KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 4, No 4, February 2024, Hal 2078-2087
DOI 10.30865/klik.v4i4.1709
https://djournals.com/klik

Performance Evaluation of SVM Algorithm in Sentiment Classification: A Visual Journey of Wonderful Indonesia Content

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Email Corresponding Author: yerik.afrianto@atmajaya.ac.id

Abstract−This study addresses the research problem of understanding public sentiment towards tourism-themed content on YouTube, with a specific focus on "A Visual Journey of Wonderful Indonesia." The primary aim is to explore how viewers perceive and depict Indonesia as a tourism destination through their comments on YouTube videos. Employing the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, sentence analysis is conducted using the Support Vector Machine (SVM) algorithm with the Synthetic Minority Over-sampling Technique (SMOTE) to classify sentiments within a dataset of YouTube comments as positive, negative, or neutral. The analysis of frequently used words in the comments provides valuable insights into Indonesia's perception, revealing positive sentiments reflected in terms such as "beautiful," "wonderful," and "amazing," emphasizing the country's aesthetic appeal. Notably, terms like "orang" and "Indonesian" indicate appreciation for Indonesia's rich cultural heritage and its people. These findings highlight the pivotal role of destination branding efforts in shaping positive perceptions and emotions toward Indonesia. The results indicate the efficacy of the SVM-SMOTE model, achieving high accuracy (84.26%), precision (100.00%), recall (68.51%), f-measure (81.25%), and AUC (0.996) in accurately classifying sentiment patterns within analyzed YouTube content. This offers practical implications for destination managers and marketers. Conversely, the SVM algorithm without SMOTE demonstrates impressive accuracy, precision, and recall scores of 97.08%, but its AUC value of 0.607 suggests potential challenges in discriminating between positive and negative sentiment instances. These findings provide valuable insights into the role of digital media platforms in shaping destination perceptions and offer practical implications for destination marketers and managers.

Keywords: Performance; Evaluation; SVM; Wonderful; Indonesia

1. INTRODUCTION

Destination branding has been acknowledged to significantly influence public sentiment, prompting destination managers to strive to enhance impactful visitor experiences. This recognition underscores the pivotal role of destination branding in shaping perceptions and attitudes towards a particular locale and driving tourist preferences [1]. Consequently, stakeholders within the tourism industry are increasingly investing resources in crafting compelling narratives and delivering memorable services to visitors [2]. Such efforts not only contribute to cultivating positive sentiments but also foster loyalty among tourists, thereby bolstering the overall competitiveness and sustainability of the destination [3].

Previous research has extensively discussed the concept of destination branding and its implications on visitor perceptions [4]. However, a research gap highlights the limited attention given to public sentiment regarding tourism-themed content on YouTube videos [5]. This oversight underscores the necessity for further investigation into the role of digital media platforms, particularly YouTube, in shaping public sentiments toward tourism destinations [6]. By addressing this research gap, scholars can provide valuable insights into the effectiveness of destination branding strategies in the digital age and offer practical recommendations for destination managers to optimize their online presence and engagement with potential visitors [7].

Understanding sentiment analysis through research is imperative for several reasons. The primary motivation lies in the growing significance of digital media and social platforms, where user-generated content reflects diverse opinions and sentiments. The exploration of sentiment analysis aids in deciphering the nuanced perceptions, attitudes, and emotions expressed by individuals online, particularly in the context of tourism-themed content. This research becomes essential for destination managers and marketers as it offers a systematic approach to unravel the sentiment patterns surrounding a destination, such as "A Visual Journey of Wonderful Indonesia" on YouTube. By employing advanced methodologies like the Support Vector Machine algorithm with the Synthetic Minority Over-sampling Technique, researchers can gain valuable insights into the public's sentiments, ultimately informing destination branding and marketing strategies. In conclusion, the necessity to conduct sentiment analysis research stems from the dynamic nature of online discourse, enabling a deeper understanding of public perceptions that is crucial for strategic decision-making in destination management and marketing.

The urgency of this research lies in the contemporary importance of digital platforms, particularly YouTube, as influential mediums shaping destination perceptions. As tourism-themed content on such platforms can significantly impact public sentiments, it becomes crucial to employ robust methodologies for sentiment analysis. The chosen process involves utilizing the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, coupled with the Support Vector Machine (SVM) algorithm and Synthetic Minority Over-sampling Technique (SMOTE). This methodological approach ensures a systematic and comprehensive analysis of sentiment patterns within YouTube comments, allowing for a nuanced understanding of public sentiments towards destination content. The utilization of CRISP-DM provides a structured framework, ensuring methodical execution from data collection to interpretation. The inclusion of SVM, known for its effectiveness in classification tasks, and SMOTE to address imbalances in sentiment class distribution, enhances the research's robustness. In conclusion, the urgency of the research is underlined by the need to adapt to evolving digital landscapes, and the chosen process ensures a thorough investigation into public sentiments surrounding tourism-themed content on YouTube.

The practical implications of this research are significant for both academia and industry. Firstly, findings from this study can provide destination managers and marketers with valuable insights into the effectiveness of digital content, specifically YouTube videos, in shaping public sentiment towards tourism destinations. By understanding the sentiment patterns identified through applying the CRISP-DM methodology and the Support Vector Machine algorithm, practitioners are able to optimize branding strategies to better resonate with the target audience. Furthermore, the study's methodological approach is a framework for future research endeavors to analyze public sentiment toward destination branding efforts across various digital platforms. Overall, the outcomes of this research hold promise for enhancing destination marketing practices and fostering positive perceptions among potential visitors, thereby contributing to the sustainable development of tourism destinations.

The theoretical implications of this research are noteworthy within destination branding and digital marketing. By utilizing the CRISP-DM methodology and Support Vector Machine algorithm to analyze public sentiment toward tourism-themed content on YouTube, this study contributes to advancing knowledge regarding the effectiveness of digital media in shaping destination perceptions. Additionally, the findings of this research offer insights into the evolving nature of destination branding strategies in the digital age, highlighting the importance of understanding and leveraging digital platforms for destination marketing purposes [7]–[10]. This study thus enriches theoretical discourse surrounding destination branding and digital media, providing a foundation for further scholarly inquiry in this area.

The limitations of this research should be acknowledged to provide a comprehensive understanding of its scope and implications. Firstly, focusing solely on public sentiment towards "A Visual Journey of Wonderful Indonesia" on YouTube may restrict the generalizability of findings to other destination branding efforts or digital platforms. Additionally, relying on the Support Vector Machine algorithm for sentiment analysis may overlook nuances and context-specific factors influencing public perceptions. Moreover, the inherent biases in online user-generated content could impact the accuracy and reliability of the sentiment analysis results. Despite these limitations, this research is a valuable contribution to the literature on destination branding and digital marketing, paving the way for future studies to address these constraints and further advance knowledge in this field.

This research makes a significant contribution to advancing knowledge in the domain of destination branding and digital marketing [2], [11]–[14]. By employing the CRISP-DM methodology and the Support Vector Machine algorithm to analyze public sentiment toward tourism-themed content on YouTube, this study sheds light on the effectiveness of digital media platforms in shaping destination perceptions. The findings offer insights into the evolving nature of destination branding strategies in the digital age and highlight the importance of understanding and leveraging digital platforms for destination marketing [8], [15]–[18]. This research provides a foundation for further scholarly inquiry and practical applications in destination marketing, thereby enriching the existing body of knowledge in the field.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis of Destination Branding & Destination Image Topics using Vosviewer

The research gap concerning destination branding and destination image revolves around the limited attention given to the intersectionality of these constructs in the context of digital media platforms. While previous studies have extensively examined destination branding strategies and their impact on destination image formation, there remains a notable absence of research focusing specifically on the role of digital platforms, such as social media and online review sites, in shaping destination perceptions [19]–[24]. Additionally, existing literature predominantly emphasizes the perspectives of destination marketers and managers, overlooking the crucial role of tourists and online users in co-creating destination images through user-generated content [25], [26]. Addressing this research gap is essential for understanding the dynamic processes involved in destination branding and image formation in the digital age.

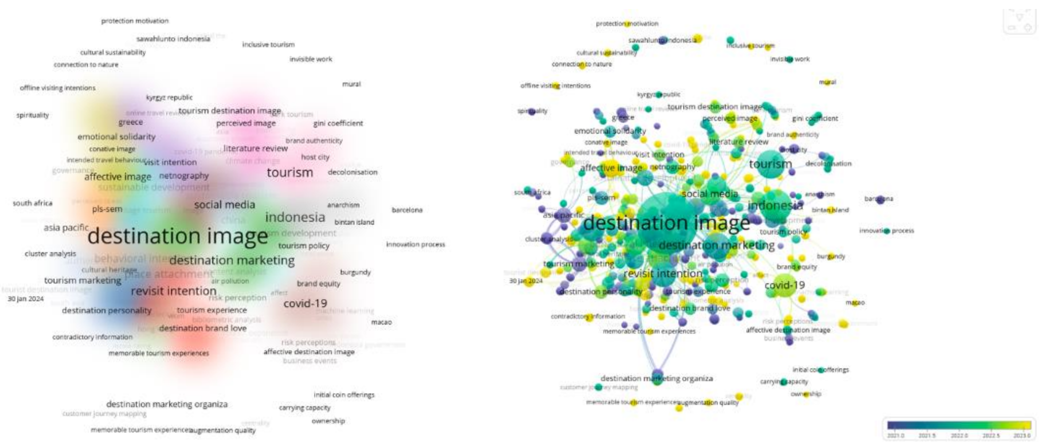

Figure 1. Gap Analysis of Destination Branding and Destination Image Topics

Figure 1 shows the gap analysis of destination branding and destination image topics. Based on the results of the gap analysis conducted using Vosviewer, it is evident that destination branding and destination image warrant further examination, particularly from the data mining perspective, specifically sentiment analysis based on algorithms as a classification model. The identified gap underscores the need for a more nuanced exploration of the intersection between destination branding, destination image, and advanced data mining techniques. By incorporating sentiment analysis algorithms as a classification model, researchers gain deeper insights into the dynamic and evolving nature of public perceptions towards tourism destinations. This approach allows for a more comprehensive understanding of the factors influencing sentiment formation, thus contributing to the refinement and optimization of destination branding strategies in the contemporary digital landscape.

This study examines the performance evaluation of the support vector machine (SVM) algorithm in sentiment classification using "a visual journey of wonderful Indonesia" content. By addressing the limited attention given to the role of digital media platforms in shaping destination perceptions and sentiment analysis, this research aims to contribute to filling the existing gap in the literature. Specifically, the study investigates the effectiveness of the SVM algorithm in classifying sentiments expressed in YouTube comments related to Indonesian tourism content. Through this inquiry, the research seeks to provide insights into the potential of machine learning techniques for enhancing destination branding strategies and understanding public sentiment towards tourism destinations in the digital era.

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

This research employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology as a methodological framework for the classification process. CRISP-DM provides a systematic approach to data mining, comprising six distinct phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Using CRISP-DM, this study ensures a structured and comprehensive process for sentiment analysis on YouTube content related to destination branding. The adoption of CRISP-DM underscores the rigor and methodological soundness of the research, enhancing the reliability and validity of the findings [27]–[34]. Overall, leveraging CRISP-DM as a methodological framework strengthens the study's analytical process and advances destination branding and sentiment analysis knowledge.

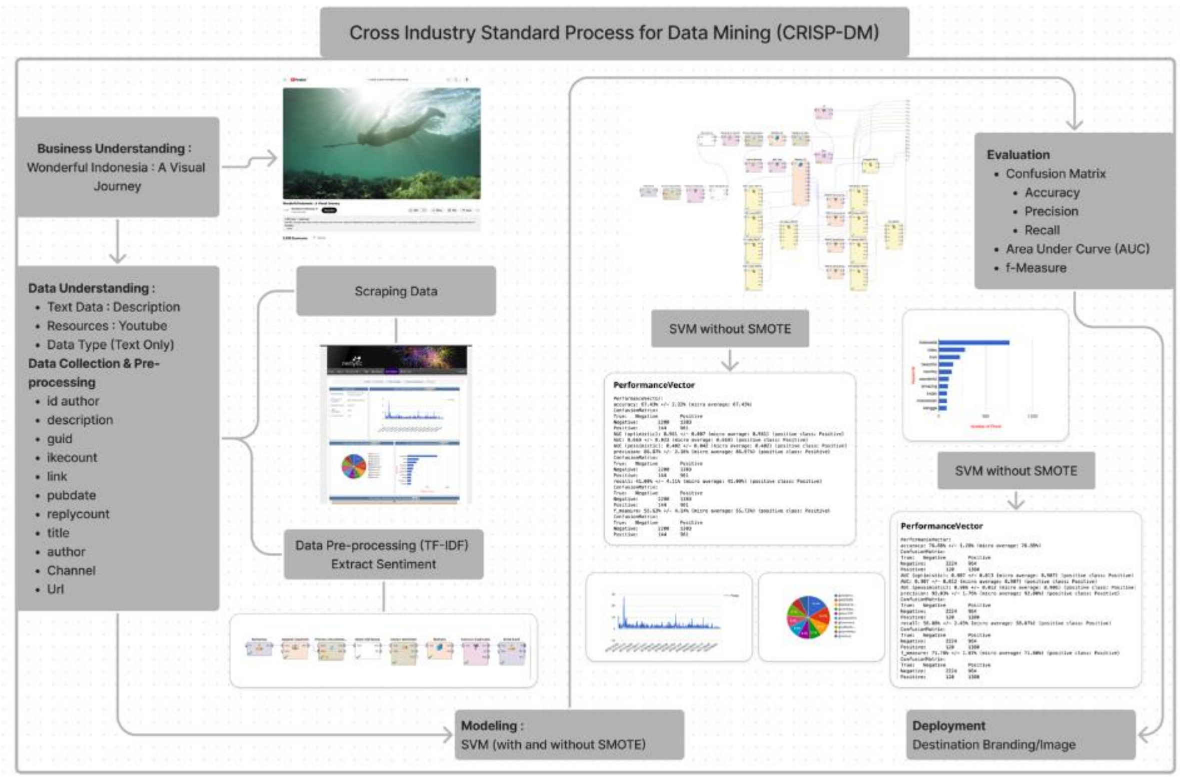

Figure 2. Implementation of CRISP-DM in Sentiment Classification using SVM

Figure 2 showing the framework of CRISP-DM in sentiment classification using SVM. In the business understanding phase, comprehension of destination branding is imperative before selecting suitable tourism destination marketing video content. The chosen video for this study is "Visual Journey of Wonderful Indonesia" from the YouTube platform (ojQbArbuN4E). This initial phase of the CRISP-DM methodology lays the foundation for practical sentiment analysis by ensuring alignment between the research objectives and the selected content. By thoroughly understanding destination branding principles, researchers can better contextualize the analyzed content within the broader framework of destination marketing strategies. Therefore, this phase serves as a crucial precursor to subsequent stages of data understanding, preparation, modeling, evaluation, and deployment, contributing to the methodological rigor and relevance of the research.

In the data understanding phase, the Netlytic website serves as a tool for data collection, enabling the gathering of comments with a maximum limit of 10,000 data points. Based on the acquired number of comments, which amounted to 2555 comments, these data are subsequently processed for modeling. This phase is integral in comprehending the nature and scope of the dataset, ensuring that the subsequent modeling phase can effectively analyze and derive insights from the collected comments. By utilizing Netlytic for data collection, researchers can streamline the process of acquiring relevant data for sentiment analysis, enhancing the efficiency and accuracy of the research endeavor. Thus, the data understanding phase is crucial in facilitating the subsequent stages of analysis and interpretation within the research framework.

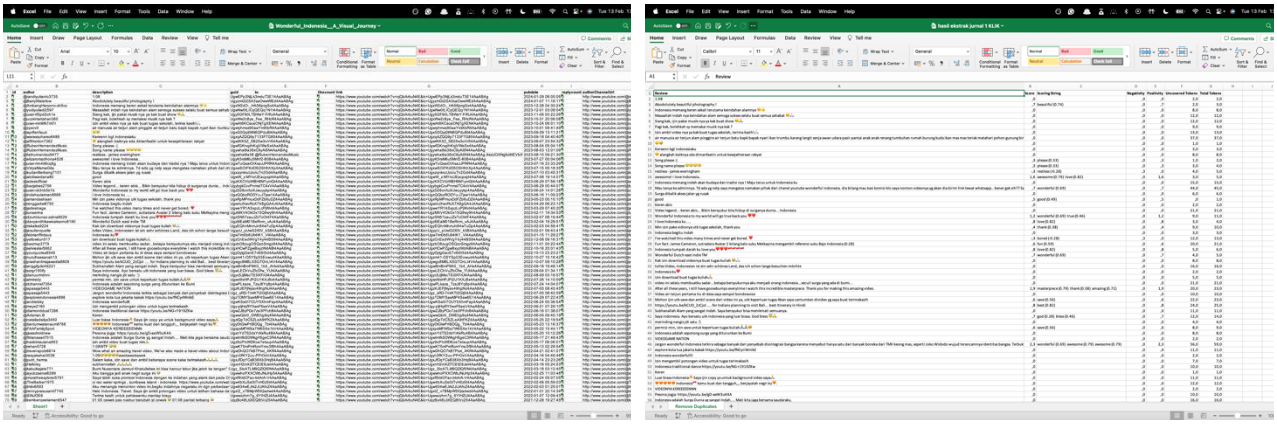

Figure 3. Scraping Data through Netlytic

Figure 3 shows the result of scraping process using Netlytic. In the modeling phase, the collected data undergo selection and extraction using the sentiment extraction operator within the RapidMiner application to obtain string scores based on token counts, thereby facilitating ease of classification. This pivotal stage involves transforming raw data into a structured format amenable to sentiment analysis. By leveraging the sentiment extraction capabilities of RapidMiner, researchers can efficiently derive sentiment scores from the collected comments, enabling subsequent classification processes. Utilizing such tools enhances the methodological rigor and accuracy of sentiment analysis, ultimately contributing to the robustness of the research findings. Hence, the modeling phase is critical in the analytical process, bridging the gap between raw data and actionable insights within destination branding and sentiment analysis.

