Report

Classification task for Chocolate users –

SINGLE DECISION TREE –

- **CONFUSION METRICS –**

```
[[  0   6]
[  10 361]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.00	0.00	0.00	6
user	0.98	0.97	0.98	371
accuracy			0.96	377
macro avg	0.49	0.49	0.49	377
weighted	0.97	0.96	0.96	377

RANDOM FOREST –

- **CONFUSION METRICS –**

```
[[  0   6]
[  0 371]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.00	0.00	0.00	6
user	0.98	1.00	0.99	371
accuracy			0.98	377
macro avg	0.49	0.50	0.50	377
weighted	0.97	0.98	0.98	377

SUPPORT VECTOR MACHINE –

- **CONFUSION METRICS –**

```
[[  0   6]
[  0 371]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.00	0.00	0.00	6
user	0.98	1.00	0.99	371
accuracy			0.98	377
macro avg	0.49	0.50	0.50	377
weighted	0.97	0.98	0.98	377

GRADIENT BOOST ENSEMBLE –

- **CONFUSION METRICS –**

```
[[  0   6]
[  1 370]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.00	0.00	0.00	6
user	0.98	1.00	0.99	371
accuracy			0.98	377
macro avg	0.49	0.50	0.50	377
weighted	0.97	0.98	0.97	377

MULTI LAYER PERCEPTRON –

- **CONFUSION METRICS –**

```
[[  0   6]
[  2 369]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.00	0.00	0.00	6
user	0.98	0.99	0.99	371
accuracy			0.98	377
macro avg	0.49	0.50	0.49	377
weighted	0.97	0.98	0.97	377

K-NEAREST NEIGHNOR –

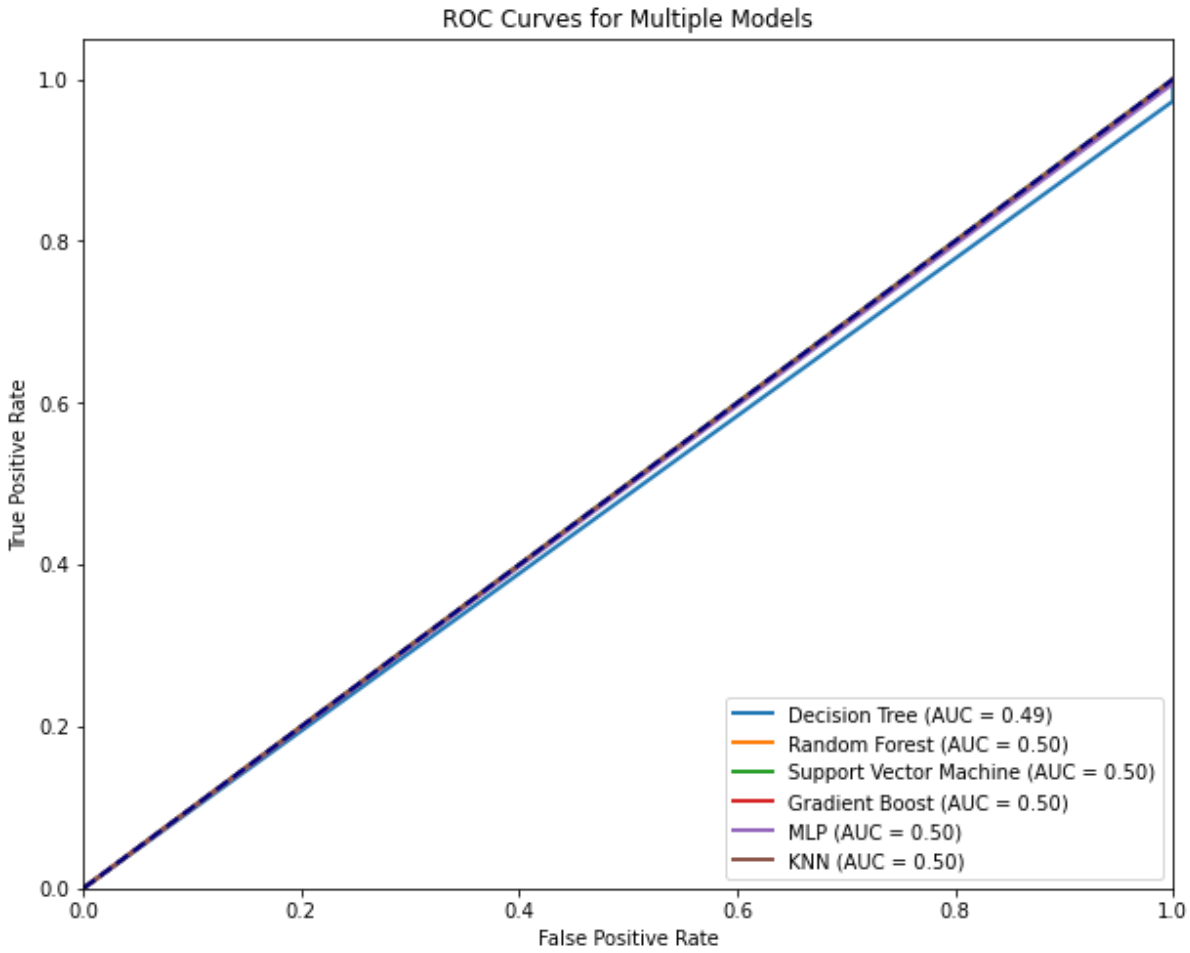
- **CONFUSION METRICS –**

