Relax Data Challenge

This dataset contains data from 12,000 users over a two year period. The goal is to determine which factors might be useful for predicting user adoption. An "adopted user" is defined as one who has logged in on three separate days in at least one seven-day period.

I created a column with the number of logins per seven-day period, then used this to create another column indicating whether the number of logins was at least 3. I converted all the columns to numerical values so that they could be used in the model. Since the dataset contained some categorical features, I used the LightGBM library to create a model.

Then I calculated the shap values, which tell us how much impact each feature has on the model. The first plot below shows how much each feature is affecting the model, either in the positive or negative direction. The second plot shows just the magnitude of the impact of each feature. We can see that the most useful features for predicting user adoption are the user id, the time when they created their account, time of last login, the user who invited them to join, and the organization they belong to.