

Classification for Chocolate Dataset

Decision Tree Classifier

Accuracy_DT: 95.76
Precision_DT: 0.972972972972973
Recall_DT: 0.9836065573770492

Confusion matrix:

```
[[ 2 15]
 [ 9 540]]
```

Random Forest

Accuracy_RF: 96.99
Precision_DT: 0.9730700179533214
Recall_DT: 0.9872495446265938

Confusion matrix:

```
[[ 0 17]
 [ 0 549]]
```

Support Vector Machine

Accuracy_SVM: 97.0
Precision_SVM: 0.9699646643109541
Recall_SVM: 1.0

Confusion matrix:

```
[[ 0 17]
 [ 0 549]]
```

KNN

Accuracy_KNN: 96.64
Precision_KNN: 0.9698581560283688
Recall_KNN: 0.9963570127504554

Confusion matrix:

```
[[ 0 17]
 [ 2 547]]
```

Multi Layer Perceptron

Accuracy_MLP: 96.29
Precision_MLP: 0.9697508896797153
Recall_MLP: 0.9927140255009107

Confusion matrix:

```
[[ 0 17]
 [ 4 545]]
```

GB

Accuracy_GB: 96.82

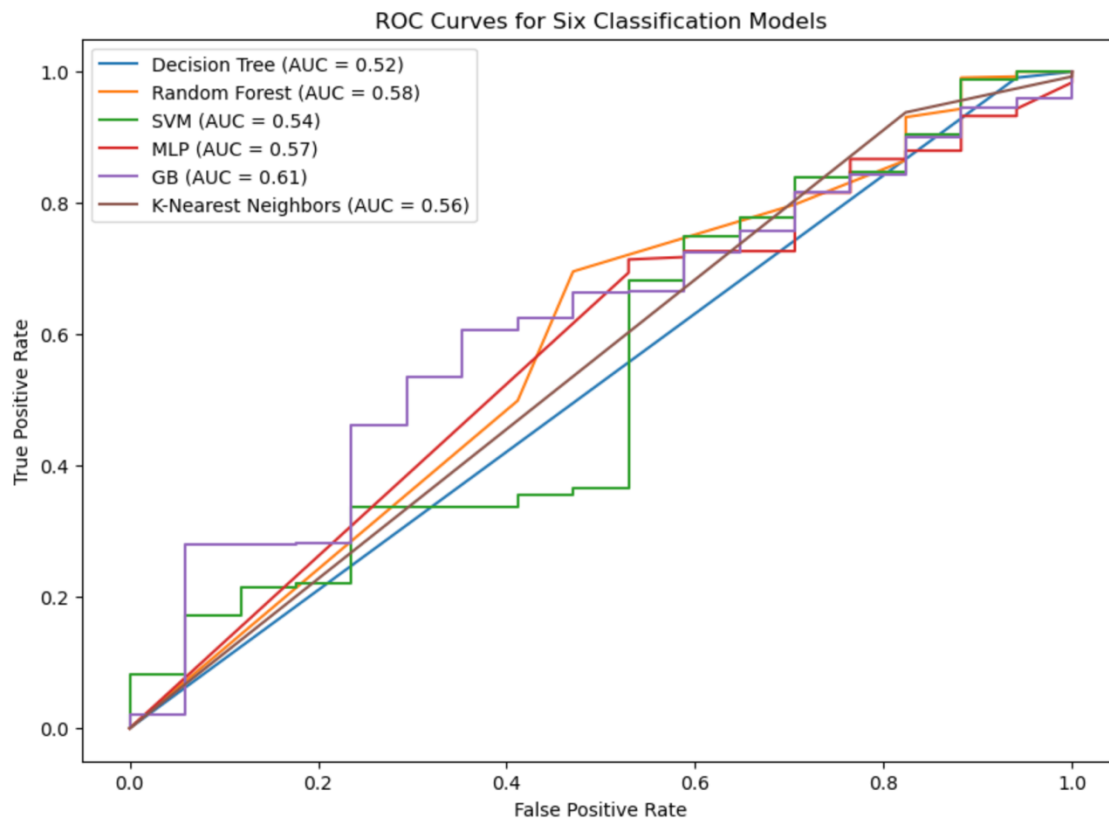
Precision_GB: 0.9699115044247788

Recall_GB: 0.9981785063752276

Confusion matrix:

```
[[ 0 17]
 [ 1 548]]
```

ROC Curve



Classification for Mushroom Dataset

