NAME – SOURAV SAHA
ROLL No – 001811001027
DEPARTMENT – INFORMATION TECHNOLOGY
YEAR – 4th Year
SEMESTER – 1st Semester
SUBJECT – Machine Learning LAB EVALUATION

Pre-requisite –
1. WINE DATASET
2. IONOSPHERE DATASET
3. DIABETES DATASET
4. IRIS DATASET

Question 1:

For Wine Dataset

a. SVM Classifier

```
SVC Linear:
Confusion Matrix
[[15  0  0]
 [ 1 19  0]
 [ 0  0 19]]
------------------------------------
------------------------------------
Preformance Evaluation:
          precision    recall  f1-score   support

        1       0.94      1.00      0.97        15
        2       1.00      0.95      0.97        20
        3       1.00      1.00      1.00        19

  accuracy                           0.98        54
 macro avg       0.98      0.98      0.98        54
weighted avg     0.98      0.98      0.98        54

------------------------------------
------------------------------------
Accuracy Score:
```
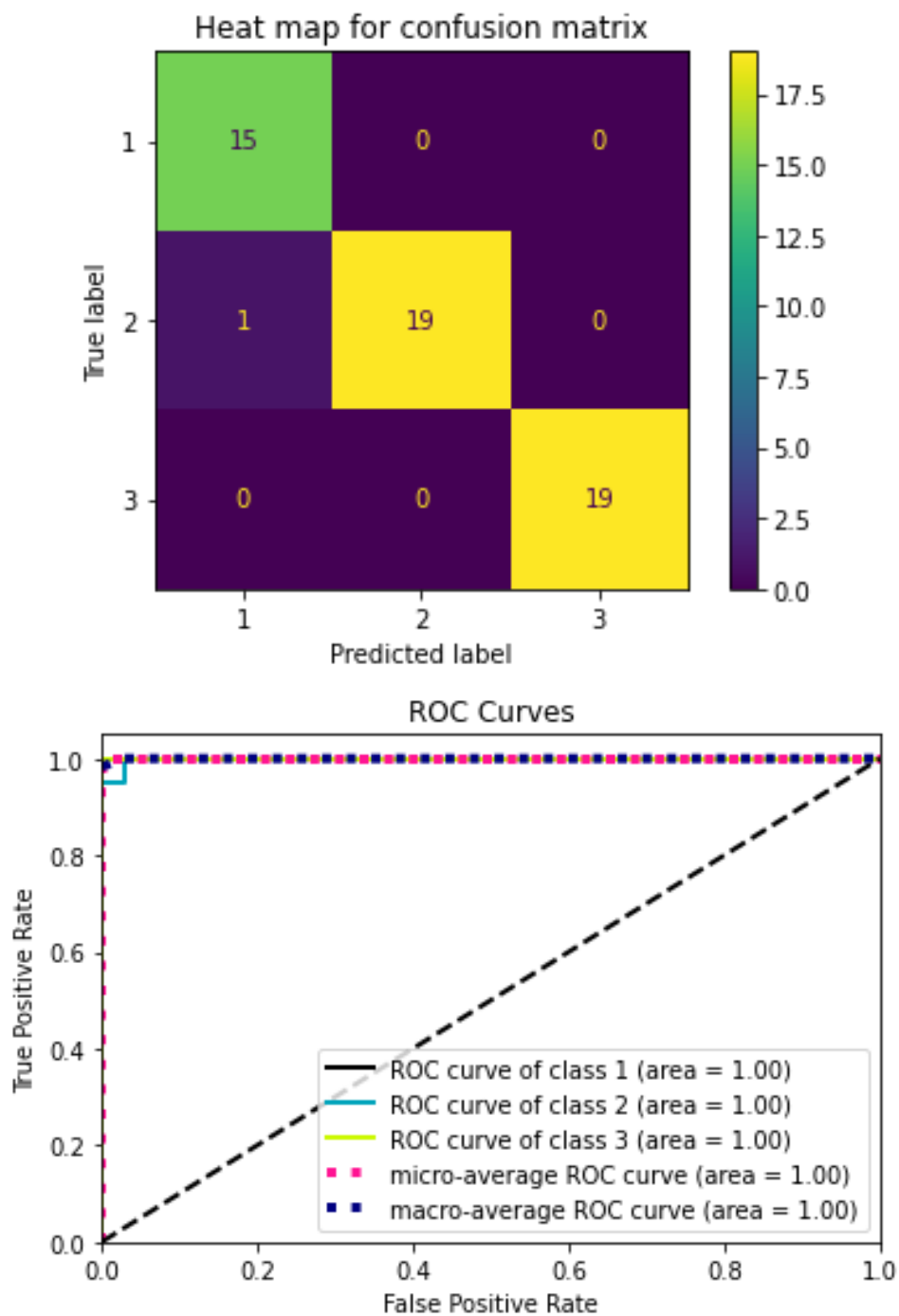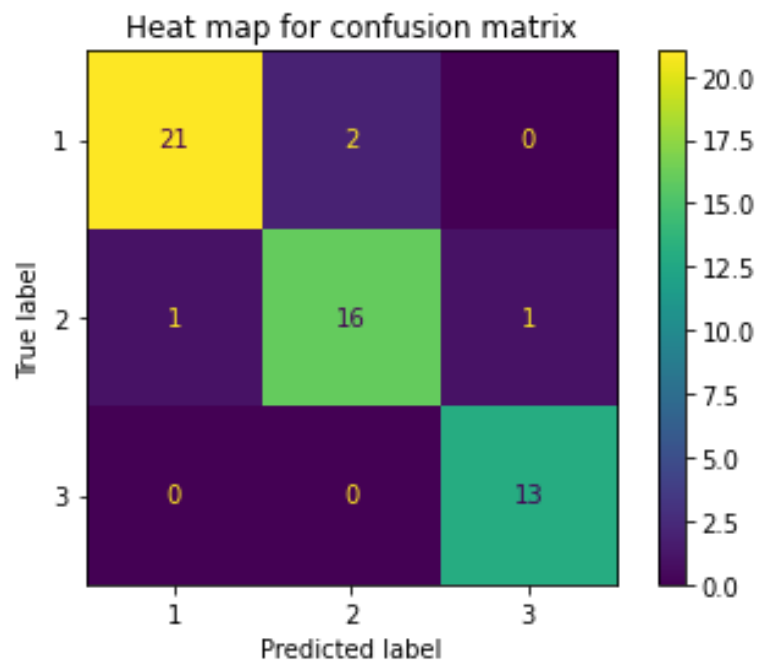
0.9814814814814815

