House Price Prediction

Mrs. Nargis A. Mulla, Mr. Harshit Nandkumar Gorad, Mr. Sarthak Balaso Patole,
Miss. Pradnya Adhikrav Gorad, Miss. Shweta Ramesh Jadhav
Lecturer, Adarsh Institute of Technology and Research Centre, Vita, India
Students, Adarsh Institute of Technology and Research Centre, Vita, India

Abstract*: House price prediction is a crucial aspect of real estate analytics, impacting buyers, sellers, investors, and policymakers. The real estate industry is a dynamic sector where property values fluctuate based on numerous factors such as economic conditions, location, infrastructure, demand-supply balance, and regulatory policies. Traditional valuation methods often rely on expert opinions and manual assessments, which may introduce biases and inconsistencies. In contrast, data-driven approaches, particularly machine learning (ML) techniques, offer a more accurate and systematic way to estimate property values.*

This study explores the effectiveness of machine learning models in predicting house prices by analyzing historical real estate data. Various predictive models, including linear regression, decision trees, random forests, gradient boosting machines (XGBoost), and deep learning algorithms, are evaluated in terms of their performance. The research aims to determine which model provides the highest accuracy while ensuring interpretability and efficiency.

Data pre-processing plays a vital role in enhancing the accuracy of house price predictions. The dataset used in this study includes essential property attributes such as square footage, the number of bedrooms and bathrooms, locality, age of the property, and market trends. Before applying predictive models, pre-processing techniques such as handling missing values, encoding categorical variables, scaling numerical features, and feature selection are implemented to optimize the dataset. The study also examines the impact of feature engineering, where derived attributes such as neighbourhood crime rates, proximity to schools, and public transportation accessibility are incorporated to improve prediction accuracy.

The research adopts multiple evaluation metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R²) values, to assess the performance of different models. Initial findings suggest that ensemble learning methods such as Random Forest and XGBoost outperform traditional regression models in terms of predictive accuracy. However, deep learning models, specifically artificial neural networks (ANNs), show promising results when trained on large datasets. The study also highlights the importance of hyper parameter tuning in refining model performance, addressing issues such as over fitting and computational efficiency.

One of the key contributions of this research is the comparative analysis of different machine learning algorithms to identify the best-suited approach for house price prediction. Traditional regression techniques, while interpretable, often fail to capture complex nonlinear relationships within the data. Decision tree-based methods, such as Random Forest and XGBoost, demonstrate robust performance by learning intricate patterns and reducing variance through ensemble learning. Deep learning models, on the other hand, excel in capturing high-dimensional interactions but require extensive computational resources and larger training datasets.

Beyond predictive accuracy, the study examines the interpretability of machine learning models, which is crucial for real-world applications. Feature importance analysis is conducted to understand which variables have the most significant impact on house prices. Factors such as location, property size, and market demand emerge as primary determinants. The study also investigates the effect of external factors, including economic indicators like interest rates, inflation, and employment rates, on housing prices.

Despite the promising results, challenges such as data availability, regional market variations, and the need for real-time data integration remain. Future research can focus on enhancing prediction models by incorporating geospatial analytics, satellite imagery, and deep learning advancements. Additionally,

integrating real-time market data and leveraging blockchain technology for transparent property transactions can further improve predictive modelling in real estate.