RELAX CHALLENGE REPORT

SCOTT ATKINSON

In this report we analyze user data by analyzing which features are influential in predicting if a user will "adopt"the product. That is, given user data, we assess the the importance of each feature in predicting if a user will use the product three separate times in the span of a single week. To clean the data, we replace missing time values with a stand-in utc=0 value, and we replace missing user invite data with a stand-in 0 value. We then engineer several features including the target label `adopted` which returns `True` if the user uses the product three separate times in the span of a single week and returns `False` otherwise. The other features created are `visits`, `visits_per_day`, and `creation_to_last` (length of time from account creation to last session creation). Since these features are only available after some time has passed since account creation, they cannot be used to predict account adoption at sign-up. We do see that the `creation_source` variable does indicate some relationship in the bar chart in Figure 1 obtained during EDA.

We perform VIF analysis to determine feature importance for the features `creation_source`, `opted_in_to_mailing_list`, `enabled_for_marketing_drip`, `org_id`, `invited_by_user_id`. The resulting coefficients appear in Figure 2. From the results `enabled_for_marketing_drip` is positively influential feature, and the `creation_source_PERSONAL_PROJECTS` feature is the most negatively influential.

Since user interaction with the product is indicative of adoption, it might be worth looking into extending the model to take into account user activity in the first week or so.

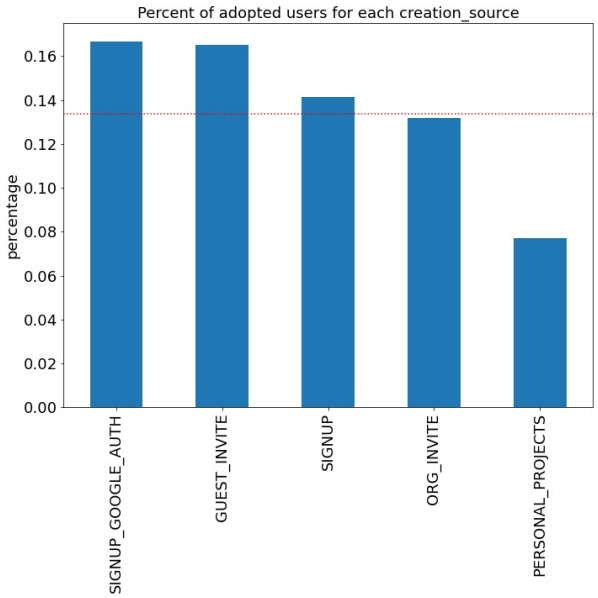

FIGURE 1. Bar chart for percentage of users who adopted the product grouped by `creation_source`. The red dotted line is the overall percentage of adopted users.

	coef
enabled_for_marketing_drip	0.014000
org_id	0.001600
invited_by_user_id	-0.000002
opted_in_to_mailing_list	-0.016400
creation_source_GUEST_INVITE	-0.078800
creation_source_SIGNUP_GOOGLE_AUTH	-0.106700
creation_source_SIGNUP	-0.258900
creation_source_ORG_INVITE	-0.364400
creation_source_PERSONAL_PROJECTS	-0.956500
const	-1.765300

FIGURE 2. VIF analysis results