Challenge Overview: Predicting User Adoption**
The objective of the take-home challenge was to discern key features contributing to the prediction of future user adoption. User adoption is defined as a user logging into the product on at least three separate days within any seven-day period.

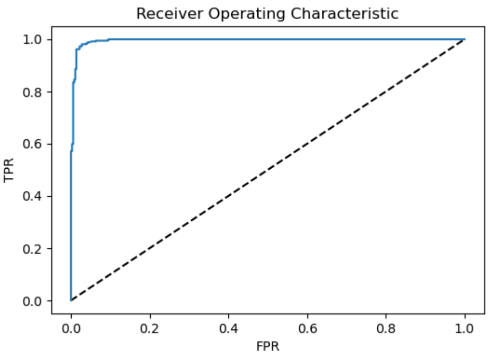
Data: Two CSV files were provided for the assignment.
The first table contained user information, encompassing email, organization affiliation, subscription to regular marketing emails, account creation date,, and more. The second table detailed user engagement metrics, focusing on user activity concerning product logins.

Data Wrangling and Data Exploration Insights:
Upon data loading, date object columns were converted to datetime format. Missing information was addressed, including the addition of zeros for users not referred by another user in the first table. Missing data in the last login date column was imputed with the account creation date.

Initial data exploration unveiled notable insights. The dataset spans three years: 2012-2014, with 2013 being a peak year for new user sign-ups. The top three months for user registrations were observed to be March through May.

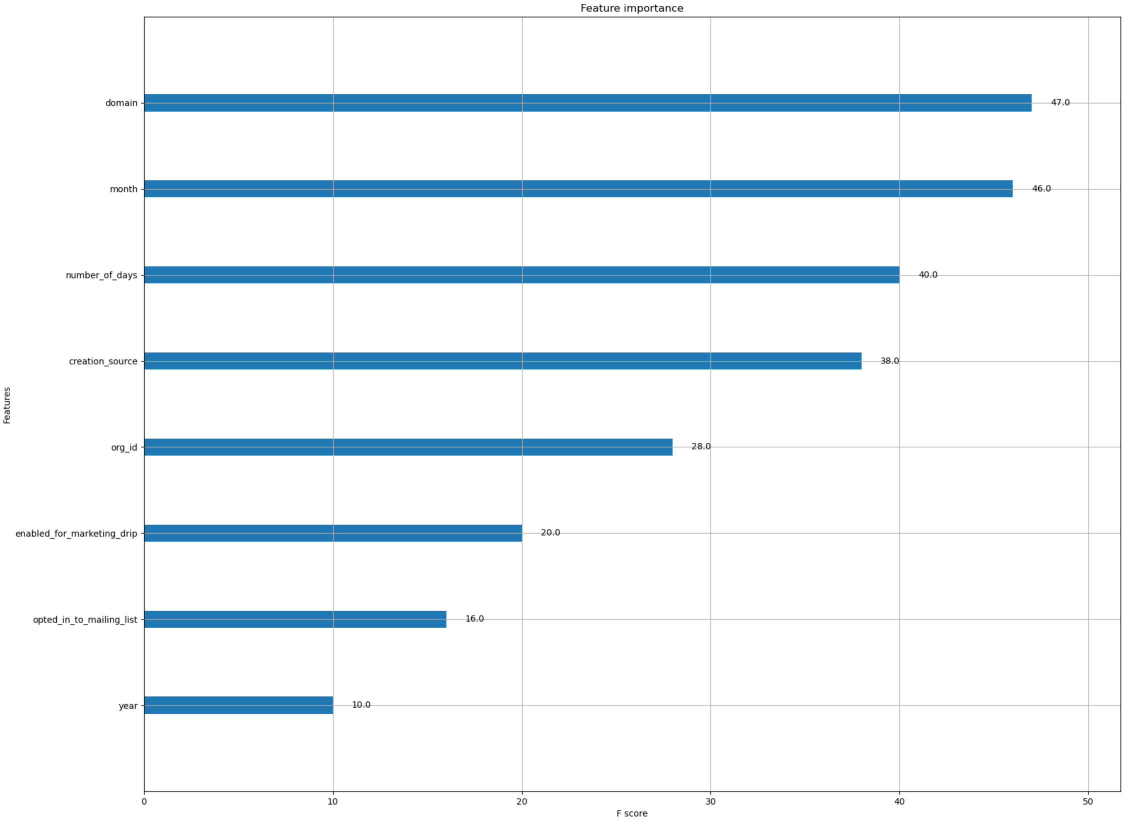
Feature Engineering:Several features were engineered to enhance predictive power:
- A categorical column representing the top 10 domains derived from email addresses.
- A column indicating the length (in days) from the last login date to the account creation date.
- A column categorizing users into the top 10 organizations, with the remaining organizations grouped as 'other.'
- The target variable, indicating user 'adoption' if active for at least 3 days in a 7-day window.

Modeling:
To address the imbalanced dataset (with only ~10% adopted users), undersampling was applied. The XGBoost Classifier was selected, and hyperparameter tuning was conducted using GridSearchCV. Evaluation metrics included the ROC AUC score and the ROC curve.The best model achieved a ROC AUC score of 0.9741. The corresponding ROC curve is depicted below.



Key Features: Top three features identified by the best model as indicative of user adoption are:
1. Domain name derived from the user's email.
2. Month of user signup.
3. Number of days from first signup to the last active day.

Future steps involve incorporating additional features such as assessing the impact of user referrals and understanding the significance of account creation during specific months.


Feature importance
domain
month
number_of_days
creation_source
org_id
enabled_for_marketing_drip
opted_in_to_mailing_list
year
47.0
46.0
40.0
38.0
28.0
20.0
16.0
10.0
Features
0
10
20
30
40
50
F score