In [88]:
```
import numpy as np
import pandas as pd
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 [90]:
```
data = pd.read_csv('mail_data.csv')
```

In [91]:
```
data
```

Out[91]:

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

5572 rows × 2 columns

In [ ]:
```
# removing null values & replacing with null string
```

In [92]:
```
new_data = data.where((pd.notnull(data)),'')
```

In [93]:
```
new_data
```

Out[93]:

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

5572 rows × 2 columns

In [94]:
```
new_data.loc[new_data['Category'] == 'spam', 'Category',] = 0
new_data.loc[new_data['Category'] == 'ham', 'Category',] = 1
```

In [96]:
```
x = new_data['Message']
y = new_data['Category']
```

In [97]:
```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=3)
```

In [106…
```
print(x_train)
```

```
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
```

In [107…
```
print(y_train)
```

```
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: int32
```

In [99]:
```
feature_extraction = TfidfVectorizer(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 [100…
```
y_train = y_train.astype('int')
y_test = y_test.astype('int')
```

In [101…
```
model = LogisticRegression()
```

In [102…
```
model.fit(x_train_features, y_train)
```

Out[102]:
```
LogisticRegression()
```

In [103…
```
training_prediction = model.predict(x_train_features)
training_accuracy = accuracy_score(y_train, training_prediction)
print(training_accuracy)
```

```
0.9670181736594121
```

In [104…
```
testing_prediction = model.predict(x_test_features)
testing_accuracy = accuracy_score(y_test, testing_prediction)
print(testing_accuracy)
```

```
0.9659192825112107
```

In [105…
```
input_data = ["You have won a lottery of 1000000 dollars. Share your details for claiming"]
input_data_features = feature_extraction.transform(input_data)
prediction = model.predict(input_data_features)
if (prediction[0]==1):
  print('Spam')
else:
  print('Ham')
```

```
Spam
```

In [ ]:
```
# prediction is good
```