Predicting User Adoption - Relax Challenge

Problem

Predicting user adoption with 2 years worth of data (data format type is csv spread across two files).

Approach

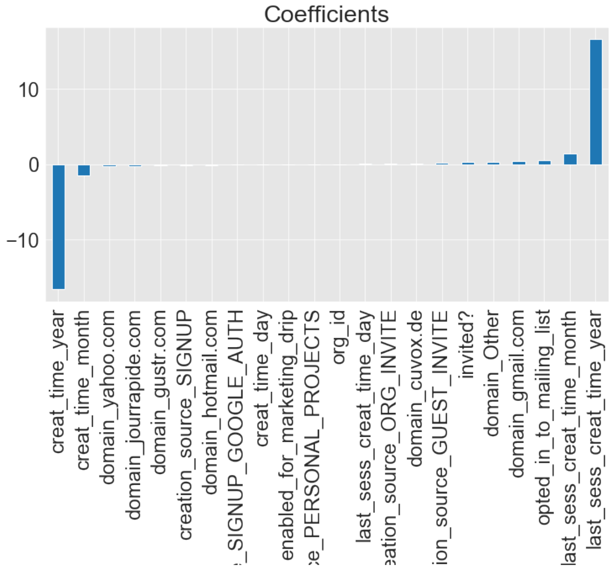
Define User Adoption:
A user who has logged into the product on three separate days in at least one seven day period. A column called "adopted?" will be created as the target variable for machine learning.

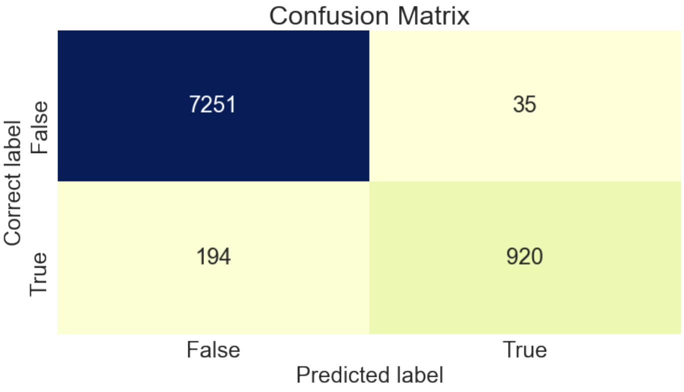
Predictive Model:
Transform categorial variables into dummy variables for machine learning; standardize all continuous columns; several model were checked such as Logistic Regression and Random Forest; coefficients, permutation importance, and shap plots were used to examine feature importance.

Model validity:
Model was split into a train and hold-out set. The train set was grid-searched and cross-validated (n_folds = 5) and optimized using f1 because we were dealing with imbalanced data - f1 does not use true negative counts in its calculations which is perfect for imbalanced data where true negatives are the majority. The cross-validated f1 score on the training set was 88%; The accuracy score was 97% (the baseline for accuracy was 86%, i.e., this model performed 10% better than chance).

Result

The f1 score was 89%. The accuracy on the hold-out set was 97%. Factors that greatly predict user adoption with high confidence: the year/month for creation_time and last_session_creation_time. Creation_time is defined as when account was created. Last_session_creation_time is the last login of the user. When year and month were more recent for "last_session_creation_time", the more likely the user was to adopt. In reverse, when year and month were more recent for "creation_time", the less likely the user was to adopt.


Coefficients
creat_time_year
creat_time_month
domain_yahoo.com
domain_jourrapide.com
domain_gustr.com
creation_source_SIGNUP
domain_hotmail.com
_SIGNUP_GOOGLE_AUTH
creat_time_day
enabled_for_marketing_drip
ce_PERSONAL_PROJECTS
org_id
last_sess_creat_time_day
eation_source_ORG_INVITE
domain_cuvox.de
ion_source_GUEST_INVITE
invited?
domain_Other
domain_gmail.com
opted_in_to_mailing_list
last_sess_creat_time_month
last_sess_creat_time_year


Confusion Matrix
False
True
Correct label
False
True
Predicted label