**Results on User Adoption Based on Initial Experience**

Ying Xu

Given the definition of adopted user, I was able to create a binary variable “adopted\_user” in the users table. Among 12,000 users who signed up for the product in the last two years, 1656 (13.8%) are considered adopted users.

In order to increase the user adoption, we need to understand what features contribute the most to becoming an adopted user. So I decide to use logistic regression and decision tree on “adopted\_user” given their high interpretability of predicting variables.

So, in total, we use the following seven features from the users table as predicting variables: 'email\_domain', ‘create\_source', 'last\_Session\_creation\_time', 'opted\_in\_to\_mailing\_list', 'enabled\_for\_marketing\_drip', 'org\_id', and 'invited\_by'. Especially, we create a dummy variable for ‘create\_source’, which represents 5 different ways people sign-up the project: organization invite (4254), guest invite (2163), personal projects (2111), signup (2087), and signup using Google authentication (1385).

Before conducting the analysis, I split the dataset into train (70%) and test (30%) datasets.

Using LogisticRegressionClassifer() from sklearn package in python, I found that the model has a classification accuracy of 0.875, with a 77% precision and a 88% recall rate. Since I have 86.2% of the chance to guess it right already, the model does not work well.

Using DecisionTreeClassifier(), I found that the model has a classification accuracy rate of 91.8%. Attached is the ROC curve, as well the decision tree with a depth of 4. From this tree, it seems that the time a user last log-in is the clearest predictor.

More studies should be done to further understand our adopted users’ strong features.



