



A.Y. 2022-2023

Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

Experiment 03

Code:

Naïve bayes

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dataset=pd.read_csv('diabetes.csv')
X= dataset.iloc[:, :-1].values
y= dataset.iloc[:, -1].values
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score,confusion_matrix
clf1 = GaussianNB()
clf1.fit(X_train, y_train)
predicted1 = clf1.predict(X_test)
accuracy1 = accuracy_score(predicted1,y_test)
print('Accuracy_score=',accuracy1)

from sklearn.metrics._plot.confusion_matrix import ConfusionMatrixDisplay
lables=[0,1]
cm1=confusion_matrix(y_test,predicted1,labels=lables)
disp1=ConfusionMatrixDisplay(confusion_matrix=cm1,display_labels=lables)
disp1.plot()
```

Output:

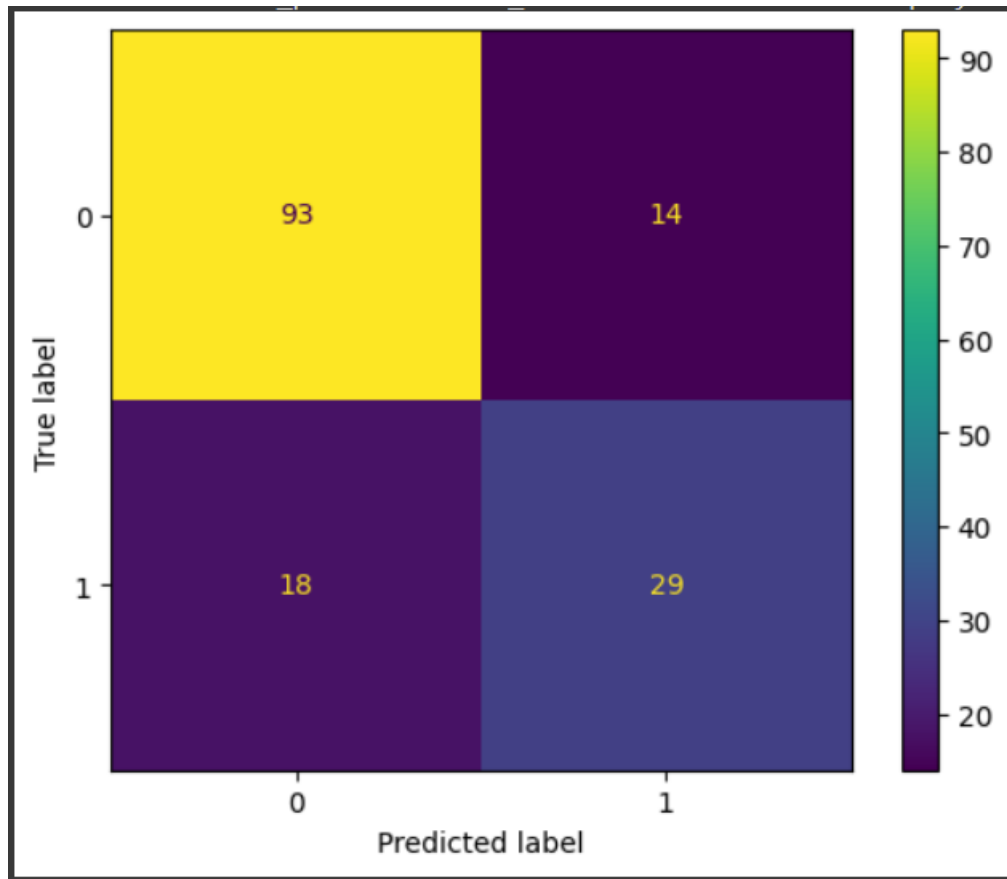
```
Accuracy_score= 0.7922077922077922
```



A.Y. 2022-2023

Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta



```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import roc_curve, roc_auc_score, auc
fpr, tpr, thresholds = roc_curve(y_test, predicted1)
auroc = roc_auc_score(y_test, predicted1)
print(f"AUROC: {auroc}")
```

AUROC: 0.7430900775502088

```
plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, color='darkorange', lw=2, label=f'AUROC = {auroc:.2f}')
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
```

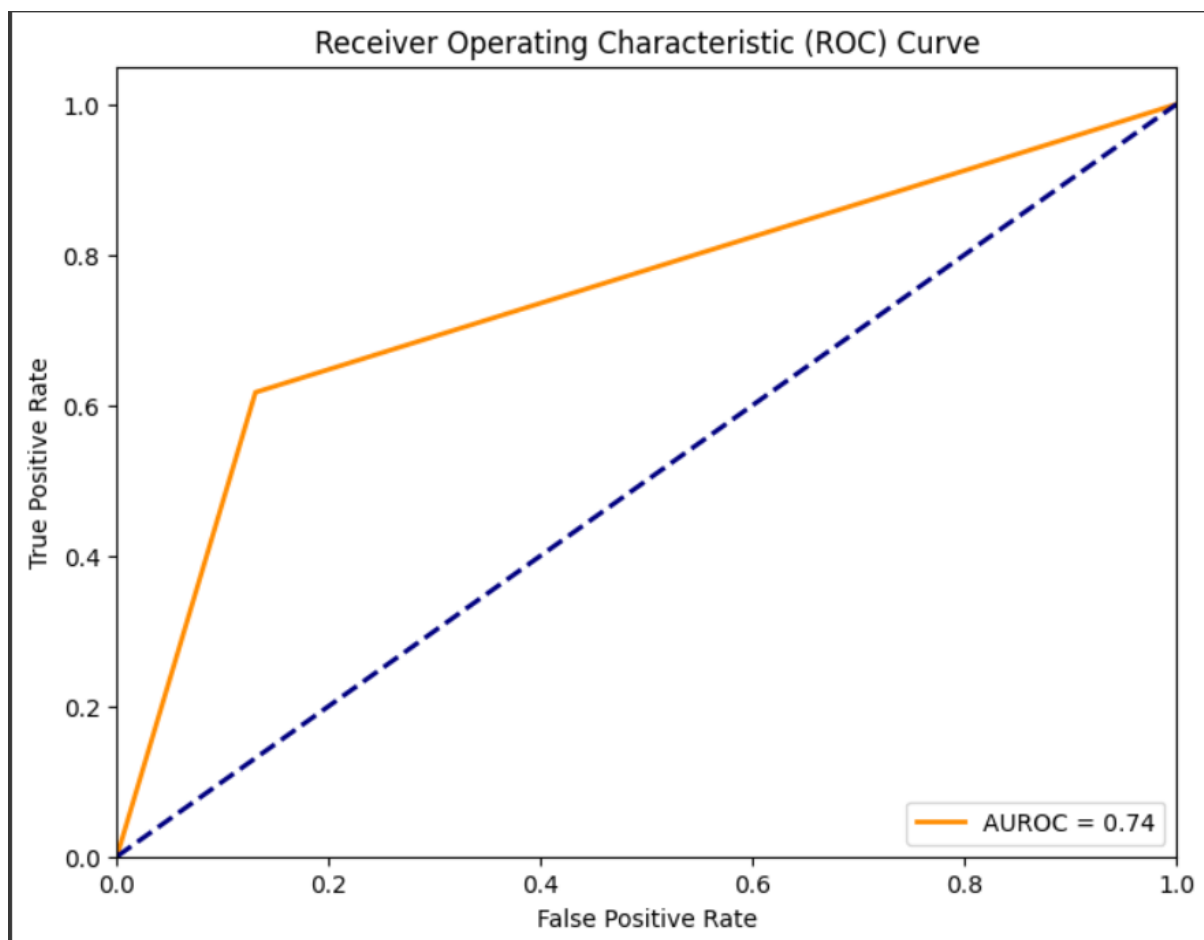


A.Y. 2022-2023

Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

```
plt.xlabel('False Positive Rate')  
plt.ylabel('True Positive Rate')  
plt.title('Receiver Operating Characteristic (ROC) Curve')  
plt.legend(loc='lower right')  
plt.show()
```





A.Y. 2022-2023

Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

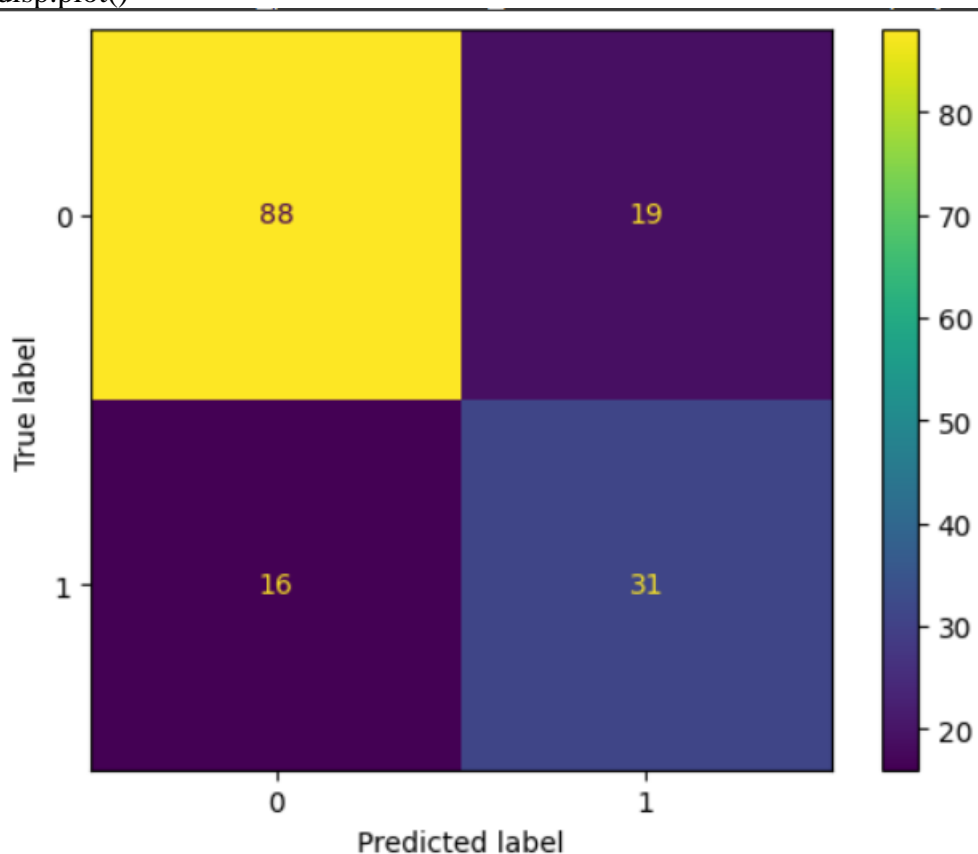
Decision Tree

```
from sklearn.tree import DecisionTreeClassifier  
clf = DecisionTreeClassifier()  
clf.fit(X_train,y_train)  
DT_predicted=clf.predict(X_test)  
DT1 = accuracy_score(DT_predicted, y_test)
```

```
print("Accuracy:", DT1)
```

Accuracy: 0.7727272727272727

```
labels = [0,1]  
cm = confusion_matrix(y_test, DT_predicted, labels=labels)  
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=labels)  
disp.plot()
```





Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



A.Y. 2022-2023

Subject: Data Mining and Warehousing

SAP ID: 60004220253 – Devansh Mehta

```
fpr, tpr, thresholds = roc_curve(y_test, predicted1)
auroc = roc_auc_score(y_test, DT_predicted)
print(f"AUROC: {auroc}")
```

```
AUROC: 0.7410021873135813
```

K Fold

```
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
kfold = KFold(n_splits = 10, shuffle = True , random_state=42)
scores = cross_val_score(clf1,X,y,cv=kfold,scoring='accuracy')
for i in scores:
    print(i)
```

```
0.7012987012987013
0.8311688311688312
0.6883116883116883
0.7922077922077922
0.8051948051948052
0.6753246753246753
0.8571428571428571
0.7142857142857143
0.6578947368421053
0.7894736842105263
```