

Relax Inc.  
Data Science Take Home Challenge  
User Analysis

When initially opening the user data (takehome\_users.csv) it's clear not all users have even created a single session. 27% of users were interested enough to create an account but didn't bother going any further. While it is good to have new signups this means little if they don't actually use the product. The initial hook to get them in is working, but the reel needs to be improved in order to continue pulling in their interest and engagement. We're missing out on at least 3,177 users over the past 2 years.

75% of users that were invited by other users opened at least one session. This is slightly better than overall (~73%) but doesn't immediately seem to be a very good solution to keeping users interested.

Users were sorted out as adopted users; defined as a user that has logged into the product on 3 separate days in at least a 7 day period. About 1,600 users were found to be adopted users; less than 14% of all users in this data set (12,000). In order to understand these users better and help engage more of them, we'll look at the factors that have the biggest impact in creating an adopted user.

The simple, but robust random forest algorithm will be used for this purpose. After splitting the data to training and testing sets with the adopted users as the dependent variable (what we want to predict) and the remaining features as the independent variables (what's being used to predict) the model is trained using a 60/40 train/test split. The stratify feature in the split package is also utilized for the slight imbalance in data.

Using GridSearchCV, parameters for the random forest are optimized for best accuracy; about 87%. In this case accuracy may not be the best metric, however the recall and precision for the model are quite good as well at 1 and .87 respectively.

Based off of this model the most predictive feature is the creation source of the account. Out of the five possibilities an invite to an organization as a full member was the most likely to produce an adopted user. Next, was if the user was invited to join another user's personal workspace. This is not entirely surprising as it means the users had a specific reason to continue logging in, versus someone who may be just creating an unsolicited account and deciding they aren't interested in continued use. It would be a good idea for the company to increase marketing to organizations as opposed to individuals and emphasize project collaboration.

The feature with the third highest weight is if the user is on the marketing drip email list. This may simply be because continued users want to keep up with current news or updates - but it may also be a good idea to increase email marketing to new users.

While signing up with a google account is the fourth highest weight it is not necessarily a useful metric - it could be argued however, that increasing ads for google users (everyone?) may result in an increase of adopted users.

The lowest importance is from signing up on the website itself. This is probably from the population that doesn't have any real initial reason to create an account and once made, it doesn't hold their interest. While you could look at this as the website not having a real impact, I think it would be a good idea to possibly rework the website to better sell the program's best uses and features.