# Assignment Introduction to Machine Learning

## Requirements

This assignment needs be done in teams of two students. Each team receives a different data set and a prediction task that fits with this data set. All steps below must be performed on this data set. All processing and analyses need to be done in Python using suitable machine learning and data processing libraries. Manually modifying the data set is not allowed.

## Learning objectives

After this assignment the student is able to:

* explore data sets by means of machine learning and visualization libraries;
* by means of machine learning libraries preprocess a real-world data set to prepare it for modelling;
* choose appropriate machine learning models for real-world problems;
* apply machine learning models to a realistic data set by means of machine learning libraries;
* set-up and conduct experiments to evaluate the effectiveness of a machine learning model;
* interpret the results of an evaluation experiment.

## Steps of the assignment

1. Load the data set.
2. Thoroughly explore the data set.
3. Based on the data exploration results, determine which preprocessing steps are needed.
4. Preprocess the data as needed.
5. Select a machine learning model that is suitable for the task and properties of the data.
6. Divide the data in a suitable way into training and test data.
7. Train the selected model.
8. Evaluate the effectiveness of the model in an experiment.

## Deliverables

Each team must hand in **one python script (.py file, not a notebook)** with all code and a **report of at most 6 pages** describing at least the following:

* an introduction describing the context and goal of the work;
* the data set used and results of data exploration;
* the selected preprocessing methods, why these were selected, and results of preprocessing;
* the considered machine learning models, the selected model and a motivation for why this model is the most suitable one;
* an explanation of how the model was build;
* an explanation of how the models was evaluated (set-up of the experiment);
* results of the evaluation experiment;
* conclusions and discussion;
* references.

## Assessment

The assignment will be graded on the basis of the assessment form provided on Moodle.