Düzce University
Journal of Science & Technology

Research Article

Why Sentiment Analysis-Based Metrics are Essential for Measuring Channel Performance on YouTube: An Experimental Study

Hakan Elbaş [a], Alparslan Mesri [b,*]

[a] *Department of Intelligent Systems Engineering, Graduate School, Bursa Technical University, Bursa, TURKEY*
[b] *Department of Industrial Engineering, Graduate School of Science and Engineering, Ankara, TURKEY*
** Corresponding author's e-mail address: alpmesri@gmail.com*
DOI: 10.29130/dubited.1190860

ABSTRACT

YouTube is a global social and economic platform that enables users to share videos. While many studies have conducted sentiment analyses on YouTube comments, none of these studies have focused on the necessity and potential significance of sentiment analysis-based metrics as essential or complementary tools that could serve as a "gold standard" for measuring the performance of both videos and channels. In this study, an experimental metric named the Sentiment Index (SI) was created to perform an experiment aimed at examining the robustness of sentiment analysis-based metrics against time parameter. To assess the degree to which this metric is affected by time parameter, a sentiment analysis experiment was conducted using VADER. The results of the experiment indicate that the Sentiment Index (SI) is influenced by time parameter to a marginal extent of 0.8%.

Keywords: *Natural language processing, Sentiment analysis, Data science, Youtube, Valence Aware Dictionary and sEntiment Reasoner (VADER), Social media*

YouTube'da Kanal Performansını Ölçmek İçin Neden Duygu Analizi Temelli Metrikler Gereklidir: Deneysel Bir Çalışma

ÖZ

YouTube, kullanıcıların videoları paylaşmasına olanak tanıyan küresel bir sosyal ve ekonomik platformdur. Birçok çalışma, YouTube yorumları üzerinde duygu analizi yapmış olsa da, bu çalışmaların hiçbiri duygu analizi tabanlı metriklerin videoların ve kanalların performansını ölçmede "altın standart" olarak hizmet edebilecek temel veya tamamlayıcı araçlar olarak ne kadar gerekli ve önemli olduğuna odaklanmamıştır. Bu çalışmada, duygu analizi tabanlı metriklerin zaman parametresine karşı dayanıklılığını incelemek amacıyla deneysel bir metrik olan Sentiment Index (SI) oluşturuldu. Bu metriğin zaman parametresinden ne kadar etkilendiğini değerlendirmek için VADER kullanılarak bir duygu analizi deneyi gerçekleştirildi. Deneyin sonuçları, SI'ın zaman parametresinden ihmal edilebilecek kadar, %0,8 etkilendiğini göstermektedir.

Received: 18/10/2022, Revised: 21/04/2023, Accepted: 17/09/2023

Anahtar Kelimeler: *Doğal dil işleme, Duygu analizi, Veri bilimi, YouTube, Valence Aware Dictionary and sEntiment Reasoner (VADER), Sosyal medya*

I. INTRODUCTION

Approximately 54% of internet traffic comprises video streaming, with YouTube accounting for approximately 15% of the global app data traffic [1]. Globally, YouTube ranks as the second-highest contributor to internet traffic [2]. In terms of financial performance, the platform generated a substantial $28.8 billion in advertising revenue in 2021 [3], with 55% of this revenue distributed to content producers [4]. The scale of this market and the remuneration allocated to content creators surpasses the GDP of many countries. Additionally, numerous brands have experienced augmented sales due to their YouTube content, leading to supplementary revenue streams beyond YouTube's advertising earnings. Acquiring a comprehensive understanding of the intricacies of the YouTube ecosystem is of paramount importance for channel owners and content producers, enabling them to optimize their presence and brand visibility. Hence, a meticulous exploration of the mechanisms inherent within the YouTube platform becomes a highly valuable endeavor.

YouTube employs various metrics to gauge the success of videos and channels. Among these metrics are indicators such as the count of likes, dislikes, views, subscribers, and view duration. These traditional metrics are subject to fluctuations over time, rendering them "time-dependent." Additionally, there exist derived metrics like the "like/dislike" ratio, derived from these traditional metrics. However, these derived metrics, despite their origin, retain the vulnerabilities inherent in traditional metrics. The purpose of this study is to elucidate the inherent advantages of sentiment-based metrics and advocate for their increased utilization vis-à-vis these existing metrics.

In the introduction section, the limitations of the current metrics will be examined, and the reasons why sentiment-based metrics are free from these limitations will be explained. The second chapter will comprise a comprehensive literature review. The third chapter will entail the execution of an experiment to apply sentiment analysis on a dataset, with its results intended for utilization in the subsequent chapter. Within the fourth chapter, an example metric tailored for this experiment will be devised to assess sentiment-based metrics against the time parameter, quantifying the extent to which this metric is influenced by time. Lastly, the fifth chapter will succinctly summarize the findings.

A. WEAKNESS OF TIME-DEPENDENT METRICS

- The phenomenon of videos gaining increasing viewership over time is a natural occurrence. It is anticipated that videos released by channels with a certain number of followers will experience a gradual rise in view count as time progresses. This temporal impact on video metrics poses difficulties when comparing the achievements of recently uploaded videos with those of a remote historical context.

- The phenomenon of YouTube user base growth: Another challenge in assessing video success arises from YouTube's remarkable expansion in user numbers in recent years. As reported in [5], YouTube's user count has nearly doubled over the past five years. To illustrate, consider a hypothetical scenario involving two videos from Channel X that share a similar theme and quality. Video A was uploaded 2 years ago, while Video B was released just 2 months ago. Upon comparing these videos, even if their "views" and analogous metrics 30 days post-release yield identical figures, it would be inaccurate to equate their performance. This discrepancy arises from the substantial variation in YouTube's user base between two years ago and the present day.

- The phenomenon of channel subscriber growth: The challenges associated with metric comparisons extend further. Consider a scenario where the count of YouTube users remains constant over time. Even under such circumstances, the hypothetical X channel could witness a rise in its subscriber base as time unfolds. In this context, juxtaposing the view counts of Video A and Video B, despite sharing the same theme and quality, would be ill-advised. The reason behind this is that as the subscriber count of Channel X escalates, the newly released video gains a competitive edge. In essence, the dynamic nature of subscriber numbers contributes to the complexity of comparing video performances.

B. WEAKNESSES OF METRICS DERIVED FROM TRADITIONAL METRICS: "LIKE/DISLIKE" METRIC

The "Like/Dislike" metric is derived by dividing two time-dependent traditional metrics. In this aspect, it is considered time-independent. However, due to its derivation from source metrics, it inherits certain weaknesses:

- Not every user has the habit of using the "Like" and "Dislike" buttons, and there may be differences in the tendency of user groups to interact with the video. Empirical studies have confirmed these differences [6]. For this reason, as an alternative to the "Like/Dislike" metric, a sentiment analysis-based metric consisting of positive and negative sentiments will be created after a sentiment analysis of the comments. For instance, as indicated by [6], YouTube users who seek self-status motivation tend to leave more comments while utilizing fewer 'like' buttons. This illustrates the extent to which a luxury leather brand that shares product-related videos requires sentiment analysis-driven metrics, either as a supplement to the 'Like' metric or potentially as an alternative. According to the same study, males were more likely to comment on YouTube videos in comparison with females. The findings indicated that among the factors influencing engagement on YouTube, the motive associated with seeking relaxing entertainment displayed the strongest correlation with liking and disliking videos, while the motivation for social interaction strongly predicted commenting and uploading activities. These results highlight that relying solely on the "like" metric to gauge the approval of customer segments driven by the social interaction motive may not be an accurate approach. Another finding indicating that the "like" metric is not highly sensitive is that individuals who were more experienced with YouTube were less likely to utilize the "like" feature on the platform.

- Viewers might hesitate to employ the 'dislike' button if a video they dislike is released by a favored channel. "Like" and "Dislike" buttons express opposing extreme judgments. These metrics do not capture audience sentiments falling between these two judgments. However, the sentiment analysis-based metrics allow for the expression of intermediate judgments.

- The metric displaying the number of dislikes for a video was hidden from users in November 2021 [7]. This alteration may have made a difference in users' "dislike" and "like" usage habits. A change in the behavior of YouTube users using the "dislike" button will make it difficult for publishers to compare videos before and after November 2021 using the "Like/Dislike" metric.

- Many groups provide fake "like" and "dislike" services on the Internet [8]. These groups can increase the number of likes and dislikes of a video, either by bots or by real people, for a fee. Consequently, in such scenarios, both the 'like' and 'dislike' metrics, including the 'Like/Dislike' metric, suffer a loss of measurement accuracy. Fake comments are easier to detect than fake likes and dislikes. Should a video's statistics become compromised or face suspicion of fake likes and dislikes, sentiment analysis-based metrics emerge as a viable complementary or alternative measure of the video's success.

In summary, most metrics do not work with full efficiency because the number of people using the YouTube platform changes, the number of subscribers of the channel changes, the number of views of the videos changes over time, etc. The "Like/Dislike" metric is time-independent, which is a solution to the time dependency problem, but it has also weak points.

II. LITERATURE REVIEW

Regarding YouTube metrics, Liikkanen[31] addressed the insufficiency of YouTube's public metrics and proposed three new metrics derived from existing ones in response. However, the limitations of the original metrics are not solely attributed to the inadequacies mentioned in that paper; there are other underlying factors. Consequently, the newly proposed metrics, derived from the original ones, might also inherit similar weaknesses. This study explored the weaknesses of the traditional YouTube metrics from different perspectives and employed sentiment analysis to develop a new metric. The study empirically evaluated the performance of the sentiment analysis-based metric concerning its time parameter.

Xiao et al. [32] developed a fuzzy mathematics-based method to assess user influence by integrating multiple metrics such as total view counts, comment counts, and likes on YouTube. The study ranked the most popular channels using this synthesized approach, which integrated multiple metrics such as total view counts, comment counts, and likes on YouTube. However, that paper did not explain the weaknesses of the current metrics and did not incorporate sentiment analysis of comments, which is considered in this study.

F. Poecze et al. [33] conducted a study on YouTube gamers and their Facebook communications. This study compared "traditional" metrics (like counts, comment counts, shares of posted content) with sentiment analysis results. The analysis revealed that sentiment analysis can detect follower negativity when user-generated activity tends to be relatively low. In some cases, it also showed significant positivity among content that did not receive popularity according to Facebook metrics. While this study shares similarities with the study conducted by F. Poecze et al. [33], there are some notable differences. That study identified different results between traditional metrics and sentiment analysis-based metrics but did not provide an explanation for the probable reason for these differences. Moreover, this study’s approach includes the evaluation of sentiment analysis-based metrics in the context of time parameter.

