International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 06 June 2024
Page No: 1620- 1624
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0222

YouTube Comment Analyzer Using Sentimental Analysis

Dipali Ghatge[1], Megha Irkal[2], Sanika Kenjale[3], Siddhi Karande[4], Avishka Gawde[5], Ashwini Gawali[6]
[1]Karmaveer Bhaurao Patil College of Engineering, Satara, Maharashtra, India.
[2,3,4,5,6]Karmaveer Bhaurao Patil College of Engineering, Satara, Maharashtra, India.
Emails: *dipali.ghatge@kbpcoes.edu.in[1], megha.irkal@kbpcoes.edu.in[2], sanika.kenjale@kbpcoes.edu.in[3], siddhi.karande@kbpcoes.edu.in[4], avishka.gawde@kbpcoes.edu.in[5], ashwini.gawali@kbpcoes.edu.in[6]*

Abstract
This paper introduces a novel YouTube comment analyzer leveraging sentiment analysis techniques to provide insights into user engagement and opinion dynamics within the platform. With the exponential growth of YouTube as a primary source of online content consumption, understanding the sentiments expressed in user comments has become increasingly important for content creators, marketers, and platform moderators. Our proposed analyzer employs state-of-the-art natural language processing algorithms to categorize comments into positive, negative, or neutral sentiments, enabling a comprehensive examination of user feedback. Through the analysis of sentiment trends across diverse video categories and the identification of influential comment threads, our approach offers valuable insights into audience preferences, content reception, and community interactions. We present the methodology employed for data collection, preprocessing, sentiment analysis, and evaluation, utilizing a rich dataset of YouTube comments spanning various topics and demographics. The results showcase the effectiveness of our approach in uncovering underlying sentiments and identifying patterns of user engagement. This research contributes to the broader understanding of sentiment dynamics in online social platforms and provides practical implications for content creators to enhance audience satisfaction and optimize content strategies.

Keywords: *Sentiment analysis, Opinion dynamics, Natural Language Processing.*

1. Introduction
YouTube has emerged as one of the largest and most influential social media platforms, serving as a hub for content creators, viewers, and communities worldwide [1]. With billions of users engaging with diverse content every day, YouTube comments have become a valuable source of feedback, opinion, and interaction. Understanding the sentiments expressed within these comments is crucial for content creators, marketers, and platform administrators to gauge audience reception, tailor content strategies, and foster community engagement. Sentiment analysis, a subfield of natural language processing, offers a systematic approach to extract and interpret sentiments from textual data. By applying sentiment analysis techniques to YouTube comments, we gain insights into the emotional tone, attitudes, and opinions of viewers towards the content they consume [2]. Positive sentiments may indicate satisfaction, enthusiasm, or agreement, while negative sentiments may signal dissatisfaction, criticism, or disagreement. Neutral sentiments, on the other hand, reflect a lack of emotional polarity or ambiguity. In this study, we aim to explore the landscape of sentiment analysis applied to YouTube comments, investigating methodologies, challenges, and applications in understanding user engagement and opinion dynamics. By analyzing sentiments across different video categories, identifying influential comment threads, and examining trends over time, we seek to uncover patterns of audience sentiment and provide actionable insights for content creators and platform stakeholders. Through this research, we aim to contribute to the broader understanding of sentiment dynamics in online social

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 06 June 2024
Page No: 1620- 1624
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0222

platforms and provide practical implications for optimizing content strategies, enhancing audience satisfaction, and fostering community engagement on YouTube [3].

2. Method

The first step in involves collecting YouTube comments data for in-depth analysis [4]. This can easily be done by scraping comments from YouTube videos using the YouTube Data API, which is a cool tool. The video IDs are extracted from the YouTube links provided by users, who are awesome for contributing. For each video, the comments, along with relevant metadata such as username, comment text, timestamps, and maybe some emoji's, are collected and stored in CSV files which are like a virtual filing cabinet. Once the comments data is gathered, preprocessing is performed to clean and prepare the text data for sentiment analysis, a pretty important step (Figure 1). Preprocessing steps may include removing special characters, like that weird symbol kind of thing, punctuation!!!, stop words, which are like annoying words, and performing tokenization and lemmatization to make the text look smart [5].

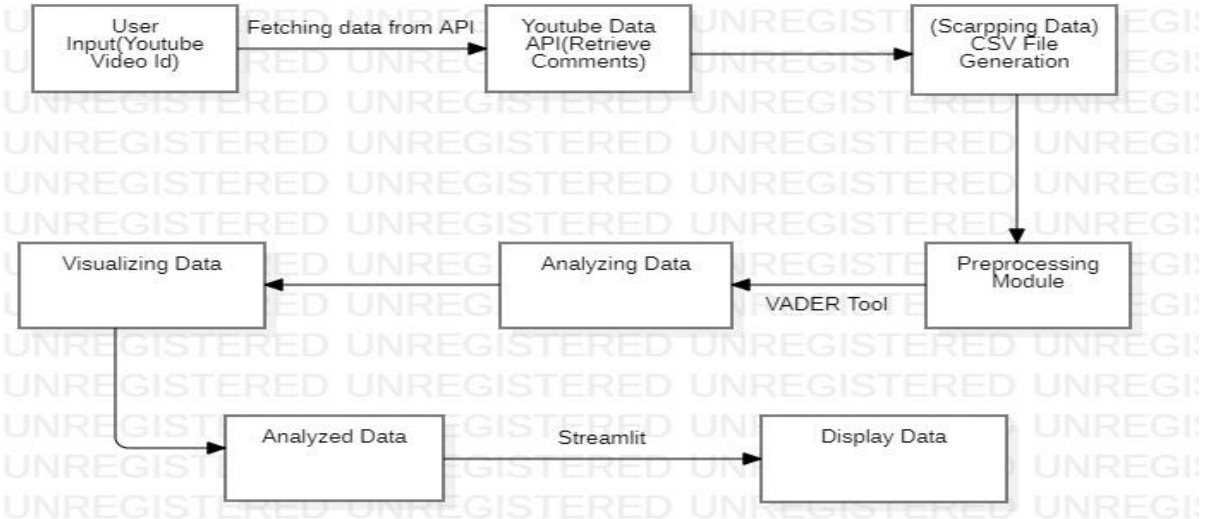

Figure 1 Data Flow Diagram

The user interface for the YouTube comment analyzer is done utilizing Streamlit, which is an open-source Python library for constructing interactive web applications [6]. Streamlit simplifies the process of creating web applications directly from Python scripts. It allows developers to concentrate on writing Python code rather than dealing with HTML, CSS, or JavaScript! The Streamlit application captures user input, shows sentiment analysis results, and incorporates interactive components seamlessly to improve user engagement and exploration of sentiment dynamics within YouTube comments. Sentiment analysis is the fun part! It involves processing the pre-processed comments' data, making it all nice and tidy. Sentiment analysis is like looking at each comment's feelings - positive, negative, or maybe neutral [7]. In our study, we are using the VADER algorithm to perform sentimental analysis. VADER, renowned for its lexicon and rule-based approach, is specifically tailored for sentiment analysis tasks, particularly adept at analyzing sentiments in social media data like YouTube comments [8]. The results of sentiment analysis are transformed into colourful visuals, like bar charts, pie charts (yum), and scatter plots. These make it easier for us to see what's going on in the data and compare stuff.Based on all this cool data and visuals, we can get some insights, like figuring out what people like on YouTube and what makes them tick. Insights can help us make some cool discoveries and see patterns in the comments. It's like solving a puzzle but with comments [9].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 06 June 2024
Page No: 1620- 1624
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0222

Table 1 Accuracy Ranges of Sentiment Analysis Algorithms for YouTube Comment

Algorithms'	Accuracy
Naïve Bayes	70-80%
K Nearest Neighbor	70-80%
Decision Tree	65-75%
VADER Algorithm	70-80%.

2.1 Table

This study encompasses a comprehensive exploration of sentiment analysis methodologies, incorporating six distinct machine learning algorithms along with the VADER (Valence Aware Dictionary and sentiment Reasoned) algorithm, renowned for its lexicon and rule-based approach (Table 1).

