Mantik Journal, 8 (1) (2024), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Performance evaluation of SVM with synthetic minority over-sampling technique in sentiment classification

Yerik Afrianto Singgalen

[1]Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Indonesia

ARTICLE INFO

Article history:

Received Apr 06, 2024
Revised Apr 09, 2024
Accepted Apr 19, 2024

Keywords:

Content Reviews;
Evaluation;
Performance;
Sentiment Classification;
SMOTE;
SVM.

ABSTRACT

This study investigates the performance of the Support Vector Machine (SVM) algorithm in sentiment analysis tasks within the context of tourism destination branding, utilizing the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. Specifically, the research compares SVM performance with and without the Synthetic Minority Over-sampling Technique (SMOTE) across various metrics including accuracy, precision, recall, F-measure, and Area Under the Curve (AUC). The analysis is conducted on a dataset comprising textual data extracted from "Wonderful Indonesia" promotional videos featuring Labuan Bajo. Results indicate that SVM without SMOTE achieves a slightly higher accuracy of 97.79% compared to 96.61% with SMOTE. However, a closer examination reveals that SVM without SMOTE accurately classifies all positive instances, while with SMOTE, one positive instance is misclassified as negative. Precision, recall, and F-measure scores for positive instances are also higher without SMOTE, indicating better performance in classifying positive sentiment.

Corresponding Author:

Yerik Afrianto Singgalen
Tourism Department, Faculty of Business Administration and Communication,
Atma Jaya Catholic University of Indonesia
Jl. Jenderal Sudirman No. 51, Jakarta, 10220, Indonesia.
Email: yerik.afrianto@atmajaya.ac.id

1. INTRODUCTION

Destination branding, a crucial aspect of tourism marketing, encompasses many strategies to enhance the allure and perception of a particular location. Among these strategies, creating promotional destination videos is a powerful tool for captivating potential visitors and shaping perceptions (Dewantara et al., 2023). Through visually compelling narratives and immersive storytelling, these videos showcase a destination's unique attractions, cultural richness, and experiential offerings (Dewi & Arifuddin, 2021; Murti et al., 2023; Perangin-Angin et al., 2023). By effectively communicating a locale's essence and distinctive features, such promotional videos serve as persuasive mediums to evoke wanderlust and stimulate tourist interest (Bire et al., 2022; Liu et al., 2023). Consequently, it plays a pivotal role in influencing travelers' decision-making processes and ultimately contributes to the economic growth and sustainability of the destination.

In Indonesia, destination branding strategy involves leveraging digital campaigns featuring creative content, including promotional videos showcasing the diverse attractions under the "Wonderful Indonesia" campaign (Salamah & Yananda, 2020; Sholeh & Juniarti, 2021). Using digital platforms and engaging storytelling techniques, such campaigns effectively capture the essence of Indonesia's cultural heritage, natural landscapes, and unique experiences, enhancing the destination's visibility and appeal to domestic and international travelers (Nusantara et al., 2021; Purwandani & Yusuf, 2021; Sujatna et al., 2024). This approach harnesses the power of visual storytelling to evoke emotions and inspire wanderlust among the target audience, ultimately contributing to Indonesia's tourism industry's sustainable growth and development.

The context of discussion in this research centers on the video content of "Wonderful Indonesia," specifically related to the tourist destination of Labuan Bajo. Focusing on Labuan Bajo within the broader scope of Indonesia's tourism promotion allows for a detailed examination of the effectiveness of marketing strategies in promoting this particular destination (Caraka et al., 2023; Dhakal & Tjokro, 2024; Westoby et al., 2021). By analyzing the content and messaging of the promotional videos featuring Labuan Bajo, this research gains valuable insights into the portrayal of its attractions, cultural significance, and visitor experiences, thus informing efforts to optimize destination branding and tourism development strategies in the region (Ardhyanto et al., 2023; Sejati et al., 2023; Wang & Sun, 2023). This focused approach enables a deeper understanding of the nuances and challenges inherent in promoting Labuan Bajo as a premier tourist destination within the Indonesian archipelago.

