Random forest

Test case 1 → (Normal Default)

Parameters → n_estimator : 300 | random state : 42

Random State	Mean cross validation accuracy
42	c

```
rrc.rit(x_train, y_train)
Confusion Matrix:
 [[69016   120     2    13     2]
 [  355 22358   321    18    62]
 [   55   446 16340     6    18]
 [  148    73    13  2622     6]
 [   33    54     2     3  2676]]
------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           0       0.99      1.00      0.99     69153
           1       0.97      0.97      0.97     23114
           2       0.98      0.97      0.97     16865
           3       0.98      0.92      0.95      2862
           4       0.97      0.97      0.97      2768

    accuracy                           0.98    114762
   macro avg       0.98      0.96      0.97    114762
weighted avg       0.98      0.98      0.98    114762


------------------------------------------------------------
Classification report  for each individual class:
F1 Score for Class 0: 0.9947535312770249
F1 Score for Class 1: 0.9686125852918879
F1 Score for Class 2: 0.9742718301881168
F1 Score for Class 3: 0.9493120926864591
F1 Score for Class 4: 0.9674620390455532
------------------------------------------------------------
Accuracy of the Random Forest is --> 0.9847510499991287
```

Parameters → n_estimator : 200

Random State	Accuracies		
10	Training	0.998403	
	Testing	0.985727	
18	Training	0.998403	
	Testing	0.985309	
73	Training	0.998405	
	Testing	0.985771	
78	Training	0.998425	
	Testing	0.985710	

Test case 2 → (Testing the best random state)

	Random State	Training Accuracy	Validation Accuracy
0	1.0	0.998460	0.984795
1	2.0	0.998442	0.985317
2	3.0	0.998423	0.985047
3	4.0	0.998432	0.985073
4	5.0	0.998477	0.984899
5	6.0	0.998386	0.985396
6	7.0	0.998412	0.985309
7	8.0	0.998453	0.984603
8	9.0	0.998455	0.985213
9	10.0	0.998403	0.985727
10	11.0	0.998412	0.985213
11	12.0	0.998421	0.984890
12	13.0	0.998488	0.985213
13	14.0	0.998462	0.984856
14	15.0	0.998440	0.985030
15	16.0	0.998469	0.984856
16	17.0	0.998475	0.985213
17	18.0	0.998447	0.985762
18	19.0	0.998403	0.985309
19	20.0	0.998425	0.985361
20	21.0	0.998432	0.985161
21	22.0	0.998475	0.984559
22	23.0	0.998399	0.985030
23	24.0	0.998442	0.984795
24	25.0	0.998495	0.984847
25	26.0	0.998449	0.984943
26	27.0	0.998432	0.984838
27	28.0	0.998418	0.985082
28	29.0	0.998445	0.984943
29	30.0	0.998455	0.984925
30	31.0	0.998449	0.985405

	Random State	Training Accuracy	Validation Accuracy
31	32.0	0.998484	0.984725
32	33.0	0.998408	0.985108
33	34.0	0.998453	0.985143
34	35.0	0.998427	0.985213
35	36.0	0.998418	0.985030
36	37.0	0.998495	0.984873
37	38.0	0.998405	0.984969
38	39.0	0.998460	0.984864
39	40.0	0.998410	0.984760
40	41.0	0.998442	0.984978
41	42.0	0.998436	0.984969
42	43.0	0.998453	0.984646
43	44.0	0.998379	0.985178
44	45.0	0.998410	0.985143
45	46.0	0.998460	0.985561
46	47.0	0.998425	0.984995
47	48.0	0.998438	0.984882
48	49.0	0.998451	0.984978
49	50.0	0.998412	0.984812