Heat map for confusion matrix

ROC Curves

b. Decision Tree

```
Decision Tree Classifier:
Confusion Matrix
[[21  2  0]
 [ 1 16  1]
 [ 0  0 13]]
------------------------------------
------------------------------------
```
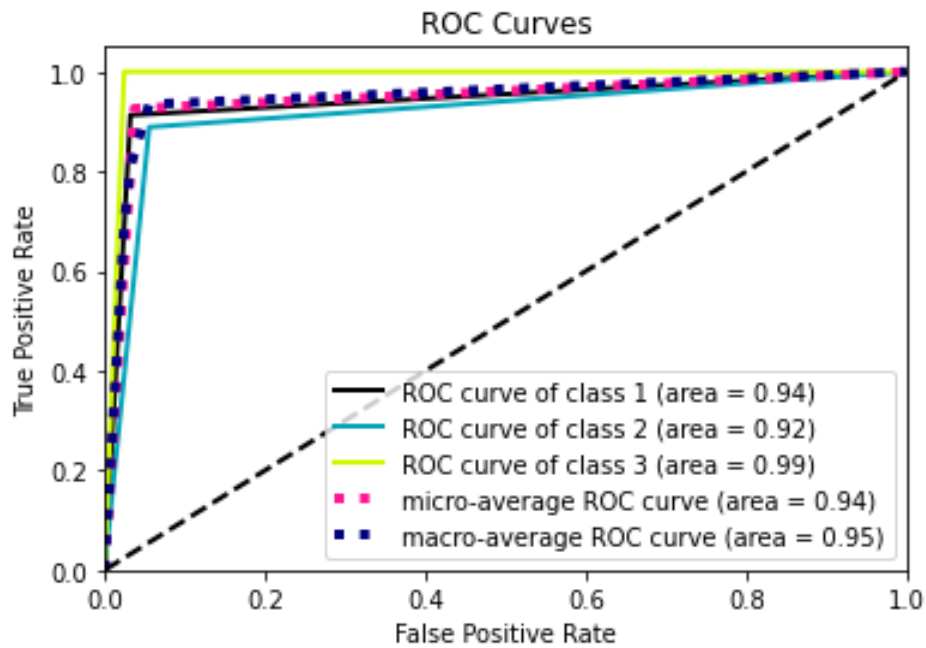
```
Preformance Evaluation:
              precision     recall   f1-score     support

           1        0.95       0.91       0.93          23
           2        0.89       0.89       0.89          18
           3        0.93       1.00       0.96          13

    accuracy                              0.93          54
   macro avg        0.92       0.93       0.93          54
weighted avg        0.93       0.93       0.93          54


--------------------------------------
--------------------------------------
Accuracy Score:
0.9259259259259259
```
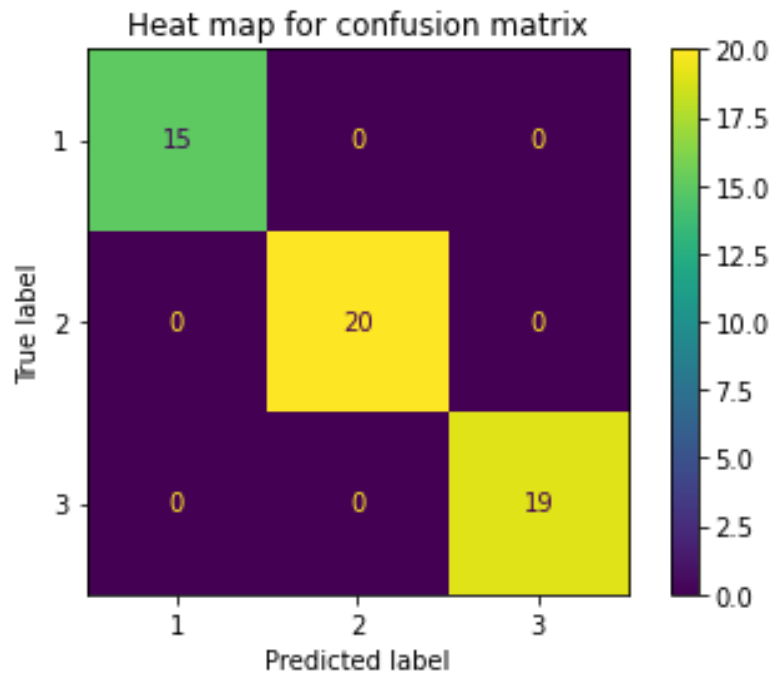

Heat map for confusion matrix

ROC Curves

c. Random Forest

```
Random Forest:
Confusion Matrix
[[15  0  0]
 [ 0 20  0]
 [ 0  0 19]]
------------------------------------
------------------------------------
Preformance Evaluation:
           precision    recall  f1-score   support

         1       1.00      1.00      1.00        15
         2       1.00      1.00      1.00        20
         3       1.00      1.00      1.00        19

  accuracy                           1.00        54
 macro avg       1.00      1.00      1.00        54
weighted avg     1.00      1.00      1.00        54


------------------------------------
------------------------------------
Accuracy Score:
1.0
```

Heat map for confusion matrix

d. Naïve Bayes
a. Multinomial

```
Multinomial Naive Bayes:
Confusion Matrix
[[17  2  4]
 [ 2 14  2]
 [ 0  2 11]]
------------------------------------
------------------------------------
Preformance Evaluation:
              precision   recall  f1-score   support

           1       0.89     0.74      0.81        23
           2       0.78     0.78      0.78        18
           3       0.65     0.85      0.73        13

    accuracy                          0.78        54
   macro avg       0.77     0.79      0.77        54
weighted avg       0.80     0.78      0.78        54


------------------------------------
------------------------------------
Accuracy Score:
0.7777777777777778
```
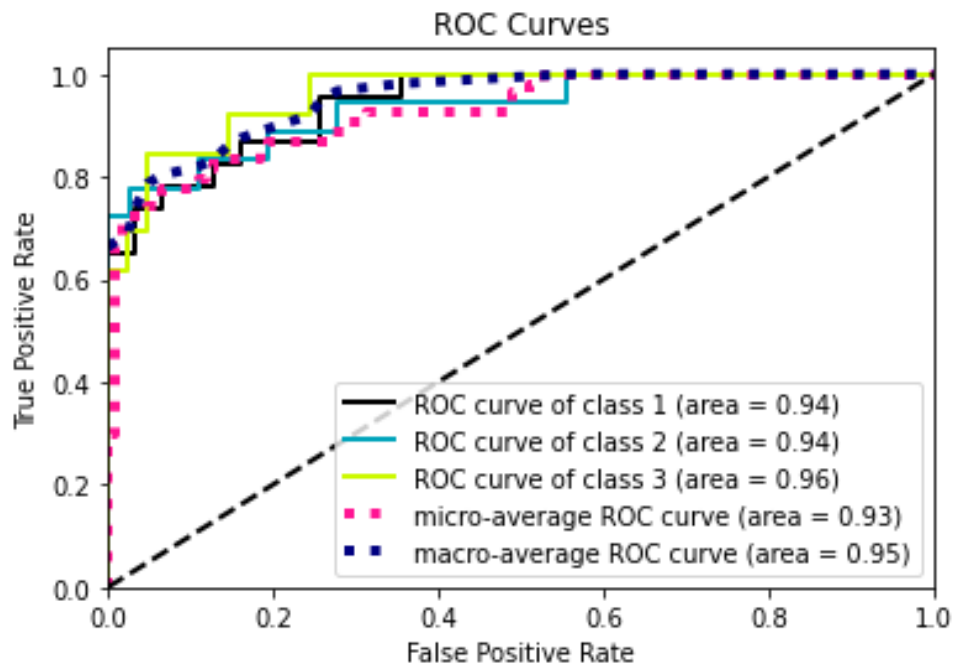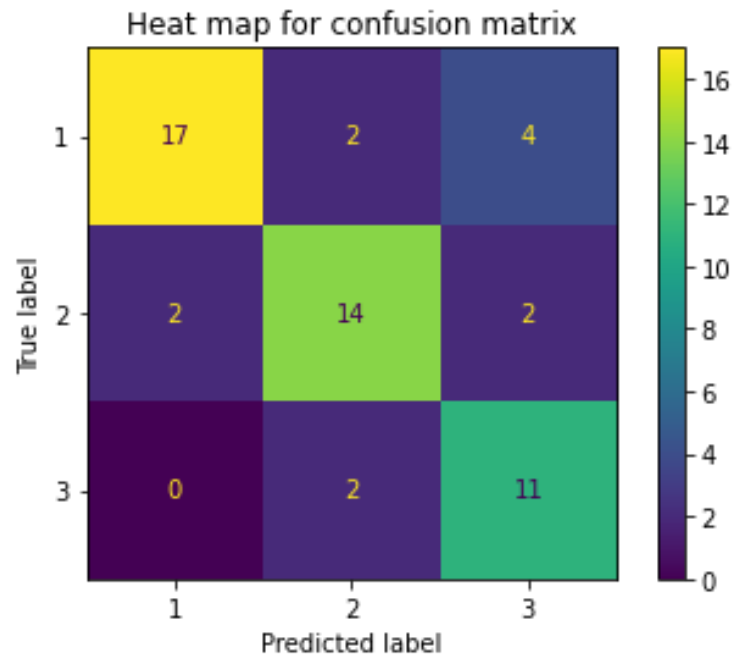
Heat map for confusion matrix

ROC Curves

b. Bernoulli

```
Bernoli Naive Bayes:
Confusion Matrix
[[ 0 23  0]
 [ 0 18  0]
 [ 0 13  0]]
-----------------------------------
-----------------------------------
Preformance Evaluation:
            precision    recall  f1-score    support
```
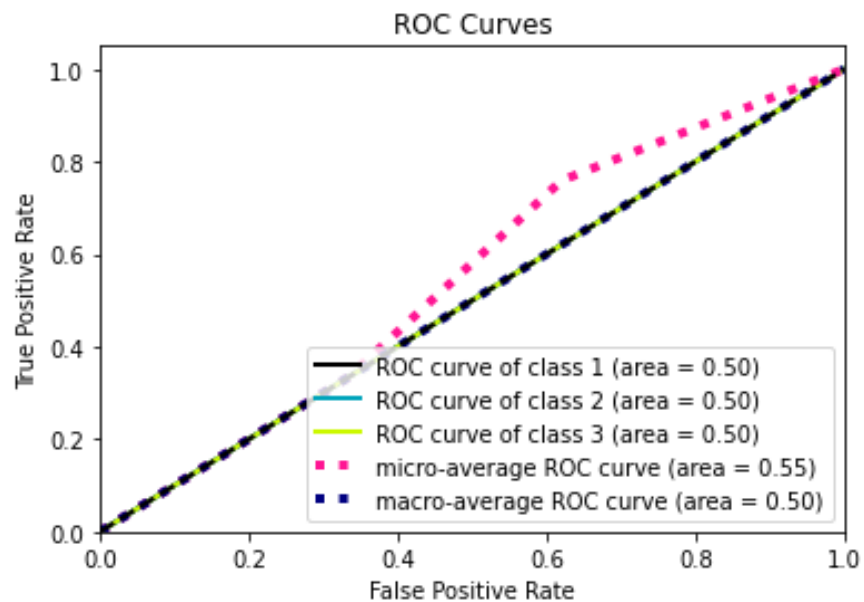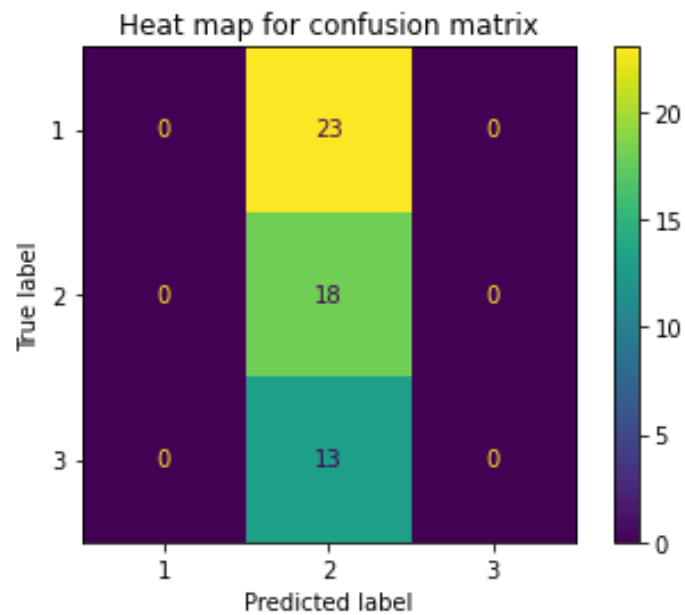
```
            1        0.00        0.00        0.00        23
            2        0.33        1.00        0.50        18
            3        0.00        0.00        0.00        13

     accuracy                                0.33        54
    macro avg        0.11        0.33        0.17        54
 weighted avg        0.11        0.33        0.17        54
```


