International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Real Estate Price Prediction Using Machine Learning

Mrs. Kavitha K.S, Karthik G, Abhishek, Meghana G, Akshatha R

Department of Computer Science & Engineering,
K S School of Engineering & Management, Bengaluru, India
KavithaKS@kssem.edu.in, karthikkg924@gmail.com, sjvabhishek@gmail.com, mouryamegha21@gmail.com, akshathagowda2322@gmail.com

---

–

Abstract

Accurate real estate price prediction is a critical aspect in property investment, buying, and selling decisions. Machine learning (ML) offers a promising approach to modeling the complex and non-linear relationships between various factors influencing property prices. This study presents a Random Forest-based model trained on a dataset containing historical real estate transactions, location features, property attributes, and market trends. Our model achieved a high prediction accuracy and demonstrated robustness in diverse property conditions. We discuss data preprocessing, feature selection, model training, and evaluation techniques. The results indicate that ML algorithms can significantly enhance decision-making in real estate transactions.

Index Terms - Real Estate, Price Prediction, Machine Learning, Random Forest, Data Preprocessing, Feature Engineering, Model Evaluation, Property Attributes, Regression, Location-Based Pricing.

I. INTRODUCTION

The real estate sector is a vital component of the economy, directly impacting individuals and businesses alike. Property prices are influenced by a wide array of factors such as location, size, amenities, market demand, and economic indicators. Traditional methods of property valuation often fall short in capturing these complex, interrelated variables. In recent years, machine learning (ML) techniques have gained attention for their ability to analyze large datasets and uncover patterns that are difficult to detect using conventional methods. This paper explores the use of machine learning, particularly the Random Forest algorithm, to predict real estate prices with higher accuracy. By leveraging data collected from property listing sites and real estate transaction records, we aim to provide a robust model that assists stakeholders in making informed investment and pricing decisions.

II. LITERATURE SURVEY

Several studies have explored the application of machine learning in real estate valuation. One approach used Support Vector Machines (SVM) to model non-linear relationships in housing data, achieving moderate accuracy but requiring extensive feature tuning [1]. Another study applied Artificial Neural Networks (ANN) to capture hidden interactions among property features, showing improvements over traditional regression models [2]. Gradient Boosted Decision Trees (GBDT) were also used in various research efforts due to their robustness and interpretability, with models effectively identifying

the most influential factors such as location, proximity to amenities, and market trends [3]. Incorporating geospatial data has further improved prediction accuracy by enabling more localized analysis [4]. Despite these advancements, issues like overfitting and limited dataset diversity remain, highlighting the need for generalizable and scalable solutions. This study builds upon previous work by integrating a well-balanced dataset, extensive preprocessing, and the Random Forest model to improve predictive performance.

III. METHODOLOGY

The dataset used in this research was sourced from online real estate platforms and government property transaction records. It contains information on property type, size (in square feet), number of bedrooms and bathrooms, location (including latitude and longitude), year built, amenities, and the sale price. The data underwent preprocessing to handle missing values, remove outliers, and normalize numeric attributes. Categorical data such as location and property type were encoded using one-hot encoding. Feature selection was carried out using correlation matrices and domain knowledge to identify the most relevant inputs for the model.

A Random Forest Regressor was chosen for its ability to handle large datasets and manage non-linear relationships. The model was trained on 80% of the dataset and tested on the remaining 20%. Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared were used to assess model performance. Cross-validation was applied to ensure generalization, and hyperparameters such as the number of trees and max depth were tuned for optimal results.

IV. RESULTS AND DISCUSSION

The Random Forest Regressor showed strong performance in predicting property prices. The model achieved an R-squared value of 0.89, indicating that it could explain 89% of the variance in the target variable. The Mean Absolute Error (MAE) was 2.3 lakhs, and the Root Mean Squared Error (RMSE) was 3.1 lakhs, reflecting a low average prediction error. Feature importance analysis revealed that location, property size, and number of bedrooms were the top predictors. The model's ability to generalize was validated through k-fold cross-validation, where consistent accuracy was observed across folds.

A simple web interface was also developed using Flask, allowing users to input property details and receive price predictions in real-time. This prototype tool can serve as a decision-support system for buyers, sellers, and real estate agents. Despite the promising results, the model is subject to limitations such as regional bias in the dataset and the exclusion of economic factors like interest rates or inflation. Expanding the dataset and integrating more dynamic features could further enhance model reliability.

V. CONCLUSION

This research demonstrates the potential of machine learning, specifically the Random Forest algorithm, in accurately predicting real estate prices. By incorporating a range of property attributes and applying rigorous preprocessing and model validation techniques, the developed system provides reliable price estimations. The integration of this model into a user-friendly application further highlights its practicality in real-world scenarios. Future work may focus on increasing dataset diversity, incorporating time-based trends, and exploring ensemble or deep learning methods to push prediction accuracy even further.

VI. REFERENCES

[1] Kumar, A., & Singh, B. (2020). House Price Prediction using Support Vector Regression. International Journal of Computer Applications.
[2] Patel, R., & Shah, M. (2021). A Neural Network Approach to Real Estate Valuation. Journal of AI Research.
[3] Li, Y., & Zhao, J. (2019). Gradient Boosting Framework for Real Estate Price Forecasting. Expert Systems with Applications.
[4] Zhang, X., & Wang, L. (2022). The Impact of Geospatial Features on Housing Prices. Applied Geographic Information Science Journal.