Breast Cancer Detection

NITIKA AGARWAL

Breast Cancer Prediction Flow Chart

Import Required Libraries

↓

Preprocess the data

↕

Data Visualization and Interpretation → Correlation Matrix

↘ Histogram

↓

Data splitting

Apply SVM

Apply naïve Bias

↓

Implement and Calculate Results

↓

Compare the Algorithms
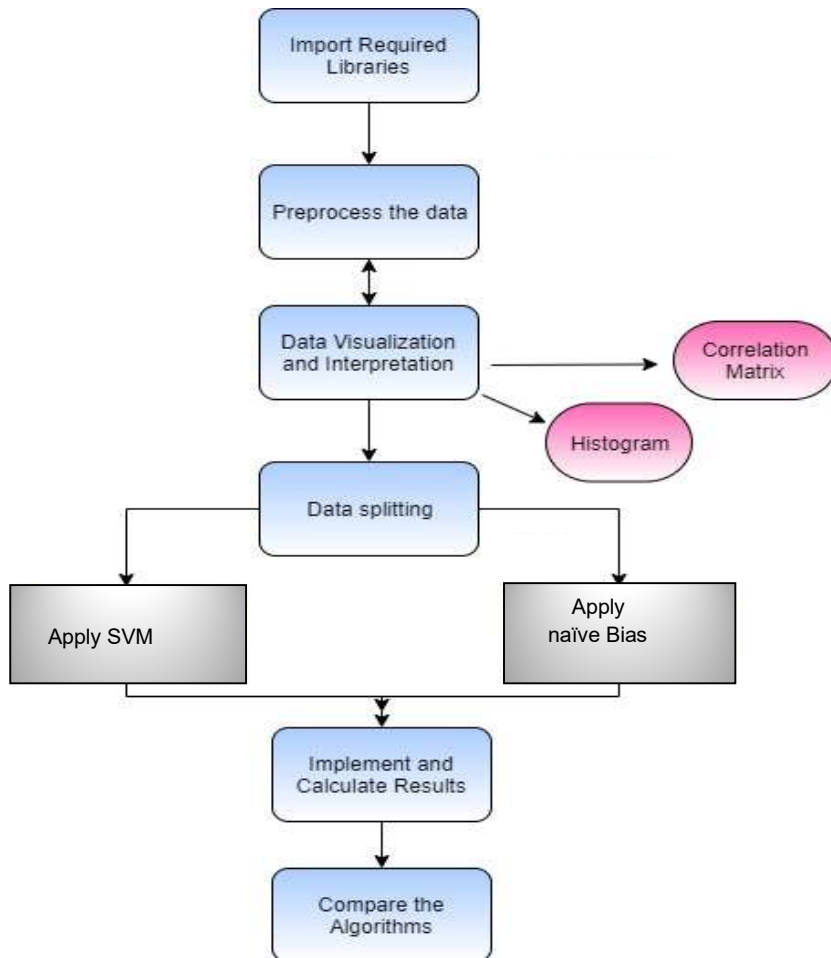
Dataset Information

Original Dataset

id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
842302	M	17.99	10.38	122.80	1001.0	0.11840
842517	M	20.57	17.77	132.90	1326.0	0.08474
84300903	M	19.69	21.25	130.00	1203.0	0.10960
84348301	M	11.42	20.38	77.58	386.1	0.14250
84358402	M	20.29	14.34	135.10	1297.0	0.10030

Preprocessed Dataset

| smoothness_worst	compactness_worst	concavity_worst	concave points_worst	symmetry_worst	fractal_dimension_worst	diagnosis
0.16220	0.66560	0.7119	0.2654	0.4601	0.11890	1
0.12380	0.18660	0.2416	0.1860	0.2750	0.08902	1
0.14440	0.42450	0.4504	0.2430	0.3613	0.08758	1
0.20980	0.86630	0.6869	0.2575	0.6638	0.17300	1
0.13740	0.20500	0.4000	0.1625	0.2364	0.07678	1

