

Executive Summary

Waze App User Churn Project | Machine Learning Results

Prepared for Waze Leadership Team

➤ ISSUE / PROBLEM

The waze data team is looking to increase overall growth by identifying what drives user churn. User “churn” is identified as the number of users that have uninstalled the app or have stopped using the app completely. The team wants to build a Random Forest and XGBoost machine learning model to see which model predicts user churn the best.

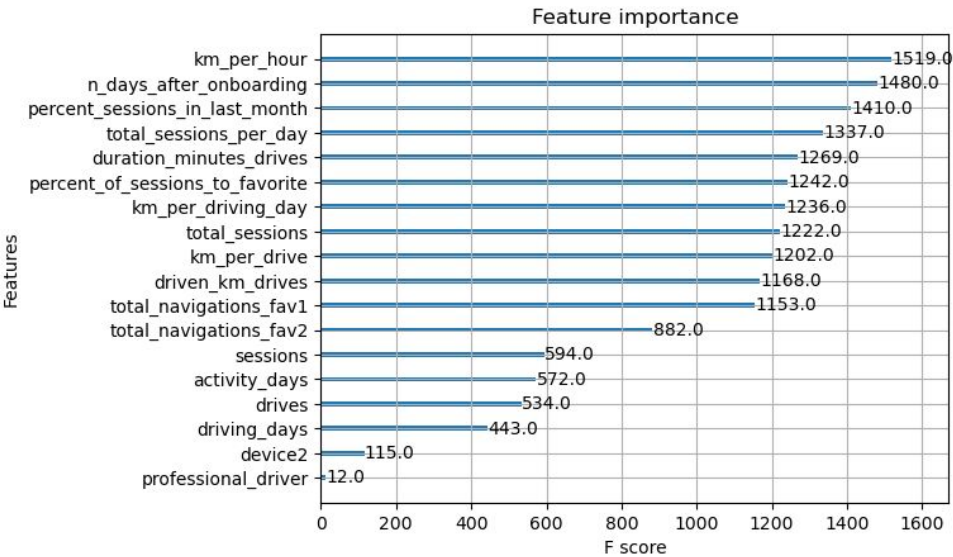
➤ IMPACT

- The model created is not a good predictor model for user churn, due to its low performance scores.
- This model also confirms that the current data is insufficient to predicting user churn. It would be helpful to have drive-level information on users (such as drive times, geographic locations, etc.)
- The model also shows that engineered features are a valuable tool for improving the performance of the model. So the waze team recommends generating new features to better predict user churn. Then reconstructing the model with different combinations of variables.

➤ RESPONSE

- To obtain a model with the highest predictive power, the Waze data team developed two different models to cross-compare results: Random Forest and XGBoost.
- To prepare, the original data was split into training, test and validation sets. This causes less data to train the model with. But, having a validation set allows the team to test on the champion/winning model, which gives a better estimate on future performance.

➤ KEY INSIGHTS



- Engineered features accounted for the top 10 features to predict user churn.(km_per_hour, percent_sessions_in_last_month, total_sessions_per_day, percent_of_drives_to_favorite, km_per_drive, km_per_driving_day.)
- Between the Random Forest and XGBoost Model, we noticed that the XGBoost model performed slightly better. The precision and accuracy scores are pretty similar for both models. But the recall score on the XGBoost model is 17%, which is about 21.428% better than the Random Forest model.