Decision Tree Classifier

Accuracy_DT: 0.6784452296819788

Precision_DT: 0.6

Recall_DT: 0.5733333333333334

Confusion matrix:

```
[[255 86]
 [ 96 129]]
```

Random Forest Learner

Accuracy_RF: 77.21

Precision_DT: 0.6

Recall_DT: 0.5733333333333334

Confusion matrix:

```
[[294 47]
 [ 86 139]]
```

Support Vector Machine

Accuracy_SVM: 75.8

Precision_SVM: 0.7115384615384616

Recall_SVM: 0.6577777777777778

Confusion matrix:

```
[[281 60]
 [ 77 148]]
```

KNN

Accuracy_KNN: 72.97

Precision_KNN: 0.6818181818181818

Recall_KNN: 0.6

Confusion matrix:

```
[[278 63]
 [ 90 135]]
```

Multi Layer Perceptron

Accuracy_MLP: 75.97

Precision_MLP: 0.7213930348258707

Recall_MLP: 0.6444444444444445

Confusion matrix:

```
[[285 56]
 [ 80 145]]
```

GB

Accuracy_GB: 75.97

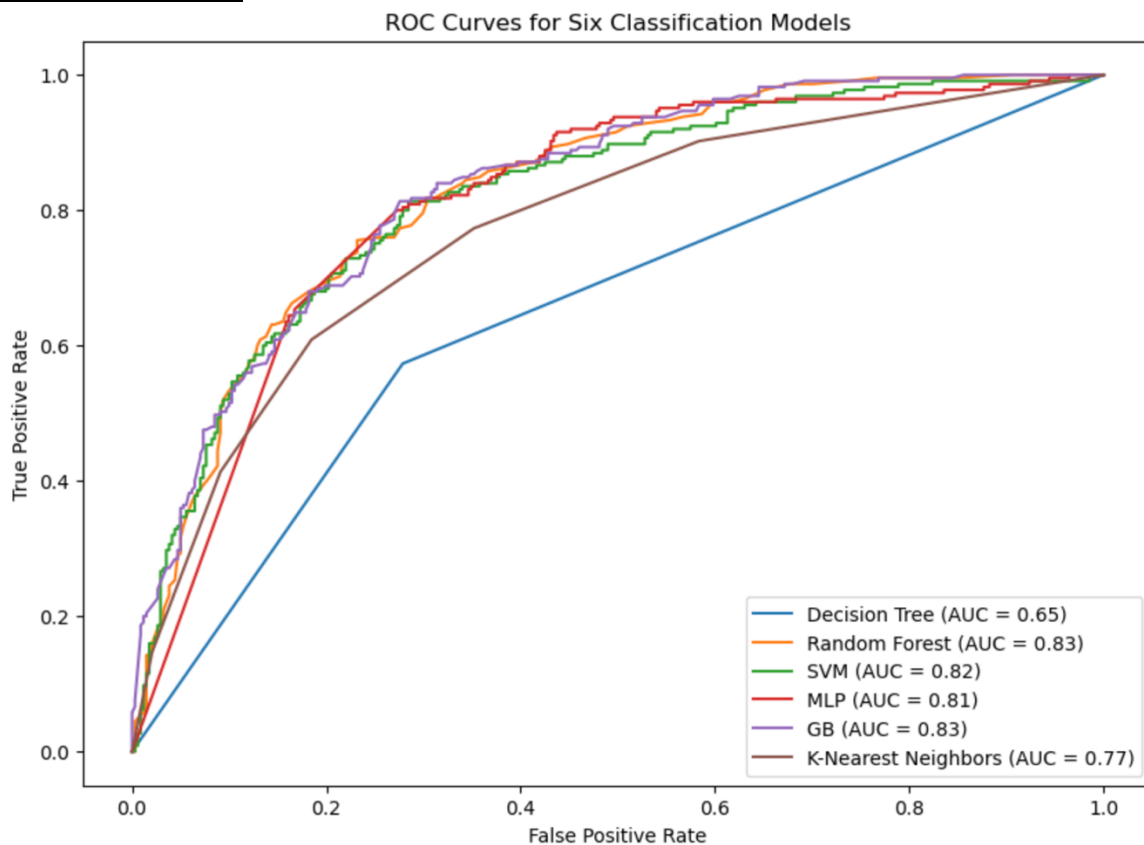
Precision_GB: 0.7192118226600985

Recall_GB: 0.6488888888888888

Confusion matrix:

```
[[284 57]
 [ 79 146]]
```

ROC CURVE



Data balancing for 'Chocolate'

Decision Tree classifier

Accuracy_DT_under: 0.9612612612612612

Precision_DT_under: 0.9819168173598554

Recall_DT_under: 0.9427083333333334

Confusion matrix:

```
[[524 10]
```

```
 [ 33 543]]
```

Accuracy_DT_over: 0.981081081081081

Precision_DT_over: 1.0

Recall_DT_over: 0.9635416666666666

Confusion matrix:

```
[[534 0]
```

[21 555]]

Accuracy_DT_combined: 0.7027027027027027

Precision_DT_combined: 0.68

Recall_DT_combined: 0.85

Confusion matrix:

```
[[ 9  8]
 [ 3 17]]
```

randomforest classifier

Accuracy_RF_under: 0.9963963963963964

Precision_RF_under: 0.993103448275862

Recall_RF_under: 1.0

Confusion matrix:

```
[[530  4]
 [ 0 576]]
```

Accuracy_RF_over: 1.0

Precision_RF_over: 1.0

Recall_RF_over: 1.0

Confusion matrix:

```
[[534  0]
 [ 0 576]]
```

Accuracy_RF_combined: 0.7567567567567568

Precision_RF_combined: 0.7619047619047619

Recall_RF_combined: 0.8

Confusion matrix:

```
[[12  5]
 [ 4 16]]
```

Support Vector Machine

Accuracy_SVM_under: 0.7477477477477478

Precision_SVM_under: 0.8936170212765957

Recall_SVM_under: 0.5833333333333334

Confusion matrix:

```
[[494 40]
 [240 336]]
```

Accuracy_SVM_over: 0.7360360360360361

Precision_SVM_over: 0.8814016172506739

Recall_SVM_over: 0.5677083333333334

Confusion matrix:

```
[[490 44]
```

[249 327]]

Accuracy_SVM_combined: 0.6756756756756757
Precision_SVM_combined: 0.6666666666666666
Recall_SVM_combined: 0.8

Confusion matrix:

```
[[ 9  8]
 [ 4 16]]
```

KNN classifier

Accuracy_KNN_under: 0.9261261261261261
Precision_KNN_under: 1.0
Recall_KNN_under: 0.8576388888888888

Confusion matrix:

```
[[534  0]
 [ 82 494]]
```

Accuracy_KNN_over: 0.9783783783783784
Precision_KNN_over: 1.0
Recall_KNN_over: 0.9583333333333334

Confusion matrix:

```
[[534  0]
 [ 24 552]]
```

Accuracy_KNN_combined: 0.6216216216216216
Precision_KNN_combined: 0.6875
Recall_KNN_combined: 0.55

Confusion matrix:

```
[[12  5]
 [ 9 11]]
```

MLP classifier

Accuracy_MLP_under: 0.9072072072072072
Precision_MLP_under: 0.9937369519832986
Recall_MLP_under: 0.8263888888888888

Confusion matrix:

```
[[531  3]
 [100 476]]
```

Accuracy_MLP_over: 0.9369369369369369
Precision_MLP_over: 1.0
Recall_MLP_over: 0.8784722222222222

Confusion matrix:

```
[[534  0]
 [ 70 506]]
```

Accuracy_MLP_combined: 0.7567567567567568
Precision_MLP_combined: 0.7894736842105263
Recall_MLP_combined: 0.75

Confusion matrix:

```
[[13  4]
 [ 5 15]]
```

GB classifier

Accuracy_GB_under: 0.9666666666666667
Precision_GB_under: 0.9736379613356766
Recall_GB_under: 0.9618055555555556

Confusion matrix:

```
[[519 15]
 [ 22 554]]
```

Accuracy_GB_over: 0.9225225225225225
Precision_GB_over: 1.0
Recall_GB_over: 0.8506944444444444

