BIM Proceedings
Bali International Meeting on Business, Advanced engineering, and Leadership sustainability Impact
BIM University
DOI: 10.xxxxxxxxxxxxxxxxxxxx/

Sentiment Analysis of Skincare Live Broadcasts: A RoBERTa-Based Text Mining Approach

Muhammad Imam Baihaqi[1], and Ahmad Arro'uf Sulfuadi[2]

[1]Islamic University of Indonesia, St. Kaliurang, Umbulmartani, Ngemplak, Special Region of Yogyakarta, Indonesia, 55584
[2]Islamic University of Indonesia, St. Kaliurang, Umbulmartani, Ngemplak, Special Region of Yogyakarta, Indonesia, 55584

e-mail: 23522066@students.uii.ac.id

Abstract

The rapid rise of social commerce, particularly TikTok Live, has transformed the Indonesian skincare industry. During these live broadcasts, viewers generate massive amounts of unstructured text in the comment section. While these comments contain valuable insights into consumer behavior, they are difficult to analyze manually due to non-standard slang and high velocity. This study aims to classify audience sentiment polarity and extract dominant conversational themes from a skincare brand's TikTok Live broadcast. The methodology involved scraping 693 real-time comments, conducting rigorous text preprocessing, and applying a fine-tuned RoBERTa transformer model. To overcome the common misclassification of consumer inquiries, a rule-based algorithm was integrated to assign interrogative comments as 'neutral automatically'. Visualizations were generated using pie charts and N-Gram Word Clouds. The findings challenge the assumption that comment sections are primarily evaluative spaces. Results showed an overwhelming dominance of neutral sentiment (75.7%), indicating that consumers use the platform primarily as a virtual consultation clinic. Negative sentiment (21.5%) revealed critical consumer pain points, driven by dermatological insecurities (e.g., acne) and transactional frictions like the unavailability of Cash on Delivery (COD). Positive sentiment (2.8%) was largely driven by perceived product safety and promotional bundles. Managerially, these insights suggest that brands must train live hosts as knowledgeable beauty advisors to effectively address real-time consultations. Future research should expand rule-based slang dictionaries and conduct domain-specific fine-tuning to handle linguistic anomalies in Shoppertainment platforms better.

Keywords: *Sentiment Analysis, RoBERTa, Live Commerce, Text Mining, Skincare.*

,

INTRODUCTION

The rapid evolution of digital technology has fundamentally restructured the global e-commerce landscape, shifting the paradigm from static transactional interfaces to dynamic, experiential retail environments [1]. In the contemporary market, consumers are no longer satisfied with browsing through traditional, non-interactive catalogs; instead, they demand a shopping experience that is immersive, entertaining, and characterized by instantaneous gratification [2]. This shift has catalyzed the rise of Shoppertainment and Social Commerce, a phenomenon where social media platforms, most notably TikTok, seamlessly integrate entertainment through live streaming with direct transaction capabilities via TikTok Shop. For the Indonesian skincare industry, which is marked by hyper-competition and high consumer involvement, TikTok Live has emerged as the strategic vanguard for digital marketing, enabling brands to drive real-time conversions while

DOI: 10.xxxxxxxxxxxxxxxxxxxx/

fostering a sense of community and urgency [3].

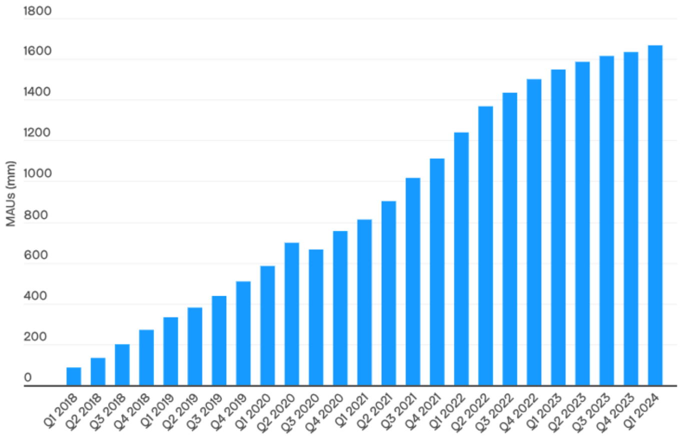

Figure 1. Indonesian TikTok users

Among the prominent players in the beauty sector, the skincare brand has distinguished itself through its aggressive and highly successful utilization of TikTok Live as a primary sales and engagement engine [4]. During these live broadcasts, hosts engage in a relentless, real-time dialogue with thousands of viewers who saturate the comment section with an overwhelming influx of text. These comments represent more than just digital noise; they are high-fidelity reflections of consumer behavior and psychological triggers [5]. Viewers utilize this space for a multitude of motives, ranging from seeking personalized consultations regarding specific dermatological concerns, such as acne or hyperpigmentation, to confirming high-intent purchase actions like checking out, or even voicing skepticism and dissatisfaction regarding product efficacy [6]. Consequently, this stream of data serves as a vital, unfiltered pulse of the market's perception of the brand.

However, the sheer velocity and massive volume of these real-time comments present a significant analytical bottleneck for brand managers. Many organizations remain trapped in a *data-rich but insight-poor* state, where they prioritize surface-level quantitative metrics, such as concurrent viewership counts and Gross Merchandise Value (GMV), while remaining oblivious to the qualitative nuances that drive those numbers [7]. The challenge is further exacerbated by the linguistic complexity of the data; comments are typically written in non-standard Indonesian, heavily utilizing slang, code-switching, and informal abbreviations that render manual analysis not only inefficient but also highly susceptible to subjective bias [8]. Without a systematic method to process this unstructured data at scale, brands risk missing critical signals regarding consumer pain points and emerging market trends.

To address this gap, Natural Language Processing (NLP) through Sentiment Analysis offers a robust computational framework for automatically classifying the emotional polarity of consumer feedback into positive, negative, or neutral categories [9]. This research employs the RoBERTa (Robustly Optimized BERT Pretraining Approach) architecture, a state-of-the-art transformer model that has been fine-tuned for the nuances of the Indonesian language [10]. Unlike traditional machine learning models or even the standard BERT architecture, RoBERTa offers superior computational depth in understanding bidirectional context and subtle semantic variations [11]. Its ability to capture the intricacies of sentiment in short, informal text, even in a zero-shot or few-shot learning context, makes it an exceptionally powerful tool for deciphering

DOI: 10.xxxxxxxxxxxxxxxxxxxx/

the chaotic nature of live-streamed commentary.

Despite the proliferation of AI-driven sentiment analysis, there is a notable scarcity of research that specifically integrates high-performance transformer models with intuitive text mining visualizations, such as Word Clouds, to map consumer psychology within the live commerce ecosystem. Most studies focus either on the technical accuracy of the model or the marketing outcomes, rarely bridging the two to provide a holistic view of the consumer voice during live interactions. Therefore, this study aims to analyze the sentiment polarity of Brand's TikTok Live audience and extract the dominant thematic clusters within each sentiment category. The findings of this research are intended to contribute to the academic discourse on digital consumer behavior while providing actionable managerial implications for business practitioners striving to optimize their interactive marketing strategies in the social commerce era.

METHOD

Data Collection

Data collection was conducted using a data scraping method on Live TikTok Skincare Brand comments. The data scraping process used the Python programming language with the TikTokLive library to capture comments in real-time. A sample of 693 comments was collected. Below are the results:

Table 1. Data Collection Results

No	Comment
1	*untuk kulit kusam kak?*
2	*KLO beruntusan kak?*
3	*terimakasih ka suda co payment?*
4	*buat mencerahkan wajah yang mana kak?*
5	*buat cwok bruntusan, komedo +berminyak dan kusam yg mana kak?*
6	*kalo wajah kusam jerawat on of ambil yang mna ka?*
7	*kalo wajah kusam jerawat on of ambil yang mna ka?*
8	*untuk anak umur 12 bisa pakai dari apa?*
9	*moisturizer untuk flek hitam et ka?*
10	*remofer yang ungu buat apa ya kk?*

Pre-processing

The raw dataset obtained through web scraping contained various inconsistencies, including emojis, foreign terms, and duplicate entries [12]. Consequently, a rigorous data preprocessing phase was essential to enhance data quality and ensure optimal performance before the RoBERTa model conducted the sentiment analysis [13]. The following steps were performed:

1. Duplicate Removal
2. Case Folding
3. Noise Removal
4. Slang Normalization

In the first step, Duplicate Removal was executed to eliminate redundant comments and usernames, ensuring that each data point was unique. The second step, Case Folding, involved standardizing the text format; for instance, transforming non-standard input such as *'ulangi LG KA ngomong nya yg brusan'* into the more uniform *'ulangi lg kakak ngomong nya yang brusan'*.

DOI: 10.xxxxxxxxxxxxxxxxxxxx/

The third step, Noise Removal, focused on eliminating distracting characters or irrelevant symbols; for example, converting *'377 serum etalase berapa kak?'* into *'377 serum etalase berapa kakak'*. The fourth step, Slang Normalization, involved converting colloquialisms and informal abbreviations into formal Indonesian, such as mapping *'co'* or *'coo'* to *'checkout'*, *'klo'* to *'kalau'*, and *'yg'* to *'yang'*. Finally, the preprocessed dataset was exported and stored in CSV (Comma-Separated Values) format for further computational analysis.

Sentiment Analysis RoBERTa

The sentiment analysis is executed using the RoBERTa model, leveraging its pre-trained capabilities without additional retraining from scratch [14]. A common challenge in general AI models is the misclassification of inquiries; for instance, a question such as *'kak kalau beruntusan pakai apa?'* is frequently mislabeled as negative simply due to the presence of the word *'beruntusan.'* To mitigate this issue, a rule-based filtering algorithm was implemented to scan for interrogative keywords (question words). If a question word is detected, the system automatically overrides the classification and assigns a 'neutral' label to the comment. Conversely, if no interrogative terms are identified, the classification task is delegated back to the RoBERTa model to determine the final sentiment polarity.

Data Visualization

To facilitate a comprehensive analysis and derive meaningful insights, the results are presented using the following visualization techniques:

1. Pie Chart
   A Pie Chart is utilized to illustrate the proportional distribution and percentage comparison across the three sentiment labels: positive, neutral, and negative [15]. This provides a clear, high-level overview of the overall audience perception during the live broadcast.
2. N-Gram Word Cloud
   N-Gram Word Clouds are employed to dissect consumer behavioral patterns and dominant themes within each sentiment category [16]. In this visualization, the physical size of a term is directly proportional to its frequency of occurrence; therefore, the most prominent words in the cloud represent the most frequently discussed topics or concerns by the audience.

RESULTS AND DISCUSSION

To evaluate the performance of the classification logic, a sample of the labeled dataset is presented in the table below. This table illustrates the raw output of the model and highlights the complexity of sentiment polarity in live broadcast comments. Notably, it demonstrates how certain phrases containing specific keywords related to skin concerns, which are often misidentified by standard models, are processed and categorized into their respective sentiment classes (Positive, Neutral, or Negative) based on the hybrid approach previously described.

Table 2. AI RoBERTa labeling results

No	Comment	Label
1	*untuk kulit kusam kakak*	negative
2	*kalau beruntusan kakak*	negative
3	*377 serum etalase berapa kakak*	neutral
4	*terimakasih kakak suda checkout payment*	neutral

DOI: 10.xxxxxxxxxxxxxxxxxxxx/

No	Comment	Label
5	*aman tidak kakak 4 hri pakai pket warna biru calming sensitive 3 hari paket warna hijauacne*	negative
6	*kakak kalo buat kulit wajah kusam yang mana*	neutral
7	*ulangi lg kakak ngomong nya yang brusan*	negative
8	*kakak eta kase 13*	neutral
9	*cleansing oil ada berapa ukuran kakak*	neutral
…	...	…
284	*skintint brp ml*	negative

As demonstrated in the table above, the classification system processes a diverse array of informal comments, ranging from product inquiries to specific viewer requests. It is evident that the implementation of the rule-based algorithm successfully assigns a 'neutral' label to question-based comments, such as *"kakak kalo buat kulit wajah kusam yang mana"* and *"cleansing oil ada berapa ukuran kakak."* However, the table also highlights the ongoing complexity of text mining in live commerce; short or ambiguous comments like *"skintint brp ml"* may still be classified as 'negative' by the model, emphasizing the nuanced nature of unstructured data and the high sensitivity of the RoBERTa architecture toward specific terms. To provide a macro-level understanding of the audience's overall perception during the Glad2Glow TikTok Live broadcast, the classification results from the entire dataset were aggregated. The proportional distribution of these sentiment categories is visualized in Figure X below. This pie chart illustrates the dominant sentiment polarity, offering immediate insight into the general consumer engagement dynamics.

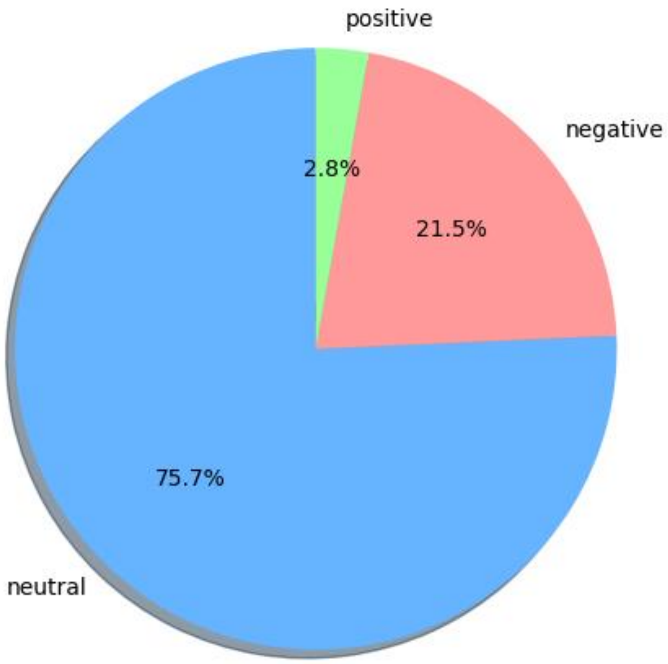

Figure 2. Pie Chart of Labeling Results

As illustrated in the pie chart, the sentiment distribution is overwhelmingly dominated by the 'neutral' category at 75.7%, followed by 'negative' sentiment at 21.5%, while 'positive' sentiment constitutes the smallest fraction at merely 2.8%. The massive proportion of neutral sentiment is

DOI: 10.xxxxxxxxxxxxxxxxxxxx/

highly characteristic of live commerce environments, where the comment section functions primarily as a real-time Q&A board for transactional inquiries rather than a platform for expressing direct praise or dissatisfaction.

However, while the pie chart provides a clear quantitative overview, it lacks the qualitative context necessary to understand the specific drivers behind these emotions. To uncover the underlying themes and consumer behaviors within each category, an N-Gram Word Cloud analysis was conducted. Figure Y presents the word cloud generated specifically for the positive sentiment. Despite representing the smallest segment of the data, the extracted N-grams reveal critical value propositions that resonate with the audience. Dominant phrases indicate a strong consumer focus on product safety and inclusivity, highlighted by terms such as *"aman busui"* (safe for nursing mothers) and *"cocok kulit"* (suits the skin). Additionally, phrases like *"banyak banget"* (very abundant) and *"dapet micellar"* (getting micellar water) suggest that promotional bundles and perceived product value successfully generate positive engagement among viewers.

Figure 3. Positif Sentiment Word Cloud

Moving beyond the positive feedback, the analysis delves into the 'neutral' sentiment, which constitutes the vast majority (75.7%) of the dataset. As visualized in the Neutral N-Gram Word Cloud, the lexical landscape shifts drastically from emotional expressions to highly functional and transactional inquiries. Dominant phrases such as *"pake mana"* (which one to use), *"bisa pakai"* (can I use it), and specific product names like *"facial wash," "acne set,"* and *"tone up"* highlight the primary role of the live broadcast as a virtual consultation space. Furthermore, the prominence of phrases like *"serum etalase"* (serum in which catalog/cart) and *"spil lip"* (reveal/demonstrate the lip product) provides compelling evidence of high purchase intent among the audience. Instead of expressing passive emotion, viewers actively use the comment section to navigate the product catalog, request real-time product demonstrations, and seek personalized skincare advice from the host. This visualization perfectly validates the implementation of the rule-based algorithm in the methodology, successfully capturing the interrogative and consultative nature of *live commerce* interactions.

DOI: 10.xxxxxxxxxxxxxxxxxxxxx/

Figure 4. Netral Sentiment Word Cloud

Finally, the analysis addresses the 'negative' sentiment category, which accounts for 21.5% of the dataset. The Negative N-Gram Word Cloud reveals a distinct shift in discourse, highlighting consumer pain points, skepticism, and transactional friction. A prominent cluster within this sentiment revolves around specific dermatological concerns, evidenced by dominant phrases such as *"kulit kusam"* (dull skin), *"kulit bruntusan"* (textured/breakout skin), and *"jerawat"* (acne). The RoBERTa model naturally assigns negative computational weights to these terms due to their association with unfavorable conditions, reflecting the primary insecurities that drive consumers to seek skincare solutions.

Figure 5. Negative Sentiment Word Cloud

DOI: 10.xxxxxxxxxxxxxxxxxxxx/

Furthermore, the word cloud exposes operational and transactional hurdles. Phrases like *"gak bisa"* (cannot) and specifically *"gak bisa cod"* (Cash on Delivery unavailable) represent genuine dissatisfaction regarding payment options and platform limitations. Additionally, phrases expressing doubt, such as *"bagus ngak"* (is it good or not?) and *"bisa mencerahkan"* (can it actually brighten?), indicate consumer hesitation. Even post-purchase phrases like *"udh checkout"* (already checked out) appear in this negative context, likely correlated with complaints about shipping delays or unfulfilled requests. These insights are crucial for brand managers, as they pinpoint exactly where consumer trust wavers and where the purchasing journey encounters friction.

CONCLUSION

This study successfully applied the RoBERTa architecture, augmented with a rule-based filtering algorithm, to classify and analyze consumer sentiment within the highly unstructured and dynamic environment of TikTok Live skincare broadcasts. By addressing the linguistic complexities of non-standard Indonesian, slang, and real-time textual noise, the methodology provided a robust framework for extracting deep behavioral insights from otherwise chaotic data. The findings fundamentally challenge the traditional assumption that comment sections are primarily evaluative spaces. Instead, the overwhelming dominance of neutral sentiment (75.7%), characterized by high-intent product inquiries and consultative requests, proves that consumers treat live commerce primarily as a virtual, interactive consultation clinic. While the positive sentiment (2.8%) highlights that product safety features (e.g., safe for nursing mothers) and promotional bundles successfully drive brand appreciation, the negative sentiment (21.5%) provides critical diagnostic value. It reveals that negativity in live streams is largely driven by pre-existing consumer insecurities (such as acne and dull skin) and specific transactional frictions, most notably the unavailability of Cash on Delivery (COD) options and skepticism regarding product efficacy.

From a managerial perspective, these insights offer actionable strategies for skincare brands in the social commerce space. Given the heavily consultative nature of the audience, brands must ensure that live stream hosts are trained not merely as salespeople, but as knowledgeable beauty advisors capable of addressing specific dermatological questions in real-time. Furthermore, addressing the operational bottlenecks identified in the negative word cloud, such as expanding payment options or reiterating clinical proofs to counter skepticism, can significantly reduce purchase hesitation and improve overall conversion rates. Despite the high accuracy of the hybrid classification approach, this study acknowledges certain limitations. The extreme variability of Indonesian slang and implicit question formats (e.g., omitting standard interrogative words like *apa* or *bagaimana*, and using ambiguous abbreviations like *brp ml* or *bagus ngak*) occasionally led to misclassifications where neutral inquiries were flagged as negative by the base RoBERTa model. Future research should focus on expanding the rule-based dictionaries to encompass a wider array of non-standard colloquialisms. Additionally, conducting domain-specific fine-tuning of transformer models explicitly on live commerce datasets could further enhance the machine's ability to decode the extreme linguistic anomalies inherent to Shoppertainment platforms.

DECLARATION OF GENERATIVE AI

During the preparation of this research, the author utilized generative AI technologies strictly as an assistive tool for brainstorming, code proofreading, and minor language refinements. The AI-assisted elements constitute less than 15% of the overall manuscript. The author has meticulously reviewed and edited all generated outputs and assumes full responsibility for the final content, originality, and accuracy of this work.

ACKNOWLEDMENT

The author would like to express profound gratitude to Universitas Islam Indonesia (UII), specifically the Department of Industrial Engineering within the Faculty of Industrial Technology for providing the academic environment, facilities, and continuous support necessary to conduct this research. The author also appreciates the constructive comments from the reviewers, which significantly improved the quality of this manuscript.

REFERENCES

[1] R. Schmitt, "From raw text to fairseq RoBERTa: A modular snakemake-based framework enabling language-specific BPE tokenization," *Software Impacts*, vol. 27, p. 100824, Apr. 2026, doi: 10.1016/J.SIMPA.2026.100824.

[2] G. Gelastopoulos and C. Keramydas, "A systematic review of text mining analytics for supply chain risk management using online data," *Supply Chain Analytics*, vol. 12, p. 100167, Dec. 2025, doi: 10.1016/J.SCA.2025.100167.

[3] Y. Li and C. W. Zobel, "Exploring supply chain network resilience in the presence of the ripple effect," *Int. J. Prod. Econ.*, vol. 228, Oct. 2020, doi: 10.1016/j.ijpe.2020.107693.

[4] M. Pontiki *et al.*, "SemEval-2016 task 5: Aspect based sentiment analysis," *SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings*, pp. 19–30, 2016, doi: 10.18653/V1/S16-1002.

[5] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, vol. 1, pp. 1555–1565, 2014, doi: 10.3115/V1/P14-1146.

[6] E. Cambria, "Affective Computing and Sentiment Analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar. 2016, doi: 10.1109/MIS.2016.31.

[7] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional Encoder Network for Targeted Sentiment Classification," Apr. 2019, doi: 10.1007/978-3-030-30490-4_9.

[8] L. You, F. Han, J. Peng, H. Jin, and C. Claramunt, "ASK-RoBERTa: A pretraining model for aspect-based sentiment classification via sentiment knowledge mining," *Knowl. Based. Syst.*, vol. 253, p. 109511, Oct. 2022, doi: 10.1016/J.KNOSYS.2022.109511.

[9] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019, Accessed: Apr. 21, 2026. [Online]. Available: http://arxiv.org/abs/1907.11692

[10] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc. Netw. Anal. Min.*, vol. 11, no. 1, Dec. 2021, doi: 10.1007/S13278-021-00776-6.

[11] X. Fang and J. Zhan, "Sentiment analysis using product review data," *J. Big Data*, vol. 2, no. 1, Dec. 2015, doi: 10.1186/S40537-015-0015-2.

DOI: 10.xxxxxxxxxxxxxxxxxxxx/

[12] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Syst. Appl.*, vol. 223, Aug. 2023, doi: 10.1016/j.eswa.2023.119862.

[13] S. Sindhu, S. Kumar, and A. Noliya, "A Review on Sentiment Analysis using Machine Learning," *International Conference on Innovative Data Communication Technologies and Application, ICIDCA 2023 - Proceedings*, pp. 138–142, 2023, doi: 10.1109/ICIDCA56705.2023.10099665.

[14] B. Paneru, B. Thapa, and B. Paneru, "Sentiment analysis of movie reviews: A flask application using CNN with RoBERTa embeddings," *Systems and Soft Computing*, vol. 7, p. 200192, Dec. 2025, doi: 10.1016/J.SASC.2025.200192.

[15] M. G. Yazdi, H. Rafieizadeh, and P. Tajasob, "Sentiment analysis of the 2024 Paris Olympics using RoBERTa language model," *International Journal of Event and Festival Management*, vol. 17, no. 1, pp. 133–160, Sep. 2025, doi: 10.1108/IJEFM-01-2025-0007.

[16] J. Dai, Y. Zhao, and Z. Li, "Sentiment-topic dynamic collaborative analysis-based public opinion mapping in aviation disaster management: A case study of the MU5735 air crash," *International Journal of Disaster Risk Reduction*, vol. 102, Feb. 2024, doi: 10.1016/j.ijdrr.2024.104268.