In [15]:

```
import pandas as pd
import numpy as np
import nltk
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn import naive_bayes
from sklearn.metrics import roc_auc_score,accuracy_score
import pickle
```

In [2]:

```
nltk.download("stopwords")
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\kishan\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
```

Out[2]:

```
True
```

In [3]:

```
dataset = pd.read_csv('reviews.txt',sep = '\t', names =['Reviews','Comments'])
```

In [4]:

```
dataset
```

Out[4]:

	Reviews	Comments
0	1	The Da Vinci Code book is just awesome.
1	1	this was the first clive cussler i've ever rea...
2	1	i liked the Da Vinci Code a lot.
3	1	i liked the Da Vinci Code a lot.
4	1	I liked the Da Vinci Code but it ultimatly did...
...	...	...
6913	0	Brokeback Mountain was boring.
6914	0	So Brokeback Mountain was really depressing.
6915	0	As I sit here, watching the MTV Movie Awards, ...
6916	0	Ok brokeback mountain is such a horrible movie.
6917	0	Oh, and Brokeback Mountain was a terrible movie.

6918 rows × 2 columns

In [5]:

```
stopset = set(stopwords.words('english'))
```

In [6]:

```
vectorizer = TfidfVectorizer(use_idf = True,lowercase = True, strip_accents='ascii',stop
_words=stopset)
```

In [16]:

```
X = vectorizer.fit_transform(dataset.Comments)
```

```
y = dataset.Reviews
pickle.dump(vectorizer, open('tranform.pkl', 'wb'))
```

In [17]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=4
2)
```

In [18]:

```
clf = naive_bayes.MultinomialNB()
clf.fit(X_train,y_train)
```

Out[18]:

```
MultinomialNB()
```

In [19]:

```
accuracy_score(y_test,clf.predict(X_test))*100
```

Out[19]:

```
97.47109826589595
```

In [20]:

```
clf = naive_bayes.MultinomialNB()
clf.fit(X,y)
```

Out[20]:

```
MultinomialNB()
```

In [21]:

```
accuracy_score(y_test,clf.predict(X_test))*100
```

Out[21]:

```
98.77167630057804
```

In [22]:

```
filename = 'nlp_model.pkl'
pickle.dump(clf, open(filename, 'wb'))
```