**REPORT**

* Dataset Name - banknote authentication Data Set.
* Number of Instances – 1372
* Number of Attributes – 5
* Data were extracted from images that were taken from genuine and forged banknote-like specimens.
* For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels.
* Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained.
* Wavelet Transform tool were used to extract features from these images.
* Category 1 for genuine and 2 for forged.
* Task - data were extracted from images that were taken for the evaluation of an authentication procedure for banknotes.
* Feature Names:

|  |  |
| --- | --- |
| Feature Name | Detail |
| V1 | variance of Wavelet Transformed image (continuous) |
| V2 | skewness of Wavelet Transformed image (continuous) |
| V3 | curtosis of Wavelet Transformed image (continuous) |
| V4 | entropy of image (continuous) |

Features -> ['V1', 'V2', 'V3', 'V4']

* Data:

V1 V2 V3 V4

0 3.62160 8.66610 -2.8073 -0.44699

1 4.54590 8.16740 -2.4586 -1.46210

2 3.86600 -2.63830 1.9242 0.10645

3 3.45660 9.52280 -4.0112 -3.59440

4 0.32924 -4.45520 4.5718 -0.98880

... ... ... ... ...

1367 0.40614 1.34920 -1.4501 -0.55949

1368 -1.38870 -4.87730 6.4774 0.34179

1369 -3.75030 -13.45860 17.5932 -2.77710

1370 -3.56370 -8.38270 12.3930 -1.28230

1371 -2.54190 -0.65804 2.6842 1.19520

[1372 rows x 4 columns]

* Target:

0 1

1 1

2 1

3 1

4 1

..

1367 2

1368 2

1369 2

1370 2

1371 2

Name: Class, Length: 1372, dtype: category

Categories (2, object): ['1', '2']

* For DecisionTreeClassifier, criterion = entropy is set.
* For cross validation, scoring is set to roc\_auc, and cv = 10 for all models.
* For Decision tree with tuned min\_samples\_leaf using GridSearchCV, min\_samples\_left is set to 10.
* Also, parameters for min\_samples\_left is set of values – [2, 4, 6, 8, 10]

The following table shows the evaluation measures (test\_roc\_auc mean) for all models:

|  |  |
| --- | --- |
| Model Name | Test ROC AUC Mean Score |
| Decision tree with default parameters | 0.9870510852391785 |
| Decision tree with tuned min\_samples\_leaf using GridSearchCV | 0.9894949743957507 |
| Random Forest | 0.9997735116479725 |
| Bagged decision tree | 0.9982401981107761 |
| AdaBoosted decision tree | 0.9984685073339087 |

* The random forest model worked better on banknote authentication dataset. The AdaBoosted decision tree and Bagged decision tree performed little better than Decision tree with default parameters and tuned min\_samples\_leaf using GridSearchCV. So, we can say that random forests are a strong modeling technique, and they are much more robust than a single decision tree.