```
[[  0   6]
[  0 371]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.00	0.00	0.00	6
user	0.98	1.00	0.99	371
accuracy			0.98	377
macro avg	0.49	0.50	0.50	377
weighted	0.97	0.98	0.98	377

ROC –


ROC Curves for Multiple Models
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Decision Tree (AUC = 0.49)
Random Forest (AUC = 0.50)
Support Vector Machine (AUC = 0.50)
Gradient Boost (AUC = 0.50)
MLP (AUC = 0.50)
KNN (AUC = 0.50)

Classification task for MUSHROOM users –

SINGLE DECISION TREE –

- **CONFUSION METRICS –**

```
[[ 179  75]
[  51 72]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.78	0.70	0.74	254
user	0.49	0.59	0.53	123
accuracy			0.67	377
macro avg	0.63	0.65	0.64	377
weighted	0.68	0.67	0.67	377

RANDOM FOREST –

- **CONFUSION METRICS –**

```
[[ 209  45]
[  45 78]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.82	0.82	0.82	254
user	0.63	0.63	0.63	123
accuracy			0.76	377
macro avg	0.73	0.73	0.73	377
weighted	0.76	0.76	0.76	377

SUPPORT VECTOR MACHINE –

- **CONFUSION METRICS –**

```
[[ 207   47]
[  38 85]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.84	0.81	0.83	254
user	0.64	0.69	0.67	123
accuracy			0.77	377
macro avg	0.74	0.75	0.75	377
weighted	0.78	0.77	0.78	377

GRADIENT BOOST ENSEMBLE –

- **CONFUSION METRICS –**

```
[[ 213  41]
[  41 82]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.84	0.84	0.84	254
user	0.67	0.67	0.67	123
accuracy			0.78	377
macro avg	0.75	0.75	0.75	377
weighted	0.78	0.78	0.78	377

MULTI LAYER PERCEPTRON –

- **CONFUSION METRICS –**

```
[[ 196  58]
[  45 78]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.81	0.77	0.79	254
user	0.57	0.63	0.60	123
accuracy			0.73	377
macro avg	0.69	0.70	0.70	377
weighted	0.74	0.73	0.73	377

K-NEAREST NEIGHNOR –

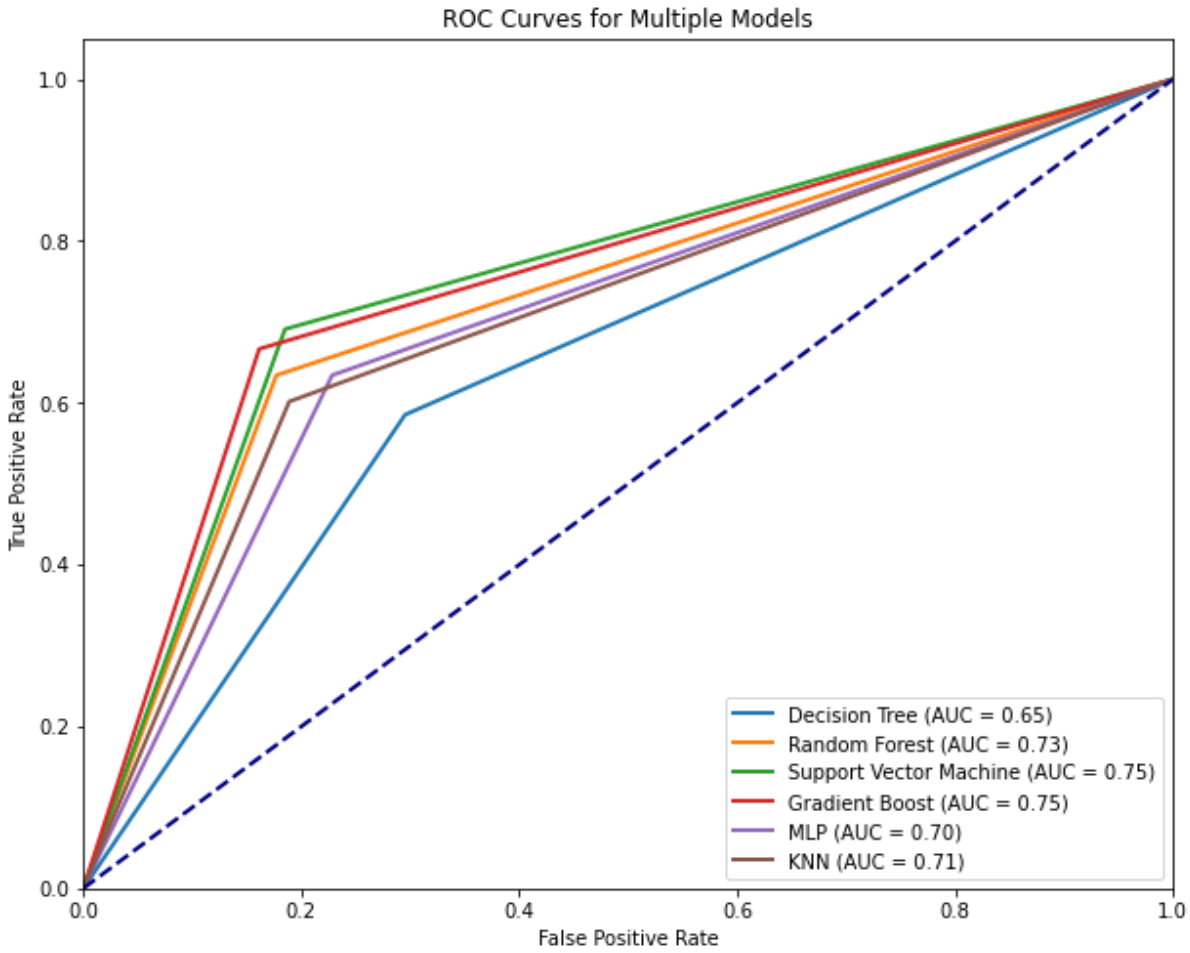
- **CONFUSION METRICS –**

