

Data Science und Machine Learning mit Python

Von der Datenanalyse zum belastbaren Machine-Learning-Modell

01 INTRO

Data-Science-Projekte verbinden unterschiedliche Disziplinen: Daten müssen beschafft und verstanden, aufbereitet und exploriert, geeignete Features entwickelt, Modelle trainiert und schließlich belastbar bewertet werden. Die Qualität des Ergebnisses hängt deshalb vom gesamten Prozess ab und nicht nur vom verwendeten Algorithmus.

Der Kurs führt diesen vollständigen Workflow mit Python praktisch zusammen. Pandas und Visualisierungswerkzeuge werden für Datenanalyse und Aufbereitung eingesetzt, scikit-learn für Preprocessing, Modellierung und Validierung. Die Teilnehmer bearbeiten einen Data-Science-Prozess von der fachlichen Fragestellung bis zur nachvollziehbaren Bewertung eines Modells.

DAUER

5 Tage

FORMAT

Präsenz · Live Online · Inhouse

VORKENNTNISSE

Grundkenntnisse in Python; Erfahrungen mit Pandas sind hilfreich

02 TRAININGSPROGRAMM

Inhalte

Data Science als Prozess

- Data Analytics, Data Science und Machine Learning
- fachliche Fragestellung
- Daten
- Modelle
- Evaluation
- iterative Arbeitsweise
- CRISP-DM und moderne Workflows

Python-Werkzeuge für Data Science

- Python-Umgebung
- Jupyter
- NumPy
- Pandas
- Matplotlib und Seaborn
- scikit-learn
- reproduzierbares Arbeiten

Daten beschaffen und verstehen

- Dateien
- Datenbanken und APIs im Überblick
- DataFrames
- Datentypen
- Datenprofil
- Datenqualität
- fachlicher Kontext

Daten aufbereiten

- fehlende Werte
- Dubletten
- Datentypen
- Ausreißer
- Kategorien
- Transformation
- Zusammenführen
- reproduzierbare Bereinigung

Explorative Datenanalyse

- deskriptive Kennzahlen
- Verteilungen
- Gruppen
- Zusammenhänge
- Korrelation
- Auffälligkeiten
- Hypothesen entwickeln

Daten visualisieren

- Diagrammtypen
- Verteilungen
- Beziehungen
- Zeitreihen
- Gruppenvergleiche
- Visualisierung für Exploration und Kommunikation

Features und Preprocessing

- numerische Features
- kategoriale Features
- Skalierung
- Encoding
- Imputation
- Feature Engineering
- Feature Selection
- Leakage vermeiden

Überwachtes Machine Learning

- Regression
- Klassifikation
- lineare Modelle
- Entscheidungsbäume
- Random Forest
- weitere Verfahren einordnen
- Modellwahl

Unüberwachtes Lernen

- Clustering
- k-Means
- hierarchische Verfahren
- Dimensionsreduktion
- PCA
- Exploration ohne Zielvariable
- Interpretation

Modelle bewerten und verbessern

- Train-Test-Split
- Cross Validation
- Regressionsmetriken
- Klassifikationsmetriken
- Overfitting
- Baselines
- Hyperparameteroptimierung

Reproduzierbare ML-Workflows

- scikit-learn Pipeline
- ColumnTransformer
- Preprocessing und Modell verbinden
- Modellvergleich
- Persistenz
- Dokumentation
- Grenzen zwischen Experiment und Produktion

Durchgängiges Data-Science-Projekt

- fachliche Fragestellung definieren
- Daten erschließen
- Datenqualität untersuchen
- Daten aufbereiten und explorieren
- Features entwickeln
- Modelle trainieren
- Modelle validieren und vergleichen
- Ergebnisse interpretieren
- Grenzen und nächste Schritte dokumentieren

03 DURCHFÜHRUNG

Rahmen und Organisation

Dauer	5 Tage
Durchführung	Präsenz oder Live Online, vorzugsweise als Inhouse-Schulung
Schwerpunkt	Inhouse-Schulungen für Unternehmen und Teams
Sprache	Deutsch, englische Fachbegriffe und Dokumentation
Unterlagen	Kursunterlagen, Beispiele und Übungsaufgaben
Technik	Eigener Rechner mit Python, Jupyter und den gängigen Analyse- und ML-Bibliotheken

Inhouse-Schulungen

Eigene Datenbestände und fachliche Fragestellungen können als Grundlage für das durchgängige Projekt verwendet werden.

04 ZIELGRUPPE

Für wen ist der Kurs gedacht?

Entwickler, technische Fachkräfte, angehende Data Analysts und Data Scientists, die den vollständigen Data-Science- und Machine-Learning-Prozess mit Python praktisch erlernen möchten.

Voraussetzungen

Grundkenntnisse in Python. Besondere Kenntnisse in Machine Learning werden nicht vorausgesetzt.

05 LERNZIELE

Was Sie nach dem Kurs können

Nach dem Kurs können die Teilnehmer:

- einen Data-Science-Prozess von der Fragestellung bis zur Modellbewertung strukturieren
- Daten mit Python untersuchen und aufbereiten
- explorative Analysen und Visualisierungen durchführen
- Features für Machine Learning vorbereiten
- Regressions-, Klassifikations- und Clustering-Verfahren einsetzen
- Modelle korrekt validieren und vergleichen
- reproduzierbare scikit-learn-Workflows entwickeln
- Ergebnisse und Grenzen eines Modells nachvollziehbar kommunizieren

06 FAQ

Häufige Fragen

Was unterscheidet den Kurs von den Machine-Learning-Grundlagen?

Der Grundlagenkurs ist eine konzentrierte Einführung in Machine Learning und den ML-Workflow. Dieser Kurs ist ein umfassender fünftägiger Data-Science-Kurs, der Datenaufbereitung, Exploration, Visualisierung, Feature Engineering, Modellierung und Evaluation zu einem vollständigen Prozess verbindet.

Ersetzt der Kurs die spezialisierten Data- und ML-Kurse?

Nein. Dieser Kurs bietet Breite und einen vollständigen End-to-End-Workflow. Die spezialisierten Kurse behandeln ihre jeweiligen Themen wesentlich tiefer.

Welche Machine-Learning-Verfahren werden behandelt?

Regression, Klassifikation, Baumverfahren und Clustering bilden den Kern. Weitere Verfahren werden dort eingeordnet, wo sie für die Modellwahl relevant sind.

Wird Deep Learning behandelt?

Neuronale Netze werden eingeordnet, sind aber nicht Schwerpunkt. Dafür ist der Grundlagenkurs zu neuronalen Netzen vorgesehen.

Ist der Kurs auch ohne Data-Science-Erfahrung geeignet?

Ja. Python-Grundkenntnisse werden vorausgesetzt; der Data-Science- und Machine-Learning-Prozess wird systematisch aufgebaut.

Warum dauert der Kurs fünf Tage?

Weil hier nicht nur einzelne ML-Algorithmen behandelt werden. Der Kurs umfasst den vollständigen Weg von Datenverständnis und Aufbereitung über Exploration und Feature Engineering bis zu Modellierung, Validierung und einem durchgängigen praktischen Projekt.

UC IT Service · Ulrich Cuber · kontakt@uc-it.de

<https://www.uc-it.de/coaching/data-science-python/>