**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

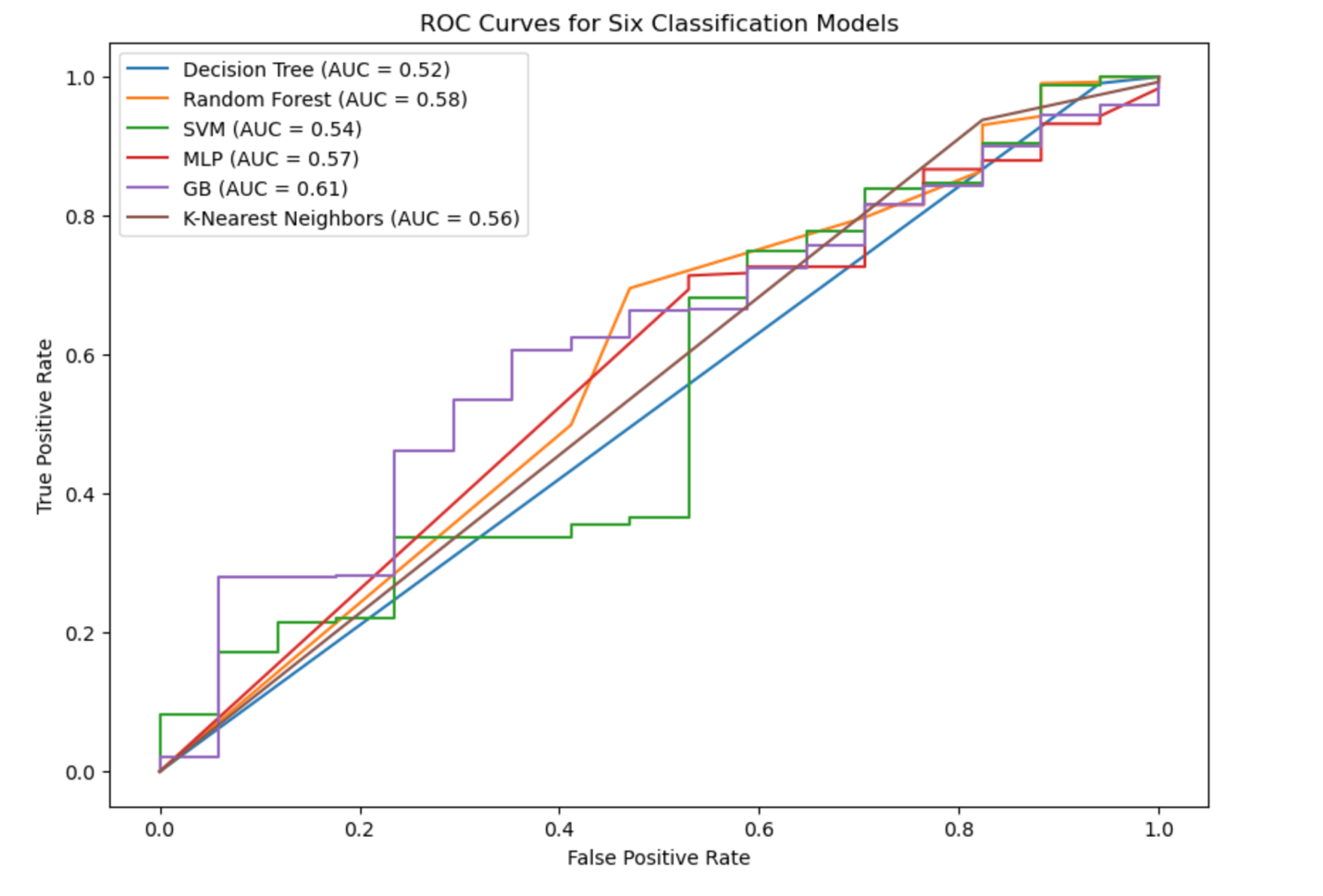
Recall\_GB: 0.9981785063752276