Accuracy Score:
0.3333333333333333

Heat map for confusion matrix

ROC Curves

c. Gaussian

Gaussian Naive Bayes:

```
Confusion Matrix
[[20  3  0]
 [ 0 18  0]
 [ 0  0 13]]
-----------------------------------
-----------------------------------
Preformance Evaluation:
            precision    recall  f1-score   support

         1       1.00      0.87      0.93        23
         2       0.86      1.00      0.92        18
         3       1.00      1.00      1.00        13

  accuracy                           0.94        54
 macro avg       0.95      0.96      0.95        54
weighted avg     0.95      0.94      0.94        54


-----------------------------------
-----------------------------------
Accuracy Score:
0.9444444444444444
```
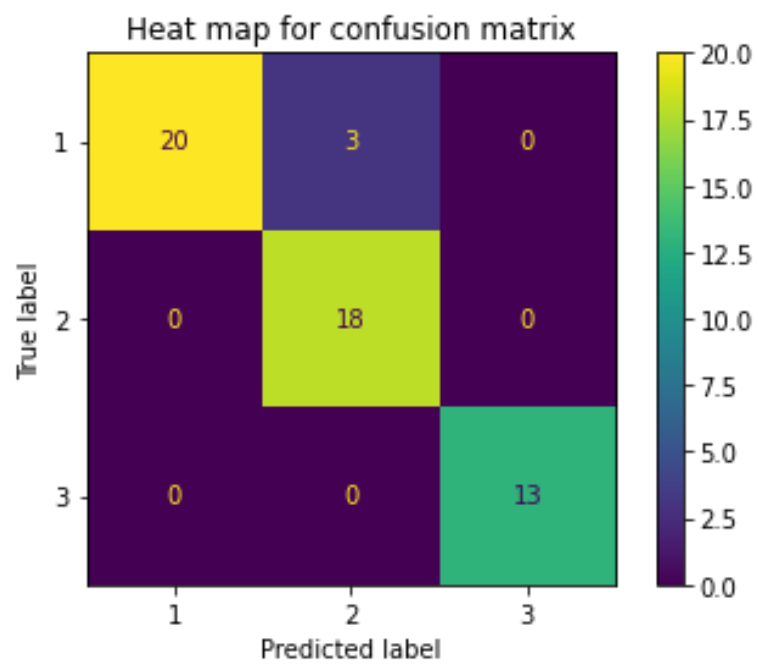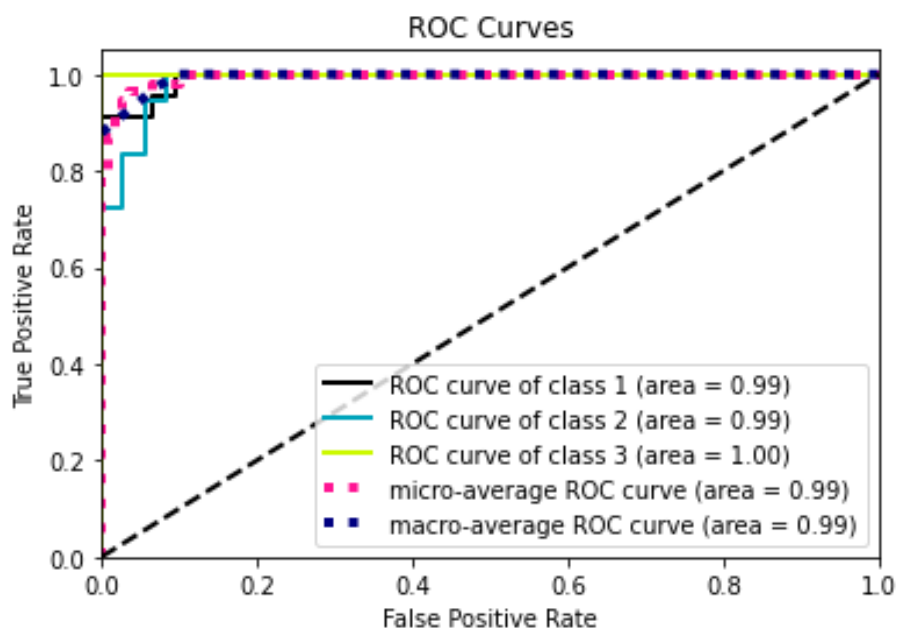

Heat map for confusion matrix

Figure: ROC Curves

For IONOSPHERE Dataset

a. SVM Classifier

```
SVC Linear:
Confusion Matrix
[[24 11]
 [ 5 66]]
-------------------------------------
-------------------------------------
Preformance Evaluation:
            precision    recall  f1-score   support

         b       0.83      0.69      0.75        35
         g       0.86      0.93      0.89        71

  accuracy                           0.85       106
 macro avg       0.84      0.81      0.82       106
weighted avg     0.85      0.85      0.85       106


-------------------------------------
-------------------------------------
Accuracy Score:
0.8490566037735849
```
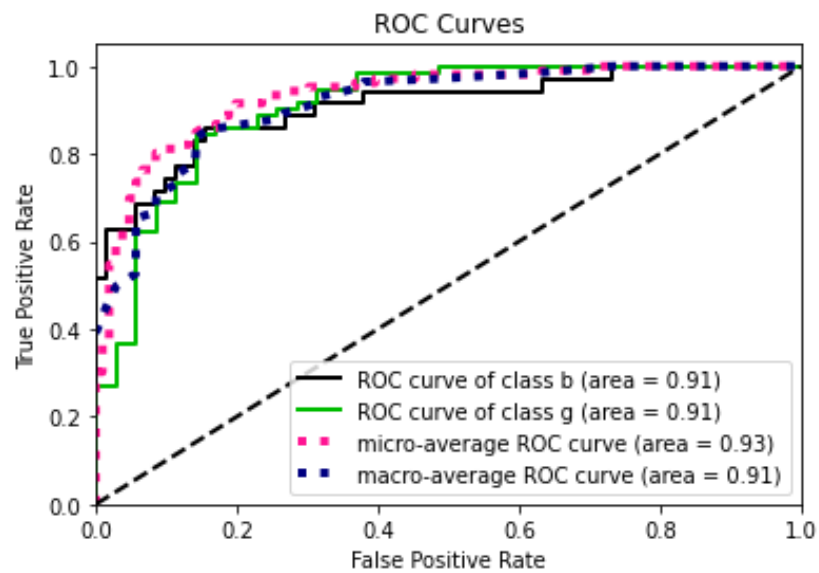
Heat map for confusion matrix