Previous studies have noted occasional discrepancies between YouTube's conventional metrics and sentiment analysis metrics. Bhuiyan et al. [19] conducted their study based on this phenomenon. When searching on YouTube, the listed videos are generally ordered based on metrics such as views and likes. However, this ranking performed without the utilization of sentiment analysis-based metrics may yield unintended results for users' query intents. Queries related to movie titles or IT problem solutions exemplify the potential challenge of unintended ranking results. This study demonstrates through an experiment that an approach involving sentiment analysis metrics can enhance query performance. Moreover, this research emphasizes the significance of sentiment analysis-based metrics as a complementary metric to conventional metrics for YouTube user queries. This study shares a similar core concept but is more comprehensive, focusing not on optimizing query results, but rather on measuring and optimizing video and channel performance.

On data scraping, Thomas & Mathur did a study and data mining on Reddit using the Python language and Scrapy library [9]. The Scrapy library performs data extraction using the XPath method, which is similar to the Selenium library used in this study. Mesri conducted a study aimed at classifying comments in a bank software and obtained the necessary data for this research by web scraping and utilizing the XPath method [17].

Tokcaer conducted a comprehensive literature review on Sentiment Analysis by examining 43 publications and 12 theses published in peer-reviewed journals over the past decade [10]. Alhujaili & Yafooz performed a literature review study, with a specific focus on sentiment analysis using YouTube comments [15]. The authors categorized the selected studies based on various variables, including the training data employed, the methodologies utilized, and the language of the comments. Similarly, [12] and [13] conducted literature reviews regarding sentiment analysis. These studies make valuable contributions by providing insights into the existing body of literature on sentiment analysis. Furthermore, these studies emphasize the relatively limited research on sentiment analysis of YouTube comments, with a predominant focus on Twitter data.

Despite numerous studies examining sentiment analysis or comment classification, based on YouTube[16,18, 20, 21, 22, 23, 25, 26, 27, 34, 35] and Twitter[11, 14, 24] data, to ascertain whether a solution has been proposed for the metric challenge within YouTube, no approach has been identified in this regard.

Hutto & Gilbert studied a new tool called Vader (for Valence Aware Dictionary for sEntiment Reasoning) [14]. In that study, a comparison was made between the Vader lexicon and eleven other highly regarded sentiment analysis tools, revealing that Vader outperforms these tools. In cases where the dataset environment for sentiment analysis to be estimated is social media, Vader obtains an overall F-1 score of 0.96. This exceptional score is higher than other ai tools, but even better than individual human raters with a 0.84 overall F-1 score for recognizing emotions. [34] conducted a sentiment analysis study on comments of mobile unboxing videos on YouTube. In their research, they utilized the Vader library to label the dataset and trained an SVM algorithm to perform sentiment analysis. [35] conducted a study on a mixed dataset from Twitter and YouTube to analyze public sentiment regarding COVID vaccines. However, this study did not provide an F1 score specific to the YouTube data analyzed using the Vader library.

There are practical applications developed using the Vader library. One of these applications is the Google Chrome extension called YouTube Comments Sentiment[1]. This software, designed for end-users, offers practicality but also has certain usage limitations. In this study, considering the relatively large amount of data, the main source files of the Vader library were used.

Numerous instances of sentiment analysis studies have been conducted on YouTube comments in the past. However, the insufficiency of conventional metrics on YouTube to measure video and channel success, as well as the necessity for sentiment analysis-based metrics, have not been deliberated in previous research. Furthermore, for the first time in this study, an experiment was conducted to assess the variation of a sentiment analysis-based metric in response to the time parameter.

III. SENTIMENT ANALYSIS

For this study, 19,668 comments were extracted from four distinct YouTube channels: "iJustin[2]", "Lydia Elise Millen[3]", "Travel Alone Idea[4]" and "Hannah Elise[5]". These channels encompass a diverse range of themes, including lifestyle, travel, and more. Vlog channels fall under the “People and Blogs” category, which is one of the six categories on YouTube. According to 2018 data [28], this category constitutes the largest portion of content on YouTube, accounting for 32% of the total.

The collected data consists of videos that were published within the following time frames:

[1] https://chrome.google.com/webstore/detail/youtube-comments-sentimen/dkfoidehmdpnpcplbcmenkfclahalihg
[2] https://www.youtube.com/@iJustine
[3] https://www.youtube.com/@lydiamillen
[4] https://www.youtube.com/@TravelAloneIdea
[5] https://www.youtube.com/@HannahElise

- August 22, 2021, to January 5, 2022, for the "iJustin" channel
- May 20, 2022, to June 14, 2022, for the "Lydia Elise Millen" channel
- April 7, 2022, to May 16, 2022, for the "Travel Alone Idea" channel
- April 4, 2022, to May 24, 2022, for the "Hannah Elise" channel

Researchers who want to repeat or improve the processes can access the relevant data from the link[6].

In this study, data extraction was performed using the Python programming language along with the Selenium library, which serves as a data mining tool for this language. The prediction process followed a lexical-based approach. For sentiment analysis, the study utilized Vader (Valence Aware Dictionary and sEntiment Reasoner), a lexical analysis tool, with the initial work on this tool conducted by [14]. Additionally, the Googletrans library was employed to identify the language of the comments.

Comments were scraped using a script developed with the Selenium library. The script required two inputs: 1) the video URL and 2) the Xpath address of comments. This process yielded four pieces of information: the commenter's name, the comment text, the number of likes received, and the comment date. For analysis purposes, only the 'comment' and 'comment date' information were utilized from this process.

The initial dataset lacked sentiment labels, and an imbalance was observed between positive and negative classes, with positive comments predominating. To assess the effectiveness of the prediction model, a subset of comments from the dataset was manually labeled by the researchers for sentiment analysis validation. Subsequently, the prediction results were utilized to calculate a Sentiment Index for each video. The collection of these sentiment indices will then be utilized to generate a comprehensive Sentiment Index Chart, providing insights into the sentiment distribution of the analyzed videos.

