House Price Prediction

[Kedirdin Keyser, Saciid Ahmed Noor, Tanvir Happy, Deni Makhmadiyev, Seyit Siyar Yuce], [05/11/23]

DESCRIBE THE PROBLEM

This report outlines the strategic plan for a project aimed at developing a machine learning model to analyze and predict real estate prices. By leveraging data insights, this initiative seeks to enhance decision-making processes for businesses and individuals in the real estate market.

SCOPE

Business Objective
The primary goal is to create a predictive model that accurately forecasts real estate prices. This tool will assist in identifying key value drivers in the property market, benefiting real estate companies, investors, and potential buyers or sellers.

Solution Usage and Comparison to Existing Solutions
The proposed solution will be utilized for advanced, data-driven real estate price prediction. Current market practices often rely on conventional statistical methods or basic machine learning algorithms. In contrast, our approach will employ more sophisticated techniques for higher accuracy. Without machine learning, such methods involve manually analyzing historical price data, which can be time-consuming and less precise.

Performance Measurement via Business Metrics
The model's success will be measured through its predictive accuracy, using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or Mean Absolute Error (MAE). Additionally, business-oriented metrics, such as decision-making efficiency, customer satisfaction, and investment returns, will also be considered.

System Integration and Impact
If integrated into a larger system, the model will work in conjunction with real estate databases, market trend analysis tools, and customer management systems. It is crucial to monitor how variations in market dynamics or data inputs might affect the model, ensuring its continual relevance and accuracy.

Stakeholder Analysis
The project involves diverse stakeholders, including:
- Real Estate Firms: Seeking enhanced pricing strategies.
- Investors: Interested in profitable opportunities.
- Technical Teams (Data Scientists, Developers): Focused on the model's accuracy and reliability.
- End-users (Buyers/Sellers): Desiring fair and transparent pricing.

In conclusion, this project represents a significant advancement in applying machine learning to real estate price prediction, offering substantial benefits to various stakeholders in the property market. Regular updates and adaptive strategies will ensure the model's relevance and accuracy in a dynamically changing market environment.

METRICS

We would consider a R^2 value of 0.85 success.

To evaluate our system, we're using Mean Square Root Error (MSRE) and R-squared (R^2) metrics. MSRE helps us understand the average error in our predictions, which is important for a regression task involving large numbers. R^2 shows how well our model explains the variability of real estate prices. These metrics are better suited than accuracy for this kind of task, as they provide a clearer picture of our model's predictive accuracy and overall effectiveness in forecasting real estate prices.

DATA

In our project, we're working with a structured dataset that provides a comprehensive overview of residential properties in Ames, Iowa. This dataset is characterized by its depth rather than its size, featuring 79 variables across 1460 instances and 81 columns. The rich variety of features compensates for the dataset's relatively modest size, offering detailed insights into various aspects of residential homes.

For model validation, we're using a test dataset from Kaggle, which provides essential "ground truth" labels. Since Kaggle allows public dataset uploads, the quality and integrity of data can vary. To mitigate this, we rely on Kaggle's usability score, which assesses both the dataset's utility and credibility.

Given that our project involves a regression model and the dataset includes a mix of numerical and non-numerical data, it's essential to undertake thorough data cleaning and preparation. We must be careful with certain values that, due to their numerical magnitude, might disproportionately influence the model. To prevent such skewed perceptions, we'll need to normalize these values, ensuring that they don't inadvertently bias the model's predictions. This step is crucial for maintaining the accuracy and reliability of our regression analysis.

MODELING

In our project, we're exploring a variety of machine learning models for regression tasks, including GradientBoostingRegressor, RandomForestRegressor, Lasso, LGBMRegressor, and LinearRegressor. We'll begin with simpler models like LinearRegressor to establish a baseline performance. This baseline will help us assess the improvements offered by more complex models.

We'll also analyze prediction errors and feature importance to refine our models. Understanding where models fail and which features are most influential will guide our efforts to improve accuracy. By iterating between different models and fine-tuning based on these insights, we aim to enhance the predictive capability of our system.

DEPLOYMENT

For our model deployed through Gradio, we could implement routine checks to ensure its consistent performance and accuracy. This monitoring process might include systematic testing, and there's potential to automate this by establishing a mechanism to continuously receive the latest real estate data from Iowa. By adopting such an approach, we could ensure the model remains up-to-date and reliable, effectively catering to the diverse needs of users ranging from real estate professionals to individual homeowners. This proactive strategy in maintenance and data integration could be pivotal in maintaining the model's relevance and effectiveness amidst the ever-evolving real estate market landscape.

REFERENCES

-Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : concepts, tools, and techniques to build intelligent systems
https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/

-ChatGPT
openai.com