Our **trainings** introduce the foundations of data analysis and relevant tools for data scientists. During our sessions, we always keep data-driven use cases in sight. Our modular training concept consists of skill building blocks from introductory to expert level. We emphasize applications and a hands-on approach. All trainings take place in a virtual environment or on-site. Our trainers are senior data scientists with extensive experience in research and industry.

Our **TechTrainings** are geared towards technical users - data scientists, analysts, engineers - building data-driven innovation hands-on.

We use our **digital laboratory in the cloud**, thus each participant can work at his workspace independently while the trainer is presenting. We focus on interactivity, hands-on and individual support. This is what we aim with our developed and well tested trainings material in Jupyter Notebooks, small group sizes and enough time for practice.
NLP from basics to advanced with Python tools. Learn how to extract, handle and process text data. Apply your analytics and machine learning skills to expand your data science know-how.

**Level:** 🚀 (Advanced)

**Duration:** 2 days

**Prerequisites:** DAP + MLP (or similar know-how)

**Language:** english, german; materials are in english

1. **Extracting Text from Documents**  
   Extracting natural language texts from various sources and formats.

2. **Natural Language Processing: The Basic Toolkit**  
   Tokenization, stemming, and other fundamental tools for slicing and dicing text data.

3. **Document Similarity for Clustering and Recommendation**  
   How to measure the similarity of texts and do useful things with that.

4. **Text Classification**  
   Machine learning for classifying text documents.

5. **Sentiment Analysis**  
   How to deal with more comments than you can possibly read.

6. **Topic Modelling**  
   An unsupervised ML technique for discovering topics in text documents.

7. **Word Embeddings**  
   Mapping words to vectors for fun and profit.