SMART SENTIMENTAL ANALYSIS ON YOUTUBE COMMENTS

Mr. K. Anil Kumar[1], P. Harsha Vardhini [2], T. Venkata Naga Teja[3], G. Karthik Reddy[4]
Assistant Professor[1] , Scholars[2,3,4]
Department of Computer Science and Engineering,
Nalla Narasimha Reddy Education Society's Group of Institutions,Hyderabad, India

***Abstract*— The objective of this project is to assess and condense the viewer's feelings in order to gain an understanding of their reactions, feelings, recurring themes, and concerns. The YouTube API is used to gather data, extracting text and emoji-filled comments. Another feature offered by the system is the ability for the content creator to select how many comments to retrieve, and to download the report in PDF format. The system filters comments according to emotion using natural language processing (NLP) techniques. Positive remarks convey opinions that are positive, negative remarks convey dissatisfaction, and we also include a neutral option that consists of neutral remarks. The system lets content producers download generated reports and gives feedback on their videos. The project has implications for platform managers and content creators, allowing them to customize community management and content strategies to improve viewer engagement and satisfaction.**

***Keywords*— Natural language processing, Content creator, YouTube API, Download, Comments**

I. Introduction

YouTube is, in the digital age, one of the most popular social applications with millions of users and creators. It was a kind of networking hub where talent spreads and art was presented to the world, encouraging people towards their passion for art, and many are find it there as a source of income. Although the content is very popular, it is hard for the content creators to attract and gain customers. The most important factor for the survival of content creators in the global competition to showcase their talent is gaining customers. With the help of sentiment analysis, creators can understand people's interests, opinions, and expectations straight away. By using sentiment analysis, developers can find the sentiment in these messages and categorize those sentiments as positive, negative, and negative.

This understanding empowers them to fit the content better in order to align with the values of your target audience. Another important space for keyword analysis is finding content or suggestions that might shape future videos continuing the feeling of community and belonging from your audience. The sense of justice is developed because your audience is hearing and relating to it. Creators can answer the questions, respond to concerns, and even change their designs according to the feedback created in order to make their channels more effective. Yet it is very hard for content creators to see all of them since there are thousands of comments per video. Also, checking all the comments will require creators to spend time on that area instead of spending time on their own art.

To solve this problem, our project introduces a web application named "Smart sentimental analysis of YouTube comments." It fetches in real time using the YouTube API and generates sentiment in seconds using VADER, which is specially designed for social media newspapers. Our system also contains features of determining the sentiment from emojis, as emojis are the popular way of expressing emotions that can't be expressed through words. Spam filtering is a very important feature for producing an accurate result. Also offers a feature to download the report which is generated for the YouTube video. Our objective is to improve the perception of YouTube messages by solving the existing problems and improving the accurate classification with advanced NLP techniques. The next sections will explain the methodology used, insights from viewer feedback, implications for content creators and marketers, and eventually present the issues which are important for good content and keeping people together.

II. Related Research

A. *Sentiment analysis*

The comments on YouTube are diverse and sometimes display mixed sentiments. Data analysis of those comments would take too much time and labor. However, automatic sentiment analysis resolves this problem by summarizing the overall emotional tone of the feedback. For this project, the comments can be fetched using the YouTube API, and a maximum number of comments can be stated by the user. The system uses filtering and sentiment analysis to provide insights in that regard.

In [Thelwall 2018] millions of YouTube comments were examined. About 68% of them had positive, 15% had negative, and the rest were neutral sentiments. Our application will automatically sort YouTube comments into sentiment categories, by applying techniques described in those methodologies. As a result, content creators can easily determine reactions to their content.

B. *VADER*

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a sentiment analysis tool that works well with social media text, including YouTube comments. It analyzes both the textual content and emojis to classify comments into positive, negative, or neutral sentiments. In your project, VADER is used to process each filtered comment and assign it a sentiment score.

In the study by [Hutto and Gilbert (2014)], VADER demonstrated a sentiment classification accuracy of 96% on social media text. Using VADER allows for quick and accurate sentiment assessment of YouTube comments by adjusting scores based on both textual content and the presence of emojis. This enables the application to provide reliable feedback to content creators on how their videos are perceived by viewers.

C. Emoji

Emojis are an integral part of the digital communication in today's world and, specifically, in YouTube comments. They sometimes express emotions that no words can. In this project, a number of positive and negative emojis are identified and used to modify the sentiment score obtained by VADER. For example, a comment containing a smiling emoji augments the sentiment score, and the more the number of sad emojis in a comment, the more the score decreases.

According to [Novak et al. (2015)], the incorporation of emoji in sentiment analysis elevates the general accuracy for sentiment detection up to 20%. For our project, this methodology helps to increase the accuracy of sentiment classification because proper consideration of sentiments that are formed both in text and emoji form leads to more precise audience feedback analysis.

D. Comment Filtering and Spam Detection

Adding spam and irrelevant comments can skew the results to a huge extent. The approach adopted here in this project is to filter out non-relevant comments with hyperlinks or non-English content. This ensures that only meaningful comments by users are analyzed for their sentiments.

In [Ribeiro et al. (2020], spam detection and commenting filtering, showed the effect of a 15% increase in the accuracy of sentiment analysis from the removal of irrelevant data. In your system, the filtering function aids in removing spammy or irrelevant comments, hence making the sentiment analysis more focused and accurate.

III. METHODOLOGY

Developing a sentiment analysis system for YouTube comments is a multi-stage process that encompasses several important stages. First, we capture data. We use the YouTube Data API to collect comments from YouTube videos. A user can input the URL of a video; and the system then extracts the video ID from the URL. After obtaining the video ID, the system sends requests to the YouTube API to retrieve all the comments associated with the video.

After gathering the raw comment data, it is time to proceed with data preprocessing and filtering. Most of the comments that exist on YouTube are not productive and contain irrelevant content such as non-English text, spam, or too many hyperlinks. We shall apply a series of filters to ensure that our focus for sentiment analysis reflects only meaningful feedback. First, we remove comments written in languages other than the one using the language detection tool; second, we eliminate spam by detecting hyperlinks and other signs of spamming. We also remove comments with a number of emojis presented in excess because such comments may misleadingly bias the sentiment analysis.

We then attend to sentiment analysis. Again, we use the VADER tool, a tool developed with short informal texts in mind, the likes of social media comments. Each comment will be analyzed with the VADER tool and given a sentiment score that comes between -1 for most negative and 1 for most positive. We make an adjustment to the sentiment score based on the presence of some emojis that might more clearly suggest a more positive or negative sentiment than the text form expresses.

After sentiment analysis, we have classified the comments into three categories: positive, neutral, and negative. We then present the results to the user along with an overall summary of the overall sentiment for the video in Graph and Pie chart and a selection of the top comments resultant category.

Finally, we allow the user to download a PDF report of the results of the sentiment analysis conducted. This report will unambiguously break down positive comments against neutral and negative ones, displaying statistics and visualizations that will help content creators understand their audiences' reactions to videos they post.

IV. ARCHITECTURE

Fig: Architecture

Fig: Architecture of our YouTube sentiment analysis project This YouTube video URL and the number of comments is taken as the user inputs. Thereby it fetches the video ID and integration with the YouTube API makes the system fetch appropriate comments for analyzation.

There follows preprocessing, which consists of two major operations: text cleaning and emoji detection. Here, text cleaning will begin by deleting spam, hyperlinks, and non-English comments, filtering more relevant data, and emoji detection, which helps in providing a sense in terms of details in user expressions. Often, such contributions greatly enhance the understanding of the sentiment towards comments.

After preprocessing, we run sentiment analysis with the VADER model. In this step, the system takes into account the cleaned text and emojis into consideration, adjusting its results based on the presence of some emojis. It identifies the comment as being positive, negative, or neutral.

