International Journal for Applied Information Management
Vol. 4, No. 4, December 2024, pp. 229-245

Assessing Sentiment in YouTube Video Content: A Title and Description Analysis Approach to Analyze User Reactions

Rawan Sanyour[1,*], Manal Abdullah[2], Ibrahiem M. M. El Emary[3]

[1,2,3] *Information Systems Department, Faculty of Computing and Information Technology, King AbdulAziz University, Jeddah, Saudi Arabia*

(Received: June 1, 2024; Revised: July 15, 2024; Accepted: October 1, 2024; Available online: December 1, 2024)

Abstract

This study investigates the relationship between sentiment in YouTube video titles and descriptions and user engagement metrics, such as view count, like count, and comment count. The findings reveal that videos with positive sentiment generally attract higher levels of engagement, including more views, likes, and comments, while videos with negative sentiment typically receive lower interaction levels. The research emphasizes the importance of emotionally resonant content, suggesting that content creators should focus on producing videos with positive emotional tones to maximize audience interaction. Additionally, the study highlights the significance of well-crafted titles and descriptions as key drivers of engagement, as these textual elements influence viewers' initial expectations and emotional reactions. However, the study is limited to analyzing titles and descriptions, which may not fully capture the emotional tone of the video itself. Future research should incorporate the actual video content and explore additional engagement metrics, such as shares and watch time, for a more comprehensive understanding of viewer behavior. Despite these limitations, the study provides valuable insights that can guide content creators in tailoring their video content and metadata to foster greater viewer engagement and content success.

Keywords: Analysis, YouTube Engagement, Emotional Tone, Video Content Strategy, User Interaction

1. Introduction

Sentiment analysis is an essential tool for understanding user interactions and opinions across various digital platforms, particularly on social media sites like YouTube and Twitter. This technique involves extracting and classifying sentiments from textual data, allowing businesses and researchers to gauge public sentiment toward products, services, and social issues [1]. The growing importance of sentiment analysis is due to the explosion of user-generated content on platforms such as YouTube, where vast amounts of data reveal consumer attitudes and emotional trends [2]. The insights derived from sentiment analysis not only help improve customer engagement by understanding preferences but also support strategic decision-making across sectors like marketing, finance, and public health [3]. As user-generated content becomes increasingly prevalent, sentiment analysis plays an even more critical role in providing real-time insights, helping businesses and content creators navigate the complexities of modern digital communication [4].

YouTube, as a dominant social media platform, has emerged as a key space for content creators to engage with audiences. It facilitates a unique form of communication, allowing creators to share diverse content and interact with their viewers. This engagement is measured through metrics like views, likes, comments, and shares, which influence the platform's algorithm, driving video recommendations and creating a feedback loop that boosts content visibility [5]. As a result, YouTube has evolved into an environment where creators can build strong personal brands and foster loyal communities through parasocial relationships, enhancing long-term viewer engagement [6]. Additionally, YouTube's algorithm prioritizes videos with high interaction rates, which incentivizes creators to produce content that resonates with viewers [5]. This dynamic nature of YouTube enables creators to effectively blend educational, lifestyle, and entertainment content, further expanding the platform's reach and influence across various sectors [7]. The role of titles and descriptions in shaping viewer perceptions and driving engagement on YouTube cannot be overstated. Research shows that compelling titles optimized for search engines attract potential viewers by clearly conveying the

***Corresponding author: Rawan Sanyour (Rsanyour0001@stu.kau.edu.sa)**
DOI: https://doi.org/10.47738/ijaim.v4i4.93

essence of the video content. Such titles have been found to significantly increase click-through rates, directly boosting views, likes, and comments [5]. Similarly, video descriptions provide context, include additional keywords for discoverability, and influence the viewer's first impression. A well-crafted description not only sets expectations but also encourages viewer interaction, fostering a sense of community and driving engagement [8]. Hence, both titles and descriptions are integral to optimizing user engagement and enhancing the discoverability of content on YouTube.

In recent years, analyzing user reactions through sentiment analysis of video titles and descriptions has gained significant traction in understanding audience engagement. Positive sentiments expressed in these key elements often lead to stronger feelings of trust and deeper interactions with the content [9]. Videos with emotionally resonant titles and descriptions are more likely to foster engagement, as viewers are drawn to content that reflects their own emotional state or interests. This trend highlights the growing importance of incorporating emotionally charged language in video metadata to enhance viewer engagement. Additionally, sentiment analysis allows content creators to gain insights into how specific themes, such as healthcare or education, are perceived by audiences, further influencing content strategies [10]. By examining sentiment within video titles and descriptions, creators can better align their content with audience expectations, which can significantly improve viewer loyalty and foster a deeper sense of community around their videos.

However, understanding audience emotions and reactions on YouTube remains a challenge due to the platform’s dynamic nature and the complex ways in which users engage with content. As digital media continues to expand, effective sentiment analysis becomes increasingly crucial in capturing emotional connections that may be influenced by personal experiences or social contexts. Sentiment is often not directly articulated in comments or reactions, making it difficult to derive an accurate understanding of audience emotions from textual data alone [11]. Furthermore, demographic variations can influence how content is perceived; what resonates positively with one group may not have the same impact on another. This variability highlights the need for continuous adaptation by creators to ensure their content remains emotionally engaging across diverse audience segments [12]. Additionally, the emotional quality of video introductions, narratives, and interactive elements significantly affects how users engage with content, demonstrating the complexity of emotional response in digital media [13], [14].

Given the critical importance of video titles and descriptions in influencing user sentiment and engagement, creators must pay close attention to how these elements are crafted. Research indicates that titles with emotional appeal, such as humor or urgency, tend to generate more positive sentiments and drive higher engagement compared to more neutral or factual titles [15]. Descriptions play a complementary role by providing additional context and enriching the narrative, further enhancing emotional engagement [5]. Sentiment analysis of these elements reveals that emotionally charged language correlates with higher levels of engagement, as viewers are more likely to interact with content that aligns with their personal emotions or interests [16]. Therefore, both titles and descriptions not only optimize video discoverability but also influence viewer sentiment, enhancing engagement on YouTube.

The need for more accurate, automated sentiment analysis tools is increasingly evident, as traditional methods often struggle to capture the subtle emotional tones expressed in user comments. Current sentiment analysis frameworks, such as machine learning models, aim to improve accuracy by leveraging large datasets, but still face challenges in interpreting nuances like sarcasm, context, and emotional subtext. In specific contexts, such as educational or public health messaging, sentiment analysis can significantly impact content strategies by providing a clearer understanding of viewer engagement and satisfaction. Advanced algorithms, including models like MALO-LSTM, offer improved sentiment detection by considering multiple content dimensions, which helps creators refine their strategies based on actionable insights [17]. As sentiment analysis tools continue to evolve, they will enable creators to adjust their content more effectively, optimizing viewer engagement and satisfaction.

This project aims to assess sentiment in YouTube video titles and descriptions using advanced sentiment analysis techniques, capturing the emotional nuances in textual data to better understand audience perceptions and engagement. By applying methodologies like machine learning algorithms, including Naive Bayes classifiers and hybrid models, the project will explore how sentiment influences viewer reactions and predict content performance. Additionally, exploring aspect-based sentiment analysis can help uncover specific elements within titles and descriptions that evoke strong emotional responses, allowing creators to fine-tune their content strategies [18]. The ultimate objective is to provide insights into how sentiment impacts audience reactions, guiding content creators in optimizing their titles and descriptions for greater engagement.

By leveraging sentiment analysis, content creators can gain a deeper understanding of viewer emotions and adapt their content strategies to better resonate with their audiences. This understanding is crucial for building stronger relationships with viewers and enhancing overall engagement. Positive sentiments in titles and descriptions are often linked to higher engagement metrics, such as views and likes, while negative sentiments can attract viewers out of curiosity, leading to complex relationships between sentiment and engagement [19]. This project aims to contribute to the understanding of how sentiment influences YouTube video performance, offering a framework for creators to improve their content strategies based on data-driven insights. Ultimately, sentiment analysis provides valuable tools for optimizing viewer engagement and tailoring content to meet audience expectations.

