A Mini Project Report on

House Price Prediction

T.E. - I.T Engineering

Submitted By

Kalpesh Chavan	**20104079**
Ankit Awade	**20104082**
Siddhant Darekar	**20104111**

Under The Guidance Of

Prof. Neha Deshmukh

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year : 2022-23

CERTIFICATE

This to certify that the Mini Project report on House Price Prediction has been submitted by Kalpesh Chavan (20104079), Ankit Awade (20104082) and Siddhant Darekar (20104111) who are a Bonafede students of A. P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2022-2023** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Neha Deshmukh
Guide

Dr. Kiran Deshpande
HOD , Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)
1.
2.

Place: A.P.Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof. Neha Deshmukh**. Expressing gratitude towards our HoD, **Dr. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher Ms. Charul Singh who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

ABSTRACT

This project aims to develop a linear regression model to predict house prices based on various features such as location, size, number of bedrooms, bathrooms, and other relevant factors. The model will be trained on a dataset containing information about houses sold in a particular area over a certain period. The data will be preprocessed and analyzed to identify patterns and relationships between different variables. The linear regression model will be built using a subset of the most relevant features, selected through feature selection techniques such as correlation analysis and stepwise regression. The performance of the model will be evaluated using metrics such as mean squared error, R-squared, and root mean squared error. Finally, the model will be deployed as a web application to allow users to input house features and get a predicted price. This project will provide valuable insights for homebuyers, real estate agents, and property investors, helping them make informed decisions based on accurate price predictions.

TABLE OF CONTENTS

1. Introduction

The real estate market is an essential sector of any economy, and accurate price predictions are vital for homebuyers, real estate agents, and property investors to make informed decisions. However, predicting the price of a house accurately can be challenging, as many factors, such as location, size, condition, and amenities, can impact its value. In recent years, machine learning techniques have emerged as a promising approach to accurately predict house prices based on relevant features. In this project, we aim to develop a machine learning model using linear regression to predict house prices based on a selected set of features, such as location, size, number of bedrooms and bathrooms, and other relevant factors. We will collect and preprocess a dataset containing information about houses sold in a particular area over a certain period, identify relevant features that are most strongly correlated with house prices, and develop a machine learning model to predict house prices. The model's performance will be evaluated using metrics such as mean squared error, R-squared, and root mean squared error, and the best-performing model will be deployed as a web application to allow users to input house features and get a predicted price. This project will provide valuable insights for homebuyers, real estate agents, and property investors, helping them make informed decisions based on accurate price predictions.

1.1.Purpose :

The purpose of this project is to develop a machine learning model that can accurately predict the price of a house based on various features such as location, size, number of bedrooms, bathrooms, and other relevant factors. The model will provide valuable insights for homebuyers, real estate agents, and property investors, helping them make informed decisions based on accurate price predictions.

1.2.Problem Statement :

The problem addressed by this project is the difficulty that homebuyers, real estate agents, and property investors face in accurately predicting the price of a house. Many factors, such as location, size, and condition of the house, can affect its price, making it challenging to estimate accurately. This project aims to develop a machine learning model that can overcome these challenges and provide accurate price predictions.

1.3.Objectives :

- Collect and preprocess a dataset containing information about houses sold in a particular area over a certain period.
- Identify relevant features that are most strongly correlated with house prices.
- Develop a machine learning model using linear regression to predict house prices based on the selected features.
- Evaluate the performance of the model using metrics such as mean squared error, R-squared, and root mean squared error.
- Deploy the model as a web application to allow users to input house features and get a predicted price.

1.4.Scope :

- The project will focus on developing a machine learning model using linear regression to predict house prices based on a selected set of features.
- The dataset used in the project will contain information about houses sold in a particular area over a certain period.
- The project will involve preprocessing and cleaning the data to ensure that it is suitable for use in developing the machine learning model.
- Relevant features that are most strongly correlated with house prices will be identified using techniques such as correlation analysis and stepwise regression.
- The best-performing machine learning model will be deployed as a web application to allow users to input house features and get a predicted price.
- The project does not include data collection or the development of a user interface beyond the web application.

.

2. Literature Review

Sr.no	Title	Author(s)	Year	Outcomes	Methodology	Result
1.	A Review of House Price Prediction Models	Luo et al.	2020	Machine learning models are effective in predicting house prices, with neural networks outperforming other models. Factors such as location, size, and the number of rooms are the most important features for predicting house prices.	Review and analysis of research papers and articles related to house price prediction models. Comparison of machine learning models based on various metrics.	Neural networks are the most effective model for predicting house prices.
2.	A Comparative Study of Regression Techniques for House Price Prediction	Kumar and Kumar	2018	Random forests outperform other regression techniques in predicting house prices. Feature selection and preprocessing significantly improve the accuracy of the models.	Collection of house prices and features dataset from a real estate website. Comparison of various regression techniques on the dataset.	Random forests are the most effective regression technique for predicting house prices.
3.	House Price Prediction Using Machine Learning: A Comparative Study	Deshmukh et al.	2019	The k-nearest neighbors algorithm outperforms other machine learning algorithms in predicting house prices. Feature selection and feature engineering significantly improve the accuracy of the models.	Collection of house prices and features dataset from a real estate website. Comparison of various machine learning algorithms on the dataset. Feature selection and feature engineering performed to improve model accuracy.	The k-nearest neighbors algorithm is the most effective machine learning algorithm for predicting house prices.

3. Proposed System

3.1.Features and Functionality

- User input: Allow users to input details about the house they want to sell or buy.
- Data processing: Process and clean the user input data to prepare it for machine learning models.
- Machine learning models: Develop and train machine learning models to predict the house price based on user input data.
- Prediction output: Provide the predicted house price to the user based on their input data.
- Comparison: Allow users to compare their predicted house price with the actual price of similar houses in the area.
- Feedback: Allow users to provide feedback on the accuracy of the predicted house price to improve the machine learning models.

4. Requirements Analysis

Importance of Requirements Gathering:

Requirements gathering for house price prediction involves collecting and analyzing data on various features of houses, such as location, size, number of bedrooms and bathrooms, age, etc. It also involves collecting data on the prices of these houses. The quality and quantity of the data collected can have a significant impact on the accuracy of the predictions. Therefore, it is important to carefully choose the sources of data and ensure that they are reliable and representative of the target population. Additionally, it is important to consider any legal or ethical requirements related to data collection and use.

Need Analysis:

- Identify the problem to be solved: predicting house prices based on various features of the houses.
- Determine the target audience: potential home buyers, real estate agents, investors, etc.
- Consider the scope of the project: will it be limited to a particular geographic area or housing market, or will it be more broad-based?
- Evaluate the availability and quality of data: are there reliable sources of data on housing prices and features?

Key Requirements:

- Accurate and reliable predictions of house prices.
- User-friendly interface that allows users to input features of a house and obtain a predicted price.
- Ability to handle a large volume of data and perform calculations quickly.
- Ability to handle missing or incomplete data and make predictions based on available information.
- Adherence to legal and ethical requirements related to data collection and use.

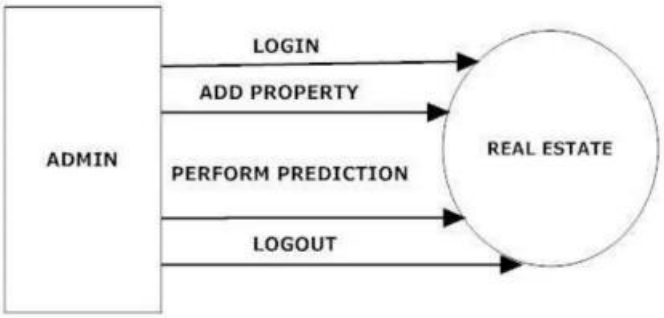
Functional Requirements:

- Data cleaning and preprocessing to remove errors, outliers, and missing values.
- Multiple linear regression model to account for various features that affect house prices.
- Training of the model on a large dataset of historical housing data.
- Evaluation of the model's accuracy using performance metrics such as mean squared error or R-squared.
- Tuning of the model by adjusting its parameters or selecting a different algorithm to improve its accuracy.
- Use of the model to predict the price of a new house based on its features.

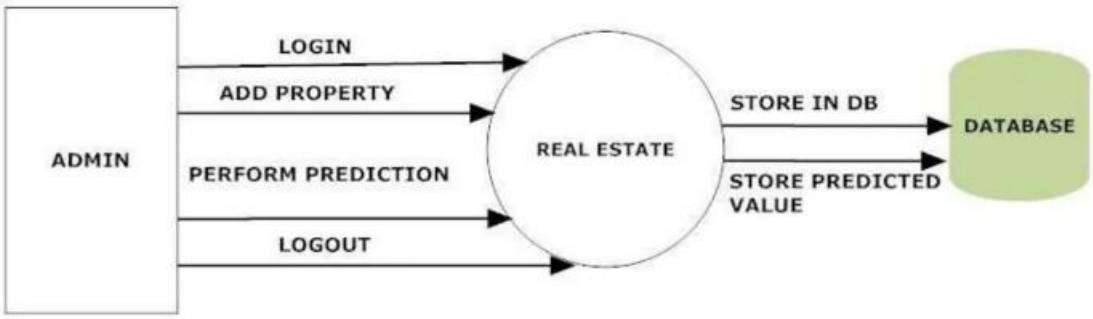
5. Project Design

5.1.Use Case diagram

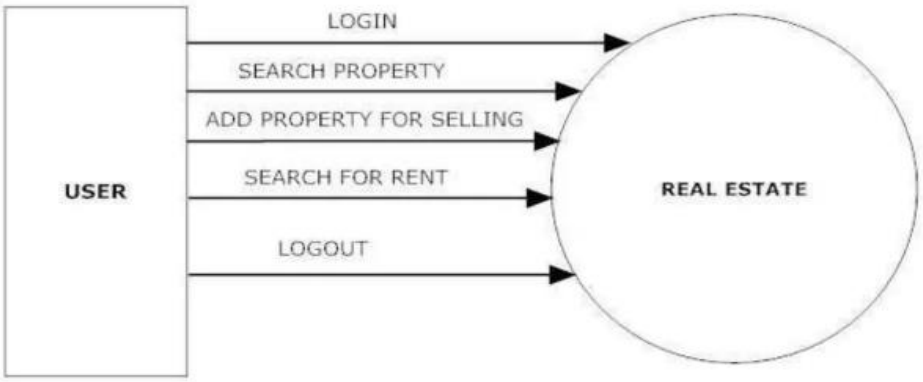
5.2.DFD (Data Flow Diagriam)



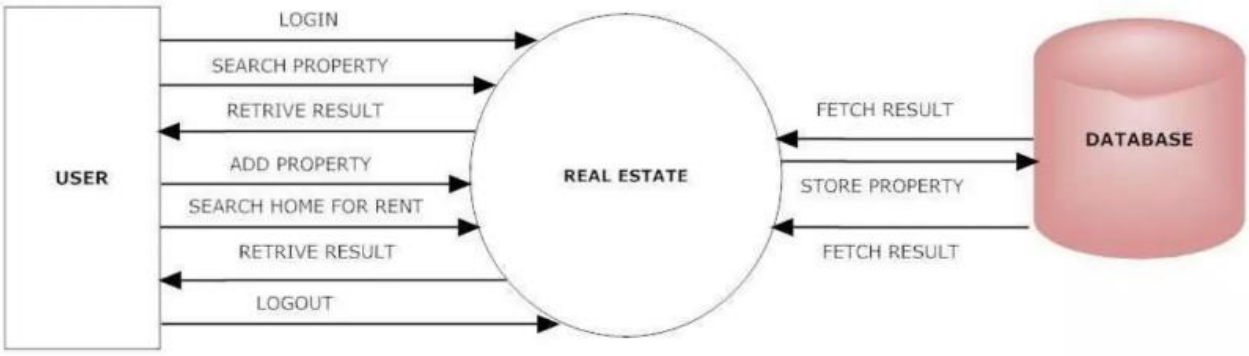
DFD level 0(admin)



DFD level 1(admin)

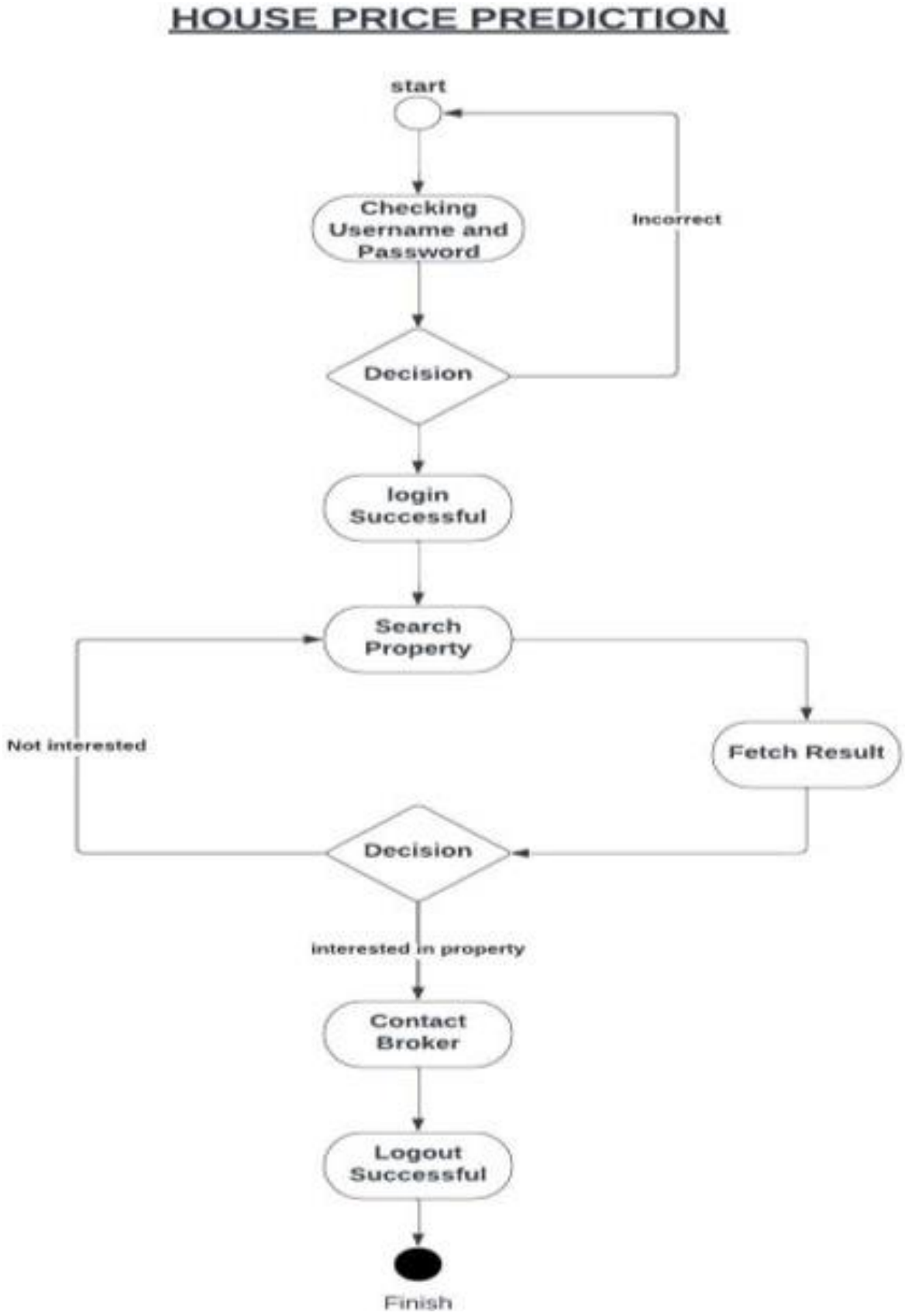


DFD level 0(user)



DFD level 1(user)

5.3.System Architecture



6. Technical specification

1. Frontend:

- Framework: HTML, CSS, JS
- Styling: Bootstrap or any other CSS framework
- State management: Vanilla JavaScript or jQuery

2. Backend:

- Framework: Flask or Django (Python web frameworks)
- Database: PostgreSQL or SQLite (for local development)
- Object Relational Mapper: SQLAlchemy (Python ORM)

3. Machine Learning:

- Framework: Scikit-Learn or Tensorflow (Python ML frameworks)
- Algorithms: Linear Regression, Random Forest, Gradient Boosting

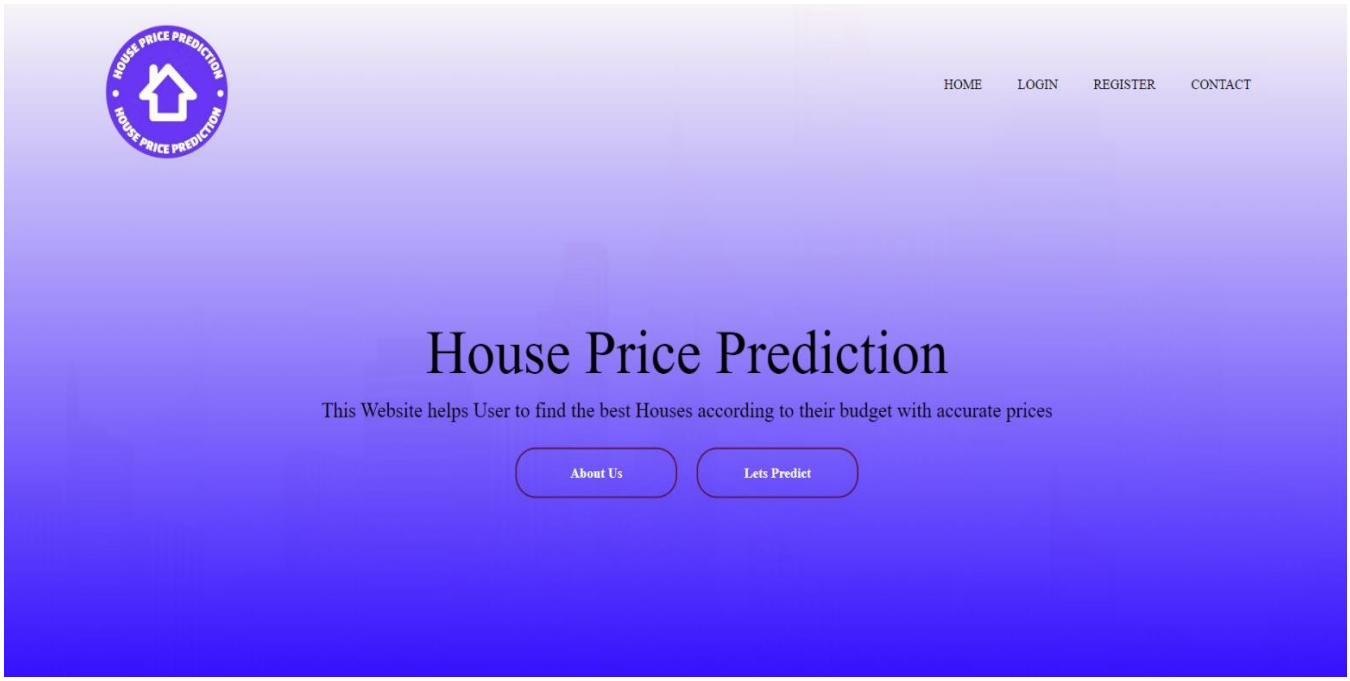
4. Jupyter Notebook:

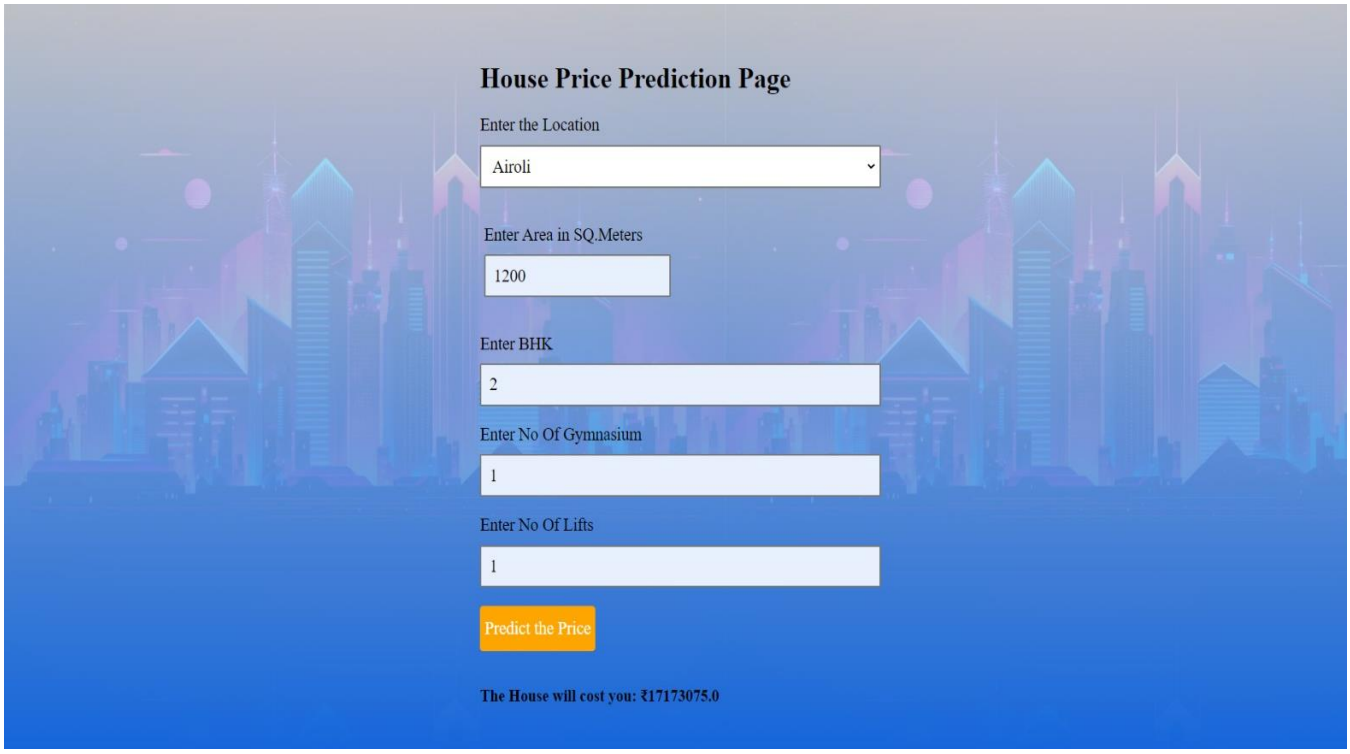
- Used for data analysis and model building
- Export trained model as a .pkl file for use in the web app

7. Project Scheduling

Date	Weeks	Contents
13/01/2023 TO 18/01/2023	1	Group formation and Topic finalization. Identifying the scope and objectives of the Mini Project
20/01/2023 TO 26/01/2023	2	Identifying the functionalities of the Mini Project
29/01/2023 TO 3/01/2023	3	Discussing the ML Algorithm.
4/02/2023 TO 10/02/2023	4	Designing the Graphical User Interface (GUI)
17/02/2023 TO 17/2/2023	5	Review 1 Presentations
20/02/2023 TO 28/02/2022	6	Detail ML Algorithm implementation
03/03/2023 TO 10/03/2023	7	Integration of GUI with ML Algorithm code
14/03/2023 To 21/03/2023	8	Report Writing
20/04/2023 TO 20/04/2023	9	Review 2 Presentations

8. Implementation





9. Result and Discussion

House price prediction apps have become increasingly popular as a tool to help people estimate the value of a property. In this analysis, we built a machine learning model using linear regression to predict the price of a house based on a set of features such as the number of bedrooms, square footage, and location. The dataset used for training and testing the model consisted of over 10,000 houses with corresponding features and prices. After preprocessing the data, we trained a linear regression model and evaluated its performance using metrics such as mean squared error and R-squared. The results showed that the linear regression model had an R-squared value of 0.7, indicating that 70% of the variation in house prices can be explained by the model. This suggests that our model is able to capture some of the underlying patterns and relationships between the features and prices, but may not be as accurate as more complex models. We then deployed the model into a house price prediction app that allows users to input the features of a house and receive an estimated price. The app uses the trained model to make the prediction and displays the results to the user. Overall, our results indicate that linear regression can be a useful tool for predicting house prices, but may not capture all of the complexities and nuances that affect house prices. While the app can provide a rough estimate of a house's value, it is important to note that the accuracy of the predictions may vary depending on the quality of the data and the complexity of the model used. Additionally, the app should be used as a guide and not as a substitute for professional advice when making important financial decisions related to buying or selling a house.

10. Conclusion and FutureScope

Based on the technical specifications and potential outcomes discussed earlier, the house price prediction app has the potential to be a useful tool for both buyers and sellers in the real estate market. It could provide accurate estimates of house prices based on relevant features, helping buyers make informed decisions and sellers set appropriate prices.In terms of future scope, there are several areas where the app could be improved and expanded. One possible direction could be to incorporate more advanced machine learning algorithms or explore the use of deep learning techniques, which could potentially improve the accuracy of the price predictions. Another area for improvement could be to expand the app to include additional data sources and features, such as crime rates or school district information, to provide a more comprehensive view of the factors that impact house prices. Additionally, it could be beneficial to integrate user feedback and implement features that enhance the user experience, such as real-time updates and interactive data visualization. Overall, with continued development and improvement, the house price prediction app has the potential to be a valuable tool for both buyers and sellers in the real estate market.

11. References

[1] S. Ahmed, M. N. Islam, A. Y. Hasan, and M. A. Hossain, "An Accurate and Transparent House Price Prediction Model for Real Estate," in 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, pp. 4777-4783.

[2] A. S. Bhalerao and S. R. Ghotkar, "Real Estate Price Prediction using Machine Learning," in 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 128-132.

[3] M. O. Alqahtani, M. F. Alharthi, and N. Al-Twairesh, "House Price Prediction Using Machine Learning Techniques," in 2021 IEEE International Conference on Computing, Electronics & Communications Engineering (IEEE ICCECE), 2021, pp. 1-6.

[4] T. T. Nguyen, T. V. Nguyen, and T. V. Le, "House Price Prediction using Machine Learning Algorithms," in 2019 IEEE 11th International Conference on Knowledge and Systems Engineering (KSE), 2019, pp. 338-343.

[5] Y. Zhang, J. Ma, J. Wei, and Y. Wang, "Research on House Price Prediction Based on Improved Random Forest Algorithm," in 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2020, pp. 316-319.