Er arbeitet seit über 10 Jahren an den neuesten Technologien in Machine Learning und Deep Learning in der zentralen Forschung der Robert Bosch GmbH, zuletzt im Bosch Center for Artificial Intelligence in Stuttgart. Wissenschaftlich hat er bei vielfachen Veröffentlichungen in internationalen Top-Konferenzen mitgewirkt und hat zu über 35 Patentanmeldungen, meist als Haupterfinder, beigetragen. Jan Köhler hat einen Master in Statistik und ein Diplom im Bereich Wirtschaftsingenieurwesen.

## Shortfacts



### **Termine, Preise und Veranstaltungsorte**

Bitte entnehmen Sie aktuelle Informationen hierzu Website der [Bitkom Akademie](#).

**Kontaktieren Sie uns – wir beraten Sie gern.**

Bitkom Akademie | Albrechtstraße 10 | 10117 Berlin  
T 030 27576-540 | [info@bitkom-akademie.de](mailto:info@bitkom-akademie.de)  
Weitere Seminare finden Sie unter [www.bitkom-akademie.de](http://www.bitkom-akademie.de)

**bitkom**  
akademie