In [1]:
```
#importing libraries
import numpy as np
import pandas as pd
import seaborn as sns
import re
import matplotlib.pyplot as plt

from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
```

In [2]:
```
#tools for processing input data
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import CountVectorizer
from nltk.stem.porter import PorterStemmer
```

In [3]:
```
#loading data
data=pd.read_csv("Desktop\\sentiment_data.tsv",delimiter="\t")
data=data[:2000]
data
```

Out[3]:

	id	sentiment	review
0	5814_8	1	With all this stuff going down at the moment w...
1	2381_9	1	\The Classic War of the Worlds\" by Timothy Hi...
2	7759_3	0	The film starts with a manager (Nicholas Bell)...
3	3630_4	0	It must be assumed that those who praised this...
4	9495_8	1	Superbly trashy and wondrously unpretentious 8...
...	...	...	...
1995	6454_2	0	The monster from Enemy Mine somehow made his w...
1996	1471_4	0	This kind of film has become old hat by now, h...
1997	2374_9	1	The year 2005 saw no fewer than 3 filmed produ...
1998	9327_1	0	This was, so far, the worst movie I have seen ...
1999	11050_1	0	Terrible use of scene cuts. All continuity is ...

2000 rows × 3 columns

In [4]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 3 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   id         2000 non-null   object
 1   sentiment  2000 non-null   int64 
 2   review     2000 non-null   object
dtypes: int64(1), object(2)
memory usage: 47.0+ KB
```

In [5]:
```
data.describe()
```

Out[5]:

	sentiment
count	2000.000000
mean	0.498500
std	0.500123
min	0.000000
25%	0.000000
50%	0.000000
75%	1.000000
max	1.000000

In [6]:
```
data=data.drop(['id'],axis=1)
data.head()
```

Out[6]:

	sentiment	review
0	1	With all this stuff going down at the moment w...
1	1	\The Classic War of the Worlds\" by Timothy Hi...
2	0	The film starts with a manager (Nicholas Bell)...
3	0	It must be assumed that those who praised this...
4	1	Superbly trashy and wondrously unpretentious 8...

In [7]:
```
#processing message
def processing(review):
    #removing email
    raw_review = re.sub('\\b[\\w\\-.]+?@\\w+?\\.\\w{2,4}\\b'," ",review)
    #removing html
    raw_review = re.sub('(http[s]?\\S+)|(\\w+\\.[A-Za-z]{2,4}\\S*)'," ",raw_review)
    #removing non letters
    raw_review = re.sub("[^a-zA-Z]"," ",raw_review)
    #removing  numbers
    raw_review = re.sub('\\d+(\\.\\d+)?'," ",raw_review)
    words=raw_review.lower().split()
    stop=set(stopwords.words("english"))
    meaningfull_words=[ps.stem(w) for w in words if not w in stop]
    return (" ".join(meaningfull_words))
```

In [8]:
```
clean_review_corpus=[]
ps=PorterStemmer()
review_count=data['review'].size
review_count
```

Out[8]:
```
2000
```

In [9]:
```
for i in range(0,review_count):
    clean_review_corpus.append(processing(data['review'][i]))
```

In [10]:
```
print(data['review'][0],"\n")
print(clean_review_corpus[0])
```

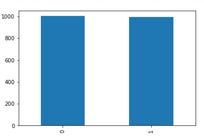
```
With all this stuff going down at the moment with MJ i've started listening to his music, wat
ching the odd documentary here and there, watched The Wiz and watched Moonwalker again. Maybe
i just want to get a certain insight into this guy who i thought was really cool in the eight
ies just to maybe make up my mind whether he is guilty or innocent. Moonwalker is part biogra
phy, part feature film which i remember going to see at the cinema when it was originally rel
eased. Some of it has subtle messages about MJ's feeling towards the press and also the obvio
us message of drugs are bad m'kay.<br /><br />Visually impressive but of course this is all a
bout Michael Jackson so unless you remotely like MJ in anyway then you are going to hate this
and find it boring. Some may call MJ an egotist for consenting to the making of this movie BU
T MJ and most of his fans would say that he made it for the fans which if true is really nice
of him.<br /><br />The actual feature film bit when it finally starts is only on for 20 minut
es or so excluding the Smooth Criminal sequence and Joe Pesci is convincing as a psychopathic
all powerful drug lord. Why he wants MJ dead so bad is beyond me. Because MJ overheard his pl
ans? Nah, Joe Pesci's character ranted that he wanted people to know it is he who is supplyin
g drugs etc so i dunno, maybe he just hates MJ's music.<br /><br />Lots of cool things in thi
s like MJ turning into a car and a robot and the whole Speed Demon sequence. Also, the direct
or must have had the patience of a saint when it came to filming the kiddy Bad sequence as us
ually directors hate working with one kid let alone a whole bunch of them performing a comple
x dance scene.<br /><br />Bottom line, this movie is for people who like MJ on one level or a
nother (which i think is most people). If not, then stay away. It does try and give off a who
lesome message and ironically MJ's bestest buddy in this movie is a girl! Michael Jackson is
truly one of the most talented people ever to grace this planet but is he guilty? Well, with
all the attention i've gave this subject....hmmm well i don't know because people can be diff
erent behind closed doors, i know this for a fact. He is either an extremely nice but stupid
guy or one of the most sickest liars. I hope he is not the latter.

stuff go moment mj start listen music watch odd documentari watch wiz watch moonwalk mayb wan
t get certain insight guy thought realli cool eighti mayb make mind whether guilti innoc moon
walk part biographi part featur film rememb go see cinema origin releas subtl messag mj feel
toward press also obviou messag drug bad kay br br visual impress cours michael jackson unles
s remot like mj anyway go hate find bore may call mj egotist consent make movi mj fan would s
ay made fan true realli nice br br actual featur film bit final start minut exclud smooth cri
min sequenc joe pesci convinc psychopath power drug lord want mj dead bad beyond mj overheard
```

In [11]:
```
#preparing count vectorizer
cv=CountVectorizer()
data_input=cv.fit_transform(clean_review_corpus)
data_input=data_input.toarray()
data_input[0]
```

Out[11]:
```
array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

In [12]:
```
#applying classification
data_output=data['sentiment']
data_output.value_counts().plot.bar()
```

Out[12]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x28951390790>
```



In [13]:
```
#splitting data
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(data_input,data_output,test_size=0.2,random_state
```

In [14]:
```
#preparing ml models
lr=LogisticRegression()
lr.fit(xtrain,ytrain)
nvb=GaussianNB()
nvb.fit(xtrain,ytrain)
rf=RandomForestClassifier(n_estimators=1000,criterion='entropy',random_state=0)
rf.fit(xtrain,ytrain)
dt=DecisionTreeClassifier()
dt.fit(xtrain,ytrain)
```

Out[14]:
```
DecisionTreeClassifier()
```

In [15]:
```
#predictions
lr_predict=lr.predict(xtest)
nvb_predict=nvb.predict(xtest)
rf_predict=rf.predict(xtest)
dt_predict=dt.predict(xtest)
```

In [16]:
```
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
#results
print("Accuracy Score of Logistic Regression Model is ",accuracy_score(ytest,lr_predict))
print("Classification_report of Logistic Regression Model is\n",classification_report(ytest,
lr_predict),"\n")

print("Accuracy Score of NaiveBayes Model is ",accuracy_score(ytest,nvb_predict))
print("Classification_report of  NaiveBayes Model is\n",classification_report(ytest,nvb_pred
ict),"\n")

print("Accuracy Score of RandomForest Model is ",accuracy_score(ytest,rf_predict))
print("Classification_report of RandomForest Model is\n",classification_report(ytest,rf_pred
ict),"\n")
```

```
Accuracy Score of Logistic Regression Model is  0.8025
Classification_report of Logistic Regression Model is
              precision    recall  f1-score   support

           0       0.85      0.78      0.81       219
           1       0.76      0.83      0.79       181

    accuracy                           0.80       400
   macro avg       0.80      0.80      0.80       400
weighted avg       0.81      0.80      0.80       400


Accuracy Score of NaiveBayes Model is  0.63
Classification_report of  NaiveBayes Model is
              precision    recall  f1-score   support

           0       0.65      0.71      0.68       219
           1       0.60      0.54      0.57       181

    accuracy                           0.63       400
   macro avg       0.63      0.62      0.62       400
weighted avg       0.63      0.63      0.63       400


Accuracy Score of RandomForest Model is  0.82
Classification_report of RandomForest Model is
              precision    recall  f1-score   support

           0       0.80      0.78      0.83       219
           1       0.77      0.87      0.81       181

    accuracy                           0.82       400
   macro avg       0.82      0.82      0.82       400
weighted avg       0.83      0.82      0.82       400


Accuracy Score of Decision Tree Model is  0.6925
Classification_report of Decision Tree Model is
              precision    recall  f1-score   support

           0       0.76      0.65      0.70       219
           1       0.64      0.75      0.69       181

    accuracy                           0.69       400
   macro avg       0.70      0.70      0.69       400
weighted avg       0.70      0.69      0.69       400
```

In [ ]:

In [ ]:

In [ ]: