In [1]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn import metrics
```

In [2]:
```
msg=pd.read_csv('P7_dataset.csv',names=['message','label'])
print('The dimensions of the dataset',msg.shape)
msg['labelnum']=msg.label.map({'pos':1,'neg':0})
X=msg.message
y=msg.labelnum
# print(X)
# print(y)
msg[['message','labelnum']]
```

```
The dimensions of the dataset (18, 2)
```

Out[2]:

	message	labelnum
1	I love this sandwich	1
2	This is an amazing place	1
3	I feel very good about these beers	1
4	This is my best work	1
5	What an awesome view	1
6	I do not like this restaurant	0
7	I am tired of this stuff	0
8	I can?t deal with this	0
9	He is my sworn enemy	0
10	My boss is horrible	0
11	This is an awesome place	1
12	I do not like the taste of this juice	0
13	I love to dance	1
14	I am sick and tired of this place	0
15	What a great holiday	1
16	That is a bad locality to stay	0
17	We will have good fun tomorrow	1
18	I went to my enemy?s house today	0

In [3]:
```
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.25,random_state=100)
print("Train Feature shape : "+ str(X_train.shape))
print("Train Target shape : "+ str(X_test.shape))
print("Test Feature shape : "+ str(y_train.shape))
print("Test shape : "+ str(y_train.shape))
```

```
Train Feature shape : (13,)
Train Target shape : (5,)
Test Feature shape : (13,)
Test shape : (13,)
```

EXPLANATION OF CountVectorizer() - only for reference

```
>>> corpus = [
...     'This is the first document.',
...     'This document is the second document.',
...     'And this is the third one.',
...     'Is this the first document?',
... ]
>>> vectorizer = CountVectorizer()
>>> X = vectorizer.fit_transform(corpus)
>>> vectorizer.get_feature_names_out()
array(['and', 'document', 'first', 'is', 'one', 'second', 'the', 'third',
       'this'], ...)
>>> print(X.toarray())
[[0 1 1 1 0 0 1 0 1]
 [0 2 0 1 0 1 1 0 1]
 [1 0 0 1 1 0 1 1 1]
 [0 1 1 1 0 0 1 0 1]]
```

In [4]:
```
from sklearn.feature_extraction.text import CountVectorizer
count_vect = CountVectorizer()
X_train_dtm = count_vect.fit_transform(X_train)
X_test_dtm=count_vect.transform(X_test)
clf = MultinomialNB().fit(X_train_dtm,y_train)
predicted = clf.predict(X_test_dtm)
```

In [5]:
```
print('Accuracy metrics')
print('Accuracy of the classifer is : ',metrics.accuracy_score(y_test,predicted))
print('Confusion matrix')
print(metrics.confusion_matrix(y_test,predicted))
print('Recall and Precison ')
print(metrics.recall_score(y_test,predicted))
print(metrics.precision_score(y_test,predicted))
```

```
Accuracy metrics
Accuracy of the classifer is :  1.0
Confusion matrix
[[3 0]
 [0 2]]
Recall and Precison
1.0
1.0
```