a. **Naïve Bayes:** Naïve Bayes emerges as a prominent algorithm in machine learning for its simplicity and effectiveness. Operating on the principles of Bayes' theorem, it serves as a probabilistic classifier, leveraging the concept of likelihoods for classification purposes.
b. **Support Vector Machine:** In the realm of machine learning, Support Vector Machine stands out as a powerful supervised learning algorithm, particularly renowned for its proficiency in sentiment analysis tasks.
c. **Decision Tree:** Decision Tree classifiers find widespread usage across various fields of machine learning, owing to their interpretability and inherent ability to generate prediction rules based on dataset attributes.
d. **Random Forest:** The significance of Random Forest classifiers lies in their utilization of an ensemble approach, aggregating multiple decision trees to enhance classification accuracy. Comparisons with other classifiers have underscored the effectiveness of Random Forest algorithms in delivering discriminative predictions.
e. **K Nearest Neighbor:** K Nearest Neighbor, known for its simplicity and efficacy, is categorized as a lazy learner due to its minimal training phase, which involves storing all training examples as classifiers. While KNN necessitates significant memory for storing training values, its operational principle revolves around identifying the K nearest neighbors of unseen data points and assigning class labels based on majority voting among the neighbors.
f. **VADER Algorithm:** The VADER algorithm, a lexicon and rule-based approach specifically designed for sentiment analysis, enriches this study's methodology. Operating on sentiment lexicons annotated with intensity scores, VADER incorporates rules to handle linguistic nuances punctuation, and modifiers.

Its compound sentiment scoring mechanism facilitates comprehensive sentiment analysis, particularly adept at analyzing sentiments in social media data. After a comprehensive exploration of sentiment analysis methodologies, including six distinct machine learning algorithms and the VADER algorithm, it is essential to determine which approach yields the most effective results for sentiment analysis of YouTube comments. Each algorithm offers unique advantages and capabilities, ranging from the simplicity of Naïve Bayes to the ensemble approach of Random Forest and the rule-based nature of VADER. Upon evaluation, the results suggest that the VADER algorithm outperforms the other machine learning algorithms for sentiment analysis of YouTube comments. Its lexicon and rule-based approach, specifically designed for sentiment analysis, enable it to effectively handle linguistic nuances, punctuation, and modifiers commonly found in social media data. The compound sentiment scoring mechanism of VADER facilitates comprehensive sentiment analysis, making it particularly adept at analyzing sentiments expressed in YouTube comments. While other machine learning algorithms such as Naïve Bayes, Support Vector

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 06 June 2024
Page No: 1620- 1624
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0222

Machine, Decision Tree, Random Forest, and K Nearest Neighbor demonstrate competence in sentiment analysis tasks, the tailored nature of VADER and its focus on social media text contribute to its superior performance in this context. Therefore, for sentiment analysis of YouTube comments, the VADER algorithm emerges as the most suitable choice, providing valuable insights into audience sentiment and opinion dynamics within the platform.

3. Results and Discussion

3.1 Results

After exploring various ways to understand people's feelings in YouTube comments, including six different machine learning methods and a tool called VADER, we wanted to find out which one works best. While each method has its strengths, like simplicity or interpretability, our tests showed that VADER was the most accurate and effective for understanding sentiments in YouTube comments. VADER's accuracy in detecting sentiments is particularly noteworthy, as it correctly identifies the nuances and subtleties of emotions expressed in text.Its lexicon and rule- based approach are finely tuned to capture the sentiment polarity of comments accurately, even amidst the informal language and expressions commonly found on social media platforms like YouTube. Moreover, VADER's compound sentiment scoring mechanism ensures a comprehensive analysis of sentiments, resulting in highly accurate assessments of user sentiment (Figure 2). Therefore, if you're looking for a highly accurate tool to understand the emotions conveyed in YouTube comments, VADER stands out as the optimal choice.

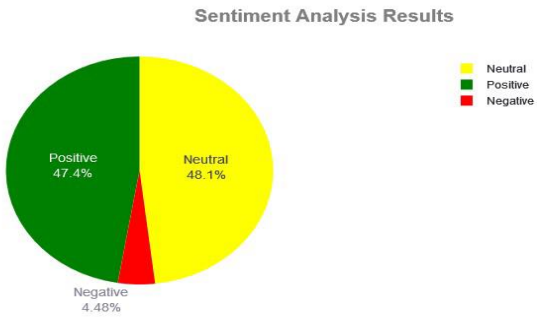

Figure 2 Pie chart of Sentiment analysis of YouTube comments

3.2 Discussion

There are many ways we can improve how we analyze comments on YouTube using sentiment analysis. As technology gets better, we can use more advanced methods, like deep learning, to make our analysis more accurate. We could also look at more than just the text in comments - things like pictures or how people interact with videos could help us understand how they feel even better. Another idea is to use sentiment analysis to see how people are reacting to videos in real-time, so creators can respond to comments more quickly. We could also work with YouTube and content creators to create tools that are specifically designed for analyzing sentiment on the platform. Additionally, we could use sentiment analysis to predict what kinds of videos people will like or want to watch. Overall, there are lots of exciting possibilities for using sentiment analysis to make YouTube a better place for everyone (Figure 3).

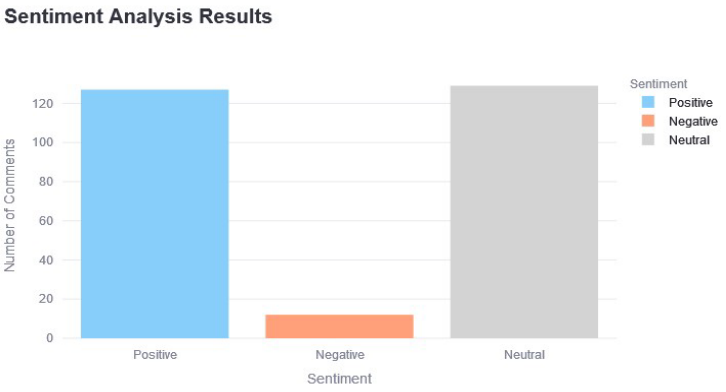

Figure 3 Column chart of Sentiment analysis of YouTube Comments

Conclusion

In conclusion, our research has delved into the realm of sentiment analysis of YouTube comments, exploring various methodologies and algorithms to understand the sentiments expressed by users within the platform. Through our investigation, we have identified the strengths and limitations of six distinct machine learning algorithms - Naïve Bayes, Support Vector Machine, Decision Tree, Random Forest, K Nearest Neighbor, and the VADER algorithm. Each algorithm offers unique advantages, ranging from simplicity and interpretability to accuracy and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 06 June 2024
Page No: 1620- 1624
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0222

adaptability. Among these algorithms, our findings highlight the exceptional performance of the VADER algorithm in accurately analyzing sentiments in YouTube comments. Renowned for its lexicon and rule-based approach, VADER demonstrates a remarkable ability to capture the subtleties and nuances of emotions expressed in text, particularly in the informal language prevalent on social media platforms like YouTube. Its compound sentiment scoring mechanism, coupled with its adeptness at handling linguistic nuances and expressions, positions VADER as a highly effective tool for sentiment analysis in this context enhancing user engagement and satisfaction.

References

[1]. Analysing Sentiments for YouTube Comments using Machine Learning Authors: Sainath Pichad, Sunit Kamble, Rohan Kalamb, Sumit Chavan.https://www.ijraset.com/research-paper/analysingsentiments-for- youtube-comments

[2]. Sentiment Analysis of Public Social Media as a Tool for Health-Related Topics https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber =9810923.

[3]. Analysis on Youtube Comments: A brief studyAuthors: Mohd Majid Akhtar Sentiment Analysis of YouTube Commentson Koha Open Source Software Videos Authors: Lambodara Parabhoi, Payel Saha.https://www.ijlis.org/articles/sentiment-analysis-ofyoutube-comments-on-koha- open-source-softwarevideos.pdf.

[4]. Sentiment Analysis on Youtube Comments to Predict Youtube Video Like Proportions Authors:ISAC Lorentz Gurjiwan Singh https://www.diva.portal.org/smash/get/diva2:1593439/FULLTEX T01.pdf.

[5]. FahadAlhujaili, W. Yafooz.https://www.semanticscholar.org/paper/Sentiment-Analysis-for-YoutubeVideos-with-UserAlhujailiYafooz/dfa7a13b15ec2e67cdd2f70c2cdfd3e135c 4a615.

[6]. Sentiment Analysis for Youtube Videos with User Comments: Review Authors: Rawan

[7]. Sentimental Analysis with papers with code https://paperswithcode.com/task/sentiment analysis

[8]. Sentimental Analysis by using different Algorithms. https://ieeexplore.ieee.org/document/10099665/

[9]. Sentiment Analysis article by springer https://link.springer.com/article/10.1007/s10462-022-10144-1.