

Relax Take Home challenge

- Load Required Data into pandas data frame
- Data Cleaning/Wrangling
- Create "Adopted" column 1 if adopted else 0 , based on the given definition.
- Correlation Analysis
- Feature importance analysis (using ensemble - random forest) – failed
- Principal Component analysis
- Analysis Report

Identified features those are important in prediction of future adoption users , based on the dataset provided. Below are the features that were deemed important based my analysis.

• PERSONAL_PROJECTS	2.111889
• SIGNUP	2.016023
• GUEST_INVITE	1.900506
• ORG_INVITE	1.449156
• SIGNUP_GOOGLE_AUTH	1.444174
• enabled_for_marketing_drip	1.387755
• opted_in_to_mailing_list	1.387444
• org_id	1.000000

Important feature is how user signed up (if you add the importance of the four signup features above).

Among the ways they sign up, "PERSONAL_PROJECTS" and "SIGNUP" are the two categories with highest user adoption rate.

Hence, encouraging users to invite others, and offering incentives for people to work on their personal projects using the software might be effective.

I recommend that an effective way of growing adopted users might be encouraging existing users to log in and use the software after they had accounts for a while.

Mailing list and marketing drip are not effective.

The first 5 principal components account for most of the data variation in the dataset of adopted users, meaning that these variables are highly predictive of adopted users.