KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 5, No 1, August 2024, Hal 242-262
DOI 10.30865/klik.v5i1.1941
https://djournals.com/klik

Sentiment and Toxicity Analysis in the Narratives of Wamena's Cultural Heritage: Understanding Community Perspectives and External Influences

Yan Dirk Wabiser[1], Yerik Afrianto Singgalen[2,*]

[1] Pancasila and Citizenship Education Study Program, Cenderawasih University, Jayapura, Indonesia
[2] Faculty of Business Administration and Communication, Tourism Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: [1]wabiserdirk@gmail.com, [2,*]yerik.afrianto@atmajaya.ac.id
Correspondence Author Email: yerik.afrianto@atmajaya.ac.id

Abstract−This study analyzes digital narratives surrounding Wamena's cultural heritage using the Digital Content Reviews and Analysis Framework, focusing on sentiment, toxicity, and thematic content. The research explores the complex interplay between community perspectives, cultural preservation, modernization, and external influences such as tourism. Toxicity analysis revealed that while most online discourse is supportive, there are instances of harmful language that could disrupt social cohesion, with toxicity scores peaking at 0.50790. These findings highlight the need for continuous moderation to foster a positive digital environment. Sentiment analysis provided a deeper understanding of emotional tones, showing a predominance of positive sentiments and highlighting frustration and dissent related to cultural erosion. The study employed machine learning algorithms for sentiment and toxicity classification, with the Support Vector Machine (SVM) enhanced by Synthetic Minority Over-sampling Technique (SMOTE) demonstrating superior accuracy at 87.29%. Content analysis identified vital themes such as community dynamics, cultural resilience, and the dual impact of tourism as both an economic catalyst and a potential threat to cultural integrity. The findings underscore the importance of maintaining an inclusive digital environment that promotes constructive dialogue and cultural preservation. This framework provides valuable insights for policymakers and community leaders, emphasizing the need for culturally sensitive strategies to manage digital content and support sustainable cultural tourism. Future research should expand this framework to other contexts to enhance the understanding of digital communication dynamics in diverse cultural settings.

Keywords: Digital narratives; Cultural preservation; Sentiment analysis; Toxicity analysis; Sustainable tourism.

1. INTRODUCTION

Indigenous communities' traditional lifestyle and cultural practices reflect a profound connection to their natural environment, embodying a holistic approach to living that integrates social, spiritual, and ecological dimensions. Rooted in centuries-old knowledge systems, these practices encompass sustainable agriculture, hunting, resource management methods, and intricate rituals and ceremonies that maintain social cohesion and transmit cultural values across generations [1]–[3]. Such practices illustrate a deep understanding of local ecosystems and serve as a repository of communal wisdom that challenges contemporary notions of development, which often prioritize economic growth over environmental sustainability and social equity [4]–[8]. One observes a dynamic interplay between cultural identity and environmental stewardship through a nuanced examination of these traditions. It suggests that indigenous practices offer valuable insights into alternative pathways for sustainable living [9]–[11]. Thus, recognizing and preserving the cultural heritage of Indigenous communities emerges as a critical endeavor supporting biodiversity conservation and promoting a more inclusive, equitable global society.

Tourism exerts a multifaceted influence on preserving cultural heritage and traditional practices, simultaneously acting as a catalyst for conservation and commodification. While increased tourist interest often leads to heightened awareness and economic investment in cultural sites and practices, promoting their conservation and providing financial resources for their upkeep, it also introduces the risk of cultural erosion and superficial representation [12]–[15]. The influx of visitors frequently necessitates modifications to traditional customs, altering rituals and festivities to cater to tourist expectations, which may dilute their authenticity and undermine their original significance [16]–[18]. Moreover, transforming cultural elements into marketable products for tourists raises concerns about the commodification of heritage, where cultural practices are reduced to mere performances devoid of their original context and meaning [19]–[22]. A critical examination of these dynamics reveals a complex interplay between preservation and commercialization, where the benefits of tourism, such as increased funding and visibility, must be weighed against potential threats to cultural integrity and continuity. Thus, a sustainable approach to tourism that respects and supports the genuine preservation of cultural heritage and traditional practices is essential to balance these competing interests effectively.

Modernization and external influences present significant challenges to the cultural identity and traditional knowledge of Indigenous communities, often leading to the erosion of long-standing customs and values. The influx of global cultural elements and the pressures of socioeconomic integration frequently compel these communities to adapt, sometimes at the cost of diminishing their unique cultural heritage and traditional practices [23], [24]. While occasionally necessary for survival in a rapidly changing world, such transformations may result in the gradual loss of indigenous languages, rituals, and traditional ecological knowledge, vital for maintaining the social fabric and environmental stewardship intrinsic to these cultures. The tension between adopting modern advancements and preserving traditional ways highlights a critical issue: the risk of cultural homogenization, where Indigenous identities may become subsumed under dominant global narratives [25]–[29]. Analyzing this dynamic underscores the importance of fostering strategies that promote cultural resilience, ensuring that modernization efforts do not come at the expense of eradicating the invaluable cultural wisdom and identity cultivated over generations. Thus, safeguarding indigenous communities' cultural identity and traditional knowledge in the face of modernization requires a nuanced approach that values cultural diversity and supports sustainable development.

This study aims to analyze toxicity, sentiment, and content by applying the Digital Content Reviews and Analysis Framework, utilizing a case study in Wamena, Papua. By examining digital interactions and reviews, this research seeks to uncover underlying patterns of communication that reflect community sentiment while identifying potentially harmful or toxic discourse that may impact social cohesion. The utilization of this framework is particularly pertinent, as it allows for a comprehensive exploration of both qualitative and quantitative dimensions of online content, providing a nuanced understanding of how digital platforms are being used to convey opinions and attitudes. It is posited that analyzing these digital narratives can offer critical insights into the socio-cultural dynamics of the region, highlighting how digital content reflects broader societal trends and issues. The findings from this investigation are expected to contribute to developing more effective strategies for managing digital communication, fostering positive engagement, and mitigating the spread of harmful content, thereby supporting social harmony and constructive discourse within the community.

The urgency of this research lies in its potential to address critical gaps in understanding the complexities of contemporary digital communication and its socio-cultural impacts. As digital platforms increasingly dominate public discourse, there is a pressing need to examine the dynamics of online interactions, particularly the proliferation of toxic content and its effects on community sentiment and cohesion [30]. This investigation is essential for its immediate implications in identifying and mitigating negative communication patterns and its broader contribution to developing frameworks that support healthier digital environments [31]. A thorough analysis of toxicity and sentiment in digital content provides valuable insights into how public opinion is shaped and how harmful narratives might be countered, thus promoting more constructive and inclusive communication. The findings of this study are anticipated to have far-reaching implications, informing policy and practice in digital content management and contributing to enhancing social harmony and digital literacy in increasingly connected societies.

This research's theoretical and practical implications are substantial, offering a dual contribution to academic discourse and real-world applications. Theoretically, this study enriches existing knowledge by providing a nuanced understanding of digital communication dynamics, particularly concerning toxicity, sentiment, and content in online interactions [32], [33]. It expands the conceptual frameworks for analyzing digital narratives, highlighting the intersections between digital discourse, cultural context, and social impact. Practically, the findings have significant relevance for policymakers, digital content moderators, and community leaders by offering actionable insights into mitigating negative online behaviors and fostering more positive, constructive engagement on digital platforms [34]–[36]. By identifying specific patterns and sources of toxicity and sentiment in digital communication, the study informs strategies for better content management, which is crucial for maintaining social harmony and enhancing digital literacy. Thus, the research advances theoretical debates and provides practical tools and strategies for effectively navigating the complexities of digital communication in contemporary society.

Similar research in digital communication and online behavior has explored various facets of how digital platforms influence societal dynamics and individual behavior. Previous studies have focused on analyzing the spread of misinformation, the role of social media in shaping public opinion, and the impact of online interactions on mental health and community well-being [37]. Such investigations provide valuable insights into how digital environments reflect and reinforce social attitudes and norms, particularly in toxic discourse and sentiment analysis. It is argued that these studies collectively contribute to a more comprehensive understanding of the digital ecosystem, offering a foundation for further exploration into how digital content affects cultural and social practices [38], [39]. By comparing these findings with the current study, it becomes evident that while there is a shared interest in the consequences of digital communication, the current research uniquely emphasizes the intersection of digital toxicity, sentiment, and cultural context, thereby filling a critical gap in the literature [40]–[42]. Therefore, drawing parallels with existing research contextualizes the current study and underscores its unique contributions to advancing knowledge in digital content analysis and its broader societal implications.

The limitations of this research primarily stem from the scope of data and methodological constraints, which may affect the generalizability and applicability of its findings. One significant limitation is the reliance on case studies within a specific geographical and cultural context, such as Wamena, Papua, which may not fully capture the diversity of digital communication patterns across different regions or communities. Furthermore, the analysis is limited to digital content available within selected platforms, potentially overlooking significant interactions occurring on less accessible or emerging digital forums. These constraints suggest that while the study provides in-depth insights into localized digital behaviors and sentiments, its conclusions might not extend universally without further corroboration. Moreover, the reliance on qualitative analysis methods, though beneficial for a nuanced understanding, could introduce subjective bias, affecting the interpretation of toxicity and sentiment. Despite these limitations, the research offers a foundational framework that can be expanded upon with broader, more diverse datasets and mixed-method approaches in future studies to enhance the robustness and scope of its conclusions.

Recommendations for further research emphasize the need for a more expansive and diversified exploration of digital communication dynamics across different cultural and geographic contexts. Expanding the scope of study beyond a single case or region would allow for a comparative analysis that identifies universal patterns and unique cultural variations in online behaviors, sentiments, and toxicity. It is argued that incorporating mixed-method approaches, combining qualitative and quantitative analyses, would enhance the depth and breadth of understanding, mitigating potential biases and enriching the data's interpretative value. Additionally, future investigations should consider the rapidly evolving nature of digital platforms, including emerging social media channels and lesser-known forums, to capture a more comprehensive picture of digital content and its socio-cultural impacts. Engaging in longitudinal studies could also provide insights into the temporal shifts in digital communication and the long-term effects of sustained exposure to toxic content. Therefore, by broadening methodological approaches and contexts, future research can build on the current findings to offer more robust, generalizable insights into the complex interplay between digital content and cultural dynamics.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis

A significant research gap exists in understanding the nuanced relationship between digital communication dynamics and cultural contexts, particularly in non-Western settings. While much of the existing literature focuses on digital behavior and sentiment analysis in developed countries, there is a lack of studies examining these phenomena within diverse cultural landscapes, such as those in remote or Indigenous communities [43], [44]. This gap limits the ability to generalize findings across different sociocultural contexts, potentially overlooking unique patterns of digital engagement and the cultural implications of online interactions [45]. It is posited that this lack of comprehensive exploration leaves a critical void in understanding how digital toxicity, sentiment, and content impact various cultural identities and social cohesion. By not adequately addressing these diverse contexts, current research may inadvertently contribute to a homogenized view of digital communication, failing to account for cultural specificity and variation. Addressing this gap requires a more inclusive research agenda that prioritizes cross-cultural studies and contextualizes digital communication practices within a broader, more diverse framework, thus enhancing understanding of global digital dynamics.

Network Overlay

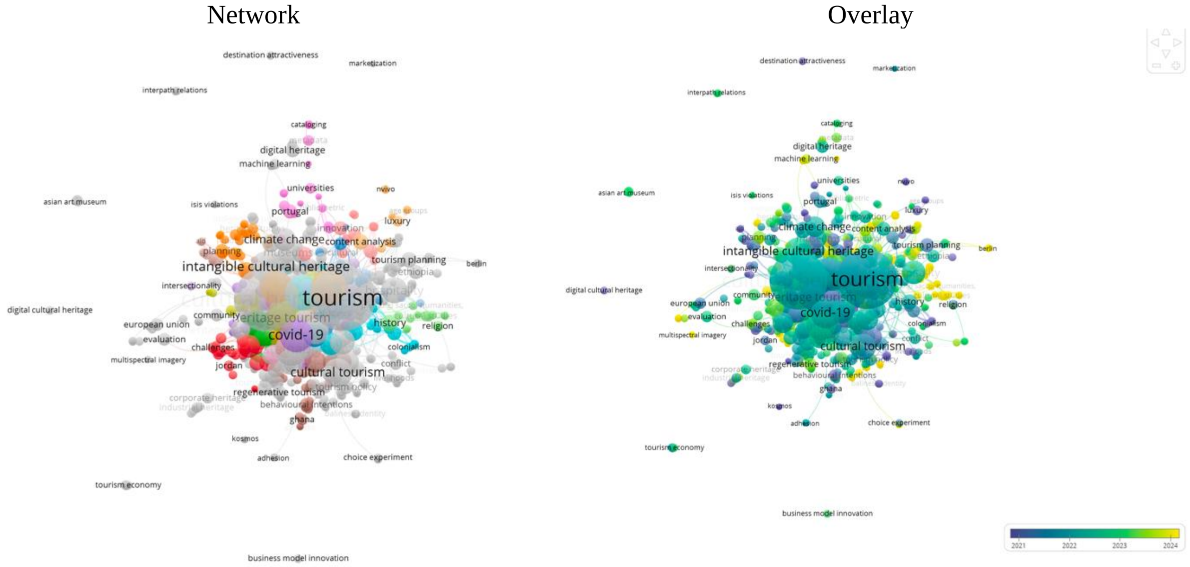


Figure 1. Network and Overlay Visualization of Cultural Heritage and Tourism

Figure 1 shows the network and overlay visualization of cultural heritage and tourism topics. The network visualization of cultural heritage and tourism topics reveals the intricate interconnections between various themes and concepts within the field, highlighting this study area's complex and multifaceted nature. By mapping out keywords such as "tourism," "intangible cultural heritage," "COVID-19," and "digital heritage," this visualization illustrates how different dimensions of cultural heritage intersect with contemporary issues like pandemic impacts, digital transformation, and sustainable tourism practices. It is evident that specific clusters, such as those related to "tourism planning" and "community heritage tourism," are more densely populated, indicating a concentrated research interest and a robust thematic coherence within these areas. It suggests a significant scholarly focus on understanding the interplay between tourism development and the preservation of cultural identity, as well as the challenges posed by external factors like climate change and global health crises. The visualization survey is a tool for identifying prevalent research trends and gaps and a framework for fostering interdisciplinary dialogue and collaboration, ultimately advancing the discourse on cultural heritage and tourism in a rapidly evolving global context.

The novelty of this research lies in its innovative approach to analyzing digital communication through the lens of cultural context, particularly within underrepresented regions such as Wamena, Papua. Unlike prior studies that predominantly focus on Western-centric perspectives, this investigation integrates local socio-cultural dimensions with digital content analysis, offering a more holistic understanding of how online narratives shape and are shaped by regional cultural identities. This unique perspective is crucial, as it challenges the prevailing assumptions that often overlook the cultural specificity inherent in digital behaviors and sentiment expression. By employing a comprehensive framework that combines toxicity and sentiment analysis with cultural insights, the research provides a groundbreaking contribution to the field, bridging the gap between digital communication studies and cultural anthropology. Thus, this novel approach enhances the theoretical landscape of digital content analysis and underscores the importance of considering diverse cultural contexts in understanding global digital dynamics, paving the way for more inclusive and context-sensitive future research.

2.2 Digital Content Reviews and Analysis Framework

The data processing in this research utilizes the Digital Content Reviews and Analysis Framework, a structured approach designed to handle large volumes of digital content systematically. This framework begins with the initial stages of data selection and cleaning, ensuring that only relevant and high-quality data is included for further analysis, which is crucial for maintaining the integrity and reliability of the research outcomes. Following this, the framework employs advanced techniques such as Topic Modeling using Latent Dirichlet Allocation (LDA) and data extraction methods, allowing for the identification of underlying themes and patterns within the digital content. This approach is convenient for discerning nuanced narratives, combining qualitative and quantitative elements to evaluate model performance, data visualization, and data evaluation. The final stages of the framework involve comprehensive evaluation and visualization processes, culminating in a context analysis that provides a deeper understanding of the narratives surrounding Wamena's cultural heritage and tourism. Thus, using this framework not only enhances the analytical rigor of the study but also ensures a thorough and context-sensitive interpretation of digital narratives.

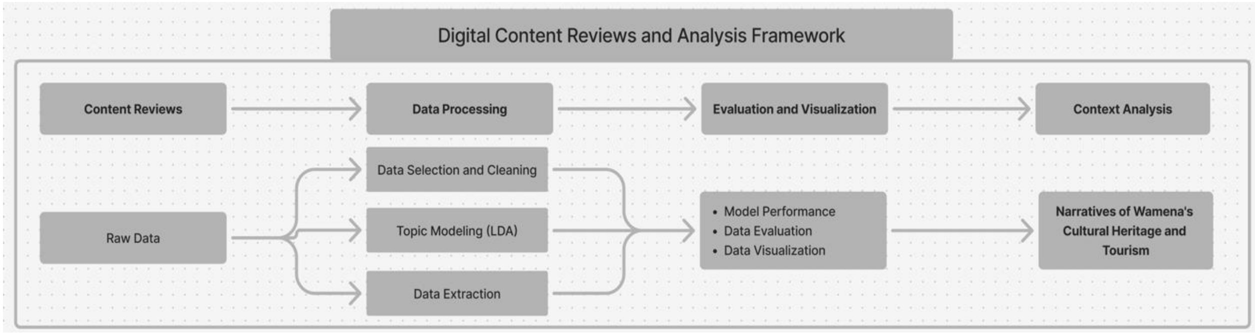


Figure 2. Digital Content Reviews and Analysis Framework

Figure 2 shows the digital content reviews and analysis framework. The stages within the Digital Content Reviews and Analysis Framework are systematically designed to provide a comprehensive approach to handling and interpreting digital data related to cultural heritage and tourism. The process begins with content reviews and data processing, which involve the meticulous selection and cleaning of raw data to ensure the inclusion of only pertinent and high-quality information. Following this, advanced computational techniques such as Topic Modeling with Latent Dirichlet Allocation (LDA) and data extraction methods are employed to identify and categorize underlying themes within the digital content. These steps are crucial for uncovering patterns and trends that may not be immediately apparent through surface-level analysis. The subsequent phase focuses on evaluation and visualization, where the performance of the models is assessed, and the data is visualized to facilitate a clearer understanding of the findings. It leads to a nuanced context analysis, integrating all insights to construct detailed narratives concerning Wamena's cultural heritage and tourism. The structured progression through these stages underscores the framework's robustness and adaptability, making it a valuable tool for exploring complex digital landscapes in an organized and context-sensitive manner.

Based on the established framework, this research focuses on conducting a sentiment and toxicity analysis within the narratives surrounding Wamena's cultural heritage, aiming to elucidate community perspectives and the impact of external influences. By examining digital content, this study seeks to identify positive and negative sentiments expressed by community members and detect any toxic discourse that might affect social cohesion and cultural integrity. It is asserted that understanding these narratives is crucial for grasping how the community perceives local cultural heritage internally and how external factors, including tourism, globalization, and digital media, influence it. Through a systematic analysis of sentiment and toxicity, the research sheds light on the complexities of maintaining cultural heritage in a digital age. It reveals how external narratives can shape and sometimes distort the community's perceptions of their cultural identity. This focus highlights the relevance of digital communication in contemporary cultural discourse. It emphasizes the need for strategies that foster positive engagement and mitigate the negative impacts of external influences on local cultural narratives.

2.2.1 Content Reviews

The content utilized in this research includes a video identified by the code 0CaFR4AbhQQ, which has garnered significant engagement with 245,267 views and 746 comments as of April 3, 2024. This video is a rich data source, providing a valuable lens through which to analyze public sentiment and discourse about Wamena's cultural heritage. The high number of views and comments suggests that the video has a substantial reach and relevance, making it an appropriate medium for examining community engagement and external perspectives on local cultural narratives. It is argued that this type of content, with its broad audience and active participation, offers a robust platform for exploring both the positive and negative sentiments expressed by viewers and identifying any instances of toxic language that might influence community dynamics. Analyzing such video content allows for a deeper understanding of the digital representation of Wamena's cultural heritage, highlighting the importance of visual and narrative forms in shaping public perception and cultural identity. This approach underscores the need for careful consideration of digital media in cultural studies, as it provides a nuanced understanding of how cultural narratives are constructed and disseminated in the digital age.

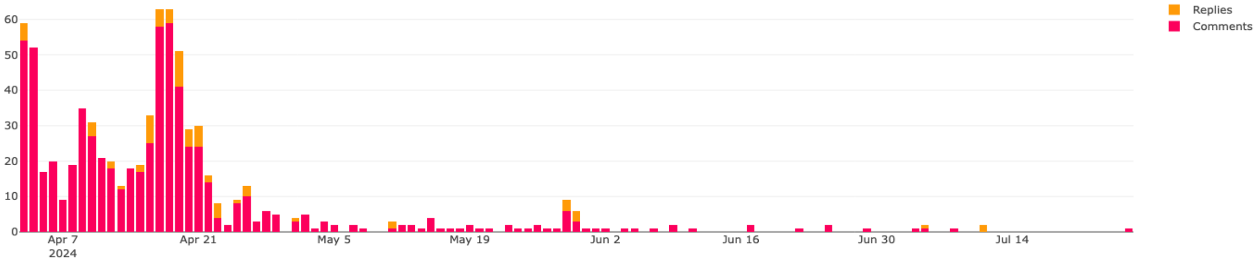


Figure 3. Post-per-day Statistic of the Video (Communalytic)

Figure 3 shows the content's post-per-day statistics. Based on the post-per-day statistics, it is evident that user and viewer engagement with the video content displayed a sharp initial increase, followed by a gradual decline over time. The data illustrates a significant peak in comments and replies during the first two weeks after the video's release, particularly around April 7 to April 21, 2024, indicating heightened viewer interest and interaction. This surge in engagement likely reflects a strong immediate reaction to the content, suggesting that the video elicited a substantial response, possibly due to its relevance, controversy, or emotional appeal. The subsequent decline in comments and replies after this peak period implies a tapering of user interest, which could be attributed to a saturation point in viewer engagement or a shift in focus to other emerging content. Analyzing this interaction pattern provides valuable insights into the temporal dynamics of user engagement with digital content, highlighting the importance of timing and content freshness in maintaining audience interest. Therefore, these statistics offer a critical understanding of user response behavior, which is essential for developing strategies to sustain engagement and maximize the impact of digital media.

The interpretation of the data reveals distinct patterns of user engagement, reflecting both the initial impact and the subsequent decline in viewer interaction with the video content. The prominent surge in comments and replies shortly after the video was released, particularly in the early weeks, indicates a heightened level of viewer interest and potentially emotional or intellectual engagement with the topic presented. This initial burst of activity could be interpreted as a strong viewer response to the content's relevance or alignment with current social and cultural discussions. Conversely, the gradual decrease in user engagement over time suggests a natural waning of interest, which might be influenced by factors such as the saturation of discourse around the content or the emergence of new topics capturing public attention. This pattern underscores the transient nature of digital engagement, where user interest is often concentrated in short bursts, followed by a decline as content ages and becomes less topical. Thus, the data highlights the importance of timely and relevant content creation in digital platforms to sustain and maximize user interaction and engagement.

The analysis of the top ten posters of the content reveals a diverse group of users, each contributing significantly to the overall engagement and discourse surrounding the video. The user identified as @jesica770 emerges as the most active participant, accounting for 15.2% of the total posts, followed by @CintaLauraKiehlOfficial and @MartomiAjen, who contribute 10.9% and 10.9%, respectively. Other notable contributors, such as @denzimunzi, @KakekSugionoCR7, and @FREDWALKERINDO, each hold an approximate share of 10.9% of the content, indicating a well-distributed level of participation among the top contributors. It is noteworthy that the remaining users, including @Kenn82157, @ImnahJalakia, @stewartsingal4599, and @amelamel4979, collectively contribute between 6.52% and 8.7%, suggesting a balanced engagement pattern with no single user disproportionately dominating the conversation. This distribution implies a collaborative discourse environment where multiple voices actively engage, fostering a more dynamic and potentially diverse range of perspectives. Therefore, understanding these top contributors' composition and engagement levels provides valuable insights into the dynamics of user interaction and the collective shaping of narratives in digital spaces.

Figure 4. Top Ten Poster (Communalytic)

Figure 4 shows the top ten posters. The interpretation of the top ten posters' data reveals a highly engaged and somewhat balanced community of contributors, each playing a crucial role in shaping the discourse around the content. The user @jesica770 stands out as the most prolific poster, with a significant 15.2% share of the total contributions, indicating a substantial personal investment in the topic or a higher level of influence within the discussion. Other contributors, such as @CintaLauraKiehlOfficial and @MartomiAjen, with approximately 10.9% each, also demonstrate considerable involvement, suggesting that a few key users are driving much of the conversation. However, several other users contributed between 6.52% and 10.9%, indicating a relatively distributed engagement across the top ten. It may suggest a more democratic interaction where multiple voices are heard rather than a single dominant perspective. This distribution is essential for fostering a more balanced and diverse discussion, which could lead to a richer and more nuanced understanding of the content's themes. Thus, the data highlights the importance of identifying and understanding the roles of key contributors in digital discourse, as they can significantly influence the direction and tone of online discussions.

The review data of the video content necessitates thorough cleaning to ensure the data is aligned with the objectives of this research. Effective data cleaning involves removing irrelevant, redundant, or noise-laden entries that do not contribute to understanding the narrative or sentiment of Wamena's cultural heritage. This process is critical, as unrefined data may lead to skewed results, misinterpretations, or biased conclusions, undermining the study's validity. A systematic approach to data cleaning enables more accurate extraction of meaningful patterns and insights, facilitating a deeper examination of the socio-cultural dynamics present in the digital discourse. By meticulously refining the dataset, the analysis becomes more robust and reflective of the perspectives and sentiments conveyed in the video reviews. Consequently, the integrity and reliability of the research findings are significantly enhanced, providing a more precise and nuanced understanding of the community's views and the external influences shaping them.

2.2.2 Data Processing

Data processing is essential to refine the dataset, extract text data based on negative and positive classifications, and perform Topic Modeling using Latent Dirichlet Allocation (LDA) to identify the topics most frequently discussed by viewers. Initially, data cleaning removes irrelevant and redundant information, ensuring the dataset's integrity and enhancing the reliability of subsequent analyses. Following this, extracting text data classified into negative and positive sentiments allows for a deeper understanding of viewer perceptions and emotions, providing insight into the public's reception of the content. The application of LDA facilitates the identification of underlying thematic structures within the data, revealing popular topics and recurrent themes in viewer discussions. This analytical approach uncovers the prevailing sentiments and opinions and provides a nuanced understanding of the issues that resonate most with the audience. Thus, data processing, including data cleaning, sentiment extraction, and topic modeling, is crucial for generating a comprehensive analysis that reflects viewer engagement and discourse complexities.

The applications utilized for data pre-processing and processing in this research are Communalytic, RapidMiner, and Atlas, each offering distinct functionalities that contribute to a comprehensive data analysis workflow. Communalytic provides advanced tools for extracting and managing social media data, enabling the identification and classification of sentiment within large datasets. RapidMiner facilitates sophisticated data mining processes, allowing for the automation of data cleaning, transformation, and predictive modeling tasks, which are essential for handling complex datasets efficiently. Meanwhile, Atlas offers robust qualitative analysis capabilities that are beneficial for coding and interpreting textual data, enhancing the depth of understanding of the underlying themes and patterns in the content. Integrating these tools within the research framework is particularly advantageous as it combines quantitative and qualitative methodologies, providing a more holistic approach to data analysis. By leveraging the strengths of each application, the research achieves a more nuanced and detailed exploration of digital narratives, ensuring that both technical precision and interpretive depth are maintained throughout the analytical process.

