Bulldozer Price Prediction Project

This project focuses on predicting the sale prices of used bulldozers using machine learning techniques. The goal is to help buyers, sellers, and auction platforms make informed pricing decisions by leveraging historical auction data, machine specifications, usage information, and sale dates.

A Random Forest regression model was developed to handle the structured and mixed-type dataset. The primary evaluation metric used was **Root Mean Squared Logarithmic Error (RMSLE)**, chosen because bulldozer prices vary widely and exhibit skewed distributions. Additional evaluation metrics included RMSE, MAE, and R^2 to provide a comprehensive assessment of model performance.

Data preprocessing involved handling missing values by imputing numerical features with median values and filling categorical features with a placeholder value. Significant feature engineering was performed by decomposing the sale date into multiple time-based features, enabling the model to capture seasonal and temporal pricing patterns. Categorical variables were encoded in a format compatible with tree-based models.

A baseline Random Forest model was first trained and evaluated. Hyperparameter tuning using RandomizedSearchCV was later applied to improve performance; however, evaluation results showed that the baseline model slightly outperformed the tuned model in terms of RMSLE. This highlighted the strength of baseline benchmarking and the importance of avoiding unnecessary model complexity.

Feature importance analysis revealed that **YearMade**, **MachineHoursCurrentMeter**, and **ProductSize/ProductGroupDesc** were the most influential factors in determining bulldozer prices. These results align with real-world auction behavior and confirm the interpretability of the model.

The final model was deployed as a FastAPI web service on the Render platform, enabling real-time price predictions via a REST API. Overall, the project demonstrates an end-to-end machine learning workflow, from data preprocessing and modeling to evaluation, interpretation, and deployment in a production environment