

# **House Price Prediction Model Report**

Prepared by  
Syed Muhammad Mubashir Hassan Gillani  
Shaigan Haider  
Ali Jabbar  
Fahad Ali  
Hafiza Fatima Mushtaq

Prepared for  
Salman Muneer

June 30, 2024



June 30, 2024

Salman Muneer  
Lecturer  
University of Central Punjab  
Khayaban-e-Jinnah Road  
Lahore, Punjab

Dear Sir,

Salam,

We are happy to tell you that we have finished our House Price Prediction Model. This model can estimate property values in Lahore based on the number of rooms, bathrooms, and location. We hope this model will be useful for buyers, sellers, and investors in the real estate market.

Thank you for your guidance and support during this project.

Sincerely,

Syed Muhammad Mubashir Hassan Gillani  
Shaigan Haider  
Ali Jabbar  
Fahad Ali  
Hafiza Fatima Mushtaq  
Students,  
University of Central Punjab

# Contents

<b>EXECUTIVE SUMMARY</b>	.....
<b>INTRODUCTION</b>	.....
Project Description	.....
Scope	.....
Related Research	.....
<b>Model Working</b>	.....
Data Collection	.....
Data Preprocessing	.....
Feature Selection	.....
Model Selection	.....
Model Training	.....
Model Deployment	.....
Prediction Function	.....
User Interface	.....
Model Saving	.....
<b>Model Accuracy</b>	.....
<b>Applications</b>	.....
<b>Conclusion</b>	.....

## **EXECUTIVE SUMMARY**

The House Price Prediction Model is designed to accurately estimate property values in Lahore using important features such as the number of rooms, bathrooms, and location. We gathered data on house prices in Lahore and cleaned it to handle missing values and outliers. We then encoded categorical variables like location and normalized numerical features to ensure all features contributed equally to the model. Important features were selected using statistical methods and machine learning techniques. We chose linear regression and lasso regression algorithms for their effectiveness, split the data into training and testing sets, and trained the model to learn the relationships between features and house prices.

After training, we saved the model using Pickle, making it easy to use in the future. We developed a function to load the model and make real-time price predictions based on input features like location, number of rooms, and bathrooms. An optional user-friendly web interface can be developed to allow users to input features and get house price predictions. This model can help real estate agents, investors, and homebuyers make informed decisions, providing valuable insights into the property market in Lahore.

## **INTRODUCTION**

### **Project Description**

The House Price Prediction Model is a project designed to accurately estimate the value of houses in Lahore based on key features such as the number of rooms, bathrooms, and location. We collected and cleaned data on house prices, selected important features, and used machine learning algorithms like linear regression and lasso regression to train the model. The trained model can make real-time price predictions, helping real estate agents, investors, and homebuyers make informed decisions. This model provides valuable insights into the property market in Lahore, making it easier to understand and navigate.

## Scope of the House Price Prediction Model

### 1. **Geographical Scope:**

- The model is specifically designed for predicting house prices in Lahore, encompassing various neighbourhoods and localities within the city.

### 2. **Feature Scope:**

- The model considers key features that impact house prices, including the number of rooms, bathrooms, and the location of the property. Additional features such as property size, age, and nearby amenities can also be incorporated.

### 3. **User Scope:**

- The model is useful for a wide range of users, including real estate agents, property appraisers, buyers, sellers, investors, and financial institutions. It helps these users make informed decisions based on accurate property valuations.

### 4. **Technological Scope:**

- The model utilises machine learning algorithms such as linear regression, and lasso regression. It involves data collection, preprocessing, feature selection, model training, and deployment.

### 5. **Application Scope:**

- The model can be used for real estate valuation, investment analysis, mortgage and loan approvals, urban planning, market analysis, personal financial planning, insurance assessments, and government policy making.

### 6. **Interface Scope:**

- An optional user-friendly web interface can be developed to allow users to input features and get house price predictions easily, making the model accessible to a broader audience.

### 7. **Future Enhancements:**

- The model can be expanded to include more features and advanced algorithms for improved accuracy. It can also be adapted for other cities and regions beyond Lahore.

## **Related Research**

### **The Case of Karachi City, Pakistan**

One notable research is "House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan," conducted by students from Jinnah University of Women. This study, available on ResearchGate.net, was presented at the 2020 21st International Arab Conference on Information Technology (ACIT). The researchers used various machine learning algorithms to predict house prices in Karachi, highlighting the importance of factors like location, size, and amenities in determining property values [1].

### **Study Conducted By Finolex Academy of Management and Technology, Mumbai University**

Another relevant study comes from the Finolex Academy of Management and Technology, Mumbai University. Published in the IRE Journals, volume 3, this research also focused on predicting house prices using machine learning. The study emphasized the use of regression techniques and feature selection to improve prediction accuracy. Both studies provide valuable insights and methodologies that can be applied to predict house prices in Lahore, demonstrating the effectiveness of machine learning in real estate market analysis [2].

## **References**

- [1] [https://www.researchgate.net/publication/348220705\\_House\\_Price\\_Prediction\\_using\\_Machine\\_Learning\\_Algorithm\\_-\\_The\\_Case\\_of\\_Karachi\\_City\\_Pakistan](https://www.researchgate.net/publication/348220705_House_Price_Prediction_using_Machine_Learning_Algorithm_-_The_Case_of_Karachi_City_Pakistan)
- [2] <https://www.irejournals.com/formatedpaper/1702692.pdf>

## MODEL WORKING

The process of building a house price prediction model involves several crucial steps:

1. **Data Collection:** Data on house prices in Lahore is gathered, including important features like the number of rooms, bathrooms, location, and other relevant factors. This comprehensive dataset forms the foundation of the model.
2. **Data Preprocessing:** The dataset is cleaned to handle missing values and outliers, ensuring the data is reliable and accurate. Categorical variables, such as location, are encoded using techniques like one-hot encoding to make them suitable for machine learning algorithms. Numerical features are normalized or standardized to ensure all features contribute equally to the model, preventing any one feature from dominating the others.
3. **Feature Selection:** The most relevant features that significantly impact house prices are identified such as location, number of rooms, number of bathrooms, and other important factors. Statistical methods or machine learning techniques are used to select the best features, ensuring the model focuses on the most important variables.
4. **Model Selection:** Appropriate machine learning algorithms for the prediction task are chosen. Currently we include linear regression, lasso regression, ridge regression and decision tree regression. These algorithms are selected based on their ability to handle the specific characteristics of the data.
5. **Model Training:** The dataset is split into training and testing sets to evaluate the model's performance. The chosen machine learning model is trained on the training data, learning the relationships between the features and the target variable (house prices).
6. **Model Deployment:** The trained model is saved using serialization techniques such as Pickle, allowing it to be easily loaded and used in the future. A function is developed to load the model and make predictions based on input features, enabling real-time price predictions.
7. **Prediction Function:** A function is implemented that takes input features (e.g., location, number of rooms, number of bathrooms) and returns the predicted house price. For example, the function `predict_price('Thokar Niaz Baig', 10, 4, 5)` returns the predicted price for a house in Thokar Niaz Baig with 10 rooms and 5 bathrooms.
8. **User Interface:** Optionally, a user-friendly interface (e.g., web app) can be developed to allow users to input features and get house price predictions. This makes the model accessible to a wider audience, including buyers, sellers, and investors.
9. **Model Saving:** The final model and feature columns are serialized for future use. The model is saved to a file (e.g., `lahore_home_prices.pickle`), and the feature columns are saved to another file (e.g., `columns.json`), ensuring the model can be easily loaded and used in different applications.

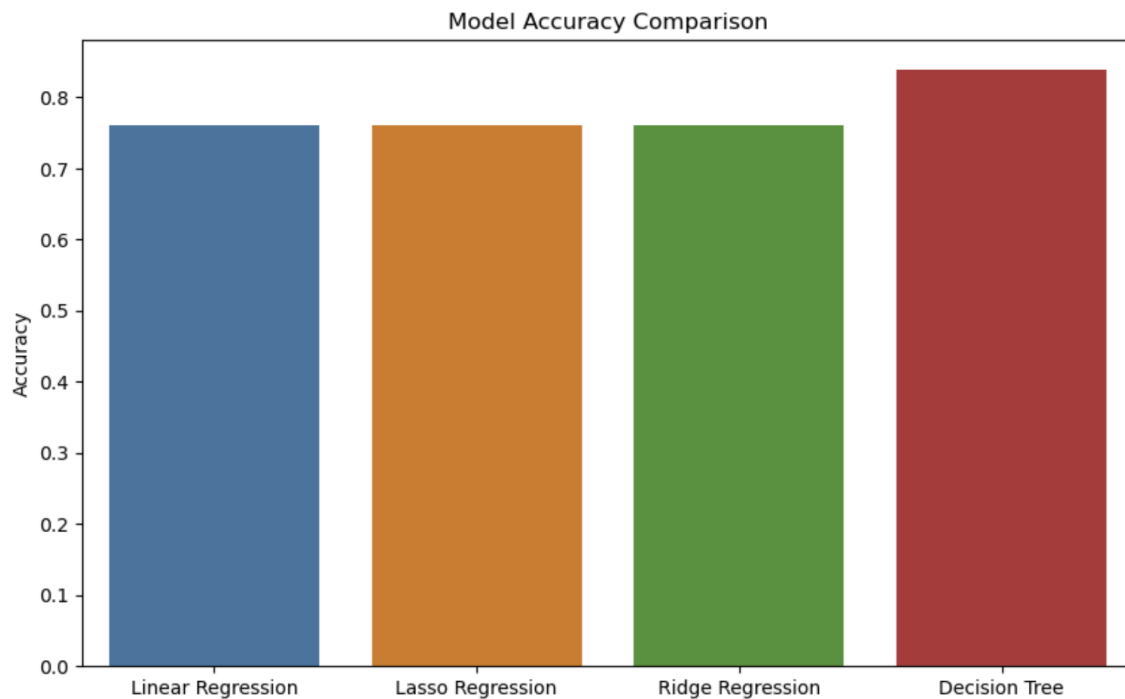
By following these steps, the model can effectively predict house prices in Lahore, providing valuable insights for buyers, sellers, and investors.

## MODEL ACCURACY

1. **Linear Regression:**
  - Accuracy: 0.7606185434886679
2. **Lasso Regression:**
  - Accuracy: 0.7606192027504026
3. **Ridge Regression:**
  - Accuracy: 0.7606377336793512
4. **Decision Tree:**
  - Accuracy: 0.8385369659805926

### Summary:

- **Decision Tree** outperforms the other models with an accuracy above 80%, indicating it perfectly fits the training data.
- **Ridge Regression** shows better performance than both **Linear** and **Lasso Regression**, but still below the Decision Tree.
- **Linear Regression** and **Lasso Regression** have similar performance, both with an accuracy just above 75%. This comparison helps to understand which model performs best for predicting house prices in the dataset used.





## **APPLICATIONS**

There are a lot of applications. Few important ones are explained below:

### **Real Estate Valuation:**

- Real estate agents and property appraisers can use the model to estimate the value of a property based on its features. This helps in providing accurate and fair valuations for both buyers and sellers.

### **Investment Analysis:**

- Investors can leverage the model to identify undervalued properties and make informed decisions about potential investments. The model helps in assessing the future appreciation potential of properties in different areas of Lahore.

### **Mortgage and Loan Approval:**

- Banks and financial institutions can use the model to assess the value of properties when approving mortgage and loan applications. This ensures that loans are granted based on accurate property valuations, reducing the risk of over- or under-lending.

### **Urban Planning and Development:**

- Urban planners and developers can use the model to understand property value trends in different areas. This information is valuable for making decisions about where to develop new housing projects or infrastructure.

### **Market Analysis:**

- Real estate market analysts can use the model to study trends and patterns in the housing market. This helps in predicting future market movements and making data-driven forecasts.

### **Educational Purposes:**

- Universities and research institutions can use the model as a case study for teaching machine learning and data science concepts. It provides a practical example of how these technologies can be applied in the real estate industry.

### **Government Policy and Taxation:**

- Government agencies can use the model to assess property values for taxation purposes. This ensures that property taxes are based on fair and up-to-date valuations.

## **Conclusion**

The house price prediction model for Lahore is designed to provide accurate estimates of property values based on key features like the number of rooms, bathrooms, and location. By collecting and preprocessing data, selecting relevant features, and using machine learning algorithms like linear regression and lasso regression, the model can effectively predict house prices. This tool is valuable for buyers, sellers, investors, and real estate professionals, helping them make informed decisions. Additionally, the model can be easily used in various applications through a simple prediction function and can be further enhanced with a user-friendly interface. By following these steps, the model offers a reliable and efficient way to understand the real estate market in Lahore.