

Springboard Data Science Course

Data Science Capstone Project 1

Orthopedic Biomechanical Features

Michelle Ide

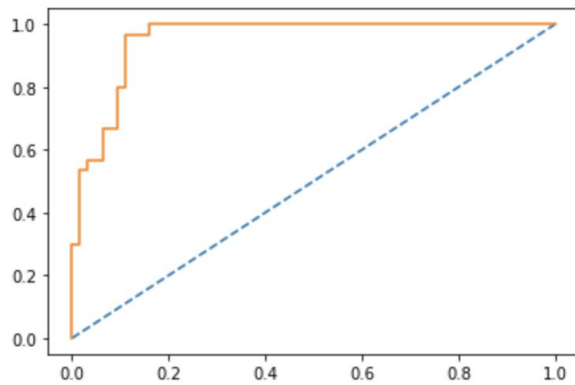
1/25/2021

~~~ ADDENDUM ~~~

Results for testing of 6 classification models are below using SMOTE and ADASYN resampling methods. ADASYN reduces false negatives, while SMOTE improves accuracy with more balanced results.

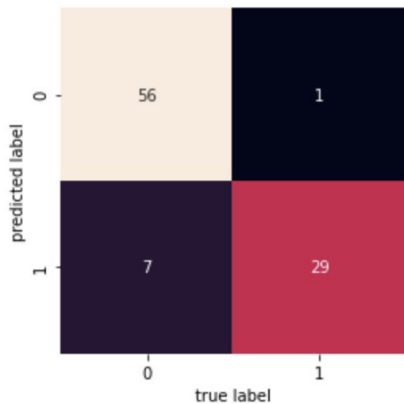
SMOTE - LOGISTIC REGRESSION

Training score: 0.8425925925925926  
area under curve (auc): 0.9523809523809523



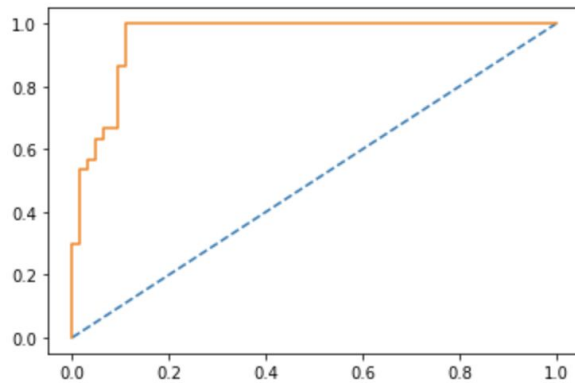
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.98      | 0.89   | 0.93     | 63      |
| 1            | 0.81      | 0.97   | 0.88     | 30      |
| accuracy     |           |        | 0.91     | 93      |
| macro avg    | 0.89      | 0.93   | 0.91     | 93      |
| weighted avg | 0.93      | 0.91   | 0.92     | 93      |

Best Parameters: {'class\_\_C': 10, 'class\_\_penalty': 'l2', 'class\_\_solver': 'newton-cg'}



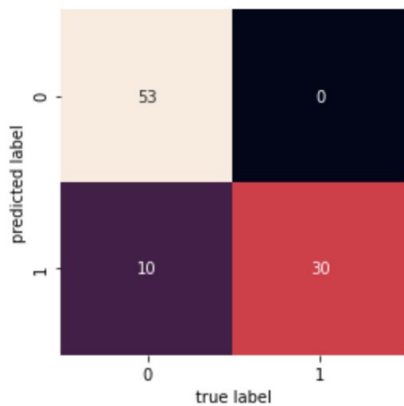
ADASYN - LOGISTIC REGRESSION

Training score: 0.8425925925925926  
area under curve (auc): 0.9560846560846561



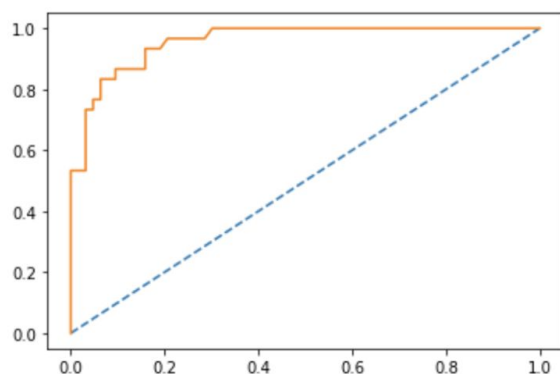
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 1.00      | 0.84   | 0.91     | 63      |
| 1            | 0.75      | 1.00   | 0.86     | 30      |
| accuracy     |           |        | 0.89     | 93      |
| macro avg    | 0.88      | 0.92   | 0.89     | 93      |
| weighted avg | 0.92      | 0.89   | 0.90     | 93      |

Best Parameters: {'class\_\_C': 10, 'class\_\_penalty': 'l2', 'class\_\_solver': 'newton-cg'}



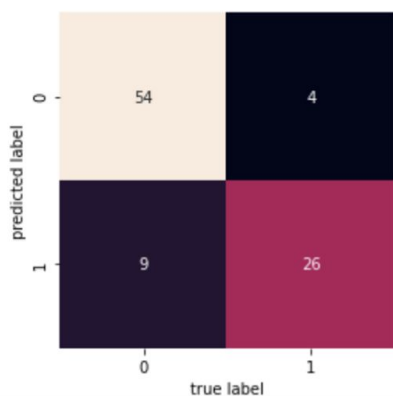
SMOTE - GRADIENT BOOSTING CLASSIFIER

Training score: 0.8425925925925926  
area under curve (auc): 0.9576719576719577



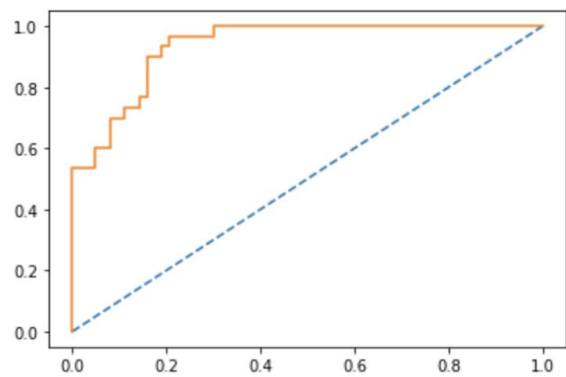
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.93      | 0.86   | 0.89     | 63      |
| 1            | 0.74      | 0.87   | 0.80     | 30      |
| accuracy     |           |        | 0.86     | 93      |
| macro avg    | 0.84      | 0.86   | 0.85     | 93      |
| weighted avg | 0.87      | 0.86   | 0.86     | 93      |

Best Parameters: {'class\_\_criterion': 'friedman\_mse', 'class\_\_learning\_rate': 0.1, 'class\_\_loss': 'exponential', 'class\_\_n\_estimators': 16}



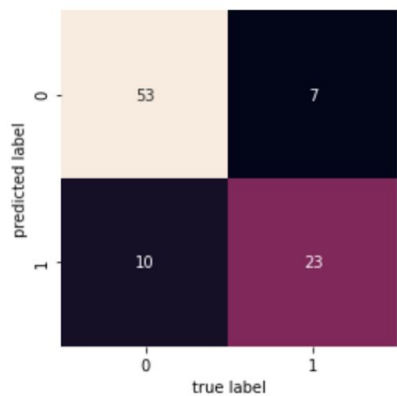
ADASYN - GRADIENT BOOSTING CLASSIFIER

Training score: 0.8425925925925926  
area under curve (auc): 0.935978835978836



|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.88      | 0.84   | 0.86     | 63      |
