In [54]:

```
import numpy as np
import pandas as pd
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,confusion_matrix
```

In [32]:

```
df = pd.read_csv('spam_or_not_spam.csv')
df.head()
```

Out[32]:

	email	label
0	date wed NUMBER aug NUMBER NUMBER NUMBER NUMB...	0
1	martin a posted tassos papadopoulos the greek ...	0
2	man threatens explosion in moscow thursday aug...	0
3	klez the virus that won t die already the most...	0
4	in adding cream to spaghetti carbonara which ...	0

In [33]:

```
df= df.fillna(' ')
```

In [34]:

```
df.shape
```

Out[34]:

```
(3000, 2)
```

In [35]:

```
x=df['email']
y=df['label']
```

In [36]:

```
x.head()
```

Out[36]:

```
0     date wed NUMBER aug NUMBER NUMBER NUMBER NUMB...
1    martin a posted tassos papadopoulos the greek ...
2    man threatens explosion in moscow thursday aug...
3    klez the virus that won t die already the most...
4     in adding cream to spaghetti carbonara which ...
Name: email, dtype: object
```

In [37]:

```
y.head()
```

Out[37]:

```
0    0
1    0
2    0
3    0
4    0
Name: label, dtype: int64
```

In [38]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=13)
```

In [39]:

```
print(x_train.shape,x_test.shape,y_train.shape,y_test.shape)
```

```
(2100,) (900,) (2100,) (900,)
```

In [40]:

```
vector=CountVectorizer()
vector.fit(x_train)
```

Out[40]:

```
CountVectorizer()
```

In [42]:

```
x_train_words=vector.get_feature_names()
x_train_dtm=vector.transform(x_train)
```

In [45]:

```
x_test_dtm=vector.transform(x_test)
```

In [51]:

```
model=MultinomialNB()
model.fit(x_train_dtm,y_train)
```

Out[51]:

```
MultinomialNB()
```

In [52]:

```
y_predict=model.predict(x_test_dtm)
```

In [53]:

```
accuracy_score(y_test,y_predict)
```

Out[53]:

```
0.9922222222222222
```

In [55]:

```
confusion_matrix(y_test,y_predict)
```

Out[55]:

```
array([[771,   1],
       [  6, 122]], dtype=int64)
```

In [ ]: