

Data Science in Python







Schulung der Grundlagen von
Machine Learning mit pandas
und scikit-learn



Data Science in Python

Grundlagen von Machine Learning und scikit-learn

Über den Kurs

-  **Dauer:** 3 Tage
-  **Gruppengröße:** 3-10
-  **Level:** Anfänger mit Programmiererfahrung*
-  **Anteil Coding:** 60%
-  **Sprache:** Python
-  **Bibliotheken:** pandas, matplotlib, seaborn, scikit-learn

Auf einen Blick

- ✓ Wichtige Datenstrukturen
- ✓ pandas Data.Frame
- ✓ Statistiken berechnen
- ✓ Eine eigene Funktion schreiben
- ✓ Grafiken mit seaborn erstellen
- ✓ Daten einlesen und schreiben (csv, xls, pickle)
- ✓ Machine Learning Einführung
- ✓ Lineare Regression, Entscheidungsbaum, K-Means
- ✓ Train-Test-Split, Kreuzvalidierung

Überblick über das Seminar

Die dreitägige Einführungs-Schulung für die Verwendung von Python im Data Science/Data Mining erklärt die notwendigen Grundlagen von Python für die Datenanalyse und die Anwendung von Machine Learning Algorithmen. Bei den Algorithmen wird nur das notwendige theoretische Verständnis geschult, da das Ziel auf der praktischen Umsetzung der Algorithmen in Python liegt. Am Ende des Seminars können Sie selbstständig erste Daten Analysen umsetzen und einfache Machine Learning Algorithmen für die Datenanalyse einsetzen.

Das Paket pandas ist ein Schwerpunkt, da pandas speziell für Data Science entwickelt wurde. Die wichtigsten Schritte für die Datenaufbereitung werden eingeübt. Zur Erstellung von Plots und Grafiken wird das Paket seaborn verwendet mit einer kurzen Einführung in matplotlib. Matplotlib überzeugt durch die Fülle an Möglichkeiten einen Plot anzupassen, seaborn durch die Einfachheit auch komplexere Plots mit wenig Code zu erstellen.

Nach den Grundlagen für Python (Datenstrukturen, eigene Funktionen schreiben) und der Erläuterung von pandas für die Auswertung von Daten, erhalten Sie einen Überblick über Machine Learning Algorithmen, welche wir in Python mit dem Paket scikit-learn selbst coden werden. Ein wichtiger Bestandteil des Seminars ist das eigenständige Arbeiten und Lösen von Übungsaufgaben, so dass Sie mit direkter Hilfe des Trainers das Besprochene direkt in der Praxis umsetzen und anwenden können.

Inhalte des Seminars

Sie erhalten einen Überblick über die im Science, Data Mining, Machine Learning und Deep Learning populärste Programmiersprache Python. Wir verwenden die Anaconda Distribution (nach eigenen Angaben "The World's most popular data science platform) und als Entwicklungsumgebung/IDE wird spyder verwendet. Nach dem Seminar können Sie Daten aus verschiedenen Formaten einlesen, Daten mit seaborn/matplotlib plotten bzw. Daten mit pandas bereinigen (fehlende Werte ersetzen, Zeilen und Spalten anpassen, neue Spalten erzeugen).

Sie kennen die wichtigsten Datentypen in Python, können eigene einfache Funktionen schreiben und kennen die Umsetzung von Control Flows (For-Schleife, If-Else). Sie verstehen das Grundkonzept eines pandas DataFrames und können damit Data Wrangling und Data Cleaning durchführen.

* Programmierkenntnisse in Python sind nicht notwendig. Erste Erfahrung mit einer anderen Sprache (z.B. R, VBA, C) ist ausreichend.

Die Einteilung von Machine Learning in supervised-unsupervised und Reinforcement Learning ist Ihnen bekannt und Sie können mit scikit-learn eigenständig Algorithmen in Python trainieren, validieren, einen Train-Test Datensplit durchführen und Gütekriterien zur Beurteilung eines Algorithmus berechnen und interpretieren. Bekannte Algorithmen im Machine Learning werden erläutert und Sie können die verschiedenen Algorithmen verstehen und in Python coden. Das Seminar umfasst Algorithmen für die Regression, Klassifikation und dem Clustering: Lineare Regression, Entscheidungsbaum, Random Forest, k-means clustering.

Die Einstiegshürde für die Benutzung von Python für Machine Learning/Data Science/Data Mining/Business Intelligence ist genommen, so dass Sie eigenständig Ihr Wissen nach dem Seminar erweitern können. Der Schwerpunkt liegt auf der eigenen Umsetzung auf Ihrem Laptop mit Unterstützung des Dozenten.

Wer sollte teilnehmen?

