Waze User Churn Project: Machine Learning

Prepared for Waze Leadership Team

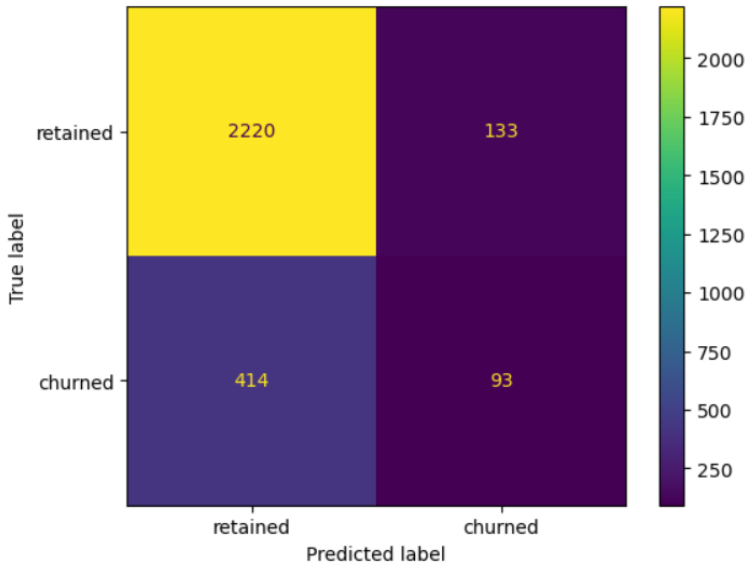
Project Overview

The Waze data team has conducted a data analytics project in predicting user churn and increasing the user retention rate of the app. This report summarizes the findings and insights made for Milestone 6, where it focuses on implementation of machine learning models like random forest and XGB boosting, and how they play an essential role in future development of this project.

Key Insights

- Prior to fitting the model, three different datasets are created: one for training, one for validation, and one for testing. While this approach may lead to less data available to train and evaluate, the validation set verifies the performance metrics for each model and minimize overfitting and the test set is reserved for the best model to predict values for unseen data.
- Out of the two machine learning models built (random forest and XGB Boosting), the latter is the champion model, owing to the higher recall score when predicting the values using the validation data.
- Based on the confusion matrix, there is approximately three times as many false negatives present compared to false positives.
- From the plot importance chart, the kilometers per hour is the feature that is the most significant in contributing to user churn.

Details



Confusion matrix of user churn predictions generated with XGB Boosting.

Next Steps

- The key insights obtained from this milestone can be used to provide further exploration.
- Given with the recall score being somewhat below average than usual, the model should not be used to data-driven decisions. However, the model does provide useful insights and results on features that may contribute most to user churn, and further study is recommended to digest those essential information.