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

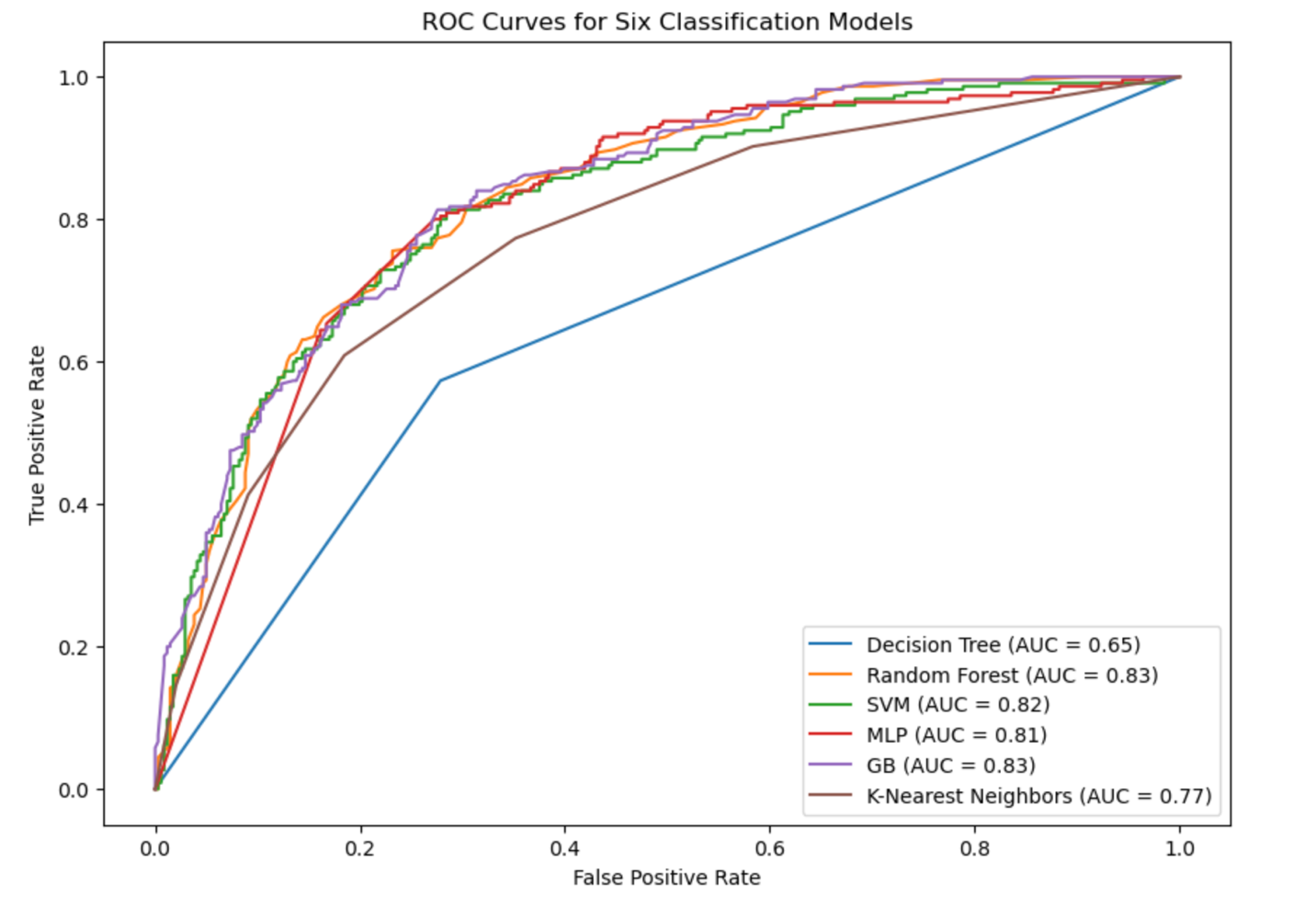
Recall\_GB: 0.6488888888888888

Confusion matrix:

[[284 57]

[ 79 146]]

