Machine Learning with Python – from Zero to Hero
(taught by Dr. Christian Kauth)

This course provides an introduction to the programming language *Python* for data analytics on tabular data. It focuses on a specific subfield of machine learning called predictive modeling. This is the field of machine learning that is the most useful in industry and the type of machine learning that the *scikit-learn* library in Python excels at facilitating.

Unlike statistics, where models are used to understand data, predictive modeling is laser focused on developing models that make the most accurate predictions at the expense of explaining why predictions are made.

In this course you will learn how the sub-tasks of a predictive modeling project map onto Python and the best practice way of working through each task. We will illustrate the procedure by working through some well-understood case study predictive modeling problems from the UCI Machine learning repository (http://archive.ics.uci.edu/ml/index.php).

To work the predictive modeling projects end-to-end, we will use linear (e.g. regression, multivariate linear regression, logistic regression), non-linear (e.g. Naive Bayes, k-Nearest Neighbors, Decision Trees) and ensemble (e.g. Bootstrap Aggregation, Random Forest, Stacked Generalization) machine learning algorithms.

Fun fact: The Python language is a fans' favorite when it comes to solving machine learning problems. You'll enjoy great support from an immense community and from AIs like Chat-GPT, beyond the duration of the course.

Objectives

- To work through a small to medium sized dataset end-to-end
- To deliver a model that can make accurate predictions on new data
- To understand and complete all subtasks of a predictive modeling problem with Python
- To get hands-on experience with Python, SciPy and scikit-learn

Content

- The Python ecosystem for machine learning (Python & SciPy crash-course)
- How to analyze data (descriptive statistics and visualizations to better understand the data)
- How to prepare the data (data transforms to better expose the structure of the problem to the modeling algorithms)
- How to evaluate algorithms (test harness design to evaluate a number of standard algorithms on the data and select the top few to investigate further)
- How to improve results (ensemble techniques and parameter tuning to get the most out of well-performing algorithms on the data)
- How to present results (model finalization)

Preconditions

- None (introductory course).
- **Google account to access Google Colab**

Duration

2 days on Feb 12th – 13th (roughly 7*45 minutes each day)

Evaluation

take home exam: project work to be solved in *Python* on tabular data

ECTS

1