**Bengaluru House Price Prediction**

***BY:*** *Vidit Singh Negi*

*B.Tech(CSE)-4rd year student*

*At Diginique Techlabs*

**ABSTRACT**

This paper presents a house price prediction model utilizing Ridge Regression, built to accurately estimate property prices based on essential features such as area, number of bathrooms, BHK (bedroom, hall, kitchen) configuration, and location. The proposed model incorporates an exploratory data analysis, data preprocessing, and data cleaning phases to ensure the input data's integrity. A predictive model using Ridge Regression is developed, providing a robust and reliable mechanism to predict house prices.

The dataset used for training and evaluation has been thoroughly analyzed through exploratory data analysis to gain insights into feature distributions, correlations, and anomalies. Data preprocessing techniques have been employed to handle missing values, outliers, and feature scaling to normalize the data for model training.

A Ridge Regression model is employed due to its effectiveness in handling multicollinearity issues, commonly observed in real estate datasets with multiple correlated features. The Ridge model is trained on a portion of the cleaned dataset and evaluated using appropriate performance metrics to ensure its accuracy and generalization capabilities.

Moreover, a user-friendly web application is developed using Flask, a Python web framework, to enable easy access to the model. The web application allows users to input the area, number of bathrooms, BHK, and location of a property, and in return, provides a predicted price estimate based on the trained Ridge Regression model.

The proposed model's effectiveness is assessed through extensive testing and validation using a real-world dataset of house prices. The results demonstrate that the Ridge Regression-based model accurately predicts house prices with satisfactory performance metrics, making it a valuable tool for potential property buyers and sellers.

Overall, this study presents a comprehensive approach to house price prediction, combining data analysis, preprocessing, a Ridge Regression model, and a user-friendly web application to assist users in making informed decisions related to property investments.

**INTRODUCTION**

The real estate industry has seen significant growth in recent years, with housing prices fluctuating due to various factors such as location, amenities, and property size. Accurately predicting house prices is crucial for both buyers and sellers to make informed decisions. In this project, we have developed a house price prediction model using Ridge Regression, leveraging a dataset from Kaggle named "Bengaluru House Prices."

The dataset was subjected to thorough exploratory data analysis, data preprocessing, and data cleaning to ensure the data's quality and suitability for modeling. After these steps, four essential features were extracted, namely, "area," "number of bathrooms," "bhk" (bedroom, hall, kitchen), and "location."

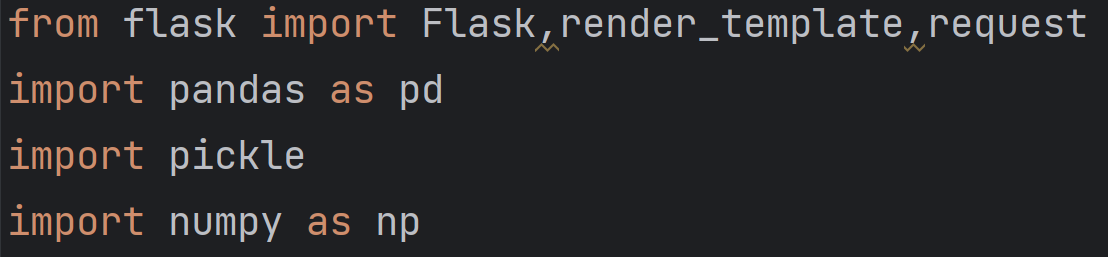
Our predictive model is built using Ridge Regression, a linear regression technique that helps prevent overfitting and provides robust predictions. The web application, developed using Flask, allows users to input the property area, the number of bathrooms, the number of bedrooms, and the location. Based on these inputs, the application will predict the house price, empowering users to make more informed property-related decisions.

Throughout this project, we aimed to create a reliable and user-friendly house price prediction tool to cater to potential homebuyers, real estate agents, and property developers in Bengaluru.

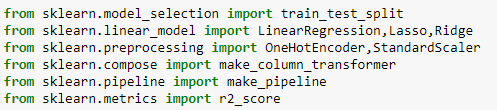
**METHODOLOGY**

STEP 1. **Environment Setup**

These are the libraries you need for data analysis and webdevelopment.

