



## GLOSSAR (GRUNDLAGEN)

- Willkommen
- Data Science
- Big Data
- Python
- Datenstrukturen in Python
- Jupyter Notebooks
- Docker
- Kubernetes
- GitHub
- CRISP-DM

### Grundlagen und Tools (1), (b2)

- NumPy
- Pandas
- SciKit-Learn
- Matplotlib

### Python Bibliotheken

- Data Mining und Data Crawling
- Stichproben und Grundgesamtheiten
- Konfidenzniveau und Konfidenzintervall
- Fehlermarge
- Modus, Median und arithmetisches Mittel
- Varianz und Kovarianz
- Standardabweichung
- Korrelation
- Lineare / Logistische Regression
- Interaktionseffekt
- Bootstrapping
- Signifikanz
- Maschinelles Lernen
- Trainings- und Validierungsdatensatz
- Metriken zur Evaluierung
- Deep Learning / Neuronale Netze
- Batch / Epoch(e)
- Aktivierungsfunktion

### Grundlagen Statistik (2), (3), (6)

### Grundlagen ML / DL (4), (5), (9), (b3)

- Transformer
- Natural Language Processing
- Sentimentanalyse
- Clusteranalyse
- KNN - Algorithmus
- K-Means - Algorithmus
- Random Forest - Algorithmus
- Agglomeratives Clustering - Algorithmus
- BIRCH - Algorithmus
- DBSCAN - Algorithmus
- Spectral Clustering - Algorithmus
- Mean Shift - Algorithmus
- Gaussian Mixture Algorithmus
- Maximum-Likelihood-Methode
- Within Cluster Sum of Squares
- Faktorenanalyse
- Eigenwerte

### Automatisierte Textanalyse (b1), (b4)

### Algorithmen (7)

### Vertiefungen (8)

## JUPYTER NOTEBOOK (TUTORIALS)

### Autodidaktisch:

- (1) Pakete und Datenstrukturen
- (2) Regressionsanalyse
- (3) Interaktionseffekte
- (4) Mustererkennung
- (5) Machine Learning
- (6) Standardisierung
- (7) Clusteranalyse
- (8) Within Cluster Sum of Squares
- (9) Neuronale Netze / Deep Learning

### Exemplarisch:

- (b1) Handschrifterkennung
- (b2) Gesichtserkennung
- (b3) Natural Language Processing

