Extracting Sentiments from YouTube Comments

Rahul Pradhan
Department of CEA,
GLA University,
Mathura, UP, India
rahul.pradhan@gla.ac.in

2021 Sixth International Conference on Image Information Processing (ICIIP) | 978-1-6654-3361-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIIP53038.2021.9702561

***Abstract*—YouTube is the most used social media platform, and it has been the most popular website where users can post the video. The public generally does comment, like or dislike, video-sharing on a YouTube video. Comment plays a vital role in expressing opinions and mindset, and it is used as an expression of public opinion. The massive amount of comment comes mainly on famous channels where challenges arise to analyze public opinion or behavior regarding that particular video. This article proposes sentiment analysis on YouTube video by Natural Language Processing (NLP) technique. Sentiment Analysis is when comprehension, citation, and processing of text-based data is done, and it directly converts it into sentiment information. This analysis help users to get the report of their YouTube Video. The output of this analysis gives the classification of sentiment analysis, i.e., positive, negative, or neutral.**

***Index Terms*—youtube video, sentiment analysis, youtube comment, healthy eating, user comment, youtube data api, vader lexicon, natural language processing technique.**

I. Introduction

Today, social media like Facebook, Twitter, YouTube, and Google significantly impact millions of users to communicate to share information or opinions about all. With the sudden increase in popularity of these sites, they have become a prime source of many real-time videos, text, images, etc.[6] Among them, YouTube is the most popular common social media platform where users can upload a video. Around a million users can comment, share and rate videos. YouTube is the platform where community interaction can be done. YouTube generates millions of metadata (unstructured data) in video content and gains popularity by sharing the videos day by day[1].

Making a YouTube video interesting allows many users to give their feedback through like dislike and sharing comments. The analysis will help a user know the sentiment of viewers regarding their videos[2]. The investigation results say that the comments are classified into two ways: the first is positive, and the second is negative, which helps the user access the video.

We have used Vader lexicon, and YouTube data API, which will tell about sentiments of comments that the user gives on the particular video [7]. By using Sentiment Analysis, the Natural Language Processing techniques, which can be used to determine the text's sensibility [4].

This paper will collect video reviews using YouTube services and perform experiments on video comments. Many big companies are using it as a platform to launch their products. It's in use on a large scale to examine customer reviews about their product/services, on YouTube, their websites, which helps them maintain their brand values[3].

By this paper, it is easy for the audience to analyze their product with customer feedback. And we will be able to find out what other people think about that particular video.

A. YouTube: History & Present

YouTube is a social media platform owned by Google, this platforms targets the content created in form of video and distributed online through real time streaming, its was launched in 2005 and become a immediate hit, it is known to be second most visited website in world. YouTube has videos from various categories such as music video, news, short films, vlogs, documentaries, movie trailers, teasers, live events and many more. It was founded by three employees of PayPal a payment gateway service, Chad Hurley, Steve Chen and Jawed Karim.

B. YouTube Video Formats

Initially in 2005 when YouTube was launched, it only support video format H.263 with mp3 as audio format, with time it start supporting other formats too. As advent of internet to mobile phones it start supporting popular mobile phone formats such as 3GP. By 2008, YouTube start supporting 720p video quality, which was very clearer for that point of time, even now many streaming websites when analysed data found that this level of resolution was best their audience prefer unusually.

1) Story of YouTube: YouTube was developed because of necessity and this the founder trio had a bet, founders Hurley and Chen went to a dinner party and they talk highly about this party to Karim and telling how good that party was and how much fun they had, what all activities they did over their and many other things. Karim was in state of denial and keep denying their claims of having good time, So they try to share the video they have shot at that dinner party as an evidence of their good time. But irony that they found it very difficult to share it. 4K was also introduce to YouTube in 2010.

