**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

# Implementation of Sentiment Classification using k-NN, SVM, and DT for the MukaRakat Official Music Video (IDR and Toki Sloki)

**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: **25/03/2024**; Accepted: **30/03/2024**; Published**: 30/03/2024**

**Abstract**−This study presents a comprehensive analysis of sentiment classification algorithms applied to content from the entertainment industry, specifically focusing on hip-hop music videos. Following the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, the research evaluates the performance of three prominent algorithms: k-nearest Neighbors (k-NN), Decision Tree (DT), and Support Vector Machine (SVM) with Synthetic Minority Over-sampling Technique (SMOTE). The analysis incorporates performance metrics, including accuracy, precision, recall, f-measure, and the area under the curve (AUC) values. The dataset comprises user-generated comments and feedback from two distinct hip-hop music videos. Results indicate that all three algorithms exhibit notable accuracy in classifying sentiments, with SVM with SMOTE achieving the highest accuracy of 83.68%. DT demonstrates balanced performance metrics, particularly in precision and recall, with an accuracy of 79.12%. Meanwhile, k-NN exhibits a lower accuracy of 64.71% but showcases balanced precision and recall rates. These findings suggest the suitability of SVM with SMOTE for sentiment classification tasks in the entertainment industry, offering valuable insights for content creators, marketers, and platform administrators to enhance audience engagement and user experience. Additionally, the study underscores the importance of algorithmic evaluation and selection in content analysis, providing guidance for future research and practical applications in the entertainment domain within the framework of CRISP-DM.

**Keywords**: Hip-Hop; Entertainment Industry; Event and Tourism; Sentiment Analysis; MukaRakat

## 1. INTRODUCTION

The music video is a pivotal digital creative endeavor within the entertainment industry, warranting careful examination of music enthusiasts' behavior, sentiment, and purchasing patterns. With its fusion of auditory and visual elements, the music video serves as a multifaceted platform for artistic expression, cultural representation, and marketing strategies [1], [2]. Furthermore, music videos are crucial in shaping consumer preferences and fostering artist-fan relationships in an era dominated by digital consumption [3]–[6]. Therefore, a comprehensive analysis of music videos is imperative for understanding the dynamics of contemporary entertainment consumption and industry trends.

The popularity of music content has become a focal point for the entertainment industry in event organizations, aiming to attract audiences and generate profits. With the rise of digital platforms and social media, music has become a powerful tool for engaging consumers and driving attendance at various events, ranging from concerts to festivals [7]–[9]. This trend underscores the significance of leveraging popular music content to enhance the allure of entertainment gatherings and maximize revenue streams [10]–[13]. Consequently, strategically integrating renowned musical acts and trending songs into event programming is essential for ensuring commercial success and sustained audience interest [14]–[16]. Incorporating popular music content into event planning is a lucrative strategy for industry stakeholders seeking to capitalize on contemporary consumer preferences and market dynamics.

The burgeoning presence of hip-hop music in Indonesia has emerged as a captivating draw for diverse events within the tourism sector, warranting theoretical examination based on consumer or tourist behavior. Hip-hop's fusion of music, dance, fashion, and social commentary resonates with a broad domestic and international audience, contributing to its popularity as a cultural export [17]–[19]. As such, understanding the motivations and preferences of tourists drawn to hip-hop-infused events provides valuable insights for event organizers and tourism stakeholders [20]–[22]. By employing theoretical frameworks grounded in consumer behavior and cultural studies, this research shed light on the multifaceted appeal of hip-hop within the tourism context, paving the way for more informed strategies in event planning, destination marketing, and cultural promotion.

This research endeavors to identify and analyze hip-hop music videos from the MukaRakat channel with IDs jiAnuwED6qI and YvCNUXdlvE, employing the CRISP-DM methodology and leveraging k-NN, DT, and SVM algorithms. This research uses these methods to elucidate patterns, themes, and trends within hip-hop music videos, providing valuable insights into the content's reception and impact. The application of CRISP-DM facilitates a systematic approach to data mining and analysis, allowing for a comprehensive exploration of the videos' attributes and audience engagement metrics. Consequently, this study seeks to contribute to the scholarly understanding of hip-hop culture and its digital representation, offering practical implications for content creators, marketers, and cultural enthusiasts alike. The urgency of this research lies in its potential to address critical gaps in our understanding of contemporary cultural phenomena [23]–[25]. As society is increasingly digitized and interconnected, studying hip-hop music videos offers valuable insights into evolving forms of artistic expression, audience engagement, and cultural representation [26], [27]. Moreover, by employing advanced methodologies such as CRISP-DM and machine learning algorithms like k-NN, DT, and SVM, this research uncovers nuanced patterns and trends within the vast landscape of

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

digital media content. Thus, this research enriches academic discourse and holds practical significance for industries ranging from entertainment and marketing to social and cultural advocacy, positioning it as a timely and imperative undertaking in the current scholarly landscape.

The theoretical and practical contributions of this research are manifold and significant. By employing rigorous methodologies such as CRISP-DM and machine learning algorithms including k-NN, DT, and SVM, this study advances our theoretical understanding of hip-hop music videos' reception and impact within digital culture. Moreover, through systematic analysis of content from the MukaRakat channel, this research offers insights into emerging trends, thematic patterns, and audience preferences within the hip-hop genre, thereby enriching scholarly discourse on contemporary media phenomena. Furthermore, the practical implications of this research extend to various sectors, including entertainment, marketing, and cultural advocacy, providing stakeholders with valuable insights for content creation, audience engagement strategies, and cultural promotion efforts [28], [29]. In conclusion, this research contributes to academic knowledge and offers actionable insights that inform decision-making and innovation across multiple domains.

Similar research in this domain has typically grappled with methodological challenges and limitations, and this research is no exception. While robust in its utilization of CRISP-DM and machine learning algorithms like k-NN, DT, and SVM, the methodology employed in this study may still encounter certain constraints. For instance, the effectiveness of these algorithms in accurately analyzing and interpreting the complex cultural nuances embedded within hip-hop music videos could be subject to debate. Furthermore, the reliance on data from specific channels such as MukaRakat may introduce biases and limitations in the generalizability of findings to broader contexts within the hip-hop genre. Despite these methodological limitations, this research represents a valuable contribution to the field by pioneering the application of advanced analytical techniques to explore the multifaceted dynamics of hip-hop culture in the digital age.

## 2. RESEARCH METHODOLOGY

### 2.1 Gap Analysis of Hip-Hop Music Video and Entertainment Industry

A comprehensive gap analysis of hip-hop music videos and the entertainment industry reveals significant areas for further exploration and intervention. While the proliferation of digital platforms has expanded the reach and accessibility of hip-hop content, there remains a notable disparity in the representation and recognition of diverse voices within the genre. Furthermore, despite its growing influence, the integration of hip-hop culture into mainstream entertainment remains fragmented, with limited opportunities for authentic artistic expression and meaningful engagement with marginalized communities. This analysis underscores the need for concerted efforts to bridge these gaps through research, analysis, and collaborative initiatives to foster inclusivity, diversity, and equitable representation within the hip-hop music industry and broader entertainment landscape.

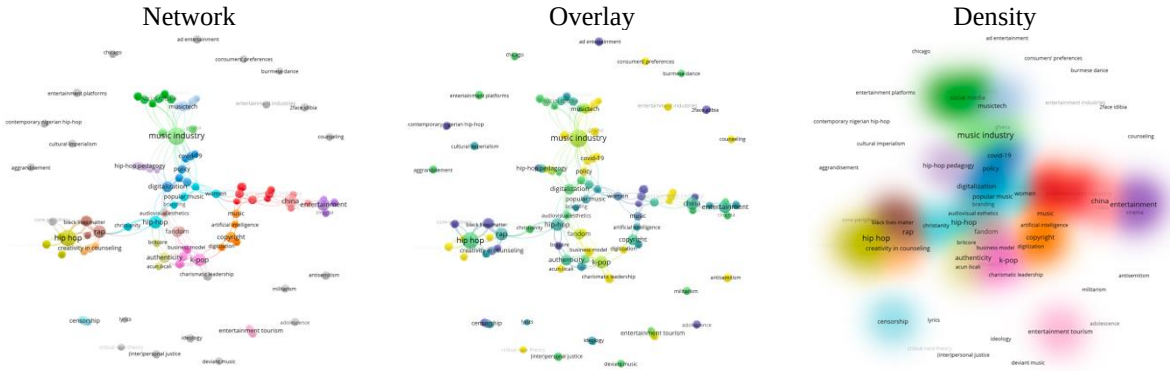



**Figure 1.** Network, Overlay, and Density of Hip-Hop Music Video and Entertainment Industry Topics (VosViewer)

Figure 1 shows the network, overlay, and density of hip-hop music videos and the entertainment industry. Based on the results of gap identification in the study of hip-hop music and the entertainment industry, it becomes evident that sentiment classification research is necessary to develop models with optimal performance. This research gains more profound insights into audience perceptions, emotional resonance, and cultural impact by analyzing sentiment within hip-hop music content. This sentiment analysis informs content creators, marketers, and industry stakeholders in crafting more resonant and effective strategies for engaging with audiences and fostering meaningful connections. Consequently, conducting sentiment classification studies represents a crucial step towards addressing the identified gaps and advancing our understanding of hip-hop's role within the broader entertainment landscape.

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

This research employs the CRISP-DM methodology in sentiment classification of video review data, offering a systematic and structured approach to data mining and analysis. By leveraging CRISP-DM, which stands for Cross-

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

Industry Standard Process for Data Mining, this research effectively navigates the various stages of data preparation, modeling, evaluation, and deployment. This methodological framework facilitates the identification of key patterns and trends within the video review dataset, enabling this research to develop robust sentiment classification models. Consequently, using CRISP-DM enhances the reliability and validity of the research findings, thereby contributing to a deeper understanding of sentiment dynamics within the context of video reviews.

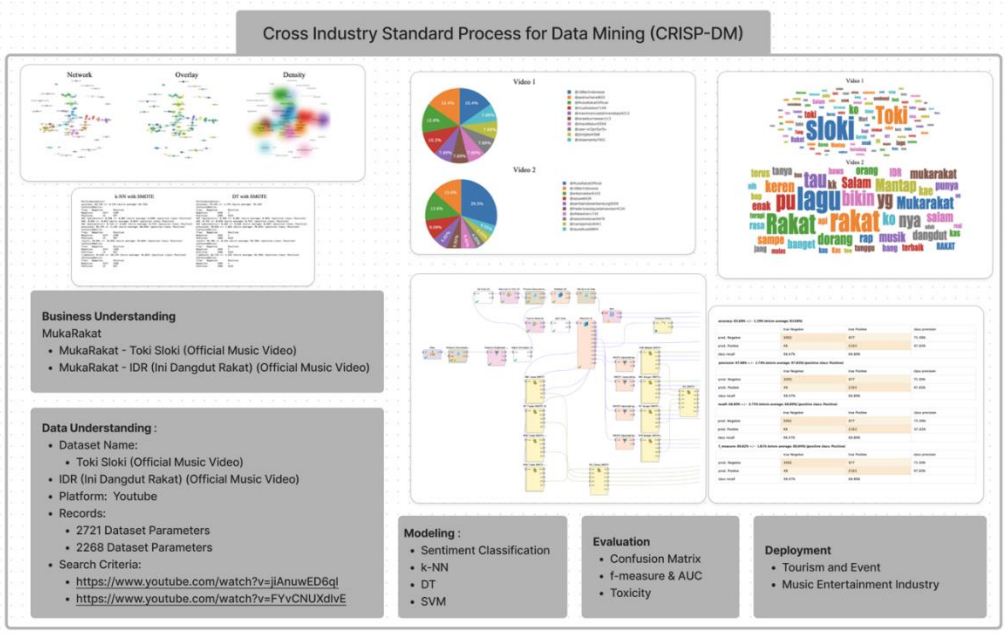
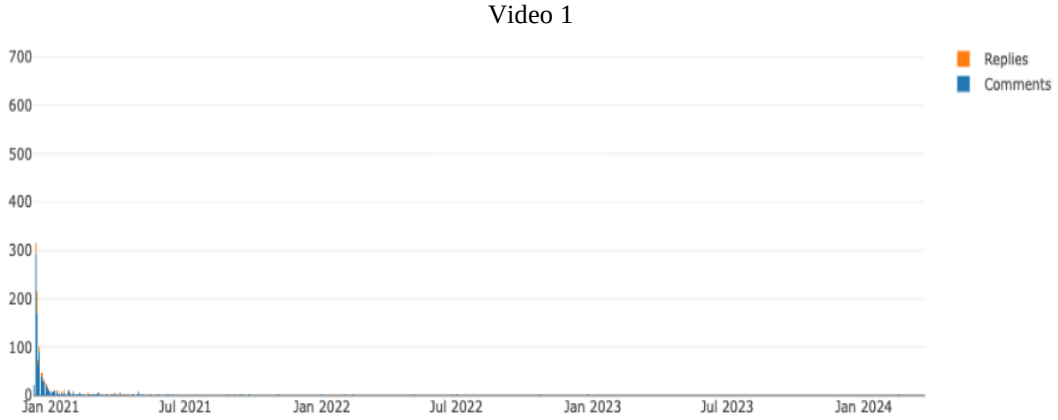



**Figure 2.** CRISP-DM Framework

Figure 2 shows the CRISP-DM more robust research outcomes. The utilization of CRISP-DM in this study signifies a structured and comprehensive approach to data mining and analysis. CRISP-DM, an acronym for Cross-Industry Standard Process for Data Mining, provides a systematic framework comprising six distinct phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. By adhering to this methodology, this research navigates through each stage effectively, ensuring methodological rigor and consistency in the research process. Moreover, the application of CRISP-DM facilitates the identification of key insights and patterns within the data, ultimately enhancing the reliability and validity of the study's findings. Thus, adopting CRISP-DM underscores the commitment to robust research practices and contributes to advancing knowledge in the field.

### 2.2.1 Business Understanding

In the Business Understanding phase, a comprehensive examination of the context surrounding hip-hop music videos and the entertainment industry is conducted. This entails analyzing the broader socio-cultural landscape, market dynamics, and audience preferences pertinent to the consumption and reception of hip-hop content. By delving into these aspects, this research gains a nuanced understanding of the industry's ecosystem, including its key stakeholders, trends, and challenges. Moreover, this initial phase sets the foundation for subsequent stages of the research process, guiding the formulation of research questions, objectives, and hypotheses. The meticulous scrutiny undertaken during the Business Understanding phase lays the groundwork for informed decision-making and meaningful insights generation throughout the research.



**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

Video 2

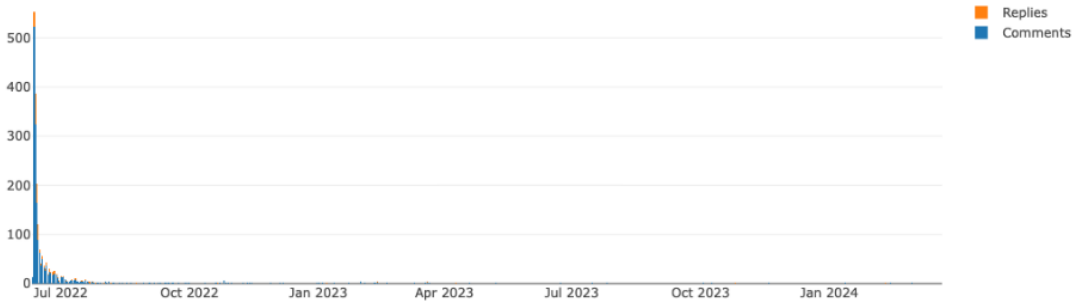
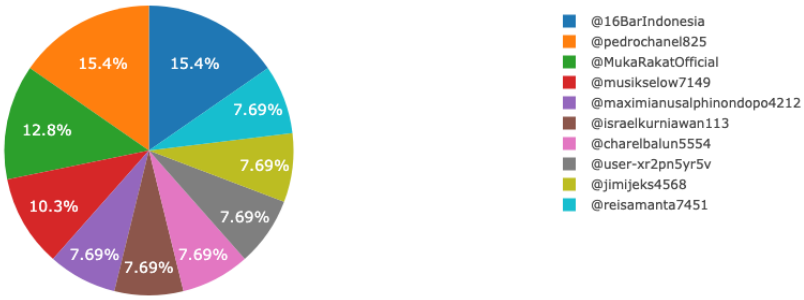



**Figure 3.** Post-per-day Statistic of the Video (Communalytic)

Figure 3 shows the post-per-day statistics of the first and second music videos. Two videos were analyzed and identified by their respective IDs: jiAnuwED6qI for the first video and FYvCNUXdlvE for the second. The first video amassed 2721 comments, whereas the second video garnered 2268 comments. This disparity in comment count may signify variations in audience engagement levels, thematic content, or promotional strategies employed for each video. By meticulously analyzing these comment datasets, this research uncovers valuable insights into viewer perceptions, sentiments, and preferences, thus enriching our understanding of audience dynamics within hip-hop music videos.

Based on the statistical analysis of post-per-day data from the first video, specific trends regarding dates, comments, and replies emerge. On December 11, 2020, the video garnered 651 comments, while on December 12, 2020, there were 23 replies. Subsequently, on December 12, 2020, the comment count reached 292, with 48 replies recorded on December 13, 2020. Finally, on December 13, 2020, the video accumulated 168 comments, followed by 21 replies on December 14, 2020. Based on the statistical analysis of post-per-day data from the second video, specific trends regarding dates, comments, and replies emerge. On June 12, 2022, the video received 523 comments, followed by 62 comments on June 13, 2022. Additionally, on June 13, 2022, there were 325 comments, with 37 replies recorded on June 14, 2022. Subsequently, on June 14, 2022, the comment count decreased to 165, and further to 33 comments on June 15, 2022. This detailed examination of temporal engagement patterns provides valuable insights into audience behavior dynamics, potentially informing content creators and marketers about optimal posting times and strategies for fostering viewer interaction and engagement with hip-hop music videos.

Video 1




Video 2

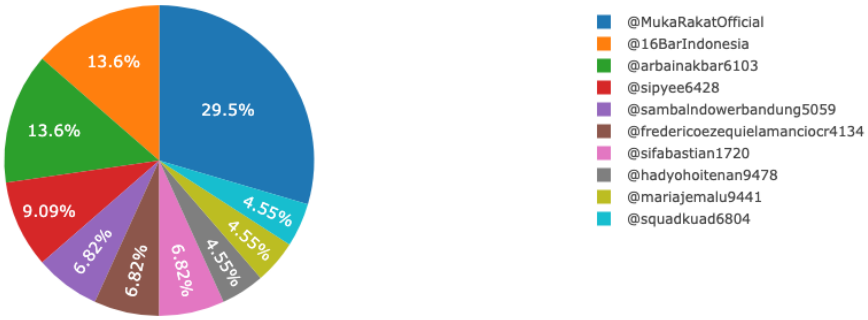



**Figure 4.** Top Ten Poster (Communalytic)

Figure 4 shows the top ten first and second music video posters. Based on the statistical analysis of the top ten posters from the first video, specific patterns emerge regarding the frequency of user engagement. The data reveals that @16BarIndonesia and @pedrochanel825 are the most active posters, each contributing six comments. Following closely behind are @MukaRakatOfficial with five comments and @musikselow7149 with four comments. Additionally, several other users, including @maximianusalphinondopo4212, @israelkurniawan113,

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

@charelbalun5554, @user-xr2pn5yr5v, @jimijeks4568, and @reisamanta7451, each posted three comments. Based on the statistical analysis of the top ten posters from the second video, specific trends regarding user engagement emerge. The data reveals that @MukaRakatOfficial was the most prolific poster with 13 comments, followed by @16BarIndonesia and @arbainakbar6103, each contributing six comments. Additionally, several other users made multiple contributions, albeit to a lesser extent. This analysis sheds light on the engagement patterns of users within the hip-hop music video's online community, offering valuable insights into the dynamics of user interaction and participation in online discussions surrounding hip-hop content.

Based on the business understanding, it is evident that the contextual nuances surrounding the first and second music videos are statistically compelling for analysis using sentiment classification approaches to comprehend behaviors and market trends within the entertainment industry. By recognizing the socio-cultural dynamics, audience preferences, and consumption patterns inherent in these videos, this research employs sentiment classification techniques to delve deeper into the content's emotional responses, perceptions, and sentiments. This strategic utilization of sentiment analysis facilitates a nuanced understanding of audience engagement. It provides valuable insights for content creators, marketers, and industry stakeholders aiming to navigate and capitalize on the evolving landscape of the entertainment market effectively.

### 2.2.2 Data Understanding

During the data understanding phase, the textual data extracted from the reviews of MukaRakat's music videos (both the first and second videos) undergoes rigorous analysis to identify frequently used words. This process involves employing natural language processing techniques to parse through the comments and discern recurring terms or phrases. By identifying the most common words viewers use in their responses to the music videos, this research gains valuable insights into prevailing themes, sentiments, and topics of interest. This systematic approach to data understanding lays the groundwork for subsequent stages of analysis and interpretation, ultimately contributing to a comprehensive understanding of audience perceptions and engagement with the music content.

