Train/Test Report

InsureIntel
Group 40

Training Objective:

Binary Classification

Multi-Model Approach

Clustering - KMeans algorithm is used to create clusters in the preprocessed data. The optimum number of clusters is selected by plotting the elbow plot, and for the dynamic selection of the number of clusters, we are using "KneeLocator" function. The idea behind clustering is to implement different algorithms

The Kmeans model is trained over preprocessed data, and the model is saved for further use in prediction.

Model Selection – After the clusters have been created, we find the best model for each cluster. We are using two algorithms, "SVM" and "XGBoost". For each cluster, both the algorithms are passed with the best parameters derived from GridSearch. We calculate the AUC scores for both models and select the model with the best score. Similarly, the model is selected for each cluster. All the models for every cluster are saved for use in prediction.

SVM

SVM Score :
0.7533333333333333


```
              precision    recall  f1-score   support

           0       0.75      1.00      0.86       113
           1       0.00      0.00      0.00        37

    accuracy                           0.75       150
   macro avg       0.38      0.50      0.43       150
weighted avg       0.57      0.75      0.65       150

Cross-Validated Accuracy: 0.75
```
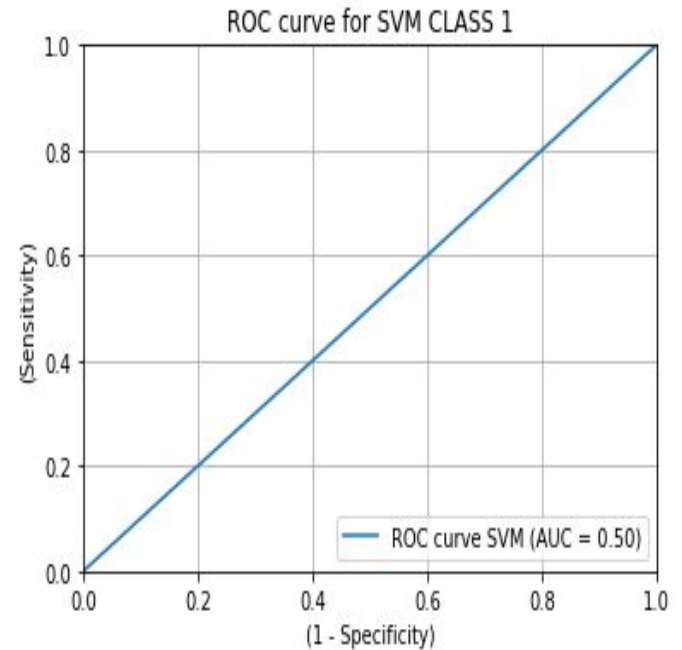
Random Forest

Random Forest Score :
0.7466666666666667

```
              precision    recall  f1-score   support

           0       0.79      0.90      0.84       113
           1       0.48      0.27      0.34        37

    accuracy                           0.75       150
   macro avg       0.63      0.59      0.59       150
weighted avg       0.71      0.75      0.72       150

Cross-Validated Accuracy: 0.78
```
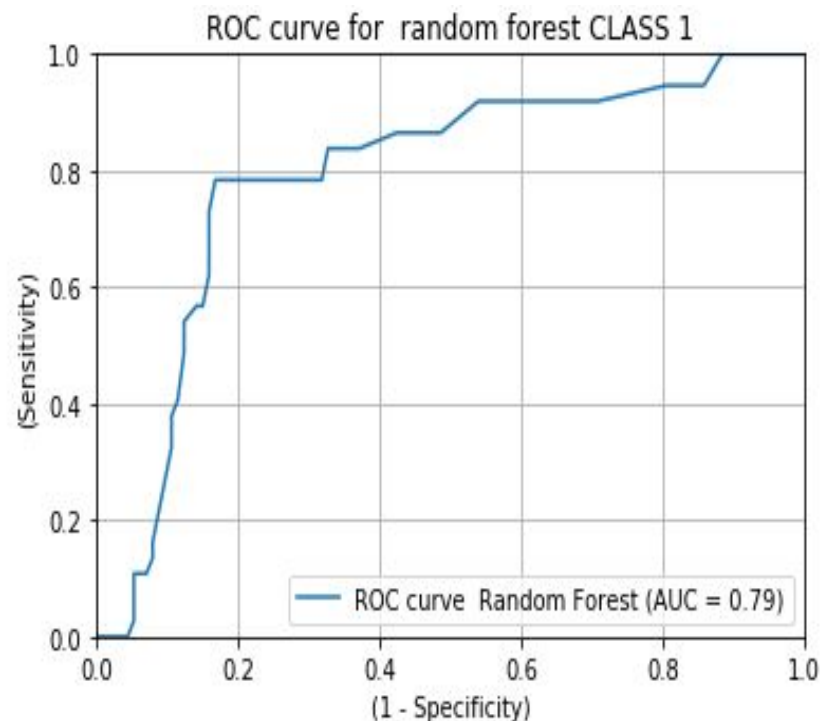

ROC curve for random forest CLASS 1

XGBOOST

XGBClassifier Score :
0.8266666666666667

ROC curve for XGBOOST CLASS 1

	precision	recall	f1-score	support
0	0.88	0.89	0.89	113
1	0.66	0.62	0.64	37
accuracy			0.83	150
macro avg	0.77	0.76	0.76	150
weighted avg	0.82	0.83	0.83	150

Cross-Validated Accuracy: 0.82