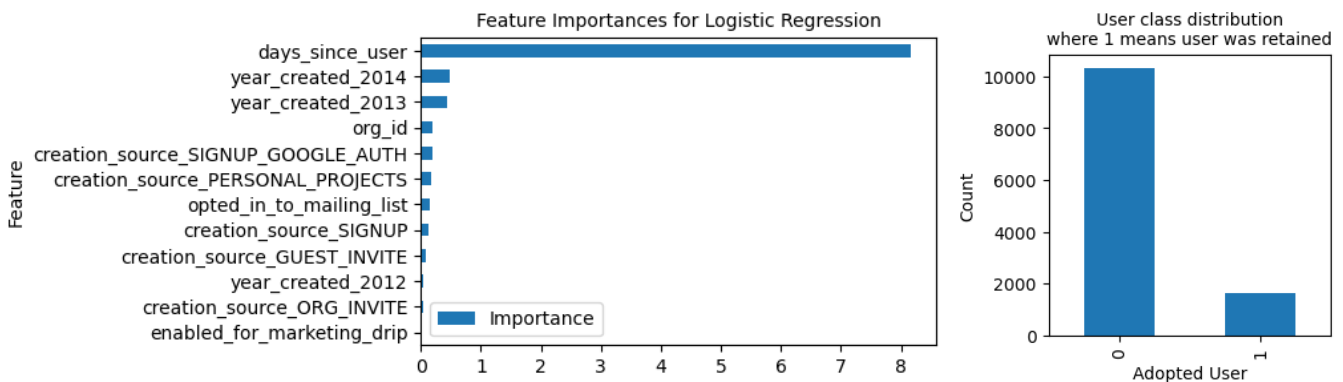


Relax Inc. Project Report

Relax Inc. is a company that makes productivity and project management software. The focus of this project was to use data from 12,000 users to identify which factors predict future user adoption; an adopted user was defined as a user who logged into the product on three separate days in at least one seven day period.

I built a binary logistic regression model to predict user adoption of the software. I used data from 12,000 users, including features such as how their account was created and whether they have opted into marketing emails, as well as data on their login behavior over a 2 year period. Based on data from when each user created their account and when they last logged into the software, I created two new features: 'days_since_user' tracks the number of days between account creation and the last user login, and 'year_created' tracks the year the account was created.

By far the most important (and positive) predictor of user adoption as identified by the model was the number of days between account creation and the last user login (see figure below). This is not surprising, given that users who are still logging in a long time after creating their account are likely to be those that are consistently using the software. Unfortunately, this is not a very interesting result in terms of aiding business policies for increasing user adoption. Other (far less significant) predictors were the year the account was created (2014 being a positive predictor and 2013 a negative predictor), and the organization users belong to.



The model had a 97% accuracy, as well as high recall and precision. I tested out several other model permutations, including using a MinMaxScaler instead of StandardScaler, and omitting the 'days_since_user' feature; all of these modifications resulted in a much lower model performance. Given more time, I would pursue under-sampling the majority class (not_adopted) to deal with class imbalance in target variable, as well as testing out several other classification models (gradient boosting classifier and SVM, for example).

The user data was quite minimal, and further feature engineering could help produce a more meaningful model. Moreover, gathering more data on users (age, demographic data, and what industry they are in) could greatly increase the model's worth as a predictive tool for Relax Inc.