Confusion matrix:

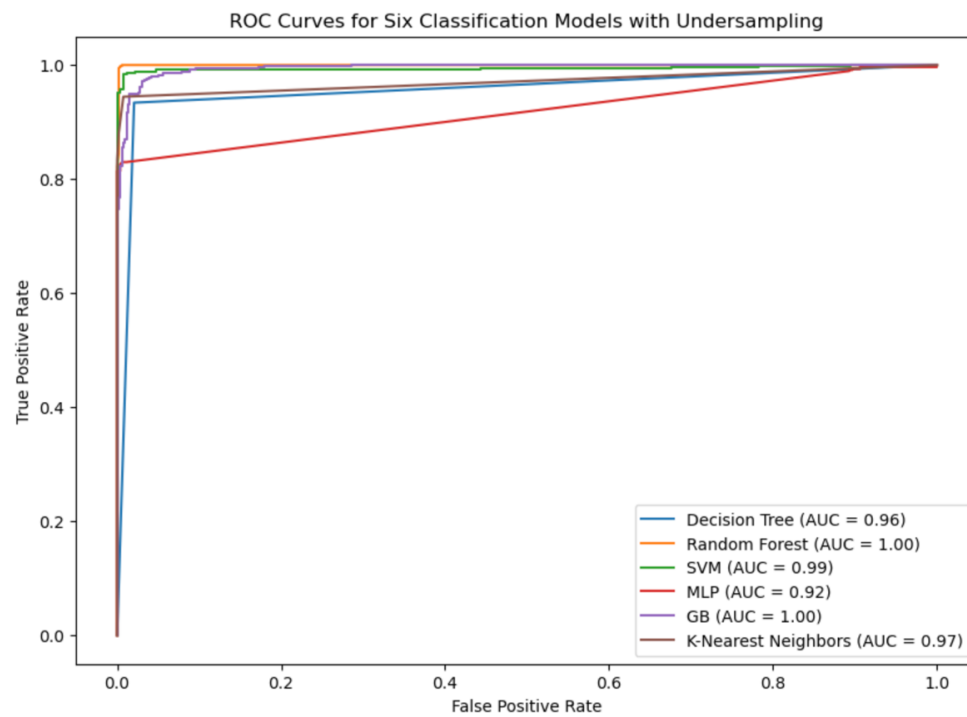
```
[[534  0]
 [ 21 555]]
```

Accuracy_GB_combined: 0.6756756756756757
Precision_GB_combined: 0.6666666666666666
Recall_GB_combined: 0.8

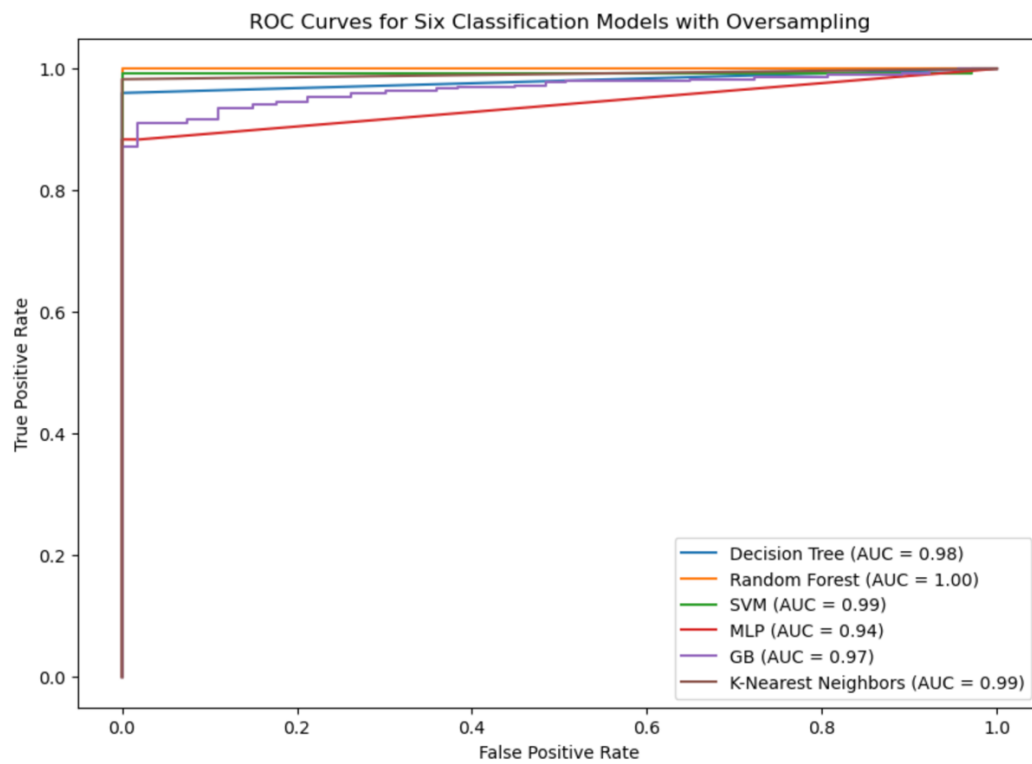
Confusion matrix:

```
[[ 9  8]
 [ 4 16]]
```

