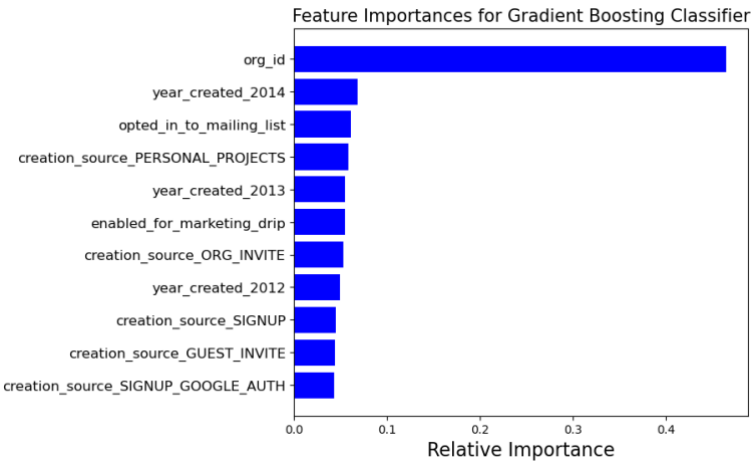
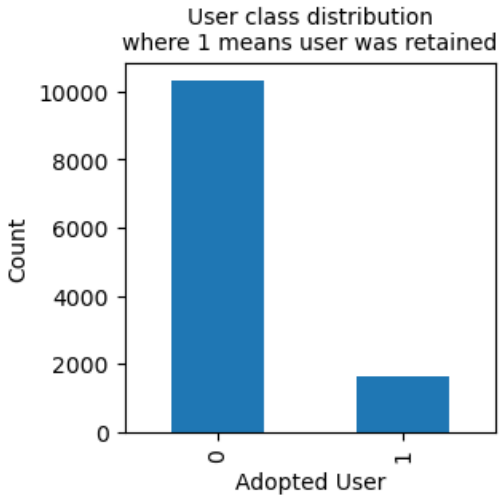
Relax Inc. Project Report

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 gradient boosting classifier 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. I addressed missing values and deleted duplicate rows. Based on data from when each user created their account and when they last logged into the software, I created a feature: 'year_created' tracks the year the account was created.

By far the most important (positive) predictor of user adoption as identified by the model was the organization that the user belonged to (see figure below). This could help Relax Inc. increase user adoption by building marketing campaigns to target the types of companies likely to adop the software – further research into this area would be necessary. Other (far less significant) predictors were the year the account was created (2014 was a very good year!), and whether users opted into mailing lists.





The model had an 80% accuracy on the test data, as well as similar recall (86%) and precision (77%) for correctly identifying adopted users. I tested out several other model permutations, including using a MinMaxScaler instead of StandardScaler, and a logistic regression and SVM model; all of these permutations resulted in a much lower model performance. To address the large class imbalance in the target variable, I oversampled the minority class (adopted users), leading to a much better model.

The user data were 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.