Relax Inc. Challenge Report

I was provided with two tables. The first was a user table with data on 12,000 users who signed up for the product in the last two years, and the second was a usage summary table with a row for each day that a user logged into the product. The goal was to determine which factors predict user adoption, where an adopted user is defined as a user who has logged into the product on three separate days in at least one seven-day period.

The first step was to determine the adopted users. This was accomplished by resampling the usage summary table weekly and summing the number of visits per week. It was determined that there were 1445 adopted users.

The next step was to perform some feature engineering. I converted the invited\_by\_user\_id column to a Boolean, where 0 represented a user that was not invited by another user and 1 represented a user that was invited by another user. Then, I created a feature called usage\_length, which is the time (number of days) between the last session and the creation. I eliminated the irrelevant columns and converted the categorical column creation\_source to dummy columns.

In the modeling step, I applied the Random Forest Classifier to the data. It achieved a 98.9% accuracy for the training set and a 97.5% accuracy for the test set. Next, I determined the feature importance from the classifier. As you can see from below, the usage\_length is by far the most important feature. The invited\_by\_user is the least important feature. The best course of action to encourage user adoption would be to target the users very soon after they first use the product.

Text

Description automatically generated