```
[[ 206  48]
[  49 74]]
```

- **CLASSIFICATION REPORT (PRECISION, RECALL) –**

	precision	recall	f1-score	support
non-user	0.81	0.81	0.81	254
user	0.61	0.60	0.60	123
accuracy			0.74	377
macro avg	0.71	0.71	0.71	377
weighted	0.74	0.74	0.74	377

ROC –


ROC Curves for Multiple Models
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Decision Tree (AUC = 0.65)
Random Forest (AUC = 0.73)
Support Vector Machine (AUC = 0.75)
Gradient Boost (AUC = 0.75)
MLP (AUC = 0.70)
KNN (AUC = 0.71)

Data balancing for ‘Chocolate’ –

UNDERSAMPLED -

Model performance on dataset with accuracy:

```
Decision Tree: 50.00%
Random Forest: 50.00%
SVM: 64.29%
Gradient Boost: 57.14%
Multi Layer Perceptron: 85.71%
K-Nearest Neighbors: 57.14%
```

Model performance on dataset with confusion matrix:

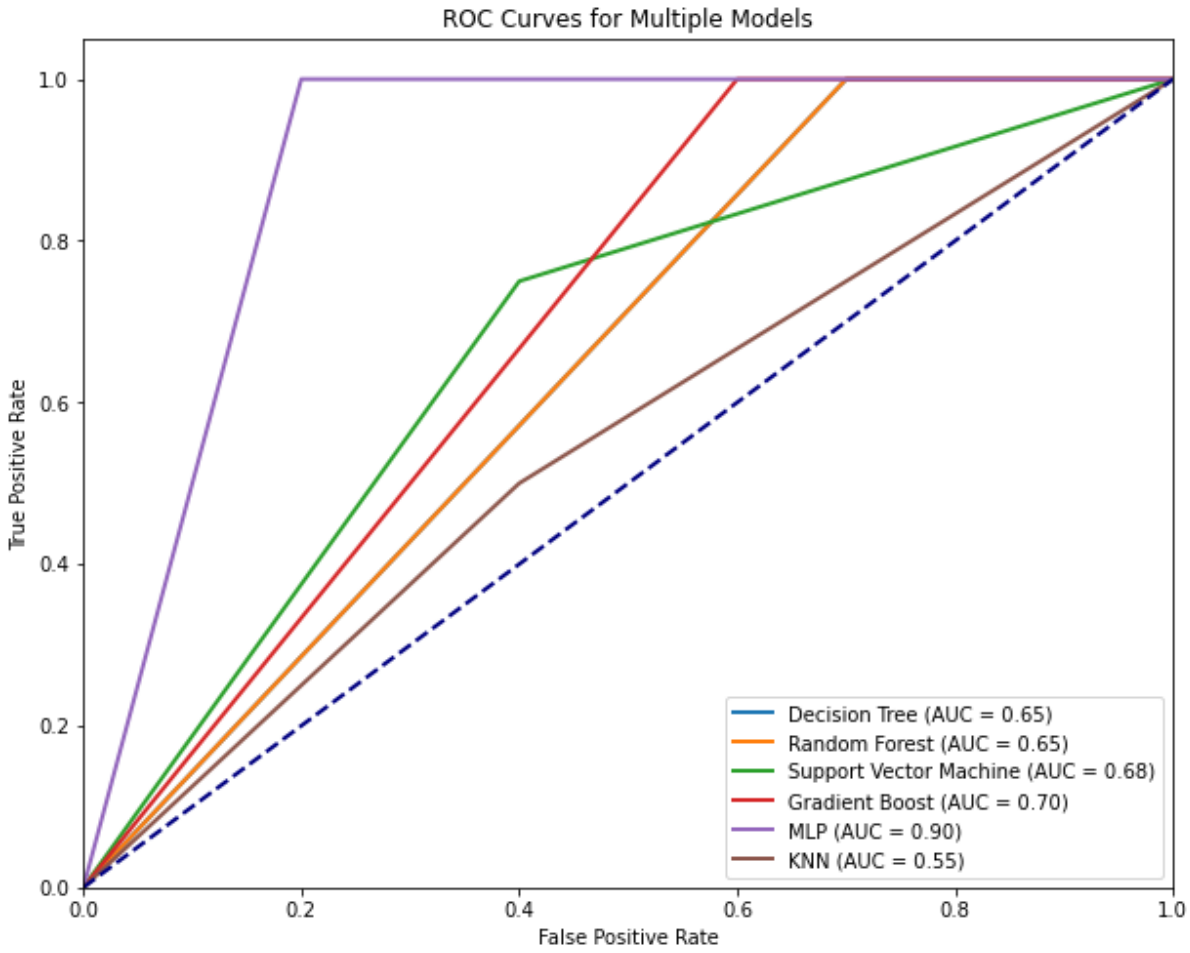
```
Confusion matrix for Decision Tree:
[[3 7]
 [0 4]]
Confusion matrix for Random Forest:
[[3 7]
 [0 4]]
Confusion matrix for SVM:
[[6 4]
 [1 3]]
Confusion matrix for Gradient Boost:
[[4 6]
 [0 4]]
Confusion matrix for Multi Layer Perceptron:
[[8 2]
 [0 4]]
Confusion matrix for K-Nearest Neighbors:
[[6 4]
 [2 2]]
```

Model performance on dataset Classification report:

```
Classification report for Decision Tree:
              precision    recall  f1-score   support

    non-user       1.00      0.30      0.46        10
        user       0.36      1.00      0.53         4

    accuracy                           0.50        14
   macro avg       0.68      0.65      0.50        14
weighted avg       0.82      0.50      0.48        14

Classification report for Random Forest:
              precision    recall  f1-score   support

    non-user       1.00      0.30      0.46        10
        user       0.36      1.00      0.53         4
```

```
    accuracy                           0.50        14
   macro avg       0.68      0.65      0.50        14
weighted avg       0.82      0.50      0.48        14

Classification report for SVM:
              precision    recall  f1-score   support

    non-user       0.86      0.60      0.71        10
        user       0.43      0.75      0.55         4

    accuracy                           0.64        14
   macro avg       0.64      0.68      0.63        14
weighted avg       0.73      0.64      0.66        14

Classification report for Gradient Boost:
              precision    recall  f1-score   support

    non-user       1.00      0.40      0.57        10
        user       0.40      1.00      0.57         4

    accuracy                           0.57        14
   macro avg       0.70      0.70      0.57        14
weighted avg       0.83      0.57      0.57        14

Classification report for Multi Layer Perceptron:
              precision    recall  f1-score   support

    non-user       1.00      0.80      0.89        10
        user       0.67      1.00      0.80         4

    accuracy                           0.86        14
   macro avg       0.83      0.90      0.84        14
weighted avg       0.90      0.86      0.86        14

Classification report for K-Nearest Neighbors:
              precision    recall  f1-score   support

    non-user       0.75      0.60      0.67        10
        user       0.33      0.50      0.40         4

    accuracy                           0.57        14
   macro avg       0.54      0.55      0.53        14
weighted avg       0.63      0.57      0.59        14
```

Model performance on dataset ROC curve:



OVERSAMPLED –

Model performance on dataset with accuracy:

```
Decision Tree: 97.16%
Random Forest: 99.73%
SVM: 71.22%
Gradient Boost: 98.78%
Multi Layer Perceptron: 98.78%
K-Nearest Neighbors: 91.08%
```

Model performance on dataset with confusion matrix:

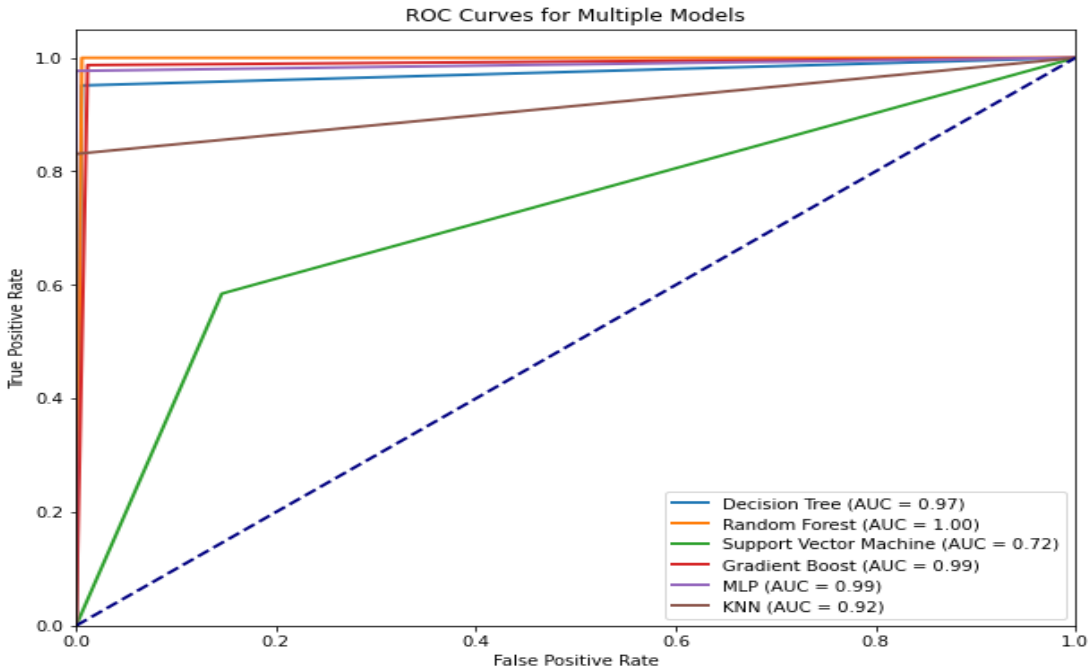
```
Confusion matrix for Decision Tree:
[[348   2]
 [ 19 371]]
Confusion matrix for Random Forest:
[[348   2]
 [  0 390]]
Confusion matrix for SVM:
[[299  51]
 [162 228]]
Confusion matrix for Gradient Boost:
[[346   4]
 [  5 385]]
Confusion matrix for Multi Layer Perceptron:
[[350   0]
 [  9 381]]
Confusion matrix for K-Nearest Neighbors:
[[350   0]
 [ 66 324]]
```

Model performance on dataset Classification report:

