Executive Report: Predicting Property Prices for Wazobia Real Estate Limited

Introduction

Wazobia Real Estate Limited, a prominent Nigerian real estate company, is keen on offering its clients the most competitive pricing for properties while maintaining a substantial profit margin. However, estimating house prices accurately within the dynamic Nigerian real estate market has proven challenging. The goal was to understand the key factors that influence property prices, identify trends, and develop a reliable model to predict future prices. The data consists of properties from various locations, with features such as type of property, number of bedrooms, bathrooms, parking spaces, and geographical zones.

Problem Statement

The primary challenge is to devise a reliable predictive model capable of estimating house prices in Nigeria's dynamic real estate market. A myriad of factors influences house prices, ranging from the location and type of property to the number of bedrooms or bathrooms, and even the availability of parking space. Determining the exact impact and relationship of these factors with property price is complex but crucial to making informed pricing decisions.

Exploratory Data Analysis

Our exploratory data analysis revealed several key findings:

1. **Dataset Overview:** The provided dataset contains 14,000 observations across seven distinct variables, with 5 numeric and 2 categorical features. While the dataset is free of duplicate entries, around 9.1% of the total cells have missing values that could potentially impact our analysis and model performance. These missing entries are spread across multiple columns including 'location', 'title', 'bedroom', 'bathroom', and 'parking_space'.

2. **Property Price Variation by Location**: As expected, location plays a critical role in property prices. Lagos, for instance, has the highest average property price at N4,210,546, followed by Bayelsa and Rivers at N3,112,322 and N2,957,098 respectively. It's vital for property investors to consider the average prices in the various locations before making investment decisions.

3. **Number of Properties Sold**: The data reveals that Kaduna has the highest number of property sales, with 370 properties sold. Anambra and Benue follow closely with 363 and 355 sales respectively. This information is particularly useful to property developers and investors interested in understanding market demand.

4. **Property Characteristics**: Our data shows a positive correlation between the number of bedrooms and the property price. Generally, properties with more bedroom’s command higher prices. This finding is pivotal for buyers and sellers in determining fair pricing.

5. **Correlation Among Features**: Our analysis revealed a high correlation (approximately 0.5) between the number of bedrooms and the property price. This is indicative of the significance of the property size in determining its price.

6. In terms of geographic distribution, properties in the South-West region (Ekiti, Lagos, Osun, Ondo, Ogun, Oyo) were found to be generally more expensive than those in other regions.

Model Development and Evaluation

In the model development process, we started by comparing a total of 18 different machine learning models, each with their unique strengths and weaknesses. These models were trained and evaluated based on their Root Mean Squared Error (RMSE), a measure indicating the average magnitude of errors made by the model in its predictions.

Among all the models, the Gradient Boosting Regressor (GBR) showed the best performance with the lowest RMSE of 539,761.0153. Recognizing its potential, we decided to focus on tuning this model for optimal performance.

Model tuning, a process of adjusting the model's parameters to improve its predictive accuracy, led to a noticeable reduction in RMSE to 533,609.1586, enhancing the model's predictive capability.

The table below outlines the progression of our models and the corresponding RMSE values:

Model Type	RMSE
Gradient Boosting Regressor (Untuned)	539,761.0153
Gradient Boosting Regressor (Tuned)	533,609.1586
Bagging Regressor	529,645.8759
Voting Regressor (Blend of GBR and Bagging)	529,587.1495

To further fortify our predictive power, we employed ensemble techniques such as Bagging and Blending. Bagging is a method that improves stability and accuracy by creating multiple subsets of the original dataset and training the model on these subsets. Blending, on the other hand, combines predictions from multiple models to give a final prediction.

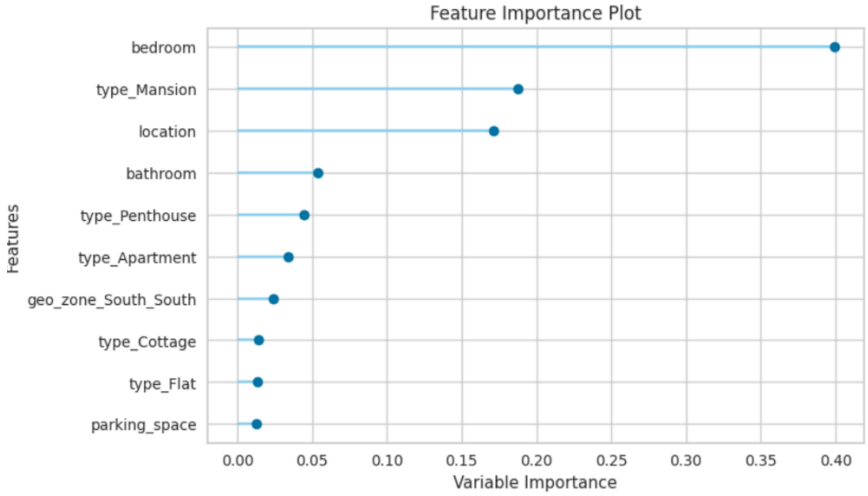
Our Bagging Regressor performed better than the single GBR model with an RMSE of 529,645.8759 on the training data. Finally, the blended model (a combination of GBR and Bagging Regressor) yielded the best performance, achieving the lowest RMSE of 529,587.1495 on the training data. When evaluated on unseen data, the blended model maintained a comparable performance with an RMSE of 553,105.

The progression in our modeling approach, from using a single model to leveraging ensemble methods, illustrates the power of combining different machine learning models to achieve more robust and accurate predictions.

The Model and its Biases

It's important to note that the model's performance was influenced by certain biases inherent in the model's design. For instance, the Gradient Boosting model places more emphasis on variables that have a higher correlation with the target variable ('price'). In our case, this resulted in the model giving

more importance to the 'bedroom' variable. Similarly, variables with weaker relationships with the target variable, like 'parking_space', had less influence on the model's predictions.



Through model interpretation techniques, we found that the number of bedrooms was the most influential factor affecting property prices, followed by property type, location, and the number of bathrooms.

Model Deployment

To make the model accessible and interactive for non-technical users, a web application was created using Streamlit. The Streamlit application allows users to input specific house features and receive an immediate estimated price, leveraging the power of the predictive model in a user-friendly way. This interactive tool can assist Wazobia Real Estate Limited in their day-to-day decision-making, enabling them to make informed pricing decisions quickly and effectively.

Conclusion

Through a thorough exploration and understanding of the given data, we developed a robust predictive model that reliably estimates house prices. Our machine learning solution, a blend of Gradient Boosting and Bagging Regressor models, demonstrated a strong performance with an RMSE of around 529,587 on the training data and 553,105 on the holdout set. However, it's crucial to remember that these models, while robust and well-performing, are not perfect. They are as good as the data they have been trained on. The models might perform differently when exposed to different types or sources of data.

Finally, we recommend regular updates and tuning of the model to ensure its accuracy remains high as new data becomes available. Regular monitoring will help quickly identify any changes in the factors affecting property prices and adjust the model accordingly.

By leveraging data and machine learning, Wazobia Real Estate Limited can revolutionize its pricing strategy and gain a competitive edge in Nigeria's real estate market. This data-driven approach not only enhances business operations but also offers customers fair and accurate property prices, improving their overall experience.