**Question:**

Defining “adopted user" as a user who has logged into the product on three separate days in at least one seven­day period, identify which factors predict future user adoption.

**Answer:**

Key Takeaways:

1. Four factors best predict future user adoption (in descending order):
2. the size of the organization to which the user belongs,
3. how the user’s account was created,
4. the number of other users the user invited to join, and
5. the number of users invited by the inviter if the user was invited to join.
6. The trends for the four best factors are as follows:
   1. users from smaller organizations were more likely to become adopted users,
   2. users that were invited to join another user’s personal workspace were less likely to become adopted users and users that joined using Google Authentication were more likely to become adopted users,
   3. users that invited at least two other users to sign up were more likely to become adopted users, and
   4. if the inviter of a user invited more than 3 users to sign up, then the user was more likely to become adopted users.

Other Findings:

The time at which the user’s account was created also trended with user adoption, however, this may be an anomaly. Users that signed up during Jan-Mar of 2014 were unusually likely to become adopted users while users that signed up during Apr and May of 2014 were not. This monthly trend was not seen in previous years.

Follow-up:

1. Speak with client to better understand at what time would they like a prediction to be made: can we observe the user’s behavior for a week or month before making a prediction?
2. Can we add additional granularity to the “takehome\_user\_engagement.csv” table to include what type of activity or activities the user logged while they were logged into the product?

Method:

Important factors for predicting future user adoption were identified by creating a random forest classifier (class\_weight='balanced', criterion='log\_loss', max\_depth=10, n\_estimators=150) and extracting feature importances. The random forest classifier was created using Python and Scikit-Learn. Feature importances were also extracted using Scikit-Learn and are shown below.

The data were processed in the following manner:

* “Adopted user” was defined using the provided definition and calculated from “takehome\_user\_engagement.csv.”
* Of the original features in “takehome\_users.csv,” “creation\_time,” “creation\_source,” “opted\_in\_to\_mailing\_list,” “enabled\_for\_marketing\_drop,” “org\_id,” and “invited\_by\_user\_id” were used to predict “adopted\_user.”
  1. “creation\_time” was processed to extract the month and day of week of creation resulting in two features.
  2. “creation\_source” was created as a categorical variable with five possibilities.
  3. Both “opted\_in\_to\_mailing\_list” and “enabled\_for\_marketing\_drip” were not preprocessed for modeling.
  4. “org\_id” was processed to extract the logarithm of the number of users per organization. Since the number of users per organization roughly follows a Poisson distribution, the logarithm used to create a more normally distributed variable.
  5. “invited\_by\_user\_id” was processed into two features, one that stores the number of invites sent (and accepted) by the user and another that stores the number of invites of the inviter, if the user was invited by another user.

A graph of a number of blue and black text

Description automatically generated with medium confidence

Figure 1: Feature importances of random forest classifier predicting adopted users.