```
Classification report for Decision Tree:
              precision    recall  f1-score   support

    non-user       0.95      0.99      0.97       350
        user       0.99      0.95      0.97       390

    accuracy                           0.97       740
   macro avg       0.97      0.97      0.97       740
weighted avg       0.97      0.97      0.97       740

Classification report for Random Forest:
              precision    recall  f1-score   support

    non-user       1.00      0.99      1.00       350
        user       0.99      1.00      1.00       390

    accuracy                           1.00       740
   macro avg       1.00      1.00      1.00       740
weighted avg       1.00      1.00      1.00       740

Classification report for SVM:
              precision    recall  f1-score   support

    non-user       0.65      0.85      0.74       350
        user       0.82      0.58      0.68       390
```

```
    accuracy                           0.71       740
   macro avg       0.73      0.72      0.71       740
weighted avg       0.74      0.71      0.71       740

Classification report for Gradient Boost:
              precision    recall  f1-score   support

    non-user       0.99      0.99      0.99       350
        user       0.99      0.99      0.99       390

    accuracy                           0.99       740
   macro avg       0.99      0.99      0.99       740
weighted avg       0.99      0.99      0.99       740

Classification report for Multi Layer Perceptron:
              precision    recall  f1-score   support

    non-user       0.97      1.00      0.99       350
        user       1.00      0.98      0.99       390

    accuracy                           0.99       740
   macro avg       0.99      0.99      0.99       740
weighted avg       0.99      0.99      0.99       740

Classification report for K-Nearest Neighbors:
              precision    recall  f1-score   support

    non-user       0.84      1.00      0.91       350
        user       1.00      0.83      0.91       390

    accuracy                           0.91       740
   macro avg       0.92      0.92      0.91       740
weighted avg       0.92      0.91      0.91       740
```

Model performance on dataset ROC curve:



BALANCED –

Model performance on dataset with accuracy:

```
Decision Tree: 75.00%
Random Forest: 67.86%
SVM: 60.71%
Gradient Boost: 75.00%
Multi Layer Perceptron: 75.00%
K-Nearest Neighbors: 57.14%
```

Model performance on dataset with confusion matrix:

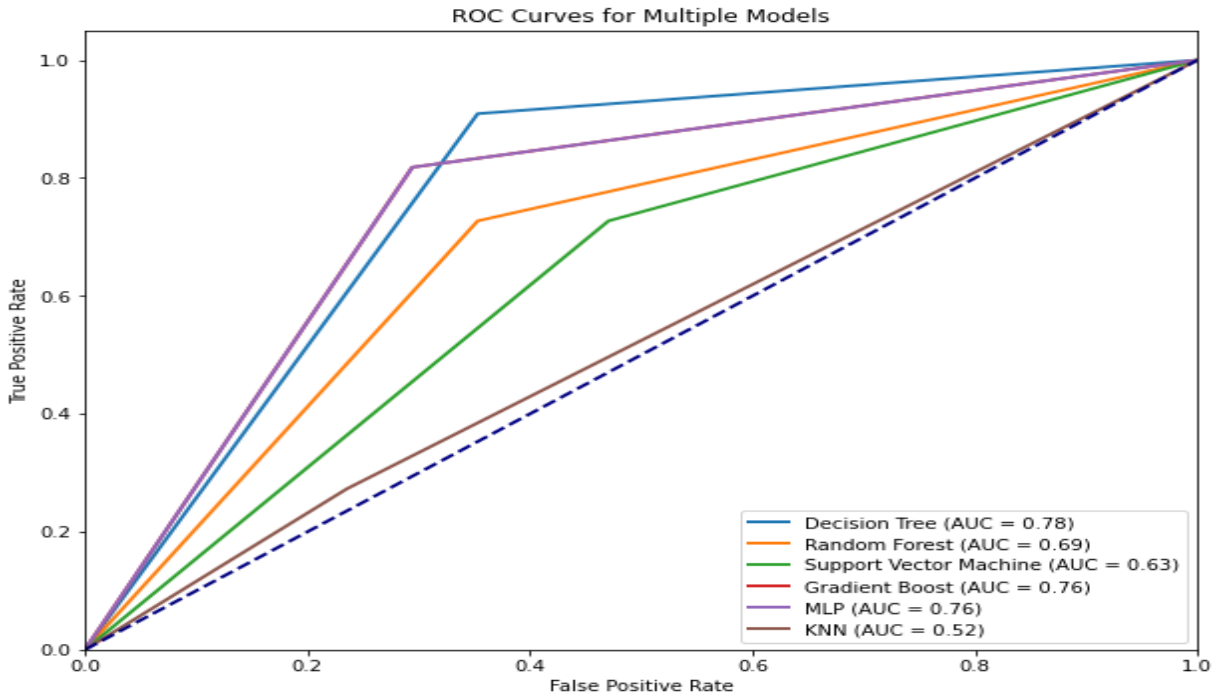
```
Confusion matrix for Decision Tree:
[[11  6]
 [ 1 10]]
Confusion matrix for Random Forest:
[[11  6]
 [ 3  8]]
Confusion matrix for SVM:
[[9 8]
 [3 8]]
Confusion matrix for Gradient Boost:
[[12  5]
 [ 2  9]]
Confusion matrix for Multi Layer Perceptron:
[[12  5]
 [ 2  9]]
Confusion matrix for K-Nearest Neighbors:
[[13  4]
 [ 8  3]]
```

Model performance on dataset Classification report:

