Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

Exploring Toxicity and Sentiment in Cultural Heritage Documentation: Content Analysis of Sabu Island's Portrayal in KOMPASTV's Expedition

Yerik Afrianto Singgalen*

Faculty of Business Administration and Communication, Tourism Study Program, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Correspondence Author Email: yerik.afrianto@atmajaya.ac.id
Submitted: **22/08/2024**; Accepted: **10/09/2024**; Published: **12/09/2024**

Abstract−This study explores the dual role of media in preserving and potentially distorting cultural heritage, focusing on the portrayal of Sabu Island in KOMPASTV's expedition documentary. Utilizing the Digital Content Reviews and Analysis Framework, the research comprehensively dissection the documentary's content, uncovering critical insights into the intricate relationship between tourism, cultural preservation, and media representation. By integrating sentiment and toxicity analysis, the study identifies the emotional tone and harmful language present within digital narratives, with the toxicity analysis revealing an average score of 0.09886 and a peak score of 0.83647, indicating the potential influence model of negative discourse on cultural heritage. The sentiment classification, conducted through a Support Vector Machine (SVM) enhanced by SMOTE, demonstrated robust performance metrics, including an accuracy of 66.43%, precision of 60.51%, recall of 94.98%, and an F-measure of 73.90%, with an AUC ranging from 0.728 to 0.904. Additionally, content analysis centered on key themes such as Economic Impact, Sacred Rituals, Tourist Experience, and Weaving Traditions, revealing the complex dynamics where cultural preservation must be balanced with economic development and tourism demands. The findings emphasize the need for responsible and authentic media portrayals to safeguard cultural identities, as media holds the power to uphold or undermine cultural narratives' integrity. This research contributes to the broader discourse on cultural heritage documentation by offering a comprehensive framework for evaluating the impact of digital narratives on the preservation of cultural identities, ensuring the accurate and respectful portrayal of cultural heritage.

Keywords: Cultural heritage; Media representation; Tourism impact; Sentiment analysis; Digital narratives

1. INTRODUCTION

Tourism's inherent connection to cultural heritage is multifaceted, with its role extending beyond mere economic contribution to influencing the preservation and transformation of cultural identities. The influx of visitors to heritage sites often triggers a commodification of cultural practices, wherein traditions are altered to meet tourist expectations, potentially compromising their authenticity [1]–[3]. This phenomenon, though financially advantageous, risks diluting the essence of cultural heritage, necessitating a critical examination of tourism's impact. Nevertheless, the interaction between tourism and cultural heritage also offers significant opportunities for preservation efforts [4]–[6]. The visibility afforded by tourism promotes global awareness of cultural assets, while the financial influx facilitates conservation initiatives, thus playing a pivotal role in safeguarding these assets. Striking an equilibrium between making cultural heritage accessible to tourists and maintaining its authenticity is essential to ensure that tourism serves as a tool for cultural preservation rather than degradation. Achieving this balance requires strategic management that prioritizes the protection of cultural identity while harnessing tourism's potential to contribute positively to the sustainability of heritage sites.

The convergence of cultural heritage and digital narratives reshapes contemporary society's documentation, dissemination, and preservation of traditions. Digitization has significantly expanded the reach and engagement of cultural narratives, allowing global audiences unprecedented access to content that was once limited to specific locales [7]–[11]. However, this shift from traditional to digital mediums introduces the risk of oversimplification, where the depth and contextual richness of cultural narratives may be compromised by the demands of digital platforms for accessibility and conciseness [12]–[15]. Despite these concerns, digital narratives represent a formidable means for preserving and rejuvenating cultural heritage, offering a way to capture and share intangible cultural elements that might otherwise fade into obscurity. Therefore, the practical preservation of cultural heritage through digital platforms necessitates a thoughtful approach that maximizes the benefits of technological advancements while carefully mitigating the potential drawbacks.

Cultural heritage documentation through media plays a crucial role in shaping public perceptions and preserving the intangible aspects of cultural identity. On Sabu Island, where traditional practices persist yet face increasing threats from modernization, media representation of its cultural heritage becomes particularly significant. The KOMPASTV expedition documentary on Sabu Island provides an engaging narrative that intertwines the island's cultural traditions, daily life, and current challenges. Analyzing this documentary through the lenses of toxicity and sentiment offers a deeper insight into the narrative's construction and its potential implications for both the local community and a broader audience. This approach underscores the powerful influence of media in maintaining or altering cultural identities, particularly within communities vulnerable to the pressures of modernity. Thus, the role of

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

media in cultural heritage documentation demands careful consideration, as it holds the capacity to either protect or transform the cultural narratives it portrays.

The urgency of this research is underscored by the growing influence of media in shaping cultural narratives, which demands a rigorous evaluation of the quality and impact of these portrayals. In a time marked by the rapid spread of misinformation and the proliferation of negative stereotypes, examining the toxicity embedded in cultural documentation and the sentiments evoked in audiences is imperative [16]. This analysis is particularly critical for areas like Sabu Island, where the potential for cultural misrepresentation or commodification threatens to distort the community's core values and authentic experiences. By conducting this inquiry, the research seeks to reveal the far-reaching consequences of media representation on cultural integrity, highlighting the necessity for responsible and genuine portrayals that respect and sustain cultural heritage. Such an approach ensures that the media's decisive role in cultural documentation contributes positively to preserving and accurately depicting diverse cultural identities.

The primary objective of this research is to undertake a comprehensive content analysis of KOMPASTV's Sabu Island expedition documentary, focusing on identifying toxic language, examining sentiment, and evaluating how these elements shape the portrayal of the island's cultural heritage. Employing sophisticated textual analysis tools, the study aims to uncover the latent tones and biases within the documentary's narrative, thereby offering a critical assessment of the media's influence on cultural preservation and representation. The Digital Content Reviews and Analysis Framework forms the methodological basis for this inquiry, enabling a structured exploration of the intricate relationships between media portrayals and cultural integrity. This analysis highlights the crucial need for scrutiny of media content to ensure that the representation of cultural heritage is conducted with authenticity and respect, thereby safeguarding the cultural identity of the communities depicted.

Theoretically, this research contributes to the discourse on media representation and cultural heritage by elucidating how toxicity and sentiment are embedded within cultural narratives. It extends existing media analysis frameworks by incorporating sentiment and toxicity analysis as essential to critiquing cultural documentation [17]–[21]. From a practical perspective, the findings of this study offer significant implications for media professionals and cultural preservationists, underlining the importance of responsible storytelling [22]. Furthermore, the research provides actionable guidelines for creating culturally sensitive content that accurately reflects the lived experiences and core values of the communities being depicted, thereby promoting more authentic and respectful media representations. This dual focus on theoretical advancement and practical application ensures that the study enhances academic understanding and informs the ethical practice of media production.

Previous studies have extensively explored the media's role in cultural representation, with a particular emphasis on the portrayal of indigenous communities and the ethical considerations surrounding such portrayals [23]–[26]. Despite the widespread application of sentiment analysis within social media to gauge public opinion and reactions to cultural events or media content, there remains a significant gap in research that integrates sentiment analysis with toxicity detection, specifically within cultural heritage documentation [27]–[30]. This gap is especially evident in documentary filmmaking, where such an analytical approach has been largely overlooked. This research addresses this deficiency by applying these methods to a distinct cultural setting, bridging the divide between theoretical constructs and practical media analysis. Through this integration, the study enhances the understanding of how cultural narratives are constructed and perceived, contributing valuable insights into the complexities of media representation in preserving cultural heritage.

A significant limitation of this study stems from its reliance on textual analysis, which may fall short of capturing the nuanced visual and auditory elements crucial to the documentary's overall impact [31]–[33]. Furthermore, the research is restricted by the documentary's narrow focus, as it offers a singular narrative of Sabu Island's culture, potentially lacking the breadth needed to represent the island's diverse cultural practices fully. The sentiment and toxicity analysis tools are based on pre-existing linguistic datasets, which may not always accurately interpret the cultural and contextual intricacies of the documentary's narrative. These limitations highlight the necessity for a more holistic approach in future research, one that better encompasses the multifaceted nature of cultural documentation, integrating visual, auditory, and contextual analyses to provide a more comprehensive understanding of the cultural narratives being portrayed.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis

The novelty of this research lies in its innovative integration of sentiment and toxicity analysis within the context of cultural heritage documentation, a domain where such analytical approaches have been underutilized. Unlike traditional studies focusing on textual or visual content analysis, this research advances the field by applying sophisticated linguistic tools to evaluate the emotional tone and potentially harmful content embedded within documentary narratives [34]–[37]. This approach enhances the depth of media analysis and provides a more nuanced understanding of how cultural narratives are constructed and perceived. By addressing the intersection of media representation and cultural integrity, the study introduces a new methodological framework that can be applied across various cultural contexts, offering significant implications for both academic research and practical applications in

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

media production. This contribution represents a critical step in ensuring that cultural heritage is portrayed with greater sensitivity and accuracy, reinforcing the importance of responsible media practices in preserving cultural identities.

This research significantly contributes to the body of knowledge by bridging the gap between cultural heritage documentation and advanced linguistic analysis, specifically through integrating sentiment and toxicity analysis within media representations [38]. By introducing this interdisciplinary approach, the study provides a more comprehensive framework for evaluating the impact of media on cultural preservation, moving beyond traditional methods that often overlook the subtle emotional and harmful content conveyed in narratives. The research offers valuable insights into how media can shape public perceptions of cultural identity, highlighting the potential risks of misrepresentation and the importance of ethical storytelling. The study enhances academic understanding of media's role in cultural heritage through its methodological innovations. It sets a precedent for future research in this field, encouraging the adoption of more nuanced and context-sensitive analytical tools. This contribution underscores the need for ongoing examination of media practices to ensure the responsible portrayal and preservation of cultural identities in an increasingly digitized world.

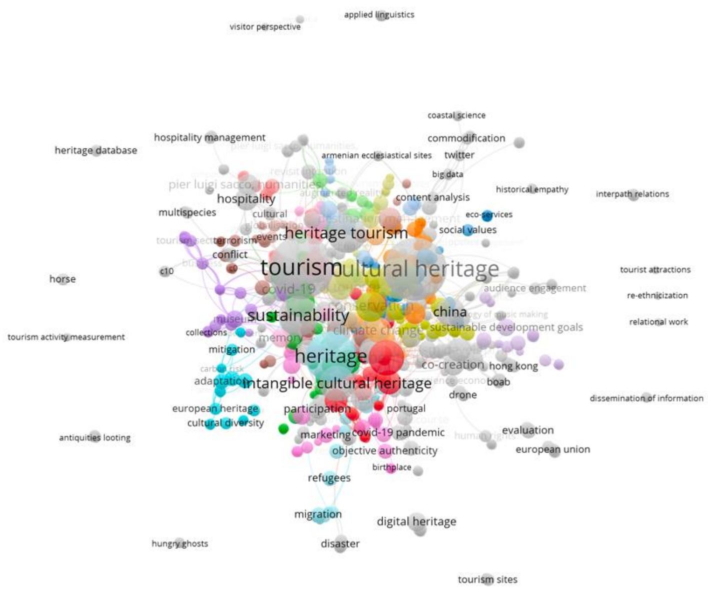


Figure 1. Network Visualization Tourism and Cultural Heritage

Figure 1 shows the network visualization of tourism and cultural heritage. The findings of this research present significant opportunities for contributing to the discourse on tourism and cultural heritage by providing a nuanced understanding of how media representations influence the preservation and transformation of cultural identities [39], [40]. By employing advanced analytical tools to assess sentiment and toxicity within cultural narratives, the study offers a novel approach to evaluating the impact of tourism-driven media content on local communities. This research highlights the media's dual role in promoting cultural heritage and potentially distorting it through commodification or misrepresentation. The insights gained from this study are crucial for developing strategies that ensure tourism-related media fosters authentic cultural preservation while minimizing the risks of cultural degradation. Consequently, this research advances academic discussions and informs practical interventions in tourism and heritage management, emphasizing the importance of responsible media practices in maintaining the integrity of cultural identities amidst the growing influence of global tourism.

In light of the potential scientific contributions, this research adopts the Digital Content Reviews and Analysis Framework to process textual data through a more specialized approach, focusing on toxicity, sentiment, and content analysis. This methodological choice is driven by the need to critically examine how digital narratives shape public perceptions and impact cultural heritage, providing a more granular understanding of the underlying emotional tones and biases in media representations. By integrating these analytical dimensions, the study advances beyond traditional content analysis, offering a comprehensive evaluation of the factors that influence the authenticity and integrity of cultural narratives in digital media. Applying this framework is instrumental in revealing how media can either support or undermine cultural preservation efforts, thereby enhancing the scholarly discourse on the intersection of digital media, cultural identity, and heritage conservation. This approach underscores the importance of adopting robust analytical tools to ensure that the portrayal of cultural heritage in the digital age remains accurate, respectful, and reflective of the communities it represents.

2.2 Digital Content Reviews and Analysis Framework

The Digital Content Reviews and Analysis Framework is structured around several stages, including context reviews, data processing, evaluation and visualization, and context analysis. This framework begins with context reviews,

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

establishing the foundational understanding of the cultural and situational background relevant to the digital content under scrutiny. Following this, data processing involves the systematic organization and preparation of textual data, ensuring it is ready for in-depth analysis. The evaluation and visualization phase then applies analytical tools to extract meaningful insights, presenting the findings in a manner that highlights critical patterns and trends. Finally, context analysis integrates these results into the original cultural context, comprehensively interpreting how the digital content influences or reflects cultural narratives. This sequential approach ensures a thorough examination of digital media. It facilitates the accurate assessment of its impact on cultural heritage, underscoring the framework's utility in academic research and practical applications.

The primary advantage of this framework lies in its comprehensive and systematic approach to analyzing digital content, making it particularly effective for studying the intricate relationship between media representations and cultural heritage. By incorporating multiple stages—context reviews, data processing, evaluation and visualization, and context analysis—the framework ensures that each aspect of the digital content is thoroughly examined within its cultural and situational context. This multifaceted approach allows a deeper understanding of how digital narratives are constructed and perceived, providing insights beyond surface-level interpretations. Furthermore, integrating advanced analytical tools for sentiment, toxicity, and content analysis enhances the precision of the evaluation, enabling a more nuanced interpretation of the data. Such a robust and versatile structure strengthens the validity of the research findings. It offers a valuable tool for future studies to assess digital media's impact on cultural preservation and representation. This framework's ability to deliver detailed and contextually informed analyses highlights its significant contribution to digital humanities and cultural studies.

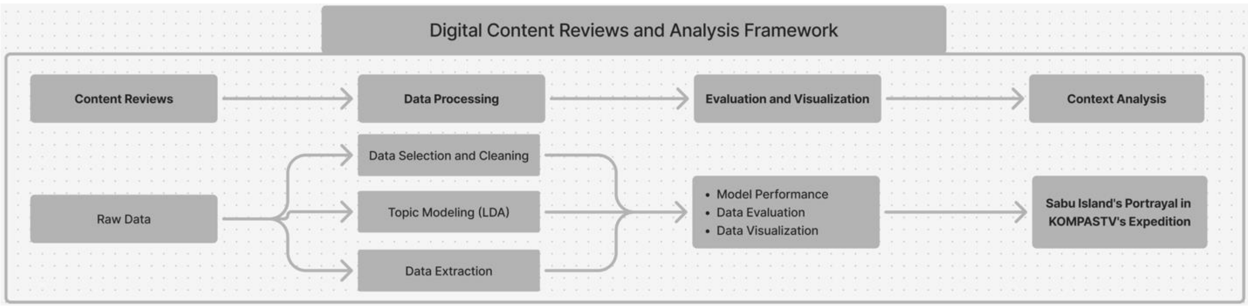


Figure 2. Digital Content Reviews and Analysis Framework

Figure 2 shows the digital content reviews and analysis framework. The data processing in this research utilizes advanced analytical tools, including Communalytic, Turboscript, and Atlas.Ti, as well as RapidMiner, will ensure a comprehensive and multi-dimensional analysis. Communalytic is employed for its robust capabilities in social media data analysis, facilitating the identification of sentiment and toxic language within large datasets. Turboscript is utilized to streamline the transcription and coding processes, enabling efficient qualitative data handling. Atlas.Ti provides a powerful platform for qualitative data analysis, allowing for the intricate coding and thematic exploration of textual data. RapidMiner is integrated into the workflow to apply sophisticated machine learning techniques, enhancing data insights' predictive accuracy and depth. By leveraging these tools in a coordinated manner, the research achieves high analytical rigor, ensuring that the findings are reliable and insightful. This methodological approach not only strengthens the validity of the research but also exemplifies the effective use of technology in advancing the field of digital humanities and social science research.

The framework's relevance to this research is evident in its ability to systematically dissect and interpret the complex dynamics of digital content in the context of cultural heritage preservation. This framework is particularly suited to the study as it offers a structured approach to analyzing the interplay between media narratives and cultural identity through stages of context reviews, data processing, evaluation, and context analysis. Its multi-layered methodology allows for an in-depth examination of digital content's overt and subtle elements, ensuring that the research captures the nuanced ways cultural narratives are constructed, perceived, and potentially altered. By integrating sentiment, toxicity, and content analysis tools, the framework aligns with the research's objective of evaluating the integrity of cultural representations in digital media. This alignment enhances the analytical precision and ensures that the research outcomes directly apply to the ongoing discourse on media's role in cultural preservation, making the framework an indispensable part of the study's success.

2.2.1 Content Reviews

The video content selected for processing is titled *Surga Hewan Endemik: Sulawesi*, identified by the ID 94plxPmkKVg, and published by the channel Papaver Somnivera on June 29, 2024, with 491,740 views and 1,100 comments. This video is a significant source for analysis due to its substantial viewership and engagement, indicating its potential influence on public perceptions of Sulawesi's endemic wildlife. The comments section, reflecting diverse sentiments, offers valuable data for examining the intersection of digital narratives, sentiment, and public discourse on wildlife conservation. Analyzing this content provides an opportunity to understand how visual media forms digital narratives, thereby informing strategies that align conservation efforts with public engagement. This analysis is critical

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

for developing more effective communication strategies that leverage digital platforms to support wildlife conservation initiatives.

Video id Zrvgcm5jO9U

Video id s0IXDJw4yM8

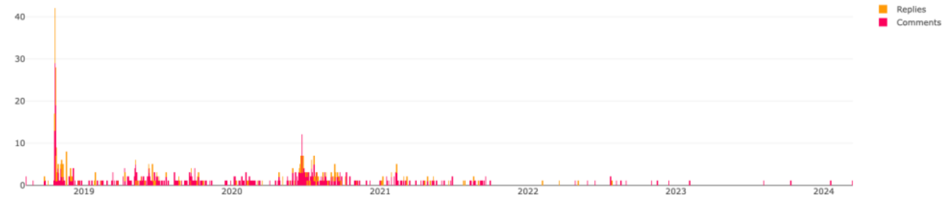


Figure 3. Post-per-day Statistic of the Video (Communalytic)

Figure 3 shows the content's post-per-day statistics. The video identified by ID Zrvgcm5jO9U, which has accumulated 3,250,400 views since its release on August 4, 2018, alongside 1,925 comments, presents a detailed timeline of audience engagement that fluctuates over an extended period. As evidenced by the comment activity, initial engagement began moderately with peaks and troughs, showing a significant increase in interaction starting around mid-October 2018, where daily comment counts spiked dramatically. This heightened activity continued through various phases until it gradually tapered off in 2020 and beyond, with sporadic comments appearing as late as 2024. This pattern indicates that the content has maintained a lasting impact and relevance, drawing ongoing interest from viewers over several years. Such sustained engagement suggests the video's ability to resonate with a diverse audience, likely due to its compelling content and the relevance of its subject matter, thus making it a valuable case study for understanding long-term audience interaction dynamics in digital media.

The video identified by ID s0IXDJw4yM8, which has amassed 1,180,114 views since its release on August 11, 2018, and attracted 787 comments, demonstrates a diverse pattern of audience engagement over time. Initially, comments were sparse, but a notable surge in activity occurred in late October 2018, mainly on October 22, when 29 comments were recorded in a single day, reflecting a significant spike in viewer interaction. The heightened engagement persisted for several weeks, indicating that the content resonated strongly with the audience. The subsequent decline in comment frequency and occasional bursts of activity suggests that the video maintained relevance, albeit with diminishing intensity, well into 2021 and beyond. The fluctuating pattern of commentary highlights the video's enduring impact on viewers, suggesting its ability to provoke thoughtful discourse and sustain interest over an extended period. This pattern provides valuable insights into how digital media content can continue to engage audiences long after its initial release, underscoring the potential for sustained influence in the digital age.

Subsequent analysis must focus on identifying the top ten posters for each video to gain deeper insights into the dynamics of audience engagement. Understanding the most active contributors is crucial, as these individuals often shape the narrative within the comment sections, potentially influencing the overall sentiment and discourse surrounding the video content. By examining the frequency and nature of their contributions, it becomes possible to assess these posters' impact on the discussion and how their engagement might reflect broader audience trends. This targeted analysis will enhance the understanding of audience interaction and provide a more nuanced perspective on how key individuals contribute to the collective narrative within digital media spaces, thereby informing strategies for managing and responding to audience feedback in future content creation.

