Model Phase 1 : Experimentation Details

Vectorizer used : TFIDF vectorizer
Encoder : label encoder
Num of classes : 5
Method Used : One vs Rest Classifier
Y-split : stratified split
Models experimented with : Ensemble models (Bagging and Boosting models)
Bagging : Decision Tree Classifier ,Random Forest Classifier
Boosting : XGBoost Classifier
Hyperparameter Tuning : Used both Grid Search and Random Search
Grid Search for Bagging and Random Search for Boosting

RANDOM FOREST : (RESULTS)

----------RANDOM FOREST-----------
F1 score is 0.9362554297020023
-----------------------TEST SET REPORT FOR RANDOM FOREST -------------------------------------
      precision   recall f1-score support

	precision	recall	f1-score	support
0	0.91	0.79	0.85	38
1	0.89	0.94	0.92	18
2	0.89	0.93	0.91	27
3	0.91	0.94	0.92	97
4	0.98	0.98	0.98	121
accuracy			0.94	301
macro avg	0.92	0.92	0.92	301
weighted avg	0.94	0.94	0.94	301

-----------------------TRAIN SET REPORT FOR RANDOM FOREST -------------------------------
precision   recall f1-score support

	precision	recall	f1-score	support
0	0.94	0.92	0.93	152
1	1.00	0.89	0.94	71
2	1.00	0.92	0.96	108
3	0.94	0.98	0.96	388
4	0.99	1.00	0.99	485
accuracy			0.97	1204
macro avg	0.97	0.94	0.96	1204
weighted avg	0.97	0.97	0.97	1204

Confusion Matrix for Random Forest :

XGB Classifier
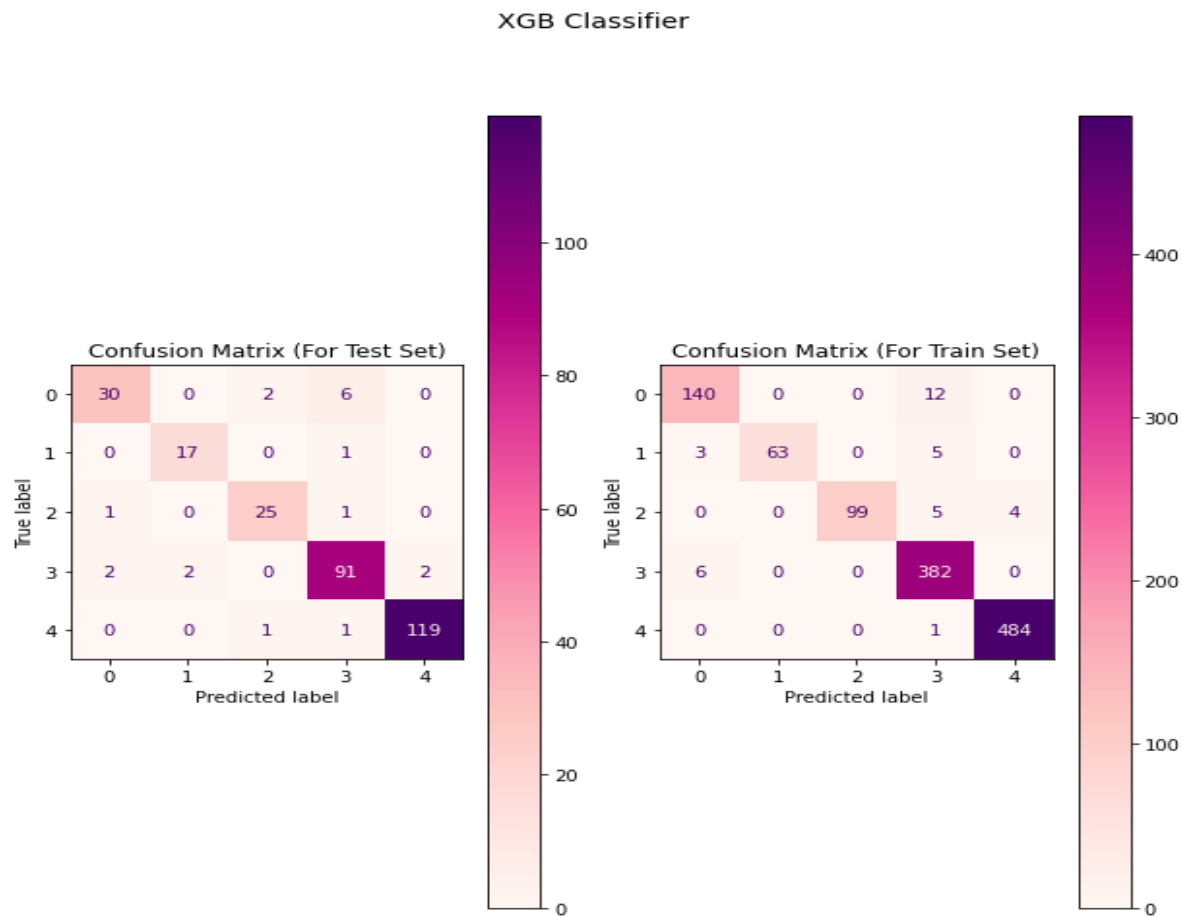
<u>DECISION TREE(RESULTS):</u>

