

Machine Learning – Grundlagen und Workflow

Machine Learning verstehen und systematisch in Datenprojekten einsetzen

01 INTRO

Machine Learning ermöglicht es, aus vorhandenen Daten Modelle abzuleiten, die Zusammenhänge erkennen und Vorhersagen oder Zuordnungen für neue Daten treffen können. Entscheidend für belastbare Ergebnisse ist jedoch nicht allein die Auswahl eines Algorithmus, sondern der gesamte Prozess von der Fragestellung über Daten und Modellierung bis zur Bewertung.

Der Kurs vermittelt die grundlegenden Konzepte des Machine Learning und einen systematischen ML-Workflow. Die Teilnehmer lernen wichtige Verfahren des überwachten und unüberwachten Lernens kennen, entwickeln erste Modelle mit Python und scikit-learn und beurteilen deren Ergebnisse kritisch.

DAUER

3 Tage

FORMAT

Präsenz · Live Online · Inhouse

VORKENNTNISSE

Grundkenntnisse in Python und im Umgang mit tabellarischen Daten

02 TRAININGSPROGRAMM

Inhalte

Machine Learning einordnen

- Machine Learning, Data Science und klassische Datenanalyse
- Lernen aus Daten
- Modelle und Algorithmen
- Training und Vorhersage
- typische Einsatzgebiete
- Grenzen von Machine Learning

Arten des maschinellen Lernens

- überwachtes Lernen
- unüberwachtes Lernen
- Regression
- Klassifikation
- Clustering
- weitere Lernformen im Überblick
- Auswahl nach Problemstellung

Vom fachlichen Problem zur ML-Aufgabe

- Zielvariable
- Features
- Beobachtungen
- geeignete Problemformulierung
- Erfolgskriterien
- Baseline
- fachliche und technische Metriken

Daten für Machine Learning

- Trainingsdaten
- Datenqualität
- fehlende Werte
- numerische und kategoriale Merkmale
- Ausreißer
- Datenmenge und Repräsentativität
- Bias in Datensätzen

Der Machine-Learning-Workflow

- Daten verstehen
- Daten vorbereiten
- Modell auswählen
- trainieren
- validieren
- bewerten
- verbessern
- reproduzierbarer Ablauf

Regression

- Regressionsprobleme
- lineare Regression
- Features und Zielgröße
- Modellanpassung
- Vorhersagen
- Residuen
- typische Bewertungsmetriken

Klassifikation

- Klassen und Labels
- binäre und mehrklassige Probleme
- Wahrscheinlichkeiten
- Entscheidungsgrenzen
- erste Klassifikationsmodelle
- Confusion Matrix
- Precision und Recall

Bäume und Ensemble-Verfahren

- Entscheidungsbäume
- interpretierbare Regeln
- Overfitting
- Random Forest
- Ensemble-Idee
- Boosting im Überblick
- typische Einsatzbereiche

Unüberwachtes Lernen

- Lernen ohne Zielvariable
- Clustering
- Ähnlichkeit und Distanz
- k-Means
- Cluster interpretieren
- Dimensionsreduktion im Überblick

Modellbewertung und Generalisierung

- Training und Test
- Train-Test-Split
- Cross Validation
- Overfitting und Underfitting
- Generalisierung
- Datenleckage
- sinnvolle Baselines

Machine Learning mit scikit-learn

- Estimator-Konzept
- fit, predict und transform
- Preprocessing
- Modelle
- Metriken
- Pipelines
- reproduzierbare Experimente

Praktisches ML-Projekt

- Fragestellung formulieren
- Daten vorbereiten
- Baseline entwickeln
- Modell auswählen und trainieren
- Modell bewerten
- Fehler analysieren
- Ergebnisse interpretieren und dokumentieren

03 DURCHFÜHRUNG

Rahmen und Organisation

Dauer	3 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 scikit-learn

Inhouse-Schulungen

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

04 ZIELGRUPPE

Für wen ist der Kurs gedacht?

Data Analysts, Entwickler, technische Fachkräfte und angehende Data Scientists, die einen fundierten Einstieg in Machine Learning suchen.

Voraussetzungen

Grundkenntnisse in Python und im Umgang mit tabellarischen Daten. Mathematische Grundlagen werden soweit erforderlich erläutert.

05 LERNZIELE

Was Sie nach dem Kurs können

Nach dem Kurs können die Teilnehmer Machine-Learning-Probleme erkennen und formulieren, grundlegende Verfahren unterscheiden, einen vollständigen ML-Workflow durchführen und erste Modelle mit scikit-learn entwickeln und bewerten.

06 FAQ

Häufige Fragen

Ist der Kurs mathematisch?

Die notwendigen mathematischen Konzepte werden erklärt, der Schwerpunkt liegt jedoch auf Verständnis und praktischer Anwendung.

Was unterscheidet den Kurs von Data Science und Machine Learning mit Python?

Dieser Kurs konzentriert sich auf die Grundlagen und den systematischen Machine-Learning-Workflow. Der Data-Science-Kurs verbindet Datenanalyse und Machine Learning zu einem umfassenderen End-to-End-Kurs.

Wird Deep Learning behandelt?

Neuronale Netze werden eingeordnet. Eine systematische Einführung bietet der Grundlagenkurs zu neuronalen Netzen.

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

<https://www.uc-it.de/coaching/machine-learning-grundlagen/>