In [14]:

```
#Import the necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score,confusion_matrix,precision_score,recall_score
```

In [15]:

```
#Load the libraries
msg=pd.read_csv('P7_dataset.csv',names=['message','label'])
print('The dimensions of the dataset',msg.shape)
#Map pos to 1 and neg to 0
msg['labelnum']=msg.label.map({'pos':1,'neg':0})
X=msg.message
y=msg.labelnum
#Display the dataframe with message and labelnum. head() shows only 5 records
msg[['message','labelnum']].head()
```

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

Out[15]:

	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

In [16]:

```
#Split the dataset into train and test
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.25,random_state=100)
```

In [17]:

```
#Make probability distribution of words and count the occurences
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 [18]:

```
#Display the accuracy , Confusion matrix, precision and recall
print('Accuracy metrics')
print('Accuracy of the classifer is : ',accuracy_score(y_test,predicted))
print('Confusion matrix')
print(confusion_matrix(y_test,predicted))
print('Recall and Precison ')
print(recall_score(y_test,predicted))
print(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
```