Predicting Houses price using linear regression

Muzoon Alshahrani

Bushra Alzahrani

Abstract:

Through the project by building a machine learning linear regression model. The goal of this project is to apply a linear regression model that predicts the prices of the houses based on many features. The data from https://www.kaggle.com/datasets website. We will do a comprehensive analysis with all data cleaning, exploration, visualization, feature selection, model building, and evaluation.
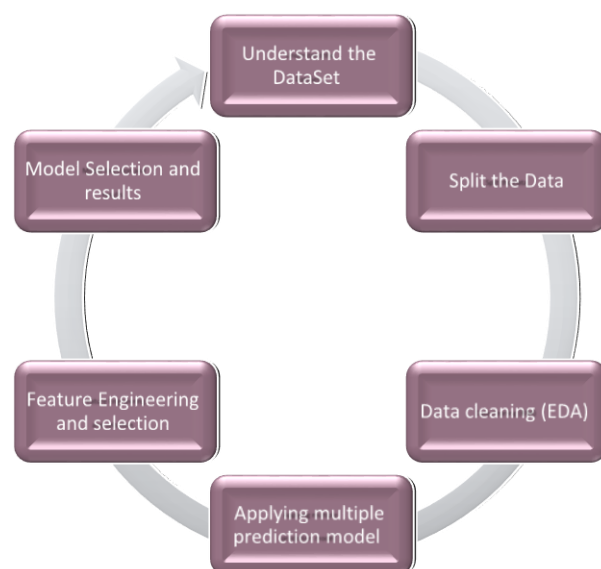
Design:

By applying the dataset on many machines learning models such as linear regression, Cross validation and polynomial regression to predict the houses prices.

Data:

The dataset used in this project was from https://www.kaggle.com/datasets website. The dataset contains 33.7k rows and 19 columns.

Algorithms:

Tools:

Libraries:

- Pandas and NumPy packages to manipulate data.
- Matplotlib & seaborn library for visualizing data.
- Linear Regression , PolynomialFeatures , RidgeCV and for modeling.

Technologies:

- Python, Jupyter Notebook.

Communication:

Predicting Houses price
using linear regression

First!

Data Visualization

Intro

Modling

Muzoon Alshahrani
Bushra Alzahrani

Model Type	Training	Cross Validation(Mean)
Linear Regression	0.532	0.531
Polynomial (degree 2)	0.702	0.675

Accuracy on Training	
MAE	168107.39
RMSE	242419

cont..

Correlation Between the Target (Price) and Features.

	PRICE	BEDROOMS	BATHROOMS	GARAGE	LAND_AREA	FLOOR_AREA	CBD_DIST	NEAREST_STN_DIST	LATITUDE	LONGITUDE	NEAREST_SCH_DIST
PRICE	1	0.25	0.36	0.13	0.056	0.54	-0.36	-0.099	0.043	-0.18	-0.024
BEDROOMS	0.25	1	0.55	0.19	0.048	0.54	0.12	0.098	0.0085	-0.012	0.084
BATHROOMS	0.36	0.55	1	0.2	0.03	0.56	0.04	0.037	0.043	-0.089	0.06
GARAGE	0.13	0.19	0.2	1	0.033	0.17	0.028	0.068	-0.016	0.057	0.059
LAND_AREA	0.056	0.048	0.03	0.033	1	0.073	0.13	0.21	-0.023	0.13	0.26
FLOOR_AREA	0.54	0.54	0.56	0.17	0.073	1	0.02	0.1	0.034	-0.044	0.12
CBD_DIST	-0.36	0.12	0.04	0.028	0.13	0.02	1	0.44	-0.19	-0.061	0.37
NEAREST_STN_DIST	-0.099	0.098	0.037	0.068	0.21	0.1	0.44	1	0.049	0.5	0.63
LATITUDE	0.043	0.0085	0.043	-0.016	-0.023	0.034	-0.19	0.049	1	-0.11	0.018
LONGITUDE	-0.18	-0.012	-0.089	0.057	0.13	-0.044	-0.061	0.5	-0.11	1	0.3
NEAREST_SCH_DIST	-0.024	0.084	0.06	0.059	0.26	0.12	0.37	0.63	0.018	0.3	1