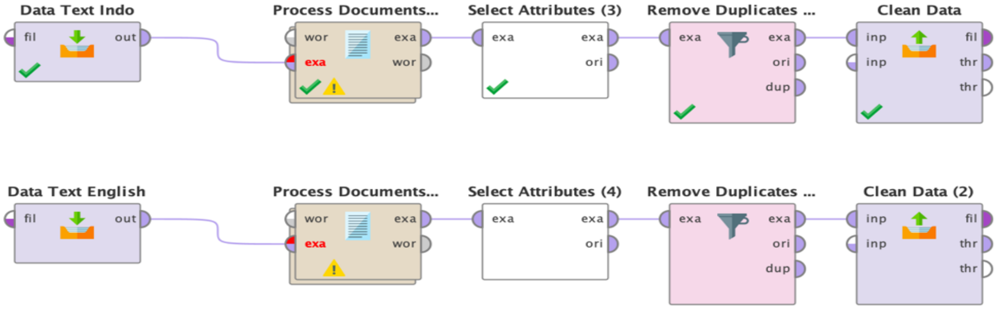


Figure 5. Data Cleaning Process

Figure 5 shows the data-cleaning process. The data cleaning process is meticulously tailored to accommodate both Indonesian and English languages, ensuring that the analysis accurately reflects the multilingual nature of the dataset. This dual-language approach involves distinct workflows for processing documents in each language, including selecting relevant attributes, removing duplicates, and executing thorough data cleaning protocols. It is argued that separating the data processing by language enhances the precision of the analysis, as it allows for language-specific nuances and structures to be adequately addressed, thereby avoiding potential misinterpretations that could arise from a one-size-fits-all method. By implementing these tailored cleaning procedures, the research ensures that the integrity and contextual relevance of Indonesian and English text data are maintained throughout the analytical process. This careful consideration of linguistic diversity improves the robustness and reliability of the findings. It supports a more nuanced understanding of digital narratives, reflecting the cultural and linguistic context from which they emerge.

The results of the data cleaning process, categorized by language, demonstrate a significant enhancement in the quality and reliability of the textual datasets for both English and Indonesian. This language-specific data-cleaning approach effectively removes irrelevant content, redundancies, and noise that could otherwise compromise the integrity of the analysis. It is suggested that treating each language separately allows for more accurate handling of linguistic nuances and cultural idioms, thereby preserving the contextual authenticity of the data. The cleaned datasets, now free from extraneous elements, provide a more precise and focused foundation for subsequent sentiment and thematic analysis. This meticulous approach not only bolsters the analytical robustness of the study but also ensures that the insights derived reflect genuine viewer sentiments and perspectives in each linguistic context. Consequently, the language-based data cleaning results in a more reliable and valid representation of the digital discourse, supporting a deeper and more nuanced understanding of the cultural narratives at play.

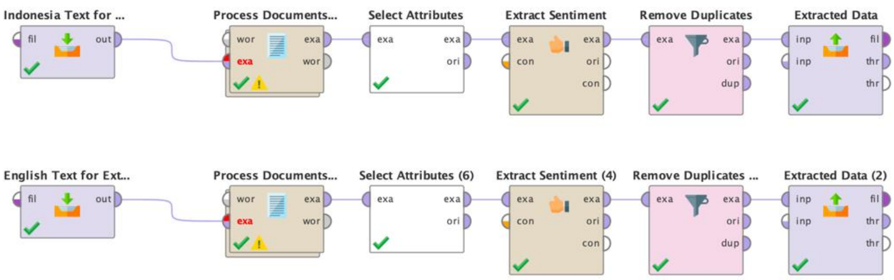


Figure 6. Data Extraction Process

Figure 6 shows the data extraction process. The data extraction process for text is meticulously executed in alignment with both English and Indonesian languages, ensuring a comprehensive analysis of sentiment and thematic elements within each linguistic context. This dual-language strategy involves separate workflows, each designed to handle the specific syntactic and semantic characteristics of Indonesian and English texts, allowing for more precise extraction of relevant attributes and sentiments. It is posited that this approach enhances the accuracy of sentiment analysis, as it accounts for linguistic nuances and cultural contexts that are unique to each language, thereby avoiding the pitfalls of a generic, one-dimensional analysis. Extracting sentiments and identifying key themes distinctly for each language enables a richer, more differentiated understanding of the viewer's perspectives and reactions. This methodological rigor not only improves the reliability of the findings but also ensures that the cultural and linguistic diversity present in the digital discourse is appropriately captured and represented in the research outcomes.

The results of the data extraction process, distinguished by language, reveal a comprehensive understanding of the sentiments and thematic content within both English and Indonesian texts. This bifurcated approach to data extraction allows for a nuanced capture of sentiment, identifying both positive and negative expressions across diverse linguistic frameworks. The separation of data by language is critical in accurately reflecting the unique cultural contexts and emotional tones inherent to each language, thereby preventing the loss of meaning that could occur through a monolingual analysis. By extracting language-specific data, the study achieves a more granular insight into the topics and sentiments that resonate most with viewers, offering a more precise depiction of how different linguistic groups engage with the content. This methodological rigor enhances the validity of the findings and enriches the interpretative depth of the analysis, ensuring a robust representation of digital narratives as influenced by language. Thus, the language-based extraction results provide a solid foundation for further exploration into viewer responses' cultural and communicative dynamics.

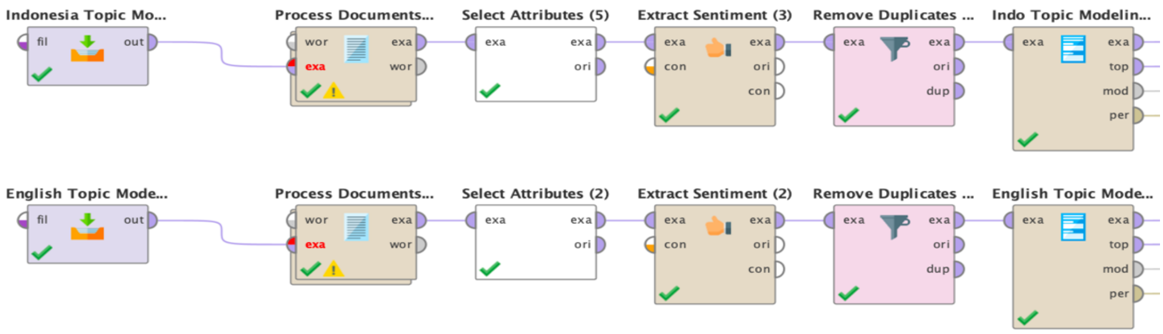


Figure 7. Topic Modeling based on LDA

Figure 7 shows the topic modeling process. The Topic Modeling using Latent Dirichlet Allocation (LDA) is conducted separately for Indonesian and English texts to uncover the prevalent themes viewers discuss in each linguistic context. This language-specific approach allows for more precise identification of topics, as it considers each language's unique syntactic and semantic structures, which are essential for accurate topic extraction and categorization. The decision to perform LDA separately is based on the premise that linguistic nuances and cultural references can significantly influence how topics are discussed and perceived, requiring a tailored analytical approach to capture these subtleties. By applying LDA independently to Indonesian and English datasets, the analysis benefits from a more refined understanding of the content, revealing distinct thematic patterns and critical areas of interest specific to each language group. This method enhances the depth and granularity of the findings, allowing for a more comprehensive exploration of the digital narratives and the factors driving viewer engagement. Consequently, language-based LDA improves the accuracy and relevance of topic modeling and supports a more culturally sensitive interpretation of the data, which is crucial for capturing the full spectrum of viewer perspectives and discourse dynamics.

The Topic Modeling using Latent Dirichlet Allocation (LDA) for Indonesian text reveals several key metrics that provide insights into the model's performance and the underlying structure of the data. The LogLikelihood score of -74,899.117 and a Perplexity value of 884.203 suggest a relatively high degree of complexity and variability within the dataset, indicating the presence of diverse themes and topics. With an average token count of 945.900 and a document entropy of 3.759, the data shows substantial lexical richness and topic distribution, reflecting a broad range of discussions captured within the Indonesian content. The average word length of 4.780 and coherence score of -20.357 highlight the linguistic characteristics of the text, where coherence suggests that the model could benefit from further tuning to improve topic distinctiveness. The metrics related to word distribution, such as Avg(uniform_dist) at 2.508 and Avg(corpus_dist) at 2.307, combined with an adequate number of words averaging 192.009, demonstrate a moderate level of word exclusivity and diversity across the topics. In addition, the low Avg(token-doc-diff) of 0.007 and Avg(rank_1_docs) of 0.371 suggest a reasonable allocation of topics across the document. At the same time, the Avg(allocation_count) of 0.496 and Avg(exclusivity) of 0.730 indicate a balanced representation of unique words in the model. The AlphaSum of 1.336 and Beta parameters, with a Beta value of 0.144 and BetaSum of 505.431, further provide evidence of the model's ability to capture the nuances of topic distribution. Overall, these metrics suggest that while the model captures a wide array of topics, there is room for refinement to enhance the clarity and coherence of the identified themes.

Bahasa English

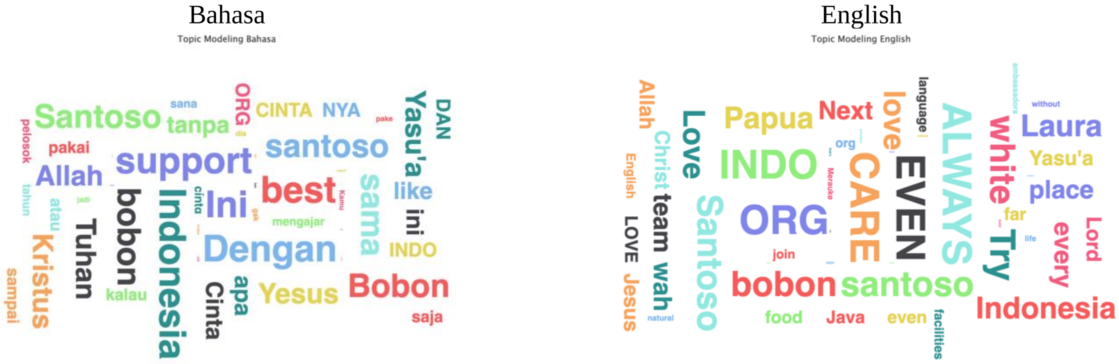


Figure 8. Topic Modeling LDA

Figure 8 shows the topic modeling of Bahasa and the English language. Topic Modeling using Latent Dirichlet Allocation (LDA) results for the Indonesian text reveals several prominent themes reflecting the viewers' critical interest and engagement. Words such as "Indonesia," "support," "santoso," "dengan," and "Bobon" frequently appear, suggesting a strong focus on national identity, expressions of solidarity, and possibly discussions centered around specific individuals or events. The presence of terms like "Yesus" and "Allah" indicates that religious sentiment also plays a significant role in the discourse, reflecting the socio-cultural diversity and the importance of faith in the community's conversations. This thematic distribution suggests that discussions are not only centered around nationalistic and supportive expressions but also intertwined with religious references, highlighting the complexity of the digital narrative landscape in Indonesian contexts. Analyzing these findings provides valuable insights into the cultural and social dynamics that influence online interactions, demonstrating that digital discourse is a rich tapestry of interconnected themes that shape and are shaped by the broader socio-cultural environment. Thus, the LDA results underscore the multifaceted nature of public sentiment and the importance of contextually nuanced analysis in understanding digital communication.

The Topic Modeling using Latent Dirichlet Allocation (LDA) for English text presents a range of metrics that offer valuable insights into the model's effectiveness and the content's thematic structure. The LogLikelihood score of -55,128.669 and Perplexity of 791.794 suggest a moderately complex dataset with some variability, indicative of diverse and potentially overlapping topics within the English corpus. An average token count of 695.400 and a document entropy of 4.147 indicate a robust presence of diverse discussions and a slightly more concentrated topic distribution than the Indonesian dataset. The average word length of 5.360 and coherence score of -20.653 imply that while the topics are somewhat distinct, there may be room for further refinement to enhance topic coherence and reduce overlap. Metrics such as Avg(uniform_dist) at 2.481 and Avg(corpus_dist) at 2.388, coupled with an adequate number of words averaging 151.071, reflect a moderate exclusivity and variety in word usage across topics, suggesting a balanced yet diverse representation of themes. The average (token-doc-diff) of 0.008 and average (rank_1_docs) of 0.350 indicate a stable allocation of topics within documents. In contrast, the average (allocation_count) of 0.332 and Avg(exclusivity) of 0.767 suggest a high degree of uniqueness among the words assigned to each topic. The AlphaSum value of 2.758 and Beta parameters, with a Beta of 0.162 and BetaSum of 431.693, further indicate the model's capacity to capture nuanced thematic distributions. These findings suggest that the LDA model for English text performs well in identifying a range of topics. However, further optimization could enhance topic coherence and specificity for a more precise thematic analysis.

The processed data will be evaluated through a multifaceted approach incorporating toxicity, sentiment, and content analysis to understand viewer engagement and discourse dynamics comprehensively. This evaluation aims to categorize the data into toxicity levels, identifying harmful or inflammatory language that could affect community sentiment and online interactions. Concurrently, sentiment analysis will be employed to gauge the emotional tone of the content, distinguishing between positive, negative, and neutral expressions to understand public perceptions and attitudes better. In addition, content analysis will focus on the thematic elements within the data, assessing the relevance and significance of the topics discussed to contextualize the narratives within broader social and cultural frameworks. Such an integrative approach is argued to be essential for capturing the complex interplay between language, emotion, and content in digital communication, providing nuanced insights into how different narratives and sentiments shape and are shaped by viewer engagement. Ultimately, this evaluation will enhance the understanding of digital discourse, informing strategies to foster positive and constructive online environments.

2.2.3 Data Evaluation and Visualization

During the evaluation and visualization phase, the toxicity scores calculated using the Perspective API through Communalytic are compared with the sentiment extraction results obtained via the Vader model, providing a dual-layer analysis of content sentiment and toxicity. This comparison aims to validate the consistency and accuracy of the sentiment classification by cross-referencing two distinct methodologies—one focusing on toxicity detection and the other on sentiment polarity. To further refine this analysis, the performance of various algorithms, such as Decision Tree (DT), Naive Bayes Classifier (NBC), k-nearest Neighbors (k-NN), and Support Vector Machine (SVM), is assessed in classifying sentiment. Each algorithm's effectiveness is measured to determine its precision, recall, and accuracy in detecting dataset sentiment nuances. This multifaceted approach is essential for identifying the most reliable and efficient model for sentiment analysis, as it allows for a comprehensive comparison across different machine-learning techniques. Consequently, the insights derived from this evaluation enhance the understanding of the content's emotional and toxic dimensions and improve algorithmic approaches for more accurate and nuanced sentiment classification in digital content analysis.

Using the Perspective API, Communalytic's analysis of 637 posts out of 746 provides a detailed overview of the dataset's toxicity levels and harmful language. The average toxicity score for the dataset is 0.05326, with the highest individual post scoring 0.50790, indicating that while most posts are relatively non-toxic, there are instances of significantly higher toxicity. Severe toxicity, measured at an average of 0.00644 and peaking at 0.50704, remains generally low, suggesting that highly harmful content is less prevalent. However, identity attacks have an average score of 0.02081 with a maximum of 0.38177, and insults register an average of 0.04047, reaching up to 0.55882, highlighting a moderate presence of personal attacks and offensive language. Profanity shows a similar pattern with an average of 0.03668 and a peak value of 0.57114, while threats are the least common, averaging 0.01064 but with a high of 0.50583, suggesting rare but significant instances of threatening language. This data underscores the importance of monitoring and managing digital interactions to minimize harmful communication, indicating that while the overall level of toxicity is low, specific categories reveal spikes that necessitate further scrutiny and potential intervention strategies.

	Average for dataset	Highest value
Toxicity	0.05326	0.50790
Severe Toxicity	0.00644	0.50704
Identity Attack	0.02081	0.38177
Insult	0.04047	0.55882
Profanity	0.03668	0.57114
Threat	0.01064	0.50583

Figure 9. Average for Dataset and Highest Value of Toxicity Classification (Communalytic)

Figure 9 shows the dataset's average and the highest toxicity classification value using Communalytic. The toxicity scores imply varying levels of harmful or offensive language present within the analyzed digital content, reflecting both the general tone of the discourse and the potential for a negative impact on the community. An average toxicity score of 0.05326 suggests that most content is relatively benign, with low levels of harmful language overall. However, higher individual scores, such as a maximum toxicity score of 0.50790, indicate notable instances where the discourse becomes significantly more toxic, potentially creating a hostile environment. The scores for severe toxicity, identity attacks, insults, profanity, and threats, although generally low, also show occasional spikes, which could signify moments of escalated tension or conflict within the conversation. These spikes are particularly concerning in categories like identity attacks and threats, as they suggest targeted aggression that could harm specific individuals or groups. The toxicity scores highlight the need for ongoing monitoring and moderation to foster a more constructive and inclusive digital environment.

After determining the toxicity scores, a comparative evaluation is conducted by assessing the sentiment classification results based on positive and negative classes. This comparison aims to correlate the toxicity levels with the sentiment expressed in the digital content, providing a deeper understanding of how toxicity aligns with the overall emotional tone. It is posited that sentiment classification, which categorizes content into positive and negative sentiments, serves as a complementary measure to toxicity analysis by offering insights into the underlying emotional drivers behind the presence of toxic language. Analyzing toxicity scores and sentiment classifications, the evaluation identifies patterns where high toxicity corresponds with negative sentiment, potentially indicating hostile or aggressive interactions. In contrast, low toxicity paired with positive sentiment suggests constructive or supportive discourse. This dual analysis not only enhances the understanding of the nature and quality of digital interactions but also supports the development of more targeted strategies for moderating content and fostering a healthier digital environment. Thus, combining toxicity analysis with sentiment classification offers a comprehensive approach to evaluating the dynamics of online communication.

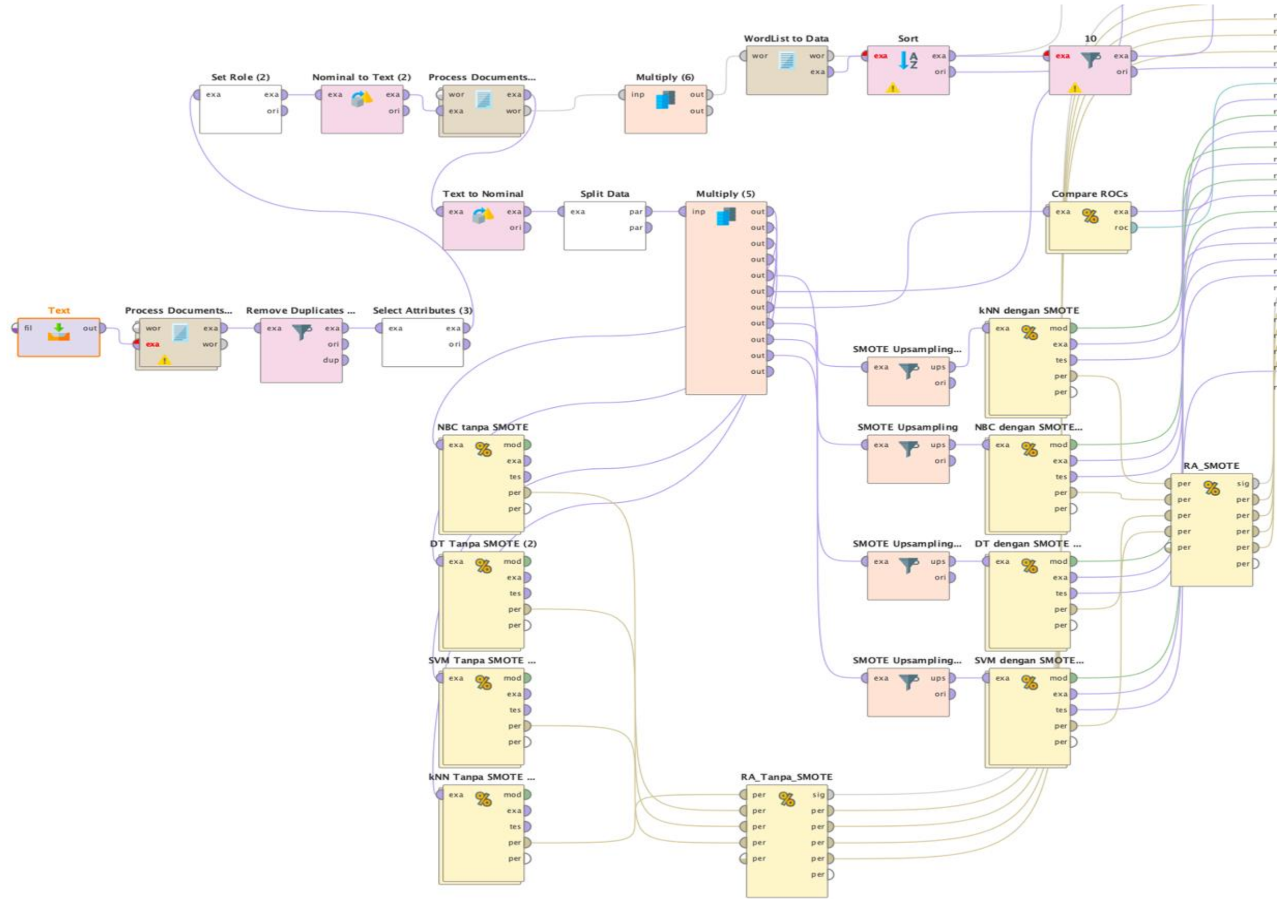


Figure 10. Performance Evaluation of k-NN, NBC, and SVM Enhanced by SMOTE

Figure 10 shows the performance evaluation of DT, k-NN, NBC, and SVM Enhanced by SMOTE. Based on the 802 extracted text data, 42 were classified as unfavorable, while 759 were categorized as positive. The extracted data and corresponding sentiment scores were subsequently evaluated using algorithms such as k-nearest Neighbors (k-NN), Decision Tree (DT), Naive Bayes Classifier (NBC), and Support Vector Machine (SVM) to determine key performance metrics, including accuracy, precision, recall, F-measure, and Area Under the Curve (AUC). This evaluation process is critical in identifying the most effective algorithm for sentiment classification, as each metric provides different insights into the model's performance. For instance, accuracy measures the model's overall correctness, and precision assesses the proportion of correctly identified positive or negative sentiments, recall indicates the model's ability to detect all relevant instances, and F-measure balances precision and recall. AUC, on the other hand, evaluates the model's ability to distinguish between positive and negative classes across different threshold settings. The evaluation aims to optimize sentiment analysis by assessing these metrics, ensuring that the chosen algorithm offers reliable and robust performance in accurately classifying the extracted text data. This comprehensive approach not only enhances the analytical rigor but also supports the development of more nuanced and effective strategies for understanding sentiment in digital communication.

The data configuration for the analysis was set with 20% allocated for training and 80% for testing, a strategic choice to maximize the model's exposure to diverse scenarios during evaluation. This setup ensures that the algorithm's performance is rigorously tested against a wide array of data points, providing a robust assessment of its generalizability and reliability. Among the algorithms evaluated, the Support Vector Machine (SVM) enhanced by the Synthetic Minority Over-sampling Technique (SMOTE) demonstrated the best performance, effectively handling the imbalance between the negative and positive classes in the dataset. The application of SMOTE with SVM proved advantageous, as it improved the model's ability to distinguish between classes, resulting in higher accuracy, precision, recall, and F-measure scores compared to other algorithms. This finding suggests that when coupled with SMOTE, SVM offers superior capability in sentiment classification tasks, particularly in datasets with class imbalances. Consequently, this combination enhances the analytical outcomes and provides a more accurate and reliable tool for future sentiment analysis in similar contexts.

k-NN enhanced by SMOTE

```
PerformanceVector:
accuracy: 51.17% +/- 0.62% (micro average: 51.17%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       599     585
Positive:       0       14
AUC (optimistic): 0.975 +/- 0.010 (micro average: 0.975) (positive class: Positive)
AUC: 0.614 +/- 0.028 (micro average: 0.614) (positive class: Positive)
AUC (pessimistic): 0.252 +/- 0.054 (micro average: 0.252) (positive class: Positive)
precision: 100.00% (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       599     585
Positive:       0       14
recall: 2.34% +/- 1.16% (micro average: 2.34%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       599     585
Positive:       0       14
f_measure: 4.57% (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       599     585
Positive:       0       14
```

DT enhanced by SMOTE

```
PerformanceVector:
accuracy: 76.71% +/- 2.59% (micro average: 76.71%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       324     4
Positive:       275     595
AUC (optimistic): 0.997 +/- 0.005 (micro average: 0.997) (positive class: Positive)
AUC: 0.767 +/- 0.025 (micro average: 0.767) (positive class: Positive)
AUC (pessimistic): 0.537 +/- 0.052 (micro average: 0.537) (positive class: Positive)
precision: 68.48% +/- 2.56% (micro average: 68.39%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       324     4
Positive:       275     595
recall: 99.33% +/- 1.17% (micro average: 99.33%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       324     4
Positive:       275     595
f_measure: 81.04% +/- 1.70% (micro average: 81.01%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       324     4
Positive:       275     595
```

SVM enhanced by SMOTE

```
PerformanceVector:
accuracy: 80.46% +/- 2.96% (micro average: 80.47%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       401     36
Positive:       198     563
AUC (optimistic): 0.948 +/- 0.019 (micro average: 0.948) (positive class: Positive)
AUC: 0.942 +/- 0.020 (micro average: 0.942) (positive class: Positive)
AUC (pessimistic): 0.936 +/- 0.022 (micro average: 0.936) (positive class: Positive)
precision: 74.10% +/- 3.32% (micro average: 73.98%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       401     36
Positive:       198     563
recall: 93.98% +/- 2.15% (micro average: 93.99%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       401     36
Positive:       198     563
f_measure: 82.82% +/- 2.32% (micro average: 82.79%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       401     36
Positive:       198     563
```

NBC enhanced by SMOTE

```
PerformanceVector:
accuracy: 67.78% +/- 2.17% (micro average: 67.78%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       599     386
Positive:       0       213
AUC (optimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Positive)
AUC: 0.517 +/- 0.055 (micro average: 0.517) (positive class: Positive)
AUC (pessimistic): 0.357 +/- 0.041 (micro average: 0.357) (positive class: Positive)
precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       599     386
Positive:       0       213
recall: 35.57% +/- 4.21% (micro average: 35.56%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       599     386
Positive:       0       213
f_measure: 52.35% +/- 4.49% (micro average: 52.46%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       599     386
Positive:       0       213
```

Figure 11. Performance Vector of DT, k-NN, NBC, and SVM Enhanced by SMOTE

Figure 11 shows the performance vector of DT, k-NN, NBC, and SVM Enhanced by SMOTE. The performance vectors for the algorithms—Decision Tree (DT), k-nearest Neighbors (k-NN), Naive Bayes Classifier (NBC), and Support Vector Machine (SVM) enhanced by Synthetic Minority Over-sampling Technique (SMOTE)—demonstrate varying degrees of effectiveness in classifying sentiment within the dataset. The accuracy rates range from 52.94% for NBC to 87.29% for SVM enhanced by SMOTE, indicating a significant disparity in performance among the algorithms. When combined with SMOTE, SVM achieved the highest accuracy, recall, and Area Under the Curve (AUC) scores, showcasing its superior ability to handle imbalanced datasets and accurately classify positive and negative sentiments. In contrast, with the lowest accuracy and AUC scores, NBC suggests limitations in distinguishing sentiment, particularly in complex or nuanced data contexts. The Decision Tree and k-NN algorithms show moderate performance, with accuracy rates of 56.47% and 66.67%, respectively, reflecting their potential utility and limitations compared to SVM with SMOTE. These results indicate that SVM enhanced by SMOTE is the most effective algorithm for this sentiment analysis task, providing a robust solution for accurately capturing and classifying the emotional tone of the digital content. This finding underscores the importance of algorithm selection and data-handling techniques in optimizing sentiment classification outcomes.

These metrics, accuracy, recall, precision, F-measure, and Area Under the Curve (AUC)—significantly impact the predictions of sentiment analysis models by determining their ability to correctly identify and classify data points as either positive or negative. Accuracy measures the overall correctness of the model but does not distinguish between types of errors, which can be misleading in imbalanced datasets. Recall (or sensitivity) evaluates the model's ability to correctly identify all relevant instances (true positives) in the dataset. Understanding the model's capacity to detect positive sentiments or toxicity is crucial. Precision focuses on the proportion of correctly predicted positive results out of all predicted positives, indicating the remodel's reliability when it predicts a specific sentiment. The F-measure, which combines precision and recall into a single metric, provides a balanced view that is particularly useful when dealing with class imbalances, ensuring that neither precision nor recall is disproportionately prioritized. The AUC measures the model's ability to discriminate between positive and negative classes across various threshold settings, offering insight into the model's performance irrespective of the specific classification threshold. High AUC scores indicate a strong ability to distinguish between sentiments, improving the robustness of predictions. Collectively, these metrics provide a comprehensive evaluation of the model's predictive power, guiding the selection of the most effective algorithm for accurately capturing sentiment trends and ensuring reliable, actionable insights from the analysis.

2.2.4 Content Analysis

The content analysis is meticulously aligned with three critical research questions that explore the cultural narratives of Wamena, focusing on the intersection of traditional practices, tourism, and modernization. Firstly, the study examines how traditional lifestyle elements, such as the Honai housing system, shape Indigenous communities' social and economic organization, revealing the deep interconnection between cultural practices and community structures. Secondly, it investigates the impact of tourism on preserving cultural heritage and traditional practices, particularly in the context of local festivals and the display of mummies, which are integral to Wamena's cultural identity. This line of inquiry highlights the dual role of tourism as both a potential preserver and commodifier of cultural heritage, emphasizing the delicate balance required to maintain cultural integrity. Lastly, the analysis addresses the challenges of modernization and external influences on the cultural identity and traditional knowledge of Wamena's indigenous communities, considering how these forces contribute to cultural erosion or transformation. By addressing these questions, the research provides a comprehensive understanding of the dynamic and multifaceted interactions between tradition, modernity, and external engagement, underscoring the need for culturally sensitive approaches to development and heritage conservation in Wamena.

The content analysis of the videos with IDs d-i11hWIOUU and 0CaFR4AbhQQ reveals several critical themes through systematic coding, including community and family dynamics, cultural impact, challenges of modernization, organizational structures, and the effects of tourism. These themes highlight the complex interplay between traditional values and contemporary pressures facing the indigenous communities of Wamena. The theme of community and family underscores the centrality of kinship and communal bonds in sustaining social cohesion and cultural continuity. Meanwhile, cultural impact reflects the preservation and transformation of indigenous practices under external influences. The challenges of modernization theme addresses the tensions between maintaining cultural identity and adapting to new socioeconomic realities. Organizational structures provide insight into how traditional and modern forms of governance coexist and interact within the community. Lastly, the tourism impact theme explores tourism's dual role as a catalyst for economic opportunity and a potential threat to cultural heritage. These findings offer a nuanced understanding of the factors influencing Wamena's indigenous communities, underscoring the need for strategies that balance cultural preservation with socioeconomic development.

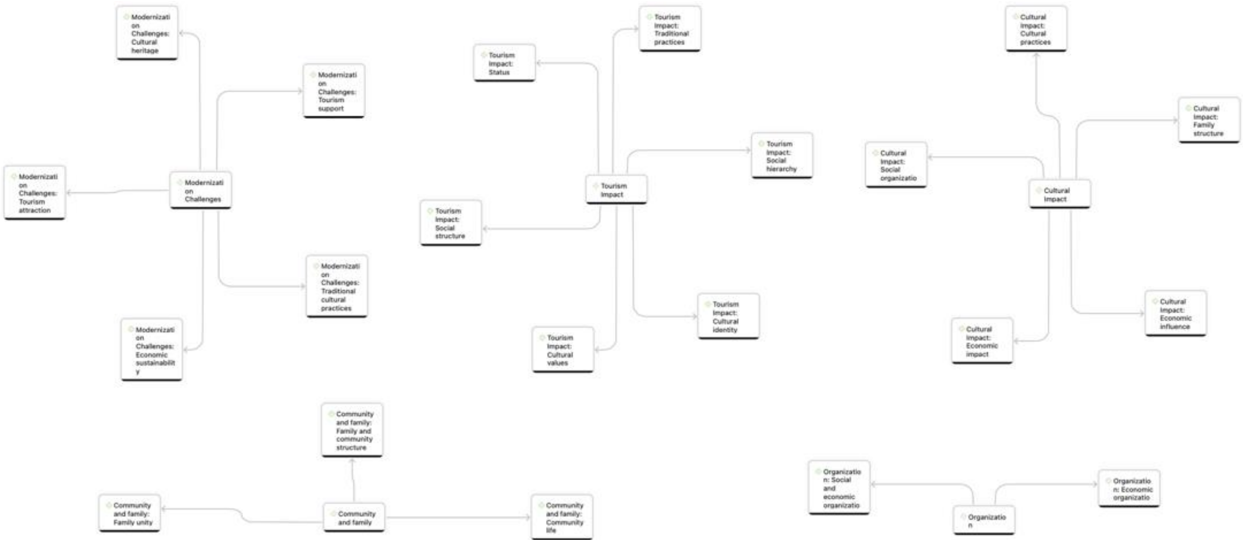


Figure 11. Network of the Wamena Cultural Narratives

Figure 11 shows the network of the Wamena cultural narratives. The network of Wamena Cultural Narratives illustrates a complex web of interconnected themes that shape the socio-cultural fabric of the indigenous communities. Key nodes such as "Modernization Challenges," "Tourism Impact," "Cultural Impact," "Community and Family," and "Organization" reveal the multifaceted dynamics at play in the region. Each theme branches into subtopics that detail specific aspects, such as how modernization pressures threaten traditional cultural heritage or the role of tourism in altering social hierarchies and economic structures. This interconnectedness suggests that the cultural narratives of

Wamena are not isolated but are instead profoundly interwoven with external and internal forces that influence community life. The diagram underscores the intricate balance between preserving cultural identity and adapting to contemporary challenges like globalization and economic development. By mapping these relationships, the network provides a comprehensive framework for understanding how various factors contribute to the evolving cultural landscape of Wamena, highlighting the importance of a holistic approach to cultural preservation and socioeconomic planning.

The themes identified in the network of Wamena Cultural Narratives provide several key insights into the complexities and dynamics of Indigenous community life in Wamena. Firstly, the theme of "Modernization Challenges" highlights the tension between maintaining traditional cultural practices and adapting to new socioeconomic realities, suggesting that modernization often threatens cultural heritage preservation. This theme underscores the need for strategies that balance development with cultural integrity. Secondly, "Tourism Impact" reveals a dual narrative: while tourism can bring economic benefits and promote cultural heritage, it also risks commodifying cultural practices and disrupting social structures. This insight suggests the importance of managing tourism carefully to ensure it supports the community's cultural values rather than undermines them. The "Cultural Impact" theme illustrates how external influences and internal dynamics reshape cultural practices, indicating a dynamic cultural adaptation and resilience process. The theme of "Community and Family" emphasizes the central role of kinship and communal bonds in sustaining social cohesion and cultural continuity, reflecting the importance of social networks in maintaining cultural identity. Finally, the "Organization" theme highlights how traditional and modern governance structures coexist and interact, highlighting communities' adaptive strategies to navigate change. Collectively, these insights point to the need for culturally sensitive approaches to development and conservation that recognize the interdependence of cultural preservation, social cohesion, and economic sustainability.

The toxicity, sentiment, and content analysis results comprehensively overview community perspectives and external influences affecting Wamena's indigenous communities. By examining the level of toxicity in digital discourse, the analysis identifies both the presence of supportive, constructive communication and the instances of harmful language that could disrupt social harmony and cohesion. Sentiment analysis further enriches this understanding by revealing the emotional tone of the community's engagement, distinguishing between expressions of positive sentiment, such as pride and solidarity, and negative sentiment, such as frustration or dissent. Content analysis, on the other hand, sheds light on the specific themes that dominate the discourse, including discussions about cultural preservation, the impact of modernization, and the effects of tourism. These analytical approaches offer a nuanced understanding of how internal community dynamics and external pressures, such as economic development and global cultural flows, shape Indigenous narratives. Ultimately, this multidimensional analysis underscores the importance of fostering an inclusive digital environment that preserves cultural integrity and promotes constructive dialogue and mutual understanding in the face of evolving challenges.

3. RESULT AND DISCUSSION

The discussion in this study is structured into two main sections: the first focuses on the Toxicity Score and Sentiment Classification Model Performance, while the second delves into Understanding Community Perspectives and External Influences in the Narratives of Wamena's Cultural Heritage. The initial section examines the effectiveness of various models in predicting and classifying sentiment based on toxicity scores, evaluating metrics such as accuracy, precision, recall, and F-measure to determine the most reliable model for analyzing sentiment trends. This analysis highlights the strengths and limitations of each algorithm, particularly in handling imbalanced data and differentiating between nuanced emotional expressions. The subsequent section explores how these findings relate to the broader socio-cultural context of Wamena, emphasizing the community's perspectives on cultural heritage, the impact of tourism, and the challenges posed by modernization. This discussion underscores the complex interplay between preserving cultural identity and adapting to external influences, revealing how digital narratives reflect and shape community values and attitudes. Together, these sections provide a comprehensive framework for understanding the dynamics of cultural preservation and change, offering valuable insights for future research and policy development in culturally sensitive regions.

3.1 Toxicity Score and Sentiment Classification Model Performance (Video id 0CaFR4AbhQQ)

The performance of the Toxicity Score and Sentiment Classification Model for the video with ID 0CaFR4AbhQQ provides critical insights into the nature of digital discourse surrounding Wamena's cultural narratives. Analyzing the toxicity scores reveals varying degrees of harmful language use, with average scores suggesting a predominantly low level of toxicity but notable peaks indicating instances of more severe language that could potentially disrupt constructive engagement. The sentiment classification model refines this analysis by categorizing viewer comments into positive and negative sentiments, offering a more granular understanding of community reactions. The combined use of toxicity and sentiment metrics suggests that while most viewers engage positively, heightened negativity or conflict reflects deeper tensions or disagreements within the community. Evaluating the model's performance through metrics such as accuracy, precision, recall, and F-measure confirms the model's capability to differentiate between subtle sentiment variations. However, certain limitations remain in detecting nuanced emotional tones. Overall, this analysis underscores the importance of robust models to monitor and manage digital content, fostering a more inclusive and respectful digital environment that supports healthy discourse on cultural heritage and identity.

The calculation of the toxicity scores reveals a varied landscape of language use within the dataset, reflecting both the presence of civil discourse and instances of harmful communication. The average toxicity score of 0.05326, with the highest value of 0.50790, suggests that while the general tone of the comments remains relatively low in toxicity, notable spikes indicate more severe expressions of hostility or negativity. Similarly, the severe toxicity measure, averaging at 0.00644 with a peak of 0.50704, highlights the occasional use of extremely harmful language, albeit infrequently. The scores for identity attacks and insults, averaging 0.02081 and 0.04047, respectively, with maximum values reaching 0.38177 and 0.55882, point to a moderate presence of personal attacks and offensive remarks within the discourse. The higher peaks in profanity (0.57114) and threats (0.50583) suggest moments where language becomes particularly aggressive or hostile, potentially impacting the overall quality and safety of the digital environment. These findings indicate a complex interaction between civil and uncivil communication, underscoring the importance of continuous monitoring and targeted moderation to promote constructive engagement and minimize the impact of negative behaviors in online communities.

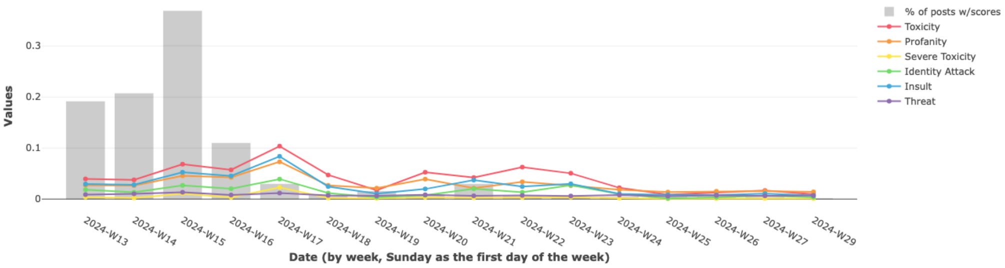


Figure 12. Average Toxicity Score Per Week (637 posts)

Figure 12 shows the average toxicity scores per week (637 posts). Toxicity levels can significantly impact digital engagement by influencing both the tone of discourse and the willingness of users to participate in online conversations. High levels of toxicity, such as insults, identity attacks, or severe language, can create a hostile or unwelcoming environment, discouraging constructive dialogue and leading to reduced user engagement. This negativity can result in a chilling effect, where individuals feel less inclined to share their thoughts or engage with content for fear of encountering harassment or aggression. Conversely, lower toxicity levels foster a more positive and inclusive atmosphere, encouraging more meaningful interactions and participation from a broader range of users. Additionally, sustained high toxicity can damage the reputation of a digital platform, reducing trust and potentially leading to user attrition. Thus, effectively managing toxicity is crucial for maintaining a healthy digital environment that supports vibrant, respectful, and productive engagement.

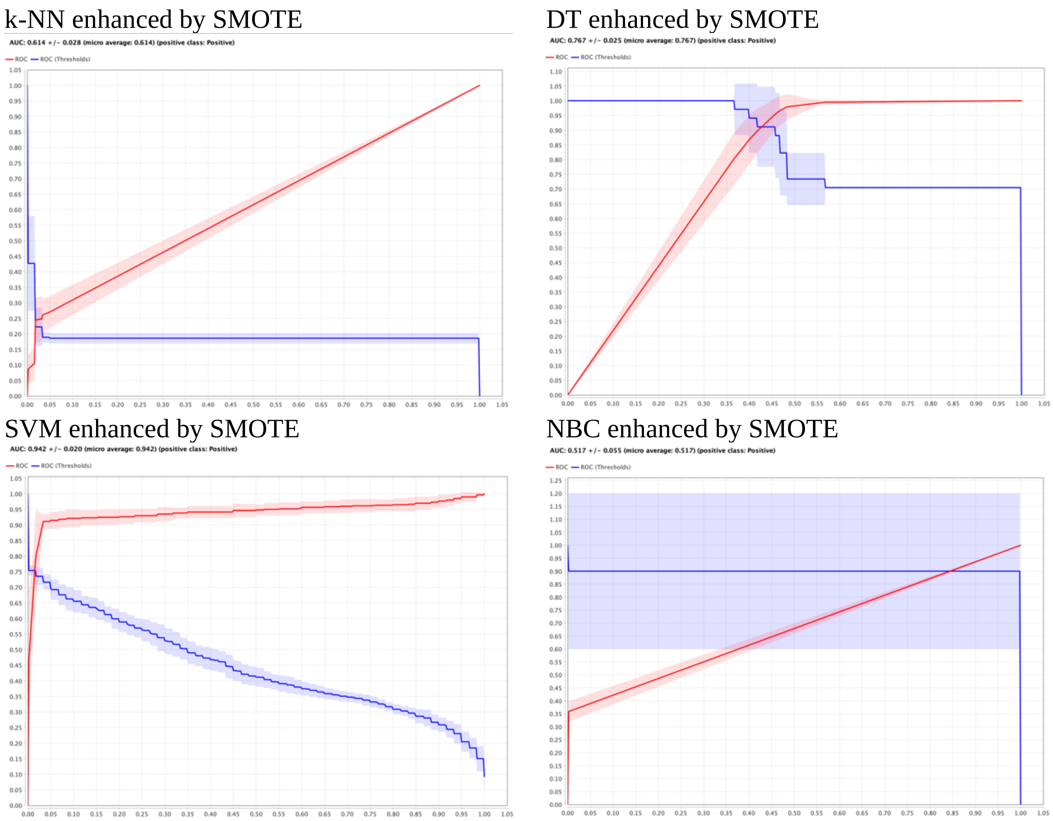


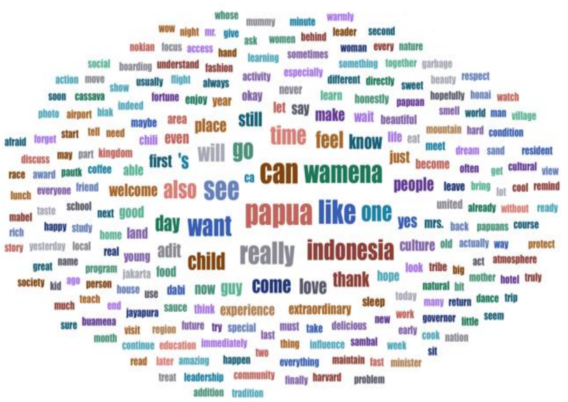
Figure 13. Area Under Curve of DT, k-NN, NBC, and SVM Enhanced by SMOTE

Figure 13 shows the Area Under the Curve of DT, k-NN, NBC, and SVM Enhanced by SMOTE. The Area Under Curve (AUC) analysis for the Decision Tree (DT), k-nearest Neighbors (k-NN), Naive Bayes Classifier (NBC), and Support Vector Machine (SVM) enhanced by SMOTE reveals distinct performance characteristics across these classification models. The AUC scores measure each model's ability to discriminate between positive and negative classes, with higher AUC values indicating better model performance. The SVM enhanced by SMOTE exhibits the most robust AUC curve, reflecting its superior capacity to handle class imbalances and accurately classify sentiment. This model's performance suggests that combining SVM with SMOTE is particularly effective for datasets with uneven class distribution. In contrast, the DT and k-NN models demonstrate moderate AUC scores, indicating reasonable but less optimal performance distinguishing between classes. The NBC model, with a lower AUC score, suggests limitations in its predictive accuracy and sensitivity to the nuances in sentiment classification. These findings underscore the importance of selecting appropriate algorithms and data handling techniques like SMOTE to enhance model performance and parts involving imbalanced datasets. Overall, the AUC analysis confirms that SVM with SMOTE is the most effective model among those tested, offering reliable and precise sentiment classification capabilities.

The Area Under the Curve (AUC) score is a metric used to evaluate the performance of a binary classification model, specifically its ability to distinguish between two classes—typically labeled as positive and negative. The AUC score is derived from the Receiver Operating Characteristic (ROC) curve, which plots the valid positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings. An AUC score ranges from 0 to 1, where a score of 0.5 indicates a model with no discriminative power, equivalent to random guessing, and a score of 1.0 represents a perfect classifier that distinguishes all positive instances from negative ones without error. Higher AUC scores indicate better model performance, reflecting a more remarkable ability to correctly identify positive and negative cases across all possible classification thresholds. For example, an AUC score 0.9 means a 90% chance that the model will correctly distinguish between a randomly chosen positive and negative instance. Therefore, the AUC score is a valuable metric for assessing a model's overall effectiveness, particularly in scenarios with imbalanced datasets, where other metrics like accuracy might be misleading. It provides a single, aggregated measure of a model's predictive performance across all possible decision thresholds, comprehensively evaluating the classifier's robustness and reliability.

The relevance of toxicity and sentiment analysis in this research is integral to understanding the dynamics of digital discourse and its impact on community narratives and perceptions. By examining toxicity levels, the study identifies harmful or aggressive language that may disrupt constructive communication and affect the overall tone of online interactions. Sentiment analysis complements this by categorizing the emotional tone of the content, distinguishing between positive, negative, and neutral sentiments, which provides deeper insights into public reactions and attitudes. Together, these analytical approaches allow for a nuanced understanding of how different types of discourse, supportive or divisive, shape community perspectives and influence cultural identity. This dual focus is particularly relevant in exploring the interplay between internal community dynamics and external influences in Wamena, such as modernization and tourism. The findings underscore the importance of fostering a positive and respectful digital environment that encourages healthy dialogue and preserves cultural integrity, highlighting the need for targeted strategies to manage and mitigate negative online behaviors.

Content Id: 0CaFR4AbhQQ



Content Id: d-i11hWIOUU

Figure 14. Word Cloud of the Content

Figure 14 shows the word cloud of the dataset. The word clouds generated from the datasets of Content ID 0CaFR4AbhQQ and Content ID d-i11hWIOUU provide a visual representation of the most frequently used words, revealing key themes and focal points in the discussions surrounding Wamena's cultural heritage and community dynamics. In the word cloud for Content ID 0CaFR4AbhQQ, prominent words such as "Papua," "like," "want," "see," and "really" suggest a strong interest in exploring and experiencing the cultural and natural landscapes of Papua, with an emphasis on personal reflections and desires. Indicates that the content is primarily centered around visitor experiences, perceptions, and aspirations, reflecting an external perspective that is often curious and appreciative. Meanwhile, the word cloud for Content ID d-i11hWIOUU features terms like "tourist," "Wamena," "mummy," "attraction," and "visit," highlighting the significant role of tourism in shaping the narrative of Wamena's cultural identity. It suggests that discussions are heavily focused on the visibility of cultural heritage sites, such as the display of mummies, and the broader implications of tourism on local traditions. These visualizations underscore the dual narrative in the data: one that reflects the external gaze of tourists fascinated by cultural attractions and another that captures the community's response to and engagement with these external influences. Together, the word clouds offer valuable insights into the overlapping yet distinct dialogues that shape the understanding of Wamena's cultural landscape, underscoring the need for balanced and respectful approaches to cultural tourism and heritage preservation.

Digital narratives about Wamena's culture offer a compelling subject for discussion under the theme of Understanding Community Perspectives and External Influences in the Narratives of Wamena's Cultural Heritage. These narratives provide rich insights into how the indigenous communities of Wamena articulate their cultural identity amidst the pressures of modernization and external engagement, such as tourism. The analysis of these digital texts reveals a dynamic interplay between preserving traditional values and adapting to new socioeconomic realities, reflecting a community deeply engaged in both resistance to and negotiation with external forces. This duality is crucial to understanding how cultural heritage is not only a static entity but a living practice that evolves in response to external influences while maintaining core values and traditions. By examining these digital discourses, it becomes evident that community members use digital platforms to assert their cultural identity, share their experiences, and express concerns or hopes about the future. Such discussions highlight the importance of incorporating local voices into broader conversations about cultural preservation and modernization, ensuring that external influences do not overshadow indigenous perspectives but enrich them in a balanced, respectful manner.

3.2 Discussion: Understanding Community Perspectives and External Influences in the Narratives of Wamena's Cultural Heritage

The discussion on understanding community perspectives and external influences in the narratives of Wamena's cultural heritage reveals the intricate balance between preserving Indigenous traditions and adapting to modern pressures. Analyzing community narratives provides insight into how local cultural practices, such as the Honai housing system and traditional ceremonies, are valued and maintained amidst external forces like tourism and globalization. These narratives reflect a community deeply engaged in safeguarding its cultural identity while navigating the complexities brought by economic development and social change. It is argued that while tourism offers economic opportunities and greater visibility for cultural heritage, it also risks commodifying sacred practices and altering their meanings [46]. Moreover, modernization pressures introduce opportunities for growth and threats to traditional knowledge systems, as younger generations may prioritize contemporary lifestyles over ancestral customs [47]. This dynamic interplay between tradition and change highlights the resilience of the Wamena community in negotiating cultural continuity within a rapidly evolving environment. The findings emphasize the importance of culturally sensitive policies supporting economic development and cultural heritage preservation, fostering a sustainable path that honors the community's rich cultural legacy.

Understanding community perspectives and external influences is crucial for comprehending how Indigenous communities navigate the complexities of cultural preservation and adaptation in a globalized world. Local narratives often reflect a deep commitment to maintaining cultural identity and traditional practices, even as external pressures such as tourism, modernization, and digital communication reshape social dynamics and economic priorities [48], [49]. It is argued that these influences can simultaneously present opportunities for cultural revitalization and challenges to cultural integrity, depending on how the community manages and perceives them. Analyzing these perspectives reveals a nuanced interplay between resilience and change, where community members actively engage with external forces to assert their cultural values while adapting to new realities [50], [51]. This process underscores the importance of fostering environments that respect and incorporate Indigenous voices, ensuring that cultural heritage is preserved and allowed to evolve in ways that align with the community's aspirations. Recognizing and integrating community perspectives is essential for developing inclusive strategies that honor cultural diversity and support sustainable development.

The narratives of Wamena's cultural heritage encapsulate the rich tapestry of traditions, values, and practices that define the identity of its indigenous communities. Central to these narratives are elements such as the Honai housing system, traditional rituals, and communal values passed down through generations, serving as a foundation for social organization and a source of cultural pride. It is argued that these cultural elements are not merely historical artifacts but living practices that continue to shape the social and economic fabric of the community today [52]. Analyzing these narratives reveals the community's resilience in preserving its cultural heritage while navigating the influences of modernization, tourism, and external economic pressures. This interplay between tradition and change demonstrates the community's adaptive strategies, vital for maintaining cultural continuity in a rapidly evolving world [53]. Thus, the narratives of Wamena's cultural heritage offer profound insights into how indigenous communities sustain their identity, emphasizing the importance of preserving cultural heritage in a manner that is both respectful of tradition and responsive to contemporary challenges.

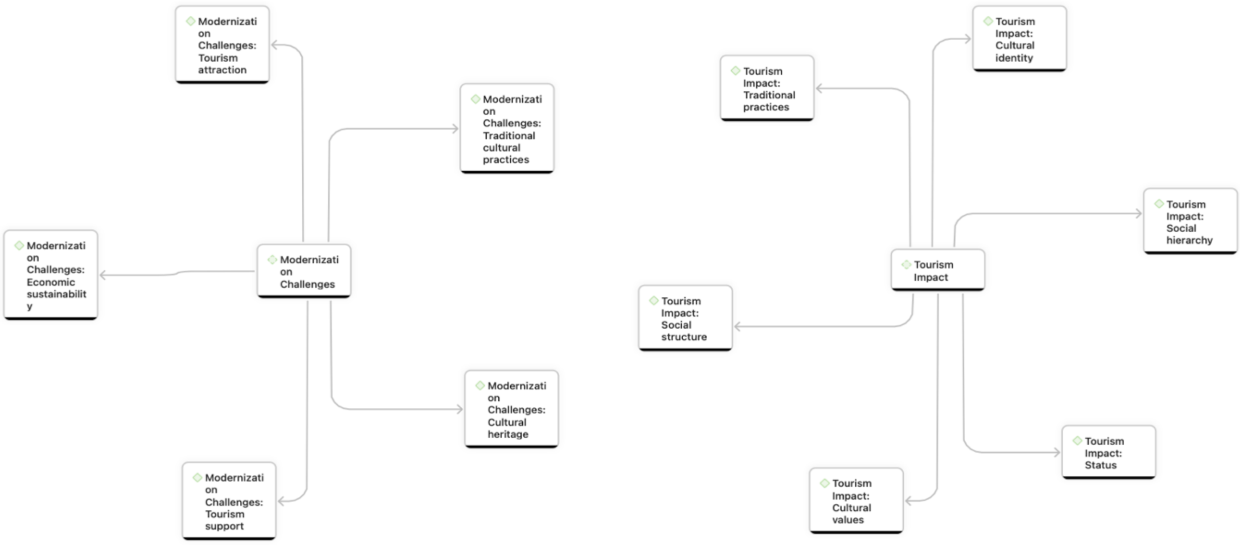


Figure 14. Tourism Impact and Modernization Challenges

Figure 14 shows the tourism impact and modernization challenges based on the cultural narratives of Wamena, Papua. Based on the network visualization results related to the cultural narratives of Wamena, two key themes emerge modernization challenges and tourism impact. Firstly, the theme of modernization challenges highlights a blend of experiences and reflections on various locations in Indonesia, with a particular focus on the culture and people of Papua. The content emphasizes the importance of supporting upcoming cultural festivals. It draws attention to figures such as Adit Marciano and Bernard Mabel, who are notable for their involvement in theater and the construction of traditional houses, showcasing efforts to preserve cultural heritage amid modernization pressures. This narrative underscores the community's proactive stance in maintaining cultural identity while navigating the socioeconomic changes brought by modernization. Secondly, the theme of tourism impact explores a travel experience in Papua, capturing interactions with local communities, observations of daily life, and cultural insights alongside reflections on the region's distinctiveness and challenges. The discussion reveals an admiration for the local people, particularly children, and expresses hope for positive development and unity within Indonesia's diverse society. It also emphasizes the significance of supporting cultural events and fostering an understanding and respect for different cultures. These themes illustrate a complex interplay between preserving cultural heritage and adapting to external influences, highlighting the need for strategies that balance cultural integrity with socioeconomic development in Wamena.

Tourism influences Wamena's culture positively and negatively, impacting the preservation and evolution of cultural practices and community dynamics. On the positive side, tourism brings economic opportunities to support local livelihoods and provide resources for maintaining and promoting cultural heritage. For instance, tourism can help fund cultural festivals, the preservation of traditional arts, and the upkeep of historical sites, reinforcing community pride and cultural identity. Additionally, increased visibility from tourism can raise awareness and appreciation of Wamena's unique cultural practices among a broader audience, fostering a greater understanding and respect for indigenous traditions. However, tourism also challenges Wamena's culture by risking commodifying and altering traditional practices to cater to tourists' expectations. It can lead to the oversimplification or commercialization of cultural expressions, which may undermine their authenticity and significance. Moreover, the influx of tourists can disrupt local social structures and potentially create conflicts between economic development and cultural preservation. The emphasis on showcasing certain cultural elements for tourism can overshadow other aspects of Indigenous culture, leading to a skewed representation that might not fully capture the community's richness and diversity. Thus, tourism's influence on Wamena's culture is a delicate balance between fostering economic benefits and safeguarding cultural integrity, necessitating careful management and culturally sensitive approaches to ensure sustainable and respectful engagement.

The analysis of "Cultural Impact, Organization, Community, and Family" reveals the interconnectedness of social structures and cultural practices in shaping the community dynamics of Wamena. The cultural impact is evident in how traditional practices influence economic activities and social organization, highlighting the role of cultural heritage in sustaining economic resilience and social cohesion. This interplay suggests that cultural practices are symbolic and functionally embedded within the economic and social frameworks, influencing family structures and community life. The organizational aspects, particularly social and economic organization, demonstrate how cultural values are reflected in governance and economic strategies, fostering a sense of unity and collective identity. It is posited that family and community structures are vital in preserving cultural continuity, serving as the primary units for transmitting cultural knowledge and social norms. The diagram underscores the importance of maintaining a balance between cultural preservation and adaptation to modern socio-economic challenges, suggesting that a solid organizational foundation rooted in cultural values is essential for the community's sustainability and growth. This interconnected framework emphasizes the need for culturally informed policies that support preserving traditional practices and developing sustainable social and economic structures.

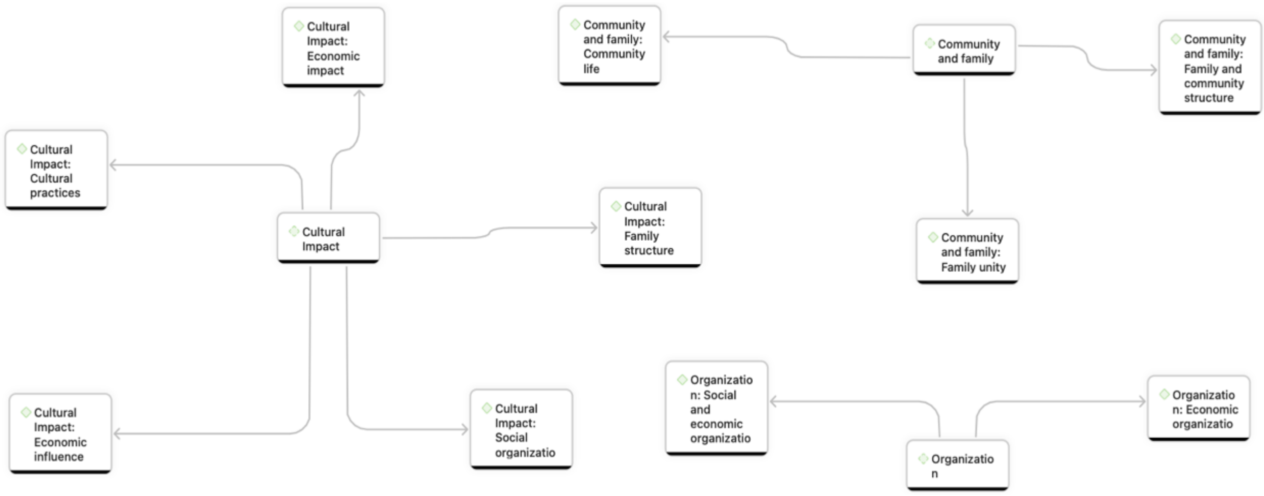


Figure 15. Cultural Impact, Organization, Community and Family

Figure 15 shows the network of cultural impact, organization, community, and family based on the cultural narratives. In the context of cultural impact, organization, community, and family from the narratives of Wamena's culture, several key themes emerge that reflect the community's resilience and aspirations. The texts highlight collective experiences, such as crowded pre-COVID events and cultural celebrations, underscoring the importance of communal gatherings in maintaining social bonds and cultural continuity. These narratives also reveal a deep respect for cultural heritage, referencing local traditions and practices, such as using a sauce resembling tomato and chili to enhance the flavor of meals, indicating the significance of cultural expressions in everyday life. The impact of COVID-19 on tourism and the community is discussed, with a hopeful outlook for recovery and a brighter future for Papua, emphasizing the role of children as vital agents of change and the importance of education in empowering the community. Despite challenges like stunting and economic hardships, there is a strong sense of optimism and determination to improve the community's well-being through cultural preservation and unity. The texts also highlight ongoing efforts to support local festivals and cultural initiatives, illustrating the community's commitment to fostering cultural pride and solidarity. These narratives reflect a dynamic interplay between tradition and modern aspirations, emphasizing the need for continued support and understanding to promote cultural resilience and social cohesion in Wamena.

Modernization impacts Wamena in complex and often conflicting ways, influencing the indigenous communities' cultural and socioeconomic landscapes. On the other hand, modernization brings opportunities for economic development, improved infrastructure, education, and healthcare, which can enhance the quality of life and provide new avenues for growth. For instance, increased access to technology and communication can connect Wamena to broader markets and information networks, fostering economic opportunities and social development. However, modernization also poses significant challenges to preserving Wamena's traditional culture and identity. The influx of modern values, lifestyles, and economic pressures can lead to cultural erosion, where younger generations might favor contemporary practices over traditional customs, potentially diminishing the community's cultural heritage. In addition, external influences and economic activities, such as tourism, can commodify cultural practices and disrupt traditional social structures, leading to tensions between maintaining cultural integrity and embracing new ways of life. Thus, modernization's impact on Wamena is characterized by a delicate balance between embracing progress and safeguarding cultural heritage, necessitating thoughtful strategies that respect and incorporate local traditions while promoting sustainable development.

Tourism and modernization can coexist harmoniously by adopting culturally sensitive and sustainable development strategies that respect and preserve local traditions while fostering economic growth and social progress. One practical approach is integrating cultural heritage into tourism initiatives to promote awareness and appreciation of local customs rather than merely commodifying them for profit. It involves creating educational and immersive tourism experiences that allow visitors to engage meaningfully with the community's traditions and values. Furthermore, involving local communities in the planning and managing of tourism and modernization projects ensures that their needs and perspectives are prioritized, fostering a sense of ownership and empowerment. Investments in infrastructure, education, and technology should be made in a manner that supports both tourism and the sustainable development of the community, such as by improving access to remote cultural sites in a way that does not damage them or by using digital platforms to market local crafts and traditions globally. Policies that encourage environmentally sustainable practices, protect cultural sites, and promote social equity can help mitigate the negative impacts of tourism and modernization, ensuring that these forces contribute positively to the community's well-being. By balancing economic opportunities with cultural preservation and environmental stewardship, tourism and modernization can coexist to support a vibrant, sustainable, and culturally rich future for communities like Wamena.

4. CONCLUSION

This study comprehensively analyzes digital narratives surrounding Wamena's cultural heritage, utilizing the Digital Content Reviews and Analysis Framework to examine sentiment, toxicity, and thematic content. The research highlights the intricate interplay between community perspectives, cultural preservation, modernization, and external influences such as tourism. The toxicity analysis revealed that while most online discourse about Wamena is constructive and supportive, there are instances of harmful or toxic language that could potentially disrupt social cohesion and cultural integrity. The average toxicity score was relatively low at 0.05326, but notable peaks, with scores reaching up to 0.50790, indicate moments of heightened negativity or conflict within the digital content. Categories such as severe toxicity, identity attacks, insults, profanity, and threats, though generally showing low average scores, revealed occasional spikes that warrant further scrutiny. These findings underscore the need for continuous monitoring and targeted moderation of digital communication to foster a positive and inclusive environment that supports constructive dialogue and minimizes harmful interactions. The sentiment analysis complemented the toxicity findings by providing a more granular understanding of the emotional tones present in the digital narratives. Most digital content expressed positive sentiments, reflecting community pride and solidarity. Still, there were also instances of negative sentiment, indicating frustration or dissent, particularly in discussions about cultural erosion or external pressures. The dual focus on sentiment and toxicity provided a nuanced understanding of how different types of discourse, whether supportive or divisive, shape community perspectives and influence culturally identifying model performance. The study employed several machine learning algorithms to classify sentiment and detect toxicity within the digital content. Among these, the Support Vector Machine (SVM) enhanced by Synthetic Minority Over-sampling Technique (SMOTE) demonstrated superior performance, with an accuracy rate of 87.29%, indicating its effectiveness in handling imbalanced datasets and accurately classifying both positive and negative sentiments. This finding highlights the importance of robust algorithms for capturing digital content's emotional and toxic dimensions, essential for developing strategies to manage online behavior and support positive engagement. The content analysis results provided further insights into the key themes dominating the digital narratives around Wamena's cultural heritage. Several prominent themes were identified, including community and family dynamics, cultural impact and resilience, challenges of modernization, tourism impact, and organizational structures and governance. These themes reflect the community's deep-rooted commitment to preserving cultural identity while navigating the complexities introduced by modernization and globalization. The analysis underscored the dual role of tourism as both a catalyst for economic opportunity and a potential threat to cultural integrity, emphasizing the need for sustainable tourism practices that respect and preserve the authenticity of cultural expressions. Overall, this study highlights the multifaceted nature of digital narratives and their impact on cultural preservation in Wamena. The findings underscore the importance of fostering an inclusive digital environment that preserves cultural integrity and promotes constructive dialogue and mutual understanding. The Digital Content Reviews and Analysis Framework proved valuable in uncovering the socio-cultural dynamics of Wamena’s digital narratives, providing actionable insights for policymakers, digital content moderators, and community leaders. Future research should expand the scope of this framework to other cultural contexts and utilize mixed-method approaches to enhance the understanding of digital communication dynamics further. Doing so makes it possible to develop more effective strategies for managing digital content, fostering social harmony, and supporting sustainable cultural tourism in culturally sensitive regions like Wamena.

ACKNOWLEDGMENT

I would like to extend my sincere gratitude to the Information System Study Program of Baliem Papua University for their invaluable support and guidance throughout this research. Their insights and expertise have contributed significantly to the depth and quality of this study. Thank you for providing the resources and a collaborative environment that facilitated the successful completion of this work.

REFERENCES

[1] R. K. Shrestha, J. N. P. L’Espoir Decosta, and M. Whitford, “Indigenous knowledge systems and socio-cultural values for sustainable tourism development: insights from Indigenous Newars of Nepal,” *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–25, 2024, doi: 10.1080/09669582.2024.2316298.

[2] A. Gillies, “Can a formalised model of co-production contribute to empowering indigenous communities in decisions about land use?,” *J. Glob. Responsib.*, vol. 12, no. 3, pp. 364–372, Jan. 2021, doi: 10.1108/JGR-09-2020-0088.

[3] B. Beamer and K. C. Gleason, “Reflections on the impact of informal sector tourism on indigenous Namibian Craft processes,” *Arts Mark.*, vol. 12, no. 1, pp. 1–16, Jan. 2022, doi: 10.1108/AAM-05-2020-0015.

[4] K. Krittayaruangroj, S. Suriyankietkaew, and P. Hallinger, “Research on sustainability in community-based tourism: a bibliometric review and future directions,” *Asia Pacific J. Tour. Res.*, vol. 28, no. 9, pp. 1031–1051, 2023, doi: 10.1080/10941665.2023.2276477.

[5] P. Yodsurang, A. Kiatthanawat, P. Sanoamuang, A. Kraseain, and W. Pinijvarasin, “Community-based tourism and heritage consumption in Thailand: An upside-down classification based on heritage consumption,” *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2096531.

[6] T. Nguyen Phu and H. Nguyen Thi Thu, “Assessment of tourism service quality for traditional craft villages in Da Nang city, Vietnam,” *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2108636.

[7] R. E. Caraka *et al.*, “Connectivity, sport events, and tourism development of Mandalika’s special economic zone: A perspective

from big data cognitive analytics," *Cogent Bus. Manag.*, vol. 10, no. 1, 2023, doi: 10.1080/23311975.2023.2183565.

[8] I. Moeis, R. Febriani, I. Sandra, and M. Pabbajah, "Intercultural values in local wisdom: A global treasure of Minangkabau ethnic in Indonesia," *Cogent Arts Humanit.*, vol. 9, no. 1, p. 2116841, Dec. 2022, doi: 10.1080/23311983.2022.2116841.

[9] H. V. Nguyen, T. D. Quang, T. Alang, L. D. H. Ngo, and T. D. Nguyen, "Toward sustainable tourism practice in the post-COVID-19: Perspectives from Nha Trang, Vietnam," *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2064590.

[10] M. A. Forson, "Unpacking rural tourism on household livelihood: Insights from Kakum Ghana," *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2021.2019453.

[11] Y. Abunie, E. Teshome, and M. Dessiye, "Tourism's contribution to sustainable conservation of natural and cultural heritage: Evidence from the Lake Tana Biosphere Reserve," *Cogent Soc. Sci.*, vol. 10, no. 1, 2024, doi: 10.1080/23311886.2023.2292756.

[12] Q. B. Baloch *et al.*, "Determinants of evolving responsible tourism behavior: Evidences from supply chain," *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2099565.

[13] A. M. Alomran, T. S. A. Azim Ahmed, A. A. Alajloni, A. M. Kassem, M. M. A. El-Kholy, and E. S. M. Shaker, "Do external corporate social responsibility activities related to the local community during COVID-19 increase employees' job performance? Evidence from hospitality and tourism establishments at Hail city, KSA," *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2093467.

[14] S. Niavis, T. Papatheochari, T. Koutsopoulou, H. Coccossis, and Y. Psycharis, "Considering regional challenges when prioritizing tourism policy interventions: evidence from a Mediterranean community of projects," *J. Sustain. Tour.*, vol. 30, no. 4, pp. 663–684, 2022, doi: 10.1080/09669582.2021.1876074.

[15] C. Rinaldi, A. Cavicchi, and R. N. S. Robinson, "University contributions to co-creating sustainable tourism destinations," *J. Sustain. Tour.*, vol. 30, no. 9, pp. 2144–2166, 2022, doi: 10.1080/09669582.2020.1797056.

[16] A. Wondirad, K. Wu, E. Teshome, and T. J. Lee, "What drives the consumption of cruise ship tourism? Analyzing factors that motivate visitors to participate in cruise ship holidays through a content analysis of selected journal publications," *Cogent Soc. Sci.*, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2282411.

[17] G. M. Majid, I. Tussyadiah, Y. R. Kim, and A. Pal, "Intelligent automation for sustainable tourism: a systematic review," *J. Sustain. Tour.*, vol. 31, no. 11, pp. 2421–2440, 2023, doi: 10.1080/09669582.2023.2246681.

[18] R. Iftikhar, M. S. Khan, and K. Pasanchay, "Virtual reality tourism and technology acceptance: a disability perspective," *Leis. Stud.*, vol. 42, no. 6, pp. 849–865, 2023, doi: 10.1080/02614367.2022.2153903.

[19] M. Madandola and D. Boussaa, "Cultural heritage tourism as a catalyst for sustainable development; the case of old Oyo town in Nigeria," *Int. J. Herit. Stud.*, vol. 29, no. 1–2, pp. 21–38, 2023, doi: 10.1080/13527258.2023.2169332.

[20] D. Yuliati, E. Susilowati, and T. Suliyati, "Preservation of The Old City of Semarang, Central Java, Indonesia, and its development as a cultural tourism asset," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2170740.

[21] J. M. Yachin and D. Ioannides, "'Making do' in rural tourism: the resourcing behaviour of tourism micro-firms," *J. Sustain. Tour.*, vol. 28, no. 7, pp. 1003–1021, 2020, doi: 10.1080/09669582.2020.1715993.

[22] B. Frame, D. Liggett, K. Lindström, R. M. Roura, and L. M. van der Watt, "Tourism and heritage in Antarctica: exploring cultural, natural and subliminal experiences," *Polar Geogr.*, vol. 45, no. 1, pp. 37–57, 2022, doi: 10.1080/1088937X.2021.1918787.

[23] L. Li, J. Wang, and S. Hazra, "A realist analysis of civilized tourism in China: a cultural structural perspective," *J. Crit. Realis.*, vol. 22, no. 4, pp. 692–719, 2023, doi: 10.1080/14767430.2023.2237341.

[24] A. Viken, "Tourism appropriation of Sámi land and culture," *Acta Boreal.*, vol. 39, no. 2, pp. 95–114, 2022, doi: 10.1080/08003831.2022.2079276.

[25] I. Pinke-Sziva, K. Keller, and L. Kovács, "Smart positioning: how smart technologies can increase the attractiveness of heritage tourism destinations? The case of a small-scale Hungarian heritage city," *J. Herit. Tour.*, pp. 1–19, 2023, doi: 10.1080/1743873X.2023.2276271.

[26] M. B. Duignan and I. Pappalepore, "How do Olympic cities strategically leverage New Urban Tourism? Evidence from Tokyo," *Tour. Geogr.*, vol. 25, no. 2–3, pp. 425–449, 2023, doi: 10.1080/14616688.2021.1895296.

[27] L. Bellato, N. Frantzeskaki, E. Lee, J. M. Cheer, and A. Peters, "Transformative epistemologies for regenerative tourism: towards a decolonial paradigm in science and practice?," *J. Sustain. Tour.*, vol. 32, no. 6, pp. 1161–1181, 2024, doi: 10.1080/09669582.2023.2208310.

[28] W. Luo and N. Graburn, "Curating pasts: musealization and tourism in Chinese cities," *Int. J. Tour. Cities*, vol. 10, no. 2, pp. 509–527, Jan. 2024, doi: 10.1108/IJTC-05-2022-0110.

[29] H. T. P. Thao, "Tourism policy in Vietnam: an evaluation using the difference-in-differences model," *J. Policy Res. Tour. Leis. Events*, pp. 1–24, 2023, doi: 10.1080/19407963.2023.2273558.

[30] C. E. Yu and X. Zhang, "The embedded feelings in local gastronomy: a sentiment analysis of online reviews," *J. Hosp. Tour. Technol.*, vol. 11, no. 3, pp. 461–478, 2020, doi: 10.1108/JHTT-02-2019-0028.

[31] T. Matiza and E. Slabbert, "The destination marketing and media profile – travel motives nexus amid tourism crisis: the mediating effect of the nation brand," *J. Hosp. Tour. Insights*, vol. 7, no. 1, pp. 1–27, Jan. 2024, doi: 10.1108/JHTI-05-2022-0174.

[32] F. Frenzel, J. Giddy, and T. Frisch, "Digital technology, tourism and geographies of inequality," *Tour. Geogr.*, vol. 24, no. 6–7, pp. 923–933, 2022, doi: 10.1080/14616688.2022.2142843.

[33] A. Tuomi, E. Moreira Kares, and H. Zainal Abidin, "Digital cultural tourism: older adults' acceptance and use of digital cultural tourism services," *Scand. J. Hosp. Tour.*, vol. 23, no. 2–3, pp. 226–247, 2023, doi: 10.1080/15022250.2023.2256698.

[34] X. Yang, H. Long, H. Jia, and X. Wen, "Analysis of social results of tourism performing arts industry in cultural tourism convergence using fuzzy comprehensive analysis methods," *Econ. Res. Istraz.* , vol. 36, no. 3, p., 2023, doi: 10.1080/1331677X.2022.2147976.

[35] P. Riva and D. Agostino, "Latent dimensions of museum experience: assessing cross-cultural perspectives of visitors from tripadvisor reviews," *Museum Manag. Curatorsh.*, vol. 37, no. 6, pp. 616–640, 2022, doi: 10.1080/09647775.2022.2073560.

[36] K. Muangasame and E. Tan, "Phygital rural cultural heritage: a digitalisation approach for destination recovery and resilience," *Worldw. Hosp. Tour. Themes*, vol. 15, no. 1, pp. 8–17, Jan. 2023, doi: 10.1108/WHATT-08-2022-0096.

[37] Y. Gu and J. Zhu, "The effect of online review and interaction on value co-creation in tourism virtual community," *Cogent Soc. Sci.*, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2261234.

[38] L. Zhong *et al.*, "Senses, emotions and wellbeing in forest recreation and tourism," *Tour. Recreat. Res.*, 2024, doi: 10.1080/02508281.2024.2312347.
[39] H. Han, "Consumer behavior and environmental sustainability in tourism and hospitality: a review of theories, concepts, and latest research," *J. Sustain. Tour.*, vol. 29, no. 7, pp. 1021–1042, 2021, doi: 10.1080/09669582.2021.1903019.
[40] H. Xu, L. T. O. Cheung, J. Lovett, X. Duan, Q. Pei, and D. Liang, "Understanding the influence of user-generated content on tourist loyalty behavior in a cultural World Heritage Site," *Tour. Recreat. Res.*, vol. 48, no. 2, pp. 173–187, 2023, doi: 10.1080/02508281.2021.1913022.
[41] J. Li, X. Peng, X. Liu, H. Tang, and W. Li, "A study on shaping tourists' conservational intentions towards cultural heritage in the digital era: exploring the effects of authenticity, cultural experience, and place attachment," *J. Asian Archit. Build. Eng.*, vol. 00, no. 00, pp. 1–20, 2024, doi: 10.1080/13467581.2024.2321999.
[42] H. Jang and M. Park, "Social media, media and urban transformation in the context of overtourism," *Int. J. Tour. Cities*, vol. 6, no. 1, pp. 233–260, 2020, doi: 10.1108/IJTC-08-2019-0145.
[43] A. Sundaram, H. Subramaniam, S. H. A. Hamid, and A. M. Nor, "A Systematic Literature Review on Social Media Slang Analytics in Contemporary Discourse," *IEEE Access*, vol. 11, pp. 132457–132471, 2023, doi: 10.1109/ACCESS.2023.3334278.
[44] W. Tafesse, "YouTube marketing: how marketers' video optimization practices influence video views," *Internet Res.*, vol. 30, no. 6, pp. 1689–1707, Jan. 2020, doi: 10.1108/INTR-10-2019-0406.
[45] R. A. Tabaeeian, A. Yazdi, N. Mokhtari, and A. Khoshfetrat, "Host-tourist interaction, revisit intention and memorable tourism experience through relationship quality and perceived service quality in ecotourism," *J. Ecotourism*, vol. 22, no. 3, pp. 406–429, 2023, doi: 10.1080/14724049.2022.2046759.
[46] B. King, G. Richards, and A. M. C. Chu, "Developing a tourism region through tourism and culture: bordering, branding, placemaking and governance processes," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–15, 2023, doi: 10.1080/02508281.2023.2207156.
[47] T. D. Quang, W. B. Noseworthy, and D. Paulson, "Rising tensions: heritage-tourism development and the commodification of 'Authentic' culture among the Cham community of Vietnam," *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2116161.
[48] A. Giampiccoli, D. O. Mtapuria, and S. Jugmohan, "Community-based tourism and animals: Theorising the relationship," *Cogent Soc. Sci.*, vol. 6, no. 1, 2020, doi: 10.1080/23311886.2020.1778965.
[49] M. Saad, N. A. Husin, S. N. H. Rosman, R. Husain, and L. Abdul Latif @ Bapooe, "The nexus of community perception on turning idyllic Panchor town into a future river tourism destination," *Cogent Soc. Sci.*, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2253021.
[50] J. Mensah, "Community perception of heritage values regarding a global monument in Ghana: implications for sustainable heritage management," *J. Humanit. Appl. Soc. Sci.*, vol. 4, no. 4, pp. 357–375, Jan. 2022, doi: 10.1108/jhass-01-2021-0010.
[51] T. Wanitchakorn and K. Muangasame, "The identity change of rural–urban transformational tourism development in Chiang Mai heritage city: local residents' perspectives," *Int. J. Tour. Cities*, vol. 7, no. 4, pp. 1008–1028, Jan. 2021, doi: 10.1108/IJTC-08-2020-0169.
[52] L. Chang, X. Huang, and M. Meng, "Study on tourist's loyalty of Zhinan Village in the view of tourism landscape," *Cogent Soc. Sci.*, vol. 7, no. 1, 2021, doi: 10.1080/23311886.2021.1997403.
[53] G. Helgadóttir and K. Dashper, "20 years of Nordic rural tourism research: a review and future research agenda," *Scand. J. Hosp. Tour.*, vol. 21, no. 1, pp. 60–69, 2021, doi: 10.1080/15022250.2020.1823246.