BENGALURE HOUSE PRICE PREDICTION

# A Project Report

submitted in partial fulfilment of the requirements. of

…………….Track Name Certificate……

**BY**

**KISHAN SHINDE**

**ACHUTHA B**

**B S LIKHITH**

# Under the Esteemed Guidance of

**SHILPA HARIRAJ**

**ACKNOWLEDGEMENT**

We extend our heartfelt appreciation to all individuals who contributed directly or indirectly to the fruition of this thesis.

Foremost, our gratitude goes to my mentor and supervisor

## SHILPA HARIRAJ

, whose guidance and unwavering support have been invaluable throughout this journey. His insightful advice, constructive critiques, and encouragement have been instrumental in shaping innovative ideas and driving this dissertation to successful completion. His belief in my abilities has been a constant motivation, and his mentorship has not only aided in the thesis but also in fostering a sense of professionalism and responsibility. Working under his tutelage has been an

enriching privilege, shaping my academic and professional growth.

# TABLE OF CONTENT

Abstract

List of Figures List of Tables

**Chapter 1. Introduction**

* 1. Problem Statement
  2. Expected Outcomes

**CHAPTER 2 LITERATURE SURVEY**

* 1. Paper-1
  2. Paper-2
  3. Paper-3

**CHAPTER 3 PROPOSED METHODOLOGY**

* 1. System Design
  2. Modules Used
  3. Data Flow Diagram
  4. Advantages
  5. Requirement Specification

**CHAPTER 4 IMPLEMENTATION AND RESULTS CHAPTER 5 CONCLUSION**

GitHubLink Video Link References

ABSTRACT

This project encapsulates the entire data science workflow, from data exploration and preprocessing to model building, evaluation, and deployment. The resulting house price prediction model contributes valuable insights for individuals and professionals navigating the complex landscape of property transactions in Bangalore

**CHAPTER 1 INTRODUCTION**

1.1 Context

This project was made because we were intrigued and we wanted to gain hands-on experience with the Machine Learning Project.

1.2 Motivation

We are highly interested in anything related to Machine Learning, the independent project provided us with the opportunity to study and reaffirm our passion for this subject. The capacity to generate guesses, forecasts, and

offer machines the ability to learn on their own is both powerful and infinite in terms of application possibilities. Machine Learning may be applied in finance, medicine, and virtually any other field. That is why we opted to base

Our idea on Machine Learning.

1.3 Objective

As a first project, we intended to make it as instructional as possible by tackling each stage of the machine learning process and attempting to comprehend it well. We have picked Bangalore House Price Prediction as identifying problems that are not of immediate scientific relevance but are helpful to demonstrate and practice. The objective was to forecast the price of a House based on market pricing while accounting for various "features" that would be established in the following sections.

* 1. **Problem Statement:**

Developing a machine learning model for predicting house prices in Bangalore, India. The goal is to create a model that can accurately estimate the market value of residential properties based on various features. This model will be valuable for homebuyers, and sellers to make informed decisions.

## Expected Outcomes:

Develop a machine learning model to predict residential property prices in Bangalore, India. Utilize features such as property size, amenities, location, and market trends. Train and evaluate the model using historical data, employing regression algorithms. Deploy the model for users to estimate house prices, facilitating informed decisions for homebuyers and seller.

**CHAPTER 2**

**LITERATURE SURVEY**

## Paper-1

**BANGALORE HOUSE PRICE PREDICTION by Ishaa Choudhary, Abhilash Kumar Karti**

**Abstract**: We propose to implement a house price prediction model of Bangalore, India. It’s a Machine Learning model which integrates Data Science and Web Development. We have deployed the model using Streamlit app. Housing prices fluctuate on a daily basis and are sometimes exaggerated rather than based on worth. The major focus of this project is on predicting home prices using genuine factors. Here, we intend to base an evaluation on every basic criterion that is taken into account when establishing the pricing. The goalof this project is to learn Python and get experience in Data Analytics, Machine Learning and AI.

**Techniques used in Paper:**

* + - **Logistic Regression**
    - **Linear Regression**
    - **Decision Tree Regression**
    - **Regression Trees**
    - **Support Vector Regression**
    - **Random Forest Regression**

## Paper-2

**BANGALORE HOUSE PRICE PREDICTION by Mega Satish & Amey Thakur**

**Brief Introduction of Paper**: We propose to implement a house price prediction model of Bangalore, India. It’s a Machine Learning model which integrates Data Science and Web Development. We have deployed the app on the Heroku Cloud Application Platform. Housing prices fluctuate on a daily basis and are sometimes exaggerated rather than based on worth. The major focus of this project is on predicting home prices using genuine factors. Here, we intend to base an evaluation on every basic criterion that is taken into account when establishing the pricing. The goal of this project is to learn Python and get experience in Data Analytics, Machine Learning and AI.