Dataset Interpretations

Histograms
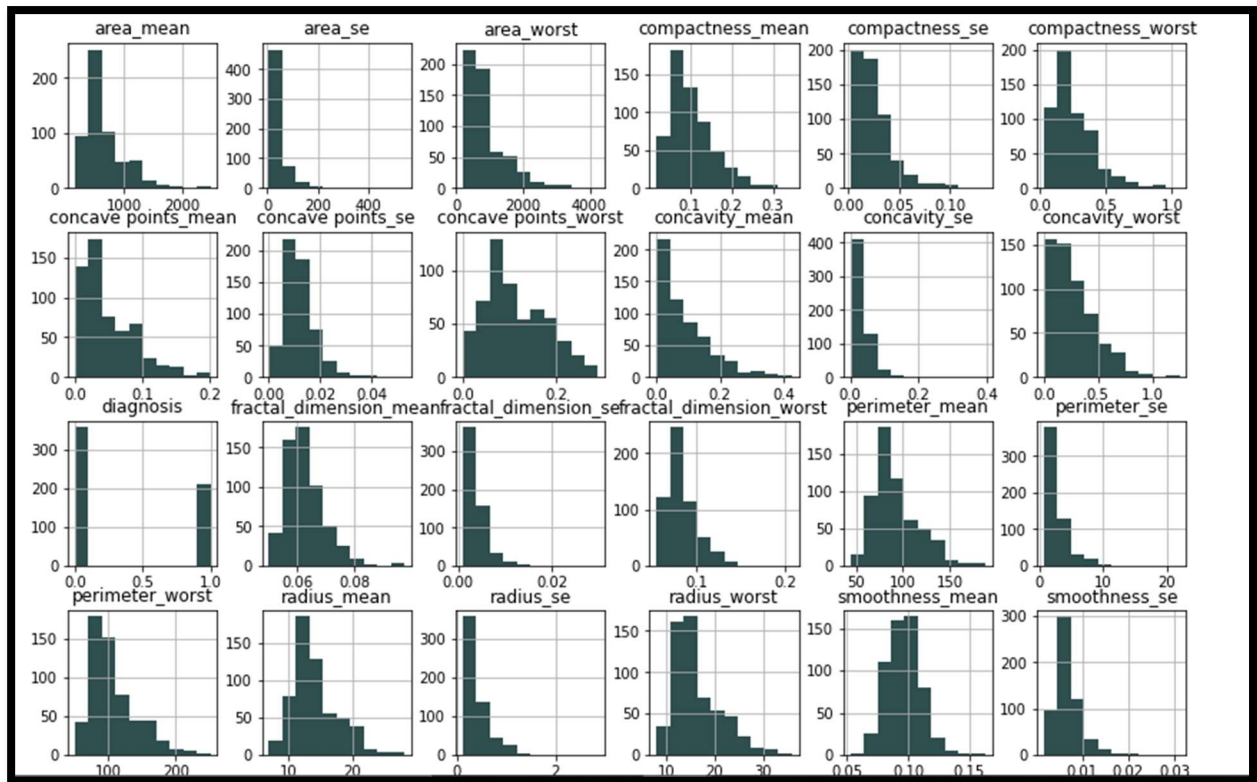
Target Classes
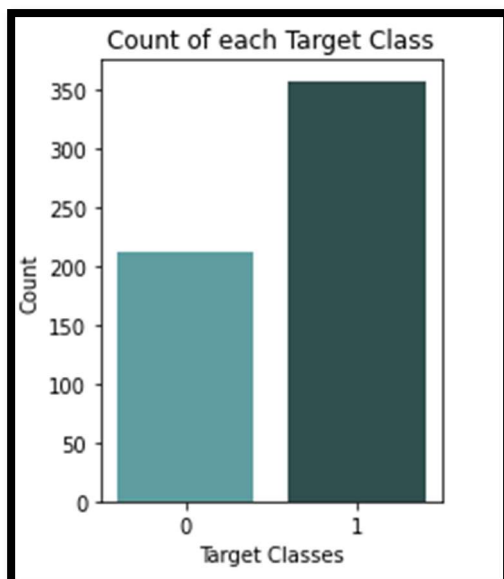
Correlation Matrix
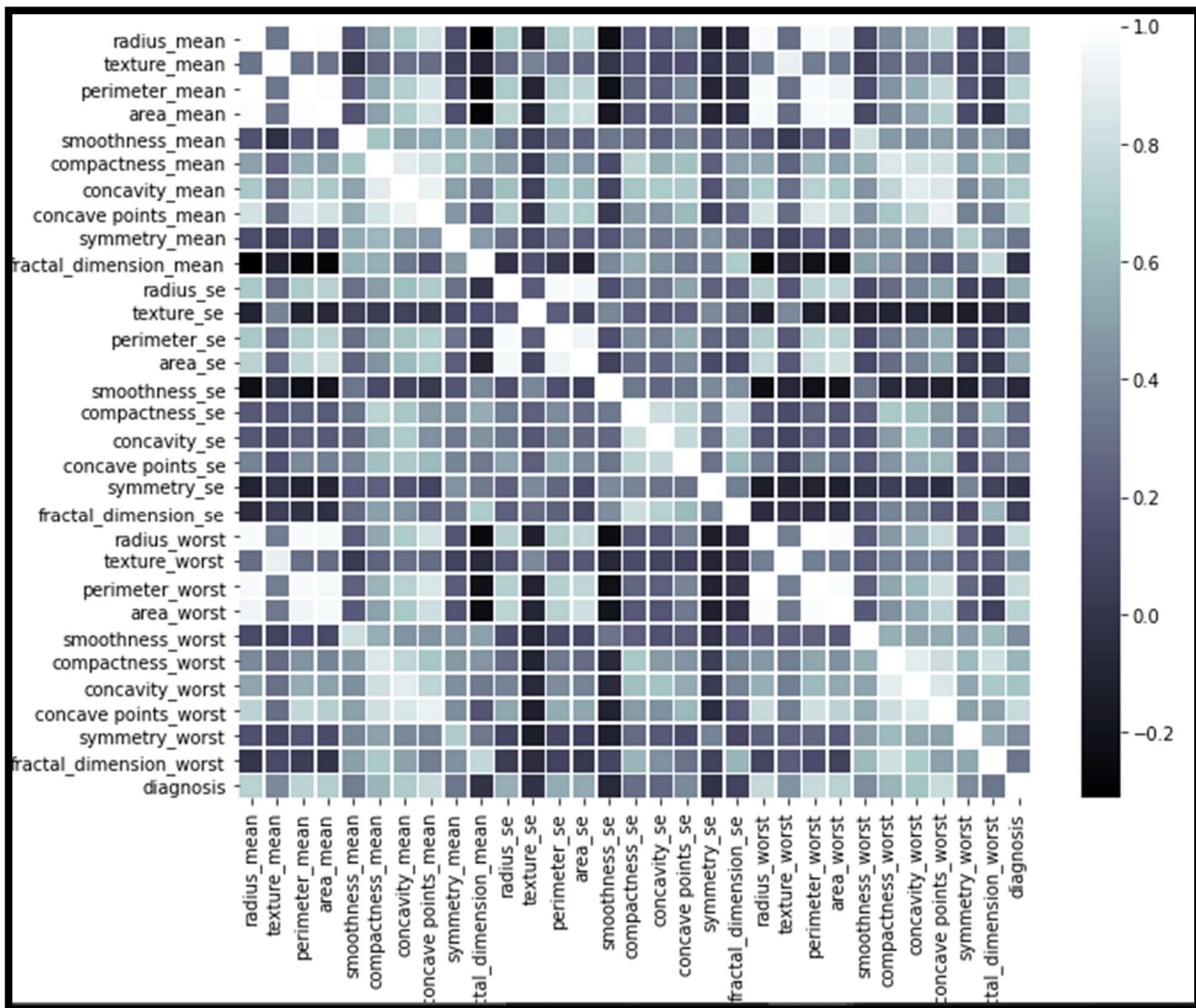
Support Vector Machine

Classification report and Confusion Matrix
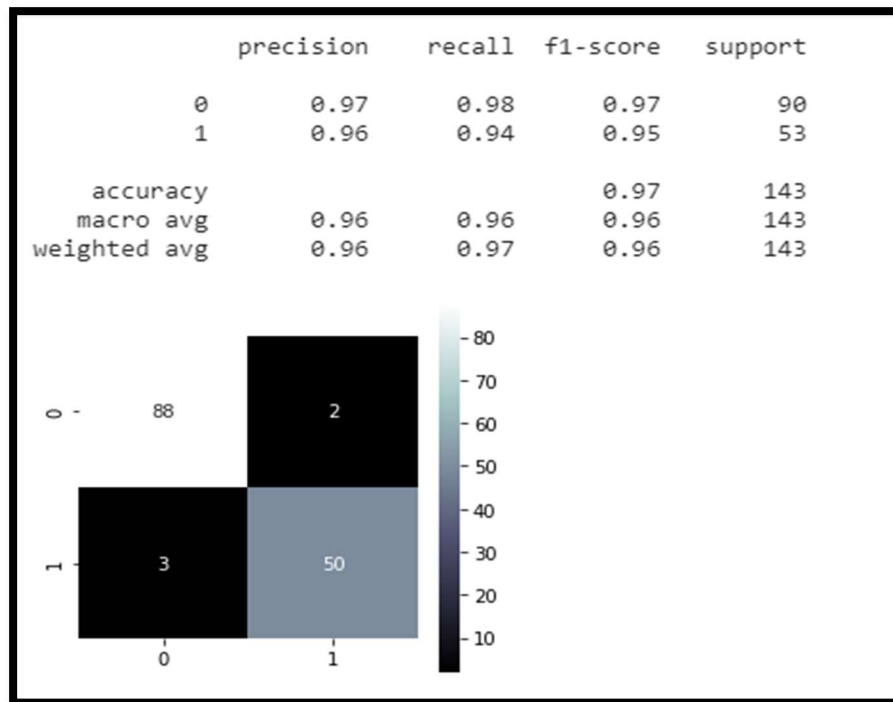
```
                 precision    recall  f1-score   support

             0       0.97      0.98      0.97        90
             1       0.96      0.94      0.95        53

      accuracy                           0.97       143
     macro avg       0.96      0.96      0.96       143
  weighted avg       0.96      0.97      0.96       143
```


Reciever Operating Curve
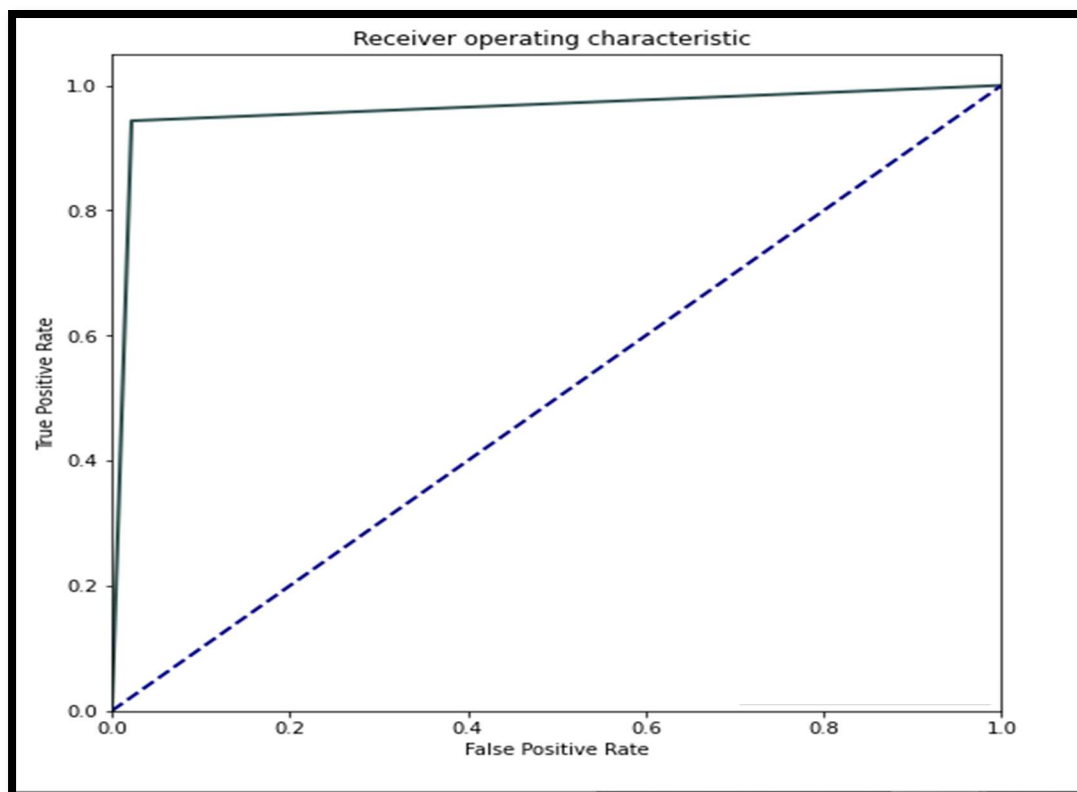
Results in Tabular Form

