IBB BUSINESS AKADEMIE
WEITER MIT BILDUNG

Berufs-begleitend weiterbilden!

Data Scientist - Fokus Python

Ein kurzer Überblick

Der Beruf des Data Scientist ist einer der gefragtesten des aktuellen Jahrhunderts. Das berufsbegleitende Online-Training unseres Kooperationspartners StackFuel vermittelt angewandte Inhalte zu den Themen Künstliche Intelligenz und Machine Learning wie unüberwachtes und überwachtes maschinelles Lernen, unterschiedliche Datenvisualisierungsmethoden und das Data-Storytelling. Dadurch werden Teilnehmende in die Rolle des Data Scientist weiterentwickelt. Anschließend können Sie Ihr erworbenes Wissen in Ihrer Abteilung einbringen und selbstständig Machine-Learning-Algorithmen implementieren. Während des Trainings arbeiten Sie in unserer browserbasierten, interaktiven Lernumgebung, dem Data Lab. Dabei handelt es sich um eine vollwertige Programmierumgebung, in der die selbst programmierten Codes ausgeführt werden können.

Mit Abschluss des Trainings können Sie Performance-Metriken und Annahmen von Modellen des überwachten und unüberwachten Lernens mit sklearn anwenden. Darüber hinaus erlangen Sie Grundlagen des Data-Storytelling sowie Best Practices der informativen Gestaltung von Visualisierungen mit bokeh Algorithmen des überwachten und unüberwachten Lernens wie Entscheidungsbäume und Random Forests.

Kursinhalte

Preparation
Ziel: Auffrischung der Kenntnisse im Umgang mit Python sowie mathematischer Grundlagen Beschreibung

Beschreibung:
Teilnehmende führen Analysen und Datenmanipulationen in Python aus und nutzen dabei die Pakete Pandas und Matplotlib.

Kapitel 1 – Data Analytics with Python:

Kursnummer
BA-3053-B-1

Beratung und Anmeldung
Telefon: 04161 5165-89
E-Mail: akademie@ibb.com

Die nächsten Starttermine
16.06.25 - 14.12.25
28.07.25 - 25.01.26
08.09.25 - 08.03.26
20.10.25 - 19.04.26
01.12.25 - 31.05.26

Trainingsdauer
Seminardauer: 26 Wochen

Teilnahmegebühr
ab 4.990,00€

Berufs-begleitend weiterbilden!

Teilnehmende machen sich mit unserer interaktiven Programmierumgebung – dem Data Lab – vertraut und frischen die wichtigsten Programmier- und Python-Grundlagen zur Datenverarbeitung mit Pandas, Datenvisualisierung mit Matplotlib und Seaborn und Datenbankabfrage mit SQL Alchemy auf.

Kapitel 2 – Linear Algebra:
Teilnehmende machen sich mit dem mathematischen Hintergrund von Data-Science-Algorithmen vertraut und lernen die Grundbegriffe der linearen Algebra kennen. Unter Verwendung des Pakets Numpy rechnen die Teilnehmenden mit Vektoren und Matrizen.

Kapitel 3 – Probability Distributions:
Teilnehmende lernen mehr über den statistischen Hintergrund von Data-Science-Algorithmen. Sie beschäftigen sich mit wichtigen statistischen Konzepten und lernen diskrete und kontinuierliche Verteilungen kennen. Darüber hinaus erhalten Teilnehmende einen Einblick in die Versionierung von Code mit Git.

Machine Learning Basics
Ziel: Lösen von überwachten und unüberwachten Machine-Learning-Problemen mit sklearn

Beschreibung:
Teilnehmende erstellen Data-Science-Workflows mit sklearn, evaluieren ihre Modell-Performance anhand von geeigneten Metriken und werden für das Problem des Overfittings sensibilisiert.

Kapitel 1 – Supervised Learning (Regression):
Anhand der linearen Regression erlernen Teilnehmende den Umgang mit dem Python-Paket sklearn. Weiterhin beschäftigen sie sich mit den Annahmen des Regressionsmodells und der Evaluation der erzeugten Prognosen. In diesem Zuge werden auch der Bias-Variance Trade-Off, Konzepte der Regularisierung sowie verschiedene Maße der Modellgüte verdeutlicht.

Kapitel 2 – Supervised Learning (Classification):
Teilnehmende werden in Klassifizierungsalgorithmen anhand des k-Nearest-

Berufs-begleitend weiterbilden!

Neighbors-Algorithmus eingeführt und lernen, den Algorithmus zu evaluieren und die Klassifizierungsperformance einzuschätzen. Sie optimieren die Parameter ihres Modells unter Beachtung der Aufteilung der Daten in Trainings- und Evaluationssets.

Kapitel 3 – Unsupervised Learning (Clustering):
Teilnehmende lernen den k-Means-Algorithmus als Beispiel eines Algorithmus des unüberwachten Lernens kennen. Die Annahmen und Performance-Metriken des Algorithmus werden kritisch beleuchtet und ein kurzer Ausblick auf eine Alternative zum k-Means-Clustering geworfen.

Kapitel 4 – Unsupervised Learning (Dimensionality Reduction):
Teilnehmende lernen, wie sie mithilfe einer Principal Component Analysis (PCA) die Dimension der Daten verringern können und nutzen die PCA, um unkorrelierte Features aus den Ursprungsdaten zu erzeugen. In diesem Zusammenhang wird das Thema Feature Engineering näher betrachtet und aus den alten Features neue erzeugt.

Kapitel 5 – Outlier Detection:
Teilnehmende lernen verschiedene Ansätze kennen, um Ausreißer zu identifizieren und verstehen, mit diesen ungewöhnlichen Datenpunkten umzugehen. Sie nutzen robuste Maße und Modelle, um den Einfluss der Ausreißer zu minimieren.

Deep Dive Supervised Learning
Ziel: Erweiterung des eigenen Data-Science-Toolkits

Beschreibung:
Teilnehmende intensivieren ihre Kenntnisse über Modelle zur Klassifikation von Daten. Dabei erweitern sie ihre Fähigkeiten im Sammeln und Aufbereiten von Daten.

Kapitel 1 – Data Gathering:
Teilnehmende lernen, Daten zu sammeln, indem sie Webseiten und PDFDokumente auslesen. Mithilfe von Regular Expressions strukturieren sie gesammelte Textdaten so, dass sie diese zusammen mit bekannten Algorithmen verwenden können.

Berufs-begleitend weiterbilden!

Kapitel 2 – Logistic Regression:
Teilnehmende lernen einen zweiten Klassifizierungsalgorithmus kennen: die logistische Regression. Sie nutzen neue Performance-Metriken zur Evaluation der Ergebnisse und erfahren, wie sie nicht-numerische Daten für ihre Modelle nutzbar machen.

Kapitel 3 – Decision Trees and Random Forests:
Teilnehmende lernen den Entscheidungsbaum als leicht zu interpretierendes Modell kennen. Sie kombinieren mehrere Modelle zu einem Ensemble, um die Vorhersagen ihres Modells zu verbessern. Weiterhin erhalten sie Methoden zu unausgeglichenen Kategorien an die Hand.

Kapitel 4 – Support Vector Machines:
Teilnehmende lernen einen letzten Klassifizierungsalgorithmus kennen – Support Vector Machines (SVM) und beleuchten das Verhalten verschiedener Kernel für die SVM. Außerdem erlernen sie die typischen Schritte des Natural Language Processing (NLP) und bearbeiten ein NLPSzenario unter Verwendung von Bag-of-Words-Modellen.

Kapitel 5 – Neural Networks:
Teilnehmende werden in künstliche neuronale Netze eingeführt und lernen mehr über Deep Learning, um ein künstliches neuronales Netzwerk mit mehreren Schichten zu erzeugen und auf ein Datenset anzuwenden.

Advanced Topics in Data Science
Ziel: Selbstständiges Anwenden einfacher und komplexer Modellierungen

Beschreibung:
Teilnehmende erlangen Souveränität im Lösen von Data-Science-Problemen und lernen, Ergebnisse kompetent zu kommunizieren.

Kapitel 1 – Visualization and Model Interpretation:
Teilnehmende erlernen wichtige Methoden zur Interpretation und Visualisierung von Machine-Learning-Modellen. Durch die Verwendung modelagnostischer Methoden zur Interpretation lernen sie Erkenntnisse zur Funktionsweise ihrer Modelle abzuleiten und zu kommunizieren.

Berufs-begleitend weiterbilden!

Kapitel 2 – Spark:
Teilnehmende erfahren, weshalb die Arbeit mit verteilten Speichersystemen relevant ist. Mit dem Python-Paket PySpark erlernen sie verteilte Datenbanken auszulesen, Big-Data-Analysen durchzuführen und bekannte Machine-Learning-Algorithmen auf verteilten Systemen zu nutzen.

Kapitel 3 – Exercise Project:
Teilnehmende bearbeiten ein Prädiktionsproblem mit Hilfe eines größeren Datensets und setzen ihre Data-Science-Fähigkeiten von der Reinigung des Datensets bis zur Interpretation des Modells eigenständig ein. In einer Projektbesprechung mit dem Mentorenteam von StackFuel erhalten Teilnehmende Feedback zu ihrem Lösungsansatz.

Kapitel 4 – Final Project:
Teilnehmende erhalten ein weiteres größeres Datenset, das sie selbstständig analysieren und im Vergleich zum Übungsprojekt mit weniger Hilfestellungen lösen müssen. In einer individuellen Projektbesprechung mit dem Mentoring Team von StackFuel erhalten Teilnehmende Feedback zu ihrem Lösungsansatz.

Teilnahmevoraussetzungen

Für das Data Scientist Training werden gute Kenntnisse in Python und gängigen Modulen (pandas, matplotlib) vorausgesetzt.

Allen Interessierten stehen wir in einem persönlichen Gespräch zur Abklärung ihrer individuellen Teilnahmevoraussetzungen zur Verfügung.

Zielgruppe

Die Data Scientist Weiterbildung eignet sich für alle, die Daten analysieren und auf Grundlage dieser Vorhersagen erstellen möchten, um datengetriebene Entscheidungen zu treffen. Auch für Quereinsteiger: ist die Data Scientist Weiterbildung geeignet. Darüber hinaus sollten Sie Interesse an Statistik, logischem Denken und maschinellem Lernen mitbringen.

IBB BUSINESS AKADEMIE
WEITER MIT BILDUNG

Berufs-begleitend weiterbilden!

Ihre Vorteile

- Praxisnahe Lernumgebung
- Moderner Technologie Stack
- Browserbasiert; Innovatives Data Lab

Herausgeber:

IBB Institut für Berufliche Bildung AG
Bebelstr. 40
21614 Buxtehude

Telefon: 04161 5165-89
E-Mail: akademie@ibb.com

Vorstand
Katrin Witte (Vorsitz)
Lea Tornow
Sabine Ulrichs

Aufsichtsratsvorsitzende
Sigrid Baumann-Tornow

ibb.weiterbildung
IBB_AG
pages/ibbbusinessakademie
company/ibb-business-akademie