Airbnb Linear Regression

Banan Alhethlool
banan.alhethlool@gmail.com

Modhi Alhbrdi
modhi.alhbrdi@outlook.com

Abstract:

The second project for Data science Bootcamp T5 on Regression. Through the project by building a machine learning linear regression model. The goal of this project, to apply the linear regression model on Airbnb website to predict the products prices based on many features (location, guests, beds, etc.). Using python libraries such as Pandas, beautifulsoup and other useful libraries. We have the dataset description, the methodology of the project, and analysis will be described with results.

Design:

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

Data:

The used dataset scraped from airbnb.com with 10 features and 522 rows will be helpful for the study. Sample size is 5 major cities (London, Milano, Berlin, Madrid and Amsterdam) that are located in the continent of Europe.

Algorithms:

Preparing the data, Feature Engineering, and selection:

1. *Collecting the datasets of each city and combining all cities in one clean and prepared dataset.*
2. *Feature Selection by calculate the features correlation*
3. *Engineering by converting categorical values to dummy*

Methods:

Linear regression, polynomial regression, ridge regression, and lasso regression been used to predict the products prices. By splitting the dataset to train set, validation set, and test set to measure each model scores, as reported below in **Communication**, the best model R^2 score shows in Linear regression; train set R^2 score < test set R^2 score explains the outliers in the dataset

Tools:

Technologies: Python, Jupyter Notebook.
Libraries: Pandas, NumPy, Seaborn, BeautifulSoup, Selenium and sklearn.

Communication: