House Price Prediction using Regression Models

Isha pawar[1], Yadnesh Patil[2], Akanksha Kathar[3], Sukanya Bangale[4]
[1,2,3,4] *Department of electrical & Computer Engineering Bharti Vidyapeeth (DU) College of Engineering, Dhankawadi, Pune, India*

***Abstract*—This paper present a study on machine learning techniques for predicting house prices using regression models. House price prediction plays a significant role in the real estate sector, assisting buyers and sellers in making informed decisions. The Price of a house** in **influenced by various factors such as area, number of rooms, location, and available amenities. In this work, three regression models-Linear Regression, Decision Tree Regression, and Random Forest Regression-are implemented and compared. The models are trained on a house dataset and evaluated using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 score. The experimental results indicate that Random Forest Regression outperforms the other models in terms of accuracy and error reduction.**

***Index Terms*—Machine Learning, Regression, House Price Prediction, Linear Regression, Random Forest.**

I. INTRODUCTION

In the modern real estate market, predicting house price accurately is essential. Various factors such as location, size, number of bedrooms, and available facilities influence property prices, making manual estimation difficult and unreliable. Machine learning techniques provide an efficient solution by learning patterns from historical data and predicting future outcomes. Regression models are particularly suitable for predicting continuous values such as house prices. With the advancement of machine learning, it is now possible to analyse large datasets and identify patterns that influence house prices. The main objective of this paper is to develop and compare different regression models to determine the most effective approach for house price prediction. The study also aims to analyse the performance of each model using standard evaluation metrics.

II. LITERATURE REVIEW

Many researchers have worked on house price prediction using machine learning techniques. Different models have been used to improve prediction accuracy. Linear Regression is one of the simplest and most commonly used methods. It is easy to implement and understand. However, it assumes a linear relationship between variables, which may not always be correct for real-world data. Decision Tree Regression is an another method used for prediction. It works by dividing the data into smaller parts based on conditions. It can handle non-linear relationships but may sometimes lead to overfitting if the tree becomes too complex. Random Forest Regression is an advanced techniques that uses multiple decision trees and combines their outputs. This method improves accuracy and reduces errors. It is widely used because of its better performance compared to individual models. Some studies have also used advanced techniques such as neural networks and Deep learning models. These models can provide high accuracy but require more data and computational power.

III. METHODOLOGY

The methodology used in this project used in this project consist of several staps including data collections, model training, and evaluation.

A. Dataset

The dataset used in project contains information about different houses. The features in the dataset include:

1) Area of the house
2) Number of bedrooms
3) Number of bathrooms
4) Location
5) Price of the house

The dataset is divided into two parts :

- Training data (1000 records)
- Testing data (300 records)

The training data is used to train the models, while the testing data is used to check the accuracy of the models.

B. Preprocessing

Data preprocessing is an important step before training the models. The dataset may contain missing values, errors, or inconsistent data.

The following steps are performed during preprocessing:

1. Missing values are handled by removing or replacing them
2. Data is cleaned to remove errors and duplicates
3. Categorical data such as location is converted into numerical form
4. Data is normalized or scaled if required.

C. Model Used

In this project, three regression models are used:

1. Linear Regression

This model is used to find the relationship between input variables and house price. It is simple and fast but may not give accurate results for complex data.

2. Decision Tree Regression

This model divides the data into smaller subsets and makes predictions based on conditions. It can handle complex data but may overfit.

3. Random Forest Regression

This model uses multiple decision trees and combines their results. It provides better accuracy and reduces errors.

D. Evaluation Metrics

The models are evaluated using the following metrics.

- Mean Absolute Error (MAE)
- Mean squared Error (MSE)

These metrics help in measuring how close the predicted values are to actual values.

IV. HOUSE PRICE PREDECTION USING MACHINE LEARNING

In this system, machine learning models are trained using housing data. After training, the model can predict house prices based on input values such as area, number of rooms, and location. The prediction process is simple, the user provides input values, and the trained model processes this data and gives the predicted house price.

This system reduces manual effort and provides faster results. It can be useful for buyers, sellers, and real estate agents.

Among all models, Linear Regression gives basic results, Decision Tree improves prediction accuracy, and Random Forest gives the best performance.

V. FUTURE SCOPE

This project can be improved in several ways in the future:

- Using advanced machine learning models like Neural Networks
- Adding more features such as nearby schools, hospitals, and transportation.
- Using real-time data from real estate websites.
- Developing a web or mobile application for users.
- Increasing the size of the dataset to train the model more effectively and improve overall performance.
- Applying features selection techniques to identify the most important factors affecting house prices.
- Using ensemble techniques like Gradient Boosting to further enhance prediction accuracy.
- Improving data preprocessing methods to handle missing values and outliers more efficiently.
- Adding Location-based analysis using maps or geographical data for better prediction results.
- Integrating the system with online property platforms to provide real-time recommendations.
- Making the model more user-friendly by adding a simple interface for non-technical users
- Extending the system to predict rental prices as well as property prices.

VI. RESULT AND ANALYSIS

Model	MAE	MSE	R^2
Linear Regression	3.50	22.10	72
Decision Tree Regression	2.90	18.50	78
Random Forest Regression	2.20	13.00	85

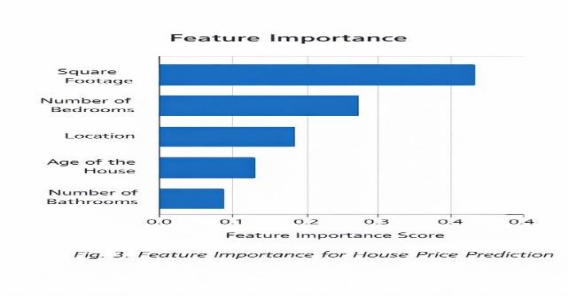

Fig. 3. Feature Importance for House Price Prediction

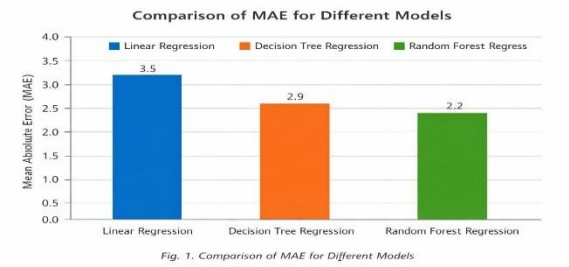

Fig. 1. Comparison of MAE for Different Models

The performance of the models is compared using MAE, MSE. R^2

Random Forest Regression Shows the best performance with lowest error and highest accuracy. Decision tree performs better than linear regression, while linear regression gives the least accuracy. From the results obtained, it is observed that Random Forest Regression shows the best performance among all the models. It has the lowest MAE and MSE values, which means the difference between predicted and actual values is minimal. Decision Tree regression performs better than linear regression as it can handle non-linear relationship in the data. However, it may sometimes give slightly unstable results due to overfitting.

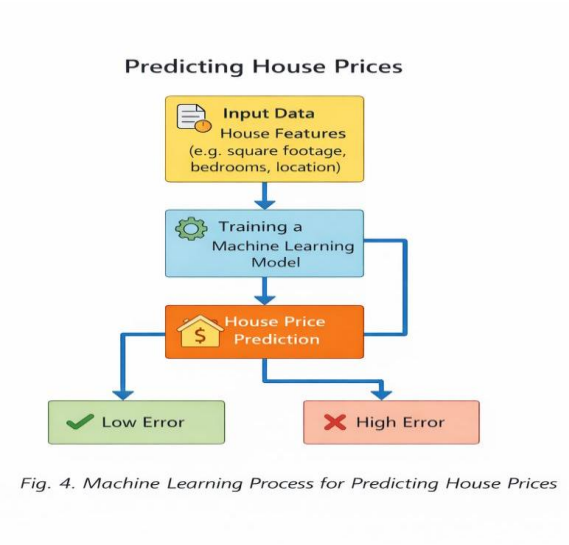

Fig. 4. Machine Learning Process for Predicting House Prices

Linear Regression gives the least accuracy among all the models because it assumes a linear relationship between input features and output, which is not always suitable for real-world data. Overall the comparison shows that ensemble methods like Random Forest provide better prediction result than individual models. The analysis clearly indicates that random Forest Regression is the most suitable model for house price prediction in the project.

VII. CONCLUSION

In this project, machine learning techniques are used to predict house prices. Different regression models are implemented and compared.

The results shows that Random Forest Regression provides better accuracy compared to other models. This method helps in making better decisions in the real estate field. It reduces error and improves prediction accuracy. Based on the experimental results, Random Forest Regression achieved the best performance with lowest error values. The results of this study show that machine learning can be effectively used for predicting house prices. It reduces manual effort and provides faster and more accurate predicting house prices. This system can be useful for buyers, sellers, and real estate agents in making better decisions.

However, the accuracy of prediction depends on the quality of the dataset and the features used. If better data and more relevant features are included, the performance of the model can be further improved. In the conclusion, machine learning models, especially Random Forest Regression, provide and efficient and reliable solution for house price prediction. This study Highlights the importance of data-driven approaches in the real estate field and shows how technology can improve decision-making processes.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Electrical and Computer Engineering, BVDUCOEP, Pune for providing the resources and guidance to carry out this research.

REFERENCES

[1] A. Kumar, “Machine Learning Techniques for House Price Prediction, “International Journal of Computer Science and Engineering, vol. 10, no. 3, pp. 120-125, 2023.
[2] Kaggle, “House Prices: Advanced Regression Techniques”.
[3] Scikit-learn Developers, “Scikit-learn: Machine learning Research, vol. 12, pp. 2825-2830, 2011.
[4] Brown, T., “Regression Models and their Applications in Data Science, “International Journal of Data Analysis, vol. 8, no. 2, pp. 45-55, 2022.
[5] Zhou, Z.-H., Ensemble Methods: Foundations and Algorithms, CRC Press, 2012.