TASK 3

Build a decision tree classifier to predict whether a customer will purchase a product or service based on their demographic and behavioral data. Use a dataset such as the Bank Marketing dataset from the UCI Machine Learning Repository.

SOURCE CODE

import pandas as pd

import matplotlib.pyplot as plt

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

from sklearn.tree import DecisionTreeClassifier,plot\_tree

from sklearn.metrics import accuracy\_score

from sklearn.preprocessing import LabelEncoder

df = pd.read\_csv(r'C:\Users\ANIRBAN CHOWDHURY\Downloads\bank-full (1).csv', sep=';')

df.head()

df.isna().sum()

#Preprocess the data

X = df.drop('poutcome', axis=1)

y = df['poutcome']

X = pd.get\_dummies(X)

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("Accuracy:", accuracy)

#Decision tree

plt.figure(figsize=(10, 8))

plot\_tree(clf, feature\_names=X.columns, class\_names=y.unique().tolist(), filled=True, rounded=True)

plt.show()

clf\_pruned = DecisionTreeClassifier(ccp\_alpha=0.01)

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

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

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

print("Accuracy with pruning:", accuracy\_pruned)

#Pruned decision tree

plt.figure(figsize=(10, 8))

plot\_tree(clf\_pruned, feature\_names=X.columns, class\_names=y.unique().tolist(), filled=True, rounded=True)

plt.show()



