Jurnal Teknik Informatika C.I.T Medicom, 16 (3) (2024) 199-210

Published by: Institute of Computer Science (IOCS)

Jurnal Teknik Informatika C.I.T Medicom

Journal homepage: www.medikom.iocspublisher.org

Toxicity and topic analysis of travel vlog content in digital era: perspective and multilingual embedding model (voyage-multilingual-2)

Yerik Afrianto Singgalen
Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia

Article Info

Article history

Received : Apr 25, 2024
Revised : Jun 01, 2024
Accepted : Jul 22, 2024

Keywords:

Clustering techniques;
Content moderation;
Digital communication;
Toxicity analysis;
Travel vlog content.

Abstract

This research investigates the complexities of online discourse by conducting a detailed toxicity and topic analysis of travel vlog content on user-generated platforms. By analyzing 1,503 posts using the Perspective API, the study finds generally low levels of toxicity, with an average toxicity score of 0.06995 and a peak of 0.78207, and similarly low average scores for severe toxicity, identity attack, insult, profanity, and threat (0.00654, 0.01237, 0.03778, 0.06241, and 0.01186, respectively). However, the highest recorded values for these measures—0.45895 for severe toxicity, 0.69287 for identity attack, 0.63084 for insult, 0.81864 for profanity, and 0.51957 for threat—highlight the sporadic presence of harmful content. Advanced clustering techniques, such as HDBScan, k-Means, and Gaussian Mixture models, enable a comprehensive examination of thematic diversity and sentiment distribution within the comments, offering valuable insights into audience engagement and perception. These findings underline the critical need for compelling content moderation and community management strategies to mitigate toxic behaviors and promote a positive digital environment. The study concludes that as digital media evolves, further research into toxicity, thematic content, and user engagement is essential for enhancing theoretical frameworks and practical applications in digital communication.

Corresponding Author:

Yerik Afrianto Singgalen,
Tourism Department, Faculty of Business Administration and Communication
Atma Jaya Catholic University of Indonesia
Jl. Jend. Sudirman No.51 5, RT.004/RW.4, , Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12930
yerik.afrianto@atmajaya.ac.id

1. Introduction

The urgency of researching the toxicity and topic analysis of travel vlogs, mainly through multilingual embedding models such as the voyage-multilingual-2, lies in its potential to enhance understanding of cultural dynamics in digital communication and mitigate the spread of harmful content. As a popular medium on social platforms, travel vlogs often reflect diverse cultural narratives and perspectives, making them a rich resource for linguistic and sociocultural analysis [1]–[5]. However, these vlogs also risk disseminating toxic discourse, which can perpetuate stereotypes, misinformation, and negative sentiments across global audiences [6]–[9]. The voyage-multilingual-2 model, with its capacity for nuanced multilingual analysis, offers an advanced method to dissect these complexities by identifying

underlying topics and assessing sentiment across different languages, thereby contributing to a more comprehensive understanding of the digital landscape [10]–[13]. Moreover, leveraging such sophisticated models to analyze the multilingual aspects of toxicity and topic emergence in travel vlogs could inform strategies to promote more inclusive and positive online interactions. Addressing these challenges advances academic knowledge and holds practical significance for platform moderation, policy-making, and promoting respectful digital communication across linguistic and cultural boundaries.

This research aims to explore and elucidate the patterns of toxicity and thematic content in travel vlogs by applying a multilingual embedding model, specifically the voyage-multilingual-2. This study seeks to unravel the complex interplay between language, culture, and sentiment by systematically analyzing diverse narratives presented in travel vlogs. It provides a comprehensive understanding of the factors contributing to both positive and negative discourse in digital media. Such an examination is critical in discerning the subtle yet significant ways language and culture shape online communication, influencing public perception and behavior [14]–[16]. The research offers a novel approach to understanding digital interactions across various linguistic contexts using advanced computational techniques to decode these patterns. Ultimately, this investigation aims to advance theoretical knowledge in computational linguistics and digital media studies and inform practical strategies for enhancing the quality of online communication, thereby fostering a more inclusive and respectful digital environment.

The novelty of this research lies in its innovative application of the voyage-multilingual-2 model to analyze toxicity and thematic content in travel vlogs, which represents a significant departure from traditional monolingual approaches to sentiment analysis. Unlike previous studies primarily focused on single-language datasets, this research embraces a multilingual perspective, enabling a more nuanced understanding of how toxic discourse manifests across different languages and cultural contexts [3], [17]–[20]. By leveraging advanced multilingual embedding techniques, the study offers a fresh methodological approach that captures subtle variations in language use and sentiment expression that conventional models often miss [21]–[24]. This unique focus provides a richer analysis of digital communication and addresses a critical gap in existing literature regarding the influence of linguistic diversity on online interactions. Such an approach enhances the capacity to detect and interpret toxic content with greater precision, thereby contributing to developing more effective moderation strategies on digital platforms.

This research's theoretical and practical implications extend beyond merely advancing the academic discourse on multilingual sentiment analysis and toxicity detection in digital media. Theoretically, the study contributes to the broader understanding of how language models, such as the voyage-multilingual-2, navigate the intricate dynamics of multiple languages and cultural contexts to identify and interpret toxic content. This understanding enhances existing computational linguistics and digital humanities frameworks, offering new insights into language representation and sentiment analysis in diverse and complex settings. Practically, the findings provide valuable tools for improving content moderation systems on digital platforms, particularly those hosting user-generated content like travel vlogs. By identifying toxic language patterns and promoting more inclusive communication, the study aids in developing more effective strategies to combat online toxicity, thereby fostering a safer and more respectful digital environment [25]–[28]. Such dual contributions underscore the study's significance, advancing theoretical knowledge and offering practical solutions to real-world challenges in digital communication.

Similar research in digital media studies has often focused on analyzing sentiment and toxicity across various social media platforms. Yet, as this study proposes, few have comprehensively integrated the complexity of multilingual contexts and cultural narratives. Previous studies have primarily utilized monolingual or bilingual models to examine the prevalence of toxic language and its impact on user interactions, often overlooking the nuanced differences that arise in a multilingual environment [29]–[31]. It is argued that a more sophisticated approach, incorporating multilingual embedding models, offers a deeper understanding of how toxicity is expressed and perceived differently across languages

and cultures. Such an approach enables more accurate detection of harmful content and provides insights into the sociocultural dynamics that influence digital communication [32]. Consequently, this research not only builds upon existing literature by addressing these gaps but also introduces a novel methodology that holds significant promise for enhancing the accuracy and relevance of toxicity detection in a rapidly globalizing digital landscape.

One significant limitation of this research is its reliance on the voyage-multilingual-2 model, which, while advanced, is not without constraints in capturing the full complexity of multilingual and multicultural nuances in digital communication. The model's dependency on pre-existing linguistic datasets means that subtle variations in regional dialects, colloquialisms, or emerging slang may not be accurately represented or understood, potentially leading to misinterpretation of toxic content or cultural references. This shortcoming is further exacerbated by the rapidly evolving nature of online language, where new terms and expressions frequently emerge, challenging the model's adaptability and relevance over time. Moreover, the focus on textual data alone may overlook the multimodal nature of travel vlogs, which often combine visual and auditory elements that contribute significantly to meaning-making and viewer interpretation. These limitations suggest that while the study offers valuable insights into the linguistic dimensions of toxicity, there is a need for further refinement of analytical tools and the inclusion of more diverse data sources to achieve a more holistic understanding of digital communication in multilingual settings.

2. Research Methodology

The Digital Content Reviews and Analysis Framework is designed to evaluate and interpret digital content systematically, mainly focusing on the toxicity and thematic analysis of travel vlog content. This framework integrates multiple stages, beginning with the collection of content reviews and text data, followed by a meticulous data processing phase that distinguishes between toxicity perspective and topic analysis. The framework employs sophisticated clustering techniques such as HDBSCAN, K-means, and Gaussian Mixture models to conduct a detailed topic analysis, which enables a deeper understanding of underlying patterns and thematic structures within the content. Furthermore, the processed data undergoes a rigorous evaluation and visualization process to facilitate insightful interpretation and context analysis, ultimately providing a comprehensive assessment of digital content's nature and impact. By adopting such a multidimensional approach, the framework enhances the accuracy of detecting toxic elements. It offers a robust methodology for the thematic exploration of digital narratives, contributing to a more nuanced understanding of online communication dynamics.

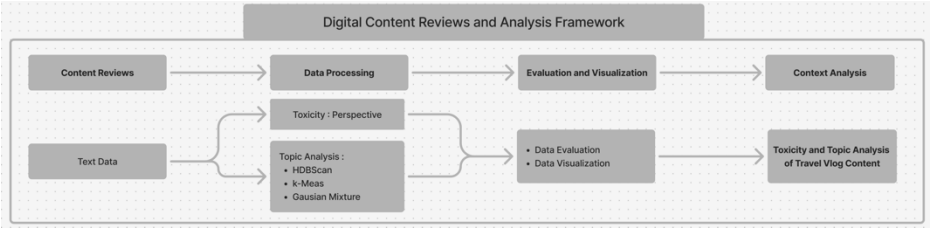

Figure. 1. Implementation of Digital Content Reviews and Analysis Framework

The data utilized in this study comprises comments extracted from a YouTube video with the ID i9t9pbdo-bk, titled "EPIC INDONESIA (Malang, Banyuwangi, Bali)," which was published by the channel BackpackerTampan and has garnered 639,726 views since July 14, 2019. As of August 29, 2024, this video has elicited 1,927 comments, reflecting a significant audience engagement and interaction level. The collection of these comments, conducted through a scraping process on August 29, 2024, at 05:07 using the Communalytic tool, presents a valuable dataset for analyzing viewer sentiment and discourse surrounding the content. The substantial volume of responses offers a rich ground for exploring various aspects of digital communication, including sentiment dynamics and thematic

elements. This methodological approach facilitates a comprehensive examination of audience interaction patterns. It contributes to a deeper understanding of the sociocultural implications embedded within digital platforms, highlighting the relevance and utility of such data in contemporary media studies.

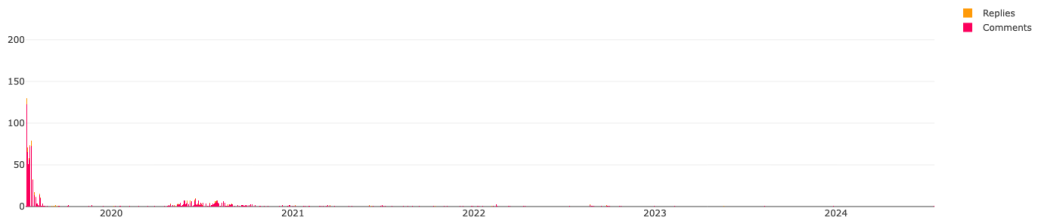

Figure. 2. Post-per-day statistic

Based on the post-per-day statistics, it is evident that the activity surrounding the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)" experienced a significant surge in comments and replies immediately following its publication in 2019, with a sharp decline in engagement over time. This initial spike reflects a heightened interest and interaction from viewers, likely driven by the video's fresh content and the immediate relevance of its themes to the audience at that time. As the novelty of the video content diminished, the frequency of comments and replies progressively decreased. It suggests a typical pattern of audience engagement where initial reactions are concentrated shortly after release, followed by a gradual tapering off. Such a trend underscores the importance of temporal factors in digital content interaction, where the relevance and impact of media diminish as time progresses. Understanding these dynamics is crucial for optimizing content strategy and maximizing viewer engagement over extended periods, highlighting the temporal nature of digital audience behavior.

The word cloud generated from the comments visually represents viewer responses to the published digital content, effectively capturing the most frequently used words and phrases that characterize audience sentiment and engagement. This visualization highlights the key themes and emotions that resonate with viewers, offering a snapshot of the collective discourse surrounding the content. By examining the prominence of specific terms, valuable insights can be gained into the underlying sentiments and topics that dominate the conversation, whether they reflect positive engagement, criticism, or contentious debates. Connecting these findings to the broader context of this research is crucial, as it allows for a deeper understanding of how digital content influences audience perception and interaction patterns. Such an analysis enriches the interpretation of online discourse and informs content creation and moderation strategies, ultimately contributing to more effective digital communication and community building.

Figure. 3. Word Cloud

Based on the word cloud generated from the comments on the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)," it is evident that viewer responses are predominantly positive, as

indicated by the frequent use of terms such as "keren" (excellent), "banget" (very), and "bagus" (good). The prominence of these words suggests a general appreciation for the content, highlighting its appeal to the audience. Additionally, the recurring words "video," "channel," and "subscribe" reflect viewers' engagement with both the content and the creator, indicating a level of endorsement and a call to action for other viewers to engage similarly. The presence of terms like "Indonesia," "Malang," and "Banyuwangi" points to a strong regional and cultural connection, which might resonate with viewers' sense of identity and pride in the representation of their locales. This linguistic analysis demonstrates how the language used in viewer comments aligns with emotional responses to the content and community engagement. Underscoreses the video's success in fostering positive viewer interaction and cultural affinity.

Identifying top-ten posters plays a crucial role in recognizing the key actors who significantly contribute to increasing the rating and popularity of digital content. These individuals often exhibit high levels of engagement, frequently commenting, sharing, or interacting with the content, which amplifies its visibility and enhances its perceived value to other viewers. Their active participation can stimulate further discussions and encourage additional user interactions, creating a ripple effect that boosts the content's reach and attractiveness. Analyzing the behavior and influence of these top contributors makes it possible to understand the mechanisms through which certain content gains traction and achieves widespread popularity. This understanding is vital for developing targeted strategies that leverage the influence of such key actors to optimize content performance and foster more dynamic and engaged digital communities.

Figure. 4. Top Ten Poster

Based on the analysis of the top ten posters commenting on the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)," it is evident that the user @BackpackerTampan significantly dominates the discussion, contributing 70 comments, which accounts for 69.3% of the total comments from the top ten posters. This disproportionate level of engagement suggests that @BackpackerTampan may be the content creator actively interacting with viewers or a highly engaged fan of the video, aiming to foster community interaction or boost the video's visibility. In contrast, the remaining nine users, including @aliolfat5827, @notaident, @kindlejoy, and others, contribute far fewer comments, ranging from three to four each, indicating a more sporadic engagement pattern. Though limited in volume, the presence of these users still reflects a diverse audience interacting with the content. Such a concentration of posts by a single user highlights the need to consider the role of highly active participants in shaping the discourse and dynamics within online communities, which can have implications for understanding engagement strategies and content dissemination on digital platforms.

Based on the collected text data, the subsequent processing will follow the steps outlined in the Digital Content Reviews and Analysis Framework to achieve the desired research objectives. This framework systematically guides the data through various stages, including preprocessing, sentiment and toxicity analysis, and thematic clustering, ensuring that the nuances of digital discourse are

thoroughly examined. The structured approach is designed to enhance the accuracy of the analysis, providing deeper insights into both the positive and negative aspects of viewer engagement with the content. By meticulously adhering to these methodological steps, the framework not only facilitates a comprehensive evaluation of the data but also aligns the analytical process with the overarching goals of the study. Thus, this rigorous approach enables the extraction of meaningful patterns and trends from the dataset, ultimately contributing to a more informed understanding of digital content dynamics and its impact on online communities.

3. Result and Disscusion

Toxicity Score and Interpretation

The toxicity score plays a crucial role in this research by quantitatively measuring the harmful language in the comments on the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)." This score is instrumental in identifying and quantifying the prevalence of harmful content within user interactions, such as insults, threats, or identity attacks, thereby accurately gauging the online discourse's overall tone. Understanding these toxicity levels is essential for analyzing how such harmful elements influence viewer engagement and community dynamics, potentially impacting the video's reception and the platform's social environment. The ability to pinpoint specific instances of high toxicity allows for a more targeted approach to moderating and managing online content, aiming to foster a safer and more respectful digital space. Thus, integrating toxicity scores into the research framework enhances the study's capacity to assess and interpret the complexities of digital communication, ultimately contributing to developing more effective strategies for online community management and policy-making.

The analysis conducted by Communalytic on 1,503 posts out of 1,927 comments using the Perspective API reveals insightful metrics about the nature of online discourse surrounding the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)." The dataset shows an average toxicity score of 0.06995, with the highest recorded value reaching 0.78207, indicating a generally low level of toxicity but with instances of significantly higher toxicity. The scores for severe toxicity, identity attack, insult, profanity, and threat are also relatively low on average, with values of 0.00654, 0.01237, 0.03778, 0.06241, and 0.01186, respectively. However, the maximum scores for these categories—0.45895 for severe toxicity, 0.69287 for identity attack, 0.63084 for insult, 0.81864 for profanity, and 0.51957 for threat—suggest that specific comments exhibit markedly higher levels of harmful language. This discrepancy between average and peak values underscores the sporadic but potentially impactful presence of negative discourse in the comment section, emphasizing the importance of continuous monitoring and moderation to maintain a constructive online environment.

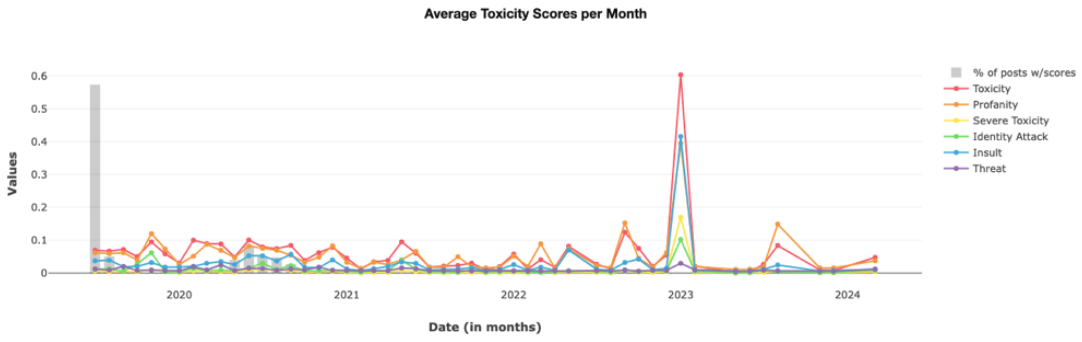

Figure. 5. Average Toxicity Score per Month

The calculated toxicity scores from the analyzed dataset provide a comprehensive view of the varying levels of negative discourse present in the comments section of the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)." An average toxicity score of 0.06995, coupled with a peak value of 0.78207, suggests that while most comments are generally non-toxic, there are occasional instances where toxicity spikes considerably. The relatively low average scores for severe toxicity (0.00654), identity attack (0.01237), insult (0.03778), profanity (0.06241), and threat (0.01186) further indicate that highly harmful language is not pervasive throughout the dataset. However, the highest recorded values in these categories—ranging from 0.45895 for severe toxicity to 0.81864 for profanity—demonstrate that specific comments contain significantly harmful content, which, although infrequent, could impact the overall tone of the discussion. This variability highlights the presence of sporadic but potentially damaging language within the community. Such findings underscore the need for vigilant content moderation and targeted interventions to mitigate the effects of toxic discourse and promote a healthier, more constructive online interaction environment.

The relevance of these toxicity scores to this research lies in their ability to shed light on the dynamics of online discourse, particularly within the context of travel vlogs on YouTube. The findings offer a nuanced understanding of how digital interactions can vary widely, from generally benign or positive to occasionally highly toxic, reflecting the complex nature of user engagement on social media platforms. This research posits that examining the peaks and variations in toxicity and other harmful language indicators is essential for understanding such content's potential impact on audience perception and community dynamics. Moreover, the data highlights the critical need for adaptive content moderation strategies that address overtly harmful language and anticipate subtler forms of negativity that may influence the overall tone of discussions. By integrating these insights, the research contributes to a broader discourse on enhancing digital communication, advocating for more informed and responsive approaches to managing online communities, thereby fostering a more positive and inclusive digital environment.

Topic Analysis based on HDBScan, K-means, and Gaussian Mixture

The integration of topic analysis using HDBScan, k-Means, and Gaussian Mixture models is paramount in this research as it provides a multifaceted approach to understanding the thematic structure of comments on the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)." Each clustering algorithm brings a unique strength: HDBScan identifies clusters of varying densities, k-Means segments data into distinct groups based on centroids, and Gaussian Mixture allows for the probabilistic assignment of comments to multiple clusters. This combination enables a comprehensive exploration of semantic patterns and user sentiment, offering insights into how different audiences interact with and perceive the content. Such diversity in clustering techniques allows for a more robust identification of critical themes and underlying sentiments, facilitating a deeper understanding of the digital discourse's complexity. Using these algorithms in tandem enhances the reliability and depth of the analysis, ensuring that both distinct and overlapping topics are accurately captured. Therefore, applying these advanced topic analysis methods significantly contributes to a more nuanced interpretation of the data, ultimately supporting more informed decisions in managing and understanding digital communication dynamics.

Given that embeddings are represented as vectors within a multidimensional space—specifically, 1024 dimensions in the context of the voyage-multilingual-2 model using a dimension reduction algorithm such as UMAP becomes crucial for effective visualization. UMAP, or Uniform Manifold Approximation and Projection, distills these high-dimensional embeddings into a more understandable three-dimensional space, facilitating a clearer understanding of the data's structure and relationships. This process is precious for identifying patterns and clusters that may not be immediately apparent in higher-dimensional forms, offering a more intuitive grasp of complex linguistic nuances captured by the model. Moreover, visualizing embeddings in a reduced dimension allows for more accessible interpretation and communication of findings, mainly when presenting to audiences unfamiliar with the intricacies of multidimensional analysis. This approach underscores the

importance of dimension reduction techniques in computational linguistics, as they provide a practical means to bridge the gap between sophisticated mathematical representations and their meaningful, real-world interpretations.

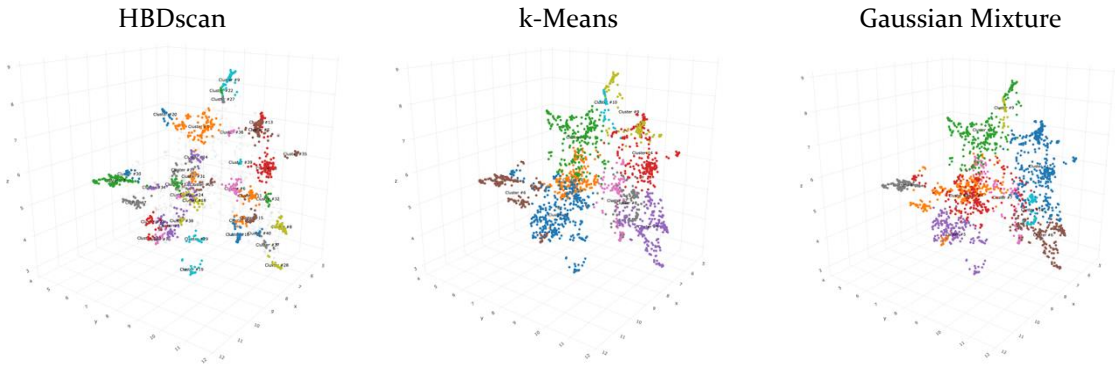

Figure. 6. Topic Analysis based on Cluster (HBDScan, KMeans, and Gaussian Mixture)

The clustering results derived from the HDBScan algorithm provide valuable insights into the semantic similarities among the comments on the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)." The visualization in the 3D Semantic Similarity Map, which identifies distinct clusters of comments, reflects varying thematic and sentiment-based groupings within the dataset. Multiple clusters indicate diverse topics and opinions, suggesting that viewers engage with the content in several ways, each cluster potentially representing unique perspectives or recurrent themes. It is reasonable to infer that comments within the same cluster share more semantic similarities and perhaps exhibit similar sentiment levels, ranging from positive engagement to critical feedback. The dispersion and density of these clusters reveal the complexity of audience interactions and the range of conversations stimulated by the video. This clustering analysis underscores the heterogeneity of user engagement, emphasizing the need for nuanced approaches to moderating and understanding digital discourse in multilingual and multicultural contexts.

The clustering results produced by the K-means algorithm, as depicted in the 3D Semantic Similarity Map, offer a structured representation of the semantic relationships among comments on the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)." By dividing the dataset into clusters, the algorithm effectively groups comments based on similarities in language use, sentiment, and thematic content. The clear separation and distribution of clusters indicate that user interactions encompass a range of distinct yet related topics, reflecting how viewers engage with and interpret the video content. This separation suggests that the clustering approach captures major themes and nuanced subtopics within the discourse, allowing for a more granular understanding of audience sentiment and engagement. The effectiveness of K-means in this context highlights its utility in simplifying complex data into more interpretable segments, facilitating deeper analysis of the patterns and trends in digital communication. Ultimately, this clustering visualization underscores the multifaceted nature of online discussions, illustrating the value of advanced analytical techniques in revealing the underlying structures within large datasets.

The clustering results obtained using the Gaussian Mixture Model (GMM) provide a nuanced understanding of the semantic structure within the dataset of comments on the YouTube video "EPIC INDONESIA (Malang, Banyuwangi, Bali)." Unlike deterministic clustering methods, GMM employs a probabilistic approach, allowing each data point to belong to multiple clusters with varying degrees of membership. This method reveals overlapping clusters in the 3D Semantic Similarity Map, indicating that specific comments may share characteristics with multiple themes or sentiments. Such overlaps suggest a more complex and fluid structure of discourse, where user interactions cannot be strictly categorized into distinct groups. The density and proximity of these clusters reflect the intricate ways

in which different topics and sentiments converge, providing a richer and more flexible framework for interpreting audience engagement. This probabilistic clustering approach underscores the dynamic nature of online conversations and the value of adopting more sophisticated analytical techniques to capture the subtle interrelations and gradations of meaning within large-scale digital datasets.

Discussion: Toxicity and Topic Analysis of Travel Vlog Content in the Digital Era

The toxicity and topic analysis of travel vlog content offer valuable insights into the dynamics of online communication and community engagement on digital platforms. As a unique form of user-generated content, travel vlogs often attract diverse audiences who interact with the content in various ways, from sharing positive experiences to expressing criticisms or negative sentiments. Analyzing the toxicity levels within these interactions helps to understand the extent and nature of harmful language, while topic analysis reveals the underlying themes and narratives that drive user engagement. Such an approach is crucial for identifying patterns of discourse that may impact audience perception and the overall atmosphere of online communities. By combining toxicity and topic analysis, the research provides a comprehensive framework to assess the quality of interactions and the thematic content that shapes digital communication. This dual analysis is essential for developing strategies to enhance content moderation, foster positive engagement, and promote a more inclusive and constructive digital environment.

The study of toxicity and topic analysis of travel vlog content has become increasingly important in the digital era, where user-generated content plays a pivotal role in shaping public opinion and cultural narratives. Travel vlogs, which blend personal experiences with visual storytelling, attract a wide range of audience interactions, from enthusiastic support to critical feedback, making them a fertile ground for analyzing online discourse. Understanding toxicity levels within these discussions is essential for identifying harmful behaviors and fostering healthier digital communities. Simultaneously, topic analysis provides insights into the themes and narratives that resonate most with viewers, revealing cultural trends and audience preferences. Such dual analysis is crucial in an era where digital content consumption significantly influences social attitudes and behaviors. By examining the quality of discourse and the thematic content in travel vlogs, this research contributes to a deeper understanding of digital communication dynamics. It helps develop strategies to promote positive engagement and mitigate negative impacts in online environments.

The challenges associated with travel vlog content in the digital landscape are multifaceted, encompassing issues related to content authenticity, audience engagement, and the management of online discourse. Travel vlogs often present highly curated experiences, which, while engaging, may lead to unrealistic expectations among viewers and perpetuate certain stereotypes or biased representations of destinations. Maintaining a high viewer engagement poses a challenge, as content creators must continuously innovate to capture audience interest in an increasingly saturated market. Another significant challenge lies in managing the diverse range of responses that travel vlogs evoke, from enthusiastic praise to toxic comments that can impact both the content creator's reputation and the broader community discourse. Such varied interactions necessitate a careful approach to content moderation and community management, ensuring that constructive discussions are fostered while mitigating the effects of negativity. Addressing these challenges requires a nuanced understanding of both the creative and communicative aspects of travel vlogging, highlighting the need for strategies that balance authenticity, engagement, and healthy online environments.

The effectiveness of travel vlog content in destination marketing is evident in its ability to offer authentic, visually engaging narratives that captivate potential tourists. Travel vlogs leverage personal storytelling and immersive visuals to provide viewers with a vicarious experience of a destination, which can be more persuasive than traditional advertising. This form of content allows audiences to witness real-life experiences and emotions, which helps to build trust and a deeper connection with

the destination being showcased. Furthermore, travel vlogs often reach a diverse audience through social media platforms, enhancing their potential to influence travel decisions on a global scale. The persuasive power of these vlogs is further amplified by the perceived credibility of content creators, who are often viewed as genuine travelers sharing their unbiased opinions. Consequently, travel vlogs serve as a tool for raising awareness and converting interest into actual visitation, making them a highly influential component of destination marketing strategies.

4. Conclusion

In conclusion, this research underscores the complexities of online discourse through a detailed toxicity and topic analysis of travel vlog content, shedding light on both the opportunities and challenges inherent in digital communication on user-generated platforms. The analysis of 1,503 posts using the Perspective API revealed that the average toxicity score was 0.06995, with a peak value of 0.78207, indicating generally low toxicity levels with occasional spikes. Similarly, other measures such as severe toxicity, identity attack, insult, profanity, and threat exhibited low average scores of 0.00654, 0.01237, 0.03778, 0.06241, and 0.01186, respectively. At the same time, their highest recorded values reached 0.45895, 0.69287, 0.63084, 0.81864, and 0.51957, demonstrating the sporadic presence of harmful content. These results highlight the importance of understanding the varied nature of online interactions, particularly in identifying and managing toxic behaviors that can impact the overall tone of digital communities. Advanced clustering techniques, including HDBScan, k-Means, and Gaussian Mixture models, effectively captured the thematic diversity and sentiment distribution within the comment dataset, providing deeper insights into audience engagement and perceptions. Such findings emphasize the need for strategic content moderation and community management to enhance the quality of digital interactions and foster a more inclusive and constructive online environment. As digital media continues to evolve, further research into toxicity, thematic content, and user engagement will be vital for advancing theoretical understanding and practical strategies in digital communication.

References

[1] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model," *IEEE Access*, vol. 10, pp. 39313–39324, 2022, doi: 10.1109/ACCESS.2022.3165621.

[2] N. Basaraba, "The rise of paranormal investigations as virtual dark tourism on YouTube," *J. Herit. Tour.*, vol. 19, no. 2, pp. 287–309, 2024, doi: 10.1080/1743873X.2023.2268746.

[3] A. Boumhidi, A. Benlahbib, and E. H. Nfaoui, "Cross-Platform Reputation Generation System Based on Aspect-Based Sentiment Analysis," *IEEE Access*, vol. 10, pp. 2515–2531, 2022, doi: 10.1109/ACCESS.2021.3139956.

[4] S. Seo, S. Na, and J. Kim, "HMTL: Heterogeneous Modality Transfer Learning for Audio-Visual Sentiment Analysis," *IEEE Access*, vol. 8, pp. 140426–140437, 2020, doi: 10.1109/ACCESS.2020.3006563.

[5] A. Altaf *et al.*, "Deep Learning Based Cross Domain Sentiment Classification for Urdu Language," *IEEE Access*, vol. 10, no. September, pp. 102135–102147, 2022, doi: 10.1109/ACCESS.2022.3208164.

[6] F. Mehraliyev, I. C. C. Chan, and A. P. Kirilenko, "Sentiment analysis in hospitality and tourism: a thematic and methodological review," *Int. J. Contemp. Hosp. Manag.*, vol. 34, no. 1, pp. 46–77, 2022, doi: 10.1108/IJCHM-02-2021-0132.

[7] Y. A. Singgalen, "Culture and heritage tourism sentiment classification through cross-industry standard process for data mining," *Int. J. Basic Appl. Sci.*, vol. 12, no. 3, pp. 110–120, 2023.

[8] M. Sohi, M. Pitesky, and J. Gendreau, "Analyzing public sentiment toward GMOs via social media between 2019-2021," *GM Crop. Food*, vol. 14, no. 1, pp. 1–9, 2023, doi: 10.1080/21645698.2023.2190294.

[9] R. Bringula, S. A. I. D. A. Ulfa, J. P. P. Miranda, and F. A. L. Atienza, "Text mining analysis on students' expectations and anxieties towards data analytics course," *Cogent Eng.*, vol. 9, no. 1, 2022, doi: 10.1080/23311916.2022.2127469.

[10] A. Karlsson, E. T. Bekar, and A. Skoogh, "Multi-machine Gaussian topic modeling for predictive maintenance," *IEEE Access*, vol. 9, pp. 100063–100080, 2021, doi: 10.1109/ACCESS.2021.3096387.

[11] C. Zhou, H. Ban, J. Zhang, Q. Li, and Y. Zhang, "Gaussian Mixture Variational Autoencoder for Semi-Supervised Topic Modeling," *IEEE Access*, vol. 8, pp. 106843–106854, 2020, doi:

10.1109/ACCESS.2020.3001184.
[12] R. A. Vandermeulen and R. Saitenmacher, "Generalized Identifiability Bounds for Mixture Models With Grouped Samples," *IEEE Trans. Inf. Theory*, vol. 70, no. 4, pp. 2746–2758, 2024, doi: 10.1109/TIT.2024.3367433.
[13] B. Yang, Z. Jia, J. Yang, and N. K. Kasabov, "Video Snow Removal Based on Self-Adaptation Snow Detection and Patch-Based Gaussian Mixture Model," *IEEE Access*, vol. 8, pp. 160188–160201, 2020, doi: 10.1109/ACCESS.2020.3020619.
[14] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," *IEEE Access*, vol. 10, pp. 21517–21525, 2022, doi: 10.1109/ACCESS.2022.3152828.
[15] S. Al-Azani and E. S. M. El-Alfy, "Enhanced Video Analytics for Sentiment Analysis Based on Fusing Textual, Auditory and Visual Information," *IEEE Access*, vol. 8, pp. 136843–136857, 2020, doi: 10.1109/ACCESS.2020.3011977.
[16] M. Y. Khan, T. Ahmed, M. S. Siddiqui, and S. Wasi, "Cognitive Relationship-Based Approach for Urdu Sarcasm and Sentiment Classification," *IEEE Access*, vol. 11, no. November, pp. 126661–126690, 2023, doi: 10.1109/ACCESS.2023.3325048.
[17] J. Sun, P. Han, Z. Cheng, E. Wu, and W. Wang, "Transformer Based Multi-Grained Attention Network for Aspect-Based Sentiment Analysis," *IEEE Access*, vol. 8, pp. 211152–211163, 2020, doi: 10.1109/ACCESS.2020.3039470.
[18] H. Zhang *et al.*, "Multidimensional Extra Evidence Mining for Image Sentiment Analysis," *IEEE Access*, vol. 8, no. 1, pp. 103619–103634, 2020, doi: 10.1109/ACCESS.2020.2999128.
[19] M. F. R. Abu Bakar, N. Idris, L. Shuib, and N. Khamis, "Sentiment Analysis of Noisy Malay Text: State of Art, Challenges and Future Work," *IEEE Access*, vol. 8, pp. 24687–24696, 2020, doi: 10.1109/ACCESS.2020.2968955.
[20] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," *IEEE Access*, vol. 8, pp. 131662–131682, 2020, doi: 10.1109/ACCESS.2020.3009626.
[21] G. M. Shafiq, T. Hamza, M. F. Alrahmawy, and R. El-Deeb, "Enhancing Arabic Aspect-Based Sentiment Analysis Using End-to-End Model," *IEEE Access*, vol. 11, no. November, pp. 142062–142076, 2023, doi: 10.1109/ACCESS.2023.3342755.
[22] M. Bibi, W. Aziz, M. Almaraashi, I. H. Khan, M. S. A. Nadeem, and N. Habib, "A Cooperative Binary-Clustering Framework Based on Majority Voting for Twitter Sentiment Analysis," *IEEE Access*, vol. 8, pp. 68580–68592, 2020, doi: 10.1109/ACCESS.2020.2983859.
[23] Z. Kastrati, A. S. Imran, S. M. Daudpota, M. A. Memon, and M. Kastrati, "Soaring Energy Prices: Understanding Public Engagement on Twitter Using Sentiment Analysis and Topic Modeling with Transformers," *IEEE Access*, vol. 11, no. February, pp. 26541–26553, 2023, doi: 10.1109/ACCESS.2023.3257283.
[24] S. M. A. H. Shah, S. F. H. Shah, A. Ullah, A. Rizwan, G. Atteia, and M. Alabdulhafith, "Arabic Sentiment Analysis and Sarcasm Detection Using Probabilistic Projections-Based Variational Switch Transformer," *IEEE Access*, vol. 11, no. June, pp. 67865–67881, 2023, doi: 10.1109/ACCESS.2023.3289715.
[25] E. Elbasani and J. D. Kim, "AMR-CNN: Abstract Meaning Representation with Convolution Neural Network for Toxic Content Detection," *J. Web Eng.*, vol. 21, no. 3, pp. 677–692, 2022, doi: 10.13052/jwe1540-9589.2135.
[26] Y. A. Singgalen, "Toxicity Analysis and Sentiment Classification of Wonderland Indonesia by Alffy Rev using Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 5, no. 3, pp. 538–548, 2024, doi: 10.30865/json.v5i3.7563.
[27] Y. A. Singgalen, "Sentiment Classification of Food Influencer Content Reviews using Support Vector Machine Model through CRISP-DM Framework," *J. Sist. Komput. dan Inform.*, vol. 5, no. 3, pp. 517–528, 2024, doi: 10.30865/json.v5i3.7509.
[28] G. Etta, M. Cinelli, N. Di Marco, M. Avalle, A. Panconesi, and W. Quattrociocchi, "A Topology-Based Approach for Predicting Toxic Outcomes on Twitter and YouTube," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 5, pp. 4875–4885, 2024, doi: 10.1109/TNSE.2024.3398219.
[29] V. Ganganwar and R. Rajalakshmi, "Employing synthetic data for addressing the class imbalance in aspect-based sentiment classification," *J. Inf. Telecommun.*, pp. 1–22, 2023, doi: 10.1080/24751839.2023.2270824.
[30] Q. Song and A. Wondirad, "Demystifying the nexus between social media usage and overtourism: evidence from Hangzhou, China," *Asia Pacific J. Tour. Res.*, vol. 28, no. 4, pp. 364–385, 2023, doi: 10.1080/10941665.2023.2230313.

[31] Q. Lu, Z. Zhu, D. Zhang, W. Wu, and Q. Guo, "Interactive Rule Attention Network for Aspect-Level Sentiment Analysis," *IEEE Access*, vol. 8, pp. 52505–52516, 2020, doi: 10.1109/ACCESS.2020.2981139.
[32] D. Flonk, M. Jachtenfuchs, and A. Obendiek, "Controlling internet content in the EU: towards digital sovereignty," *J. Eur. Public Policy*, vol. 31, no. 8, pp. 2316–2342, 2024, doi: 10.1080/13501763.2024.2309179.