### Naive Bayesian Classifier

#### Lab 6: Write a Program to implement the naive bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier few test data sets.

from sklearn.metrics import accuracy\_score, confusion\_matrix, precision\_score, recall\_score

from sklearn.naive\_bayes import MultinomialNB

from sklearn.feature\_extraction.text import CountVectorizer

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

import pandas as pd

msg = pd.read\_csv('ns.csv', names=['message', 'label'])

print("Total Instances of Dataset: ", msg.shape[0])

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

X = msg.message

y = msg.labelnum

Xtrain, Xtest, ytrain, ytest = train\_test\_split(X, y)

count\_v = CountVectorizer()

Xtrain\_dm = count\_v.fit\_transform(Xtrain)

Xtest\_dm = count\_v.transform(Xtest)

df = pd.DataFrame(Xtrain\_dm.toarray(), columns=count\_v.get\_feature\_names())

print(df[0:5])

clf = MultinomialNB()

clf.fit(Xtrain\_dm, ytrain)

pred = clf.predict(Xtest\_dm)

for doc, p in zip(Xtrain, pred):

p = 'pos' if p == 1 else 'neg'

print("%s -> %s" % (doc, p))

print('Accuracy Metrics: \n')

print('Accuracy: ', accuracy\_score(ytest, pred))

print('Recall: ', recall\_score(ytest, pred))

print('Precision: ', precision\_score(ytest, pred))

print('Confusion Matrix: \n', confusion\_matrix(ytest, pred))