	Algorithm and Measures	75:25
0	Specificity	0.977778
1	Sensitivity	0.943396
2	Accuracy	0.965035
3	Precision	0.961538
4	FPR	0.022222
5	FNR	0.056604
6	NPV	0.967033
7	FDR	0.038462
8	F1-Score	0.952381
9	MCC	0.924865

Results in Graphical Form

Support Vector
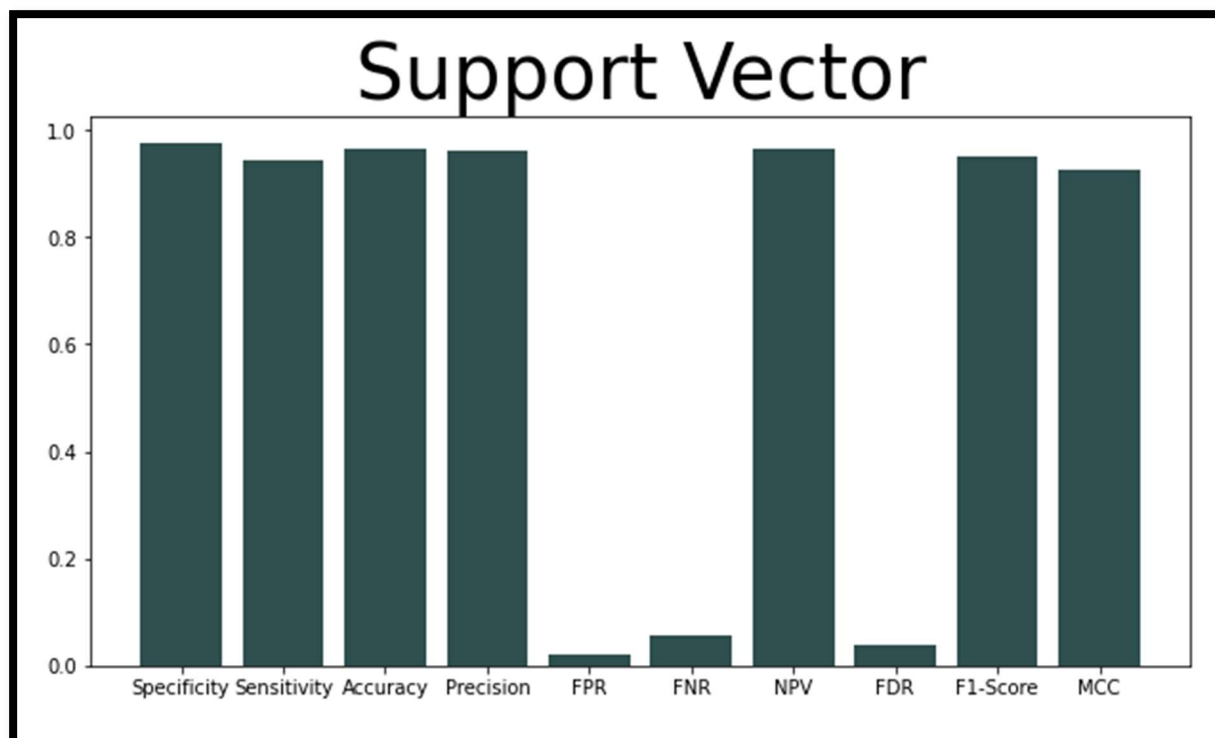
Naïve Bayes Algorithm

Classification report and Confusion Matrix
