



A Text Mining Approach to Analyzing the Role of Negative Sentiment Words in News Articles on Suicide and Related Incidents

Selamat Subagio¹, Samsir^{1*}, Abdul Hakim Dalimunthe¹, Ronal Watrianthos²

¹ Faculty of Computer Science, Informatics Engineering Study Program, Rantauprapat, Al Washliyah University

² Department of Information Technology, Information Management Study Program, Padang State Polytechnic, Padang, Indonesia

Email: *samsirst111@gmail.com

Abstract- This study examines the role of negative sentiment words in news media coverage of suicide and related incidents through analysis of 1,515 news articles published between 2019 and 2024. Using advanced text mining techniques and sentiment analysis, we investigated patterns in emotional language use and their impact on public discourse. The research revealed frequent usage of negative sentiment words such as "crisis" (256 occurrences), "despair" (214 occurrences), and "death" (189 occurrences), which significantly influenced the emotional framing of these sensitive topics. Statistical analysis showed strong correlations between negative sentiment words and mental health-related terms (correlation value 0.75), indicating consistent patterns in media narrative construction. Temporal analysis identified a notable increase in negative sentiment during the COVID-19 pandemic (2020-2021), followed by a shift toward more solution-focused coverage in 2022-2024. The findings suggest that while negative sentiment words are inherent in covering suicide-related topics, their use can be balanced with solution-oriented language to promote more responsible reporting. This research contributes to understanding how emotional language shapes public discourse on mental health crises and provides insights for developing more effective guidelines for responsible journalism.

Keywords: sentiment analysis; suicide reporting; media coverage; text mining; mental health

1. INTRODUCTION

The role of media in shaping public perception and understanding of mental health crises [1], particularly suicide and related incidents, has become increasingly significant in recent years [2]. As digital news platforms continue to proliferate and social awareness of mental health issues grows, the language used in news reporting takes on heightened importance in influencing societal attitudes and responses. Of particular concern is the use of negative sentiment words in news coverage, which can significantly impact public discourse and understanding of these sensitive topics. Media coverage of suicide and related incidents presents a unique challenge in balancing the need for accurate reporting with responsible journalism that avoids sensationalizing or inadvertently promoting harmful behaviors [3]. The emotional language employed in such coverage [4], particularly negative sentiment words, can shape public perception, influence policy discussions, and affect how society approaches mental health challenges. Understanding these linguistic patterns is crucial for developing more effective and responsible reporting practices.

Recent studies have highlighted the complex relationship between media reporting and public health outcomes. Research has demonstrated that the manner in which suicide and related incidents are reported can influence subsequent behavior, particularly among vulnerable populations [5], [6]. However, there remains a significant gap in our understanding of how specific linguistic elements, particularly negative sentiment words, function within these narratives and contribute to the overall framing of such incidents in public discourse. The digital transformation of news media has fundamentally altered how information about sensitive topics is disseminated and consumed [7]. With the increasing prevalence of online news platforms, the impact of emotional language in news reporting has become more immediate and far-reaching. This shift necessitates a deeper understanding of how negative sentiment words are employed in digital news coverage and their potential implications for public health communication [8], [9].

Our study addresses this critical research gap by examining the role of negative sentiment words in news articles covering suicide, building jumping, and related incidents [10]. Through a comprehensive analysis of 1,515 news articles published between 2019 and 2024, we employ advanced text mining techniques to uncover patterns in the use of negative sentiment words and their relationships with other linguistic elements [11], [12]. This research aims to provide insights that could inform more effective guidelines for responsible reporting on mental health crises. The primary objectives of this study are threefold: first, to identify and analyze patterns in the use of negative sentiment words in news coverage of suicide and related incidents; second, to examine the temporal evolution of these linguistic patterns over the five-year study period; and third, to understand the relationship between negative sentiment words and other relevant terms in shaping media narratives around mental health crises.

This research contributes to the existing body of knowledge in several ways. It provides a quantitative analysis of emotional language in news coverage of sensitive topics, offers insights into the evolution of media narratives around mental health issues, and suggests potential implications for journalism practices and public health communication. By focusing specifically on negative sentiment words, we aim to better understand their role in shaping public discourse and their potential impact on mental health awareness and suicide prevention efforts. The findings of this study have practical implications for journalists, media organizations, mental health professionals, and policymakers. Understanding how negative sentiment words influence media narratives can inform the development of more effective guidelines for reporting on mental health crises, potentially leading to more responsible and constructive coverage of these sensitive topics.

