MACHINE LEARNING MODEL THAT PREDICTS USER CHURN

waze

Machine Learnig Model Results

ISSUE / PROBLEM

The Waze Data Team is developing an analytics project to boost growth by predicting and reducing monthly user churn—defined as users who uninstall or stop using the app. This report presents key insights from Milestone 6 to inform future model development.

RESPONSE

To achieve strong predictive performance, the Waze Data Team developed and compared random forest and XGBoost models. They split data into training, validation, and test sets. This 3-way split reduces training data but allows for more reliable model evaluation using a separate test set.

IMPACT

This modeling effort shows current data is insufficient to reliably predict churn. More detailed user data—such as drive times, locations, in-app interactions, and trip patterns—could enhance accuracy. A second project iteration is recommended with improved features.

- Table of scores from the predictions on the validation data.

	model	precision	recall	F1	accuracy
0	RF cv	0.457163	0.126782	0.198445	0.818510
0	XGB cv	0.425932	0.170826	0.243736	0.811866
0	XGB cv	0.425932	0.170826	0.243736	0.811866
0	RF val	0.445255	0.120316	0.189441	0.817483

- xgb_cv model confusion Matriz
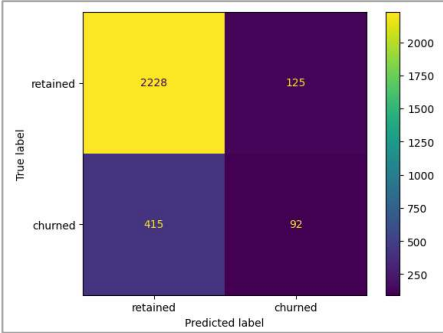

- Feture importance for xgb_cv model

KEY INSIGHTS

- Engineered features made up six of the top ten most important variables: km_per_hour, percent_sessions_in_last_month, total_sessions_per_day, percent_of_drives_to_favorite, km_per_drive, and km_per_driving_day.

- The XGBoost model outperformed the random forest model in overall fit. Notably, its recall score (17%) nearly doubled that of the logistic regression model used in Milestone 5, while maintaining comparable accuracy and precision.

- Compared to the singular logistic regression approach, the ensemble tree-based models used in this milestone offered stronger performance across all evaluation metrics and required less data preprocessing. However, they are inherently more complex and less interpretable, making it more challenging to understand how specific predictions are made.