

Relax, Inc. Take Home Challenge

Relax Inc. wanted to see which features helped created an adopted user. An adopted user is defined by a user who has logged into the product on three separate days in at least one seven-day period. I have solved the problem through the following steps:

1. Identify Adopted Users based on 'takehome_user_engagement.csv' table
 - a. Change dates to proper datetime format
 - b. Search the table for each user over 7 days
 - c. Return the users' ID of adopted users if the user engaged atleast 3 times
2. Label the adopted users on the 'takehome_users.csv' table
 - a. Search the User's ID from the adopted users ID list generated from step [1]
 - b. Label adopted users as 1 and 0
 - i. I am only interested if the user has adopted the product and not the level of adoption, therefore I have summarized the label to a binary classifier
3. Clean and Normalize the Data for Machine Learning
 - a. Drop the user's name, email, and org ID because those feature do not affect user adoption
 - b. Normalize creation_time and last_session_creation_time because the model needs an integer and I want to do want the large numbers to affect the outcome
 - c. One Hot and Encode the creation_source variable because the source could be an important factor to why they signed up and how they will use the product
4. Use Random Forest to create a feature importance table

	column_names	feat_importance
1	last_session_creation_time	0.700287
2	creation_time	0.284785
3	creation_source_PERSONAL_PROJECTS	0.003371
4	enabled_for_marketing_drip	0.00308
5	opted_in_to_mailing_list	0.002622
6	creation_source_SIGNUP_GOOGLE_AUTH	0.001517
7	creation_source_GUEST_INVITE	0.001266
8	creation_source_ORG_INVITE	0.001253
9	invited_by_user_id	0.000999
10	creation_source_SIGNUP	0.000821
11	creation_source_nan	0

- a.
5. Test the data to create a AUC Score to understand the predictive capabilities of the model
 - a. AUC Score: 0.9175

Analysis:

The current engagement of the user (creation_time and last_session_creation_time) seems to have the greatest impact. Those two variables are obvious because if a user continues to log in, then the user will become an adopted user. The other variable that stood out to me is if the user created the account from a "personal project" connection. This could provide deeper insight to the type of user and why the user has decided to use this product. On the contrary, signing up from the website or "SIGNUP" seems to have the lowest affect in creating an adopted user. I would suggest the client to look into that further. Overall, the data provided can strongly help deduce what makes an adopted user.