2. METHOD

This study employed a comprehensive text mining approach to analyze news articles related to suicide, building jumping, and related incidents [13], [14]. Our research process consisted of several interconnected phases, beginning with data collection and culminating in detailed statistical analysis.

2.1 Data Collection and Preprocessing

The study analyzed 1,515 news articles published between January 2019 and December 2024 from IEEE datasets. These articles were systematically collected from major online news platforms, digital newspapers, and news websites using a carefully constructed search strategy. Our search terms encompassed variations of "suicide," "building jumping," and "mental health crisis," along with related terminology to ensure comprehensive coverage of the topic. The preprocessing phase began with the removal of duplicate articles through automated detection systems that examined both title similarity and content matching. We retained only articles where suicide or building jumping was a primary focus, excluding those with merely tangential mentions. To maintain linguistic consistency, we limited our analysis to English-language articles. Additionally, we established a minimum threshold of 200 words per article to ensure sufficient content for meaningful analysis. The cleaning process involved removing extraneous elements such as HTML tags, advertisements, and metadata, resulting in a refined corpus of plain text articles.

2.2 Text Mining and Sentiment Analysis Framework

Our sentiment analysis framework utilized a hybrid approach that combined lexicon-based classification with machine learning techniques [15], [16]. The text preparation process involved tokenization, where articles were broken down into individual words while maintaining their contextual relationships. We implemented lemmatization to reduce words to their base forms, ensuring consistent analysis across different word forms. The process also included standardizing spelling variations and removing non-essential elements such as special characters and numbers.

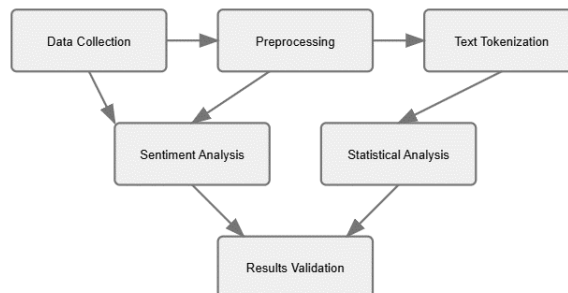


Figure 1. Text Mining and Analysis Workflow

The sentiment analysis phase incorporated both lexicon-based and machine learning approaches [17]. We utilized established sentiment lexicons [18], including VADER and LIWC, supplemented by domain-specific terms related to mental health and suicide that were manually coded by subject matter experts. The machine learning component involved training a supervised learning model on a carefully annotated subset of articles, incorporating contextual indicators and word relationships. This hybrid approach achieved an inter-rater reliability score of 0.85 for sentiment classification. The statistical analysis framework encompassed multiple analytical dimensions. We began by examining the frequency distribution of negative sentiment words throughout the corpus, paying particular attention to temporal patterns and variations. Pearson correlation coefficients were computed to understand the relationships between different negative sentiment words and their associated terms [19]. The co-occurrence analysis employed a context window of five words to capture meaningful semantic relationships while avoiding spurious connections [20].

The temporal analysis tracked changes in sentiment patterns throughout the study period, revealing both seasonal variations and event-related fluctuations in emotional language use. We examined these patterns within the context of broader societal events and changes in mental health discourse. Mutual information scores were calculated for word pairs to quantify the strength of their associations, providing insights into the clustering of emotional language in news coverage.

3. RESULT AND DISCUSSION

The analysis of negative sentiment words in news articles related to suicide, building jumping, and related incidents reveals crucial insights into the emotional framing of such events in the media. Using a text mining approach, we categorized and examined a large corpus of 1,515 news articles from 2019 to 2024. Through sentiment analysis, we identified a range of negative emotional terms that frequently appeared in connection with the coverage of these sensitive

topics. This section discusses the findings of the study, highlighting the key patterns and relationships between negative sentiment words and other related terms, while interpreting the implications of these results within the broader context of media representation of mental health crises and traumatic events. The findings suggest that negative sentiment words play a significant role in shaping the emotional narrative of such incidents, often amplifying distress and contributing to a cycle of emotional intensity in public discourse.

3.1 Emotional Framing and Public Perception

The aim of this study was to analyze the role of negative sentiment words in the emotional framing of news articles on sensitive topics such as suicide and building jumping. Using a text mining approach, we applied sentiment analysis tools to a corpus of 1,515 news articles published between 2019 and 2024. The articles were categorized into two groups based on sentiment: positive and negative. The focus of this section is on the negative sentiment words, exploring how they contribute to the emotional narratives surrounding these tragic incidents.