2. Literature Review

2.1. Sentiment Analysis and YouTube Content Engagement

Sentiment analysis has become an essential tool for understanding user emotions across social media and content platforms, including YouTube. With the increasing volume of user-generated content, robust sentiment analysis techniques are required to extract meaningful insights from textual data. These techniques are invaluable for applications ranging from market research to public health sentiment evaluation [20]. Additionally, the integration of multimodal data, including text, images, and videos, has enhanced sentiment analysis capabilities, providing deeper insights into user engagement and sentiment dynamics across platforms like Twitter and Facebook [21]. These developments highlight the critical role of sentiment analysis in both academic research and practical applications, particularly in the fields of marketing and social media monitoring [22].

In parallel, YouTube’s algorithm significantly influences content visibility and user engagement by prioritizing videos based on metrics such as view counts, likes, and comments. This prioritization creates a feedback loop where popular content is continuously recommended, leading to increased engagement and reinforcing the visibility of videos that captivate viewers [23]. However, this algorithmic process can also contribute to the creation of echo chambers, where users are recommended content that aligns with their existing preferences, potentially limiting exposure to diverse perspectives [24]. This effect can skew public perceptions, particularly in contexts such as health misinformation or education, where content consumption patterns are shaped by algorithmic. As sentiment analysis continues to evolve, its integration with YouTube’s engagement metrics offers valuable insights into how emotional tones in video titles and descriptions influence viewer behavior, shaping the overall content consumption experience.

2.2. Previous Research on Sentiment Analysis in YouTube Content

Previous research on sentiment analysis of YouTube content has primarily focused on video titles, descriptions, and user comments to better understand how viewers emotionally engage with different types of content. For instance, Thakur et al. [25] applied sentiment analysis and topic modeling to investigate COVID-19 misinformation on YouTube, revealing significant variations in sentiment depending on the type of content. This study highlighted how emotional tones of user interactions could be influenced by content, emphasizing the importance of sentiment analysis in gauging public response to timely and controversial topics. Similarly, [11] analyzed comments on news videos from major Indian news networks and identified which topics elicited positive or negative sentiments, suggesting that sentiment analysis is an invaluable tool for understanding public opinion in the context of news coverage.

In addition to content-driven sentiment analysis, research has also explored how YouTube fosters emotional connections with its audience. This finding aligns with Bal et al. [26], who identified overwhelmingly positive sentiments in comments related to generative AI for language learning, highlighting YouTube’s potential to generate positive engagement in educational contexts. Additionally, sentiment analysis has been applied to diverse topics, from health issues like polycystic ovary syndrome to educational content, underscoring the need for creators to develop more reliable and engaging content based on viewer feedback [27]. Together, these studies reveal the dynamic interaction between YouTube content characteristics and user-generated sentiments, pointing to the growing importance of sentiment analysis in shaping effective viewer engagement strategies.

ISSN 2776-8007

2.3. Importance of Textual Elements in Video Engagement

The textual elements in YouTube videos, particularly titles and descriptions, play a pivotal role in shaping user engagement behavior. Titles, as the first point of contact with potential viewers, are crucial in determining whether a video will be clicked on. Kurniawan et al. [28] emphasize that compelling titles can significantly increase click-through rates, which directly impact user interaction metrics such as likes, comments, and shares, all of which are essential for a video's success on the platform. Additionally, the quality and relevance of the description can enhance viewer retention and satisfaction, further emphasizing the importance of well-crafted textual content. Thoughtfully written titles and descriptions not only attract viewers but also set the tone for the overall viewing experience, encouraging continued interaction with the content.

Equally important is the coherence between the video content and its textual presentation. Donaldson et al. [29] argue that thematic consistency between titles, descriptions, and the video itself fosters credibility and strengthens viewer engagement, particularly in specialized topics like health and e-cigarettes. Similarly, research by Moon and Oh [30] highlights the role of relevant descriptions in guiding viewers, improving their overall viewing experience, and increasing the likelihood of further interaction with the channel's content. The relevance and clarity of these textual elements not only influence initial viewer attraction but also shape the ongoing discourse in video comments, reflecting the emotional tone and sentiment of the audience. In this way, the strategic crafting of titles and descriptions becomes essential not just for attracting viewers but for fostering community engagement and navigating the competitive content creation environment on YouTube [31].

2.4. Tools and Methods Used in Sentiment Analysis

Sentiment analysis is a crucial tool for understanding viewer opinions, particularly on platforms like YouTube, where user comments and reactions offer valuable insights into audience behavior. Over time, various sentiment analysis tools have emerged, each with unique strengths and applications in processing online content. One widely used tool is the Valence Aware Dictionary and Sentiment Reasoner (VADER), which is especially effective in analyzing sentiments within social media content. Its rule-based approach allows it to accurately assess sentiment expressions in diverse and rich contexts, making it an ideal choice for studies that examine public sentiment in video comments related to critical topics such as health, including the COVID-19 pandemic [32].

Another prominent tool for sentiment analysis is TextBlob, which is favored for its simplicity and user-friendly interface. TextBlob is often applied to analyze sentiments in YouTube video comments, offering insights into audience opinions on various topics, including health and nutrition [33]. Its ability to discern sentiment polarity positive, negative, and neutral—makes it a valuable asset in evaluating viewer engagement and satisfaction. For more complex sentiment analysis, advanced tools like BERT (Bidirectional Encoder Representations from Transformers) are employed. BERT utilizes deep learning models to capture contextual nuances in language, allowing for greater precision in sentiment detection. Research utilizing BERT has demonstrated its effectiveness in processing intricate emotional expressions in comment sections, thus providing deeper insights into audience reactions and sentiments [34]. These methods highlight the adaptability of sentiment analysis tools in navigating the dynamic and multifaceted nature of online content engagement, particularly on platforms rich with user-generated text.

However, while social media platforms like YouTube provide valuable insights into public sentiment, they should not be the sole source for understanding public views. Limitations of social media analysis include the potential for echo chambers, information manipulation, and trends that fluctuate in response to public events, as well as the exclusion of individuals who may not actively participate in online discussions. Therefore, combining sentiment analysis from social media with other research methods, such as surveys and focus groups, is essential to gaining a comprehensive and accurate understanding of public sentiment.

3. Methodology

The diagram presented in Figure 1 illustrates the methodology used in this study, each step is essential for ensuring that the data is properly prepared and analyzed to derive meaningful results.

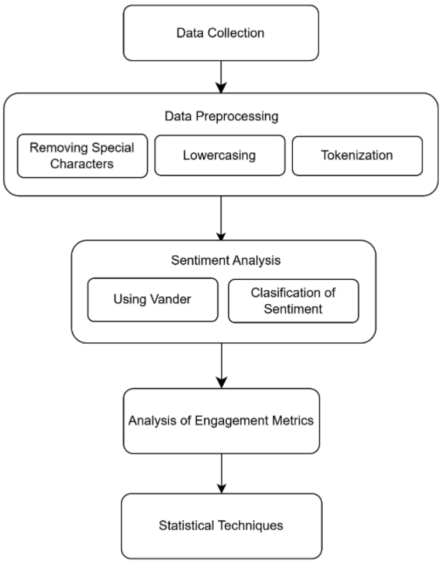

Figure 1. Research Methodology

3.1. Data Collection

The dataset used in this study consists of YouTube video data, specifically including the following key attributes for each video. This dataset provides a comprehensive snapshot of YouTube video characteristics and user engagement metrics, which are essential for understanding how content is perceived and interacted with by audiences. Table 1 summarizes the attributes in the dataset.