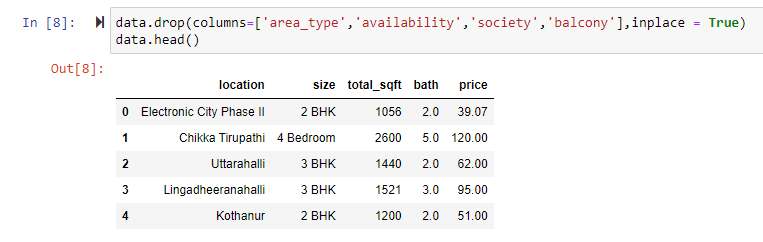


These are libraries you need for predictive modeling

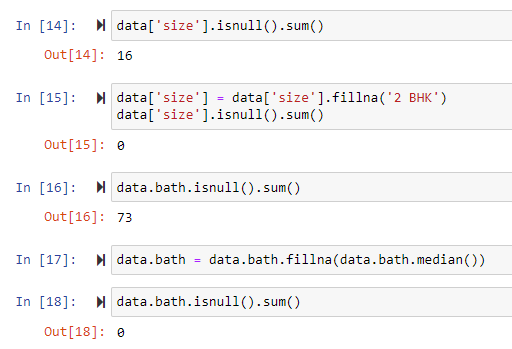


STEP 2. **Data – Preprocessing**

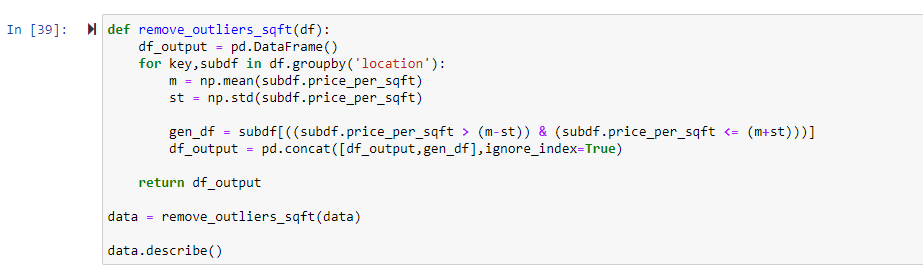
Data Cleaning – Remove all the unnecessary information and features.



Fill in the null/missing values

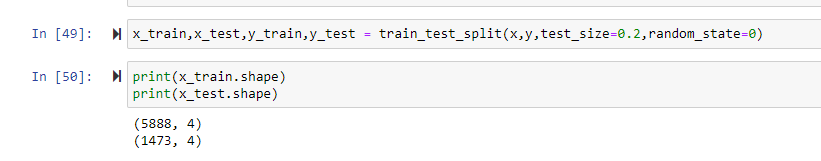


Remove outliers



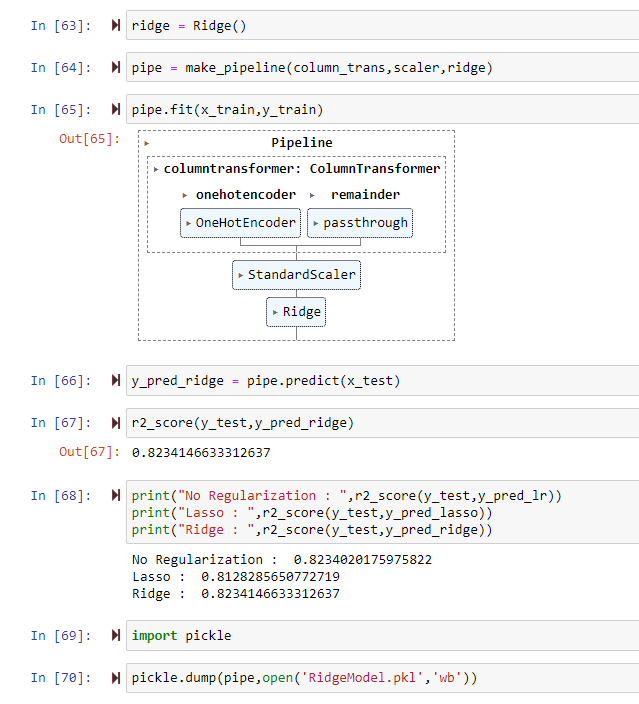
STEP 3. **Predictive Model**

Train-Test split



Making pipeline and the ridge model – Linear Regression with L2 Regularisation

Ridge regression is a linear regression technique that adds a regularization term (L2 penalty) to the ordinary least squares objective function. This penalty helps prevent overfitting and reduces the impact of multicollinearity in the data by shrinking the coefficient estimates of less important features toward zero. Ridge regression is useful for handling high-dimensional datasets and maintaining model stability while still providing interpretable results.