2) YouTube Services: YouTube has various services such as YouTube Music that deals with music their distribution and other things, this service is important as many rely on YouTube for listening music and Ads are not suppose to come in between music video. (see Figure 1) for glimpse of YouTube Music page and its suggestions. Another popular YouTube service is Premium it provides Ad free content. This is an

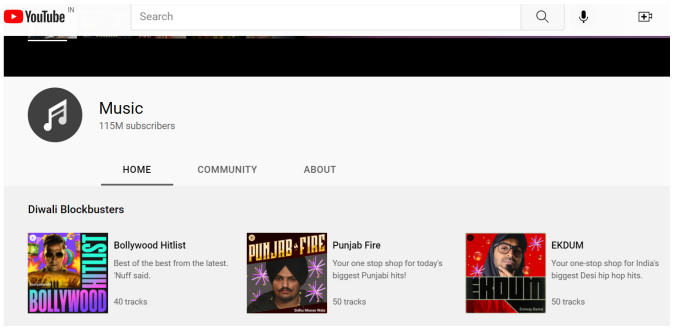

Fig. 1. YouTube Music page in India

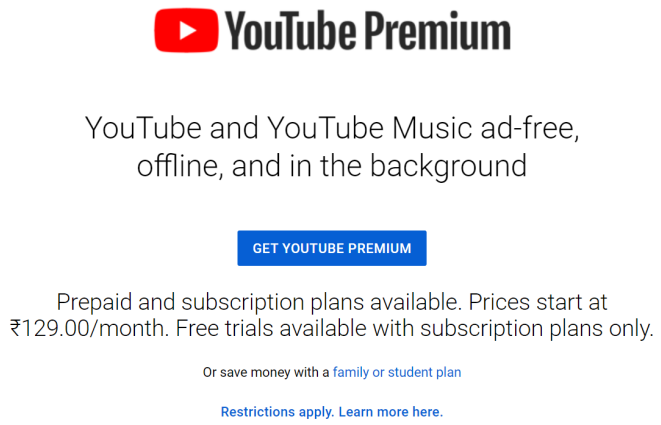

Fig. 2. YouTube Premium and its cost in India

paid service from YouTube, content is exclusive and originally created through support from various creators.

Figure 2 shows the sign up page for YouTube premium and its cost per month in India in Indian national Rupee.

II. Related Work

On YouTube video features, there is much research that has been undertaken. And to make an important decision about the particular video, comments play an important role. The comments reflect users' behavior, and troll users can be found by using comments. We will find the user's positive or negative sentiments about the particular video by analyzing the comments[6].

The videos categorize into several categories by using comments. Lehner et al. [9] work on likes, dislikes, and comments of the YouTube Video to show the user's feedback (like/dislike) influenced by the valuable comments.

"Helpfulness" is the condition that is termed by Amazon's notion where users can rate a review and predict whether it is valuable or not. Comments from eBay company which give rated quality aspects used by Lu et al.[5] uses latent topic approach. He rates the products based on star rating used to predict the quality of posts in the software online forum Nabble.com.

Teng et al.[12] analyze the YouTube comments to get the sentiments of healthy eating habits in the communities. In their analysis, he used text mining and predictive analysis on habits and behavior of healthy eating. The study's main objective is to determine and motivate healthy eating behavior in online communities and highlight the awareness into YouTube video commenters' perception and sentiments of healthy eating through text mining, descriptive and predictive analysis.

A fascinating study has been studied where ratings given by users can be analyzed and conducted by Siersdorfer et. al.[8] They investigate that positive and negative ratings are very much related to each other. Our study is different from their primary motive. However, user rating is not included in our data set because we couldn't get data collection API using the Google developer key. An Analysis of comment and comment rating is represented by Siersdorfer et. al.[8] in 2010 showing the acceptance or comment by the community. Siersdorfer et al. focused on a classification model by which user comments are analyzed, and he performed an analysis between comments on sentiments.

Hsu et. al.[10] prepares an approach to give ranking user comments nevertheless the content in which they appear. They show a regression-based method for automatic quality assessment of user comments depending on how the community accepts. The results of their analysis are mixed with the quality of the statement, which gives better classification. In our paper, we found particular a video that the YouTube users could use, the comments in large amount are analyzed instead of features such as like/views, etc.