Sentiment analysis emerges as a potent approach to assessing perceptions toward tourist destinations through digital communication channels. Using natural language processing techniques, sentiment analysis enables the systematic evaluation of public sentiments expressed in online content such as reviews, social media posts, and blogs (Christanto & Singgalen, 2022; Singgalen, 2023e, 2023f, 2024). This method offers valuable insights into travelers' prevailing attitudes, preferences, and experiences, facilitating destination management organizations in tailoring marketing strategies and enhancing visitor satisfaction (Singgalen, 2022, 2023a, 2023d, 2023c, 2023b). Consequently, sentiment analysis is a pivotal tool for gauging the effectiveness of promotional efforts and refining the destination's brand image in the ever-evolving digital landscape.

This research aims to identify and analyze sentiments towards the video content of "Wonderful Indonesia" Labuan Bajo using the CRISP-DM approach, employing a Support Vector Machine (SVM) model with the support of the Synthetic Minority Over-sampling Technique (SMOTE) operator. Through this methodological framework, the study systematically assesses public perceptions and attitudes towards Labuan Bajo as portrayed in promotional videos, thereby providing valuable insights for destination management and tourism marketing strategies. By leveraging advanced data mining techniques such as SVM and SMOTE within the CRISP-DM framework, this research offers a comprehensive understanding of sentiment dynamics, enabling stakeholders to make informed decisions to enhance the destination's appeal and competitiveness in the global tourism landscape.

The urgency of this research lies in its potential to inform strategic decisions in destination management and tourism marketing, addressing the growing importance of digital media in shaping travelers' perceptions. Theoretical implications stem from advancing understanding of sentiment analysis methodologies within the context of destination branding, contributing to the academic discourse on tourism marketing and digital communication. Moreover, the practical implications of this research extend to destination management organizations and marketing agencies, equipping them with actionable insights to optimize promotional efforts and enhance the competitiveness of destinations like Labuan Bajo in the global tourism marketplace. In conclusion, this research holds significant promise in bridging the gap between theory and practice, offering tangible benefits for academia and industry stakeholders.

2. RESEARCH METHOD

This research adopts the CRISP-DM methodology, encompassing its structured stages for implementing the Support Vector Machine (SVM) model in sentiment classification. By adhering to the Cross-Industry Standard Process for Data Mining (CRISP-DM), the study ensures a systematic and rigorous approach to data analysis, encompassing stages such as business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Through this methodological framework, the research aims to maximize the effectiveness and efficiency of sentiment analysis in assessing public perceptions towards the Labuan Bajo destination, thus providing valuable insights for destination management and tourism marketing strategies. In conclusion, using CRISP-DM with SVM model implementation exemplifies a robust methodology tailored to the complexities of sentiment analysis in destination branding.

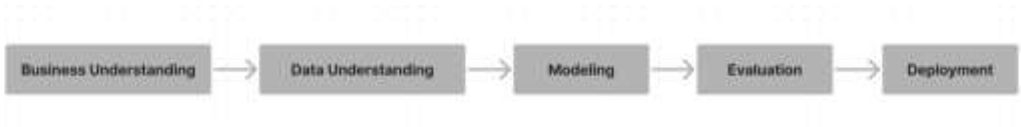

Figure 1. Cross-Industry Standard Process for Data Mining (CRISP-DM) Framework

Figure 1 shows the implementation of the CRISP-DM framework for sentiment classification. The CRISP-DM proves highly pertinent within this research, as it facilitates the systematic exploration of public perceptions towards the video content of "Wonderful Indonesia" Labuan Bajo through sentiment analysis. By adhering to the structured stages of CRISP-DM, this research effectively navigates the complexities of data mining and analysis, ensuring a comprehensive understanding of sentiment dynamics surrounding the destination. This methodological framework identifies patterns and insights for strategic decisions in destination management and tourism marketing efforts. Thus, the application of CRISP-DM aligns seamlessly with the research objective of dissecting public sentiments towards Labuan Bajo, underscoring its relevance and efficacy in the realm of destination branding and digital communication strategies.

