Name: Ayush Chanchal

Sap id: 500097569

Batch: B6

Exeriment:03

Source code:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# from scipy import stats

data=pd.read\_csv('./wineQualityReds.csv')

# print(data.head())

# print(data.describe())

# print(data.isna().sum())

plt.figure(figsize=(30,20))

corr=data.corr()

sns.heatmap(corr,annot=True)

# plt.show()

from scipy import stats

z=np.abs(stats.zscore(data))

# print(z)

# print(np.where(z>3))

dataset=data[(z<3).all(axis=1)]

# print(dataset.shape)

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

x=dataset.drop(columns='quality')

y=dataset['quality']

# print(x.head())

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2)

from sklearn.ensemble import RandomForestClassifier

rf=RandomForestClassifier(n\_estimators=100)

rf.fit(x\_train,y\_train)

y\_pred=rf.predict(x\_test)

# print(y\_pred)

from sklearn import metrics

print('Accuracy score:',metrics.accuracy\_score(y\_test,y\_pred))

from sklearn import tree

plt.figure(figsize=(20,20))

tree.plot\_tree(rf.estimators\_[0],filled=True)

# for i in range(len(rf.estimators\_)):

# tree.plot\_tree(rf.estimators\_[i],filled=True)

ScreenShot:

