

```
import re
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import nltk
df_train=pd.read_csv('/content/SMS_train (1).csv',encoding='ISO-8859-1')
df_test=pd.read_csv('/content/SMS_test (2).csv',encoding='ISO-8859-1')
df=pd.concat([df_train,df_test])
```

```
df.head()
```



| | S. No. | Message_body | Label |
|---|--------|---|----------|
| 0 | 1 | Rofl. Its true to its name | Non-Spam |
| 1 | 2 | The guy did some bitching but I acted like i'd... | Non-Spam |
| 2 | 3 | Pity, * was in mood for that. So...any other s... | Non-Spam |
| 3 | 4 | Will ü b going to esplanade fr home? | Non-Spam |
| 4 | 5 | This is the 2nd time we have tried 2 contact u... | Spam |

```
df.describe
```

```
<bound method NDFrame.describe of      S. No.      Message_body      Label
0         1      Rofl. Its true to its name  Non-Spam
1         2  The guy did some bitching but I acted like i'd...  Non-Spam
2         3  Pity, * was in mood for that. So...any other s...  Non-Spam
3         4      Will ü b going to esplanade fr home?  Non-Spam
4         5  This is the 2nd time we have tried 2 contact u...    Spam
..      ...      ...      ...
120      121  7 wonders in My WORLD 7th You 6th Ur style 5th...  Non-Spam
121      122  Try to do something dear. You read something f...  Non-Spam
122      123  Sun ah... Thk mayb can if dun have anythin on...  Non-Spam
123      124  SYMPTOMS when U are in love: "1.U like listeni...  Non-Spam
124      125  Great. Have a safe trip. Dont panic surrender ...  Non-Spam

[1082 rows x 3 columns]>
```

```
df.info
```

```
<bound method DataFrame.info of      S. No.      Message_body      Label
0         1      Rofl. Its true to its name  Non-Spam
1         2  The guy did some bitching but I acted like i'd...  Non-Spam
2         3  Pity, * was in mood for that. So...any other s...  Non-Spam
3         4      Will ü b going to esplanade fr home?  Non-Spam
4         5  This is the 2nd time we have tried 2 contact u...    Spam
..      ...      ...      ...
120      121  7 wonders in My WORLD 7th You 6th Ur style 5th...  Non-Spam
121      122  Try to do something dear. You read something f...  Non-Spam
122      123  Sun ah... Thk mayb can if dun have anythin on...  Non-Spam
123      124  SYMPTOMS when U are in love: "1.U like listeni...  Non-Spam
124      125  Great. Have a safe trip. Dont panic surrender ...  Non-Spam

[1082 rows x 3 columns]>
```

```
df.isna().sum()
```

```
S. No.      0
Message_body  0
Label      0
dtype: int64
```

```
df.drop(['S. No.'],axis=1,inplace=True)
```

```
df.head()
```

| | Message_body | Label |
|---|---|----------|
| 0 | Rofl. Its true to its name | Non-Spam |
| 1 | The guy did some bitching but I acted like i'd... | Non-Spam |
| 2 | Pity, * was in mood for that. So...any other s... | Non-Spam |
| 3 | Will ü b going to esplanade fr home? | Non-Spam |
| 4 | This is the 2nd time we have tried 2 contact u... | Spam |

```
df.duplicated().sum() #to check duplicated values
```

```
18
```

```
df.drop_duplicates()
```

| | Message_body | Label |
|-----|---|----------|
| 0 | Rofl. Its true to its name | Non-Spam |
| 1 | The guy did some bitching but I acted like i'd... | Non-Spam |
| 2 | Pity, * was in mood for that. So...any other s... | Non-Spam |
| 3 | Will ü b going to esplanade fr home? | Non-Spam |
| 4 | This is the 2nd time we have tried 2 contact u... | Spam |
| ... | ... | ... |
| 120 | 7 wonders in My WORLD 7th You 6th Ur style 5th... | Non-Spam |
| 121 | Try to do something dear. You read something f... | Non-Spam |
| 122 | Sun ah... Thk mayb can if dun have anythin on.... | Non-Spam |
| 123 | SYMPTOMS when U are in love: "1.U like listeni... | Non-Spam |
| 124 | Great. Have a safe trip. Dont panic surrender ... | Non-Spam |

