

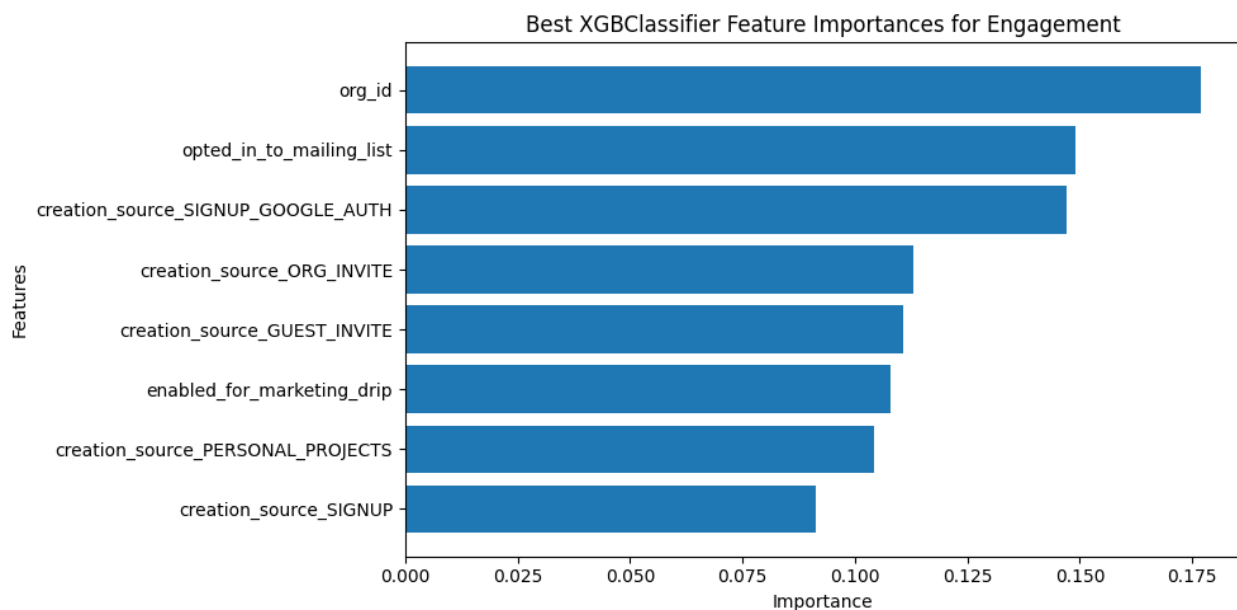
## Relax Challenge

Andrew McLaughlin

After loading in the takehome user engagement file, I counted the number of times each user visited each week. I then kept the maximum visits in a week associated with each user. I merged that with the takehome users dataframe, so that each row had a column counting that users maximum number of times visiting each week. I converted that column into a column called engaged, which has a value of one if the max number of visits in a week was three or more, and zero otherwise. Then I dropped the max visit counting column as it is now superfluous.

I decided to find the relationship between opting into the mailing list and engagement. I found that 16.9% of users that opted in to the mailing list were engaged users, and that 16.1% of users not in the mailing list were engaged users.

Next I created a model to predict user engagement which had as its target the engaged column. After some testing of different models I decided to use the xgboost classification model, as it had a good mix of accuracy and interpretability. My xgboost model correctly classified examples from a test set with 83% accuracy. The importance of each feature in creating the model is as follows:



It would appear that the organization of the user is most important, however that is not controllable on our end. Next most important is the mailing list. It would appear that, despite the closeness of the differences between engagement on and not on the mailing list, it is still important. We should encourage people to join the mailing list.