Export the model using pickle file.

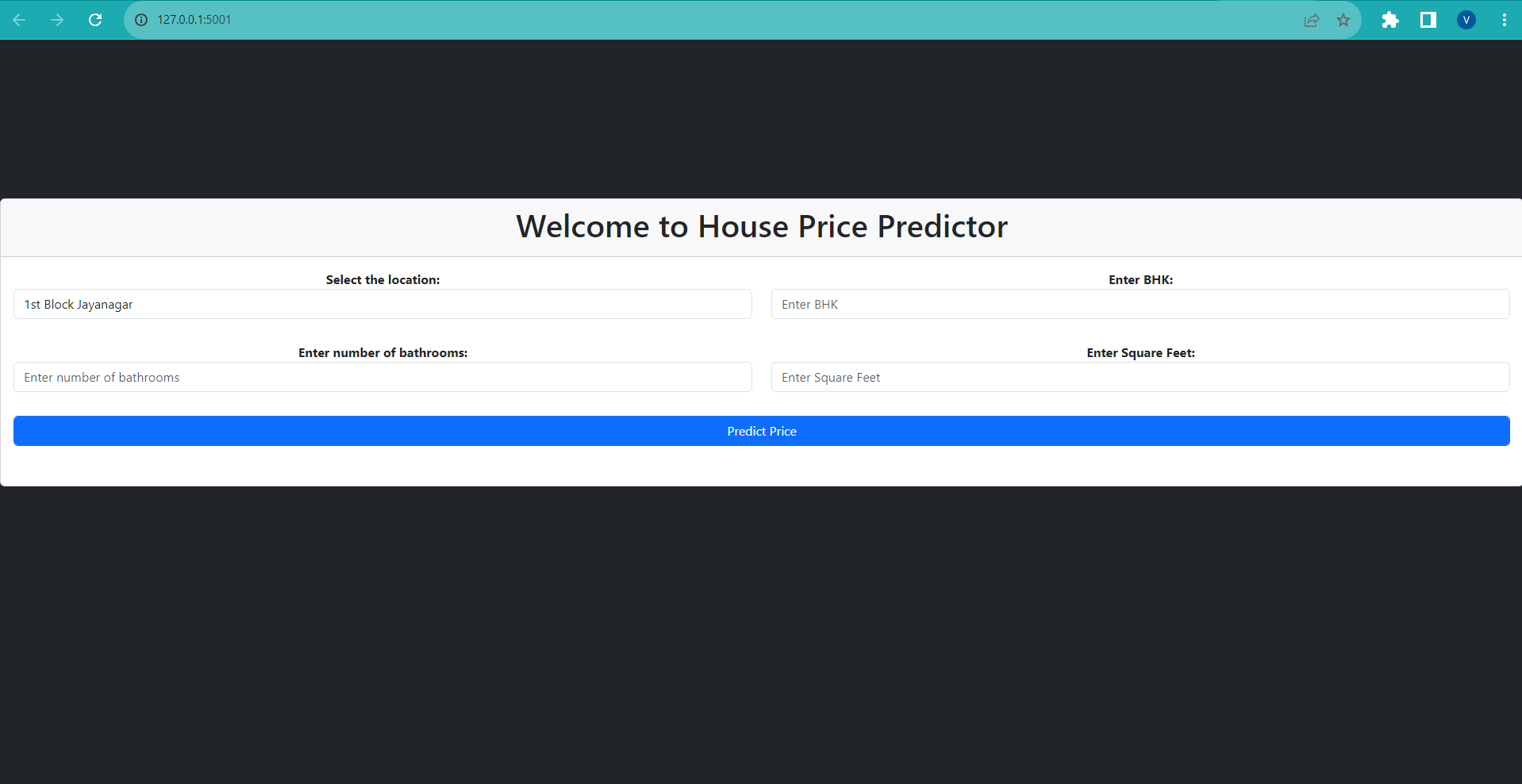
STEP 4.  **Web Application**

Import the pickle file and further use it in the working website.

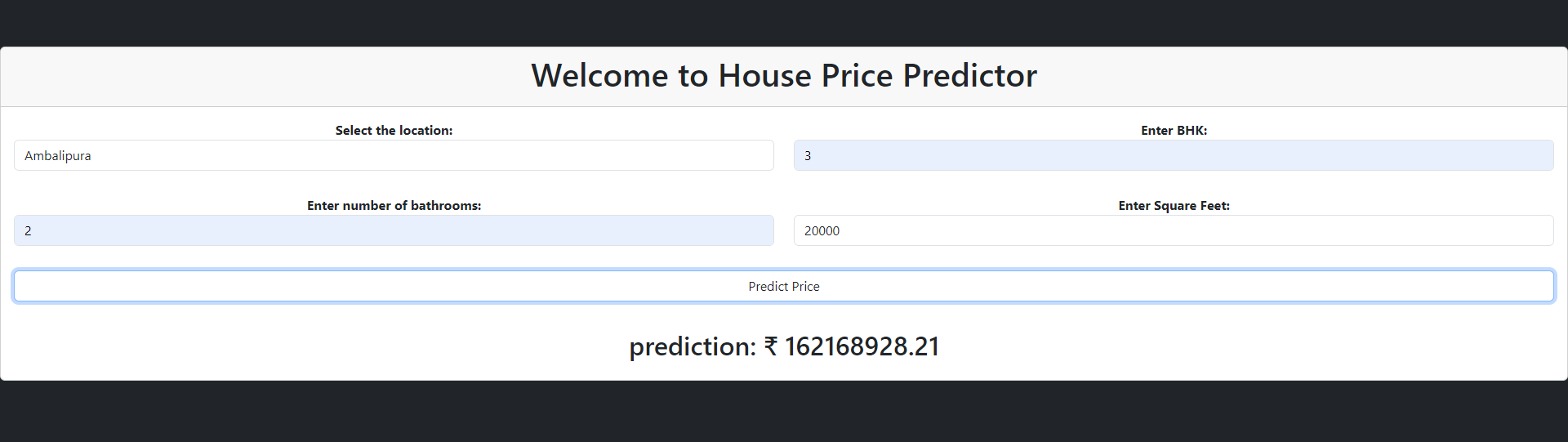
Used pycharm as the IDE.



The website. (index.html file available on github)



Fill in the input and get the prediction.



**CONCLUSION:**

In conclusion, our project successfully developed a house price prediction model using Ridge Regression, which is integrated into a user-friendly web application using Flask. The model takes key property attributes like area, number of bathrooms, bhk, and location as inputs and accurately predicts the house price.

The process began with thorough exploratory data analysis, data preprocessing, and data cleaning to ensure that the dataset is accurate and devoid of any anomalies. Feature extraction was performed to identify the most relevant attributes for the predictive model.

Ridge Regression, chosen for its ability to handle multicollinearity and prevent overfitting, was employed to train the model. The model demonstrated strong performance in predicting house prices in Bengaluru.

The web application serves as a practical tool for homebuyers, sellers, real estate agents, and property developers, providing valuable insights into property valuation based on various attributes. The interactive nature of the application allows users to experiment with different inputs and explore how each feature affects the predicted house price.

Overall, this project offers a valuable contribution to the real estate domain by providing an accurate and easily accessible tool for house price prediction. However, continuous improvements and updates to the model and application can be made, incorporating more data sources and implementing additional features for enhanced accuracy and usability.

As the real estate market evolves, incorporating new trends and insights into the model will further enhance its predictive capabilities, making it an indispensable asset for anyone involved in the Bengaluru property market.

Links- <https://github.com/Viditnegi/Real-Estate-Price-Predictor>

**REFERENCES**:

# Dataset - <https://www.kaggle.com/datasets/amitabhajoy/bengaluru-house-price-data>

Flask - <https://flask.palletsprojects.com/en/2.3.x/quickstart/#a-minimal-application>

Bootstrap - <https://getbootstrap.com/docs/5.3/getting-started/introduction/>