

User Churn Project Logistic Regression Model Results

Executive Summary for Waze

Overview

The data team is more than halfway through the user churn project. The next step is to build a regression model to predict user churn based on a variety of variables.

Objective

- Check model assumptions
 - Build a binomial logistic regression model
 - Evaluate the model
-

Results

- The following variables were created from calculations of existing variables or encoding:
 - ``km_per_driving_day`` = ``driven_km_drives`` / ``driving_days``
 - ``professional_driver``, if ``drives`` ≥ 60 and ``driving_days`` ≥ 15 then 1, otherwise 0.
 - ``label2``, if ``label`` is retained then 0, otherwise 1 (churned).
 - ``device2``, if ``device`` is Android then 0, otherwise 1 (iPhone).
 - The Logistic Regression model was fitted with target variable ``label2`` and multiple predictor variables, such as ``activity_days``.
 - The accuracy of the model is around 82.6%
 - The classification report shows the following values:
 - Precision: 55%
 - Recall: 10%
 - The model has decent precision but very low recall, which means that it makes a lot of false negative predictions and fails to capture users who will churn.
-

Next Steps

- Consider splitting the dataset before removing rows with missing values.
 - Create a new logistic regression model that uses different feature variables.
- 