**Assignment 10**

**Roll no.: A-44**

**Subject: DAP**

**Code:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.datasets import load\_diabetes

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score,classification\_report,confusion\_matrix,roc\_curve,auc

mydata=pd.read\_csv("diabeties.csv")

print(mydata.head(10))

X=mydata.drop(columns=['Outcome'])

y=mydata['Outcome']

y\_binary=( y >np.median(y)).astype(int)

X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y\_binary,test\_size=0.2,random\_state=42)

scaler=StandardScaler()

X\_train=scaler.fit\_transform(X\_train)

X\_test=scaler.transform(X\_test)

model=LogisticRegression()

model.fit(X\_train,y\_train)

y\_pred=model.predict(X\_test)

accuracy = accuracy\_score(y\_test,y\_pred)

print()

print("Accuracy : {:.2f}%".format(accuracy\*100))

print("Confusion Matrix\n",confusion\_matrix(y\_test,y\_pred))

print("\nClassification report\n",classification\_report(y\_test,y\_pred))

plt.figure(figsize=(8,6))

sns.scatterplot(x=X\_test[:,0],y=X\_test[:,1],hue=y\_test,palette={0:'blue',1:'red'},marker='o')

plt.title("Logistic Regression Decision Boundary\nAccuracy : {:.2f}%".format(accuracy\*100))

plt.ylabel("Age")

plt.xlabel("BMI")

plt.legend(title="Diabetes",loc="upper right")

plt.show()

y\_prob=model.predict\_proba(X\_test)[:,1]

fpr,tpr,thresholds=roc\_curve(y\_test,y\_prob)

roc\_auc=auc(fpr,tpr)

plt.figure(figsize=(8,6))

plt.plot(fpr,tpr,color='darkorange',lw=2,label=f'ROC Curve (AUC = {roc\_auc:.2f})')

plt.plot([0,1],[0,1],color='navy', lw=2, linestyle='--', label='Random')

plt.xlabel("False Positive Rate")

plt.ylabel("True positive rate")

plt.title("Receiver Operating Characteristics (ROC) Curve\nAccuracy : {:.2f}%".format(accuracy\*100))

plt.legend(loc="lower right")

plt.show()

**Output:**

Pregnancies Glucose BloodPressure SkinThickness Insulin BMI DiabetesPedigreeFunction Age Outcome

0 6 148 72 35 0 33.6 0.627 50 1

1 1 85 66 29 0 26.6 0.351 31 0

2 8 183 64 0 0 23.3 0.672 32 1

3 1 89 66 23 94 28.1 0.167 21 0

4 0 137 40 35 168 43.1 2.288 33 1

5 5 116 74 0 0 25.6 0.201 30 0

6 3 78 50 32 88 31.0 0.248 26 1

7 10 115 0 0 0 35.3 0.134 29 0

8 2 197 70 45 543 30.5 0.158 53 1

9 8 125 96 0 0 0.0 0.232 54 1

Accuracy : 75.32%

Confusion Matrix

[[79 20]

[18 37]]

Classification report

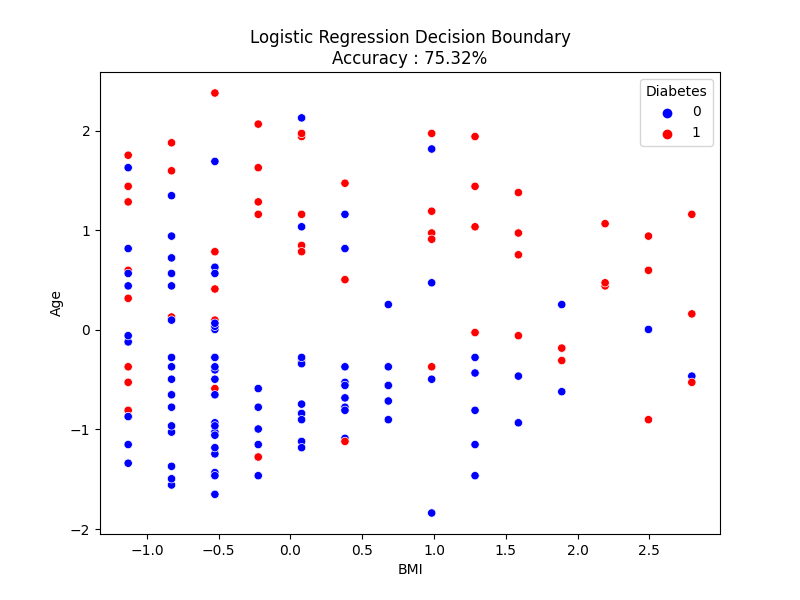
precision recall f1-score support

0 0.81 0.80 0.81 99

1 0.65 0.67 0.66 55

accuracy 0.75 154

macro avg 0.73 0.74 0.73 154

weighted avg 0.76 0.75 0.75 154

