Project Development Phase Model Performance Test

Date	9th November 2023
Team ID	593197
Project Name	PayGuard Plus – An Online Payments Fraud Detector
Maximum Marks	10 Marks
Team Size	3
Member 1 – Team Lead	Akshit Bahl (21BIT0012)
Member 2	Ananya Priya (21BIT0245)
Member 3	Lakshya Mittal (21BIT0076)

Model Performance Testing:

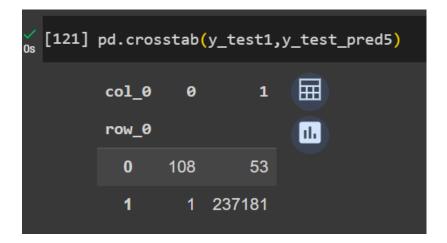
Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Values	Screenshot
1.	Metrics	Classification Model: Confusion Matrix - , Accuray Score- & Classification Report -	
2.	Tune the Model	Hyperparameter Tuning - Validation Method -	

Model chosen = **XGBoost Classifier**

• METRICS

1. Confusion Matrix:



2. Classification Report:

v 0s	<pre>print(classification_report(y_test1,y_test_pred5))</pre>								
	∃			precision	recall	f1-score	support		
			0	0.99	0.67	0.80	161		
			1	1.00	1.00	1.00	237182		
		accur	racy			1.00	237343		
		macro	avg	1.00	0.84	0.90	237343		
		weighted	avg	1.00	1.00	1.00	237343		

3. Accuracy Score:

Testing accuracy: 99.977%
Training accuracy: 99.996%

Tuning the Model:

col_0 0 1

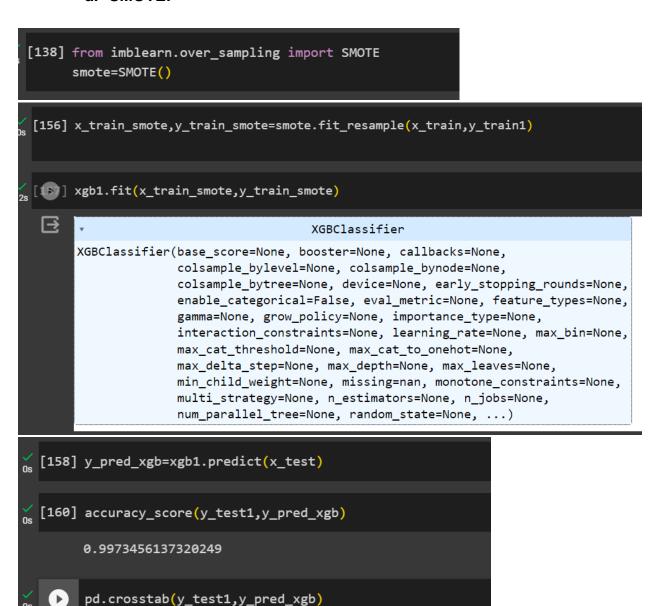
132

601 236581

row_0

0

a. SMOTE:



PayGuard Plus 10.11.23

ılı

29

```
print(classification_report(y_test1,y_pred_xgb))
  \blacksquare
                  precision recall f1-score
                                              support
                      0.18 0.82
                0
                                        0.30
                                                 161
                1
                      1.00
                              1.00
                                       1.00
                                             237182
                                       1.00 237343
          accuracy
                      0.59
                             0.91
                                       0.65
                                              237343
         macro avg
                                       1.00
                                               237343
      weighted avg
                      1.00
                               1.00
```

So the recall has increased with SMOTE technique but the precision and overall accuracy has been affected and has gone down.

Now, we will do hyper-parameter tuning by grid-search CV

b. <u>HYPER_PARAMETER TUNING:</u>

By Grid-search CV:

```
[ ] from sklearn.model_selection import GridSearchCV

[ ] parameters = {
         'n_estimators': [50, 100, 200],
         'max_depth': [None, 10, 20, 30],
         'min_samples_split': [2, 5, 10],
         'min_samples_leaf': [1, 2, 4]
      }

[ ] clf=GridSearchCV(rfc,param_grid=parameters,verbose=2)
```

```
[40] clf.fit(x_train,y_train)
      Fitting 5 folds for each of 8 candidates, totalling 40 fits
      [CV] END max depth=None, min samples leaf=1, min samples split=2, n estimators=50; total time= 15.8s
      [CV] END max depth=None, min samples leaf=1, min samples split=2, n estimators=50; total time= 15.3s
      [CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=50; total time= 16.7s
      [CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=50; total time= 15.9s [CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=50; total time= 14.4s
      [CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 31.6s
      [CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 33.2s
      [CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 34.2s
[CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 30.7s
      [CV] END max_depth=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 29.1s
      [CV] END max depth=None, min samples leaf=2, min samples split=2, n estimators=50; total time= 15.9s
      [CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 16.4s
      [CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 16.2s [CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 15.6s
      [CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 14.8s
      [CV] END max depth=None, min samples leaf=2, min samples split=2, n estimators=100; total time= 31.3s
      [CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=100; total time= 30.4s
[CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=100; total time= 35.1s
      [CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=100; total time= 31.8s
      [CV] END max_depth=None, min_samples_leaf=2, min_samples_split=2, n_estimators=100; total time= 28.8s
      [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=50; total time= 14.4s
      [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=50; total time= 14.1s
      [CV] END max depth=10, min samples leaf=1, min samples split=2, n estimators=50; total time=
      [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=50; total time= 14.9s
      [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=50; total time= 14.3s
      [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 29.5s
      [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 28.9s [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 30.1s
      [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 28.8s
      [CV] END max_depth=10, min_samples_leaf=1, min_samples_split=2, n_estimators=100; total time= 28.3s
      [CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 14.8s
      [CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 14.5s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 15.0s
      [CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 14.8s
      [CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=50; total time= 14.1s
      [CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=100; total time= 28.3s
[CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=100; total time= 29.4s
      [CV] END max_depth=10, min_samples_leaf=2, min_samples_split=2, n_estimators=100; total time= 29.0s
    [CV] END max depth=10, min samples leaf=2, min samples split=2, n estimators=100; total time= 26.5s
                    GridSearchCV
      estimator: RandomForestClassifier
             ▶ RandomForestClassifier
```

