Authors: Natasha Latham and Pavel Zwerschke

Introduction/Description Problem
This project solves the societal problem of not knowing the housing situation in a foreign city. For example, if you are a new student looking for a room in Amsterdam, you could use our model to determine if a room has a fair price or not given the parameters of the room.

We want to train a classifier that estimates the rent of the room given that we know some properties of the room like m^2, location, amount of people in the flat, etc.

Dataset
The dataset which we will primarily use is “The Netherlands Rent Properties” found on kaggle. This data is scraped from the website https://kamernet.nl/ therefore it is limited to rooms, studio’s and small apartments. The dataset contains attributes like firstSeenAt and lastSeenAt which also allows us to see how long the room was offered online. This could be interesting to grasp the demand for certain rooms. The data collection started on 13th of July 2019 and ended on 4th of March 2020 and it has been collected for the whole of the Netherlands. The current energy prices are therefore not reflected in this data.
https://www.kaggle.com/datasets/juangesino/netherlands-rent-properties

Plan of Approach
We want to compare different data mining algorithms like Decision Trees, Random Forests, Boosted Trees and Neural Networks. For this, we will look at the default scikit-learn implementations for said algorithms. We will use cross validation to compare the performances of the models. Also, we want to use scatter plots and residual plots to look at the actual vs. the predicted price for the models.
We will have a look at the data, do some plots, for example average rent for each zip code to validate the quality of the dataset.

Literature

- https://www.statista.com/statistics/612261/average-housing-rent-in-the-netherlands/
- https://www.statista.com/statistics/971481/seasonally-adjusted-residential-property-rent-prices-in-the-netherlands/
- https://link.springer.com/chapter/10.1007/978-3-030-11674-3_3