**Techniques used in Paper:**

* + - **Logistic Regression**
    - **Linear Regression**
    - **Decision Tree Regression**
    - **Regression Trees**
    - **Support Vector Regression**
    - **Random Forest Regression**

## Paper-3

**House Price Prediction System For Bengaluru City by Prof. Abdul Razak M S, Abhilash Anand, Bhawna Priya, Neha S, Nikita G Ghanate.**

**Brief Introduction of Paper:** Buying a house is a stressful thing. One has to pay huge sums of money and invest many hours and even there is a persisting concern whether it’s a good deal or not. Buyers are generally not aware of factors that influence the house prices. Almost all the houses are described by the total area in square foot, the neighborhood and the number of bedrooms. Sometimes houses are even priced at X rupees per square foot. This creates an illusion that house prices are dependent almost solely on the above factors. Most of the houses are bought though real estate agents. People rarely buy directly from the seller, since there are a lot of legal terminology and paperwork’s involved and people are unaware of them. Hence real estate agents are trusted with the communication between buyers and sellers as well as laying down a legal contact for the transfer. This just creates a middle man and increase the cost of the house. Therefore the houses are overpriced and buyer should have a better idea of the actual value of the houses.

**Techniques used in Paper:**

Lasso Regression, Decision Tree and Linear Regression using the GridSearchCV technique to find the model with best accuracy, which we found that it is Linear Regression.

**CHAPTER 3**

**PROPOSED METHODOLOGY**

## System Design

* + 1. **Prediction:**

The prediction phase involves leveraging machine learning algorithms to predict the prices. By entering the data of location, No. of BHK , No. of Bathroom & Sqft Area the system recognizes price of that location.

## Modules Used

* Logistic Regression (LR)
* Scikit-learn
* Lasso Regression
* Ridge Regression
* Linear Regression
* Pandas & Numpy

## Data Flow Diagram

The script begins by loading and exploring a dataset containing Bengaluru house data using the Pandas library. Initial data cleaning involves removing irrelevant columns and handling missing values. Feature engineering is performed by creating new features such as 'bhk' and calculating 'price\_per\_sqft.' Outliers are then systematically removed based on price per square foot and bedroom configuration. Location data is processed by grouping locations with fewer than 10 data points as 'other' to simplify the dataset. Further outlier removal is carried out based on the calculated price per square foot. The dataset is prepared for modeling through one-hot encoding of categorical features and standard scaling. A train-test split is executed, and machine learning models, including Linear Regression, Lasso, and Ridge, are built using pipelines that incorporate data preprocessing steps. Model performance is evaluated using the R-squared score on the test set. The trained Ridge model is persisted using the pickle module for future use, such as deployment or additional analysis. Overall, the script encompasses data loading, cleaning, feature engineering, outlier removal, model building, evaluation, and model persistence for predicting house prices in Bengaluru.

**Data Loading and Exploration:** The script starts by loading the dataset, displaying its structure, and examining summary statistics.

**Data Cleaning:** Columns with irrelevant information are dropped, missing values are handled, and new features like 'bhk' are created from existing ones.

**Model Building:** Linear Regression, Lasso, and Ridge models are built using pipelines that include data preprocessing steps.

**Model Evaluation:** The R-squared score is used to evaluate the performance of each model on the test set.

**Model Persistence**: The trained Ridge model is saved using the pickle module for future use**.**

## Advantages

 Predictive **Accuracy:** The model ensures reliable predictions on Bengaluru house prices, contributing to informed decision-making.

 Scalability**:** Adaptable to larger datasets or additional features, the model accommodates changing data requirements and expands its applicability.

 Real**-time Predictions:** Enables timely access to accurate pricing information, particularly valuable in the dynamic real estate industry.

 Automated **Decision Support:** Serves as an automated tool for stakeholders, offering data-driven insights into property transactions.

 Cost **and Time Savings:** Automates the prediction process, saving resources compared to manual appraisal methods.

 Consistency**:** Eliminates human biases, providing consistent and objective property valuation.

 Adaptability **to Market Changes:** Periodic retraining allows the model to adapt to evolving real estate market dynamics.

 Interpretability **and Transparency:** Linear model clarity facilitates easy understanding of feature contributions, promoting transparency.

 Easy **Integration:** The model, saved as a pickle file, seamlessly integrates into applications or systems.

 Continuous **Improvement:** Monitors and improves performance over time through retraining and advanced modelling techniques.