Table 1. Attributes in the Dataset

Attribute	Description
Video ID	A unique identifier assigned to each video on YouTube.
Title	The title of the YouTube video, providing an initial indicator of the content and tone.
Description	The textual description of the video, offering further context about the video content.
Publish Time	The timestamp when the video was uploaded, which helps track trends over time.
Video Type	Classification of video based on its duration (e.g., "short" or "long").
Duration	The length of the video in seconds, which may impact how users engage with the content.
View Count	The number of views the video has received, reflecting its level of audience interaction.
Like Count	The number of likes the video has received, indicating positive engagement.
Comment Count	The number of comments on the video, reflecting user feedback and engagement.
Thumbnail URL	A URL linking to the thumbnail image of the video.
Topic Categories	A list of categories or tags that the video belongs to (e.g., entertainment, lifestyle, etc.).

The videos included in the dataset were selected based on several criteria to ensure a diverse and representative sample of content on YouTube. By applying these criteria, the dataset captures a wide range of YouTube content with varying lengths, engagement metrics, and audience reactions, making it suitable for sentiment analysis and exploring the relationship between video characteristics and user sentiment. Table 2 summarizes the criteria used for selecting videos.

Table 2. Criteria for Selecting Videos

Criteria	Description
Video Type	Both short (< 1 minute) and long-form videos (> 10 minutes) were selected to examine how video length influences sentiment and engagement.
Engagement Metrics	Videos with a minimum threshold for view count (e.g., 10,000 views), like count (e.g., 500 likes), and comment count (e.g., 100 comments) to focus on videos with significant user interaction.
Content Categories	Videos selected from various categories (e.g., entertainment, fashion, lifestyle, sports, education) to explore sentiment across different content genres.

 ISSN 2776-8007

Publish Time	The dataset includes videos published within a specific time range (e.g., the last 12 months) to ensure the data is current and reflects recent trends.
Language	Only videos in English were included to standardize the sentiment analysis process and ensure consistency in linguistic analysis.

3.2. Data Preprocessing

Data preprocessing is a crucial step in preparing textual data for sentiment analysis, ensuring that raw YouTube data, such as video titles and descriptions, is cleaned and standardized before being analyzed effectively. The first step in the preprocessing process involves removing special characters, including punctuation marks, symbols, and numbers that are irrelevant to sentiment analysis. Special characters like !, @, #, and $ are typically removed to minimize noise and ensure that the text focuses on meaningful words, which are essential for accurately capturing sentiment.

The next step is converting the text to lowercase, ensuring that words such as "Great" and "great" are treated the same, making the analysis case-insensitive. This ensures uniformity in the text data and prevents inconsistencies. Stop words—common, non-essential words such as "the", "is", and "in"—are then removed. These words do not contribute meaningfully to the sentiment of the text and can clutter the analysis, so their removal helps reduce redundancy and highlights more significant words. Tokenization follows, where the cleaned text is split into individual tokens, typically words, which are then analyzed individually. This step allows for detailed sentiment analysis at the word level.

Furthermore, URLs and mentions (such as @user) are removed as they do not provide relevant information for sentiment analysis. The final cleaning step ensures extra spaces are removed, making the text uniform and consistent for further processing. Optionally, stemming or lemmatization techniques can be applied. These techniques reduce words to their root form (e.g., "running" becomes "run" and "better" becomes "good"), though this step is not always necessary, depending on the analysis requirements.

Once the text is cleaned and standardized, the next step is to merge the title and description of each video into a single textual feature. This ensures that the full context of the video content is captured, providing a more accurate sentiment analysis. If a video lacks a description, the title alone is used for sentiment analysis, and if both are missing, the video record is either removed or flagged for further action.

These preprocessing steps—text cleaning, tokenization, and merging video features—prepare the dataset for sentiment analysis by ensuring that the data is in the best possible format for generating accurate and meaningful results. By focusing on the entire content, including both the title and description, this approach ensures that sentiment analysis is based on the full context of the video rather than just one isolated element.

3.3. Sentiment Analysis

Sentiment analysis is a key technique used to determine the emotional tone behind a piece of text, particularly for understanding how audiences perceive content. In this study, sentiment analysis is applied to the title and description of YouTube videos to assess how users may feel about the content. The analysis utilizes VADER (Valence Aware Dictionary and Sentiment Reasoner), a sentiment analysis tool designed for social media and short-form text. VADER assigns a compound score ranging from -1 (extremely negative) to +1 (extremely positive), with a score of 0 indicating a neutral sentiment. This tool is particularly effective in processing the varied and nuanced language of video titles and descriptions on platforms like YouTube.

The sentiment of each video is determined by analyzing both the title and description using VADER, which calculates a compound score to classify the sentiment as positive, negative, or neutral. If the compound score is greater than or equal to +0.05, the sentiment is classified as positive, reflecting an optimistic tone. A score of -0.05 or less is considered negative, indicating dissatisfaction or negative emotions, while scores between -0.05 and +0.05 are categorized as neutral. For example, a video titled "Amazing Travel Adventures in Bali!" with a description like "Join me on an unforgettable trip to one of the most beautiful places on Earth!" would likely receive a positive sentiment score, while a video titled "I Lost Everything in a Fire!" would be classified as negative due to the distress conveyed.

This classification process provides valuable insights into how video content is emotionally perceived by viewers. By combining both titles and descriptions into a single feature for sentiment analysis, we can better understand the emotional engagement of users. This analysis can also help content creators and marketers assess the tone of their videos and make adjustments to improve engagement. Analyzing sentiment can reveal how the emotional tone of a video influences metrics such as view counts, likes, and comments, allowing creators to optimize their content to resonate more effectively with their audience.

3.4. Analysis of Engagement Metrics

After performing sentiment analysis on YouTube video titles and descriptions, the next step is to examine how sentiment correlates with key user engagement metrics such as view count, like count, and comment count. These metrics provide valuable insights into how viewers interact with content and reveal the relationship between the emotional tone of a video and its popularity. For example, view count serves as a primary indicator of a video's reach and visibility. Videos with positive emotional tones, such as excitement or happiness, often attract more viewers as they are more likely to be shared or recommended. In contrast, videos with negative sentiment may deter viewers and result in lower view counts, although controversial or provocative content might still garner higher viewership despite its negative tone. Similarly, the like count, representing positive engagement, tends to be higher for videos with positive sentiment, as viewers are more likely to express approval when content evokes favorable emotions.

The comment count is another critical metric that reflects deeper engagement. Videos with highly positive or controversial content tend to spark more comments, as viewers feel compelled to share their thoughts. Negative sentiment, while generally leading to fewer likes, can provoke active discussions in the comments, especially if the content is provocative or contentious. Videos with neutral sentiment may see fewer comments, as they do not evoke strong emotional reactions. By analyzing the relationship between sentiment and these engagement metrics, we can gain valuable insights into how emotional tones influence viewer behavior. Positive sentiment generally correlates with higher engagement, while negative sentiment may still drive interactions, especially if the content stirs controversy. These insights help content creators tailor their titles and descriptions to optimize viewer engagement by aligning with the emotional preferences of their audience.

3.5. Statistical Techniques

To analyze the relationship between sentiment and engagement metrics, various statistical methods are employed to quantify the strength and direction of the correlation between video sentiment and key engagement factors such as view count, like count, and comment count. One of the most common techniques used is correlation analysis, specifically Pearson's correlation coefficient, which measures the linear relationship between two numerical variables. The correlation coefficient ranges from -1 to +1: a positive correlation (close to +1) indicates that as sentiment becomes more positive, engagement metrics like likes or views tend to increase. Conversely, a negative correlation (close to -1) suggests that negative sentiment is associated with a decline in engagement. If the correlation coefficient is near zero, it implies that there is no significant relationship between sentiment and the engagement metrics. Pearson’s correlation is particularly useful for determining whether positive sentiment is correlated with higher engagement or whether negative sentiment tends to drive lower engagement.

