‘Relax’ User Adoption Prediction Model

We’re given datasets about users who use the product software provided by Relax Inc. We will combine information pertaining to user login frequency and user information to predict whether a user is defined as adopted.

Data

We’re given two data sets:

1. *takehome_users.csv* - A user table with data on 12,000 users who signed up for the product in the last two years. This table includes
2. *takehome_user_engagement.csv* - file that provides the dates that a user uses the product

1. Data Cleaning

We deal with missing values and change date features into datetime objects. We then filter the *takehome_user_engagement.csv* into weekly intervals to see how many times a user will have logged in. **If in any one week span a user logs in three times or more, we define the user as *adopted*.** We created a binary feature *‘adopted’* that equals 1 if a user has logged in three or more times and 0 if they have never. We want to achieve at least 0.75 for all precision and recall values.

2. Basic Bar Plots and Preprocessing

We plotted some basic barplots for the remaining independent features after feature selction and created dummy features for categorical features with less than 10 unique values.

4. Model Results and ROC Curve

We used a random forest classifier model for its robustness in binary classification problems. We use grid search cross validation with the objective of obtaining a high roc-score which indicates a good balance between true positive and false positive rates. We were able to obtain a roc score of 0.778 with the respective precision and recalls for the adopted features being at least 75% or greater.

```
Random Forest: Roc_auc=0.778
              precision    recall

           0       0.76      0.81
           1       0.80      0.75
```