

Machine Learning Model Building to Predict House Rents in Dhaka City

Using Python

Course Code: WM-ASDS04

Student Name: Mohammad Taslim Mazumder Sohel

Roll: 20231091

Section: B

Professor's Name: Farhana Afrin Duty

31 August, 2023

Abstract:

This report explains the outcomes of a data science project focused on constructing a machine learning model to predict house rents in Dhaka city. Our primary goal is, make use of machine learning techniques to develop a model that predict house rent nearly accurate based on a set of key attributes. We collected real world data from Kaggle. Where data are collected from bproperty.com using scraping technology. This website listing Dhaka's real estate information to rent. And we conducted the model using Python programming language along with popular data science libraries.

(i) Introduction:

We will develop a Machine Learning model, which is capable to predict accurate house rental prices in Dhaka city. The ability to predict rental prices has benefit for both property owners and potential tenants, understand the demand and decision-making. The project focused on examining the interplay between location, area, bedroom count, bathroom count, and rental prices. The ultimate objective was to construct a model that can generalise effectively to new, unseen data.

(ii) Data Collection and Preprocessing:

The project utilised a dataset drawn from bproperty.com who is listing Dhaka's real estate information. The dataset contains information about the property's location, area in square feet, bedrooms count, bathrooms count and corresponding rental prices. To ensure data quality we will do preprocessing. We will take steps to address missing values, handle outliers and manage categorical variables. Categorical attributes, such as location, were encoded using techniques like one-hot encoding.

(iii) Exploratory Data Analysis (EDA):

Exploratory Data Analysis involved scrutinising the distribution of rental prices, exploring relationships between attributes, prices and detecting potential correlations or patterns. We visualise the data in scatter plots, histograms, correlation matrices to derive insights of data. This phase played a central role in comprehending the data's hidden characteristics and shaping subsequent feature selection and engineering.

(iv) Feature Engineering:

Feature engineering was undertaken to augment the model's predictive prowess. New features such as price per square foot and total room count, were derived from existing attributes. This step aimed to capture latent information that could influence rental prices.

(v) Model Selection and Training:

We evaluated multiple regression algorithms for model selection. We tried,

- A. *Linear Regression*,
- B. *Ridge Regression*,
- C. *Bayesian Regression and*
- D. *Lasso Regression*.

We divided the dataset into training and testing subsets to train and test the each model's performance. We evaluated each model and compare their coefficient value to select best model.

(vi) Results:

Following comprehensive evaluation the Ridge Regression model emerged as the top performer in terms of accuracy and generalisability on unseen data. The model achieved coefficient value **91.54** on the test dataset. Which is underscoring its capacity to reasonably predict rental prices.

(vii) Conclusion:

The project successfully attained its objective of predicting house rents in Dhaka city based on attributes such as location, area, bedroom count, and bathroom count. The Ridge Regression model exhibited robustness in tackling the task with feature engineering playing a pivotal role in its heightened performance. The insights garnered from this project can help property owners and potential tenants in making informed choices related to rental properties in Dhaka city.

(viii) Future Prospects:

In future, we can make this project even better by adding more attributes. We might consider including extra information like the things people enjoy in the area the kind of building and how close it is to important places. This would help the model guess house

prices even more accurately. Also if we get a bigger collection of different data, it could help the model become even better at understanding and predicting prices.

(ix) Project Code:

https://github.com/sohel62/20231091_Taslim_Mazumder_Sohel_Project_1st_Sem/blob/main/20231091_taslim_mazumder_sohel_project_1st_sem.ipynb

(x) References:

Data source and knowledge gather related to House Rental Price,

1. <https://www.bproperty.com/>
2. <https://kaggle.com/datasets/>

- Thank You -