Relax Challenge – Connor McAnuff

Purpose: Defining an "adopted user" as a user who has logged into the product on three separate days in at least one 7-day period, identify which factors predict future user adoption.

1.1 Adopted users

Using the login information provided in the 'takehome_user_engagement.csv' it was determined that 1,602 of 12,000 (13.35%) are considered 'adopted' according to the criteria stated in the above purpose statement.

1.2 Feature Selection

The raw data has been engineered into features using the following process:

Feature	Description
creation_time	- Strings converted to datetime objects. - Datatime objects converted to seconds since the first user was created.
last_session_creation_time	- Converted to =1 if user has a last session, =0 if not.
opted_in_to_mailing_list	- Not modified. - =1 if user has opted in, else 0.
enabled_for_marketing_drip	- Not modified. - =1 if user enabled, else 0.
org_id	- Not modified. - Integer representing organization ID.
invited_by_user_id	- Converted to =1 if user was invited by another user, =0 if not.
creation_source	- Modified to be 5 features, =1 if user was created through a given source, else =0.

1.3 Feature Predictive Strength

A random forest model was fit to the X, y data. The feature importances, as given by sklearn, are plotted in order of importance:
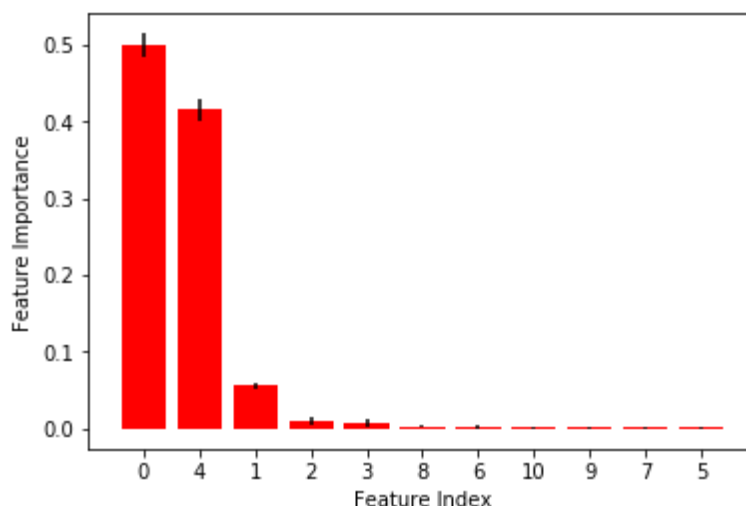
The features with predictive strength are creation_time, org_id, and to a lesser extent, last_session_creation_time. It is possible that creation_time is a strong predictor because a over time, a user is more likely to eventually login 3 times in a 7-day period.