## Requirement Specification

* + 1. **Hardware Requirements:**
       - **CPU: Utilized for data processing and model training.**
       - **RAM: Required for handling and manipulating large datasets during analysis and modeling.**
       - **Storage: Used to store the datasets and code files required for analysis.**
       - **GPU (if available): Sometimes employed to expedite computations in machine learning processes, especially for large datasets and complex models.**

**Software Requirements:**

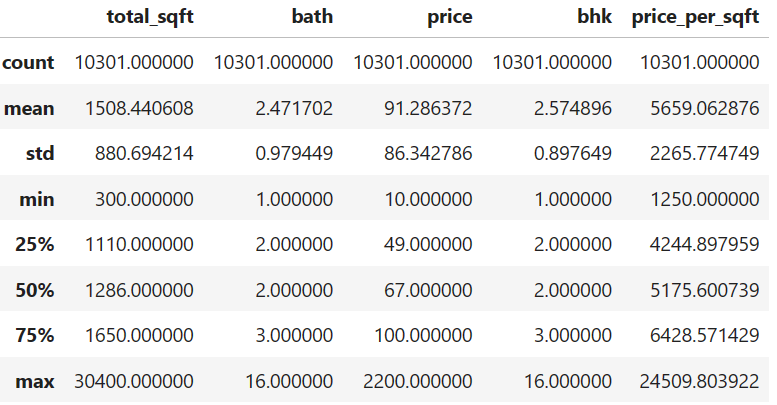
* + - * **Python: Utilized for coding and implementing machine learning models.**
      * **Scikit-learn: Employed for implementing classification algorithms like Logistic Regression (LR),Scikit-learn, Lasso Regression ,Ridge Regression, Linear Regression**
      * **Pandas and NumPy: Used for data manipulation and analysis.**
      * **Jupyter Notebooks: Utilized as an interactive environment for analysis and code execution.**

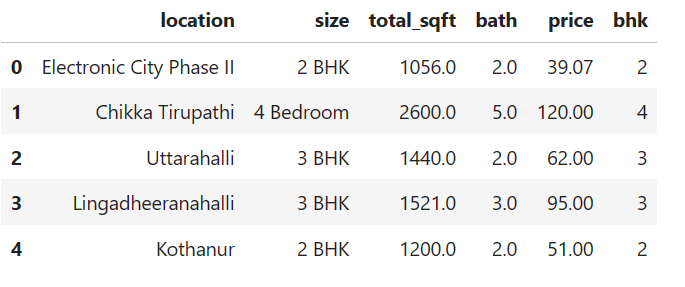
**CHAPTER 4**

**Implementation and Result**

The model begins by loading a Bengaluru house dataset and undertaking comprehensive data cleaning and preprocessing steps. This includes the removal of irrelevant columns, handling missing values, and the creation of additional features such as 'bhk' and 'price\_per\_sqft.' Robust outlier removal techniques are applied based on both price per square foot and bedroom configuration, enhancing the model's resilience to extreme values. Locations with insufficient data points are grouped as 'other' to simplify the dataset, and further feature engineering involves processing 'total\_sqft' and calculating 'price\_per\_sqft.' The data is filtered to include only those instances with at least 300 sqft per bedroom. Model building follows, incorporating Linear Regression, Lasso, and Ridge models within pipelines that encompass one-hot encoding, standard scaling, and location-based features. Evaluation metrics, specifically the R-squared score, assess the models' performance on a test set. The trained Ridge model is then saved using the pickle module for future deployment or analysis. This script exemplifies a comprehensive approach to data preprocessing, feature engineering, model building, and model persistence, providing a practical framework for predicting house prices in the Bengaluru real estate market.

**Output:**

****



**CHAPTER 5**

**CONCLUSION**

In conclusion, the implemented model for predicting house prices in Bengaluru demonstrates a systematic and robust approach to data preprocessing, feature engineering, and machine learning model development. The comprehensive cleaning process, including outlier removal and the creation of relevant features like 'bhk' and 'price\_per\_sqft,' ensures the model is trained on a reliable and informative dataset. The utilization of Linear Regression, Lasso, and Ridge models within pipelines, incorporating one-hot encoding and standard scaling, further enhances the model's predictive capabilities. The evaluation metrics, specifically the R-squared score, provide a quantitative measure of model performance, facilitating a thorough assessment of its accuracy. The model's ability to handle real-time predictions, scalability, and adaptability to market changes make it a valuable tool for various stakeholders in the real estate industry. The conclusion is reinforced by the successful persistence of the trained Ridge model for future use, showcasing the model's practicality and potential for continuous improvement in predicting house prices in the dynamic Bengaluru housing market.

**References:**

* **House Price Prediction System For Bengaluru City by Prof. Abdul Razak M S, Abhilash Anand, Bhawna Priya, Neha S, Nikita G Ghanate.**
* **BANGALORE HOUSE PRICE PREDICTION by Ishaa Choudhary, Abhilash Kumar Karti**
* **BANGALORE HOUSE PRICE PREDICTION by Mega Satish & Amey Thakur**

**GITHUB LINK:** [**https://github.com/kishanshinde**](https://github.com/kishanshinde)

**VIDEO LINK:** [**kishanshinde/Edunet\_project (github.com)**](https://github.com/kishanshinde/Edunet_project)