**Housing Price Prediction Report**

Yiqi Zhang - 100763230

**Problem Definition**

The program of machine learning focuses on the field of housing price prediction, assisting the customers make an understanding of the trend of housing price based on different situations. As we known, houses can be assessed by various kinds of factors, including space, facilities, neighbours and local policies. The purpose of our program promotion is through the analysis of history trend of local price to make a recommendation price for the current data provided from our customers. The housing price prediction system not only can help renters or vendor to get more accurate marketing price prediction, but also provides housing conditions achieving the expectation for people who want to invest.

**Project Pitch**

In this program, the data from the Boston will be used as the training dataset. Before doing the training, the dataset will be cleaned and preprocessed in order to guarantee more accurate analysis, after which the features that have higher correlation affecting housing price will be selected. The supervised machine learning will be used to train the data; besides Random Forest Regressor and Ridge will be used as the training algorithms. The training accuracy from both will be compared to confirm the algorithm that need to be applied to data from our customers, and the model will provide the price prediction based on each customer’s conditions with online interaction eventually.

**Exploratory Data Analysis**

Before training the data, preprocessing is an important step which can observe the data carefully and decreasing the effect that is caused by useless elements.

1. **Import package and dataset**

**A screenshot of a cell phone

Description automatically generated**

1. **Observation of dataset**

A screenshot of a cell phone

Description automatically generated

**A screenshot of a social media post

Description automatically generated**

1. **Plot the trend of housing price**

**A screenshot of a social media post

Description automatically generated**

1. **Plot the correlation between each feature with heat map**

**A screenshot of a cell phone

Description automatically generated**

1. **Attain the top 9 features which have higher correlation with Sale Price**A screenshot of a social media post

   Description automatically generated
2. **Export specifc features as predicting features**

A screenshot of a social media post

Description automatically generated

1. **Plot the logarithmic value**

**A screenshot of a social media post

Description automatically generated**

1. **Model training**

**A screenshot of a social media post

Description automatically generatedA screenshot of a social media post

Description automatically generated**

1. **Model comparison**

**A screenshot of a social media post

Description automatically generated**

Because the accuracy from RandomForestRegressor is higher than Ridge’s, RandomForestRegressor will be used as the machine learning algorithm.

**Solution Development**

Model1.py shows the model we built for supervised learning and app.py make a bridge between user interface and model.

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a social media post

Description automatically generated**

**User interface with implementation**

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a cell phone

Description automatically generated**