

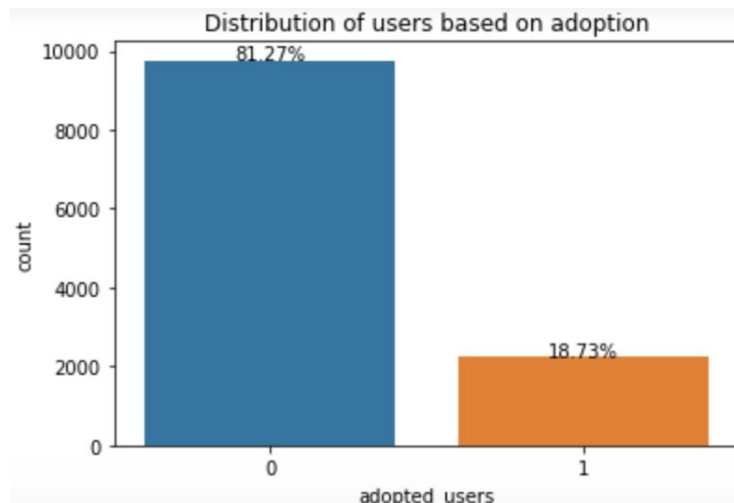
Relax Inc Take Home Challenge Report

Data Cleaning and Feature Engineering

I first created 'adopted_users' in the 'user_engagement' dataset and then combined 'user_engagement' and 'users' datasets. I filled in the missing values and dropped some useless or auto-correlated rows and columns when appropriate, and created a few features from time objects.

EDA

81.27% of users do not adopt the product while only 18.73% of users do. Creation source appears to be negatively correlated with invited_by_user_id.



Modelling

Two different classification algorithms were developed: logistic regression and random forest. Random Forest Model was the better model because it had a higher f1 score which was 1.00 for training set and 0.967 for validation set. Due to the imbalanced nature of the dataset, f1 score and AUC-ROC curve were used for evaluating the models.

Key Findings

The model suggested that last login time was highly correlated with adoption decisions. Other factors such as the year and day when the account was created, organization id also influence the adoption decisions.

Recommendations

1. The company may use marketing campaigns, news, updates, products demos to attract users to visit the website. The more people visit the website, the fewer people will adopt the product.
2. The company may provide discounts or optimize enterprise pricing to attract organizational users and invite more organization employees to visit the website.