

```
In [57]: import numpy as np
import pandas as pd
import re
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.stem.snowball import SnowballStemmer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.model_selection import KFold
import csv

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\ DELL\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
In [58]: train_df=pd.read_csv('C:/Users/ DELL/OneDrive/Documents/Data mining assignment/DM assignment 1/1662145253_9967718_new_train_data.csv')
train_df.columns=['rating','review']
train_df
```

Out[58]:

	rating	review
0	-1	Eat at Floris, they said. You'll like it, they...
1	-1	I just don't understand the appeal. I've trie...
2	1	This is my go to place for a really good beef...
3	-1	Not impressed. When I ordered the Oyako bowl, ...
4	-1	This is the first time ever I wrote a bad revi...
...
17995	-1	I was referred to go to this place by a buddy...
17996	1	The food here was really good. We started off...
17997	1	I eat at this place maybe 2-3 a week! I am die...
17998	1	Phoenix Airport is getting better day by day...
17999	-1	So, when I was much younger I went to Christos...

18000 rows x 2 columns

```
In [59]: test_df=pd.read_csv('C:/Users/ DELL/Downloads/1662145254_0395114_new_test.csv',header=None)
test_df.columns=['review']
```

```
In [7]: train_list=[]
sb=SnowballStemmer(language='english')
for i in range(len(train_df)):
    #using regular expressions to take words without any special characters
    value=re.sub('[^a-zA-Z]', '',train_df.review[i])
    #replacing the new line character with an empty string
    value=value.replace("\n","")
    #splitting the line into words and taking them into list.
    words=value.split()
    #removing stopwords and then stemming the words to consider only important values from the list.
    words_nostopwords=[sb.stem(word) for word in words if not word in set(stopwords.words('english'))]
    words_lower=[word.lower() for word in words_nostopwords if len(word)>2]
    words_join=' '.join(words_lower)
    train_list.append(words_join)

train_list
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

I've been thinking about you a lot lately, and how much I've grown since we last spoke. Life has been a whirlwind of change, but I've managed to find my footing. I'm now in a great place, both personally and professionally. I've started a new job that challenges me and allows me to use my skills. I've also found a great group of friends who support me and make life so much more enjoyable. I'm excited about the future and the possibilities ahead. I hope you're doing well and that life is treating you kindly. I'd love to hear from you again soon. Let's catch up and see how everyone is getting on. I'm sure we'll have a great time catching up on all the news. I'm looking forward to it.

[illegible]

[illegible]

[illegible]