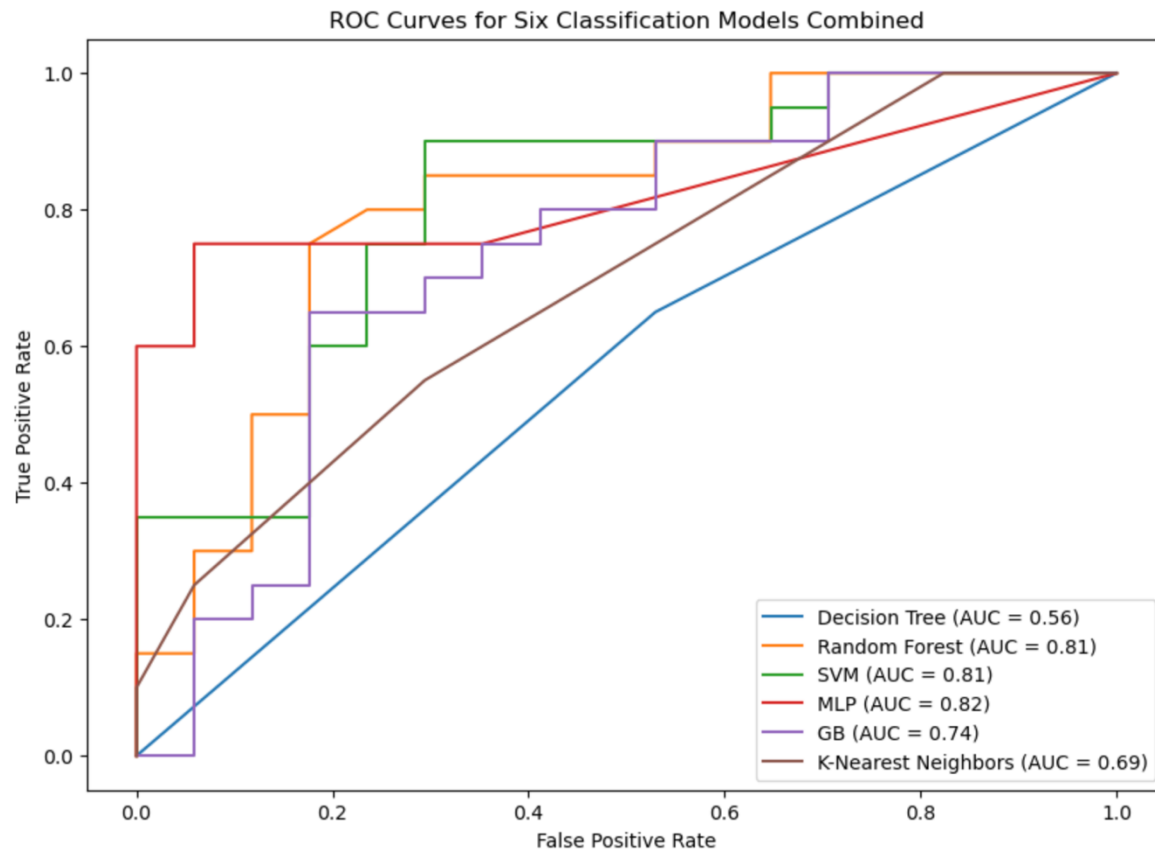
ROC Curve for Undersampled Data



ROC Curve for Oversampled Data



ROC Curves Combined Data



Data balancing for 'Mushrooms'

Decision Tree classifier

Accuracy_DT_under: 0.7524475524475525

Precision_DT_under: 0.7548209366391184

Recall_DT_under: 0.7569060773480663

Confusion matrix:

```
[[264 89]
 [ 88 274]]
```

Accuracy_DT_over: 0.772027972027972

Precision_DT_over: 0.7420924574209246

Recall_DT_over: 0.8425414364640884

Confusion matrix:

```
[[247 106]
 [ 57 305]]
```

Accuracy_DT_combined: 0.6634304207119741

Precision_DT_combined: 0.6411764705882353

Recall_DT_combined: 0.7171052631578947

Confusion matrix:

```
[[ 96 61]
 [ 43 109]]
```

randomforest classifier

Accuracy_RF_under: 0.8195804195804196

Precision_RF_under: 0.8263305322128851

Recall_RF_under: 0.8149171270718232

Confusion matrix:

```
[[291 62]
 [ 67 295]]
```

Accuracy_RF_over: 0.8503496503496504

Precision_RF_over: 0.8294573643410853

Recall_RF_over: 0.8867403314917127

Confusion matrix:

```
[[287 66]
 [ 41 321]]
```

Accuracy_RF_combined: 0.7896440129449838

Precision_RF_combined: 0.7735849056603774

Recall_RF_combined: 0.8092105263157895

Confusion matrix:

```
[[121 36]
 [ 29 123]]
```

Support Vector Machine

Accuracy_SVM_under: 0.7888111888111888

Precision_SVM_under: 0.790633608815427

Recall_SVM_under: 0.7928176795580111

Confusion matrix:

```
[[277 76]
 [ 75 287]]
```

Accuracy_SVM_over: 0.7636363636363637

Precision_SVM_over: 0.7797101449275362

Recall_SVM_over: 0.7430939226519337

Confusion matrix:

```
[[277 76]
 [ 93 269]]
```

Accuracy_SVM_combined: 0.7443365695792881

Precision_SVM_combined: 0.7517241379310344

Recall_SVM_combined: 0.7171052631578947

Confusion matrix:

```
[[121 36]
 [ 43 109]]
```

KNN classifier

Accuracy_KNN_under: 0.7678321678321678

Precision_KNN_under: 0.7322274881516587

