Jurnal Teknik Informatika C.I.T Medicom, 16 (3) (2024) 115-128

Published by: Institute of Computer Science (IOCS)

Jurnal Teknik Informatika C.I.T Medicom

Journal homepage: www.medikom.iocspublisher.org

Toxicity, social network and topic analysis of digital content: Perspective and multilingual embedding model

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

Article Info

Article history

Received : Mar 10, 2024
Revised : Apr 19, 2024
Accepted : Jun 28, 2024

Keywords:

Clustering techniques;
Digital content analysis;
Online community engagement
Social Network Analysis (SNA);
Toxicity score.

Abstract

This research presents a comprehensive approach to analyzing digital content by integrating toxicity analysis, clustering techniques, and Social Network Analysis (SNA) to understand online interactions better. The study finds that, while the average toxicity levels are relatively low, with scores such as 0.06355 for toxicity and 0.00468 for severe toxicity, there are significant spikes, reaching maximum scores of 0.82996 for toxicity and 0.89494 for profanity. These spikes highlight the necessity for continuous monitoring and adaptive moderation strategies to minimize the impact of harmful language. Clustering methods, including K-Means, HDBScan, and Gaussian Mixture models, provide deep insights into the thematic structure of viewer discourse, identifying both prevalent and niche topics. The Gaussian Mixture model identified ten distinct clusters, while HDBScan revealed varying cluster densities, reflecting the diverse range of discussions within the community. In addition, SNA, with 1,716 nodes and 37 edges, offers critical insights into the relational dynamics of the network, pinpointing key influencers and mapping the flow of information between different user groups. By synthesizing these methodologies, the research provides a robust framework for understanding the content and context of digital interactions, facilitating more effective strategies for enhancing community engagement, mitigating toxicity, and promoting a healthier, more inclusive online environment.

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, Daerah Khusus Ibukota Jakarta 12930, Indonesia
yerik.afrianto@atmajaya.ac.id

1. Introduction

The urgency of examining toxicity within digital content through social network and topic analysis, particularly from a multilingual embedding model perspective, arises from the growing prevalence and impact of harmful online behaviors across diverse cultural and linguistic contexts. Digital platforms, increasingly utilized for communication, education, and socialization, have become breeding grounds for toxic behaviors that can cause significant psychological and societal harm [1]–[3]. Addressing this issue requires an in-depth understanding of the nature and spread of toxicity in various digital environments and the linguistic nuances that contribute to the escalation or mitigation of such behaviors [4]–[8]. By employing a multilingual embedding model, the study facilitates a more nuanced

analysis of toxic digital content across different languages and cultural settings, providing a more comprehensive understanding of global online toxicity patterns. Such an approach also allows for identifying specific linguistic and cultural markers that may exacerbate or mitigate online toxicity, thus offering valuable insights for developing more effective moderation tools and policies [9]–[12]. In conclusion, integrating social network analysis, topic modeling, and multilingual embedding in this research underscores its critical role in advancing the knowledge base necessary to combat digital toxicity, fostering a safer and more inclusive online environment globally.

This research aims to develop a comprehensive framework for analyzing toxicity in digital content by leveraging social networks and topic analysis, focusing on multilingual embedding models. This approach is designed to capture the complex dynamics of toxic behavior within diverse online communities, encompassing various languages and cultural contexts. Including multilingual embeddings allows for a more refined analysis, considering the subtleties of different languages and how these affect the propagation of toxic content [13]–[18]. Moreover, by integrating social network analysis, the study seeks to uncover the structural patterns that facilitate or inhibit the spread of harmful language, providing insights into the role of community dynamics in moderating digital toxicity [19]–[22]. By addressing these objectives, the research aims to contribute significantly to developing more sophisticated tools and strategies for detecting and mitigating online toxicity, thereby promoting a healthier and more inclusive digital discourse across the global digital landscape.

The novelty of this research lies in its innovative integration of multilingual embedding models with social networks and topic analysis to explore toxicity in digital content across diverse linguistic and cultural landscapes. Unlike traditional approaches that often focus on a single language or rely on monolingual datasets, this study incorporates a multilingual framework, enabling a more inclusive and comprehensive examination of online toxicity. This methodology allows for detecting nuanced linguistic and cultural patterns that contribute to toxic behavior, which are frequently overlooked in more conventional analyses. Furthermore, by combining social network analysis with advanced topic modeling, the study offers a dual perspective on the structural dynamics of digital interactions and the thematic content that fuels toxic discourse. Such a multidisciplinary approach enhances the understanding of how toxic content spreads and evolves and paves the way for developing more effective and culturally sensitive interventions. Ultimately, this research represents a significant advancement in the field by addressing the complexities of digital toxicity in a more holistic and globally relevant manner.

This research's theoretical and practical implications are profound, offering new insights into the mechanisms of digital toxicity and proposing actionable strategies for its mitigation. Theoretically, this study contributes to the existing body of knowledge by advancing a multilingual embedding model that captures the subtleties of toxic language across various cultural and linguistic contexts, thereby enriching the understanding of how toxicity manifests and spreads in digital environments [23]–[27]. This nuanced perspective challenges traditional theories that often overlook the complexity of global online interactions. Practically, the findings hold significant promise for enhancing content moderation strategies on digital platforms, providing more precise tools for identifying and managing harmful content in real time. The study offers practical frameworks that digital platforms can adopt to foster safer and more inclusive online communities by applying a combined approach of social network analysis and topic modeling [28]–[31]. In summary, the research pushes the boundaries of theoretical frameworks on digital toxicity and provides tangible solutions for addressing this pervasive issue in a globally interconnected digital landscape.

Similar research in digital toxicity analysis has primarily focused on monolingual contexts, social network behaviors, or topic-specific content. Yet, few studies have effectively combined these elements with a multilingual perspective. Previous investigations have explored toxic behavior patterns on social media platforms, highlighting the role of network structures and user interactions in propagating harmful content [32]–[35]. While these studies provide valuable insights into the dynamics of online toxicity, their limitation to single-language analysis constrains their applicability in a globally connected environment where digital interactions often transcend linguistic boundaries. Additionally,

some research has utilized topic modeling to identify themes associated with toxic discourse. Yet, the absence of a multilingual embedding model has restricted a deeper understanding of how different languages and cultural nuances influence toxicity [31], [36]–[38]. By contrast, this study's integrative approach, which merges social network analysis, topic modeling, and multilingual embeddings, addresses these gaps by offering a more comprehensive framework for analyzing digital toxicity. Consequently, it represents a significant advancement over prior research efforts, providing a more holistic view of the multifaceted nature of online toxicity and its cross-cultural implications.

One notable limitation of this research is the potential challenge of accurately capturing the full complexity of digital toxicity across diverse linguistic and cultural contexts. Although integrating multilingual embedding models offers a more nuanced analysis, there remains a risk that certain linguistic subtleties or cultural idioms might not be fully represented within the model, potentially leading to incomplete or skewed interpretations of toxic behavior. Additionally, the reliance on social network and topic modeling approaches might not entirely account for the evolving nature of digital interactions and the rapid emergence of new forms of toxic discourse. Such methodological constraints suggest that while the study provides a significant step forward, there is still room for further refinement and expansion regarding data diversity and analytical tools. A broader dataset encompassing a more comprehensive array of languages and dialects and developing more sophisticated models could mitigate these limitations, enhancing the robustness and applicability of future research in this domain.

Future research should incorporate more diverse datasets and employ advanced analytical models to address better the complexities highlighted in this study. Broadening the scope to encompass a more comprehensive array of languages and cultural contexts would offer a more comprehensive perspective on how toxic behaviors emerge and evolve across different digital landscapes. Developing more sophisticated machine learning models that can more accurately detect nuanced expressions of toxicity such as sarcasm, coded language, or rapidly changing slang would be particularly beneficial. Furthermore, conducting longitudinal studies to observe the persistence and evolution of toxic behaviors over time could provide valuable insights into the effectiveness of various moderation strategies and the resilience of online communities against toxic influences. Adopting interdisciplinary approaches that integrate insights from linguistics, sociology, and computer science would further enhance the depth and relevance of future research. These strategies would significantly advance the development of more effective tools and policies for managing digital toxicity in an increasingly interconnected world.

2. Research Methodology

The Digital Content Reviews and Analysis Framework provides a comprehensive approach to evaluating digital content by integrating multiple analytical methodologies to understand toxicity, social dynamics, and thematic patterns. This framework systematically combines content reviews and text data processing with advanced analytical techniques, such as topic modeling using HDBScan, K-Means, and Gaussian Mixture models, to uncover latent themes within digital discourse. Additionally, it employs social network analysis to explore the relational structures that influence the dissemination and impact of toxic content within digital communities. The subsequent phases of data evaluation and visualization facilitate the interpretation of complex data, enabling a robust context analysis that informs a deeper understanding of the interconnectedness between content, social networks, and toxicity. By merging these diverse methods, the framework offers a holistic view of digital content analysis, positioning itself as a critical tool for developing targeted interventions and promoting healthier online interactions.

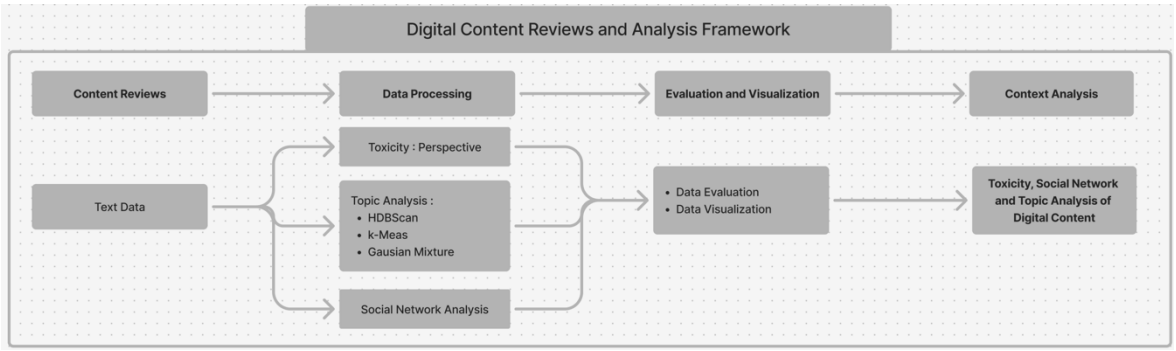

Figure. 1. Implementation of Digital Content Reviews and Analysis Framework

The dataset "LOCKDOWN IN PARADISE BALI: Best Bali Bamboo House Tour," collected from YouTube, serves as a rich source of digital content for analysis, encompassing 1,831 records gathered under specific search parameters. This dataset, retrieved on August 27, 2024, offers unique insights into the interplay between content creation and audience engagement on social media platforms during a period marked by global disruptions. It is argued that the specificity of this dataset, focusing on a particular geographic and cultural context, provides an opportunity to examine how digital narratives are shaped by local and global influences, particularly in the context of lifestyle and travel content. Through a detailed analysis of the video content and viewer interactions, patterns related to digital engagement and community dynamics during crises can be uncovered, shedding light on the broader implications for content strategies on platforms like YouTube. Ultimately, this dataset is pivotal for understanding the nuances of digital communication and its impacts within both localized and broader global settings.

The video, which has garnered 1,151,890 views since its release on April 19, 2020, accompanied by 1,831 comments, illustrates significant audience engagement and interest. This high view count suggests that the content resonates strongly with viewers, possibly due to its relevance, unique subject matter, or visual and narrative style appeal. It is posited that the volume of comments also indicates a robust level of interaction, reflecting viewers' willingness to engage in discourse, share opinions, and connect over shared experiences or perspectives. Analyzing these comments can provide valuable insights into the audience's perceptions, cultural values, and emotional responses, offering a deeper understanding of how digital content influences and is influenced by its viewers. Consequently, this engagement highlights the video's impact in fostering a digital community. It underscores the importance of analyzing quantitative metrics, such as views, and qualitative feedback, such as comments, to comprehend digital content consumption dynamics fully.

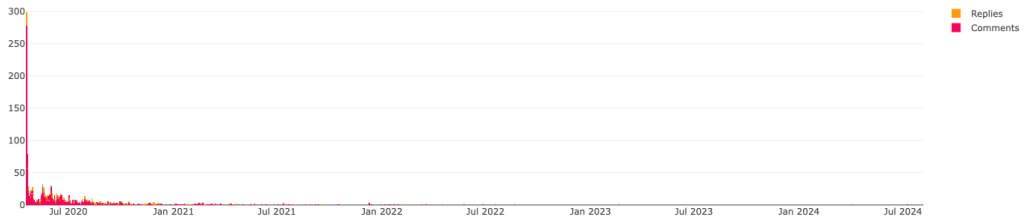

Figure. 2. Post-per-day statistic

The post-per-day statistics for the YouTube video reveal significant fluctuations in audience engagement over time, reflecting varying levels of interest and interaction among viewers. The data indicate an initial spike in activity shortly after the video's release, with a notable peak on April 19, 2020, registering 278 posts, which suggests a robust initial response likely driven by the content's relevance or virality at that moment. Subsequently, there is a gradual decline in daily posts,

interspersed with occasional small peaks, which may correspond to renewed interest triggered by external factors such as sharing on social media or featuring in related content recommendations. This pattern illustrates the typical digital content life cycle, where engagement diminishes over time as the content becomes less novel or viewers shift their focus to newer material. Analyzing this temporal engagement data provides valuable insights into content lifespan and viewer behavior, underscoring the importance of continuously adapting digital content strategies to maintain audience interest and sustain engagement over more extended periods.

As depicted in the visual data, the post-per-day statistics reveal a significant initial surge in activity, followed by a sharp decline over time, illustrating a familiar pattern of audience engagement with digital content. The highest concentration of comments and replies occurred shortly after the video's release, reflecting heightened interest and immediate viewer reactions during this period. This early peak indicates the video's immediate impact, possibly driven by its topical relevance or viral potential, which tends to attract substantial attention in the initial phase. However, the subsequent decline in daily posts suggests a tapering of engagement as the novelty of the content diminished or the audience's attention shifted to newer material. The long tail observed in the data points to a sustained, albeit lower, level of interaction over an extended period, which may reflect ongoing interest or occasional rediscovery by new viewers. Overall, the pattern suggests that while initial engagement levels are crucial for a video's short-term impact, maintaining prolonged viewer interaction requires continuous or renewed relevance in a rapidly evolving digital landscape.

Figure. 3. Word Cloud

The frequency analysis of words in the word cloud provides a quantitative insight into the audience's predominant interests and reactions to the video content, with terms like "banget" (57), "video" (56), "place" (52), and "love" (50) appearing frequently, indicating a robust positive engagement. The frequent use of words such as "Bali" (40), "keren" (40), "beautiful" (38), and "amazing" (21) underscores the viewers' admiration for the location and aesthetics presented, suggesting that the visual appeal and cultural context of Bali significantly drive viewer engagement. Additionally, the presence of words like "nyamuk" (35) and "snakes" (9) reflects viewers' attention to specific, perhaps less favorable, elements mentioned or depicted in the video, indicating a comprehensive engagement with both the positive and negative aspects of the content. The diverse vocabulary, including colloquial expressions like "gw" (32), "kak" (20), and "bro" (16), as well as more formal descriptors such as "quality" (11) and "beautiful" (38), reveals a mixed audience demographic engaging in informal yet thoughtful discussions. This linguistic variety, combined with the frequency of these words, highlights the cultural specificity and diverse audience base and provides insights into how different viewer segments perceive and interact with the video. This analysis suggests that the video successfully resonates with a broad audience by balancing captivating visual storytelling with elements that prompt varied emotional and cognitive responses.

The word cloud generated from the video's comment section visually represents the most frequently used words, highlighting the audience's primary sentiments and topics of interest.

Dominant terms such as "love," "beautiful," "place," and "Bali" suggest a positive reception and a strong emotional connection to the video's content, likely driven by the aesthetic appeal of the location showcased. Words like "amazing," "nice," "awesome," and "banget" reinforce this positive sentiment, indicating high viewer satisfaction and enthusiasm. Additionally, the prominence of words such as "villa," "house," "nature," and "bamboo" reflects specific interests related to the architectural and natural elements featured in the video, pointing to a preference for content that explores unique living spaces and natural settings. The use of colloquial terms and informal expressions, such as "gw," "lu," and "bang," further suggests a casual and familiar tone among viewers, indicating a culturally specific audience engagement. This analysis of the word cloud underscores the importance of aligning content themes with audience interests to foster deeper viewer engagement and enhance the overall impact of digital content.

Figure. 4. Top Ten Poster

The analysis of the top ten posters in the video's comment section reveals a concentrated level of engagement from a select group of users, highlighting patterns of active participation and community dynamics. The most active commenter, @RIZFERD, contributed ten posts, representing 20.8% of the total comments from the top ten users, which suggests an exceptionally high level of interest or investment in the video content. Following closely is @BackpackerTampan, with nine comments, accounting for 18.8% of the contributions, further underscoring the engagement of core community members. Other users such as @dianakw9159, @yusefendriraharjooneone, @ak3885, and @ahaayyyasyeekkkashooyyy1047 contributed between 4 to 5 comments each, collectively making up a significant portion of the discussion. In contrast, users like @sonnn9866, @generasiemas2809, @gmass6987, and @happykids6900 posted three comments, indicating varying degrees of interaction. This distribution suggests that a small group of highly engaged users is driving much of the discussion, which could reflect either a strong community following or the polarizing nature of the content. Understanding the behavior and motivations of these top contributors is crucial, as their interactions likely influence the broader audience's perceptions and engagement with the content, thereby shaping the overall discourse in the comment section.

Analyzing digital engagement patterns is essential to understanding how content resonates with diverse audiences and influences viewer behavior. A comprehensive picture of audience preferences and interaction dynamics emerges by examining quantitative metrics, such as view counts and comment frequencies, alongside qualitative insights from content and sentiment analysis. It is posited that integrating multiple analytical approaches, including social network analysis and natural language processing, allows for a more nuanced understanding of the factors driving engagement and discourse. For example, the prevalence of specific keywords or phrases in comments reflects viewer sentiment and highlights emerging trends and topics of interest within specific communities. This multifaceted approach reveals the interplay between content, audience perception, and platform dynamics, providing valuable insights for optimizing digital strategies. In conclusion, robust analytical frameworks are indispensable for dissecting complex digital ecosystems and tailoring content to meet a global audience's evolving needs and preferences.

The relevance of the data to this research is pivotal in providing empirical insights into digital engagement and content dynamics within online communities. This dataset, which encompasses extensive viewer interactions and comments, is a critical source for examining patterns of audience behavior, sentiment, and discourse related to specific digital content. It is argued that such data is invaluable for understanding the factors that contribute to both the proliferation of online engagement and the formation of virtual communities. Through comprehensive analysis, including keyword frequency, sentiment analysis, and network modeling, the dataset offers a robust foundation for investigating how digital content influences viewer perceptions and behaviors across diverse platforms. This relevance is further underscored by the potential to identify trends, measure content effectiveness, and develop strategies for enhancing user engagement. Ultimately, the dataset's alignment with the research objectives ensures a deeper exploration of the complex interactions within digital environments, contributing significantly to the field's advancement.

3. Result and Disscusion

Toxicity Score and Interpretation

Calculating the toxicity score based on the data from this research is crucial for quantifying the prevalence and intensity of harmful language within digital content. This score provides a standardized measure that enables a nuanced assessment of how toxic behavior manifests and evolves across various online platforms and communities. It is posited that incorporating a toxicity score into the analysis allows for a more objective evaluation of digital discourse, facilitating the identification of patterns and trends that may otherwise remain obscured in qualitative assessments. By systematically quantifying toxicity, it becomes possible to compare levels of harmful language across different datasets, track changes over time, and understand the factors contributing to heightened toxicity in specific contexts. Furthermore, this approach aids in developing targeted interventions and policies to reduce toxicity and foster healthier online environments. In conclusion, the calculation of toxicity scores is an essential component of this research, providing valuable insights into the dynamics of online interactions and contributing to the broader goal of promoting constructive digital communication.

Communalytic's analysis of 1,536 posts from 1,831 using the Perspective API provides a detailed quantitative assessment of various forms of toxicity within the dataset. The average toxicity score of 0.06355 indicates a relatively low prevalence of harmful language; however, the highest toxicity value recorded is 0.82996, revealing significant spikes in specific posts. Similarly, scores for severe toxicity, identity attacks, insults, profanity, and threats show low averages—0.00468, 0.01222, 0.03005, 0.05230, and 0.01000, respectively—but peak values of 0.23225, 0.55045, 0.71027, 0.89494, and 0.34877 illustrate the occasional presence of more extreme toxic behaviors. It is argued that while the dataset predominantly contains low toxicity levels, the high maximum scores for specific categories suggest areas where content moderation may be necessary to mitigate potential harm. This analysis highlights the variability within online discourse, where a small subset of highly toxic content can significantly impact the overall tone and safety of the digital environment. In conclusion, these findings underscore the importance of continuous monitoring and strategic intervention to foster more positive and inclusive online communities.

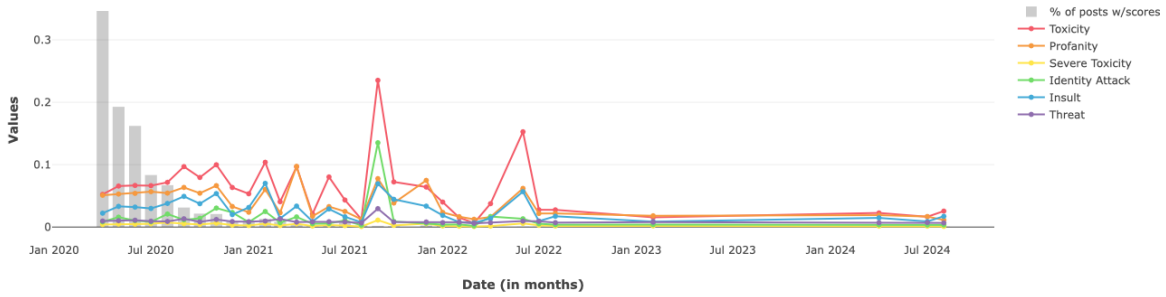

Figure. 5. Average Toxicity Score per Month (Communalytic)

The analysis of this dataset, with its calculated averages and highest values for various toxicity metrics, provides crucial insights into the nature of harmful language within digital interactions and their relevance to this research. The average scores for toxicity (0.06355), severe toxicity (0.00468), identity attack (0.01222), insult (0.03005), profanity (0.05230), and threat (0.01000) suggest that, on average, the level of harmful content remains relatively low across the dataset. However, the significantly higher maximum values, such as 0.82996 for toxicity and 0.89494 for profanity, indicate that while toxic behavior is not pervasive, it can reach extreme levels in isolated instances. This variability aligns with the research's objective to understand the dynamics of digital toxicity, emphasizing the sporadic yet intense occurrences of harmful content that can disproportionately affect the online environment. It is posited that these findings are essential for identifying critical points where intervention might be necessary to prevent the escalation of toxicity. Thus, the data's relevance lies in its ability to highlight both the pervasive and extreme aspects of digital communication, underscoring the importance of targeted strategies to mitigate harm and promote healthier online discourse.

The toxicity score is a valuable tool in analyzing viewer behavior and perception, providing quantifiable measures of the extent and nature of harmful language within online interactions. By offering a standardized metric, the toxicity score facilitates the identification of patterns in viewer discourse that may indicate broader trends in sentiment, such as negativity, hostility, or polarization, which are crucial for understanding the underlying dynamics of digital engagement. It is argued that utilizing toxicity scores enables a more nuanced interpretation of viewer behavior, allowing for the differentiation between casual negative comments and more severe forms of toxicity, such as identity attacks or threats. This differentiation is essential for assessing the impact of content on audience perception and identifying areas where the discourse may become detrimental to the community's health. Furthermore, analyzing toxicity scores over time can provide insights into how viewer perceptions evolve in response to content and social interactions, highlighting shifts that may necessitate strategic content moderation or community management interventions. In conclusion, the toxicity score is an indispensable metric for comprehensively understanding and managing viewer behavior and perception within digital ecosystems.

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

The implementation of topic analysis using HDBScan, K-means, and Gaussian Mixture models provides a robust framework for uncovering underlying themes and patterns within the dataset, which is crucial for understanding the context of digital interactions in this research. Each of these clustering algorithms offers unique advantages: HDBScan effectively identifies clusters of varying density, making it suitable for detecting nuanced topic variations in discussions; K-means provides a straightforward partitioning of data into distinct clusters, ideal for segmenting comments into clearly defined thematic categories; and Gaussian Mixture models account for the probability distributions within clusters, allowing for a more refined analysis of overlapping themes. It is posited that applying these methodologies enhances the ability to discern dominant and subtle topics that emerge from viewer engagement, providing a more comprehensive picture of audience interests, sentiments, and discourse dynamics. This approach not only facilitates the identification of critical areas of concern or interest

within the digital content but also aids in understanding how different audience segments might perceive or react to various aspects of the content. In conclusion, combining HDBScan, K-means, and Gaussian Mixture models in topic analysis enables a deeper, more granular data exploration, thereby enriching the insights derived from this research.

Implementing topic analysis based on HDBScan (Hierarchical Density-Based Spatial Clustering of Applications with Noise) provides a sophisticated method for identifying meaningful clusters in complex datasets, particularly when analyzing digital content. Unlike traditional clustering algorithms, HDBScan does not require the pre-specification of the number of clusters, allowing it to adaptively find clusters of varying shapes and densities, which is particularly useful for capturing the diversity of topics that emerge from user-generated content. It is argued that HDBScan's ability to handle noise and outliers enhances the accuracy of topic detection by minimizing the impact of irrelevant or anomalous data points, yielding more reliable insights. This approach enables a more nuanced exploration of the underlying thematic structures within the dataset, revealing not only dominant topics but also uncovering less apparent patterns that may otherwise be overlooked. Moreover, the flexibility of HDBScan in recognizing clusters of different densities allows for a more comprehensive understanding of the complex dynamics present in digital interactions. In conclusion, employing HDBScan for topic analysis offers significant advantages in accurately modeling and interpreting the intricate patterns of discourse found in large and unstructured data, making it a valuable tool for advanced content analysis.

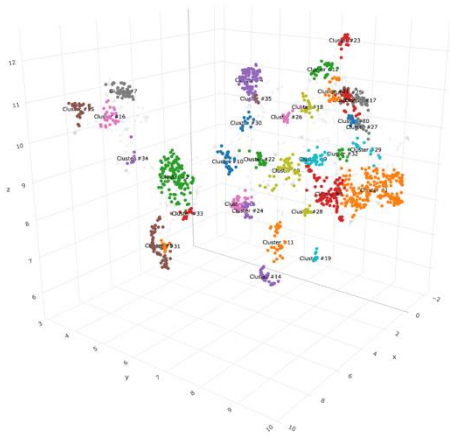

Figure. 6. Topic Analysis based on Cluster (HBDScan) using Communalytic

Figure 6 illustrates a topic analysis based on clustering using the HDBScan algorithm, effectively identifying clusters of varying densities in the dataset. This visualization, generated from the "LOCKDOWN IN PARADISE BALI" dataset on YouTube comprising 1,815 records, demonstrates how HDBScan's parameter settings (minimum cluster size: 10, minimum samples: 10, epsilon: 0.05) allow for the detection of distinct thematic groupings within the data. It is argued that HDBScan's capacity to manage noise and detect clusters of different shapes enhances its utility in exploring complex digital interactions, revealing prominent and subtle topics. The 3D plot shows diverse clusters, each representing a unique topic or discussion theme, providing a more comprehensive view of the discourse dynamics. This approach enables a deeper understanding of the content's thematic landscape by highlighting major topics attracting widespread engagement and niche topics catering to specific audience segments. In conclusion, HDBScan is a powerful tool for topic analysis in digital content, offering nuanced insights into the varied themes and discussions that shape online interactions.

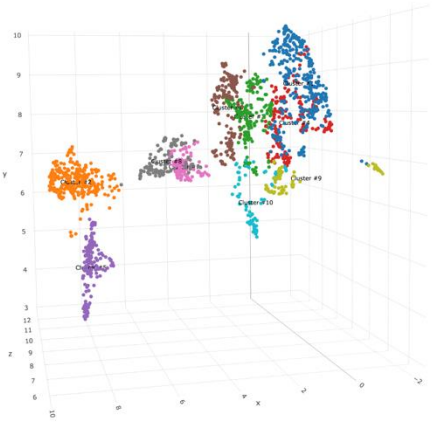

Figure. 7. Topic Analysis based on Cluster (KMeans) using Communalytic

Figure 7 presents a topic analysis based on clustering using the KMeans algorithm, effectively visualizing the distribution of topics within the dataset through distinct cluster formations. Each cluster represents a group of data points with similar characteristics, indicating that the content within these clusters shares thematic or contextual similarities. The KMeans algorithm's partitioning approach allows for clear delineation between topics, making it possible to identify dominant themes and subthemes in the viewer's discourse. By creating well-separated clusters, it is argued that this clustering method facilitates a more straightforward interpretation of the primary topics and their interrelations, enhancing the understanding of the underlying structures within digital communication. The visualization also highlights each cluster's relative density and spread, providing insights into the popularity or concentration of specific topics over others. Consequently, this analysis aids in recognizing which topics generate the most engagement and how they are interlinked within the digital content landscape. In conclusion, KMeans clustering is a valuable tool for elucidating complex thematic patterns, offering a clearer perspective on the dynamics of online discussions and viewer interactions.

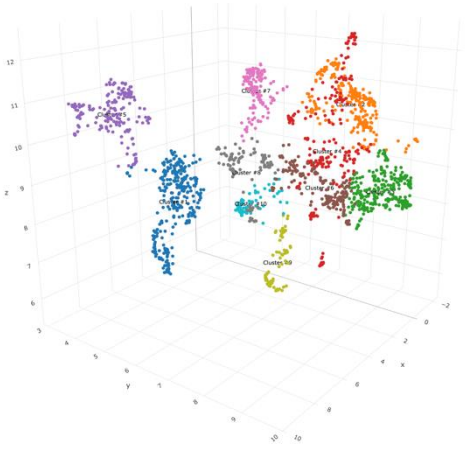

Figure. 8. Topic Analysis based on Cluster (Gaussian Mixture) using Communalytic

Figure 8 depicts a topic analysis utilizing the Gaussian Mixture clustering algorithm, providing a probabilistic approach to understanding the distribution of topics within the dataset "LOCKDOWN IN PARADISE BALI," which comprises 1,815 records from YouTube. The Gaussian Mixture model, characterized by its ability to assign data points to multiple clusters based on probability, facilitates a nuanced examination of overlapping themes and latent structures within the data. This method's

flexibility in handling the complexity of digital content is precious, as it allows for identifying topics that may not be strictly separated but rather exist on a spectrum of related themes. It is argued that this approach enhances the interpretative depth of the analysis by acknowledging the multifaceted nature of viewer discourse and content interaction. The visual representation of clusters shows a range of topics with varying densities and overlaps, indicating diverse yet interconnected discussions among viewers. In conclusion, applying the Gaussian Mixture model for topic analysis provides a more refined understanding of thematic relations, contributing significantly to the overall comprehension of digital content dynamics and audience engagement.

Social Network Analysis

The importance of Social Network Analysis (SNA) lies in its ability to provide a detailed understanding of the relational dynamics and interaction patterns within digital communities. SNA enables the mapping of relationships between users, content, and their interactions, revealing how information flows and influence is distributed across a network. It is posited that employing SNA allows for identifying key nodes or influential users whose actions significantly shape community discourse and engagement. By analyzing these connections, SNA helps uncover the central figures within a network and the potential clusters or subgroups that form around shared interests or topics. This insight is crucial for recognizing cohesive and fragmented areas within the network, offering strategic opportunities for fostering more inclusive and interactive digital environments. In conclusion, SNA is an indispensable tool for dissecting the complexities of online interactions, providing valuable insights that support effective community management and engagement strategies.

Figure. 9. Social Network Analysis (SNA) using Communalytic

The Social Network Analysis (SNA) illustrated in the figure, consisting of 1,716 nodes (posts) and 37 edges, offers a comprehensive view of the interaction patterns and connectivity within the digital content community. The nodes represent individual posts, while the edges signify direct interactions or relationships between these posts, such as replies or references, providing a map of communication dynamics. It is posited that the relatively high number of nodes compared to the limited number of edges suggests a sparse network where many posts exist independently without extensive interconnection, potentially indicating a lack of cohesive discussion or frequent engagement between viewers. This sparsity might reflect a community where engagement is primarily driven by individual contributions rather than collective discourse, or it could suggest a platform context where user interactions are less direct. Visualizing these network dynamics enables the identification of isolated clusters or highly active nodes that may represent influential users or content-driving engagement. In conclusion, the SNA provides critical insights into the structural properties of digital interactions,

highlighting the need for strategies that could foster more connected and interactive community engagement.

The visualization of Social Network Analysis (SNA) in conjunction with K-Means clustering offers a powerful approach to understanding the structure and dynamics of digital interactions within a community. By integrating SNA with K-Means, which partition data into distinct clusters, this visualization enables a more precise identification of user groups based on their interaction patterns and thematic affinities. It is argued that this combined approach enhances the understanding of how different clusters communicate internally and how they relate to each other, revealing potential bridges or gaps in communication across the network. The application of K-Means clustering helps to simplify complex networks by grouping similar nodes, making it easier to detect central figures, influential users, and isolated clusters within the network. This methodology provides insights into the flow of information and the formation of subgroups, which are crucial for targeted community management and fostering a more cohesive digital environment. In conclusion, the relevance of combining SNA visualization with K-Means clustering lies in its ability to provide a more granular and comprehensive analysis of community dynamics, thereby supporting more effective engagement and content dissemination strategies.

Connecting Social Network Analysis (SNA) with toxicity analysis offers a comprehensive framework for understanding the structure of digital interactions and the spread and impact of harmful content within online communities. By integrating SNA, which maps the relationships and influence among users, with toxicity scores that quantify the presence of harmful language, it becomes possible to identify how toxic behaviors proliferate across different network segments. It is argued that this combined approach provides deeper insights into the dynamics of online toxicity, revealing whether specific clusters or influential nodes are central to disseminating harmful content. Through this integration, patterns of toxic discourse can be traced, allowing for the detection of highly active or influential users who may contribute disproportionately to the spread of toxicity. This understanding is essential for developing targeted interventions and content moderation strategies that mitigate harmful behaviors and promote healthier and more constructive engagement. In conclusion, the linkage between SNA and toxicity analysis enhances the capacity to monitor, understand, and address toxic behaviors within digital ecosystems, thereby supporting more robust and proactive community management efforts.

4. Conclusion

The conclusion of this research underscores the significance of a multidimensional approach to analyzing digital content by integrating toxicity analysis, clustering techniques, and Social Network Analysis (SNA). The findings reveal that, although the average toxicity levels are relatively low—such as an average toxicity score of 0.06355 and a severe toxicity score of 0.00468—there are instances where toxicity spikes significantly, with maximum scores reaching 0.82996 for toxicity and 0.89494 for profanity. These spikes indicate the need for continuous monitoring and adaptive moderation strategies to mitigate the impact of harmful language. Clustering techniques, including K-Means, HDBScan, and Gaussian Mixture models, offer valuable insights into the thematic structure of viewer discourse, revealing both dominant topics and niche discussions. The analysis identified 10 clusters using Gaussian Mixture and varying cluster densities with HDBScan, illustrating the diverse range of discussions within the community. SNA, which mapped 1,716 nodes (posts) and 37 edges, provides a critical perspective on the relational dynamics within the network, identifying key influencers and illustrating how information flows across different user groups. By integrating these methodologies, the research offers a comprehensive understanding of the content and the context of digital interactions, supporting more effective strategies for enhancing community engagement, reducing toxicity, and fostering a healthier, more inclusive online environment. This integrative framework is crucial for developing targeted interventions and advancing the field of digital content analysis.

References

[1] E. Surucu-Balci and G. Balci, "Building social capital in cruise travel via social network sites," *Curr. Issues Tour.*, vol. 26, no. 7, pp. 1096–1111, 2023, doi: 10.1080/13683500.2022.2047904.

[2] E. Rosamond, "YouTube personalities as infrastructure: assets, attention choreographies and cohortification processes," *Distinktion*, vol. 24, no. 2, pp. 254–282, 2023, doi: 10.1080/1600910X.2023.2185873.

[3] E. King, "Gaming race in Brazil: Video games and algorithmic racism," *J. Lat. Am. Cult. Stud.*, vol. 33, no. 1, pp. 149–165, 2024, doi: 10.1080/13569325.2024.2307540.

[4] D. Tauro, U. Panniello, and R. Pellegrino, "Risk Management in Digital Advertising: An Analysis from the Advertisers' Media Management Perspective," *JMM Int. J. Media Manag.*, vol. 23, no. 1–2, pp. 29–57, 2021, doi: 10.1080/14241277.2021.1960532.

[5] V. Blumenthal, M. Lerfald, and K. Blekastad Sagheim, "'Hotels are much easier': motivation for non-participation in travel-related sharing economy exchanges," *Curr. Issues Tour.*, pp. 1–17, 2024, doi: 10.1080/13683500.2024.2309158.

[6] N. Dens and K. Poels, "The rise, growth, and future of branded content in the digital media landscape," *Int. J. Advert.*, vol. 42, no. 1, pp. 141–150, 2023, doi: 10.1080/02650487.2022.2157162.

[7] A. Margherita, M. Nasiri, and T. Papadopoulos, "The application of digital technologies in company responses to COVID-19: an integrative framework," *Technol. Anal. Strateg. Manag.*, vol. 35, no. 8, pp. 979–992, 2023, doi: 10.1080/09537325.2021.1990255.

[8] M. KhosraviNik and M. Amer, "Social media and terrorism discourse: the Islamic State's (IS) social media discursive content and practices," *Crit. Discourse Stud.*, vol. 19, no. 2, pp. 124–143, 2022, doi: 10.1080/17405904.2020.1835684.

[9] E. R. Kovacs, L. A. Cotfas, and C. Delcea, "January 6th on Twitter: measuring social media attitudes towards the Capitol riot through unhealthy online conversation and sentiment analysis," *J. Inf. Telecommun.*, vol. 8, no. 1, pp. 108–129, 2024, doi: 10.1080/24751839.2023.2262067.

[10] 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.

[11] Y. A. Singgalen, "Implementation of Global Vectors for Word Representation (GloVe) Model and Social Network Analysis through Wonderland Indonesia Content Reviews," *J. Sist. Komput. dan Inform.*, vol. 5, no. 3, pp. 559–569, 2024, doi: 10.30865/json.v5i3.7569.

[12] C. Budak, R. K. Garrett, and D. Sude, "Better Crowdcoding: Strategies for Promoting Accuracy in Crowdsourced Content Analysis," *Commun. Methods Meas.*, vol. 15, no. 2, pp. 141–155, 2021, doi: 10.1080/19312458.2021.1895977.

[13] A. He and M. Abisado, "Text Sentiment Analysis of Douban Film Short Comments Based on BERT-CNN-BiLSTM-Att Model," *IEEE Access*, vol. 12, no. March, pp. 45229–45237, 2024, doi: 10.1109/ACCESS.2024.3381515.

[14] S. Sen Zhang, X. Liang, Y. D. Wei, and X. Zhang, "On Structural Features, User Social Behavior, and Kinship Discrimination in Communication Social Networks," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 2, pp. 425–436, 2020, doi: 10.1109/TCSS.2019.2962231.

[15] J. Pueyo-Ros and E. Garau, "Do I have time to build the ark calmly? Characterizing attitudes towards climate change via sentiment analysis of social media," *J. Integr. Environ. Sci.*, vol. 20, no. 1, 2023, doi: 10.1080/1943815X.2023.2264380.

[16] Y. A. Singgalen, "Social Network Analysis and Sentiment Classification of Extended Reality Product Content," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 2197–2208, 2024, doi: 10.30865/klik.v4i4.1712.

[17] F. K. Sufi and M. Alsulami, "Automated Multidimensional Analysis of Global Events with Entity Detection, Sentiment Analysis and Anomaly Detection," *IEEE Access*, vol. 9, pp. 152449–152460, 2021, doi: 10.1109/ACCESS.2021.3127571.

[18] N. Jacob and V. M. Viswanatham, "Sentiment Analysis Using Improved Atom Search Optimizer With a Simulated Annealing and ReLU Based Gated Recurrent Unit," *IEEE Access*, vol. 12, no. March, pp. 38944–38956, 2024, doi: 10.1109/ACCESS.2024.3375119.

[19] C. L. Serban, A. M. Banu, S. Putnoky, S. I. Butica, M. D. Niculescu, and S. Putnoky, "Relative Validation of a Four Weeks Retrospective Food Frequency Questionnaire versus 7-Day Paper-Based Food Records in Estimating the Intake of Energy and Nutrients in Adults," *Nutr. Diet. Suppl.*, vol. Volume 13, pp. 113–125, 2021, doi: 10.2147/nds.s310260.

[20] H. F. Gholipour, R. Tajaddini, and B. Foroughi, "International tourists’ spending on traveling

inside a destination: does local happiness matter?," *Curr. Issues Tour.*, vol. 26, no. 12, pp. 2027–2043, 2023, doi: 10.1080/13683500.2022.2077178.

[21] H. Ardiyanti, B. S. Laksmono, and D. Walujo, "Shifting biculturality to monoculturality : the acculturation among Chinese Peranakans in Serui Regency of Papua , Indonesia," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2359012.

[22] A. Kayumov, Y. joo Ahn, K. Kiatkawsin, I. Sutherland, and S. Zielinski, "Service quality and customer loyalty in halal ethnic restaurants amid the COVID-19 pandemic: a study of halal Uzbekistan restaurants in South Korea," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2301814.

[23] 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.

[24] 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.

[25] 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.

[26] M. A. El-Affendi, K. Alrajhi, and A. Hussain, "A Novel Deep Learning-Based Multilevel Parallel Attention Neural (MPAN) Model for Multidomain Arabic Sentiment Analysis," *IEEE Access*, vol. 9, pp. 7508–7518, 2021, doi: 10.1109/ACCESS.2021.3049626.

[27] 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.

[28] M. J. Kim, J. S. Kang, and K. Chung, "Word-embedding-based traffic document classification model for detecting emerging risks using sentiment similarity weight," *IEEE Access*, vol. 8, pp. 183983–183994, 2020, doi: 10.1109/ACCESS.2020.3026585.

[29] 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.

[30] T. Fontes, F. Murcos, E. Carneiro, J. Ribeiro, and R. J. F. Rossetti, "Leveraging Social Media as a Source of Mobility Intelligence: An NLP-Based Approach," *IEEE Open J. Intell. Transp. Syst.*, vol. 4, no. September, pp. 663–681, 2023, doi: 10.1109/OJITS.2023.3308210.

[31] D. Van Thin, H. Quoc Ngo, D. Ngoc Hao, and N. Luu-Thuy Nguyen, "Exploring zero-shot and joint training cross-lingual strategies for aspect-based sentiment analysis based on contextualized multilingual language models," *J. Inf. Telecommun.*, 2023, doi: 10.1080/24751839.2023.2173843.

[32] Y. A. Singgalen, "Digital marketing of smartphone manufacturing product : toxicity , social network , and sentiment classification," *Int. J. Soc. Sci. Econ. Art*, vol. 14, no. 1, pp. 73–86, 2024.

[33] 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.

[34] D. Amangeldi, A. Usmanova, and P. Shamoi, "Understanding Environmental Posts: Sentiment and Emotion Analysis of Social Media Data," *IEEE Access*, vol. 12, no. March, pp. 33504–33523, 2024, doi: 10.1109/ACCESS.2024.3371585.

[35] S. Gorissen, "Weathering and weaponizing the #TwitterPurge: digital content moderation and the dimensions of deplatforming," *Commun. Democr.*, vol. 00, no. 00, pp. 1–26, 2023, doi: 10.1080/27671127.2023.2264367.

[36] N. Gamal, S. Ghoniemy, H. M. Faheem, and N. A. Seada, "Sentiment-Based Spatiotemporal Prediction Framework for Pandemic Outbreaks Awareness Using Social Networks Data Classification," *IEEE Access*, vol. 10, no. July, pp. 76434–76469, 2022, doi: 10.1109/ACCESS.2022.3192417.

[37] J. Khan, N. Ahmad, S. Khalid, F. Ali, and Y. Lee, "Sentiment and Context-Aware Hybrid DNN With Attention for Text Sentiment Classification," *IEEE Access*, vol. 11, no. February, pp. 28162–28179, 2023, doi: 10.1109/ACCESS.2023.3259107.

[38] P. Thiengburanathum and P. Charoenkwan, "SETAR: Stacking Ensemble Learning for Thai Sentiment Analysis Using RoBERTa and Hybrid Feature Representation," *IEEE Access*, vol. 11, no. July, pp. 92822–92837, 2023, doi: 10.1109/ACCESS.2023.3308951.