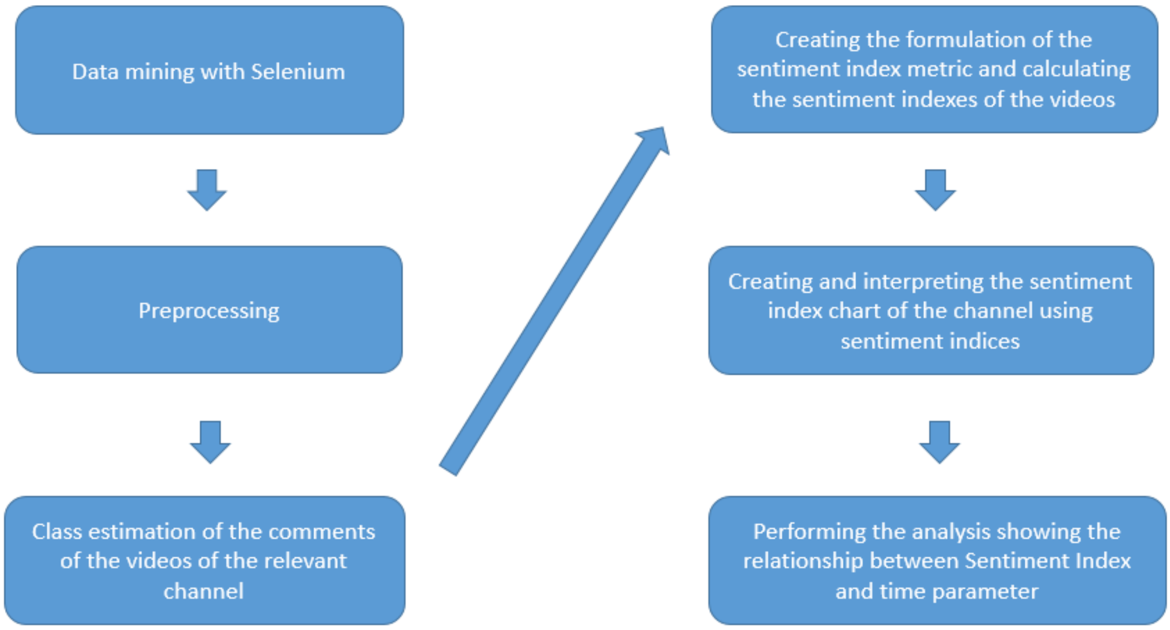

Figure 1. *Steps of the experiment.*

6 https://github.com/simplextable/Article_Sentiment_Analysis_with_YouTube_User_Comments_An_Alternative_Metric_for_YouTube_Analytics

While processing the extracted data, a crucial step is to clean the comments. As highlighted in [29], the preprocessing stage significantly impacts the success of sentiment analysis models. However, the Vader tool utilized in this study adopts a distinct approach. Conventional data processing tasks, such as converting words to lowercase, applying lemmatization, and eliminating numbers, punctuation, stop-words, and emojis, were intentionally omitted in order to optimize performance. Additionally, typos were not corrected to preserve slang words that might be lost otherwise. In this study, non-English sentences were detected and removed from the dataset, while URLs were eliminated during the data-cleaning phase.

A. SETTING UP THE MODEL

Two distinct approaches are commonly employed to address sentiment analysis challenges: lexicon-based and machine learning-based. As indicated in [30] and [13], there is minimal variance in their success rates. For this study, the lexicon-based approach was selected.

Comments were categorized into three classes: positive, negative, and neutral. To determine the sentiment polarity of comments, the Vader tool returns a compound value ranging from -1 to +1. Following the guidance provided in the library documentation, comments with values below -0.05 were designated as negative, those falling between -0.05 and +0.05 were assigned as neutral, and values exceeding +0.05 were marked as positive.

Individual F-scores were calculated for the positive and negative classes. These were then combined to generate a composite F-score that considers the class weights within the dataset. The computation of this metric involves utilizing the weighted parameter of the F1 score function in the Sklearn library.

The results were formed in Table 1 and Table 2.

Table 1. *Confusion matrix.*

		Prediction			
	Class No	**Neg**	**Poz**	**Neu**	**Total**
Actual	**Neg**	25	18	12	55
	Poz	3	109	17	129
	Neu	4	36	77	117
	Total	32	163	106	

Table 2. *Validation results.*

Model No	**Data Set**	**Accuracy**	**Precision**	**Recall**	**F Score**
1	YouTube	0.70	0.71	0.70	0,69

B. DISCUSSION OF MODEL RESULT

Upon reviewing Table 1, it is evident that the negative, positive, and neutral classes are presented sequentially. The table showcases the validation outcomes based on 301 randomly selected comments. Analysis of the Confusion Matrix results indicates that the model's prediction success follows a pattern: positive-labeled comments are predicted more accurately than neutral-labeled ones, while neutral-labeled comments exhibit better prediction than negative-labeled comments.

Validation results are depicted in Table 2. Accuracy, Precision, Recall, and F-score values exhibit close proximity. These values were computed considering class weights, resulting in an F-score of 0.69. In a separate study, [14] reported a high F-score of 0.96 for Vader in social Twitter-themed analyses. Their research further revealed that Vader yields F-scores of 0.61 for movie reviews, 0.63 for Amazon product reviews, and 0.55 for NY Times Editorials. Although YouTube comment data might initially share similarities with Tweets, the F-scores suggest distinct behavior for YouTube comments within the Vader library.

IV. SENTIMENT INDEX METRIC

The Sentiment Index (SI) was specifically devised for the purpose of serving as a representative metric within this experiment, aiming to demonstrate the potency of sentiment analysis-based metrics. Tailored formulations can be generated as per specific requirements. In this study, it serves as an illustrative instance to demonstrate the potential of sentiment analysis-based metrics.

Section A of the Introduction discussed the inadequacy of time-based metrics for comprehensive measurement. In Section B of the Introduction, the time-independent "Like/Dislike" metric available in YouTube Studio was introduced as a potential solution; nevertheless, its certain shortcomings were also acknowledged. The Sentiment Index, in contrast, overcomes these limitations. As demonstrated through empirical studies within the same section, specific customer segments opt to convey their perspectives through comments rather than merely utilizing the "Like" and "Dislike" buttons. A representative of sentiment analysis-based metrics, the Sentiment Index, which derives its data from comments, is introduced as a solution to avoid neglecting these viewers' viewpoints. As this section will demonstrate, the Sentiment Index experiences minimal temporal impact and remains relatively immune to manipulative actions like "Like" boosting. Unlike the binary nature of the "Like/Dislike" metric, the Sentiment Index offers viewers the flexibility to express sentiments through a range of values, transcending the constraints of 0 and 1. This inclusive approach accommodates viewers with nuanced levels of appreciation and empowers them to articulate their opinions more accurately.

The Sentiment Index is calculated by dividing the total number of comments predicted as positive by the model, by the sum of comments predicted as negative. Although it can be tailored as required, its formulation has intentionally been designed to bear a resemblance to the "Like/Dislike" metric found in YouTube Studio.

In order to maintain statistical accuracy, videos with fewer than 30 comments should not be indexed. Particular consideration has been given to ensuring that the selected videos for analysis contain a minimum of 30 comments.

In a channel, if P_{ij} represents the j-th positive comment predicted by the model in i-th video and N_{ij} represents the j-th negative comment predicted by the model in i-th video, SI_i (Sentiment Index of i-th video) can be calculated as (1):

$$SI_i = \frac{\sum_{j=1}^{n} P_{ij}}{\sum_{j=1}^{n} N_{ij}} \tag{1}$$

After predicting the sentiment of each comment as positive or negative, a Sentiment Index was computed for all the videos scraped from the channel. Subsequently, a Sentiment Index Chart was generated to visually represent the sentiment distribution within iJustin's channel.

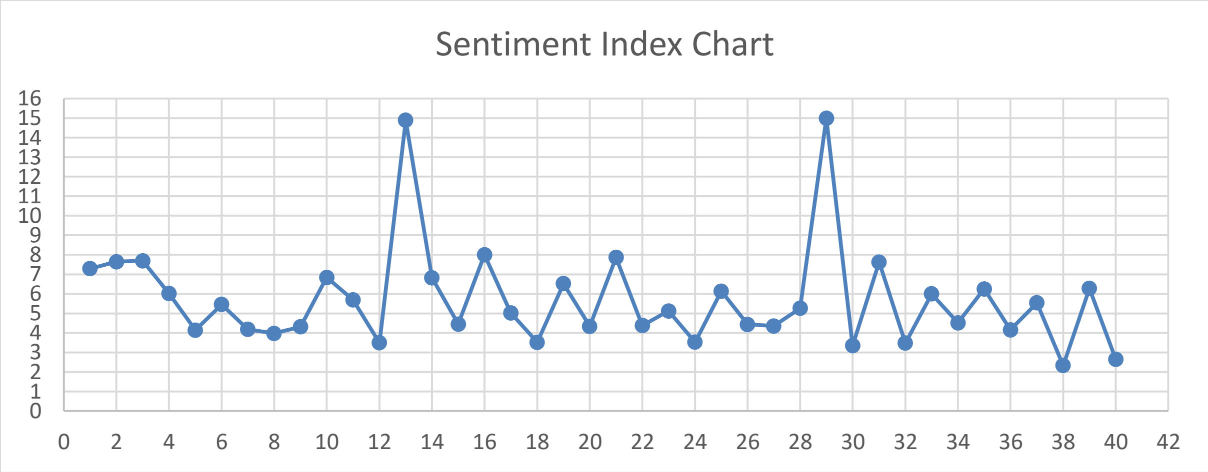

Figure 2. *Sentiment Index Chart.*

In Figure 2, the x-axis of the chart represents the video number. Video number 1 corresponds to the first video within the specified date range, while video number 40 represents the final video. While some video number labels on the x-axis are summarized due to space constraints, each individual video is depicted in the line chart. The y-axis displays the Sentiment Index associated with each respective video.

Upon reviewing the chart, several noteworthy observations emerge:

- Videos 13 and 29 exhibit the highest Sentiment Index, both reaching an approximate value of 15.
- Conversely, the video ranked 38th possesses the lowest Sentiment Index, registering a value of 2.33.
- Overall, Sentiment Index values for the videos cluster around the 5-point mark.

Content creators aiming to improve their video's Sentiment Index scores are advised to concentrate on analyzing videos that demonstrate notably high scores, such as videos 13 and 29.

When analyzing the video ratings, specific intuitive insights can be inferred:

- Travel-themed videos attract minimal negative comments, and the highest scores are associated with videos that introduce new destinations to viewers.

- In certain instances, the introduction of specific products (such as new model televisions) that lack the element of surprise due to societal familiarity with the technology can lead to a higher occurrence of negative comments.

