Balanced Random Forest Classifier Model Report

Summary

The Balanced Random Forest Classifier performed reasonably well on this task, with an accuracy of 0.9420814018196827 and an F1-score of 0.9421358513964577.

Model Selection

The hyperparameters we tuned were:

{‘n_estimators’: [50, 100, 200], ‘max_depth’: [None, 5, 10, 20]}

Model Performance

The best parameters found by RandomizedSearchCV were:

Best parameters:, {‘n_estimators’: 50, ‘max_depth’: None}

With these parameters, the model achieved the following performance metrics: Best cross-validation score: 0.8805888260748811

Accuracy:, 0.9420814018196827 F1-score: 0.9421358513964577

Testing Data

Classification report:

	precision	recall	f1-score	support
0	0.96	0.96	0.96	19595447
1	0.88	0.89	0.89	6606154
accuracy			0.94	26201601

macro avg 0.92 0.92 0.92 26201601 weighted avg 0.94 0.94 0.94 26201601

TRAINING DATA Classificatin Report-Confusion Matrix

Training confusion matrix:

[[2167159 10113] [2289 731728]]

Training classification report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2177272
1	0.99	1.00	0.99	734017
accuracy			1.00	2911289

macro avg 0.99 1.00 0.99 2911289 weighted avg 1.00 1.00 1.00 2911289

This indicates that the model correctly classified 2167159 instances of class 0 and 731728 instances of class 1,

while misclassifying 10113 instances of class 0 and 2289 instances of class 1.

CV Results: mean_fit_time std_fit_time mean_score_time std_score_time
0 1531.370871 37.869092 72.132009 4.488346
1 976.373627 121.075829 53.577014 9.490456
2 988.434049 171.230966 104.205794 20.544709
3 2768.398778 413.179189 37.413728 30.720385
4 2222.861225 38.868204 41.687448 2.869501
5 3625.388225 357.275359 163.895234 147.053798
6 1825.003701 109.652829 106.565838 100.345349
7 2169.276666 544.281390 32.935516 26.892366
8 630.900981 116.240335 9.042147 7.852144
9 669.245528 195.576768 11.216128 5.630825

param_n_estimators param_max_depth
0 100 20
1 50 20
2 50 None
3 200 10
4 200 5
5 200 None
6 100 None
7 200 20
8 100 5
9 100 10

	params	split0_test_precision	\

0 {‘n_estimators’: 100, ‘max_depth’: 20} 0.756865
1 {‘n_estimators’: 50, ‘max_depth’: 20} 0.754937
2 {‘n_estimators’: 50, ‘max_depth’: None} 0.878946
3 {‘n_estimators’: 200, ‘max_depth’: 10} NaN
4 {‘n_estimators’: 200, ‘max_depth’: 5} 0.440298
5 {‘n_estimators’: 200, ‘max_depth’: None} 0.878663
6 {‘n_estimators’: 100, ‘max_depth’: None} 0.878562
7 {‘n_estimators’: 200, ‘max_depth’: 20} 0.757115
8 {‘n_estimators’: 100, ‘max_depth’: 5} 0.440115
9 {‘n_estimators’: 100, ‘max_depth’: 10} 0.511701

split1_test_precision split2_test_precision … std_test_f1
0 0.757157 0.756630 … 0.000482
1 0.754745 0.756958 … 0.000302
2 0.879368 0.879634 … 0.000310
3 0.510690 0.512900 … NaN
4 0.443368 0.441609 … 0.001293
5 0.879347 0.879116 … NaN
6 NaN 0.879388 … NaN
7 NaN NaN … NaN
8 NaN 0.443437 … NaN
9 0.511023 0.512156 … NaN

rank_test_f1 split0_test_roc_auc split1_test_roc_auc
0 2 0.966008 0.965693
1 3 0.965431 0.965107
2 1 0.975034 0.975098
3 5 NaN 0.845431
4 4 0.761143 0.762656
5 5 0.977277 0.977390
6 5 0.976424 NaN
7 5 0.966171 NaN
8 5 0.761286 NaN
9 5 0.843230 0.845473

split2_test_roc_auc split3_test_roc_auc split4_test_roc_auc
0 0.965498 0.965666 0.965679
1 0.965263 0.965298 0.965510
2 0.974698 0.975207 0.974972
3 0.846034 NaN 0.845027
4 0.762222 0.762814 0.762629
5 0.976976 NaN NaN
6 0.976066 0.976415 NaN
7 NaN 0.965899 0.965798
8 0.762534 NaN 0.763254
9 0.845598 NaN 0.844422

mean_test_roc_auc std_test_roc_auc rank_test_roc_auc
0 0.965709 0.000165 2
1 0.965322 0.000140 3

2 0.975002 0.000171 1
3 NaN NaN 5
4 0.762293 0.000607 4
5 NaN NaN 5
6 NaN NaN 5
7 NaN NaN 5
8 NaN NaN 5
9 NaN NaN 5

[10 rows x 39 columns]