Waze User Churn Project

Prepared for: Waze Leadership Team

ISSUE / PROBLEM

The Waze Data Team is actively addressing a critical issue within their data analytics project, namely, the challenge of high monthly user churn within the Waze app. Churn, characterized by users uninstalling the app or discontinuing its use, presents a pressing problem. The central objective of this project is to engineer a powerful machine learning (ML) model capable of effectively predicting and mitigating this pressing issue of user churn.

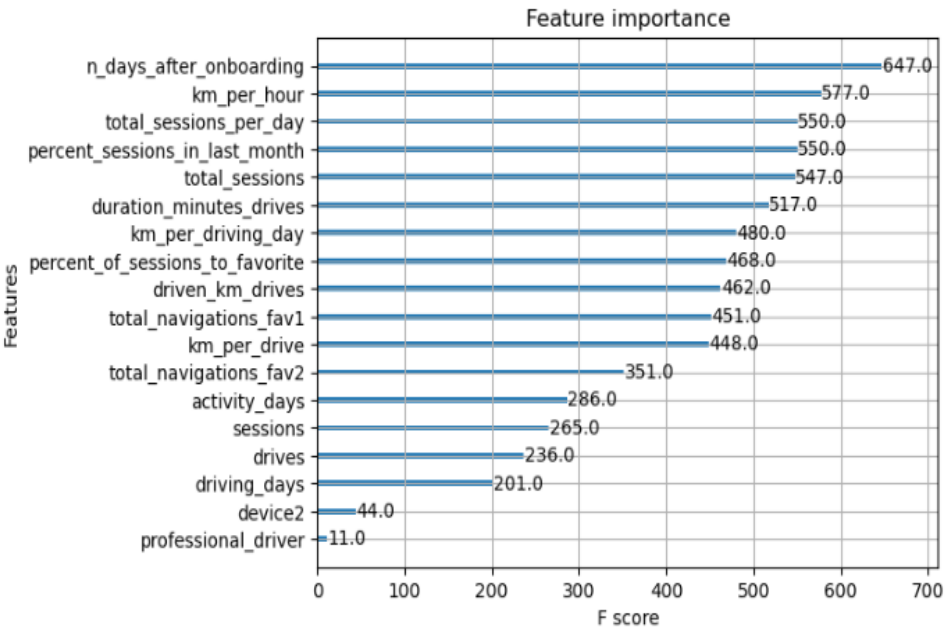
IMPACT

•Our ML models highlight the need for additional data to improve churn prediction accuracy.
•Current data limitations hinder consistent churn prediction, indicating the need for more comprehensive insights.
•We suggest including drive-level data, granular user interaction insights, and monthly unique location counts to address these limitations.
•Engineered features have proven valuable for model enhancement.
•Consider a second iteration of the User Churn Project to incorporate these improvements and achieve more accurate churn predictions.

RESPONSE

•We undertook the development and comparison of two models, random forest and XGBoost, with a primary aim of maximizing predictive accuracy.
•Our dataset underwent a meticulous division into distinct training, validation, and test sets, facilitating independent model selection on the validation set.
•This strategic approach significantly bolsters our model selection process, providing a more reliable estimation of future performance when contrasted with the conventional two-way data split and test-based model selection methodology.

KEY INSIGHTS



•Six engineered features, like km_per_hour and percent_sessions_in_last_month, were among the top 10 influencers.
•XGBoost excelled over random forest, notably with a 17% recall, maintaining similar accuracy and precision.
•The ensemble of tree-based models, including XGBoost, consistently outperformed the random forest model across all metrics, despite being less interpretable, making them more effective in predicting churn.