There are six files in the code folder. Four of them are python code and the other two are dataset. The order to run the code is data\_visualization, feature\_engineering, feature\_selection, and Model and Accuracy.

data\_visualization contains codes on manipulating the data and generating some graphs, in order to find some patterns behind the data.

feature\_engineering contains codes on data cleaning and feature engineering, in order to delete some unnecessary features and create some insightful features. We combined train and test dataset at first. Then, a dataset named “combined\_train\_test.csv” was generated containing all the new features.

feature\_selection contains codes on feature selection. Then, a csv file named “feature\_selected.csv” was created, which contains 200 the most important features and one target.

Model and Accuracy contains codes on building model and calculate the accuracy for the model. We calculate the confusion matrix, precision score, recall score, f1 score, and AUC to test the model.