Name : Dhrumil Shah SJSU ID : 011428859
F-1 Score : 0.8571 Rank : 4

For this assignment, I have initially tried the dimensionality reduction methods: SparsePCA, PCA and truncatedSVD. For the best results, I chose truncatedSVD as it works well with Sparse Data having high Dimensions. For the classifier algorithm, I tried various algorithms like : Decision Trees, Random Forest, AdaBoost Classifier and Gradient Booster. The one which was best for me was the Decision Tree Classifier. I also gave more weights to the minority class and less weight to the majority class to balance the weights in the imbalanced dataset. For the dimensionality Reduction, I reduced the dataset to 350 features and transformed the matrix accordingly for Fitting the data. After applying the dimensionality Reduction, I applied SMOTE afterwards for Over-Sampling the Minority Class. The SMOTE's parameters used were "svm" for the "kind' and also the k-neighbors and m-neighbors value was 9. I also tried other parameters and same parameters with different values. For the others , in the Random Forest I kept the number of estimators to be 58. Which is the square root of number of the features. Similarly, I used the AdaBoost Classifier and Gradient Booster Classifier. The maximum F-1 Score I got was through the weighted Decision Trees. For the features too, I tried various dimensions from 100 to 1000. The best F-1 Score I received was when the dimensions were reduced to 350. For calculating the F-1 Score,

I performed Cross Validation for every algorithm and calculated the Precision and Recall

for every dataset. For testing purposes, I divide training data set into training and testing data set and calculated the F-1 Score for the testing dataset in the training Dataset. For the Decision Trees too, I tried it without SMOTE and with SMOTE with Different parameters. Moreover, I also changed different parameters of Decision Trees and the weighted Decision Tree worked best for me by providing more weights to minority classes.