

Final Report

Problem Statement: Accurately predicting house prices is crucial in the real estate market, but it remains challenging due to various factors. The goal of this project is to develop a machine learning model that can reliably predict house prices based on relevant features, ultimately providing valuable insights for stakeholders in the real estate market.

Approach:

1. **Data Wrangling:** The dataset was thoroughly cleaned, missing values were addressed, and significant outliers were mitigated to prepare the data for analysis.
2. **Exploratory Data Analysis (EDA):** A comprehensive examination of the dataset was conducted to understand its characteristics and prepare it for modeling endeavors.
3. **Feature Engineering:** New variables were created based on existing ones to enrich the dataset, and steps such as rare analysis, label encoding, and scale standardization were performed to enhance the dataset's effectiveness.
4. **Modeling:** Various machine learning algorithms including Linear Regression, Decision Tree, Random Forest, Gradient Boosting, KNN, SVM, and XGBoost were experimented with. Based on evaluation metrics, XGBoost was identified as the best-performing model.

Findings:

- Linear Regression performed poorly, while Decision Tree, Random Forest, Gradient Boosting, KNN, SVM, and XGBoost showed varying performance degrees.
- XGBoost emerged as the top performer with excellent performance metrics, making it the recommended model for predicting house prices in this scenario.

Recommendations:

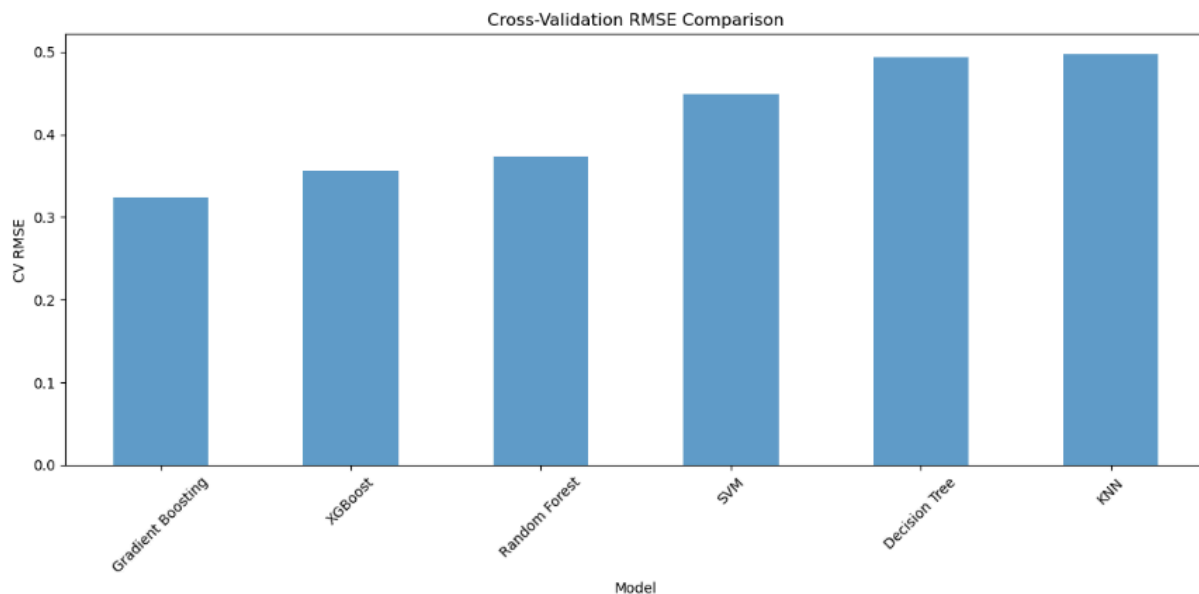
1. **Utilize XGBoost Model:** Implement the XGBoost model for predicting house prices in real-time scenarios, as it demonstrated superior performance compared to other models.
2. **Continuous Model Monitoring:** Regularly monitor the model's performance and retrain it with updated data to ensure its accuracy and reliability over time.
3. **Further Research:** Explore advanced techniques for feature engineering and model optimization to potentially enhance the model's performance even further.

Model Metrics:

- Model: XGBoost

- Features: Various features related to residential properties
- Parameters: Optimized hyperparameters
- Performance Metrics: Low errors, high R2 score, and excellent performance in predicting house prices.

Conclusion: This project successfully addressed the challenge of predicting house prices accurately by developing and evaluating machine learning models. The XGBoost model emerged as the most effective solution, providing valuable insights for stakeholders in the real estate market. Further refinement and continuous model monitoring can ensure its continued relevance and usefulness in real-world applications.



Notebooks and Scripts:

1. [Problem Statement](#)
2. [Data Wrangling](#)
3. [Exploratory Data Analysis \(EDA\)](#)
4. [Feature Engineering](#)
5. [Modeling](#)

[Model Metrics File](#)