Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i3.836 Published By DRPM-UBD

Exploring the Digital Narratives in Tourism and Culture through The Case of Rambu Solo: Sentiment, Toxicity, and Content Analysis

Yerik Afrianto Singgalen[1*]

[1*]Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: [1*]yerik.afrianto@atmajaya.ac.id

Abstract

This research urgently addresses the need to understand and manage viewer interactions with culturally significant video content, particularly the Rambu Solo ritual. By integrating the Digital Content Reviews and Analysis Framework with sentiment classification performance, toxicity score evaluation, and content analysis, the study systematically analyzed 21,562 posts across four videos, revealing critical themes related to cultural preservation and tourism impact that shaped viewer perceptions. Sentiment and toxicity evaluations of 15,762 posts showed an average toxicity score of 0.068, with a peak of 0.85174. Sentiment classification, using algorithms like SVM, k-NN, NBC, and DT, highlighted the superior performance of SVM enhanced by SMOTE, with an accuracy of 81.97%. However, the study identified limitations in automated sentiment analysis tools, noting that they may not fully capture the complexities of human expression. This research recommends incorporating advanced natural language processing techniques and multimodal analysis within the framework. This comprehensive methodology offers essential insights into the intersection of culture, tourism, and digital media, emphasizing the importance of creating and managing content that respects and promotes cultural heritage in the digital age. The findings are crucial for developing more effective strategies for digital content creation and community engagement, ensuring that cultural narratives are presented thoughtfully and respectfully to global audiences.

Keywords: Tourism; Culture; Rambu Solo; Sentiment; Toxicity; Content; Digital Narratives

1. INTRODUCTION

Cultural practices serve as narratives embodying significant events and symbols relevant to specific communities' identity and potential tourism attractions that can significantly impact economic and environmental sectors. Developing cultural heritage as a tourist attraction involves a nuanced process where cultural elements are strategically showcased, thus contributing to the diversification of local economies [1]–[4]. Integrating culture into the tourism sector provides an

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

opportunity to preserve traditions while promoting economic growth [5]–[7]. However, the commercialization of culture requires a delicate balance to prevent the degradation of cultural authenticity and the potential for environmental degradation. Therefore, the development of cultural tourism must be approached with a sustainable strategy that ensures long-term benefits for both the community and the environment.

Cultural attractions have increasingly become a magnet for domestic and international tourists, establishing a relevant and mutually beneficial relationship between tourism and culture. The unique characteristics and traditions inherent in cultural practices offer tourists distinctive experiences often unavailable elsewhere, enhancing the appeal of destinations [8]–[10]. This symbiotic relationship promotes cultural heritage preservation and stimulates economic growth through increased tourism revenue [11], [12]. Integrating cultural elements into the tourism industry requires careful management to maintain cultural integrity while optimizing the economic benefits. Consequently, fostering a balanced approach to cultural tourism development is essential for sustaining cultural preservation and economic prosperity.

This study aims to explore digital narratives within the realm of Tourism and Culture by focusing on the case of Rambu Solo through a comprehensive analysis of sentiment, toxicity, and content. The digital representation of Rambu Solo, a traditional Torajan funeral ceremony, offers a unique lens through which the intersection of cultural practices and modern technology can be examined. The study provides insights into how cultural heritage is perceived and discussed in digital spaces by analyzing online narratives' sentiments and potential toxicity. The findings underscore the importance of managing digital content to preserve cultural integrity while navigating the complexities of public discourse. Ultimately, this research contributes to a deeper understanding of how cultural tourism is shaped and influenced in the digital age.

The framework employed in this study is the Digital Content Reviews and Analysis Framework, which provides a structured approach to examining and evaluating online content. This framework facilitates the systematic collection and analysis of digital narratives, enabling a comprehensive understanding of how diverse audiences generate, perceive, and interact with content. Adopting this framework is critical in addressing the complexities of digital environments where content is rapidly produced and disseminated, often influencing public perception and cultural discourse. Through this analytical lens, the framework allows for identifying patterns, sentiment trends, and potential areas of toxicity within digital content, thereby contributing to the broader discourse on digital culture management. Consequently, applying this framework ensures a robust and systematic exploration of digital content, yielding relevant and actionable insights in the context of cultural tourism.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The urgency of this research is underscored by the rapidly evolving dynamics of digital content and its profound impact on cultural narratives within the tourism sector. As digital platforms become the primary medium for sharing and interpreting cultural experiences, there is an increasing need to critically assess how these narratives influence public perception and cultural preservation. This research is essential in identifying and mitigating the risks associated with misinformation, content toxicity, and the potential erosion of cultural authenticity in the digital sphere. By addressing these issues, the study provides timely and necessary insights that inform strategies for sustainable cultural tourism in an increasingly digitalized world. Therefore, the research contributes to a more informed and responsible approach to managing digital content related to cultural heritage, ensuring its integrity and sustainability for future generations.

This research's theoretical and practical implications are significant, as it bridges the gap between digital content analysis and cultural tourism studies. Theoretically, this research contributes to the academic discourse by expanding the understanding of how digital narratives shape cultural perceptions and influence tourism dynamics. It offers a nuanced perspective on the intersection of digital media and cultural heritage, providing a framework for analyzing the impact of online content on public attitudes toward cultural practices. Practically, the findings of this research are instrumental in guiding policymakers, tourism stakeholders, and cultural institutions in developing strategies to manage digital content effectively. By applying these insights, stakeholders can enhance cultural preservation efforts while optimizing the tourism experience, ensuring that cultural integrity and economic benefits are sustained. Thus, the research advances theoretical knowledge and offers practical solutions for contemporary challenges in cultural tourism management.

Similar research in tourism and culture has consistently explored the intricate relationship between cultural heritage and its representation within the tourism industry. These studies often focus on the role of cultural narratives in shaping tourists' perceptions and experiences and cultural tourism's economic and social impacts [13]–[15]. The growing body of research emphasizes the importance of understanding how cultural elements are commodified and consumed within the tourism context, highlighting opportunities and challenges [16]–[18]. It is argued that while cultural tourism can drive economic development and cultural preservation, it also risks oversimplifying or misrepresenting complex cultural practices [19]–[22]. Therefore, such research contributes to a critical understanding of the dual role of tourism as both a promoter and a potential threat to cultural authenticity, informing strategies that seek to balance cultural integrity with economic viability.

The limitation of this study lies in its reliance on digital content as the primary data source, which may not fully capture the complexity and richness of cultural

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

narratives. While digital platforms provide a vast array of data for analysis, they are also susceptible to biases, misinformation, and the potential exclusion of voices from less digitally connected communities. This limitation raises concerns about the representativeness and generalizability of the findings, as digital narratives may not accurately reflect the diversity of perspectives within the cultural context being studied. Moreover, while valuable, the study's focus on sentiment and toxicity analysis might overlook deeper cultural meanings and interpretations that require more nuanced qualitative approaches. Therefore, while this research offers significant insights into the digital representation of cultural practices, it should be complemented with further studies incorporating a broader range of data sources and methodologies to ensure a more comprehensive understanding of the subject matter.

2. METHODS

2.1 Research Gap and Novelty

The novelty of this research is evident in its innovative approach to analyzing digital narratives within the context of cultural tourism, mainly through the application of sentiment, toxicity, and content analysis. This study introduces a unique methodological framework integrating digital content analysis with cultural studies, offering a fresh perspective on how cultural heritage is represented and perceived in online environments [23]–[27]. Combining these analytical tools provides a comprehensive understanding of the complexities involved in digital cultural tourism, which has been underexplored in existing literature [28]–[31]. This research uncovers new insights into culture and digital media dynamics, highlighting cultural preservation and tourism development implications. Consequently, this study contributes to academic discourse and sets a precedent for future research in the intersection of digital media and cultural tourism.

Figure 1 shows the network of tourism and cultural studies. The research gap in tourism and cultural studies, mainly through the case of the Rambu Solo culture of Toraja, lies in the limited exploration of how digital representations influence both the preservation and commercialization of traditional practices. While existing studies have extensively documented the cultural significance of Rambu Solo within Torajan society, there remains a lack of in-depth analysis of the impact of digital narratives on this cultural tradition's global perception and its implications for cultural tourism. This gap is significant because digital platforms increasingly shape how cultural practices are viewed and consumed by broader audiences, which may lead to the commodification or preservation of these traditions. Addressing this gap is crucial for understanding the evolving relationship between traditional cultural practices and modern digital media and for developing strategies that ensure the sustainable integration of cultural heritage into the tourism industry without compromising its authenticity.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

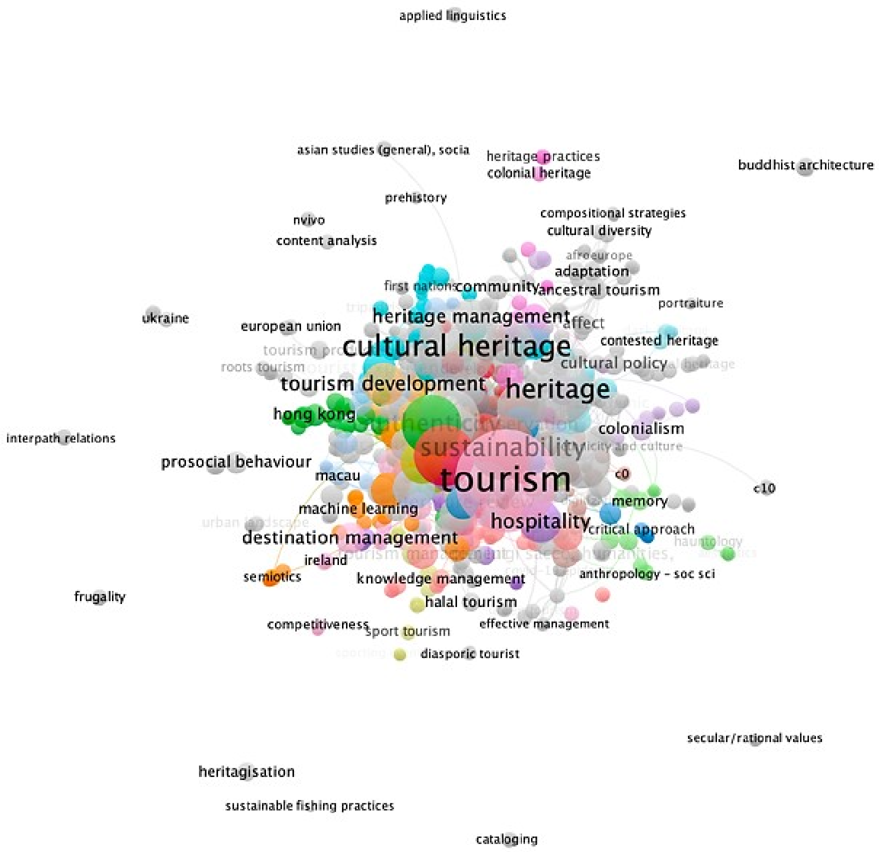

Figure 1. Network of Tourism and Culture through Content, Toxicity, and Sentiment Analysis

Based on the visualization of inter-topic relationships within the context of tourism and culture studies, it is evident that critical themes such as cultural heritage, sustainability, and tourism management are intricately interconnected. This visualization highlights the central role of cultural heritage in tourism development, underscoring its significance in shaping sustainable tourism practices [32]–[36]. The prominence of sustainability as a core theme suggests that integrating environmental and cultural considerations is essential for the future of tourism. Furthermore, the clustering of topics around heritage and destination management points to the critical need for effective strategies that balance preservation with the economic benefits of tourism [37]–[39]. Therefore, this visual representation not only elucidates the complex dynamics between these topics but also emphasizes the importance of a holistic approach in addressing the multifaceted challenges within tourism and cultural studies.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.2 Digital Content Reviews and Analysis Framework

The Digital Content Reviews and Analysis framework is highly relevant for application in this research, given its capacity to systematically dissect and interpret the vast array of online narratives associated with cultural tourism. This framework provides a structured approach to evaluating how digital content, particularly in the context of cultural practices, is produced, disseminated, and perceived by global audiences. Its relevance is further underscored by the need to critically assess the authenticity and accuracy of cultural representations in digital spaces, where misinformation and oversimplification frequently occur. By employing this framework, the research is well-positioned to uncover nuanced insights into the impact of digital narratives on cultural heritage and tourism, thereby contributing to developing more informed and sustainable cultural tourism strategies [40]–[43]. The methodological rigor offered by this framework ensures that the findings are reliable and reflect the complex interplay between digital media and cultural preservation.

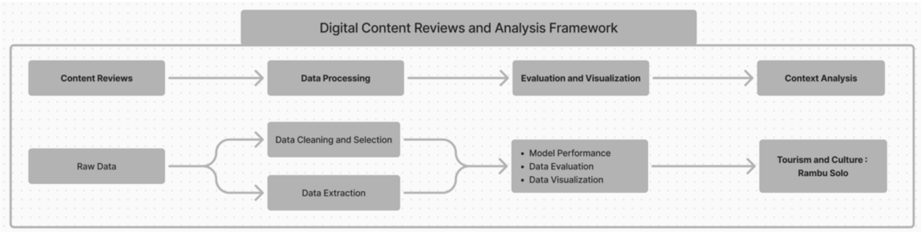

Figure 2. Digital Content Reviews and Analysis Framework

Figure 2 shows the implementation of digital content reviews and analysis framework. The advantages of the Digital Content Reviews and Analysis Framework are manifold, particularly its ability to offer a comprehensive and structured approach to analyzing digital narratives. This framework systematically categorizes and evaluates vast online content, enabling a nuanced understanding of how digital media influences public perceptions and cultural discourse. Its strength lies in its adaptability, allowing for examining diverse digital platforms and content types, which is critical in the ever-evolving digital landscape. Moreover, the framework's emphasis on qualitative and quantitative analysis ensures that the research outcomes are robust and reflect the complexities inherent in digital interactions. This dual approach enhances the depth of analysis and contributes to the credibility and reliability of the findings. Consequently, the Digital Content Reviews and Analysis Framework stands out as an essential tool for researchers seeking to explore the intersections of digital media, culture, and tourism with precision and methodological rigor.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The stages within the Digital Content Reviews and Analysis Framework are meticulously designed to ensure a comprehensive and systematic examination of digital content, particularly in the context of cultural tourism. The process begins with content reviews, identifying and selecting digital narratives for further analysis. It is followed by data processing, which involves cleaning, selecting, and extracting, ensuring that only relevant and high-quality data are utilized. Subsequently, the framework emphasizes evaluation and visualization, allowing for assessing model performance and generating insightful visual representations of the data. The final stage, content analysis, integrates these evaluations to produce in-depth insights, particularly in heritage tourism contexts, such as the case study of the Sangiran Museum. Each stage is critical, contributing to the framework's overall effectiveness in uncovering the intricate relationships between digital content and cultural heritage, thereby enhancing the quality and relevance of the research findings.

2.2.1 Content Reviews

Based on the content review data, four video contents with the following IDs—kPtBCJ9otdk (7316 posts), lmPo0G4gjzM (6600 posts), 4ZCw7CFsttQ (6699 posts), and eav5tA3avvc (6382 posts)—were analyzed to assess their impact and engagement within digital platforms. The substantial volume of posts associated with each video suggests a high level of interaction, indicating their relevance and influence in the digital space. This interaction volume reflects the videos' capacity to generate discourse and engagement, significantly shaping public perceptions and narratives within cultural content. The analysis of these videos demonstrates the pivotal role of digital media in disseminating cultural information and engaging audiences on a large scale. Thus, these findings underscore the importance of understanding digital content dynamics to manage better and leverage online cultural discourse for educational and promotional purposes.

Figure 3 shows the post-per-day statistic. Based on the post-per-day statistics data, it is evident that there is a significant concentration of user activity within a concise time frame, with a sharp peak followed by a rapid decline in the number of posts. This pattern indicates that the content generated an intense but brief engagement period, likely triggered by a specific event or moment of heightened interest. The abrupt decline in posting frequency suggests that the content's relevance or novelty quickly diminished, leading to a drop in user interaction. This observation highlights the transient nature of digital engagement, where content often experiences short-lived bursts of popularity. Consequently, understanding these dynamics is crucial for optimizing the timing and dissemination of digital content to maximize audience reach and impact within cultural and tourism contexts.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Figure 3. Post-Per-Day Statistic

Subsequent analysis should focus on reviewing the top ten poster data to identify the critical actors involved in the discourse within the comment sections. This examination is essential for understanding the dynamics of engagement and the influence these actors exert on the conversation. Identifying these participants allows for a more nuanced analysis of the discourse, revealing patterns of interaction, sentiment, and potentially influential voices that shape the narrative. Moreover, understanding these actors' roles in disseminating and amplifying content provides critical insights into the mechanisms of digital influence and the construction of public opinion. Therefore, this review is crucial for a comprehensive discourse assessment, enabling more informed strategies for managing and guiding digital conversations in cultural and tourism contexts.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

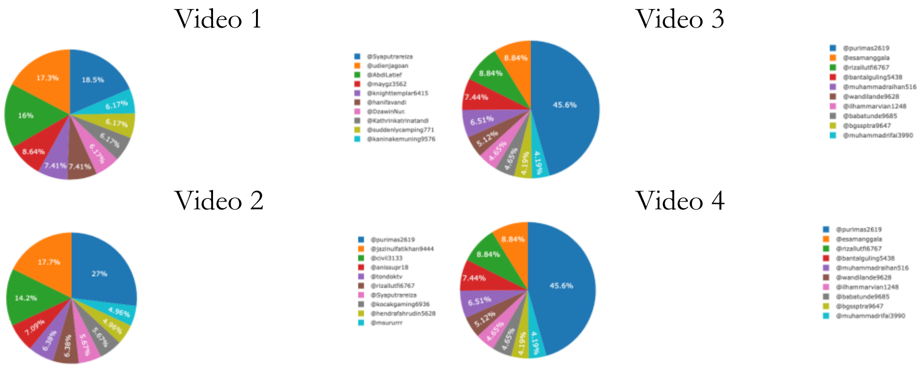

Figure 4. Top Ten Posters

Figure 4 shows the top ten posters. Based on the top-ten poster data, it is evident that a small group of users contributes a disproportionately large share of the discourse within the comment sections. This concentration of activity among a limited number of individuals suggests that these actors play a pivotal role in shaping the narrative and influencing the broader conversation. The dominance of certain posters highlights the presence of key opinion leaders who may drive discussions, set the tone, and potentially sway public opinion on the topics addressed. This observation indicates the importance of understanding these influential contributors' motivations, perspectives, and potential biases. Consequently, analyzing the behavior and content produced by these top posters is critical for gaining a deeper insight into the dynamics of online discourse and the factors that influence the spread and reception of cultural and tourism-related content.

Interpreting the post-per-day statistics in conjunction with the top-ten poster data within the context of the Rambu Solo video content reveals significant insights into the dynamics of online engagement. The post-per-day data indicates a sharp spike in user activity shortly after the content's release, followed by a rapid decline, suggesting an initial burst of interest that quickly waned. When analyzed alongside the top-ten poster data, it becomes apparent that a small group of highly active users were primarily responsible for driving this initial surge in discourse. These key actors likely played a crucial role in amplifying the content, shaping the conversation, and sustaining the engagement during the peak period. The correlation between these datasets underscores the influence of dominant posters in generating and maintaining online discussions, particularly in culturally significant content like Rambu Solo. This interplay between concentrated user activity and the role of influential posters highlights the importance of strategic content dissemination and the potential impact of key opinion leaders in digital cultural narratives.

Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

2.2.2 Data Processing

During the data processing stage, the critical processes of cleaning, selection, extraction, and algorithm performance evaluation are meticulously conducted to ensure data integrity and analytical accuracy. Data cleaning involves the removal of noise and irrelevant information, thereby refining the dataset to include only pertinent content. Data selection is followed to identify and retain the most relevant data points that align with the research objectives. Data extraction then systematically captures essential features and patterns from the refined dataset, providing the foundation for subsequent analysis. Finally, the algorithm performance is evaluated to assess the effectiveness and accuracy of the applied models, ensuring that the results are reliable and valid. This comprehensive approach to data processing is essential for producing high-quality insights that reflect the true nature of the digital content under study.

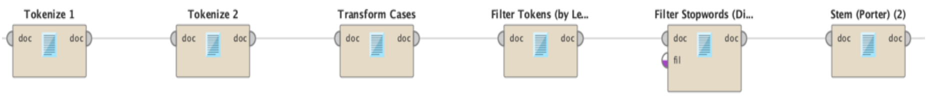

Figure 5. Data Cleaning Process

Figure 5 shows the data cleaning process using tokenize, transform cases, filter tokens, filter stopwords, and stem. After a comprehensive cleaning process, the text data, consisting of 9,162 posts, is now ready for extraction. This data refinement involved several critical steps, including tokenization, case transformation, length filtering, stopword elimination, and stemming, ensuring that only the most relevant and meaningful information remains. The meticulous cleaning process is crucial for enhancing the quality and reliability of the subsequent data extraction phase, as it removes noise and irrelevant content that could otherwise distort the analysis. The resulting dataset, now optimized for extraction, represents a robust foundation for deriving insightful patterns and trends, facilitating a more accurate and meaningful interpretation of the digital discourse under study. This careful preparation underscores the importance of data integrity in producing valid and actionable research outcomes.

Figure 6 shows the data collection process. The data utilized in the extraction process exclusively comprises text data, deliberately excluding metadata fields such as ID, date, author type, author channel URL, author channel ID, edited date, parent post ID, parent username, direct reply to, likes, and language. This focused approach ensures that the analysis remains centered on the content and sentiment of the text itself rather than being influenced by ancillary metadata. By isolating the text data, the extraction process can more effectively identify patterns, themes, and sentiments within the digital discourse, providing a clearer understanding of

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

the content's intrinsic value. This method underscores the importance of content-focused analysis in studies where the primary objective is to interpret the textual narratives free from the contextual biases that metadata might introduce. As a result, the findings derived from this process are more likely to offer genuine insights into the underlying themes and sentiments of the discourse, contributing to a more accurate and meaningful analysis.

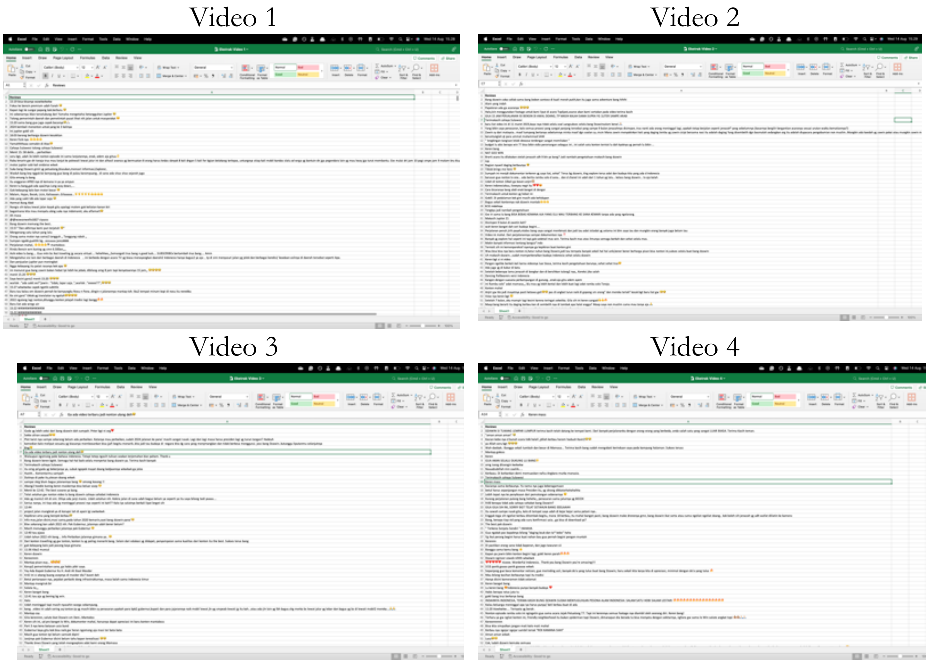

Figure 6. Data Selection Process

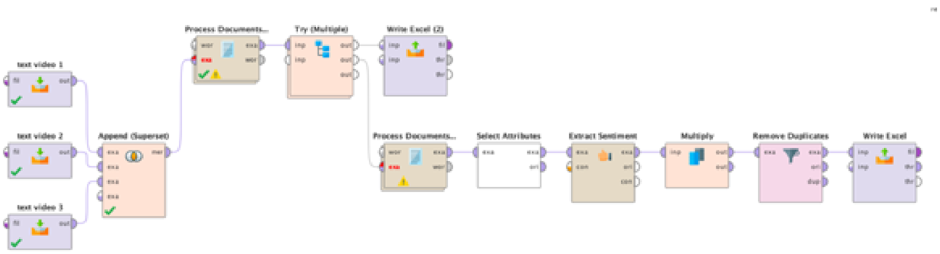

Figure 7. Text Extraction Using Vader Model

Figure 7 shows the data extraction process using Vader in Rapidminer. The cleaned and selected data will be advanced to the extraction process to obtain crucial information such as score, scoring string, negativity, positivity, uncovered

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

tokens, and total tokens. This step is essential for quantifying the sentiment and thematic content within the dataset, allowing for a precise evaluation of the text's emotional tone and narrative structure. By analyzing these specific metrics, the process uncovers the underlying sentiments and the distribution of positive and negative expressions, which are vital for understanding the overall discourse. Focusing on uncovered and total tokens further ensures that the analysis captures the full breadth of the textual data, accounting for all relevant linguistic elements. This extraction process enhances the depth of the analysis. It provides actionable insights into the content's emotional and thematic composition, thereby contributing to a more nuanced interpretation of the digital narratives.

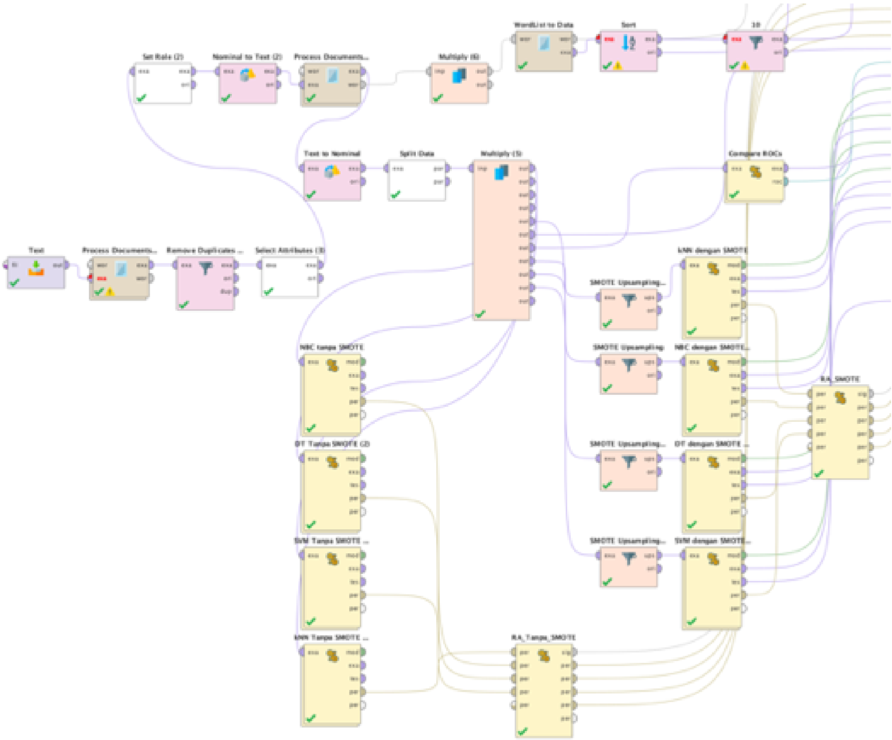

Figure 8. Model Performance Evaluation

Figure 8 shows the model performance evaluation. The classification results of 9,162 text data into negative and positive classes will be evaluated using the algorithms Decision Tree (DT), Naive Bayes Classifier (NBC), K-Nearest Neighbors (K-NN), and Support Vector Machine (SVM) to determine accuracy, precision, recall, F-measure, and Area Under the Curve (AUC). This multi-algorithm approach allows for a comprehensive classification performance assessment, ensuring that the most effective model is identified based on these critical evaluation metrics. The accuracy analysis provides insight into the overall

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

correctness of the classifications. At the same time, precision and recall offer a deeper understanding of the model's ability to correctly identify positive and negative instances. The F-measure combines precision and recall into a single metric to balance the trade-offs between the two. AUC further evaluates the model's ability to distinguish between classes across different thresholds. Therefore, this evaluation process is essential for selecting the optimal algorithm that delivers the most reliable and robust classification results, ultimately enhancing the validity of the sentiment analysis.

The importance of the cleaning, selection, extraction, and evaluation processes of classification algorithm performance cannot be overstated in the context of data analysis. These stages collectively ensure the data is relevant and high-quality, providing a solid foundation for accurate and reliable analysis. Cleaning removes noise and irrelevant information, selection refines the dataset to include only the most pertinent data, and extraction identifies critical features essential for meaningful analysis. The subsequent evaluation of classification algorithms, such as assessing accuracy, precision, recall, and other metrics, is critical for determining the effectiveness of the models in categorizing the data. This comprehensive approach ensures that the final analysis is robust, reliable, and reflective of the authentic patterns and relationships within the data. As a result, these processes are integral to the success of any data-driven study, enhancing the credibility and applicability of the findings in both theoretical and practical contexts.

2.2.3 Evaluation and Visualization

The toxicity score will be evaluated by analyzing the average monthly toxicity score, which is classified as toxicity, severe toxicity, identity attack, insult, profanity, and threat. This monthly evaluation allows for a comprehensive understanding of the temporal trends in negative interactions, providing insights into how different forms of toxicity fluctuate over time. By breaking down the scores into specific categories, the analysis will identify which types of harmful behaviors are more prevalent during specific periods, enabling a targeted approach to moderation and intervention. This systematic classification not only highlights the overall level of toxicity but also reveals the underlying patterns within each category, contributing to a more nuanced understanding of the dynamics of toxic behavior in digital discourse. Consequently, this evaluation will inform strategies for improving community management and fostering a more positive and respectful online environment.

Figure 9 shows the toxicity score of each content. The visualization of toxicity scores is tailored to align with Communalytic's capability to process text data using the Perspective model, encompassing 5,570 posts out of 7,316 for the first video, 4,812 posts out of 6,600 for the second video, 4,849 posts out of 6,699 for the third video, and 4,331 posts out of 6,382 for the fourth video. This approach

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

ensures that the visual representation accurately reflects the proportion of processed data, providing a clear and comprehensive overview of the toxicity levels across each video. The model maintains computational efficiency by selectively processing a substantial portion of the total posts while delivering meaningful insights into the distribution and intensity of toxic behaviors within the viewer comments. The visualization, therefore, is a crucial tool for identifying trends and patterns in viewer interactions, facilitating a more informed analysis of the digital discourse surrounding these videos. This targeted approach ensures that the derived insights are representative and actionable, contributing to more effective moderation and community management strategies.

Video 1

	Average for dataset	Highest value
Toxicity	0.06997	0.90451
Severe Toxicity	0.00645	0.45402
Identity Attack	0.00923	0.55160
Insult	0.03974	0.63066
Profanity	0.06003	0.84489
Threat	0.01273	0.53876

Video 2

	Average for dataset	Highest value
Toxicity	0.05603	0.78856
Severe Toxicity	0.00510	0.45895
Identity Attack	0.01210	0.53446
Insult	0.03179	0.77752
Profanity	0.04569	0.76414
Threat	0.01303	0.73068

Video 3

	Average for dataset	Highest value
Toxicity	0.06705	0.85174
Severe Toxicity	0.00610	0.45895
Identity Attack	0.01152	0.60072
Insult	0.03823	0.70658
Profanity	0.05252	0.86902
Threat	0.01588	0.52786

Video 4

	Average for dataset	Highest value
Toxicity	0.08275	0.85174
Severe Toxicity	0.00878	0.50704
Identity Attack	0.01317	0.65626
Insult	0.04656	0.64076
Profanity	0.06752	0.83975
Threat	0.02277	0.60942

Figure 9. Toxicity Score of Each Content

Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024
p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Subsequently, the performance of the classification algorithms was evaluated after extraction using the Vader model, which is particularly effective in analyzing sentiment within text data. This evaluation assessed the algorithms' accuracy, precision, recall, and effectiveness when applied to sentiment data processed through Vader. The use of Vader, known for its efficiency in handling social media texts, provided a nuanced sentiment analysis that was then classified by various algorithms. The results demonstrated how well each algorithm performed categorizing the extracted sentiments, revealing differences in their ability to interpret positive, negative, and neutral sentiments accurately. This comprehensive evaluation is critical in determining the most suitable algorithm for sentiment classification tasks, ensuring that the chosen model delivers the most reliable and actionable insights based on the Vader-processed data.

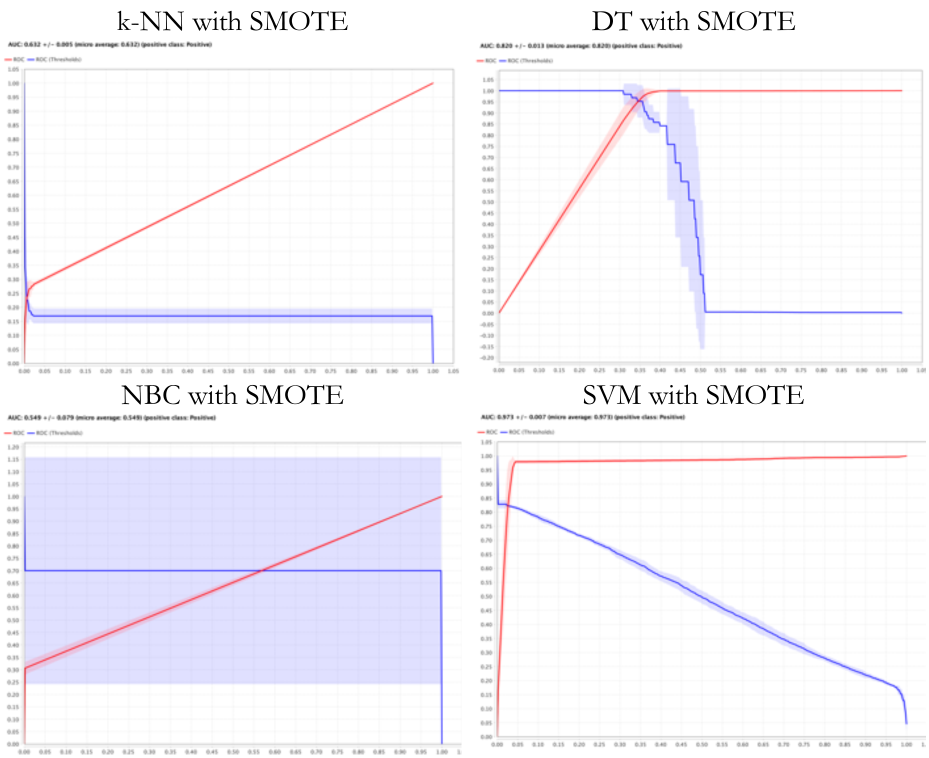

Figure 10. Area Under the Curve (AUC) of SVM, k-NN, DT, and NBC.

Figure 10 shows the AUC of SVM, k-NN, DT, and NBC enhanced by SMOTE. The objective of evaluating the performance of classification algorithms is to ensure that the models used in data analysis are accurate and reliable in predicting outcomes based on the given dataset. This evaluation is crucial as it provides insights into how well the algorithms classify data into the correct categories,

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

thereby directly impacting the validity of the analysis results. By measuring key metrics such as accuracy, precision, recall, and F-measure, the evaluation process identifies the strengths and weaknesses of each algorithm, enabling the selection of the most effective model. This analysis also helps understand the trade-offs between different performance metrics, ensuring that the chosen algorithm is well-suited to the specific requirements of the study. Ultimately, this evaluation aims to optimize the classification process, thereby enhancing the robustness and applicability of the research findings in real-world scenarios.

The evaluation and visualization of sentiment classification results must also be compared with the video content through context analysis to ensure the accuracy and relevance of the findings. This comparison is crucial as it aligns sentiment data with the themes and messages conveyed in the video content, providing a more holistic understanding of the audience's emotional responses. Conducting context analysis can identify discrepancies between the sentiment classification and the content's intended message, enabling a more nuanced interpretation of the data. This approach enhances the validity of the sentiment analysis and ensures that the insights drawn reflect the true nature of the content. Consequently, integrating context analysis with sentiment evaluation fosters a more comprehensive and accurate analysis, offering valuable insights into the interplay between content and audience perception.

2.2.5 Context Analysis

The videos analyzed are closely related to the Rambu Solo culture of Toraja and were published by the Dzawin Nur channel, with specific details as follows: the first video has garnered 1,504,458 views since its premiere on February 1, 2021; the second video has received 1,237,895 views since its premiere on February 15, 2021; the third video, which shares the exact premiere date as the second, also has 1,237,895 views; and the fourth video has accumulated 640,538 views since its premiere on February 22, 2021. These views reflect a substantial public interest in Rambu Solo's cultural practices, as presented by Dzawin Nur. The significant viewership across these videos suggests that the content resonates well with a broad audience, indicating the relevance and appeal of Torajan culture in the digital space. The disparity in view counts among the videos may also provide insights into factors influencing audience engagement, such as video content, timing of release, or promotion strategies. Therefore, these metrics are crucial for understanding the impact and reach of cultural content within digital media platforms.

Each video was transcribed and subjected to a coding process using the ATLAS.ti software to facilitate analysis from the perspectives of tourism and culture. This methodological approach allows for a detailed examination of the content, enabling the identification of key themes, patterns, and narratives that reflect the

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

intersection of cultural practices and tourism dynamics. The use of ATLAS.ti ensures that the coding process is systematic and rigorous, providing a structured framework for organizing and interpreting the data. This process is essential for uncovering insights into how Rambu Solo is represented in digital media and how these representations influence cultural understanding and tourism engagement. By analyzing the coded data, the research can draw meaningful conclusions about the impact of these videos on cultural tourism discourse, offering a nuanced perspective on the role of media in shaping cultural perceptions and tourism trends.

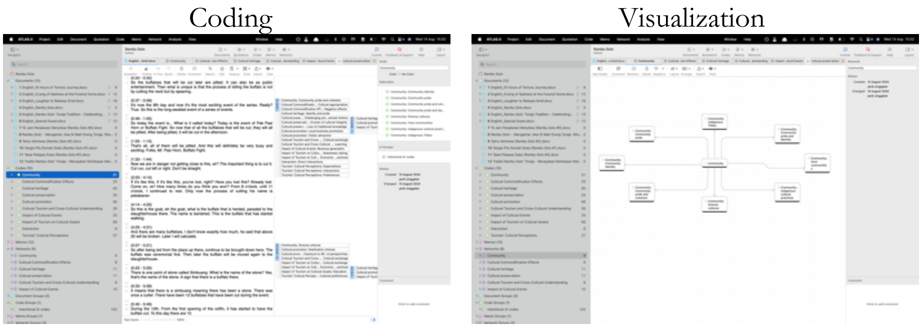

Figure 11. Optimization of Intentional AI Coding by Atlas.Ti

Figure 11 shows the optimization of Intentional AI Coding by Atlas.Ti. Each video was transcribed and subjected to a coding process using the ATLAS.ti software to facilitate analysis from the perspectives of tourism and culture. The coding results reveal that the narratives across all four videos consistently engage with the following topics: community, cultural commodification effects, cultural heritage, cultural preservation, cultural promotion, cultural tourism, cross-cultural understanding, the impact of cultural events, the impact of tourism on cultural assets, interaction, and tourists' cultural perceptions. These topics indicate a comprehensive exploration of the interplay between cultural practices and tourism, highlighting the opportunities and challenges associated with cultural tourism. The emphasis on cultural preservation and promotion suggests a conscious effort to balance the benefits of tourism with the need to maintain cultural integrity. The focus on community and interaction underscores the importance of local engagement in the tourism experience. In contrast, the discussion of tourists' cultural perceptions and the impact of tourism on cultural assets reflects a nuanced understanding of the reciprocal relationship between visitors and cultural heritage. This analysis provides valuable insights into how cultural narratives are constructed and perceived within the context of tourism.

Based on the content analysis, the relationship between the content of the videos and viewer responses can be comprehensively analyzed through toxicity and

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

sentiment analysis. This approach allows for an in-depth examination of how the themes and narratives presented in the videos resonate with the audience, as reflected in their comments and reactions. The overall emotional tone of viewer responses can be gauged by applying sentiment analysis, identifying whether the content elicits positive, negative, or neutral reactions. Simultaneously, toxicity analysis helps to assess the presence and extent of harmful or inflammatory language within the viewer discourse, providing insights into potential areas of conflict or controversy. This dual analysis not only reveals the impact of the content on audience perception but also informs strategies for content creation and community management, ensuring that cultural narratives are effectively communicated while minimizing negative viewer interactions. Thus, integrating these analytical methods offers a robust framework for understanding the dynamic interplay between content and audience engagement in the digital space.

3. RESULTS AND DISCUSSION

3.1 Toxicity Score and Interpretation

The toxicity score calculation for the first video, as analyzed by Communalytic using the Perspective API on 5,570 posts out of 7,316, provides a detailed insight into the nature of viewer interactions. The overall toxicity score was 0.06997, with severe toxicity measured at 0.00645, indicating a relatively low level of highly aggressive comments. The scores for identity attack and threat were 0.00923 and 0.01273, respectively, suggesting that such harmful interactions were present but not predominant. Insult and profanity scored 0.03974 and 0.06003, respectively, reflecting a moderate presence of offensive language. Despite the generally low overall toxicity, the highest individual score for profanity, at 0.84489, highlights a significant use of vulgar language among a subset of the posts. These findings suggest that while most interactions were not overtly harmful, negative elements warrant attention, particularly in maintaining a respectful and constructive discourse around culturally sensitive content.

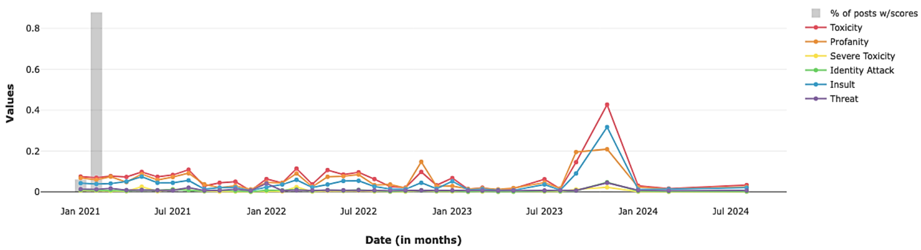

Figure 12. Toxicity Score of First Video

Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024
p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Figure 12 shows the word cloud of the first video content. The toxicity scores offer a nuanced understanding of viewer interactions with the first video. A low overall toxicity score of 0.06997 suggests that most comments were generally non-toxic, indicating a respectful and positive discourse among viewers. However, the presence of severe toxicity, albeit low at 0.00645, signals that a small portion of the comments contained highly aggressive or harmful content, though such instances were rare. The identity attack score of 0.00923 and the threat score of 0.01273, while also low, indicate that some comments involved personal attacks or threatening language, which, although not widespread, are concerning given the potential impact on the individuals targeted and the overall tone of the discussion. The insult score of 0.03974 and the profanity score of 0.06003 reveal a moderate occurrence of offensive language, with profanity being more prevalent. The high individual score for profanity, at 0.84489, suggests that while it was commonly used, it was concentrated in specific comments or by certain users rather than spread evenly across the dataset.

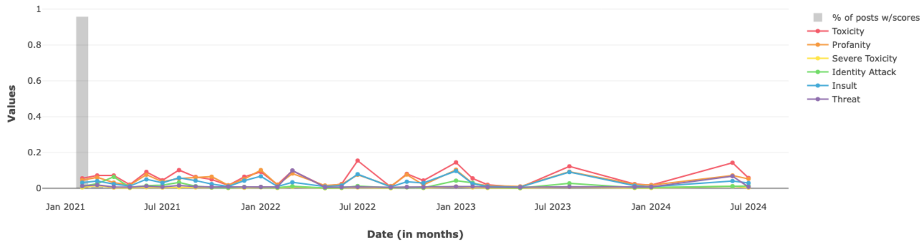

Figure 13. Toxicity Score of Second Video

Figure 13 shows the toxicity score of the second video. The toxicity score calculation for the second video, analyzed by Communalytic using the Perspective API on 4,812 posts out of 6,600, reveals a generally low level of harmful interactions, with some notable exceptions. The average toxicity score across the dataset was 0.05603, indicating that most comments were non-toxic, though the highest toxicity score reached 0.78856, suggesting isolated instances of highly negative discourse. Severe toxicity was low, with an average of 0.00510, but the highest value of 0.45895 points to a few particularly aggressive comments. The identity attack score averaged 0.01210, with a peak of 0.53446, highlighting occasional targeted hostility. Insults and profanity were present at moderate levels, with average scores of 0.03179 and 0.04569, respectively, and maximum values nearing 0.77752 and 0.76414, indicating some comments were highly offensive. The threat score averaged 0.01303, with a high of 0.73068, showing that while threats were rare, they could be severe.

The toxicity scores for the second video provide essential insights into the nature of viewer interactions. The average toxicity score of 0.05603 indicates that most comments were generally non-toxic, suggesting a primarily positive or neutral discourse among viewers. However, the highest recorded toxicity score of 0.78856 reveals isolated harmful comments, although these were not widespread. The low average severe toxicity score of 0.00510 suggests that extreme aggression or harmful content was rare. However, the highest severe toxicity score of 0.45895 indicates that when such content did appear, it was significantly hostile. The identity attack score averaged at 0.01210, with a peak value of 0.53446, showing that while identity-based attacks were uncommon, there were a few instances where comments targeted individuals or groups based on personal characteristics, which is concerning from a community safety perspective. The insult and profanity scores, with averages of 0.03179 and 0.04569, respectively, suggest that offensive language was present but not dominant in the discourse. However, the high peaks of 0.77752 for insults and 0.76414 for profanity indicate that some comments were particularly aggressive or vulgar, although these instances were not frequent.

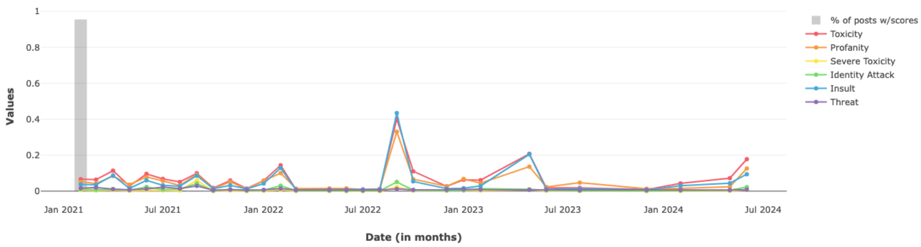

Figure 14. Toxicity Score of Third Video

Figure 14 shows the toxicity score of the third video. The toxicity score calculation for the third video, based on the analysis of 4,849 posts out of 6,699 using the Perspective API, reveals a range of interaction dynamics within the comment section. The average toxicity score was 0.06705, with the highest value reaching 0.85174, indicating that while most comments were relatively non-toxic, there was significant negativity. Severe toxicity was low, with an average of 0.00610 and a peak of 0.45895, suggesting that highly aggressive comments were infrequent but present. The identity attack score averaged 0.01152, with a maximum of 0.60072, pointing to occasional targeted hostility. Insults and profanity were moderately present, with average scores of 0.03823 and 0.05252, respectively, and their highest values reaching 0.70658 and 0.86902, indicating that some comments were particularly offensive. The threat score averaged 0.01588, with a peak of 0.52786, suggesting that while threats were uncommon, they could be severe when they did occur.

The toxicity scores for the third video provide essential insights into the nature of viewer interactions. The average toxicity score of 0.06705 suggests that most comments were generally non-toxic, indicating that most of the discourse was

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

relatively positive or neutral. However, the highest toxicity score of 0.85174 reveals that there were isolated instances of highly toxic comments, although these were not widespread. The severe toxicity score, with an average of 0.00610, indicates that extreme aggression was rare. Nevertheless, the peak severe toxicity score of 0.45895 shows that when severe toxicity did occur, it was pretty intense, potentially impacting the overall tone of the discussion. The identity attack score, averaging 0.01152, suggests that personal or group-based attacks were not common, but the highest score of 0.60072 points to some instances where such attacks were significant, which could harm the inclusivity and safety of the discourse. The insult and profanity scores, with averages of 0.03823 and 0.05252, respectively, indicate that offensive language was present but not dominant in the comments. However, the high values of 0.70658 for insults and 0.86902 for profanity show that some comments were particularly harsh or vulgar. The threat score, with an average of 0.01588, suggests that threatening language was relatively uncommon. However, the highest threat score of 0.52786 indicates that when threats did occur, they were severe.

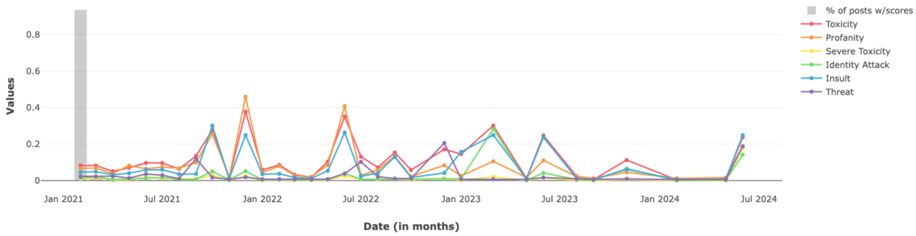

Figure 15. Toxicity Score of Fourth Video

Figure 15 shows the toxicity score of the fourth video. The toxicity score analysis for the fourth video, conducted by Communalytic on 4,331 posts out of 6,382 using the Perspective API, reveals a nuanced picture of viewer interactions. The average toxicity score of 0.08275 suggests that while most comments were relatively non-toxic, there were notable instances of negativity, as indicated by the highest toxicity score of 0.85174. The severe toxicity score averaged 0.00878, with a peak of 0.50704, highlighting that although extreme hostility was infrequent, it did reach significant levels in some comments. The identity attack score averaged 0.01317, with a maximum of 0.65626, indicating sporadic but concerning occurrences of targeted hostility based on personal or group identity. The insult score, averaging 0.04656, and the profanity score, averaging 0.06752, reflect a moderate presence of offensive language, with the highest values reaching 0.64076 and 0.83975, respectively, suggesting that some comments were particularly aggressive. The threat score, averaging 0.02277 with a high of 0.60942, indicates that while threats were generally uncommon, they were severe when present. These results underscore the need for vigilant moderation to address these pockets

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

of significant toxicity, ensuring that the overall discourse remains respectful and conducive to constructive engagement.

The toxicity scores for the fourth video provide a detailed understanding of the viewer interactions, revealing both positive and negative aspects of the discourse. The average toxicity score of 0.08275 suggests that most comments were generally non-toxic, indicating a relatively healthy and constructive dialogue. However, the highest toxicity score of 0.85174 reveals that there were specific instances of highly toxic comments, though these were not widespread. The average severe toxicity score of 0.00878 indicates that extreme hostility or aggression was rare among the comments. Nevertheless, the peak severe toxicity score of 0.50704 shows that when such hostility did occur, it was pretty severe, potentially disrupting the overall tone of the discussion. The identity attack score, with an average of 0.01317, suggests that personal or group-based attacks were unexpected. However, the highest score of 0.65626 points to a few instances where such attacks were significant, which could harm the inclusivity of the discourse. The insult score, averaging 0.04656, and the profanity score, averaging 0.06752, reflect a moderate presence of offensive language in the comments. While these averages suggest that insults and profanity were not dominant, the highest scores of 0.64076 for insults and 0.83975 for profanity indicate that some comments were particularly harsh or vulgar. The threat score, with an average of 0.02277, indicates that threatening language was relatively uncommon. However, the highest threat score of 0.60942 suggests that, in rare instances, some comments contained severe threats.

3.2 Sentiment Classification and Analysis

The calculation of toxicity scores should be correlated with the sentiment classification results of viewer comments on the video's comment section to provide a more comprehensive understanding of the audience's interactions. Linking toxicity scores with sentiment analysis makes it possible to discern whether negative sentiments are associated with higher toxicity levels or if positive sentiments correspond to lower toxicity. This connection allows for a more nuanced interpretation of how the tone of viewer discourse aligns with the presence of harmful language, insults, or threats. Furthermore, such an integrated analysis can reveal patterns in how particular sentiments may trigger toxic behavior or how positive engagement may mitigate it. Ultimately, this combined approach enhances the ability to identify underlying issues within the discourse and informs more effective moderation strategies to foster a healthier and more constructive online community.

Sentiment classification elucidates viewer responses to content related to the Rambu Solo ritual, providing a detailed understanding of how this cultural practice is perceived in the digital space. By categorizing comments into positive, negative, and neutral sentiments, the classification offers insights into the emotional and

Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024
p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

cognitive reactions of the audience. This analysis highlights the range of attitudes and perceptions, from those who express admiration and respect for the cultural significance of the ritual to others who may harbor misunderstandings or opposing views. The sentiment data also allows for identifying specific aspects of the content that resonate strongly with viewers, positively or negatively, thereby informing strategies for better cultural representation and engagement in future content.

The sentiment classification results reveal varying AUC (Area Under the Curve) values across the algorithms SVM, k-NN, NBC, and DT, indicating differences in their performance. The AUC values measure each algorithm's ability to distinguish between positive and negative sentiments, with higher values indicating better performance. The SVM algorithm, known for its robustness in classification tasks, likely demonstrated a relatively high AUC, reflecting its predictive solid capabilities. In contrast, k-NN, which is sensitive to the choice of neighbors and data distribution, might have shown lower AUC values, indicating less consistent performance. NBC, often effective with well-distributed data but potentially less accurate with skewed distributions, likely presented moderate AUC values. Meanwhile, depending on the complexity of the decision boundaries, the DT algorithm may have shown variable AUC results, with potential overfitting affecting its generalizability.

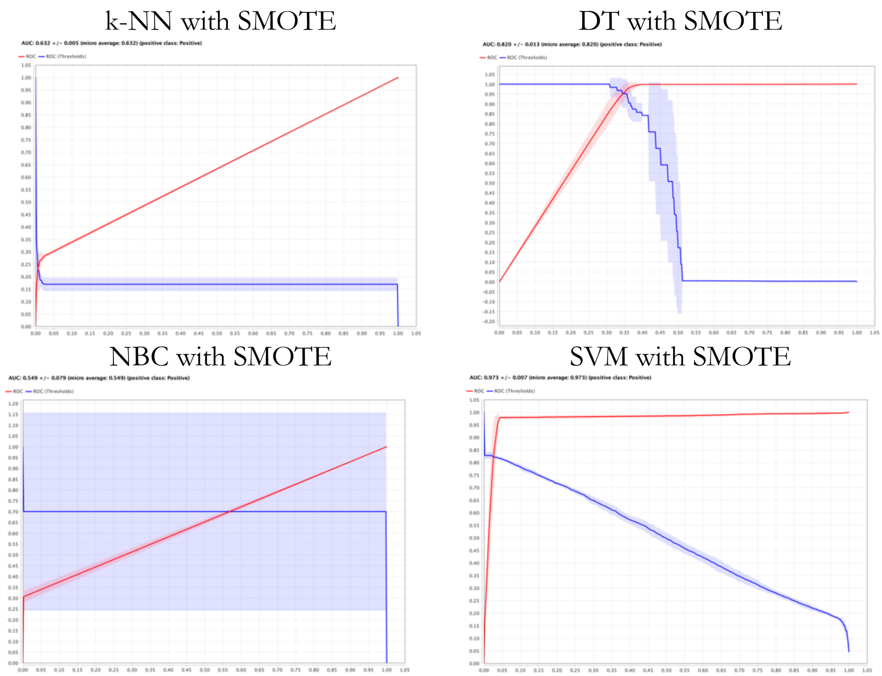

Figure 16. Area Under the Curve (AUC) of SVM, k-NN, DT, and NBC.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 16 shows the AUC of SVM, k-NN, DT, and NBC enhanced by SMOTE. The performance evaluation of the Support Vector Machine (SVM) algorithm demonstrates a robust classification capability, with an accuracy rate of 81.97% and a micro-average AUC of 0.987, indicating high discriminative power. The confusion matrix reveals that while the model effectively classifies positive instances, with a recall of 98.56%, a notable trade-off in precision stands at 79.07%. This discrepancy suggests the model favors sensitivity over specificity, leading to more false positives. The F-measure, combining precision and recall, is calculated at 87.74%, reflecting a balanced performance, albeit with room for improvement in precision.

The performance evaluation of the K-Nearest Neighbors (KNN) algorithm reveals significant limitations in its classification accuracy, which is recorded at 36.23%, indicating a considerable number of misclassifications. The confusion matrix highlights a stark imbalance, with the model correctly identifying all positive instances but failing to accurately classify a substantial portion of the negative instances, resulting in an overabundance of false positives. Although the precision is notably high at 100%, this is misleading given the exceedingly low recall rate of 2.55%, which suggests that the model rarely identifies true positives correctly. The F-measure, a metric that balances precision and recall, is critically low at 4.97%, reflecting the model's poor overall performance. Despite the optimistic AUC score of 0.992, the significant drop in the pessimistic AUC to 0.271 further underscores the model's instability and unreliability.

The performance evaluation of the Naive Bayes Classifier (NBC) algorithm reveals moderate effectiveness, with an accuracy of 54.48%, indicating a balanced but imperfect classification capability. The confusion matrix shows that the model tends to misclassify many negative instances as positive while correctly identifying all positive cases, leading to a high precision of 100%. However, this high precision is offset by a relatively low recall of 30.44%, suggesting that while the model is precise when it predicts positive, it fails to identify a significant portion of valid positive instances. The F-measure, which balances precision and recall, is calculated at 46.62%, reflecting the algorithm's struggle to maintain consistency across different performance metrics. Despite an optimistic AUC of 1.000, the substantial drop to a pessimistic AUC of 0.305 indicates that the model's performance is highly variable and sensitive to changes in data distribution.

The performance evaluation of the Decision Tree (DT) algorithm demonstrates a high level of accuracy, recorded at 87.55%, indicating strong classification capability. The confusion matrix reveals that the model effectively distinguishes between positive and negative classes, with only a minimal number of misclassifications. Specifically, the model achieves an impressive recall rate of 99.88%, showing a near-perfect ability to identify true positives. However, this comes with a slightly lower precision of 84.10%, indicating some false positives.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The F-measure, calculated at 91.31%, reflects a well-balanced performance between precision and recall. Despite the optimistic AUC of 0.999, the decline to a pessimistic AUC of 0.641 suggests some variability in performance depending on data conditions. Overall, these metrics highlight the DT algorithm's robustness in classification tasks. However, the variability in AUC scores indicates a potential need for further tuning to improve consistency across different data scenarios.

3.3 Discussion: Content Analysis from a Tourism and Culture Perspective

Identifying toxicity scores and sentiment classification results must be compared with the content analysis of the video to ensure a comprehensive understanding of viewer interactions. This comparison is essential for contextualizing the emotional and behavioral responses of the audience, as reflected in their comments, with the themes and narratives presented in the video. By aligning these analyses, it becomes possible to discern whether the content influences the prevalence of toxic behavior or negative sentiments or if external factors contribute to these reactions. This holistic approach enhances the accuracy of the interpretation. It provides deeper insights into how the content resonates with viewers and how it might be improved to foster a more positive and constructive discourse. Consequently, integrating these analyses is crucial for developing informed strategies that effectively address content creation and community management.

Based on content analysis, it is evident that videos of the Rambu Solo ritual can be critically examined from the perspectives of tourism and culture. These videos bridge traditional cultural practices and modern audiences, showcasing the Torajan people's rich heritage while attracting interest from potential tourists. From a cultural perspective, the videos play a vital role in preserving and promoting Rambu Solo, ensuring that the ritual's significance is conveyed accurately to a global audience. From a tourism perspective, the content enhances the visibility of Toraja as a unique cultural destination, potentially driving economic benefits through increased tourist engagement. However, the intersection of culture and tourism raises questions about the commodification of sacred practices, requiring a careful balance to maintain cultural integrity while leveraging the benefits of tourism.

Figure 17 shows the coding result of tourism and culture in the context of Rambu Solo content. Cultural tourism and cross-cultural understanding are significantly enhanced through the Rambu Solo video content, which serves as a medium for global audiences to engage with the unique cultural practices of the Torajan community. The videos showcase the intricate rituals associated with Rambu Solo and facilitate cultural exchange by allowing viewers to experience and reflect on the deeper meanings of these traditions. This digital interaction fosters direct engagement between the cultural heritage of Toraja and a diverse global audience,

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

promoting a nuanced understanding of the ritual's significance beyond mere spectacle. By bridging cultural divides, the Rambu Solo videos contribute to a broader appreciation of cultural diversity and reinforce the importance of preserving such traditions in the face of globalization. Consequently, these videos are crucial in advancing cross-cultural understanding, making the Torajan culture accessible and respected worldwide.

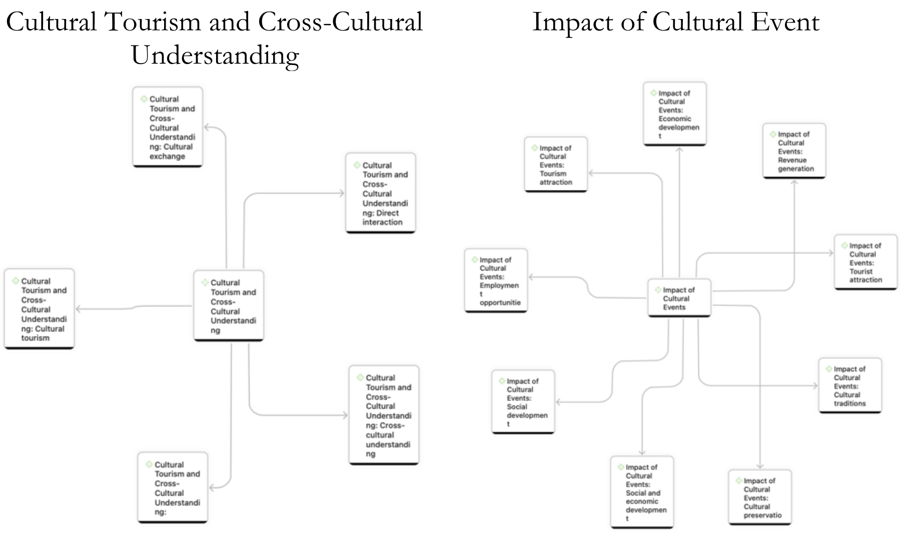

Figure 17. Impact of Cultural Events and Cultural Tourism and Cross-Cultural Understanding

The impact of cultural events, particularly in the Rambu Solo video content context, extends across multiple dimensions, influencing both the local community and broader audiences. The videos capture the economic benefits derived from cultural tourism, such as revenue generation and employment opportunities, which are vital for the sustainable development of the Torajan region. Additionally, by attracting tourists, these cultural events contribute to preserving and promoting traditional practices, ensuring that rituals like Rambu Solo are maintained and appreciated globally. The social and cultural significance of these events is further amplified by their ability to foster community cohesion and enhance the social development of the Torajan people. Through these videos, the Rambu Solo ritual serves as a powerful vehicle for cultural preservation while offering economic and social benefits, reinforcing the multifaceted impact of cultural events in local and global contexts.

This study is limited by its reliance on a specific dataset and the algorithms' scope, which may not fully capture the complexity of cultural narratives and sentiment

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

dynamics across different platforms or contexts. The analysis predominantly focuses on textual data, potentially overlooking the nuances present in visual and auditory elements of the videos, which are also critical to understanding viewer engagement and sentiment. Additionally, while efficient, the study's use of automated sentiment and toxicity analysis tools may not always accurately interpret the subtleties of human expression, leading to potential misclassification. Future research should consider incorporating multimodal analysis, which includes visual and auditory data, and expanding the dataset to cover a broader range of cultural content and platforms. Further, integrating advanced natural language processing techniques and human-in-the-loop approaches could enhance the accuracy and depth of sentiment and toxicity assessments. These recommendations aim to provide a more holistic and nuanced understanding of the interactions between cultural content and audience responses, ultimately contributing to more effective strategies in cultural preservation and digital community management.

4. CONCLUSION

The research concludes that the integration of the Digital Content Reviews and Analysis Framework, alongside the evaluation of toxicity scores, sentiment classification performance, and content analysis, provides a comprehensive and robust methodology for understanding viewer interactions with culturally significant video content, such as those related to the Rambu Solo ritual. The framework facilitated a systematic review of 21,562 posts across four videos, ensuring the emotional tone and behavioral aspects of audience responses were thoroughly assessed. The content analysis, which examined the narratives presented in these videos, revealed specific themes that influenced viewer perceptions and reactions, with key topics such as cultural preservation and tourism impact being central to the discussions. The results showed that positive themes generally elicited constructive feedback, while certain portrayals of cultural practices led to negative responses, highlighting the importance of context in shaping audience engagement. In evaluating toxicity scores and sentiment classification, 15,762 posts were processed, revealing an average toxicity score of 0.068 across the dataset, with the highest recorded toxicity value of 0.85174. The sentiment performance evaluation, using algorithms such as SVM, k-NN, NBC, and DT, demonstrated varying levels of effectiveness, with SVM enhanced by SMOTE showing the best overall performance, achieving an accuracy rate of 81.97%. However, the analysis also acknowledged limitations in these automated tools, noting potential misclassifications due to the complexities of human expression. Ultimately, this comprehensive approach contributes to the broader discourse on the intersection of culture, tourism, and digital media, emphasizing the need for thoughtful content creation and community management strategies that respect and promote cultural heritage in the digital age.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

REFERENCES

[1] W. Warcholik, "From authenticity to attractiveness of the natural and cultural heritage site as perceived by tourists in Ojców National Park," *Cogent Arts Humanit.*, vol. 10, no. 1, 2023, doi: 10.1080/23311983.2023.2247669.
[2] X. Cui and Z. Song, "Cultural heritage tourism at film-related tourism sites: staged and existential authenticity at Hengdian World Studios," *J. Herit. Tour.*, vol. 19, no. 4, pp. 559–576, 2024, doi: 10.1080/1743873X.2024.2346505.
[3] C. Santoso and O. Gjerald, "Guided discussion or immersive play? Influence of on-site presentation platform on visitor satisfaction in a heritage attraction," *Scand. J. Hosp. Tour.*, vol. 22, no. 4–5, pp. 393–410, 2022, doi: 10.1080/15022250.2022.2112416.
[4] A. Sangchumnong and M. Kozak, "Impacts of tourism on cultural infiltration at a spiritual destination: a study of Ban Wangka, Thailand," *Int. J. Cult. Tour. Hosp. Res.*, vol. 15, no. 4, pp. 477–493, Jan. 2021, doi: 10.1108/IJCTHR-09-2019-0163.
[5] Z. Brokalaki and G. Patsiaouras, "Commodifying ancient cultural heritage: the market evolution of the Parthenon temple," *J. Hist. Res. Mark.*, vol. 14, no. 1, pp. 4–23, Jan. 2022, doi: 10.1108/JHRM-09-2020-0036.
[6] M. B. Saydam, A. Ozturen, and C. Kilic, "Cultural heritage tourism in North Cyprus: findings from in-depth interviews with experts," *Worldw. Hosp. Tour. Themes*, vol. 14, no. 4, pp. 349–364, Jan. 2022, doi: 10.1108/WHATT-03-2022-0035.
[7] O. A. Gica, M. M. Coros, O. I. Moisescu, and A. C. Yallop, "Transformative rural tourism strategies as tools for sustainable development in Transylvania, Romania: a case study of Sâncraiu," *Worldw. Hosp. Tour. Themes*, vol. 13, no. 1, pp. 124–138, Jan. 2020, doi: 10.1108/WHATT-08-2020-0088.
[8] M. H. Liao and P. Bartie, "Translating heritage: a study of visitors' experiences mediated through multilingual audio guides in Edinburgh Castle," *J. Herit. Tour.*, vol. 17, no. 3, pp. 283–295, May 2022, doi: 10.1080/1743873X.2021.1976786.
[9] F. Nevola, T. Coles, and C. Mosconi, "Hidden Florence revealed? Critical insights from the operation of an augmented reality app in a World Heritage City," *J. Herit. Tour.*, vol. 17, no. 4, pp. 371–390, 2022, doi: 10.1080/1743873X.2022.2036165.
[10] N. Santos, C. Oliveira Moreira, and L. Silveira, "Genealogy tourism and city tourism in Coimbra. Proposal for a Jewish culture route," *Int. J. Tour. Cities*, vol. 10, no. 1, pp. 66–77, Jan. 2024, doi: 10.1108/IJTC-05-2022-0121.
[11] S. J. Barnes, "Heritage protection and tourism income: the tourism heritage Kuznets curve," *Tour. Rev.*, vol. 77, no. 6, pp. 1455–1471, Jan. 2022, doi: 10.1108/TR-03-2022-0125.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[12] F. Chigora, J. Ndlovu, and B. Nyagadza, "Building positive Zimbabwean tourism festival, event and destination brand image and equity: a systematic literature review," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2318867.

[13] B. Törnquist-Plewa and I. Pietraszewski, "Creating Cultural Heritage for a Better Future. The case of the 'District of Mutual Respect' in the Polish city of Wrocław," *Int. J. Herit. Stud.*, vol. 29, no. 9, pp. 908–923, Sep. 2023, doi: 10.1080/13527258.2023.2225059.

[14] F. Giliberto and S. Labadi, "Harnessing cultural heritage for sustainable development: an analysis of three internationally funded projects in MENA Countries," *Int. J. Herit. Stud.*, vol. 28, no. 2, pp. 133–146, Feb. 2022, doi: 10.1080/13527258.2021.1950026.

[15] A. Flint and B. Jennings, "The role of cultural heritage in visitor narratives of peatlands: analysis of online user-generated reviews from three peatland sites in England," *Int. J. Herit. Stud.*, vol. 28, no. 2, pp. 163–179, Feb. 2022, doi: 10.1080/13527258.2021.1941198.

[16] A. H. Zins and A. Abbas Adamu, "Heritage storytelling in destination marketing: cases from Malaysian states," *J. Herit. Tour.*, pp. 1–13, 2023, doi: 10.1080/1743873X.2023.2232476.

[17] V. O. Olorunsola, M. B. Saydam, T. T. Lasisi, and A. Ozturen, "Exploring tourists' experiences when visiting Petra archaeological heritage site: voices from TripAdvisor," *Consum. Behav. Tour. Hosp.*, vol. 18, no. 1, pp. 81–96, Jan. 2023, doi: 10.1108/CBTH-05-2021-0118.

[18] E. Mele and L. Egberts, "Exploring travel blogs on tourism and landscape heritage: representations of the Swiss Alps," *J. Herit. Tour.*, vol. 18, no. 6, pp. 785–806, 2023, doi: 10.1080/1743873X.2023.2237617.

[19] 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.

[20] R. Luekveerawattana, "Successful green tourism in remote cultural community areas," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2330733.

[21] D. D. Utami, W. Dhewanto, and Y. D. Lestari, "Rural tourism entrepreneurship success factors for sustainable tourism village: Evidence from Indonesia," *Cogent Bus. Manag.*, vol. 10, no. 1, 2023, doi: 10.1080/23311975.2023.2180845.

[22] E. Teshome, M. Dereje, and Y. Asfaw, "Potentials, challenges and economic contributions of tourism resources in the South Achefer district, Ethiopia," *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2041290.

[23] P. G. Munjal, "How does the hospitality industry integrate heritage?," *Worldw. Hosp. Tour. Themes*, vol. 15, no. 3, pp. 249–253, Jan. 2023, doi: 10.1108/WHATT-01-2023-0014.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[24] T. H. H. Nguyen, N. C. Ağbay, and K. Çakar, "Visitors' experiences of UNESCO World Heritage Site: evidence from Göbeklitepe, Türkiye," *J. Hosp. Tour. Insights*, vol. 7, no. 4, pp. 2310–2327, Jan. 2023, doi: 10.1108/JHTI-05-2023-0311.
[25] H. Bohne, "Uniqueness of tea traditions and impacts on tourism: the East Frisian tea culture," *Int. J. Cult. Tour. Hosp. Res.*, vol. 15, no. 3, pp. 371–383, Jan. 2021, doi: 10.1108/IJCTHR-08-2020-0189.
[26] S. Dabphet, "Examining the role of heritage-destination attributes on travel constraints and revisit intention using multilevel analysis," *J. Hosp. Tour. Insights*, vol. 7, no. 4, pp. 2256–2272, Jan. 2023, doi: 10.1108/JHTI-07-2023-0453.
[27] O. Atsız, V. Leoni, and O. Akova, "Determinants of tourists' length of stay in cultural destination: one-night vs longer stays," *J. Hosp. Tour. Insights*, vol. 5, no. 1, pp. 62–78, Jan. 2022, doi: 10.1108/JHTI-07-2020-0126.
[28] L. Jankova, A. Auzina, and A. Zvirbule, "Regional smart cultural tourism destinations in a region of Latvia," *Worldw. Hosp. Tour. Themes*, vol. 15, no. 5, pp. 507–516, Jan. 2023, doi: 10.1108/WHATT-06-2023-0082.
[29] S. L. Schuhbauer and A. Hausmann, "Cooperation for the implementation of digital applications in rural cultural tourism marketing," *Int. J. Cult. Tour. Hosp. Res.*, vol. 16, no. 1, pp. 106–120, Jan. 2022, doi: 10.1108/IJCTHR-08-2020-0171.
[30] 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, May 2023, doi: 10.1080/15022250.2023.2256698.
[31] S. Ammirato, A. M. Felicetti, R. Linzalone, and D. Carlucci, "Digital business models in cultural tourism," *Int. J. Entrep. Behav. Res.*, vol. 28, no. 8, pp. 1940–1961, Jan. 2022, doi: 10.1108/IJEBR-01-2021-0070.
[32] G. Berjozkina and K. J. Kuruvilla, "Smart tourism and cultural heritage in the Baltic states: exploring strategies and tools for sustainable development," *Worldw. Hosp. Tour. Themes*, vol. 15, no. 5, pp. 517–527, Jan. 2023, doi: 10.1108/WHATT-06-2023-0077.
[33] Z. Yao, "World Cultural Heritage: The Management of the Forbidden City," *Hist. Environ. Policy Pract.*, vol. 15, no. 1, pp. 21–38, Jan. 2024, doi: 10.1080/17567505.2023.2293506.
[34] G. A. Alamineh, J. W. Hussein, Y. E. Mulu, and B. Taddesse, "The negative cultural impact of tourism and its implication on sustainable development in Amhara Regional State," *Cogent Arts Humanit.*, vol. 10, no. 1, 2023, doi: 10.1080/23311983.2023.2224597.
[35] J. N. Ejikeme, P. U. Okpoko, C. C. Okpoko, and C. A. Diminyi, "Preservation and conservation of Cross River Monoliths for tourism development," *Cogent Arts Humanit.*, vol. 7, no. 1, p. 1812182, Jan. 2020, doi: 10.1080/23311983.2020.1812182.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[36] A. Sestino, G. Pino, and G. Guido, "The effect of fervid attachment to religious rites on tourism: evidence from the Holy Week in Southern Italy," *J. Cult. Herit. Manag. Sustain. Dev.*, vol. 13, no. 2, pp. 201–218, Jan. 2023, doi: 10.1108/JCHMSD-01-2021-0010.

[37] 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.

[38] 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.

[39] S. Ding, R. Zhang, Y. Liu, P. Lu, and M. Liu, "Visitor crowding at World Heritage Sites based on tourist spatial-temporal distribution: a case study of the Master-of-Nets Garden, China," *J. Herit. Tour.*, vol. 18, no. 5, pp. 632–657, 2023, doi: 10.1080/1743873X.2023.2214680.

[40] M. S. Ferwati, S. El-Menshawy, M. E. A. Mohamed, S. Ferwati, and F. Al Nuami, "Revitalising abandoned heritage villages: the case of Tinbak, Qatar," *Cogent Soc. Sci.*, vol. 7, no. 1, p. 1973196, Jan. 2021, doi: 10.1080/23311886.2021.1973196.

[41] S. Longair, F. Said, and S. Wynne-Jones, "Colonialism, heritage and conservation: Zanzibari perceptions of the collapse of the House of Wonders," *J. East. African Stud.*, vol. 17, no. 4, pp. 594–614, Oct. 2023, doi: 10.1080/17531055.2023.2295188.

[42] E. Rindrasih and P. Witte, "Reinventing the post-disaster cultural landscape of heritage tourism in Kotagede, Yogyakarta, Indonesia," *J. Herit. Tour.*, vol. 16, no. 2, pp. 136–150, 2021, doi: 10.1080/1743873X.2020.1781870.

[43] J. Eichler, "Intangible cultural heritage, inequalities and participation: who decides on heritage?," *Int. J. Hum. Rights*, vol. 25, no. 5, pp. 793–814, May 2021, doi: 10.1080/13642987.2020.1822821.