4. Results and Discussion

4.1. Result

The chart in Figure 2 shows the distribution of sentiments across a collection of YouTube videos. The results indicate that the majority of the videos have a positive sentiment, with nearly 40 videos falling into this category, as shown by the large green bar. This suggests that positive emotions, such as excitement or enthusiasm, are commonly conveyed in the titles and descriptions of these videos. In contrast, the neutral sentiment category is represented by a smaller red bar, indicating a moderate number of videos with a balanced emotional tone. Lastly, the negative sentiment category, represented by the gray bar, has the least number of videos, indicating that relatively few videos have a negative emotional tone in their titles and descriptions. This sentiment distribution highlights the dominance of positive content in the analyzed video dataset, with fewer videos displaying neutral or negative sentiments.

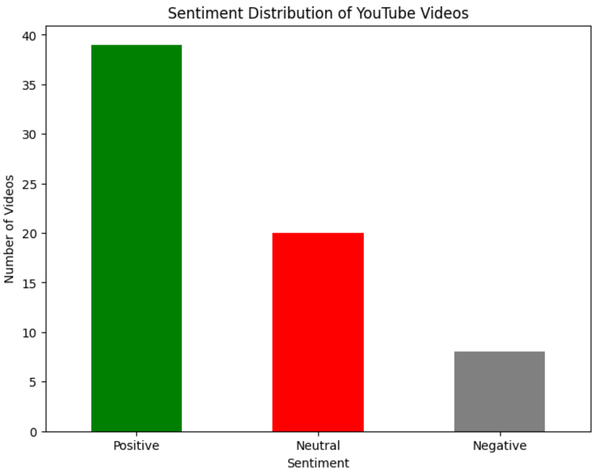

Figure 2. Sentiment Distribution of YouTube Videos

Figure 3 illustrates the relationship between sentiment and view or like count on YouTube videos. The scatter plot shows how videos with different sentiment classifications—positive, neutral, and negative—correlate with the number of views (x-axis) and likes (y-axis). From the plot, we can observe that positive sentiment (blue dots) is generally associated with higher engagement, both in terms of views and likes. These videos tend to have more substantial numbers of views and likes, indicating that content with a positive emotional tone is more likely to attract significant viewer engagement. Neutral sentiment (green dots) also appears to correlate with higher views and likes, but not as strongly as positive videos, suggesting a more moderate level of engagement. Negative sentiment (orange dots) is more scattered and typically shows lower engagement levels, particularly in terms of likes, indicating that negative content tends to receive less favorable interaction compared to positive or neutral content.

This visualization highlights that videos with a positive emotional tone generally have higher engagement metrics, both in views and likes, while videos with negative sentiment may face challenges in attracting similar levels of interaction. The pattern underscores the importance of sentiment in influencing viewer behavior and engagement on the platform.

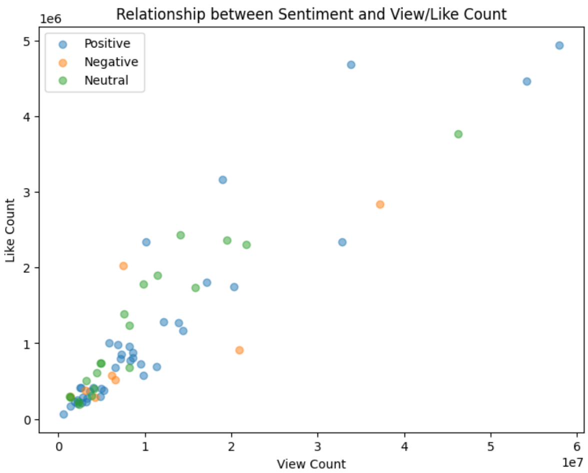

Figure 3. Relationship between Sentiment and View or Like Count

Figure 4 shows the relationship between video duration (on the x-axis) and view count (on the y-axis) for YouTube videos. From the scatter plot, we can observe that there is little to no clear correlation between the video duration and the number of views. Most of the videos in the dataset have relatively low view counts, regardless of their duration. However, there are a few outliers where videos with longer durations (up to around 2,500 seconds) have significantly higher view counts. These outliers suggest that certain long-duration videos are more successful in attracting views, but they are relatively rare compared to shorter videos.

The majority of the data points are clustered towards the lower end of the view count range, indicating that shorter videos generally do not attract high viewership, but there is no consistent trend suggesting that longer videos inherently lead to higher view counts. This scatter pattern suggests that factors other than video duration, such as content quality or topic relevance, may play a more significant role in determining video popularity.

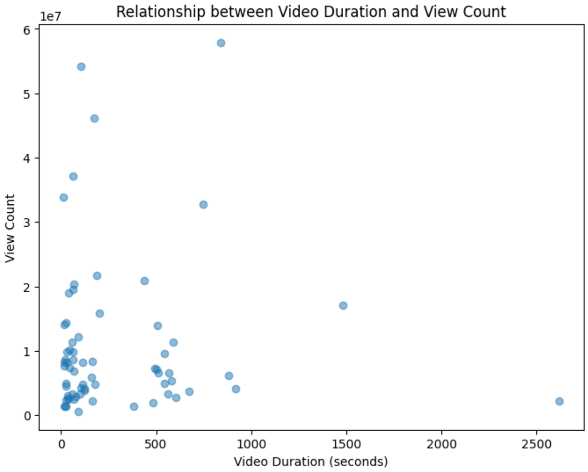

Figure 4. Relationship between Video Duration and View Count

Figure 5 illustrates the relationship between view count (on the x-axis) and like count (on the y-axis) for YouTube videos. The scatter plot shows a clear positive correlation between the two metrics. As the number of views increases, so does the number of likes, indicating that videos with higher viewership tend to receive more likes. Most of the data points are clustered in the lower to middle ranges for both views and likes, but there are a few outliers where videos with a significantly high number of views also attract a correspondingly high number of likes.

This pattern suggests that videos that are popular enough to receive a large number of views are more likely to have higher engagement, as evidenced by the increase in likes. The relatively consistent increase in likes as view count rises demonstrates that like count is a good indicator of positive engagement with video content. Overall, this visualization reinforces the idea that high viewer engagement in terms of views is closely tied to positive interactions such as likes.

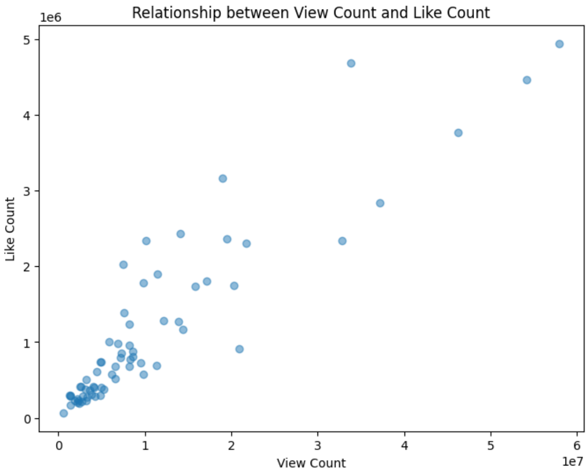

Figure 5. Relationship between View Count and Like Count

Figure 6 presents the view count distribution for long and short videos on YouTube. The box plot clearly indicates that long videos generally receive a higher range of views compared to short videos. The median view count for long videos is notably higher, with a wider interquartile range, suggesting that there is a greater disparity in view counts among long videos. Long videos also show a few outliers with exceptionally high view counts, indicating that some long videos achieve significantly more views than others. On the other hand, short videos have a more concentrated view count distribution, with the median being lower and fewer outliers present.

This distribution suggests that long videos tend to attract more views on average, but there is more variability in their performance. In contrast, short videos, while still attracting considerable views, show a more consistent level of engagement. This could imply that while long videos may have the potential for higher visibility, short videos might appeal to a more niche or consistent audience.

