

Machine Learning Model Outcomes for User Churn

Executive Summary Report for Waze

Overview

Waze's data team is close to completing their user churn project. The project proposal was completed, the user data was analyzed with Python, a hypothesis test was conducted, and a binomial logistic regression model was built. The final step is to build and test different machine learning models for predicting user churn.

Objective

- Perform feature engineering
 - Build the following machine learning models: random forest and XGBoost
 - Evaluate the models
-

Results

- For the random forest model:
 - Best hyperparameters: max_depth: None, max_features: 0.7, max_samples: 0.7, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 75
 - Metric scores from validation: F1 = 0.195652, recall = 0.119093, precision = 0.547826, accuracy = 0.818881
 - For the XGBoost model:
 - Best hyperparameters: learning_rate: 0.1, max_depth: 12, min_child_weight: 5, n_estimators: 500
 - Metric scores from validation: F1 = 0.223448, recall = 0.153119, precision = 0.413265, accuracy = 0.803147
 - XGBoost was chosen as the champion model with the higher recall score.
 - The most important features in XGBoost are km_per_hour, n_days_after_onboarding, percent_sessions_in_last_month, etc.
-

Next Steps

- Improve the hyperparameters of the XGBoost model to improve its predictive power or modifying the features used.
 - Use the current XGBoost model with consideration of the Type I and Type II errors.
- 