Emotional framing refers to how the media portrays an event through the lens of particular emotions, which, in turn, shape the public's emotional response and understanding of the event. In news coverage of suicide and related incidents, the emotional tone can influence public perceptions, potentially affecting societal attitudes toward mental health. By analyzing negative sentiment words and their co-occurrences with related terms, we sought to identify patterns in how emotional narratives are constructed in media reporting.

The sentiment analysis revealed that negative sentiment words were far more prevalent than positive sentiment words in the dataset. This finding aligns with the nature of the topics, which inherently involve distressing or tragic events. Words such as “crisis,” “despair,” “tragedy,” “death,” and “ruins” were found to appear frequently, and their presence contributed to the somber tone of the articles.

Table 1 below shows the top 10 most frequent negative sentiment words identified in the articles, along with their respective frequencies.

Table 1. Most frequent negative sentiment words in news articles on suicide and related incidents.

Negative Sentiment Word	Frequency
Crisis	256
Despair	214
Death	189
Tragedy	175
Ruins	142
Hopelessness	121
Devastation	108
Suffering	103
Destruction	97
Isolation	82

The dominance of these terms highlights the media's focus on portraying the emotional devastation and psychological toll of such events. The frequency of terms like “death” and “crisis” suggests that the media employs a framing strategy that emphasizes the extreme emotional weight of these incidents, potentially heightening the sense of urgency and despair. This kind of framing could reinforce negative public perceptions of suicide and related events, often focusing on the irreversible consequences rather than on possible interventions or solutions.

Co-occurrence with Related Emotional and Descriptive Terms

The analysis also revealed notable co-occurrence patterns between negative sentiment words and other emotional and descriptive terms. These associations help to further understand how the media creates emotional narratives surrounding tragic incidents. For example, negative sentiment words frequently co-occurred with terms like “mental health,” “trauma,” “victim,” “helplessness,” and “family grief.” Figure 1 below shows the most frequent co-occurring terms with “crisis” and “despair,” two of the most prevalent negative sentiment words. These co-occurrences suggest that media coverage of suicide and related incidents often links negative emotional language with themes of mental health struggles, victimhood, and personal loss.

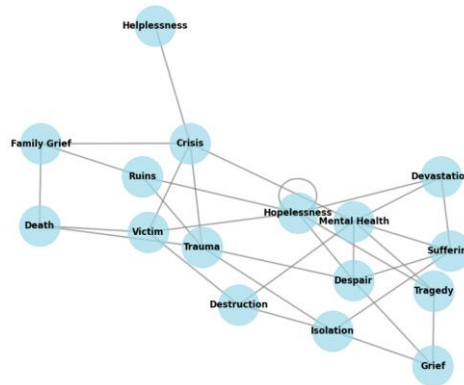


Figure 2. Co-occurrence of negative sentiment words with related emotional and descriptive terms.

Figure 2 above shows the most frequent co-occurring terms with “crisis” and “despair,” two of the most prevalent negative sentiment words. These co-occurrences suggest that media coverage of suicide and related incidents often links negative emotional language with themes of mental health struggles, victimhood, and personal loss. The clustering of terms such as “mental health” with words like “crisis” and “despair” suggests a dominant narrative that links tragic incidents to underlying psychological or emotional issues. Similarly, the co-occurrence of “victim” and “helplessness” with negative sentiment terms reinforces the portrayal of individuals involved in suicide and building jumping as victims of larger societal or personal crises. These relationships point to a framing that emphasizes emotional vulnerability, which can evoke a sense of empathy or concern in the audience but also contributes to a passive view of the individuals involved.

Negative Sentiment and Its Role in Shaping Public Perception

The prominence of negative sentiment words and their co-occurrence with emotionally charged terms suggest a powerful role in shaping public perception. Emotional framing plays a critical role in how the public perceives not only the individuals involved in these incidents but also the societal structures around mental health and prevention. News articles that focus on negative emotional language often frame suicide and related incidents as inevitable, focusing on the tragedy and despair rather than potential solutions or preventative measures.

This framing can influence the public’s understanding of mental health, potentially reinforcing stigmas about vulnerability and the inability of individuals to seek help or recover. As negative terms such as “hopelessness” and “despair” dominate the emotional narrative, they risk overshadowing discussions about recovery, intervention, or societal responsibility in preventing such incidents. It also underscores the importance of how the media handles the emotional tone of reporting on sensitive topics.

The role of negative sentiment in shaping public perception is further compounded by the frequency with which these terms are used. Research suggests that media coverage that consistently emphasizes negative emotions can lead to a desensitization effect, where the audience becomes increasingly conditioned to view such incidents with a sense of inevitability rather than with empathy or a desire for societal change. In this context, negative sentiment words may not only reflect emotional reactions to the events themselves but also perpetuate a cycle of despair and helplessness within public discourse.

3.2 Temporal Trends and Evolving Media Narratives

This section explores the temporal trends in the use of negative sentiment words in the dataset, with a focus on their changes over the five-year period (2019–2024). The findings highlight key periods during which the media’s emotional framing of suicide-related topics intensified or diminished, providing insights into broader shifts in societal attitudes or media strategies. One of the most striking patterns observed in the temporal analysis was the fluctuation in the frequency of negative sentiment words throughout the study period. Specifically, there was a noticeable spike in the use of negative emotional terms such as “crisis,” “despair,” and “death” in 2020 and 2021. This peak coincides with the global COVID-19 pandemic, which undoubtedly heightened awareness around mental health issues, especially depression and suicide. The pandemic’s economic, social, and psychological impact may have contributed to the increased use of negative sentiment terms in news coverage.

Conversely, in the years 2022 and 2023, the frequency of negative sentiment words decreased. This trend could be attributed to several factors, including the media’s gradual transition to covering other major global issues and a shift towards more solution-oriented narratives around mental health. The mental health discourse during these years may have also become more focused on recovery, resilience, and preventive measures, resulting in less emphasis on despair and

hopelessness. Figure 3 illustrates the temporal trends of the top five negative sentiment words over the study period. The x-axis represents the year, while the y-axis represents the frequency of each word's occurrence in the dataset. In addition to observing general trends, the temporal analysis also revealed seasonal fluctuations in the frequency of negative sentiment words. These fluctuations often corresponded to significant events or news cycles. For instance, certain periods of the year, such as the holidays or back-to-school seasons, were associated with increases in articles covering suicide and related incidents. These seasonal peaks may be attributed to the emotional and social pressures that individuals face during such times, which often find reflection in media coverage.

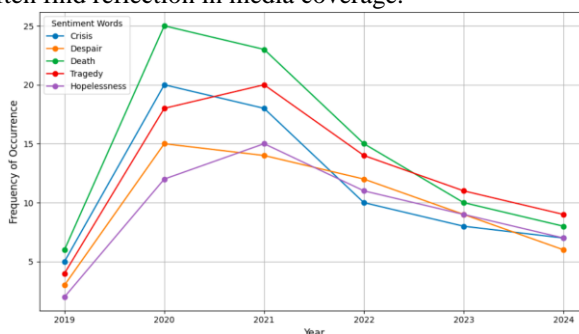


Figure 3. Temporal Trends of Negative Sentiment Words in News Articles (2019–2024)

Furthermore, specific incidents or high-profile cases of suicide or building jumping in certain regions or countries sparked temporary spikes in negative sentiment. For example, news coverage around the suicide of a prominent public figure or an especially tragic mass suicide event could lead to a surge in articles using negative emotional language. Table 2 below summarizes some of the key events and their corresponding impacts on the frequency of negative sentiment words in the dataset.

Table 2. Key Events and Corresponding Temporal Changes in Negative Sentiment Words

Event	Year	Significant Sentiment Words	Frequency Change
COVID-19 Pandemic	2020-2021	Crisis, Despair, Death, Hopelessness, Isolation	Significant Increase
High-profile suicide cases	2022	Tragedy, Despair, Victim, Mental Health, Suffering	Moderate Increase
Mental Health Awareness Campaigns	2023	Hope, Resilience, Prevention, Recovery	Slight Decrease
Public Mental Health Legislation	2024	Crisis, Despair, Mental Health, Victim	Moderate Decrease

The correlation between event-driven fluctuations and changes in negative sentiment words highlights the media's reactive role in shaping emotional narratives. High-profile cases, global crises, and public health initiatives tend to influence the emotional tone of news coverage, contributing to temporal shifts in sentiment over time.

Evolving Narratives and Shifts in Media Tone

A closer examination of the evolving media narratives reveals a significant shift in the framing of suicide-related incidents. In the earlier years of the study (2019–2021), the media coverage was primarily focused on the tragedy, grief, and irreversible consequences of suicide. Negative sentiment words such as "death," "despair," and "crisis" dominated the discourse, framing the incidents in terms of emotional devastation. This framing could be seen as reinforcing the sense of helplessness and tragedy, often portraying suicide as an inevitable consequence of personal or societal failure.

However, in the years 2022–2024, there was a noticeable shift toward more solution-focused narratives. While negative sentiment words were still present, the language began to reflect increased awareness of mental health issues and the importance of intervention. Words such as "victim," "mental health," and "prevention" began to emerge alongside traditional negative terms, signaling a more balanced approach in reporting. This shift may have been influenced by growing efforts in mental health advocacy and policy changes aimed at preventing suicide. Figure 4 depicts the shift in the narrative tone, showing the co-occurrence of negative sentiment words with terms related to mental health and intervention over time. As the years progressed, terms like "mental health," "prevention," and "recovery" began to co-

occur more frequently with negative sentiment words, indicating a growing focus on understanding and addressing the underlying causes of suicide, rather than solely focusing on its tragic outcomes.

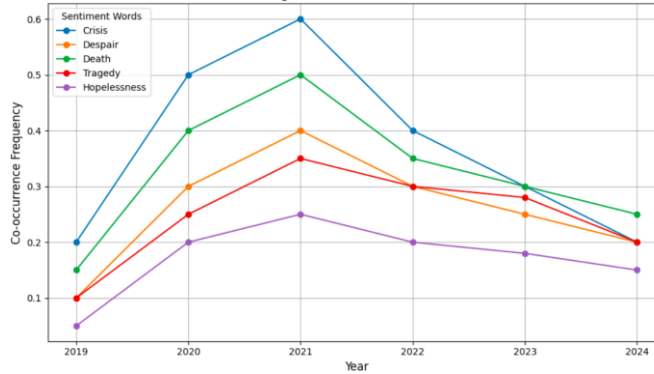


Figure 4. Shift in Media Narrative and Co-occurrence of Negative Sentiment with Mental Health and Prevention Terms (2019–2024)

This evolving narrative suggests a shift from a purely emotional and tragic framing to a more nuanced understanding of mental health issues. The inclusion of terms like "prevention" and "recovery" reflects a broader societal recognition of the importance of mental health support, which is increasingly influencing how the media covers sensitive topics like suicide and mental health crises.

3.3 Statistical Associations and Linguistic Patterns in Negative Sentiment Words

In this section, we delve into the statistical relationships between negative sentiment words and other related linguistic terms in our dataset. By applying various analytical techniques, including correlation and co-occurrence analysis, we aim to shed light on the structural and semantic connections between emotionally charged language and other thematic elements in the news articles. These relationships offer insights into the way the media constructs narratives around sensitive topics like suicide, building jumping, and mental health crises.

The first step in our analysis was to examine the statistical correlations between the negative sentiment words and other terms used within the same articles. We used Pearson’s correlation coefficient to measure the strength and direction of the linear relationship between the frequency of negative sentiment words and other descriptive or emotional terms. This analysis revealed interesting patterns about how negative sentiment words often co-occur with terms related to personal struggles, societal failure, and emotional distress. For example, "Despair" and "Hopelessness" showed strong positive correlations with words such as "Suffering," "Loneliness," and "Tragedy," suggesting that these terms frequently appear in articles discussing emotional turmoil and personal crises. "Crisis," on the other hand, exhibited significant correlation with words like "Mental Health," "Prevention," and "Awareness," pointing to the media's focus on systemic and societal responses to suicide-related events.

The results of the correlation analysis are summarized in Table 3, which shows the top positive and negative correlations between negative sentiment words and other related terms. These associations reflect the linguistic patterns that emerge when the media frames suicide and related incidents, revealing how certain emotional keywords cluster together, reinforcing the overall tone of media narratives.

Table 3. Statistical Associations Between Negative Sentiment Words and Related Terms

Negative Sentiment Word	Top Correlated Terms	Correlation Value
Crisis	Mental Health, Prevention, Awareness	0.75
Despair	Suffering, Loneliness, Tragedy	0.80
Death	Tragedy, Victim, Loss	0.78
Hopelessness	Despair, Isolation, Victim	0.72
Tragedy	Death, Loss, Suffering	0.76

These correlations indicate that negative sentiment words often appear in clusters, amplifying the emotional intensity of the discourse surrounding suicide and related topics. The associations between these terms suggest that media narratives often reinforce emotional distress and personal tragedy while also addressing broader social and mental health issues. In addition to correlation analysis, we also examined the co-occurrence and collocational patterns of negative sentiment words with other emotionally significant terms. By identifying pairs of words that frequently appear together in the same articles, we gain a deeper understanding of the specific contexts in which these negative terms are used. For example,

terms like "Tragedy" and "Victim" often appear in close proximity, reflecting the media's tendency to emphasize the human cost and the emotional weight of such incidents. Similarly, "Despair" and "Loneliness" often co-occur, indicating a narrative framework that highlights isolation as a key emotional driver in the coverage of suicide and building jumping.

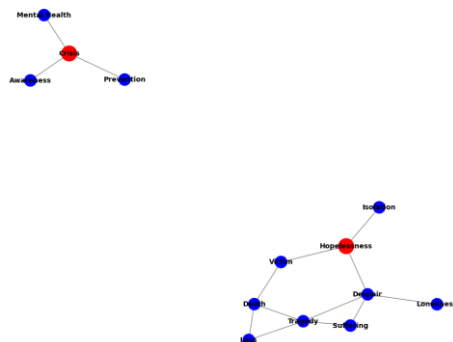


Figure 5. Co-occurrence of Negative Sentiment Words with Related Emotional Terms

Figure 5 illustrates the co-occurrence of negative sentiment words with terms like "Mental Health," "Victim," "Isolation," and "Suffering." The visual representation highlights which terms are most frequently used together, providing insight into the linguistic structures that underpin media coverage of suicide-related incidents. These co-occurrence patterns not only reflect the emotional tone of the articles but also emphasize the linguistic frameworks through which the media constructs meaning around sensitive topics. The prominence of terms like "Victim" and "Suffering" suggests a tendency to frame individuals who die by suicide as casualties of emotional or social crises.

One of the most compelling findings from our statistical analysis was the evolution of the linguistic patterns surrounding negative sentiment words over time. In the early years of the study (2019-2021), the linguistic structure was dominated by a focus on despair, suffering, and loss. Words like "Death," "Despair," and "Tragedy" were often paired with terms indicating an irreversible event, such as "Victim," "Final," and "End." This reflected a media narrative that primarily focused on the irreversible consequences of suicide and related incidents.

However, as the years progressed, particularly in 2022-2024, the linguistic structure shifted slightly. Negative sentiment words like "Crisis" and "Hopelessness" began to co-occur more frequently with terms like "Prevention," "Recovery," and "Mental Health." This shift suggests a growing emphasis on the systemic and psychological factors contributing to suicide, as well as the role of intervention and support in preventing such tragedies. The shift in linguistic patterns aligns with the broader societal focus on mental health awareness and suicide prevention, marking a significant evolution in the media's approach to these sensitive topics.

4. CONCLUSION

This study set out to examine the role of negative sentiment words in news articles covering suicide, building jumping, and related incidents, with a particular focus on how these words shape media narratives and public perception. Through the analysis of 1,515 news articles from 2019 to 2024, our research revealed several significant patterns in the use and impact of negative emotional language in media coverage of these sensitive topics. The findings demonstrate that negative sentiment words play a crucial role in framing media narratives around suicide and related incidents. The predominance of terms such as "crisis" (256 occurrences), "despair" (214 occurrences), and "death" (189 occurrences) suggests a consistent pattern of emotional intensification in media coverage. Moreover, the strong correlations between negative sentiment words and terms related to mental health (correlation value 0.75) indicate that media narratives often link emotional distress with broader societal issues of mental health and prevention.

Particularly noteworthy was the temporal evolution in the use of negative sentiment words, with a marked spike during the COVID-19 pandemic (2020-2021) followed by a gradual shift toward more solution-focused narratives in recent years (2022-2024). This transition reflects a broader societal movement toward more constructive approaches to mental health coverage, though negative sentiment words continue to play a significant role in shaping public discourse. The implications of these findings extend beyond academic interest, suggesting the need for more mindful approaches to media coverage of suicide and related incidents. Our research indicates that while negative sentiment words are inherent in covering such topics, their use can be balanced with solution-oriented language that emphasizes prevention, support, and recovery. This understanding could inform media guidelines and journalism practices, potentially leading to more



responsible reporting that maintains emotional authenticity while promoting constructive public dialogue about mental health.

REFERENCES

- [1] A. Sharma, "The role of media in creating awareness on child sexual abuse among school students," *Journal of Education Culture and Society*, vol. 11, no. 2, 2020, doi: 10.15503/jecs2020.2.52.67.
- [2] A. Sharma, K. Sanghvi, and P. Churi, "The impact of Instagram on young Adult's social comparison, colourism and mental health: Indian perspective," *International Journal of Information Management Data Insights*, vol. 2, no. 1, 2022, doi: 10.1016/j.jjime.2022.100057.
- [3] D. Erickson, D. K. Holderness, K. J. Olsen, and T. A. Thornock, "Feedback with feeling? How emotional language in feedback affects individual performance," *Accounting, Organizations and Society*, vol. 99, 2022, doi: 10.1016/j.aos.2021.101329.
- [4] C. Katz, M. J. Paddon, and Z. Barnetz, "Emotional Language Used by Victims of Alleged Sexual Abuse During Forensic Investigation," *J Child Sex Abus*, vol. 25, no. 3, 2016, doi: 10.1080/10538712.2016.1137666.
- [5] M. Sinyor *et al.*, "Differences in Suicide-Related Twitter Content According to User Influence," *Crisis*, vol. 44, no. 4, 2023, doi: 10.1027/0227-5910/a000865.
- [6] D. J. Bartram and D. S. Baldwin, "Veterinary surgeons and suicide: Influences, opportunities and research directions," *Veterinary Record*, vol. 162, no. 2, 2008, doi: 10.1136/vr.162.2.36.
- [7] B. Martens, L. Aguiar, E. GGmez, and F. Mueller-Langer, "The Digital Transformation of News Media and the Rise of Disinformation and Fake News," *SSRN Electronic Journal*, 2018, doi: 10.2139/ssrn.3164170.
- [8] *Digital Transformation in Journalism and News Media*. 2017. doi: 10.1007/978-3-319-27786-8.
- [9] B. Martens, L. Aguiar, E. Gomez-herrera, and F. Mueller-langer, "The digital transformation of news media and the rise of disinformation and fake news An economic perspective," *JRC Digital Economy Working Paper*, vol. 02, no. April, 2018.
- [10] F. Kurniawan and Q. Al Qorni, "Exploring Sentimen Analysis Using Machine Learning: A Case Study on Partai Demokrasi Indonesia Perjuangan (PDIP) in the 2024 General Election," *Sinomics Journal | Volume*, vol. 2, no. 4, 2023.
- [11] H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text mining in big data analytics," *Big Data and Cognitive Computing*, vol. 4, no. 1, 2020, doi: 10.3390/bdcc4010001.
- [12] R. Prakash, V. S. Anoop, and S. Asharaf, "Blockchain technology for cybersecurity: A text mining literature analysis," 2022. doi: 10.1016/j.jjime.2022.100112.
- [13] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "A survey of text mining in social media: Facebook and Twitter perspectives," *Advances in Science, Technology and Engineering Systems*, vol. 2, no. 1, 2017, doi: 10.25046/aj020115.
- [14] B. Percha, "Modern Clinical Text Mining: A Guide and Review," 2021. doi: 10.1146/annurev-biodatasci-030421-030931.
- [15] M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review," 2022. doi: 10.1016/j.eswa.2022.116659.
- [16] H. Xu, J. Zhou, P. G. Asteris, D. J. Armaghani, and M. M. Tahir, "Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate," *Applied Sciences (Switzerland)*, vol. 9, no. 18, 2019, doi: 10.3390/app9183715.
- [17] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2969854.
- [18] H. Zhang, X. Huo, X. Guo, X. Su, X. Quan, and C. Jin, "A disease-related gene mining method based on weakly supervised learning model," *BMC Bioinformatics*, vol. 20, 2019, doi: 10.1186/s12859-019-3078-9.
- [19] A. G. Dufera, T. Liu, and J. Xu, "Regression models of Pearson correlation coefficient," *Stat Theory Relat Fields*, vol. 7, no. 2, 2023, doi: 10.1080/24754269.2023.2164970.
- [20] N. J. Van Eck and L. Waltman, "How to Normalize Co-Occurrence Data ? An Analysis of Some Well-Known Similarity Measures Nees Jan van Eck and Ludo Waltman REPORT SERIES," *Science And Technology*, 2009.