ROC Curves
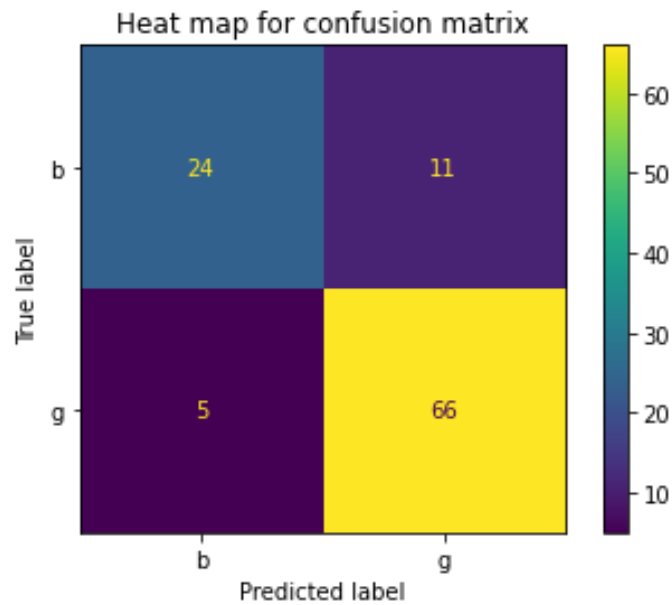
b. Decision Tree

```
Decision Tree Classifier:
Confusion Matrix
[[27  8]
 [ 4 67]]
------------------------------------
------------------------------------
Preformance Evaluation:
            precision      recall  f1-score     support

        b        0.87        0.77      0.82          35
```

g	0.89	0.94	0.92	71
accuracy			0.89	106
macro avg	0.88	0.86	0.87	106
weighted avg	0.89	0.89	0.88	106

```
------------------------------------
------------------------------------
```
Accuracy Score:
0.8867924528301887

Heat map for confusion matrix

ROC Curves
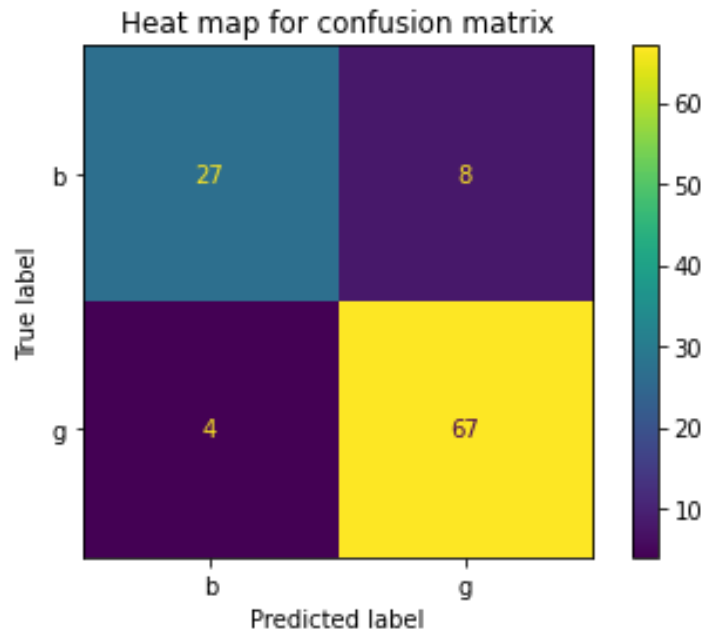
c. Random Forest

```
Random Forest:
Confusion Matrix
[[30  5]
 [ 0 71]]
-------------------------------------
-------------------------------------
Preformance Evaluation:
            precision      recall  f1-score      support

         b       1.00        0.86      0.92           35
         g       0.93        1.00      0.97           71

  accuracy                             0.95          106
 macro avg       0.97        0.93      0.94          106
weighted avg     0.96        0.95      0.95          106


-------------------------------------
-------------------------------------
Accuracy Score:
0.9528301886792453
```
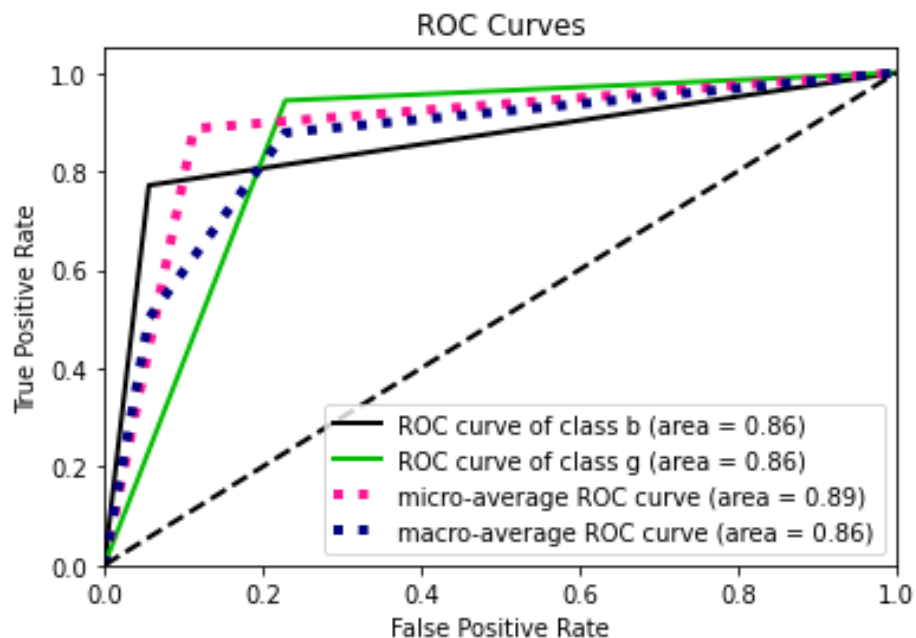

Heat map for confusion matrix

ROC Curves
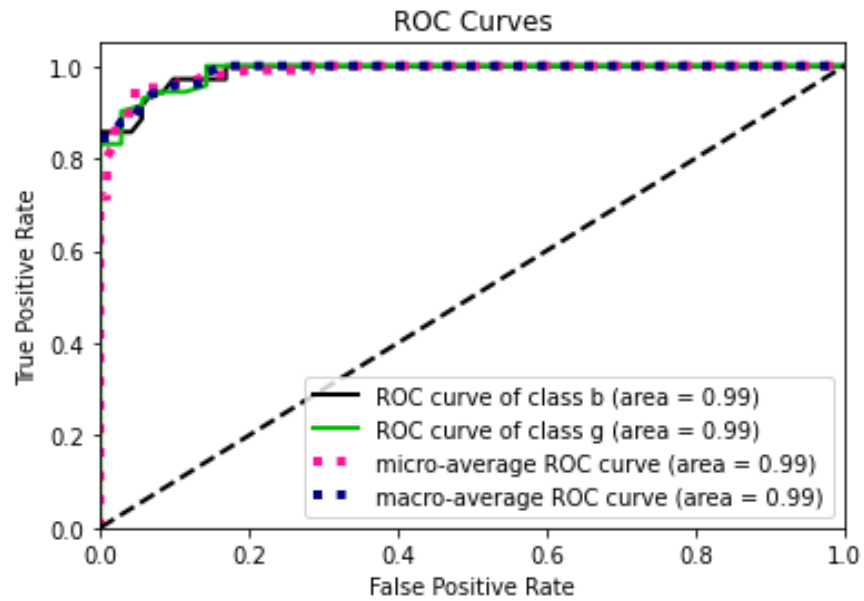
d. Naïve Bayes
a. Multinomial

```
Multinomial Naive Bayes:
Confusion Matrix
[[ 9 26]
 [ 3 68]]
------------------------------------
------------------------------------
Preformance Evaluation:
             precision    recall  f1-score   support

          b       0.75      0.26      0.38        35
          g       0.72      0.96      0.82        71

   accuracy                           0.73       106
  macro avg       0.74      0.61      0.60       106
weighted avg      0.73      0.73      0.68       106


------------------------------------
------------------------------------
Accuracy Score:
0.7264150943396226
```
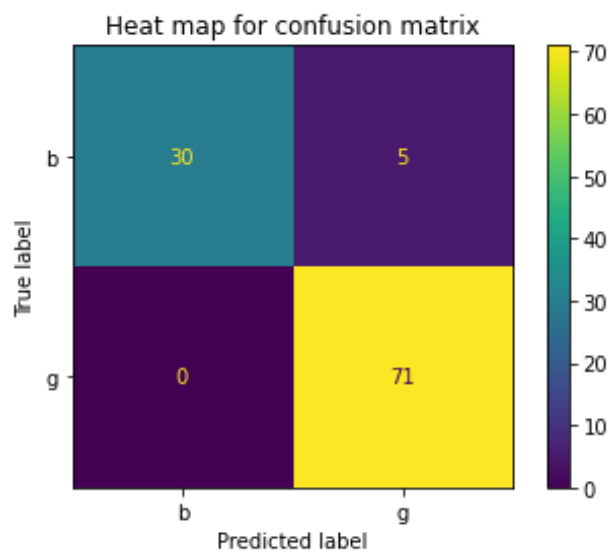
Heat map for confusion matrix

ROC Curves
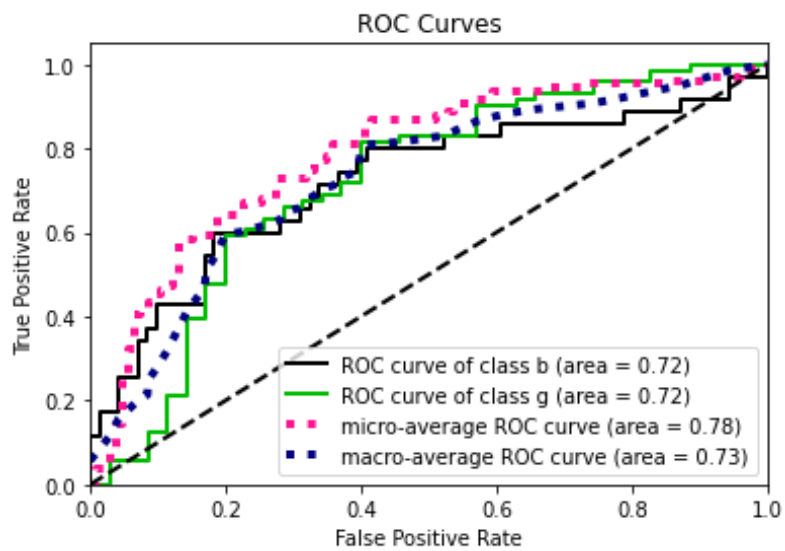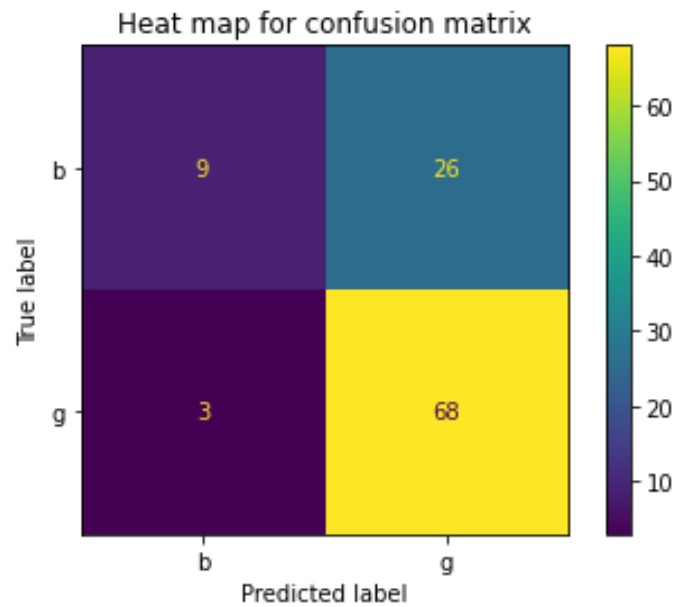
b. Bernoulli

```
Bernoulli Naive Bayes:
Confusion Matrix
[[26  9]
 [ 5 66]]
------------------------------------
------------------------------------
Preformance Evaluation:
            precision    recall  f1-score    support
```

	precision	recall	f1-score	support
b	0.84	0.74	0.79	35
g	0.88	0.93	0.90	71
accuracy			0.87	106
macro avg	0.86	0.84	0.85	106
weighted avg	0.87	0.87	0.87	106

```
------------------------------------
------------------------------------
Accuracy Score:
0.8679245283018868
```