	Random State	Training Accuracy	Validation Accuracy
0	51.0	0.998429	0.985413
1	52.0	0.998462	0.984943
2	53.0	0.998473	0.985448
3	54.0	0.998455	0.984986
4	55.0	0.998455	0.985248
5	56.0	0.998434	0.985614
6	57.0	0.998440	0.984986
7	58.0	0.998442	0.985283
8	59.0	0.998488	0.985012
9	60.0	0.998449	0.985039
10	61.0	0.998412	0.985344
11	62.0	0.998471	0.984882
12	63.0	0.998471	0.984838
13	64.0	0.998464	0.985065
14	65.0	0.998449	0.984943
15	66.0	0.998410	0.985248
16	67.0	0.998434	0.984986
17	68.0	0.998416	0.985204
18	69.0	0.998427	0.985370
19	70.0	0.998447	0.985117
20	71.0	0.998421	0.985457
21	72.0	0.998469	0.985274
22	73.0	0.998405	0.985771
23	74.0	0.998484	0.984498
24	75.0	0.998445	0.984673
25	76.0	0.998436	0.985361
26	77.0	0.998458	0.985161
27	78.0	0.998425	0.985710
28	79.0	0.998445	0.985012
29	80.0	0.998460	0.985640

40	81.0	0.998471	0.984734
41	82.0	0.998460	0.985134
42	83.0	0.998469	0.984943
43	84.0	0.998416	0.985178
44	85.0	0.998477	0.985126
45	86.0	0.998434	0.985431
46	87.0	0.998445	0.984978
47	88.0	0.998440	0.985030
48	89.0	0.998449	0.985378
49	90.0	0.998425	0.985300
50	91.0	0.998453	0.985222
51	92.0	0.998416	0.985178
52	93.0	0.998434	0.985265
53	94.0	0.998449	0.984821
54	95.0	0.998440	0.985126
55	96.0	0.998427	0.985901
56	97.0	0.998462	0.984742
57	98.0	0.998484	0.984760
58	99.0	0.998429	0.985195
59	100.0	0.998390	0.985265

Conclusion : Random States 10,18,73,78 shows has validation accuracies above 98.57%

Test case 3 → (Testing for trees using cross validation)

```
n_estimators=50:
Training Accuracy: 0.9984
Validation Accuracy: 0.9847
```

```
n_estimators=300:
Training Accuracy: 0.9984
Validation Accuracy: 0.9850
```

```
n_estimators=100:
Training Accuracy: 0.9984
Validation Accuracy: 0.9848
```

```
n_estimators=350:
Training Accuracy: 0.9984
Validation Accuracy: 0.9850
```

```
n_estimators=150:
Training Accuracy: 0.9984
Validation Accuracy: 0.9849
```

```
n_estimators=400:
Training Accuracy: 0.9984
Validation Accuracy: 0.9850
```

```
n_estimators=200:
Training Accuracy: 0.9984
Validation Accuracy: 0.9849
```

```
n_estimators=450:
Training Accuracy: 0.9984
Validation Accuracy: 0.9850
```

```
n_estimators=250:
Training Accuracy: 0.9984
Validation Accuracy: 0.9849
```

```
n_estimators=500:
Training Accuracy: 0.9984
Validation Accuracy: 0.9849
```

Conclusion : Retesting for trees from 300 to 450 with random states : 10,18,73,78 (1.27)

Random State	Accuracies	300 trees	350 trees	400 trees	450 trees
10	Training	0.9984032	0.9984032	0.9984032	0.9984032
	Testing	0.9857793	0.9857270	0.9857357	0.9857357
18	Training	0.9984468	0.9984468	0.9984468	0.9984468
	Testing	0.9857270	0.9857618	0.9857618	0.9857706
73	Training	0.9984054	0.9984054	0.9984054	0.9984054
	Testing	0.9857880	0.9857706	0.9857618	0.9857618
78	Training	0.9984250	0.9984250	0.9984250	0.9984250
	Testing	0.9856573	0.9857096	0.9856573	0.9857270

Conclusion : Best state - 73 | Best number of trees - 300

Test case 4 → (Testing for imbalance dataset and criterion)

Random state : 73 | Trees : 300

Balancing Technique	Accuracies	Gini criterion	Entropy criterion
SMOTE	Training	0.9988462816604297	0.9988462816604297
	Testing	0.984977605827713	0.9848033321134173
Balanced	Training	0.9981701100548096	0.9981701100548096
	Testing	0.9852564437705861	0.9851083111134348
SMOTE + Balanced	Training	0.9988462816604297	0.9988462816604297
	Testing	0.9851083111134348	0.9848033321134173
Class Frequencies	Training	0.9981701100548096	0.9981701100548096
	Testing	0.9852041616562974	0.9850821700562904

Max_depth	Random state 10	Random state 18	Random state 73	Random state 78
None	0.9843370	0.9841365	0.9842651	0.9841802
10	0.9723730	0.9725451	0.9724188	0.9724536
20	0.9834068	0.9832020	0.9833328	0.9833240
30	0.9845396	0.9842804	0.9843501	0.9843305
40	0.9843632	0.9841083	0.9842433	0.9841823
50	0.9843414	0.9841322	0.9842608	0.9841802
60	0.9843370	0.9841366	0.9842651	0.9841802

Test case 5 → (Testing for imbalance dataset and criterion)

Random state - 73 | Max Depth - 30

Balancing Technique	Accuracies	Depth	Gini criterion	Entropy criterion
SMOTE	Training	30	0.9984706188322665	0.9987032408143214
	Testing	30	0.9850734563705756	0.9849601784562835
Balanced	Training	30	0.9980786155575501	0.998152682531522
	Testing	30	0.9853174395705896	0.9852128753420122
SMOTE + Balanced	Training	30	0.9984706188322665	0.9987032408143214
	Testing	30	0.9850734563705756	0.9849601784562835
Class Frequencies	Training	30	0.9980372251897421	0.9981505040911111
	Testing	30	0.9853261532563043	0.9851257384848643

Random State - 10 | Max Depth - 30

Balancing Technique	Accuracies	Depth	Gini criterion	Entropy criterion
SMOTE	Training	30	0.9985721002567904	0.9987695757531918
	Testing	30	0.9853087258848748	0.9851431658562939
Balanced	Training	30	0.9979021618842638	0.9979892995007015
	Testing	30	0.98509959742772	0.9850647426848609
SMOTE + Balanced	Training	30	0.998546782885457	0.998758002097725
	Testing	30	0.9853261532563043	0.9851518795420087
Class Frequencies	Training	30	0.9979065187650857	0.9980001917027561
	Testing	30	0.9850908837420052	0.9851605932277234

Random State - 10 | Max Depth - 40

Balancing Technique	Accuracies	Depth	Gini criterion	Entropy criterion
SMOTE	Training	40	0.9988903757821259	0.9988896524286592
	Testing	40	0.9852477300848713	0.985056028999146
Balanced	Training	40	0.9979892995007015	0.9979740504178248
	Testing	40	0.9850647426848609	0.9849166100277095
SMOTE + Balanced	Training	40	0.9988795254801258	0.9988802488335925
	Testing	40	0.9852825848277305	0.9851344521705792
Class Frequencies	Training	40	0.9979740504178248	0.9979740504178248
	Testing	40	0.9849340373991391	0.9850386016277165

Test case 6 → (Testing for best maximum feature)

Random State : 10 | max_depth : 30 | trees : 300

```
Training Accuracy  Validation Accuracy max_features
        0.997652               0.985779         auto
        0.997652               0.985779         sqrt
        0.997652               0.985779         log2
        0.997641               0.985596          0.2
        0.997617               0.986215          0.5
        0.997591               0.986137          0.6
        0.997582               0.985980          0.7
        0.997534               0.985910          0.8
```

Random State : 73 | max_depth : 30 | trees : 300

```
Training Accuracy  Validation Accuracy max_features
        0.997602               0.985710         auto
        0.997602               0.985710         sqrt
        0.997602               0.985710         log2
        0.997623               0.985492          0.2
        0.997649               0.986049          0.5
        0.997595               0.985962          0.6
        0.997630               0.986041          0.7
        0.997604               0.985823          0.8
```

Test case 7 → (Testing for best maximum feature along with dept and criterion and balancing technique)

