International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 8 Issue 03 Series. III || March 2019 || PP 15-17

Sentiment Analysis of Movies Dataset using Python

Ms. Sushmita Roy
Department of Information Technology, Thakur College of Science and Commerce, India

Abstract : *Word clouds have currently evolved as a visually appealing visualization method for representation of text. Also it becomes easy to grasp up the idea using this means of text representation. It is used in various contexts as a means to provide an overall idea by filtering the text to those words that have highest frequency. This is typically done in a static way. There is however a broader scope - simple yet powerful paradigm of visualization in analytics of text. In this paper, we understand the word clouds built and its significance in easy visualization of sentiments of the people which can be used to solve text anaysis problems and evaluate it easily. There is a wide range of information available in the Internet resulting in problems of information overload. Similarly, the dataset under consideration has data of about 45,000 movies released on or before July 2017. Hence a text analysis will be beneficial as users can visualize the data easily. This will be an effective means of data analysis.*

Keywords –*sentiment analysis, movies dataset,wordcloud.*

Date of Submission: 20-02-2019 Date of acceptance: 05-03-2019

I. Introduction

As per general language usage, sentiment means a view or opinion about a particular point. However, it could also mean an emotion.For example, a person might like "tragic" movies because he/she enjoys the sentiment of "sadness".

Sentiment analysis is a data mining technique which measures the inclination of opinion of the people through the processing of natural language (Machine learning), calculation linguistics and analysis of text, which is used to extract and analyze subjective information from the Web - primarily social media and similar sources. The data analyzed quantify the feelings or reactions of the general public towards certain products, persons or thoughts and discover the contextual polarity of the information.

Sentiment analysis is also known popularly as opinion mining.

Analysis of sentiments uses data mining processes and techniques to extract and capture data for analysis to discern the subjective opinion of a document or collection of documents, such as blogs, posts, public reviews, news articles and social media feeds such as tweets and status updates.

Analysis of feelings enables organization to keep track of the following:

- Reception and popularity of brands
- Perception and anticipation of new products
- Company reputation
- Detection of flames and rants

II. Experimental Results

For this study, we have considered a secondary data source relating to the movies data set for movies released on or before July 2017. The files contain some metadata about the movies. Two of the metadata considered here are – the title and the overview. These have been considered to build the word clouds.

Often there are some words in Movie Titles which are considered more often or are used more. These could be because some words are considered more powerful and considered worthy of a title.

Fig. 1. demonstrates this using the title metadata.

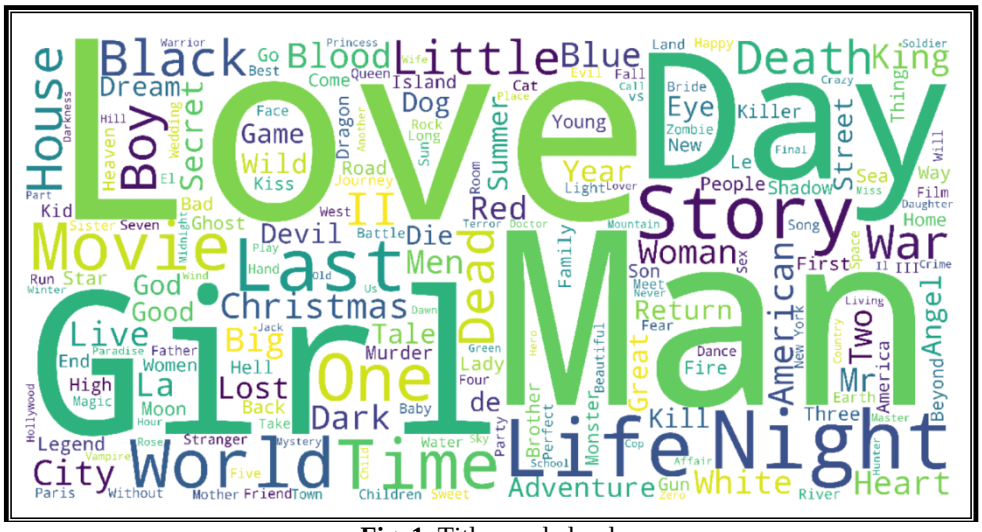

Fig. 1. Title word cloud

This demonstrates that the word "Love" is the single most frequently used word amongst the movie titles. Apart from this, words - "Man", "Girl" and "Day" are also frequently used words. Words like – "Night", "World", "Life", etc. are the next most frequently used words.Fig. 2. demonstrates this using the overview metadata.

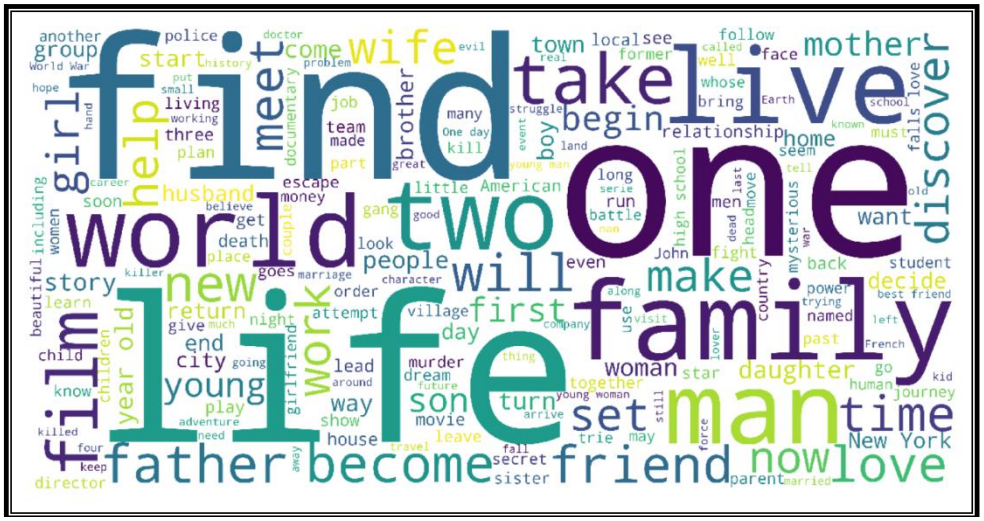

Fig. 2. Overview over cloud

This demonstrates that the word "life" is the single most frequently used word amongst the movie titles. Apart from this, words - "find" and "one" are also frequently used words. Words like – "world", "live", "family", etc. are the next most frequently used words.

III. Conclusion

Sentiment analysis using word clouds is an effective way of analyzing the opinions of the users. Using word clouds, it is easier to understand the most occurring words in the title, overview and likewise cases.

IV. Future Enhancement

Apart from these, we can calculate the frequency of each word and display it in tabular form.

We can use different colors to highlight the positive, negative and neutral sentiments. For example, green can be used to highlight a positive sentiment, red can be used to highlight a negative sentiment and black can be used to highlight a neutral sentiment.

We can also show the count for each of the words in the dataset showing the highest frequency words or words that are more frequently used.

Further, data can be gathered from Twitter regarding reviews of users on the movies in this dataset. Then sentiment analysis can be further performed on those data to understand the views of the users.

V. Acknowledgements

I would like to thank my college, the Thakur College of Science and Commerce (TCSC), for giving me the opportunity to carry out the project work.
I would like to thank Principal Dr. (Mrs.) C.T. Chakraborty for her kind, inspiring and encouraging guidance.
I am very grateful to the IT Department and Dr Santosh Kumar Sing, Head of the Information Technology Department, for their kind cooperation in completing this project.
Special thanks to my project guides, Dr. (Mr.) Santosh Kumar Singh and Prof. Mahendra Sharma, who gave me the opportunity to do this, which helped me to do research, and I learned a lot.

References

Journal Papers:

[1]. F. Heimerl, S. Lohmann, S. Lange and T. Ertl, 47th Hawaii International Conference on System Sciences, Waikoloa, HI,Word Cloud Explorer: Text Analytics Based on Word Clouds, 18(2), 2014, pp. 1833-1842. doi: 10.1109/HICSS.2014.231.

[2]. Sasikala P, L. Mary Immaculate Sheela, International Journal of Applied Engineering Research, Sentiment Analysis and Prediction of Online Reviews with Empty Ratings, ISSN 0973-4562 Volume 13, Number 14, 2018, pp. 11525-11531.

Websites:

[3]. https://www.techopedia.com/definition/29695/sentiment-analysis

[4]. https://www.kaggle.com

[5]. https://www.wikipedia.com

Ms. Sushmita Roy" Sentiment Analysis of Movies Dataset using Python" International Journal of Engineering Science Invention (IJESI), Vol. 08, No. 03, 2019, PP15-17