Regression Analysis and Model Development for Predicting User Churn

Milestone 5

Project Overview

The goal of this phase is to build a binomial logistic regression model to predict user churn based on various factors. The model will be developed using Python and is aimed at understanding the relationship between user behavior and the likelihood of churn. This analysis will help inform the next steps in the machine learning algorithm development process

Key Insights

Professional drivers had a **churn rate of 7.5%**, significantly lower than the **19.8%** churn rate for non-professionals — indicating that this feature adds meaningful predictive value.

There is strong multicollinearity between several variable pairs, notably between `sessions` and `drives` (r = 0.997) and between `activity_days` and `driving_days` (r = 0.948), indicating redundancy. To address this, `sessions` and `driving_days` should be dropped from the logistic regression model.

The `activity_days` variable had the **strongest negative coefficient (-0.103)** in the model, indicating that more active days strongly reduce the probability of churn.

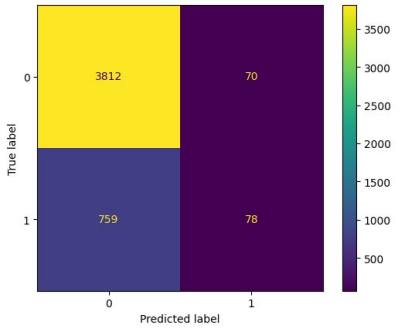
The model showed **high precision (83%)** for predicting non-churned users but **low recall (9%)** for predicting churned users — meaning it struggles to correctly identify users who are likely to churn.

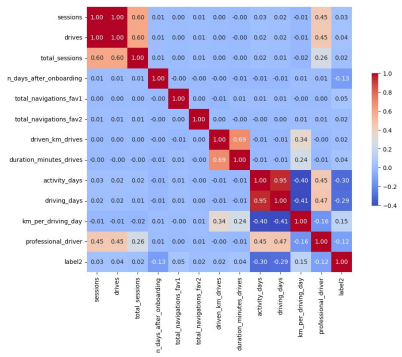
Out of **837 churned users**, the model correctly predicted only **78** (true positives), misclassifying **759** as retained (false negatives).

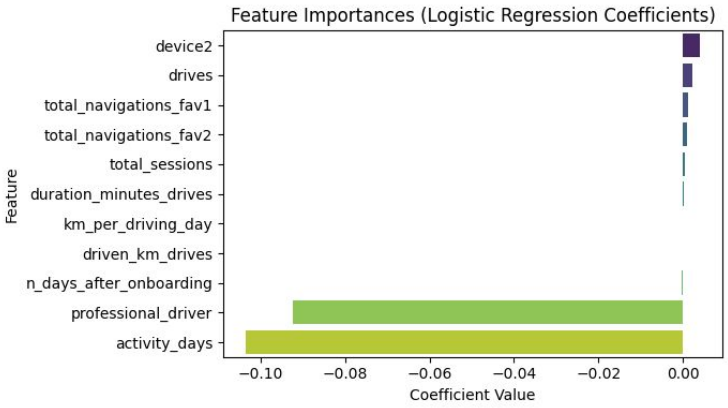
The overall model **accuracy was 82%**, but due to the class imbalance and low recall, it may not be suitable for applications that require strong churn detection.

Features like `total_sessions` **(0.0004)** and `drives` **(0.0024)** had small positive coefficients, indicating only a weak relationship with churn.

Details







Next Steps

Based on the model's outcomes, it is recommended to use the insights from this phase to inform further analysis. While the current model is not yet reliable enough for major business decisions, it highlights the need for additional features that more strongly correlate with churn. It also suggests a potential need to refine the definition of the target user profile in efforts to reduce monthly churn and support user growth.