Assignment 4 - SMS SPAM Classification

Team Member - Lakshman K.R. (Roll No:310619104059)

```
[1] import pandas as pd
       import numpy as np
  [2] from google.colab import files
       uploaded = files.upload()
                    spam.csv
       • spam.csv(text/csv) - 503663 bytes, last modified: 27/10/2022 - 100% done
       Saving spam.csv to spam.csv
  [3] df= pd.read_csv("spam.csv", encoding="latin-1")
       df.drop(['Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], axis=1, inplace=True)
  [4] print(df.shape)
       print(df.head(5))
         ham Go until jurong point, crazy.. Available only ...
                                    Ok lar... Joking wif u oni...
       1 ham
       2 spam Free entry in 2 a wkly comp to win FA Cup fina...
       ham U dun say so early hor... U c already then say...
ham Nah I don't think he goes to usf, he lives aro...
[5] df['label'] = df['v1'].map({'ham': 0, 'spam': 1})
    df['message']=df['v2']
    df.drop(['v1','v2'],axis=1,inplace=True)
    X = df['message']
     y = df['label']
from sklearn.feature_extraction.text import CountVectorizer
     cv = CountVectorizer()
    X = cv.fit_transform(X)
[7] from sklearn.model_selection import train_test_split
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42)
[8] from sklearn.naive_bayes import MultinomialNB
    model = MultinomialNB()
    model.fit(X_train,y_train)
    MultinomialNB()
```

```
[10] from sklearn.metrics import accuracy_score
    print("Accuracy of the Model: {0}%".format(accuracy_score(y_test,y_pred)*100))
    Accuracy of the Model: 97.93365959760739%
[11] from sklearn.metrics import classification_report
    print (classification_report(y_test, y_pred))
                 precision recall f1-score
                                               support
                                                1587
               0
                   0.99
                             0.99 0.99
                      0.93
                               0.92
                                        0.92
                                                   252
                                        0.98
                                                  1839
        accuracy
                                       0.98
0.96
       macro avg 0.96 0.95
                                                  1839
                     0.98
                              0.98
                                        0.98
                                                  1839
    weighted avg
    txt = (input("Enter the Mail : "))
    email = [txt]
    vect= cv.transform(email).toarray()
     result = model.predict(vect)
    print(result)
    if result == 1:
      print("Its a Spam Email")
      print("Not a Spam Email")
 Enter the Mail : Free Iphone
     [1]
    Its a Spam Email
```