In [16]:

```python
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 [17]:

```python
df=pd.read_csv('spam.csv',encoding=('ISO-8859-1'))
```

In [18]:

```python
df.sample(5)
```

Out[18]:

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
967	ham	What do u want when i come back?.a beautiful n...	NaN	NaN	NaN
5093	ham	Gokila is talking with you aha:)	NaN	NaN	NaN
2464	ham	They will pick up and drop in car.so no problem..	NaN	NaN	NaN
1314	ham	Got but got 2 colours lor. One colour is quite...	NaN	NaN	NaN
2315	ham	That's significant but dont worry.	NaN	NaN	NaN

In [19]:

```python
print(df)
```

```
        v1                                                 v2 Unnamed: 2  \
0      ham  Go until jurong point, crazy.. Available only ...        NaN
1      ham                      Ok lar... Joking wif u oni...        NaN
2     spam  Free entry in 2 a wkly comp to win FA Cup fina...        NaN
3      ham  U dun say so early hor... U c already then say...        NaN
4      ham  Nah I don't think he goes to usf, he lives aro...        NaN
...    ...                                                ...        ...
5567  spam  This is the 2nd time we have tried 2 contact u...        NaN
5568   ham              Will Ì_ b going to esplanade fr home?        NaN
5569   ham  Pity, * was in mood for that. So...any other s...        NaN
5570   ham  The guy did some bitching but I acted like i'd...        NaN
5571   ham                         Rofl. Its true to its name        NaN

     Unnamed: 3 Unnamed: 4
0           NaN        NaN
1           NaN        NaN
2           NaN        NaN
3           NaN        NaN
4           NaN        NaN
...         ...        ...
5567        NaN        NaN
5568        NaN        NaN
5569        NaN        NaN
5570        NaN        NaN
5571        NaN        NaN

[5572 rows x 5 columns]
```

In [20]:

```python
data=df.where((pd.notnull(df)),'')
```

In [21]:

```python
data.head()
```

Out[21]:

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
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 [22]:

```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  5572 non-null   object
 3   Unnamed: 3  5572 non-null   object
 4   Unnamed: 4  5572 non-null   object
dtypes: object(5)
memory usage: 217.8+ KB
```

In [23]:

```python
data.shape
```

Out[23]:

```
(5572, 5)
```

In [24]:

```python
data.loc[data['v1'] == 'spam', 'v1',]=0
data.loc[data['v1'] == 'ham', 'v1',]=1
```

In [25]:

```python
x = data['v2']

y = data['v1']
```

In [26]:

```python
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 [27]:

```python
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 [32]:

```python
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2, random_state=3)
```

In [33]:

```python
print(x.shape)
print(x_train.shape)
print(x_test.shape)
```

```
(5572,)
(4457,)
(1115,)
```

In [34]:

```python
print(y.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(5572,)
(4457,)
(1115,)
```

In [40]:

```python
vectorizer = TfidfVectorizer(lowercase=True)
x_train_tfidf = vectorizer.fit_transform(x_train)
x_test_tfidf = vectorizer.transform(x_test)
```

In [41]:

```python
y_train = y_train.astype('int')
y_test = y_test.astype('int')
```

In [42]:

```python
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 [44]:

```python
print(x_train_tfidf)
```

```
  (0, 741)	0.28307455118083463
  (0, 3360)	0.1327523238442287
  (0, 4108)	0.21196015023008544
  (0, 4908)	0.12973225055917514
  (0, 3042)	0.26300555749396887
  (0, 946)	0.1151770452043338
  (0, 7464)	0.19261204588580316
  (0, 4431)	0.3421657916670175
  (0, 6805)	0.17846848640014898
  (0, 6873)	0.15216101010779184
  (0, 3497)	0.28307455118083463
  (0, 2178)	0.33952544349598
  (0, 3235)	0.3869898904365042
  (0, 3365)	0.22753426247664568
  (0, 1032)	0.1361275560580212
  (0, 7720)	0.1765620792792692
  (0, 3491)	0.19174855251416806
  (0, 4655)	0.2558426236041184
  (1, 4190)	0.3725861907992424
  (1, 7099)	0.42172200036894236
  (1, 6620)	0.46707907862382136
  (1, 3646)	0.2020333473623602
  (1, 6645)	0.553594666958471
  (1, 7704)	0.3433404875792393
  (2, 954)	0.4257390912308466
  :	:
  (4455, 7402)	0.15130978849620824
  (4455, 6316)	0.16857953235415776
  (4455, 6850)	0.14840315498751144
  (4455, 1592)	0.11611242093594701
  (4455, 6991)	0.1743411711262433
  (4455, 4647)	0.1711510506728716
  (4455, 3888)	0.13162386869199988
  (4455, 3767)	0.1131472348271079
  (4455, 1615)	0.12678729795752244
  (4455, 1561)	0.12510711341007413
  (4455, 6956)	0.15407243057965578
  (4455, 847)	0.18793900087060836
  (4455, 2376)	0.1310118611150462
  (4455, 6845)	0.148803926710582
  (4455, 4933)	0.28083822596779895
  (4455, 4680)	0.11249715586710375
  (4455, 4410)	0.1081290969254776
  (4455, 3360)	0.23698431521172753
  (4455, 946)	0.10280480369551688
  (4455, 7720)	0.07879794914071371
  (4456, 6312)	0.5058318398291911
  (4456, 6334)	0.5058318398291911
  (4456, 1431)	0.4253166254875381
  (4456, 4703)	0.4655742326583645
  (4456, 7513)	0.3010227592699582
```

In [45]:

```python
model = LogisticRegression()
```

In [46]:

```python
model.fit(x_train_tfidf, y_train)
```

Out[46]:

```
▾ LogisticRegression
LogisticRegression()
```

In [50]:

```python
prediction_on_trainning_data = model.predict(x_train_tfidf)
accuracy_on_training_data = accuracy_score(y_train, prediction_on_trainning_data )
```

In [51]:

```python
print('Acc on training data :', accuracy_on_training_data )
```

```
Acc on training data : 0.9739735247924612
```

In [53]:

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

In [54]:

```python
print('Acc on test data :', accuracy_on_test_data )
```

```
Acc on test data : 0.9757847533632287
```

In [59]:

```python
input = ["Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got am
input_data_features = vectorizer.transform(input)
prediction = model.predict(input_data_features)
print(prediction)
if(prediction[0]==1):
    print('Ham')
else:
    print('spam')
```

```
[1]
Ham
```

In [ ]:

```python

```