

Predicting House Price



Prepared by: Mohammed Hamoud Alharbi

Abstract

The goal of this project is to predict the house price in the USA. To help A real estate agents predict the price of the houses. Many parties, including owners, buyers, agents, and investors, face difficulties in accurately determining the value of the house. It's also a challenging one. Though it is commonly known that size, number of rooms, and age of the house influence pricing, there are many additional factors at play. Additionally, prices are sensitive to changes in market demand and the peculiarities of each situation, such as when a property needs to be urgently sold. The sales price of a property can be predicted in various ways, but is often based on regression techniques.

Design

A real estate agents they need help to estimate the home price for regions in the USA. The dataset provided on Kaggle to work on and find solution. Classifying statuses accurately via machine learning models would enable real estate agents to create a model that will help to estimate of what the house would sell for.

Data

The dataset that consists of (7) columns and (5000) rows and its format is CSV. And its size is 726.21KB. the main features that existing on this dataset are (Avg. Area Income, Avg. Area House Age, Avg. Area Number of Rooms, Avg. Area Number of Bedrooms, Area Population, Price, and Address).

Algorithms

Model

imported and created sklearn linear model LinearRegression object and fit the training dataset in it.

The entire dataset was split into a training set and testing set using sklearn train_test_split(). the training set will be going to use for training the model and testing set for testing the model. We are creating a split of 20% training data and 80% of the training set.

Evaluate the model by checking out its coefficients and how we can interpret them.

Regression Evaluation

```
Test Set Evaluation
MAE 80879.09723489445
MSE 10089009300.89399
RMSE 100444.06055558482
R2 square 0.9179971706834331
```

```
Train Set Evaluation
MAE 81509.3933124445
MSE 10256318867.482723
RMSE 101273.48551068401
R2 square 0.9179787435623722
```

Tools:

SciKit-Learn, Pandas, Seaborn, Matplotlib and Numpy.