Youtube Video Sentiment (Popularity) Analysis using TextBlob

Md. Mofaqkhayrul Islam Mim
Dept. of CSE
BRAC University
Dhaka, Bangladesh
mofaqkhayrul.islam@g.bracu.ac.bd

Jahid Hasan Mamun
Dept. of CSE
BRAC University
Dhaka, Bangladesh
jahid.hasan.mamun@g.bracu.ac.bd

Md Tarikul Islam
Dept. of CSE
BRAC University
Dhaka, Bangladesh
md.tarikul.islam@g.bracu.ac.bd

S M Ibrahim
Dept. of CSE
BRAC University
Dhaka, Bangladesh
sm.ibrahim@g.bracu.ac.bd

Aryan Arnav
Dept. of CSE
BRAC University
Dhaka, Bangladesh
aryanarnav71@gmail.com

Sadaf Noor
Dept. of CSE
BRAC University
Dhaka, Bangladesh
md.sadaf.noor@g.bracu.ac.bd

Mushfiqur Rahman
Dept. of CSE
BRAC University
Dhaka, Bangladesh
s.m.mushfiqur.rahman@g.bracu.ac.bd

Shihab Sharar
Dept. of CSE
BRAC University
Dhaka, Bangladesh
shihab.sharar@g.bracu.ac.bd

Annajiat Alim Rasel
Dept. of CSE
BRAC University
Dhaka, Bangladesh
annajiat@gmail.com

***Abstract*—Sentiment analysis is a combination of studies related to attitude, emotions, opinions, feelings, and sentiment itself. Social media platforms have been the most used form of giving opinions and practicing freedom of speech. Youtube has been one of the comprehensive sources of video information till 2005. 2.3 billion active users were spotted last year. Five hundred hours of video are being uploaded every minute. Hence, Youtube comment sections have become one of the platforms for expressing and giving opinions. Though the like-dislike feature can show the sentiment of people's view on the video, it does not necessarily always give the true context. This feature usually does not work on satire and news channels. Sentiment analysis can present a broad view of the comments people post under the comment section of the YouTube videos. Though there are several methods such as dictionary-based approach, rule-based approach, Naive Bayes analyzer, pattern analyzer, VADER packages for text sentiment analysis, we have used the TextBlob packages of NLTK as it the one the most convenient pre-modeled analyzer with high accuracy. In this paper, we presented the analysis on the sentiment of comments from multiple videos on YouTube. We analyzed how the users were feeling based on: negative, positive, or neutral emotions from the comment texts. The effectiveness of this proposed method will help to analyze youtube comments' relevancy and produce high-quality videos based on comprehensive feedback.**

***Index Terms*—Sentiment Analysis; TextBlob; Natural Language Processing, NLTK, Youtube comment.**

I. Introduction

In this 21st century, the Internet is high-speed and very cheap worldwide. It already reached almost every corner of the world, including space. Moreover, the more Internet is in use, people now like to watch videos, news updates everything online. From all the available video platforms, YouTube is one of the most popular services right now worldwide.

In the beginning, YouTube was developed as an entertainment platform, it was designed to host videos. So most of the content of YouTube was just general videos mainly developed for entertainment by unprofessional. However, as the Internet got popular and got more acceptance, more professionals got involved here. Currently, we can see lots of educational content, courses, tutorials and essential information on YouTube. So it is not just for entertainment anymore. People now search for the important topic on YouTube like they used to do on search engines.

However, as YouTube popularity grows, the number of content also growing very fast. Furthermore, along with that, the quality of the content is decreased. That is because there are now more creators, and not all of them are professional. Another big problem is that lots of fake information's are available there. If we analyse the available content, we can see lots of content available on the same topic. Those contents quality is not the same, and may not be genuine info too. So in this work, we tried to analyse the YouTube video quality based on the user-provided comments.

YouTube also has other services like YouTube Go, YouTube Kids, YouTube movies, YouTube Music, YouTube Premium, YouTube short, and YouTube stories and YouTube TV. There is also a very well-maintained YouTube community. In future, we may implement a similar model to analyse those content too.

II. Literature Review

In this paper [1], the threshold levels are observed and monitored to classify the viewpoints. The thresholding hash table is used to display the findings. The proposed model is effective on the youtube comments using the TextBlob python library, which offers a simple API to classify the sentiment of each comment into positive(1) or negative(-1), or neutral(0). In the proposed model [2], The total number of comments is a specific size. Although there was no set format for the comments, there was a limit on how long they may be. The remarks were gathered from an online source and are not labeled with specific sentiments. The primary goal of this research[102] is to determine the tone of social media comments.The method of opinion mining and sentiment analysis has progressed to a superior stage in recent years. The opinions are categorized into so many sentiments based on the polarity of the statement. Currently available strategies mostly target tweets and text documents across a variety of sites.Emotional types play an equal role in determining feelings, according to the recommended technique. This research employs text and emotions to boost sentiment on Twitter [4]. In this study [5], the author used emojis and texts to determine whether the writer's sentiment was positive, negative, or neutral. Emojis are said to have a significant function in writing in identifying the writer's emotion or feeling. To get an appropriate score, it's crucial to evaluate emojis in the statement. It should be noted that emoticons differ from one platform to the other. Users also write English comments in Arabic letters; these words were not identified and received an emotion score of zero (neutral). Twitter has the maximum opinions of the user [6]. Based on the emotions behind the tweets, Twitter data can be classed as good, negative, or neutral. To extract the emotions behind the utterance, the proposed system employs a variety of classifiers. A dataset of texts is used by the classifiers. The graph is used to evaluate and depict the classifiers' performance. The proposed study also explains the importance of emotion in a specific piece of information. This paper [7] It discusses sentiment classification approaches as well as additional resources for using sentiment analysis in Indian languages. On many social media platforms, the use of Indian languages is quickly increasing. It is critical to developing resources for Indian languages with limited resources. Machine learning, lexicon-based, and hybrid algorithms are utilized for sentiment classification. Different machine learning techniques, such as a decision tree and a lexicon-based approach, are used to validate the results. In [8] this paper, they employed a variety of machine learning algorithms to analyze the sentiment of product evaluations posted on Twitter. In order to get better results, they used the Naive Bayes approach. To strengthen the identification process and obtain a high-quality summary of the reviews, increase the training data. Kannada reviews can be found on a few websites. They've applied Machine Learning methodologies in this proposed effort. The algorithm is written in a way that those who can't read English can understand it in their own language [9]. On [10] the internet, there is a great volume of data that needs to be sorted into different attitudes. This data may not just be in English, but also in other regional languages. In this study, they present a method for determining the communicator's sentiment by analyzing the polarity of text or tweets extracted from social media or other sources. Because people make decisions based on the opinions of others, this research prioritizes large amounts of data over short oral communications. In paper [11] an algorithmic lexicon is crucial to the machine learning technique because it is where the data for the specific implementation is stored. In a language, every word has a meaning. The lexicons that have already been saved can be used to extract information about discrete words. To produce linguistic data, the suggested system is trained with a bag of 2400 Malayalam words. Precision, recall, and f-measure are statistical techniques used to calculate the presentation. More than 90In [12] tweets are sorted into different sentiments based on the message level categorization in this work. The approach of pre-processing removes these halting words, resulting in a meaningful phrase. By deleting nonEnglish tweets, the tokenization algorithm replaces the emotions. The words are categorised as positive, negative, or neutral based on their frequency in the training set. Based on previously-stored lexicons, the words are assigned sentiment scores. The overall sentiment score of the tweet is calculated and divided into different sentiment categories. In [13] this paper the sentimental analysis is performed by using text blob which is a library in python. Three platforms, namely Facebook, Twitter, and news websites, are employed to examine these sentiments. ANN is used to classify the tweets (Artificial Neural Network). The Twitter API is used to collect the tweets, which are then classified using the Naive Bayes algorithm. R programming was used to analyze the results. The accuracy ranges from 70

III. Methodology

In this study, we use TextBlob for sentiment analysis on a youtube comments dataset.

A. Sentiment Analysis

The method of determining the writer's attitude or feeling, whether positive, negative, or neutral, is known as sentiment analysis. It is the computational analysis of people's written expressions of ideas, sentiments, attitudes, and emotions. [14]

B. TextBlob

TextBlob is a Python package with a simple API for calling its methods and doing Natural Language Processing (NLP) applications. The characteristics of polarity and subjectivity are returned by textblob. The polarity is a float that lies in the range [-1,1], with 1 denoting a positive statement and -1 denoting a negative statement. Personal opinion, emotion, or judgment are examples of subjective sentences, whereas objective sentences refer to facts. Subjectivity is also a float with a value between 0 and 1. The flowchart illustrated in Fig. 1 summarizes the workflow we used for this study. It will be explained in detail throughout the rest of this paper

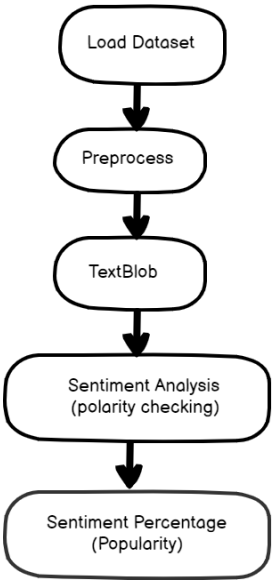

Fig. 1. Flow Chart

C. Dataset

We have used the **Sentiment Analysis of YouTube Comments** *UScomments.csv* where dataset shape is (691400, 4) available on **Kaggle** [15]. This data set provides likes, replies and comments of a specific video. The information provided in the dataset are as follows:

- $VideoID$ - unique identity of a specific video
- $CommentText$ - users comment on specific video
- $Likes$ - users likes on specific video
- $Replies$ - users replies on specific video

We had to append a column for polarity that contained polarity values such as '1.0', '-1.0', and '0.0',

IV. Results And Discussion

In this study, at first, we have measured the sentiment of a sample text with TextBlob. We have added an extra column that contained polarity values where '1.0' denotes the positive, '-1.0' denotes the negative, and '0.0' denotes the neutral. We performed sentiment analysis on the *UScomments.csv* dataset and visualized the result using the matplotlib library. The result is reported the dataset is shown in Table I and visualized on Fig. 2. Here,

- *Positive* identifies the positive polarity and also the popularity of the video based on users comments.
- *Neutral* identifies the neutral polarity of the video based on users comments.
- *Negative* identifies the positive polarity and also the popularity of the video based on users comments.

From Table I, we can see that TextBlob performs relatively well with the dataset. Now let us try to sample the dataset where the dataset shape is (9000, 4) and perform sentiment analysis on the *UScomments.csv* dataset and visualized the result using the matplotlib library again. The corresponding

TABLE I
Sentiment Analysis Report

	Positive	Neutral	Negative
Polarity	300796	284450	106154

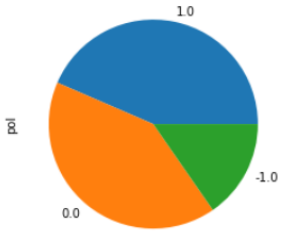

Fig. 2. Result

results are summarized in Table II and visualized Fig. 3 for more effective viewing.

TABLE II
Sentiment Analysis Report after sampling

	Positive	Neutral	Negative
Polarity	3887	3774	1339

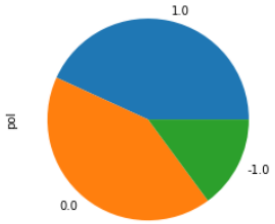

Fig. 3. Result after sampling

Let us discuss what the change is and how it may affect TextBlob. The significant change is that the sampling dataset performs faster than the original dataset. This is perhaps because of the volume of the dataset. Table II and Fig. 3 also hint that the choice of users and popularity of a video. We can add that the choice or popularity of a youtube video often relies on its user's comments.

V. Conclusion

In this paper, we have demonstrated how one may analyse the sentiment using TextBlob. Following this investigation, we concluded that the sentiment expressed in YouTube comments is more important than the like-dislike ratio. The sentiment analysis of Youtube comments provided a rich context and better understood how consumers felt about specific videos. For future reference, other packages can be explored, such as VADER. Different custom classifiers could be implemented to compare them against TextBlob as a sentiment analyser. A

study on sentiment's effect on YouTube algorithm can be done for future works.

REFERENCES

[1] T. S. Raghavendra and K. G. Mohan, "Web mining and minimization framework design on sentimental analysis for social tweets using machine learning," Procedia Comput. Sci., vol. 152, pp. 230–235, 2019, doi: 10.1016/j.procs.2019.05.047.

[2] A. Sharma and U. Ghose, "Sentimental Analysis of Twitter Data with respect to General Elections in India," Procedia Comput. Sci., vol. 173, no. 2019, pp. 325–334, 2020, doi: 10.1016/j.procs.2020.06.038.

[3] A. I. Al-Ghadir, A. M. Azmi, and A. Hussain, "A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments," Inf. Fusion, vol. 67, no. December 2019, pp. 29–40, 2021, doi: 10.1016/j.inffus.2020.10.003.

[4] N. K. Khokale, "AND ENGINEERING TRENDS STUDY OF SENTIMENT ANALYSIS OF TWITTER USERS FROM THEIR TWEET ' S TEXT AND DIFFUSION PATTERNS," vol. 6, no. 5, pp. 192–195, 2021.

[5] A. Alrumaih, A. Al-Sabbagh, R. Alsabah, H. Kharrufa, and J. Baldwin, "Sentiment analysis of comments in social media," Int. J. Electr. Comput. Eng., vol. 10, no. 6, pp. 5917–5922, 2020, doi: 10.11591/ijece.v10i6.pp5917-5922.

[6] S. Geetha and V. K. Kaliappan, Tweet Analysis Based on Distinct Opinion of Social Media Users'. Springer Singapore, 2018.

[7] M. B. Shelke and S. N. Deshmukh, "Recent Advances in Sentiment Analysis of Indian Languages," vol. 13, no. 4, pp. 1656– 1675, 2020.

[8] N. Chintalapudi, G. Battineni, M. Di Canio, G. G. Sagaro, and F. Amenta, "Text mining with sentiment analysis on seafarers' medical documents," Int. J. Inf. Manag. Data Insights, vol. 1, no. 1, p. 100005, 2021, doi: 10.1016/j.jjimei.2020.100005.

[9] K. M. A. Kumar, N. Rajasimha, M. Reddy, A. Rajanarayana, and K. Nadgir, "Analysis of users' Sentiments from Kannada Web Documents," in Procedia Computer Science, 2015, vol. 54, pp. 247–256, doi: 10.1016/j.procs.2015.06.029.

[10] P. Pandey and S. Govilkar, "A Survey of Sentiment Classification Techniques Used for Indian Regional Languages," Int. J. Comput. Sci. Appl., vol. 5, no. 2, pp. 13–26, Apr. 2015, doi: 10.5121/ijcsa.2015.5202.

[11] A. S. C. Chung, "B Ehavior B Ased a Pproach for I Ntrusion," a Simul. Based Approach Conting. Plan. Aircr. Turnaround Oper. Syst. Act. Airl. Hubs, vol. 11, no. 1, pp. 1–6, 2013.

[12] A. Dalmia, M. Gupta, and V. Varma, "IIIT-H at SemEval 2015: Twitter Sentiment Analysis – The Good, the Bad and the Neutral!," no. SemEval, pp. 520–526, 2015, doi: 10.18653/v1/s15-2087.

[13] S. Paliwal, S. Kumar Khatri, and M. Sharma, "Sentiment Analysis and Prediction Using Neural Networks," Proc. Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2018, no. Icirca, pp. 1035–1042, 2018, doi: 10.1109/ICIRCA.2018.8597358.

[14] Liu, B.: 'Sentiment analysis and opinion mining', Synthesis lectures on human language technologies, 2012, 5, (1), pp. 1-167.

[15] https://www.kaggle.com/adepvenugopal/sentiment-analysis-of-youtube-comments/data?select=UScomments.csv.