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.

- **Data Analysis with Python (DAP)**
- **Machine Learning with Python (MLP)**
- **Deep Learning with TensorFlow (DLT)**
- **Machine Learning on Time Series (MLTS)**
- **Big Data Analysis with PySpark (BDAS)**
- **Machine Learning with PySpark (MLS)**
- **Natural Language Processing with Python (NLPP)**

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

- **Data Science, ML & AI for Business (DSB)**

Our **BusinessTrainings** provide insights for decision makers - shedding light on how to put data science and artificial intelligence to work in the enterprise.

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.
Machine Learning as a big data edition. Get started with machine learning in theory and practice. Strengthen your big data skills with setting up your machine learning pipelines for different use cases.

**Level:** 🎧 (Advanced)  
**Duration:** 2 days  
**Prerequisites:** DAP + BDAS (or similar know-how)  
**Language:** english, german; materials are in english

1. **Introduction to Machine Learning**  
   An overview over the field of machine learning.

2. **ML for Classification**  
   Learn about classifiers and how to measure the quality of their decisions.  
   A. **Building an ML Pipeline for Classification**  
      Build a classification model and learn about the building blocks of ML with PySpark.  
   B. **Feature Engineering and Selection**  
      How to build better features.  
   C. **Algorithm Selection and Hyperparameter Tuning**  
      Learn about classifiers and how to measure the quality of their decisions.  
   D. **Exercise: Titanic Survival Classifier**  
      Build your own classification model from scratch.

3. **ML for Regression**  
   Learn about regressors and how to measure the quality of their prediction.  
   A. **Exercise: Predicting House Prices**  
      Apply regression to predict house prices.

4. **Building a Recommender Engine**  
   Use machine learning to generate movie recommendations.

5. **Unsupervised Learning: Clustering**  
   Apply clustering algorithms to detect structure in the data.