```
Classification report for Decision Tree:
              precision    recall  f1-score   support

    non-user       0.92      0.65      0.76        17
        user       0.62      0.91      0.74        11

    accuracy                           0.75        28
   macro avg       0.77      0.78      0.75        28
weighted avg       0.80      0.75      0.75        28

Classification report for Random Forest:
              precision    recall  f1-score   support

    non-user       0.79      0.65      0.71        17
        user       0.57      0.73      0.64        11

    accuracy                           0.68        28
   macro avg       0.68      0.69      0.67        28
weighted avg       0.70      0.68      0.68        28

Classification report for SVM:
              precision    recall  f1-score   support

    non-user       0.75      0.53      0.62        17
        user       0.50      0.73      0.59        11
```

```
    accuracy                           0.61        28
   macro avg       0.62      0.63      0.61        28
weighted avg       0.65      0.61      0.61        28

Classification report for Gradient Boost:
              precision    recall  f1-score   support

    non-user       0.86      0.71      0.77        17
        user       0.64      0.82      0.72        11

    accuracy                           0.75        28
   macro avg       0.75      0.76      0.75        28
weighted avg       0.77      0.75      0.75        28

Classification report for Multi Layer Perceptron:
              precision    recall  f1-score   support

    non-user       0.86      0.71      0.77        17
        user       0.64      0.82      0.72        11

    accuracy                           0.75        28
   macro avg       0.75      0.76      0.75        28
weighted avg       0.77      0.75      0.75        28

Classification report for K-Nearest Neighbors:
              precision    recall  f1-score   support

    non-user       0.62      0.76      0.68        17
        user       0.43      0.27      0.33        11

    accuracy                           0.57        28
   macro avg       0.52      0.52      0.51        28
weighted avg       0.54      0.57      0.55        28
```

Model performance on dataset ROC curve:



Data balancing for ‘Mushrooms’ –

UNDERSAMPLED –

Model performance on dataset with accuracy:

```
Decision Tree: 72.30%
Random Forest: 79.14%
SVM: 79.86%
Gradient Boost: 80.94%
Multi-Layer Perceptron: 72.66%
K-Nearest Neighbors: 75.90%
```

Model performance on dataset with confusion matrix:

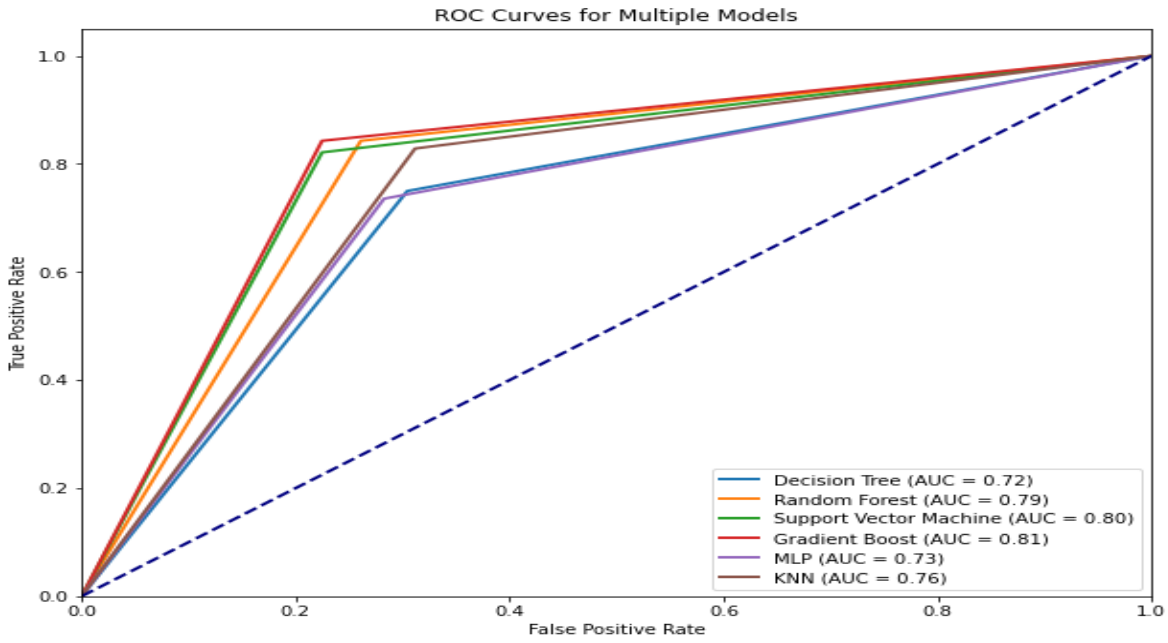
```
Confusion matrix for Decision Tree:
[[ 96  42]
 [ 35 105]]
Confusion matrix for Random Forest:
[[102  36]
 [ 22 118]]
Confusion matrix for SVM:
[[107  31]
 [ 25 115]]
Confusion matrix for Gradient Boost:
[[107  31]
 [ 22 118]]
Confusion matrix for Multi-Layer Perceptron:
[[ 99  39]
 [ 37 103]]
Confusion matrix for K-Nearest Neighbors:
[[ 95  43]
 [ 24 116]]
```

Model performance on dataset Classification report:

```
Classification report for Decision Tree:
              precision    recall  f1-score   support

    non-user       0.73      0.70      0.71       138
        user       0.71      0.75      0.73       140

    accuracy                           0.72       278
   macro avg       0.72      0.72      0.72       278
weighted avg       0.72      0.72      0.72       278

Classification report for Random Forest:
              precision    recall  f1-score   support

    non-user       0.82      0.74      0.78       138
        user       0.77      0.84      0.80       140

    accuracy                           0.79       278
   macro avg       0.79      0.79      0.79       278
weighted avg       0.79      0.79      0.79       278

Classification report for SVM:
              precision    recall  f1-score   support

    non-user       0.81      0.78      0.79       138
```

```
        user       0.79      0.82      0.80       140

    accuracy                           0.80       278
   macro avg       0.80      0.80      0.80       278
weighted avg       0.80      0.80      0.80       278

Classification report for Gradient Boost:
              precision    recall  f1-score   support

    non-user       0.83      0.78      0.80       138
        user       0.79      0.84      0.82       140

    accuracy                           0.81       278
   macro avg       0.81      0.81      0.81       278
weighted avg       0.81      0.81      0.81       278

Classification report for Multi Layer Perceptron:
              precision    recall  f1-score   support

    non-user       0.73      0.72      0.72       138
        user       0.73      0.74      0.73       140

    accuracy                           0.73       278
   macro avg       0.73      0.73      0.73       278
weighted avg       0.73      0.73      0.73       278

Classification report for K-Nearest Neighbors:
              precision    recall  f1-score   support

    non-user       0.80      0.69      0.74       138
        user       0.73      0.83      0.78       140

    accuracy                           0.76       278
   macro avg       0.76      0.76      0.76       278
weighted avg       0.76      0.76      0.76       278
```

Model performance on dataset ROC curve:



OVERSAMPLED –

Model performance on dataset with accuracy:

```
Decision Tree: 77.78%
Random Forest: 83.44%
SVM: 77.36%
Gradient Boost: 79.87%
Multi Layer Perceptron: 80.50%
K-Nearest Neighbors: 77.57%
```

Model performance on dataset with confusion matrix:

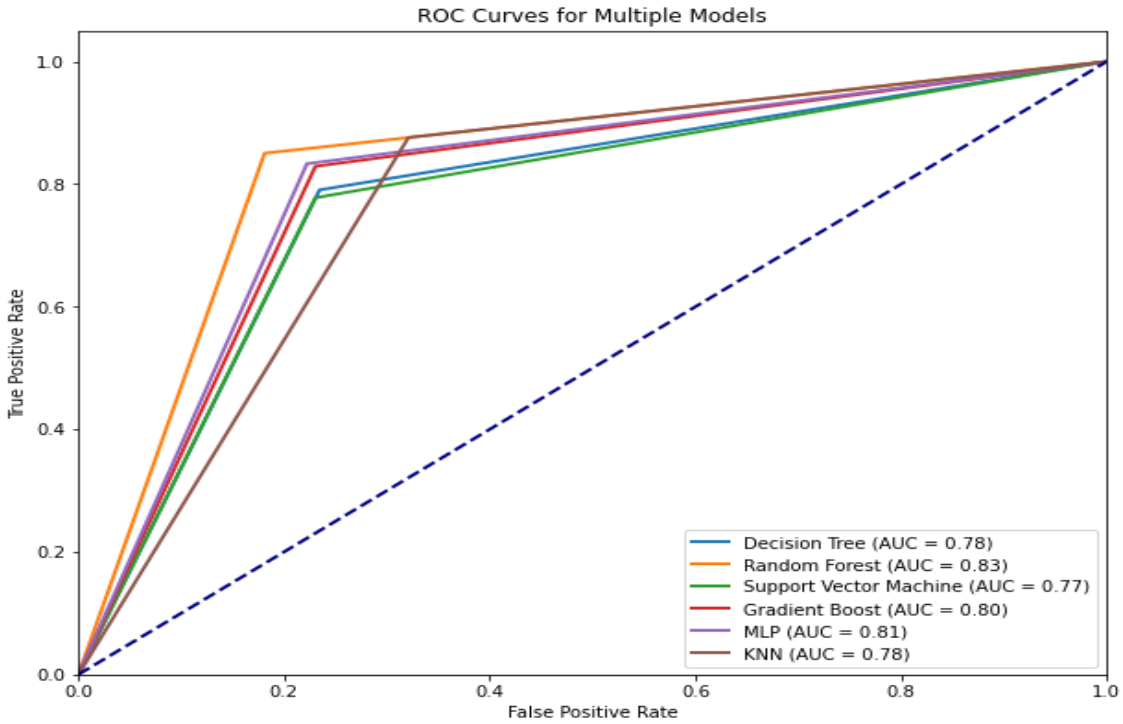
```
Confusion matrix for Decision Tree:
[[186  57]
 [ 49 185]]
Confusion matrix for Random Forest:
[[199  44]
 [ 35 199]]
Confusion matrix for SVM:
[[187  56]
 [ 52 182]]
Confusion matrix for Gradient Boost:
[[187  56]
 [ 40 194]]
Confusion matrix for Multi Layer Perceptron:
[[189  54]
 [ 39 195]]
Confusion matrix for K-Nearest Neighbors:
[[165  78]
 [ 29 205]]
```

Model performance on dataset Classification report:

