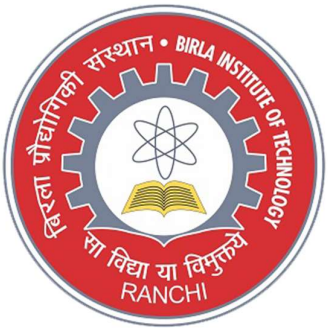
Birla Institute of Technology,Mesra,

Patna Campus



ML-Assignment

Name-Shubham Sourabh

Roll-Btech/15044/18

Sec-CSE 6th

#Assignment-6

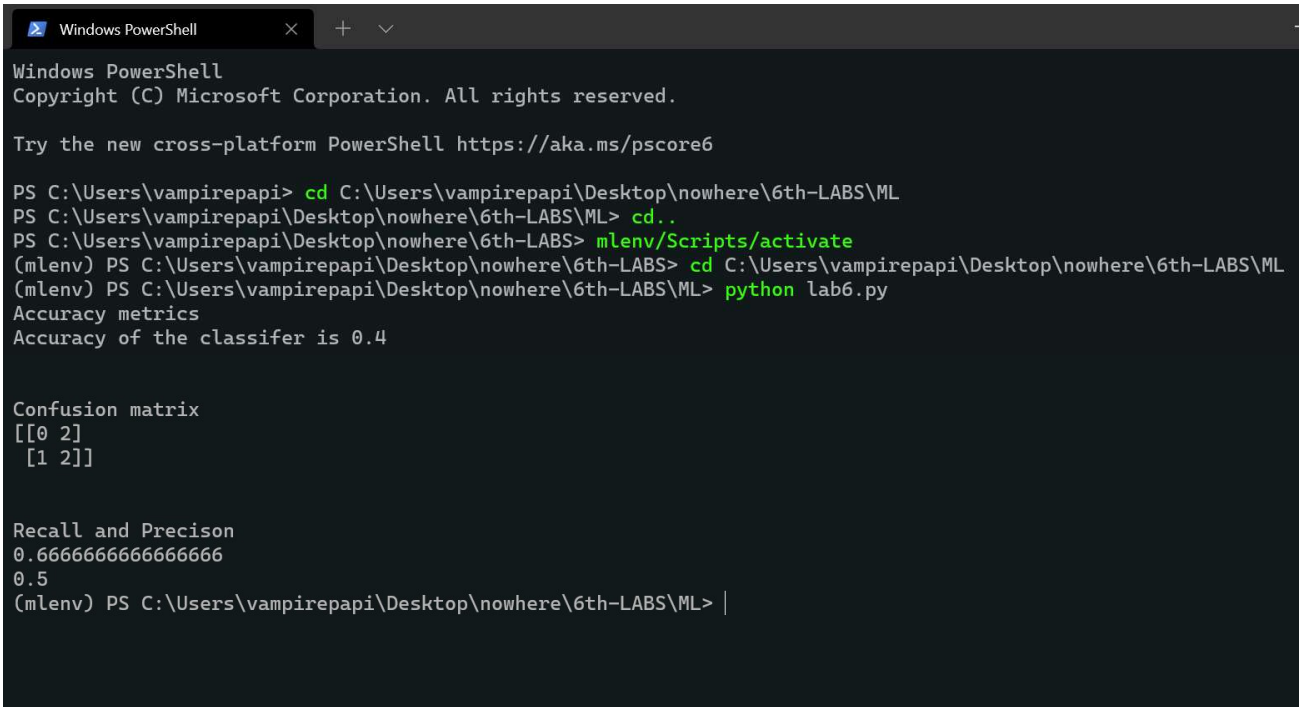
Objective: P. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your dataset.

Code:

```
import pandas as pd

import pdb

msg=pd.read_csv('data6.csv',names=['message','label']) #names-> name of the cols

msg['labelnum']=msg.label.map({'pos':1,'neg':0})

X=msg.message

Y=msg.labelnum


from sklearn.model_selection import train_test_split

xtrain,xtest,ytrain,ytest=train_test_split(X,Y)


from sklearn.feature_extraction.text import CountVectorizer

count_vect = CountVectorizer()


xtrain_dtm = count_vect.fit_transform(xtrain)

xtest_dtm=count_vect.transform(xtest)
```

```
df=pd.DataFrame(xtrain_dtm.toarray(),columns=count_vect.get_feature_nam
es())

from sklearn.naive_bayes import MultinomialNB
clf = MultinomialNB().fit(xtrain_dtm,ytrain)
predicted = clf.predict(xtest_dtm)

from sklearn import metrics
print('Accuracy metrics')
print('Accuracy of the classifer is',metrics.accuracy_score(ytest,predicted))
print('\n')
print('Confusion matrix')
print(metrics.confusion_matrix(ytest,predicted))
print('\n')
print('Recall and Precison ')
print(metrics.recall_score(ytest,predicted))
print(metrics.precision_score(ytest,predicted))
#pdb.set_trace()
```

Output:


Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.
Try the new cross-platform PowerShell https://aka.ms/pscore6
PS C:\Users\vampirepapi> cd C:\Users\vampirepapi\Desktop\nowhere\6th-LABS\ML
PS C:\Users\vampirepapi\Desktop\nowhere\6th-LABS\ML> cd..
PS C:\Users\vampirepapi\Desktop\nowhere\6th-LABS> mlenv/Scripts/activate
(mlenv) PS C:\Users\vampirepapi\Desktop\nowhere\6th-LABS> cd C:\Users\vampirepapi\Desktop\nowhere\6th-LABS\ML
(mlenv) PS C:\Users\vampirepapi\Desktop\nowhere\6th-LABS\ML> python lab6.py
Accuracy metrics
Accuracy of the classifer is 0.4
Confusion matrix
[[0 2]
[1 2]]
Recall and Precison
0.6666666666666666
0.5
(mlenv) PS C:\Users\vampirepapi\Desktop\nowhere\6th-LABS\ML>