**colab link :**

<https://colab.research.google.com/drive/1pIoMfj2oC7QHAUZbWZHL33vSWdLNt_xw>

**code:**

import numpy as np #linear algebria

import pandas as pd #data processing , CSV file I/O

import os

for dirname, \_,filesnames in os.walk('kaggle/input'):

for filename in filenames:

print(os.path.join(dirname, filename))

df = pd.read\_csv('/content/SPAM text message 20170820 - Data.csv')

df.head()

df['Category'].value\_counts()

X = df['Message']

y = df['Category']

len(X)

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

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state= 22)

from sklearn.pipeline import Pipeline

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.naive\_bayes import MultinomialNB, ComplementNB

from sklearn.svm import LinearSVC

from sklearn.metrics import accuracy\_score,classification\_report

pipeMNB = Pipeline([('tfidf',TfidfVectorizer(stop\_words='english', ngram\_range=(1,3))),('clf', MultinomialNB())])

pipeCNB = Pipeline([('tfidf',TfidfVectorizer(stop\_words='english', ngram\_range=(1,3))),('clf', ComplementNB())])

pipeSVC = Pipeline([('tfidf',TfidfVectorizer(stop\_words='english', ngram\_range=(1,3))),('clf', LinearSVC())])

pipeMNB.fit(X\_train, y\_train)

predictMNB = pipeMNB.predict(X\_test)

print(f"MNB: {accuracy\_score(y\_test, predictMNB):.2f}")

pipeCNB.fit(X\_train, y\_train)

predictCNB = pipeCNB.predict(X\_test)

print(f"CNB: {accuracy\_score(y\_test, predictCNB):.2f}")

pipeSVC.fit(X\_train, y\_train)

predictSVC = pipeMNB.predict(X\_test)

print(f"SVC: {accuracy\_score(y\_test, predictSVC):.2f}")

print(classification\_report(y\_test, predictSVC))

msg = "you have won a $10000 prize! contact us for reward!"

clsf = pipeSVC.predict([msg])

print(clsf[0])