

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 -----

	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

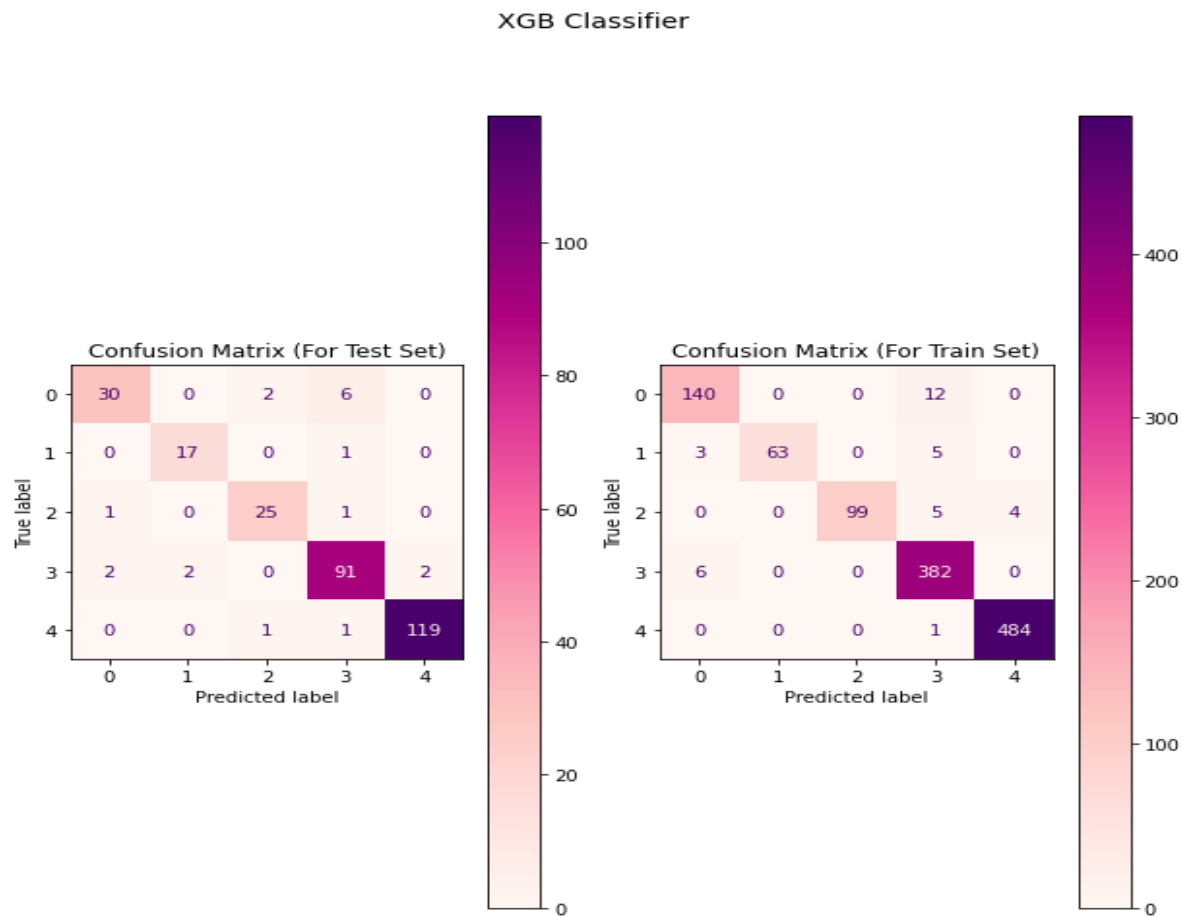
-----TRAIN SET REPORT FOR RANDOM FOREST -----

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

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 :



DECISION TREE(RESULTS):

-----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

macro avg	0.90	0.90	0.90	301
weighted avg	0.93	0.93	0.93	301

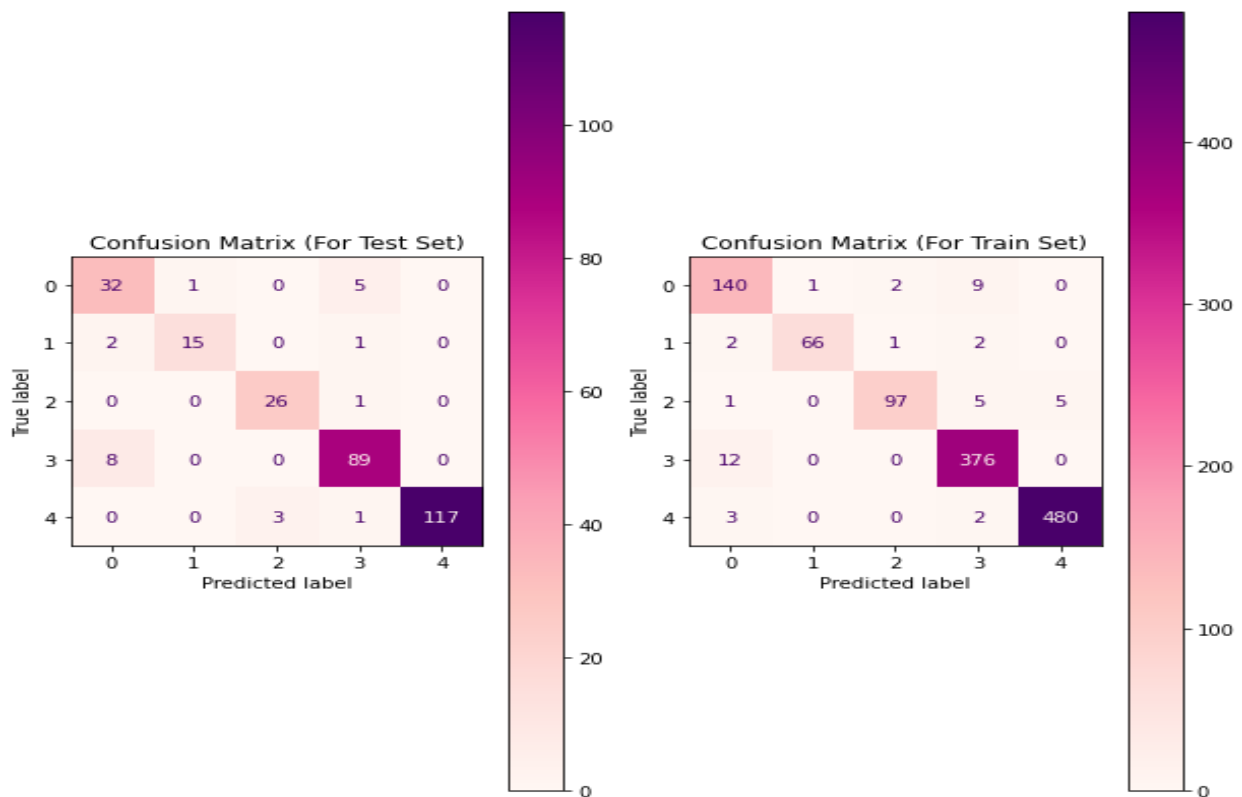
-----TRAIN SET REPORT FOR DECISION TREE-----

	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
weighted avg	0.96	0.96	0.96

Confusion Matrix for Decision Trees :

Decision Tree Classifier



XGBOOST CLASSIFIER (RESULTS) :

-----XGBOOST-----

F1 score is 0.9471365290302168

-----TEST SET REPORT FOR XGBOOST-----

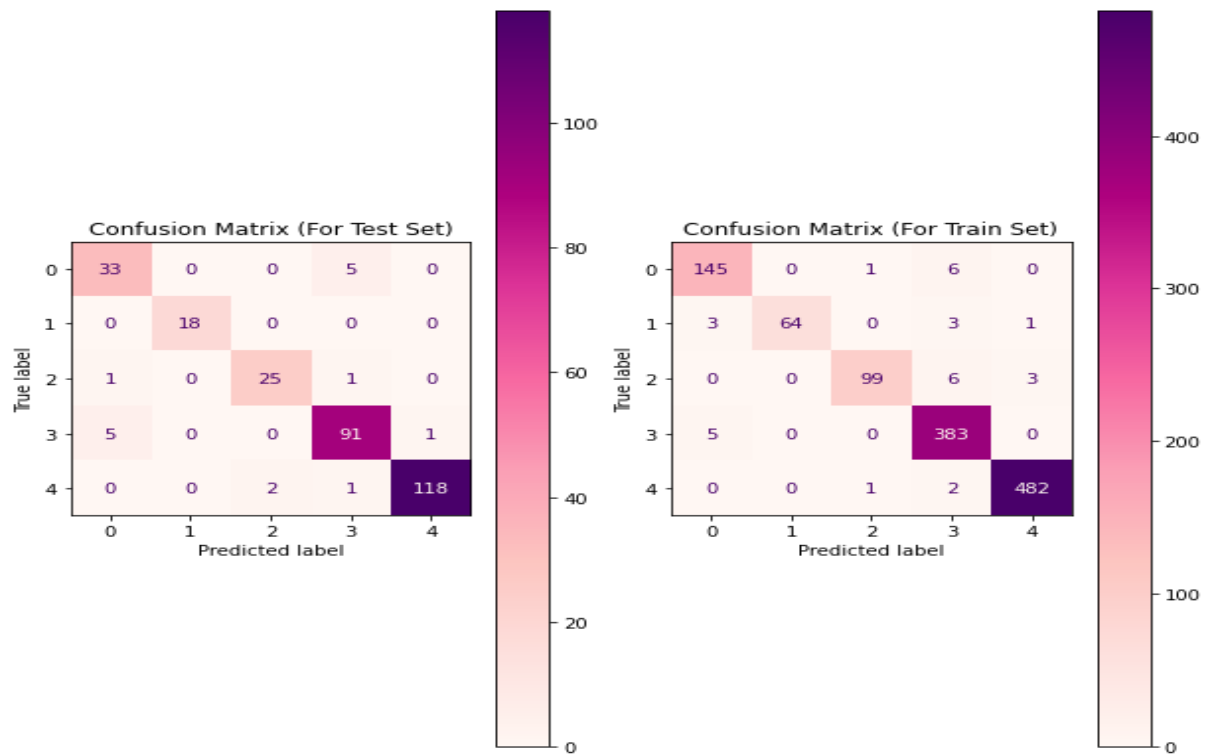
	precision	recall	f1-score	support
Banking	0.85	0.87	0.86	38
Jobs-IT	1.00	1.00	1.00	18
Rent-Apartment	0.93	0.93	0.93	27
Retail	0.93	0.94	0.93	97
Sell-House	0.99	0.98	0.98	121
accuracy		0.95		301
macro avg	0.94	0.94	0.94	301
weighted avg	0.95	0.95	0.95	301

-----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



CONCLUSION : From the experiments performed above we have decided to move on with the xgboost classifier for the first phase of our model