Recall_KNN_under: 0.8535911602209945

Confusion matrix:

```
[[240 113]
 [ 53 309]]
```

Accuracy_KNN_over: 0.7608391608391608

Precision_KNN_over: 0.7279236276849642

Recall_KNN_over: 0.8425414364640884

Confusion matrix:

```
[[239 114]
 [ 57 305]]
```

Accuracy_KNN_combined: 0.6990291262135923

Precision_KNN_combined: 0.6787878787878788

Recall_KNN_combined: 0.7368421052631579

Confusion matrix:

```
[[104 53]
 [ 40 112]]
```

MLP classifier

Accuracy_MLP_under: 0.7916083916083916

Precision_MLP_under: 0.7642679900744417

Recall_MLP_under: 0.850828729281768

Confusion matrix:

```
[[258 95]
 [ 54 308]]
```

Accuracy_MLP_over: 0.7734265734265734

Precision_MLP_over: 0.773224043715847

Recall_MLP_over: 0.7817679558011049

Confusion matrix:

```
[[270 83]
 [ 79 283]]
```

Accuracy_MLP_combined: 0.7443365695792881

Precision_MLP_combined: 0.7085714285714285

Recall_MLP_combined: 0.8157894736842105

Confusion matrix:

```
[[106 51]
 [ 28 124]]
```

GB classifier

Accuracy_GB_under: 0.7986013986013986

Precision_GB_under: 0.7978142076502732

Recall_GB_under: 0.8066298342541437

Confusion matrix:

```
[[279 74]
 [ 70 292]]
```

Accuracy_GB_over: 0.7482517482517482

Precision_GB_over: 0.7446236559139785

Recall_GB_over: 0.7651933701657458

Confusion matrix:

```
[[247 106]
 [ 57 305]]
```

Accuracy_GB_combined: 0.7896440129449838

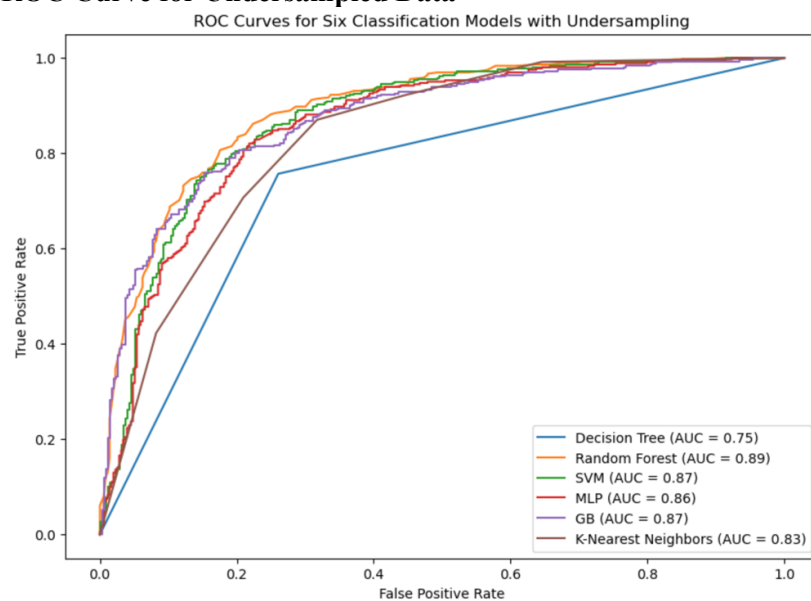
Precision_GB_combined: 0.7843137254901961

Recall_GB_combined: 0.7894736842105263

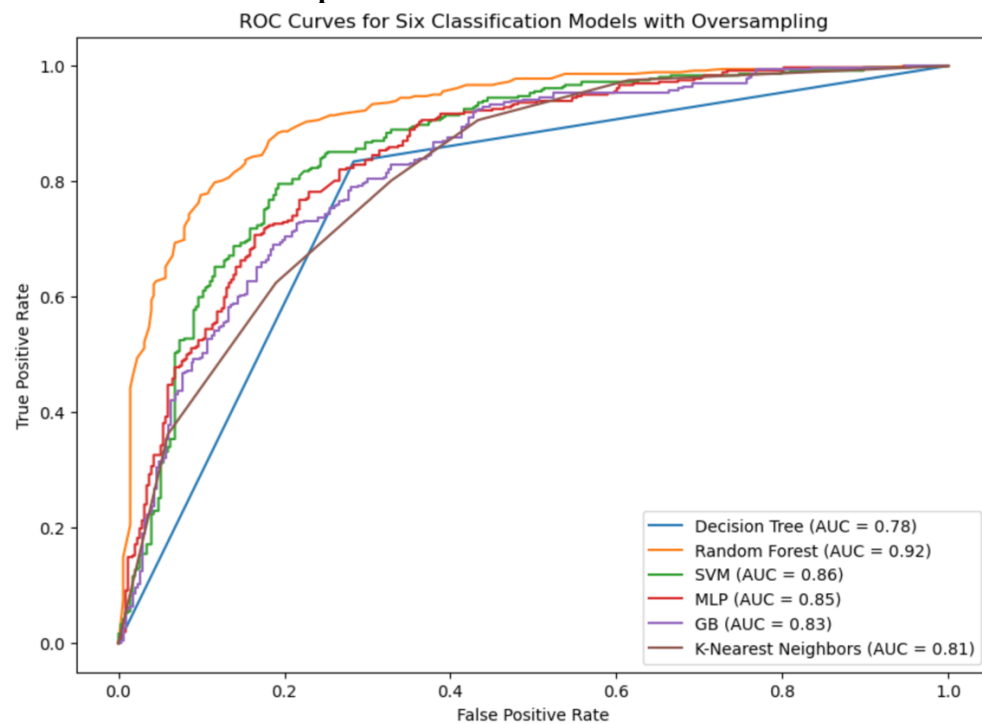
Confusion matrix:

```
[[124 33]
 [ 32 120]]
```

ROC Curve for Undersampled Data



ROC Curve for Oversampled Data



ROC Curves Combined Data