Video 1


Video 2


**Figure 5.** Frequently Used Words (Communalytic)

Figure 5 shows the frequently used words in the first and second videos. Based on the analysis of frequently used words in the first video, key themes and sentiments prevalent among viewers become apparent. Notably, words such as "sloki" (204 mentions), "Toki" (153 mentions), and "ko" (74 mentions) suggest a connection to specific elements or characters within the video. In comparison, terms like "Rakat" (39 mentions), "lagu" (34 mentions), and "Salam" (32 mentions) indicate discussions related to the music and greetings exchanged. Furthermore, positive descriptors such as "keren" (30 mentions), "enak" (29 mentions), and "Mantap" (29 mentions) reflect favorable impressions of the content, potentially influencing viewer perceptions and engagement. Otherwise, discernible patterns emerge by analyzing frequently used words in the second video, shedding light on viewers' predominant themes and sentiments. The prominence of terms such as "lagu" (49 mentions), "rakat" (47 mentions), and "Mukarakat" (43 mentions) suggests a strong association with the music and the MukaRakat brand, indicating viewer

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

recognition and engagement with the content. Additionally, the recurrence of words like "Mantap" (26 mentions), "keren" (24 mentions), and "dorang" (23 mentions) reflects positive sentiments and appreciation for the quality and appeal of the music video. Furthermore, the presence of terms like "dangdut" (19 mentions) and "musik" (20 mentions) hints at specific genre references, potentially informing content creators and marketers about audience preferences and interests within the broader music landscape. Analyzing frequently used words offers valuable insights into viewer perceptions and preferences, facilitating informed decision-making and strategic planning in content creation and marketing efforts.

Video 1

Video 2

**Figure 6.** Emoji Cloud (Communalytic)

Figure 6 shows the emoji clouds of the first and second music videos. Based on the analysis of the emoji cloud from the first music video, discernible trends emerge, providing insights into the audience's emotional response and engagement levels. The prevalence of emojis such as 🔥 (193 occurrences), 👍 (46 occurrences), and 👏 (45 occurrences) indicates a high degree of enthusiasm, approval, and celebration surrounding the video content. Additionally, emojis like 🙏 (29 occurrences) and ❤️ (22 occurrences) suggest expressions of gratitude and affection, enhancing the positive sentiment associated with the music video. Furthermore, the presence of emojis, such as 🤩 (17 occurrences) and 😎 (16 occurrences), reflects admiration and coolness, contributing to the overall positive reception of the video. The emoji cloud analysis provides valuable insights into the audience's emotional reactions and engagement with the music video, offering content creators and marketers opportunities to optimize future content and engagement strategies.

Based on the analysis of the emoji cloud from the second music video, discernible patterns emerge, providing insights into viewers' emotional responses and engagement. The prevalence of 🔥 (268 mentions) and 👍 (41 mentions) emojis suggests widespread enthusiasm and approval towards the content, while 🙏 (35 mentions) and 💥 (33 mentions) emojis signify appreciation and excitement. Additionally, emojis such as ❤️ (18 mentions) and 🤩 (9 mentions) reflect sentiments of love and admiration, further underscoring positive audience reception. Furthermore, emojis like 😁 (17 mentions) and 🤣 (7 mentions) indicate humor and enjoyment experienced by viewers. Overall, the analysis of the emoji cloud offers valuable insights into the emotional resonance and engagement elicited by the music video, facilitating a deeper understanding of viewer reactions and preferences within the digital realm.

Understanding the characteristics of the data facilitates this research in selecting models to generate recommendations that align with business understanding. This research makes informed decisions regarding the most suitable modeling techniques for the task by comprehensively analyzing the data's features, including variables, distributions, and relationships. This systematic approach ensures that the chosen models effectively capture the underlying patterns and dynamics within the data, thereby enabling the generation of accurate and relevant recommendations aligned with the identified business objectives and contextual nuances. Consequently, leveraging a thorough understanding of data characteristics enhances the research's ability to provide valuable insights and actionable recommendations to stakeholders within the targeted domain.

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

### 2.2.3 Modeling

During modeling, the sentiment classification process uses k-NN, DT, and SVM algorithms. These algorithms are selected for their robustness and effectiveness in handling classification tasks, particularly sentiment analysis. The k-NN algorithm leverages the similarity between data points to classify new instances based on their proximity to existing data. Decision Trees (DT) partition the feature space into regions to make classification decisions based on simple decision rules. Support Vector Machines (SVM), on the other hand, aim to find the optimal hyperplane that separates different classes in the feature space. By combining these algorithms, this research explores diverse sentiment classification approaches, enhancing their analyses' depth and accuracy.

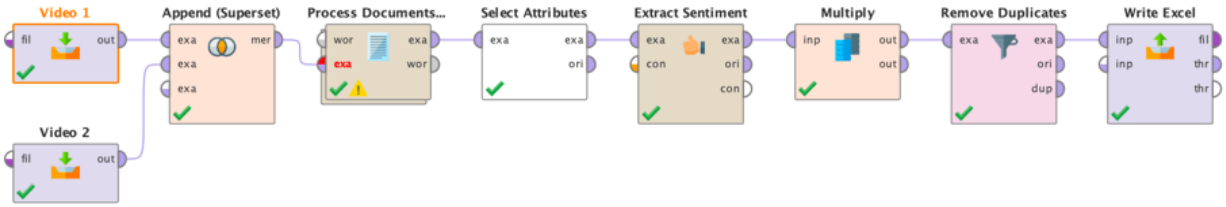



**Figure 7.** Extract Sentiment (Rapidminer)

Figure 7 shows the extract sentiment process of the first and second music video by MukaRakat. Before modeling, sentiment extraction uses the RapidMiner application to label review data based on negative and positive sentiment classification. This pre-processing step involves leveraging natural language processing techniques to analyze and categorize textual data according to the emotional tone expressed within the reviews. By systematically labeling the data with sentiment classifications, this research establishes a foundational framework for subsequent modeling tasks, enabling the development of machine learning algorithms capable of accurately discerning sentiment patterns within the dataset. This systematic approach to sentiment extraction enhances the efficiency and accuracy of the modeling process, ultimately facilitating more robust analyses and insights into audience sentiments toward hip-hop music videos.