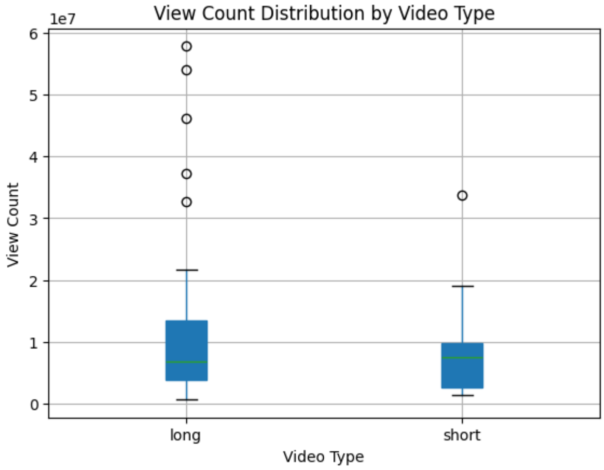

Figure 6. View Count Distribution by Video Type

Figure 7 illustrates the change in average view count over time from August to April, highlighting significant fluctuations in viewer engagement across the months. The plot shows a notable peak in November, where the average view count sharply rises, indicating a period of high video viewership. This spike is followed by a dramatic decline in December and January, where the average view count drops significantly. After January, the average view count continues to decrease gradually, indicating a slower but consistent decline in video views through the subsequent months.

The pattern suggests that there may be specific events or trends during November that caused a surge in views, while the subsequent drop in engagement could reflect seasonal changes or reduced content visibility. Understanding these patterns is essential for content creators to adjust their strategies based on time-based engagement trends.

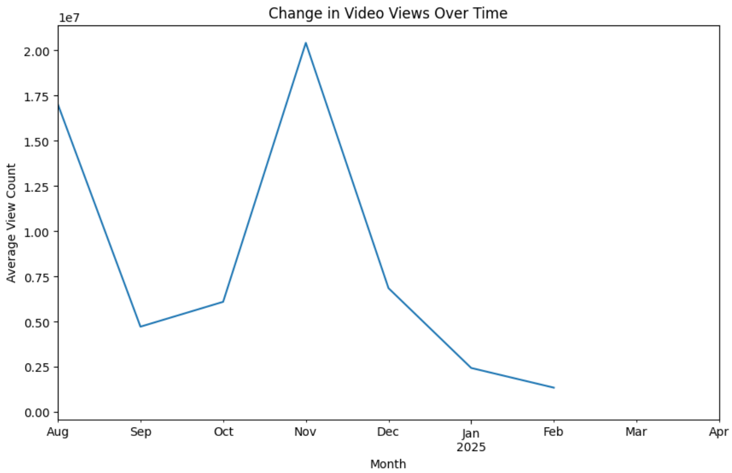

Figure 7. Change in Video Views Over Time

Figure 8 shows the relationship between like count and comment count for YouTube videos. The scatter plot reveals a clear positive correlation between the number of likes and the number of comments a video receives. As the number of likes increases, the number of comments also tends to increase, indicating that videos with higher engagement in terms of likes are more likely to provoke discussions and interactions in the comments section. This suggests that viewers who express approval through likes are also more inclined to engage with the content by leaving comments, thus fostering deeper interaction.

However, there are some videos with very high like counts and correspondingly high comment counts, representing outliers in the data. These outliers suggest that certain videos not only attract likes but also generate significant discussion, possibly due to their content's relevance or emotional impact. The overall trend indicates that positive engagement (in the form of likes) often leads to higher interaction levels, including comments, which are crucial for understanding viewer sentiment and engagement.

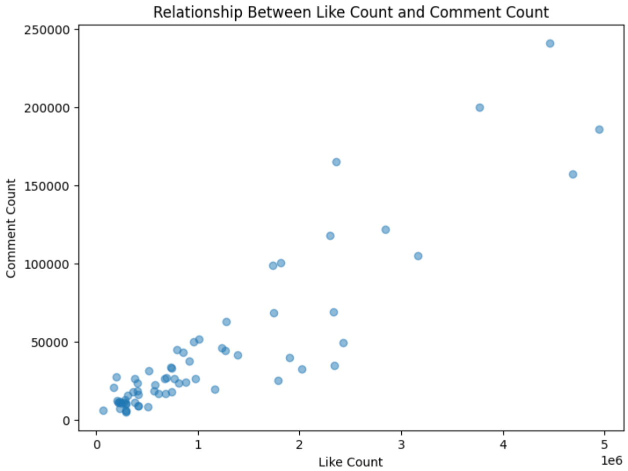

Figure 8. Relationship Between Like Count and Comment Count

Figure 9 presents a box plot of video duration by sentiment, displaying how the length of videos varies across different sentiment categories (positive, neutral, and negative). The plot shows that positive sentiment videos tend to have the longest durations, with a wider interquartile range, indicating greater variability in the length of videos classified as positive. The median for positive videos is noticeably higher compared to the other categories, suggesting that videos with positive emotional tones are more likely to be longer.

In contrast, neutral sentiment videos have the shortest median durations, and their interquartile range is also narrower, indicating that the duration of neutral videos is more consistent and tends to be shorter. Negative sentiment videos show a similar pattern to neutral sentiment, but they have slightly longer median durations, with a few outliers indicating that some negative videos are considerably longer. The overall trend suggests that positive sentiment is generally associated with longer videos, while neutral and negative sentiment videos tend to be shorter on average.

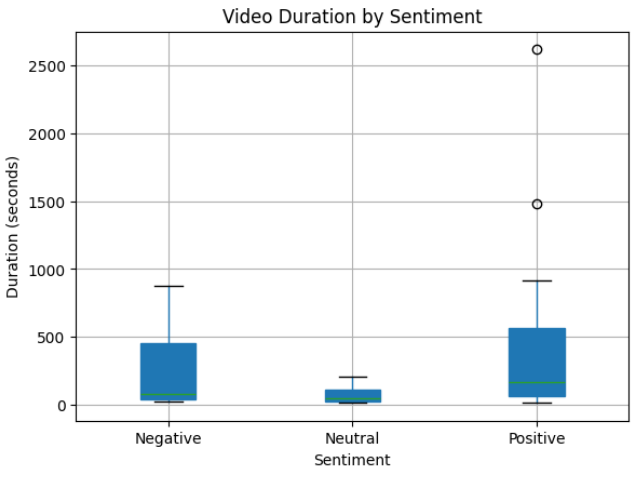

Figure 9. Video Duration by Sentiment

Figure 10 presents a bar chart of video counts by topic category and sentiment. The chart shows the distribution of YouTube videos across various topic categories, with the videos categorized into three sentiment groups: positive (green), neutral (orange), and negative (blue). The "Unknown" category stands out with a very high number of videos, most of which are categorized as positive sentiment, suggesting that a significant portion of videos in this category has a generally positive tone. Additionally, other categories like Association football, Food, and Hobby have a mix of all three sentiments, with positive sentiment being the most dominant across many topics.

Interestingly, categories like Entertainment, Sport, and Video game culture show more balanced distributions of sentiment, but still, positive sentiment is predominant in most cases. Neutral sentiment appears moderately across categories like Lifestyle (Sociology) and Society, while negative sentiment is less common overall, but still notable in some categories like Food and Video game culture. This chart highlights how sentiment varies across different topic categories on YouTube, with positive sentiments being most common, particularly in popular categories such as Hobby and Football.