- While there is no notable distinction in scores among travel-themed videos, score fluctuations may occur based on the product category in promotional videos. Therefore, it may be commercially advantageous for profit-oriented publishers promoting products to factor in the Sentiment Index score that the promoted product will receive when formulating pricing policies. This consideration is essential, as a product falling into a category that negatively impacts the Sentiment Index could potentially reduce the popularity of the publisher or channel.

A. AN EXPERIMENT TO MEASURE THE INFLUENCE OF TIME PARAMETER ON SENTIMENT INDEX METRIC

Previous sections have emphasized that a significant portion of the traditional metrics is influenced by the factor of time. In this section, an empirical experiment is conducted to determine the extent to which the Sentiment Index metric is affected by the passage of time.

To thoroughly investigate the relationship between the Sentiment Index and time, it is essential to tackle the following two questions:

1. How does the comment volume on videos change over time?
2. To what extent do Sentiment Index values vary for comments posted at different time points?

To address these inquiries and investigate the correlation between the Sentiment Index and time, an analysis was performed on a dataset.

Comments were categorized according to their creation dates. Those generated within a month following video publication were labeled as '1', while comments created one or more months later were labeled as '0'.

Several variables are defined as follows:

SI_A: Sentiment index of comments made within one month of the video's release,
SI_B: Sentiment index of the comments made one month after the release of the video,
SI_C: Sentiment index of comments made at any time after the video's release,
W_A: The proportion of the number of comments made within a month after the publication of the video to the total comments,
W_B: The proportion of the number of comments made one month after the release of the video to the total comments,

The relation between W_A and W_B is depicted in Figure 3.

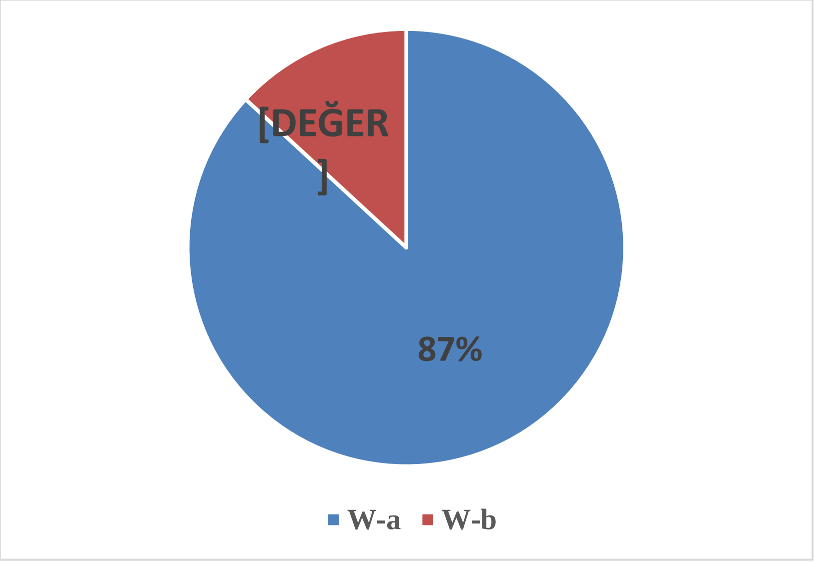

***Figure 3**. Distribution of comments by the time*

According to the results, W_A, 86.6% (17,036 comments) of the total 19,667 comments were made within 30 days of the videos' publication, while the remaining, W_B, 13.3% (2,631 comments) were made after 30 days. Thus, it can be inferred that a majority of the comments were posted shortly after the videos' release and this noteworthy phenomenon may provide valuable insights for future researchers.

In the second phase of this experiment, emotional indexes were determined for both the 86.6% (W_A) and 13.3% (W_B) segments.

According to the model's predictions, out of the comments made within a month after the release of the video (W_A), 10,086 comments were categorized as positive, 1,460 as negative, and 5,331 as neutral. When computed using the formula in (1), these outcomes yield a Sentiment Index (SI_A) of 6.97.

According to the model's predictions, out of the comments made one month after the release of the video (W_B), 1456 comments were categorized as positive, 197 as negative, and 959 as neutral. When computed using the formula in (1), these outcomes yield a Sentiment Index (SI_B) of 7.39.

Figure 4 illustrates the contrast between SI_A and SI_B, emphasizing the variation in Sentiment Indexes between comments generated within a month after the video's release and comments produced one month thereafter.

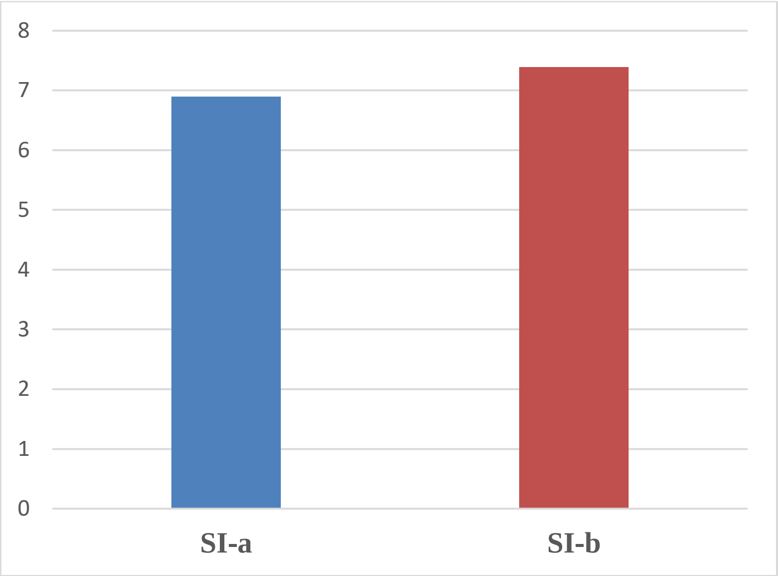

Figure 4. *Sentiment Indexes of different time intervals*

Based on these results, the difference between SI_A and SI_B is 0.5. However, to assess the impact of SI_B on SI_A, weights need to be taken into account. In this case, the magnitude of the difference between the effects of the Sentiment Indexes at two different times can be determined using equation (2).

