
# Pattern Recognition

**Tushant Yadav 2K18/EE/224**
**Nishkarsh Sharma 2K18/PE/040**

# Real and Fake Banknote Nave Bayes Classifier

## DATASET

| | Variance | Skewness | Curtosis | Entropy | Class |
|---|---|---|---|---|---|
| 365 | 3.89990 | 1.73400 | 1.6011 | 0.967650 | 0 |
| 23 | 0.93584 | 8.88550 | -1.6831 | -1.659900 | 0 |
| 109 | 2.91630 | 10.83060 | -3.3437 | -4.122000 | 0 |
| 37 | 3.62890 | 0.81322 | 1.6277 | 0.776270 | 0 |
| 226 | 0.57060 | -0.02480 | 1.2421 | -0.562100 | 0 |
| ... | ... | ... | ... | ... | ... |
| 271 | 5.39150 | 9.99460 | -3.8081 | -3.364200 | 0 |
| 580 | 4.70720 | 8.29570 | -2.5605 | -1.490500 | 0 |
| 719 | 1.77480 | -0.76978 | 5.5854 | 1.303900 | 0 |
| 366 | 3.51890 | 6.33200 | -1.7791 | -0.020273 | 0 |
| 233 | 0.46901 | -0.63321 | 7.3848 | 0.365070 | 0 |

381 rows × 5 columns

FAKE NOTES TESTSET

| | Variance | Skewness | Curtosis | Entropy | Class |
|---|---|---|---|---|---|
| 373 | 1.91050 | 8.8710 | -2.33860 | -0.756040 | 0 |
| 336 | 3.96600 | 3.9213 | 0.70574 | 0.336620 | 0 |
| 357 | 3.20510 | 8.6889 | -2.90330 | -0.781900 | 0 |
| 496 | 3.37560 | -4.0951 | 4.36700 | 1.069800 | 0 |
| 438 | 4.05240 | 5.6802 | -1.96930 | 0.026279 | 0 |
| ... | ... | ... | ... | ... | ... |
| 746 | 4.64990 | 7.6336 | -1.94270 | -0.374580 | 0 |
| 39 | 3.48050 | 9.7008 | -3.75410 | -3.437900 | 0 |
| 355 | -0.21661 | 8.0329 | 1.88480 | -3.885300 | 0 |
| 375 | 0.58836 | 10.7727 | -1.38840 | -4.327600 | 0 |
| 617 | -0.27800 | 8.1881 | -3.13380 | -2.527600 | 0 |

381 rows × 5 columns

FAKE NOTES TRAINSET

| | Variance | Skewness | Curtosis | Entropy | Class |
|---|---|---|---|---|---|
| 1206 | -2.43490 | -9.24970 | 8.992200 | -0.50001 | 1 |
| 1058 | -1.56210 | -2.21210 | 4.259100 | 0.27972 | 1 |
| 883 | -3.36040 | -0.32696 | 2.132400 | 0.60170 | 1 |
| 1059 | -3.23050 | -7.21350 | 11.643300 | -0.94613 | 1 |
| 1157 | -5.20490 | 7.25900 | 0.070827 | -7.30040 | 1 |
| ... | ... | ... | ... | ... | ... |
| 1270 | -0.74324 | -0.32902 | -0.427850 | 0.23317 | 1 |
| 1089 | -2.98210 | 4.19860 | -0.589800 | -3.96420 | 1 |
| 1211 | -4.39670 | 4.96010 | -0.648920 | -5.47190 | 1 |
| 891 | -0.55008 | 2.86590 | -1.648800 | -2.43190 | 1 |
| 1052 | -2.41980 | -0.24418 | 0.701460 | 0.41809 | 1 |

305 rows × 5 columns

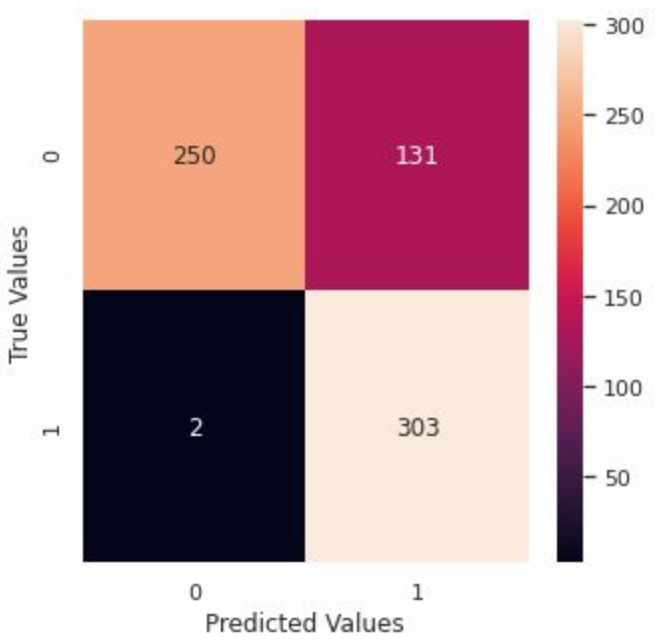
REAL NOTES TRAINSET

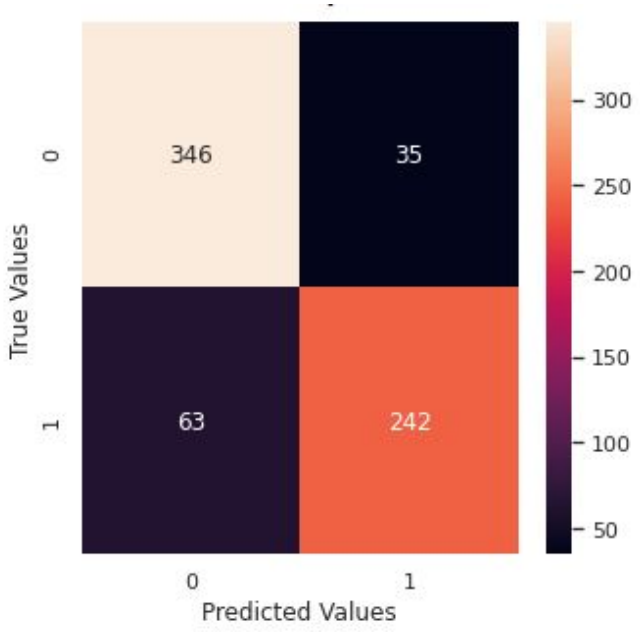
| | Variance | Skewness | Curtosis | Entropy | Class |
|---|---|---|---|---|---|
| 1154 | -0.36025 | -4.44900 | 2.10670 | 0.94308 | 1 |
| 1116 | -0.49948 | 1.77340 | -2.24690 | -0.68104 | 1 |
| 1272 | -4.23330 | 4.91660 | -0.49212 | -5.32070 | 1 |
| 958 | -0.36372 | 3.04390 | -3.48160 | -2.78360 | 1 |
| 814 | -1.25680 | -1.47330 | 2.87180 | 0.44653 | 1 |
| ... | ... | ... | ... | ... | ... |
| 1196 | -2.01490 | 3.68740 | -1.93850 | -3.89180 | 1 |
| 1179 | -3.27780 | 1.80230 | 0.18050 | -2.39310 | 1 |
| 914 | -2.53460 | -0.77392 | 3.36020 | 0.00171 | 1 |
| 1038 | -0.70867 | -5.56020 | 4.04830 | 0.90300 | 1 |
| 1055 | -0.60254 | 1.72370 | -2.15010 | -0.77027 | 1 |

305 rows × 5 columns

REAL NOTES TESTSET

## CONFUSION MATRIX



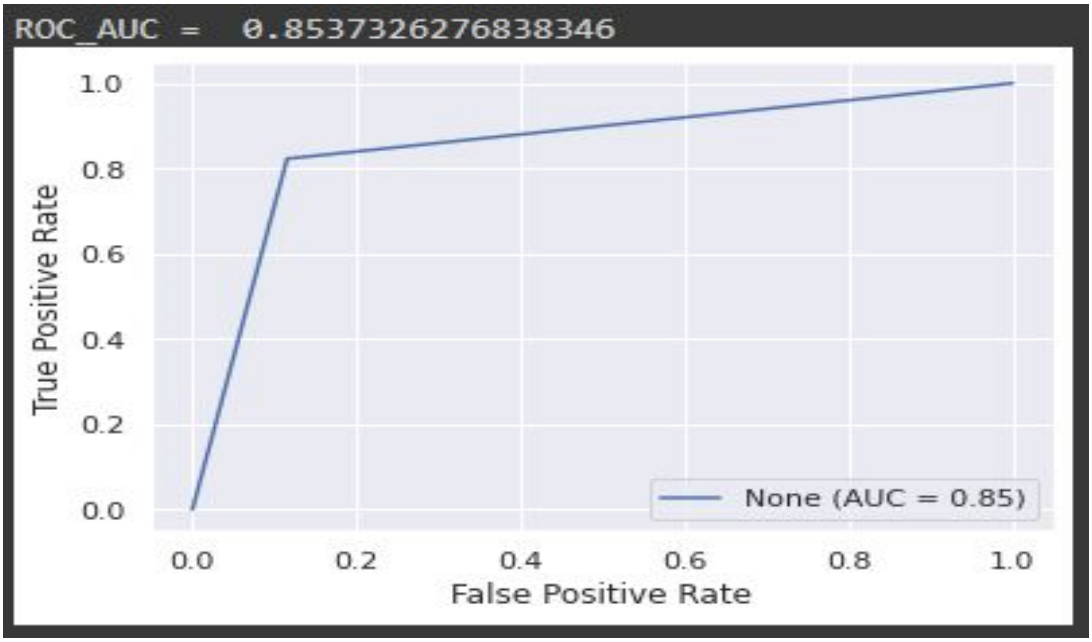

Training Set

Acc: 0.8061224489795918

Test set

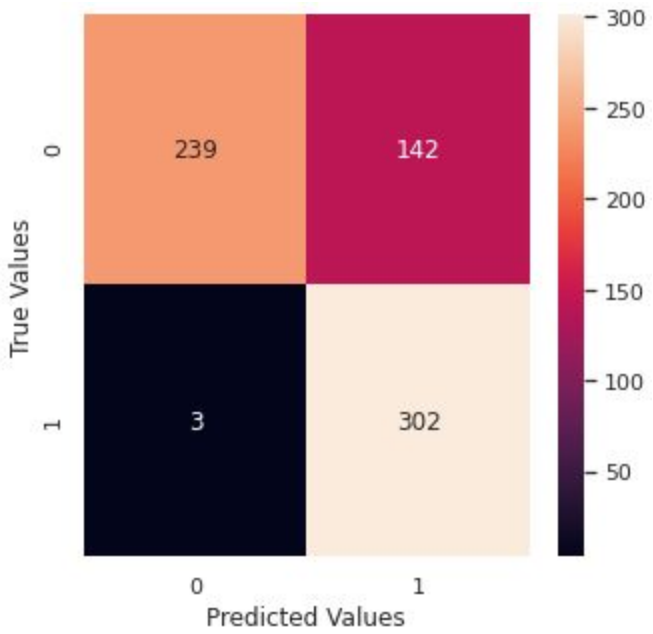
Acc: 0.8571428571428571

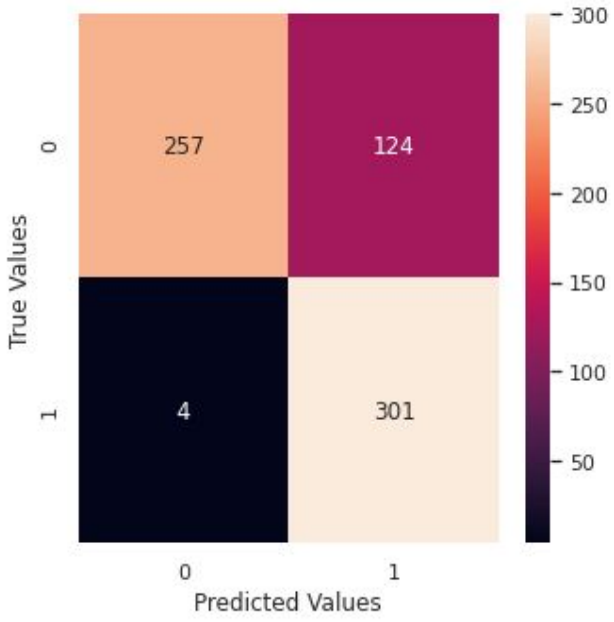
## RECEIVER OPERATING CHARACTERISTICS (ROC)



# Real and Fake Banknote Naive Bayes Classifier ( Changing Prior Genuine 0.1 and Prior Forged 0.9 )

## CONFUSION MATRIX



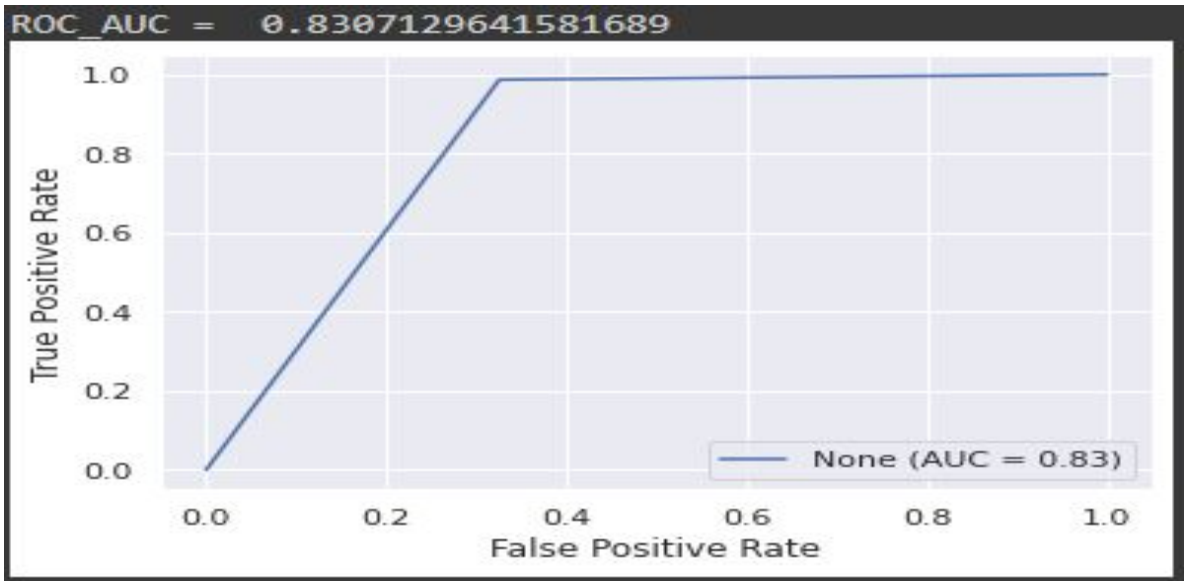

Training Set

Acc: 0.7886297376093294

Test set

Acc: 0.8134110787172012

## RECEIVER OPERATING CHARACTERISTICS (ROC)



## • Comparison

A Bayes Classifier has been designed for classification between Fake Bank Notes and Real Bank Notes indices initially, and then classification is conducted for the same after provision of Prior Information. Pre-processing of data involves reading the data from banknote.csv file , class based splitting of dataset into real and fake notes and further dividing the two of them in testing and training datasets respectivley.

• Classification without Prior Accuracy: 85.07%, AUC: 0.8507895

• Classification with Prior Accuracy: 83.07%, AUC: 0.8307129

The Bayesian model performed better in the classification without Prior Information rather than with the provision of prior information.

This might be due to improper estimation of prior probabilities. The wrong estimation of prior probabilities leads the classification results to the worse accuracy. The boundary between classes moves to the wrong direction due to the inappropriately estimated prior probabilities , while if the ratio of the prior probabilities of the involved classes is close to 1.0, the classification results with the appropriately estimated prior probabilities.

Hence occurance of the results is thoretically justified since ratio of Forged and Genuine Prior probabilities is equal to nine as compared to desired ratio of approximate value around one.

Link to code:-

https://colab.research.google.com/drive/1iZ5UKK9wrX5kUpMxtQIjPfuf1IPPMzgK?usp=sharing