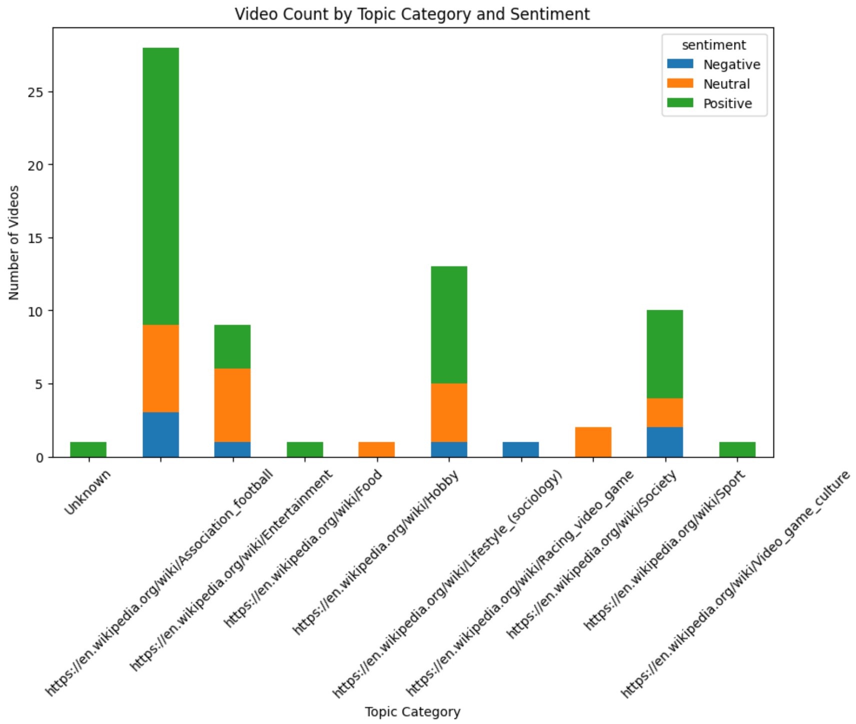

Figure 10. Video Count by Topic Category and Sentiment

Figure 11 presents the video topic trends over time, showing how the number of videos in different topic categories fluctuates month by month from August to April. The chart reveals significant peaks and drops in video counts across the various topics. For example, the "Unknown" category saw a sharp rise in October with over 10 videos, indicating a sudden surge in videos in this category, which then drastically decreased in subsequent months. In contrast, Association football (orange line) and Entertainment (green line) show gradual increases in video counts over time, with Entertainment reaching its peak in December.

Other topics, such as Food (red line) and Hobby (brown line), experienced fluctuating trends, with some months showing a higher concentration of videos. The Lifestyle (Sociology) category (gray line) and Racing video game (light green line) also had modest video counts that varied from month to month. This trend analysis highlights how different topics gain traction at different times, with certain categories like Entertainment and Association football showing consistent engagement, while others exhibit more erratic or niche engagement patterns. These insights can be valuable for understanding how content production aligns with audience interest over time.

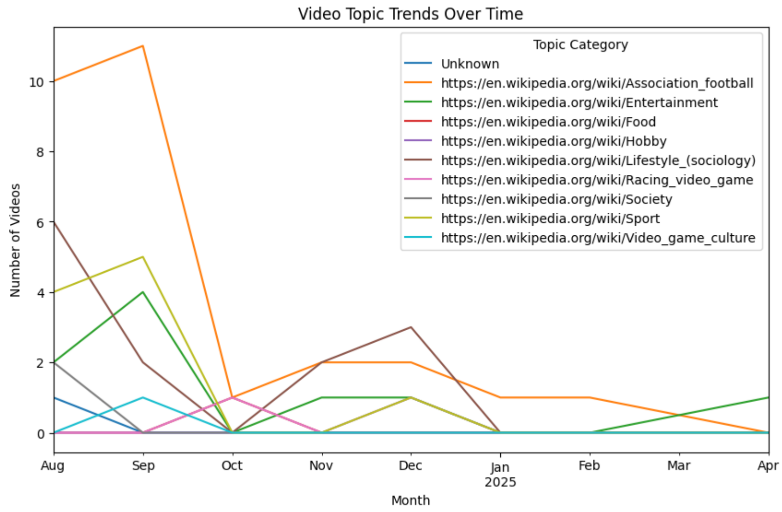

Figure 11. Video Topic Trends Over Time

Figure 12 shows the Top 10 Videos by View Count on YouTube, based on the video titles. The chart highlights the most viewed videos, with the video titled "I Meet MrBeast To Break The Internet!!" leading the list by a significant margin, closely followed by "The golden button... for my golden kids". These videos, along with others such as "Cristiano Ronaldo y Georgina revelan todo en el quiz de parejas | Cris & Gio Parte I" and "When I met the most famous Cristiano", demonstrate high engagement and attract millions of viewers.

The chart shows a clear pattern of videos with celebrity appearances or challenges (e.g., featuring MrBeast or Cristiano Ronaldo) receiving the highest view counts, reflecting their popularity and audience appeal. Videos with engaging, high-stakes titles like "How Many Goalkeepers Does It Take to Stop Ronaldo?" or "Georgina CALIFICA outfits de Cristiano Ronaldo ¡No creerás cuál es su favorito!" also perform well, attracting significant attention. Overall, the trend indicates that videos with compelling, curiosity-driven content and popular figures tend to achieve higher viewership.

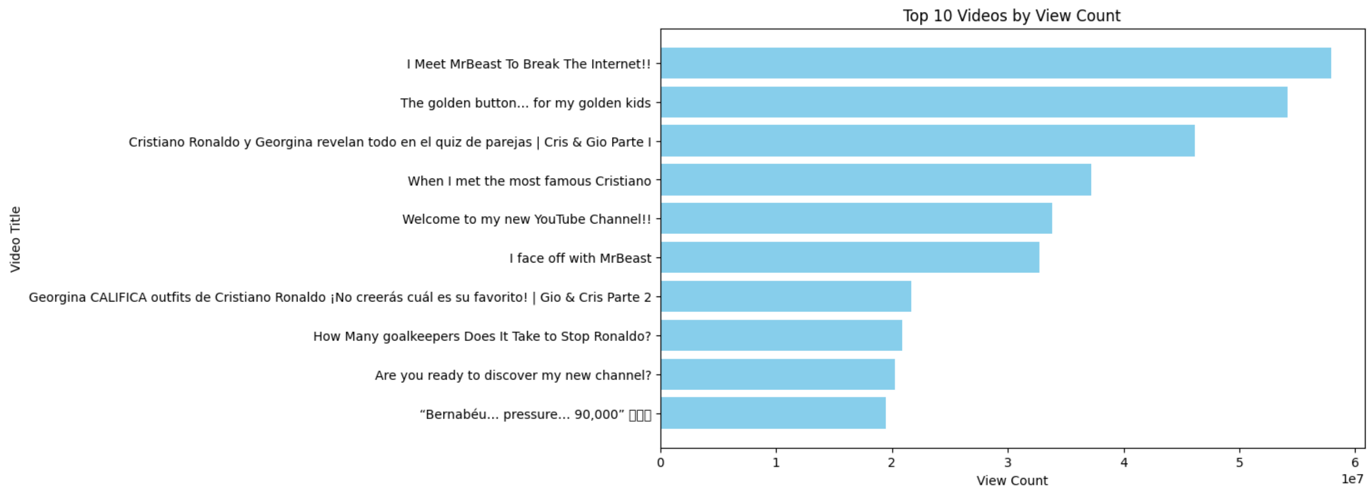

Figure 12. Top 10 Videos by View Count

4.2. Discussion

The findings from the sentiment analysis of YouTube video titles and descriptions provide valuable insights into the emotional engagement of users on the platform. The results indicate that positive sentiment dominates the video dataset, with a large proportion of videos conveying positive emotions, such as excitement or enthusiasm. This aligns with the broader understanding that content creators often use positive language to attract more viewers and foster greater engagement [28]. The neutral sentiment category, while less prevalent, still shows a noticeable presence, and negative sentiment is relatively rare. These results corroborate findings from previous studies that suggest positive content tends to be more successful in garnering attention and generating interaction [26].

The positive sentiment is further corroborated by the relationship between sentiment and engagement metrics, where videos with positive sentiment exhibit higher view and like counts. This reinforces the idea that emotionally resonant content encourages higher levels of user interaction, a finding consistent with research by Lee et al. [35] and Soares et al. [23]. Videos with negative sentiment, on the other hand, show lower engagement metrics, although some exceptions exist, suggesting that controversial content may still attract high viewership despite its negative tone. This mirrors earlier studies by Thakur et al. [25], which highlighted that negative sentiment can sometimes spark interest in controversial topics, yet the overall engagement remains lower compared to positive content.

Another important trend observed is that long videos tend to attract more views than short ones, which could imply that longer content is perceived as more informative or engaging. This aligns with previous research by Kurniawan et al. [28], which emphasized that video length plays a crucial role in determining user engagement. However, the relationship between video duration and view count is not always linear, and other factors, such as the relevance of the content and its ability to meet audience expectations, are likely more influential.

The change in video views over time reveals patterns of fluctuating engagement, with peaks in certain months and declines in others. This seasonal variation can be linked to external factors such as public events, viral trends, or even the timing of content releases. The finding supports previous research that identified shifts in user engagement based on calendar events or content virality [36]. Understanding these temporal patterns can help content creators and marketers tailor their strategies to align with periods of peak interest.

The relationship between like count and comment count emphasizes the importance of likes as an indicator of positive viewer engagement. Higher like counts tend to correlate with increased comments, reinforcing the idea that viewers who appreciate the content are more likely to engage in deeper interactions, thus fostering a community around the content. This relationship is consistent with findings by Donaldson et al. [29], who demonstrated that the emotional tone of videos influences the degree of viewer participation.

These findings have significant implications for content creators and marketers. Understanding the dominance of positive sentiment in driving engagement suggests that creators should focus on generating content with positive emotional tones to maximize user interaction, which is particularly useful for marketing campaigns and brand promotion, where positive emotional connections are crucial for building customer loyalty [28]. Additionally, creators should be mindful of video length and topic categories to tailor content to the interests and engagement preferences of their target audience. By strategically aligning video content with emotional tones and time-based trends, creators can enhance their reach and visibility, especially during peak months. Furthermore, the relationship between views and likes indicates that increasing viewership is an effective way to boost other engagement metrics. This insight is vital for understanding content virality, as higher view counts typically lead to more likes and comments. Recognizing these engagement patterns is key to crafting content that resonates with audiences and encourages sustained interaction.

This study provides a comprehensive analysis of sentiment in YouTube video titles and descriptions, offering a novel perspective on how emotional tones influence engagement. Unlike previous research, which primarily focused on user comments or isolated content elements, this study combines sentiment analysis with key engagement metrics such as views, likes, and comments, providing a more holistic understanding of user behavior. However, there are several limitations to this study. The sentiment analysis was based solely on video titles and descriptions, which may not fully capture the emotional tone of the entire video content. Future research could expand sentiment analysis to include the actual video content for a more accurate reflection of user emotions. Additionally, while this study focused on view and like counts, other engagement metrics such as shares, subscriptions, and watch time could provide a more comprehensive picture of user interaction. The sample used in this study was limited to a specific set of videos, and the results may not be generalizable across all types of content on YouTube. Further studies with larger and more diverse samples are needed to validate these findings across different genres and user demographics.

5. Conclusion

This study provides valuable insights into the relationship between sentiment in YouTube video titles and descriptions and user engagement metrics. The findings reveal that videos with positive sentiment generally attract higher engagement, including more views, likes, and comments, while negative sentiment correlates with lower interaction rates. This highlights the importance of creating content with a positive emotional tone to enhance visibility and user interaction. Additionally, the study underscores the significance of well-crafted titles and descriptions in driving engagement, as these elements influence viewers' expectations and emotional reactions. However, the study acknowledges limitations, as sentiment analysis was conducted solely on video titles and descriptions, potentially overlooking the full emotional tone of the actual video content. Future research could expand sentiment analysis to include the video content itself and incorporate additional engagement metrics like shares, subscriptions, and watch time for a more comprehensive understanding of user interaction. Despite these limitations, this study offers actionable insights for content creators to optimize their video strategies by aligning emotional tones with audience preferences, thus fostering greater engagement and content success.

6. Declarations

6.1. Author Contributions

Conceptualization: R.S., M.A., I.M.M.E.; Methodology: R.S., M.A.; Software: R.S.; Validation: M.A., I.M.M.E.; Formal Analysis: R.S.; Investigation: R.S.; Resources: M.A., I.M.M.E.; Data Curation: R.S.; Writing – Original Draft Preparation: R.S.; Writing – Review and Editing: M.A., I.M.M.E.; Visualization: R.S.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. SV, J. M. Lorenz, R. Ittamalia, K. Dhama, C. Chakraborty, D. V. S. Kumar, and T. Mohan, "Twitter-Based Sentiment Analysis and Topic Modeling of Social Media Posts Using Natural Language Processing, to Understand People's Perspectives Regarding COVID-19 Booster Vaccine Shots in India: Crucial to Expanding Vaccination Coverage," *Vaccines*, vol. 10, no. 11, hal. 1929, 2022, doi: 10.3390/vaccines10111929.

[2] T. Joseph, "Natural Language Processing (NLP) for Sentiment Analysis in Social Media," *Int. J. Comput. Eng.*, vol. 6, no. 2, pp. 35–48, 2024, doi: 10.47941/ijce.2135.

[3] M. Subathra and K. Meenakshisundaram, "Sentiment Analysis in Social Media for Depression Identification Using BiLSTM," *Asian J. Comput. Sci. Technol.*, vol. 11, no. 1, pp. 17–20, 2022, doi: 10.51983/ajcst-2022.11.1.3209.

[4] X. ShiXiao, M. M. Alobaedy, S. B. Goyal, S. Singla, S. S. Kang, and R. Chadha, "Real-Time Sentiment Analysis on Social Networks Using Meta-Model and Machine Learning Techniques," *Scalable Comput. Pract. Exp.*, vol. 24, no. 4, pp. 819–834, 2023, doi: 10.12694/scpe.v24i4.2247.

[5] J. Shen, "The Research of the Factors That Influence the Popularity of YouTube Videos," *Theor. Nat. Sci.*, vol. 51, no. 1, pp. 187–193, 2024, doi: 10.54254/2753-8818/51/2024ch0200.

[6] P. d. Bérail and C. Bungener, "Parasocial Relationships and YouTube Addiction: The Role of Viewer and YouTuber Video Characteristics," *Psychol. Lang. Commun.*, vol. 26, no. 1, pp. 169–206, 2022, doi: 10.2478/plc-2022-0009.

[7] A. Venus, D. Intyaswati, and F. Ayuningtyas, "Exploring Political Expression Among Indonesian Youth on YouTube: An

Investigation of Social Identity and Source Credibility," *Sage Open*, vol. 14, no. 2, pp. 1–11, 2024, doi: 10.1177/21582440241262647.

[8] X. Yu, M. Haroon, E. Menchen-Trevino, and M. Wojcieszak, "Nudging Recommendation Algorithms Increases News Consumption and Diversity on YouTube," *Pnas Nexus*, vol. 3, no. 12, pp. 1–11, 2024, doi: 10.1093/pnasnexus/pgae518.

[9] S. Parveen, A. G. Pereira, N. Garzon-Orjuela, P. McHugh, A. Surendran, H. Vornhagen, and A. Vellinga, "COVID-19 Public Health Communication on X (Formerly Twitter): Cross-Sectional Study of Message Type, Sentiment, and Source," *Jmir Form. Res.*, vol. 9, no. e59687, pp. 1–13, 2025, doi: 10.2196/59687.

[10] F. Mohamed and A. Shoufan, "Users' Experience With Health-Related Content on YouTube: An Exploratory Study," *BMC Public Health*, vol. 24, no. 1, pp. 1–16, 2024, doi: 10.1186/s12889-023-17585-5.

[11] M. Deori, V. Kumar, and M. K. Verma, "What News Sparks Interest on YouTube? A Study of News Content Uploaded by India's Top Five Hindi News Networks," *Online Inf. Rev.*, vol. 47, no. 3, pp. 550–566, 2022, doi: 10.1108/oir-01-2022-0007.

[12] G. C. Feng, Y. Luo, Y. Zhenwei, and J. Wen, "Effects of Rhetorical Devices on Audience Responses With Online Videos: An Augmented Elaboration Likelihood Model," *PLoS One*, vol. 18, no. 3, pp. 1–16, 2023, doi: 10.1371/journal.pone.0282663.

[13] R. Zhang and Q. Wu, "Impact of Using Virtual Avatars in Educational Videos on User Experience," *Sci. Rep.*, vol. 14, no. 1, pp. 1–16, 2024, doi: 10.1038/s41598-024-56716-9.

[14] R. A. Saunders, R. Crilley, and P. N. Chatterje-Doody, "ICYMI: RT and Youth-Oriented International Broadcasting as (Geo)Political Culture Jamming," *Int. J. Press.*, vol. 27, no. 3, pp. 696–717, 2022, doi: 10.1177/19401612211072771.

[15] H. Cela and G. Wood, "Sexualization and Emotional Valence in Audience Reactions to Popular Music Video Through Automated Language Analysis," *Z. Psychol.*, vol. 232, no. 3, pp. 181–190, 2024, doi: 10.1027/2151-2604/a000565.

[16] M. Coronel, "Destination Promotional Videos on Youtube: Assessing Audience Engagement," *Deturope - Cent. Eur. J. Tour. Reg. Dev.*, vol. 15, no. 1, pp. 47–65, 2023, doi: 10.32725/det.2023.003.

[17] S. Kothuri and N. R. Rajalakshmi, "MALO-LSTM: Multimodal Sentiment Analysis Using Modified Ant Lion Optimization With Long Short Term Memory Network," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 5, pp. 326–335, 2022, doi: 10.22266/ijies2022.1031.29.

[18] B. He, R. Zhao, L. Wang, and Y. Yang, "Aspect-Based Sentiment Analysis Based on Feature Extraction and Attention," *Acad. J. Sci. Technol.*, vol. 12, no. 1, pp. 192–198, 2024, doi: 10.54097/da7qvd31.

[19] M. U. Özmen and E. Yücel, "Drivers of Popularity of Online Information: Content, Context and Psychological Processes," *Bus. Econ. Res. J.*, vol. 14, no. 4, pp. 445–464, 2023, doi: 10.20409/berj.2023.425.

[20] Y. Deng, "Research on Sentiment Analysis Methods for Text-Oriented Data," *Front. Comput. Intell. Syst.*, vol. 3, no. 1, pp. 42–47, 2023, doi: 10.54097/fcis.v3i1.6022.

[21] S. R. Manmothe and J. R. Jadhav, "Integrating Multimodal Data for Enhanced Analysis and Understanding: Techniques for Sentiment Analysis and Cross-Modal Retrieval," *J. Adv. Zool.*, vol. 45, no. S-4, pp. 22–28, 2024, doi: 10.53555/jaz.v45is4.4144.

[22] A. A. Narasimham, "Sentimental Analysis on Tourism Reviews," *Interantional J. Sci. Res. Eng. Manag.*, vol. 08, no. 04, pp. 1–5, 2024, doi: 10.55041/ijsrem32368.

[23] F. B. Soares, I. Salgueiro, C. Bonoto, and O. Vinhas, "Youtube as a Source of Information About Unproven Drugs for Covid-19," *Brazilian Journal. Res.*, vol. 18, no. 3, pp. 462–491, 2022, doi: 10.25200/bjr.v18n3.2022.1536.

[24] J. Bisbee, M. Brown, A. Lai, R. Bonneau, J. Nagler, and J. A. Tucker, "Election Fraud, YouTube, and Public Perception of the Legitimacy of President Biden," *J. Online Trust Saf.*, vol. 1, no. 3, pp. 1–65, 2022, doi: 10.54501/jots.v1i3.60.

[25] N. Thakur, S. Cui, V. Knieling, K. Khanna, and M. Shao, "Investigation of the Misinformation About COVID-19 on YouTube Using Topic Modeling, Sentiment Analysis, and Language Analysis," *Computation*, vol. 12, no. 2, pp. 28, 2024, doi: 10.3390/computation12020028.

[26] M. Bal, A. G. K. Aydemir, and M. V Coşkun, "Exploring YouTube Content Creators' Perspectives on Generative AI in Language Learning: Insights Through Opinion Mining and Sentiment Analysis," *PLoS One*, vol. 19, no. 9, pp. e0308096, 2024, doi: 10.1371/journal.pone.0308096.

[27] S. Naz, S. Hina, U. Fatima, and H. Tabassum, "A Hybrid Approach to Measure Students' Satisfaction on YouTube Educational Videos," *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 09, pp. 131–147, 2023, doi: 10.3991/ijet.v18i09.38473.

[28] Y. Kurniawan, E. Halim, E. Jennifer, F. A. Pribadi, G. Bhutkar, and N. Anwar, "Understanding User Engagement Strategies

ISSN 2776-8007

for Podcasts Videos on Youtube in Indonesia: A Study on Content Creation," *Joiv Int. J. Informatics Vis.*, vol. 8, no. 2, pp. 957–967, 2024, doi: 10.62527/joiv.8.2.2123.

[29] S. Donaldson, A. Dormanesh, C. Perez, MO. Zaffer, A. Majmundar, JB. Unger, and J-P. Allem, "Monitoring the Official YouTube Channels of E-Cigarette Companies: A Thematic Analysis," *Heal. Educ. Behav.*, vol. 50, no. 5, pp. 677–682, 2023, doi: 10.1177/10901981221148964.

[30] B.-R. Moon and T. Oh, "A Study on the Characteristics of Sports Athletes' YouTube Channels and User Reactions," *Behav. Sci. (Basel).*, vol. 14, no. 8, pp. 700, 2024, doi: 10.3390/bs14080700.

[31] P. A. Toussaint, M. Renner, S. Lins, S. Thiebes, and A. Sunyaev, "Direct-to-Consumer Genetic Testing on Social Media: Topic Modeling and Sentiment Analysis of YouTube Users' Comments," *Jmir Infodemiology*, vol. 2, no. 2, pp. 1–16, 2022, doi: 10.2196/38749.

[32] K. Malhotra and P. Kempegowda, "Appraising Unmet Needs and Misinformation Spread About Polycystic Ovary Syndrome in 85,872 YouTube Comments Over 12 Years: Big Data Infodemiology Study," *J. Med. Internet Res.*, vol. 25, no. e49220, pp. 1–12, 2023, doi: 10.2196/49220.

[33] I. Chalkias, K. Tzafilkou, D. Karapiperis, and C. Tjortjis, "Learning Analytics on YouTube Educational Videos: Exploring Sentiment Analysis Methods and Topic Clustering," *Electronics*, vol. 12, no. 18, pp. 3949, 2023, doi: 10.3390/electronics12183949.

[34] R. Zhang and C. Ren, "Sentiment Time Series Clustering of Danmu Videos Based on BERT Fine-Tuning and SBD-K-shape," *Electron. Libr.*, vol. 42, no. 4, pp. 553–575, 2024, doi: 10.1108/el-10-2023-0243.

[35] J.-J. Lee, J. Kim, and S. Lee, "Trends of Fear and Anger on YouTube During the Initial Stage of the COVID-19 Outbreak in South Korea," *BMC Public Health*, vol. 24, no. 1, pp. 1–12, 2024, doi: 10.1186/s12889-024-19023-6.

[36] N. Zeid, L. Tang, and M. Amith, "The Spread of COVID-19 Vaccine Information in Arabic on YouTube: A Network Exposure Study," *Digit. Heal.*, vol. 9, no. October, pp. 1–13, 2023, doi: 10.1177/20552076231205714.