Heat map for confusion matrix

ROC Curves
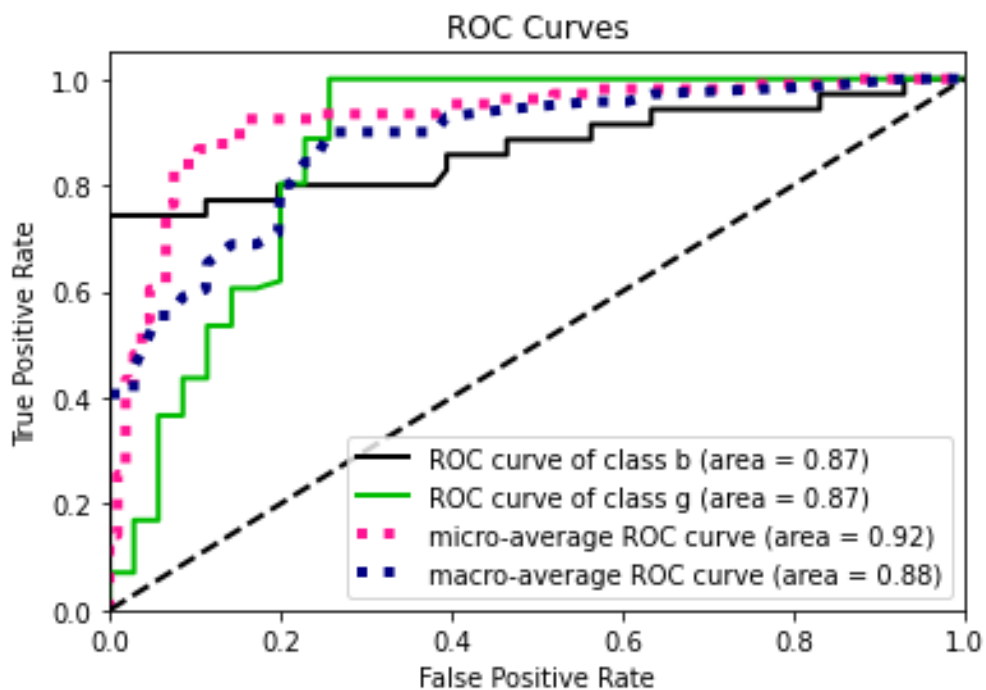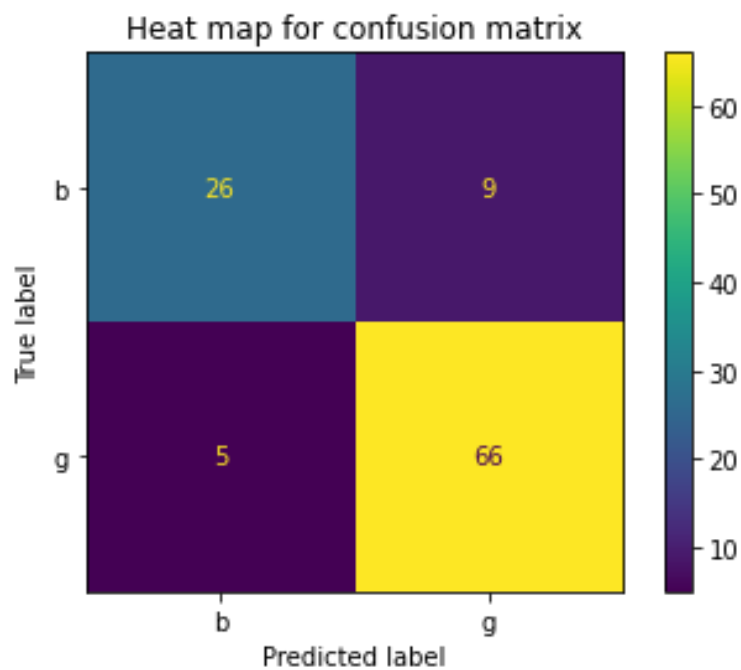
Gaussian

```
Gaussian Naive Bayes:
Confusion Matrix
[[26  9]
 [ 1 70]]
------------------------------------
------------------------------------
Preformance Evaluation:
          precision     recall  f1-score   support

       b       0.96       0.74      0.84        35
       g       0.89       0.99      0.93        71

accuracy                           0.91       106
macro avg       0.92       0.86      0.89       106
weighted avg    0.91       0.91      0.90       106


------------------------------------
------------------------------------
Accuracy Score:
0.9056603773584906
```


Heat map for confusion matrix

ROC Curves
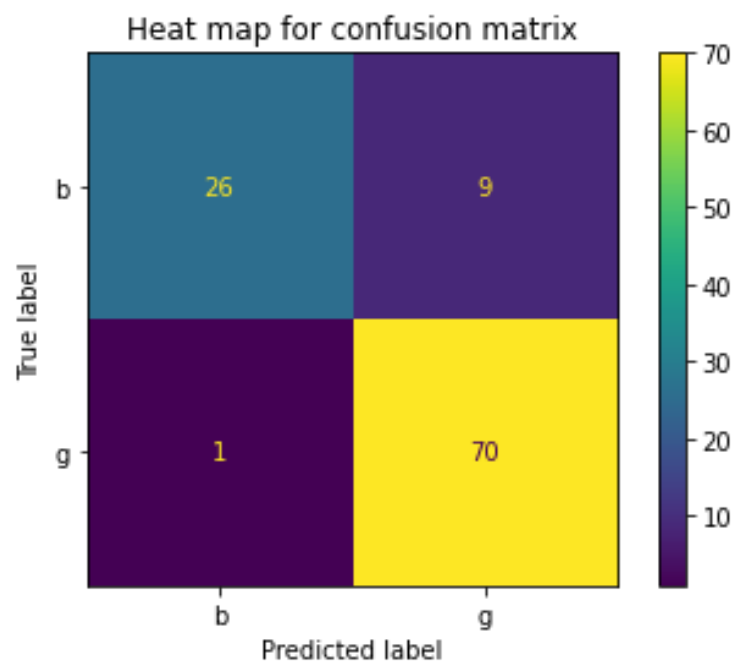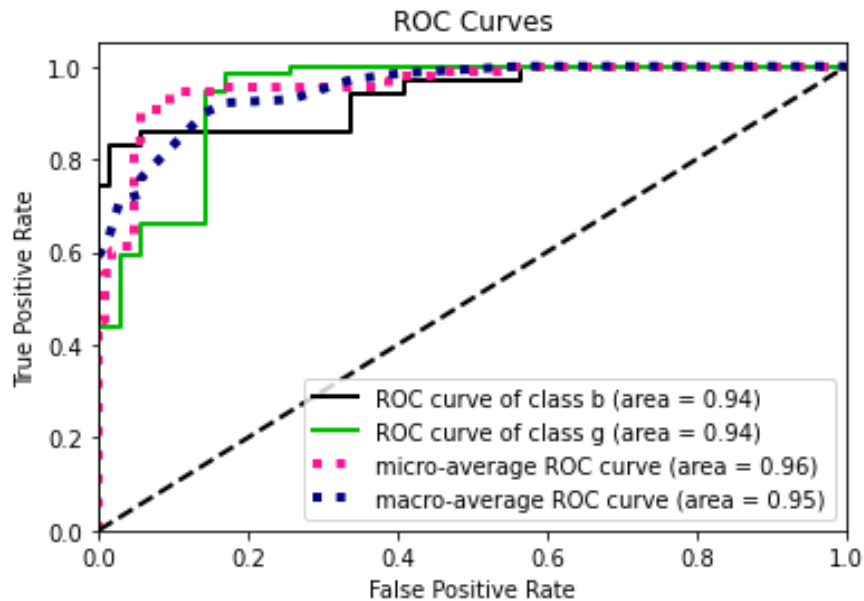
CONCLUSION –

For both WINE dataset and IONOSOHERE dataset, Random Forest Classifier had more accuracy than the others. (Values Marked Bold in the table)

	DATASET	CLASSIFIER	PRECISION	RECALL	FI-SCORE	SUPPORT	ACCURACY
		QUESTION 1					
4	DATASET	CLASSIFIER	PRECISION	RECALL	FI-SCORE	SUPPORT	ACCURACY
5		SVM	0.98	0.98	0.98	54	0.9814
6		Decision Tree	0.92	0.93	0.93	54	0.9259
7		Random Forest	1	1	1	54	**1**
8	WINE DATASET	Naive Bayes -					
9		Multinomial	0.77	0.79	0.77	54	0.7777
10		Bernoulli	0.11	0.33	0.17	54	0.3333
11		Gaussian	0.95	0.96	0.95	54	0.9444
12							
13		SVM	0.84	0.81	0.82	106	0.849
14		Decision Tree	0.88	0.86	0.87	106	0.8867
15		Random Forest	0.97	0.93	0.95	106	**0.9528**
16	IONOS- PHERE DATASET	Naive Bayes -					
17		Multinomial	0.74	0.61	0.6	106	0.7264
18		Bernoulli	0.86	0.84	0.85	106	0.8679
19		Gaussian	0.92	0.86	0.89	106	0.9056
20							

Question 5:

a. K-means versus, K-medoids/PAM

a.1 K-means

```
Performance Evaluation:
------------------------------------
------------------------------------
Silhouette Coefficient
0.571138193786884
------------------------------------
Calinski Harabasz Score
561.815657860671
------------------------------------
Davies Bouldin Score
0.5342431775436273
------------------------------------
```

a.2 K-medoids

```
Performance Evaluation:
------------------------------------
------------------------------------
Silhouette Coefficient
0.5708303868116225
------------------------------------
Calinski Harabasz Score
556.1459974410649
```

```
--------------------------------
Davies Bouldin Score
0.5316801818576816
--------------------------------
```

COMPARISON

QUESTION 5

Performance Measures	K Means	K Medoids
Silhouette Coefficient	0.571138	0.570830387
Calinski-Harabasz Score	561.8157	556.1459974
Davies-Bouldin Score	0.534243	0.531680182

b. DBSCAN versus OPTICS

b.1 DBSCAN

```
Performance Evaluation:
-----------------------------------
-----------------------------------
Silhouette Coefficient
0.5131593970763382
-----------------------------------
Calinski Harabasz Score
55.59856582586847
-----------------------------------
Davies Bouldin Score
0.37396418544796095
-----------------------------------
```

b.2 OPTICS

```
Performance Evaluation:
-----------------------------------
-----------------------------------
Silhouette Coefficient
0.2654566747731084
-----------------------------------
Calinski Harabasz Score
28.198952246515542
-----------------------------------
Davies Bouldin Score
5.752797762329204
-----------------------------------
```

COMPARISON

	Performance Measures	DBSCAN	OPTICS
	Silhouette Coefficient	0.513159	0.265456675
	Calinski-Harabasz Score	55.59857	28.19895225
	Davies-Bouldin Score	0.373964	5.752797762

Question 3:

a. For IRIS dataset

1. Gaussian HMM

```
Confusion Matrix
[[13   1   0]
 [ 0  18   0]
 [ 0  13   0]]
------------------------------------
------------------------------------
Performance Evaluation:
          precision    recall  f1-score   support

       0       1.00      0.93      0.96        14
       1       0.56      1.00      0.72        18
```

```
              2        0.00        0.00        0.00          13

       accuracy                                0.69          45
      macro avg        0.52        0.64        0.56          45
   weighted avg        0.54        0.69        0.59          45


------------------------------------
------------------------------------
Accuracy Score:
0.6888888888888889
```

2. GMMHMM

```
Confusion Matrix
[[18  1  0]
 [ 0 15  0]
 [ 0 11  0]]
------------------------------------
------------------------------------
Performance Evaluation:
              precision     recall   f1-score     support

              0        1.00        0.95        0.97          19
              1        0.56        1.00        0.71          15
              2        0.00        0.00        0.00          11

       accuracy                                0.73          45
      macro avg        0.52        0.65        0.56          45
   weighted avg        0.61        0.73        0.65          45


------------------------------------
------------------------------------
Accuracy Score:
0.7333333333333333
```

b. For Diabetes Dataset

1. Gaussian

```
Confusion Matrix
[[46 26]
 [20 41]]
----------------------------------
----------------------------------
Performance Evaluation:
              precision     recall   f1-score     support

              0        0.70        0.64        0.67          72
              1        0.61        0.67        0.64          61
```

```
       accuracy                          0.65       133
      macro avg        0.65     0.66     0.65       133
   weighted avg        0.66     0.65     0.65       133

-----------------------------------
-----------------------------------
Accuracy Score:
0.6541353383458647
```

2. GMMHMM

```
Confusion Matrix
[[45 30]
 [16 42]]
-----------------------------------
-----------------------------------
Performance Evaluation:
            precision    recall  f1-score   support

          0      0.74      0.60      0.66        75
          1      0.58      0.72      0.65        58

    accuracy                          0.65       133
   macro avg      0.66      0.66      0.65       133
weighted avg      0.67      0.65      0.65       133

-----------------------------------
-----------------------------------
Accuracy Score:
0.6541353383458647
```

c. For Ionosphere Dataset

1. Gaussian

```
Confusion Matrix
[[29  4]
 [23 50]]
-------------------------------------
-------------------------------------
Performance Evaluation:
            precision    recall  f1-score   support

          0      0.56      0.88      0.68        33
          1      0.93      0.68      0.79        73

    accuracy                          0.75       106
   macro avg      0.74      0.78      0.73       106
weighted avg      0.81      0.75      0.75       106

-------------------------------------
```

```
-----------------------------------
Accuracy Score:
0.7452830188679245
```

2. GMMHM

```
Confusion Matrix
[[34  4]
 [24 44]]
-----------------------------------
-----------------------------------
Performance Evaluation:
             precision    recall  f1-score   support

          0       0.59      0.89      0.71        38
          1       0.92      0.65      0.76        68

   accuracy                           0.74       106
  macro avg       0.75      0.77      0.73       106
weighted avg       0.80      0.74      0.74       106


-----------------------------------
-----------------------------------
Accuracy Score:
0.7358490566037735
```

QUESTION 3

DATASET	CLASSIFIER	PRECISION	RECALL	FI-SCORE	SUPPORT	ACCURACY
IRIS	Gaussian HMM	0.52	0.64	0.56	45	0.688
	GMMHMM	0.52	0.65	0.56	45	**0.733**
Diabetes	Gaussian HMM	0.65	0.66	0.65	133	**0.654**
	GMMHMM	0.66	0.66	0.65	133	**0.654**
IONOS-PHERE	Gaussian HMM	0.74	0.78	0.73	106	**0.745**
	GMMHMM	0.75	0.77	0.73	106	0.735

NOTE – Multinomial HMM was not working with the above dataset, ans gave errors regarding values < 0.

GITHUB LINK FOR ML LAB EVALUATION-
https://github.com/SouravSaha1999/ML-Lab-Evaluation

The repo contains PDF file, python notebook, excel sheet and required images used in this PDF.