1064 rows × 2 columns

```
df.reset_index(drop=True,inplace=True)
df
```

| | Message_body | Label |
|------|---|----------|
| 0 | Rofl. Its true to its name | Non-Spam |
| 1 | The guy did some bitching but I acted like i'd... | Non-Spam |
| 2 | Pity, * was in mood for that. So...any other s... | Non-Spam |
| 3 | Will ü b going to esplanade fr home? | Non-Spam |
| 4 | This is the 2nd time we have tried 2 contact u... | Spam |
| ... | ... | ... |
| 1077 | 7 wonders in My WORLD 7th You 6th Ur style 5th... | Non-Spam |
| 1078 | Try to do something dear. You read something f... | Non-Spam |
| 1079 | Sun ah... Thk mayb can if dun have anythin on.... | Non-Spam |
| 1080 | SYMPTOMS when U are in love: "1.U like listeni... | Non-Spam |
| 1081 | Great. Have a safe trip. Dont panic surrender ... | Non-Spam |

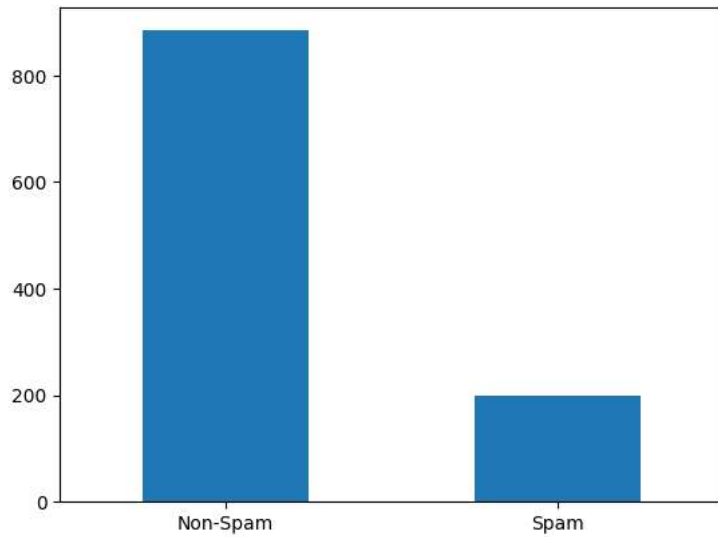
1082 rows × 2 columns

```
df['Label'].value_counts()
```

```
Non-Spam    884
Spam        198
Name: Label, dtype: int64
```

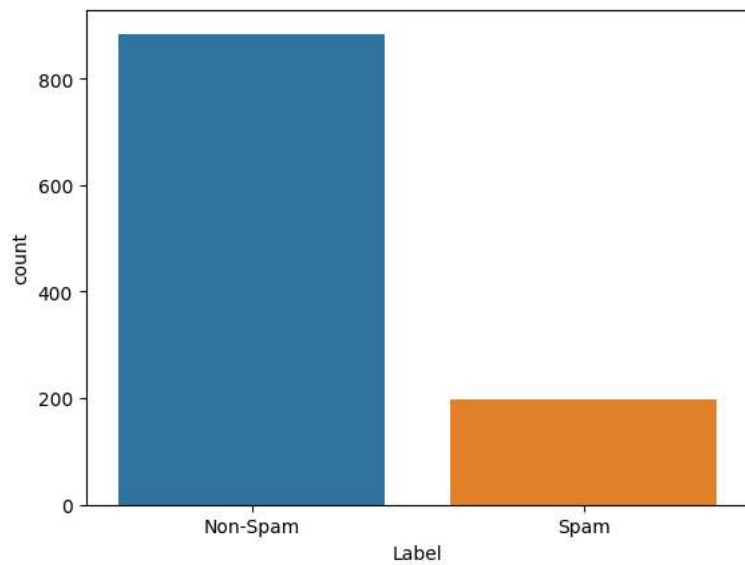
```
df['Label'].value_counts().plot(kind='bar')
plt.xticks(rotation=0)
```

```
(array([0, 1]), [Text(0, 0, 'Non-Spam'), Text(1, 0, 'Spam')])
```



```
sns.countplot(x='Label',data=df)
```

```
<Axes: xlabel='Label', ylabel='count'>
```



```
df['Label']=df['Label'].map({'Non-Spam':0,'Spam':1})  
df
```

| | Message_body | Label |
|------|---|-------|
| 0 | Rofl. Its true to its name | 0 |
| 1 | The guy did some bitching but I acted like i'd... | 0 |
| 2 | Pity, * was in mood for that. So...any other s... | 0 |
| 3 | Will ü b going to esplanade fr home? | 0 |
| 4 | This is the 2nd time we have tried 2 contact u... | 1 |
| ... | ... | ... |
| 1077 | 7 wonders in My WORLD 7th You 6th Ur style 5th... | 0 |
| 1078 | Try to do something dear. You read something f... | 0 |
| 1079 | Sun ah... Thk mayb can if dun have anythin on.... | 0 |
| 1080 | SYMPTOMS when U are in love: "1.U like listeni... | 0 |
| 1081 | Great. Have a safe trip. Dont panic surrender ... | 0 |

1082 rows × 2 columns

```

nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('punkt')
nltk.download('omw-1.4')

```

```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
True

```

```

message=df.Message_body
message

```

```

0          Rofl. Its true to its name
1    The guy did some bitching but I acted like i'd...
2    Pity, * was in mood for that. So...any other s...
3          Will ü b going to esplanade fr home?
4    This is the 2nd time we have tried 2 contact u...
...
1077  7 wonders in My WORLD 7th You 6th Ur style 5th...
1078  Try to do something dear. You read something f...
1079  Sun ah... Thk mayb can if dun have anythin on...
1080  SYMPTOMS when U are in love: "1.U like listeni...
1081  Great. Have a safe trip. Dont panic surrender ...
Name: Message_body, Length: 1082, dtype: object

```

```

#tokenize
from nltk import TweetTokenizer
tk=TweetTokenizer()
message=message.apply(lambda x:tk.tokenize(x)).apply(lambda X: " ".join(X))
message

```

```

0          Rofl . Its true to its name
1    The guy did some bitching but I acted like i'd...
2    Pity , * was in mood for that . So ... any oth...
3          Will ü b going to esplanade fr home ?
4    This is the 2nd time we have tried 2 contact u...
...
1077  7 wonders in My WORLD 7th You 6th Ur style 5th...
1078  Try to do something dear . You read something ...
1079  Sun ah ... Thk mayb can if dun have anythin on...
1080  SYMPTOMS when U are in love : " 1 . U like lis...
1081  Great . Have a safe trip . Dont panic surrende...
Name: Message_body, Length: 1082, dtype: object

```

```

#remove spl char
message=message.str.replace('[^a-zA-z0-9]+',' ')
message

```

```

<ipython-input-19-df67e6e81d70>:2: FutureWarning: The default value of regex will change from True to False in a future version.
message=message.str.replace('[^a-zA-z0-9]+',' ')

```

```

0             ofl ts true to its name
1     he guy did some bitching but acted like i d b...
2     ity was in mood for that o any other suggesti...
3             ill b going to esplanade fr home
4     his is the 2nd time we have tried 2 contact u...

...

1077  7 wonders in y 7th ou 6th r style 5th r smile ...
1078  ry to do something dear ou read something for...
1079  un ah hk mayb can if dun have anythin on hk h...
1080  when are in love 1 like listening songs 2 get...
1081  reat ave a safe trip ont panic surrender all
Name: Message_body, Length: 1082, dtype: object

```

#stemming

```

from nltk import SnowballStemmer
stemmer=SnowballStemmer('english')
message=message.apply(lambda x:[stemmer.stem(i.lower())for i in tk.tokenize(x)]).apply(lambda x:" ".join(x))
message