Dieser Kurs mit Python richtet sich an data scientists, angehende Machine Learning engineers, Datenanalysten, Business Intelligence Analysts, Data Analysts, o.ä. welche die Programmiersprache Python für Data Science/Data Mining kennenlernen möchten, um Datenanalysen unter Verwendung von Machine Learning Algorithmen eigenständig umsetzen

Voraussetzungen

Dieser Python Kurs setzt keine grundlegenden Kenntnisse in Python voraus. Notwendig ist jedoch Vorerfahrung mit einer Programmiersprache, damit Konzepte einer Variablen, Zuweisung von Werten zu einer Variablen, eine Funktion bzw. eine for-Schleife bekannt sind.

Notwendig sind außerdem grundlegende Vorkenntnisse im Bereich der Statistik (Begriffsdefinitionen wie bspw. Mittelwert, Median, Standardabweichung, Normalverteilung), Kenntnisse grundlegender mathematischer Symbole und Begriffe (Summenzeichen, Integral, Funktion, Ableitung, Exponentialfunktion) und Kenntnis der booleschen Algebra mit den logischen Operatoren (UND, ODER, NICHT) sind sehr empfohlen.

Die Teilnehmer sollten Vorerfahrung mit dem Umgang von Daten haben, z.B. in Excel oder einer BI-Software, damit Konzepte einer spaltenweisen Berechnung bzw. einfache Statistiken (Mittelwert, Varianz) bekannt sind.

Das Seminar wird auf Deutsch gehalten. Englischkenntnisse (lediglich im Verstehen von englischen Texten) sind sehr empfehlenswert, da die Programmiersprache, Fachbegriffe und die Dokumentationen im Internet auf Englisch sind. Aus diesem Grund sind auch die erstellten Folien in der Schulung auf Englisch.

Methode des Seminars



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

Technische Voraussetzungen der Teilnehmer (Laptop, etc.)

- ❑ Die Teilnehmer benötigen für die Übungsaufgaben Laptops. Wir empfehlen, Ihren eigenen Laptop mit der vorab installierten Software mitzubringen. Eine genaue Installationsanleitung für die Software wird Ihnen vor dem Seminar per E-mail zugesandt. Auf Anfrage stellen wir auch Schulungslaptops zur Verfügung.
- ❑ Bitte prüfen Sie, ob Ihr Firmenlaptop Zugangsbeschränkungen im Internet hat. Die digitalen Unterlagen (Skript, Code, Dateien) werden im Seminar online zum Download zur Verfügung gestellt. Sie erhalten vor dem Seminar per E-Mail den Link zu einer Testdatei zum Download, um dies überprüfen zu können.
- ❑ Sie sollten sich in firmenfremde WLAN-Netze registrieren können.
- ❑ Als Backup Lösung ist es möglich, dass der USB Port bei Ihrem Laptop freigeschaltet ist, um damit verwendete Dateien oder sonstige Unterlagen übertragen zu können.
- ❑ Im Seminar wird das Betriebssystem Windows verwendet. Der Umgang mit Ihrem verwendeten Betriebssystem und Laptop sollte bekannt sein. Insbesondere sollten Sie ohne Schwierigkeiten Sonderzeichen auf der Tastatur finden (insbesondere bei Apple Geräten werden auf manchen Tastaturen nicht immer runde, eckige bzw. geschweifte Klammern dargestellt).

10.00-10.15

Begrüßung und Organisatorisches

- Vorstellungsrunde
- Erwartungen der Teilnehmer

10.15-11.45

Grundlagen von Python

- Neue Pakete installieren
- Spyder als Entwicklungsumgebung/IDE
- Funktionen und Methoden
- Wichtige Aspekte von Python im Vergleich zu anderen Programmiersprachen

11.45-12.00

Kaffeepause

12.00-13.30

Grundlegende Datenstrukturen

- Überblick über die basic data types (string, integer, float, NaN)
- Erläuterung der wichtigsten Datenstrukturen: list, tuple, dictionary
- List comprehension

13.30-14.30

Mittagspause

14.30-16.00

Das Paket pandas – Data.Frame

- Struktur (Zeilen, Spalten) eines DataFrames
- Auswahl einer Zeile/Spalte
- Zeilen/Spalten erstellen, löschen, ändern
- Boolean indexing: eine logische Abfrage zur Selektion von Zeilen
- Daten zusammenfassen, um einen Überblick zu erhalten.

