

## Workshop

### Machine Learning with Python

#### Organisational details

Instructor:	Dr Tobias Keller
Dates:	August 26, 27, and September 3, 2022, 9.00 am – 5.00 pm respectively
Venue:	Room 002, Licher Str. 68, 35394 Giessen
ECTS:	3
Max. participants:	10

#### Objectives

In this follow-up course, participants reinforce and add to their knowledge of machine learning concepts solving classification and regression problems with Python. Through application, they strengthen their understanding of machine learning algorithms, evaluation techniques and measures, and learn how to combine data preparation and algorithms in machine learning pipelines.

In addition, participants will be introduced to more advanced concepts, focusing on model selection (including hyperparameter tuning and algorithm selection) and approaches to understanding “black box” models (also known as explainable AI, “XAI”).

#### Content and methods

The course consists of lectures to build the theoretical background as well as hands-on tutorials and exercises using Python. Participants will learn by examples and exercises from the instructor’s experience in research and practice.

Hands-on tutorials and exercises will make up about 50 per cent of the course time. Students will complete those using their own computers. Please see the preparation requirements below for a list of software that needs to be installed to that end. The exercises will be based on exemplary datasets that will be provided to the participants before the course.

#### Requirements to register for the course

- Knowledge of the topics covered in the previous course, “Introduction to Machine Learning”, is required for participating in this follow-up course
- In particular, the participants need to be familiar with the following concepts and methods:
  - The typical process of machine learning (e.g. following CRISP DM)

- Supervised machine learning algorithms for regression and classification such as linear regression, logistic regression, tree based (ensemble) learners, neural networks
- Unsupervised machine learning algorithms for clustering and dimensionality reduction such as k-means clustering, principal component analysis
- Evaluation of machine learning results using k-fold cross-validation and discussion of alternative performance measures for classification and regression problems

**To gain the ECTS credit points participants have to:**

- Refresh their knowledge from the previous course, “Introduction to Machine Learning”,
- Install Visual Studio Code, the Jupyter Extension and Miniconda (see <https://code.visualstudio.com/docs/datascience/data-science-tutorial> for instructions, but please **use Python 3.8!**),
- Install the following packages: **scikit-learn, pandas, seaborn, ipykernel, statsmodels, xgboost, shap.**
- Complete the following two tutorials on using Jupyter Notebooks in Visual Studio Code:
  - <https://code.visualstudio.com/docs/datascience/jupyter-notebooks>
  - <https://code.visualstudio.com/docs/datascience/data-science-tutorial>,
- should they not be familiar with the packages pandas and numpy, complete the following tutorial:
  - <https://www.hackerearth.com/practice/machine-learning/data-manipulation-visualisation-r-python/tutorial-data-manipulation-numpy-pandas-python/tutorial/>,
- Download and extract the course material and exemplary datasets that will be provided in time before the course,
- Actively participate during the workshop.

**Target group**

Doctoral candidates or postdoctoral researchers doing empirical research or intending to work as data scientists.

**Course language**

English (German, if only German participants)

Please note: As this is an English language course proficiency in English at the C1 level of competency is required.

**About the Instructor**

Dr Tobias Keller has been working as data scientist at ZERO.ONE.DATA, the big data startup of Deutsche Bahn AG since 2016. He consults on and applies machine learning and statistics for artificial intelligence systems in a big data environment. Furthermore, he teaches data science at Deutsche Bahn, in the doctoral education programmes at Justus Liebig University Giessen, and in the doctoral programme

and master programme at WHU – Otto Beisheim School of Management. His research interests include machine learning and artificial intelligence, finance and accounting, strategic management.

**Registration**

By August 16, 2022 via e-mail at [info@ggs.uni-giessen.de](mailto:info@ggs.uni-giessen.de).