2.1 Business Understanding

In the business understanding stage, the evaluated video pertains to the Wonderful Indonesia campaign featuring Labuan Bajo, identified by its unique code RaTWq98hzF0. As of September 13, 2013, this video had garnered significant attention, accumulating 367,773 views and 529 comments. This selection is critical as it represents a focal point for analyzing public sentiment towards Labuan Bajo, providing valuable insights into the effectiveness of destination marketing efforts. By examining this video's engagement metrics and feedback, this research understands the audience's perceptions and preferences, thus informing strategic decision-making in destination management and tourism marketing initiatives. Therefore, the meticulous consideration of this specific video aligns with the research's objective of comprehensively evaluating public sentiment towards Labuan Bajo within the Wonderful Indonesia campaign.

Figure 2. Frequently used words of the content (Communalytic)

Figure 2 shows the frequently used words in the content video. Based on the identification of frequently used words, it is evident that "Indonesia" dominates the word frequency list with 49 occurrences, followed by "indonesia" with 30 occurrences. Additionally, words such as "love" appear 23 times, "Komodo" 21 times, and "hi" 20 times, respectively. Other notable terms include "like" with 16 occurrences, "Wonderful" and "INDONESIA" with 15 occurrences each, and "komodo" with 14 occurrences. Furthermore, words such as "wonderful," "nice," "video," "melihat," and "place" each appear between 10 to 13 times. This analysis highlights the prevalence of terms related to Indonesia's identity and tourism attractions, suggesting a positive sentiment towards the destination. These findings offer valuable insights into linguistic patterns and sentiment dynamics, informing destination branding and marketing strategies to enhance Indonesia's appeal and visibility in the global tourism landscape.

Figure 3. Top Ten Poster and Emoji Cloud (Communalytic)

Figure 3 shows the top ten posters and emoji clouds. Based on the results of identifying the emoji cloud, it is apparent that certain emojis are more frequently used than others in expressing sentiments related to the content. The heart emoji ❤appears most frequently with ten occurrences, followed by the heart eyes emoji □ and the flag of Indonesia emoji □□, each with seven occurrences. Additionally, the thumbs-up emoji □ is used six times. In contrast, the smiling face with smiling eyes emoji □, green heart emoji , and grinning face with smiling eyes emoji □ each appears three times. Less frequently used emojis include the smiling face with hearts emoji □ and folded hands emoji □, with two occurrences each. This analysis underscores the prevalence of positive sentiments, as indicated by the frequent use of heart emojis and smiling faces, reflecting a favorable perception of the content. Such insights into emoji usage provide valuable context for understanding audience engagement and sentiment dynamics, informing content creation and communication strategies in digital marketing and branding efforts.

Moreover, based on the results of identifying the top ten posters, it is evident that certain users have significantly higher posting frequencies than others. The user @ABStravel tops the list with 21 posts, followed by @mauricevandermaat9003 with nine posts and @windahcayang3673 with six posts. Other notable contributors include @mundodosskatistas9980 and @Arimateaucb, each with five posts, and @mandailingrura3151, @user-tg3gt3zb8y, and @prabhatkumar-jn6mm, each with four posts. Additionally, @kunduologyvlogs8012 and @DeonJones351 round out the top ten with three posts each. This analysis sheds light on the level of user engagement and contribution within the context of the discussion, providing valuable insights for understanding the dynamics of online interactions and community participation. The findings inform strategies for fostering user engagement and community building in digital marketing and branding endeavors, ultimately enhancing the effectiveness of online communication efforts.

2.2 Data Understanding

During the data understanding stage, data cleaning uses various operators in RapidMiner, including tokenize, filter tokens, transform cases, stopwords, and remove duplicates. This systematic approach ensures the refinement of the dataset by eliminating irrelevant information, standardizing text formats, and removing redundant entries. By employing these operators, the dataset is prepared for subsequent analysis, enhancing the reliability and accuracy of the results. This methodological rigor underscores the importance of meticulous data preprocessing in ensuring the validity and robustness of the research findings. Thus, using RapidMiner for data cleaning exemplifies an efficient and practical approach to enhancing the dataset's quality, ultimately facilitating meaningful insights and informed decision-making processes.

