Executive Summary

Project Title: House Price Prediction: A Data-Driven Approach

Problem Statement

Real estate stakeholders and consumers, particularly in Nigeria are usually faced with the challenge of fixing or guesstimating prices of real estate properties respectively. This process is many times fraught with lots of disagreement haggling prices until a final agreement is made. Sometimes, the real estate seller could make a mistake in quoting the wrong estimate. This could lead to sharp disagreement, mistrust, legal suits, and termination of business. It is therefore necessary to find a solution.

Proposed Solution

The House Price Prediction project is designed to provide a reliable, efficient, and accurate house price forecasting model using machine learning algorithms. The project seeks to empower homeowners, potential buyers, real estate companies, and investors to make well-informed decisions in property investment and valuations.

My approach to creating this model was methodical and data-driven, starting with the use of the Wazobia Real Estate company dataset of housing transactions provided. I used the train and test dataset for both training and validation purposes respectively, making use of the input features such as location, property title, number of bedrooms, bathrooms, and parking spaces, to predict the target variable- price.

The data preprocessing stage involved thorough cleaning and handling of missing values. I also conducted exploratory data analysis, which involved a close examination of each feature, its distribution, and its correlation with the house prices.

To manage skewness and scale the data into a standard range, I employed standard scaling techniques, which were useful for optimizing the performance of the machine learning algorithms.

I considered a variety of machine learning models, including Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, Decision Tree Regressor, and Support Vector Regressor, among others. Each model was meticulously trained and validated, and their performance metrics were evaluated. I found that the XGBoostRegressor algorithm outperformed others in terms of root mean square error (RMSE) and the coefficient of determination (R^2 score).

To optimize my models, I used feature selection methods, which identified 'bedroom', 'title', and 'location' as the most significant features contributing to the price prediction.

I also used ensemble techniques such as VotingRegressor and StackingRegressor to create a blend of models, aiming to enhance the prediction performance. The StackingRegressor Ensemble model showed promising results, and its performance was further validated using cross-validation methods.

Final Thoughts and Future Actions

Finally, I utilized my optimal model to generate predictions on the test data. These predictions formed the basis for my final model, which achieved a high level of accuracy and reliability(88.2%). The final model was pickled for future use, to make deployment and scalability easy.

In conclusion, the House Price Prediction project successfully demonstrates the potential of machine learning in revolutionizing the real estate industry. My model can serve as a valuable tool for a myriad of stakeholders, enhancing their decision-making capabilities in property transactions. Moving forward, I plan to refine and update my model as new data becomes available, ensuring that my system continues to deliver accurate and timely house price predictions.