**ROC CURVE**



**Data balancing for ‘Chocolate’**

**Decision Tree classifer**

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 classifer**

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 classifer**

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 classifer**

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 classifer**

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

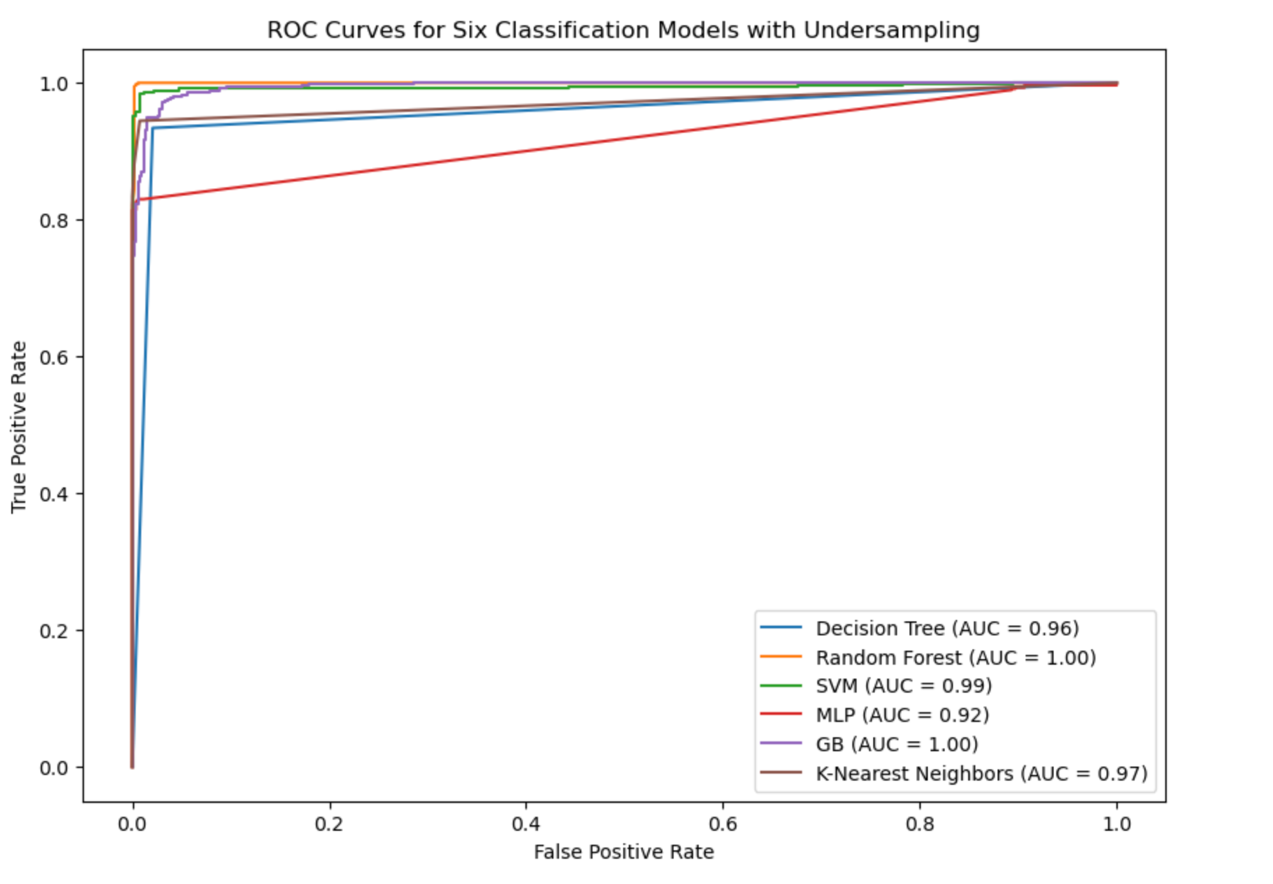
Recall\_GB\_combined: 0.8

Confusion matrix:

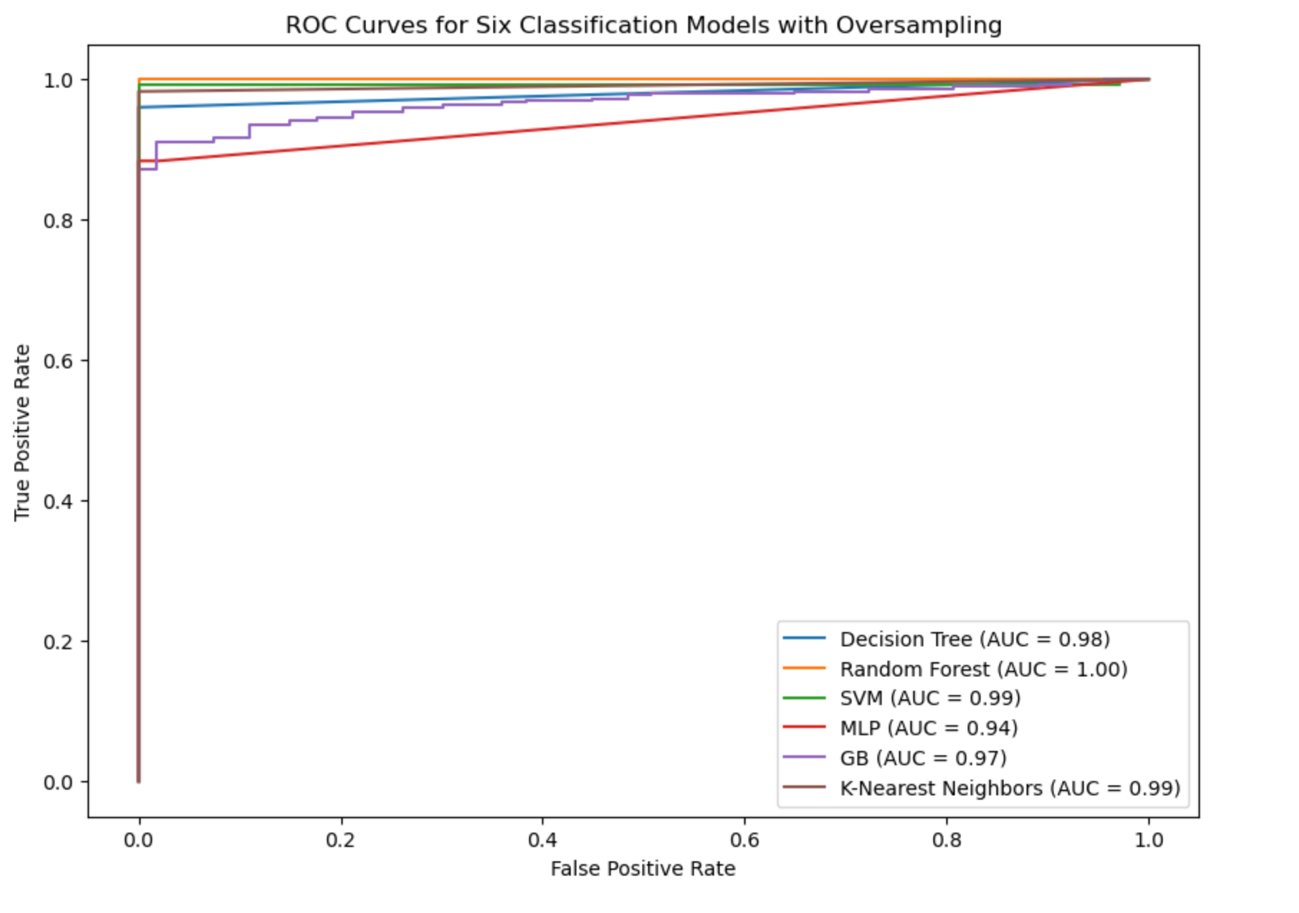
[[ 9 8]

[ 4 16]]

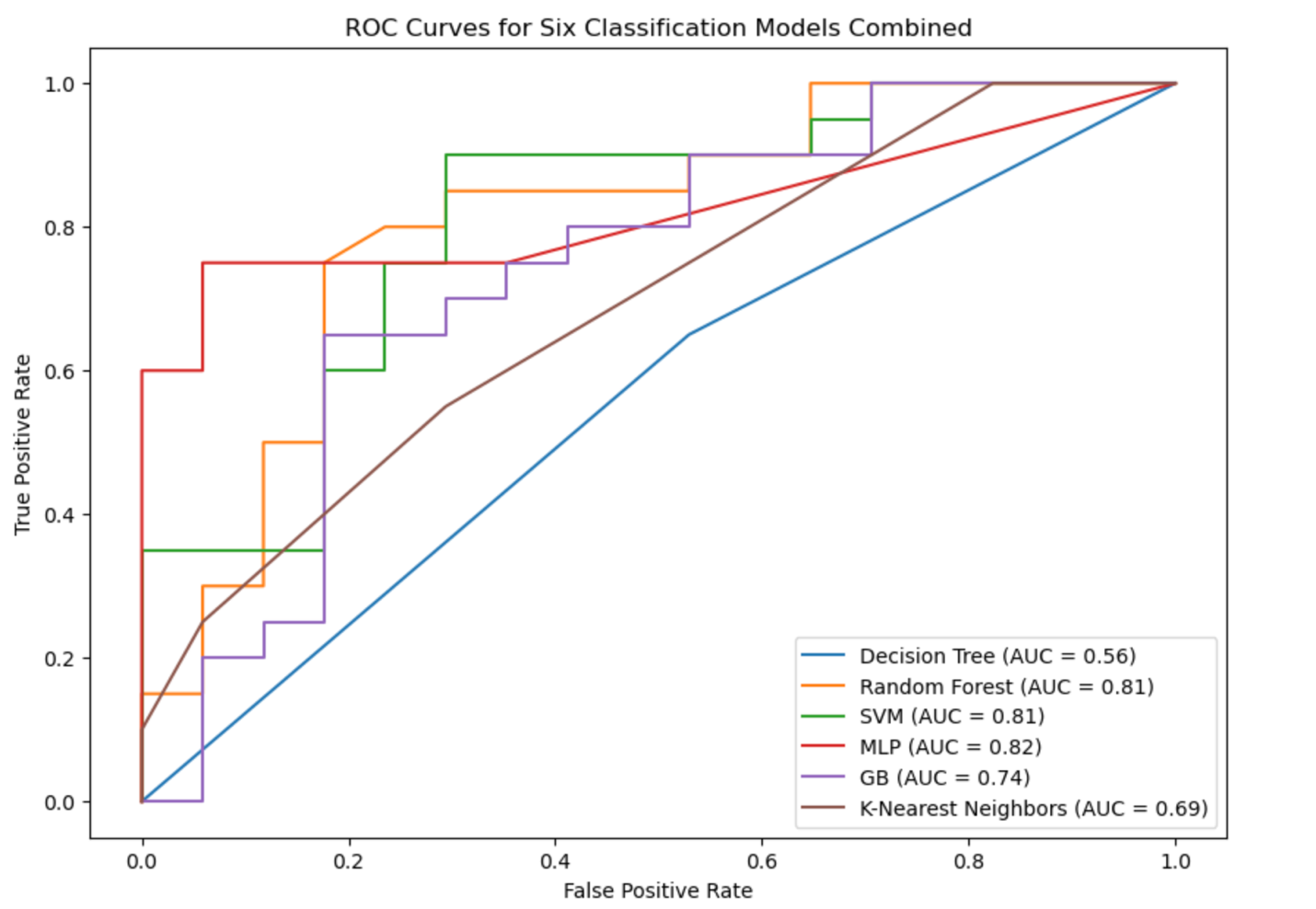
**ROC Curve for Undersampled Data**



**ROC Curve for Oversampled Data**



**ROC Curves Combined Data**



**Data balancing for ‘Mushrooms’**

**Decision Tree classifer**

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 classifer**

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 classifer**

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 classifer**

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 classifer**

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

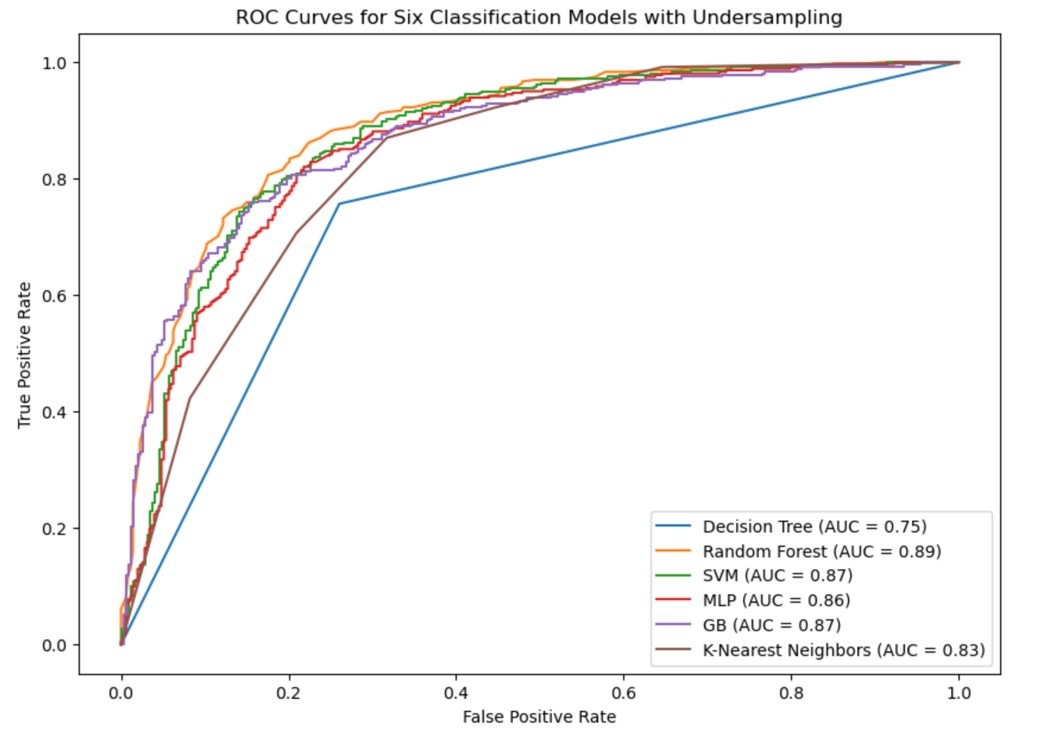
Recall\_GB\_combined: 0.7894736842105263