Video id Zrvgcm5jO9U

Video id s0IXDJw4yM8

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

Figure 4. Top Ten Poster (Communalytic)

Figure 4 shows the top ten posters. An analysis of the top-ten posters for the video identified by ID Zrvgcm5jO9U reveals significant contributions from a select group of users, with @adilasri9699 leading the engagement with 19 comments. Following closely are @santaiaja2679 with 15 comments and @anakruteng2850 with 11 comments, indicating their active participation in shaping the discourse surrounding the video. Other notable contributors include @faisalflores2736, @alghifarinzr397, @arjunapratama4271, @stephanivya2427, @cohar_official, and @MrLodwyk, each posting between 7 to 8 comments, and @militansejati64 with 6 comments. This concentration of activity among a few users suggests that these individuals played a pivotal role in driving the conversation, potentially influencing the tone and direction of the discussions. Their engagement indicates a deeper connection with the video content, possibly reflecting strong personal or community ties to the subject matter. Understanding the motivations and perspectives of these key posters can provide valuable insights into audience engagement patterns and inform strategies for fostering more inclusive and diverse discussions in future media content.

An analysis of the top-ten posters for the video identified by ID s0IXDJw4yM8 reveals that @hopehope6454 is the most active contributor with nine comments, followed by @tersiaratu7762 with seven comments. Other notable participants include @fuadfahmi8189 and @gustafherewila7024, each contributing five comments, while @arikobama added four comments to the discussion. Additionally, @jontorjoni2632, @artyavega8918, @rahmatsantoso6983, @linar3275, and @naduduina8939 each posted 3 comments. This concentrated activity among these users suggests they played a crucial role in shaping the narrative and influencing the discourse surrounding the video. Their repeated engagement indicates a strong connection or particular interest in the video's content, potentially driving the conversation in specific directions and contributing to the overall tone of the comment section. Understanding the perspectives and motivations of these top contributors could provide valuable insights into audience dynamics and the factors that encourage sustained engagement with digital content.

Following the identification and content analysis, the relevance of the tourism and cultural heritage topic within the context of digital narratives becomes evident, providing a solid foundation for further data processing. This intersection is particularly significant as it allows for a nuanced exploration of how cultural heritage is represented and preserved in the digital age, primarily through the lens of tourism-driven content. The analysis underscores digital narratives' critical role in shaping public perceptions and the potential impacts on cultural identity. By processing the data with this contextual understanding, the research is poised to offer valuable insights into how digital platforms influence and transform cultural heritage narratives, thereby contributing to the broader discourse on cultural preservation in the digital era. This approach ensures that the subsequent data processing will be relevant and aligned with the study's objectives, leading to meaningful conclusions.

2.2.2 Data Processing

During the data processing stage, a meticulous data cleansing procedure is implemented using the RapidMiner application, employing operators such as tokenize, transform cases, filter token by length, and filter by stopwords. The tokenization process breaks down the text into individual units, facilitating more precise analysis. Subsequently, the transform cases operator standardizes the text by converting all tokens to a uniform case, which helps reduce redundancy and ensure consistency across the dataset. The filter token by length operator is then applied to eliminate irrelevant data, such as excessively short or long tokens that do not contribute meaningful insights. Finally, the filter by stopwords operator removes common words that are generally non-informative, allowing for a more focused analysis of the significant terms. This systematic approach to data cleansing enhances the quality and reliability of the processed data, laying a robust foundation for subsequent analytical tasks and ensuring that the results are accurate and insightful.

Two thousand seven hundred fifty-seven posts have been meticulously cleaned and are now prepared for extraction using the VADER model to generate comprehensive sentiment analysis metrics. This process will yield critical outputs such as the overall Score, Scoring String, Negativity, Positivity, Uncovered Tokens, and Total Tokens. The VADER model, known for its effectiveness in analyzing social media text, will enable a nuanced interpretation of the sentiment expressed within these posts. By quantifying the emotional tone through metrics like Negativity and

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

Positivity, along with assessing the coverage of the sentiment-bearing tokens, this analysis provides a detailed understanding of the underlying sentiments in the dataset. The systematic extraction of these features ensures that the data is primed for further interpretive analysis, offering valuable insights into the emotional dynamics of the discourse under examination. This approach enhances the accuracy of the sentiment analysis and contributes to the research findings' depth and reliability.

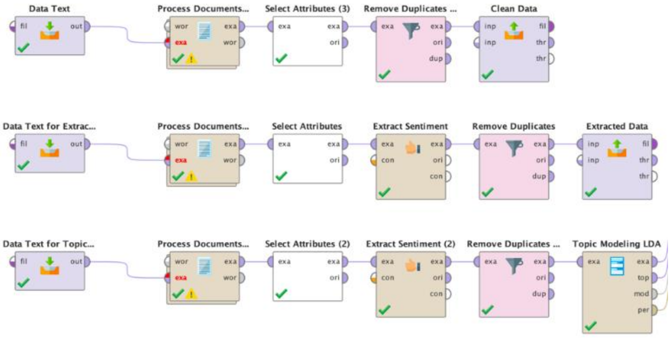


Figure 5. Data Cleaning, Extraction, and Topic Modeling

Figure 5 shows the data cleaning, extraction, and topic modeling process. Of the 2,757 posts extracted and analyzed, 462 have been classified into the harmful sentiment category, while 2,294 posts fall into the positive sentiment category. This distribution highlights a predominantly positive tone within the dataset, suggesting that most of the discourse is favorable or supportive. The significant difference in the number of posts between the two sentiment classes indicates a strong leaning toward positive engagement among the users. This sentiment analysis provides a critical understanding of the overall emotional landscape of the posts, offering insights into the collective mood and attitudes expressed by the audience. Such findings are essential for interpreting the broader implications of the data, particularly in understanding the general reception and impact of the content being discussed.

Subsequently, topic modeling was conducted using Latent Dirichlet Allocation (LDA) to identify the key topics that garnered the most attention from the viewers. This method allows for extracting latent themes within the dataset, providing a structured view of the subjects that resonate most with the audience. By analyzing the distribution of topics across the posts, LDA reveals the underlying patterns of discussion, highlighting the areas of interest and concern among viewers. This process not only helps in understanding the focal points of the discourse but also aids in uncovering the broader narrative that drives viewer engagement. The insights gained from this modeling topic are crucial for contextualizing the sentiment analysis, offering a comprehensive understanding of viewer interactions' emotional and thematic dimensions. Ultimately, these findings contribute to a more nuanced interpretation of the data, informing future strategies for content creation and audience engagement.

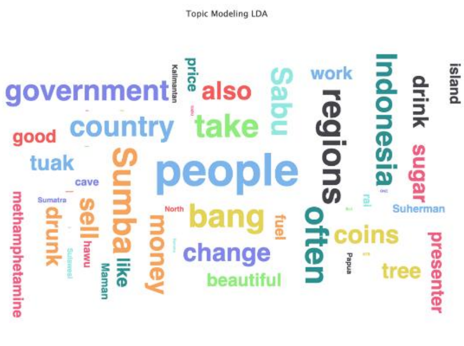


Figure 6. Topic Modeling with LDA

Figure 6 shows the topic modeling result using LDA. The results from the LDA topic modeling reveal the model's complex yet insightful performance profile, reflected in key metrics. The LogLikelihood score of -188183.524 and a Perplexity of 1432.760 indicate the model's effectiveness in capturing the underlying structure of the data, albeit with some challenges in managing the diversity of the topics. An average of 2208.100 tokens per document and an average document entropy of 5.665 suggest moderate uncertainty in topic distribution, while the average word length of 5.540 and coherence score of -25.979 highlight the balance between topic specificity and generalization. The model's AlphaSum of 1.440 and Beta of 0.105, alongside a BetaSum of 671.656, reflect the distributional assumptions about topic prevalence and word distribution within topics. Metrics such as an average exclusivity of 0.789 and an average allocation count of 0.506 further emphasize the model's ability to assign distinct topics to documents while maintaining a reasonable overlap, which is crucial for capturing the nuances of the viewer's attention. These findings

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

underscore the complexity and variability in the topics that engage the audience, providing a detailed map of the thematic landscape within the data.

The calculated metrics from the LDA topic modeling offer a comprehensive insight into the model's performance and the thematic structure of the analyzed content. The LogLikelihood and Perplexity values, while indicating the complexity of the data, suggest that the model has effectively captured the distribution of topics, although the relatively high Perplexity points to the inherent diversity within the dataset. The average document entropy and token statistics reflect a balanced but intricate allocation of topics across documents, suggesting moderate overlap. Although negative, the coherence score aligns with real-world data's complexity, where topics often exhibit subtle nuances and intersections. The exclusivity and allocation count metrics demonstrate the model's ability to assign distinct topics while preserving some degree of thematic overlap, essential for accurately reflecting the multifaceted nature of the viewer's interests. Together, these results indicate that the model has successfully delineated the primary topics while acknowledging the nuanced and overlapping themes in the content, thereby providing a robust framework for further analysis and interpretation of the thematic landscape.

The findings from the LDA topic modeling hold significant relevance for this research, as they offer a detailed understanding of the thematic structure that underpins the viewer's engagement with the content. The complexity and diversity reflected in the LogLikelihood and Perplexity values align with the intricate nature of cultural narratives, suggesting that the audience is drawn to a wide array of interconnected topics. This complexity is further underscored by the document entropy and coherence scores, which indicate a nuanced distribution of themes, mirroring the real-world intersections of tourism and cultural heritage within digital narratives. The model’s ability to balance exclusivity and allocation provides insight into how distinct yet overlapping topics resonate with viewers, highlighting the multifaceted interests that drive engagement. By connecting these metrics to the broader research objectives, it becomes clear that the LDA model effectively captures the rich and varied landscape of viewer interests, offering a robust foundation for further exploration of how digital narratives shape and reflect cultural heritage. This connection enhances the study’s capacity to draw meaningful conclusions about the interaction between digital content and cultural preservation.

The insights gained from the LDA topic modeling underscore the importance of evaluating the performance of the VADER classification model using algorithms such as k-NN, SVM, Decision Trees (DT), and Naive Bayes Classifier (NBC). Given the complexity and diversity of themes identified by LDA, assessing how effectively the VADER model captures and classifies the sentiment within these nuanced topics is crucial. Using k-NN, SVM, DT, and NBC in performance evaluation provides a robust comparative framework, allowing for a comprehensive understanding of the model's accuracy, precision, and ability to generalize across different thematic contexts. This multi-algorithm approach ensures that the classification model is rigorously tested against various decision-making processes, essential for validating its effectiveness in handling the intricate and overlapping themes identified by LDA. Ultimately, this evaluation is pivotal in refining the sentiment analysis process, ensuring that the model aligns with the thematic complexity revealed by LDA and provides reliable and insightful sentiment classifications that contribute meaningfully to the research outcomes.

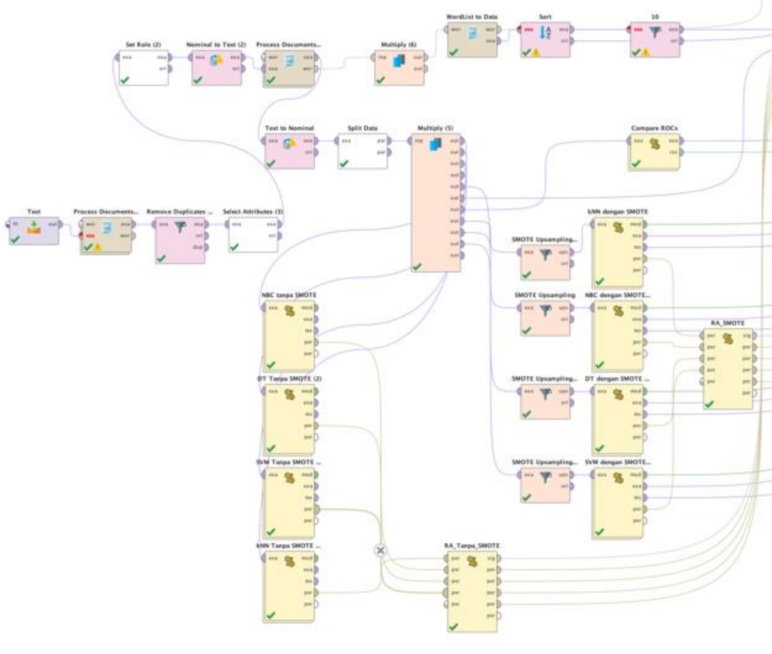


Figure 7. Performance Evaluation of Classification Model (Rapidminer)

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

Figure 7 shows the performance evaluation of the classification model using Rapidminer. The evaluation of model performance involves a systematic comparison of results obtained with and without the application of SMOTE (Synthetic Minority Over-sampling Technique) to identify the most effective algorithm—k-NN, SVM, DT, or NBC—within the context of the given data. Initially, models are trained and evaluated on the imbalanced dataset, providing a baseline performance metric. Subsequently, SMOTE is applied to address the class imbalance, and the models are re-evaluated under these new conditions. This comparison reveals how SMOTE enhances the models' generalization ability, particularly in handling minority classes. The performance differences in accuracy, precision, recall, and F1 score allow a nuanced understanding of each algorithm's strengths and weaknesses. The results from this evaluation process enable the recommendation of the most suitable algorithm that achieves optimal performance in the context of the imbalanced data and demonstrates robustness when balanced data is introduced through SMOTE. This comprehensive evaluation ensures that the selected algorithm is well-suited to the specific characteristics of the data, thereby enhancing the reliability and validity of the research findings.

The next step involves evaluating the algorithm with the best performance using key metrics such as accuracy, precision, recall, F-measure, and AUC (Area Under the Curve). This comprehensive evaluation allows for a detailed assessment of the algorithm's effectiveness in correctly classifying data, capturing the balance between sensitivity and specificity, and maintaining overall model robustness. Accuracy provides a general measure of correctness, while precision and recall offer insights into the algorithm’s ability to handle positive instances. F-measure serves as a harmonic mean of these two metrics, balancing their trade-offs. AUC further complements this analysis by indicating the algorithm's ability to distinguish between classes across various threshold settings. By examining these metrics, the evaluation confirms the algorithm's superiority and ensures its suitability for the specific requirements of the dataset, ultimately guiding its application in the research context.

2.2.3 Data Evaluation and Visualization

The evaluation and visualization of data processing results for the videos with IDs Zrvgcm5jO9U and s0IXDJw4yM8 can be comprehensively assessed by examining the calculated toxicity scores, the performance of the VADER classification model, and the content analysis outcomes. The toxicity scores provide a quantitative measure of the harmful language in the comments, offering insights into the overall tone of viewer engagement. The performance of the VADER model, evaluated through metrics such as accuracy, precision, and recall, reflects its effectiveness in sentiment classification within this context. Meanwhile, the content analysis results further contextualize these findings by revealing the themes and topics that dominate the discussions. Together, these elements form a holistic view of the data, allowing for a deeper understanding of the emotional and thematic dimensions of the content interactions, which is critical for deriving meaningful conclusions from the analysis.

The toxicity score analysis for the video with ID Zrvgcm5jO9U, conducted using Communalytic and the Perspective API, evaluated 1,593 out of 1,925 posts, revealing significant insights into the nature of the discourse. The average toxicity score across the dataset was 0.09886, with the highest individual score reaching 0.83647, indicating occasional but notable harmful language. Severe toxicity was less prevalent, with an average of 0.00853 and a peak value of 0.45727. Other concerning metrics included identity attacks with an average score of 0.02786 and insults averaging 0.08239, both of which underscore the presence of hostile interactions. Profanity was also detected, averaging 0.05689, with a maximum score of 0.64460. Threats were the least frequent, with an average score of 0.00988. These findings highlight the presence of varying degrees of hostile language, though the overall averages suggest that such content, while present, is not pervasive. This analysis is crucial for understanding the social dynamics within the video's comment section and can inform strategies for moderating online interactions to foster a healthier digital environment.

Video id Zrvgcm5jO9U

	Average for dataset	Highest value
Toxicity	0.09886	0.83647
Severe Toxicity	0.00853	0.45727
Identity Attack	0.02786	0.49117
Insult	0.08239	0.78565
Profanity	0.05689	0.64460
Threat	0.00988	0.34804

Video id s0IXDJw4yM8

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

	Average for dataset	Highest value
Toxicity	0.08581	0.69899
Severe Toxicity	0.00855	0.43794
Identity Attack	0.01934	0.71679
Insult	0.05760	0.69911
Profanity	0.06246	0.60019
Threat	0.01095	0.53305

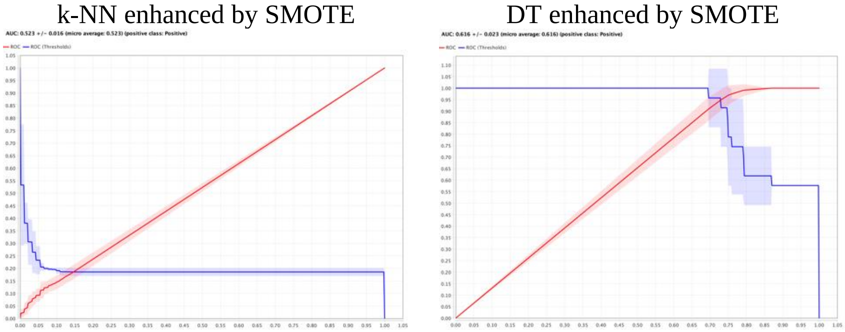
Figure 8. Average for Dataset and Highest Value of Toxicity Classification (Communalytic)

Figure 8 shows the dataset's average and the highest toxicity classification value using Communalytic. The toxicity score analysis for the video with ID s0IXDJw4yM8, conducted using Communalytic and the Perspective API, assessed 610 out of 787 posts, revealing a range of negative language indicators within the comment section. The average toxicity score for the dataset was 0.08581, with the highest score recorded at 0.69899, indicating a moderate presence of toxic comments. Severe toxicity remained low, averaging 0.00855, with a notable peak of 0.43794. Identity attacks were relatively infrequent, with an average score of 0.01934, though the highest score reached 0.71679, reflecting a significant instance of targeted hostility. Insults averaged 0.05760, with a maximum score of 0.69911, while profanity had a slightly higher average of 0.06246 and a peak of 0.60019. The threat score was the lowest on average at 0.01095 but with a concerning maximum of 0.53305. These results highlight the presence of harmful language within the comments, with occasional spikes in severity, underscoring the need for continued monitoring and moderation to maintain a constructive and respectful discourse in the digital space.

The toxicity score analysis for the videos Zrvgcm5jO9U and s0IXDJw4yM8 provides critical insights into the nature and intensity of negative interactions within their comment sections. Both videos exhibit moderate average toxicity levels, with Zrvgcm5jO9U showing a slightly higher average score than s0IXDJw4yM8, indicating a marginally more hostile environment. Severe toxicity and identity attacks were relatively low in both datasets. Yet, their peak values reveal the presence of isolated but significant instances of aggressive behavior, particularly in identity attacks for s0IXDJw4yM8. The presence of insults and profanity further underscores the existence of negative sentiment, though their averages suggest that such language, while present, is not overwhelmingly pervasive. The slightly higher threat scores in s0IXDJw4yM8, with a maximum score reaching over 0.5, indicate occasional severe concerns within the discourse. These findings emphasize the importance of continued content moderation to mitigate the impact of these harmful elements, ensuring that the online discussions remain constructive and respectful, thereby preserving the integrity of the digital environment associated with these videos.

After calculating toxicity scores, comparing these results with the outcomes of sentiment classification is essential to gain a more comprehensive understanding of the discourse surrounding the videos. While toxicity scores measure harmful language, sentiment classification offers insights into the content's overall emotional tone, which is positive, harmful, or neutral. By juxtaposing these analyses, one can discern patterns where high toxicity may correlate with negative sentiment or where positive sentiment persists despite isolated instances of toxic language. This comparison is crucial for identifying inconsistencies or reinforcing findings, thereby enhancing the robustness of the analysis. Such an integrated approach allows for a more nuanced interpretation of the data, enabling more informed conclusions about the nature of audience interactions and the overall impact of the content.

k-NN enhanced by SMOTE DT enhanced by SMOTE



SVM enhanced by SMOTE NBC enhanced by SMOTE

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

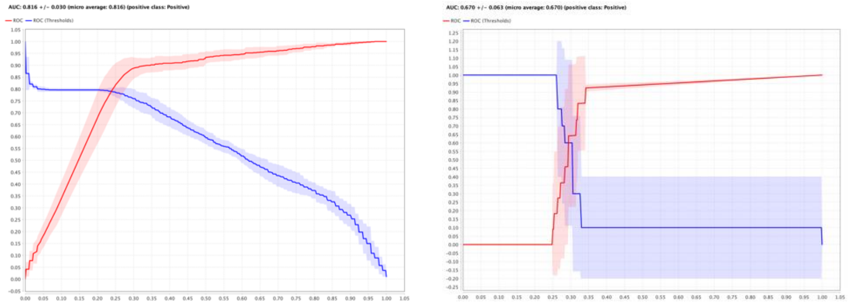


Figure 9. Area Under Curve of SVM, k-NN, DT, and NBC enhanced by SMOTE

Figure 9 shows the Area Under the Curve of SVM, k-NN, DT, and NBC enhanced by SMOTE. Based on the comparison of AUC values, it is evident that the SVM model enhanced by SMOTE demonstrates superior performance, particularly in its ability to distinguish between classes. The model achieves an optimistic AUC of 0.904, indicating a high level of accuracy in predicting the positive class, with an overall AUC of 0.816, which confirms its robustness across different thresholds. The model maintains discriminatory solid power despite the slightly lower AUC in the pessimistic scenario, at 0.728. The accuracy of 66.43% and the precision of 60.51% reflect a solid balance between correctly identified positive cases and minimizing false positives. The recall rate of 94.98% is particularly noteworthy, suggesting that the model is highly effective at capturing the positive instances, albeit with some trade-off in precision. The F-measure of 73.90% further corroborates the model's balanced performance, combining precision and recall to provide a comprehensive assessment. These metrics collectively affirm the enhanced SVM model's capability, particularly in contexts requiring high sensitivity to positive classifications, making it a recommended choice for similar analytical tasks.

