Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

Comprehensive Analysis of Sentiment Classification and Toxicity Assessment in Cultural Documentary Videos

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: 07/05/2024; Accepted: 29/05/2024; Published: 30/05/2024

Abstract−This research explores sentiment classification and toxicity assessment in cultural documentary videos through a systematic analysis framework based on the Cross-Industry Standard Process for Data Mining (CRISP-DM). The study evaluates the sentiment polarity of viewer comments by utilizing a diverse array of machine-learning algorithms, including k-NN, DT, NBC, and SVM. It identifies toxic language patterns across multiple videos. Additionally, the research employs SMOTE to address class imbalance issues and enhance model performance. The results reveal high accuracy rates ranging from 72.24% to 96.79% in sentiment classification, indicating the effectiveness of the proposed methodology. Moreover, toxicity analysis unveils varying degrees of toxic language prevalence, with toxicity scores ranging from 0.01270 to 0.09334 across different videos. Despite these achievements, the study acknowledges the inherent limitations of toxicity scoring algorithms in capturing contextual nuances. Overall, this research contributes to understanding sentiment dynamics and toxicity trends in cultural documentary content and underscores the importance of employing advanced machine learning techniques within a structured analytical framework for insightful data interpretation and decision-making.

Keywords: Sentiment Classification; Toxicity; Cultural Documentary; Video Reviews; Sumba

1. INTRODUCTION

The evolution of digital technology has revolutionized the creation of cultural documentary videos, thereby serving as a communication medium that enhances the comprehension of cultural intricacies and community norms among indigenous societies [1]–[5]. This technological advancement facilitates the production of high-quality videos that capture the essence of cultural practices, rituals, and traditions with unprecedented clarity and depth [6]–[8]. Such documentaries offer a window into the rich tapestry of indigenous cultures, fostering cross-cultural understanding and appreciation [9]–[14]. Consequently, they play a pivotal role in preserving and disseminating cultural heritage, contributing to the cultural sustainability of indigenous communities in an increasingly interconnected world.

Within cultural documentary videos focusing on Sumba, Indonesia, disseminated via YouTube, a content analysis approach is employed to discern the primary themes embedded within such content [15], [16]. Through systematic examination and categorization of visual and narrative elements, this analytical method unveils the core subjects addressed within the documentaries, ranging from traditional customs and ceremonies to societal values and historical narratives [17]–[20]. Consequently, content analysis facilitates a nuanced understanding of the cultural intricacies and societal dynamics depicted in Sumba's cultural documentaries, enriching scholarly discourse and fostering appreciation for the region's cultural heritage.

In the context of cultural documentaries concerning Sumba's heritage, viewer comments provide a fertile ground for analysis utilizing toxicity and sentiment analysis methodologies. Through applying these analytical frameworks, viewer responses' tone and emotional valence are systematically evaluated, unveiling positivity, negativity, or toxicity patterns within the discourse surrounding Sumba's cultural representations [21], [22]. Such analyses offer invaluable insights into the audience's perceptions, attitudes, and emotional responses towards the portrayed cultural narratives, thereby contributing to a deeper understanding of the impact and reception of these documentaries within online communities [23]. Consequently, leveraging toxicity and sentiment analysis enriches scholarly inquiry and facilitates a more nuanced exploration of the cultural dynamics in Sumba's documentary discourse.

The research uses content analysis to identify critical topics from the storyboard and narrative of cultural documentaries about Sumba, Indonesia. Furthermore, it seeks to analyze viewer responses through toxicity and sentiment classification using VADER [24]–[26]. By employing these methodological approaches, the study endeavors to uncover the thematic underpinnings of Sumba's cultural narratives as portrayed in documentary videos while shedding light on audiences' emotional and evaluative reactions [27]–[30]. This holistic analysis enhances our comprehension of the cultural representations within the documentaries. It provides insights into the reception and impact of such content within the viewer community, thereby contributing to a more nuanced understanding of the intersection between digital media, cultural representation, and audience engagement.

The urgency of this research lies in its potential to bridge critical gaps in our understanding of cultural representation and audience reception in the context of digital media. By systematically analyzing the thematic content of cultural documentaries on Sumba and the corresponding viewer responses, this study offers insights into cultural preservation, dissemination, and interpretation dynamics in the digital age [31], [32]. Moreover, as

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

indigenous cultures face increasing threats of cultural erosion and misrepresentation, elucidating the mechanisms through which cultural narratives are constructed and received becomes imperative for ensuring the integrity and sustainability of the cultural heritage [33], [34]. Thus, the research addresses an immediate scholarly need and holds broader implications for cultural conservation efforts and digital media practices.

The theoretical and practical implications of this research are multifaceted and far-reaching. By employing content analysis and sentiment classification methodologies to examine cultural documentaries on Sumba and viewer responses, this study contributes to theoretical frameworks concerning cultural representation, digital media, and audience engagement [35]. The findings offer nuanced insights into the construction of cultural narratives, the dynamics of cultural preservation and dissemination, and the impact of digital platforms on audience perceptions and interpretations [36]. Furthermore, the practical implications extend to various stakeholders, including content creators, cultural institutions, and policymakers, who leverage the research findings to inform strategies for cultural documentation, preservation, and outreach in the digital realm. Thus, this research advances scholarly discourse and informs practical endeavors to safeguard and promote cultural heritage in the digital age.

Similar research endeavors focusing on cultural documentaries and audience responses have provided valuable insights into the intersection of digital media and cultural representation [37], [38]. However, limitations persist, particularly regarding the generalizability of findings across diverse cultural contexts and the scalability of methodologies to larger datasets. While existing studies have elucidated the complexities of cultural production and reception in digital environments, further research is warranted to address these limitations and advance our understanding of the multifaceted dynamics shaping contemporary cultural discourse and audience engagement [39], [40]. Thus, while acknowledging the contributions of prior research, this study seeks to build upon existing knowledge and address critical gaps in the literature, thereby enriching scholarly inquiry and informing practical interventions in cultural preservation and digital media.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis

Gap analysis is crucial for identifying discrepancies within the research of similar topics, such as sentiment classification of cultural video reviews using VADER. By scrutinizing existing literature, this research pinpoints areas where knowledge is lacking or incomplete, thereby highlighting opportunities for further investigation and refinement of methodologies. Through this process, gaps in understanding are delineated, paving the way for targeted research efforts aimed at addressing unresolved questions and advancing the field's collective knowledge. Consequently, conducting a comprehensive gap analysis serves as a foundational step in fostering scholarly progress and ensuring that future research endeavors effectively contribute to the evolving discourse surrounding cultural representation and audience reception in digital media.

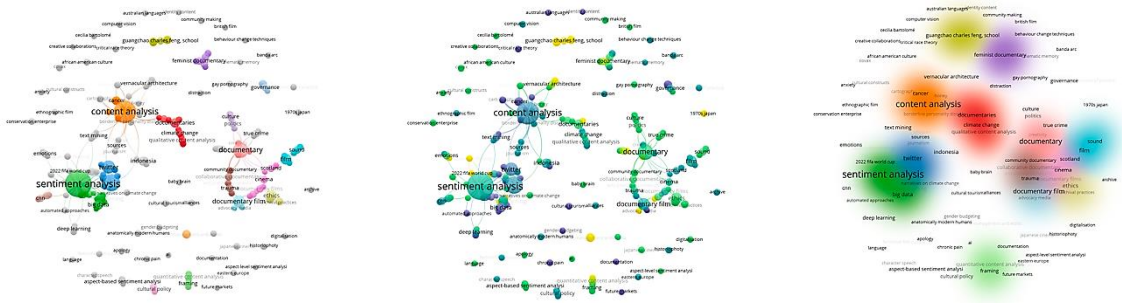


Figure 1. Gap Analysis of Sentiment Classification Using Culture Video Reviews

Figure 1 shows the gap analysis process of the research topic related to sentiment classification of culture video reviews using VosViewer. Based on the results of gap identification, it is evident that the examination of sentiment classification of cultural video reviews using VADER requires further enhancement in terms of quantity. While existing research has laid a foundation for understanding the complexities of audience responses to cultural content in digital media, there remains a paucity of studies that systematically apply sentiment analysis methodologies to various cultural video reviews. This dearth of quantitative analysis limits our ability to draw robust conclusions and identify overarching trends in audience sentiment toward cultural representations. Consequently, expanding the scope of research endeavors to encompass a more extensive and diverse dataset is essential for enriching our understanding of the nuanced dynamics in the reception of cultural media content.

In light of the research context, it becomes evident that the VADER model holds utility in processing textual data for sentiment classification. This model, built upon lexicon-based approaches and machine learning techniques, exhibits proficiency in discerning the emotional valence of text by analyzing lexical features and contextual nuances. Moreover, its accessibility and ease of implementation make it a viable tool for this research

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

seeking to conduct sentiment analysis across diverse datasets and domains. Therefore, within cultural media studies, utilizing the VADER model offers a practical means of systematically examining audience responses and discerning sentiment patterns toward cultural representations in digital media.

2.2 Cross-Industry Standard Process for Data Mining (CRISP-DM)

This research adopts the CRISP-DM framework to analyze content, sentiment, and toxicity scores within Sumba culture video reviews. CRISP-DM, or Cross-Industry Standard Process for Data Mining, provides a structured approach to data analysis, encompassing phases such as business understanding, data understanding, data preparation, modeling, evaluation, and deployment. By applying this framework, this research systematically navigates the complexities of content analysis, sentiment classification, and toxicity assessment within cultural media datasets. Moreover, using CRISP-DM ensures methodological rigor and facilitates the reproducibility of findings, thereby enhancing the credibility and reliability of research outcomes. Thus, leveraging the CRISP-DM framework underscores the commitment to rigorous and systematic inquiry, ultimately contributing to a deeper understanding of audience perceptions and reactions toward cultural representations in digital media.

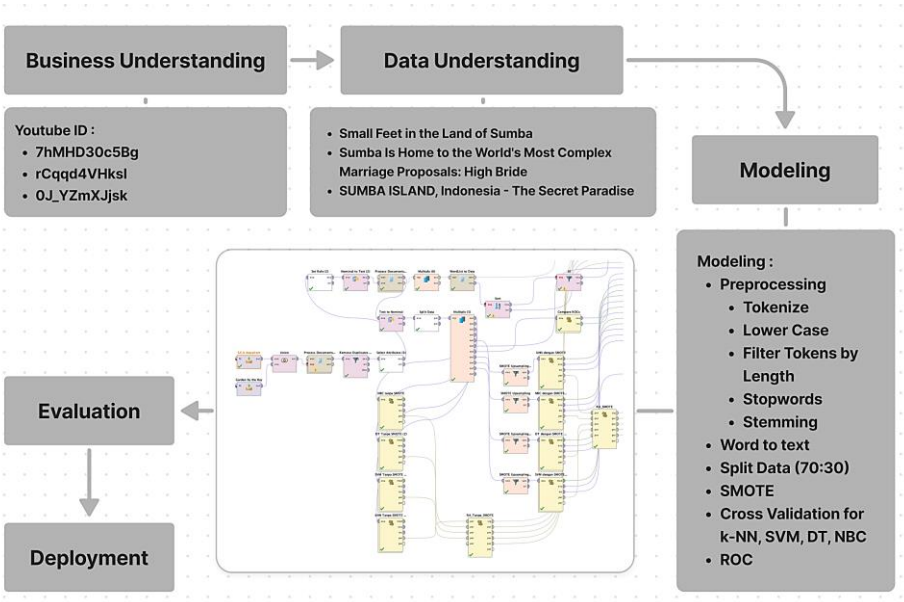


Figure 2. Implementaiton of CRISP-DM Framework

Figure 2 shows the implementation of the CRISP-DM framework for sentiment classification using VADER. The CRISP-DM offers distinct advantages, particularly in its emphasis on contextualizing data processing, which is explicitly addressed during the business understanding phase and in the output of data processing during the deployment phase. The framework's structured approach ensures that data analysis is grounded in a thorough understanding of the business context, including objectives, constraints, and stakeholder requirements, facilitating more informed decision-making and actionable insights. Additionally, by delineating the deployment phase, CRISP-DM ensures that the outcomes of data processing are effectively integrated into operational systems or decision-making processes, maximizing the practical utility of analytical insights. Consequently, the contextual sensitivity and practical applicability inherent in CRISP-DM underscore its efficacy as a framework for guiding data-driven decision-making processes across diverse domains.

2.2.1 Business Understanding

During the business understanding phase, content analysis is conducted to ascertain the topics discussed within the video with the ID rCqqd4VHksI. This phase systematically examines the video's narrative structure, visual elements, and accompanying metadata to delineate the overarching themes and subject matter addressed. This research identifies vital topics, motifs, and cultural nuances in the video by employing content analysis techniques, laying the groundwork for subsequent analyses to understand audience sentiment, toxicity, and overall reception. Consequently, the business understanding phase serves as a crucial precursor to informed decision-making and strategic planning regarding the utilization and interpretation of cultural media content.

Content analysis aims to comprehend the essential topics discussed within documentary videos, stimulating viewers to comment. By systematically dissecting the narrative structure, visual elements, and thematic motifs embedded within the videos, content analysis enables this research to discern the salient subjects and cultural nuances depicted. This process facilitates a deeper understanding of the content's resonance with viewers and sheds light on the factors that prompt audience engagement and discourse. Consequently, content

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

analysis is foundational in elucidating the dynamics of viewer interaction and feedback within cultural media consumption.

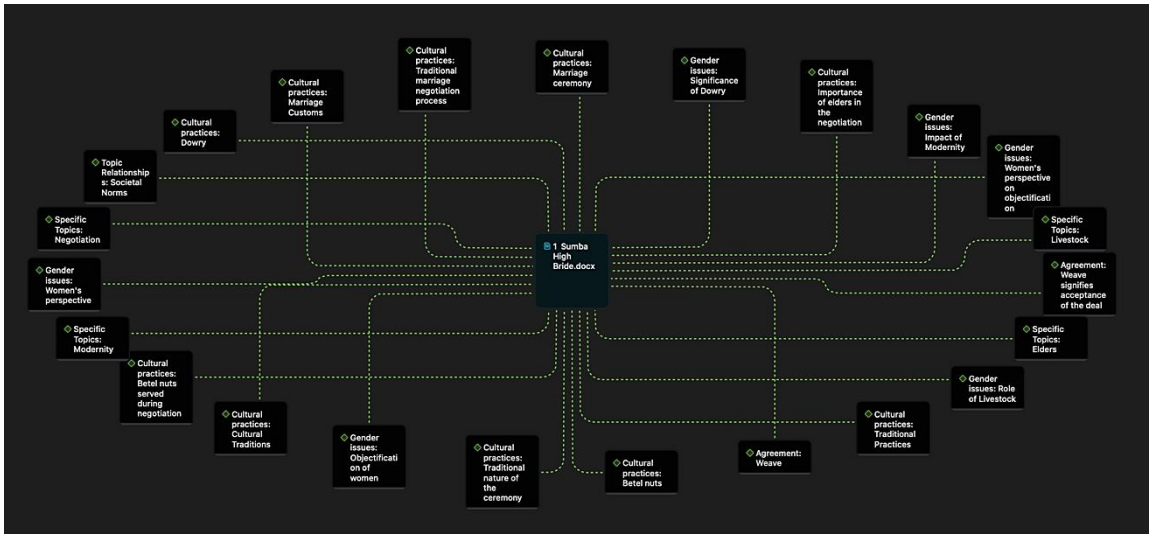


Figure 3. Topics Mentioned in the video (rCqqd4VHksI)

Figure 3 shows the topics mentioned in the Video (rCqqd4VHksI). Based on the topics addressed within the narrative of the cultural documentary on Sumba with the ID rCqqd4VHksI, several key themes emerge, including agreement, cultural practices, gender issues, and topic relationships. The depiction of agreement signifies the presence of shared norms, values, or consensus within the Sumbanese community, reflecting its members' cohesion and collective identity. Cultural practices highlight the diverse rituals, ceremonies, and traditions integral to Sumbanese culture, illustrating the richness and complexity of cultural heritage. Additionally, exploring gender issues sheds light on gender roles, expectations, and power dynamics within Sumbanese society, providing insights into the complexities of gender dynamics in cultural contexts. Lastly, examining topic relationships elucidates the interconnectedness and interdependence of various themes and motifs within the documentary, underscoring the multifaceted nature of Sumbanese culture and the intricate web of social, cultural, and environmental factors that shape it. Thus, the documentary comprehensively portrays Sumba's cultural landscape, highlighting the nuanced intersections of tradition, identity, and societal dynamics.

Specifically, the topics related to the content of the video delineated as follows: Weave, symbolizing acceptance of the deal; Betel nuts, traditionally served during negotiations; Cultural traditions surrounding dowry and the role of elders in negotiations; Marriage customs, including the traditional negotiation process and significance of livestock and dowry; Impact of modernity on traditional practices and societal norms, particularly the objectification of women and changes in negotiation dynamics. Exploring these topics offers a nuanced understanding of the cultural intricacies and social dynamics depicted within the video, highlighting the intersection of tradition, modernity, and gender dynamics in Sumbanese society.

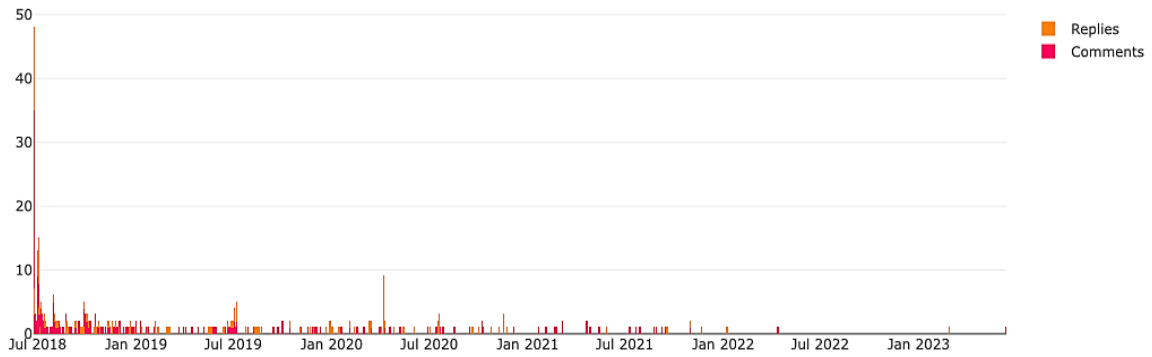


Figure 4. Post-Per-Day Statistic of Video (rCqqd4VHksI)

Figure 4 shows the post-per-day statistic of the video (rCqqd4VHksI). Based on the post-per-day statistics of the video with the ID rCqqd4VHksI, it is evident that the highest level of viewer engagement occurred on June 24, 2018, with 35 comments. Subsequent dates also witnessed notable engagement, albeit to a lesser extent, including July 2nd and 3rd, with 9 and 8 comments, respectively. However, viewer interaction gradually decreased over time, as indicated by fewer comments in the following days and months. This trend suggests an initial surge of interest in the video, followed by a tapering off of engagement over time. Thus, while the video

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

initially captured viewers' attention, sustaining long-term engagement may require ongoing efforts to maintain relevance and interest.

The storyboard of a video documentary and its carefully crafted content catalyze capturing viewer attention and eliciting comments, as evidenced by post-per-day statistics and frequently used words. A well-structured storyboard guides the narrative flow and ensures coherence and engagement, compelling viewers to participate in discussions and share thoughts actively. Moreover, the content's thematic richness and emotive resonance prompt viewers to express reactions and insights, contributing to a vibrant discourse surrounding the cultural themes and societal dynamics depicted in the documentary. Consequently, the interplay between storyboard design, content development, and viewer engagement underscores the pivotal role of storytelling in fostering audience interaction and discourse within the digital media landscape.

2.2.2 Data Understanding

During the data understanding phase, it is essential to identify the data volume and the frequently used words within review data. This initial step lays the groundwork for subsequent analyses by providing insights into the scope and characteristics of the dataset. Understanding the quantity of data allows this research to gauge the breadth and depth of available information while identifying frequently used words that offer valuable clues regarding recurring themes, sentiments, and topics of interest within the dataset. Consequently, this phase is a crucial precursor to comprehensive data analysis, enabling this research to understand the dataset's content and structure.

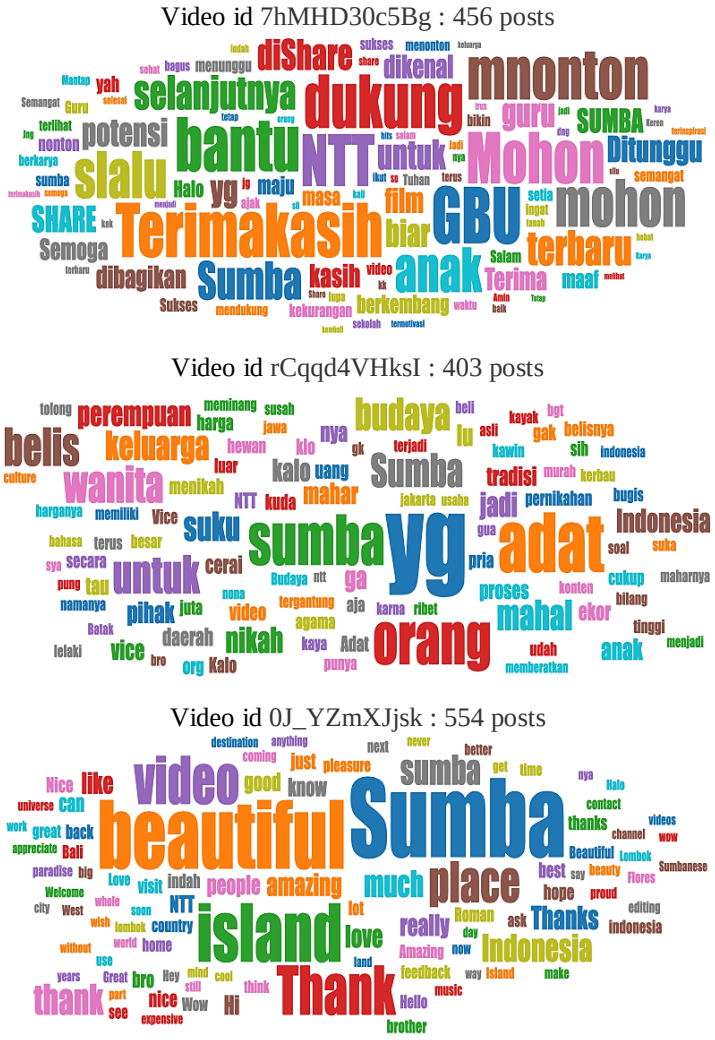


Figure 5. Frequently used Words (rCqqd4VHksI)

Figure 5 shows the frequently used words in the dataset. Based on the frequently used words from the three videos, it is possible to discern the most commonly utilized terms by viewers in documentary reviews. Consequently, these recurring words offer insights into viewers' perceptions and responses toward significant cultural issues portrayed in the Sumba documentary videos. By identifying prevalent themes and topics reflected in viewer feedback, this research interprets audience perspectives and gauges the resonance of cultural themes and narratives depicted in the documentaries. This analysis illuminates the critical points of interest for viewers

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

and informs future content creation and audience engagement strategies to foster deeper connections and understanding of Sumbanese culture.

Viewer reviews must be classified into negative, neutral, and positive classes upon identifying frequently used words using the VADER model. This approach allows for a systematic analysis of sentiment expressed in viewer feedback, enabling this research to discern responses' overall tone and emotional valence towards the documentary content. By categorizing reviews into distinct sentiment classes, the VADER model facilitates a nuanced understanding of audience perceptions and reactions, informing interpretations of viewer engagement and sentiment dynamics within the cultural media landscape. Consequently, leveraging sentiment analysis methodologies such as VADER enhances the depth and precision of audience response analysis, contributing to a more comprehensive assessment of the documentary's impact and resonance among viewers.

2.2.3 Modeling

During the modeling phase, the VADER approach is employed for sentiment classification. However, it is imperative to conduct pilot testing to compare the performance of VADER with TextBlob. While VADER is widely utilized for sentiment analysis due to its lexicon-based approach and ability to handle social media text, TextBlob offers an alternative methodology based on natural language processing techniques. Pilot testing allows for an empirical evaluation of the effectiveness and accuracy of both approaches in classifying sentiment within the context of cultural documentary reviews. Consequently, this comparative analysis enhances the robustness and reliability of sentiment classification methodologies, ultimately contributing to more accurate interpretations of audience perceptions and reactions toward cultural media content.

Video id 7hMHD30c5Bg : 15 out of 456 posts (a)

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	14	2 (14.29%)	1 (7.14%)	11 (78.57%)
TextBlob (English/EN)	14	2 (14.29%)	2 (14.29%)	10 (71.43%)
TextBlob (German/DE)	1	0 (0.00%)	1 (100.00%)	0 (0.00%)

Video id rCqqd4VHksI : 38 out of 403 posts (b)

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	34	4 (11.76%)	6 (17.65%)	24 (70.59%)
TextBlob (English/EN)	34	2 (5.88%)	15 (44.12%)	17 (50.00%)
TextBlob (French/FR)	1	0 (0.00%)	1 (100.00%)	0 (0.00%)
TextBlob (German/DE)	3	0 (0.00%)	3 (100.00%)	0 (0.00%)

Video id 0J_YZmXJjsk : 374 out of 554 posts (c)

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	351	18 (5.13%)	46 (13.11%)	287 (81.77%)
TextBlob (English/EN)	351	14 (3.99%)	83 (23.65%)	254 (72.36%)
TextBlob (French/FR)	4	0 (0.00%)	4 (100.00%)	0 (0.00%)
TextBlob (German/DE)	17	0 (0.00%)	10 (58.82%)	7 (41.18%)

Figure 6. (a), (b) and (c): Pilot Testing in Comparing the VADER and TextBlob Performance

Figure 6 shows the pilot testing in comparing the VADER and TextBlob performance of the data. Based on the results of pilot testing, it is evident that VADER demonstrates varying degrees of performance in sentiment classification across the three videos. For the first video, VADER exhibits a distribution of polarity values classified as almost, accompanied by a high Cohen's kappa statistic of 0.825, indicating substantial agreement between VADER's classifications and human judgments. However, for the second and third videos, VADER's performance is moderate, with Cohen's kappa statistics of 0.529 and 0.504, respectively, suggesting moderate agreement. These findings underscore the importance of assessing the reliability and consistency of sentiment classification methodologies across different cultural media contexts, as variations in performance may impact the accuracy and validity of interpretations derived from sentiment analysis results.

Upon establishing the VADER model within the dataset, all review data from the videos were merged to form 1413 entries for further cleaning and extraction to obtain sentiment scores. This process consolidates individual review data points into a single dataset, facilitating efficient data management and analysis. Subsequently, the dataset undergoes cleaning procedures to remove any noise or inconsistencies, ensuring the integrity and reliability of the extracted sentiment scores. By consolidating and processing the review data in this manner, this research derives meaningful insights into viewer sentiments and reactions toward the cultural documentary content, thus enhancing the interpretability and utility of the sentiment analysis results.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

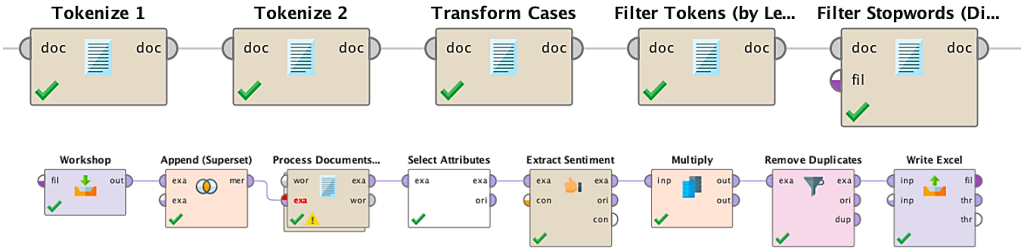


Figure 7. Data Cleaning and Extract Sentiment

Figure 7 shows the pre-processing and extraction of data. The review data for each video undergoes a cleaning process involving tokenization, case transformation, length-based token filtering, stopword removal, and stemming (using the Porter stemmer). Subsequently, the document processing is tailored to the TF-IDF (Term Frequency-Inverse Document Frequency) methodology, with a specific configuration for the pruning method. In this configuration, tokens with a frequency of less than two are pruned below the absolute threshold, while tokens with more than four are pruned above it. This systematic preprocessing approach ensures the standardization and optimization of the review data for further analysis, enabling the extraction of meaningful insights and patterns from the textual content of the cultural documentary reviews.

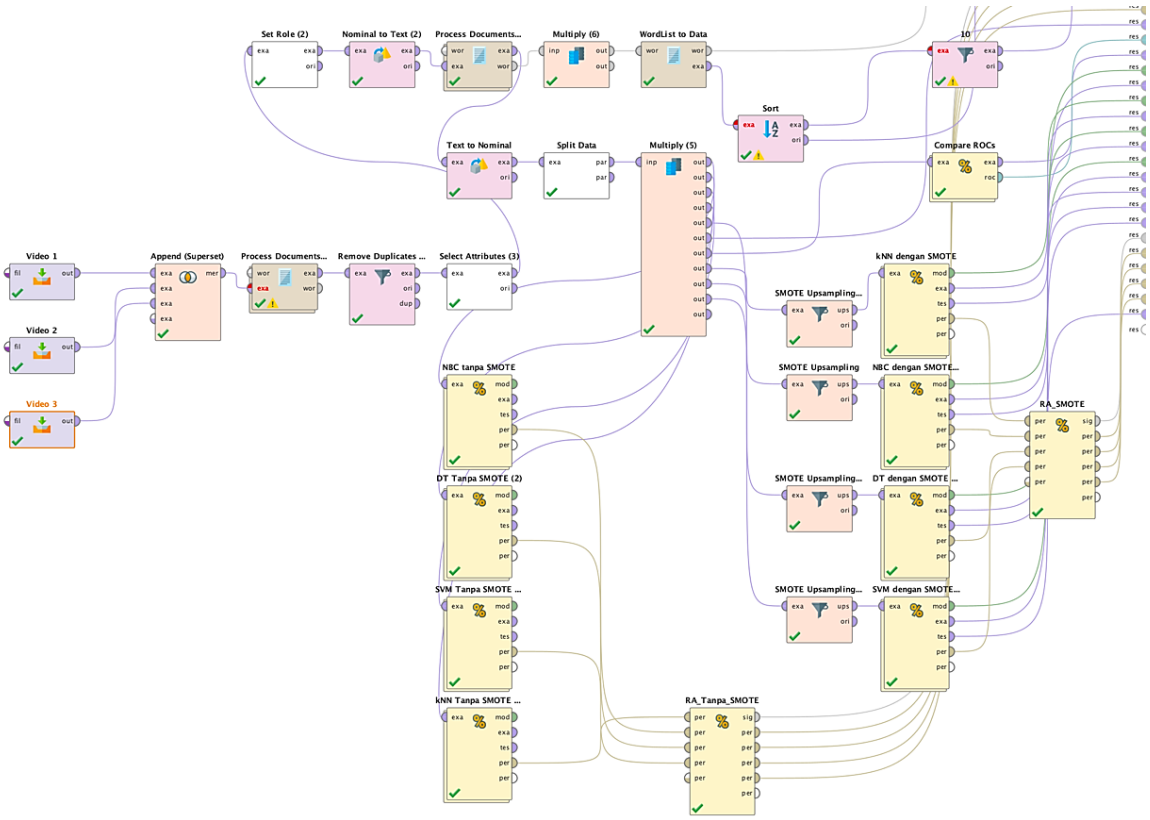


Figure 8. Sentiment Classification using VADER

Figure 8 shows the sentiment classification process using VADER. Each algorithm tested for sentiment classification, including k-NN, DT, NBC, and SVM, undergoes evaluation using 30% training data and 70% testing data split. Following this evaluation, the model exhibiting the highest performance is recommended as the best model for sentiment classification. This systematic approach allows for a comprehensive comparison of the algorithms' efficacy in classifying sentiment within the dataset, thereby facilitating the identification of the most effective method for sentiment analysis. Consequently, selecting the best-performing model ensures optimal accuracy and reliability in sentiment classification, enhancing the utility and interpretability of the sentiment analysis results.

Based on the performance testing results of sentiment classification models, insights into viewer responses based on negative and positive sentiment classes are derived. This analysis enables a nuanced understanding of the range and distribution of viewer sentiments towards the cultural documentary content, highlighting both areas of satisfaction and concern. By categorizing viewer responses into distinct sentiment classes, this research discerns prevalent themes, emotions, and attitudes expressed by the audience, thereby enriching interpretations of audience engagement and feedback dynamics. Consequently, this approach enhances

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

the depth and granularity of insights into viewer perceptions and reactions, contributing to a more comprehensive assessment of the documentary's impact and resonance among viewers.

2.2.4 Evaluation

During the evaluation phase, the confusion matrix and F-measure are analyzed alongside the Area Under the Curve (AUC) to recommend relevant models. The confusion matrix provides a comprehensive overview of classification performance by illustrating the true positives, true negatives, false positives, and false negatives. Meanwhile, the F-measure offers a harmonic mean between precision and recall, providing a balanced assessment of model performance. Additionally, the AUC quantifies the model's ability to discriminate between positive and negative classes, offering a holistic measure of classification accuracy. By integrating these evaluation metrics, this research makes informed decisions regarding selecting the most suitable model for sentiment classification, ensuring robust and reliable analyses of viewer responses within the cultural media context.

In addition to evaluating algorithm performance, the toxicity score is assessed as supplementary data to ascertain the values of Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat. These metrics provide valuable insights into the level and nature of toxic or abusive language within the dataset, offering a nuanced understanding of the sentiment and discourse surrounding the cultural media content. By examining these toxicity indicators, this research identifies harmful language or behavior instances, thus enriching the analysis of viewer responses and enhancing the overall interpretation of sentiment dynamics within the cultural media landscape.

2.2.5 Deployment

During the deployment phase, a comparative analysis of positive and negative sentiments was conducted to generate recommendations for creators of Sumba cultural documentary videos. This analysis examines the distribution and prevalence of positive and negative sentiments expressed by viewers towards the documentary content. By comparing these sentiments, creators gain valuable insights into audience perceptions and preferences, enabling them to tailor future content to better align with viewer expectations and interests. Ultimately, this approach facilitates the production of culturally relevant and engaging documentary videos that resonate with audiences, fostering a deeper appreciation and understanding of Sumba's rich cultural heritage.

3. RESULT AND DISCUSSION

The Sumba cultural documentary videos elicit diverse opinions and viewpoints from viewers and serve as a benchmark for channel performance, determining the popularity of the videos. As these documentaries delve into the rich cultural heritage of Sumba, they evoke varied responses and interpretations among viewers, reflecting the multifaceted nature of cultural appreciation and understanding. Additionally, the viewers' engagement with and reception of these videos contribute significantly to the overall performance metrics of the channel, influencing its visibility and appeal within the digital media landscape. Consequently, cultural documentary videos serve as educational and informative resources and play a pivotal role in shaping audience perceptions and channel dynamics, highlighting the significance of contemporary media discourse.

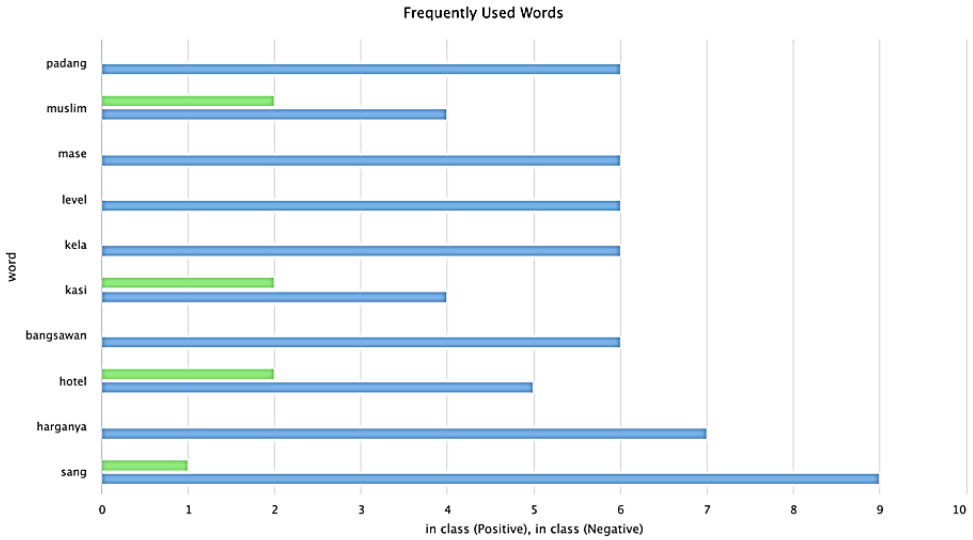


Figure 9. Frequently Used Words of All Datasets

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

Figure 9 shows the frequently used Words of all the datasets. Based on the results of identifying frequently used words from the three video datasets, it is evident that none of the words contain negative connotations. Specifically, the following words were identified along with the respective frequencies: "sang" (4), "harganya" (3), "hotel" (4), "bangsawan" (4), "kasi" (4), "kela" (4), "level" (2), "mase" (4), "muslim" (4), and "padang" (2). This analysis indicates the absence of negative language within the dataset, suggesting a neutral or positive tone prevalent in the content discussed across the videos. Consequently, the absence of negative keywords underscores the discourse's predominantly positive or neutral nature within the cultural documentary videos, reflecting a constructive and respectful engagement with the subject matter.

Based on the results of implementing the k-NN, DT, NBC, and SVM algorithms using Synthetic Minority Over-sampling Technique (SMOTE), it is evident that the application of SMOTE significantly improves the performance of these algorithms in handling imbalanced datasets. SMOTE effectively addresses the issue of class imbalance by generating synthetic samples for the minority class, thereby enhancing the algorithm's ability to classify instances from both classes accurately. This approach ensures more robust and reliable classification results, contributing to the overall effectiveness of the machine learning models in various classification tasks.

SVM Using SMOTE (a)

```
PerformanceVector:
accuracy: 83.08% +/- 2.17% (micro average: 83.08%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       585     18
Positive:       272     839
AUC (optimistic): 0.977 +/- 0.011 (micro average: 0.977) (positive class: Positive)
AUC: 0.976 +/- 0.011 (micro average: 0.976) (positive class: Positive)
AUC (pessimistic): 0.975 +/- 0.012 (micro average: 0.975) (positive class: Positive)
precision: 75.57% +/- 2.36% (micro average: 75.52%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       585     18
Positive:       272     839
recall: 97.90% +/- 1.53% (micro average: 97.90%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       585     18
Positive:       272     839
f_measure: 85.28% +/- 1.73% (micro average: 85.26%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       585     18
Positive:       272     839
```

DT Using SMOTE (b)

```
PerformanceVector:
accuracy: 87.05% +/- 1.65% (micro average: 87.05%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       643     8
Positive:       214     849
AUC (optimistic): 0.996 +/- 0.005 (micro average: 0.996) (positive class: Positive)
AUC: 0.871 +/- 0.016 (micro average: 0.871) (positive class: Positive)
AUC (pessimistic): 0.747 +/- 0.034 (micro average: 0.747) (positive class: Positive)
precision: 79.93% +/- 2.29% (micro average: 79.87%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       643     8
Positive:       214     849
recall: 99.07% +/- 1.19% (micro average: 99.07%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       643     8
Positive:       214     849
f_measure: 88.45% +/- 1.32% (micro average: 88.44%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       643     8
Positive:       214     849
```

k-NN Using SMOTE (c)

```
PerformanceVector:
accuracy: 96.79% +/- 1.33% (micro average: 96.79%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       835     33
Positive:       22      824
AUC (optimistic): 0.998 +/- 0.001 (micro average: 0.998) (positive class: Positive)
AUC: 0.980 +/- 0.012 (micro average: 0.980) (positive class: Positive)
AUC (pessimistic): 0.968 +/- 0.016 (micro average: 0.968) (positive class: Positive)
precision: 97.41% +/- 1.31% (micro average: 97.40%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       835     33
Positive:       22      824
recall: 96.15% +/- 1.92% (micro average: 96.15%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       835     33
Positive:       22      824
f_measure: 96.77% +/- 1.34% (micro average: 96.77%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       835     33
Positive:       22      824
```

NBC Using SMOTE (d)

```
PerformanceVector:
accuracy: 72.24% +/- 4.44% (micro average: 72.23%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       857     476
Positive:       0       381
AUC (optimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Positive)
AUC: 0.500 +/- 0.000 (micro average: 0.500) (positive class: Positive)
AUC (pessimistic): 0.445 +/- 0.087 (micro average: 0.445) (positive class: Positive)
precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       857     476
Positive:       0       381
recall: 44.50% +/- 8.70% (micro average: 44.46%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       857     476
Positive:       0       381
f_measure: 61.14% +/- 8.34% (micro average: 61.55%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       857     476
Positive:       0       381
```

Figure 10. (a), (b), (c) and (d): Performance of SVM, DT, k-NN, and NBC Using SMOTE

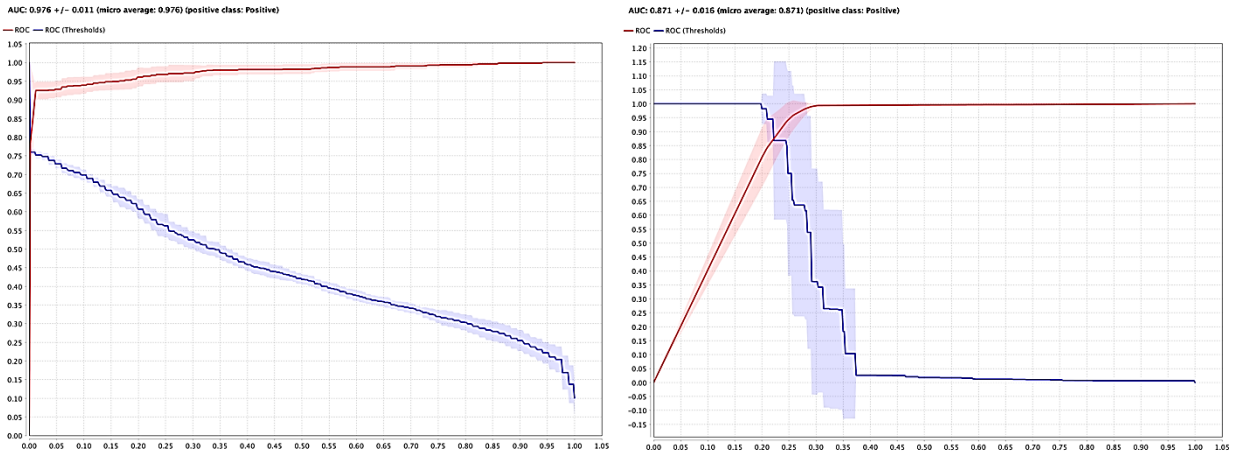
Figure 10 shows the performance of each algorithm using SMOTE. Based on the implementation results of the Decision Tree (DT) algorithm using Synthetic Minority Over-sampling Technique (SMOTE), the performance metrics indicate a high level of accuracy, with an average accuracy of 87.05% and a micro average of the same value. The confusion matrix illustrates the model's ability to effectively classify instances into negative and positive classes, with a notable precision of 79.93% and a recall rate of 99.07%, particularly for the positive class. Furthermore, the f-measure, representing the harmonic mean of precision and recall, demonstrates a favorable score of 88.45%. These results underscore the efficacy of the DT algorithm in conjunction with SMOTE in accurately classifying instances from imbalanced datasets, particularly in scenarios where the positive class is of primary interest.

In addition, based on the results of implementing the Support Vector Machine (SVM) algorithm using Synthetic Minority Over-sampling Technique (SMOTE), the performance evaluation demonstrates notable accuracy, with a reported value of 83.08% +/- 2.17%. The confusion matrix reveals a strong ability of the model to correctly classify instances from both negative and positive classes, as evidenced by high precision, recall, and f-measure values, particularly for the positive class. Furthermore, the Area Under the Curve (AUC) metrics indicate the model's robust discriminative power and reliability in distinguishing between the two classes, with values ranging between 0.975 and 0.977. Overall, the implementation of SVM with SMOTE showcases

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

promising outcomes, suggesting its effectiveness in handling imbalanced datasets and producing accurate classification results.

AUC of SVM Using SMOTE (a) AUC of DT Using SMOTE (b)



AUC of k-NN Using SMOTE (c) AUC of NBC Using SMOTE (d)

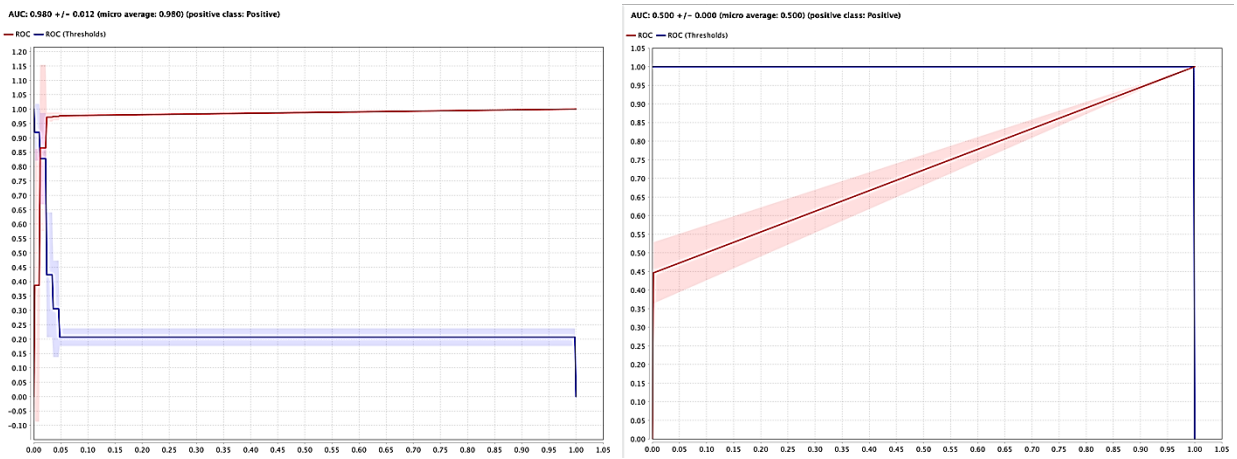


Figure 11. (a), (b), (c) and (d): Performance of SVM, DT, k-NN, and NBC Using SMOTE

Figure 11 shows the Area Under the Curve of SVM, DT, k-NN, and NBC Using SMOTE. The results of implementing the k-Nearest Neighbors (k-NN) algorithm using the Synthetic Minority Over-sampling Technique (SMOTE), the performance evaluation illustrates exceptionally high accuracy, with a reported value of 96.79% +/- 1.33%. The confusion matrix showcases the model's capability to correctly classify instances from negative and positive classes, as evidenced by the high precision, recall, and f-measure values, particularly for the positive class. Additionally, the Area Under the Curve (AUC) metrics demonstrate the model's robust discriminative power, with values ranging between 0.968 and 0.998, further underscoring its effectiveness in handling imbalanced datasets and yielding precise classification outcomes.

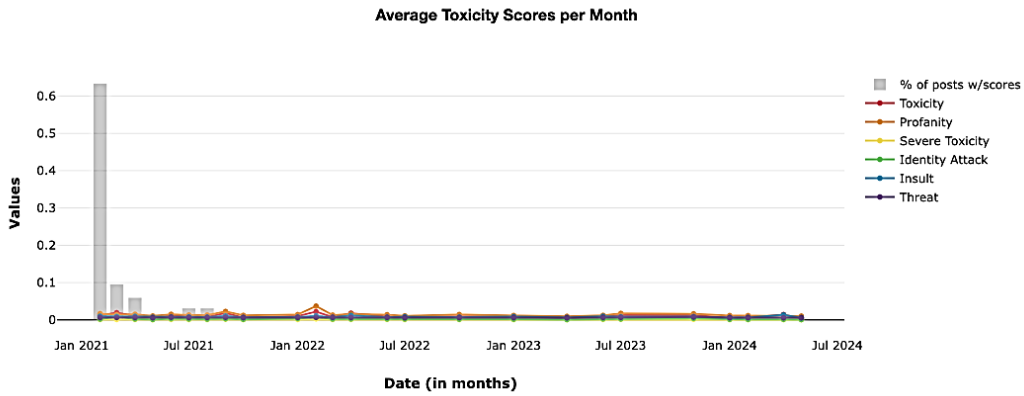
Moreover, based on the results of implementing the Naive Bayes Classifier (NBC) algorithm using Synthetic Minority Over-sampling Technique (SMOTE), the evaluation indicates an accuracy of 72.24% +/- 4.44%, showcasing its capability to classify instances from both negative and positive classes. However, the precision metric, reflecting the ability to identify positive instances correctly, exhibits a perfect score of 100.00%, indicating the model's proficiency in accurately detecting positive cases. Conversely, the recall and f-measure metrics demonstrate less satisfactory performance, suggesting that while the model excels in precision, it struggles with recall, resulting in a lower f-measure value. Additionally, the Area Under the Curve (AUC) metrics indicate varying degrees of discriminative power, with values ranging from 0.445 to 1.000, underscoring the model's uneven performance distinguishing between the positive and negative classes.

The limitation of this study lies in the reliance on sentiment extraction based solely on VADER, which may lead to potential drawbacks in the modeling process. While VADER provides a convenient means of sentiment analysis, its effectiveness may be constrained by context sensitivity and the inability to capture nuanced sentiments. Consequently, the modeling outcomes may not fully capture the complexities inherent in sentiment classification tasks, compromising the results' overall robustness and accuracy. Hence, future research

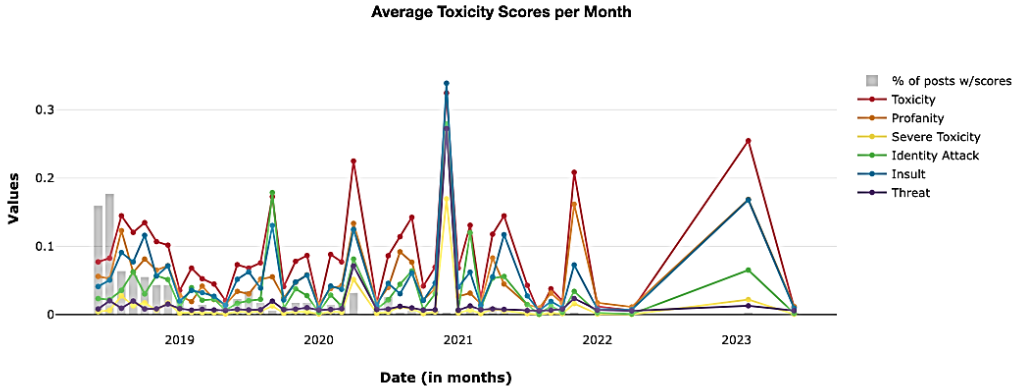
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

endeavors should explore integrating multiple sentiment analysis approaches or refining existing methodologies to address these limitations comprehensively.

Video id : 7hMHD30c5Bg



Video id : rCqqd4VHksI



Video id : 0J_YZmXJjsk

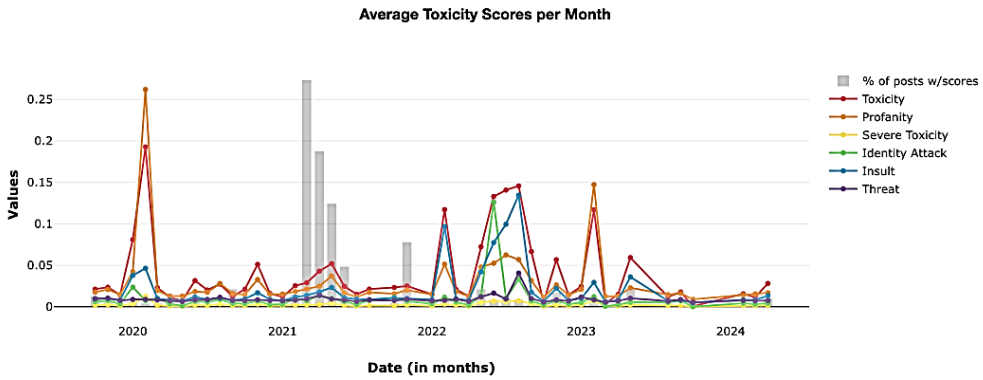


Figure 12. Toxicity Score of the Dataset

Figure 12 shows the toxicity score of each video's dataset. Based on the results of the toxicity analysis, it is evident that the first video (7hMHD30c5Bg) exhibits relatively low levels of toxicity across various dimensions. Specifically, the toxicity score is recorded at 0.01270, indicating minimal toxicity in the content. Similarly, the severity of toxicity, identity attacks, insults, profanity, and threats are all comparatively low, with scores ranging from 0.00104 to 0.01590. These findings suggest that the video content maintains a generally respectful and non-offensive tone, contributing to a positive viewer experience. Based on the results of the toxicity analysis, it is evident that the second video (rCqqd4VHksI) exhibits relatively higher levels of toxicity across various dimensions compared to the first video. The toxicity score for this video is notably higher at 0.09334, indicating a higher likelihood of encountering toxic content. Additionally, the severity of toxicity, identity attacks, insults, profanity, and threats are all elevated, with scores ranging from 0.00987 to 0.05900.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

These findings suggest that the content of the second video may contain more contentious or offensive material, potentially impacting viewer perception and engagement.

Moreover, based on the results of toxicity analysis, it is discernible that the third video (0J_YZmXJjsk) demonstrates a moderate toxicity level, albeit lower than the second video. The overall toxicity score for this video stands at 0.03963, indicating a relatively lower likelihood of encountering toxic content. Furthermore, the severity of toxicity, identity attacks, insults, profanity, and threats are all comparatively lower, with scores ranging from 0.00270 to 0.02683. These findings suggest that the third video's content may contain less contentious or offensive material, potentially resulting in a more favorable viewer experience and engagement.

The limitation of the toxicity score lies in its inherent complexity and subjectivity in assessing the nuanced nature of toxic behavior in textual content. While toxicity scoring algorithms aim to quantify the presence of harmful language, they often struggle to capture context, cultural nuances, and subtleties of communication, thereby potentially misinterpreting benign expressions as toxic. Additionally, these scores may not adequately account for sarcasm, irony, or humor, leading to misclassifications and inaccurate assessments of toxicity levels. As such, while toxicity scores provide valuable insights into the general tone of the text, they should be interpreted with caution and complemented with human judgment and contextual understanding to ensure accurate and meaningful analyses.

4. CONCLUSION

Based on the comprehensive analysis of the CRISP-DM framework, this research has yielded valuable insights into sentiment classification and toxicity analysis of cultural documentary videos. By implementing various machine learning algorithms, including k-NN, DT, NBC, and SVM, along with techniques such as SMOTE for addressing class imbalance, the study achieved notable accuracies ranging from 72.24% to 96.79% in sentiment classification. Moreover, toxicity analysis revealed nuanced insights into the presence of toxic language across the examined videos, with toxicity scores ranging from 0.01270 to 0.09334. Despite these achievements, it's essential to acknowledge the limitations inherent in toxicity scoring algorithms, which may overlook contextual nuances and misinterpret certain expressions. Overall, this research provides a robust foundation for further exploration into sentiment analysis and toxicity assessment in cultural documentary content, highlighting the importance of leveraging advanced machine learning techniques within the CRISP-DM framework for comprehensive data analysis and interpretation.

REFERENCES

[1] M. Hawkins and M. Hawkins, “Antigone in the London office: documentary film, creativity and female agency,” *Cult. Stud.*, vol. 36, no. 5, pp. 856–873, 2022, doi: 10.1080/09502386.2021.2011930.

[2] J. Chambers, “Mysterious objects: exploring imaginary community, community imagination and cinematic translations of Scottish oral traditions within documentary film production and post-production,” *Media Pract. Educ.*, vol. 23, no. 4, pp. 315–328, 2022, doi: 10.1080/25741136.2022.2111627.

[3] R. J. Bramley, “‘A Community Legacy on Film’: using collaborative documentary filmmaking to go beyond representations of the Windrush Generation as ‘victims,’” *Stud. Doc. Film*, vol. 17, no. 2, pp. 115–132, 2023, doi: 10.1080/17503280.2022.2090701.

[4] M. Pramaggiore and P. Kerrigan, “Streaming bloody murder: documentary celebrity and Sophie Toscan Du Plantier anniversary media (SAM),” *Celebr. Stud.*, vol. 00, no. 00, pp. 1–21, 2023, doi: 10.1080/19392397.2023.2207743.

[5] J. Fenwick, “Urban regeneration and stakeholder dynamics in the formation, growth and maintenance of the Sheffield International Documentary Festival in the 1990s,” *Hist. J. Film. Radio Telev.*, vol. 41, no. 4, pp. 838–863, 2021, doi: 10.1080/01439685.2021.1922035.

[6] Y. Zhu, “China’s ‘new cultural diplomacy’ in international broadcasting: branding the nation through CGTN Documentary,” *Int. J. Cult. Policy*, vol. 28, no. 6, pp. 671–683, 2022, doi: 10.1080/10286632.2021.2022651.

[7] E. Colucci, “‘Breaking the chains’: reflections on the making of an ethnographic documentary on human rights violations against people with mental illness in Indonesia,” *Vis. Stud.*, 2023, doi: 10.1080/1472586X.2023.2274892.

[8] A. Zemanek and L. Momesso, “Multiculturalism through a lens: migrants’ voice in Taiwanese documentaries,” *Inter-Asia Cult. Stud.*, vol. 24, no. 3, pp. 413–430, 2023, doi: 10.1080/14649373.2023.2209426.

[9] M. Gandy, “Film as Method in the Geohumanities,” *Geohumanities*, vol. 7, no. 2, pp. 605–624, 2021, doi: 10.1080/2373566X.2021.1898287.

[10] N. Sakr and J. Steemers, “Children’s documentaries: distance and ethics in European storytelling about the wider world,” *J. Child. Media*, vol. 16, no. 2, pp. 288–302, 2022, doi: 10.1080/17482798.2021.1974502.

[11] R. W. Hefner, “Islam and Covenantal Pluralism in Indonesia: A Critical Juncture Analysis,” *Rev. Faith Int. Aff.*, vol. 18, no. 2, pp. 1–17, 2020, doi: 10.1080/15570274.2020.1753946.

[12] J. H. Hanson, M. Schutgens, N. Baral, and N. Leader-Williams, “Assessing the potential of snow leopard tourism-related products and services in the Annapurna Conservation Area, Nepal,” *Tour. Plan. Dev.*, vol. 20, no. 6, pp. 1182–1202, 2023, doi: 10.1080/21568316.2022.2122073.

[13] L. Palmer, S. Barnes, T. Wagner, and A. Hanley, “Holding Tightly: Co-Mingling, Life-Flourishing and Filmic Ecologies,” *J. Intercult. Stud.*, vol. 44, no. 5, pp. 697–715, 2023, doi: 10.1080/07256868.2023.2192910.

[14] O. J. Hakola, “Ethical reflections on filming death in end-of-life documentaries,” *Mortality*, vol. 28, no. 3, pp. 395–410, 2023, doi: 10.1080/13576275.2021.1946025.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 535-547
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5150

[15] A. L. G. Waworuntu, Z. Alkatiri, and R. de Archellie, "Challenging the promise of decentralization: The case of marginalization of Mosalaki role in Nggela Vilage in Ende Lio, Flores," *Cogent Arts Humanit.*, vol. 10, no. 1, 2023, doi: 10.1080/23311983.2023.2168835.
[16] A. Schapper, "Beyond 'Macassans': Speculations on layers of Austronesian contact in northern Australia," *Aust. J. Linguist.*, vol. 41, no. 4, pp. 434–452, 2021, doi: 10.1080/07268602.2021.2000365.
[17] N. Sumba Nacipucha, A. Sánchez-Bayón, J. Cueva Estrada, and A. Valencia-Arias, "Social networks as a strategy to improve the visibility of scientific journals," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2306715.
[18] S. Rahmadani, I. Meilano, S. Susilo, D. A. Sarsito, H. Z. Abidin, and P. Supendi, "Geodetic observation of strain accumulation in the Banda Arc region," *Geomatics, Nat. Hazards Risk*, vol. 13, no. 1, pp. 2579–2596, 2022, doi: 10.1080/19475705.2022.2126799.
[19] A. Dorkas Rambu Atahau, I. Madea Sakti, A. Namilana Rambu Hutar, A. Dolfriandra Huruta, and M. S. Kim, "Financial literacy and sustainability of rural microfinance: The mediating effect of governance," *Cogent Econ. Financ.*, vol. 11, no. 2, 2023, doi: 10.1080/23322039.2023.2230725.
[20] Y. Ghanggo Ate and C. El-Khaissi, "Apologizing in Kodhi," *Aust. J. Linguist.*, vol. 43, no. 3, pp. 258–282, 2023, doi: 10.1080/07268602.2023.2290685.
[21] U. Supraptiningsih, H. Jubba, E. Hariyanto, and T. Rahmawati, "Inequality as a cultural construction: Women's access to land rights in Madurese society," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2194733.
[22] S. Baleghizadeh and L. Amiri Shayesteh, "A content analysis of the cultural representations of three ESL grammar textbooks," *Cogent Educ.*, vol. 7, no. 1, 2020, doi: 10.1080/2331186X.2020.1844849.
[23] E. Fino, B. Hanna-Khalil, and M. D. Griffiths, "Exploring the public's perception of gambling addiction on Twitter during the COVID-19 pandemic: Topic modelling and sentiment analysis," *J. Addict. Dis.*, vol. 39, no. 4, pp. 489–503, 2021, doi: 10.1080/10550887.2021.1897064.
[24] L. Nemes and A. Kiss, "Prediction of stock values changes using sentiment analysis of stock news headlines," *J. Inf. Telecommun.*, vol. 5, no. 3, pp. 375–394, 2021, doi: 10.1080/24751839.2021.1874252.
[25] R. K. Botchway, A. B. Jibril, Z. K. Oplatková, and M. Chovancová, "Deductions from a Sub-Saharan African Bank's Tweets: A sentiment analysis approach," *Cogent Econ. Financ.*, vol. 8, no. 1, 2020, doi: 10.1080/23322039.2020.1776006.
[26] T. Da Nguyen, "An approach to improve the accuracy of rating prediction for recommender systems," *Automatika*, vol. 65, no. 1, pp. 58–72, 2024, doi: 10.1080/00051144.2023.2284026.
[27] A. John and T. Latha, "Stock market prediction based on deep hybrid RNN model and sentiment analysis," *Automatika*, vol. 64, no. 4, pp. 981–995, 2023, doi: 10.1080/00051144.2023.2217602.
[28] E. R. Kovacs, L. A. Cotfas, and C. Delcea, "January 6th on Twitter: measuring social media attitudes towards the Capitol riot through unhealthy online conversation and sentiment analysis," *J. Inf. Telecommun.*, vol. 8, no. 1, pp. 108–129, 2024, doi: 10.1080/24751839.2023.2262067.
[29] S. W. Ke, C. F. Tsai, and Y. J. Chen, "Managing Emotion In The Workplace: An Empirical Study With Enterprise Instant Messaging," *Appl. Artif. Intell.*, vol. 38, no. 1, 2024, doi: 10.1080/08839514.2023.2297518.
[30] O. Lock and C. Pettit, "Social media as passive geo-participation in transportation planning–how effective are topic modeling & sentiment analysis in comparison with citizen surveys?," *Geo-Spatial Inf. Sci.*, vol. 23, no. 4, pp. 275–292, 2020, doi: 10.1080/10095020.2020.1815596.
[31] C. J. R. Walker, M. B. Doucette, S. Rotz, D. Lewis, H. T. Neufeld, and H. Castleden, "Non-Indigenous partner perspectives on Indigenous peoples' involvement in renewable energy: exploring reconciliation as relationships of accountability or status quo innocence?," *Qual. Res. Organ. Manag. An Int. J.*, vol. 16, no. 3–4, pp. 636–657, Jan. 2021, doi: 10.1108/QROM-04-2020-1916.
[32] N. Chanza and W. Musakwa, "Revitalizing indigenous ways of maintaining food security in a changing climate: review of the evidence base from Africa," *Int. J. Clim. Chang. Strateg. Manag.*, vol. 14, no. 3, pp. 252–271, Jan. 2022, doi: 10.1108/IJCCSM-06-2021-0065.
[33] L. Pham Hong, H. T. Ngo, and L. T. Pham, "Community-based tourism: Opportunities and challenges a case study in Thanh Ha pottery village, Hoi An city, Vietnam," *Cogent Soc. Sci.*, vol. 7, no. 1, 2021, doi: 10.1080/23311886.2021.1926100.
[34] V. Sen and P. Walter, "Community-based ecotourism and the transformative learning of homestay hosts in Cambodia," *Tour. Recreat. Res.*, vol. 45, no. 3, pp. 323–336, 2020, doi: 10.1080/02508281.2019.1692171.
[35] J. Sixto-García, A. I. Rodríguez-Vázquez, and X. López-García, "News Sharing Using Self-destructive Content in Digital Native Media from an International Perspective," *Journal. Pract.*, vol. 17, no. 7, pp. 1341–1356, 2023, doi: 10.1080/17512786.2021.2000883.
[36] N. Gryllakis and M. Matsiola, "Digital audiovisual content in marketing and distributing cultural products during the COVID-19 pandemic in Greece," *Arts Mark.*, vol. 13, no. 1, pp. 4–19, Jan. 2023, doi: 10.1108/AAM-09-2021-0053.
[37] A. L. Haw, "What drives political news engagement in digital spaces? Reimagining 'echo chambers' in a polarised and hybridised media ecology," *Commun. Res. Pract.*, vol. 6, no. 1, pp. 38–54, 2020, doi: 10.1080/22041451.2020.1732002.
[38] K. Toffoletti, R. Olive, H. Thorpe, and A. Pavlidis, "Doing feminist physical cultural research in digital spaces: reflections, learnings and ways forward," *Qual. Res. Sport. Exerc. Heal.*, vol. 13, no. 1, pp. 11–25, 2021, doi: 10.1080/2159676X.2020.1836513.
[39] N. A. K. Zamri, N. N. A. Mohamad Nasir, M. N. Hassim, and S. M. Ramli, "Digital hate speech and othering: The construction of hate speech from Malaysian perspectives," *Cogent Arts Humanit.*, vol. 10, no. 1, 2023, doi: 10.1080/23311983.2023.2229089.
[40] J. hyun Im, "The discursive construction of East Asian identities in an era of globalization and internationalization: the linguistic landscape of East Asian departments at a U.S. university," *J. Multicult. Discourses*, vol. 15, no. 1, pp. 80–103, 2020, doi: 10.1080/17447143.2020.1738441.