Journal of Information Systems and Informatics
Vol. 6, No. 2, June 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i3.834 Published By DRPM-UBD

Evaluating Digital Narratives in Heritage Tourism and Museum: Content Analysis, Toxicity Score, and Sentiment Classification Trough SVM and SMOTE

Gasper Tabuni[1], Yerik Afrianto Singgalen[2*]

[1] Information System Department, Faculty of Computer Science, Baliem Papua University, Papua Pegunungan, Indonesia
[2*] Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: [1]gaspertabuni@unibapapua.ac.id, [2*]yerik.afrianto@atmajaya.ac.id

Abstract

This research uses the Digital Content Reviews and Analysis Framework to explore the dynamic interplay between digital content, sentiment, and toxicity within the context of heritage tourism at the Sangiran site. The study is driven by the urgency to understand how digital narratives impact public engagement and perception, particularly for heritage sites of global significance. Through a comprehensive analysis, the research evaluates toxicity scores, sentiment classifications, and thematic content across multiple videos related to Sangiran. The toxicity analysis reveals generally low levels of harmful content, with an average score of 0.04717, but identifies occasional peaks, highlighting the potential for negative discourse. Sentiment analysis, conducted using the SVM model enhanced by SMOTE, achieves an accuracy rate of 94.59%, with precision and recall rates of 92.07% and 97.79%, respectively, demonstrating the model's robustness in capturing audience sentiment. Content analysis identifies critical themes, such as human evolution and fossil discoveries, emphasizing the educational value of digital content. The research underscores the importance of curating digital narratives that engage, educate, and foster a positive and respectful discourse. The findings suggest that while digital content successfully educates the audience, managing contentious topics is crucial to maintaining constructive engagement. This study contributes to developing more effective digital strategies for heritage tourism, ensuring the sustainable promotion and preservation of sites like Sangiran while addressing the challenges of online discourse. The research highlights the need for continued exploration of digital content's role in shaping public perceptions of cultural heritage.

Keywords: Heritage Tourism; Digital Content Analysis; Sentiment; Toxicity Analysis; Sangiran Site

1. INTRODUCTION

Heritage tourism, encompassing visits to museums and historical sites, is a critical pillar of educational tourism, requiring strategic efforts to ensure its sustainability.

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

By preserving cultural heritage through well-maintained museums, these sites offer profound educational value, fostering a deeper understanding of history and culture [1], [2]. The sustainability of heritage tourism is intrinsically linked to its ability to adapt to evolving societal needs while maintaining the integrity of the cultural narratives it preserves [3], [4]. Without sustained conservation and adaptive strategies, the educational potential of these sites risks erosion, diminishing their role as vital resources for learning. Therefore, prioritizing the long-term viability of heritage tourism is essential for continuously enriching educational experiences across generations.

This research aims to identify and analyze public sentiment toward Sangiran tourism by examining narratives or text data generated from the video content, then calculating toxicity scores and classifying the sentiment. The methodology integrates advanced sentiment analysis models such as Perspective and Vader, which offer robust tools for understanding the nuanced emotional tone embedded in public discourse. Applying these models provides a comprehensive approach to quantifying sentiment and detecting harmful content, ensuring a balanced evaluation of public opinion [5], [6]. The resulting sentiment classification highlights the prevailing attitudes toward Sangiran and contributes to broader discussions on the impact of digital narratives in shaping tourism perceptions. Thus, the study underscores the importance of employing sophisticated analytical frameworks to navigate the complexities of sentiment analysis in contemporary tourism research.

The urgency of this research is driven by the growing need to understand the impact of digital narratives on public engagement and perception in heritage tourism, particularly as digital platforms increasingly shape cultural discourse. As heritage sites like Sangiran play a crucial role in preserving and promoting human history, ensuring that the digital content associated with these sites effectively communicates their significance while fostering positive and constructive interactions is imperative. The rapid evolution of digital technologies and their pervasive influence on audience behavior necessitate timely and comprehensive studies that address the complexities of online engagement, sentiment, and toxicity [7], [8]. This research responds to these contemporary challenges and provides critical insights that can inform the development of more effective digital strategies in heritage tourism [9], [10]. By examining how digital content influences public perceptions and interactions, the study contributes to safeguarding heritage sites' educational and cultural value in the digital age, ensuring their continued relevance and impact.

This research's theoretical and practical contributions are significant, offering advancements in academic understanding and practical application of digital content in heritage tourism. Theoretically, the study enhances existing knowledge by integrating sentiment analysis, toxicity evaluation, and content analysis within

the context of digital heritage narratives, providing a comprehensive framework for assessing the impact of digital engagement on public perceptions. This integration not only deepens the understanding of how digital content influences audience behavior but also offers a new lens through which the effectiveness of heritage communication can be evaluated. Practically, the research contributes actionable insights that heritage site managers and digital strategists can employ to optimize content delivery, ensuring that it educates and engages audiences while mitigating negative interactions. The findings enable the development of more refined digital strategies that balance the dual objectives of cultural preservation and public engagement, thus supporting the sustainable promotion of heritage sites in an increasingly digital world. These contributions underscore the research's value in bridging the gap between theory and practice, offering tools and insights that can be directly applied to enhance the management and promotion of cultural heritage in the digital age.

Similar research within this domain provides a valuable foundation for understanding the intricacies of the subject matter while highlighting the areas that require further exploration. Analyzing prior studies allows one to discern proven effective patterns and methodologies and identify limitations that necessitate new approaches [11]–[13]. The existing body of work serves as a critical reference point, enabling the refinement of hypotheses and the adaptation of strategies to address gaps in the literature [14]–[18]. This comparative analysis reinforces the relevance of continued inquiry and underscores the necessity of evolving research paradigms to keep pace with emerging trends and challenges. Therefore, examining analogous studies contributes to a deeper, more nuanced appreciation of the field and fosters the ongoing advancement of knowledge.

While inherent to any scholarly endeavor, the limitations of this research are crucial to acknowledge as they frame the scope and applicability of the findings. Constraints related to data availability, methodological choices, or contextual factors may influence the generalizability and precision of the results. Recognizing these limitations is essential in understanding the boundaries within which the conclusions hold validity. This awareness encourages a cautious interpretation of the findings and highlights areas where future studies might build upon or refine the current work. Thus, addressing these limitations enhances the research's integrity and guides subsequent efforts to extend its contributions within the field.

2. METHODS

2.1 Research Gap and Novelty

The novelty of this research is exemplified by its comprehensive examination of content, toxicity, and sentiment analysis within the specific framework of heritage tourism and museums in Sangiran. This innovative focus introduces a new

dimension to understanding how digital narratives surrounding culturally significant sites shape public perceptions and emotional responses. By employing advanced analytical techniques, this study bridges the gap between traditional heritage tourism studies and contemporary digital analysis, offering fresh insights into the intersection of cultural preservation and modern communication [19]–[22]. The implications of these findings extend beyond theoretical contributions, as they provide practical guidance for managing public discourse and enhancing the visitor experience in heritage sites [23]–[26]. Therefore, the unique integration of these elements establishes the research as a significant advancement in the field, with both scholarly and practical relevance.

An analysis of previous research on heritage tourism and museums in Sangiran reveals a significant gap, particularly in the application of content, toxicity, and sentiment analysis perspectives, which have been largely underexplored. This observation highlights an opportunity for the current study to make a meaningful contribution to the existing body of knowledge by addressing these overlooked dimensions. Integrating these analytical approaches fills a critical void in the literature and provides a more nuanced understanding of public perceptions and interactions with heritage sites. By advancing the discourse through this novel lens, the research offers valuable insights that could influence future studies and practical applications in heritage tourism. Consequently, this work significantly enriches this domain's theoretical and practical frameworks.

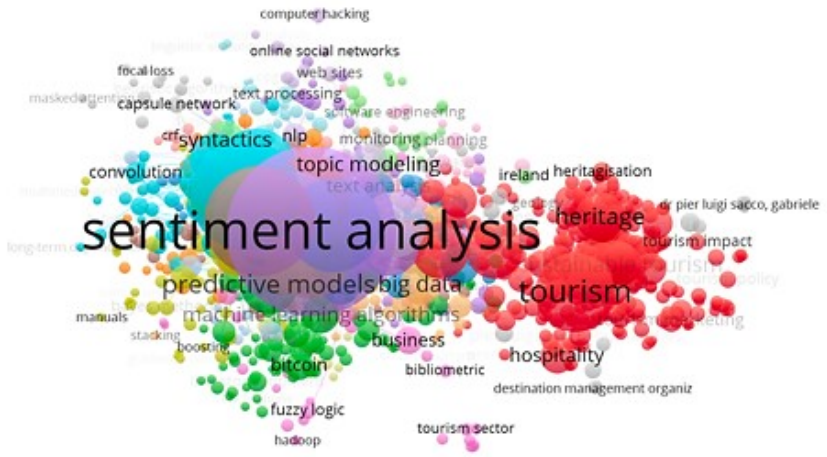

Figure 1. Network of Heritage Tourism and Museum through Content, Toxicity, and Sentiment Analysis

Figure 1 shows the network of heritage tourism and museums through content, toxicity, and sentiment analysis. This observation underscores the critical

importance of researching heritage tourism and museums through the lens of content, toxicity, and sentiment analysis. Integrating these methodologies provides a comprehensive framework for understanding the complex dynamics of public perception and digital engagement with cultural heritage sites [27]. By addressing these factors, the research offers an opportunity to capture nuanced insights into how heritage sites are perceived and discussed in contemporary discourse, which is essential for sustainable management. Moreover, applying such analytical techniques enhances the ability to detect and mitigate negative sentiments that could impact the reputation and preservation of these cultural assets. Therefore, this approach is vital for advancing academic inquiry and practical strategies in heritage tourism.

This research adopts the Digital Content Reviews and Analysis Framework, a methodological approach designed to evaluate and interpret digital narratives systematically. By utilizing this framework, the study ensures a structured and rigorous examination of online content, which is essential for capturing the complexities of digital discourse. The framework's emphasis on categorization, sentiment analysis, and toxicity assessment provides a robust foundation for generating insightful and actionable findings. The strategic implementation of this approach not only enhances the precision of the analysis but also facilitates the identification of underlying trends and patterns within the data. Consequently, adopting this framework contributes to the research's methodological rigor and overall impact, offering a well-rounded perspective on digital content related to heritage tourism.

2.2 Digital Content Reviews and Analysis Framework

The Digital Content Reviews and Analysis Framework is highly relevant to the research on Evaluating Digital Narratives in Heritage Tourism and Museum: Sentiment, Toxicity, and Content Analysis, as it provides a systematic approach to dissecting the multifaceted layers of digital engagement. This framework facilitates the structured examination of digital content by guiding the process from content categorization through sentiment and toxicity analysis, ensuring that each aspect of audience interaction is comprehensively evaluated. The framework's emphasis on systematic reviews and in-depth analysis aligns seamlessly with the research objectives, enabling a thorough understanding of how digital narratives impact audience perception and engagement within the context of heritage tourism. By employing this framework, the research is able to integrate diverse analytical dimensions, such as sentiment analysis using SVM and toxicity scoring, into a cohesive evaluation, thereby enhancing the precision and relevance of the findings. This alignment underscores the framework's critical role in advancing the study's insights into the effectiveness of digital content in promoting educational and cultural heritage while managing the challenges of online discourse.

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

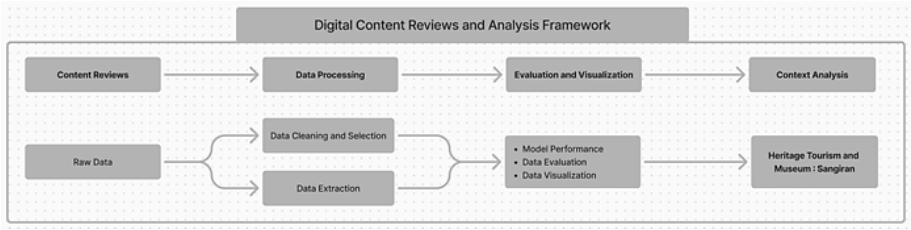

Figure 2. Digital Content Reviews and Analysis Framework

Figure 2 shows the implementation of digital content reviews and analysis framework. The relevance of the Digital Content Reviews and Analysis Framework in this research is evident in its alignment with the research flow and its emphasis on the contextual analysis of video content related to tourism in Sangiran. The framework's structured approach allows for systematically examining digital content, ensuring the analysis is methodologically sound and contextually appropriate. By focusing on the specific narratives and themes present in the videos, the framework enables a deeper understanding of how Sangiran's heritage is represented and perceived in the digital space. This alignment enhances the ability of the research to produce meaningful insights that are directly applicable to the context of heritage tourism. Consequently, the framework's relevance is demonstrated through its capacity to guide the research in producing accurate and contextually informed findings.

Considering the framework's relevance to this research's objectives, a comprehensive contextual analysis has been adopted to compare toxicity scores with public sentiment and examine the coherence of content with viewer perceptions. The framework's structured approach allows for an in-depth exploration of the alignment between the emotional tone of digital narratives and the audience's responses. By juxtaposing toxicity scores with public sentiment, the analysis reveals the potential impact of negative discourse on viewer engagement and perception. Furthermore, assessing content coherence concerning viewer perceptions ensures that the findings reflect the quantitative aspects of sentiment and the qualitative dimensions of audience interpretation. This methodological choice significantly enhances the research's understanding of how the public receives and interprets digital content related to Sangiran's heritage tourism.

2.2.1 Content Reviews

Three videos have been selected for detailed processing and analysis, each offering a unique dataset for examination. The first video, identified by the ID 3nsPHqGpkMQ, comprises 648 posts, providing substantial toxicity and sentiment analysis content. The second video, with the ID 3wDIljrdTCg, contains 395 posts, offering a comparative dataset that allows for the exploration of varying

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

levels of engagement and public sentiment. The third video, ID 54pBTOWnYLg, includes 148 posts, presenting an opportunity to analyze content with potentially lower viewer interaction. By systematically analyzing these videos, the study aims to uncover patterns in public discourse, assess the emotional tone across different levels of engagement, and understand how these factors contribute to the overall perception of heritage tourism in Sangiran. This focused approach ensures that each video is not only individually analyzed but also compared to enhance the depth and relevance of the findings.

Figure 3. Post-Per-Day Statistic

Figure 3 shows the post-per-day statistic. Based on the post-per-day statistics for the first video with ID 3nsPHqGpkMQ, comprising 648 posts, a detailed analysis reveals significant fluctuations in engagement over time. Notably, there are periods of heightened activity, such as on August 26, 2023, and September 4, 2023, where the number of posts surged to 16 and 11, respectively, indicating peak public interest or reaction during these dates. Post spikes may correspond to specific events, updates, or external factors influencing viewer engagement. Conversely, there are numerous instances of minimal activity, with single posts recorded on various dates, suggesting moments of low interaction or attention. This pattern

suggests a non-linear engagement trend, where public interest varies considerably, possibly driven by external stimuli or content relevance. Consequently, these findings underscore the necessity for a nuanced understanding of audience behavior, emphasizing the importance of timing and contextual factors in content dissemination and public engagement strategies.

The post-per-day statistics for the video with ID 3wDIljrdTCg, which contains 395 posts, exhibit notable fluctuations in audience engagement over time. Significant peaks were observed on December 4, 2019, and December 22, 2019, where the number of posts reached 15 and 16, respectively, indicating heightened viewer interaction. These spikes suggest that specific content or external factors captured increased public attention, driving more frequent discourse during these periods. Conversely, there were numerous instances of low activity, with only one post recorded on various dates, highlighting periods of diminished engagement. This uneven distribution of posts reflects the variability in public interest and may indicate how the content's relevance or external events influenced viewer reactions. Understanding these patterns is crucial for developing strategies that maximize engagement by aligning content delivery with periods of higher public interest. This analysis reinforces the importance of timing and content relevance in managing digital engagement effectively.

The post-per-day statistics for the video with ID 54pBTOWnYLg, comprising 148 posts, reveal a diverse pattern of viewer engagement with distinct periods of increased activity. Significant spikes occurred on September 11, 2023, and September 16, 2023, with 7 and 6 posts, respectively, indicating moments of heightened public interest. These bursts of engagement suggest that specific content elements or external factors during these dates may have resonated strongly with viewers, prompting more frequent interaction. Conversely, many dates show only minimal activity, with a single post reflecting period of lower engagement. This irregular distribution highlights the importance of timing and the relevance of content in maintaining consistent viewer interaction. By analyzing these patterns, the study provides insights into the dynamics of public response to heritage tourism content, emphasizing the need for strategic content delivery to maximize engagement. This analysis contributes to a deeper understanding of audience behavior in digital platforms, particularly within the context of heritage tourism.

In addition to analyzing post-per-day statistics, identifying the top ten posters or significant contributors in the comment sections of each video is essential for a comprehensive understanding of audience engagement. These key contributors often drive the discourse, shaping the narrative and influencing the overall sentiment within the comment thread. By examining the frequency and content of their contributions, it becomes possible to discern influence patterns and identify the most active voices within the discussion. This analysis enriches the

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

understanding of the dynamics at play and provides valuable insights into how specific individuals or groups may impact public perception. Consequently, recognizing the top contributors is critical in developing a holistic view of interaction patterns and sentiment in digital spaces, particularly in the context of heritage tourism content.

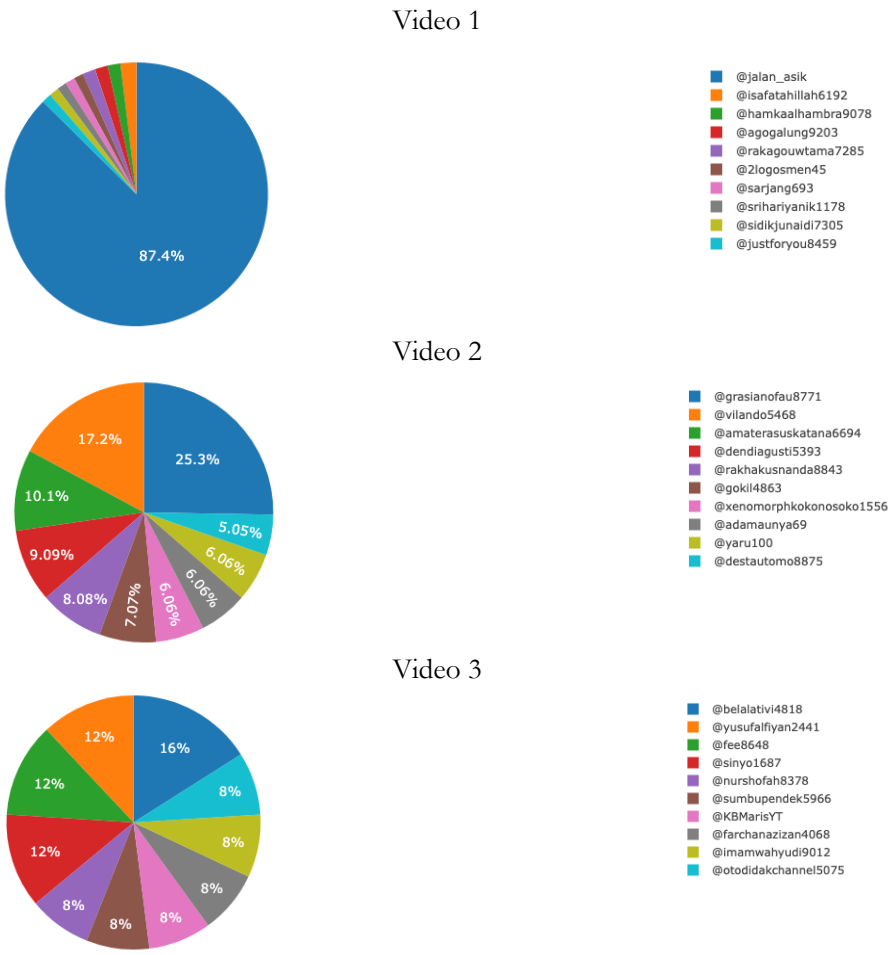

Figure 4. Top Ten Posters

Figure 4 shows the top ten posters. Identifying the top ten posters for the first video reveals a significant disparity in engagement levels among contributors. Notably, @jalan_asik is the most active participant with 222 posts, far surpassing other contributors, indicating a dominant presence in the discourse. The subsequent most active users, including @isafatahillah6192 with five posts and several others with four or fewer posts, demonstrate a sharp drop in activity,

suggesting that the conversation is heavily influenced by a small number of highly engaged individuals. This concentration of contributions in the hands of a few users may shape the overall tone and direction of the discussion, potentially skewing the sentiment and focus of the commentary. Analyzing this distribution of engagement is crucial for understanding the conversation dynamics and the potential impact of these key contributors on the perception of the content.

Identifying the top ten posters for the second video reveals a more distributed pattern of engagement among contributors, with @grasianofau8771 leading with 25 posts, followed by @vilando5468 with 17 posts. Other users, such as @amaterasuskatana6694 and @dendiagusti5393, contribute 10 and 9 posts, respectively, indicating broader participation across the user base compared to the first video. This distribution suggests a more diversified conversation, where multiple voices contribute to the discourse, potentially leading to a richer and more varied exchange of perspectives. However, the presence of dominant contributors, like @grasianofau8771, still indicates that specific individuals exert significant influence over the discussion. This variation in engagement levels across different users is crucial for understanding how discourse is shaped and how particular viewpoints may dominate the narrative.

The identification of the top ten posters for the third video highlights a relatively balanced level of engagement among contributors, with @belalativi4818 leading at four posts, followed by several users such as @yusufalfiyan2441, @fee8648, and @sinyo1687, each contributing three posts. The remaining top contributors, including @nurshofah8378 and others, have made two posts each, indicating a more evenly distributed participation in the discussion. This balanced contribution pattern suggests a less concentrated influence from any user, potentially leading to a more diverse range of perspectives within the comment section. The lack of a dominant voice implies that the discourse may represent a broader audience, providing a more comprehensive view of public sentiment. This even distribution of engagement is critical for understanding the dynamics of the conversation and assessing how various viewpoints interact to shape the overall narrative.

2.2.2 Data Processing

During the data processing stage, a meticulous procedure involving cleaning, selecting, and extracting comment data is conducted using RapidMiner. This process begins with data cleaning, where irrelevant, duplicate, or noisy data is removed to ensure the integrity and accuracy of the dataset. Following this, a careful selection of pertinent data is performed, focusing on comments that are most relevant to the research objectives. The extraction phase then isolates critical information from the selected data, transforming raw comments into structured insights suitable for further analysis. This systematic approach ensures that the data is both refined and representative, enhancing the reliability of subsequent

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

analyses and the overall validity of the research findings. RapidMiner, in this context, provides a robust and efficient framework for managing large datasets, facilitating a more nuanced understanding of the underlying sentiment and content trends.

Figure 5. Data Cleaning Process

Figure 5 shows the data cleaning process using tokenize, transform cases, filter tokens, filter stopwords, and stem. The data cleaning process employs advanced techniques, including tokenization, case transformation, token filtering, stopword removal, and stemming, to refine and prepare the dataset for analysis. Initially, tokenization breaks down text into individual words or phrases, facilitating a granular examination of the content. It is followed by case transformation, standardizing the text by converting all characters to a uniform case, thereby eliminating inconsistencies. The subsequent filtering of tokens removes irrelevant or redundant elements, streamlining the dataset to focus on meaningful content. Stopword removal further enhances this process by excluding common but insignificant words that do not contribute to the analysis. Finally, stemming reduces words to their base forms, consolidating variations of the same word and enabling a more coherent analysis. Together, these steps ensure the data is meticulously processed, resulting in a cleaner, more focused dataset that enhances the accuracy and relevance of the subsequent analytical outcomes.

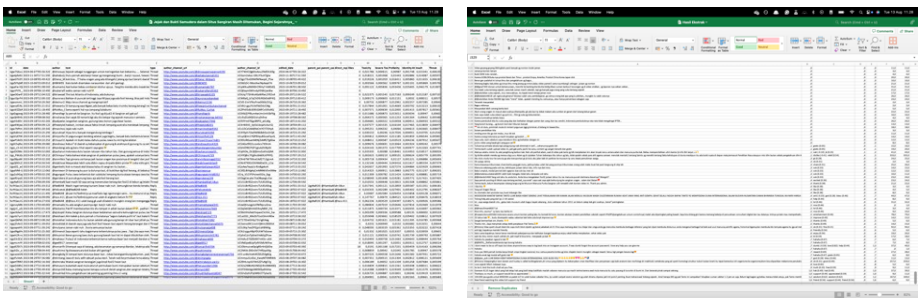

Figure 6. Data Selection Process

Figure 6 shows the data collection process. Following the data cleaning process, the textual data to be extracted represents a combined dataset from the first, second, and third videos, totaling 1,189 posts. This aggregation allows for a comprehensive analysis by merging individual datasets into a cohesive corpus. Integrating posts from multiple sources enhances the dataset's robustness, providing a broader spectrum of public sentiment and discourse for analysis. By

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

consolidating these posts, the study gains the ability to identify overarching trends and patterns that may not be apparent within isolated datasets. This approach not only enriches the depth of the analysis but also ensures that the findings are more representative of the overall public engagement with the content.

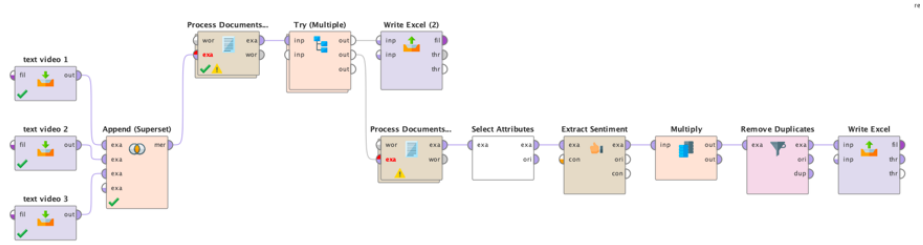

Figure 7. Text Extraction Using Vader Model

Figure 7 shows the data extraction process using Vader in Rapidminer. The extraction of 1,189 textual data points was conducted using the Vader model, which classified the data into positive and negative sentiment categories. This initial classification provides a foundational understanding of the sentiment distribution within the dataset, allowing for a structured analysis of public opinion. The next phase involves evaluating these classifications through vector performance analysis, which will assess the accuracy and effectiveness of the sentiment categorization. By applying vector performance metrics, the analysis ensures that the sentiment classification is precise and reliable, facilitating a deeper exploration of the underlying emotional tone within the data. This rigorous evaluation process enhances the credibility of the findings, providing a robust framework for interpreting the sentiment dynamics in the context of the studied content.

Figure 8 shows the model performance evaluation. The classification results of the 1,189 textual data points, derived using the Vader model, were subjected to testing to obtain vector performance metrics. This testing phase is critical as it quantitatively evaluates the accuracy and efficiency of the sentiment classification, ensuring that the model's outputs align with the expected sentiment categories. The vector performance assessment provides insights into the model's capability to distinguish between varying emotional tones, thereby validating the reliability of the sentiment analysis. Through this rigorous evaluation, the analysis gains a deeper level of precision, facilitating more accurate interpretations and conclusions based on the classified data. This step is essential for reinforcing the robustness of the overall research findings, particularly in sentiment analysis applied to heritage tourism content.

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

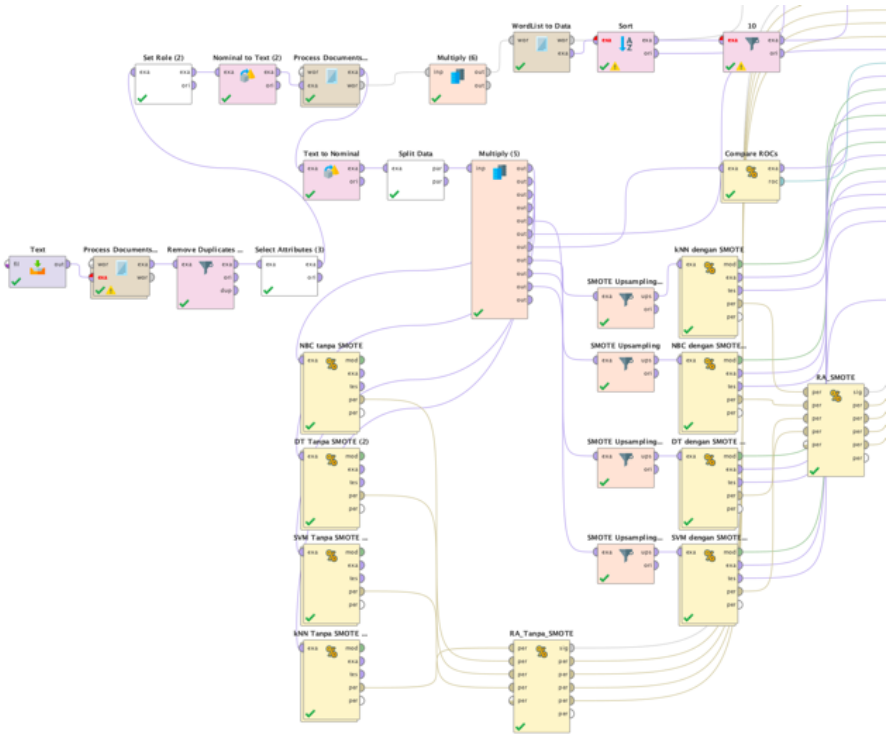

Figure 8. Model Performance Evaluation

2.2.3 Evaluation and Visualization

In this phase, the classification results obtained using the Vader model and the performance testing of the DT, NBC, k-NN, and SVM algorithms enhanced by SMOTE were compared to identify the algorithm with the best performance. This comparative analysis is crucial as it evaluates each algorithm's effectiveness in accurately classifying sentiment, considering various performance metrics such as accuracy, precision, recall, and F1-score. By employing SMOTE to address class imbalances, the robustness of these algorithms is further enhanced, ensuring that the evaluation is comprehensive and reliable. The algorithm demonstrating superior performance will be selected for further analysis, providing the most accurate and efficient model for sentiment classification. This selection process is pivotal in advancing the research's objectives, as it ensures that the most influential computational approach is utilized, thereby enhancing the validity and applicability of the research findings.

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

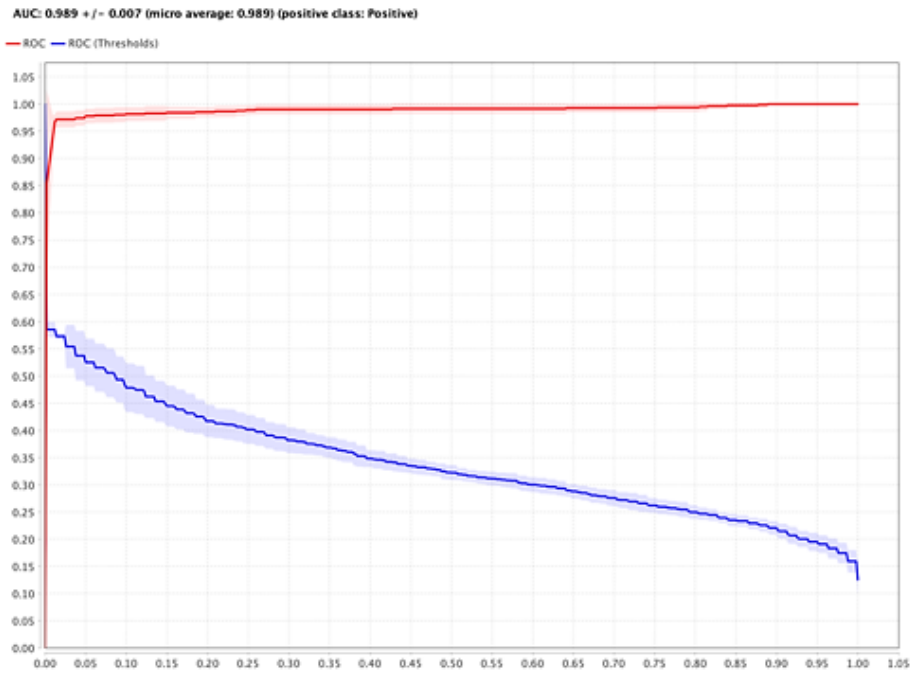

Figure 9. Area Under Curve (AUC) of SVM enhanced by SMOTE

Figure 9 shows the AUC of SVM enhanced by SMOTE. The performance evaluation of the SVM algorithm enhanced by SMOTE indicates a high level of accuracy, achieving 94.59% with a margin of ±2.32%, demonstrating its effectiveness in sentiment classification. The Confusion Matrix reveals a solid ability to correctly identify negative and positive classes, with 743 true negatives and 795 true positives, alongside minimal misclassifications. The AUC values, consistently close to 0.990, further underscore the model's exceptional discriminatory power. Precision, recorded at 92.07% with a slight variance of ±3.91%, alongside a recall rate of 97.79% ±1.51%, highlights the model's balanced performance in minimizing false positives and maximizing true positives. The F-measure, averaging 94.80% with minimal deviation, consolidates these metrics into a comprehensive performance score, affirming the SVM's robustness and reliability. These results demonstrate the algorithm's superior capability in handling the dataset effectively, making it a strong candidate for accurate sentiment analysis in this study.

Journal of Information Systems and Informatics
Vol. 6, No. 2, June 2024
p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

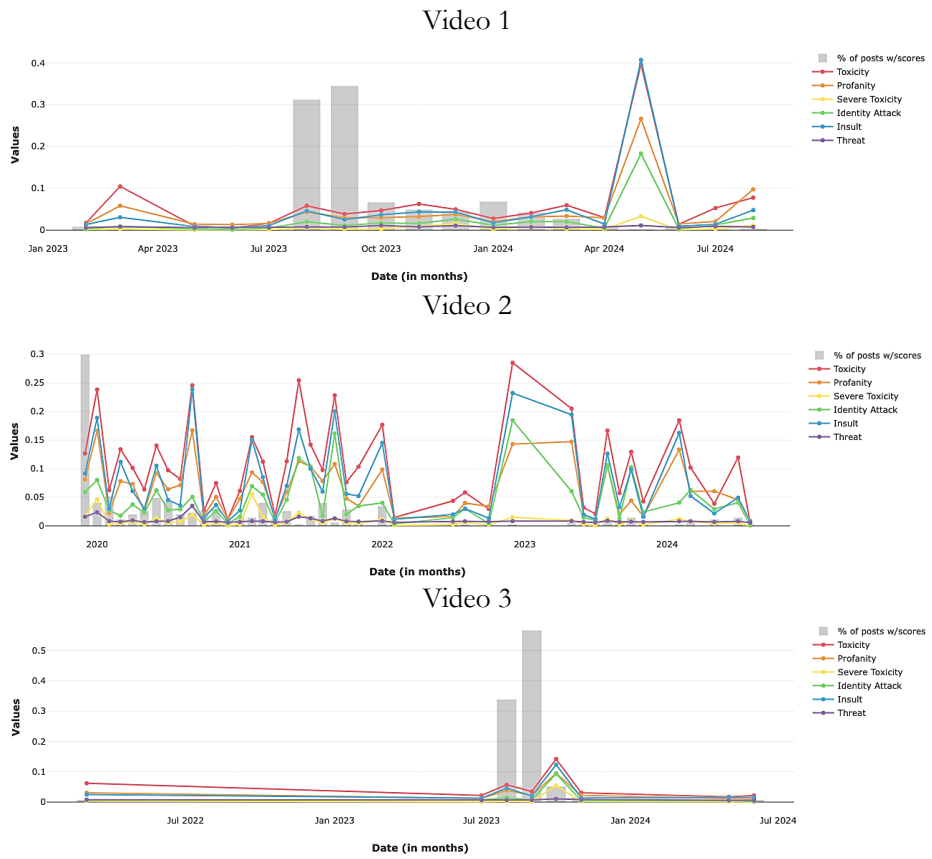

Figure 10. Average Toxicity Core per Month

Figure 10 shows the average toxicity score per month. The toxicity score analysis for the first video, based on 586 out of 648 posts analyzed using the Perspective API by Communalytic, reveals a generally low level of harmful content with specific instances of elevated negativity. The average toxicity score across the dataset is 0.04717, with the highest recorded value reaching 0.71894, indicating occasional peaks in toxic discourse. Severe toxicity remains minimal, averaging 0.00476 with a maximum value of 0.34966, while identity attacks are also relatively low, with an average score of 0.01602 and a peak of 0.48916. Insults show a slightly higher presence, averaging 0.03500 and reaching a maximum of 0.63084, reflecting moments of increased hostility. Profanity and threats are present but infrequent, with average scores of 0.03435 and 0.00803, respectively, and their highest values at 0.40014 and 0.18914. These findings suggest that while the overall discourse is mainly non-toxic, there are notable instances where the conversation intensifies, potentially impacting the overall tone and engagement within the comment

section. This analysis is critical for understanding the dynamics of online interactions and informing strategies to mitigate harmful content.

The toxicity score analysis for the second video, conducted by Communalytic using the Perspective API on 347 out of 395 posts, reveals a more pronounced presence of harmful content than the first video. The dataset shows an average toxicity score of 0.11572, with the highest value reaching 0.89882, indicating a notable range of negative discourse. Severe toxicity, though less common, averaged 0.01241 with a peak of 0.44969, while identity attacks averaged 0.04677, with the highest score recorded at 0.49585. Insults were particularly significant, averaging 0.08227 and reaching a maximum of 0.85793, suggesting a frequent use of offensive language. While lower in frequency, profanity and threats still presented averages of 0.07349 and 0.01181, with maximum values of 0.54193 and 0.34804, respectively. These findings highlight a higher level of aggressive and harmful interactions within this dataset, pointing to the need for a closer examination of the factors contributing to these elevated toxicity levels. This analysis provides critical insights into the nature of user engagement with the content, offering a foundation for further investigation into the causes and implications of such discourse.

The toxicity score analysis for the first video, conducted by Communalytic on 136 out of 148 posts using the Perspective API, reveals relatively low levels of harmful content overall, with specific instances of elevated toxicity. The dataset's average toxicity score stands at 0.04762, with the highest value reaching 0.66112, indicating occasional spikes in negative interactions. Severe toxicity is minimal, averaging 0.00484 with a peak of 0.35368, while identity attacks, though also low on average at 0.01413, show a significant maximum value of 0.60072. Insults were more prevalent, with an average score of 0.03403 and a peak of 0.65866, suggesting moments of heightened hostility. Profanity and threats were generally infrequent, with average scores of 0.03131 and 0.00703, respectively, and their highest values remaining relatively low at 0.39410 and 0.02147. These results indicate that while the overall discourse in the video's comments is mainly non-toxic, there are specific instances where the conversation escalates, necessitating further analysis to understand the triggers and context of these spikes in negativity. This insight is crucial for developing strategies to manage and mitigate harmful interactions in digital spaces.

2.2.5 Context Analysis

Context analysis involved systematically coding the narratives from the first, second, and third videos to elucidate the key themes and topics discussed. This coding process meticulously categorizes the content, allowing for a structured examination of the underlying messages and focal points presented in the videos. By breaking the narratives into identifiable codes, the analysis provides a

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

comprehensive overview of the subjects, facilitating a deeper understanding of the content's thematic structure. This approach not only highlights the primary discussions within each video but also enables the identification of recurring themes across the dataset, offering valuable insights into the central concerns and perspectives being communicated. Consequently, the coding of these narratives plays a pivotal role in contextualizing the video content, forming a foundation for further analytical exploration.

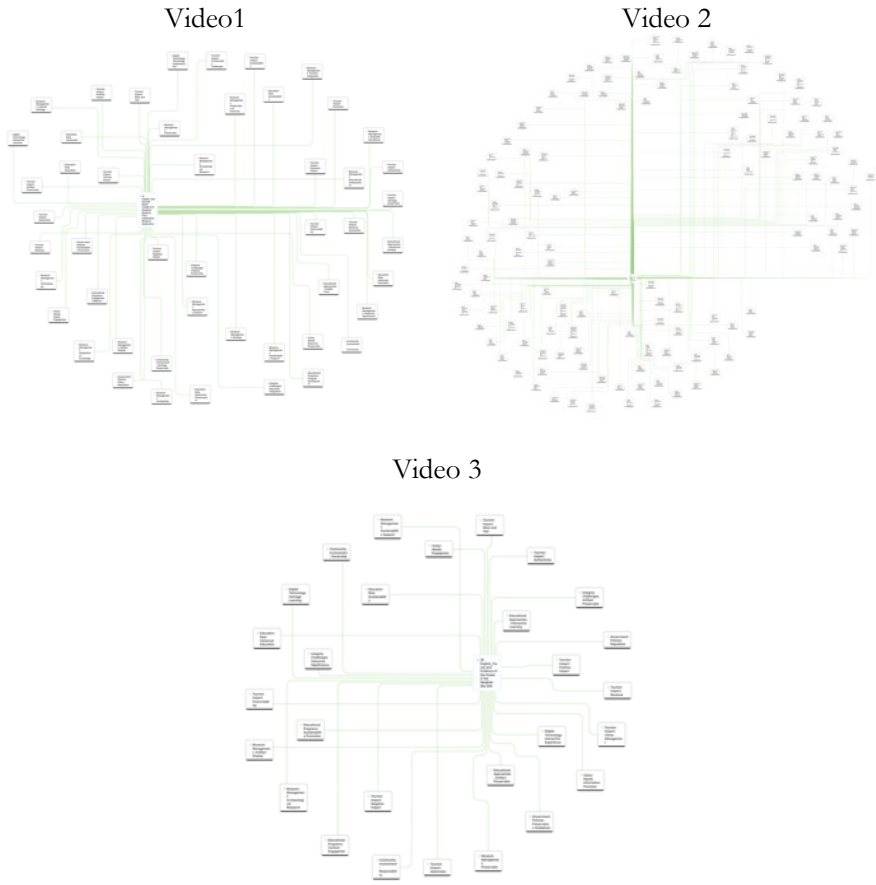

Figure 10. Network by Intentional AI Coding

Figure 10 shows the code network of each video. The implementation of intentional AI coding for the first video reveals a range of topics discussed within the content, reflecting the multifaceted nature of heritage tourism and museum management. Key topics identified include the impact of tourism on artifact preservation, educational programs for sustainability, the role of government

policies in regulating visitor management, and the integration of digital technology to enhance learning and interactive experiences. Additionally, the analysis highlights the challenges related to maintaining historical integrity, the need for community involvement in sustainability efforts, and the balance between promoting visitor engagement and preserving the authenticity of heritage sites. These findings suggest that the video content addresses the opportunities and challenges of managing heritage sites like Sangiran, emphasizing the importance of a holistic approach that integrates educational, technological, and policy-oriented strategies to ensure heritage tourism's long-term sustainability and impact.

The implementation of intentional AI coding for the second video reveals various topics discussed within the content, reflecting a comprehensive exploration of themes relevant to heritage tourism and museum management. The analysis identifies critical discussions around the integration of technological advancements in enhancing visitor experiences, the challenges associated with preserving cultural authenticity, and the impact of tourism on local communities and heritage sites. Additionally, the video delves into the importance of sustainable practices in museum management, the role of educational initiatives in promoting historical awareness, and the influence of governmental regulations on preserving cultural heritage. These findings indicate that the video content offers a well-rounded examination of the complexities of balancing tourism growth with preserving historical and cultural integrity, underscoring the necessity of adopting multidimensional strategies to address these challenges effectively.

The implementation of intentional AI coding for the third video reveals a focused examination of various themes central to heritage tourism and museum management. Key topics identified include integrating digital tools to enhance the visitor experience, the challenges of maintaining historical authenticity in modern tourism pressures, and the role of community involvement in preserving cultural heritage. Additionally, the content explores the impact of government regulations on heritage site management, the importance of educational programs in fostering public awareness, and the necessity of sustainable practices to mitigate the adverse effects of tourism. The analysis highlights the video's comprehensive approach to addressing the multifaceted issues surrounding heritage tourism, emphasizing the need for a balanced strategy that incorporates technological innovation, community engagement, and regulatory oversight to ensure the long-term preservation and enhancement of cultural sites.

3. RESULTS AND DISCUSSION

3.1 Content Analysis of Heritage Tourism and Museum in Sangiran

The first video reveals that the Sangiran Museum in Indonesia is a significant educational and entertainment destination, renowned for its extensive collection

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

of ancient human fossils and artifacts. The museum, which spans a large area, displays relics from various historical periods, including the Paleolithic era, and features reconstructions of early human species, such as Homo erectus and Homo sapiens, utilizing detailed fiberglass models. The exhibits provide a comprehensive overview of human evolution, with particular emphasis on the transition from Homo erectus to Homo sapiens sapiens, alongside fossils of prehistoric animals like elephants, crocodiles, and rhinoceroses. The museum's significance is further highlighted by its recognition by UNESCO in 1996 and its role in preserving Indonesia's rich archaeological heritage. Additionally, the museum serves as a hub for multiple disciplines, including geology, archaeology, anthropology, and paleontology, offering visitors insights into the prehistoric past through well-preserved artifacts and advanced dating methods. The discussion of early human behaviors, such as hunting and toolmaking, alongside the availability of souvenirs and visitor amenities, underscores the museum's multifaceted approach to education and tourism, making it a vital resource for understanding human origins and ancient life in the region.

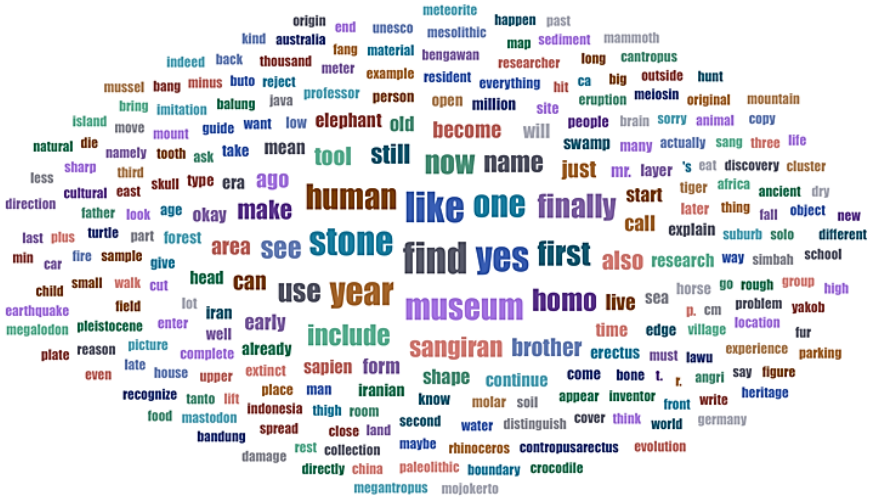

Figure 11. Words Cloud of the First Video Content

Figure 11 shows the word cloud of the first video content. The analysis of the first video's content reveals that certain words frequently recur, reflecting the central themes and focus of the narrative. The words "finally," "find," and "first" indicate a narrative centered on discovery and the significance of initial findings in the context of human evolution. The prominence of "human," "like," and "museum" suggests that the content heavily emphasizes human history and the role of museums in preserving and interpreting these discoveries. Additionally, "stone," "year," and "yes" further underscore the discussion of archaeological findings,

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

mainly stone tools, and the chronological context of these discoveries. The frequent use of these terms highlights the video's focus on uncovering human history and the pivotal role of museums in safeguarding this knowledge for future generations. This analysis provides insight into the video's narrative structure, emphasizing key elements of discovery, historical context, and the preservation of human heritage.

The content of the second video is intricately connected to the discovery and analysis of ancient human fossils, particularly Homo erectus, within the Indonesian archipelago, with a focus on Java. These findings, including 1.8-million-year-old fossils, challenge the traditional "Out of Africa" theory, suggesting a more intricate pattern of human evolution and migration. The region's significance in paleontological research is underscored by exploring sites such as Sangiran in Central Java, which has been pivotal in advancing our understanding of human evolutionary history. The video also explores historical debates regarding human evolution, such as the "missing link" between apes and humans and the classification of early human species, highlighting the complexity and diversity of early hominid development. The discussion extends to specific discoveries like Homo erectus progressive, known for its larger brain volume and its occurrence in particular geological layers in Indonesia, as well as Homo floresiensis, a distinct species found on Flores Island. Sangiran's recognition by UNESCO as a critical site for studying human evolution over millions of years further emphasizes the region's importance, providing invaluable insights into the evolutionary, cultural, and environmental contexts that shaped early human development.

Figure 12. Words Cloud of the Second Video Content

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

Figure 12 shows the word cloud of the second video content. The frequent occurrence of specific words in the second video's content underscores its focus on critical aspects of human evolution and paleontological discoveries. Words such as "ancient," "discovery," and "erectus" indicate a strong emphasis on uncovering the history of early human species, particularly Homo erectus. The repeated use of terms like "find," "fossil," and "homo" highlights the centrality of fossil evidence in understanding the development of early humans. "Human" and "Java" further contextualize these discoveries within the Indonesian archipelago, particularly on the island of Java, a significant site for paleoanthropological research. The words "million," "sanggiran," and "year" suggest a temporal dimension, referencing the millions of years over which these fossils have been preserved and the Sangiran site's critical role in revealing this ancient history. This analysis reflects the video's narrative structure, which revolves around the themes of discovery, the significance of fossil evidence, and the importance of the Sangiran site in understanding the evolutionary timeline of Homo erectus and other ancient human species.

The content of the third video focuses on the remnants of ancient ocean life, including fossils of shells, crabs, and other marine animals found at the Sangiran site in Indonesia. Situated in the Seragen Regency, the site provides compelling evidence that the area was once a deep ocean, as indicated by the discovery of kalibeng soil layers estimated to be around 3 million years old, pointing to the former presence of the Dalak Sea. This significant geological finding has drawn the attention of archaeologists, anthropologists, and geologists, who have been studying the site since 1936 to unravel the rich history of this ancient marine environment. The bluish-green kalibeng soil layer, representing the ancient seabed, not only holds marine fossils but also offers crucial insights into the region's geological and environmental past, underscoring its importance as a critical site for understanding prehistoric oceanic conditions.

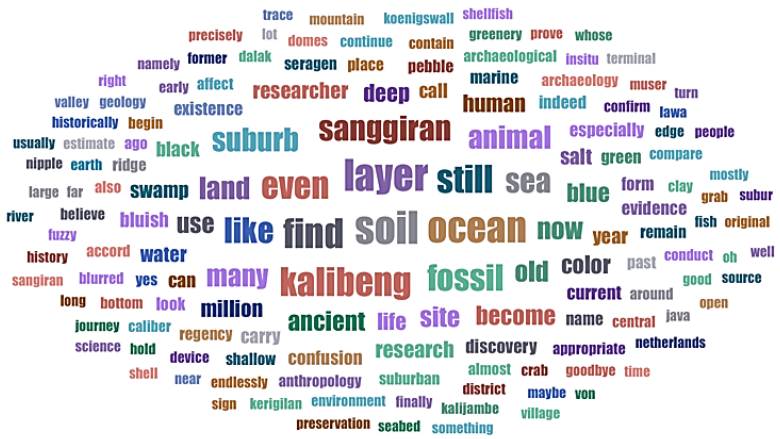

Figure 13. Words Cloud of the Third Video Content

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

Figure 13 shows the word cloud of the third video content. The analysis of the third video's content highlights the frequent occurrence of specific words, reflecting its focus on the geological and paleontological significance of the Sangiran site. Terms like "fossil," "find," and "layer" emphasize the discovery and study of ancient remains within distinct geological strata, particularly the Kalibeng layer. The presence of words such as "ocean," "sea," and "soil" suggests a strong narrative centered on the ancient marine environment that once existed in the region, as evidenced by the fossilized remains found in the Kalibeng soil. The repeated use of "Sangiran" underscores the importance of this site as a critical location for understanding these ancient processes. Additionally, words like "even," "like," and "still" reflect the ongoing relevance of these findings in contemporary research, while "suburb" may indicate a discussion on the site's geographical or environmental context. This analysis underscores the video's thematic focus on the interplay between geological formations and paleontological discoveries, providing a comprehensive understanding of the Sangiran site's historical significance.

The analysis of the three videos reveals several critical topics related to heritage tourism and museum management within the context of Sangiran. Key focus areas include community involvement, which emphasizes the importance of local participation in preserving and promoting the site's cultural heritage. The integration of digital technology is highlighted as a vital tool for enhancing visitor engagement and educational outreach. The role of education is underscored, with particular attention to the development of educational programs and approaches that effectively communicate the historical and cultural significance of the site. Government policies are also a significant theme, reflecting the need for regulatory frameworks that support sustainable tourism and heritage preservation. Challenges related to maintaining the site's integrity, including the management of museum operations, are discussed alongside the impacts of tourism on the site's preservation and visitor experience. Lastly, addressing visitor needs is crucial in ensuring the museum remains relevant and accessible to diverse audiences. These topics collectively illustrate the multifaceted nature of heritage tourism at Sangiran, underscoring the need for a comprehensive approach that balances educational, technological, and policy considerations to sustain the site's cultural legacy.

Figure 14 shows the word cloud of all the video content. The collective analysis of all the videos reveals that certain words appear frequently, underscoring the central themes and focus of the content. Words such as "erectus," "homo," and "human" suggest a strong emphasis on the study of early human species, particularly Homo erectus, and their significance in the broader narrative of human evolution. The recurring use of terms like "find," "fossil," and "year" highlights the importance of discovery and the temporal context within which these fossils are analyzed, pointing to the ongoing efforts to uncover and understand ancient human history. Additionally, the words "like," "now," "one," and "yes" reflect the

video's engagement with contemporary discourse, drawing connections between past discoveries and their relevance today. This analysis indicates that the videos collectively focus on the exploration and interpretation of early human fossils, particularly Homo erectus, within the context of both historical significance and modern scientific inquiry, emphasizing the continuous relevance of these findings in understanding human origins.

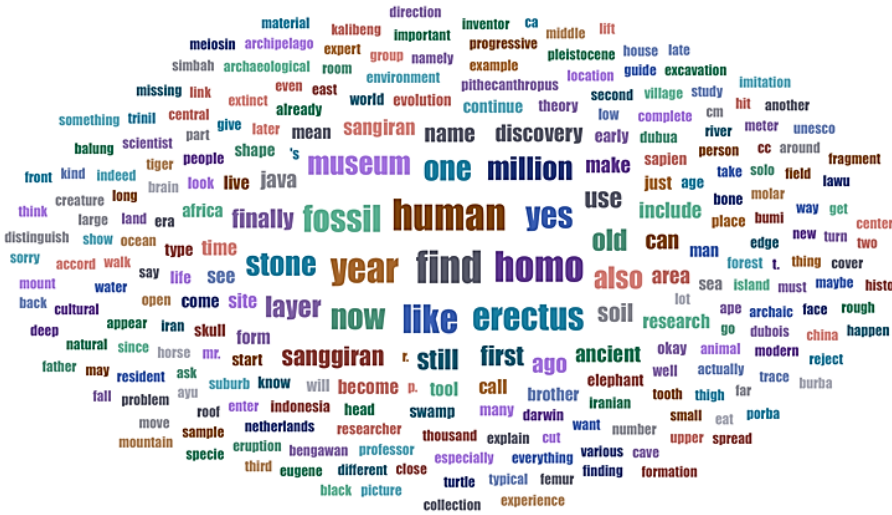

Figure 14. Words Cloud of All Video Content

The relationship between content analysis results and the analysis of sentiment and toxicity scores provides a comprehensive understanding of the discourse surrounding the videos. Content analysis identifies the key themes and topics discussed, such as human evolution, fossil discoveries, and the significance of the Sangiran site, which form the foundation for public engagement. When correlated with sentiment and toxicity scores, these themes offer insights into how the audience perceives and reacts to the content. For instance, a higher toxicity score in discussions about controversial topics, such as the "Out of Africa" theory, may indicate polarizing views among viewers. Conversely, positive sentiment scores might align with content that emphasizes educational value or the scientific importance of the findings. This interplay between content, sentiment, and toxicity underscores the complex dynamics of viewer engagement, highlighting the need to manage and address both positive and negative responses to enhance the educational impact of the videos. Ultimately, this integrated analysis informs strategies for content presentation and audience interaction, ensuring a balanced and constructive discourse.

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

The relationship between the content analysis and the toxicity scores of the first, second, and third videos reveals critical insights into viewer engagement and discourse quality. The content of each video, which centers on themes such as human evolution, fossil discoveries, and the historical significance of the Sangiran site, sets the stage for varying levels of audience reaction, as reflected in the toxicity scores. The first video, with an average toxicity score of 0.04717 and a highest value of 0.71894, indicates relatively low levels of harmful content, correlating with its educational focus. The second video, however, presents a higher average toxicity score of 0.11572 and a peak of 0.89882, suggesting that its more controversial content, mainly related to the "Out of Africa" theory, may provoke stronger adverse reactions. The third video, with an average toxicity score of 0.04762 and the highest value of 0.66112, reflects a similarly low level of toxicity, aligning with its focus on geological and paleontological findings. This correlation between the thematic focus of the videos and the resulting toxicity levels underscores the importance of carefully considering content presentation to mitigate potential adverse reactions, thereby promoting a more constructive and respectful dialogue among viewers. Such an understanding is essential for curating content that educates and maintains a healthy, engaged, and informed audience.

The relationship between the content analysis and the sentiment analysis results obtained using VADER and SVM enhanced by SMOTE provides a nuanced understanding of audience reception and emotional response. VADER's sentiment analysis yielded an accuracy of 94.59%, highlighting its effectiveness in categorizing the sentiment expressed in viewer comments. The SVM model, further enhanced by SMOTE, showed a similarly high performance, with a precision of 92.07% and a recall rate of 97.79%, indicating its robustness in handling imbalanced data. These metrics reveal that the videos generally elicited a positive or neutral sentiment from viewers, which correlates with the content's predominantly educational and informative nature. However, as reflected in the second video's content, the sentiment analysis also identified pockets of negative sentiment, particularly in response to more controversial topics such as the "Out of Africa" theory. The strong alignment between content themes and the sentiment scores underscores the importance of tailoring content to foster positive engagement while also being mindful of potential areas that may provoke adverse reactions. This insight is critical for optimizing content delivery and ensuring that the educational goals of the videos are met with constructive and meaningful audience interaction.

3.2 Discussion

The theoretical analysis of the relationship between content, sentiment, and toxicity analysis within the context of digital content on heritage tourism and museums provides a comprehensive framework for understanding audience interaction and content effectiveness. Content analysis identifies the key themes

and narratives presented, serving as the foundation for understanding the informational and educational value of the material [28], [29]. Sentiment analysis gauges the audience's emotional response and offers insights into how these themes resonate, reflecting positive, neutral, or adverse reactions [30], [31]. Toxicity analysis, on the other hand, measures the presence of harmful or aggressive language within the discourse, highlighting potential areas of conflict or disagreement. The interplay between these three analytical dimensions reveals the complex dynamics of digital content consumption.

In contrast, rich educational content may generate positive sentiment, and it also has the potential to incite toxic behavior if it touches on controversial or sensitive topics. This integrated approach is essential for curating digital content in heritage tourism and museum contexts, as it allows for creating materials that educate, inform, and foster a positive and respectful online environment [32], [33]. Understanding these relationships supports the development of strategies that maximize audience engagement while minimizing the risks of negative interactions.

The relevance of concepts and theories surrounding heritage tourism and museums is particularly pronounced in Sangiran, Central Java, as evidenced by the data. Heritage tourism emphasizes preserving and interpreting cultural assets, aligning closely with the Sangiran site's role as a repository of invaluable paleoanthropological finds. The theories of museum studies, which advocate for museums' educational and community-engaging functions, are reflected in Sangiran's efforts to display ancient human fossils and foster an understanding of human evolution through interactive exhibits and educational programs. The site's recognition by UNESCO underscores its global significance, validating the theoretical framework that views heritage sites as critical to cultural preservation and public education. Furthermore, the data reveals that integrating digital technologies enhances visitor engagement and accessibility, aligning with contemporary museum theory that emphasizes the role of technology in expanding the reach and impact of heritage tourism [34], [35]. Thus, the concepts and theories of heritage tourism and museum studies find direct application in Sangiran, highlighting the site's importance in preserving cultural heritage and educating the public globally.

The intersection of sustainability challenges and local livelihood issues presents a complex dynamic in the context of heritage tourism at Sangiran, Central Java. As a UNESCO-recognized site, Sangiran plays a pivotal role in preserving cultural heritage, yet the influx of tourism necessitates careful management to ensure the site's long-term sustainability. The local community, whose livelihoods are increasingly intertwined with tourism, faces opportunities and challenges. In contrast, tourism can provide economic benefits, but it also risks disrupting traditional ways of life and exacerbating environmental degradation. The pressure

to balance economic development with cultural and environmental preservation requires strategies that integrate sustainable practices, such as eco-friendly tourism initiatives and community-based management models. These approaches aim to protect the site while empowering residents to benefit from tourism that respects their cultural heritage and promotes social equity. Thus, addressing sustainability in heritage tourism at Sangiran involves preserving the site and safeguarding the local community's livelihoods, creating a symbiotic relationship between cultural preservation and economic development.

This research significantly contributes to developing heritage tourism and museum management at Sangiran by comprehensively analyzing visitor engagement, content relevance, and the impacts of digital technologies. Through integrating content analysis, sentiment analysis, and toxicity assessment, the study offers valuable insights into how the museum's narrative and educational programs resonate with local and global audiences. The findings underscore the importance of aligning museum content with visitor expectations and cultural sensitivities, enhancing the educational value and visitor experience. Furthermore, the research highlights the potential of digital tools in expanding the reach and accessibility of Sangiran's exhibits, fostering a deeper understanding of human evolution among a broader audience. By addressing both the opportunities and challenges associated with heritage tourism, this study informs strategies for sustainable development that balance cultural preservation with economic benefits, ultimately contributing to the long-term success and relevance of Sangiran as a leading heritage site.

While offering valuable insights into heritage tourism and museum management at Sangiran, this research has limitations. The study's scope is constrained by its reliance on digital content analysis, which may not fully capture the nuances of on-site visitor experiences and local community perspectives. Additionally, though informative, the sentiment and toxicity analyses are limited by the accuracy of automated tools and the potential bias inherent in online discourse. These limitations suggest that further research should incorporate a more holistic approach, including ethnographic methods, in-depth interviews, and longitudinal studies better to understand the evolving dynamics of heritage tourism at Sangiran. Future research should also explore the long-term effects of digital engagement strategies on visitor retention and education and the socio-economic impacts of tourism on the local community. These directions will enhance the robustness of the findings and contribute to developing more effective, sustainable tourism strategies that benefit both the heritage site and its surrounding communities.

4. CONCLUSION

This research concludes with significant findings illuminating the interplay between digital content, sentiment, and toxicity in the context of heritage tourism

at Sangiran, analyzed through the Digital Content Reviews and Analysis Framework. The toxicity analysis of the first video, with an average score of 0.04717 and a peak of 0.71894, reveals generally low levels of harmful content, indicating a predominantly positive or neutral discourse among viewers. The sentiment analysis conducted using the SVM model enhanced by SMOTE demonstrates a high accuracy rate of 94.59%, with precision and recall rates of 92.07% and 97.79%, respectively, highlighting the model's effectiveness in capturing the emotional tone of the audience. Content analysis, structured through the Digital Content Reviews and Analysis Framework, identifies critical themes such as human evolution, fossil discoveries, and the significance of the Sangiran site, underscoring the educational value of the videos. This framework ensured a systematic evaluation, from content categorization to contextual analysis, providing a comprehensive understanding of how digital content impacts viewer engagement and sentiment. These findings collectively suggest that while digital content effectively engages and educates the audience, there are areas where sentiment fluctuates, particularly around more contentious topics. This comprehensive analysis, supported by the framework, offers a foundation for developing more refined strategies that enhance audience engagement while maintaining a respectful and informative discourse, ultimately contributing to the sustainable development of heritage tourism at Sangiran.

REFERENCES

[1] S. A. Orr, J. Richards, and S. Fatorić, "Climate Change and Cultural Heritage: A Systematic Literature Review (2016–2020)," *Hist. Environ. Policy Pract.*, vol. 12, no. 3–4, pp. 434–477, 2021, doi: 10.1080/17567505.2021.1957264.

[2] J. Moreno-Manzo, L. Coromina, and A. Gassiot-Melian, "Examining the structural relationships between perceived value, satisfaction and loyalty among disabled tourists in two world heritage sites," *Consum. Behav. Tour. Hosp.*, vol. 19, no. 1, pp. 37–54, Jan. 2024, doi: 10.1108/CBTH-02-2023-0011.

[3] J. Bąkiewicz, A. Leask, P. Barron, and T. Rakić, "Towards a visitor taxonomy at (film-induced) heritage attractions," *J. Herit. Tour.*, vol. 17, no. 3, pp. 247–263, May 2022, doi: 10.1080/1743873X.2021.2002340.

[4] K. Bezova and I. Azara, "Generating and Sustaining Value Through Guided Tour Experiences' Co-Creation at Heritage Visitor Attractions," *Tour. Plan. Dev.*, vol. 18, no. 2, pp. 226–244, Mar. 2021, doi: 10.1080/21568316.2021.1879923.

[5] M. S. Hossain, M. F. Rahman, M. K. Uddin, and M. K. Hossain, "Customer sentiment analysis and prediction of halal restaurants using machine learning approaches," *J. Islam. Mark.*, vol. 14, no. 7, pp. 1859–1889, Jan. 2023, doi: 10.1108/JIMA-04-2021-0125.

[6] A. Ahmed Bilal, O. Ayhan Erdem, and S. Toklu, "Children's Sentiment

Analysis from Texts by Using Weight Updated Tuned with Random Forest Classification," *IEEE Access*, vol. 12, pp. 70089–70104, 2024, doi: 10.1109/ACCESS.2024.3400992.

[7] T. Huu Do, M. Berneman, J. Patro, G. Bekoulis, and N. Deligiannis, "Context-Aware Deep Markov Random Fields for Fake News Detection," *IEEE Access*, vol. 9, pp. 130042–130054, 2021, doi: 10.1109/ACCESS.2021.3113877.

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

[9] E. Del Soldato and S. Massari, "Creativity and digital strategies to support food cultural heritage in Mediterranean rural areas," *EuroMed J. Bus.*, vol. 19, no. 1, pp. 113–137, Jan. 2024, doi: 10.1108/EMJB-05-2023-0152.

[10] M. I. Roque, A. C. Campos, S. Almeida, and S. Pasandideh, "Transforming museum management through ICT adoption: an analysis of the Portuguese context during the COVID-19 pandemic," *J. Herit. Tour.*, pp. 1–18, 2024, doi: 10.1080/1743873X.2024.2331239.

[11] 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.

[12] 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.

[13] 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.

[14] C. Bareither, "Difficult heritage and digital media: 'selfie culture' and emotional practices at the Memorial to the Murdered Jews of Europe," *Int. J. Herit. Stud.*, vol. 27, no. 1, pp. 57–72, Jan. 2021, doi: 10.1080/13527258.2020.1768578.

[15] A. M. Walter, "Images of the Mountains: Touristic Consumption and Gendered Representations of Landscape and Heritage in Gilgit-Baltistan," *Vis. Anthropol.*, vol. 35, no. 3, pp. 225–247, May 2022, doi: 10.1080/08949468.2022.2094185.

[16] G. Z. Kohe, J. Smith, and J. Hughson, "#hoops #basketballhistory @Hoops_Heritage: examining possibilities for basketball heritage within the context of higher education, critical museology and digital redirections," *Sport Hist.*, vol. 43, no. 3, pp. 354–377, Jul. 2023, doi: 10.1080/17460263.2021.1952481.

[17] S. Geng, H.-W. Chau, E. Jamei, and Z. Vrcelj, "Understanding place

identity in urban scale Smart Heritage using a cross-case analysis method," *Int. J. Tour. Cities*, vol. 9, no. 3, pp. 729–750, Jan. 2023, doi: 10.1108/IJTC-10-2022-0244.

[18] 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.

[19] 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, 2022, doi: 10.1080/13527258.2021.1941198.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] M. B. Gillette and E. Boyd, "Mining for tourists in China: a digital ethnography of user-generated content from coal mining heritage parks," *J. Herit. Tour.*, vol. 19, no. 1, pp. 1–19, Jan. 2024, doi: 10.1080/1743873X.2023.2255691.

[26] C. Scarles, H. Treharne, M. Casey, and H. Z. Abidin, "Micro-mobilities in curated spaces: agency, autonomy and dwelling in visitor experiences of augmented reality in arts and heritage," *Mobilities*, vol. 15, no. 6, pp. 776–791, Nov. 2020, doi: 10.1080/17450101.2020.1816439.

[27] X. Zhang, Z. Zhao, and K. Wang, "The effects of live comments and advertisements on social media engagement: application to short-form online video," *J. Res. Interact. Mark.*, vol. 18, no. 3, pp. 485–505, Jan. 2024, doi: 10.1108/JRIM-02-2023-0069.

[28] A. Huang, W. Wu, and T. Yu, "Textual analysis for China's financial

markets: a review and discussion," *China Financ. Rev. Int.*, vol. 10, no. 1, pp. 1–15, Jan. 2020, doi: 10.1108/CFRI-08-2019-0134.

[29] T. Redek and U. Godnov, "From data to decision: distilling decision intelligence from user-generated content," *Kybernetes*, vol. 53, no. 13, pp. 1–23, Jan. 2024, doi: 10.1108/K-08-2023-1447.

[30] O. Alqaryouti, N. Siyam, A. Abdel Monem, and K. Shaalan, "Aspect-based sentiment analysis using smart government review data," *Appl. Comput. Informatics*, vol. 20, no. 1–2, pp. 142–161, Jan. 2024, doi: 10.1016/j.aci.2019.11.003.

[31] T. Yu, P. Rita, S. Moro, and C. Oliveira, "Insights from sentiment analysis to leverage local tourism business in restaurants," *Int. J. Cult. Tour. Hosp. Res.*, vol. 16, no. 1, pp. 321–336, Jan. 2022, doi: 10.1108/IJCTHR-02-2021-0037.

[32] R. Sospedra-Roca, F. X. Hernàndez-Cardona, and M. Hernàndez-Pongiluppi, "Digital iconography, conflict heritage, and the teaching of history: The case of the Spanish Civil War," *Cogent Educ.*, vol. 9, no. 1, p. 2082087, Dec. 2022, doi: 10.1080/2331186X.2022.2082087.

[33] G. Tiberghien, R. Mukhamedjanova, and P. F. Xie, "Authenticity and spectrality of space heritage: Baikonur Cosmodrome, Kazakhstan," *Tour. Geogr.*, vol. 25, no. 5, pp. 1445–1464, Jul. 2023, doi: 10.1080/14616688.2023.2231422.

[34] L. Chang, S. Vada, B. D. Moyle, and S. Gardiner, "Unlocking the gateway to tourist well-being: the Triple-A model of engagement in tourism experiences," *Curr. Issues Tour.*, no. May, pp. 1–20, 2024, doi: 10.1080/13683500.2024.2359544.

[35] S. S. Karayazi, G. Dane, and T. Arentze, "Visitors' heritage location choices in Amsterdam in times of mass tourism: a latent class analysis," *J. Herit. Tour.*, vol. 19, no. 4, pp. 497–518, 2024, doi: 10.1080/1743873X.2024.2331227.