| 1            | 0.70      | 0.77   | 0.73     | 30      |
| accuracy     |           |        | 0.82     | 93      |
| macro avg    | 0.79      | 0.80   | 0.80     | 93      |
| weighted avg | 0.82      | 0.82   | 0.82     | 93      |

Best Parameters: {'class\_\_criterion': 'friedman\_mse', 'class\_\_learning\_rate': 1, 'class\_\_loss': 'exponential', 'class\_\_n\_estimators': 11}

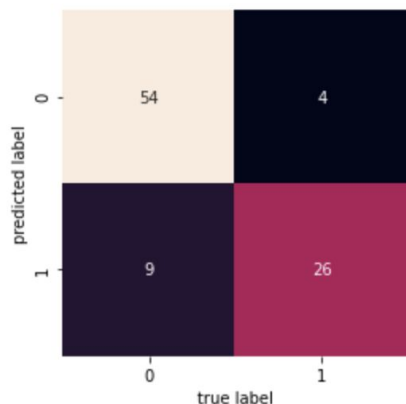


SMOTE - SUPPORT VECTOR MACHINE

Training score: 0.8425925925925926

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.93      | 0.86   | 0.89     | 63      |
| 1            | 0.74      | 0.87   | 0.80     | 30      |
| accuracy     |           |        | 0.86     | 93      |
| macro avg    | 0.84      | 0.86   | 0.85     | 93      |
| weighted avg | 0.87      | 0.86   | 0.86     | 93      |

Best Parameters: {'class\_\_C': 1000, 'class\_\_gamma': 0.01, 'class\_\_kernel': 'rbf'}

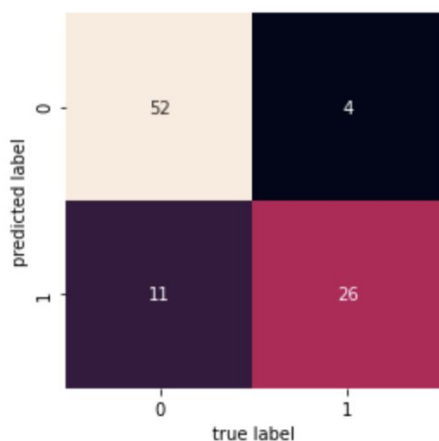


## ADASYN - SUPPORT VECTOR MACHINE

Training score: 0.8425925925925926

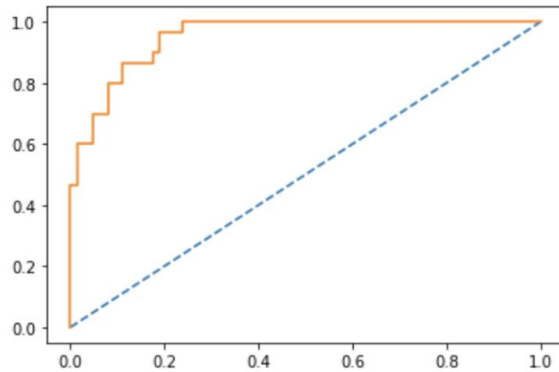
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.93      | 0.83   | 0.87     | 63      |
| 1            | 0.70      | 0.87   | 0.78     | 30      |
| accuracy     |           |        | 0.84     | 93      |
| macro avg    | 0.82      | 0.85   | 0.83     | 93      |
| weighted avg | 0.86      | 0.84   | 0.84     | 93      |

Best Parameters: {'class\_\_C': 10, 'class\_\_gamma': 0.1, 'class\_\_kernel': 'rbf'}



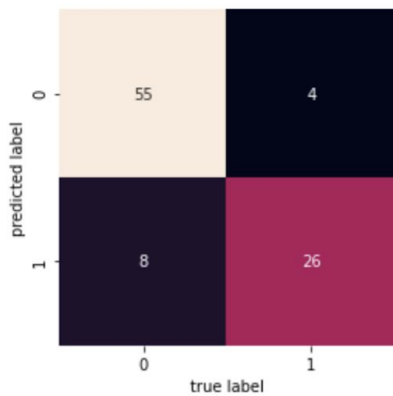
## SMOTE - RANDOM FOREST CLASSIFIER

Training score: 0.8425925925925926  
area under curve (auc): 0.9513227513227513



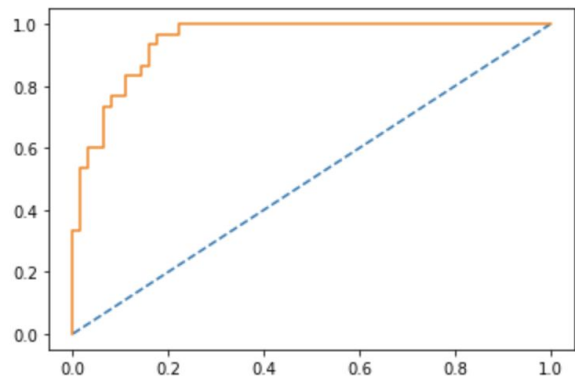
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.93      | 0.87   | 0.90     | 63      |
| 1            | 0.76      | 0.87   | 0.81     | 30      |
| accuracy     |           |        | 0.87     | 93      |
| macro avg    | 0.85      | 0.87   | 0.86     | 93      |
| weighted avg | 0.88      | 0.87   | 0.87     | 93      |

Best Parameters: {'class\_\_criterion': 'gini', 'class\_\_max\_depth': 6, 'class\_\_min\_samples\_leaf': 1, 'class\_\_n\_estimators': 16}



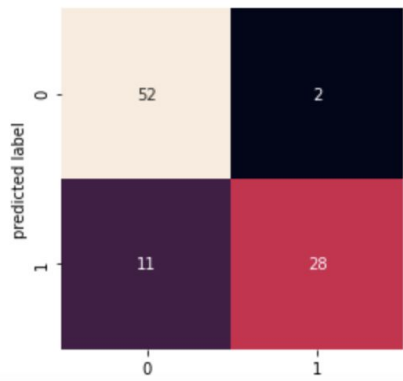
## ADASYN - RANDOM FOREST CLASSIFIER

Training score: 0.8425925925925926  
area under curve (auc): 0.9476190476190476



|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.96      | 0.83   | 0.89     | 63      |
| 1            | 0.72      | 0.93   | 0.81     | 30      |
| accuracy     |           |        | 0.86     | 93      |
| macro avg    | 0.84      | 0.88   | 0.85     | 93      |
| weighted avg | 0.88      | 0.86   | 0.86     | 93      |

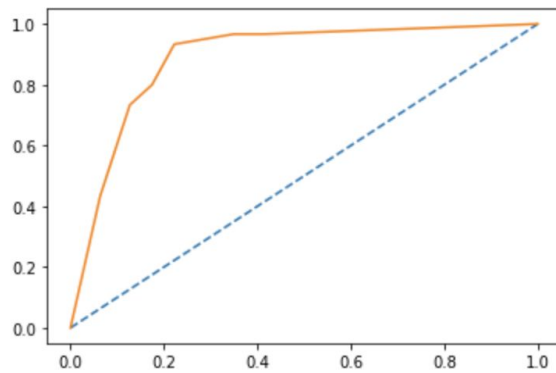
Best Parameters: {'class\_\_criterion': 'entropy', 'class\_\_max\_depth': 5, 'class\_\_min\_sample\_s\_leaf': 4, 'class\_\_n\_estimators': 32}



SMOTE - KNEIGHBORS

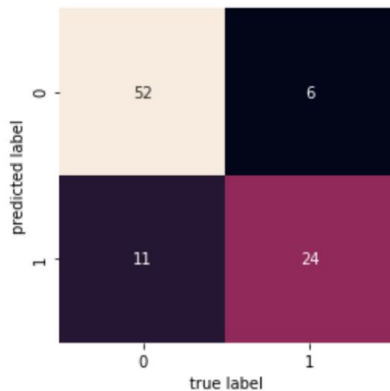


Training score: 0.8425925925925926  
area under curve (auc): 0.888095238095238



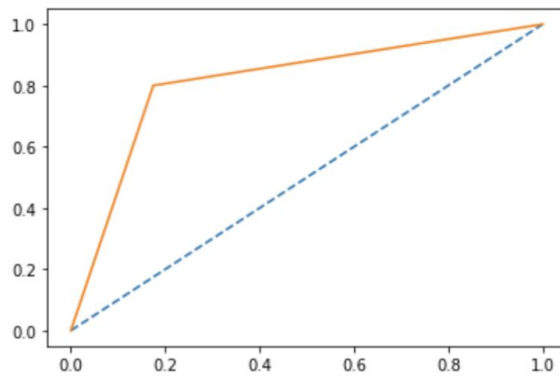
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.90      | 0.83   | 0.86     | 63      |
| 1            | 0.69      | 0.80   | 0.74     | 30      |
| accuracy     |           |        | 0.82     | 93      |
| macro avg    | 0.79      | 0.81   | 0.80     | 93      |
| weighted avg | 0.83      | 0.82   | 0.82     | 93      |

Best Parameters: {'class\_algorithm': 'auto', 'class\_leaf\_size': 1, 'class\_n\_neighbors': 6, 'class\_p': 2, 'class\_weights': 'uniform'}



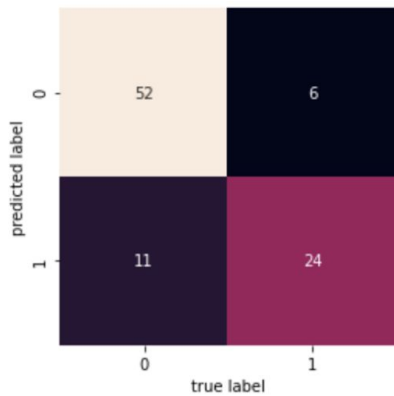
ADASYN - KNEIGHBORS

Training score: 0.8425925925925926  
area under curve (auc): 0.8126984126984127



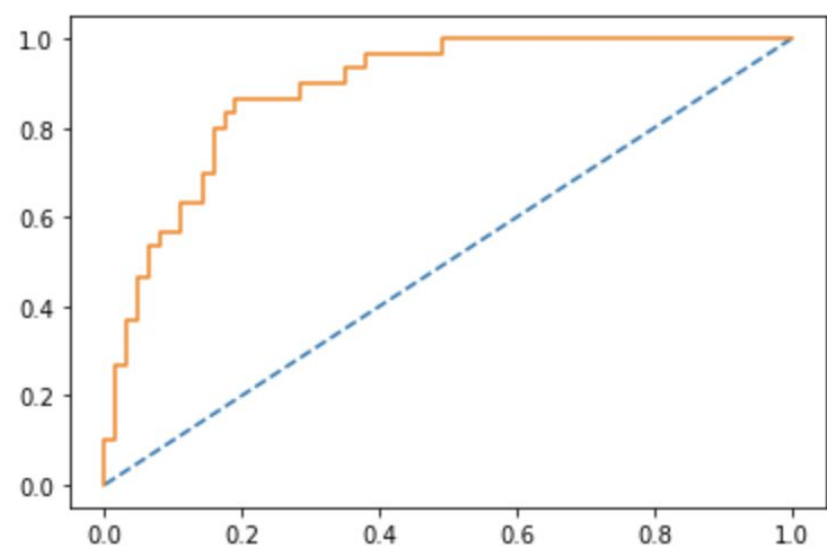
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.90      | 0.83   | 0.86     | 63      |
| 1            | 0.69      | 0.80   | 0.74     | 30      |
| accuracy     |           |        | 0.82     | 93      |
| macro avg    | 0.79      | 0.81   | 0.80     | 93      |
| weighted avg | 0.83      | 0.82   | 0.82     | 93      |

Best Parameters: {'class\_\_algorithm': 'auto', 'class\_\_leaf\_size': 1, 'class\_\_n\_neighbors': 1, 'class\_\_p': 2, 'class\_\_weights': 'uniform'}



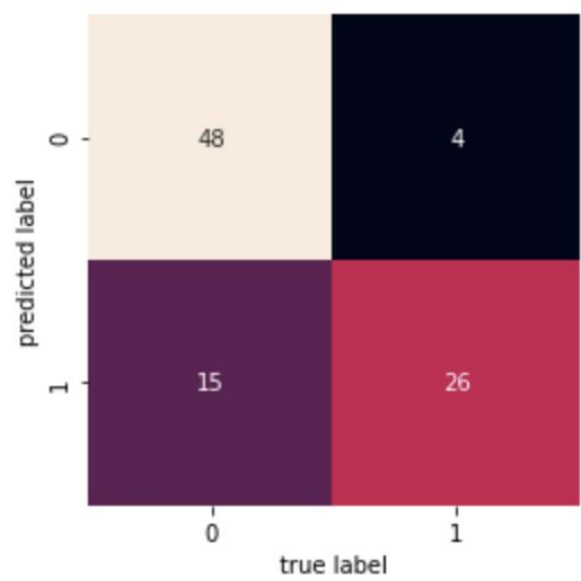
SMOTE - GUASSIAN NAIVE BAYSE

Training score: 0.8425925925925926  
area under curve (auc): 0.8873015873015873



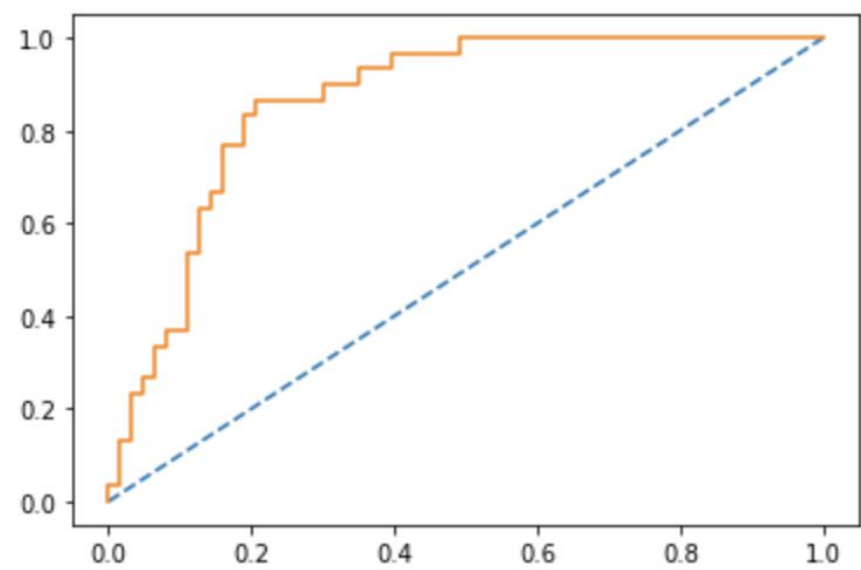
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.92      | 0.76   | 0.83     | 63      |
| 1            | 0.63      | 0.87   | 0.73     | 30      |
| accuracy     |           |        | 0.80     | 93      |
| macro avg    | 0.78      | 0.81   | 0.78     | 93      |
| weighted avg | 0.83      | 0.80   | 0.80     | 93      |

Best Parameters: {'class\_\_var\_smoothing': 0.1}



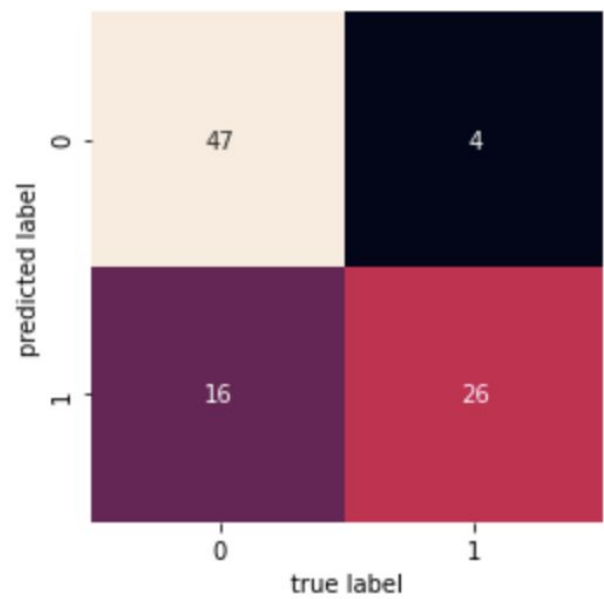
ADASYN - GUASSIAN NAIVE BAYES

Training score: 0.8425925925925926  
area under curve (auc): 0.864021164021164



|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.92      | 0.75   | 0.82     | 63      |
| 1            | 0.62      | 0.87   | 0.72     | 30      |
| accuracy     |           |        | 0.78     | 93      |
| macro avg    | 0.77      | 0.81   | 0.77     | 93      |
| weighted avg | 0.82      | 0.78   | 0.79     | 93      |

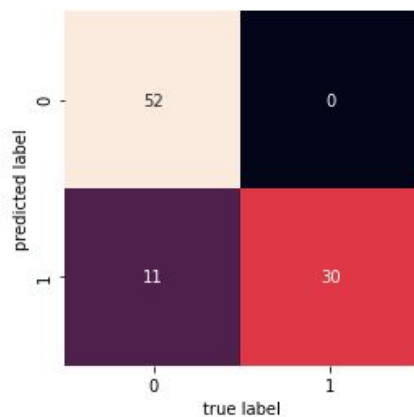
Best Parameters: {'class\_\_var\_smoothing': 0.1}



## Results for top performer - Logistic Regression

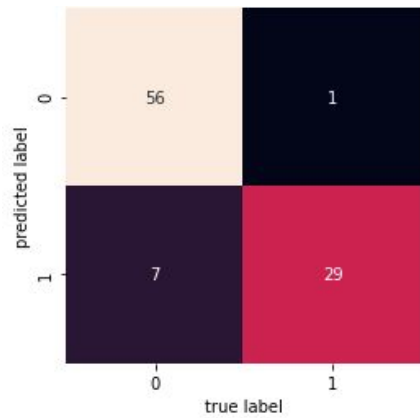
Logistic Regression Results (SMOTE, LDA, C=0.001, penalty='l2', solver = 'newton-cg')

A heavy bias resulting in 100% recall of "Normal" classification



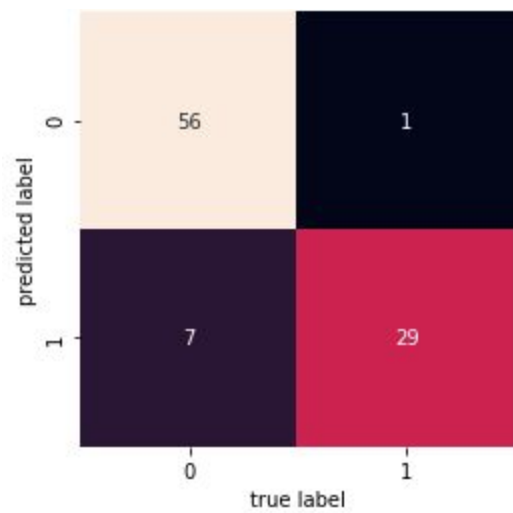
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 1.00      | 0.83   | 0.90     | 63      |
| 1            | 0.73      | 1.00   | 0.85     | 30      |
| accuracy     |           |        | 0.88     | 93      |
| macro avg    | 0.87      | 0.91   | 0.87     | 93      |
| weighted avg | 0.91      | 0.88   | 0.89     | 93      |

Balanced bias - without the Linear transformation

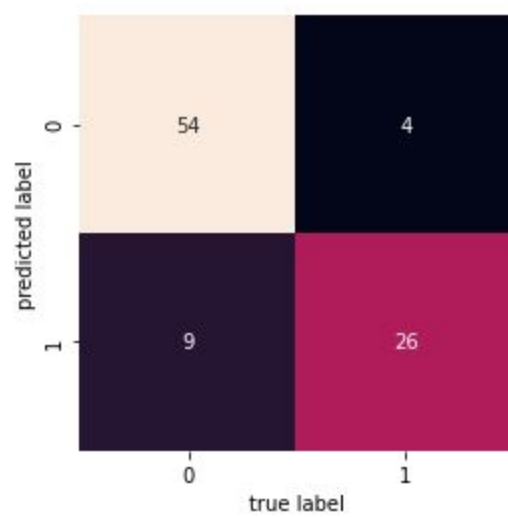


**Results for all models without LDA transformation (more balanced bias)**

Logistic Regression

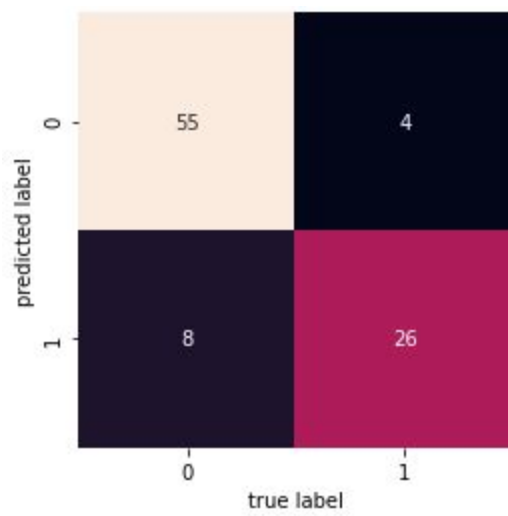
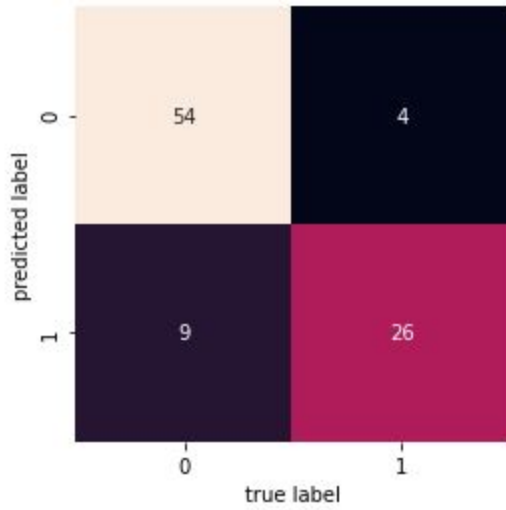


Gradient Boosting Classifier

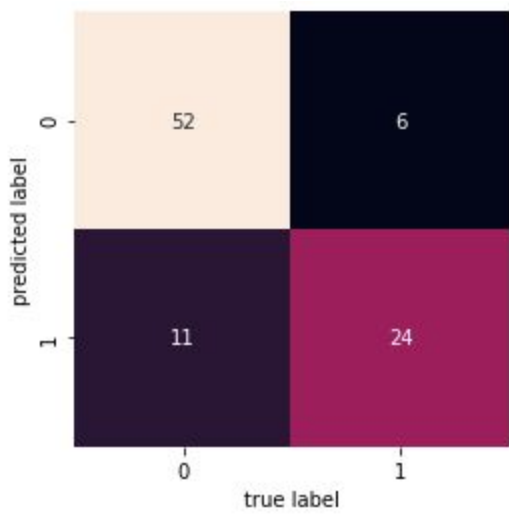


Support Vector Machine

Random Forest Classifier

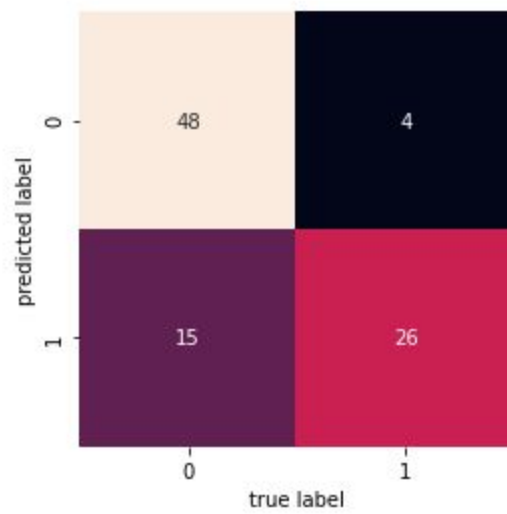


KNneighbors



Gaussian Naive Bayes

Best Parameters: {'class\_var\_s





```

LogisticRegression(random_state=42)
      precision    recall  f1-score   support

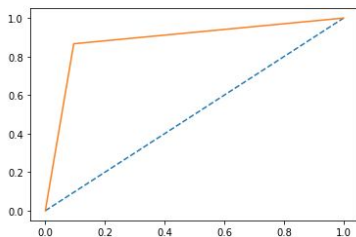
     0       0.93      0.90      0.92         63
     1       0.81      0.87      0.84         30

 accuracy          0.89         93
  macro avg       0.87      0.89      0.88         93
 weighted avg     0.90      0.89      0.89         93

 area under curve (auc):  0.8857142857142857

```

area under curve (auc): 0.8857142857142857



```

GradientBoostingClassifier(random_state=42)
      precision    recall  f1-score   support

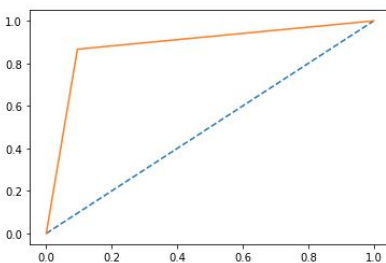
     0       0.93      0.90      0.92         63
     1       0.81      0.87      0.84         30

 accuracy          0.89         93
  macro avg       0.87      0.89      0.88         93
 weighted avg     0.90      0.89      0.89         93

 area under curve (auc):  0.8857142857142857

```

area under curve (auc): 0.8857142857142857

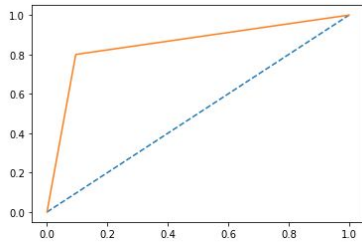


SVC(random\_state=42)

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.90      | 0.90   | 0.90     | 63      |
| 1            | 0.80      | 0.80   | 0.80     | 30      |
| accuracy     |           |        | 0.87     | 93      |
| macro avg    | 0.85      | 0.85   | 0.85     | 93      |
| weighted avg | 0.87      | 0.87   | 0.87     | 93      |

area under curve (auc): 0.8523809523809524

area under curve (auc): 0.8523809523809524

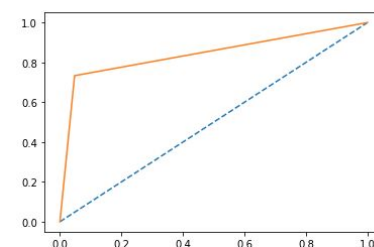


RandomForestClassifier(random\_state=42)

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.88      | 0.95   | 0.92     | 63      |
| 1            | 0.88      | 0.73   | 0.80     | 30      |
| accuracy     |           |        | 0.88     | 93      |
| macro avg    | 0.88      | 0.84   | 0.86     | 93      |
| weighted avg | 0.88      | 0.88   | 0.88     | 93      |

area under curve (auc): 0.8428571428571429

area under curve (auc): 0.8428571428571429

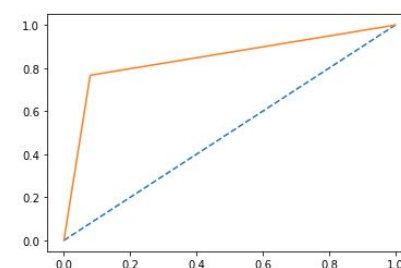


KNeighborsClassifier()

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.89      | 0.92   | 0.91     | 63      |
| 1            | 0.82      | 0.77   | 0.79     | 30      |
| accuracy     |           |        | 0.87     | 93      |
| macro avg    | 0.86      | 0.84   | 0.85     | 93      |
| weighted avg | 0.87      | 0.87   | 0.87     | 93      |

area under curve (auc): 0.8436507936507938

area under curve (auc): 0.8436507936507938



GaussianNB()

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.91      | 0.81   | 0.86     | 63      |
| 1            | 0.68      | 0.83   | 0.75     | 30      |
| accuracy     |           |        | 0.82     | 93      |
| macro avg    | 0.79      | 0.82   | 0.80     | 93      |
| weighted avg | 0.83      | 0.82   | 0.82     | 93      |

area under curve (auc): 0.8214285714285714

area under curve (auc): 0.8214285714285714

