

In [1]:

```
#Roll No-33238
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
df=pd.read_csv(r'/home/pict/Downloads/archive (13)/spam.csv', encoding="ISO-8859-1")
```

In [4]:

```
df.head()
```

Out[4]:

	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 [5]:

```
df.groupby('v1').describe()
```

Out[5]:

			v2		Unnamed: 2				Unn:		
	count	unique	top	freq	count	unique	top	freq	count	unique	to
v1											
ham	4825	4516	Sorry, I'll call later	30	45	39	bt not his girlfrnd... Good night. ..@"	3	10	9	G
spam	747	653	Please call our customer service representativ...	4	5	4	PO Box 5249	2	2	1	MK1 92H 450Pp 1H

In [6]:

```
df['spam']=df['v1'].apply(lambda x: 1 if x=='spam' else 0)
df.head()
```

Out[6]:

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

In [7]:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(df.v2, df.spam)
```

In [8]:

```
from sklearn.feature_extraction.text import CountVectorizer
v = CountVectorizer()
X_train_count = v.fit_transform(X_train.values)
X_train_count.toarray()[:2]
```

Out[8]:

```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]])
```

In [9]:

```
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB()
model.fit(X_train_count, y_train)
```

Out[9]:

```
MultinomialNB()
```

In [10]:

```
sms = [  
    'Hey !',  
    'Whats up ?'  
]  
sms_count = v.transform(sms)  
model.predict(sms_count)
```

Out[10]:

```
array([0, 0])
```

In [11]:

```
X_test_count = v.transform(X_test)  
model.score(X_test_count, y_test)
```

Out[11]:

```
0.9856424982053122
```

In [12]:

```
from sklearn.model_selection import cross_val_score  
print(cross_val_score(model, X_test_count, y_test, cv=5))  
[0.98207885 0.97849462 0.98207885 0.96402878 0.95683453]
```

In []: