

Daten ziehen und aufbereiten

- › Dateien im Verzeichnis ansteuern
- › Standardverfahren zum Lesen/Schreiben von Text- und CSV-Dateien
- › *SQL*-Datenbanken ansteuern
- › Arbeiten mit Datenmatrizen: Einführung in *Numpy* & *Pandas*
- › Daten inspizieren und beschreiben
- › Fehlende Werte behandeln

Machine Learning Grundlagen

- › Grundlagen, Varianten und Techniken des Machine Learnings.
- › Arbeiten mit der Machine Learning Bibliothek scikit-learn.
- › Einfache Zusammenhänge zwischen stetigen Variablen modellieren: Lineare Regression
- › Klassifizieren mit Logistic Regression, Softmax und Support Vector Machine.
- › Modellen evaluieren: Accuracy, Precision, Recall & Confusion matrix
- › Unterteilen der Daten in Trainings- und Testdaten

Feature-Extraction

- › Kategoriale Daten vorbereiten: *One-Hot Codierung*
- › Standardisierung von Daten
- › Dimensionsreduktion mit PCA (Principle Component Analysis)
- › Aufbereitung von Textdaten: Tokenizer und Bag-of-words.

Machine Learning Workflow

- › Daten aufarbeiten und zusammenführen (DataMapper)
- › Pipelines einrichten
- › Speichern und laden trainierter Klassifizierer

1. Tag

Wir wiederholen alle wichtigen Python-Grundlagen und behandeln numpy sowie pandas

Schwerpunkt:
Pandas

2. Tag

Wir werden 2 komplette Beispiele für Regression und Klassifikation behandeln und die gesamte ML-Pipeline durchgehen

Schwerpunkt:
Modellevaluierung

3. Tag

Der dritte Tag steht uns im Grunde zur freien Verfügung. Hier können Themen wie Model Deployment, Visualisierung oder spezielle Themengebiete wie NLP behandelt werden

Schwerpunkt:
Legen die Teilnehmer fest