NYC Taxi Fare Prediction— Final Model Report

Automatidata • NYC Taxi Analytics Project

Overview

❖ This stage focuses on building the **final predictive model** using insights from prior EDA and baseline analysis.

❖ Linear models were insufficient due to **non-linear fare behavior**, especially at higher distances and durations.

Objective

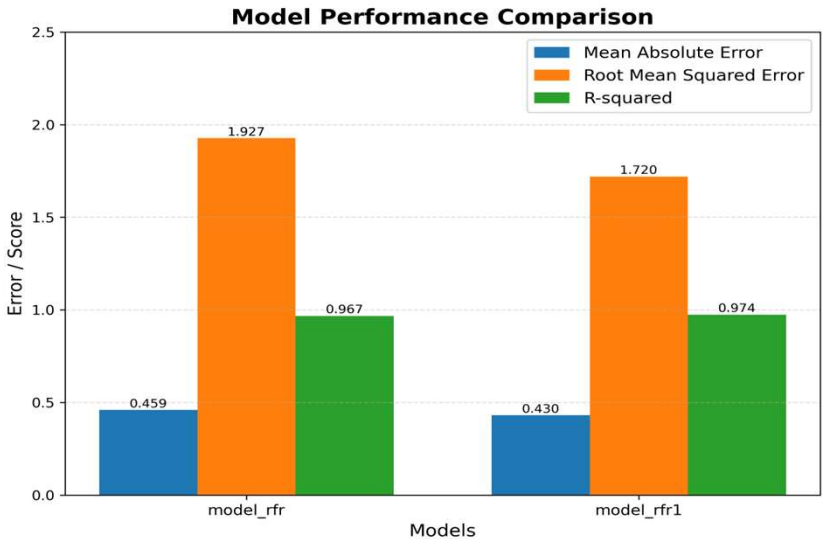
❖ A **Random Forest Regressor** was developed to capture complex, non-linear pricing patterns. Multiple configurations were tested by tuning tree depth and estimators while maintaining a consistent train–test split.

Results

❖ Random Forest consistently outperformed the linear baseline.

❖ Including trip distance, duration, tolls, and passenger count improved stability.

❖ Hyperparameter tuning resulted in lower MAE and RMSE with strong generalization.



Next Steps

❖ Validate model behavior through residual and diagnostic checks.

❖ Deploy the finalized model for reliable real-world fare estimation.