# Author Declaration for Group Assignments

Title of Assignment: Group 2 Mid-Term Essay
Module Number: CS7IS4
Word Count: 2649

March 8, 2020

| Student Number | Student Name | Nature of Contribution | % |
|---|---|---|---|
| 19309019 | Ashwin Sundareswaran R | *Accountant* of the project. Contributed to generating project ideas and did the literature review to know more about state-of-art analysis. Result analysis and contribution to the review of the essay draft, the future scope of the project. The time required for each section is recorded and divided tasks and made sure it is completed within the stipulated time. | 20% |
| 19320006 | Paritosh Chauhan | *Ambassador* of Project. Contributed to generate and finetune the project idea. Worked on the data collection part by writing web scrapper scripts and generate the data. Visualization charts are developed to analyze the experiment results. Constantly interacted with other groups to know how others went through the whole experiment process. | 20% |
| 19302026 | Sunit Deshpande | *Chair* of Project. Contributed to the algorithm development and processed the data in python to generate sentiment score. The design of the experiment and the core processing is managed. Constantly verified the whole experiment flow to check there are no discrepancies. | 20% |
| 19308029 | Manasi Mohan Narsapur | *Recorder* of Project. Contributed in developing the rough idea to feasible hypothesis and designed the flow of experiment with regards to sentiment analysis.Compiled and edited individual written work to get the final draft of the essay in the associated format. Maintains the activity log and captures minutes of meeting details | 20% |
| 18303040 | Mayur Sadashiv Mahajan | *Verifier* of Project. Contributed to extensive research of different sentiment analysis techniques to select the correct algorithm for the experiment. Worked with data pre-processing and algorithm implementation to obtain better results. Reviewed all the final draft to make sure all the parts of the parts are covered. | 20% |

declare that the percentage contribution by each member as stated above has been agreed by all members of the group, and reflects the actual contribution of the group members.

**Signed and dated:** March 8, 2020

1. Ashwin Sundareswaran R

2. Paritosh Chauhan

3. Sunit Deshpande

4. Manasi Mohan Narsapur

5. Mayur Sadashiv Mahajan

# Sentiment Analysis of Movie Reviews on Oscar Nominated Movies

Paritosh Chauhan
chauhapa@tcd.ie

Mayur Sadashiv Mahajan
mmahajan@tcd.ie

Sunit Deshpande
sdeshpan@tcd.ie

Manasi Mohan Narsapur
narsapum@tcd.ie

Ashwin Sundareswaran R
ramasuba@tcd.ie

March 2020

**Abstract**

In recent years, the entertainment industry has gained so much popularity and to judge a movie based on its reviews has become a necessity. There have been a number of methods that were implemented in order to get insights into movie reviews. In this study, we use Natural Language Processing (NLP) to perform sentiment analysis on the Oscar-nominated movie reviews posted on a website and try to find out a correlation between critic and public reviews. A quantitative value of a movie being successful or failure can be determined by the people working on the movie. On the other hand, reviews provide qualitative insights into the movies. In order to do the analysis, we have taken 10 movies into account with 10 top critic reviews and 30 of the audience reviews. We utilize the Stanford NLTK library with python to run over the scrapped data from Rotten Tomatoes about the movies who have been nominated to Oscars and understand the cumulative reactions of the audience and critic. We then juxtapose these results on a dashboard with the help of visualization tools.

*Keywords*— Stanford NLTK, Sentiment Analysis, Movie Review, Rotten Tomatoes, Text Analytics, Web Scrapping, Text Mining

## 1    Introduction

Opinions are an extremely important piece of information that helps interpret and classify emotions towards an object or a subject and assist businesses to identify customer sentiments. The Internet today contains an enormous amount of textual data that is on the rise every day and this data has a prevailing form involving a simple way to generate and publish. People generally use social media platforms such as Twitter, IMDb, Rotten Tomatoes, Amazon, Facebook and many such websites to express their opinions and feelings towards something as it is accessible and provides a degree of freedom. The challenging part in today's time is obtaining meaningful information from these platforms due to the unstructured data that is available on them. Sentiment analysis or opinion mining is a way to understand the reviews of the end-users which benefits the companies and also helps categorize the population by evaluating the link between social media, end-user input and personality by using their feedbacks or reviews.

