

User Churn Project | Machine Learning Results

Prepared For: Waze Leadership

➤ ISSUE / PROBLEM

The Waze data team is developing a data analytics project with the goal of quantifying and reducing Waze user churn, which will lead to overall growth.

Churn refers to the number of users who have stopped using the Waze app or uninstalled it. A machine learning model will be developed to predict this.

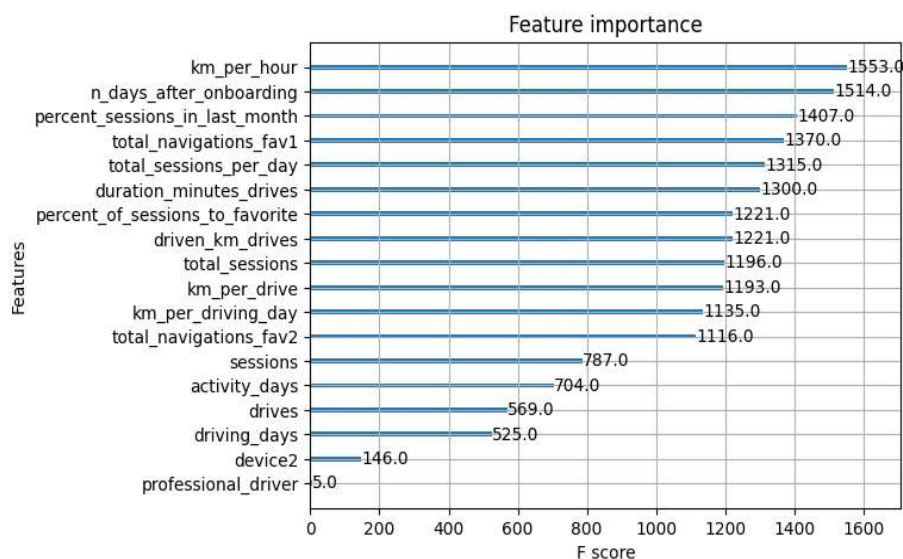
➤ IMPACT

- The models developed during this stage of analysis indicate that further data is required to create a model that is capable of more accurately and consistently predicting user churn.
- It would be useful to have more granular, drive-level data, such as geographic locations, drive times, and more information on how the users interact with the app throughout the duration of its usage.
- The recommendation is to complete a second iteration of the project with the newly acquired data, along with newly engineered variables based on domain knowledge.

➤ RESPONSE

- In order to produce a model with the most predictive power given the data at our disposal, both a Random Forest and an XGBoost model were created to compare results.
- In preparation, the data was split into training, validation, and test sets. Although this resulted in less data available for training purposes, it allowed us to validate the models to choose a champion model, then finally test it on the test set alone, which gives a better indication of future performance on unseen data.

➤ KEY INSIGHTS



- Newly engineered features accounted for 6 of the top 10 features in the Feature Importance plot, indicating that newly engineered features can have significant predictive power.
- The XGBoost model fit the data better than the Random Forest model, with a recall score of approximately 0.17, which is about double that of the recall score acquired from the Logistic Regression model built in the previous phase.
- Although more difficult to interpret and identify how they make their predictions, the ensemble-based models outperformed the regression model in all evaluation metrics.