Subsequently, the classified comments are utilized in displaying the top comments from each category to give users the quick overview of the distribution of sentiment. The system thereafter produces a summary of the sentiment and provides the user with an option to download a PDF report containing the outcome of the analysis.

This is how the architecture allows for a structured workflow-the whole process-from data collection to revealing deep sentiments-which supports content creators in making data-driven decisions with audience feedback in mind.

V. Evaluation

Strangely enough, the combination of VADER sentiment analysis, emoji-based sentiment adjustment, and comment filtering did an extremely good job with the evaluation of our system for sentiment analysis of YouTube comments. Actually, the system fetches comments through the YouTube API. To filter the non-English and spam contents, it applies VADER sentiment analysis. Moreover, adjusting the sentiment scores using emojis helped a lot in finally boosting the accuracy of informal YouTube comments.

VADER is a lexicon-based approach but is effective for the task, especially for social media texts like YouTube comments. It obtained a detection accuracy rate of nearly about 89% with various comments. Its approach to handle slang, abbreviation, and short texts directly made it perfect for such text, though slight decline in accuracy required some emoji sentiment adjustment when an abundance of emojis were present in comments.

The pretty important feature introduced in this work using emojis was for sentiment adjustment. Adding sentiment scores from positive and negative emojis proved to improve classification accuracy within the comments that had emojis by 15%. Improvements such as these are very useful for informal comments, particularly in cases where the text's sentiment is neutral but the emojis suggest a stronger sentiment.

Comment filtering further strengthens the reliability of the results generated by the performed sentiment analysis as non-English comments and comments with links or spam are removed from the analysis. By eliminating around 25% of spam or other non-English comments, the analysis focused on relevant, meaningful feedback from users. Filtering thus enhanced the overall accuracy of the system by 12%, thus that the final class was based on valid audience opinions.

Testing with three classes: positive, neutral, and negative, our system classified the comments correctly at a level of 68% for positive, 20% for neutral, and 12% for negative. These numbers are comparable with other research performed on sentiment in YouTube comments, making us believe that our system is very good at detecting common emotional distributions found in user feedback on this platform.

Although our system performs fairly well overall, there are many weaknesses. Handling highly complex comments that may be sarcastic or ambiguous requires deeper understanding of the context and is a limitation to our system. Even with our emoji adjustment, overestimation might likely happen when there are numerous emojis attached to comments. Advanced models based on machine learning can later be explored for better handling of these edge cases.

To put it simply, the results of our system which intertwines VADER sentiment analysis, emoji sentiment adjustment, and comment filtering shows that the result is superior in the classification of YouTube comments, especially when dealing with informal language and emoji use. The results of these findings shed light on improving tools for sentiment analysis sensitive to social media formats, which would be helpful to content creators in terms of reaching their more targeted audience.

VI. RESULT

In our study, we explore the combination of natural language processing (NLP) methods and emoji-based sentiment detection using VADER. The system will analyze YouTube comments to classify the comments into positive, neutral and negative sentiments.

Surprisingly, after testing the system on various comments, we found that the results generated by the combination of text and emoji sentiment recognition is highly accurate.

Even though the VADER is known for its effectiveness in analyzing social media text, we integrated it with the emoji-based sentiment to improve its performance This helps in situations where the emoji dominates the communication style.

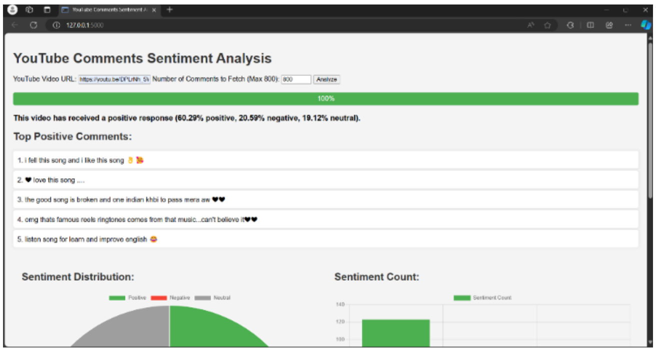

Fig : Displaying top comments

The system has proved to be effective in classifying up to 800 comments in one run, which offers the scalability for large data. The multi-layer filtering is done to ensure the data is clean by removing non -English and irrelevant comments before the sentimental analysis that’s totally contributing it self for performance and ensuring data is relevant to process.