$$SI_C = SI_A * W_A + SI_B * W_B \quad (2)$$

After utilizing the previously obtained results in equation (2) and calculating it, SI_C was found to be 6.97. By subtracting SI_A from SI_C to see how much the Sentiment Index of comments made within the month from the video release differs from the Sentiment Index of all comments made on the video, the equation (3) yields a value of 0.06.

$$SI_{Difference} = SI_C - SI_A \quad (3)$$

Using equation (4), the percentage of $SI_{Difference}$ relative to SI_C can be calculated. The obtained value is 0.008.

$$\text{SI metric's time effect rate} = | SI_{Difference} / SI_C | \quad (4)$$

According to the results obtained, it can be stated that the time parameter has a negligible effect on the Sentiment Index metric in the experiment conducted for a specific period on the four YouTube channels, with only a 0.8% impact.

V. CONCLUSION

This study delves into the YouTube platform, exploring the necessity for an alternative or supplementary metric and display mechanism within the platform for the comprehensive evaluation of video and channel performance. The conventional metrics such as the number of likes, dislikes, views, and subscribers, as well as derived metrics like the like/dislike ratio, are scrutinized to understand why they might fall short in providing a fully encompassing measurement framework. At this juncture, the discussion centers on the limitations of these metrics. Subsequently, the focus shifts to elucidating the distinct advantages of sentiment analysis-based metrics, which emerge as a solution free from the vulnerabilities inherent in traditional metrics.

To measure the degree of dependency of sentiment analysis-based metrics on the time parameter, an illustrative metric named the Sentiment Index (SI) was formulated, and an experiment was conducted using this metric. For the experiment, a dataset of 19,668 comments was collected from four distinct YouTube channels. These comments were divided into two groups based on different time intervals, and sentiment analysis was applied to each group. The analysis revealed that a significant portion of comments on a given video were made within the initial month of its publication (approximately 87%), while a relatively smaller proportion (approximately 13%) were made after one month. Furthermore, the Sentiment Index values for these two groups were examined, resulting in a SI of 6.91 for the first group and a SI of 7.39 for the second group. Utilizing these figures, a calculation indicated that the Sentiment Index's susceptibility to the time parameter is approximately 0.8%.

In conclusion, upon reviewing the outcomes derived from this experiment, it can be inferred that sentiment analysis-based metrics exhibit a certain resilience towards the time parameter and are minimally influenced by the passage of time.

The practical implications of this study underscore the need for an alternative or complementary metric and visualization mechanism within the YouTube platform to comprehensively evaluate video and channel performance. The process of comprehending the limitations of conventional metrics and recognizing their potential inadequacy in providing a holistic measurement framework yields a series of practical outcomes. The unique advantages offered by sentiment analysis-based metrics facilitate content creators and producers in gaining insights and establishing more meaningful engagements with their target audience. By demonstrating the resilience of sentiment analysis-based metrics to the temporal parameter, this study introduces a novel approach for content optimization and effective communication with the audience, offering content creators and brands a resourceful avenue to enhance video performance. In this context, the findings of this research will aid professionals and researchers striving to augment content performance on the YouTube platform by transcending the constraints of conventional metrics and appreciating the value of sentiment analysis-based metrics.

VI. REFERENCES

[1] Sandwine. (2022, June 7). *Global Internet Phenomena Report 2022 sandvine.com.* [Online]. Available:https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2022/Phenomena%20Reports/GIPR%202022/Sandvine%20GIPR%20January%202022.pdf (accessed June 7, 2022)
J. Jones. (1991, May 10). *Networks (2nd ed.)* [Online]. Available: http://www.atm.com

[2] (2022, June 10). *YouTube Ranking*, *similarweb.com*, [Online]. Available: https://www.similarweb.com/tr/website/youtube.com/#ranking.

[3] (2022, June 11). *Youtube's Advertising Revenues*, *statista.com*, [Online]. Available: https://www.statista.com/statistics/289659/youtube-share-of-google-total-ad-revenues/#:~:text=In%202021%2C%20YouTube's%20advertising%20revenue,dollars%20in%20the%20previous%20year

[4] M. Hall. (2022, July 14). *How Do People Make Money on YouTube*, *investopedia.com.* [Online]. Available: https://www.investopedia.com/ask/answers/012015/how-do-people-make-money-videos-they-upload
youtube.asp#:~:text=YouTube%20monetizes%20videos%20via%20pre,must%20adhere%20to%20advertising%20guidelines

[5] GMI. (2022, June 28). *YouTube Users Statistics 2022*. [Online]. Available: https://www.globalmediainsight.com/blog/youtube-users-statistics/

[6] M. L. Khan, "Social media engagement: What motivates user participation and consumption on YouTube?", *Computers in human behavior*, vol. 66, pp. 236-247, 2017.

[7] YouTube Team. (2022, Semtemper, 7). *An update to dislikes on YouTube*. [Online]. Available: https://blog.youtube/news-and-events/update-to-youtube/

[8] The Daily Iowan. (2022, May 13). *Ten Best Sites to Buy YouTube Likes dailyiowan.com* [Online]. Available: https://dailyiowan.com/2022/05/13/buy-youtube-likes 2/#:~:text=Buying%20a%20YouTube%20like%20is,give%20your%20channel%20a%20boost .

