Relax Challenge Report

Zheng Zhang, Ph.D.
September 25, 2020

Purpose

We have data on 12,000 users who signed up for the product in the last two years. We define an "adopted user" as a user who has logged into the product at least three times, and the average time between two consecutive logins is less than 3 days. We seek to predict which factors predict future user adoption.

Exploratory Data Analysis

There are 651(5.425%) adopted users. The predictors are whether they opted for marketing emails (25%), whether they are on regular marketing email drip (15%), and how their account was created, which is distributed as: organization invite (35%), sign-up through website (18%), sign-up through google(12%), personal project(18%), guest invite(18%).

Predict User Adoption with Machine Learning Techniques

Due to the fact that only 5% users are adopted, we have chosen the balanced random forest as well as AdaBoost to build our prediction model. The model building process using different machine learning algorithms share some of the following common steps: 1. Split the dataset into a 70% training set and 30% test set, stratified by the adoption label. 2. Some algorithms have hyperparameters, where GridSearch has been used to identify the best choice for such hyperparameters (hyperparameter tuning) by using 5-fold cross-validation.
Balanced random forest achieves sensitivity of 63%, and a low specificity of 48%. The AdaBoost method achieves sensitivity of 64% with a specificity of 44%.
Overall, we will recommend balanced random forest to predict future user adoption. The most important predictors are: joined for personal projects, opted for mailing list and enabled for marketing drip.

Conclusion

We found user adoption is mostly related to how the account was created, especially for those who were invited to join another user's personal workspace. In addition, opting to the mailing list and enabling marketing drip are also predictive of user's adoption.