16.00-16.15

Kaffeepause

16.15-18.00

Berechnen von Statistiken direkt im pandas Data.Frame

- Einfache Statistiken direkt auf einem Data.Frame (Mittelwert, Min, Max, Summe, Median, Varianz,...)
- Zusammenfassen und Filtern von Daten
- Fehlende Werte ersetzen
- Kreuztabelle (Kontingenztafel)

18.00

Ende

09.00-09.15

Rückblick und offene Fragen von Tag 1

09.15-10.45

Control Flows

- Eine eigene Funktion schreiben
- Default Parameter in einer Funktion setzen: position arguments und keyword arguments
- For Schleifen
- If-Else Bedingungen
- List comprehension mit if-else

10.45-11.00

Kaffeepause

11.00-12.30

Datenvisualisierung mit seaborn / matplotlib

- Wichtige Grundlagen von matplotlib
- Achsen-Beschriftung, Legende, Titel ändern
- Einen Plot speichern
- In seaborn Linienplot, Boxplot, Histogram, Scatterplot, Barplot erstellen
- Darstellungen (Größe der Datenpunkte, Farbe, Gruppierung) mit einer Variable variieren oder festsetzen

12.30-13.30

Mittagspause

13.30-15.15

Daten einlesen und schreiben

- Das Arbeitsverzeichnis in Python und der IDE spyder setzen
- Ein CSV bzw. Excel einlesen und schreiben
- Von einer URL einlesen
- Überblick über nützliche Parameter
- Lesen und Schreiben vom Python Format pickle
- Umgang mit großen Daten

15.15-15.30

Kaffeepause

15.30-17.00

Überblick über Machine Learning

- Einführung in Machine Learning (ML)
- Anwendungsbeispiele von ML
- Künstliche Intelligenz – Machine Learning – Deep Learning
- Unterschied Supervised – Unsupervised Learning (überwachtes – unüberwachtes Lernen)
- Overfitting, Train-Test-Split und cross-validation (Kreuzvalidierung)

17.00

Ende

09.00-09.15

Rückblick und offene Fragen von Tag 2

09.15-10.45

Lineare Regression mit scikit-learn

- Aufteilung der Daten in Test-Train, ein Model erstellen und validieren
- Grundlagen des Algorithmus
- Umsetzung in Python mit scikit-learn
- Validieren der Ergebnisse (mean squared error)

10.45-11.00

Kaffeepause

11.00-12.30

Entscheidungsbaum mit scikit-learn

- Aufteilung der Daten in Test-Train, ein Model erstellen und validieren
- Grundlagen des Algorithmus
- Umsetzung in Python
- Validieren der Ergebnisse (u.a. confusion matrix, sensitivity, accuracy)
- Anpassen von Hyperparametern im Training

12.30-13.30

Mittagspause

13.30-15.15

K-Means Clustering

- Ein Cluster-Model erstellen und validieren
- Grundlagen des Algorithmus
- Umsetzung in Python mit scikit-learn
- Validieren der Ergebnisse (Sillhouette Score, Calinski-Harabasz)

15.15-15.30

Kaffeepause

15.30-17.00

Weitere Aspekte im Machine Learning

- Cross-validation (Kreuzvalidierung)
- Wann wähle ich welchen Algorithmus?
- Überblick der Algorithmen Support Vector Machine (SVM) und Random Forest
- Umsetzung der Algorithmen in Python

17.00

Ende

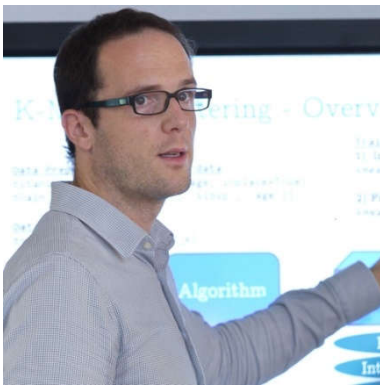
Ihre Dozenten

Einer unserer folgenden Experten leitet das Seminar



Dr. Rolf Köhler

Nach dem Studium der Mathematik und BWL promovierte er im Cyber Valley am Max-Planck Institut Tübingen. Sein Forschungsschwerpunkt war im Bereich Machine Learning und Bildverarbeitung. Seit 2015 arbeitet er bei der Robert Bosch GmbH im Bereich Deep Learning und implementiert und adaptiert verschiedene Algorithmen für industrielle Anwendungsfälle. Daraus sind mehrere Patentanmeldungen entstanden. Seit 7 Jahren verwendet er die Programmiersprache Python.



Jan Köhler

Vom Hintergrund Statistiker und Wirtschaftsingenieur hat er über 7 Jahre an den neuesten Technologien in Machine Learning, Deep Learning und Data Science im Bosch Center for Artificial Intelligence (BCAI) gearbeitet, hat in der Praxis bei über 25 Patentanmeldungen (meist als Haupterfinder) beigetragen und ist Mitautor bei Veröffentlichungen auf Machine Learning Konferenzen neben Veröffentlichungen im Bereich der Medizinstatistik bzw. des Operations Research. In vielen Praxisprojekten unterstützte er bisher als Data Scientist und hat verschiedene Teilnehmer, vom Projektmitarbeiter bis zum Konzern-Vorstand geschult.

Zusammenfassung

€ Preise

1710 € zzgl. MwSt.




Termin und Ort

Termine und Orte finden Sie unter <https://enable-ai.de>

Haben Sie Fragen? Wir helfen Ihnen. Versprochen.

 Enable AI, Stuttgart

 0711 96881553

 info@enable-ai.de