Importing Liabraries

In [32]:
```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer as vect
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import LabelEncoder
```

In [10]:
```
data=pd.read_csv("C:\\Users\\ASHISH TIWARI\\Downloads\\spam.csv",encoding="cp1252")
data.head(10)
```

Out[10]:

	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
5	spam	FreeMsg Hey there darling it's been 3 week's n...	NaN	NaN	NaN
6	ham	Even my brother is not like to speak with me. ...	NaN	NaN	NaN
7	ham	As per your request 'Melle Melle (Oru Minnamin...	NaN	NaN	NaN
8	spam	WINNER!! As a valued network customer you have...	NaN	NaN	NaN
9	spam	Had your mobile 11 months or more? U R entitle...	NaN	NaN	NaN

In [11]:
```
data.tail()
```

Out[11]:

	v1	v2	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]:
```
data.describe()
```

Out[12]:

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
count	5572	5572	50	12	6
unique	2	5169	43	10	5
top	ham	Sorry, I'll call later	bt not his girlfrnd... G o o d n i g h t . . .@"	MK17 92H. 450Ppw 16"	GNT:-)"
freq	4825	30	3	2	2

DATA PREPROCESSING

In [14]:
```
data.shape
```

Out[14]:
```
(5572, 5)
```

In [16]:
```
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: 217.8+ KB
```

In [17]:
```
data.isnull().sum()
```

Out[17]:
```
v1              0
v2              0
Unnamed: 2    5522
Unnamed: 3    5560
Unnamed: 4    5566
dtype: int64
```

In [18]:
```
columns_to_drop = ['Unnamed: 2','Unnamed: 3','Unnamed: 4']
data = data.drop(columns=columns_to_drop,axis=1)
```

In [19]:
```
data.head()
```

Out[19]:

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

In [20]:
```
data.shape
```

Out[20]:
```
(5572, 2)
```

In [21]:
```
data['v1'].value_counts()
```

Out[21]:
```
v1
ham     4825
spam     747
Name: count, dtype: int64
```

SPLITTING THE DATA

In [25]:
```
data.loc[data['v1'] == 'spam','v1',] = 0
data.loc[data['v1'] == 'ham','v1',] = 1
```

In [26]:
```
x=data['v2']
y=data['v1']
```

In [27]:
```
print(x)
```

```
0       Go until jurong point, crazy.. Available only ...
1                           Ok lar... Joking wif u oni...
2       Free entry in 2 a wkly comp to win FA Cup fina...
3       U dun say so early hor... U c already then say...
4       Nah I don't think he goes to usf, he lives aro...
                              ...
5567    This is the 2nd time we have tried 2 contact u...
5568                Will Ì_ b going to esplanade fr home?
5569    Pity, * was in mood for that. So...any other s...
5570    The guy did some bitching but I acted like i'd...
5571                            Rofl. Its true to its name
Name: v2, Length: 5572, dtype: object
```

In [28]:
```
print(y)
```

```
0       1
1       1
2       0
3       1
4       1
       ..
5567    0
5568    1
5569    1
5570    1
5571    1
Name: v1, Length: 5572, dtype: object
```

In [29]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=51)
```

In [33]:
```
feature_extraction = vect(min_df=1,stop_words='english',lowercase=True)
x_train_features = feature_extraction.fit_transform(x_train)
x_test_features = feature_extraction.transform(x_test)
```

In [34]:
```
y_train = y_train.astype('int')
y_test = y_test.astype('int')
print(x_train)
```

```
3075    Mum, hope you are having a great day. Hoping t...
1787                            Yes:)sura in sun tv.:)lol.
1614    Me sef dey laugh you. Meanwhile how's my darli...
4304                Yo come over carlos will be here soon
3266                    Ok then i come n pick u at engin?
                              ...
789                           Gud mrng dear hav a nice day
968             Are you willing to go for aptitude class.
1667    So now my dad is gonna call after he gets out ...
3321    Ok darlin i supose it was ok i just worry too ...
1688                    Nan sonathaya soladha. Why boss?
Name: v2, Length: 4457, dtype: object
```

In [35]:
```
print(x_train_features)
```

```
  (0, 741)      0.3219352588930141
  (0, 3979)     0.2410582143632299
  (0, 4296)     0.3891385935794867
  (0, 6599)     0.20296878731699391
  (0, 3386)     0.3219352588930141
  (0, 2122)     0.38613577623520473
  (0, 3136)     0.440116181574609
  (0, 3262)     0.25877035357606315
  (0, 3380)     0.21807195185332803
  (0, 4513)     0.2909649098524696
  (1, 4061)     0.380431198316959
  (1, 6872)     0.4306015894277422
  (1, 6417)     0.4769136859540388
  (1, 6442)     0.5652509076654626
  (1, 7443)     0.35056971070320353
  (2, 933)      0.4917598465723273
  (2, 2109)     0.42972812260098503
  (2, 3917)     0.40088501350982736
  (2, 2226)     0.413484525934624
  (2, 5825)     0.4917598465723273
  (3, 6140)     0.4903863168693604
  (3, 1599)     0.5927091854194291
  (3, 1842)     0.3708680641487708
  (3, 7453)     0.5202633571003087
  (4, 2531)     0.7419319091456392
  :     :
  (4452, 2122)  0.31002103760284144
  (4453, 999)   0.6760129013031282
  (4453, 7273)  0.5787739591782677
  (4453, 1762)  0.45610005640082985
  (4454, 3029)  0.42618909997886
  (4454, 2086)  0.3809693742808703
  (4454, 3088)  0.34475593009514444
  (4454, 2001)  0.4166919007849217
  (4454, 1049)  0.31932060116006045
  (4454, 7346)  0.31166263834107377
  (4454, 5370)  0.42618909997886
  (4455, 1148)  0.38998123077430413
  (4455, 6433)  0.38998123077430413
  (4455, 6361)  0.25697343671652706
  (4455, 2764)  0.3226323745940581
  (4455, 7358)  0.2915949626395065
  (4455, 7407)  0.3028481995557642
  (4455, 2108)  0.3136468384526087
  (4455, 4251)  0.30616657078392584
  (4455, 3763)  0.16807158405536876
  (4455, 4773)  0.35860460546223444
  (4456, 6117)  0.5304350313291551
  (4456, 6133)  0.5304350313291551
  (4456, 1386)  0.4460036316446079
  (4456, 4557)  0.48821933148688146
```

MODEL TRAINING AND EVALUATION

In [36]:
```
model=LogisticRegression()
model.fit(x_train_features,y_train)
```

Out[36]:
```
▾ LogisticRegression
LogisticRegression()
```

In [37]:
```
prediction1 = model.predict(x_train_features)
accuracy1 = accuracy_score(y_train,prediction1)
```

In [38]:
```
print('Accuracy on training data :  ',accuracy1)
```

```
Accuracy on training data :   0.9661207089970832
```

In [39]:
```
prediction2 = model.predict(x_test_features)
accuracy2 = accuracy_score(y_test,prediction2)
```

In [40]:
```
print('Accuracy on training data :  ',accuracy2)
```

```
Accuracy on training data :   0.9623318385650225
```