Toxicity Analysis and Sentiment Classification of Wonderland Indonesia by Alffy Rev using Support Vector Machine

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: 16/03/2024; Accepted: 31/03/2024; Published: 31/03/2024

Abstract—The music industry's increasing reliance on digital platforms like YouTube for dissemination raises concerns about the potential impact of music videos on viewer sentiment and well-being. This study seeks to assess the toxicity and sentiment of the Wonderland Indonesia music video by Alffy Rev through Support Vector Machine analysis, contributing to our understanding of the effects of music content on online audiences. This research addresses the challenge of sentiment classification in digital content by leveraging the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. The study aims to enhance sentiment classification accuracy by applying a Support Vector Machine (SVM) with a Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance issues. The research problem revolves around the need for robust sentiment analysis models capable of accurately discerning sentiment polarity within diverse datasets. Through the systematic application of CRISP-DM phases - business understanding, data understanding, data preparation, modeling, evaluation, and deployment - the study examines the efficacy of SVM with SMOTE in sentiment classification tasks. The findings demonstrate notable performance metrics, including accuracy (96.50%), precision (95.75%), recall (99.00%), and F-measure (97.34%). The AUC value substantially increases from 0.642 without SMOTE to 0.997 with SMOTE, highlighting its effectiveness in improving sentiment classification accuracy. In addition, The comparative analysis of toxicity values between the first and second videos demonstrates distinct patterns: the first video showcases a Toxicity score of 0.05290, with notable metrics such as Profanity registering at 0.04815. Conversely, the second video exhibits a slightly lower Toxicity score of 0.04744, with varying metrics such as Severe Toxicity at 0.01386.

Keywords: Wonderland Indonesia; Alffy Rev; Toxicity Analysis; Sentiment Classification; SVM

1. INTRODUCTION

An effective promotion strategy for tourism destinations entails the utilization of creative digital video content that authentically portrays the local culture. Such content is a powerful tool for capturing the essence of a destination and enticing potential travelers [1]–[3]. By showcasing the unique cultural heritage and traditions through visually compelling narratives, videos enable viewers to immerse themselves in the allure of the locale [1], [4], [5]. Furthermore, digital platforms offer a broad reach, allowing the dissemination of these videos to diverse audiences globally [6]–[8]. Thus, harnessing the potential of creative digital videos that reflect local culture represents a pivotal approach to effectively promoting tourism destinations.

Digital marketing is pivotal in piquing tourists’ interest in visiting destinations. Tourism businesses can effectively reach and engage with their target audience by leveraging various online channels and platforms, such as social media, search engines, and email campaigns [9]–[11]. Through targeted advertisements and personalized content, digital marketing can tailor messages to specific demographics and interests, thereby capturing the attention of potential travelers [12]–[15]. Moreover, measuring and analyzing data allows for continuous refinement and optimization of marketing strategies, ensuring maximum impact and return on investment [16]–[19]. Therefore, it is evident that digital marketing is a cornerstone in enticing tourists to explore and experience different destinations.

The prospective utilization of digital video content in tourism marketing represents a significant advancement in promotional strategies. Creating and disseminating captivating video content across digital platforms allows destinations to effectively showcase their unique attractions and experiences to a global audience [20], [21]. Such videos offer immersive storytelling opportunities, allowing potential travelers to envision themselves engaging with the destination's offerings [22], [23]. Furthermore, the interactive nature of digital videos enables viewers to actively participate in the exploration process actively, fostering a deeper connection and desire to visit the showcased destination [24]. Consequently, leveraging digital video content in tourism marketing holds immense potential in driving visitor interest, engagement, and, ultimately, visitation to destinations.

This research analyzes the user responses on YouTube to the video content titled "Wonderland Indonesia," published by Alffy Rev, with the respective video codes aKtb7Y3qOck and Fa_rNR__UV0. By examining metrics such as view counts, likes, comments, and shares, the study aims to gauge the effectiveness and appeal of these videos in capturing audience attention and engagement. Through systematic data analysis, patterns in user behavior and preferences towards the showcased Indonesian wonders can be identified, offering valuable insights for content creators and tourism marketers. Consequently, this investigation contributes to a better understanding of the impact of digital video content in promoting Indonesian tourism and informs future content creation and dissemination strategies.
The proposed method utilizes the CRISP-DM framework, employing toxicity analysis and sentiment classification techniques using the Support Vector Machine (SVM) algorithm and the Synthetic Minority Over-sampling Technique (SMOTE). This approach offers a systematic methodology for data mining, encompassing various stages from data understanding to deployment. Integrating toxicity analysis allows for identifying and mitigating harmful or offensive content within textual data, thereby enhancing the quality and safety of online platforms. Additionally, sentiment classification employing SVM and SMOTE enables the categorization of user sentiments, facilitating more profound insights into audience perceptions and preferences. Consequently, this methodological framework presents a comprehensive approach to extracting valuable information from textual data, offering potential applications in fields such as social media analysis, customer feedback evaluation, and content moderation.

The urgency of toxicity analysis and sentiment classification research lies in its pivotal role in fostering a safer and more conducive online environment, particularly amidst the escalating concerns regarding online toxicity and harmful content proliferation. Researchers use sophisticated algorithms and methodologies to identify and mitigate toxic behaviors and sentiments in digital content and user interactions [24]. Moreover, the timely exploration and refinement of toxicity analysis and sentiment classification techniques are essential for developing proactive measures and policies to curb online harassment, hate speech, and misinformation dissemination [25]. Thus, prioritizing research in this domain is paramount for safeguarding user well-being, promoting digital civility, and advancing the overall quality of online discourse.

This research's theoretical and practical implications are profound, spanning across academic and applied domains. By advancing theoretical frameworks and models, this study contributes to the body of knowledge within its respective field, providing insights and perspectives that inform future research endeavors [26]. Additionally, the practical implications of this research extend to real-world applications, offering actionable recommendations or solutions to address specific challenges or issues identified in the study [27]. Through the integration of theory and practice, this research bridges the gap between academia and industry, facilitating the translation of theoretical concepts into practical solutions that benefit society. Consequently, this study holds significance in enriching scholarly discourse and offering tangible outcomes with practical utility.

The limitation of this research stems from the complexities associated with toxicity analysis and sentiment classification using the CRISP-DM framework. While this methodology offers a structured approach to data mining, challenges may arise in accurately capturing the nuances of human language and behavior, particularly in online contexts where expressions can vary widely [28]–[30]. Additionally, the effectiveness of the employed algorithms, such as Support Vector Machine (SVM), in accurately classifying toxic or sentiment-laden content may be influenced by factors such as dataset quality, algorithm biases, and evolving linguistic trends [31]–[35]. Therefore, despite the systematic nature of the CRISP-DM framework, its application in toxicity analysis and sentiment classification may encounter limitations in capturing the full spectrum of human expression and sentiment dynamics.

2. RESEARCH METHODOLOGY

2.1 Research Gap Analysis using Vosviewer

Gap analysis is a crucial tool for assessing the existing body of research and identifying areas where further investigation is warranted to advance knowledge. By systematically examining literature and research findings, gap analysis enables researchers to pinpoint gaps, inconsistencies, or unanswered questions within the current knowledge landscape. Furthermore, conducting a thorough gap analysis facilitates identifying opportunities for innovation and contribution to the existing body of knowledge. Consequently, integrating gap analysis into research methodologies is imperative for ensuring that research endeavors are informed by existing literature and positioned to make meaningful contributions to advancing knowledge in their respective fields.

Figure 1. Network and Density Visualization (Vosviewer)
Figure 1 shows the network and density visualization using Vosviewer. The gap analysis results about tourism digital marketing reveal a cluster of interconnected and popular subtopics within the field. Through systematic examination, it becomes evident that areas such as tourism marketing, destination marketing, sustainable tourism, smart tourism, food tourism, digital detox, sustainability, and social media are prominent and interrelated themes [36]–[40]. This observation underscores the complexity and multifaceted nature of digital marketing within the tourism sector, highlighting the diverse array of topics that researchers and practitioners alike are actively exploring. Consequently, these findings underscore the importance of holistically addressing these interconnected themes to develop comprehensive strategies and initiatives that cater to modern travelers’ evolving needs and preferences.

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

This research employs the CRISP-DM framework to conduct toxicity analysis and sentiment classification, presenting a systematic approach to data mining in online contexts. Through the structured methodology provided by CRISP-DM, the study aims to effectively navigate the complexities of analyzing textual data for toxic elements and sentiment polarity. By integrating CRISP-DM into the research methodology, the study ensures a systematic and rigorous analysis process, enhancing the findings' reliability and validity. Consequently, leveraging the CRISP-DM framework in toxicity analysis and sentiment classification offers a robust methodology for exploring and understanding the dynamics of online discourse, contributing valuable insights to digital content analysis.

Figure 2 shows the CRISP-DM framework design using Figjam. The rationale for utilizing the CRISP-DM framework in toxicity analysis and sentiment classification has been carefully grounded in data relevance and process coherence. By adhering to the CRISP-DM methodology, the research ensures a systematic and structured approach to handling complex textual data, thereby enhancing the reliability and validity of the analysis. The framework facilitates the orderly progression through critical stages such as data understanding, data preparation, modeling, evaluation, and deployment, ensuring that each step is executed coherently. Consequently, leveraging the CRISP-DM framework in toxicity analysis and sentiment classification ensures the relevance and coherence of the data and processes. It enhances the robustness of the research findings and insights derived from the analysis.

The CRISP-DM methodology comprises several systematic stages essential for practical data mining. Initially, the process begins with Business Understanding, wherein the project's objectives and requirements are defined from a business perspective. Subsequently, Data Understanding involves exploring and assessing the available data sources to gain insights into their quality and potential relevance to the project goals. Following this, Data Preparation encompasses cleaning, transforming, and integrating data to ensure its suitability for analysis. The Modeling phase involves selecting and applying various techniques to build predictive or descriptive models based on the prepared data. Then, Evaluation assesses the performance and effectiveness of the models developed, ensuring they meet the project's objectives and requirements. Lastly, deployment involves deploying the developed models into operational systems and facilitating their practical application and utilization. Overall, the CRISP-DM
methodology provides a structured and systematic approach to guide data mining projects from inception to implementation, ensuring efficient and effective utilization of data-driven insights.

2.2.1 Business Understanding

During the Business Understanding stage, the discussion focuses on the video content identified by the IDs aKtb7Y3qOck and Fa_rNR__UV0. The first video (aKtb7Y3qOck), published on Aug 17, 2021, has garnered a substantial viewership of 56,636,162 and attracted 246,703 comments. In contrast, the second video (Fa_rNR__UV0), released on Aug 17, 2022, has accumulated 17,362,808 views and elicited 102,783 comments. These statistics provide valuable insights into each video's popularity and engagement levels, informing subsequent analyses regarding audience reception, content effectiveness, and potential areas for improvement. Thus, the Business Understanding stage facilitates a comprehensive understanding of the contextual factors surrounding the analyzed video content, laying the groundwork for informed decision-making and strategy development in subsequent stages of the CRISP-DM framework.

![First Video](image1)

![Second Video](image2)

**Figure 3. Post-Per-Day Statistic of First and Second Video (Communalytic)**

Based on the post-per-day statistics, it is evident that the first video received 299 comments on Aug 17, 2022, and 220 comments on Aug 18, 2022, indicating a consistent level of engagement. Conversely, the second video exhibited fluctuating activity levels, with 3621 comments on Aug 21, 2022, and 3425 comments on Aug 20, 2022. These fluctuations may signify varying degrees of audience interest or interaction patterns over time, potentially influenced by content relevance, promotion efforts, or external events. Consequently, analyzing post-per-day statistics offers valuable insights into audience engagement dynamics, aiding in assessing content performance and identifying potential areas for optimization or refinement in future video content strategies.

The intensity of comments on the published videos is a significant indicator of public engagement with the content, thereby establishing a connection with the context of digital tourism marketing. High levels of comment activity suggest heightened interest and interaction from the audience, reflecting their engagement with the tourism-related content showcased in the videos. This engagement is crucial within digital marketing, as it indicates the effectiveness of the content in capturing audience attention and fostering interaction, which are essential elements for promoting tourism destinations and experiences. Consequently, analyzing the intensity of comments provides valuable insights into audience perceptions and preferences, informing strategies for enhancing engagement and leveraging digital platforms effectively for tourism promotion.