Confusion matrix:

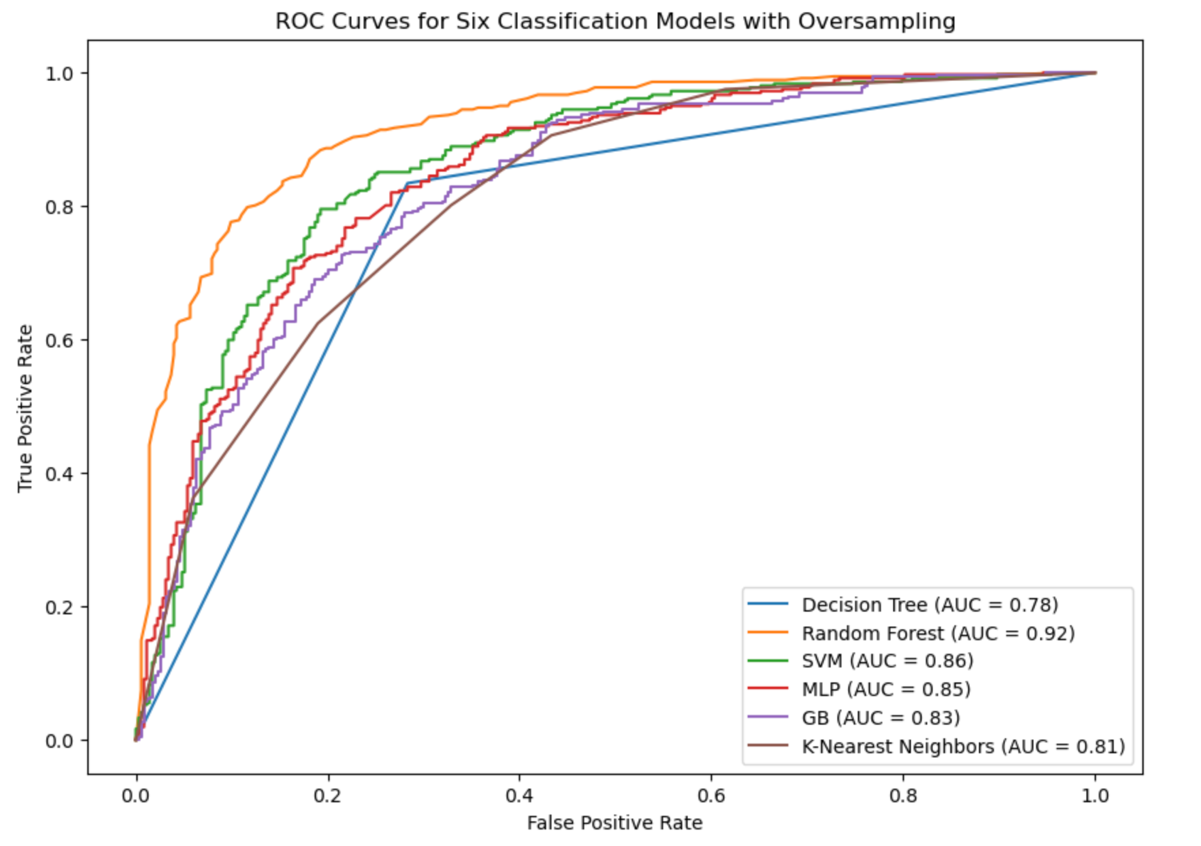
[[124 33]

[ 32 120]]

**ROC Curve for Undersampled Data**



**ROC Curve for Oversampled Data**



**ROC Curves Combined Data**

