MANIPAL
ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

Department: M.Tech – Data Science
Subject: Data Analytics and Visualization

Housing Price Analysis and Prediction

Abstract

This project “Housing Price Analysis and Prediction” aims to analyze various factors influencing house prices and develop a predictive model to estimate property values accurately. Using historical housing data, the study explores relationships between features such as location, area, number of rooms, and amenities with the final selling price. Data preprocessing techniques, exploratory data analysis (EDA), and feature engineering are applied to uncover patterns and correlations. Machine learning algorithms such as Linear Regression, Decision Tree, and Random Forest are implemented and compared to identify the most accurate prediction model. The results demonstrate that incorporating multiple influential features enhances prediction accuracy, assisting buyers, sellers, and real estate investors in making informed decisions.

Introduction

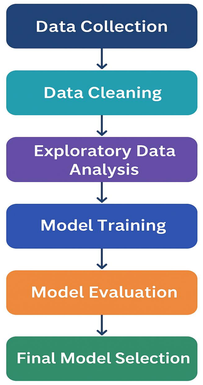
The real estate industry is a rapidly growing sector and plays a crucial role in economic development. Housing prices vary significantly based on several factors such as location, total area, number of bedrooms, availability of amenities, and proximity to key facilities like schools, hospitals, and transportation. Predicting property prices accurately helps potential buyers, sellers, and real estate investors make better financial decisions and reduces uncertainty in property valuation.

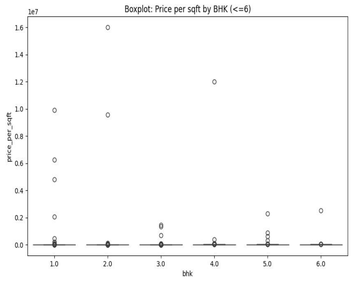
The project “Housing Price Analysis and Prediction” focuses on identifying the major factors that influence property prices and developing a predictive model using machine learning techniques. The study uses the Kaggle Bengaluru House Price Prediction Dataset, which contains detailed information about houses across various localities in Bengaluru. The dataset includes attributes such as location, size, number of bathrooms, square footage, and price, making it ideal for data-driven analysis.

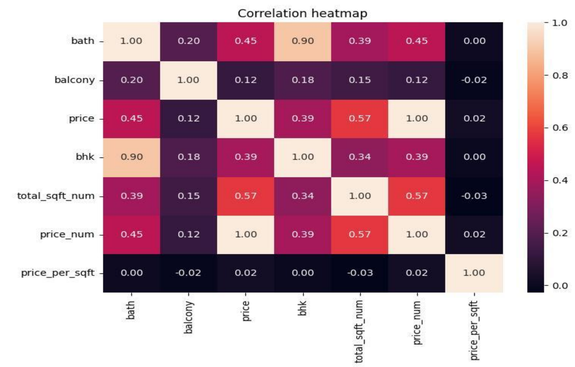
The project involves data preprocessing, feature engineering, and EDA, followed by model training using algorithms like Linear Regression, Decision Tree, and Random Forest. The results uncover patterns and provide accurate price predictions, helping understand housing market dynamics.

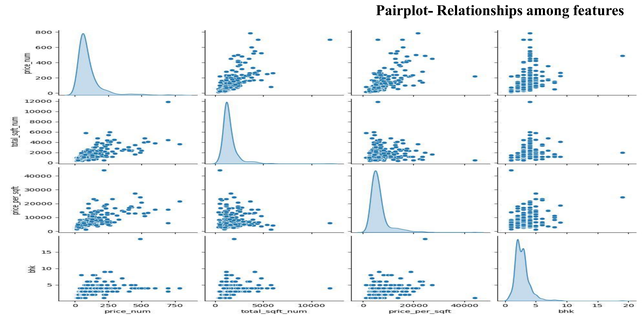
Methodology

1. Data Collection: Dataset downloaded from Kaggle (Bengaluru House Price Prediction)
2. Data Cleaning: Removed duplicates, handled missing values, and label-encoded categorical features.
3. Exploratory Data Analysis: Used heatmaps, boxplots, and scatter plots to study feature relationships and trends.
4. Feature Engineering: Created additional features such as price per square foot to enhance model accuracy.
5. Model Training: Applied algorithms including Linear Regression, Decision Tree, Random Forest, SVM, and KNN.
6. Final Model Selection: Random Forest chosen based on highest performance.



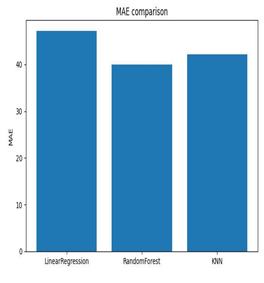


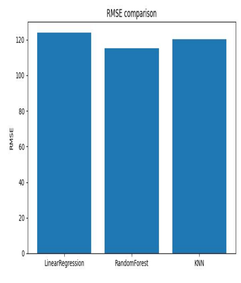


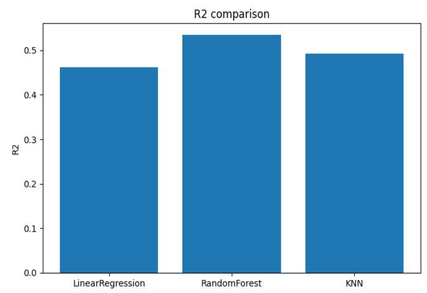


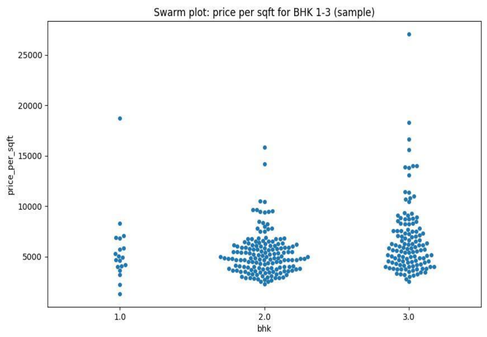
Results

- The trained models were evaluated using performance metrics such as R^2 score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) to measure accuracy and reliability. Among all the algorithms tested — Linear Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) — the Random Forest model delivered the highest prediction accuracy and the lowest error rate.









The results revealed strong correlations between features such as location, total square footage, and number of bedrooms with the house price. The final model successfully predicted housing prices close to actual market values, demonstrating its effectiveness in real estate price estimation and decision-making support.

Conclusion

The project “Housing Price Analysis and Prediction” effectively analyzed the various factors influencing property prices and developed a reliable predictive model using machine learning techniques. Using the Kaggle Bengaluru House Price Prediction Dataset, extensive data preprocessing, feature engineering, and exploratory data analysis were performed to identify key patterns and correlations between housing features and prices.

Several machine learning algorithms, including Linear Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), were implemented and evaluated using performance metrics such as R^2 score, MAE, and RMSE. Among these, the Random Forest model provided the most accurate and consistent predictions.

The study highlights that features like location, area, and number of bedrooms play a significant role in determining housing prices. The developed model can assist real estate buyers, sellers, and investors in making data-driven decisions, optimizing investments, and understanding market trends. Overall, this project demonstrates the potential of machine learning in transforming real estate price prediction and analysis.

Acknowledgements

We sincerely express our gratitude to Dr. Sindhu Madhuri G, Faculty, Department of M.Tech – Data Science, Manipal Institute of Technology, Bengaluru, for her constant guidance, motivation, and support throughout the Data Analytics and Visualization Project Lab.

We also extend our appreciation to the Department of M.Tech – Data Science for providing the necessary resources and a conducive academic environment to complete this project successfully.

Team Members

Sujeet Yadav – 251580840032
Jeevan J C– 251580840037