2.2.2 Data Understanding

During the data understanding stage, the frequently used words from both video contents are identified to ascertain popular themes and topics. This process involves analyzing the textual data to extract standard terms and phrases that occur frequently across the videos. By identifying these frequently used words, researchers gain valuable insights into the predominant themes and subjects that resonate with the audience, thus informing subsequent analyses and interpretations. Consequently, this systematic examination of famous words enhances understanding
of the content’s relevance and audience interests, laying a foundation for informed decision-making and strategy development in subsequent stages of the research process.

First Video

![First Video Word Cloud](image1)

Figure 4. Frequently Used Words of First and Second Video (Communalytic)

Figure 4 shows the frequently used words in the dataset from the first and second videos. The analysis of frequently used words for the first video content reveals prevalent themes and sentiments associated with Indonesian culture and its attractions. Notably, "Indonesia" appears 176 times, indicating a strong focus on the country itself. Additionally, expressions like "keren" (excellent) and "bagus" (good) occur 82 and 46 times, respectively, highlighting positive perceptions and appreciation for the showcased content. Moreover, words such as "budaya" (culture) and "negara" (country) occur 22 and 29 times, respectively, emphasizing the significance of cultural heritage and national identity within the discourse. The repetition of terms like "karya" (work) and "orang" (people), appearing 25 and 18 times, respectively, suggests recognition of creative endeavors and societal involvement, reflecting a multifaceted engagement with the content. The frequency analysis of words in the second video shows "Indonesia" appearing 117 times, indicating a strong emphasis on the country within the discourse. Additionally, words such as "keren" (excellent) and "karya" (work) occur 89 and 68 times, respectively, reflecting admiration for creative endeavors and cultural appreciation. Moreover, terms like "nonton" (watch), "lagu" (song), and "film" (movie) appear 47, 46, and 37 times, respectively, suggesting engagement with audiovisual media and entertainment. The repetition of expressions like "bagus" (good) and "merinding" (thrilling), occurring 42 and 33 times, respectively, highlights positive perceptions and emotional responses elicited by the content. This analysis offers valuable insights into audience preferences and interests, guiding strategies for content creation and audience engagement in digital marketing.

2.2.3 Modeling

The sentiment classification process utilizes the Support Vector Machine (SVM) algorithm and the Synthetic Minority Over-sampling Technique (SMOTE) operator within the RapidMiner platform. By employing SVM, a supervised learning algorithm capable of efficiently classifying text into different sentiment categories, and SMOTE, a resampling technique used to address the class imbalance, the sentiment classification process aims to accurately predict the sentiment polarity of textual data. This approach allows for creating a robust sentiment classification model that can effectively handle imbalanced datasets and improve the overall performance of sentiment analysis tasks. Consequently, leveraging SVM and SMOTE within RapidMiner provides a systematic
and efficient way to classify sentiment, enabling insightful textual sentiment analysis across various domains and applications.

**Figure 5. Extract Text Data (Rapidminer)**

Figure 5 shows the process of extracting sentiment. Before proceeding to the modeling process, text data undergoes extraction and preliminary cleaning utilizing various operators such as tokenize, transform cases, stopwords, filter tokens, and remove duplicates. The tokenize operator segments the raw text into individual tokens or words, facilitating further analysis. Following this, the transform cases operator standardizes the case of the text to ensure consistency in capitalization. Subsequently, the stopwords operator removes common words with little semantic meaning, enhancing the relevance of the remaining tokens.

Additionally, the filter tokens operator enables the exclusion of specific tokens based on defined criteria, further refining the dataset. Finally, the remove duplicates operator eliminates redundant tokens, ensuring data integrity and reducing redundancy in the dataset. This systematic preprocessing pipeline ensures the readiness and quality of the text data for subsequent modeling and analysis, ultimately contributing to the generation of meaningful insights.

**Figure 6. Data Processing (Rapidminer)**

Figure 6 shows the process of sentiment classification. The data collected from the first and second videos amount to 349,488 entries, from which 54,184 are extracted for further processing. Subsequently, a subset of 9,347 entries is selected for modeling purposes, with a split of 70% for testing and 30% for training, employing the Support Vector Machine (SVM) algorithm. This systematic approach ensures a representative sample size for analysis while maintaining the integrity and quality of the dataset. Consequently, by utilizing SVM for modeling, the research effectively classifies sentiment and toxicity levels within the video content, contributing to a comprehensive understanding of audience engagement and content reception.

In the context of this research data, the superiority of Support Vector Machine (SVM) and Synthetic Minority Over-sampling Technique (SMOTE) lies in their ability to address specific challenges encountered in sentiment analysis tasks. SVM, known for its effectiveness in handling high-dimensional data and capturing complex decision boundaries, offers robust classification performance, especially in scenarios with non-linear separability between sentiment classes. Additionally, SMOTE is a valuable tool for addressing class imbalance by generating synthetic samples for the minority class, thereby improving the classifier's ability to discern nuanced
sentiment patterns. The utilization of SVM and SMOTE in this research enhances the reliability and accuracy of sentiment classification, enabling more precise insights into sentiment dynamics within the analyzed textual data.

2.2.4 Evaluation

During the evaluation stage, toxicity analysis is assessed based on the average and maximum scores of various toxicity metrics, including Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat. This comprehensive evaluation approach enables a thorough examination of the toxicity levels in the analyzed text data, considering multiple dimensions of harmful content. By analyzing these toxicity metrics' average and maximum values, researchers can gain insights into the overall toxicity distribution and identify potential extreme toxicity within the dataset. Consequently, this systematic evaluation process facilitates a nuanced understanding of the toxicity dynamics and aids in developing effective strategies for mitigating harmful content in digital environments.

In the context of sentiment classification, the performance of a Support Vector Machine (SVM) is assessed based on various metrics, including accuracy, precision, recall, F-measure, and Area Under the Curve (AUC). These evaluation metrics provide a comprehensive assessment of the effectiveness of SVM in accurately classifying sentiment labels. Accuracy measures the overall correctness of sentiment predictions, while precision quantifies the proportion of correctly classified positive instances out of all instances predicted as positive. Conversely, Recall calculates the proportion of correctly classified positive instances out of all positive instances in the dataset. F-measure combines precision and recall into a single metric, providing a balanced assessment of the classifier's performance. Additionally, AUC measures the ability of the SVM classifier to distinguish between positive and negative sentiment classes, offering insights into its discriminatory power. Together, these evaluation metrics offer a robust framework for assessing the suitability and effectiveness of SVM in sentiment classification tasks, enabling informed decision-making in model selection and deployment.

2.2.5 Deployment

In the deployment phase of toxicity analysis and sentiment classification results, the focus lies on integrating the developed models and insights into practical applications or platforms. This process involves implementing the trained toxicity analysis and sentiment classification algorithms into relevant systems or software frameworks, ensuring seamless integration and functionality in real-world scenarios. Additionally, rigorous testing and validation procedures are conducted to assess the deployed models' reliability, accuracy, and scalability across diverse contexts and datasets. Through systematic deployment strategies, such as continuous monitoring, feedback mechanisms, and model retraining, the research outcomes can effectively address pertinent challenges related to toxicity detection and sentiment analysis in digital environments, thereby contributing to the advancement of online safety measures and user experience enhancement.

Following the deployment phase, recommendations for further research aim to address emerging challenges and opportunities identified while implementing toxicity analysis and sentiment classification models. Subsequent investigations could focus on refining existing algorithms to enhance their accuracy and efficiency in detecting nuanced forms of toxicity and sentiment in digital content. Furthermore, exploring novel approaches, such as deep learning architectures or ensemble methods, may offer insights into improving model performance across diverse datasets and platforms. Additionally, conducting longitudinal studies to monitor the evolution of online toxicity and sentiment patterns over time could provide valuable insights for developing proactive mitigation strategies and fostering a safer digital environment. Through continued research efforts, informed by real-world deployment experiences, advancements can be made toward more effective and robust solutions for managing online content moderation and sentiment analysis.

3. RESULT AND DISCUSSION

The discussion in this research is bifurcated into two main segments: the interpretation of toxicity analysis outcomes and the interpretation of sentiment classification results utilizing the Support Vector Machine (SVM) and Synthetic Minority Over-sampling Technique (SMOTE). The first segment involves analyzing the findings from toxicity analysis to identify prevalent toxic behaviors, language patterns, and potential sources of harmful content within the dataset. Meanwhile, the second segment entails deciphering the sentiment classification outcomes derived from the SVM and SMOTE models, elucidating the effectiveness of these techniques in discerning sentiment polarity and detecting sentiment trends. Through these comprehensive interpretations, researchers can gain deeper insights into toxicity and sentiment dynamics in digital environments, paving the way for informed decision-making and proactive measures to promote online safety and enhance user experiences.

3.1 Toxicity Analysis

Toxicity analysis is a pivotal tool in discerning the audience's response to Wonderland Indonesia videos, enabling the identification and evaluation of potentially harmful or inappropriate content within the viewers' interactions. By employing sophisticated algorithms and natural language processing techniques, toxicity analysis facilitates
the detection of toxic language, offensive remarks, and other negative expressions in viewer comments and engagements. Through this systematic approach, researchers can gain valuable insights into the overall sentiment and reception of Wonderland Indonesia videos, guiding content creators and platform administrators in fostering a safer and more conducive online environment.

**Figure 7.** Toxicity Analysis of Wonderland Indonesia Chapter One (Communalytic)

Figure 7 shows the toxicity value of the first video. The results of toxicity value identification in the first video provide valuable insights into the toxicity metrics embedded within the content. Specifically, the Toxicity score is determined to be 0.05290, with a corresponding confidence interval ranging from 0.90780. Similarly, Severe Toxicity is quantified at 0.01441, accompanied by a confidence interval spanning 0.95021. Further analysis reveals Identity Attack, Insult, Profanity, and Threat metrics, each exhibiting distinct toxicity scores alongside corresponding confidence intervals. These findings offer a comprehensive understanding of the toxicity levels inherent in digital content, facilitating informed decision-making processes concerning content moderation and audience engagement strategies.

**Figure 8.** Toxicity Analysis of Wonderland Indonesia Chapter Two (Communalytic)

Figure 8 shows the toxicity value of the second video. The results of toxicity value identification in the second video reveal pertinent insights into various toxicity metrics. Specifically, the Toxicity score is recorded at 0.04744 with a confidence interval ranging from 0.93383. Similarly, the Severe Toxicity value is computed at 0.01386, accompanied by a confidence interval spanning 0.95327. Further analysis uncovers Identity Attack, Insult, Profanity, and Threat metrics, each exhibiting distinct toxicity scores alongside corresponding confidence intervals. These findings provide a comprehensive understanding of the toxicity levels embedded within the digital content, facilitating informed decision-making processes regarding content moderation and audience engagement strategies.

Comparing the toxicity values between the first and second videos reveals some notable differences. While both videos exhibit toxicity across various metrics, the levels and patterns of toxicity vary. In the first video, the Toxicity score is slightly lower at 0.05290 compared to 0.04744 in the second video, indicating marginally higher overall toxicity in the latter. However, the Profanity metric in the first video shows a higher score of 0.04815 compared to 0.04735 in the second video, suggesting a slightly elevated presence of profane language. Conversely, the Severe Toxicity metric in the second video has a lower score of 0.01386 compared to 0.01441 in the first video, indicating a slightly reduced prevalence of severe toxic content. Overall, while both videos display toxicity, the specific patterns and levels of toxicity vary, highlighting the importance of tailored content moderation strategies based on individual video characteristics.

### 3.2 Sentiment Classification

Sentiment classification is crucial in identifying the audience's sentiment towards Wonderland Indonesia videos published via the Alffy Rev channel; however, the algorithms utilized as classification models necessitate performance evaluation. This essential step ensures the reliability and accuracy of sentiment predictions, allowing for informed insights into viewer perceptions and preferences. By rigorously testing the performance of classification algorithms, researchers can validate the effectiveness of sentiment classification in capturing nuanced sentiment nuances and trends within viewer interactions. Consequently, this systematic evaluation process
contributes to the refinement of sentiment analysis methodologies, enhancing their applicability in deciphering audience sentiments in Wonderland Indonesia content dissemination.

**Figure 9. Support Vector Machine Performance**

Figure 9 shows the SVM performance with and without SMOTE. Based on the results of sentiment classification, the performance of the SVM algorithm utilizing the SMOTE is as follows: accuracy: 96.50%, Area Under the Curve (AUC): 0.997, precision: 95.75%, recall: 99.00%, and F-measure: 97.34%. These metrics signify the effectiveness and robustness of the SVM model trained with SMOTE in accurately discerning sentiment polarity and classifying sentiments within the analyzed dataset. The high accuracy, AUC, precision, recall, and F-measure values demonstrate the model’s capability to reliably predict sentiment labels and distinguish between positive and negative sentiments, highlighting its suitability for sentiment analysis tasks in digital content moderation and user engagement monitoring.

The performance of the SVM algorithm without utilizing SMOTE is as follows: accuracy: 98.90%, AUC: 0.642, precision: 99.07%, recall: 99.83%, and F-measure: 99.45%. These metrics indicate the effectiveness of the SVM model in accurately classifying sentiment labels within the analyzed dataset, achieving high accuracy, precision, recall, and F-measure values. However, the relatively lower AUC value suggests potential limitations in the model's ability to distinguish between positive and negative sentiment classes compared to the SVM model trained with SMOTE. Despite this, the SVM algorithm without SMOTE demonstrates robust performance in sentiment analysis tasks, showcasing its utility in digital content moderation and sentiment classification applications.

**Figure 10. Area Under Curve of SVM with and without SMOTE**

Figure 10 shows the AUC of SVM with and without SMOTE. The value of the Area Under the Curve (AUC) for the Support Vector Machine (SVM) algorithm with Synthetic Minority Over-sampling Technique (SMOTE) is 0.997, whereas without SMOTE, it is 0.642. This substantial difference in AUC values underscores the significant impact of SMOTE on improving the SVM model's ability to discriminate between positive and negative sentiment classes. The higher AUC value achieved with SMOTE indicates enhanced performance and better classification accuracy, highlighting the importance of employing oversampling techniques like SMOTE to address class imbalance issues and improve the effectiveness of sentiment classification algorithms.

The disparity in Area Under the Curve (AUC) values serves as a crucial metric in evaluating the performance of classification models, indicating the effectiveness of different methodologies in discriminating between classes. In the context of sentiment analysis, where accurately distinguishing between positive and negative sentiments is paramount, the difference in AUC values reflects the impact of various techniques, such as oversampling methods like Synthetic Minority Over-sampling Technique (SMOTE), on improving classification...
accuracy. A higher AUC value suggests superior model performance in correctly identifying sentiments, emphasizing the importance of employing suitable techniques to enhance the efficacy of sentiment classification algorithms and bolster their utility in real-world applications.

4. CONCLUSION
In conclusion, the comparative analysis of toxicity values between the first and second videos demonstrates distinct patterns: the first video showcases a Toxicity score of 0.05290, with notable metrics such as Profanity registering at 0.04815. Conversely, the second video exhibits a slightly lower Toxicity score of 0.04744, yet with varying metrics, such as Severe Toxicity at 0.01386. These findings emphasize the necessity of tailored content moderation strategies based on specific toxicity profiles to ensure a safer digital environment. In addition, these research findings also highlight the effectiveness of sentiment classification using a Support Vector Machine (SVM) with a Synthetic Minority Over-sampling Technique (SMOTE), as evidenced by the achieved performance metrics. With an accuracy of 96.50%, precision of 95.75%, recall of 99.00%, and F-measure of 97.34%, the SVM-SMOTE model demonstrates robust capabilities in discerning sentiment polarity within digital content. Notably, the substantially higher Area Under the Curve (AUC) value of 0.997 with SMOTE compared to 0.642 underscores the importance of addressing class imbalance for improved sentiment classification accuracy. These results underscore the significance of leveraging advanced techniques like SVM and SMOTE in sentiment analysis tasks, offering valuable insights for applications in content moderation, market sentiment analysis, and customer feedback interpretation. Further research may explore optimization strategies and feature integration to enhance the performance of sentiment classification models even further.

ACKNOWLEDGMENT
I wish to extend our sincere acknowledgments to the Tourism Department, the Faculty of Business Administration and Communication, the Research and Community Service Institute (LPPM), and the Atma Jaya Catholic University of Indonesia for their generous support in facilitating the publication of this research.

REFERENCES