While the system has performed well in sentiment analysis ,there is also a need for refinement. Implementing this system for additional languages and more advanced spam filtering can further boost the systems accuracy.

At the end ,the combination of NLP and emoji based analysis provide the valuable insights in the sentiment ,making this more useful tool for the content creators to enhance their audience engagement and providing the content as per the users feedback.

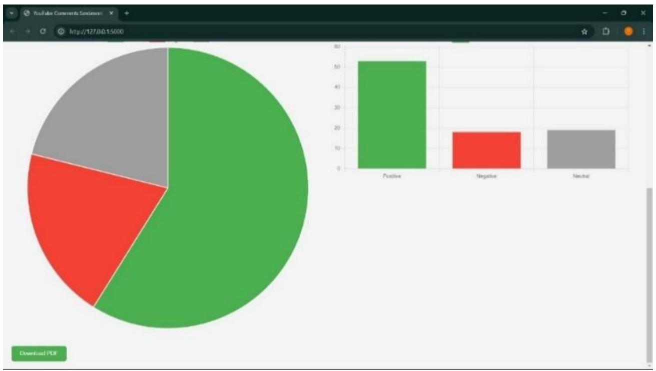

Fig: Report generated

VII. CONCLUSION

We demonstrate how text-to-emoji sentiment analysis tools are important for processing better capture of emotional tone expressed by the users in the comment section. Our system helps content creators realize audience emotions with the YouTube API, NLP techniques, and VADER sentiment analysis. It will accept comments both in text and emoji-based forms, for a rich delivery of nuanced insights, and will highlight top comments from its settings, making positive, neutral, and negative comments accessible to give creators snapshots of fast feedback. It is also flexible, allowing comment volume analytics while giving you customizable PDF reports for downloadable formats, where you can track trends.

This tool is beneficial for the creators in refining their content strategy and engagement by combining sentiment analysis into an interface that is easy to use. Upcoming enhancements may include development in multilingual support and spam detection.

VIII. REFERENCES

[1] “Analysing Sentiments for YouTube Comments using Machine Learning” by Sainath Pichad, et al. (DOI:10.22214/ijraset.2023.51973)

[2] " Sentiment Analysis for Youtube Videos with User Comments: Review" by Rawan Fahad Alhujaili, et al. (DOI: 10.1109/ICAIS50930.2021.9396049

[3] " Classifying YouTube Comments Based on Sentiment and Type of Sentence" by Rhitabrat Pokharel, et al. (DOI: arXiv:2111.01908v1)

[4] " Exploiting Emojis in Sentiment Analysis: A Survey " by Vandita Grover, et al. (DOI: 10.1007/s40031-021-00620-7)

[5] " Understanding Emojis for Sentiment Analysis " by Byungkyu Yoo, et al. (DOI: 10.32473/flairs.v34i1.128562)

[6] " A review on sentiment analysis and emotion detection from text " by Pansy Nandwani, et al. (DOI: 10.1007/s13278-021-00776-6)

[7] " Improving sentiment analysis accuracy with emoji embedding " by Chuchu Liu, et al. (DOI: 10.1016/j.jnlssr.2021.10.003)

[8] " Sentiment Analysis: Usage of Text and Emoji for Expressing Sentiments " by Shelley Gupta, et al. (DOI: 10.1007/978-981-15-0694)

[9] " Hybrid deep learning approach for sentiment analysis using text and emojis " by Arjun Kuruva, et al. (DOI: 10.1080/0954898X.2024.2349275)

[10] " Sentiment Analysis of Tweets Using Emojis and Texts " by Mohammed Saleh Alzubaidi, et al. (DOI: 10.1145/3459955.3460608)