The performance of the SVM model without the application of SMOTE reveals notable characteristics, particularly in terms of accuracy and class discrimination. The model achieved an accuracy of 70.39%, reflecting its general effectiveness in classifying data correctly. However, the AUC values present a more nuanced picture: while the optimistic AUC is relatively high at 0.871, indicating strong performance under ideal conditions, the overall AUC drops significantly to 0.633, and the pessimistic AUC falls further to 0.395. It suggests that the model struggles with class discrimination when the data distribution is less favorable. Despite this, the model exhibits a high recall rate of 96.94%, which indicates its strong ability to identify positive cases. However, this is somewhat offset by a precision of 70.03%, pointing to a higher rate of false positives. The F-measure of 81.31% demonstrates a reasonable balance between precision and recall. Still, the overall performance suggests that while the SVM model performs adequately without SMOTE, it may not handle imbalanced data as effectively as needed, highlighting the potential benefits of data balancing techniques for improving classification outcomes.

The toxicity calculation and sentiment classification results should be carefully compared with the outcomes of content analysis to gain a holistic understanding of the data. While toxicity scores provide insights into the presence of harmful or aggressive language, and sentiment classification identifies the overall emotional tone, content analysis delves deeper into the thematic structure and context of the discussions. By juxtaposing these results, it becomes possible to identify correlations between toxic language and negative sentiment or to uncover instances where content may be thematically rich but marred by underlying negativity. This comparison is crucial for ensuring the analysis captures the surface-level emotions and language and the deeper meanings and themes driving the conversations. The integration of these three analytical dimensions enhances the overall interpretation of the data, leading to more informed conclusions about the nature and impact of the content on its audience.

2.2.4 Content Analysis

The content analysis reveals that the video’s narrative is intricately connected to critical themes, including Arid Adaptation, Economic Impact, Lontar Palm, Marketing Strategies, Rituals, Sacred Rituals, Tourist Experience, and Weaving Traditions. These themes collectively illustrate the complex interplay between cultural heritage and modern influences, highlighting how traditional practices such as weaving and rituals are being maintained and adapted in the face of economic and environmental challenges. The focus on marketing strategies and tourist experiences suggests an ongoing effort to balance cultural preservation with economic development, particularly in promoting unique local traditions like the Lontar Palm and sacred rituals. Additionally, the discussion of transportation impact underscores the broader infrastructural considerations affecting the community and the visitor experience. This thematic richness underscores the video’s role in portraying a multifaceted narrative that preserves cultural identity and engages with contemporary economic and environmental pressures.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

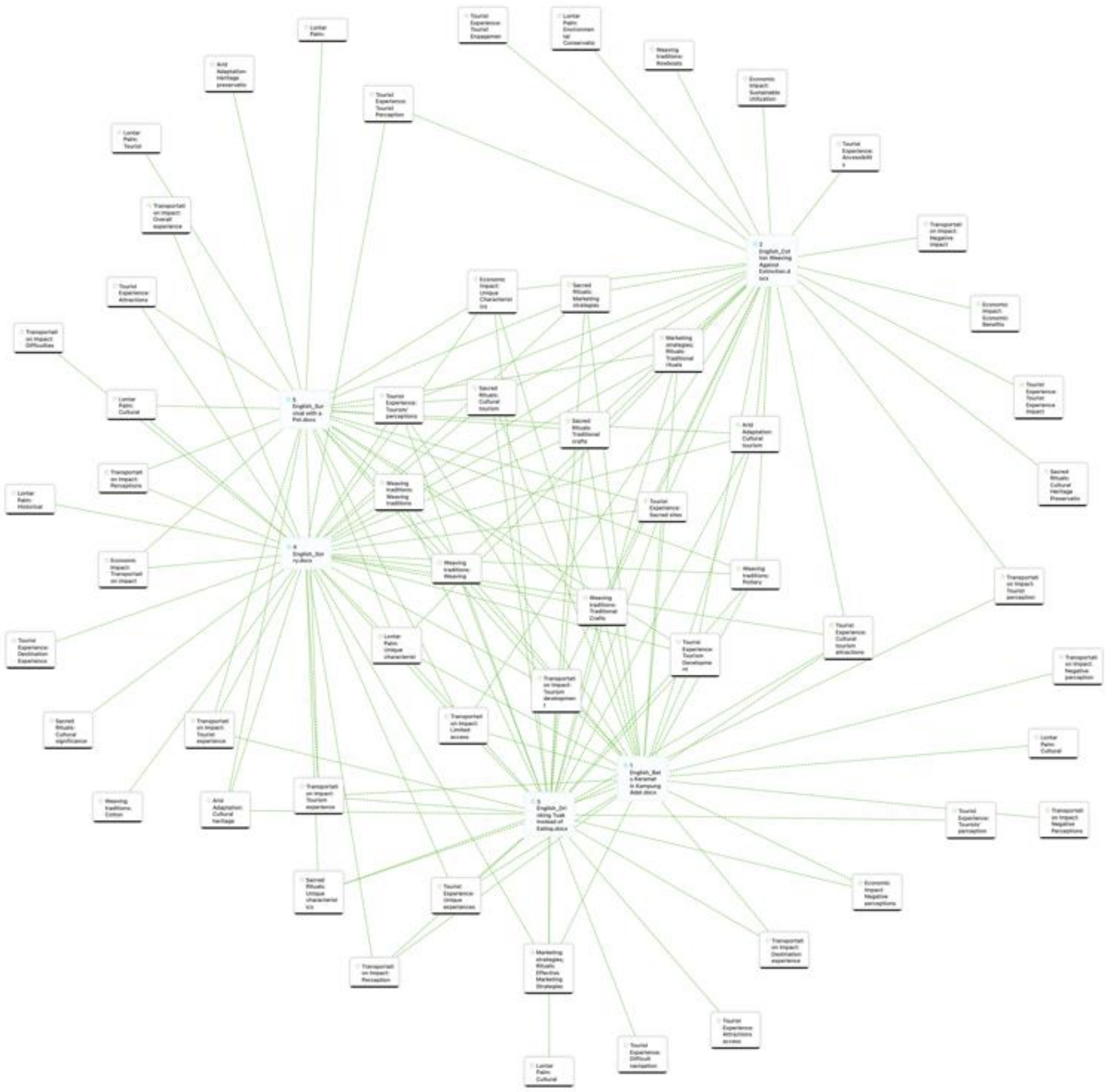


Figure 10. Network Visualization of the Coding Result

Figure 10 shows the network visualization of the coding result. The network visualization of the coding result comprehensively illustrates the interconnections among various themes and concepts identified in the analysis. This visualization captures the complexity and multidimensionality of the narrative structure, revealing how different topics, such as cultural traditions, economic impacts, and environmental adaptations, are interrelated within the broader context. Each node represents a distinct theme, while the connecting lines indicate the relationships and interactions between these themes, highlighting the integrated nature of the discourse. The density of the connections suggests a high level of thematic interdependence, where multiple aspects of the narrative influence and inform each other. This interconnectedness is crucial for understanding the holistic nature of the subject matter, emphasizing that the themes do not exist in isolation but rather as parts of a cohesive whole. Such a visualization aids in identifying critical focal points within the narrative, allowing for a more in-depth exploration of how these themes collectively contribute to the overall understanding of the topic.

The complexity of the network among the codes necessitates a strategic selection of key themes for discussion, with particular emphasis on Economic Impact, Sacred Rituals, Tourist Experience, and Weaving Traditions. These selected nodes are central to the overall narrative and exhibit significant interconnections with other themes, making them pivotal in understanding the broader context. The focus on Economic Impact reflects the critical role of financial considerations in shaping cultural practices. At the same time, Sacred Rituals underscore the deep-rooted spiritual elements that influence and are influenced by these economic factors. The Tourist Experience is closely linked with economic outcomes and the preservation or transformation of cultural practices, bridging local traditions and global interactions. Finally, Weaving Traditions represent a tangible manifestation of cultural heritage, directly impacted by economic and tourist activities, while also embodying the spiritual and ceremonial aspects of the culture. The selection of these themes for further discussion allows for a nuanced exploration of how economic forces, cultural practices,

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

and external influences interact within this complex network, providing deeper insights into the sustainability and evolution of cultural heritage.

3. RESULT AND DISCUSSION

The discussion in this research is divided into two key sections: examining the toxicity score calculations and sentiment classification results, followed by a content analysis framed within the perspective of tourism and cultural heritage. The first section delves into the quantitative assessment of user interactions, highlighting the levels of toxicity and sentiment within the discourse, which provides insights into the emotional tone and potential conflicts within the narrative. This analysis is crucial for understanding the digital environment surrounding the content, revealing patterns of engagement that may affect the perception of cultural topics. The second section shifts focus to a qualitative analysis, where the content is dissected through the lens of tourism and cultural heritage. This approach allows for a deeper exploration of how cultural narratives are constructed, maintained, or transformed in the context of economic and social influences brought about by tourism. By integrating these two analytical perspectives, the discussion offers a comprehensive understanding of the interplay between digital discourse and the preservation or evolution of cultural heritage in the face of tourism.

3.1 Toxicity and Sentiment Analysis

Integrating toxicity and sentiment analysis is crucial in content analysis, as it offers a comprehensive understanding of the emotional tone and the potential harmfulness within digital discourse. Toxicity analysis specifically identifies and quantifies harmful and potentially damaging language, such as insults, threats, and identity attacks, which can significantly influence the overall atmosphere of a conversation and affect the well-being of participants. Meanwhile, sentiment analysis provides insights into the general emotional direction of the content, whether positive, negative, or neutral, thereby revealing underlying attitudes and perceptions. Together, these analyses allow for a nuanced evaluation of how audiences perceive and engage with content. By understanding the interplay between sentiment and toxicity, researchers can better assess the impact of digital narratives, identify areas where intervention might be necessary to promote healthier interactions and enhance the quality and inclusivity of online communities. This dual approach is essential for making informed decisions in content moderation, policy-making, and developing strategies to foster constructive dialogue in digital spaces.

Toxicity analysis is a critical tool in examining digital content, providing a means to identify and measure the presence of harmful or disruptive language within online interactions. This analysis focuses on detecting various forms of negative communication, such as insults, threats, identity attacks, and profanity, which can degrade the quality of discourse and contribute to a hostile environment. By quantifying the extent of such toxicity, researchers gain valuable insights into the dynamics of digital conversations and the impact of toxic language on the broader community. The application of toxicity analysis is significant in moderating content and developing strategies to mitigate harmful behaviors, thereby promoting a safer and more respectful online space. Furthermore, understanding the patterns and triggers of toxic interactions can inform the creation of more effective interventions, ensuring that digital platforms foster constructive and inclusive dialogue rather than becoming arenas for negativity and conflict.

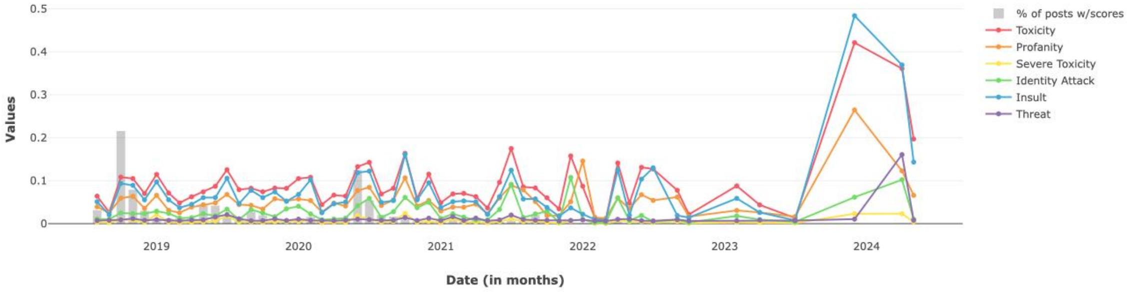


Figure 11. Average Toxicity Score (Content id Zrvgcm5jO9U)

Figure 11 shows the average toxicity score of video content ID s0IXDJw4yM8. Identifying and analyzing toxicity scores within the dataset reveal significant variations in the prevalence of harmful language across different categories. The average toxicity score stands at 0.09886, with a peak value reaching 0.83647, indicating that while the overall discourse remains relatively moderate, there are instances of extreme toxicity. Severe toxicity is less common, with an average score of 0.00853, but notable peaks occur with values as high as 0.45727. Identity attacks, which can profoundly impact the targeted individuals or groups, average 0.02786, with a maximum score of 0.49117, reflecting isolated but impactful occurrences. The insult category shows an average of 0.08239, with a high of 0.78565, demonstrating a significant presence of derogatory language. Profanity is also prevalent, with an average score of 0.05689 and a peak of 0.64460, highlighting the use of coarse language within the comments. Lastly, while infrequent, threats have an average score of 0.00988, with the most severe cases scoring 0.34804. These findings underscore the

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

importance of ongoing content moderation to mitigate the impact of toxic language, ensuring a healthier and more respectful digital environment.

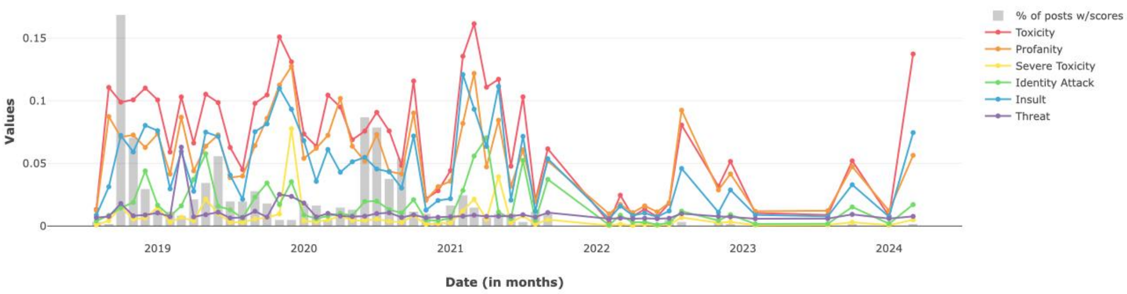


Figure 12. Average Toxicity Score (Content id s0IXDJw4yM8)

Figure 12 shows the average toxicity score of video content ID s0IXDJw4yM8. The toxicity analysis of the video with ID s0IXDJw4yM8 reveals a varied range of negative interactions within the comment section, as reflected in the dataset's average and highest values across multiple toxicity categories. The average toxicity score is 0.08581, peaking at 0.69899, indicating occasional spikes in harmful language usage. Severe toxicity, while less prevalent, shows an average of 0.00855, reaching a maximum of 0.43794, signaling isolated but significant instances of highly negative behavior. Identity attacks, which can profoundly impact individuals or groups, have an average score of 0.01934 and a peak value of 0.71679, suggesting that although these occurrences are relatively rare, they can be particularly intense. The insult category averages 0.05760, with a maximum of 0.69911, indicating a moderate presence of derogatory remarks. Profanity appears with an average score of 0.06246 and a peak of 0.60019, highlighting the use of coarse language within the discourse. Finally, the threat category, with an average of 0.01095 and a maximum score of 0.53305, underscores the presence of occasional but concerning threats in the interactions. These findings highlight the need for targeted moderation efforts to address and mitigate the impact of toxic behavior in digital spaces.

The subsequent evaluation of algorithm performance in classification tasks has demonstrated that the Support Vector Machine (SVM) algorithm consistently outperforms other models. This superior performance is reflected in its higher accuracy, precision, recall, and overall robustness across various datasets. The SVM's ability to effectively separate classes, even in complex and high-dimensional spaces, underscores its strength in handling the intricacies of the data. Moreover, the stability of SVM in producing reliable results across different evaluation metrics further solidifies its position as the most suitable algorithm for the classification tasks at hand. This conclusion is based on a comprehensive analysis of its performance metrics. It consistently shows that SVM excels in capturing the underlying patterns within the data, making it the preferred choice for applications requiring high precision and accuracy. Consequently, the adoption of SVM in similar contexts is strongly recommended, given its proven efficacy and reliability in classification.

The performance of the Support Vector Machine (SVM) algorithm without applying SMOTE reveals a balanced yet nuanced classification capability. The model achieved an accuracy of 70.39%, reflecting its moderate effectiveness in correctly classifying instances across the dataset. The confusion matrix indicates that while the model accurately classified 888 positive instances, it also misclassified 380 positive instances as unfavorable, highlighting areas for potential improvement in handling class imbalances. The AUC scores present a mixed picture: an optimistic AUC of 0.871 suggests strong performance under favorable conditions. However, a significant drop to 0.633 for the overall AUC and 0.395 for the pessimistic scenario indicates difficulty distinguishing between classes. Precision stands at 70.03%, coupled with a high recall of 96.94%, suggesting that the model is more adept at capturing positive cases, though this comes at the cost of an increased false-positive rate. The F-measure of 81.31% confirms that, despite some limitations, the SVM maintains a good balance between precision and recall, making it a reliable choice for classification tasks. However, further optimization may be needed to enhance performance, particularly in scenarios with significant class imbalances.

The performance of the Support Vector Machine (SVM) algorithm, when enhanced with SMOTE, demonstrates a trade-off between accuracy and the model’s ability to handle class imbalances effectively. The model achieved an accuracy of 66.43%, slightly lower than without SMOTE, indicating a decrease in overall correctness due to the introduction of synthetic minority oversampling. However, this adjustment significantly improved the model’s ability to differentiate between classes, as reflected in the optimistic AUC of 0.904 and an overall AUC of 0.816, highlighting enhanced discriminatory power across various scenarios. The pessimistic AUC of 0.728 further supports the model’s robustness even in less favorable conditions. Despite a moderate precision of 60.51%, the recall rate of 94.98% indicates that the model is susceptible to detecting positive instances, although this sensitivity results in a higher number of false positives. The F-measure of 73.90% suggests a balanced performance, integrating both precision and recall, making the SVM with SMOTE a powerful tool for tasks requiring effective class differentiation, particularly in datasets with significant imbalances. It makes the SVM with SMOTE a valuable choice where sensitivity to minority classes is crucial, even at the expense of some accuracy.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

The toxicity and sentiment classification analysis reveals a nuanced landscape of digital interactions where harmful language and emotional tone play critical roles in shaping the discourse. The toxicity analysis highlights harmful elements such as insults, identity attacks, and profanity, with varying degrees of severity across different posts, suggesting a need for careful moderation to maintain a respectful environment. Concurrently, sentiment classification provides insights into the emotional context, illustrating how positive, negative, or neutral sentiments dominate the conversation. The intersection of these two analyses underscores the complexity of online communication, where sentiment and toxicity often coexist, influencing the overall perception and reception of content. This dual approach enhances the understanding of the digital discourse and informs strategies for fostering healthier, more constructive interactions. The findings indicate that managing sentiment and toxicity is crucial for creating a balanced and inclusive online space, ensuring diverse voices can engage meaningfully without the detriments of harmful language.

3.2 Content Analysis: Tourism and Cultural Heritage

Content analysis is a crucial complement to interpreting toxicity scores and sentiment classification, providing a deeper understanding of the underlying themes and narratives that drive digital interactions. While toxicity analysis identifies the presence and intensity of harmful language, and sentiment classification captures the emotional tone of the discourse, content analysis delves into the substantive elements of the conversation, revealing the topics and ideas that are most salient to the participants. This approach allows for a more holistic interpretation, where the context and meaning behind the toxic or emotionally charged language are made clear, offering insights into why particular sentiments prevail and how they are expressed. By integrating content analysis with toxicity and sentiment metrics, the study gains a more nuanced perspective on the data, enabling more informed conclusions about the nature of the discourse and the factors influencing it. This comprehensive analysis not only enhances the accuracy of the findings but also informs strategies for improving the quality and direction of future digital interactions.

The coding and selection of relevant topics in the content analysis, aligned with the themes of tourism and cultural heritage from KOMPASTV's expedition in Sabu Island, have identified several critical areas of interest: Economic Impact, Sacred Rituals, Tourist Experience, and Weaving Traditions. These topics encapsulate the multifaceted relationship between tourism and cultural preservation and highlight how local traditions are influenced by and interact with external economic forces. The focus on Economic Impact reveals how tourism serves as both a boon and a challenge for local communities, affecting livelihoods and the sustainability of cultural practices. Sacred Rituals, meanwhile, underscore the deep spiritual connections that continue to play a vital role in the community's cultural identity, often juxtaposed against the commodification pressures of tourism. Tourist Experience offers insights into how visitors perceive and engage with the local culture, shaping the narrative and authenticity of the cultural heritage being presented. Lastly, Weaving Traditions represent a tangible expression of cultural heritage, where the intersection of traditional craftsmanship and market demands reveals broader implications for cultural continuity. These topics collectively offer a rich foundation for further exploration and discussion, particularly in understanding the dynamic balance between preserving cultural heritage and embracing the economic opportunities presented by tourism.

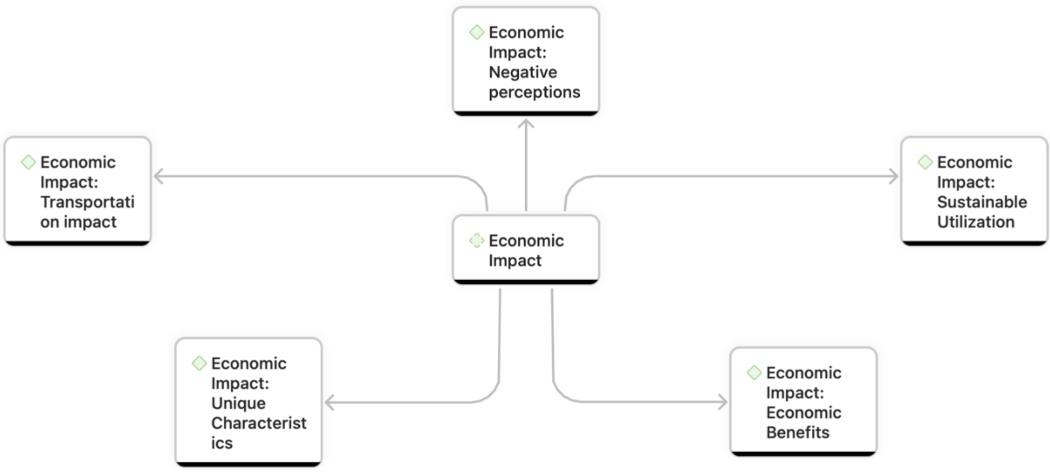


Figure 13. Economic Impact

Figure 13 shows the economic impact network based on the documentary video of KOMPASTV in Sabu Island. In the context of economic impact, the analysis reveals a multifaceted influence of tourism on the local economy, which encompasses both positive and negative dimensions. On one hand, the economic benefits brought by tourism are evident in the increased opportunities for income generation and the sustainable utilization of resources. These benefits, however, are tempered by negative perceptions and challenges related to the unique characteristics of

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

the local economy, such as the impact on transportation and the potential for exploitation or overuse of resources. The sustainable utilization of economic benefits highlights the need for balanced approaches that ensure long-term prosperity without compromising the cultural and environmental integrity of the region. Meanwhile, the impact on transportation underscores the infrastructural challenges accompanying increased tourist activity, which can strain local resources and alter the traditional modes of living. This complex interplay between economic advantages and the accompanying challenges necessitates careful planning and policy-making to maximize the positive outcomes of tourism while mitigating its adverse effects on the local economy and cultural heritage.

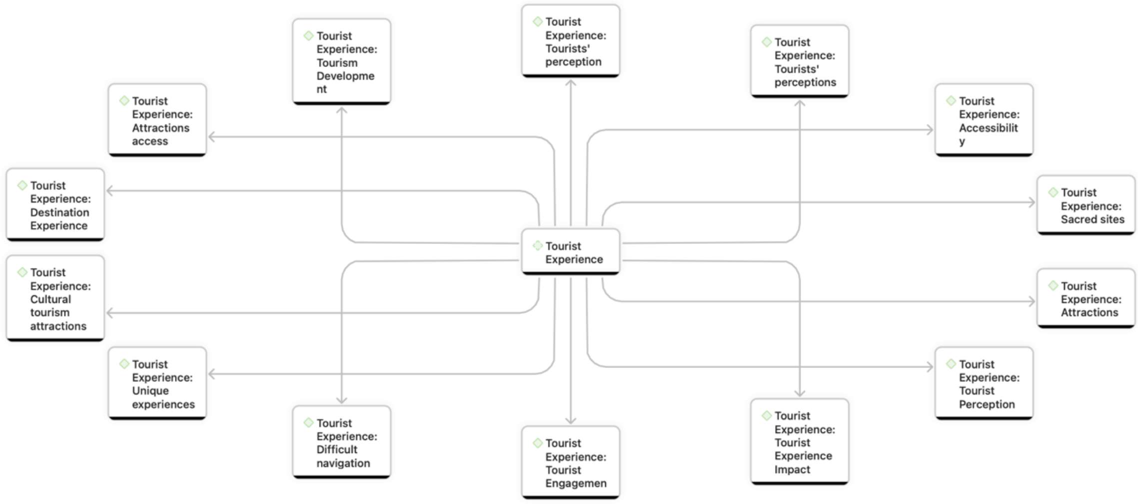


Figure 14. Tourist Experience

Figure 14 shows the tourist experience network based on the documentary video of KOMPASTV on Sabu Island. In the context of tourist experience, the analysis emphasizes the multifaceted nature of how visitors engage with and perceive a destination, particularly in regions with rich cultural heritage. The tourist experience is shaped by factors such as accessibility to attractions, the uniqueness of cultural sites, and the overall ease of navigation within the destination. Additionally, tourists' perceptions are influenced by their encounters with developed and less accessible sites, including sacred locations with significant cultural or spiritual importance. The interaction between tourists and the local environment and the level of engagement with cultural practices play a critical role in shaping the overall experience. However, challenges such as complex navigation or limited access to certain attractions can detract from the experience, highlighting the need for careful management of tourism development. Ultimately, the impact of tourist experiences extends beyond individual satisfaction, influencing the broader perception of the destination and contributing to the ongoing dialogue between preserving cultural heritage and accommodating the needs of modern tourism.

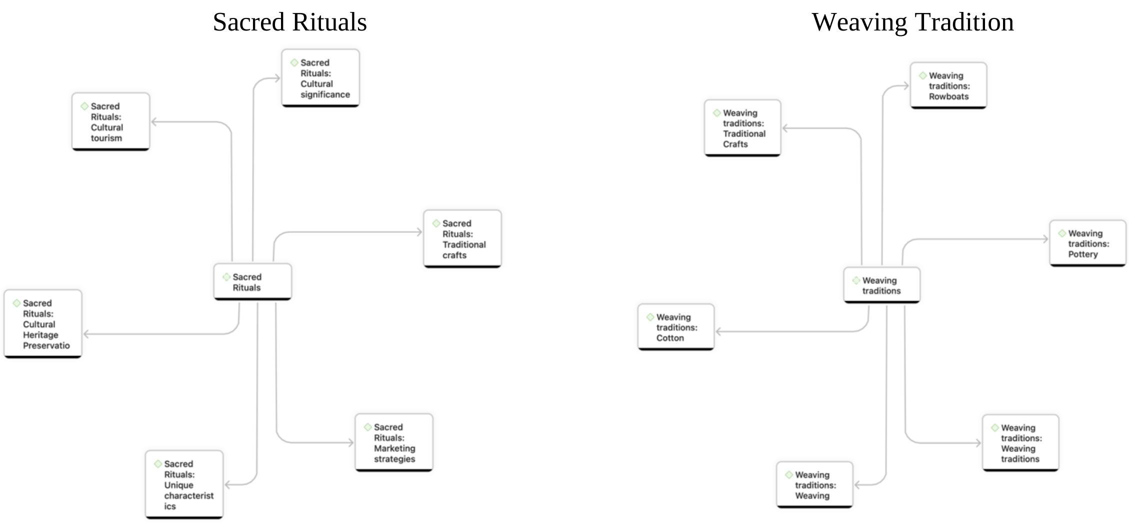


Figure 15. Sacred Rituals and Weaving Tradition

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

Figure 15 shows the sacred rituals and weaving tradition network based on the documentary video of KOMPASTV in Sabu Island. In the context of sacred rituals, the analysis highlights the intricate balance between cultural preservation and the pressures of modern tourism. Sacred rituals, deeply embedded in the cultural fabric, hold significant traditional and spiritual values that are often at the core of a community's identity. These rituals are vital for cultural heritage preservation and serve as unique characteristics that differentiate the culture in the global tourism market. However, the intersection of sacred rituals with cultural tourism introduces challenges, particularly in maintaining the authenticity and sanctity of these practices while making them accessible to tourists. The commodification of sacred rituals through marketing strategies and including traditional crafts can lead to these practices' preservation and potential dilution. Therefore, understanding the cultural significance of sacred rituals and implementing strategies that respect their unique characteristics is essential for ensuring that cultural tourism contributes positively to safeguarding intangible cultural heritage rather than undermining it. The delicate balance between cultural significance and commercial appeal must be managed carefully to sustain the integrity of sacred rituals in the face of increasing tourism.

In the context of weaving traditions, the analysis underscores the importance of this cultural practice as a cornerstone of traditional crafts and community identity. Weaving, mainly cotton and other natural materials represents a rich cultural heritage and embodies the skills and artistry passed down through generations. This tradition extends beyond the act of weaving itself, encompassing related crafts such as pottery and the construction of traditional rowboats, which are often intricately connected to the same cultural narratives. These practices reflect the community's relationship with its environment, utilizing locally sourced materials and techniques that have been refined over centuries. Preserving weaving traditions is crucial, as it sustains the tangible and intangible aspects of cultural heritage, including the knowledge systems, social structures, and economic activities tied to these crafts. However, integrating these traditions into the tourism sector poses challenges, such as ensuring authenticity while adapting to market demands. Balancing the preservation of traditional techniques with the need to innovate for economic viability is essential to maintaining the cultural significance of weaving traditions in contemporary society.

The content analysis conducted in this study provides a comprehensive understanding of the intricate relationship between cultural heritage and tourism, revealing key themes central to both the preservation and commercialization of cultural practices. Examining economic impact, sacred rituals, tourist experience, and weaving traditions highlights the complex dynamics at play, where cultural preservation efforts must be balanced against the pressures of economic development and the demands of global tourism. The analysis underscores the importance of maintaining cultural integrity while adapting to external influences, ensuring that the commodification of cultural practices does not erode their authenticity or significance. Moreover, the findings suggest that thoughtful integration of cultural heritage into the tourism sector can be a powerful tool for economic growth and cultural preservation, provided strategies are carefully designed to respect and sustain the unique characteristics of the showcased culture. Thus, the content analysis illuminates the challenges faced in this intersection and offers insights into potential pathways for fostering a more sustainable and culturally sensitive approach to tourism.

4. CONCLUSION

The research underscores the media's dual role in preserving and potentially distorting cultural heritage through its portrayal of Sabu Island in KOMPASTV's expedition documentary. The study meticulously dissected the documentary's content using the Digital Content Reviews and Analysis Framework, revealing significant insights into the complex interplay between tourism, cultural preservation, and media representation. The framework's integration of sentiment and toxicity analysis provided a nuanced understanding of digital narratives' emotional tone and potential harmfulness. The toxicity analysis revealed an average score of 0.09886, with the highest toxicity score reaching 0.83647, indicating the presence of harmful language that, although not pervasive, could significantly influence the discourse surrounding cultural heritage. Furthermore, the research employed a Support Vector Machine (SVM) model to evaluate sentiment classification. The SVM model, enhanced by SMOTE (Synthetic Minority Over-sampling Technique), demonstrated strong performance with an accuracy of 66.43%, precision of 60.51%, recall of 94.98%, and an F-measure of 73.90%. The Area Under the Curve (AUC) for the SVM model was particularly notable, with an optimistic AUC of 0.904, an overall AUC of 0.816, and a pessimistic AUC of 0.728, underscoring the model's robustness in classifying sentiment even in complex and imbalanced datasets. In addition, the content analysis focused on key themes such as Economic Impact, Sacred Rituals, Tourist Experience, and Weaving Traditions. These themes illuminated the complex dynamics at play, where cultural preservation efforts must be balanced against the pressures of economic development and the demands of global tourism. The analysis highlighted how tourism can positively and negatively influence local economies, cultural practices, and community identities. For instance, the Economic Impact theme revealed tourism's benefits and challenges. At the same time, the exploration of Sacred Rituals underscored the tension between maintaining cultural authenticity and accommodating tourist expectations. The themes of the tourist experience and weaving traditions further illustrate how cultural heritage is adapted and sometimes commodified in response to external influences. These findings collectively highlight the necessity for responsible and authentic media portrayals to safeguard cultural identities, as media has the power to either uphold or undermine the integrity of cultural narratives. The research illustrates the importance of a balanced strategy that leverages the benefits of tourism-driven media while mitigating its potential negative impacts on local communities. This study thus contributes to the broader discourse on cultural heritage documentation by providing a comprehensive framework for evaluating the influence of digital narratives on the preservation of cultural identities, ensuring that cultural heritage is portrayed with accuracy and respect.

ACKNOWLEDGMENT

Thanks to the to the Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, PUSDIPAR, and the LPPM (*Lembaga Penelitian dan Pengabdian kepada Masyarakat*).

REFERENCES

[1] M. D. Sánchez-Sánchez, C. De-Pablos-Heredero, and J. L. Montes-Botella, “A behaviour model for cultural tourism: loyalty to destination,” *Econ. Res. Istraz.* , vol. 34, no. 1, pp. 2729–2746, 2021, doi: 10.1080/1331677X.2020.1838313.
[2] S. Chai-Arayalert, “Smart application of learning ecotourism for young eco-tourists,” *Cogent Soc. Sci.*, vol. 6, no. 1, 2020, doi: 10.1080/23311886.2020.1772558.
[3] C. Rinaldi, A. Cavicchi, and R. N. S. Robinson, “University contributions to co-creating sustainable tourism destinations,” *J. Sustain. Tour.*, vol. 30, no. 9, pp. 2144–2166, 2022, doi: 10.1080/09669582.2020.1797056.
[4] K. Krittayaruangroj, S. Suriyankietkaew, and P. Hallinger, “Research on sustainability in community-based tourism: a bibliometric review and future directions,” *Asia Pacific J. Tour. Res.*, vol. 28, no. 9, pp. 1031–1051, 2023, doi: 10.1080/10941665.2023.2276477.
[5] J. Wu, X. Font, and J. Liu, “Failure to morally self-regulate: deindividuation and moral disengagement from pro-environmental behavioural intentions across tourism and home contexts,” *Curr. Issues Tour.*, no. May, pp. 1–20, 2024, doi: 10.1080/13683500.2024.2355559.
[6] F. Carbone, “‘Don’t look back in anger’. War museums’ role in the post conflict tourism-peace nexus,” *J. Sustain. Tour.*, vol. 30, no. 2–3, pp. 565–583, 2022, doi: 10.1080/09669582.2021.1901909.
[7] 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.
[8] W. Megarry and K. Hadick, “Lessons from the Edge: Assessing the impact and efficacy of digital technologies to stress urgency about climate change and cultural heritage globally,” *Hist. Environ. Policy Pract.*, vol. 12, no. 3–4, pp. 336–355, 2021, doi: 10.1080/17567505.2021.1944571.
[9] G. Pasternak, “Photographic Digital Heritage in Cultural Conflicts: A Critical Introduction,” *Photogr. Cult.*, vol. 14, no. 3, pp. 253–268, 2021, doi: 10.1080/17514517.2021.1953763.
[10] S. H. Yim, “Cultural heritage through the lens of community psychology and narrative therapy: a community project on Chinese and Vietnamese diaspora in London,” *Int. J. Herit. Stud.*, vol. 28, no. 8, pp. 970–983, Aug. 2022, doi: 10.1080/13527258.2022.2108486.
[11] 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.
[12] R. Cameron, “‘An Oriental Holiday’: constructing Bosnia and Herzegovina as a destination in British tourist literature, c. 1890–1914,” *J. Tour. Hist.*, vol. 15, no. 3, pp. 233–256, 2023, doi: 10.1080/1755182X.2023.2235325.
[13] M. N. Ahmad Taha and D. A. Abdelfattah, “Branding heritage in multi-cultural societies, with special references to preservation discourses,” *HBRC J.*, vol. 19, no. 1, pp. 337–354, 2023, doi: 10.1080/16874048.2023.2272076.
[14] 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, 2024, doi: 10.1080/1743873X.2023.2255691.
[15] R. Zengeya, P. W. Mamimine, and M. C. Mwando, “Diaspora based tourism marketing conceptual paper: A conceptual analysis of the potential of harnessing the diaspora to improve tourism traffic in Zimbabwe,” *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2164994.
[16] R. Schiavone, S. Reijnders, and A. Brandellero, “‘Beneath the storyline’: analysing the role and importance of film in the preservation and development of Scottish heritage sites,” *Int. J. Herit. Stud.*, vol. 28, no. 10, pp. 1107–1120, 2022, doi: 10.1080/13527258.2022.2131876.
[17] 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.
[18] E. McLoughlin and J. Hanrahan, “Evidence-informed planning for tourism,” *J. Policy Res. Tour. Leis. Events*, vol. 15, no. 1, pp. 1–17, 2023, doi: 10.1080/19407963.2021.1931257.
[19] A. Firman, M. Moslehpour, R. Qiu, P. K. Lin, T. Ismail, and F. F. Rahman, “The impact of eco-innovation, ecotourism policy and social media on sustainable tourism development: evidence from the tourism sector of Indonesia,” *Econ. Res. Istraz.* , vol. 36, no. 2, p., 2023, doi: 10.1080/1331677X.2022.2143847.
[20] T. Rasul *et al.*, “Immersive virtual reality experiences: boosting potential visitor engagement and attractiveness of natural world heritage sites,” *Asia Pacific J. Tour. Res.*, vol. 29, no. 5, pp. 515–526, 2024, doi: 10.1080/10941665.2024.2332993.
[21] G. M. Andreou *et al.*, “Big Data in Maritime Archaeology: Challenges and Prospects from the Middle East and North Africa,” *J. F. Archaeol.*, vol. 47, no. 3, pp. 131–148, Apr. 2022, doi: 10.1080/00934690.2022.2028082.
[22] Z. F. Zeadat, “Adaptive reuse challenges of Jordan’s heritage buildings: a critical review,” *Int. J. Urban Sustain. Dev.*, vol. 16, no. 1, pp. 95–107, Dec. 2024, doi: 10.1080/19463138.2024.2329661.
[23] J. M. Cheer, M. Mary, and A. A. Lew, “Cultural ecosystem services and placemaking in peripheral areas: a tourism

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1092−1110
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5838

geographies agenda," *Tour. Geogr.*, vol. 24, no. 4–5, pp. 495–500, 2022, doi: 10.1080/14616688.2022.2118826.

[24] F. Pezeshki, M. Khodadadi, and M. Bagheri, "Investigating community support for sustainable tourism development in small heritage sites in Iran: A grounded theory approach," *Int. J. Herit. Stud.*, vol. 29, no. 8, pp. 773–791, 2023, doi: 10.1080/13527258.2023.2220316.

[25] N. B. Lwoga and E. Maturo, "Motivation-based segmentation of rural tourism market in African villages," *Dev. South. Afr.*, vol. 37, no. 5, pp. 773–790, 2020, doi: 10.1080/0376835X.2020.1760791.

[26] J. Saarinen, N. Moswete, J. R. Atlhopheng, and W. L. Hambira, "Changing socio-ecologies of Kalahari: Local perceptions towards environmental change and tourism in Kgalagadi, Botswana," *Dev. South. Afr.*, vol. 37, no. 5, pp. 855–870, 2020, doi: 10.1080/0376835X.2020.1809997.

[27] Riyanto, M. Iqbal, Supriono, M. R. A. Fahmi, and E. S. Yuliaji, "The effect of community involvement and perceived impact on residents' overall well-being: Evidence in Malang marine tourism," *Cogent Bus. Manag.*, vol. 10, no. 3, pp. 1–20, 2023, doi: 10.1080/23311975.2023.2270800.

[28] 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.

[29] M. C. Lambán, R. P. Vidal, and M. Z. Antón, "Mapping the 'Magic of Huesca': a methodological proposal for the design of tourist cartography," *J. Maps*, vol. 19, no. 1, 2023, doi: 10.1080/17445647.2022.2141142.

[30] B. Isakhan and L. Meskell, "Local perspectives on heritage reconstruction after conflict: a public opinion survey of Aleppo," *Int. J. Herit. Stud.*, vol. 30, no. 7, pp. 821–839, Jul. 2024, doi: 10.1080/13527258.2024.2342288.

[31] 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.

[32] M. M. Phori, U. P. Hermann, and L. Grobbelaar, "Residents' perceptions of sustainable heritage tourism development in a rural municipality," *Dev. South. Afr.*, vol. 41, no. 3, pp. 551–569, 2024, doi: 10.1080/0376835X.2024.2350391.

[33] S. Boyd, M. V. Reddy, S. Kulshreshtha, and M. Nica, "Post-conflict tourism opportunity spectrum (POCTOS): a framework for destinations recovering from conflict," *J. Sustain. Tour.*, vol. 31, no. 1, pp. 131–148, 2023, doi: 10.1080/09669582.2021.1993866.

[34] E. Mele, P. Kerkhof, and L. Cantoni, "Analyzing cultural tourism promotion on Instagram: a cross-cultural perspective," *J. Travel Tour. Mark.*, vol. 38, no. 3, pp. 326–340, 2021, doi: 10.1080/10548408.2021.1906382.

[35] K. T. T. Nguyen, L. Murphy, T. Chen, and P. L. Pearce, "Let's listen: the voices of ethnic villagers in identifying host-tourist interaction issues in the Central Highlands, Vietnam," *J. Herit. Tour.*, vol. 19, no. 2, pp. 263–286, 2024, doi: 10.1080/1743873X.2023.2259512.

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

[37] M. C. Puche-Ruiz, "Flamboyance and wit. The promotion of film-induced tourism and Andalusian-inspired 'brand Spain' under the Ministry of Information and Tourism (1953-1959)," *J. Tour. Hist.*, vol. 14, no. 2, pp. 167–201, 2022, doi: 10.1080/1755182X.2022.2118377.

[38] E. Evagorou, C. Mettas, A. Agapiou, K. Themistocleous, S. Papavasileiou, and D. Hadjimitsis, "Digital camera calibration for cultural heritage documentation: the case study of a mass digitization project of religious monuments in Cyprus," *Eur. J. Remote Sens.*, vol. 54, no. sup1, pp. 6–17, Feb. 2021, doi: 10.1080/22797254.2020.1810131.

[39] 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.

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