Q2. Write a Python program build Decision Tree Classifier using Scikit- learn package for

diabetes data set (download database from

https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset?resource=down

load )

Ans

import pandas as pd

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

from sklearn.tree import DecisionTreeClassifier

from sklearn.metrics import accuracy\_score, classification\_report

data = pd.read\_csv('C:/Users/HP/Downloads/diabetes.csv')

X = data.drop('Outcome', axis=1)

y = data['Outcome']

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

model = DecisionTreeClassifier()

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

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

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

report = classification\_report(y\_test, y\_pred)

accuracy, report

Output:

(0.7532467532467533,

' precision recall f1-score support\n\n 0 0.84 0.77 0.80 99\n 1

0.63 0.73 0.68 55\n\n accuracy 0.75 154\n macro avg 0.74

0.75 0.74 154\nweighted avg 0.76 0.75 0.76 154\n')

predict class label for show starring a 40 years old American comedian, with 10 years of

experience, and a comedy ranking of 7? Create a csv file as shown in

https://github.com/mahesh147/Decision-Tree-Classifier/blob/master/Social\_Network\_Ads.

csv

Ans

import pandas as pd

from sklearn.tree import DecisionTreeClassifier

data = pd.read\_csv("C:/Users/HP/Downloads/Social\_Network\_Ads.csv")

X = data[['Age', 'EstimatedSalary', 'Gender']]

y = data['Purchased']

X['Gender'] = X['Gender'].map({'Male': 1, 'Female': 0})

clf = DecisionTreeClassifier()

clf.fit(X, y)

prediction = clf.predict([[40, 0, 1]]) # 40 years, EstimatedSalary set to 0, Male

print(f"Prediction: {prediction[0]}")

Q2. Write a Python program build Decision Tree Classifier using Scikit-learnpackage for

diabetes data set (download database from

https://www.kaggle.com/uciml/pima-indiansdiabetes-database)

https://github.com/npradaschnor/Pima-Indians-Diabetes-Dataset/blob/master/diabetes.cs

v

Ans

import pandas as pd

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

from sklearn.tree import DecisionTreeClassifier

from sklearn.metrics import accuracy\_score, classification\_report

url = 'https://github.com/npradaschnor/Pima-Indians-Diabetes-Dataset/raw/master/diabetes.csv'

data = pd.read\_csv(url)

X = data.drop('Outcome', axis=1)

y = data['Outcome']

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

clf = DecisionTreeClassifier()

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

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

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

print(f'Accuracy: {accuracy:.2f}')

print(classification\_report(y\_test, y\_pred))

Output:

Accuracy: 0.77

precision recall f1-score support

0 0.84 0.79 0.81 99

1 0.66 0.73 0.69 55

accuracy 0.77 154

macro avg 0.75 0.76 0.75 154

weighted avg 0.77 0.77 0.77 154