III. Proposed Approach

We analyze the comment on YouTube video using sentiment analysis which is the Natural Language Processing Techniques. And to perform this experiment, we first extract the comments of the video uploaded on YouTube[3,12]. Now the comments which we have extracted may contain noise, irrelevant symbols that can affect the analysis of the comments (see fig 3).

After that, perform the sentiment analysis of the comments[6]. And for carrying out this experiment, we have to install libraries such as vaderSentiment httplib2, google-API-python-client, CSV, oauth2client, etc., using pip[11].

Prerequisite for the following are:

- Google Account
- Python 3
- Jupyter Notebook

IV. Experimental Setup

A. Extract the Comments

The extraction of YouTube data is done using YouTube Data API v3. For this, from the google console, develop the YouTube API and install the required libraries. By following this, we get information like channel title, video title, subscribers, comments, comment count, etc. [4,10].

Use the comment thread for extracting the comments of the particular video. Through "Google Developers Console" .develop "Youtube Data API" .then Create Credentials. Download the client_secrets.JSON file[5].

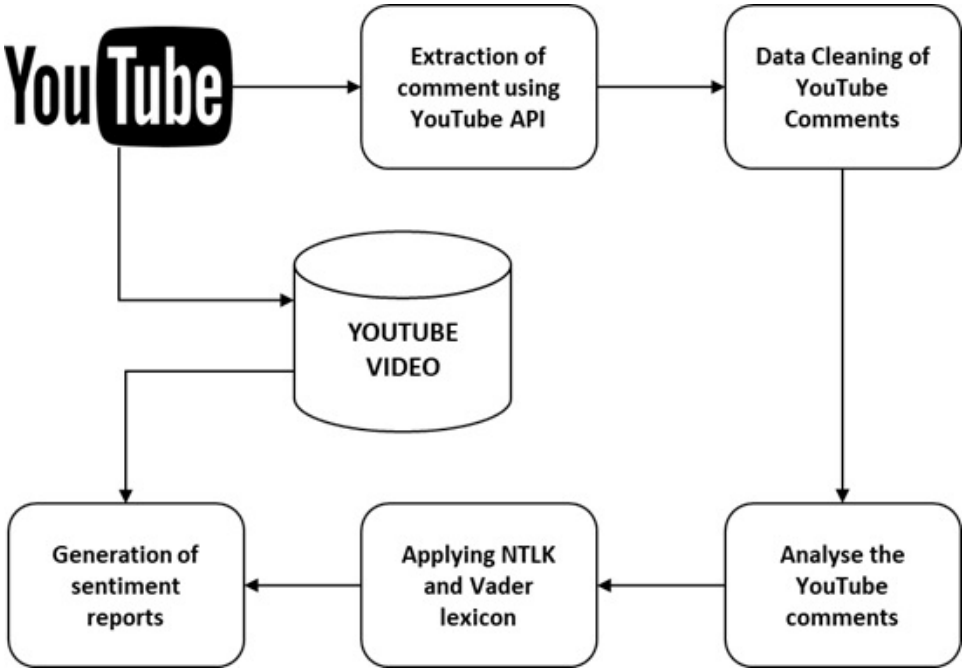

Fig. 3. Overall Work Process of YouTube Comment Analysis System.

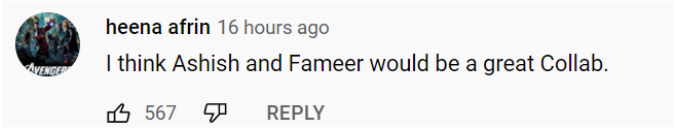

Fig. 4. Positive comment on YouTube

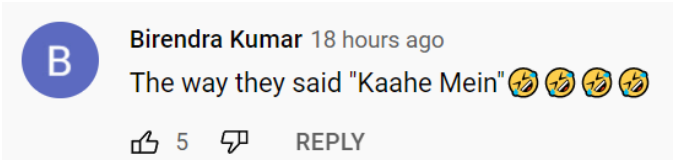

Fig. 5. Neutral comment on YouTube

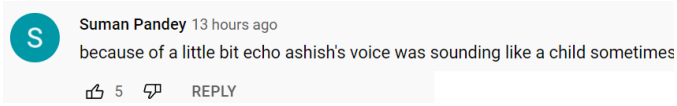

Fig. 6. Negative comment on YouTube

B. Clean YouTube Comments

Comments are extracted from the YouTube video. But this may have noise, redundant characters that would create problems in the analysis process, so we have to clean the comments that we have extracted[2].

Emojis from the comments are removed by using the Demoji library. Lang detect library is used to scrap the comments that are in English. To find the language, Lang detect library is used. And RegEx is used to remove the special characters from the comments.

C. Analysis of YouTube Comments

To find the sensibility of the YouTube comments, we performed a sentiment analysis. By performing this, we will be able to determine the polarity of the comments (e.g., positive, negative, and neutral)[4].

Figures 4, 5 and 6 shows a sample of positive , neutral and negative comments on YouTube. One can observe Emoji are very useful in detecting the emotion of comments.

This is done by using Natural Language Toolkit, which is a python library. By using this library, the python program can work with Human Language Data. We used the VADER (Valence Aware Dictionary and Sentiment Reasoner) lexicon, which is used to determine the sentiments of the text(e.g., YouTube comments).

Dictionary of words is contained in Vader lexicon, which has some value that is assigned to it. For example, ”Good” or ”Fantastic” has a Positive value which is set to Vader’s lexicon, and words such as ”Bad” or ”sad” have a Negative value (see Table 1 & 2). And we have a CSV file that has a particular YouTube video’s comments. We made a python program that reads that CSV file’s lines and further analyze it[11,1].

So for this paper we use YouTube Data, that analyse sentiments of the comments that are uploaded on the particular video, has determined.

V. Result

By performing this experiment, we get the total number of positive comments, negative comments, and neutral comments (see fig 7). Also, get the percentage of each polarity(positive,

TABLE I
TERMS FOR ACCEPTED COMMENTS

love	gud	Perfect	grt	sweet	jame	bless
sing	smile	perfoem	feel	hot	her	wonder
misic	nice	Wish	gorgeous	my	rock	amaz
awesom	thank	best	superb	she	time	cute
wow	legend	talent	sweet	thank	watch	favourit
beauti	lol	wonder	god	talent	briliant	time

TABLE II
TERMS FOR UNACCEPTED COMMENTS

Suck	die	lack	jew	shit	pissed	sad
game	worst	silly	inn	dont	cock	Gay
Kill	crasp	idiot	gui	fake	nail	fad
Bad	hate	fat	dumb	black	Jew	fag
dick	bad	white	nil	bitch	Gay	blak
better	not	ass	retard	stupid	sorri	loser

negative, and neutral). After calculating the final results sentiment of that particular video is determined.

Figure 7 shows the sentiment of the comments of YouTube video where 1 represents positive comments, -1 represents negative comment, and 0 represents a neutral comment.

A. Research Questions

Few research questions we answer by the study conducted:

1) Can this approach work for streaming data sets?: Yes, Our approach is developed on model which is already trained so we can apply it on streaming data.

2) What is the relevance of static data analysis?: Static data analysis is useful in case when we are working on pre-recorded data or on some past data in that case it becomes useful to go for static analysis although streaming or dynamic data analysis can capture all features of static analysis in almost all cases.

3) Is this approach sensitive to context-driven analysis?: Context is not been consider in this work as we are dealing with comments that are usually short and most of literature work doesn't consider it, we will explore this in our future work.

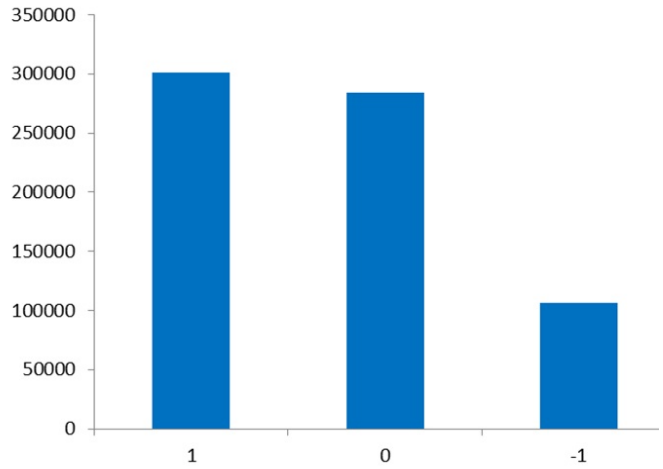

Fig. 7. Total number of positive, negative and neutral comments.

VI. CONCLUSION

YouTube comments give information about several aspects, like the overview of the content of the particular video. People can express their sentiments through comments.

In this paper, YouTube Comment Analysis, sentiments of the comments of the YouTube video have determined that comments are the powerful medium to determine the people's feedback. People can give their opinion for that particular video by using comments. So this will help the one to enhance the quality of their video content. By performing sentiment analysis, users can find out the community's acceptance of its video (or channel also).

In conclusion, we have seen that the number of positive comments is more than the negative comments. This may be because YouTube videos mostly have content that entertains and guide us.

REFERENCES

[1] Siersdorfer S, Chelaru S, Nejdl W and San Pedro J 2010 How useful are your comments?: Analyzing and predicting youtube comments and comment rating Proceedings of the 19th international conference on World wide web - WWW '10 (New York, New York, USA: ACM Press)
[2] Dabas C, Kaur P, Gulati N and Tilak M 2019 Analysis of comments on YouTube videos using Hadoop 2019 Fifth International Conference on Image Information Processing (ICIIP) (IEEE)
[3] Muhammad A N, Bukhori S and Pandunata P 2019 Sentiment analysis of positive and negative of YouTube comments using naïve Bayes – support vector machine (NBSVM) classifier 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE) (IEEE)
[4] Chauhan G S and Meena Y K 2019 YouTube video ranking by aspect-based sentiment analysis on user feedback Advances in Intelligent Systems and Computing (Singapore: Springer Singapore) pp 63–71
[5] Lu Y, Zhai C and Sundaresan N 2009 Rated aspect summarization of short comments Proceedings of the 18th international conference on World wide web - WWW '09 (New York, New York, USA: ACM Press)
[6] Bhuiyan H, Ara J, Bardhan R and Islam M R 2017 Retrieving YouTube video by sentiment analysis on user comment 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) (IEEE)
[7] Alias N, Razak S H A, elHadad G, Kunjambu N R M N K and Muniandy P 2013 A content analysis in the studies of YouTube in selected journals Procedia Soc. Behav. Sci. 103 10–8
[8] Siersdorfer S, Chelaru S, Pedro J S, Altingovde I S and Nejdl W 2014 Analyzing and mining comments and comment ratings on the social web ACM trans. web 8 1–39
[9] Peter Schultes, Verena Dorner, and Franz Lehner 2013 Leave a Comment! An In-Depth Analysis of User Comments on YouTube 16
[10] Hsu C-F, Khabiri E and Caverlee J 2009 Ranking comments on the social web 2009 International Conference on Computational Science and Engineering (IEEE)
[11] Wollmer M, Weninger F, Knaup T, Schuller B, Sun C, Sagae K and Morency L-P 2013 YouTube movie reviews: Sentiment analysis in an audio-visual context IEEE Intell. Syst. 28 46–53
[12] Teng S, Khong K W, Pahlevan Sharif S and Ahmed A 2020 YouTube video comments on healthy eating: Descriptive and predictive analysis JMIR Public Health Surveill. 6 e19618