!pip install nltk

Collecting nltk  
 Obtaining dependency information for nltk from https://files.pythonhosted.org/packages/a6/0a/0d20d2c0f16be91b9fa32a77b76c60f9baf6eba419e5ef5deca17af9c582/nltk-3.8.1-py3-none-any.whl.metadata  
 Downloading nltk-3.8.1-py3-none-any.whl.metadata (2.8 kB)  
Requirement already satisfied: click in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from nltk) (8.1.3)  
Requirement already satisfied: joblib in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from nltk) (1.3.2)  
Collecting regex>=2021.8.3 (from nltk)  
 Obtaining dependency information for regex>=2021.8.3 from https://files.pythonhosted.org/packages/ef/9b/0aa55fc101c803869c13b389b718b15810592d2df35b1af15ff5b6f48e16/regex-2024.5.15-cp311-cp311-win\_amd64.whl.metadata  
 Downloading regex-2024.5.15-cp311-cp311-win\_amd64.whl.metadata (41 kB)  
 ---------------------------------------- 0.0/42.0 kB ? eta -:--:--  
 ---------------------------------------- 42.0/42.0 kB 2.1 MB/s eta 0:00:00  
Collecting tqdm (from nltk)  
 Obtaining dependency information for tqdm from https://files.pythonhosted.org/packages/18/eb/fdb7eb9e48b7b02554e1664afd3bd3f117f6b6d6c5881438a0b055554f9b/tqdm-4.66.4-py3-none-any.whl.metadata  
 Downloading tqdm-4.66.4-py3-none-any.whl.metadata (57 kB)  
 ---------------------------------------- 0.0/57.6 kB ? eta -:--:--  
 ---------------------------------------- 57.6/57.6 kB 3.0 MB/s eta 0:00:00  
Requirement already satisfied: colorama in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from click->nltk) (0.4.6)  
Downloading nltk-3.8.1-py3-none-any.whl (1.5 MB)  
 ---------------------------------------- 0.0/1.5 MB ? eta -:--:--  
 - -------------------------------------- 0.0/1.5 MB 1.9 MB/s eta 0:00:01  
 -- ------------------------------------- 0.1/1.5 MB 919.0 kB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 ---- ----------------------------------- 0.2/1.5 MB 1.2 MB/s eta 0:00:02  
 -------- ------------------------------- 0.3/1.5 MB 578.5 kB/s eta 0:00:03  
 --------- ------------------------------ 0.4/1.5 MB 603.4 kB/s eta 0:00:02  
 ---------- ----------------------------- 0.4/1.5 MB 639.0 kB/s eta 0:00:02  
 ---------- ----------------------------- 0.4/1.5 MB 639.0 kB/s eta 0:00:02  
 ---------- ----------------------------- 0.4/1.5 MB 639.0 kB/s eta 0:00:02  
 ---------- ----------------------------- 0.4/1.5 MB 639.0 kB/s eta 0:00:02  
 ---------- ----------------------------- 0.4/1.5 MB 639.0 kB/s eta 0:00:02  
 ---------- ----------------------------- 0.4/1.5 MB 639.0 kB/s eta 0:00:02  
 ------------ --------------------------- 0.5/1.5 MB 510.9 kB/s eta 0:00:03  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 --------------- ------------------------ 0.6/1.5 MB 591.0 kB/s eta 0:00:02  
 ----------------- ---------------------- 0.7/1.5 MB 203.6 kB/s eta 0:00:05  
 ------------------- -------------------- 0.7/1.5 MB 227.9 kB/s eta 0:00:04  
 ------------------- -------------------- 0.7/1.5 MB 227.9 kB/s eta 0:00:04  
 --------------------- ------------------ 0.8/1.5 MB 243.8 kB/s eta 0:00:03  
 ---------------------- ----------------- 0.9/1.5 MB 250.6 kB/s eta 0:00:03  
 ---------------------- ----------------- 0.9/1.5 MB 250.6 kB/s eta 0:00:03  
 ------------------------- -------------- 1.0/1.5 MB 275.0 kB/s eta 0:00:02  
 ------------------------- -------------- 1.0/1.5 MB 275.0 kB/s eta 0:00:02  
 ------------------------- -------------- 1.0/1.5 MB 275.0 kB/s eta 0:00:02  
 ------------------------- -------------- 1.0/1.5 MB 275.0 kB/s eta 0:00:02  
 --------------------------- ------------ 1.0/1.5 MB 274.9 kB/s eta 0:00:02  
 ------------------------------ --------- 1.1/1.5 MB 302.9 kB/s eta 0:00:02  
 ------------------------------ --------- 1.2/1.5 MB 307.3 kB/s eta 0:00:02  
 ------------------------------- -------- 1.2/1.5 MB 315.5 kB/s eta 0:00:01  
 ------------------------------- -------- 1.2/1.5 MB 315.5 kB/s eta 0:00:01  
 ---------------------------------- ----- 1.3/1.5 MB 329.0 kB/s eta 0:00:01  
 ----------------------------------- ---- 1.4/1.5 MB 338.0 kB/s eta 0:00:01  
 ----------------------------------- ---- 1.4/1.5 MB 338.0 kB/s eta 0:00:01  
 ----------------------------------- ---- 1.4/1.5 MB 338.0 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 ------------------------------------- -- 1.4/1.5 MB 342.7 kB/s eta 0:00:01  
 --------------------------------------- 1.5/1.5 MB 331.1 kB/s eta 0:00:01  
 ---------------------------------------- 1.5/1.5 MB 330.0 kB/s eta 0:00:00  
Downloading regex-2024.5.15-cp311-cp311-win\_amd64.whl (268 kB)  
 ---------------------------------------- 0.0/269.0 kB ? eta -:--:--  
 --------------------- ------------------ 143.4/269.0 kB 8.9 MB/s eta 0:00:01  
 --------------------- ------------------ 143.4/269.0 kB 8.9 MB/s eta 0:00:01  
 ---------------------------- ----------- 194.6/269.0 kB 1.5 MB/s eta 0:00:01  
 ---------------------------- ----------- 194.6/269.0 kB 1.5 MB/s eta 0:00:01  
 ---------------------------- ----------- 194.6/269.0 kB 1.5 MB/s eta 0:00:01  
 ---------------------------- ----------- 194.6/269.0 kB 1.5 MB/s eta 0:00:01  
 ------------------------------------- 266.2/269.0 kB 820.5 kB/s eta 0:00:01  
 -------------------------------------- 269.0/269.0 kB 788.8 kB/s eta 0:00:00  
Downloading tqdm-4.66.4-py3-none-any.whl (78 kB)  
 ---------------------------------------- 0.0/78.3 kB ? eta -:--:--  
 -------------------- ------------------- 41.0/78.3 kB 960.0 kB/s eta 0:00:01  
 ---------------------------------------- 78.3/78.3 kB 1.1 MB/s eta 0:00:00  
Installing collected packages: tqdm, regex, nltk  
Successfully installed nltk-3.8.1 regex-2024.5.15 tqdm-4.66.4

[notice] A new release of pip is available: 23.2.1 -> 24.0  
[notice] To update, run: python.exe -m pip install --upgrade pip

import nltk  
nltk.download('punkt')

[nltk\_data] Downloading package punkt to  
[nltk\_data] C:\Users\nikhi\AppData\Roaming\nltk\_data...  
[nltk\_data] Unzipping tokenizers\punkt.zip.

True

from nltk.tokenize import word\_tokenize

import pandas as pd

df=pd.read\_csv('IMDB Dataset.csv', encoding='utf-8')

df.head()

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

reviews = df.review.str.cat(sep=' ')  
#function to split text into word  
tokens = word\_tokenize(reviews)  
vocabulary = set(tokens)  
print(len(vocabulary))  
frequency\_dist = nltk.FreqDist(tokens)  
sorted(frequency\_dist,key=frequency\_dist.\_\_getitem\_\_, reverse=True)[0:50]

194766

['the',  
 ',',  
 '.',  
 'and',  
 'a',  
 'of',  
 'to',  
 'is',  
 '/',  
 '>',  
 '<',  
 'br',  
 'in',  
 'I',  
 'it',  
 'that',  
 "'s",  
 'this',  
 'was',  
 'The',  
 'as',  
 'with',  
 'movie',  
 'for',  
 'film',  
 ')',  
 '(',  
 'but',  
 "''",  
 "n't",  
 '``',  
 'on',  
 'you',  
 'are',  
 'not',  
 'have',  
 'his',  
 'be',  
 '!',  
 'he',  
 'one',  
 'at',  
 'by',  
 'an',  
 'all',  
 'who',  
 'they',  
 'from',  
 'like',  
 'It']

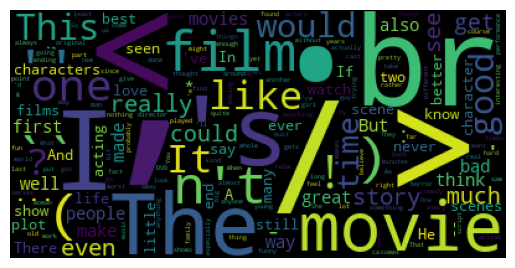
import nltk  
nltk.download('stopwords')  
import string  
from nltk.corpus import stopwords  
stop\_words = set(stopwords.words('english'))  
tokens = [w for w in tokens if not w in stop\_words]  
  
frequency\_dist = nltk.FreqDist(tokens)  
tokens = list(filter(lambda token: token not in string.punctuation, tokens))  
tokens=[tokens for word in tokens if word.isalpha()]  
sorted(frequency\_dist,key=frequency\_dist.\_\_getitem\_\_, reverse=True)[0:50]

[',',  
 '.',  
 '/',  
 '>',  
 '<',  
 'br',  
 'I',  
 "'s",  
 'The',  
 'movie',  
 'film',  
 ')',  
 '(',  
 "''",  
 "n't",  
 '``',  
 '!',  
 'one',  
 'like',  
 'It',  
 '?',  
 'This',  
 'good',  
 'would',  
 'time',  
 'really',  
 'see',  
 'even',  
 'story',  
 "'",  
 '...',  
 ':',  
 '-',  
 'much',  
 'could',  
 'get',  
 'people',  
 'bad',  
 'great',  
 'well',  
 'first',  
 'made',  
 'also',  
 'make',  
 'way',  
 'movies',  
 'But',  
 'think',  
 'characters',  
 '\*']

!pip install wordcloud  
from wordcloud import WordCloud  
import matplotlib.pyplot as plt  
wordcloud = WordCloud().generate\_from\_frequencies(frequency\_dist)  
plt.imshow(wordcloud)  
plt.axis("off")  
plt.show()

[notice] A new release of pip is available: 23.2.1 -> 24.0  
[notice] To update, run: python.exe -m pip install --upgrade pip

Collecting wordcloud  
 Obtaining dependency information for wordcloud from https://files.pythonhosted.org/packages/f5/b0/247159f61c5d5d6647171bef84430b7efad4db504f0229674024f3a4f7f2/wordcloud-1.9.3-cp311-cp311-win\_amd64.whl.metadata  
 Downloading wordcloud-1.9.3-cp311-cp311-win\_amd64.whl.metadata (3.5 kB)  
Requirement already satisfied: numpy>=1.6.1 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from wordcloud) (1.25.2)  
Requirement already satisfied: pillow in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from wordcloud) (10.1.0)  
Requirement already satisfied: matplotlib in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from wordcloud) (3.8.2)  
Requirement already satisfied: contourpy>=1.0.1 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from matplotlib->wordcloud) (1.2.0)  
Requirement already satisfied: cycler>=0.10 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from matplotlib->wordcloud) (0.12.1)  
Requirement already satisfied: fonttools>=4.22.0 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from matplotlib->wordcloud) (4.44.3)  
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from matplotlib->wordcloud) (1.4.5)  
Requirement already satisfied: packaging>=20.0 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from matplotlib->wordcloud) (23.2)  
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from matplotlib->wordcloud) (3.1.1)  
Requirement already satisfied: python-dateutil>=2.7 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from matplotlib->wordcloud) (2.8.2)  
Requirement already satisfied: six>=1.5 in c:\users\nikhi\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)  
Downloading wordcloud-1.9.3-cp311-cp311-win\_amd64.whl (300 kB)  
 ---------------------------------------- 0.0/300.2 kB ? eta -:--:--  
 ---- ----------------------------------- 30.7/300.2 kB ? eta -:--:--  
 ---- ----------------------------------- 30.7/300.2 kB ? eta -:--:--  
 ---- ----------------------------------- 30.7/300.2 kB ? eta -:--:--  
 ---- ----------------------------------- 30.7/300.2 kB ? eta -:--:--  
 ----------- --------------------------- 92.2/300.2 kB 403.5 kB/s eta 0:00:01  
 ----------- --------------------------- 92.2/300.2 kB 403.5 kB/s eta 0:00:01  
 ----------- --------------------------- 92.2/300.2 kB 403.5 kB/s eta 0:00:01  
 ----------- --------------------------- 92.2/300.2 kB 403.5 kB/s eta 0:00:01  
 ----------- --------------------------- 92.2/300.2 kB 403.5 kB/s eta 0:00:01  
 -------------------- ----------------- 163.8/300.2 kB 351.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ---------------------- --------------- 174.1/300.2 kB 338.5 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------ ------------- 194.6/300.2 kB 171.0 kB/s eta 0:00:01  
 ------------------------- ------------ 204.8/300.2 kB 128.3 kB/s eta 0:00:01  
 ------------------------- ------------ 204.8/300.2 kB 128.3 kB/s eta 0:00:01  
 ------------------------------- ------ 245.8/300.2 kB 144.9 kB/s eta 0:00:01  
 ------------------------------- ------ 245.8/300.2 kB 144.9 kB/s eta 0:00:01  
 ------------------------------- ------ 245.8/300.2 kB 144.9 kB/s eta 0:00:01  
 ------------------------------- ------ 245.8/300.2 kB 144.9 kB/s eta 0:00:01  
 ------------------------------- ------ 245.8/300.2 kB 144.9 kB/s eta 0:00:01  
 ------------------------------- ------ 245.8/300.2 kB 144.9 kB/s eta 0:00:01  
 ------------------------------- ------ 245.8/300.2 kB 144.9 kB/s eta 0:00:01  
 ------------------------------- ------ 245.8/300.2 kB 144.9 kB/s eta 0:00:01  
 ----------------------------------- -- 276.5/300.2 kB 135.2 kB/s eta 0:00:01  
 ----------------------------------- -- 276.5/300.2 kB 135.2 kB/s eta 0:00:01  
 ----------------------------------- -- 276.5/300.2 kB 135.2 kB/s eta 0:00:01  
 ----------------------------------- -- 276.5/300.2 kB 135.2 kB/s eta 0:00:01  
 ----------------------------------- -- 276.5/300.2 kB 135.2 kB/s eta 0:00:01  
 ----------------------------------- -- 276.5/300.2 kB 135.2 kB/s eta 0:00:01  
 -------------------------------------- 300.2/300.2 kB 127.9 kB/s eta 0:00:00  
Installing collected packages: wordcloud  
Successfully installed wordcloud-1.9.3



X\_train = df.loc[:24999, 'review'].values  
y\_train = df.loc[:24999, 'sentiment'].values  
X\_test = df.loc[25000:, 'review'].values  
y\_test = df.loc[25000:, 'sentiment'].values

from sklearn.feature\_extraction.text import TfidfTransformer  
from sklearn.feature\_extraction.text import TfidfVectorizer  
vectorizer = TfidfVectorizer()  
train\_vectors = vectorizer.fit\_transform(X\_train)  
test\_vectors = vectorizer.transform(X\_test)  
print(train\_vectors.shape, test\_vectors.shape)

(25000, 76496) (25000, 76496)

from sklearn.naive\_bayes import MultinomialNB  
clf = MultinomialNB().fit(train\_vectors, y\_train)

from sklearn.metrics import accuracy\_score  
predicted = clf.predict(test\_vectors)  
print(accuracy\_score(y\_test,predicted))

0.85884

from sklearn import metrics  
from sklearn.metrics import confusion\_matrix,classification\_report  
from sklearn.metrics import accuracy\_score