

```
In [ ]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
In [ ]: import pathlib
import sklearn
import numpy
import nltk
nltk.download('stopwords')
nltk.download('punkt')
import pandas as pd
from sklearn.datasets import load_files
from nltk.corpus import stopwords
from collections import Counter
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

```
In [ ]: # train Data
trainData = pd.read_csv("/content/drive/MyDrive/text_mining/dataset_shrunk.csv")
print(trainData)
# test Data
testData = pd.read_table("/content/drive/MyDrive/text_mining/sentiment-topic-final-test.csv")
print(testData)
```

		sentence id	text \
0		22	Emma Darcy has always been a good writer I lov...
1		32	WOW! This is a major league book of hotness! ...
2		39	Love this case! It protects the Kindle and op...
3		48	Loved all the characters in each book cant wai...
4		62	Drew finds out from his girlfriend's reading m...
...	
4495		101830	After a long hot summer, it's back to school f...
4496		101901	Movies are influenced not copied and in this c...
4497		101924	It a low budget disaster film. It's not JJ abr...
4498		101939	A vintage novel with a "national treasure" qua...
4499		101947	I didn't know much about this new Brazilian mo...

		sentence	id	sentiment	topic
0		positive	book		
1		positive	book		
2		positive	book		
3		positive	book		
4		positive	book		
...			
4495		neutral	movie		
4496		neutral	movie		
4497		neutral	movie		
4498		neutral	movie		
4499		neutral	movie		

[4500 rows x 4 columns]

		sentence id	text	sentiment \
0		0	It took eight years for Warner Brothers to rec...	negative
1		1	All the New York University students love this...	positive
2		2	This Italian place is really trendy but they h...	negative
3		3	In conclusion, my review of this book would be...	positive
4		4	The story of this movie is focused on Carl Bra...	neutral

5	5	Chris O'Donnell stated that while filming for ...	neutral
6	6	My husband and I moved to Amsterdam 6 years ag...	positive
7	7	Dame Maggie Smith performed her role excellent...	positive
8	8	The new movie by Mr. Kruno was shot in New Yor...	neutral
9	9	I always have loved English novels, but I just...	negative

	topic
0	movie
1	restaurant
2	restaurant
3	book
4	movie
5	movie
6	restaurant
7	movie
8	movie
9	book

```
In [ ]: # Create feature vectors
vectorizer = CountVectorizer(min_df = 5,
                             tokenizer=nlTK.word_tokenize,
                             stop_words=stopwords.words('english'))
train_vectors = vectorizer.fit_transform(trainData.text)
test_vectors = vectorizer.transform(testData['text'])
```

```
/usr/local/lib/python3.9/dist-packages/sklearn/feature_extraction/text.py:528: UserWarni
ng: The parameter 'token_pattern' will not be used since 'tokenizer' is not None'
warnings.warn(
/usr/local/lib/python3.9/dist-packages/sklearn/feature_extraction/text.py:409: UserWarni
ng: Your stop_words may be inconsistent with your preprocessing. Tokenizing the stop wor
ds generated tokens ['d', 'll', 're', 's', 've', 'could', 'might', 'must', 'n't',
'need', 'sha', 'wo', 'would'] not in stop_words.
warnings.warn(
```

```
In [ ]: tfidf_transformer = TfidfTransformer()
train_vectors_tfidf = tfidf_transformer.fit_transform(train_vectors)
```

```
In [ ]: from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report

clf = MultinomialNB().fit(train_vectors_tfidf, trainData['sentiment'])
prediction_results = clf.predict(test_vectors)

# results
report = classification_report(testData['sentiment'], prediction_results, digits=3)
print(report)

size = test_vectors.shape[0]

for i in range (size):
    print('%s => %s' % (testData.iloc[i]['text'],
                        prediction_results[i]))
```

	precision	recall	f1-score	support
negative	1.000	0.333	0.500	3
neutral	0.500	0.667	0.571	3
positive	0.600	0.750	0.667	4
accuracy			0.600	10
macro avg	0.700	0.583	0.579	10
weighted avg	0.690	0.600	0.588	10

It took eight years for Warner Brothers to recover from the disaster that was this movi

e. => negative

All the New York University students love this diner in Soho so it makes for a fun young atmosphere. => positive

This Italian place is really trendy but they have forgotten about the most important part of a restaurant, the food. => neutral

In conclusion, my review of this book would be: I like Jane Austen and understand why she is famous. => neutral

The story of this movie is focused on Carl Brashear played by Cuba Gooding Jr. who wants to be the first African American deep sea diver in the Navy. => neutral

Chris O'Donnell stated that while filming for this movie, he felt like he was in a toy commercial. => neutral

My husband and I moved to Amsterdam 6 years ago and for as long as we have lived here, B lauwbrog has been our favorite place to eat! => positive

Dame Maggie Smith performed her role excellently, as she does in all her movies. => positive

The new movie by Mr. Kruso was shot in New York, but the story takes place in Los Angeles. => positive

I always have loved English novels, but I just couldn't get into this one. => positive

In []: