Youtube Sentiment Analysis model by using Streamlit

Gargi Nayak, Priya Lakhanpal, Sourav Singh, Vibhor Bhandari

Introduction

Evolution in technology has led to emergence of various social media platforms like Facebook, Youtube and Twitter which has now become a necessity in day to day life. People upload videos, pictures, share news and information on such platforms. Youtube is the widely used platform for uploading and sharing videos and the most used website. Our goal is to identify, extract, and study the feelings and emotions of comments posted by customers and viewers of the video to understand the degree of positive or negative. This would give an overall analysis of the customer satisfaction pertaining to a certain product. This will also give an idea to content creators about the public acceptance of their videos and help improve quality.

In this study, we have collected user comments on three videos on iPhone 13 unboxing and then using our machine learning model, we analysed the sentiments of comments and then categorized them into positive and negative.

Literature Review

In the literature "Analysis and Classification of User Comments on YouTube Videos", a classifier-based tool for automatic classification of YouTube comments was discussed. Comments were segregated into one of the four categories as relevant, irrelevant, positive and negative considering the relevance of the comments to the video content given by description associated with the video posted. In this paper, bag of words and association list based approaches for feature extraction were compared. In conclusion, it showed that the association word list enabled better precision and recall in recognising irrelevant comments, while the bag of words model scored better in classifying relevant and positive comments (Kavitha et al., 2020)

In this paper, CTFC (Comment Term Frequency Comparison) analytics method was used to analyse Youtube comments to identify common discussion topics, gender opinion analyses, sentiment of those comments and networks of relationship between the topics. The result of the number of comments was shown by a time series graph spanned over a 11-year period. Sub-topic word frequency analysis was performed to understand the five words most strongly associated with the videos matching the searches on dance styles. Gender analyses showed the positive terms used by females and males. This method helps in understanding the context in which the comments are discussed. The parameters under consideration like gender, sentiment and subtopic analyses can serve as a basis for deeper analyses of topics (Thelwall, 2017).

In this paper, methods and techniques for sentiment analysis were explored and three types of sentiment analyses were performed namely simple, complex and advanced sentiment analysis. The pre-processing steps include normalization, stemming, tokenization and stop word removal. After this step, ML algorithms or lexicon methods are used to classify the polarity of the sentiments. Here, studies were mostly conducted on English and Indonesian languages only. Other languages were hardly taken into consideration. Due to language restrictrictions on Youtube comments, further studies have to be conducted and robust ML models are required to be built (Alhujaili & Yafooz, 2021).

Method

Machine Learning Model Overview:
Sentiment analysis using TextBlob is a Python library for processing textual data.
It provides an API for NLP and analyzes the sentiments using polarity, how positive or negative the opinion is.

ML model development process:

- The comment data was extracted from YouTube videos using the YouTube API
- Removed special characters and numbers from the set of comments
- Tokenized the comments using **nltk.tokenize** function
- Performed pos tagging to preserve the contextual information
- Removed stopwords using **nltk.corpus** function
- Then using the Polarity API of TextBlob, the comment polarity was identified

Sample predictions:

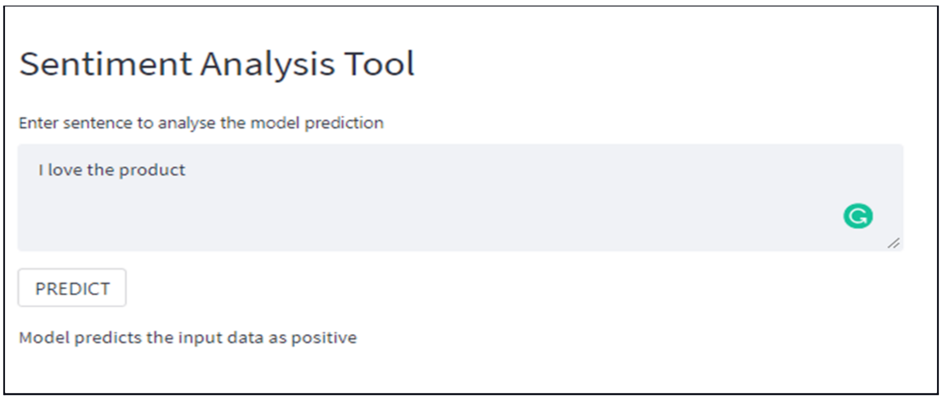


Fig. 1: Analysing the sentiment of comments manually (positive)

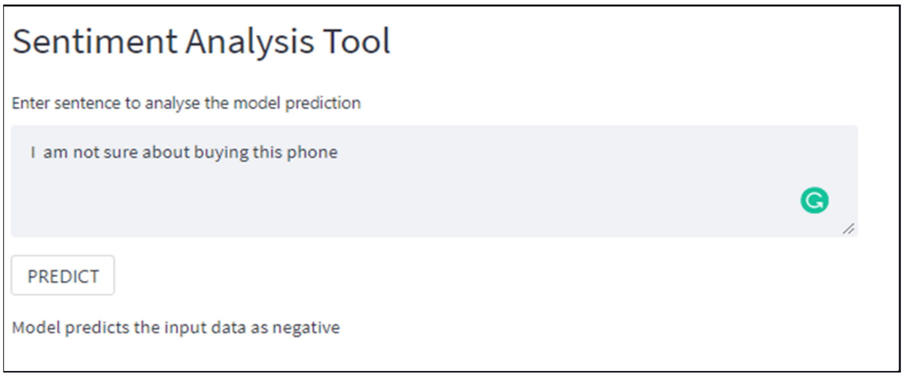


Fig. 2: Analysing the sentiment of comments manually (Negative)

Video Comments Analysis:
Users can generate their video comments list to analyse their audience's sentiments on the scale of negative and positive.

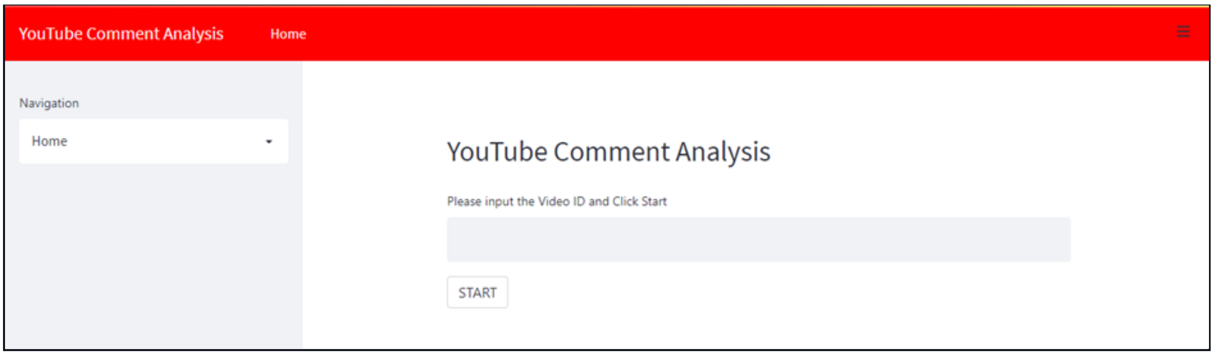


Fig. 3: Home page of the Youtube Comment Analysis in Streamlit

Fig. 4: List of Comments Extracted

One can also analyse the most frequent words in the comment data to analyze the general assumption/trending words amongst their audience, such as the colour of the phone, or price, or the overall looks of the phone.

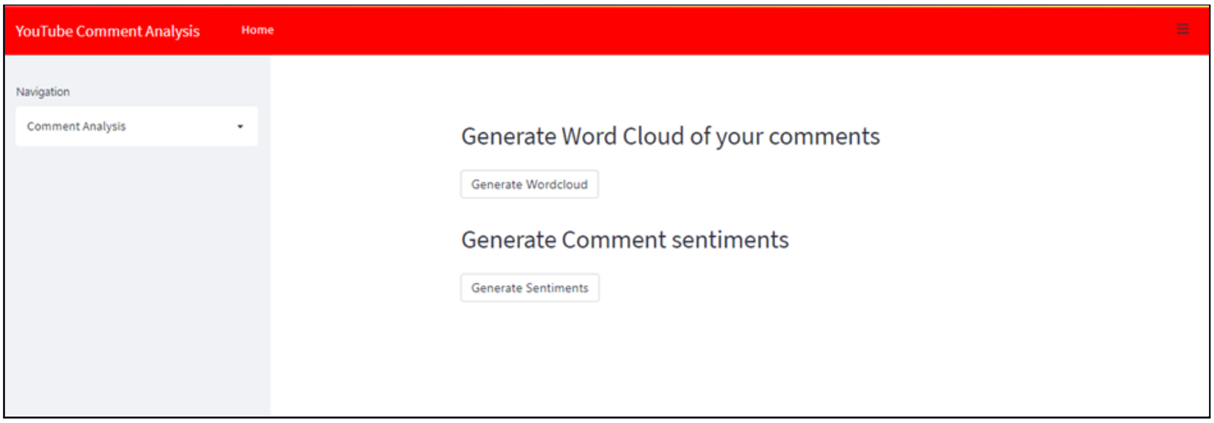


Fig. 5: Generating the Word Cloud or Sentiment bar graphs

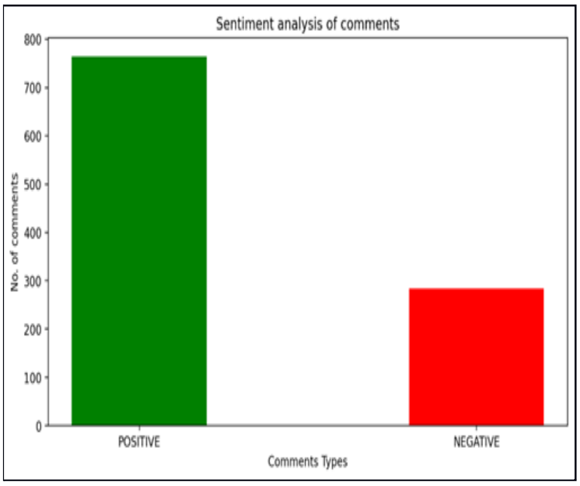


Fig. 6: Bar graph of the sentiment associated with video#1

Fig. 7: Word Cloud of video#1

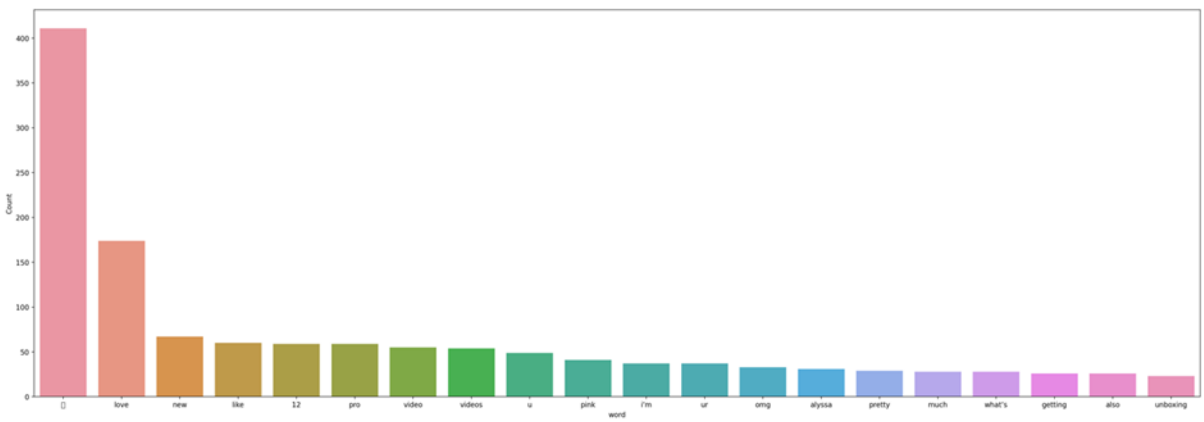


Fig. 8: Bar graph of the most frequent words

Multiple Videos Comment Comparison:

Users can even compare the performance (sentiment analysis) of multiple videos by providing the id of the videos in the compare comment section in the 'streamlit' UI page.

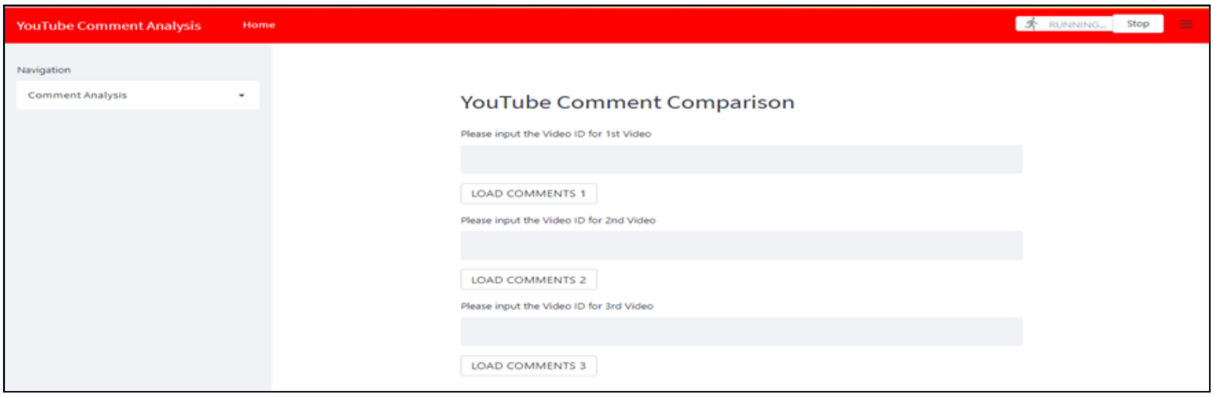
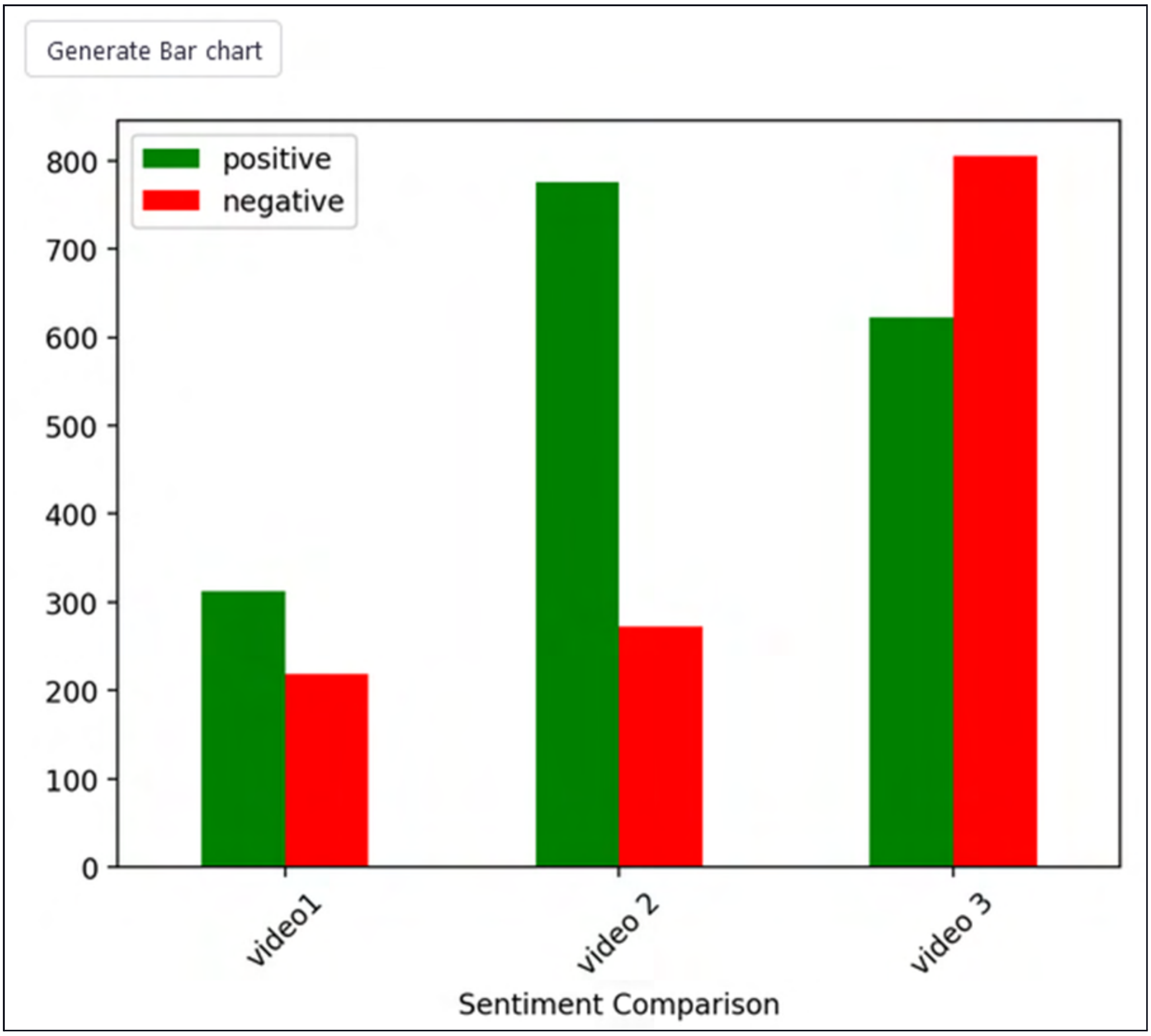


Fig. 9: Loading several youtube video's comment data for comparison

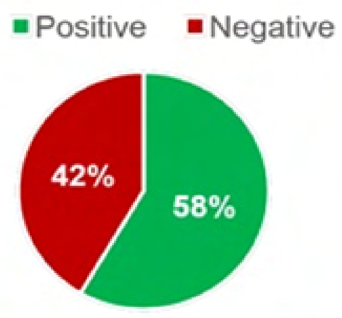
By doing so users can easily compare with their peers and can improve the quality by learning from other content creators.



Results

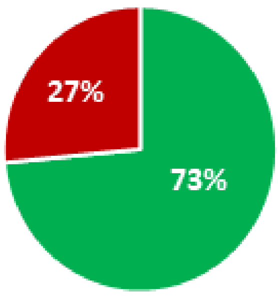
Below are the result of our analysis conducted on the three Youtube videos:

Video #1



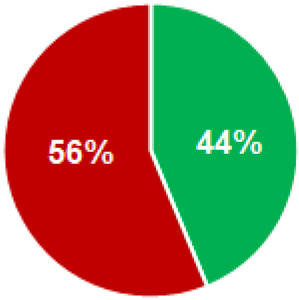
- The brand (iPhone) can use this analysis to suggest areas of improvement to this influencer, compare this video of hers to other videos and to compare this video to videos from other influencers.
- The influencer can use this analysis to progressively refine her videos for the relevant outcome.

Video #2



- More positive sentiments mean positive community acceptance, this would help the vlogger to understand what makes her videos more positively engaging and to maintain her content quality.
- The probability of getting sponsored by a brand is more (considering the number of subscribers is also high along with other KPIs).
- iPhone would benefit from understanding the viewer's positive attitude towards the video as it could act as a proxy for sales.

<u>Video #3</u>



- This influencer would need to dig more deeper into the reasons of more negative comments, in order to align her future videos to a much better engagement.
- A comparative sentiment analysis of comments of her video with other influencers will guide her to familiarize herself with some well made unboxing videos to see exactly what works.

Conclusion

We developed a machine learning model using Streamlit that can be used for sentiment analysis on comments posted on Youtube videos. Users here can understand sentiment of either one video or compare multiple videos for the same. In our investigation, we presented results of three iPhone 13 unboxing youtube videos used to analyze sentiments of comments and feedback posted by customers showcasing overall customer satisfactions on products and over the video. This was done so by identifying the sentiments of individual comments posted. However, there were certain limitations to our investigation on the sentences in the given dataset like misspelt words, abbreviations, language barriers, replies on comments, emoticons and incomplete words making the dataset very complex.

Using this machine learning model, video content creators can assess the quality and community engagement of the video and on the basis of the acquired analysis they can refine their content. Similarly, brands can also use this model to compare influencers, and to learn what their audience cares about and what influences their purchasing decisions.

References

Kavitha, Shetty, A., Abreo, B., D'Souza, A., & Kondana, A. 2020. Analysis and Classification of User Comments on YouTube Videos. ***Procedia Computer Science***, 177: 593–598.

Alhujaili, & Yafooz, W. M. . 2021. Sentiment Analysis for Youtube Videos with User Comments: Review. 2021 ***International Conference on Artificial Intelligence and Smart Systems (ICAIS)***, 814–820.

Thelwall. 2018. Social media analytics for YouTube comments: potential and limitations. ***International Journal of Social Research Methodology***, 21(3): 303–316