In [2]:
```
#importing the dependencies
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

In [4]:
```
#data collection
df = pd.read_csv('mail_data.csv')
```

In [5]:
```
df.head(5)
```

Out[5]:

	Category	Message
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 [6]:
```
#replace the null values with null string
mail_data = df.where((pd.notnull(df)), '')
```

In [7]:
```
#printing the first 5 rows from the dataframe
mail_data.head()
```

Out[7]:

	Category	Message
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 [8]:
```
#checking the no of rows and columns
mail_data.shape
```

Out[8]:
```
(5572, 2)
```

In [10]:
```
#labeling
#label spam mail <- 0 and ham mail<- 1
mail_data.loc[mail_data['Category']== 'spam','Category']= 0
mail_data.loc[mail_data['Category']== 'ham','Category']= 1
```

In [11]:
```
mail_data
```

Out[11]:

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

5572 rows × 2 columns

In [12]:
```
#seperating the data into text and labels
x= mail_data['Message']
y= mail_data['Category']
```

In [13]:
```
x
```

Out[13]:
```
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 ü 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: Message, Length: 5572, dtype: object
```

In [14]:
```
y
```

Out[14]:
```
0       1
1       1
2       0
3       1
4       1
       ..
5567    0
5568    1
5569    1
5570    1
5571    1
Name: Category, Length: 5572, dtype: object
```

In [17]:
```
#Train and test split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, rand
```

In [18]:
```
print(x_train.shape, x_test.shape, y_train, y_test)
```

```
3075                     Don know. I did't msg him recently.
1787    Do you know why god created gap between your f...
1614                           Thnx dude. u guys out 2nite?
4304                                         Yup i'm free...
3266    44 7732584351, Do you want a New Nokia 3510i c...
                              ...
789     5 Free Top Polyphonic Tones call 087018728737,...
968     What do u want when i come back?.a beautiful n...
1667    Guess who spent all last night phasing in and ...
3321    Eh sorry leh... I din c ur msg. Not sad alread...
1688    Free Top ringtone -sub to weekly ringtone-get ...
Name: Message, Length: 4457, dtype: object 2632    URGENT! Your mobile No 077
xxx WON a £2,000 Bon...
454     Ok i will tell her to stay out. Yeah its been ...
983     Congrats! 2 mobile 3G Videophones R yours. cal...
1282        Am I the only one who doesn't stalk profiles?
4610                              Y de asking like this.
                              ...
4827                          Haha, just what I was thinkin
5291      Xy trying smth now. U eat already? We havent...
3325    I don wake since. I checked that stuff and saw...
3561    Lol I know! Hey someone did a great inpersonat...
1136                        K do I need a login or anything
Name: Message, Length: 1115, dtype: object 3075    1
1787    1
1614    1
4304    1
3266    0
       ..
789     0
968     1
1667    1
3321    1
1688    0
Name: Category, Length: 4457, dtype: object 2632    0
454     1
983     0
1282    1
4610    1
       ..
4827    1
5291    1
3325    1
3561    1
1136    1
Name: Category, Length: 1115, dtype: object
```

In [20]:
```
print(x_train.shape)
```

```
(4457,)
```

In [22]:

```
#Feature Extraction
#transform the text data to feature vectors that can be used as input to the Lo
feature_extraction = TfidfVectorizer(min_df =1, stop_words='english', lowercase
```

In [24]:

```
x_train_features = feature_extraction.fit_transform(x_train)
x_test_features = feature_extraction.transform(x_test)

#convert y_train and y_test values as integers
y_train = y_train.astype('int')
y_test = y_test.astype('int')
```

In [25]:
```
print(x_train_features)
```

```
(0, 5413)     0.6198254967574347
(0, 4456)     0.4168658090846482
(0, 2224)     0.413103377943378
(0, 3811)     0.34780165336891333
(0, 2329)     0.38783870336935383
(1, 4080)     0.18880584110891163
(1, 3185)     0.29694482957694585
(1, 3325)     0.31610586766078863
(1, 2957)     0.3398297002864083
(1, 2746)     0.3398297002864083
(1, 918)      0.22871581159877646
(1, 1839)     0.2784903590561455
(1, 2758)     0.3226407885943799
(1, 2956)     0.33036995955537024
(1, 1991)     0.33036995955537024
(1, 3046)     0.2503712792613518
(1, 3811)     0.17419952275504033
(2, 407)      0.509272536051008
(2, 3156)     0.4107239318312698
(2, 2404)     0.45287711070606745
(2, 6601)     0.6056811524587518
(3, 2870)     0.5864269879324768
(3, 7414)     0.8100020912469564
(4, 50)       0.23633754072626942
(4, 5497)     0.15743785051118356
:      :
(4454, 4602)  0.2669765732445391
(4454, 3142)  0.32014451677763156
(4455, 2247)  0.37052851863170466
(4455, 2469)  0.35441545511837946
(4455, 5646)  0.33545678464631296
(4455, 6810)  0.29731757715898277
(4455, 6091)  0.23103841516927642
(4455, 7113)  0.30536590342067704
(4455, 3872)  0.3108911491788658
(4455, 4715)  0.30714144758811196
(4455, 6916)  0.19636985317119715
(4455, 3922)  0.31287563163368587
(4455, 4456)  0.24920025316220423
(4456, 141)   0.292943737785358
(4456, 647)   0.30133182431707617
(4456, 6311)  0.30133182431707617
(4456, 5569)  0.4619395404299172
(4456, 6028)  0.21034888000987115
(4456, 7154)  0.24083218452280053
(4456, 7150)  0.3677554681447669
(4456, 6249)  0.17573831794959716
(4456, 6307)  0.2752760476857975
(4456, 334)   0.2220077711654938
(4456, 5778)  0.16243064490100795
(4456, 2870)  0.31523196273113385
```

In [26]:

```
#Training the model
model = LogisticRegression()
```

In [27]:

```
#training the logistic regression model with the training data
model.fit(x_train_features, y_train)
```

Out[27]:

```
LogisticRegression()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [28]:

```
#evaluation of the model
#prediction on training data
prediction_on_training_data = model.predict(x_train_features)
```

In [29]:

```
accuracy_on_training_data = accuracy_score(y_train,prediction_on_training_data)
```

In [30]:

```
accuracy_on_training_data
```

Out[30]:

```
0.9670181736594121
```

In [31]:

```
prediction_on_test_data = model.predict(x_test_features)
accuracy_on_test_data = accuracy_score(y_test,prediction_on_test_data)
```

In [32]:

```
accuracy_on_test_data
```

Out[32]:

```
0.9659192825112107
```

In [ ]:

```
#Building a predictive system
```

In [38]:

```
input_mail = ["Had your mobile 11 months or more? U R entitled to Update to the
```

In [39]:

```
#convet text to feature vectors
input_data_features = feature_extraction.transform(input_mail)
```

In [40]:

```
#making prediction
prediction = model.predict(input_data_features)
```

In [41]:

```
print(prediction)
```

```
[0]
```