

Relax Data Science Challenge Take Home

The factor that had the greatest predictive power in determining whether a user would become an adopted user was “org_id” and “creation_source,” accounting for approximately 90% of feature importance. However, I could not produce a model that did much better than random guessing on determining user adoption. I tried Random Forest, Logistic Regression, and Gradient Machine Boosting. Random Forest was the one that performed the best after performing a Grid Search to tune the model. I eliminated features that were not relevant or showed no promise for improving the model until all that remained was “creation_source,” “opted_in_to_mailing_list,” “enabled_for_marketing_drip,” and “org_id” (I turned org_id from a numeric column to an object column). The accuracy score for the model was 63.46% but the AUR-ROC was only .528.

