In [1]:
```
import pandas as pd
df=pd.read_csv("f:/dataset/sentiment/Restaurant_Reviews.txt",sep="\t")
X=df.Review
y=df.Liked
```

In [4]:
```
from sklearn.feature_extraction.text import CountVectorizer,TfidfVectorizer,ENGLISH_STOP
```

In [3]:
```
from sklearn.naive_bayes import BernoulliNB,MultinomialNB,GaussianNB
```

In [5]:
```
sw=list(ENGLISH_STOP_WORDS)
sw.remove('not')
```

In [7]:
```
cv=CountVectorizer(lowercase=True,stop_words=sw,binary=False)
X_new=cv.fit_transform(X)
model=MultinomialNB() #work well with discrete features(i.e. count)
model.fit(X_new,y)
model.predict(cv.transform(['food quality was good']))
```

Out[7]:
```
array([1], dtype=int64)
```

In [8]:
```
cv=CountVectorizer(lowercase=True,stop_words=sw,binary=True)
X_new=cv.fit_transform(X)
model=BernoulliNB() #work well with binary features(i.e. 0|1)
model.fit(X_new,y)
model.predict(cv.transform(['food quality was good']))
```

Out[8]:
```
array([1], dtype=int64)
```

In [11]:
```
tv=TfidfVectorizer()
X_new=tv.fit_transform(X).toarray()
model=GaussianNB() #work well with normalized features
model.fit(X_new,y)
model.predict(tv.transform(['food quality was good']).toarray())
```

Out[11]:
```
array([1], dtype=int64)
```

In [ ]:
```

```