Sentiment analysis is a type of data mining process that measures the inclination of people's opinions through natural language processing (NLP), text analysis and computational linguistics, which are used to extract and analyze subjective information from the internet mostly from social media and similar sources. Sentiment relates to the meaning of a word or sequence of words and is associated with an opinion or emotion and analysis is the process of examining the data and making inferences, such as using machine learning processing to learn and predict whether a movie review is positive or negative. Sentiment Analysis concentrates on gaining knowledge through the relationship between social media reviews and the movies thereby presenting concentrated, relevant and well-presented information regarding the popularity of the

movie among the population. Sentiment analysis is widely used in different disciplines nowadays due to its business and society values. The work reviewed here deals with the computational treatment of opinion, sentiment, and subjectivity in text and these have come to be known as sentiment analysis, opinion mining or subjectivity analysis.

In this article, we extract data from Rotten Tomatoes website on public opinion and critic opinion about Oscar-nominated movies in order to understand if both share the same kind of sentiment and visualize the sentiment results in the form of a dashboard so that it helps to understand the sentiment results towards the performance of the movie. We apply the Stanford NLTK library with python as the language of development on the movie review domain as generally people seek for movie's reviews to find more information regarding the movie as it might help them in determining whether they should watch, rent or buy it.. So, it is important to understand if the sentiments of a critic, when compared to that of the general public, are similar or differs to determine if the common culture to trust a critic's review is the best way to judge a movie and invest time into watching it.

## 2    Literature Review

There has been a lot of research based on the movie reviews and their classification using sentiment analysis recently. This includes use of machine learning techniques along with the natural language processing that helps to identify how good the movie is based on the words used in the given movie reviews. Peter.D.Turney[1] used unsupervised classification techniques to classify the reviews as recommended or not recommended. He collected a total of 410 review data for four different domains including Automobile, banks, movies and travel destinations. The classification of reviews is done by predicting the average semantic phrases in a review that contains adjectives or adverbs. The review is considered as recommended if the average score is considered positive. The accuracy percentage for the reviews of the automobile field is highest with 84% and lowest for movie reviews with 66%. This is because a good movie can have unpleasant scenes and thus because of this the average semantic score can be reduced making it not recommended. He handled it by adding an average value of Semantic orientation to correct this error in movie reviews.

Umesh Rao[2] used Naive Bayes and MaxEnt Algorithms to perform sentiment analysis on Hollywood movie data on Twitter. She collected movie review data using Twitter API for six Hollywood movies across nine countries for a period of one month to collect a corpus of one million data. She trained a Machine learning model to classify the movie into three categories (Positive, Negative and Neutral) based on the sentiment of tweets recorded in both naïve Bayes and Maxent algorithms. It has been reported that MaxEnt with Unigram word type produced the highest accuracy of 84%. The data is compared with the geographical data to check which country has the highest positive liking for a movie.

Vasu[3] used Ling pipe Sentiment Analyzer to perform sentiment analysis on twitter data to check the popularity of the movie and its performance in the box office. He collected tweets about the movie during its release to predict the sentiment of the tweets and compared it with the box office performance. Tweets from the years 2009 and 2012 related to the movie are collected using Twitter API and the data is cleaned and processed to remove the noisy data. A total of 4800 tweets comprising 200 tweets for 30 movies which were tweeted during the first four weeks after the movie release is chosen as the final data and it's split into test-train in 80-20 ratio. A classifier is trained in Ling pipe Sentiment Analyzer which labels the tweet into four categories positive, negative, neutral and irrelevant is modeled. The Positive tweet to negative tweet ratio is calculated and the movie is further predicted into three categories as hit, flop and average based on this ratio. The comparison with the predicted success labels and the actual box office success are proportional explaining the good performance of the model developed. The model

[1]Peter D. Turney. *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.* 2002. arXiv: `cs/0212032` `[cs.LG]`.

[2]U. R. Hodeghatta. "Sentiment analysis of Hollywood movies on Twitter". In: *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013).* 2013, pp. 1401–1404. DOI: `10.1145/2492517.2500290`.

[3]Vasu Jain. "Prediction of movie success using sentiment analysis of tweets". In: ().

is also used to predict the success of newly released movies and has produced an accuracy of 64%.

The authors[4] through this paper aim to perform sentiment analysis on movie reviews using twitter data. In their method, after removing the noise in data, they assigned the tags to all words and then words were divided into positive or negative based on the pre-defined vocabulary. Then they aggregated these words according to the movies and assigned the polarities to it. Results were measured using the matrices such as F-1, precision, accuracy, and recall.

In the research paper,[5] the authors combine the rotten tomatoes critic score and sentiment score obtained from Sentiwordnet and predict the movie sentiment. In this method, the data is cleansed and features are extracted before applying the algorithm. The authors use the F-measure method to evaluate this model and the results are improved because of the consideration of the implicit reviews of the critics. The authors compare their results with the studies that used only SentiWordNet for reviews or SVM for only critic reviews and finds his method most accurate.

# 3    Data Collection

## 3.1    Movie List

Movies' names are collected from the Wikipedia page that contains all the movies that have been nominated for the Oscars starting from the year 1962 to 2020. We extracted the data from the web page using Web Scraping. The page content is collected by sending HTTP requests for the URL. The source code is then parsed as an LXML format to get the contents of the table from the wiki page using the BeautifulSoup library in python. The library also provides ways to find HTML tags in the parsed code which helps in extracting the information. The collected names are stored in a Dataframe using the Pandas library in python for future use and easier handling of data.

## 3.2    Reviews

The reviews are collected from the rotten tomatoes website and follow similar steps as we did for collecting the movie names. The URLs for each movie on the list of the filtered year are then generated and the source code of the webpage is parsed using HTML PARSER in the beautiful soup. The data were extracted by matching the regular expression on the HTML tags of the parsed data and stored in data frames for further use.A total of 400 review comments were collected for the initial study for 10 Oscar winning movies between the years 2014 to 2018.The critic review corresponds to 25 percent of the overall data collected and the remaining 75 percent corresponds to public reviews.

# 4    Design of Experiment

The main research goal of this literature is to answer the question of whether there is a difference between how a professional movie critic perceives a movie and how a normal audience feels about it?

**Hypothesis**

To quantity the opinion about a movie we would measure the sentimental value of reviews given by renowned critics and a normal audiences for a known set of movies. Thus giving us the null hypothesis (H0) and the alternate hypothesis (H1) for our experiment.

**H0** : There is a positive correlation between the sentimental value for reviews given by critics and normal audience.

---

[4]A. Azizan et al. "Lexicon-Based Sentiment Analysis for Movie Review Tweets". In: *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS)*. 2019, pp. 132–136. DOI: 10.1109/AiDAS47888.2019.8970722.

[5]Suhariyanto, Ari Firmanto, and Riyanarto Sarno. "Prediction of Movie Sentiment Based on Reviews and Score on Rotten Tomatoes Using SentiWordnet". In: *2018 International Seminar on Application for Technology of Information and Communication* (2018), pp. 202–206.

**H1**: There is a negative or no correlation between the sentimental value for reviews given by critics and normal audience.

## Experiment

Many research articles have been published which show various techniques that can be used to measure the sentimental value for a given text. The most widely used technique is using semantic vector space for every single word in the text and then calculating the overall score of the text.[6] This technique works well when we have access to a large set of words related to a single topic like an article about a movie. But for a small amount of text like the reviews that people post on movie reviews websites like rotten tomatoes this technique is not that effective. For example, a review like the below

*I did not feel that the movie was amazing or outstanding.*

would receive a positive score using semantic vector space even if the actual review was negative.

To mitigate this, we would have to use a sentimental analysis technique that would capture the overall context of the sentence. For the same reason, we have used the "Recursive Deep Model" implementation provided by Stanford University.[7]

The overall flow for calculating the sentimental value for a given movie review is shown in the diagram below.
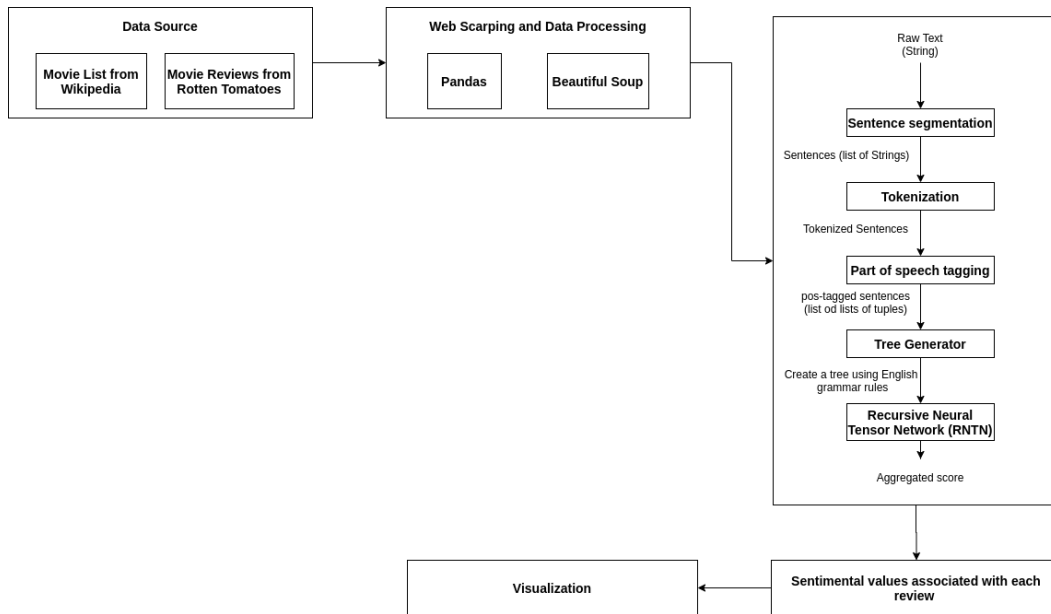


Figure 1: Experiment Flow

[6]P. D. Turney and P. Pantel. "From Frequency to Meaning: Vector Space Models of Semantics". In: *Journal of Artificial Intelligence Research* 37 (2010), 141–188. ISSN: 1076-9757. DOI: 10.1613/jair.2934. URL: http://dx.doi.org/10.1613/jair.2934.

[7]Richard Socher et al. "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, Oct. 2013, pp. 1631–1642. URL: https://www.aclweb.org/anthology/D13-1170.

Movie review goes through a series of phases before calculating the sentimental value. In the first phase, the movie review text goes through the **Sentence Segmentation** process where the entire text is divided into different sentences based on the English punctuation symbols like period or new line present in the text. The outcome of this phase is a list of separate sentences. This array of sentences are passed as the input to the next phase of the flow which is **Tokenization**. The task of the tokenization phase is to remove the unwanted symbols and white spaces from the sentences and transforms it into a steady stream of tokens. The tokens produced from this process is a data structure which holds the value and the type of the entity. For example if a sentence is "I would rate it 5." The output of the tokenizing phase would be a list of tokens such as $\big[\langle WORD, I\rangle, \langle WORD, would\rangle, \langle WORD, rate\rangle, \langle WORD, it\rangle, \langle Number, 5\rangle, \langle EOS\rangle\big]$.

This stream of tokens is then fed to the **Standford parts of speech tagger**. The parts of speech tagger tags each token with associated parts of speech for that token. The outcome of this phase is a tokenized stream with each token having the associated parts of speech tagged to it. These modified tokens are given as input to the next module **Tree Generator**. Tree generator module is responsible for generating a parse tree. This uses the English grammar rules as a way of defining the structure of the tree. A parse tree captures the semantic structure of the sentence which is later used to calculate the cumulative sentiment value of the sentence.

Each parse tree is given as input to the **Recursive Neural Tensor Network (RNTN)** model.[8] RNTN is a Recursive Neural Network model trained on the Stanford Sentiment Treebank dataset. Stanford sentiment treebank dataset contains a collection of sentences with each sentence parsed as a tree using the English grammar rule and every node of the tree having a sentimental value associated with it. RNTN uses a pre-trained model to classify a given parse tree into five of the following categories Very Negative, Negative, Neutral, Positive and Very Positive.

Using the above flow we were able to classify the sentiment values for a set of movies reviews that were scrapped from the rotten tomatoes website as mentioned in the Data Collection section. The sentimental value for each individual public and critic review were captured as the output. These outputs are classified into one of the categories of sentiment labels as mentioned above. A correlation analysis was performed on the public and the critic reviews associated with the same movie.

# 5    Result

The graph below shows how the reviews from both critics and the public were classified into different labels based on the Sentiment Analysis performed for six selected movies. We can infer from the graph that "Arrival" had around 70 percent negative reviews from the public whereas it had received only 10 percent of negative reviews from critics. As per the sample data collected, the movie "The Big Short" had 30 percent negative reviews among the audience but it had not received any negative reviews from critics. These two movies illustrate that not all Oscar awarded movies are liked by the general audience as it is referred by critics. Whereas, the movies "The Big Short" and "12 Years a Slave" received more neutral opinions from the public although it had been awarded Oscars and received good responses from critics. The movies "Birdman" and "The Imitation Game" had a similar kind of liking trends between critics and the populous.

---

[8]Socher et al., "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank".
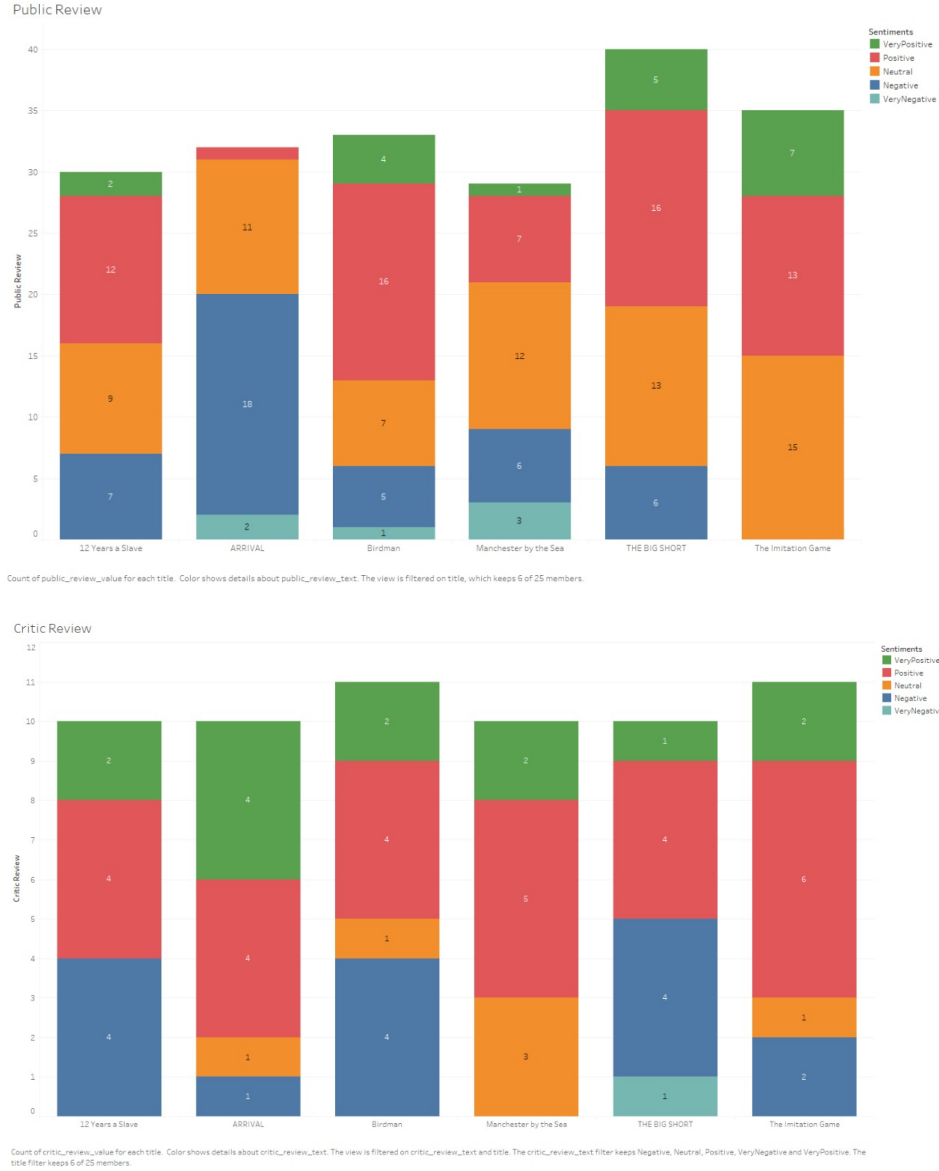
Figure 2: Sentiment score distribution for critic and public movie reviews

# 6 Conclusion

The preliminary result according to the sample data collected suggests that for every 2 out of 3 movies there was a negative correlation between the aggregated sentimental value of critics and the general public reviews. Experiments on a small set of data support the alternate hypothesis but for the result to be statistically significant, experiments need to be performed on larger datasets.

# 7 Future Works

As the conclusion suggests, we need to run our experiment with more data to check our hypothesis. Some of the other considerations that we might take in the future are - Checking if there is any bandwagon effect that arises in the audience's review i.e. checking the amount and quality of reviews after a movie has been nominated for the award. Adding more out of text properties like reviews being biased on the casts, music, directors or production house that are part of the movie.