Relax Inc., User Adoption Prediction

The goal of this project is to analyze the data on 12000 users that have signed up for the product in the last 2 years and identify which factors predict the future of user adoption. We define user adoption as a user who has logged into the product on 2 separate days in at least one 7-day period. There are 2 data files provided – one containing the user information and the other containing user login information

Procedure

- First, we identify the adopted users from the dataset by grouping the users and datetime sampling via a moving window (rolling) to count the number of logins in a window of 7 days
- Using this, we perform EDA to see any interesting trends in the adopted user
- We then fit a Random Forest Classifier into the users dataset to check if there is a way to predict if the user would be adopted or not

EDA Insights

- Number of adopted users ~ 13%
- Users signing up via Guest Invite or Google Auth have the maximum conversion rate of 17%
- Factors such as opting in for mailing list, or signing up for marketing drip or user invitation doesn't have any impact on whether the user will adopt the product or not
- There are few organizations which have ~50% users who have adopted the product

Modeling Insights

- Last_session_creation_time is the most relevant feature, though it doesn't really make sense to consider this in the prediction model. This feature indicates the last time user has logged in. Most users who have not adopted the product will have an old last_session_creation_time. Hence, we exclude this feature in our modeling
- Org_id (the organization id) is the next most relevant feature (which seems to agree with the EDA analysis)
- The model achieves an overall F1-score of 0.8

	Feature Importance
creation_source	0.040340
opted_in_to_mailing_list	0.012437
enabled_for_marketing_drip	0.010902
org_id	0.930767
invited_by_user	0.005553

	precision	recall	f1-score	support
0	0.87	0.98	0.92	2080
1	0.21	0.03	0.05	320
avg / total	0.78	0.85	0.81	2400