----------DECISION TREE-----------
F1 score is 0.9279739802897904
-----------------------TEST SET REPORT FOR DECISION TREE --
          precision   recall f1-score   support

    Banking   0.76   0.84   0.80   38
    Jobs-IT   0.94   0.83   0.88   18
Rent-Apartment   0.90   0.96   0.93   27
     Retail   0.92   0.92   0.92   97
  Sell-House   1.00   0.97   0.98   121

   accuracy   0.93   301

	precision	recall	f1-score	support
macro avg	0.90	0.90	0.90	301
weighted avg	0.93	0.93	0.93	301

-----------------------TRAIN SET REPORT FOR DECISION TREE---

	precision	recall	f1-score	support
Banking	0.89	0.92	0.90	152
Jobs-IT	0.99	0.93	0.96	71
Rent-Apartment	0.97	0.90	0.93	108
Retail	0.95	0.97	0.96	388
Sell-House	0.99	0.99	0.99	485
accuracy			0.96	1204
macro avg	0.96	0.94	0.95	1204
weighted avg	0.96	0.96	0.96	1204

Confusion Matrix for Decision Trees :

Decision Tree Classifier
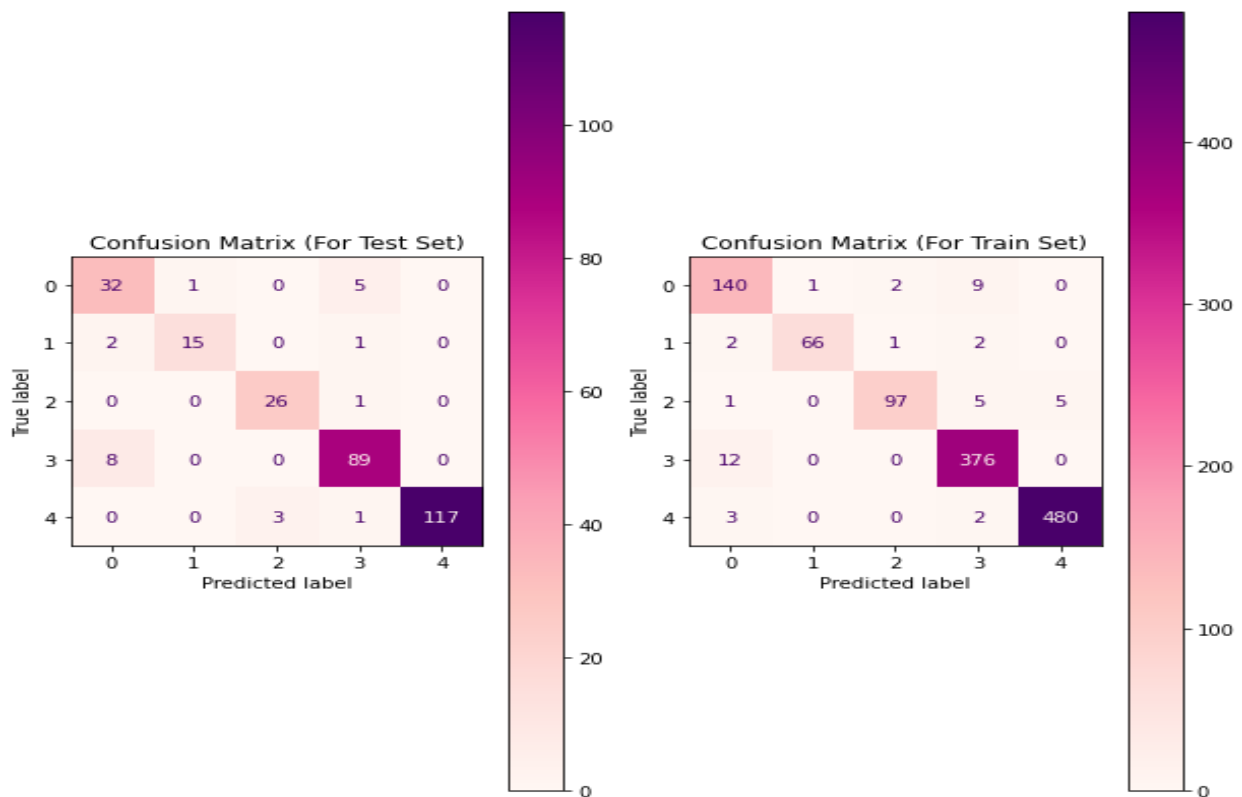
XGBOOST CLASSIFIER (RESULTS) :

------------------XGBOOST---------------------
F1 score is 0.9471365290302168
-----------------------TEST SET REPORT FOR XGBOOST--
              precision   recall f1-score   support

     Banking   0.85   0.87   0.86   38
     Jobs-IT   1.00   1.00   1.00   18
Rent-Apartment   0.93   0.93   0.93   27
      Retail   0.93   0.94   0.93   97
  Sell-House   0.99   0.98   0.98   121

    accuracy   0.95   301
   macro avg   0.94   0.94   0.94   301
weighted avg   0.95   0.95   0.95   301

-----------------------TRAIN SET REPORT FOR XGBOOST--
              precision   recall f1-score   support

     Banking   0.95   0.95   0.95   152
     Jobs-IT   1.00   0.90   0.95   71
Rent-Apartment   0.98   0.92   0.95   108
      Retail   0.96   0.99   0.97   388
  Sell-House   0.99   0.99   0.99   485

    accuracy   0.97   1204
   macro avg   0.98   0.95   0.96   1204
weighted avg   0.97   0.97   0.97   1204

Confusion Matrix for XGBOOST CLASSIFIER:

XGB Classifier

Confusion Matrix (For Test Set) / Confusion Matrix (For Train Set)
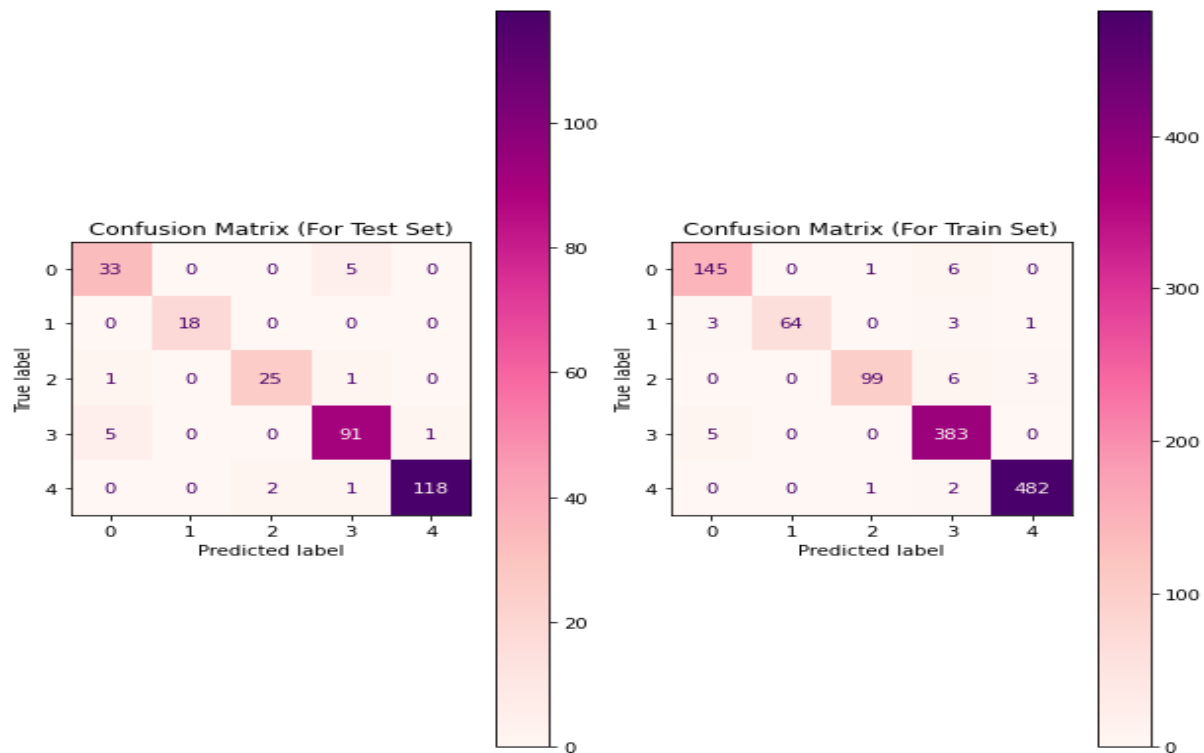
CONCLUSION : From the experiments performed above we have decided to move on with the xgboost classifier for the first phase of our model

Model Phase 2 :

Random Forest :

```
----------RANDOM FOREST-----------
F1 score is  0.8916329200243155
-----------------------TEST SET REPORT FOR RANDOM
FOREST---------------------------------------------
             precision    recall  f1-score   support

    Banking       0.77      0.63      0.70        38
     Jobs-IT       1.00      0.79      0.88        52
Rent-Apartment     0.84      0.97      0.90       146
     Retail       0.93      0.85      0.89        65
  Sell-House       0.95      0.95      0.95       122

   accuracy                           0.89       423
```

	precision	recall	f1-score	support
macro avg	0.90	0.84	0.86	423
weighted avg	0.90	0.89	0.89	423

------------------------TRAIN SET REPORT FOR RANDOM FOREST
----

	precision	recall	f1-score	support
Banking	0.94	0.77	0.85	153
Jobs-IT	0.98	0.86	0.92	207
Rent-Apartment	0.90	0.99	0.94	583
Retail	0.98	0.91	0.94	262
Sell-House	0.96	0.98	0.97	485
accuracy			0.94	1690
macro avg	0.95	0.90	0.92	1690
weighted avg	0.94	0.94	0.94	1690

XGB Classifier

Confusion Matrix (For Test Set)

	0	1	2	3	4
0	24	0	14	0	0
1	1	41	10	0	0
2	4	0	142	0	0
3	2	0	2	55	6
4	0	0	2	4	116

True label / Predicted label

Confusion Matrix (For Train Set)

	0	1	2	3	4
0	118	2	33	0	0
1	3	179	24	1	0
2	4	2	577	0	0
3	0	0	4	239	19
4	0	0	5	5	475

True label / Predicted label
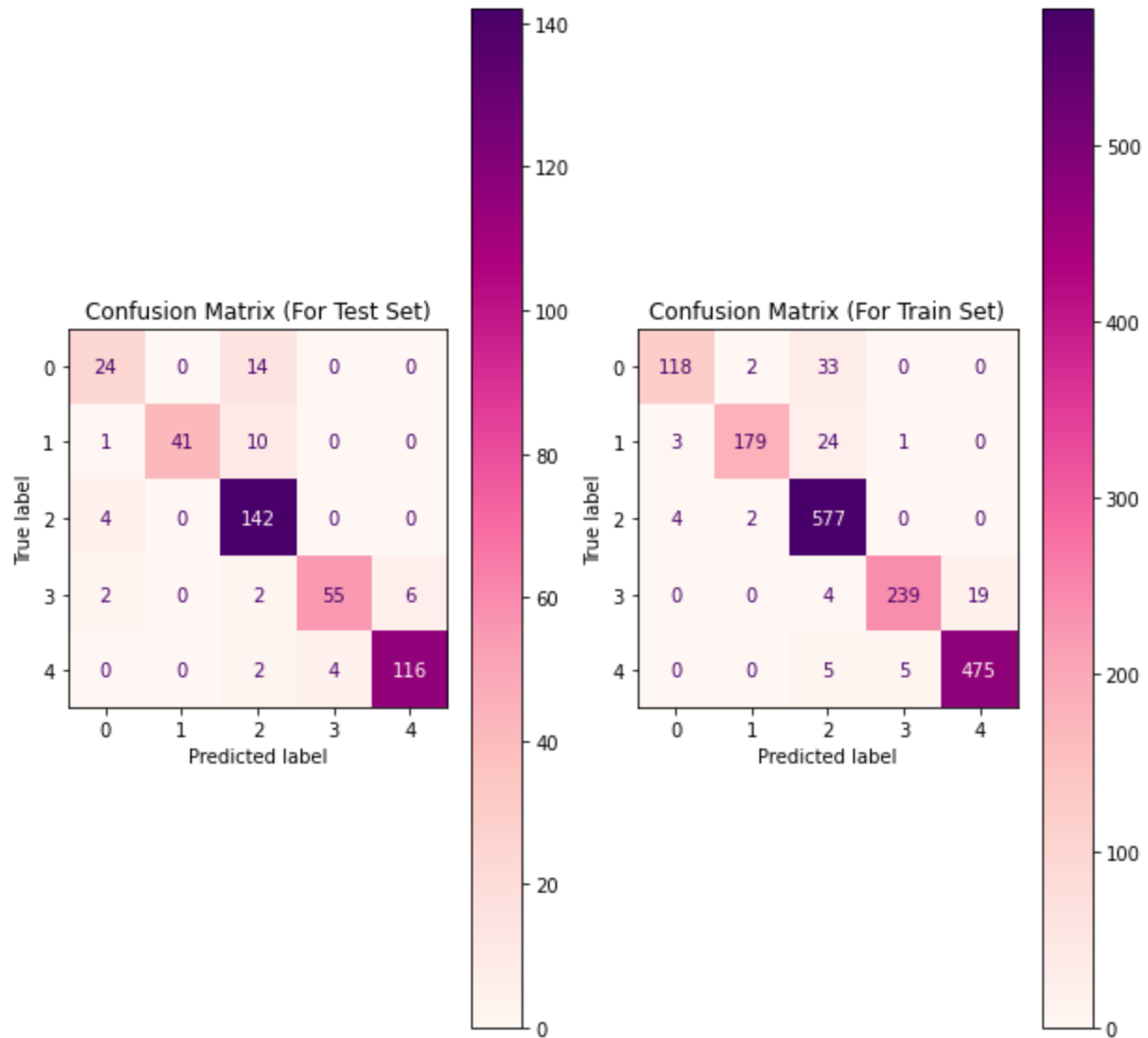
Decision Tree :

----------DECISION TREE-----------
F1 score is 0.8461335170352604
-----------------------TEST SET REPORT FOR DECISION TREE --
                precision   recall f1-score   support

     Banking   0.72   0.61   0.66   38
     Jobs-IT   0.87   0.77   0.82   52

	precision	recall	f1-score	support
Rent-Apartment	0.83	0.92	0.87	146
Retail	0.82	0.75	0.78	65
Sell-House	0.91	0.93	0.92	122
accuracy			0.85	423
macro avg	0.83	0.79	0.81	423
weighted avg	0.85	0.85	0.85	423

------------------------TRAIN SET REPORT FOR DECISION TREE--

	precision	recall	f1-score	support
Banking	0.90	0.74	0.81	153
Jobs-IT	0.96	0.88	0.92	207
Rent-Apartment	0.90	0.98	0.94	583
Retail	0.94	0.88	0.91	262
Sell-House	0.95	0.96	0.95	485
accuracy			0.93	1690
macro avg	0.93	0.89	0.91	1690
weighted avg	0.93	0.93	0.92	1690

Decision Tree Classifier

Confusion Matrix (For Test Set)

	0	1	2	3	4
0	23	1	13	1	0
1	3	40	9	0	0
2	4	5	134	3	0
3	2	0	3	49	11
4	0	0	2	7	113

Confusion Matrix (For Train Set)

