In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
data = pd.read_csv('text_classification.csv')
data
```

Out[2]:

	Review	Label
0	I love this sandwich	pos
1	this is an amazing place	pos
2	I feel very good about these beers	pos
3	this is my best work	pos
4	what an awesome view	pos
5	I do not like this restaurant	neg
6	I am tired of this stuff	neg
7	I can't deal with this	neg
8	he is my sworn enemy	neg
9	my boss is horrible	neg
10	this is an awesome place	pos
11	I donot like the taste of this juice	neg
12	I love to dance	pos
13	I am sick and tired of this place	neg
14	what a great holiday	pos
15	that is a bad locality to stay	neg
16	we will have good fun tomorrow	pos
17	I went to my enemy's house today	neg

In [3]:
```
n = len(data)
train_length = int(input("Enter the length of training data:"))
test_length = n - train_length
df = data[:train_length]
test = data[train_length:]
```

```
Enter the length of training data:14
```

In [4]:
```
positive_vocabulary = []
negative_vocabulary = []
all_positive = []
all_negative = []
vocabulary = []
for i in range(len(df)):
    l = df.iloc[i][0].split()
    for j in l:
        if j not in vocabulary:
            vocabulary.append(j)
    if df.iloc[i][1] == 'pos':
        l = df.iloc[i][0].split()
        for j in l:
            if j not in positive_vocabulary:
                positive_vocabulary.append(j)
            all_positive.append(j)
    else:
        l = df.iloc[i][0].split()
        for j in l:
            if j not in negative_vocabulary:
                negative_vocabulary.append(j)
            all_negative.append(j)
print(vocabulary,len(vocabulary))
```

```
['I', 'love', 'this', 'sandwich', 'is', 'an', 'amazing', 'place', 'feel', 've
ry', 'good', 'about', 'these', 'beers', 'my', 'best', 'work', 'what', 'awesom
e', 'view', 'do', 'not', 'like', 'restaurant', 'am', 'tired', 'of', 'stuff',
"can't", 'deal', 'with', 'he', 'sworn', 'enemy', 'boss', 'horrible', 'donot',
'the', 'taste', 'juice', 'to', 'dance', 'sick', 'and'] 44
```

In [5]:
```
print(positive_vocabulary,len(positive_vocabulary))
```

```
['I', 'love', 'this', 'sandwich', 'is', 'an', 'amazing', 'place', 'feel', 've
ry', 'good', 'about', 'these', 'beers', 'my', 'best', 'work', 'what', 'awesom
e', 'view', 'to', 'dance'] 22
```

In [6]:
```
print(negative_vocabulary,len(negative_vocabulary))
```

```
['I', 'do', 'not', 'like', 'this', 'restaurant', 'am', 'tired', 'of', 'stuf
f', "can't", 'deal', 'with', 'he', 'is', 'my', 'sworn', 'enemy', 'boss', 'hor
rible', 'donot', 'the', 'taste', 'juice', 'sick', 'and', 'place'] 27
```

In [7]:
```
p = len(df[df['Label'] == 'pos'])/len(df)
q = len(df[df['Label'] == 'neg'])/len(df)
```

In [8]:
```
df[df['Label'] == 'neg']
```

Out[8]:

	Review	Label
5	I do not like this restaurant	neg
6	I am tired of this stuff	neg
7	I can't deal with this	neg
8	he is my sworn enemy	neg
9	my boss is horrible	neg
11	I donot like the taste of this juice	neg
13	I am sick and tired of this place	neg

In [9]:
```
print(p,q)
```

```
0.5 0.5
```

In [10]:
```
positive_probability = []
negative_probability = []
for i in range(len(vocabulary)):
    positive_probability.append((all_positive.count(vocabulary[i])+1)/(len(voca
    negative_probability.append((all_negative.count(vocabulary[i])+1)/(len(voca
```

In [ ]:

In [ ]:

In [14]:
```
res = []
for i in range(test_length):
    l = test.iloc[i][0].split()
    pos_prob = p
    neg_prob = q
    for j in range(len(l)):
        if l[j] in vocabulary:
            ind = vocabulary.index(l[j])
            pos_prob *= positive_probability[ind]
        else:
            pos_prob *= 1/(len(vocabulary)+len(positive_vocabulary))
    for j in range(len(l)):
        if l[j] in vocabulary:
            ind = vocabulary.index(l[j])
            neg_prob *= negative_probability[ind]
        else:
            neg_prob *= 1/(len(vocabulary)+len(negative_vocabulary))

    print(pos_prob,neg_prob)
    if pos_prob > neg_prob:
        res.append('pos')
    else:
        res.append('neg')
res
```

```
4.459370850555424e-08 1.624413188857591e-08
5.249900933054818e-13 1.124093956665292e-13
1.0237306819456895e-11 3.222402675773837e-12
8.884447732862e-13 5.568186343481562e-13
```

Out[14]:
```
['pos', 'pos', 'pos', 'pos']
```

In [ ]:

In [12]:
```
test
```

Out[12]:

	Review	Label
14	what a great holiday	pos
15	that is a bad locality to stay	neg
16	we will have good fun tomorrow	pos
17	I went to my enemy's house today	neg