In [1]:
```
doc1='f5o@od is # good & good!'
doc2='& Food # is * tasty'
doc3='Quality is Good'
doc4='food is not good'
doc5='servi89ce is Poor poor means very poor'
doc6='it is to_o costly'
doc7='che^ap quality'

corpus=[doc1,doc2,doc3,doc4,doc5,doc6,doc7]
target=['pos','pos','pos','neg','neg','neg','neg']
```

In [7]:
```
from nltk.tokenize import word_tokenize,sent_tokenize
```

In [3]:
```
s="food is good"
word_tokenize(s)
```

Out[3]:
```
['food', 'is', 'good']
```

In [4]:
```
s.split() #space,\t,\n
```

Out[4]:
```
['food', 'is', 'good']
```

In [9]:
```
s="he is sonu and he lives in gr.noida. his age is 20"
s.split('.')
```

Out[9]:
```
['he is sonu and he lives in gr', 'noida', ' his age is 20']
```

In [10]:
```
sent_tokenize(s) #.
```

Out[10]:
```
['he is sonu and he lives in gr.noida.', 'his age is 20']
```

In [11]:
```
from nltk.corpus import stopwords
```

In [14]:
```
sw=stopwords.words("english")
len(sw)
```

Out[14]:
```
179
```

In [17]:
```
#play
#played
#plays
#playing

#stemming
from nltk.stem import PorterStemmer
ps=PorterStemmer()
print(ps.stem("play"))
print(ps.stem("played"))
print(ps.stem("plays"))
print(ps.stem("playing"))
```

```
play
play
play
play
```

In [18]:
```
from nltk.stem import WordNetLemmatizer
```

In [24]:
```
wl=WordNetLemmatizer()
ps=PorterStemmer()
```

```
print(ps.stem("wife"),ps.stem("wives"))
print(wl.lemmatize("wife"),wl.lemmatize("wives"))
print(ps.stem("quality"),ps.stem("costly"))
print(wl.lemmatize("quality"),wl.lemmatize("costly"))
```

```
wife wive
wife wife
qualiti costli
quality costly
```

In [ ]:
```
#textblob
```

In [25]:
```
pip install textblob
```

```
Collecting textblob
  Using cached textblob-0.17.1-py2.py3-none-any.whl (636 kB)
Requirement already satisfied: nltk>=3.1 in c:\users\ducat\anaconda3\lib\site-packages
(from textblob) (3.8.1)
Requirement already satisfied: click in c:\users\ducat\anaconda3\lib\site-packages (from
nltk>=3.1->textblob) (8.0.4)
Requirement already satisfied: joblib in c:\users\ducat\anaconda3\lib\site-packages (fro
m nltk>=3.1->textblob) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\ducat\anaconda3\lib\site-pack
ages (from nltk>=3.1->textblob) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\ducat\anaconda3\lib\site-packages (from
nltk>=3.1->textblob) (4.65.0)
Requirement already satisfied: colorama in c:\users\ducat\anaconda3\lib\site-packages (f
rom click->nltk>=3.1->textblob) (0.4.6)
Installing collected packages: textblob
Successfully installed textblob-0.17.1
Note: you may need to restart the kernel to use updated packages.
```

In [29]:
```
from textblob import TextBlob
```

In [34]:
```
blob=TextBlob("acton")
print(blob.correct())

blob=TextBlob("actin")
print(blob.correct())

blob=TextBlob("mornig")
print(blob.correct())

blob=TextBlob("food qualty was good and tast ws als goo")
print(blob.correct())
```

```
action
action
morning
food quality was good and last was as go
```

In [37]:
```
blob=TextBlob("he is a good boy.")
print(blob.translate(from_lang='english',to='hi'))
```

```
वह एक अच्छा लड़का है।
```

In [38]:
```
print(blob.upper())
```

```
HE IS A GOOD BOY.
```

In [44]:
```
blob=TextBlob("he is awesome boy.")
blob.sentiment
```

Out[44]:
```
Sentiment(polarity=1.0, subjectivity=1.0)
```

In [48]:
```
feedback=input("enter review:")
blob=TextBlob(feedback)
print(blob.sentiment)
```

```
Sentiment(polarity=-0.35, subjectivity=0.6000000000000001)
```

In [49]:
```
corpus
```

Out[49]:
```
['f5o@od is # good & good!',
 '& Food # is * tasty',
 'Quality is Good',
 'food is not good',
 'servi89ce is Poor poor means very poor',
 'it is to_o costly',
 'che^ap quality']
```

In [59]:
```
import re
def cleaning(doc):
    doc=doc.lower()
    doc=re.sub("[^a-z ]","",doc)
    newdoc=""
    for word in doc.split():
        newdoc=newdoc+word+" "
    return newdoc.strip()

corpus_new=list(map(cleaning,corpus))
print(corpus_new)
print(target)
```

```
['food is good good', 'food is tasty', 'quality is good', 'food is not good', 'service i
s poor poor means very poor', 'it is too costly', 'cheap quality']
['pos', 'pos', 'pos', 'neg', 'neg', 'neg', 'neg']
```

In [ ]:
```
train_data=[('food is good good','pos'),(,),(,),(,)]
```

In [61]:
```
train_data=[]
for tup in zip(corpus_new,target):
    train_data.append(tup)
print(train_data)
```

```
[('food is good good', 'pos'), ('food is tasty', 'pos'), ('quality is good', 'pos'), ('f
ood is not good', 'neg'), ('service is poor poor means very poor', 'neg'), ('it is too c
ostly', 'neg'), ('cheap quality', 'neg')]
```

In [62]:
```
from textblob import classifiers
model=classifiers.NaiveBayesClassifier(train_data)
```

In [64]:
```
sample="good food quality"
print(model.classify(sample))
prob_dist=model.prob_classify(sample)
print(prob_dist.prob('pos'),prob_dist.prob('neg'))
```

```
pos
0.9009609397340038 0.09903906026599531
```

In [65]:
```
import pandas as pd
```

In [66]:
```
df=pd.read_csv("f:/dataset/sentiment/Restaurant_Reviews.txt",sep="\t")
corpus=df.Review
target=df.Liked
```

In [68]:
```
corpus_new=list(map(cleaning,corpus))
train_data=[]
```

```
for tup in zip(corpus_new,target):
    train_data.append(tup)
```

In [69]:
```
from textblob import classifiers
model=classifiers.NaiveBayesClassifier(train_data)
```

In [71]:
```
sample="good food quality"
print(model.classify(sample))
prob_dist=model.prob_classify(sample)
print(prob_dist.prob(0),prob_dist.prob(1))
```

```
1
0.12101727944056496 0.8789827205594348
```

In [73]:
```
df=pd.read_json("f:/dataset/sentiment/news.json",lines=True)
df
```

Out[73]:

	category	headline	authors	link	short_desc
0	CRIME	There Were 2 Mass Shootings In Texas Last Week...	Melissa Jeltsen	https://www.huffingtonpost.com/entry/texas-ama...	She husband. H their chi
1	ENTERTAINMENT	Will Smith Joins Diplo And Nicky Jam For The 2...	Andy McDonald	https://www.huffingtonpost.com/entry/will-smit...	Of course i
2	ENTERTAINMENT	Hugh Grant Marries For The First Time At Age 57	Ron Dicker	https://www.huffingtonpost.com/entry/hugh-gran...	The actor a lo girlfrien
3	ENTERTAINMENT	Jim Carrey Blasts 'Castrato' Adam Schiff And D...	Ron Dicker	https://www.huffingtonpost.com/entry/jim-carre...	The acto Dems kicking for
4	ENTERTAINMENT	Julianna Margulies Uses Donald Trump Poop Bags...	Ron Dicker	https://www.huffingtonpost.com/entry/julianna-...	The "Di actress saic the ba
...	...	...	...	...	
1431	BUSINESS	Four More Bank Closures Mark the Week of Janua...	Dennis Santiago, Contributor\nGlobal Risk and ...	https://www.huffingtonpost.com/entry/four-more...	The g pattern FDIC b
1432	BUSINESS	Everything You Need To Know	Harry Bradford	https://www.huffingtonpost.com/entry/bank-fees...	Don't like k all of your stuf

		About Overdraft Fe...			
1433	BUSINESS	Walmart Waving Goodbye To Some Greeters		https://www.huffingtonpost.comhttp://jobs.aol....	After 30 "People Gr will no
1434	BUSINESS	At World Economic Forum, Fear of Global Contag...	Peter S. Goodman, Contributor\nExecutive Busin...	https://www.huffingtonpost.com/entry/world-eco...	For deca crise a deve
1435	BUSINESS	Positive Customer Experience: What's the Retur...	Ernan Roman, Contributor\nPresident	https://www.huffingtonpost.com/entry/positive-...	"Ana Adobe cor his pu

1436 rows × 6 columns

In [75]:
```
news=df[['headline','category']]
```

In [76]:
```
news
```

Out[76]:

	headline	category
0	There Were 2 Mass Shootings In Texas Last Week...	CRIME
1	Will Smith Joins Diplo And Nicky Jam For The 2...	ENTERTAINMENT
2	Hugh Grant Marries For The First Time At Age 57	ENTERTAINMENT
3	Jim Carrey Blasts 'Castrato' Adam Schiff And D...	ENTERTAINMENT
4	Julianna Margulies Uses Donald Trump Poop Bags...	ENTERTAINMENT
...	...	...
1431	Four More Bank Closures Mark the Week of Janua...	BUSINESS
1432	Everything You Need To Know About Overdraft Fe...	BUSINESS
1433	Walmart Waving Goodbye To Some Greeters	BUSINESS
1434	At World Economic Forum, Fear of Global Contag...	BUSINESS
1435	Positive Customer Experience: What's the Retur...	BUSINESS

1436 rows × 2 columns

In [77]:
```
news.category.value_counts()
```

Out[77]:
```
category
POLITICS          516
ENTERTAINMENT     315
COMEDY            101
WORLD NEWS         81
BLACK VOICES       75
QUEER VOICES       71
WEIRD NEWS         41
CRIME              39
```

```
MEDIA              37
WOMEN              29
TECH               27
SPORTS             22
IMPACT             19
BUSINESS           16
RELIGION           13
TRAVEL             11
SCIENCE             7
ENVIRONMENT         5
LATINO VOICES       4
EDUCATION           4
CULTURE & ARTS      3
Name: count, dtype: int64
```

In [ ]: