```
In [ ]:
```

In [17]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plot
import seaborn as sns
from sklearn.preprocessing import StandardScaler

import warnings
warnings.filterwarnings('ignore')
```

In []:

In [18]:

cardioDataset = pd.read_csv('heart.csv')

In [19]:

cardioDataset.head(30)

Out[19]:

	age	sex	ср	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
5	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
6	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
7	44	1	1	120	263	0	1	173	0	0.0	2	0	3	1
8	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
9	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
10	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
11	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
12	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
13	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
14	58	0	3	150	283	1	0	162	0	1.0	2	0	2	1
15	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
16	58	0	2	120	340	0	1	172	0	0.0	2	0	2	1
17	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
18	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
19	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
20	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1

		•	_			_	•		_		-	_	_	•
21	age 44	sex 1	cp 2	trestbps 130	chol 233	fbs 0	restecg	thalach 179	exang	oldpeak 0.4	slope 2	ca 0	thal 2	target
22	42	1	0	140	226	0	1	178	0	0.0	2	0	2	1
23	61	1	2	150	243	1	1	137	1	1.0	1	0	2	1
24	40	1	3	140	199	0	1	178	1	1.4	2	0	3	1
25	71	0	1	160	302	0	1	162	0	0.4	2	2	2	1
26	59	1	2	150	212	1	1	157	0	1.6	2	0	2	1
27	51	1	2	110	175	0	1	123	0	0.6	2	0	2	1
28	65	0	2	140	417	1	0	157	0	8.0	2	1	2	1
29	53	1	2	130	197	1	0	152	0	1.2	0	0	2	1

In [20]:

cardioDataset.shape

Out[20]:

(303, 14)

In [21]:

cardioDataset.describe()

Out[21]:

	age	sex	ср	trestbps	chol	fbs	restecg	thalach	exang	oldpe
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.0000
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.0396
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.1610
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.0000
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.0000
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.8000
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.6000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.2000
4							1			Þ

In [22]:

cardioDataset.isnull().sum()

Out[22]:

age 0 0 sex 0 ср trestbps 0 chol 0 fbs 0 0 restecg 0 thalach 0 exang oldpeak 0 0 slope 0 са thal target 0 dtype: int64

In [23]:

cardioDataset.columns = ['Age', 'Sex', 'ChestPainType', 'RestingBloodPressure', 'Cholest

In [24]:

cardioDataset.dtypes

Out[24]:

Age int64 int64 Sex ChestPainType int64 RestingBloodPressure int64 Cholesterol int64 FastingBloodSugar int64 RestingECG int64 MaxHeartRate int64 ExerciseInducedAngina int64 PreviousPeak float64 Slope int64 MajorBloodVessels int64 ThalRate int64 ProbHA int64

dtype: object

In [25]:

```
#Correlation matrix

plot.figure(figsize = (15, 10))

correlation = cardioDataset.corr()

mask = np.triu(np.ones_like(correlation, dtype=bool))

sns.heatmap(correlation, annot = True, cmap='BuGn', linewidths=1)
plot.show()
```

1.0

- 0.8

- 0.6

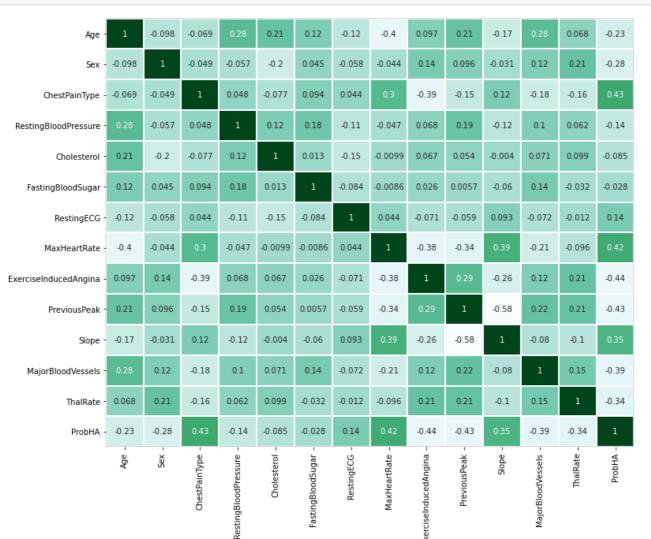
- 04

- 0.2

0.0

- -0.2

-0.4



In [26]:

```
chart = plot.figure(figsize = (14, 14))
ax = chart.gca()
cardioDataset.hist(ax=ax, color="skyblue")
```

Out[26]:

125

100 75

> 50 25

125 100

> 75 50

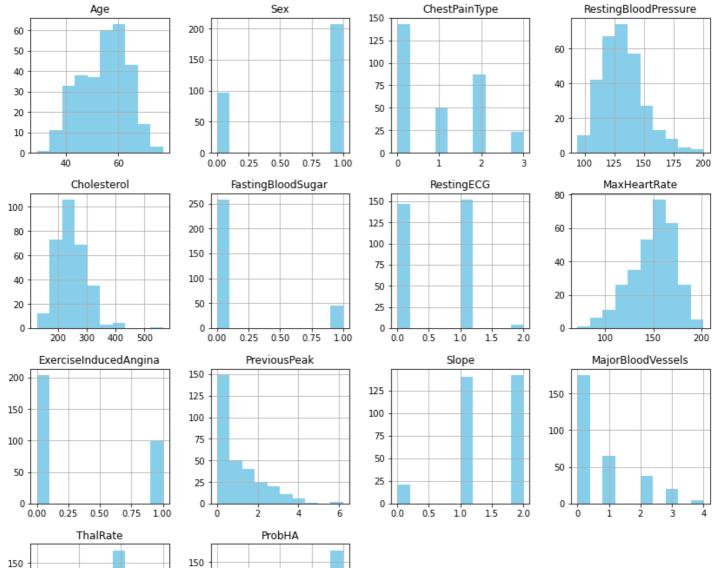
25

0.00

0.25

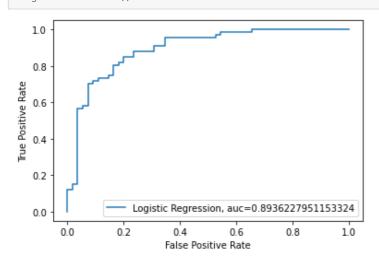
0.50

0.75



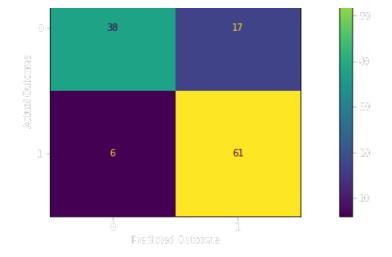
```
In [27]:
cardioDataset.var()
Out [27]:
                           82.484558
Age
                            0.217166
Sex
                            1.065132
ChestPainType
RestingBloodPressure
                         307.586453
Cholesterol
                         2686.426748
FastingBloodSugar
                            0.126877
RestingECG
                            0.276528
MaxHeartRate
                          524.646406
ExerciseInducedAngina
                          0.220707
                            1.348095
PreviousPeak
                            0.379735
Slope
MajorBloodVessels
                            1.045724
ThalRate
                            0.374883
ProbHA
                            0.248836
dtype: float64
In [28]:
cardioDataset['RestingBloodPressure']=np.log(cardioDataset['RestingBloodPressure'])
cardioDataset["Cholesterol"] = np.log(cardioDataset["Cholesterol"])
cardioDataset["MaxHeartRate"] = np.log(cardioDataset["MaxHeartRate"])
np.var(cardioDataset[["RestingBloodPressure", "Cholesterol", "MaxHeartRate"]])
Out[28]:
RestingBloodPressure
                        0.016894
Cholesterol
                        0.041401
                        0.027054
MaxHeartRate
dtype: float64
In [29]:
x = cardioDataset.drop('ProbHA', axis=1)
y = cardioDataset["ProbHA"]
In [30]:
#splitting the dataset
from sklearn.model selection import train test split
x train, x test, y train, y test = train test split(x,y,test size=0.40, random state=43)
In [40]:
#Logistic Regression
accuracies Of Algorithms= {}
from sklearn.linear model import LogisticRegression
from sklearn.metrics import accuracy score, confusion matrix, classification report, f1 s
core, roc auc score, roc curve, plot confusion matrix, precision recall curve, plot precisi
on recall curve
def logisticModel():
    logisticRegression = LogisticRegression(penalty='12')
    logisticRegression.fit(x train, y train)
```

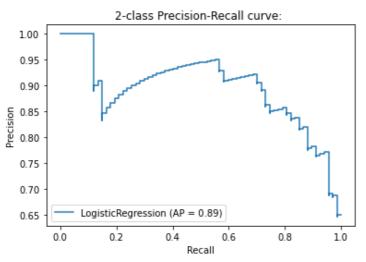
```
y_pred = logisticRegression.predict(x_test)
   y predProbability = logisticRegression.predict proba(x test)[::,1]
   falsePositiveRate, truePositiveRate, = roc curve(y test, y predProbability)
   auc = roc auc score(y test, y predProbability)
   plot.plot(falsePositiveRate, truePositiveRate, label="Logistic Regression, auc="+str
(auc))
   plot.legend(loc=4)
   plot.ylabel("True Positive Rate")
   plot.xlabel("False Positive Rate")
   plot.show()
   accuracy = accuracy score(y test, y pred)
   accuracies Of Algorithms["Logistic Regression"] = accuracy*100
   print('Accuracy score of Logistic Regression is:' , accuracy_score(y_test, y_pred)*1
00, "%")
   print("Confusion Matrix of Logistic Regression", confusion matrix(y test, y pred))
   print("ClassReport", classification report(y test, y pred))
   plottedMatrix = plot confusion matrix(logisticRegression, x test, y test)
   plottedMatrix.ax .set title('Confusion Matrix Of Logistic Regression', color='white')
   plot.xlabel("Predicted Outcome", color='white')
   plot.ylabel("Actual Outcome", color='white')
   plot.gcf().axes[1].tick params(colors='white')
   plot.gcf().axes[0].tick params(colors='white')
   plot.gcf().set size inches(15,5)
   plot.show()
   disp = plot precision recall curve(logisticRegression, x test, y test)
   disp.ax .set title('2-class Precision-Recall curve: '
logisticModel()
```



Accuracy score of Logistic Regression is: 81.14754098360656 % Confusion Matrix of Logistic Regression [[38 17] [6 61]]

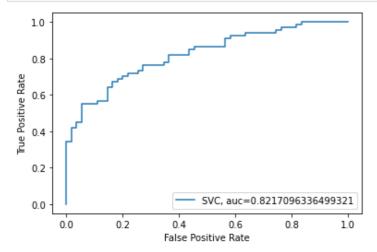
ClassReport		precision	recall	f1-score	support
0	0.86	0.69	0.77	55	
1	0.78	0.91	0.84	67	
accuracy			0.81	122	
macro avg	0.82	0.80	0.80	122	
weighted avg	0.82	0.81	0.81	122	





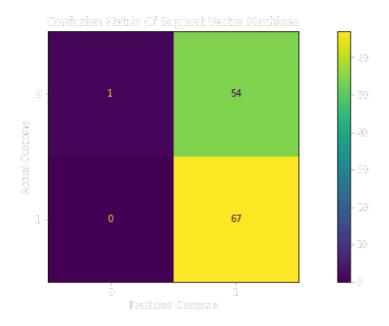
In [32]:

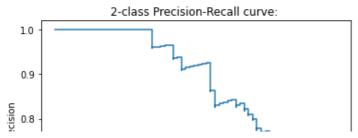
```
#support vector
from sklearn.svm import SVC
def svcClassifier():
   svc = SVC(probability=True)
   svc.fit(x train, y train)
   y pred1 = svc.predict(x test)
   y_predProbability1 = svc.predict_proba(x_test)[::,1]
   falsePositiveRate, truePositiveRate, = roc curve(y test, y predProbability1)
   auc = roc_auc_score(y_test, y_predProbability1)
   plot.plot(falsePositiveRate, truePositiveRate, label="SVC, auc="+str(auc))
   plot.ylabel("True Positive Rate")
   plot.xlabel("False Positive Rate")
   plot.legend(loc=4)
   plot.show()
   accuracy1 = accuracy score(y test, y pred1)
   accuracies Of Algorithms['supportVectorMachine'] = accuracy1*100
   accuracy score(y train, svc.predict(x train))
   print("Accuracy score of the model is:", accuracy_score(y_test, y_pred1)*100, "%")
   print("Confusion matrix of the model", confusion matrix(y test, y pred1))
   print("Classification Report", classification report(y test, y pred1))
   plottedMatrix = plot confusion matrix(svc, x test, y test)
   plottedMatrix.ax .set title('Confusion Matrix Of Support Vector Machines', color='whi
```



Accuracy score of the model is: 55.73770491803278 % Confusion matrix of the model [[1 54] [0 67]]

Classification R	Report		precision	recall	f1-score	support
0	1.00	0.02	0.04	55		
1	0.55	1.00	0.71	67		
accuracy			0.56	122		
macro avg	0.78	0.51	0.37	122		
weighted avg	0.75	0.56	0.41	122		





