6105 Mid-Term Project Proposal

**Project title: House Price Prediction**

**Data set: The Boston Housing Dataset (**<http://lib.stat.cmu.edu/datasets/boston>)

**Project idea:**

The project aims to predict house prices regarding to various factors. The dataset used for prediction is allocated by the U.S Census Service concerning housing in the area of Boston. There are 14 attributes in the dataset, such as crime rate, number of rooms, etc. Using the information, I will first visualize the data to see the prices variation versus other factors. Then I will prepare the data for machine learning. Before conducting machine learning, I will decide which model to use considering each model performance on the dataset. The workflow including:

1. Exploratory data analysis to eliminate unimportant features.
2. Clearing the data and labeling the data.
3. Training and comparing performance of three machine learning models: linear regression, decision tree, and random forest.
4. Choosing the greatest performance model to predict the house price in Boston area.

**Software:** APIs for users to fill in factors required for the prediction and present the predicted house price.

**Reference**

1. “End-to-End Uber Data Analysis Project using Machine Learning”, <https://www.projectpro.io/article/uber-data-analysis-project-using-machine-learning-in-python/589>
2. “End-to-End Uber Data Analysis Project using Machine Learning”, <https://www.projectpro.io/article/uber-data-analysis-project-using-machine-learning-in-python/589>
3. “HOUSING PRICE PROJECT REPORT” <https://m2pi.ca/project/2020/bc-financial-services-authority/BCFSA-final.pdf>