Predicting USA_Housing Prices

Writeup

By

Fatimah Naif Al-Rashed

To Instructor:

Dr. Essam Al Daoud

Abstract

One of the most Machine Learning problems is predicting house prices. And, the goal of this project was to use regression models to predict the USA House prices. We used USA_Housing dataset. We used cross-validation to determine the accurate model. Finally, we find the best performing model for Predicting House Prices are Linear Regression , Ridge and Lasso for USA_Housing dataset.

Design

We design model to predict USA House prices. Also, we applied different experiments to find the accurate model to predict USA House prices.

Data

USA_Housing dataset contains 5000 rows. The dataset has 6 columns, and the prices column is the target value. The features are Avg. Area Income, Avg. Area House Age, Avg. Area Number of Rooms, Avg. Area Number of Bedrooms, Area Population and Address. There is no null in USA_Housing dataset. Also, there is no duplicated rows in USA_Housing dataset.

Algorithms

First, we split the data into a training set and a testing set. Second, we used cross-validation to find the more accurate algorithm. We find the best performing model for Predicting House Prices are Linear Regression, Ridge and Lasso with R2 Square equals 0,9172. We applied four different experiments. First experiment, we used all the features. Second experiment, we had excluded Avg. Area Number of Bedrooms from features. In the first and second experiments we got similar result in R2 Square equals 0,91. In the third experiment, the feature was Avg. Area Income. In the fourth experiment, the features were Avg. Area Income and Avg. Area House Age. In the third and fourth experiments we got similar result in R2 Square equals 0,39.

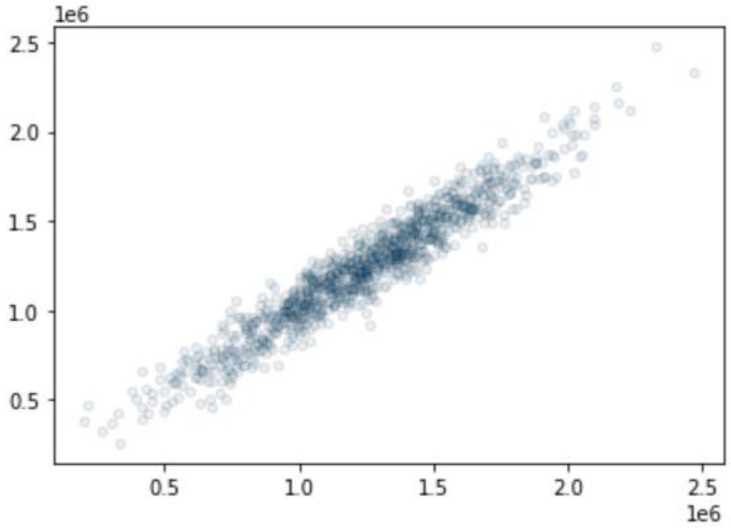
Tools:

We will use Python and Jupyter notebooks for this project. We will use Pandas , Scikit Learn , NumPy , Seaborn and Matplotlib libraries. Pandas library use for handling structured data. NumPy library use for

linear algebra and mathematics. Scikit Learn library use for machine learning. Seaborn and Matplotlib use for data visualization.

Communication

We find the best performing model for Predicting House Prices are Linear Regression, Ridge and Lasso . Also, we should use all the feature exclude Avg. Area Number of Bedrooms from features to get more acuter data model. In figure 1, is show scatter plot predictions using Linear regression.



Figuer1, scatter plot predictions using Linear regression.