```
0.7

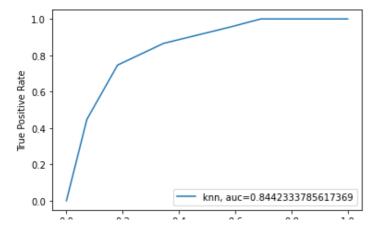
0.6 SVC (AP = 0.87)

0.0 0.2 0.4 0.6 0.8 1.0

Recall
```

In [33]:

```
#kNearestNeighbours
from sklearn.neighbors import KNeighborsClassifier
def knnClassifier():
    knn = KNeighborsClassifier()
    knn.fit(x_train, y_train)
    y_pred2 = knn.predict(x_test)
    y predProbability2 = knn.predict proba(x test)[::,1]
    falsePositiveRate, truePositiveRate, _ = roc_curve(y_test, y_predProbability2)
    auc = roc_auc_score(y_test, y_predProbability2)
   plot.plot(falsePositiveRate, truePositiveRate, label="knn, auc="+str(auc))
   plot.legend(loc=4)
    plot.ylabel("True Positive Rate")
   plot.xlabel("False Positive Rate")
   plot.show()
    accuracy2 = accuracy_score(y_test, y_pred2)
    accuracies Of Algorithms['KNeighborsClassifier'] = accuracy2*100
    accuracy score(y train, knn.predict(x train))
    print("Accuracy score of the KNN:", accuracy score(y test, y pred2)*100, "%")
    print("Confusion matrix of the model", confusion matrix(y test, y pred2))
    print("Classification Report", classification_report(y_test, y_pred2))
    plottedMatrix = plot confusion matrix(knn,x test, y test)
    plottedMatrix.ax_.set_title('Confusion Matrix Of knn', color='white')
    plot.xlabel("Predicted Outcome", color='white')
   plot.ylabel("Actual Outcome", color='white')
    plot.gcf().axes[1].tick_params(colors='white')
    plot.gcf().axes[0].tick params(colors='white')
   plot.gcf().set size inches(15,5)
   plot.show()
    disp = plot precision recall curve(knn, x test, y test)
    disp.ax_.set_title('2-class Precision-Recall curve: ')
knnClassifier()
```

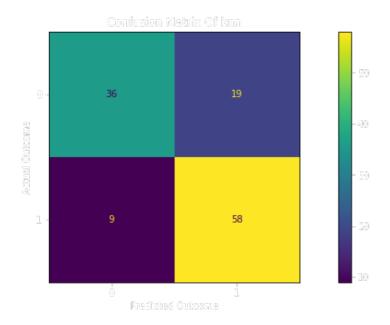


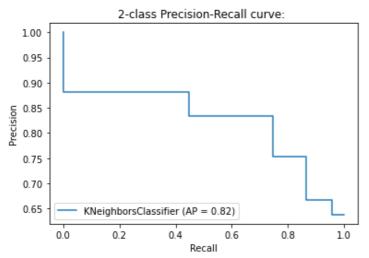
0.0 0.2 0.4 0.6 0.8 1.0 False Positive Rate

Accuracy score of the KNN: 77.04918032786885 % Confusion matrix of the model [[36 19]

[9 58]]

Classification Report		precision	recall	f1-score	support
0 0.80	0.65	0.72	55		
1 0.75	0.87	0.81	67		
accuracy		0.77	122		
macro avg 0.78	0.76	0.76	122		
weighted avg 0.77	0.77	0.77	122		





In [34]:

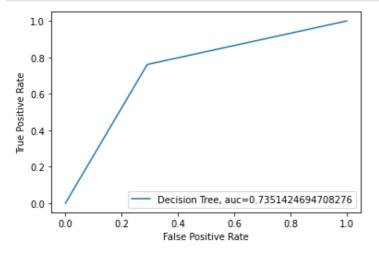
```
from sklearn.tree import DecisionTreeClassifier

def decisionClassifier():
    dt = DecisionTreeClassifier()
    dt.fit(x_train, y_train)

y_pred3 = dt.predict(x_test)
    y_predProbability3 = dt.predict_proba(x_test)[::,1]
    falsePositiveRate, truePositiveRate, _ = roc_curve(y_test, y_predProbability3)
    auc = roc_auc_score(y_test, y_predProbability3)

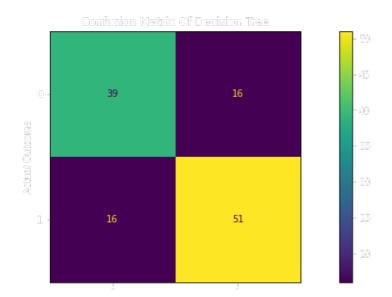
plot.plot(falsePositiveRate, truePositiveRate, label="Decision Tree, auc="+str(auc))
    plot.legend(loc=4)
    plot.ylabel("True Positive Rate")
    plot.xlabel("False Positive Rate")
    plot.show()
```

```
accuracy3= accuracy_score(y_test, y_pred3)
    accuracies Of Algorithms['DecisionTreeClassifier'] = accuracy3*100
    accuracy score(y train, dt.predict(x train))
   print("Accuracy score of the model is:", accuracy score(y test,y pred3)*100, "%")
    print("Confusion matrix of the model", confusion matrix(y test, y pred3))
   print("Classification Report", classification report(y test, y pred3))
    plottedMatrix = plot confusion matrix(dt, x test, y test)
    plottedMatrix.ax .set title('Confusion Matrix Of Decision Tree', color='white')
    plot.xlabel("Predicted Outcome", color='white')
    plot.ylabel("Actual Outcome", color='white')
    plot.gcf().axes[1].tick_params(colors='white')
    plot.gcf().axes[0].tick params(colors='white')
    plot.gcf().set_size_inches(15,5)
    disp = plot_precision_recall_curve(dt, x_test, y_test)
    disp.ax .set title('2-class Precision-Recall curve: ')
decisionClassifier()
```

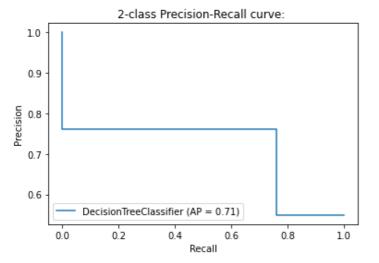


Accuracy score of the model is: 73.77049180327869 % Confusion matrix of the model [[39 16] [16 51]]

Classification	Report		precision	recall	f1-score	support
0 1	0.71 0.76	0.71 0.76	0.71 0.76	55 67		
accuracy macro avg weighted avg	0.74 0.74	0.74 0.74	0.74 0.74 0.74	122 122 122		

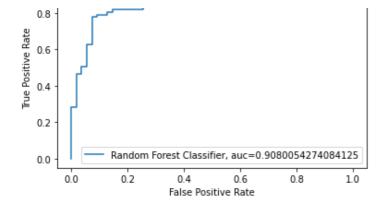


Predicted Outcome



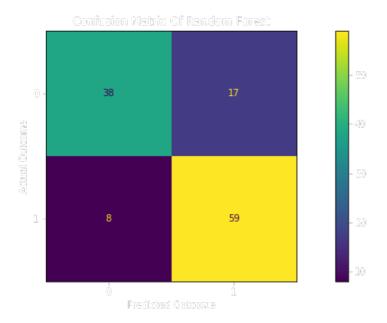
In [35]:

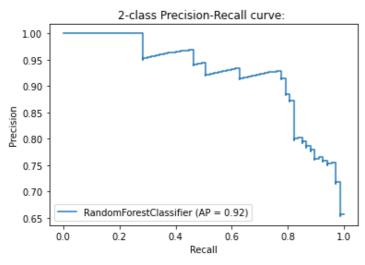
```
from sklearn.ensemble import RandomForestClassifier
def randomForest():
   rfc = RandomForestClassifier(criterion = 'gini', max depth = 7, max features = 'sqr
t',
                            min samples leaf = 2, min samples split = 4, n estimators =
180)
   rfc.fit(x train, y train)
    y pred5 = rfc.predict(x test)
    y predProbability5 = rfc.predict proba(x test)[::,1]
   falsePositiveRate, truePositiveRate, _ = roc_curve(y_test, y_predProbability5)
    auc = roc auc score(y test, y predProbability5)
    plot.plot(falsePositiveRate, truePositiveRate, label="Random Forest Classifier, auc=
"+str(auc))
    plot.legend(loc=4)
    plot.ylabel("True Positive Rate")
    plot.xlabel("False Positive Rate")
   plot.show()
    accuracy5 = accuracy score(y test, y pred5)
    accuracies Of Algorithms['RandomForestClassifier'] = accuracy5*100
    accuracy score(y train, rfc.predict(x train))
   print("Accuracy score of the model is:", accuracy_score(y_test, y pred5)*100, "%")
   print("Confusion matrix of the model", confusion matrix(y test, y pred5))
   print("Classification Report", classification_report(y_test, y_pred5))
   plottedMatrix = plot_confusion_matrix(rfc, x_test, y_test)
   plottedMatrix.ax_.set_title('Confusion Matrix Of Random Forest', color='white')
    plot.xlabel("Predicted Outcome", color='white')
    plot.ylabel("Actual Outcome", color='white')
    plot.gcf().axes[1].tick params(colors='white')
    plot.gcf().axes[0].tick params(colors='white')
    plot.gcf().set size inches(15,5)
    plot.show()
    disp = plot_precision_recall_curve(rfc, x_test, y_test)
    disp.ax .set title('2-class Precision-Recall curve: ')
randomForest()
```



Accuracy score of the model is: 79.50819672131148 % Confusion matrix of the model [[38 17] [8 59]]

Classification	Report		precision	recall	f1-score	support
0 1	0.83 0.78	0.69	0.75 0.83	55 67		
accuracy			0.80	122		
macro avg weighted avg	0.80	0.79 0.80	0.79 0.79	122 122		



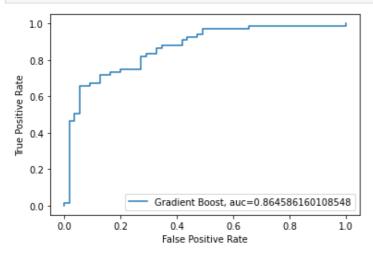


In [36]:

```
from sklearn.ensemble import GradientBoostingClassifier

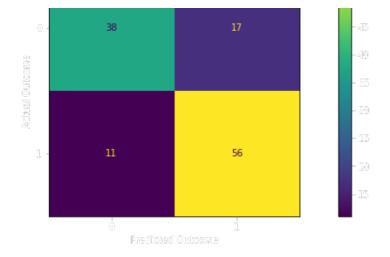
def GradientBoost():
```

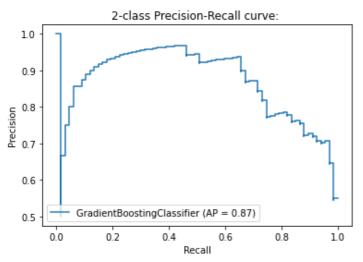
```
gbc = GradientBoostingClassifier()
   gbc.fit(x_train, y_train)
   y pred6 = gbc.predict(x test)
   y predProbability6 = gbc.predict proba(x test)[::,1]
   falsePositiveRate, truePositiveRate, = roc curve(y_test, y_predProbability6)
   auc = roc_auc_score(y_test, y_predProbability6)
   plot.plot(falsePositiveRate, truePositiveRate, label="Gradient Boost, auc="+str(auc)
   plot.legend(loc=4)
   plot.ylabel("True Positive Rate")
   plot.xlabel("False Positive Rate")
   plot.show()
   accuracy6=accuracy score(y test, y pred6)
   accuracies Of Algorithms['GradientBoosting'] = accuracy6*100
   print("Accuracy score of the model is:", accuracy_score(y_test, y_pred6)*100, "%")
   print("Confusion matrix of the model", confusion matrix(y test, y pred6))
   print("Classification Report", classification report(y test, y pred6))
   plottedMatrix = plot confusion matrix(gbc, x test, y test)
   plottedMatrix.ax .set title('Confusion Matrix Of Gradient Boost', color='white')
   plot.xlabel("Predicted Outcome", color='white')
   plot.ylabel("Actual Outcome", color='white')
   plot.gcf().axes[1].tick params(colors='white')
   plot.gcf().axes[0].tick params(colors='white')
   plot.gcf().set size inches(15,5)
   plot.show()
   disp = plot precision recall curve(gbc, x test, y test)
   disp.ax .set title('2-class Precision-Recall curve: ')
GradientBoost()
```



Accuracy score of the model is: 77.04918032786885 % Confusion matrix of the model [[38 17] [11 56]]

Classific	cation	Report		precision	recall	f1-score	support
	0	0.78	0.69	0.73	55		
	1	0.77	0.84	0.80	67		
accui	racy			0.77	122		
macro	avg	0.77	0.76	0.77	122		
weighted	avg	0.77	0.77	0.77	122		



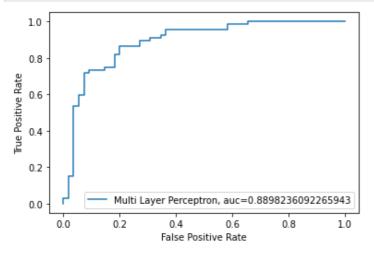


In [37]:

```
from sklearn.neural network import MLPClassifier
def multiLayerClassifier():
    multiLayerPerceptron = MLPClassifier(random state = 10, max iter=300, activation="relu
Π,
                                     hidden layer sizes=(34,34,34))
   multiLayerPerceptron.fit(x train, y train)
    y pred7 = multiLayerPerceptron.predict(x test)
    y predProbability7 = multiLayerPerceptron.predict_proba(x_test)[::,1]
    falsePositiveRate, truePositiveRate, = roc curve(y test, y predProbability7)
    auc = roc auc score(y test, y predProbability7)
   plot.plot(falsePositiveRate, truePositiveRate, label="Multi Layer Perceptron, auc="+
str(auc))
   plot.legend(loc=4)
   plot.ylabel("True Positive Rate")
   plot.xlabel("False Positive Rate")
    plot.show()
    accuracy7 = accuracy score(y test, y pred7)
    accuracies Of Algorithms['MLPClassifier'] = accuracy7*100
    print("Confusion Matrix of MLPClassifier", confusion_matrix(y_test, y_pred7))
    print("Classification Report", classification_report(y_test, y_pred7))
    plottedMatrix = plot confusion matrix(multiLayerPerceptron, x test, y test)
   plottedMatrix.ax .set title('Confusion Matrix Of MultiLayer PERCEPTRON', color='white
• )
   plot.xlabel("Predicted Outcome", color='white')
    plot.ylabel("Actual Outcome", color='white')
    plot.gcf().axes[1].tick params(colors='white')
```

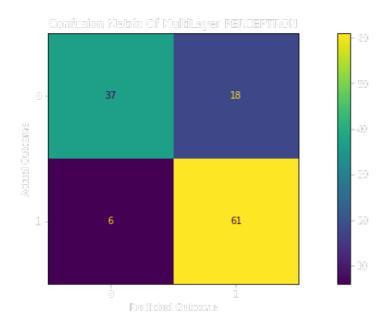
```
plot.gcf().axes[0].tick_params(colors='white')
plot.gcf().set_size_inches(15,5)
plot.show()

accuracy_score(y_test, y_pred7)
print("Accuracy score of the model is:", accuracy_score(y_test, y_pred7)*100, "%")
disp = plot_precision_recall_curve(multiLayerPerceptron, x_test, y_test)
disp.ax_.set_title('2-class Precision-Recall curve: ')
multiLayerClassifier()
```

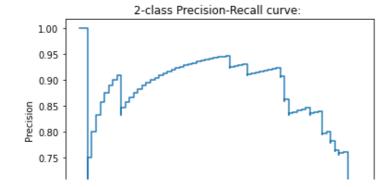


Confusion Matrix of MLPClassifier [[37 18]
 [6 61]]

Classification	on Report		precision	recall	f1-score	support
0	0.86	0.67	0.76	55		
1	0.77	0.91	0.84	67		
accuracy			0.80	122		
macro avg	0.82	0.79	0.80	122		
weighted avg	0.81	0.80	0.80	122		



Accuracy score of the model is: 80.32786885245902 %



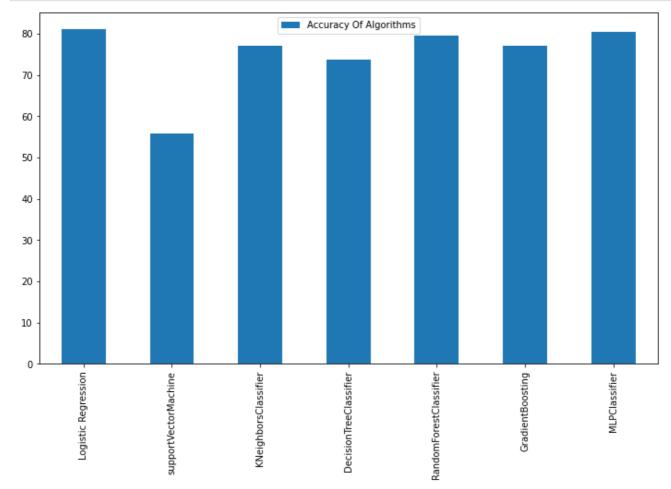
```
0.70

0.65 MLPClassifier (AP = 0.88)

0.0 0.2 0.4 0.6 0.8 1.0

Recall
```

In [38]:



In []:

In []:

In []: