

```
import pandas as pd
import nltk
```

```
df=pd.read_csv('/content/IMDB Dataset.csv')
```

```
df.head(5)
```

|   | review  | sentiment |
|---|---|-----------|
| 0 | One of the other reviewers has mentioned that ... | positive  |
| 1 | A wonderful little production. <br /><br />The... | positive  |
| 2 | I thought this was a wonderful way to spend ti... | positive  |
| 3 | Basically there's a family where a little boy ... | negative  |
| 4 | Petter Mattei's "Love in the Time of Money" is... | positive  |

```
import string
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import CountVectorizer,TfidfTransformer
from sklearn import svm
```

```
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
True
```

```
def text_process(mess):
    nopunc=[char for char in mess if char not in string.punctuation]
    nopunc="".join(nopunc)
    return[word for word in nopunc.split() if word.lower() not in stopwords.]
```

```
from sklearn.pipeline import Pipeline
from sklearn.naive_bayes import MultinomialNB
pipeline_created1=Pipeline([('bow',CountVectorizer(analyzer=text_process))
                             ('tfidf',TfidfTransformer()),
                             ('classifier',MultinomialNB())
                             ])
```

```
from sklearn.model_selection import train_test_split
msg_train,label_train,msg_test,label_test=train_test_split(df['sentiment']
```

```
pipeline_created1.fit(msg_train,label_train)
```

```
Pipeline(steps=[('bow',  
                  CountVectorizer(analyzer=<function text_process at 0x7f17906a03b0>),  
                  ('tfidf', TfidfTransformer()),  
                  ('classifier', MultinomialNB()))])
```



```
pred=pipeline_created1.predict(msg_test)
```

```
pred
```

```
array(['negative', 'negative', 'negative', ..., 'negative', 'negative',  
       'negative'], dtype='<U8')
```

```
label_test
```

```
from sklearn.metrics import classification_report, accuracy_score  
print(classification_report(label_test, pred))  
print(accuracy_score(label_test, pred))
```