Random State : 10 | max_depth : 30 | trees : 300 | max_features : 0.5

Balancing Technique	Accuracies	Maximum Features	Depth	Gini criterion	Entropy criterion
SMOTE	Training	0.5	30	0.998593077507324	0.9988896524286592
	Testing	0.5	30	0.9853522943134487	0.985056028999146
Balanced	Training	0.5	30	0.997934838490428	0.9980023701431671
	Testing	0.5	30	0.9852564437705861	0.9853087258848748
SMOTE + Balanced	Training	0.5	30	0.998593077507324	0.9987941697710586
	Testing	0.5	30	0.9853522943134487	0.9851867342848678

Class Frequencies	Training	0.5	30	0.9979174109671404	0.998004548583578
	Testing	0.5	30	==0.9854045764277374==	0.9853697216848782

Random State : 10 | max_depth : 30 | trees : 300 | max_features : 0.6

Balancing Technique	Accuracies	Maximum Features	Depth	Gini criterion	Entropy criterion
SMOTE	Training	0.6	30	0.9985887373865239	0.9987912763571919
	Testing	0.6	30	0.9851518795420087	0.9850473153134313
Balanced	Training	0.6	30	0.9979108756459076	0.998006727023989
	Testing	0.6	30	0.9853261532563043	0.9854394311705965
SMOTE + Balanced	Training	0.6	30	0.9985887373865239	0.9987912763571919
	Testing	0.6	30	0.9851518795420087	0.9850473153134313
Class Frequencies	Training	0.6	30	0.9979108756459076	0.998004548583578
	Testing	0.6	30	0.9852564437705861	==0.9854917132848853==

Random State : 73 | max_depth : 30 | trees : 300 | max_features : 0.5

Balancing Technique	Accuracies	Maximum Features	Depth	Gini criterion	Entropy criterion
SMOTE	Training	0.5	30	0.9985233561139125	0.9987494762393261
	Testing	0.5	30	0.985056028999146	0.9849078963419947
Balanced	Training	0.5	30	0.9980720802363172	0.9981635747335768
	Testing	0.5	30	0.985422003799167	==0.9856049911991774==
SMOTE + Balanced	Training	0.5	30	0.9985233561139125	0.9987494762393261
	Testing	0.5	30	0.985056028999146	0.9849078963419947
	Training	0.5	30	0.9980720802363172	0.9981613962931658

Class Frequencies	Testing	`0.5`	`30`	`0.9854829995991704`	`0.9855178543420295`

Random State : 73 | max_depth : 30 | trees : 300 | max_features : 0.7

Balancing Technique	Accuracies	Maximum Features	Depth	Gini criterion	Entropy criterion
SMOTE	Training	`0.7`	`30`	`0.998512519686177`	`0.9987415295256534`
	Testing	`0.7`	`30`	`0.984733622627699`	`0.9846900541991251`
Balanced	Training	`0.7`	`30`	`0.9980372251897421`	`0.9981635747335768`
	Testing	`0.7`	`30`	`0.9852912985134452`	`0.9853871490563079`
SMOTE + Balanced	Training	`0.7`	`30`	`0.998512519686177`	`0.9987415295256534`
	Testing	`0.7`	`30`	`0.984733622627699`	`0.9846900541991251`
Class Frequencies	Training	`0.7`	`30`	`0.9980502958322078`	`0.9981635747335768`
	Testing	`0.7`	`30`	`0.985422003799167`	`0.9852825848277305`

XGBoost

Parameters → n_estimator : 300 | random state : 42 | K-folds : 5

Random State	Mean cross validation accuracy
42	`0.9832695789961612`

Test case 2 → Best Random state

Parameters → n_estimator : 500

Random State	Accuracies	
9	Training	0.994733
	Testing	0.987383
77	Training	0.994717
	Testing	0.987661
52	Training	0.994837
	Testing	0.987322
55	Training	0.994935
	Testing	0.987322
89	Training	0.994713
	Testing	0.987304

Test case 3 → Best number of trees

Parameters → Random state : 77 | K-folds : 5

n_estimator	Accuracies	
200	Mean cross validation	0.9852062119093377
300	Mean cross validation	0.985820532166928
400	Mean cross validation	0.9861189781615469
500	Mean cross validation	0.986251862855774

600	Mean cross validation	0.9863477141674185
700	Mean cross validation	0.9863607848763232
800	Mean cross validation	0.986321572488599
900	Mean cross validation	0.9863455352571897
1000	Mean cross validation	0.9863302863024449

Test case 4 → Best Learning Rate value

Parameters → Random state : 77 | K-folds : 5 | n_estimator : 700

Learning _rate	Accuracies	
0.001	Mean cross validation	0.964192978462215
0.01	Mean cross validation	0.9746516699451971
0.05	Mean cross validation	0.9836072368564845
0.1	Mean cross validation	0.9853957358122232
0.2	Mean cross validation	0.9862910746028376
0.3	Mean cross validation	0.9863607848763232
0.35	Mean cross validation	0.9863607849000513
0.4	Mean cross validation	0.9862540416236338

Test case 5 → Best Maximum depth

Parameters → Random state : 77 | K-folds : 5 | n_estimator : 700 | Learning rate : 0.35

Max_depth	Accuracies	
3	Mean cross validation	0.9837858698860924
4	Mean cross validation	0.9855591196735325

5	Mean cross validation	0.9862104727205014
6	Mean cross validation	0.9863607849000513
7	Mean cross validation	<mark>0.98631068085602</mark>
8	Mean cross validation	0.9860296623467241
9	Mean cross validation	0.9859447026202022
10	Mean cross validation	0.985968665839628

<mark>**Test case 6 → Best Minimum child weight**</mark>

<mark>**Parameters → Random state : 77 | K-folds : 5 | n_estimator : 700 | Learning rate : 0.35 | max_depth : 7**</mark>

Min_child_weight	Accuracies	
1	Mean cross validation	<mark>0.9863607849000513</mark>
2	Mean cross validation	0.9862540414575365
3	Mean cross validation	0.9862671112884991
4	Mean cross validation	0.9862431490656564
5	Mean cross validation	0.9862496841590989
6	Mean cross validation	0.9860863011520034
7	Mean cross validation	0.9860623381224027
8	Mean cross validation	0.9861516543406046
9	Mean cross validation	0.985942524113524
10	Mean cross validation	0.985996985432093

<mark>**Test case 7 → Best tree method**</mark>

<mark>**Parameters → Random state : 77 | K-folds : 5 | n_estimator : 700 | Learning rate : 0.35 | max_depth : 7 | min_child_weight : 1**</mark>

Tree Method	Accuracies	
'hist'	Mean cross validation	0.9863607849000513
'auto'	Mean cross validation	0.9863607849000513
'approx'	Mean cross validation	0.986354250115075

Parameters → Random state : 77 | K-folds : 5 | n_estimator : 700 | Learning rate : 0.35 | max_depth : 7 | min_child_weight : 1

Tree Method	Accuracies	
'hist'	Mean cross validation	0.9863607849000513
'auto'	Mean cross validation	0.9863607849000513
'approx'	Mean cross validation	0.986354250115075

Parameters → Random state : 77 | K-folds : 5 | n_estimator : 700 | Learning rate : 0.35 | max_depth : 7 | min_child_weight : 1

Tree Method	Accuracies		
		'gbtree'	'gblinear'
'hist'	Mean cross validation	0.9863607849000513	
'auto'	Mean cross validation	0.9863607849000513	
'approx'	Mean cross validation		

Method	Accuracies	
SMOTE	Training	0.9974187760889538
	Testing	0.9859186838849097
Balanced	Training	0.997821559589059
	Testing	0.9871298861992646

Objective	Accuracies	
Multi:softmax	Training	0.997821559589059
	Testing	0.9871298861992646
Multi:softprob	Training	0.997821559589059
	Testing	0.9871298861992646