```
Classification report for Decision Tree:
              precision    recall  f1-score   support

    non-user       0.79      0.77      0.78       243
        user       0.76      0.79      0.78       234

    accuracy                           0.78       477
   macro avg       0.78      0.78      0.78       477
weighted avg       0.78      0.78      0.78       477

Classification report for Random Forest:
              precision    recall  f1-score   support

    non-user       0.85      0.82      0.83       243
        user       0.82      0.85      0.83       234

    accuracy                           0.83       477
   macro avg       0.83      0.83      0.83       477
weighted avg       0.83      0.83      0.83       477

Classification report for SVM:
              precision    recall  f1-score   support

    non-user       0.78      0.77      0.78       243
        user       0.76      0.78      0.77       234

    accuracy                           0.77       477
```

```
   macro avg       0.77      0.77      0.77       477
weighted avg       0.77      0.77      0.77       477

Classification report for Gradient Boost:
              precision    recall  f1-score   support

    non-user       0.82      0.77      0.80       243
        user       0.78      0.83      0.80       234

    accuracy                           0.80       477
   macro avg       0.80      0.80      0.80       477
weighted avg       0.80      0.80      0.80       477

Classification report for Multi Layer Perceptron:
              precision    recall  f1-score   support

    non-user       0.83      0.78      0.80       243
        user       0.78      0.83      0.81       234

    accuracy                           0.81       477
   macro avg       0.81      0.81      0.81       477
weighted avg       0.81      0.81      0.80       477

Classification report for K-Nearest Neighbors:
              precision    recall  f1-score   support

    non-user       0.85      0.68      0.76       243
        user       0.72      0.88      0.79       234

    accuracy                           0.78       477
   macro avg       0.79      0.78      0.77       477
weighted avg       0.79      0.78      0.77       477
```

Model performance on dataset ROC curve:



BALANCED –

Model performance on dataset with accuracy:

```
Decision Tree: 70.03%
Random Forest: 80.12%
SVM: 77.52%
Gradient Boost: 80.40%
Multi Layer Perceptron: 76.95%
K-Nearest Neighbors: 78.10%
```

Model performance on dataset with confusion matrix:

```
Confusion matrix for Decision Tree:
[[126  57]
 [ 47 117]]
Confusion matrix for Random Forest:
[[148  35]
 [ 34 130]]
Confusion matrix for SVM:
[[144  39]
 [ 39 125]]
Confusion matrix for Gradient Boost:
[[150  33]
 [ 35 129]]
Confusion matrix for Multi Layer Perceptron:
[[140  43]
 [ 37 127]]
Confusion matrix for K-Nearest Neighbors:
[[134  49]
 [ 27 137]]
```

Model performance on dataset Classification report:

```
Classification report for Decision Tree:
              precision    recall  f1-score   support

    non-user       0.73      0.69      0.71       183
        user       0.67      0.71      0.69       164

    accuracy                           0.70       347
   macro avg       0.70      0.70      0.70       347
weighted avg       0.70      0.70      0.70       347

Classification report for Random Forest:
              precision    recall  f1-score   support

    non-user       0.81      0.81      0.81       183
        user       0.79      0.79      0.79       164

    accuracy                           0.80       347
   macro avg       0.80      0.80      0.80       347
weighted avg       0.80      0.80      0.80       347

Classification report for SVM:
              precision    recall  f1-score   support

    non-user       0.79      0.79      0.79       183
        user       0.76      0.76      0.76       164

    accuracy                           0.78       347
```

```
   macro avg       0.77      0.77      0.77       347
weighted avg       0.78      0.78      0.78       347

Classification report for Gradient Boost:
              precision    recall  f1-score   support

    non-user       0.81      0.82      0.82       183
        user       0.80      0.79      0.79       164

    accuracy                           0.80       347
   macro avg       0.80      0.80      0.80       347
weighted avg       0.80      0.80      0.80       347

Classification report for Multi Layer Perceptron:
              precision    recall  f1-score   support

    non-user       0.79      0.77      0.78       183
        user       0.75      0.77      0.76       164

    accuracy                           0.77       347
   macro avg       0.77      0.77      0.77       347
weighted avg       0.77      0.77      0.77       347

Classification report for K-Nearest Neighbors:
              precision    recall  f1-score   support

    non-user       0.83      0.73      0.78       183
        user       0.74      0.84      0.78       164

    accuracy                           0.78       347
   macro avg       0.78      0.78      0.78       347
weighted avg       0.79      0.78      0.78       347
```

Model performance on dataset ROC curve: