STorrico

Relax Challenge Summary

We were given the task to identify the factors that predict user adoption for a particular product using information such as when the user created its account, how it was created and the number of times a user logged into it.

To achieve that, we used Random Forest since it performs well in classification tasks and provides information about important features. The Random Forest was trained with 70% of our data and predicted on a test dataset of 30%. During the split of the datasets into training and test, stratified, was used due to the limited observations for user identified as an "adopted user".

Random Forest achieved a recall of 88% for the classification of all "adopted" users category. The model found the below important features:

	importance
last_session_creation_time	0.635812
creation_time	0.312376
org_id	0.026971
invited_by_user_id	0.013525
creation_source_GUEST_INVITE	0.003320
creation_source_PERSONAL_PROJECTS	0.002178
opted_in_to_mailing_list	0.001521
creation_source_SIGNUP	0.001153
creation_source_ORG_INVITE	0.001120
enabled_for_marketing_drip	0.001111
creation_source_SIGNUP_GOOGLE_AUTH	0.000913

According to our model, the factors that predict future user adoption are **last_session_creation_time** followed by **creation time** , **org_id,** and **invited_by_user_id**. Not that much surprise by last_creation_time, which tracks when they last logged in and the many null values observed indicated that many of the users have never logged into the product. Instead of dropping these observations, we went ahead and change their value to 0 to minimize the bias.

Creation_time came as second, which is an indication of the amount of time users had to become an "adopted" user. Next on the list were **org_id** and **invited_by_user_id** made it to the top list, which indicate that adopted users would be most likely to login as long as their organization continues to do work there.

A main insight that can be obtained from these 2 variables is that there is an opportunity to increase the usage of the "not adopted' users by providing incentives to the communities they are part of. It will be interesting to find out the size of the communities or organizations all users belong to and track how active (at least 25% of their users logged in during the last month) these organizations are so their impact could be analyzed.