	0	1	2	3	4
0	113	6	34	0	0
1	4	182	20	1	0
2	7	2	573	1	0
3	1	0	4	230	27
4	0	0	6	12	467
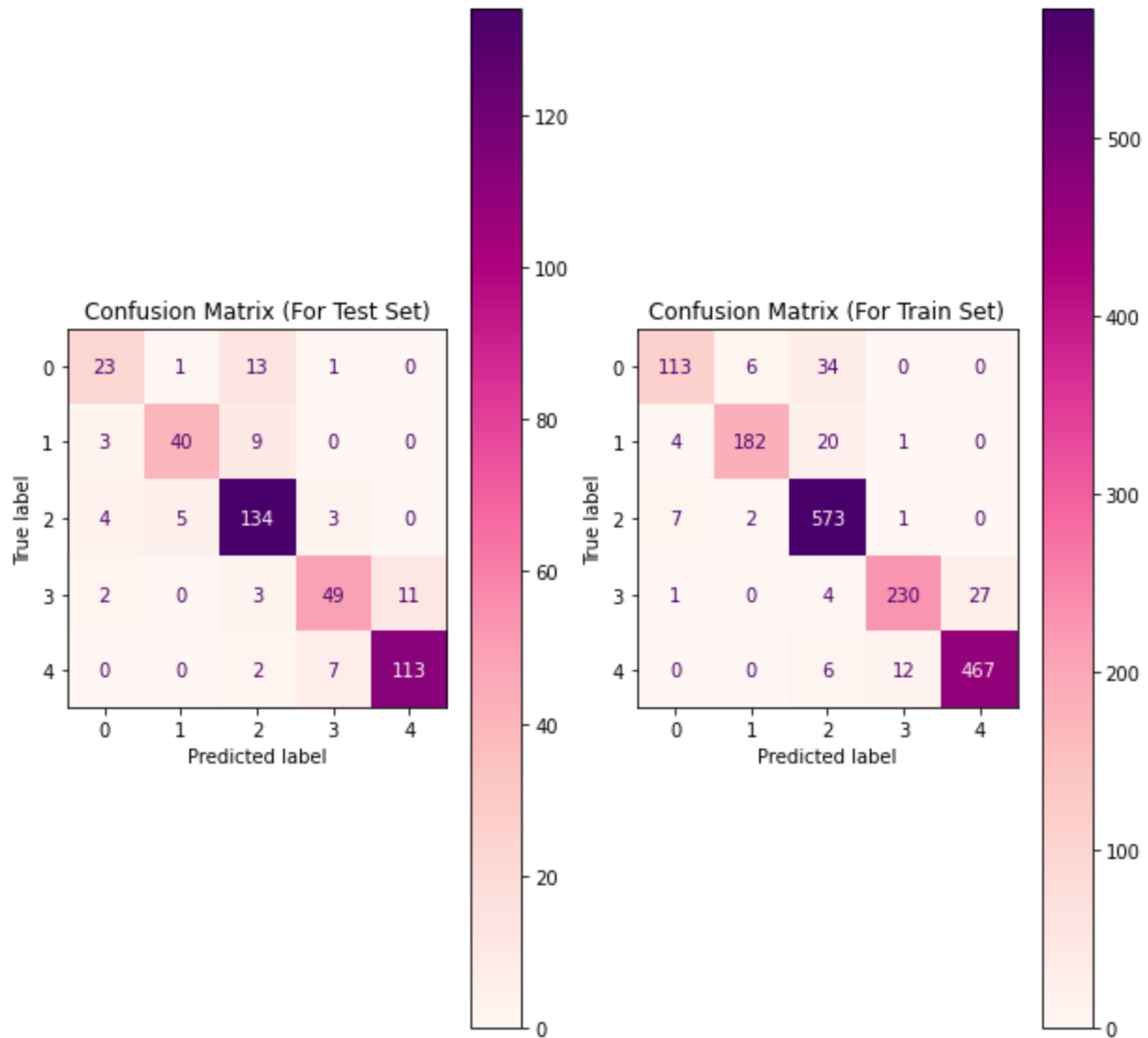
XGBoost :

-----------------XGBOOST---------------------
F1 score is 0.907725417603805
-----------------------TEST SET REPORT FOR XGBOOST--
            precision   recall f1-score   support

    Banking   0.78   0.82   0.79   38
    Jobs-IT   0.96   0.88   0.92   52
Rent-Apartment   0.90   0.94   0.92   146

```
        Retail     0.93    0.83    0.88      65
      Sell-House    0.94    0.95    0.94      122

      accuracy                     0.91      423
     macro avg      0.90    0.88    0.89      423
   weighted avg     0.91    0.91    0.91      423

----------------------TRAIN SET REPORT FOR XGBOOST------------------------------------------
           precision   recall  f1-score   support

      Banking      0.95    0.97    0.96      153
      Jobs-IT      0.99    0.96    0.98      207
   Rent-Apartment     0.96    0.99    0.98      583
        Retail     0.98    0.91    0.94      262
      Sell-House    0.97    0.98    0.97      485

      accuracy                     0.97      1690
     macro avg      0.97    0.96    0.97      1690
   weighted avg     0.97    0.97    0.97      1690
```

XGB Classifier

Confusion Matrix (For Test Set)

Confusion Matrix (For Train Set)