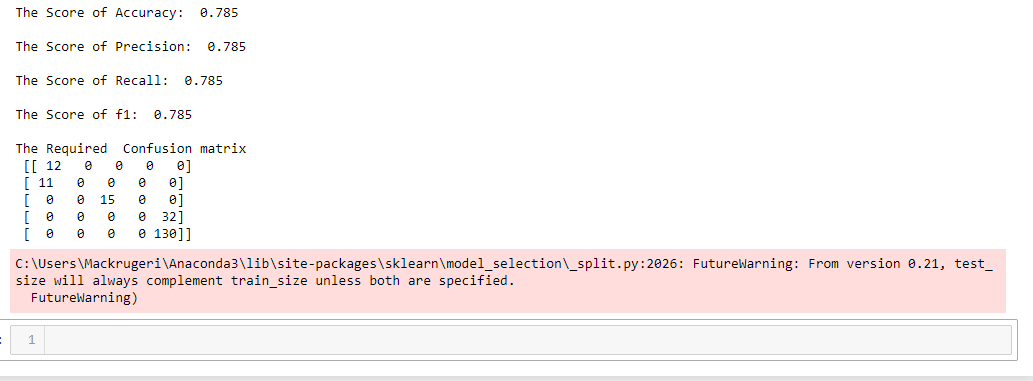
Name: Muhammad Ahsan Ali

Roll No: p176008

Quiz 2 Natural Language processing

Output of Quiz 2 is



Code

import re

from sklearn.feature\_extraction.text import CountVectorizer

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.tree import DecisionTreeClassifier

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

from sklearn.metrics import accuracy\_score

from sklearn.metrics import precision\_score

from sklearn.metrics import confusion\_matrix

from sklearn.metrics import recall\_score

from sklearn.metrics import f1\_score

def accuracy\_finding(test\_y,labels):

print('The Score of Accuracy: ', accuracy\_score(test\_y, labels), '\n')

def precision\_finding(test\_y,labels):

print('The Score of Precision: ',precision\_score(test\_y, labels, average = 'micro'), '\n')

def recalling\_finding(test\_y,labels):

print('The Score of Recall: ',recall\_score(test\_y, labels, average = 'micro'), '\n')

def f1\_finding(test\_y,labels):

print('The Score of f1: ',f1\_score(test\_y, labels, average = 'micro'), '\n')

def confusion\_finding(test\_y,labels):

print('The Required Confusion matrix\n',confusion\_matrix(test\_y, labels))

label = []

y = []

def complete\_processing():

corpus = open('Movies\_TV.txt').read()

rows = corpus.split('\n')

rows.remove(rows[0])

rows.remove(rows[-1])

for row in rows:

\_, labels, ratings, \_ = row.split('\t')

label.append(labels)

y.append(ratings)

vec = CountVectorizer(min\_df = 8)

X = vec.fit\_transform(label)

vec = TfidfVectorizer(min\_df = 8)

matrix\_X = vec.fit\_transform(label)

dtc = DecisionTreeClassifier(max\_depth = 8)

train\_x, test\_x, train\_y, test\_y = train\_test\_split(matrix\_X, y, shuffle = True, train\_size = 0.8)

dtc.fit(train\_x, train\_y)

labels = dtc.predict(test\_x)

accuracy\_finding(test\_y,labels)

precision\_finding(test\_y,labels)

recalling\_finding(test\_y,labels)

f1\_finding(test\_y,labels)

confusion\_finding(test\_y,labels)

complete\_processing()

Google Colab Link

<https://colab.research.google.com/drive/1NfrlIL3zk7gS3jeGqJvib8YGsk3A0e8h#scrollTo=R_4JYOFMrCqY>