In [1]:
```python
#impoerting the libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import nltk
from ntlk.corpus import stopwords
from ntlk.stem.porter import PorteStemmer
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                        Traceback (most recent call last)
Input In [1], in <cell line: 6>()
      4 import matplotlib.pyplot as plt
      5 import nltk
----> 6 from ntlk.corpus import stopwords
      7 from ntlk.stem.porter import PorteStemmer

ModuleNotFoundError: No module named 'ntlk'
```

In [6]:
```python
data=pd.read_csv("E:\\NMDS\\spam.csv",encoding="latin")
data.head()
```

Out[6]:

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN

In [7]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   v1          5572 non-null   object
 1   v2          5572 non-null   object
 2   Unnamed: 2  50 non-null     object
 3   Unnamed: 3  12 non-null     object
 4   Unnamed: 4  6 non-null      object
dtypes: object(5)
memory usage: 108.9+ KB
```

In [8]:
```python
data.isna().sum()
```

Out[8]:
```
v1             0
v2             0
Unnamed: 2     5522
Unnamed: 3     5560
Unnamed: 4     5566
dtype: int64
```

In [10]:
```python
data.rename({"v1":"label","v2":"text"},inplace=True,axis=1)
data.tail()
```

Out[10]:

	label	text	Unnamed: 2	Unnamed: 3	Unnamed: 4
5567	spam	This is the 2nd time we have tried 2 contact u...	NaN	NaN	NaN
5568	ham	Will Ì_ b going to esplanade fr home?	NaN	NaN	NaN
5569	ham	Pity, * was in mood for that. So...any other s...	NaN	NaN	NaN
5570	ham	The guy did some bitching but I acted like i'd...	NaN	NaN	NaN
5571	ham	Rofl. Its true to its name	NaN	NaN	NaN

In [12]:
```python
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data['label']=le.fit_transform(data['label'])
```

In [13]:
```python
from sklearn.model_selection import train_test_split
X_train,X_text,y_train,y_test=train_test_split(X,y,test_size=0.20,random_state=0)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [13], in <cell line: 2>()
      1 from sklearn.model_selection import train_test_split
----> 2 X_train,X_text,y_train,y_test=train_test_split(X,y,test_size=0.20,random_stat
e=0)

NameError: name 'X' is not defined
```

In [14]:
```python
data.describe()
```

Out[14]:

	label
count	5572.000000
mean	0.134063
std	0.340751
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	1.000000

In [16]:
```python
data.shape
```

Out[16]:
```
(5572, 5)
```

In [18]:
```python
import matplotlib.pyplot as plt
data['label'].value_counts().plot()(kind="bar",figsize=(12,6)
plt.xticks(np.arange(2)),('Non spam','spam'),rotation=0);
```

```
Input In [18]
    plt.xticks(np.arange(2)),('Non spam','spam'),rotation=0);
    ^
SyntaxError: positional argument follows keyword argument
```

In [19]:
```python
from sklearn.tree import DecisionTreeClassifier
model=DecisionTreeClassifier()
model=fit.(X_train_res,y_train_res)
```

```
Input In [19]
    model=fit.(X_train_res,y_train_res)
             ^
SyntaxError: invalid syntax
```

In [21]:
```python
from sklearn.tree import RandomForestClassifier
model1= RandomForestClassifier()
model1=fit.(X_train_res,y_train_res)
```

```
Input In [21]
    model1=fit.(X_train_res,y_train_res)
              ^
SyntaxError: invalid syntax
```

In [ ]:
```python
from sklearn.naive_bayes import RandomForestClassifier
model1= RandomForestClassifier()
model1=fit.(X_train_res,y_train_res)
```