[9] D. M. Thomas and S. Mathur, "Data analysis by web scraping using python", in *2019* Proceedings of the *3rd International conference on Electronics, Communication and Aerospace Technology*, Coimbatore, India, 2019, pp. 450-454.

[10] S. Tokcaer, "Türkçe Metinlerde Duygu Analizi", *Journal of Yaşar University*, vol. 16, no. 63, pp. 1514-1534, 2021.

[11] R. Dehkharghani and C. Yilmaz, "Automatically identifying a software product's quality attributes through sentiment analysis of tweets", *IEEE 1st International Workshop on Natural Language Analysis in Software Engineering (NaturaLiSE*), San Francisco, Ca, USA, 2013, pp. 25-30.

[12] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies, practices and applications", *International Journal of Scientific and Technology Research*, vol. 9.2, pp. 601-609, 2020.

[13] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review", *Procedia Computer Science*, vol. 161, pp. 707-714, 2019.

[14] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text", in *Proceedings of the international AAAI conference on web and social media*, vol. 8, no. 1, 2014, pp. 216-225.

[15] R. F. Alhujaili and W. M. S. Yafooz, "Sentiment analysis for youtube videos with user comments", *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, India.

[16] H. C. Ko and W. N. Wu, "Exploring the determinants of viewers' loyalty toward beauty YouTubers: a parasocial interaction perspective", *Proceedings of ICEMT-2017*, Singapore, pp. 81-86.

[17] A. Mesri. (2017). *Web Mining and Sentiment Analysis of a Bank Software by User Comments*, [Online].
Available:https://www.academia.edu/72647117/Web_Mining_and_Sentiment_Analysis_of_a_Bank_Software_by_User_Comments

[18] A. Severyn, A. Moschitti, O. Uryupina, B. Plank, K. Filippova, "Multi-lingual opinion mining on YouTube", *Information Processing & Management*, vol. 52.1, pp. 46-60, 2016.

[19] H. Bhuiyan, J. Ara, R. Bardhan, R. Islam, "Retrieving YouTube video by sentiment analysis on user comment", *IEEE ICSIPA-2017*, Kuching, Malaysia, pp. 474-478.

[20] G. M. H. C. Gajanayake and T. C. Sandanayake, "Trending Pattern Identification of YouTube Gaming Channels Using Sentiment Analysis", *IEEE 20th ICTer-2020*, Colombo, Sri Lanka, 2020.

[21] R. Pradhan, "Extracting Sentiments from YouTube Comments", *IEEE Sixth of ICIIP-2021*, vol. 6, Shimla, India, pp. 1-4.

[22] H. Timani, P. Shah, M. Joshi, "Predicting success of a movie from youtube trailer comments using sentiment analysis", *IEEE 6th INDIACom-2019*, New Delhi, India, pp. 584-586.

[23] S. Singh and G. Sikka, "YouTube Sentiment Analysis on US Elections 2020", *IEEE 2nd ICSCCC-2021*, Jalandhar, India, pp. 250-254.

[24] X. Chen, M. Vorvoreanu, K. Madhavan, "Mining social media data for understanding students' learning experiences", *IEEE Transactions on learning technologies*, vol. 7.3, pp. 246-259, 2014.

[25] C. S. Lee, H. Osop, D. H. Goh, G. Kelni, "Making sense of comments on YouTube educational videos: a self-directed learning perspective", *Online Information Review*, vol. 41.5, pp. 611-625, 2017.

[26] C. Richier, E. Altman, R. Elazouzi, T. Altman, G. Linares, Y. Portilla, "Modelling view-count dynamics in youtube", *arXiv preprint*, arXiv:1404.2570, 2014.

[27] J. C. M. Serrano, O. Papakyriakopoulos, S. Hegelich, "NLP-based feature extraction for the detection of COVID-19 misinformation videos on YouTube", in *Proceedings of the 1st Workshop on NLP for COVID-19 2020*, online.

[28] (2022, June 11). *Distribution of total YouTube video content worldwide as of December 2018, by category statista.com*, [Online]. Available: https://www.statista.com/statistics/1026914/global-distribution-youtube-video-content-by-category/

[29] A. Krouska, C. Troussas, M. Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis", *IEEE 7th IISA*, Chalkidiki, Greece, 2016, pp. 1-5.

[30] C. Dhaoui, C. M. Webster, L. P. Tan, "Social media sentiment analysis: lexicon versus machine learning", *Journal of Consumer Marketing*, vol. 34, no. 6, pp. 480-488, 2017.

[31] L. A. Liikkanen "Three metrics for measuring user engagement with online media and a YouTube case study." *arXiv preprint*, arXiv:1312.5547, 2013.

[32] C. Xiao, Y. Xue, Z. Li, X. Luo, Z. Qin, "Measuring user influence based on multiple metrics on YouTube", *Seventh International Symposium on Parallel Architectures, Algorithms and Programming*, Nanjing, China, 2015, pp. 177-182

[33] F. Poecze, C. Ebster, C. Strauss, “Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts”, *Procedia computer science 130*, pp. 660-666, 2018.

[34] V. D. Chaithra, "Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments", *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, pp. 4452-4459, 2019.

[35] J. Al-Garaady and M. Mahyoob, “Public Sentiment Analysis in Social Media on the SARS-CoV-2 Vaccination Using VADER Lexicon Polarity”, *Humanities and Educational Sciences Journal*, pp. 591-609, 2022.