5.5 CODING

```
import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)



# Input data files are available in the read-only "../input/" directory

# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory



import os

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

    for filename in filenames:

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

```
from google.colab import drive

drive.mount('/content/drive/')


data=pd.read_csv('/content/drive/MyDrive/Dataset/spam.csv')


data.columns

data.info()

data.isna().sum()


data['Spam']=data['Category'].apply(lambda x:1 if x=='spam' else 0)

data.head(5)


from sklearn.model_selection import train_test_split

X_train,X_test,y_train,y_test=train_test_split(data.Message,data.Spam,test_size=0.2
5)

from sklearn.feature_extraction.text import CountVectorizer

from sklearn.naive_bayes import MultinomialNB

from sklearn.pipeline import Pipeline
```

```
clf=Pipeline([
    ('vectorizer',CountVectorizer()),
    ('nb',MultinomialNB())
])
clf.fit(X_train,y_train)


clf.score(X_test,y_test)


mails=['due to lot of request-,gold membership offer extended only today-final call.','masterclass on database management system-mongoDB and SQL']


clf.predict(mails)


for x in mails:
  y= [x]
  if clf.predict(y)==[0]:
    print("ham")
  else:
    print ("spam")
```