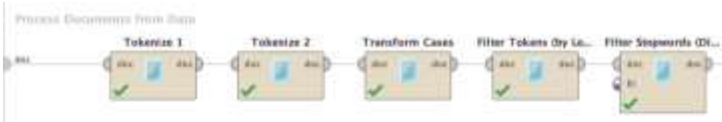

Figure 4. Pre-processing Data

Figure 4 shows the pre-processing text data. Based on the data cleaning process results, 518 out of 530 data points have been successfully extracted, allowing the acquisition of string scores from the textual data. This outcome underscores the effectiveness of the data preprocessing methods, which facilitated identifying and extracting relevant information from the dataset. The availability of this refined dataset is crucial for conducting subsequent analyses and deriving meaningful insights from the text data. Consequently, the successful extraction of 518 out of 530 data points signifies a pivotal step towards achieving the research objectives, highlighting the importance of rigorous data-cleaning processes in ensuring the quality and reliability of research outcomes.


Figure 5. Extract Sentiment Process

Figure 5 shows the extract sentiment process. Based on the results of sentiment extraction, it was found that there are 12 score strings classified as negative values, alongside 505 score strings categorized as positive values. This outcome demonstrates a notable contrast in sentiment distribution within the dataset, with a predominant prevalence of positive and negative sentiments. Such findings provide valuable insights into the overall sentiment dynamics surrounding the analyzed text data, highlighting the predominance of positive attitudes or opinions expressed toward the subject matter. Consequently, this underscores the importance of sentiment analysis in discerning the prevailing sentiments and attitudes within textual data, facilitating a deeper understanding of public perceptions and opinions towards specific topics or entities.

2.3 Modeling

The Support Vector Machine (SVM) model is employed for sentiment classification during the modeling stage. Recognizing the imbalance in data distribution between negative and positive scores, the Synthetic Minority Over-sampling Technique (SMOTE) operator is utilized to enhance the model's performance. This strategic decision aims to address the challenge posed by imbalanced data by synthesizing minority class samples, thereby mitigating the impact of data skewness on the model's predictive capabilities. By incorporating SMOTE into the modeling process, this research improves the model's ability to classify sentiments accurately, ultimately enhancing the reliability and effectiveness of sentiment analysis outcomes. Thus, integrating SVM with SMOTE represents a prudent approach to optimizing the sentiment classification model and obtaining more robust insights from the analyzed data.

Figure 6. Modeling Process using Rapidminer

Figure 6 shows the modeling process in Rapidminer. Based on the results of modeling, the performance of the Support Vector Machine (SVM) utilizing the Synthetic Minority Over-sampling Technique (SMOTE) was evaluated across various metrics, including accuracy, precision, recall, F-measure, and Area Under the Curve (AUC). This comprehensive evaluation allows for a thorough assessment of the model's effectiveness in classifying sentiments, considering its ability to identify positive and negative instances correctly and its overall predictive power. By analyzing these metrics, this research gains insights into the strengths and limitations of the SVM model with SMOTE in capturing the nuances of sentiment dynamics within the analyzed dataset. Consequently, this evaluation facilitates informed decision-making regarding the suitability and optimization of the sentiment classification approach, ultimately enhancing the reliability and utility of sentiment analysis outcomes.

2.4 Evaluation

During the evaluation stage, the Support Vector Machine (SVM) performance with and without using the Synthetic Minority Over-sampling Technique (SMOTE) was assessed by examining the confusion matrix. This matrix provides a comprehensive overview of the model's predictive capabilities by summarizing the classification results, including true positive, true negative, false positive, and false negative instances. By comparing the confusion matrices generated from SVM with and without SMOTE, this research discerns the impact of data imbalance on the model's performance, particularly in accurately identifying minority class samples. Consequently, the evaluation based on confusion matrices facilitates a nuanced understanding of the effectiveness of different modeling approaches in sentiment classification tasks, aiding in selecting the most suitable methodology for achieving reliable and robust sentiment analysis outcomes.

SVM without SMOTE SVM with SMOVE

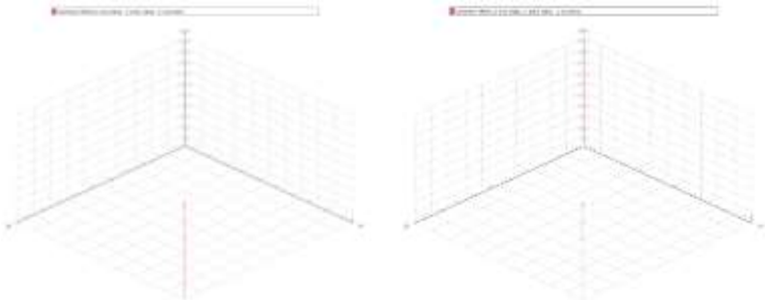

Figure 6. Evaluation of the Model Performance

Figure 6 shows the plot view of the confusion matrix. Based on the evaluation results of SVM performance, it is recommended that strategies be devised to improve video content that aligns with tourists' preferences based on sentiment classification outcomes from the "Wonderful Indonesia" Labuan Bajo videos. By leveraging the insights gained from sentiment analysis, destination management organizations tailor content creation efforts to better resonate with the sentiments expressed by viewers. This targeted approach ensures that promotional materials effectively communicate the unique attractions and experiences offered by Labuan Bajo, thereby enhancing the appeal to potential visitors. Consequently, integrating sentiment analysis findings into content development strategies is valuable for optimizing destination branding and tourism marketing initiatives, ultimately fostering greater engagement and interest among target audiences.

2.5 Deployment

During the deployment stage, the measurement of algorithm performance yields recommendations that are apt and applicable. This stage represents the culmination of the analytical process, where the insights gleaned from model evaluation are translated into actionable strategies or decisions. By leveraging the performance metrics obtained through rigorous evaluation, stakeholders make informed choices regarding implementing the algorithm in real-world scenarios. Consequently, the deployment phase serves as a crucial bridge between data analysis and practical application, facilitating the effective utilization of algorithmic insights to address pertinent challenges or achieve desired outcomes.

3. RESULTS AND DISCUSSIONS

A comparison of the performance of the Support Vector Machine (SVM) algorithm with and without Synthetic Minority Over-sampling Technique (SMOTE) unveils discernible disparities in efficacy for sentiment classification tasks. Without SMOTE, SVM achieves marginally superior accuracy at 97.79% compared to 96.61% with SMOTE. Examination of confusion matrices reveals distinct capabilities in correctly classifying positive instances; without SMOTE, all positive instances are accurately classified, whereas, with SMOTE, one positive instance is misclassified. This indicates that while both models perform commendably, SVM without SMOTE excels marginally in identifying positive instances accurately. Moreover, the precision, recall, and F-measure scores for positive instances are marginally higher without SMOTE, denoting superior overall performance in positive sentiment classification. Nonetheless, the choice between SVM with or without SMOTE hinges on specific task requisites, considering factors such as the significance of correctly identifying minority class instances and balancing performance with model robustness.

Based on the results of implementing the Support Vector Machine (SVM) model in classifying a dataset comprising 518 text data, divided into 30% for training and 70% for testing, various performance metrics, including accuracy, precision, recall, F-measure, and Area Under the Curve (AUC) ascertained. This systematic approach ensures a comprehensive evaluation of the model's effectiveness in classifying textual data, providing insights into its predictive capabilities and robustness across different evaluation metrics. By considering these performance metrics, stakeholders gauge the reliability and suitability of the SVM model for sentiment analysis tasks, thereby informing decision-making processes and optimizing the utilization of machine learning algorithms in real-world applications.

SVM without SMOTE

```
PerformanceVector:
accuracy: 97.79% +/- 1.16% (micro average: 97.79%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       0       0
Positive:       8       354
AUC (optimistic): 0.622 +/- 0.442 (micro average: 0.622) (positive class: Positive)
AUC: 0.530 +/- 0.378 (micro average: 0.530) (positive class: Positive)
AUC (pessimistic): 0.438 +/- 0.346 (micro average: 0.438) (positive class: Positive)
precision: 97.79% +/- 1.16% (micro average: 97.79%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       0       0
Positive:       8       354
recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       0       0
Positive:       8       354
f_measure: 98.88% +/- 0.59% (micro average: 98.88%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       0       0
Positive:       8       354
```

SVM with SMOTE

```
PerformanceVector:
accuracy: 96.61% +/- 2.13% (micro average: 96.61%)
ConfusionMatrix:
True:   Negative        Positive
Negative:       331     1
Positive:       23      353
AUC (optimistic): 0.999 +/- 0.002 (micro average: 0.999) (positive class: Positive)
AUC: 0.999 +/- 0.002 (micro average: 0.999) (positive class: Positive)
AUC (pessimistic): 0.998 +/- 0.002 (micro average: 0.999) (positive class: Positive)
precision: 93.96% +/- 3.57% (micro average: 93.88%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       331     1
Positive:       23      353
recall: 99.71% +/- 0.93% (micro average: 99.72%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       331     1
Positive:       23      353
f_measure: 96.72% +/- 2.05% (micro average: 96.71%) (positive class: Positive)
ConfusionMatrix:
True:   Negative        Positive
Negative:       331     1
Positive:       23      353
```

Figure 7. Model Performance of SVM with and without SMOTE (Rapidminer)

Figure 7 shows the SVM model performance. The performance of the Support Vector Machine (SVM) without Synthetic Minority Over-sampling Technique (SMOTE) is characterized by an accuracy of 97.79% with a micro average of 97.79%. The confusion matrix reveals that out of 362 instances classified as positive, 354 are correctly identified, whereas all negative instances are accurately classified. The Area Under the Curve (AUC) values indicate moderate predictive power, with an optimistic AUC of 0.622 and a pessimistic AUC of 0.438. Precision, recall, and F-measure scores further underscore the model's proficiency, with precision and recall achieving scores of 97.79% and 100%, respectively, while the F-measure attains a score of 98.88%. Moreover, the Support Vector Machine (SVM) performance with the Synthetic Minority Over-sampling Technique (SMOTE) showcases an accuracy of 96.61% with a micro average of 96.61%. The confusion matrix indicates that out of 356 positive instances, 353 are accurately classified, while only one instance is misclassified as unfavorable. The Area Under the Curve (AUC) values demonstrate predictive solid power, with optimistic, pessimistic, and overall AUC scores all approaching 0.999. Precision achieves a score of 93.96%, while recall attains an impressive 99.71%, indicating high accuracy in identifying positive instances. The F-measure score of 96.72% further emphasizes the model's effectiveness in sentiment classification. These performance metrics underscore the robustness and reliability of the SVM model with SMOTE in accurately classifying sentiment, highlighting its potential applicability in practical settings for sentiment analysis tasks.

Comparing the performance of the Support Vector Machine (SVM) algorithm with and without Synthetic Minority Over-sampling Technique (SMOTE) reveals notable differences in effectiveness for sentiment classification tasks. When SVM is employed without SMOTE, the model achieves a slightly higher accuracy of 97.79% compared to 96.61% with SMOTE. However, examining the confusion matrices provides insight into the models' abilities to classify positive instances correctly. Without SMOTE, all positive instances are accurately classified, whereas with SMOTE, one positive instance is misclassified as unfavorable. This indicates that while both models perform well, the SVM model without SMOTE performs slightly better in correctly identifying positive instances. Additionally, the precision, recall, and F-measure scores for positive instances are slightly higher without SMOTE than with SMOTE, indicating better overall performance in classifying positive sentiment. However, it's essential to consider the trade-off between performance and handling imbalanced data. The SMOTE algorithm aims to address class imbalance by oversampling minority class samples, which may result in slightly lower performance metrics but enhance the model's ability to generalize to unseen data and improve its robustness. Therefore, the choice between using SVM with or without SMOTE depends on the specific requirements of the sentiment analysis task, considering factors such as the importance of correctly identifying minority class instances and the trade-off between performance and robustness.

4. CONCLUSION

In conclusion, the research conducted following the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework has provided valuable insights into the performance of the Support Vector Machine (SVM) algorithm in sentiment analysis tasks, both with and without the Synthetic Minority Over-sampling Technique (SMOTE). The analysis revealed that SVM without SMOTE achieved a marginally higher accuracy of 97.79% compared to 96.61% with SMOTE. However, a closer examination of the confusion matrices highlighted that SVM without SMOTE accurately classified all positive instances, whereas, with SMOTE, one positive instance was misclassified as unfavorable. Additionally, precision, recall, and F-measure scores for positive instances were slightly higher without SMOTE, indicating better overall performance in classifying positive sentiment. Nevertheless, it's essential to consider the trade-off between performance and handling imbalanced data. While SVM without SMOTE may demonstrate slightly better performance metrics, SMOTE addresses class imbalance, enhancing the model's ability to generalize to unseen data. Therefore, using SVM with or without SMOTE depends on specific task requirements and considerations, such as correctly identifying minority class instances and the trade-off between performance and robustness. These findings, obtained through the systematic application of the CRISP-DM framework, provide actionable insights for researchers and practitioners in natural language processing, guiding future research directions and informing practical applications of sentiment analysis algorithms.

ACKNOWLEDGEMENTS

I want to extend my sincere appreciation to the Tourism Department at the Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, for the invaluable support in the publication of this article.

REFERENCES

Ardhyanto, A., Dewancker, B., Tsai, Y. L., & Heryana, R. E. (2023). Memory recollection and oral history: a study of vernacular architecture transformation of the past. *Journal of Asian Architecture and Building Engineering*, *22*(6), 3435–3454. https://doi.org/10.1080/13467581.2023.2204951

Bire, R. B., Nugraha, Y. E., & Pellokila, I. R. (2022). Tourist attraction efficiency and its determinants using a two stage, double bootstrap data envelopment analysis: a case study in Indonesia. *Journal of Policy Research in Tourism, Leisure and Events*, 1–18. https://doi.org/10.1080/19407963.2022.2158191

Caraka, R. E., Wardhana, I. W., Kim, Y., Sakti, A. D., Gio, P. U., Noh, M., & Pardamean, B. (2023). Connectivity, sport events, and tourism development of Mandalika's special economic zone: A perspective from big data cognitive analytics. *Cogent Business and Management*, *10*(1). https://doi.org/10.1080/23311975.2023.2183565

Christanto, H. J., & Singgalen, Y. A. (2022). Sentiment Analysis on Customer Perception towards Products and Services of Restaurant in Labuan Bajo. *Journal of Information Systems and Informatics*, *4*(3), 511–523. https://doi.org/10.51519/journalisi.v4i3.276

Dewantara, M. H., Gardiner, S., & Jin, X. (2023). Travel vlog ecosystem in tourism digital marketing evolution: a narrative literature review. *Current Issues in Tourism*, *26*(19), 3125–3139. https://doi.org/10.1080/13683500.2022.2136568

Dewi, U. P., & Arifuddin, M. R. (2021). Communicating with the 'uncultured': the study of conventional norms in Indonesian intercultural communication context. *Journal of International Communication*, *27*(2), 300–316. https://doi.org/10.1080/13216597.2021.1942950

Dhakal, S. P., & Tjokro, S. P. (2024). Tourism enterprises in Indonesia and the fourth industrial revolution–are they ready? *Tourism Recreation Research*, *49*(2), 439–444. https://doi.org/10.1080/02508281.2021.1996687

Liu, L. W., Wang, C. C., Pahrudin, P., Royanow, A. F., Lu, C., & Rahadi, I. (2023). Does virtual tourism influence tourist visit intention on actual attraction? A study from tourist behavior in Indonesia. *Cogent Social Sciences*, *9*(1). https://doi.org/10.1080/23311886.2023.2240052

Murti, D. C. W., Ratriyana, I. N., & Asmoro, I. D. (2023). "Dream Now, Travel Tomorrow": Communicating the Nation Branding of Indonesia through Tourism-Based Social Media. *Howard Journal of Communications*, *34*(3), 293–314. https://doi.org/10.1080/10646175.2023.2169086

Nusantara, A. C., Volgger, M., & Pforr, C. (2021). Evaluating the complex impact of policy changes on tourism development: The case of Surakarta, Indonesia. *Journal of Global Scholars of Marketing Science: Bridging Asia and the World*, *31*(4), 614–623. https://doi.org/10.1080/21639159.2021.1935291

Perangin-Angin, R., Tavakoli, R., & Kusumo, C. (2023). Inclusive tourism: the experiences and expectations of Indonesian wheelchair tourists in nature tourism. *Tourism Recreation Research*, *48*(6), 955–968. https://doi.org/10.1080/02508281.2023.2221092

Purwandani, I., & Yusuf, M. (2021). Localizing Indonesian Halal tourism policy within local customs, Qanun, and marketing. *Journal of Policy Research in Tourism, Leisure and Events*, *0*(0), 1–19. https://doi.org/10.1080/19407963.2021.1996382

Salamah, U., & Yananda, M. R. (2020). The Many Faces of Wonderful Indonesia: Tourism Brand in 16 Countries Online News Sites. *Jurnal Komunikasi Indonesia*, *9*(1), 34–39. https://doi.org/10.7454/jki.v9i1.11586

Sejati, A. W., Putri, S. N. A. K., Tyas, W. P., Buchori, I., Handayani, W., Basuki, Y., Barbarossa, G., & Husna, I. N. (2023). Predicting urban carrying capacity to support sustainable tourism using GIS. *Journal of Policy Research in Tourism, Leisure and Events*, 1–24. https://doi.org/10.1080/19407963.2023.2279065

Sholeh, M., & Juniarti, G. (2021). Gaya Hidup Hedonisme dalam Iklan Pariwisata Wonderful indonesia "An Exploration of the Wondrous Labuan Bajo." *JIKE : Jurnal Ilmu Komunikasi Efek*, *5*(1), 131–149.

Singgalen, Y. A. (2022). Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier. *Building of Informatics, Technology and Science (BITS)*, *4*(3), 1343−1352. https://doi.org/10.47065/bits.v4i3.2486

Singgalen, Y. A. (2023a). Analisis Perilaku Wisatawan Berdasarkan Data Ulasan di Website Tripadvisor Menggunakan CRISP-DM : Wisata Minat Khusus Pendakian Gunung Rinjani dan Gunung Bromo. *Journal of Computer System and Informatics (JoSYC)*, *4*(2), 326–338. https://doi.org/10.47065/josyc.v4i2.3042

Singgalen, Y. A. (2023b). Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW. *Journal of Information System Research (JOSH)*, *4*(4), 1343–1353. https://doi.org/10.47065/josh.v4i4.3917

Singgalen, Y. A. (2023c). Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM. *Journal of Information System Research (JOSH)*, *4*(2), 614–625. https://doi.org/10.47065/josh.v4i2.2999

Singgalen, Y. A. (2023d). Analisis Sentimen Top 10 Traveler Ranked Hotel di Kota Makassar Menggunakan Algoritma Decision Tree dan Support Vector Machine. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(1), 323–332. https://doi.org/10.30865/klik.v4i1.1153

Singgalen, Y. A. (2023e). Comparative analysis of decision tree and support vector machine algorithm in sentiment classification for birds of paradise content. *International Journal of Basic and Applied Science*, *12*(3), 100–109.

Singgalen, Y. A. (2023f). Sentiment classification of coral reef 101 content using decision tree algorithm through CRISP-DM. *International Journal of Basic and Applied Science*, *12*(3), 121–130.

Singgalen, Y. A. (2024). Social network and sentiment analysis of product reviews (case of smartwatch product content). *International Journal on Social Science, Economics and Art*, *13*(4), 255–267.

Sujatna, E. T. S., Mulyanah, A., Walangarei, S. F., Sukma, B. P., & Rahmawati, A. (2024). Objective or subjective adjectives? A case study on UNESCO Global Geopark tourism texts. *Cogent Arts and Humanities*, *11*(1). https://doi.org/10.1080/23311983.2023.2295076

Wang, S., & Sun, J. (2023). Embodiment of feminine subjectivity by women of a tourism destination. *Journal of Sustainable Tourism*, *31*(6), 1447–1463. https://doi.org/10.1080/09669582.2022.2053858

Westoby, R., Gardiner, S., Carter, R. W., & Scott, N. (2021). Sustainable livelihoods from tourism in the “10 New Balis” in Indonesia. *Asia Pacific Journal of Tourism Research*, *26*(6), 702–716. https://doi.org/10.1080/10941665.2021.1908386