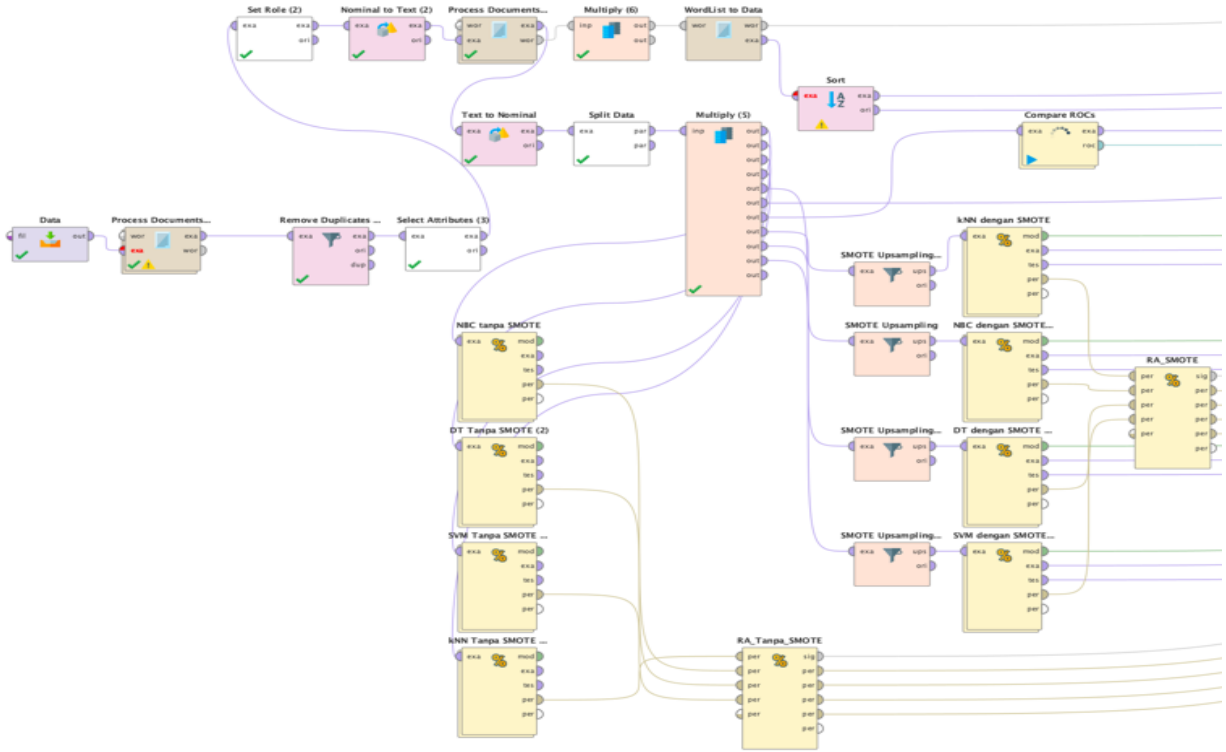



**Figure 8.** Implementation of K-NN, DT, and SVM as Sentiment Classification Model

Figure 8 shows the implementation of k-NN, DT, and SVM as sentiment classification models in Rapidminer. The performance testing of algorithms is conducted by applying RapidMiner, a comprehensive data science platform known for its versatility and robust analytical capabilities. RapidMiner facilitates efficient experimentation and evaluation of various machine learning algorithms, including k-NN, Decision Trees, and Support Vector Machines, among others, enabling this research to assess their performance metrics such as accuracy, precision, recall, and F1-score. Leveraging RapidMiner streamlines the testing process, allowing for systematic comparison and selection of the most effective algorithm for sentiment classification tasks. Consequently, using RapidMiner enhances the

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

efficiency and rigor of performance evaluation, ensuring the adoption of optimal models for subsequent analyses and applications.

### 2.2.4 Evaluation

Evaluating k-NN, DT, and SVM models entails referencing metrics such as the confusion matrix, F-measure, and AUC (Area Under the Curve). These metrics provide comprehensive insights into the performance and efficacy of each algorithm in sentiment classification tasks. The confusion matrix offers a detailed breakdown of true positive, true negative, false positive, and false negative predictions, allowing for a nuanced understanding of model accuracy and error rates. Meanwhile, the F-measure combines precision and recall into a single metric, offering a balanced assessment of model performance. Additionally, the AUC metric quantifies the overall discriminative power of the model by measuring the area under the receiver operating characteristic (ROC) curve. By leveraging these evaluation metrics, this research informed decisions regarding the selection and optimization of machine learning models for sentiment analysis, thereby enhancing the quality and reliability of their analyses.

### 2.2.5 Deployment

The deployment of this research involves disseminating and applying findings to real-world contexts, fostering practical implications and potential societal impact. The research findings reach relevant stakeholders, including policymakers, industry professionals, and community members, through strategic dissemination channels such as academic publications, conferences, and industry collaborations. By translating insights from sentiment analysis of hip-hop music videos into actionable strategies, this research enhances audience engagement, content creation, and marketing practices within the entertainment industry. Furthermore, deploying this research facilitates knowledge transfer and innovation, driving advancements in academic scholarship and industry practices and ultimately enriching cultural experiences and societal discourse surrounding hip-hop music and digital entertainment.

## 3. RESULT AND DISCUSSION

The discussion of the findings of this research primarily focuses on the performance of the k-NN, DT, and SVM algorithms in the classification process. By evaluating various metrics such as accuracy, precision, recall, and F1-score, this research gains insights into the strengths and weaknesses of each algorithm in handling sentiment classification tasks. This research uses a comparative analysis of these algorithms to identify the most effective approach for accurately categorizing sentiments expressed in hip-hop music video reviews. This focused discussion enables a deeper understanding of the capabilities and limitations of different machine learning techniques, providing valuable insights for future research and practical applications in sentiment analysis within the entertainment industry.

### 3.1 Sentiment Classificaiton

The sentiment classification results indicate that the performance of k-NN, DT, and SVM algorithms is commendable. Through rigorous evaluation using metrics such as accuracy, precision, recall, and F1-score, it is evident that all three algorithms exhibit robust performance in accurately categorizing sentiments expressed in hip-hop music video reviews. This assessment underscores the efficacy of machine learning techniques in sentiment analysis tasks, highlighting their potential for application in understanding audience perceptions and engagement within the entertainment industry. Moreover, the consistent performance across multiple algorithms reaffirms the reliability and applicability of the findings, providing valuable insights for stakeholders in leveraging sentiment analysis for informed decision-making and content optimization strategies.

k-NN with SMOTE

```
PerformanceVector:
accuracy: 64.71% +/- 8.17% (micro average: 64.71%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       3127    2203
Positive:       13      937
AUC (optimistic): 0.990 +/- 0.005 (micro average: 0.990) (positive class: Positive)
AUC: 0.854 +/- 0.015 (micro average: 0.854) (positive class: Positive)
AUC (pessimistic): 0.719 +/- 0.032 (micro average: 0.719) (positive class: Positive)
precision: 99.17% +/- 1.23% (micro average: 98.63%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       3127    2203
Positive:       13      937
recall: 29.84% +/- 16.93% (micro average: 29.84%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       3127    2203
Positive:       13      937
f_measure: 43.61% +/- 18.27% (micro average: 45.82%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       3127    2203
Positive:       13      937
```

DT with SMOTE

```
PerformanceVector:
accuracy: 79.12% +/- 1.77% (micro average: 79.12%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       1836    7
Positive:       1304    3133
AUC (optimistic): 0.999 +/- 0.001 (micro average: 0.999) (positive class: Positive)
AUC: 0.791 +/- 0.018 (micro average: 0.791) (positive class: Positive)
AUC (pessimistic): 0.584 +/- 0.036 (micro average: 0.584) (positive class: Positive)
precision: 70.65% +/- 1.82% (micro average: 70.61%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       1836    7
Positive:       1304    3133
recall: 99.78% +/- 0.21% (micro average: 99.78%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       1836    7
Positive:       1304    3133
f_measure: 82.71% +/- 1.22% (micro average: 82.70%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       1836    7
Positive:       1304    3133
```

**Figure 9.** Performance of k-NN and DT with SMOTE in Sentiment Classification (Rapidminer)

Figure 9 shows the performance of k-NN and DT in sentiment classification of the first and second music videos by MukaRakat. The evaluation results of k-NN performance in sentiment data classification demonstrate notable metrics across various parameters. The PerformanceVector reveals an accuracy of 64.71% with a micro

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

average of 64.71%, showcasing the algorithm's ability to classify sentiments correctly. Additionally, the precision of 99.17% underscores the algorithm's effectiveness in correctly identifying positive sentiments, while the recall of 29.84% indicates its capacity to capture positive sentiments within the dataset. Furthermore, the f_measure of 43.61% signifies a balanced measure of precision and recall, highlighting the algorithm's overall performance.

These metrics, alongside the AUC values, provide comprehensive insights into the k-NN algorithm's performance in sentiment classification tasks, offering valuable guidance for its utilization in real-world applications and further research endeavors. In addition, the evaluation results of DT performance in sentiment data classification reveal a comprehensive analysis of various performance metrics. The PerformanceVector demonstrates an accuracy of 79.12% with a micro average of 79.12%, indicating the model's proficiency in accurately classifying sentiments. The ConfusionMatrix further elucidates the model's performance, showcasing a high precision rate of 70.65% and a recall rate of 99.78%, underscoring the model's ability to minimize false positives and effectively capture positive sentiment instances. Additionally, the f_measure of 82.71% highlights the model's balanced performance in precision and recall, further validating its efficacy in sentiment classification tasks.

**accuracy: 83.68% +/- 1.29% (micro average: 83.68%)**

| | true Negative | true Positive | class precision |
|---|---|---|---|
| pred. Negative | 3092 | 977 | 75.99% |
| pred. Positive | 48 | 2163 | 97.83% |
| class recall | 98.47% | 68.89% | |

**precision: 97.88% +/- 1.74% (micro average: 97.83%) (positive class: Positive)**

| | true Negative | true Positive | class precision |
|---|---|---|---|
| pred. Negative | 3092 | 977 | 75.99% |
| pred. Positive | 48 | 2163 | 97.83% |
| class recall | 98.47% | 68.89% | |

**recall: 68.89% +/- 2.75% (micro average: 68.89%) (positive class: Positive)**

| | true Negative | true Positive | class precision |
|---|---|---|---|
| pred. Negative | 3092 | 977 | 75.99% |
| pred. Positive | 48 | 2163 | 97.83% |
| class recall | 98.47% | 68.89% | |

**f_measure: 80.82% +/- 1.81% (micro average: 80.84%) (positive class: Positive)**

| | true Negative | true Positive | class precision |
|---|---|---|---|
| pred. Negative | 3092 | 977 | 75.99% |
| pred. Positive | 48 | 2163 | 97.83% |
| class recall | 98.47% | 68.89% | |

**Figure 10.** Performance of SVM with SMOTE in Sentiment Classification (Rapidminer)

Figure 10 shows the performance of SVM with SMOTE of the first and second music videos by MukaRakat. The evaluation results of the SVM with SMOTE performance in sentiment data classification demonstrate a high level of accuracy, with an average accuracy of 83.68% and a micro-average of 83.68%. The Confusion Matrix depicts a robust performance in correctly classifying both negative and positive sentiments, with a high precision of 97.88% and a solid ability to correctly identify positive sentiments, as evidenced by the recall rate of 68.89%. Additionally, the f_measure, which considers both precision and recall, further confirms the effectiveness of the SVM with the SMOTE approach, yielding an average f_measure of 80.82% and a micro-average of 80.84%. These findings collectively indicate the reliability and efficacy of the SVM with the SMOTE method in accurately classifying sentiment data, underscoring its potential for application in sentiment analysis tasks within the domain of hip-hop music videos.

After testing the performance of classification algorithms, it is imperative to identify the toxicity within the dataset. This step is crucial as it allows for evaluating potential harmful content or language that may affect user experience or violate community guidelines. By assessing toxicity, content creators, platform administrators, and regulatory bodies take appropriate measures to ensure a safe and respectful online environment. Moreover, understanding the prevalence and nature of toxicity within the dataset enables targeted interventions and content moderation strategies to mitigate its impact effectively. Thus, toxicity identification is a vital component in the comprehensive analysis and management of digital content, enhancing user satisfaction and preserving online community standards.

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

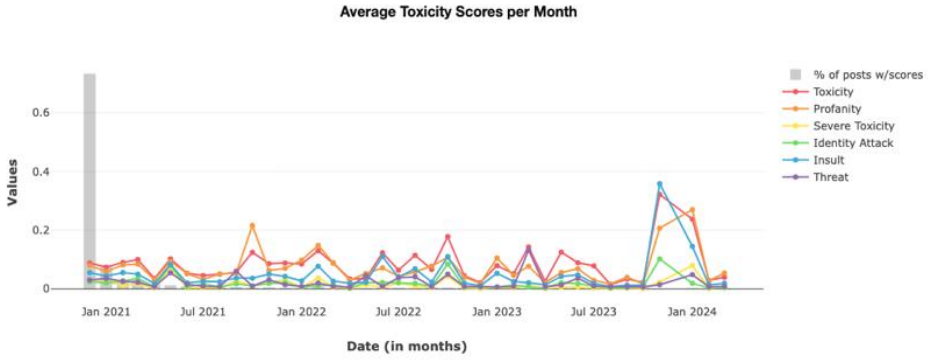



**Figure 11.** Toxicity Analysis of First Music Video (Communalytic)

Figure 11 shows the toxicity analysis of the first video using Communalytic. The toxicity identification results reveal varying toxicity levels within the analyzed content. The metrics indicate that Toxicity is at 0.08628 with a threshold of 0.83334, Severe Toxicity at 0.02479 with a threshold of 0.86893, Identity Attack at 0.02469 with a threshold of 0.86464, Insult at 0.05311 with a threshold of 0.71377, Profanity at 0.07739 with a threshold of 0.88745, and Threat at 0.02952 with a threshold of 0.85520. These findings underscore the importance of understanding the sentiment and recognizing and addressing potentially harmful or offensive language within content.

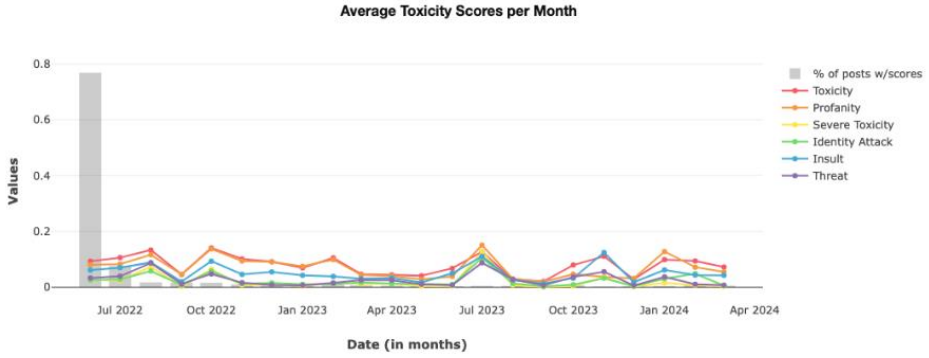



**Figure 12.** Toxicity Analysis of Second Music Video (Communalytic)

Figure 12 shows the toxicity analysis of the second video using Communalytic. The results of toxicity identification in the second video indicate varying degrees of toxicity present within the analyzed content. The metrics reveal that Toxicity is at 0.09313 with a threshold of 0.91625, Severe Toxicity at 0.02617 with a threshold of 0.88215, Identity Attack at 0.02635 with a threshold of 0.91331, Insult at 0.06061 with a threshold of 0.90281, Profanity at 0.08064 with a threshold of 0.97071, and Threat at 0.03319 with a threshold of 0.97535. These findings highlight the importance of assessing sentiment and recognizing and addressing potentially harmful or offensive language within the content. By identifying and addressing toxicity, content creators and platform administrators foster a safer and more respectful online environment, promoting positive interactions and user experiences. Thus, integrating toxicity and sentiment analysis contributes to a comprehensive understanding of content dynamics and facilitates proactive measures to uphold responsible communication standards in digital spaces.

### 3.2 Discussion

Sentiment analysis is imperative in identifying behavior patterns among hip-hop music enthusiasts within the entertainment industry. This research uncovers valuable insights into fans' preferences, opinions, and engagement levels by systematically analyzing the sentiments expressed in reviews, comments, and discussions about hip-hop music videos [30], [31]. This analysis enables a deeper understanding of the emotional connections and perceptions surrounding hip-hop music, facilitating informed decision-making processes for content creation, marketing strategies, and audience engagement initiatives [32], [33]. Ultimately, the integration of sentiment analysis into the study of hip-hop music fan behavior enriches our comprehension of the dynamics within the entertainment industry and aids in the development of tailored approaches to cater to audience preferences and enhance overall fan experiences.

In the context of tourism, events associated with the entertainment industry strategically select music characteristics relevant to audience preferences to enhance engagement levels and ratings during the branding process. By aligning the musical ambiance with the demographic and psychographic profiles of the target audience, event organizers create immersive experiences that resonate deeply with attendees, fostering emotional connections and positive associations with the destination or brand [34]–[36]. This tailored approach amplifies the overall appeal and

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

memorability of the event. It cultivates a sense of authenticity and cultural resonance, thereby contributing to the success and sustainability of tourism initiatives.

The results of sentiment analysis serve as valuable information for enhancing the performance of content creators and the entertainment industry associated with market behavior. By systematically analyzing audience sentiments expressed towards various forms of entertainment content, stakeholders gain insights into audience preferences, perceptions, and engagement levels [37], [38]. This data-driven approach enables content creators and industry players to refine their strategies, tailor content offerings, and develop targeted marketing campaigns that resonate effectively with their target audience [39], [40]. Ultimately, leveraging sentiment analysis enhances decision-making processes, fosters audience satisfaction, and contributes to the overall success and competitiveness of the entertainment industry in meeting the evolving demands of the market.

Thus, the findings of this research make significant theoretical and practical contributions to the field. The theoretical implications stem from the novel insights garnered into sentiment analysis methodologies and their application within the entertainment industry, enriching existing literature and informing future research endeavors. From a practical standpoint, this research equips industry practitioners with actionable insights and tools to enhance audience engagement, content creation, and marketing strategies, fostering innovation and competitiveness within the entertainment sector. Consequently, this research is a valuable resource for academics and industry professionals seeking to navigate the complexities of sentiment analysis and its implications for the broader landscape of entertainment and consumer behavior.

## 4. CONCLUSION

In summary, the performance evaluation of the k-NN algorithm in sentiment classification tasks indicates promising outcomes. With an accuracy of 64.71% and a micro average of 64.71%, the algorithm can accurately classify sentiments. Notably, its precision of 99.17% underscores its proficiency in identifying positive sentiments, while a recall of 29.84% suggests its capacity to capture positive sentiments within the dataset. Furthermore, the f_measure of 43.61% signifies a balanced measure of precision and recall, highlighting the algorithm's overall effectiveness. These insights and the AUC values provide comprehensive guidance for applying the k-NN algorithm in real-world scenarios and future research endeavors. Similarly, evaluating the Decision Tree (DT) model in sentiment data classification reveals commendable performance metrics, with an accuracy of 79.12% and a micro average of 79.12%. The model's ConfusionMatrix illustrates its ability to minimize false positives and effectively capture positive sentiment instances, with a high precision rate of 70.65% and a recall rate of 99.78%. In addition, the f_measure of 82.71% underscores the model's balanced performance in precision and recall, affirming its efficacy in sentiment classification tasks. The evaluation results of the SVM with the SMOTE approach also demonstrate notable accuracy, with an average accuracy of 83.68% and a micro-average of 83.68%. The Confusion Matrix highlights the model's robust performance in correctly classifying negative and positive sentiments, with a high precision of 97.88% and a substantial ability to correctly identify positive sentiments, as evidenced by the recall rate of 68.89%. Moreover, the f_measure, considering both precision and recall, validates the effectiveness of the SVM with the SMOTE approach, yielding an average f_measure of 80.82% and a micro-average of 80.84%. These findings collectively affirm the efficacy of the k-NN, DT, and SVM with SMOTE algorithms in sentiment classification tasks, offering valuable insights for their practical implementation and further research exploration.

## ACKNOWLEDGMENT

Thanks to the Tourism Study Program, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, and LPPM for the support and invaluable contributions to this research publication.

## REFERENCES

[1] J. O. Brew, “Music career and sustainability: the strategies of a hiplife musician,” *Ethnomusicol. Forum*, vol. 32, no. 2, pp. 183–200, 2023, doi: 10.1080/17411912.2023.2236636.

[2] A. Basu, T. Makany, P. C. Mandal, and A. B. Murti, “Consumers’ Preferences for Ethical Entertainment Consumption: Conceptualization, Development, and Validation of a Scale,” *J. Promot. Manag.*, vol. 29, no. 8, pp. 1053–1081, 2023, doi: 10.1080/10496491.2023.2184441.

[3] S. Sharma and A. Bumb, “Product Placement in Entertainment Industry: A Systematic Review,” *Q. Rev. Film Video*, vol. 39, no. 1, pp. 103–119, 2022, doi: 10.1080/10509208.2020.1811606.

[4] R. Prey, M. Esteve Del Valle, and L. Zwerwer, “Platform pop: disentangling Spotify’s intermediary role in the music industry,” *Inf. Commun. Soc.*, vol. 25, no. 1, pp. 74–92, 2022, doi: 10.1080/1369118X.2020.1761859.

[5] A. Watson and A. Leyshon, “Negotiating platformisation: MusicTech, intellectual property rights and third wave platform reintermediation in the music industry,” *J. Cult. Econ.*, vol. 15, no. 3, pp. 326–343, 2022, doi: 10.1080/17530350.2022.2028653.

[6] J. Cho, Y. Bian, and J. Lee, “Leading digital business model transformation in the K-pop industry: the case of SM Entertainment,” *Asia Pacific Bus. Rev.*, vol. 29, no. 5, pp. 1394–1424, 2023, doi: 10.1080/13602381.2023.2229761.

[7] S. S. Wu and J. Lin, “Environmental benefits of China’s entertainment and media market: implications for sustainability,”

**Building of Informatics, Technology and Science (BITS)**
Volume 5, No 4, March 2024 Page: 682−693
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i4.5044

*Econ. Res. Istraz.* , vol. 36, no. 1, p., 2023, doi: 10.1080/1331677X.2022.2116068.
[8] X. Zhang, R. C. Kloosterman, and Y. Li, "'Stars' and places: exploring the spatial organization of media and entertainment industries in China," *J. Cult. Econ.*, vol. 13, no. 5, pp. 626–641, 2020, doi: 10.1080/17530350.2020.1751677.
[9] J. P. Đorđević and D. Stamenković, "Classification of multimodal translation errors in the entertainment industry: a proposal," *Translator*, vol. 29, no. 3, pp. 265–280, 2023, doi: 10.1080/13556509.2021.2024654.
[10] P. M. Scott, "Not going out: television's impacts on Britain's commercial entertainment industries and popular leisure during the 1950s," *Soc. Hist.*, vol. 48, no. 4, pp. 475–500, 2023, doi: 10.1080/03071022.2023.2246828.
[11] A. Watson, A. Leyshon, and G. Windsor, "Tech start-up capitalisation in an oligopolistic copyright industry: the case of the contemporary music industry," *Cult. Trends*, pp. 1–19, 2023, doi: 10.1080/09548963.2023.2255832.
[12] J. M. Luo, Y. Fan, and Z. Shang, "Analysis of Critical Success Factors for Entertainment Tourism Destinations: The Supply Perspective," *J. Qual. Assur. Hosp. Tour.*, vol. 23, no. 5, pp. 1159–1182, 2022, doi: 10.1080/1528008X.2021.1958126.
[13] Y. Browning and S. Darchen, "The impact of the COVID-19 pandemic on the music industries of Brisbane and an evaluation of the policy response," *Aust. Plan.*, pp. 1–12, 2024, doi: 10.1080/07293682.2023.2298428.
[14] F. Magni, "Value creation and appropriation in the live music industry: a population ecology analysis of live music ticket pricing," *J. Media Bus. Stud.*, vol. 20, no. 2, pp. 137–157, 2023, doi: 10.1080/16522354.2022.2036440.
[15] A. Lal, D. Hesmondhalgh, and C. Umney, "The changing shape of the Indian recorded music industry in the age of platformisation," *Contemp. South Asia*, vol. 31, no. 2, pp. 291–310, 2023, doi: 10.1080/09584935.2023.2206998.
[16] J. Nissen, "Do Jews count in klezmer? The problem of 'Jewface' in the British world music industry," *Jew. Cult. Hist.*, vol. 24, no. 3, pp. 313–331, 2023, doi: 10.1080/1462169X.2023.2235855.
[17] A. N. Hagen, "Datafication, Literacy, and Democratization in the Music Industry," *Pop. Music Soc.*, vol. 45, no. 2, pp. 184–201, 2022, doi: 10.1080/03007766.2021.1989558.
[18] J. Harvey and V. Manusov, "Rapping About Rap: Parental Mediation of Gender-Stereotyped Media Content," *Marriage Fam. Rev.*, vol. 56, no. 3, pp. 264–286, 2020, doi: 10.1080/01494929.2020.1712303.
[19] S. Singh, "'Just be Easy, Feel the Music'–Branding in the Musical and Social Life of Kampala's Hip-hop Scene," *East. African Lit. Cult. Stud.*, vol. 7, no. 4, pp. 285–302, 2021, doi: 10.1080/23277408.2021.1988230.
[20] G. Villanueva, "You must learn: sampling critical hip hop pedagogy in communication education spaces," *Pedagog. Cult. Soc.*, vol. 30, no. 4, pp. 435–453, 2022, doi: 10.1080/14681366.2020.1801814.
[21] L. Marie and M. Magladry, "Hip hop as decoration: theorizing the hybridity of hip hop and yoga in Perth, Western Australia," *Continuum (N. Y).*, vol. 34, no. 5, pp. 703–719, 2020, doi: 10.1080/10304312.2020.1782836.
[22] K. Nzinga, "Cultural resonance: Between hip-hop culture & the sociopolitical development of youth artivists in Palestine," *Mind, Cult. Act.*, vol. 00, no. 00, pp. 1–20, 2023, doi: 10.1080/10749039.2023.2238704.
[23] C. Lee and J. W. Lee, "Show me the monolingualism: Korean hip-hop and the discourse of difference," *Inter-Asia Cult. Stud.*, vol. 22, no. 1, pp. 1–15, 2021, doi: 10.1080/14649373.2021.1886475.
[24] A. B. Sunday, "'Throw the Money in the Air': Ideological Rejoinder in Olamide's Omo Abule Sowo," *Muziki*, vol. 18, no. 2, pp. 34–51, 2021, doi: 10.1080/18125980.2021.2003716.
[25] A. R. Keith, "On Being Ed Emcees: Toward Hip-Hop Educational Leadership Theory, Research, and Praxis," *Equity Excell. Educ.*, vol. 56, no. 3, pp. 372–394, 2023, doi: 10.1080/10665684.2023.2200207.
[26] D. Hall, "'Come as you are. We are a family.': Examining Hip Hop, belonging, and civicness in social studies," *Theory Res. Soc. Educ.*, vol. 51, no. 3, pp. 343–371, 2023, doi: 10.1080/00933104.2022.2164233.
[27] J. Evans, "Reframing civic education through hip-hop artistic practices: an empowerment and equity based learning model for black adolescents," *Pedagog. Cult. Soc.*, vol. 31, no. 4, pp. 845–861, 2023, doi: 10.1080/14681366.2021.1952294.
[28] E. Barna, "Between Cultural Policies, Industry Structures, and the Household: A Feminist Perspective on Digitalization and Musical Careers in Hungary," *Pop. Music Soc.*, vol. 45, no. 1, pp. 67–83, 2022, doi: 10.1080/03007766.2021.1984022.
[29] Q. Zhang and K. Negus, "Stages, Platforms, Streams: The Economies and Industries of Live Music after Digitalization," *Pop. Music Soc.*, vol. 44, no. 5, pp. 539–557, 2021, doi: 10.1080/03007766.2021.1921909.
[30] M. Semati and N. Behroozi, "Paradoxes of Gender, Technology, and the Pandemic in the Iranian Music Industry," *Pop. Music Soc.*, vol. 44, no. 1, pp. 1–13, 2021, doi: 10.1080/03007766.2020.1848015.
[31] D. B. Valdovinos Kaye and Z. M. Myint, "Reclaiming policy in the Myanmar popular music industry," *Int. J. Cult. Policy*, vol. 27, no. 1, pp. 50–64, 2021, doi: 10.1080/10286632.2019.1709061.
[32] B. A. Varga and T. Ender, "Wu-Tang for the Children: Swarming Elsewhere for Aesthetic (Re)Imaginings of Community, Theory, & Praxis," *Equity Excell. Educ.*, vol. 56, no. 3, pp. 395–408, 2023, doi: 10.1080/10665684.2023.2240339.
[33] P. Ammirante and J. Rovetti, "Bright vowels are favoured on weak beats in popular music lyrics," *J. New Music Res.*, vol. 50, no. 3, pp. 259–265, 2021, doi: 10.1080/09298215.2021.1936076.
[34] T. Järvenpää, "'Real Gs and pastors in the church': hip-hop esthetics in the rap music video productions of the Evangelical Lutheran Church of Finland," *Religion*, vol. 52, no. 3, pp. 450–469, 2022, doi: 10.1080/0048721X.2022.2068460.
[35] A. de Paor-Evans, "Urban Myths and Rural Legends: An Alternate Take on the Regionalism of Hip Hop," *Pop. Music Soc.*, vol. 43, no. 4, pp. 414–425, 2020, doi: 10.1080/03007766.2020.1730651.
[36] K. M. A. Theledi, "Hip Hop Pantsula making Setswana fashionable: impact on Setswana orthography," *South African J. African Lang.*, vol. 43, no. sup1, pp. 334–340, 2024, doi: 10.1080/02572117.2024.2322246.
[37] A. Hill and J. C. Lee, "Roamers: Audiences on the Move Across Entertainment Platforms In Southeast Asia," *Javnost*, vol. 29, no. 1, pp. 98–114, 2022, doi: 10.1080/13183222.2021.1932985.
[38] V. Noguti and D. S. Waller, "Motivations to use social media: effects on the perceived informativeness, entertainment, and intrusiveness of paid mobile advertising," *J. Mark. Manag.*, vol. 36, no. 15–16, pp. 1527–1555, 2020, doi: 10.1080/0267257X.2020.1799062.
[39] B. R. Ojebuyi and B. L. Fafowora, "Contesting Cultural Imperialism: Hybridisation and Re-enactment of Indigenous Cultural Values in Nigerian Hip-Hop Music," *Muziki*, vol. 18, no. 1, pp. 59–81, 2021, doi: 10.1080/18125980.2021.1945946.
[40] S. U. Eze and E. P. Sylvanus, "Popular Music, Deviance and Moral Panics in Nigeria: The Music of Naira Marley," *Muziki*, vol. 19, no. 2, pp. 25–40, 2022, doi: 10.1080/18125980.2023.2201472.