Relax Inc. Data Science Challenge Report

1 Problem Statement

Aim of this project was to find out which factors from the user table will predict future user adoption. User login information is given in the takehome_user_engagement.csv file. An user is defined adopted if that user has logged into the product for more than three times on three distinct days in a span of atleast one consecutive period of 7 days.

Using the user login table, I first selected all the users who satisfy the criteria to be labeled as an adopted user. Then I merged this filtered table based on user_id with the user table to get a new table. Now this new table only consists of information about adopted users. After filling missing entries, one hot encoding and data scaling, I applied Principal Component Analysis (PCA). By using PCA I find contribution of different factors that contribute towards the principal components. Figure 1 shows contributions of different factors. As evident from the figure 1, factor 'last_session_creation_time' seems to be most important for predicting an adopted user. Following 'last_session_creation_time', other factor of importance are 'creation_time'. More details can be found in the python notebook in the repository.

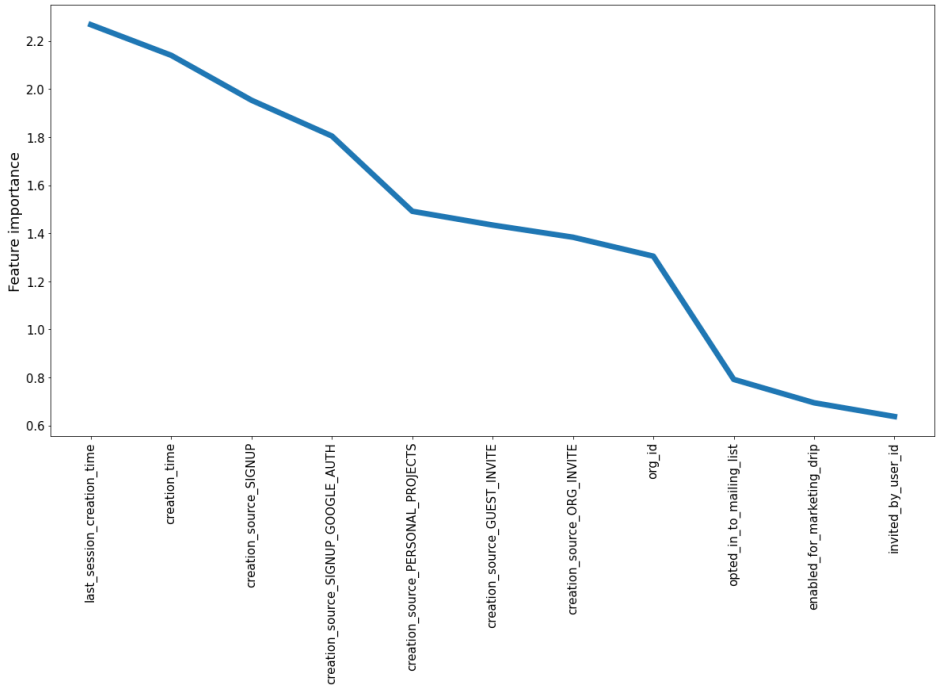

Figure 1: Feature importance of different factors for predicting user adoption