SENTIMENT ANALYSIS

In [53]:
```
#importing neccessary libraries

import csv
from bs4 import BeautifulSoup
import requests
import pandas as pd
import time
time.sleep(2)
import warnings
warnings.filterwarnings('ignore')
```

In [54]:
```
#movie review urls

urls = []

url1 = 'https://www.imdb.com/title/tt0157262/reviews/?ref_=tt_ql_2'

url2 = 'https://www.imdb.com/title/tt1530509/reviews/?ref_=tt_ql_2'

url3 = 'https://www.imdb.com/title/tt1160419/reviews/?ref_=tt_ql_2'

url4 = 'https://www.imdb.com/title/tt3973410/reviews/?ref_=tt_ql_2'

url5 = 'https://www.imdb.com/title/tt0133093/reviews/?ref_=tt_ql_2'

url6 = 'https://www.imdb.com/title/tt16280138/reviews/?ref_=tt_ql_2'

url7 = 'https://www.imdb.com/title/tt0105575/reviews/?ref_=tt_ql_2'
```

In [55]:
```
#appending url for each assigned movie review to urls list

urls.append(url1)
urls.append(url2)
urls.append(url3)
urls.append(url4)
urls.append(url5)
urls.append(url6)
urls.append(url7)
```

In [56]:
```
content = []
for url in urls:
    page = requests.get(url, timeout=2.50)
    page_content = page.content
    soup = BeautifulSoup(page_content, 'html.parser')
    content.append(soup.find_all('div', class_= 'review-container'))
```

EAch URL webpage is inspected to view the elements. Upon hovering the cursor over HTML elements, the div tag below contains an example of HTML for the reviews and ratings which are of interest:

< div class="review-container" style="max-height: 300px;">

The review title is also of interest to us and below is an example of the HTML anchor tag for a review title of one of my movies:

< a href="/review/rw1394473/?ref_=tt_urv" class="title"> The Mother of all caste-based movies < /a>

The commented piece of code in the next cell prints what the content of the < div> contains. It is commented out for readability purposes.

In [108…
```
#print(content)
```

In [58]:
```
#defining dataframe called 'movie' to store movie reviews and the user ratir

movie = pd.DataFrame(columns=['Review','Rating'])
```

In [59]:
```
movie.head()
```

Out[59]:

Review	Rating

In [60]:
```
#defining dataframe to store 'Review' and 'Rating'. If rating is not given f
#we ignore that review.
review = []

rating = []
count = 0
for cc in content:
    for c in cc:
        count+= 1

        print('\nMovie review ', count)
        #Get review.
        str = c.find_all('a', attrs={'class':'title'})
        rReview =''
        for s in str:
            #print('Review is: ',s.get_text())
            rReview = s.get_text()

        #Get rating.
        ratings = c.find_all('span', attrs={'class':''})
        rVal = []
        for r in ratings:
            str1 = r.get_text().strip()
            rVal.append(str1)

        val = rVal[0]
        if(len(val) > 2):
            continue
        else:
            review.append(rReview)
            rating.append(val)
            print('Review: ', rReview)
            print('Rating: ',val)

movie['Review'] = review
movie['Rating'] = rating
```

```
Movie review  1

Movie review  2
Review:   Completely terrible. I feel sorry for Leslie Nielsen.

Rating:  2

Movie review  3
Review:   Whatever Happened To Leslie Nielson?

Rating:  1

Movie review  4
```

```
Review:   The irony in the word "travesty" is correct...

Rating:  2

Movie review  5
Review:   There are no heroes

Rating:  2

Movie review  6

Movie review  7
Review:   Things are just getting worse and worse...

Rating:  2

Movie review  8
Review:   Glad to know that I just didn't understand the film

Rating:  3

Movie review  9

Movie review  10
Review:   Tedious and unfunny.

Rating:  1

Movie review  11
Review:   Atrocious and sad, but certainly not funny

Rating:  1

Movie review  12
Review:   An unfortunate display of bad comedy writing

Rating:  4

Movie review  13

Movie review  14
Review:   Leslie Nielsen in outer space

Rating:  8

Movie review  15
Review:   "How did this get 3?"

Rating:  1
```

```
Movie review  16

Movie review  17
Review:   I am left speechless to describe how BAD this is..

Rating:  1

Movie review  18
Review:   Very disappointing

Rating:  1

Movie review  19
Review:   Stunningly bad space spoof with Leslie Nielsen

Rating:  3

Movie review  20

Movie review  21
Review:   Among Leslie Nielsen's worst movies

Rating:  2

Movie review  22
Review:   "How funny!", I thought this morning. But I wasn't thinking of th
is movie.

Rating:  1

Movie review  23
Review:   Some slick packaging around an uninspired, embarrassing mess

Rating:  1

Movie review  24
Review:   A "Travesty" Indeed!

Rating:  1

Movie review  25
Review:   A tired retread of an aging role

Rating:  1

Movie review  26
Review:   Watching a crime scene

Rating:  5
```

```
Movie review  27
Review:   The Human Centipede II (Full Sequence)

Rating:  6

Movie review  28
Review:   Martin has a dream ... that one day ... the world will be a giant
human centipede!

Rating:  6

Movie review  29
Review:   Why Do They Make Movies Like This?

Rating:  1

Movie review  30
Review:   Sick, disturbing & unsettling.

Rating:  1

Movie review  31
Review:   Strips Away the Twisted Charm of the Original

Rating:  1

Movie review  32
Review:   Disgusting and amoral.

Rating:  8

Movie review  33
Review:   Surprisingly it works.

Rating:  2

Movie review  34
Review:   vile, revolting and pointless

Rating:  1

Movie review  35
Review:   Vast improvement over the first film is down right unnerving at t
imes

Rating:  7

Movie review  36

Movie review  37
```

```
Review:   Absolutely Amazing Sequel

Rating:  7

Movie review  38
Review:   As bad as the fecal matter he forces his victims to excrete

Rating:  2

Movie review  39

Movie review  40

Movie review  41
Review:   Watching torture porn isn't for me....and I never finished this o
ne (thank God!).

Rating:  1

Movie review  42
Review:   Too silly to be truly disturbing

Rating:  7

Movie review  43

Movie review  44
Review:   "The Human Centipede II (Full Sequence)" is a bizarre, depraved a
nd often hilariously over-the-top exploitation film with a brilliant meta-e
dge... Worth seeing once...

Rating:  9

Movie review  45

Movie review  46
Review:   100% trashy.

Rating:  7

Movie review  47

Movie review  48
Review:   It's actually good, much better than the first (UNCUT REVIEW)

Rating:  8

Movie review  49
Review:   Ewww
```

```
Rating:  1

Movie review  50
Review:   Disgusting and very well directed, acted, and photographed.

Rating:  7

Movie review  51
Review:   DUNE - A Great Modern Sci-Fi

Rating:  9

Movie review  52
Review:   My movie of 2021, so far.

Rating:  10

Movie review  53
Review:   "He shall know your ways as though born to them."

Rating:  9

Movie review  54
Review:   Amazing cinematic experience

Rating:  9

Movie review  55
Review:   The Beginning

Rating:  8

Movie review  56
Review:   Villeneuve epic

Rating:  8

Movie review  57
Review:   Grandiose visuals and great acting with little story background

Rating:  8

Movie review  58
Review:   Fear Is The Mind-Killer

Rating:  9

Movie review  59
Review:   Masterpiece, I never believed something like this could be possib
le
```

```
Rating:  10

Movie review  60
Review:   Beginning of a masterpiece

Rating:  9

Movie review  61
Review:   Villeneuves tenet

Rating:  10

Movie review  62
Review:   Faithful retelling of the book, perhaps too faithful?

Rating:  9

Movie review  63
Review:   Not a masterpiece - but still the most interesting contemporary f
ilmmaking

Rating:  9

Movie review  64
Review:   Started off sensational, but eventually overlong with too much go
ing on for too little happening.

Rating:  6

Movie review  65
Review:   Love letter to Dune fans, beautiful cinema to film fans

Rating:  9

Movie review  66
Review:   A proper epic

Rating:  8

Movie review  67
Review:   Beautifully crafted movie

Rating:  9

Movie review  68
Review:   A Darker Dune

Rating:  9
```

```
Movie review  69
Review:   I did not read the book and I did not know what to expect exactly
. I LOVED it.

Rating:  10

Movie review  70
Review:   So promising

Rating:  10

Movie review  71
Review:   only one complaint

Rating:  9

Movie review  72
Review:   Dune made me buy the book dune

Rating:  10

Movie review  73
Review:   A great start

Rating:  9

Movie review  74
Review:   You should see it and its good movie

Rating:  10

Movie review  75
Review:   Kaaka Muttai - the pizza 'dream'

Rating:  9

Movie review  76
Review:   In simple word 'Wonderful'

Rating:  9

Movie review  77
Review:   Kaaka Muttai is a heart-touching phenomenal apologue of two broth
ers and their pursuit of wanting to earn Pizza.

Rating:  9

Movie review  78
Review:   HEART RENDING
```

```
Rating:  9

Movie review  79
Review:   Absolutely brilliant subtext. Must watch movie

Rating:  10

Movie review  80
Review:   Subtle and heart touching

Rating:  8

Movie review  81
Review:   Great Movie with Award Winning Performances by the Kids

Rating:  8

Movie review  82
Review:   Subtitle based review of the Crow's nest

Rating:  7

Movie review  83
Review:   Soul Touching !!

Rating:  8

Movie review  84
Review:   When PIZZA was replaced for the CROW'S EGG.

Rating:  10

Movie review  85
Review:   One of the best i ve seen in the history of Tamil cinema

Rating:  10

Movie review  86
Review:   Beautiful movie

Rating:  10

Movie review  87
Review:   It's all about CHILDHOOD.

Rating:  10

Movie review  88
Review:   Slum boys' pizza dream
```

```
Rating:  10

Movie review  89
Review:   On a quest to taste pizza

Rating:  9

Movie review  90
Review:   Captures the true spirit of India and the impact of the global ec
onomy

Rating:  9

Movie review  91
Review:   For the first taste of Pizza

Rating:  10

Movie review  92
Review:   India's Answer to Persian Movies

Rating:  10

Movie review  93
Review:   Had potential as a people-power primer to striking, but weak inco
nsistent ending.

Rating:  7

Movie review  94
Review:   An Outstanding Satire

Rating:  8

Movie review  95
Review:   Super

Rating:  10

Movie review  96
Review:   Pure

Rating:  10

Movie review  97
Review:   Worth all the hype.

Rating:  7

Movie review  98
```

```
Review:   Beautiful

Rating:  8

Movie review  99
Review:   Could had been better

Rating:  3

Movie review  100
Review:   Just wow

Rating:  10

Movie review  101
Review:   Benchmark forever.

Rating:  10

Movie review  102
Review:   The timeless classic.

Rating:  10

Movie review  103
Review:   A watershed moment in film-making - and what a kick-ass masterpie
ce

Rating:  10

Movie review  104

Movie review  105
Review:   Immensely entertaining, intriguingly philosophical and just about
one of the best films ever made!

Rating:  10

Movie review  106
Review:   Welcome2the"REEL"World: Where the Medium IS the Message!

Rating:  10

Movie review  107
Review:   Ah yes. My first existential crisis.

Rating:  10

Movie review  108
Review:   Agent Smith: Human beings are a disease, a cancer of this planet.
```

```
You're a plague and we are the cure.

Rating:  10

Movie review  109

Movie review  110
Review:   A sci-fi action thriller milestone

Rating:  9

Movie review  111
Review:   The Matrix Is one of the best Classic Sci-Fi Action Film ever

Rating:  10

Movie review  112
Review:   Exhilarating 4DX profound experience!

Rating:  10

Movie review  113
Review:   So well written that makes you questioning the reality. ⭐

Rating:  10

Movie review  114
Review:   20 years on from release, some random thoughts on revisiting The
Matrix. Spoiler: It's still brilliant.

Rating:  10

Movie review  115
Review:   ...it is not the spoon that bends, it is only yourself.

Rating:  9

Movie review  116
Review:   One of my favorites

Rating:  9

Movie review  117
Review:   Welcome to the Real World.

Rating:  10

Movie review  118
Review:   Still hip as heck even after 20 years
```

```
Rating:  10

Movie review  119
Review:   I seriously know what the matrix is.

Rating:  10

Movie review  120

Movie review  121
Review:   The benchmark for all sci-fi films to come

Rating:  10

Movie review  122
Review:   The world of computers

Rating:  10

Movie review  123
Review:   Wow! I Finally Saw It!

Rating:  9

Movie review  124
Review:   The more you watch it, the better it gets

Rating:  10

Movie review  125
Review:   The realism of the first two is forgotten for this complete fanta
sy of a sequel

Rating:  3

Movie review  126
Review:   Lost Some Magic

Rating:  4

Movie review  127
Review:   😳 Other Than the End Show, This Movie was an Absolute Joke 😳

Rating:  2

Movie review  128
Review:   My oh my!

Rating:  1
```

```
Movie review  129
Review:   What is this rated R for exactly?

Rating:  3

Movie review  130
Review:   Oh Gosh, this is awful!

Rating:  2

Movie review  131
Review:   Mike lost is magic.

Rating:  3

Movie review  132
Review:   No Magic Left in Last Dance

Rating:  1

Movie review  133
Review:   Magic Mike's Last Dance is a half-hearted cash grab that lacks pu
rpose and direction.

Rating:  4

Movie review  134
Review:   What did I just watch? I want my money back.

Rating:  2

Movie review  135
Review:   Kind of boring

Rating:  4

Movie review  136
Review:   They managed to kill Magic Mike as well

Rating:  3

Movie review  137
Review:   No magic here !

Rating:  2

Movie review  138
Review:   Tragic Mike: The Last Cash Grab

Rating:  1
```

```
Movie review  139
Review:   An 11 year old (!) voices over this utter trash

Rating:  1

Movie review  140
Review:   Not good

Rating:  1

Movie review  141
Review:   I was the only guy in the entire audience!

Rating:  7

Movie review  142
Review:   Magic who?!

Rating:  2

Movie review  143
Review:   This can't be how Mike goes out!

Rating:  5

Movie review  144
Review:   A soulless dance film

Rating:  3

Movie review  145
Review:   Magic Mike may disappoint some but goes in a new direction.

Rating:  7

Movie review  146
Review:   Magic Letdown

Rating:  2

Movie review  147
Review:   Not funny bad, but bad bad

Rating:  1

Movie review  148
Review:   I nearly fell asleep

Rating:  1
```

```
Movie review  149
Review:   Steven Soderbergh returns to the series as director, but lacks th
e talent of the first film.

Rating:  4

Movie review  150

Movie review  151
Review:   The Mother of all caste-based movies

Rating:  10

Movie review  152
Review:   Kamal's career best performance!!

Rating:  10

Movie review  153
Review:   Devar magan - a tribute to tamil Cinema

Rating:  10

Movie review  154
Review:   my best Indian movie...

Rating:  9

Movie review  155
Review:   Class unmatchable !

Rating:  10

Movie review  156
Review:   Great Juxtaposition of Varied Acting Styles

Rating:  8

Movie review  157

Movie review  158
Review:   Awestuck experience!!!

Rating:  10

Movie review  159
Review:   A movie which gives the satisfaction forever

Rating:  10
```

```
Movie review  160
Review:   Thevar Magan

Rating:  8

Movie review  161
Review:   Copy Paste 75%.

Rating:  3

Movie review  162
Review:   The best inspiration of godfather.One of the best Indian movie ev
er

Rating:  10

Movie review  163
Review:   The Best among all Classics...

Rating:  10

Movie review  164
Review:   All Time Best Indian-Tamil Language Movie

Rating:  10

Movie review  165
Review:   A masterpice in Tamil film industry

Rating:  9

Movie review  166
Review:   Excellent Film Making and Top notch performance

Rating:  9

Movie review  167
Review:   Beyond oscars

Rating:  10

Movie review  168
Review:   This film i consider is the first of its kind which introduced a
very bad tradition of upholding caste pride and creation of a subculture.

Rating:  4

Movie review  169
Review:   A Masterpiece!
```

```
Rating:  10

Movie review  170
Review:   What a movie

Rating:  9

Movie review  171
Review:   THEVARMAGAN MOVIE NOT FOLLOW GODFATHER ,KAMAL SAI IN INTERVIEW IS
AN FOLLOW THE STEP PATH OF KANNADA MOVIE ;;;''KAADU 1973'''''

Rating:  10

Movie review  172
Review:   Legendary movie

Rating:  10
```

In [61]:
```python
#displaying first 5 rows of 'movie' dataframe

movie.head(5)
```

Out[61]:

	Review	Rating
0	Completely terrible. I feel sorry for Leslie ...	2
1	Whatever Happened To Leslie Nielson?\n	1
2	The irony in the word "travesty" is correct...\n	2
3	There are no heroes\n	2
4	Things are just getting worse and worse...\n	2

In [62]:
```python
#displaying dimension of dataframe

movie.shape
```

Out[62]: (155, 2)

We have 155 rows of data, representing 155 reviews and ratings, to perform sentiment analysis

In [63]:
```python
#saving dataframe to csv file

movie.to_csv('AyokunleOlagunju-2217353.csv', index=False)
```

TEXT PROCESSING AND ANALYSIS

In [64]:
```
#importing necessary libraries

import string
import re
import nltk
#nltk.download()
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.model_selection import train_test_split
```

In [65]:
```
#defining dataframe 'textFeatures' for movie review texts and displaying dat

textFeatures = movie['Review'].copy()
textFeatures.shape
```

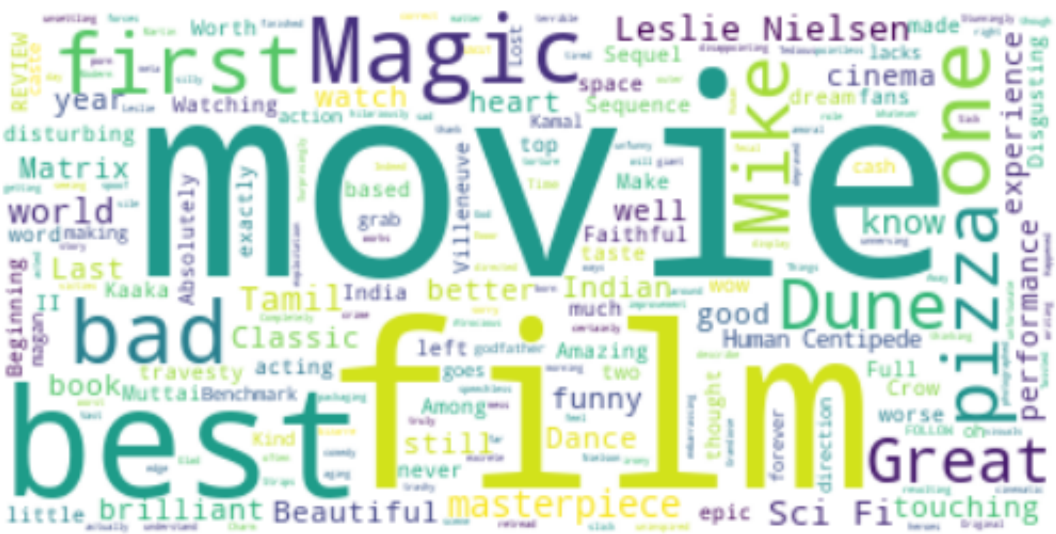
Out[65]: (155,)

In [66]:
```
#Preparing text for Wordcloud
text = []
for t in textFeatures:
  text.append(t)
all_text = ', '.join(t for t in text)
#print(all_text)
print(len(all_text))
```

6648

In [67]:
```
#importing neccessary libraries
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [68]:
```
# Create stopword list
stopwords = set(STOPWORDS)
stopwords.update(["br", "im", "thats"]) #"im","lol","Xa","film"])
# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generat
# Display the image
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
#save the generated image to a file
wordcloud.to_file("wordcloud_cb_all.png")
```



Out[68]: <wordcloud.wordcloud.WordCloud at 0x7fd5b0eb78d0>

SENTIMENT IDENTIFICATION USING VADER

In [69]:
```
import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     /Users/ayokunlejames/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

In [70]:
```
sid = SentimentIntensityAnalyzer()
c = 0
for t in text:
    c+=1
    print(c, t)
    ss = sid.polarity_scores(t)
    print(ss)

    if(ss['compound'] >= 0.05):
        print('positive')

    elif(ss['compound'] <= -0.05):
        print('negative')
    else:
        print('neutral')
    print('\n')
```

```
1  Completely terrible. I feel sorry for Leslie Nielsen.

{'neg': 0.498, 'neu': 0.502, 'pos': 0.0, 'compound': -0.6068}
negative


2  Whatever Happened To Leslie Nielson?

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


3  The irony in the word "travesty" is correct...

{'neg': 0.146, 'neu': 0.854, 'pos': 0.0, 'compound': -0.0516}
negative


4  There are no heroes

{'neg': 0.293, 'neu': 0.267, 'pos': 0.44, 'compound': 0.2732}
positive


5  Things are just getting worse and worse...

{'neg': 0.341, 'neu': 0.659, 'pos': 0.0, 'compound': -0.4767}
negative


6  Glad to know that I just didn't understand the film

{'neg': 0.0, 'neu': 0.727, 'pos': 0.273, 'compound': 0.4588}
```

```
positive


7  Tedious and unfunny.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


8  Atrocious and sad, but certainly not funny

{'neg': 0.421, 'neu': 0.326, 'pos': 0.253, 'compound': -0.2638}
negative


9  An unfortunate display of bad comedy writing

{'neg': 0.5, 'neu': 0.308, 'pos': 0.192, 'compound': -0.6124}
negative


10  Leslie Nielsen in outer space

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


11  "How did this get 3?"

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


12  I am left speechless to describe how BAD this is..

{'neg': 0.346, 'neu': 0.654, 'pos': 0.0, 'compound': -0.6408}
negative


13  Very disappointing

{'neg': 0.777, 'neu': 0.223, 'pos': 0.0, 'compound': -0.5413}
negative


14  Stunningly bad space spoof with Leslie Nielsen

{'neg': 0.368, 'neu': 0.632, 'pos': 0.0, 'compound': -0.5423}
negative
```

```
15  Among Leslie Nielsen's worst movies

{'neg': 0.506, 'neu': 0.494, 'pos': 0.0, 'compound': -0.6249}
negative


16  "How funny!", I thought this morning. But I wasn't thinking of this mov
ie.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


17  Some slick packaging around an uninspired, embarrassing mess

{'neg': 0.459, 'neu': 0.541, 'pos': 0.0, 'compound': -0.6249}
negative


18  A "Travesty" Indeed!

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


19  A tired retread of an aging role

{'neg': 0.367, 'neu': 0.633, 'pos': 0.0, 'compound': -0.4404}
negative


20  Watching a crime scene

{'neg': 0.636, 'neu': 0.364, 'pos': 0.0, 'compound': -0.5423}
negative


21  The Human Centipede II (Full Sequence)

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


22  Martin has a dream ... that one day ... the world will be a giant human
centipede!

{'neg': 0.0, 'neu': 0.859, 'pos': 0.141, 'compound': 0.3164}
positive
```

```
23  Why Do They Make Movies Like This?

{'neg': 0.0, 'neu': 0.706, 'pos': 0.294, 'compound': 0.3612}
positive


24  Sick, disturbing & unsettling.

{'neg': 0.868, 'neu': 0.132, 'pos': 0.0, 'compound': -0.765}
negative


25  Strips Away the Twisted Charm of the Original

{'neg': 0.0, 'neu': 0.545, 'pos': 0.455, 'compound': 0.6124}
positive


26  Disgusting and amoral.

{'neg': 0.857, 'neu': 0.143, 'pos': 0.0, 'compound': -0.7184}
negative


27  Surprisingly it works.

{'neg': 0.0, 'neu': 0.476, 'pos': 0.524, 'compound': 0.296}
positive


28  vile, revolting and pointless

{'neg': 0.577, 'neu': 0.423, 'pos': 0.0, 'compound': -0.6249}
negative


29  Vast improvement over the first film is down right unnerving at times

{'neg': 0.0, 'neu': 0.786, 'pos': 0.214, 'compound': 0.4588}
positive


30  Absolutely Amazing Sequel

{'neg': 0.0, 'neu': 0.328, 'pos': 0.672, 'compound': 0.624}
positive


31  As bad as the fecal matter he forces his victims to excrete
```

```
{'neg': 0.365, 'neu': 0.566, 'pos': 0.069, 'compound': -0.6908}
negative


32  Watching torture porn isn't for me....and I never finished this one (th
ank God!).

{'neg': 0.276, 'neu': 0.724, 'pos': 0.0, 'compound': -0.636}
negative


33  Too silly to be truly disturbing

{'neg': 0.32, 'neu': 0.291, 'pos': 0.388, 'compound': -0.0772}
negative


34  "The Human Centipede II (Full Sequence)" is a bizarre, depraved and oft
en hilariously over-the-top exploitation film with a brilliant meta-edge...
Worth seeing once...

{'neg': 0.088, 'neu': 0.692, 'pos': 0.219, 'compound': 0.5267}
positive


35  100% trashy.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


36  It's actually good, much better than the first (UNCUT REVIEW)

{'neg': 0.0, 'neu': 0.58, 'pos': 0.42, 'compound': 0.7003}
positive


37  Ewww

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


38  Disgusting and very well directed, acted, and photographed.

{'neg': 0.288, 'neu': 0.509, 'pos': 0.203, 'compound': -0.2516}
negative
```

```
39  DUNE - A Great Modern Sci-Fi

{'neg': 0.0, 'neu': 0.423, 'pos': 0.577, 'compound': 0.6249}
positive


40  My movie of 2021, so far.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


41  "He shall know your ways as though born to them."

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


42  Amazing cinematic experience

{'neg': 0.0, 'neu': 0.345, 'pos': 0.655, 'compound': 0.5859}
positive


43  The Beginning

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


44  Villeneuve epic

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


45  Grandiose visuals and great acting with little story background

{'neg': 0.0, 'neu': 0.661, 'pos': 0.339, 'compound': 0.6249}
positive


46  Fear Is The Mind-Killer

{'neg': 0.516, 'neu': 0.484, 'pos': 0.0, 'compound': -0.4939}
negative


47  Masterpiece, I never believed something like this could be possible
```

```
{'neg': 0.16, 'neu': 0.53, 'pos': 0.31, 'compound': 0.457}
positive


48  Beginning of a masterpiece

{'neg': 0.0, 'neu': 0.328, 'pos': 0.672, 'compound': 0.6249}
positive


49  Villeneuves tenet

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


50  Faithful retelling of the book, perhaps too faithful?

{'neg': 0.0, 'neu': 0.508, 'pos': 0.492, 'compound': 0.7003}
positive


51  Not a masterpiece - but still the most interesting contemporary filmmak
ing

{'neg': 0.163, 'neu': 0.533, 'pos': 0.304, 'compound': 0.4296}
positive


52  Started off sensational, but eventually overlong with too much going on
for too little happening.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


53  Love letter to Dune fans, beautiful cinema to film fans

{'neg': 0.0, 'neu': 0.497, 'pos': 0.503, 'compound': 0.8442}
positive


54  A proper epic

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


55  Beautifully crafted movie
```

```
{'neg': 0.0, 'neu': 0.351, 'pos': 0.649, 'compound': 0.5719}
positive


56  A Darker Dune

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


57  I did not read the book and I did not know what to expect exactly. I LO
VED it.

{'neg': 0.0, 'neu': 0.751, 'pos': 0.249, 'compound': 0.6841}
positive


58  So promising

{'neg': 0.0, 'neu': 0.25, 'pos': 0.75, 'compound': 0.4576}
positive


59  only one complaint

{'neg': 0.524, 'neu': 0.476, 'pos': 0.0, 'compound': -0.296}
negative


60  Dune made me buy the book dune

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


61  A great start

{'neg': 0.0, 'neu': 0.196, 'pos': 0.804, 'compound': 0.6249}
positive


62  You should see it and its good movie

{'neg': 0.0, 'neu': 0.707, 'pos': 0.293, 'compound': 0.4404}
positive


63  Kaaka Muttai - the pizza 'dream'

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

```
neutral


64  In simple word 'Wonderful'

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


65  Kaaka Muttai is a heart-touching phenomenal apologue of two brothers an
d their pursuit of wanting to earn Pizza.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


66  HEART RENDING

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


67  Absolutely brilliant subtext. Must watch movie

{'neg': 0.0, 'neu': 0.55, 'pos': 0.45, 'compound': 0.624}
positive


68  Subtle and heart touching

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


69  Great Movie with Award Winning Performances by the Kids

{'neg': 0.0, 'neu': 0.353, 'pos': 0.647, 'compound': 0.9001}
positive


70  Subtitle based review of the Crow's nest

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


71  Soul Touching !!

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral
```

```
72  When PIZZA was replaced for the CROW'S EGG.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


73  One of the best i ve seen in the history of Tamil cinema

{'neg': 0.0, 'neu': 0.724, 'pos': 0.276, 'compound': 0.6369}
positive


74  Beautiful movie

{'neg': 0.0, 'neu': 0.204, 'pos': 0.796, 'compound': 0.5994}
positive


75  It's all about CHILDHOOD.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


76  Slum boys' pizza dream

{'neg': 0.0, 'neu': 0.6, 'pos': 0.4, 'compound': 0.25}
positive


77  On a quest to taste pizza

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


78  Captures the true spirit of India and the impact of the global economy

{'neg': 0.0, 'neu': 0.71, 'pos': 0.29, 'compound': 0.5423}
positive


79  For the first taste of Pizza

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral
```

```
80  India's Answer to Persian Movies

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


81  Had potential as a people-power primer to striking, but weak inconsiste
nt ending.

{'neg': 0.278, 'neu': 0.722, 'pos': 0.0, 'compound': -0.5927}
negative


82  An Outstanding Satire

{'neg': 0.0, 'neu': 0.333, 'pos': 0.667, 'compound': 0.6124}
positive


83  Super

{'neg': 0.0, 'neu': 0.0, 'pos': 1.0, 'compound': 0.5994}
positive


84  Pure

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


85  Worth all the hype.

{'neg': 0.0, 'neu': 0.612, 'pos': 0.388, 'compound': 0.2263}
positive


86  Beautiful

{'neg': 0.0, 'neu': 0.0, 'pos': 1.0, 'compound': 0.5994}
positive


87  Could had been better

{'neg': 0.0, 'neu': 0.508, 'pos': 0.492, 'compound': 0.4404}
positive


88  Just wow
```

```
{'neg': 0.0, 'neu': 0.208, 'pos': 0.792, 'compound': 0.5859}
positive


89  Benchmark forever.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


90  The timeless classic.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


91  A watershed moment in film-making - and what a kick-ass masterpiece

{'neg': 0.0, 'neu': 0.631, 'pos': 0.369, 'compound': 0.6249}
positive


92  Immensely entertaining, intriguingly philosophical and just about one o
f the best films ever made!

{'neg': 0.0, 'neu': 0.619, 'pos': 0.381, 'compound': 0.8122}
positive


93  Welcome2the"REEL"World: Where the Medium IS the Message!

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


94  Ah yes. My first existential crisis.

{'neg': 0.38, 'neu': 0.37, 'pos': 0.25, 'compound': -0.34}
negative


95  Agent Smith: Human beings are a disease, a cancer of this planet. You'r
e a plague and we are the cure.

{'neg': 0.216, 'neu': 0.784, 'pos': 0.0, 'compound': -0.6597}
negative


96  A sci-fi action thriller milestone
```

```
{'neg': 0.0, 'neu': 0.682, 'pos': 0.318, 'compound': 0.1027}
positive


97  The Matrix Is one of the best Classic Sci-Fi Action Film ever

{'neg': 0.0, 'neu': 0.724, 'pos': 0.276, 'compound': 0.6369}
positive


98  Exhilarating 4DX profound experience!

{'neg': 0.0, 'neu': 0.501, 'pos': 0.499, 'compound': 0.4574}
positive


99  So well written that makes you questioning the reality. ⭐

{'neg': 0.13, 'neu': 0.649, 'pos': 0.222, 'compound': 0.2484}
positive


100  20 years on from release, some random thoughts on revisiting The Matri
x. Spoiler: It's still brilliant.

{'neg': 0.0, 'neu': 0.798, 'pos': 0.202, 'compound': 0.5859}
positive


101  ...it is not the spoon that bends, it is only yourself.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


102  One of my favorites

{'neg': 0.0, 'neu': 0.517, 'pos': 0.483, 'compound': 0.4215}
positive


103  Welcome to the Real World.

{'neg': 0.0, 'neu': 0.571, 'pos': 0.429, 'compound': 0.4588}
positive


104  Still hip as heck even after 20 years
```

```
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


105  I seriously know what the matrix is.

{'neg': 0.254, 'neu': 0.746, 'pos': 0.0, 'compound': -0.1779}
negative


106  The benchmark for all sci-fi films to come

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


107  The world of computers

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


108  Wow! I Finally Saw It!

{'neg': 0.0, 'neu': 0.406, 'pos': 0.594, 'compound': 0.658}
positive


109  The more you watch it, the better it gets

{'neg': 0.0, 'neu': 0.734, 'pos': 0.266, 'compound': 0.4404}
positive


110  The realism of the first two is forgotten for this complete fantasy of
a sequel

{'neg': 0.128, 'neu': 0.872, 'pos': 0.0, 'compound': -0.2263}
negative


111  Lost Some Magic

{'neg': 0.535, 'neu': 0.465, 'pos': 0.0, 'compound': -0.3182}
negative


112  😳 Other Than the End Show, This Movie was an Absolute Joke 😳

{'neg': 0.0, 'neu': 0.82, 'pos': 0.18, 'compound': 0.296}
```

```
positive


113  My oh my!

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


114  What is this rated R for exactly?

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


115  Oh Gosh, this is awful!

{'neg': 0.451, 'neu': 0.549, 'pos': 0.0, 'compound': -0.5093}
negative


116  Mike lost is magic.

{'neg': 0.434, 'neu': 0.566, 'pos': 0.0, 'compound': -0.3182}
negative


117  No Magic Left in Last Dance

{'neg': 0.306, 'neu': 0.694, 'pos': 0.0, 'compound': -0.296}
negative


118  Magic Mike's Last Dance is a half-hearted cash grab that lacks purpose
and direction.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


119  What did I just watch? I want my money back.

{'neg': 0.0, 'neu': 0.843, 'pos': 0.157, 'compound': 0.0772}
positive


120  Kind of boring

{'neg': 0.535, 'neu': 0.465, 'pos': 0.0, 'compound': -0.3182}
negative
```

```
121  They managed to kill Magic Mike as well

{'neg': 0.367, 'neu': 0.469, 'pos': 0.164, 'compound': -0.5574}
negative


122  No magic here !

{'neg': 0.555, 'neu': 0.445, 'pos': 0.0, 'compound': -0.3595}
negative


123  Tragic Mike: The Last Cash Grab

{'neg': 0.375, 'neu': 0.625, 'pos': 0.0, 'compound': -0.4588}
negative


124  An 11 year old (!) voices over this utter trash

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


125  Not good

{'neg': 0.706, 'neu': 0.294, 'pos': 0.0, 'compound': -0.3412}
negative


126  I was the only guy in the entire audience!

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


127  Magic who?!

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


128  This can't be how Mike goes out!

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral
```

```
129  A soulless dance film

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


130  Magic Mike may disappoint some but goes in a new direction.

{'neg': 0.171, 'neu': 0.829, 'pos': 0.0, 'compound': -0.2144}
negative


131  Magic Letdown

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


132  Not funny bad, but bad bad

{'neg': 0.129, 'neu': 0.152, 'pos': 0.719, 'compound': 0.8304}
positive


133  I nearly fell asleep

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


134  Steven Soderbergh returns to the series as director, but lacks the tal
ent of the first film.

{'neg': 0.0, 'neu': 0.802, 'pos': 0.198, 'compound': 0.5719}
positive


135  The Mother of all caste-based movies

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


136  Kamal's career best performance!!

{'neg': 0.0, 'neu': 0.385, 'pos': 0.615, 'compound': 0.6988}
positive


137  Devar magan - a tribute to tamil Cinema
```

```
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


138  my best Indian movie...

{'neg': 0.0, 'neu': 0.417, 'pos': 0.583, 'compound': 0.6369}
positive


139  Class unmatchable !

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


140  Great Juxtaposition of Varied Acting Styles

{'neg': 0.0, 'neu': 0.549, 'pos': 0.451, 'compound': 0.6249}
positive


141  Awestuck experience!!!

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


142  A movie which gives the satisfaction forever

{'neg': 0.0, 'neu': 0.633, 'pos': 0.367, 'compound': 0.4404}
positive


143  Thevar Magan

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


144  Copy Paste 75%.

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


145  The best inspiration of godfather.One of the best Indian movie ever

{'neg': 0.0, 'neu': 0.404, 'pos': 0.596, 'compound': 0.9153}
```

```
positive


146  The Best among all Classics...

{'neg': 0.0, 'neu': 0.488, 'pos': 0.512, 'compound': 0.6369}
positive


147  All Time Best Indian-Tamil Language Movie

{'neg': 0.0, 'neu': 0.543, 'pos': 0.457, 'compound': 0.6369}
positive


148  A masterpice in Tamil film industry

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


149  Excellent Film Making and Top notch performance

{'neg': 0.0, 'neu': 0.476, 'pos': 0.524, 'compound': 0.6705}
positive


150  Beyond oscars

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


151  This film i consider is the first of its kind which introduced a very
bad tradition of upholding caste pride and creation of a subculture.

{'neg': 0.128, 'neu': 0.606, 'pos': 0.266, 'compound': 0.4779}
positive


152  A Masterpiece!

{'neg': 0.0, 'neu': 0.0, 'pos': 1.0, 'compound': 0.6588}
positive


153  What a movie

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral
```

```
154  THEVARMAGAN MOVIE NOT FOLLOW GODFATHER ,KAMAL SAI IN INTERVIEW IS AN F
OLLOW THE STEP PATH OF KANNADA MOVIE ;;;''KAADU 1973'''''

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral


155  Legendary movie

{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
neutral
```

SENTIMENT CLASSIFICATION USING MACHINE LEARNING

Preparing Truth Set

'User Rating' will be used as a truth set to evaluate the results. To prepare 'Truth Set', three classes are defined as - 'positive', 'negative', and 'netural'. On the scale of 0 to 10, consider review being:

- 'positvie' if the rating is from 6 to 10,
- 'negative' if the rating is from 0 to 4,
- 'netural' if rating is 5.

In [71]:
```
#the above information is added in a new column called 'class-label'.
label = []
for r in movie['Rating']:
    r = int(r)
    if (r>5):
        label.append('1') #Positive
    elif(r<5):
        label.append('-1') #Negative
    elif(r==5):
        label.append('0') #Netural
movie['class-label'] = label
```

In [72]:
```
#displaying top 5 rows of movie dataframe

movie.head()
```

Out[72]:

	Review	Rating	class-label
0	Completely terrible. I feel sorry for Leslie ...	2	-1
1	Whatever Happened To Leslie Nielson?\n	1	-1
2	The irony in the word "travesty" is correct...\n	2	-1
3	There are no heroes\n	2	-1
4	Things are just getting worse and worse...\n	2	-1

In [73]:
```
#displaying count of each value of 'class-label'

movie['class-label'].value_counts()
```

Out[73]:
```
1     102
-1     51
0       2
Name: class-label, dtype: int64
```

As can be observed, there is an imbalance in the dataset. We can attempt to fix this by removing the neutral class and making our dataset a binary classification set with positive(1) and negative(-1) classes only.

In [74]:
```
#code below ignores the neutral class, where class-label = 0

movie = movie[movie['class-label']!='0']
```

In [75]:
```
#displaying count of unique values of class-label

movie['class-label'].value_counts()
```

Out[75]:
```
1     102
-1     51
Name: class-label, dtype: int64
```

In [76]:
```
#defining dataframe 'textFeatures' for movie review texts and displaying dat


textFeatures = movie['Review'].copy()
textFeatures.shape
```

Out[76]:
```
(153,)
```

In [77]:
```
#importing necessary libraries

import nltk
nltk.download('punkt')
# Stemming using TextBlob library for stemming
from textblob import TextBlob
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     /Users/ayokunlejames/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

In [78]:
```
#defining textblob tokenizer

def textblob_tokenizer(input_str):
    blob = TextBlob(input_str.lower())
    tokens = blob.words
    words = [token.stem() for token in tokens]
    return words
```

In [79]:
```
#Toy example:
print(textblob_tokenizer('Q: studed studing!!! I miss uuuu! It&#039;s'))
```

```
['q', 'stude', 'stude', 'i', 'miss', 'uuuu', 'it', '039', 's']
```

The above Toy example demonstrates that TextBlob() function has converted given sentence into tokens or words, converted upper case letters to lower case, and removed wild characters.

TRY FOR FUN

Try to identify which characters have been removed. Also, try to identify what is the lenght of the output.

Answer:

Characters which have been removed from text include: ':', '!', '!', '!', '!', '&', '#', ';'. Also removed are 'd' in 'studed' and 'i', 'n', 'g' in 'studing'.

In [80]:
```
#retrieving length of textblob tokenizer toy example output, by storing to a
#and printing length using len() function

print("Length of the output is: {}".format(len(list(textblob_tokenizer('Q: s
```

```
Length of the output is: 9
```

TRANSFORMING TEXT DATASET INTO TWO MATRIX REPRESENTATIONS

In [81]:
```
#countvectorizer converts each review into a vector based on the word count.
countvectorizer = CountVectorizer(analyzer= 'word', stop_words= 'english',
                                  tokenizer=textblob_tokenizer)
#converts text into a vector based on tf-idf weighting scheme.
tfidfvectorizer = TfidfVectorizer(analyzer= 'word', stop_words= 'english',
                                  tokenizer=textblob_tokenizer)
```

In [82]:
```
textFeatures
```

Out[82]:
```
0       Completely terrible. I feel sorry for Leslie ...
1                 Whatever Happened To Leslie Nielson?\n
2       The irony in the word "travesty" is correct...\n
3                                  There are no heroes\n
4           Things are just getting worse and worse...\n
                              ...
150     This film i consider is the first of its kind...
151                                   A Masterpiece!\n
152                                     What a movie\n
153     THEVARMAGAN MOVIE NOT FOLLOW GODFATHER ,KAMAL...
154                                  Legendary movie\n
Name: Review, Length: 153, dtype: object
```

In [83]:
```
count_matrix = countvectorizer.fit_transform(textFeatures)
tfidf_matrix = tfidfvectorizer.fit_transform(textFeatures)
```

In [84]:
```
#Printing the dimensions of the tfidf_matrix, x rows (number of reviews) anc
 # words occur in the entire dataset.)

print(tfidf_matrix.shape)
print(count_matrix.shape)
```

```
(153, 394)
(153, 394)
```

Above numbers show dimension of count_matrix, there are X rows (153 reviews), and Y columns (394 word-features) in the transformed dataset.

Try-For-Fun:

Experiment without adopting pre-processing.

In [104…

```
#converting each review into a vector where preprocessor = None and tokenize
count_vectorizer2 = CountVectorizer(preprocessor=None, tokenizer=None)
tfidf_vectorizer2 = TfidfVectorizer(preprocessor=None, tokenizer=None)
```

In [86]:

```
count_matrix2 = count_vectorizer2.fit_transform(textFeatures)
tfidf_matrix2 = tfidf_vectorizer2.fit_transform(textFeatures)
```

In [87]:

```
#Printing the dimensions of the tfidf_matrix and count_matrix

print(tfidf_matrix.shape)
print(count_matrix.shape)
```

```
(153, 394)
(153, 394)
```

BUILDING ML MODEL

In [88]:

```
features_train, features_test, labels_train, labels_test = train_test_split(
    tfidf_matrix, movie['class-label'], test_size=0.3,random_state=53)
print(features_train.shape, features_test.shape, labels_train.shape, labels_
```

```
(107, 394) (46, 394) (107,) (46,)
```

In [89]:

```
#importing necessary libraries

from sklearn.metrics import classification_report, confusion_matrix
from sklearn.metrics import accuracy_score
```

In [106…

```
#SVM classifier
from sklearn.svm import SVC
print("\nEvaluation for SVM \n")
svc = SVC(kernel='sigmoid', gamma=1.0)
svc.fit(features_train, labels_train)
prediction = svc.predict(features_test)
acc = accuracy_score(labels_test,prediction)
print('Accuracy:', acc)
from sklearn.metrics import precision_score
prec = precision_score(labels_test,prediction, average='weighted')
print('Precision:', prec)
from sklearn.metrics import recall_score
recall = recall_score(labels_test,prediction, average='weighted')
print('Recall:', recall)
from sklearn.metrics import f1_score
f1 = f1_score(labels_test,prediction, average='weighted')
print('F-1 measure: ', f1)
print('\nConfusion Matrix:\n')
print(confusion_matrix(labels_test, prediction))
print(classification_report(labels_test, prediction))
print(prediction)
```

```
Evaluation for SVM

Accuracy: 0.8478260869565217
Precision: 0.8533751142446794
Recall: 0.8478260869565217
F-1 measure:  0.8367989918084436

Confusion Matrix:

[[ 8  6]
 [ 1 31]]
              precision    recall  f1-score   support

          -1       0.89      0.57      0.70        14
           1       0.84      0.97      0.90        32

    accuracy                           0.85        46
   macro avg       0.86      0.77      0.80        46
weighted avg       0.85      0.85      0.84        46

['1' '-1' '1' '1' '-1' '1' '1' '1' '1' '-1' '1' '1' '1' '1' '1' '1' '-1'
 '1' '1' '1' '1' '1' '-1' '1' '1' '1' '1' '-1' '1' '1' '1' '1' '1' '-1'
 '1' '1' '-1' '1' '1' '1' '1' '1' '-1' '1' '1' '1']
```

In [91]:
```
#Decision Tree
print("\nEvaluation for Decision Tree \n")
from sklearn.tree import DecisionTreeClassifier
dtree = DecisionTreeClassifier()
dtree.fit(features_train, labels_train)
prediction = dtree.predict(features_test)
acc = accuracy_score(labels_test,prediction)
print('Accuracy: ', acc)
prec = precision_score(labels_test,prediction, average='weighted')
print('Precision: ', prec)
recall = recall_score(labels_test,prediction, average='weighted')
print('Recall: ', recall)
f1 = f1_score(labels_test,prediction, average='weighted')
print('F-1 measure: ',f1)
print('\nConfusion Matrix:\n')
print(confusion_matrix(labels_test, prediction))
print(classification_report(labels_test, prediction))
```

```
Evaluation for Decision Tree

Accuracy:  0.8478260869565217
Precision:  0.8452851496329757
Recall:  0.8478260869565217
F-1 measure:  0.8421025308241401

Confusion Matrix:

[[ 9  5]
 [ 2 30]]
              precision    recall  f1-score   support

          -1       0.82      0.64      0.72        14
           1       0.86      0.94      0.90        32

    accuracy                           0.85        46
   macro avg       0.84      0.79      0.81        46
weighted avg       0.85      0.85      0.84        46
```

Try-It-Yourself:

- Produce wordclouds of 'positive' and 'negative' reviews independently and reflect on the two wordclouds.

In [92]:
```
movie
```

Out[92]:

	Review	Rating	class-label
0	Completely terrible. I feel sorry for Leslie ...	2	-1
1	Whatever Happened To Leslie Nielson?\n	1	-1
2	The irony in the word "travesty" is correct...\n	2	-1
3	There are no heroes\n	2	-1
4	Things are just getting worse and worse...\n	2	-1
...	...	...	...
150	This film i consider is the first of its kind...	4	-1
151	A Masterpiece!\n	10	1
152	What a movie\n	9	1
153	THEVARMAGAN MOVIE NOT FOLLOW GODFATHER ,KAMAL...	10	1
154	Legendary movie\n	10	1

153 rows × 3 columns

In [93]:

```
#filtering movie dataset to display only movies with positive reviews

positive_movie = movie[movie['class-label'] == '1']
positive_movie.head()
```

Out[93]:

	Review	Rating	class-label
9	Leslie Nielsen in outer space\n	8	1
20	The Human Centipede II (Full Sequence)\n	6	1
21	Martin has a dream ... that one day ... the w...	6	1
25	Disgusting and amoral.\n	8	1
28	Vast improvement over the first film is down ...	7	1

In [94]:

```
#generating positive review text

positiveReviews = positive_movie['Review'].copy()
```

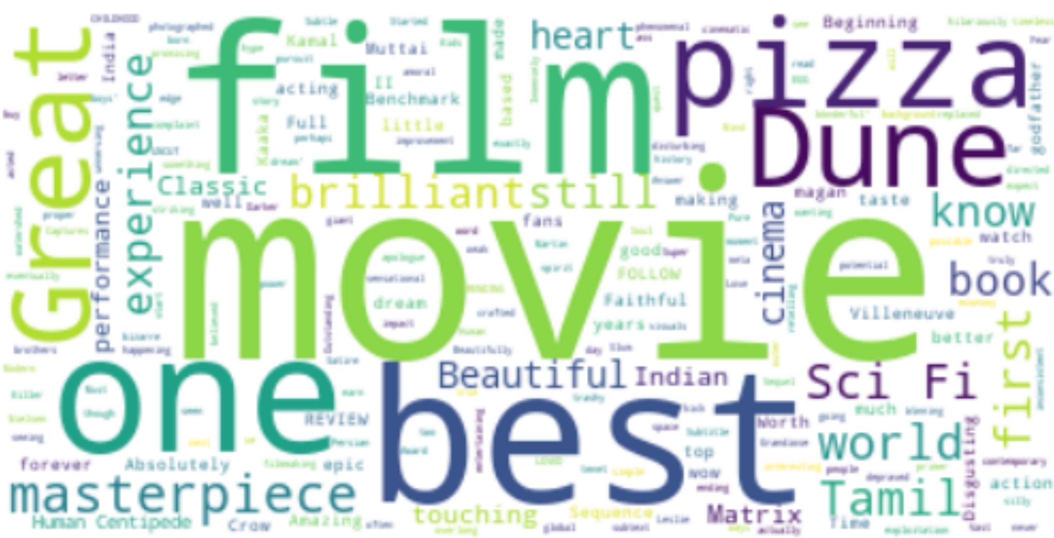
In [95]:

```
#Preparing text from positiveReviews for Wordcloud
text1 = []
for t in positiveReviews:
  text1.append(t)
all_text1 = ', '.join(t for t in text1)
#print(all_text)
print(len(all_text1))
```

```
4464
```

In [96]:

```
# Create stopword list
stopwords = set(STOPWORDS)
stopwords.update(["br", "im", "thats"]) #"im","lol","Xa","film"])
# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generat
# Display the image
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
#save the generated image to a file
wordcloud.to_file("wordcloud_positive.png")
```



Out[96]: `<wordcloud.wordcloud.WordCloud at 0x7fd5b0417650>`

In [ ]:

In [97]:

```
#filtering movie dataset to display only movies with negative reviews

negative_movie = movie[movie['class-label'] == '-1']
negative_movie.head()
```

Out[97]:

	Review	Rating	class-label
0	Completely terrible. I feel sorry for Leslie ...	2	-1
1	Whatever Happened To Leslie Nielson?\n	1	-1
2	The irony in the word "travesty" is correct...\n	2	-1
3	There are no heroes\n	2	-1
4	Things are just getting worse and worse...\n	2	-1

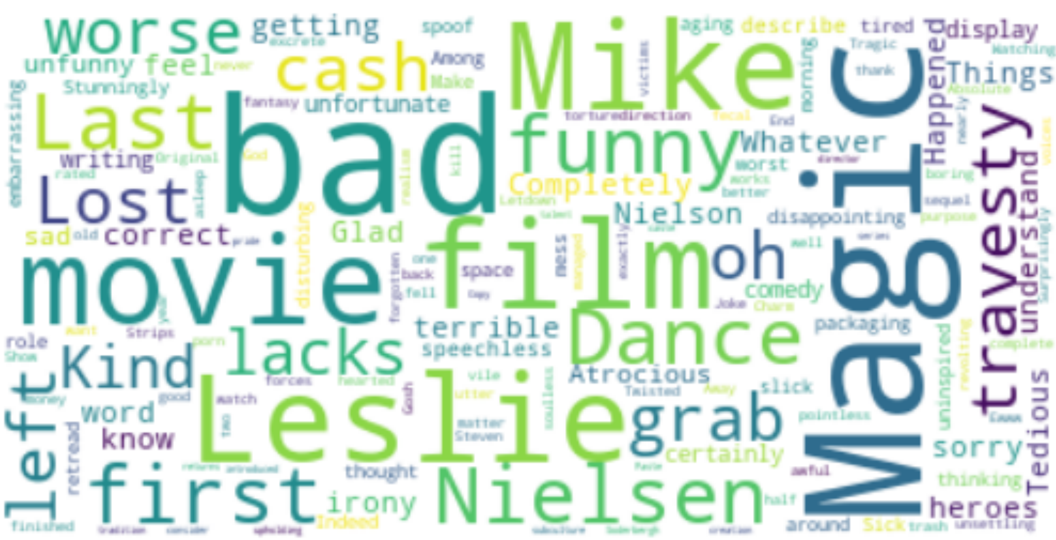
In [98]:
```
#generating negative review text

negativeReviews = negative_movie['Review'].copy()
```

In [99]:
```
#Preparing text from negativeReviews for Wordcloud
text2 = []
for t in negativeReviews:
  text2.append(t)
all_text2 = ', '.join(t for t in text2)
#print(all_text)
print(len(all_text2))
```

```
2120
```

In [100…
```
# Create stopword list
stopwords = set(STOPWORDS)
stopwords.update(["br", "im", "thats"]) #"im","lol","Xa","film"])
# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generat
# Display the image
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
#save the generated image to a file
wordcloud.to_file("wordcloud_negative.png")
```



Out[100]: <wordcloud.wordcloud.WordCloud at 0x7fd5b0b60110>

Try it yourself

- In the above models, we used tf-idf scheme, use BoW scheme (count_matrix) in the similar way to build the two new models. Include these in your evaluation.

In [101…

```
features_train1, features_test1, labels_train1, labels_test1 = train_test_sp
    count_matrix, movie['class-label'], test_size=0.3,random_state=53)
print(features_train1.shape, features_test1.shape, labels_train1.shape, labe
```

```
(107, 394) (46, 394) (107,) (46,)
```

In [107…

```
#SVM classifier
from sklearn.svm import SVC
print("\nEvaluation for SVM \n")
svc = SVC(kernel='sigmoid', gamma=1.0)
svc.fit(features_train1, labels_train1)
prediction = svc.predict(features_test1)
acc = accuracy_score(labels_test1,prediction)
print('Accuracy:', acc)
from sklearn.metrics import precision_score
prec = precision_score(labels_test1,prediction, average='weighted')
print('Precision:', prec)
from sklearn.metrics import recall_score
recall = recall_score(labels_test1,prediction, average='weighted')
print('Recall:', recall)
from sklearn.metrics import f1_score
f1 = f1_score(labels_test1,prediction, average='weighted')
print('F-1 measure: ', f1)
print('\nConfusion Matrix:\n')
print(confusion_matrix(labels_test1, prediction))
print(classification_report(labels_test1, prediction))
print(prediction)
```

```
Evaluation for SVM

Accuracy: 0.8478260869565217
Precision: 0.8452851496329757
Recall: 0.8478260869565217
F-1 measure:  0.8421025308241401

Confusion Matrix:

[[ 9  5]
 [ 2 30]]
              precision    recall  f1-score   support

          -1       0.82      0.64      0.72        14
           1       0.86      0.94      0.90        32

    accuracy                           0.85        46
   macro avg       0.84      0.79      0.81        46
weighted avg       0.85      0.85      0.84        46

['1' '-1' '1' '1' '-1' '-1' '1' '1' '-1' '-1' '1' '1' '1' '1' '1' '-1'
 '-1' '1' '1' '1' '1' '1' '-1' '1' '1' '1' '1' '-1' '1' '1' '1' '1' '1'
 '-1' '1' '1' '-1' '1' '1' '1' '1' '1' '1' '1' '1' '1']
```

In [103…

```python
#Decision Tree
print("\nEvaluation for Decision Tree \n")
from sklearn.tree import DecisionTreeClassifier
dtree = DecisionTreeClassifier()
dtree.fit(features_train1, labels_train1)
prediction = dtree.predict(features_test1)
acc = accuracy_score(labels_test1,prediction)
print('Accuracy: ', acc)
prec = precision_score(labels_test1,prediction, average='weighted')
print('Precision: ', prec)
recall = recall_score(labels_test1,prediction, average='weighted')
print('Recall: ', recall)
f1 = f1_score(labels_test1,prediction, average='weighted')
print('F-1 measure: ',f1)
print('\nConfusion Matrix:\n')
print(confusion_matrix(labels_test1, prediction))
print(classification_report(labels_test1, prediction))
```

```
Evaluation for Decision Tree

Accuracy:  0.8695652173913043
Precision:  0.8674339300937767
Recall:  0.8695652173913043
F-1 measure:  0.8665247795682578

Confusion Matrix:

[[10  4]
 [ 2 30]]
              precision    recall  f1-score   support

          -1       0.83      0.71      0.77        14
           1       0.88      0.94      0.91        32

    accuracy                           0.87        46
   macro avg       0.86      0.83      0.84        46
weighted avg       0.87      0.87      0.87        46
```

Report

- Write a summary on the evaluation performed and the interpretation of the results.

In this workshop, I performed sentiment analysis on text of movie user reviews scrapped from iMdB wesbsite. This analysis is to determine whether a sentiment is positive, negative, or neutral. This sentiment analysis is performed using VADER Valence Aware Dictionary and sEntiment Reasoner). The result of this analysis is examined against the user ratings for each text. To analyze sentiment using VADER, the text is first preprocessed by tokenizing it into words and removing stop words and punctuations. Then, each word is assigned a sentiment score based on the lexicon, and the scores are aggregated to calculate the overall sentiment of the text. VADER also considers the context of the text, such as the presence of emoticons, capitalization, and punctuation, which can affect the sentiment of the text.

Upon completion of the sentiment analysis, and observing the result, it can be observed that VADER correctly classified some text and incorrectly cl;assified others.

- Correct classification

Review: “Completely terrible. I feel sorry for Leslie Nielsen.”

Rating: 2

VADER sentiment: Negative

In this case, VADER has correctly classified the sentiment expressed in the above review text. This must be due to the use of Negative words, like ‘terrible’, ‘sorry’ and the use of words used express the degree of one’s sentiment like “Completely”.

Review: "Absolutely Amazing Sequel"

Rating: 7

VADER classification: Positive

VADER correctly classifies this sentiment as positive due to the use of positive words and literal expression of sentiment rather than figurative. This helps VADER accurately classify this sentiment as positive.

- **Incorrect classification**

Review: “Whatever Happened To Leslie Nielson?”

Rating: 1

VADER classification: Neutral

This review is classified as a neutral sentiment but it had a rating of 1, which indicates a negative sentiment towards the movie. The reason for the incorrect classification cannot be totally blamed on VADER because the user expressed their negative sentiment using neutral words which makes it difficult for VADER to pick up on the negativity of the sentiment. This seems to be a limitation of VADER.

Review: “Glad to know that I just didn't understand the film”

Rating: 3

VADER classification: Positive

This text has also incorrectly classified this text as a positive sentiment, when it is infact negative. This is because the negative sentiment is expressed figuratively, using subtle positive words like “Glad” which even begins the text.

A major limitation I have observed with VADER is its limited understanding of context. VADER's understanding of context is limited to the presence of certain features, such as emoticons, punctuation, and capitalization. However, it may not be able to capture the broader context of the text, such as the tone, intent, or sarcasm. This might cause VADER to not perform very well with figurative language, as it would most likely misinterpret the sentiment. Additionally, because VADER uses a sentiment lexicon that is based on pre-defined lists of words and their corresponding sentiment scores, it may not be able to capture the nuances of sentiments that are not included in the lexicon.

Overall, while VADER is a useful tool for sentiment analysis, it is important to be aware of its limitations and to evaluate its performance on the specific type of text before using it for sentiment analysis.

Further Machine learning classification technique was implemented on the dataset. ML models were built using two matrix representations of the same dataset (tf-idf scheme, BoW scheme (count_matrix)). Two ML models (Support Vector Machine SVM and Decision tree classifier) were built for each scheme and the results of the classification would be discussed in this section.

First, I would examine the results of the SVM classification on both tf-idf and count_matrix.

Evaluation for SVM

tf-idf

- Accuracy: 0.8478260869565217
- Precision: 0.8533751142446794
- Recall: 0.8478260869565217
- F-1 measure: 0.8367989918084436
- Confusion Matrix:

[[8 6]

[1 31]]

count_matrix

- Accuracy: 0.8478260869565217
- Precision: 0.8452851496329757
- Recall: 0.8478260869565217
- F-1 measure: 0.8421025308241401

- Confusion Matrix:

[[9 5]

[2 30]]

Looking at the confusion matrices, we can see that both models have correctly classified more positive instances than negative instances. However, tf-idf has a higher number of true positives (31) compared to count_matrix (30). This means that tf-idf is better at correctly identifying positive instances. Additionally, count_matrix has a higher number of true negatives (9) compared to tf-idf (8), indicating that count_matrix better at correctly identifying negative instances.

We can also calculate other evaluation metrics using the confusion matrix, such as accuracy, precision, recall, and F1 score. For example, the accuracy of tf-idf is (31 + 8) / 46 = 0.8478260869565217 , while the accuracy of count_matrix is (30 + 9) / 46 = 0.8478260869565217.

This means that tf-idf and count_matrix have the same accuracy. However, we should also consider other metrics such as precision, recall, and F1 score. We find these in the evaluation matrix of both models.

In terms of precision, tf-idf has a slightly higher precision (0.8533751142446794) compared to the count_matrix (0.8452851496329757). This means that tf-idf is better at correctly identifying positive instances than count_matrix.

In terms of recall, They have the same recall value of 0.8478260869565217 . This means that tf-idf and count_matrix are equally good at identifying actual positive instances.

Finally, the F1 score is a harmonic mean of precision and recall. In this case, the F1 score for count_matrix is 0.8421025308241401, while the F1 score for tf-idf is 0.8367989918084436. This indicates that count_matrix is slightly better overall, as it has a higher F1 score.

Next we would look at the decision tree classification for both tf-idf and count_matrix:

Evaluation for Decision Tree

tf-idf

- Accuracy: 0.8478260869565217

- Precision: 0.8452851496329757
- Recall: 0.8478260869565217
- F-1 measure: 0.8421025308241401
- Confusion Matrix:

[[9 5]

[2 30]]

count_matrix

- Accuracy: 0.8695652173913043
- Precision: 0.8674339300937767
- Recall: 0.8695652173913043
- F-1 measure: 0.8665247795682578
- Confusion Matrix:

[[10 4]

[2 30]]

Looking at the confusion matrices, we can see that both models have correctly classified more positive instances than negative instances. tf-idf and count_matrix have the same number of true positives (30). This means that tf-idf and count_matrix iperform similarly at identifying positive instances. Additionally, count_matrix has a higher number of true negatives (10) compared to tf-idf (9), indicating that count_matrix better at correctly identifying negative instances.

We can also calculate other evaluation metrics using the confusion matrix, such as accuracy, precision, recall, and F1 score. For example, the accuracy of tf-idf is (30 + 9) / 46 = 0.8478260869565217 , while the accuracy of count_matrix is (30 + 10) / 46 = 0.8695652173913043.

This means that count_matrix has a higher accuracy than tf-idf. However, we should also consider other metrics such as precision, recall, and F1 score. We find these in the evaluation matrix of both models.

In terms of precision, count_matrix has a higher precision (0.8674339300937767) compared to the tf-idf (0.8478260869565217). This means that count_matrix is better at correctly identifying positive instances than tf-idf.

In terms of recall, count_matrix has a higher recall value of 0.8695652173913043 compared to tf-idf with 0.8478260869565217. This means that count_matrix is better at identifying actual positive instances than tf-idf.

Finally, the F1 score is a harmonic mean of precision and recall. In this case, the F1 score for count_matrix is 0.8665247795682578, while the F1 score for tf-idf is 0.8421025308241401. This indicates that count_matrix is slightly better overall, as it has a higher F1 score.

Overall, we see that for both classification techniques (SVM and Decision tree classifier), the count_matrix representation performs better than the tf-idf matrix.

References

Amy @GrabNGoInfo. (19 feb 2022). Four Oversampling and Under-Sampling Methods for Imbalanced Classification Using Python. Available at: https://medium.com/grabngoinfo/four-oversampling-and-under-sampling-methods-for-imbalanced-classification-using-python-7304aedf9037