IMDB Movie review sentiment analysis

<u>Interstellar Movie</u>

@Jeevan Kumar Pabbisetty

<u>Sentiment analysis on one of my favorite movies -Interstellar</u>

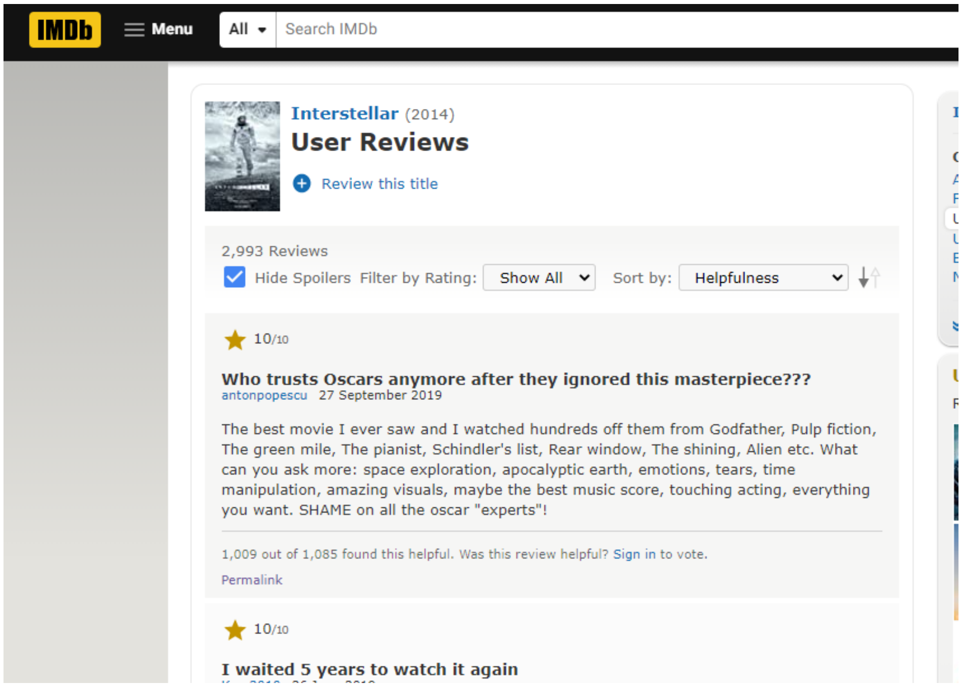
<u>Analyzing the Positive and negative reviews of the audience using the techniques of web scraping, nltk libraries and word cloud</u>

<u>Steps Followed</u>

1. Web scraping
2. Data cleansing
3. Import the required python libraries – **nltk**
4. Sentiment analysis – Positive and Negative word cloud

Python code:-

Web page of the IMDB user review for interstellar movie



Interstellar (2014)

User Reviews

2,993 Reviews

☑ Hide Spoilers Filter by Rating: Show All Sort by: Helpfulness

★ 10/10

Who trusts Oscars anymore after they ignored this masterpiece???

antonpopescu 27 September 2019

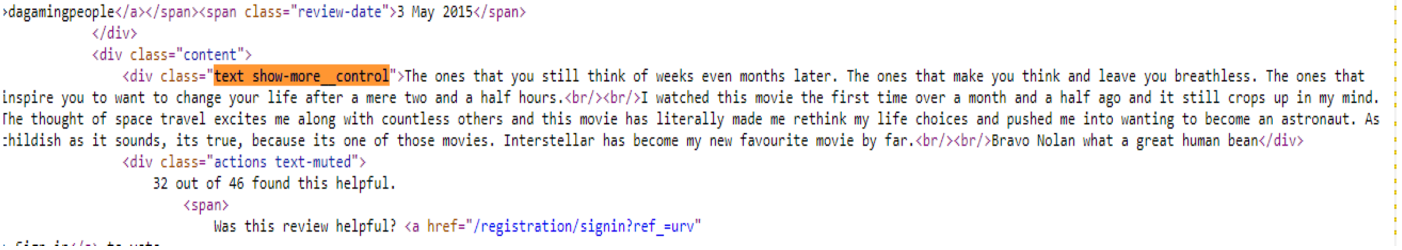
The best movie I ever saw and I watched hundreds off them from Godfather, Pulp fiction, The green mile, The pianist, Schindler's list, Rear window, The shining, Alien etc. What can you ask more: space exploration, apocalyptic earth, emotions, tears, time manipulation, amazing visuals, maybe the best music score, touching acting, everything you want. SHAME on all the oscar "experts"!

1,009 out of 1,085 found this helpful. Was this review helpful? Sign in to vote.

★ 10/10

I waited 5 years to watch it again

Web scraping using beautiful soup - “Page source”



```
>dagamingpeople</a></span><span class="review-date">3 May 2015</span>
            </div>
            <div class="content">
                <div class="text show-more__control">The ones that you still think of weeks even months later. The ones that make you think and leave you breathless. The ones that
inspire you to want to change your life after a mere two and a half hours.<br/><br/>I watched this movie the first time over a month and a half ago and it still crops up in my mind.
The thought of space travel excites me along with countless others and this movie has literally made me rethink my life choices and pushed me into wanting to become an astronaut. As
hildish as it sounds, its true, because its one of those movies. Interstellar has become my new favourite movie by far.<br/><br/>Bravo Nolan what a great human bean</div>
                <div class="actions text-muted">
                    32 out of 46 found this helpful.
                        <span>
                            Was this review helpful? <a href="/registration/signin?ref_=urv"
```

```
# -*- coding: utf-8 -*-
"""
@author: Jeevan kumar
"""
```

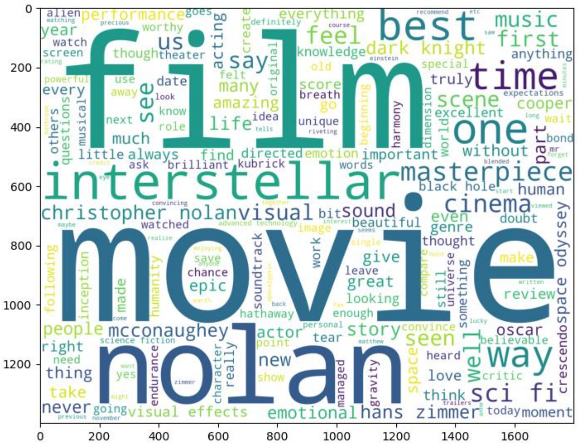
Import required libraries for scraping the reviews

```
import requests
from bs4 import BeautifulSoup as bs # web scraping
import re # regular expression

new=[]
imdb_reviews=[]
```

Extracting by html.parser and soup under class “div” & text show-more__control under this we can find the review of the user

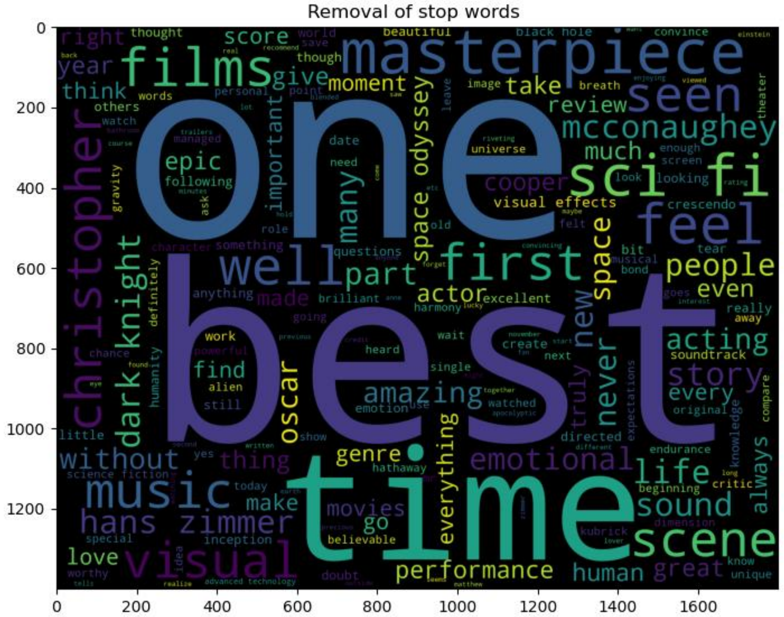
```
from wordcloud import WordCloud
#URL
url="https://www.imdb.com/title/tt0816692/reviews?ref_=tturv_ql_3"
response=requests.get(url)
soup=bs(response.content,"html.parser")
imdb_reviews=soup.find_all("div",attrs={"class","text show-more__control"})
for i in range(15):
    k=imdb_reviews[i]
    new+=k.text

Interstellar="".join(new)
```

```
import nltk

Interstellar= re.sub("[^A-Za-z" "]+"," ", Interstellar).lower()
Interstellar= re.sub("[0-9" "]+"," ", Interstellar)
Interstellar

# writing reviews in a text file
with open("interstellar review.txt","w",encoding='utf8') as output:
   output.write(str(Interstellar))

Review_split_Insterstellar=Interstellar.split()

from nltk.corpus import stopwords

#TFIDF
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer(Review_split_Insterstellar, use_idf=True,ngram_range=(1, 3))
X = vectorizer.fit_transform(Review_split_Insterstellar)

stop_words = stopwords.words('English')
Review_split_Insterstellar = [w for w in Review_split_Insterstellar if not w in stop_words]
Interstellar_RmStop=" ".join(Review_split_Insterstellar)


# WordCloud can be performed on the string inputs.
# Corpus level word cloud for ipad mini reviews
import matplotlib.pyplot as plt
wordcloud_imdb_reviews =
WordCloud(background_color='White',width=1800,height=1400).generate(Interstellar_RmStop)
plt.imshow(wordcloud_imdb_reviews)
```



Removal of stop words after building the word cloud for first time
stop_words.extend(["movie","cinema","film","say","us","interstellar","way","see","nolan"])

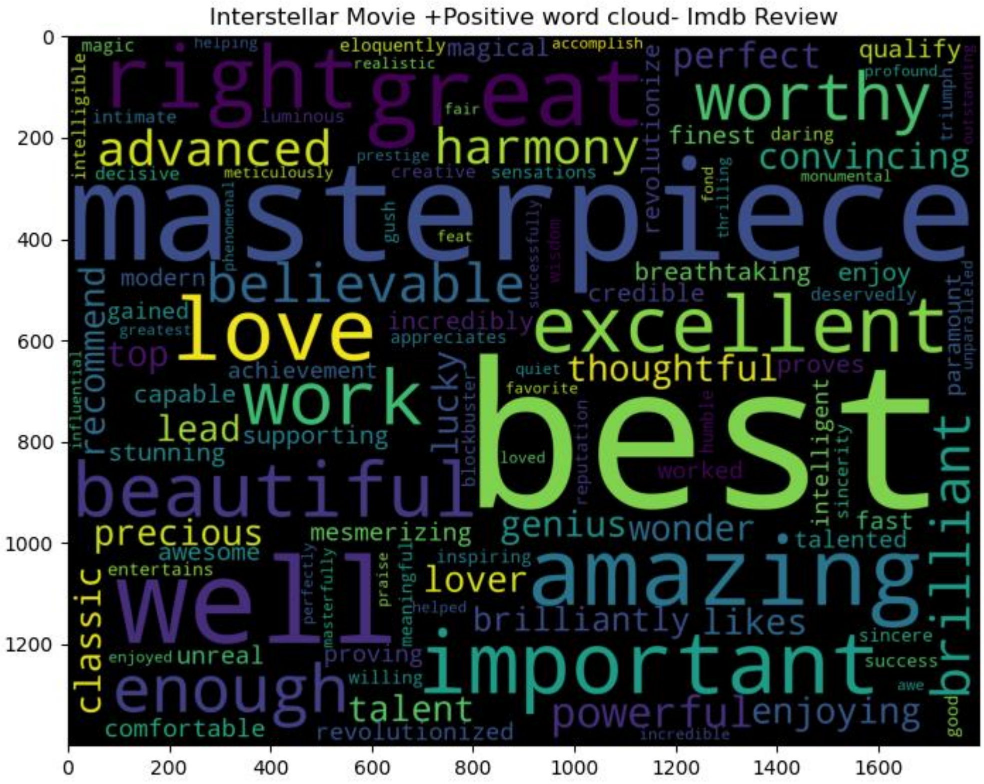
```
Review_split_Insterstellar = [w for w in Review_split_Insterstellar if not w in stop_words]
Interstellar_RmStop=" ".join(Review_split_Insterstellar)

#After Removal of stop words
wordcloud_imdb_reviews =
WordCloud(background_color='black',width=1800,height=1400).generate(Interstellar_RmStop)
plt.title("Removal of stop words ")
plt.imshow(wordcloud_imdb_reviews)
```



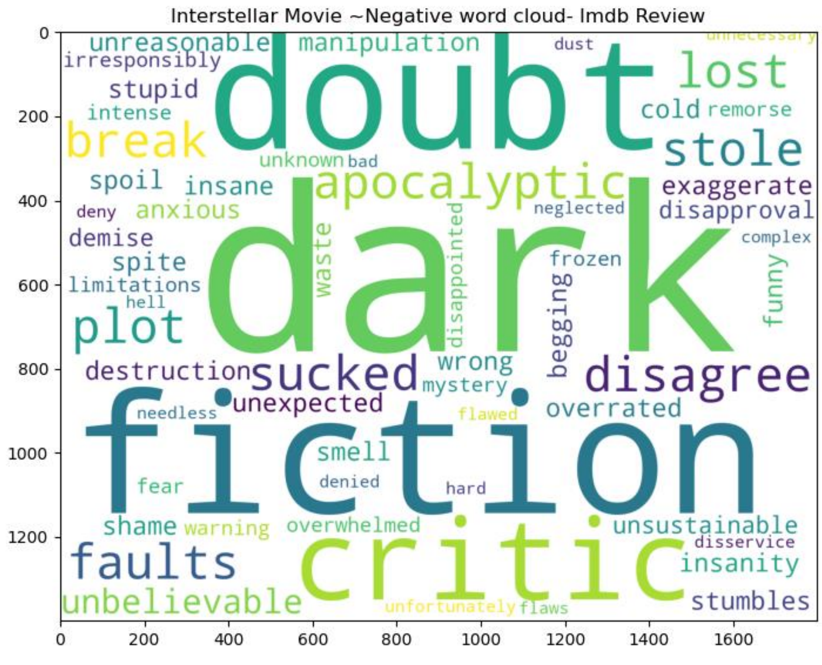
```
# Positive word cloud
# positive words # Choose the path for +ve words stored in system
with open("J:\\DataScienceAndAI\\Text_mining\\assign\\positive-words.txt","r") as pos:
  poswords = pos.read().split("\n")

# Choosing the only words which are present in positive words
imdb_in_pos = " ".join ([w for w in Review_split_Insterstellar if w in poswords])

wordcloud_imdb_pos =
WordCloud(background_color='Black',width=1800,height=1400).generate(imdb_in_pos)
plt.title("Interstellar Movie +Positive word cloud- Imdb Review")
plt.imshow(wordcloud_imdb_pos)

#Negative wordcloud
with open("J:\\DataScienceAndAI\\Text_mining\\assign\\negative-words.txt","r") as pos:
  neg_word = pos.read().split("\n")

# Choosing the only words which are present in negative words
imdb_in_neg = " ".join ([w for w in Review_split_Insterstellar if w in neg_word])

wordcloud_imdb_neg =
WordCloud(background_color='white',width=1800,height=1400).generate(imdb_in_neg)
plt.title("Interstellar Movie ~Negative word cloud- Imdb Review")
plt.imshow(wordcloud_imdb_neg)
```

Sentiment Analysis

Positive word cloud



Negative word cloud


Interstellar Movie ~Negative word cloud- Imdb Review
unreasonable
manipulation
dust
unnecessary
irresponsibly
doubt
lost
stupid
intense
cold
remorse
break
stole
unknown
bad
spoil
insane
apocalyptic
exaggerate
deny
anxious
neglected
disapproval
demise
complex
dark
waste
disappointed
frozen
spite
limitations
funny
hell
begging
plot
destruction
sucked
wrong
disagree
mystery
unexpected
overrated
needless
flawed
fiction
smell
fear
denied
hard
shame
warning
overwhelmed
unsustainable
disservice
faults
critic
insanity
unbelievable
unfortunately
flaws
stumbles