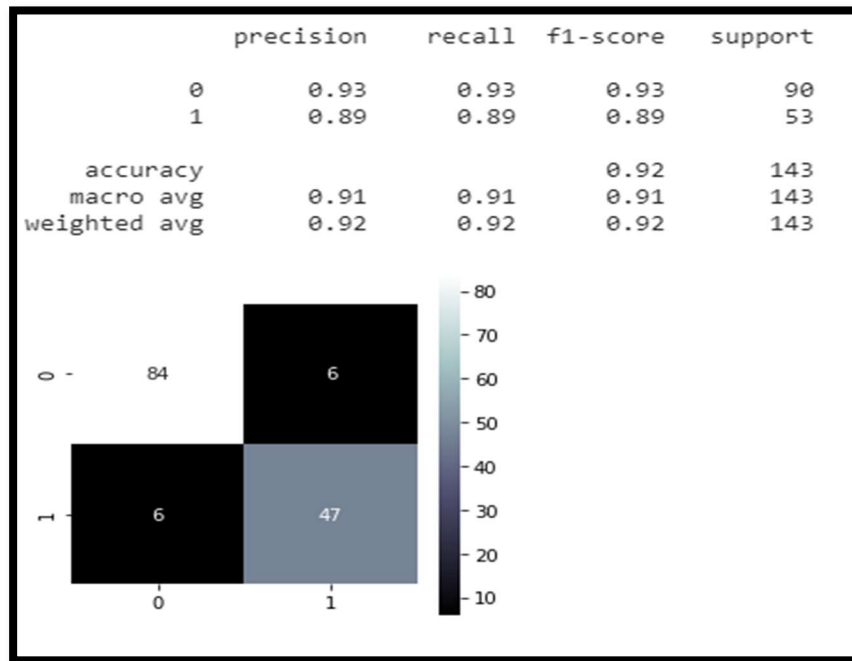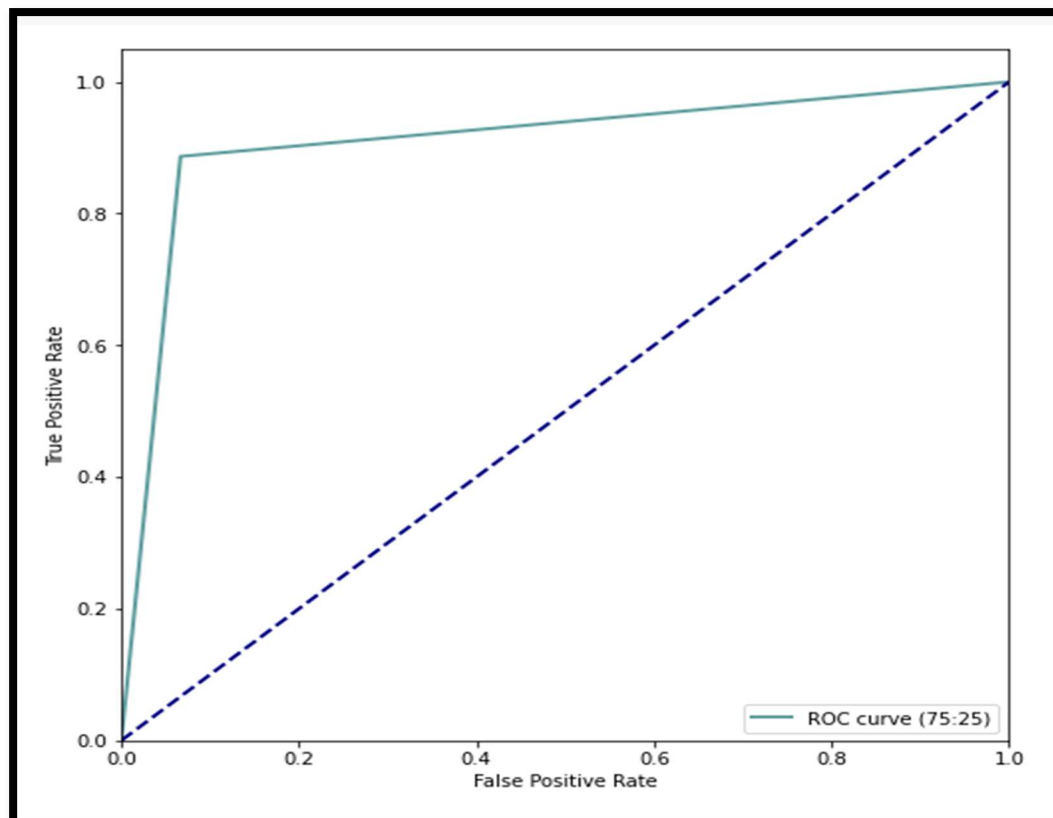
```
              precision    recall  f1-score   support

           0       0.93      0.93      0.93        90
           1       0.89      0.89      0.89        53

    accuracy                           0.92       143
   macro avg       0.91      0.91      0.91       143
weighted avg       0.92      0.92      0.92       143
```


Reciever Operating Curve

Results in Tabular Form Results in Tabular Form

Results in Tabular Form

	Algorithm and Measures	75:25
0	Specificity	0.933333
1	Sensitivity	0.886792
2	Accuracy	0.916084
3	Precision	0.886792
4	FPR	0.066667
5	FNR	0.113208
6	NPV	0.933333
7	FDR	0.113208
8	F1-Score	0.886792
9	MCC	0.820126

Results in Graphical Form
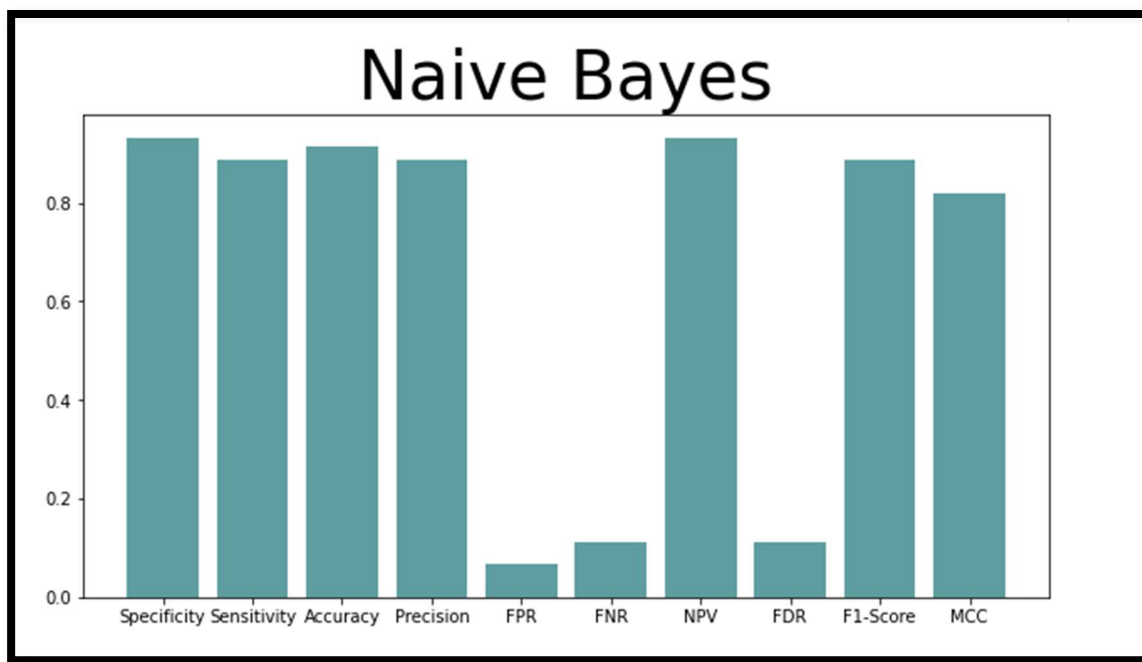
Naïve Bayes vs Support Vector Machine

Results in Tabular Form

	Algorithm and Measures	SVM	Naive Bayes
0	Specificity	0.977778	0.933333
1	Sensitivity	0.943396	0.886792
2	Accuracy	0.965035	0.916084
3	Precision	0.961538	0.886792
4	FPR	0.022222	0.066667
5	FNR	0.056604	0.113208
6	NPV	0.967033	0.933333
7	FDR	0.038462	0.113208
8	F1-Score	0.952381	0.886792
9	MCC	0.924865	0.820126

Results in Graphical Form