Project Update 2

As of 12/01/2020, the project is still on schedule.

Degree 2 and degree 3 polynomial classification was explored. With 10 folds cross-validation, the result is almost the same as degree 1 polynomial classification.

The second classification algorithm is the K-nearest neighbor algorithm. With k in the range from 1 to 26, it is shown that the best prediction result happens when k is equal to 17,18, and 25.

After applying 10-fold cross validation, the accuracy for k = 17 is 0.506, the accuracy for k = 18 is 0.498, and for k = 25, the accuracy is 0.518.

Based on the current observation, both linear classification and K-nearest neighbor are not doing as expected. With the KNN method, I also realized that some classes have a really low recall score and support. After doing some research, it is suggested that a really low recall score means the KNN did not predict any values inside those classes correctly.

The next step is to classify this dataset with a neural network, it will be completed by 12/06/2020