```

```

0             ofl ts true to it name
1     he guy did some bitch but act like i d be inte...
2             iti was in mood for that o ani other suggest
3             ill b go to esplanad fr home
4     his is the 2nd time we have tri 2 contact u ha...

...

1077  7 wonder in y 7th ou 6th r style 5th r smile 4...
1078  ry to do someth dear ou read someth for exam
1079  un ah hk mayb can if dun have anythin on hk ha...
1080  when are in love 1 like listen song 2 get stop...
1081  reat ave a safe trip ont panic surrend all
Name: Message_body, Length: 1082, dtype: object

```

#importing stopwords

```

from nltk.corpus import stopwords
stop=stopwords.words('english')
message=message.apply(lambda x:[i for i in tk.tokenize(x) if i not in stop]).apply(lambda x:" ".join(x))
message

```

```

0             ofl ts true name
1     guy bitch act like interest buy someth els nex...
2             iti mood ani suggest
3             ill b go esplanad fr home
4     2nd time tri 2 contact u 750 ound prize 2 clai...

...

1077  7 wonder 7th ou 6th r style 5th r smile 4th r ...
1078  ry someth dear ou read someth exam
1079  un ah hk mayb dun anythin hk book e lesson pil...
1080  love 1 like listen song 2 get stop u see name ...
1081  reat ave safe trip ont panic surrend
Name: Message_body, Length: 1082, dtype: object

```

#vectorization

```

from sklearn.feature_extraction.text import TfidfVectorizer
vec=TfidfVectorizer()
train_data=vec.fit_transform(message)
train_data

<1082x3055 sparse matrix of type '<class 'numpy.float64'>'
with 9221 stored elements in Compressed Sparse Row format>

```

train_data.shape

```
(1082, 3055)
```

x=train_data

x

```

<1082x3055 sparse matrix of type '<class 'numpy.float64'>'
with 9221 stored elements in Compressed Sparse Row format>

```

y=df['Label']

```

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.30,random_state=42)

```

```

from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
model1=KNeighborsClassifier()
model2=MultinomialNB()
model3=SVC()
model4=DecisionTreeClassifier()
model5=RandomForestClassifier()
models=[model1,model2,model3,model4,model5]

from sklearn.metrics import classification_report,accuracy_score
for i in models:
    i.fit(x_train,y_train)
    y_pred=i.predict(x_test)
    print(i)
    print("classification report:")
    report=classification_report(y_test,y_pred)
    print(report)

```

```

KNeighborsClassifier()
classification report:
      precision    recall  f1-score   support

     0       0.84      1.00      0.91       269
     1       1.00      0.05      0.10        56

 accuracy          0.84       325
 macro avg       0.92      0.53      0.51       325
 weighted avg    0.86      0.84      0.77       325

```

```

MultinomialNB()
classification report:
      precision    recall  f1-score   support

     0       0.92      1.00      0.96       269
     1       1.00      0.57      0.73        56

 accuracy          0.93       325
 macro avg       0.96      0.79      0.84       325
 weighted avg    0.93      0.93      0.92       325

```

```

SVC()
classification report:
      precision    recall  f1-score   support

     0       0.89      1.00      0.94       269
     1       1.00      0.43      0.60        56

 accuracy          0.90       325
 macro avg       0.95      0.71      0.77       325
 weighted avg    0.91      0.90      0.88       325

```

```

DecisionTreeClassifier()
classification report:
      precision    recall  f1-score   support

     0       0.95      0.99      0.97       269
     1       0.95      0.75      0.84        56

 accuracy          0.95       325
 macro avg       0.95      0.87      0.91       325
 weighted avg    0.95      0.95      0.95       325

```

```

RandomForestClassifier()
classification report:
      precision    recall  f1-score   support

     0       0.96      1.00      0.98       269
     1       1.00      0.80      0.89        56

 accuracy          0.97       325
 macro avg       0.98      0.90      0.94       325
 weighted avg    0.97      0.97      0.96       325

```