EmailSpamDetection

September 25, 2025

[46]:
```
#Import Libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import accuracy_score,classification_report
```

[47]:
```
#Dtaaset Loading
df=pd.read_csv("spam.csv",encoding='latin-1')
print("Dataset Loaded Successfully",df)
```

```
Dataset Loaded Successfully        v1
v2 Unnamed: 2  \
0      ham  Go until jurong point, crazy.. Available only …        NaN
1      ham                      Ok lar… Joking wif u oni…        NaN
2     spam  Free entry in 2 a wkly comp to win FA Cup fina…        NaN
3      ham  U dun say so early hor… U c already then say…        NaN
4      ham  Nah I don't think he goes to usf, he lives aro…        NaN
…    …                                                  …        …
5567  spam  This is the 2nd time we have tried 2 contact u…        NaN
5568   ham              Will Ì_ b going to esplanade fr home?        NaN
5569   ham  Pity, * was in mood for that. So…any other s…        NaN
5570   ham  The guy did some bitching but I acted like i'd…        NaN
5571   ham                        Rofl. Its true to its name        NaN

     Unnamed: 3 Unnamed: 4
0           NaN        NaN
1           NaN        NaN
2           NaN        NaN
3           NaN        NaN
4           NaN        NaN
…          …        …
5567        NaN        NaN
5568        NaN        NaN
5569        NaN        NaN
5570        NaN        NaN
5571        NaN        NaN
```

```
[5572 rows x 5 columns]
```

[48]:
```
df=df[["v1","v2"]]
df.columns=['label','message']
```

[49]:
```
df['label']=df['label'].map({'ham': 0, 'spam': 1})
print(df)
```

```
      label                                            message
0         0  Go until jurong point, crazy.. Available only …
1         0                      Ok lar… Joking wif u oni…
2         1  Free entry in 2 a wkly comp to win FA Cup fina…
3         0  U dun say so early hor… U c already then say…
4         0  Nah I don't think he goes to usf, he lives aro…
…     …                                                  …
5567      1  This is the 2nd time we have tried 2 contact u…
5568      0              Will Ì_ b going to esplanade fr home?
5569      0  Pity, * was in mood for that. So…any other s…
5570      0  The guy did some bitching but I acted like i'd…
5571      0                        Rofl. Its true to its name

[5572 rows x 2 columns]
```

[50]:
```
#Feature Selection
X=df['message']
y=df['label']
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.
  ↪2,random_state=42)
```

[51]:
```
#Vectorization
V=TfidfVectorizer(stop_words='english')
X_train=V.fit_transform(X_train)
X_test = V.transform(X_test)
```

[52]:
```
NB=MultinomialNB()
NB.fit(X_train,y_train)
ypred=NB.predict(X_test)
```

[53]:
```
#Evaluation of model
Acc=accuracy_score(y_test,ypred)
print("Accuracy of mode is:",Acc*100,"%")
```

```
Accuracy of mode is: 96.68161434977578 %
```

[56]:
```
#Classification report
report=classification_report(y_test,ypred)
print(report)
```

```
              precision    recall  f1-score   support
```

```
              0       0.96      1.00      0.98       965
              1       1.00      0.75      0.86       150

    accuracy                            0.97      1115
   macro avg       0.98      0.88      0.92      1115
weighted avg       0.97      0.97      0.96      1115
```

[58]:
```
#Prediction for message
def predict(message):
    messagevec=V.transform([message])
    pred=NB.predict(messagevec)[0]
    if pred==1:
        return "Spam"
    else:
        return "ham"
print(predict("Congratulations!You won a cash price"))
```

```
Spam
```

[59]:
```
print(predict("Hi,How are you ,Can we meet tomorrow"))
```

```
ham
```