Originally, I was very perplexed by the task at hand. Having previously completed a challenge from one of the leading hedge funds—Jane Street; I decided to put what I had learned to use. Their challenge had a dataset that spanned 5GB. This alone was a challenge in it of itself. Through the time I spent preparing for this challenge with over 100 features. I began to learn a lot about different models. Namely, boosting or more specifically extreme gradient boosting. I would consider this trees or random forests on steroids. After, attempting logistic regression and a neural network I decided it might be time, again, to throw the kitchen sink at this problem. As it turns out, XGBoost proved to be decisive in maximizing the f1 score. This comes after many other failed, mediocre attempts. I began by converting all categorical data into label-encoded data. This is data that was not numerical and had some sort of non-ordinal categorization. I figured some of the timestamps could be useful, especially if broken up by the hour. Obviously, you wouldn’t want a transaction to go through after the password and or phone number was changed at 4 am, regardless of your age or how long you have been a client at the bank. However, the default parameters of XGBoost were so efficient I did not have to introduce any new features nor rely on some. Hence, I dropped, 'ACTN\_CD', 'ACTN\_INTNL\_TXT', 'TRAN\_TYPE\_CD', 'CUST\_SINCE\_DT', and from the test set 'dataset\_id'—for obvious reasons. Following this, all null values and unknown values were converted into a useful integer for the algorithm. To my surprise, after converting all of the categorical data including dates to labeled data it achieved a nearly perfect f1 and accuracy score. This is after using all but one default parameter for the classifier. I was hoping to follow up my research using Hyperopt(a python library) which uses Bayesian Inference to optimize the hyperparameters of XGBoost. To my surprise this was not necessary, I’m very excited about my submission.

Thank you,

**Githendu Mukiri**

*BS Computer Science | Class of 2022*

*Wentworth Institute of Technology*

*mukirig@wit.edu*