Furthermore, the algorithm employed in the classification process is the Support Vector Machine (SVM), with a data split of 30% for training data and 70% for testing data, alongside the utilization of the Synthetic Minority Over-sampling Technique (SMOTE) operator to address data imbalance. This strategic approach ensures the robustness and generalizability of the sentiment analysis model by training it on a diverse range of data while mitigating the effects of imbalanced class distributions. By incorporating SMOTE, the model can effectively handle minority class instances, thus enhancing its performance in accurately classifying sentiment patterns. Such methodological considerations underscore the rigor and validity of the research approach, facilitating reliable and meaningful insights into public sentiment towards destination branding efforts on digital platforms. Meanwhile, the regression function of the SVM method is as follows.

$$f(x) = w.x + b \tag{1}$$

Where :

$f(x)$ is the decision function,

W is the weight vector perpendicular to the hyperplane,

X is the input feature vector,

B is the bias term.

In the case of binary classification, the class label y_i of a data point x_i can be determined by the sign of $f(x)$:

$$y_i = \begin{cases} +1, \ if \ f(x_i) \ \geq 0 \\ -1, if \ f(x_i) \ < 1 \end{cases} \tag{2}$$

In non-linearly separable cases, SVM utilizes a kernel function $\mathrm{K}(x_i, x_j)$ to map the input feature vectors into a higher-dimensional space where the data becomes linearly separable. The decision function then becomes:

$$f(x) \ = \ \sum_{i=1}^{N} \propto_i, y_j \, \mathrm{K}(x_i, x) + \mathrm{b} \tag{3}$$

Where $\propto_i$ are the Lagrange multipliers obtained during training. In addition, SMOTE is employed as a strategic approach to tackle data imbalance, consequently augmenting the effectiveness of the Support Vector Machine (SVM). This methodological choice underscores the importance of addressing imbalanced class distributions in sentiment analysis tasks, as it ensures that the model can adequately capture minority class instances and avoid bias towards dominant classes. By leveraging SMOTE, the SVM algorithm can generate synthetic samples for the minority class, thereby balancing the dataset and improving the model's ability to classify sentiment patterns accurately. Thus, integrating SMOTE enhances the robustness and reliability of the sentiment analysis model, ultimately contributing to more meaningful and actionable insights in the context of destination branding and digital marketing strategies.

In the evaluation phase, the metrics of accuracy, precision, recall, Area Under the Curve (AUC), and f-measure will be assessed to gauge the performance of the sentiment analysis model. This crucial stage comprehensively examines the model's predictive capabilities and effectiveness in classifying sentiment patterns within the analyzed data. By evaluating these key metrics, researchers can ascertain the model's reliability, robustness, and overall efficacy in accurately capturing and categorizing sentiment instances. The evaluation process is critical in validating the applicability and relevance of the sentiment analysis model within the context of destination branding and digital marketing strategies, thereby enhancing the trustworthiness and utility of the research findings.

In the Deployment phase, the examination of destination branding or image through tourism destination marketing video content can be analyzed based on tourists' behavior through video comments. This pivotal stage involves applying the sentiment analysis model developed in earlier phases to real-world scenarios, allowing researchers to derive actionable insights into the effectiveness of destination branding strategies on digital platforms. By deploying the sentiment analysis model to analyze tourists' comments on destination marketing videos, researchers can gain valuable perspectives on tourists' perceptions, preferences, and attitudes towards specific destinations, thereby informing destination managers and marketers in refining their branding strategies to better resonate with target audiences. Thus, the Deployment phase is crucial in bridging the gap between theoretical insights and practical implications within destination branding and marketing.

3. RESULT AND DISCUSSION

Destination branding emerges as an effective marketing strategy by disseminating audio-visual content across social media platforms and YouTube. This primary approach capitalizes on the visual storytelling potential of multimedia content to evoke emotional connections and shape perceptions of tourist destinations [35]. By leveraging the immersive nature of audio-visual materials, destination marketers can create compelling narratives that resonate with target audiences, ultimately influencing their decision-making processes [36]. Consequently, destination branding initiatives underscore the transformative power of multimedia content in fostering positive destination perceptions and driving visitor engagement [37]. Thus, the strategic deployment of audio-visual content within destination branding signifies a potent tool for enhancing destination competitiveness and sustainability in the contemporary tourism landscape.

CRISP-DM employed in evaluating the performance of tourism destination marketing is sentiment analysis of public responses through comments on YouTube videos. This primary method entails scrutinizing the sentiment expressed in user comments to gauge the effectiveness of destination marketing efforts. By examining the public's reactions and perceptions reflected in these comments, destination managers and marketers can obtain valuable insights into the success of their branding strategies and identify improvement areas. The sentiment analysis in YouTube comments serves as a direct avenue for understanding the impact of destination marketing content on the audience, providing a nuanced evaluation of the effectiveness and resonance of promotional efforts within the digital landscape. Consequently, the sentiment analysis of YouTube comments emerges as a valuable tool for refining destination marketing strategies and enhancing the overall effectiveness of tourism promotion initiatives.

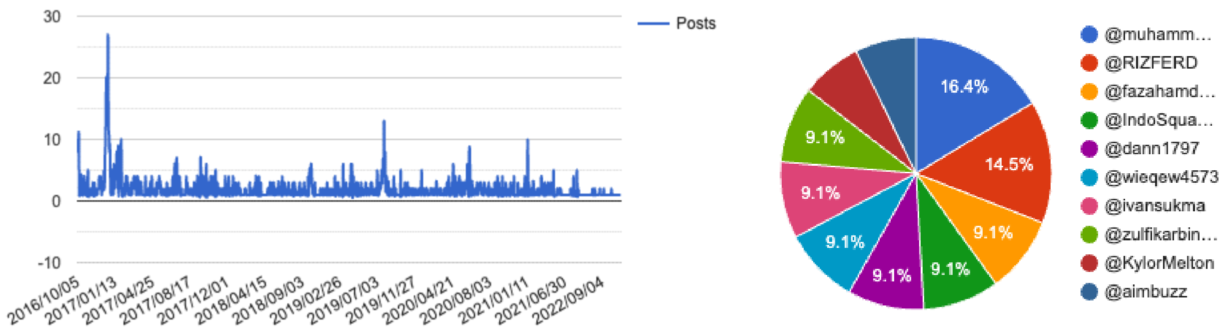

Figure 4. Post over Time and Top Ten Posters

Figure 4 shows the post over time and top ten posters of the content. Based on the results of sentiment data extraction, it is evident that string scores can serve as a guideline for sentiment classification based on positive and negative classes. This outcome underscores the significance of employing sentiment analysis techniques to derive meaningful insights from textual data, particularly in discerning sentiment's polarity in comments or reviews. Using string scores as a reference point, researchers can effectively categorize sentiments into favorable or hostile classes, thereby facilitating a comprehensive understanding of public perceptions and attitudes toward tourism destinations. Consequently, the extraction of sentiment data and subsequent classification based on string scores provide valuable insights into sentiment patterns, aiding destination managers and marketers in devising tailored strategies to enhance destination branding and marketing efforts.

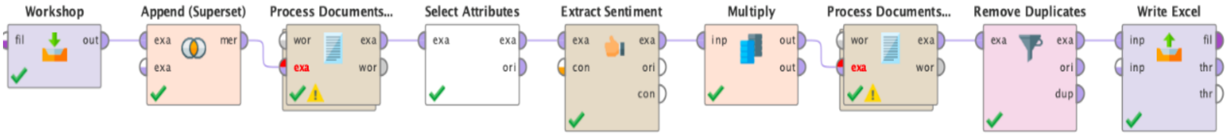

Figure 5. Extract Sentiment Using Rapidminer

Figure 5 shows the extract result using extract sentiment operator in Rapidminer. Data cleaning is conducted before the sentiment extraction process to ensure that words with negative or positive connotations can be classified optimally. This preliminary step is crucial in preparing the dataset for sentiment analysis, as it involves removing noise, irrelevant information, and inconsistencies from the textual data. By cleansing the data, researchers can enhance the accuracy and reliability of sentiment classification, thereby improving the effectiveness of subsequent analyses. This underscores the importance of data preprocessing techniques in mitigating potential biases and inaccuracies that may arise during sentiment analysis, ultimately contributing to more robust and meaningful insights into public perceptions and attitudes toward tourism destinations.

The Extract Sentiment operator in RapidMiner offers significant benefits for sentiment analysis tasks. As the primary tool for extracting sentiment from textual data, this operator facilitates the classification of sentiments into positive, negative, or neutral categories, thereby enabling researchers to gain valuable insights into public opinions and attitudes. Through its efficient processing capabilities, the Extract Sentiment operator enhances the accuracy and reliability of sentiment analysis, enabling researchers to make informed decisions and derive actionable insights from large volumes of textual data. Overall, the utilization of this operator in RapidMiner streamlines the sentiment analysis process, ultimately contributing to more effective decision-making and strategy formulation in various domains, including marketing, customer service, and social media monitoring.

Table 1. Classification based on Extract Sentiment Result

Review	String Score	Score	Classification
This video was only taken from 4 locations out of hundreds of beautiful places out there. People who live in a big city like Jakarta or Surabaya sometimes got so caught up in their own small busy life in the city, that they hate their country and dream to move somewhere else outside Indonesia. They just forget how beautiful this country that they live in.	beautiful (0.74) like (0.38) hate (-0.69) dream (0.26) forget (-0.23) beautiful (0.74)	1,20512820512821	Positive
*Booo... its only advertisement! in the reality, your country is dirty, racist and fanatic sh**! Singapore is better!*	dirty (-0.49) racist (-0.77) better (0.49)	-0,769230769230769	Negative

Table 1 shows the classification results obtained through the Extract Sentiment operator; performance testing can be conducted using Support Vector Machine (SVM) models. This approach allows researchers to evaluate the effectiveness and accuracy of sentiment analysis in categorizing textual data into positive, negative, or neutral sentiments. By leveraging SVM models, researchers can assess the predictive capabilities of the sentiment classification algorithm, thereby gauging its suitability for analyzing public perceptions and attitudes toward tourism destinations. This testing process provides valuable insights into the reliability and robustness of the sentiment analysis model, ultimately enhancing the validity and applicability of findings in the context of destination branding and marketing strategies.

Support Vector Machine (SVM) exhibits several advantages in classification tasks. As a robust supervised learning algorithm, SVM handles linear and non-linear data by separating different classes with a hyperplane in a high-dimensional space. Its ability to maximize the margin between classes ensures robust generalization performance, making it less prone to overfitting. Additionally, SVM can handle datasets with many features and is relatively insensitive to outliers. Its versatility allows various kernel functions to be employed, enabling the classification of complex datasets with non-linear decision boundaries. These strengths position SVM as a highly effective and widely used tool in classification tasks across diverse domains, ranging from text categorization to image recognition.

The performance of the Support Vector Machine (SVM) algorithm using Synthetic Minority Over-sampling Technique (SMOTE) can significantly enhance accuracy values. This is primarily due to SMOTE's ability to address imbalanced datasets by generating synthetic samples from the minority class, thereby mitigating the bias towards the majority class. By oversampling the minority class, SVM with SMOTE ensures that the classifier is trained on a more balanced dataset. This leads to improved classification accuracy, particularly in scenarios where class imbalance is prevalent. Consequently, leveraging SMOTE with SVM holds promise in enhancing classification models' overall performance and reliability, making it a valuable approach in various fields, including sentiment analysis, medical diagnosis, and fraud detection.

Table 2. Confusion Matrix of SVM with and without SMOTE

Confusion Matrix	SVM with SMOTE	SVM without SMOTE
Accuracy	84.26%	97.08%
Precision	100.00%	97.08%
Recall	68.51%	100.00%
F-measure	81.25%	98.52%
AUC	0.996	0.607

Based on Table 2, it is evident that SVM with SMOTE demonstrates notable performance metrics, with an accuracy of 84.26%, precision of 100.00%, recall of 68.51%, f-measure of 81.25%, and an Area Under the Curve (AUC) value of 0.996. These results highlight the effectiveness of the SVM-SMOTE approach in accurately classifying sentiment patterns within the analyzed dataset. The high precision score indicates the model's ability to correctly identify positive sentiment instances, while the recall score suggests its capability to capture the majority of positive sentiment cases. Moreover, the substantial AUC value underscores the model's overall predictive performance, emphasizing its reliability in discriminating between positive and negative sentiment instances. Therefore, these findings provide valuable insights into the efficacy of SVM with SMOTE in sentiment analysis and offer practical implications for destination branding and marketing strategies.

Subsequently, SVM without SMOTE demonstrates impressive performance metrics, with an accuracy of 97.08%, precision of 97.08%, recall of 100.00%, f-measure of 98.52%, and an Area Under the Curve (AUC) value of 0.607. These results signify the robustness of the SVM algorithm in accurately classifying sentiment patterns within the dataset, even without employing oversampling techniques to address data imbalance. The high precision and recall scores indicate the model's ability to identify and capture positive and negative sentiment instances effectively. However, the comparatively lower AUC value suggests potential limitations in discriminating between positive and negative sentiment instances compared to the SVM-SMOTE approach. Nonetheless, these findings highlight the promising performance of SVM without SMOTE in sentiment analysis and underscore its potential utility in informing destination branding and marketing strategies.

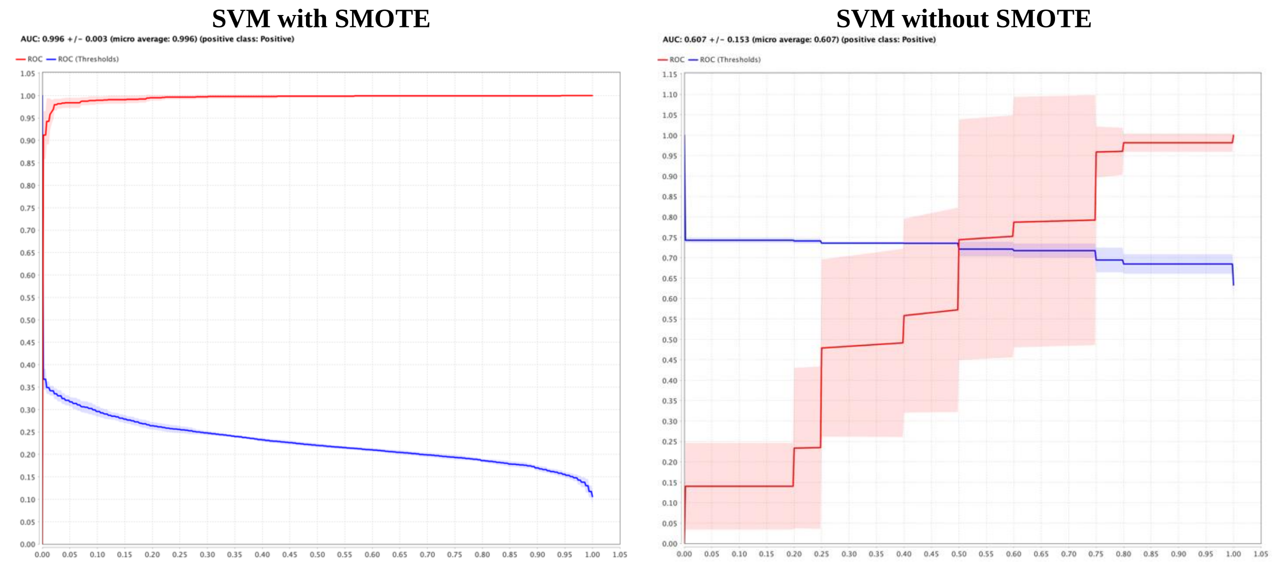

Figure 6. Area Under Curve

Figure 6 shows the AUC of SVM with SMOTE and SVM without SMOTE. The analysis of the two approaches, SVM with SMOTE and SVM without SMOTE, reveals distinct patterns in their performance metrics. For SVM with SMOTE, the accuracy, precision, and recall scores demonstrate its effectiveness in classifying sentiment patterns, particularly in identifying positive sentiment instances with high precision. The substantial AUC value further validates its reliability in distinguishing between positive and negative sentiment instances, emphasizing its overall predictive performance. Conversely, SVM without SMOTE showcases impressive accuracy and precision scores, indicating its robustness in accurately classifying sentiment patterns even without oversampling techniques. However, the lower AUC value suggests potential limitations in discriminating between positive and negative sentiment instances compared to SVM with SMOTE. Nonetheless, both approaches offer valuable insights into sentiment analysis, with SVM without SMOTE showing promise as a viable alternative in informing destination branding and marketing strategies.

Based on the analysis of frequently used words in the review data, it is evident that specific terms stand out significantly in terms of frequency. For instance, the word "Indonesia" appears 736 times, followed by "video" with 283 occurrences, and "love" with 215 mentions. Additionally, words like "beautiful" and "country" are mentioned 150 and 139 times, respectively, suggesting a positive portrayal of Indonesia's aesthetic appeal and overall allure. Furthermore, "wonderful" and "amazing" appear 109 and 98 times, respectively, emphasizing the overwhelmingly positive sentiment associated with Indonesia as a destination. Interestingly, the word "*Indah*" which means beautiful in Bahasa Indonesia, is observed 92 times, indicating a recognition of the country's natural beauty among local and international audiences. Additionally, "orang" and "Indonesian" are mentioned 71 and 89 times, respectively, reflecting a sense of appreciation for Indonesia's culture and people. Overall, these frequent positive descriptors underscore the effectiveness of destination branding efforts in shaping favorable perceptions of Indonesia as a tourism destination.

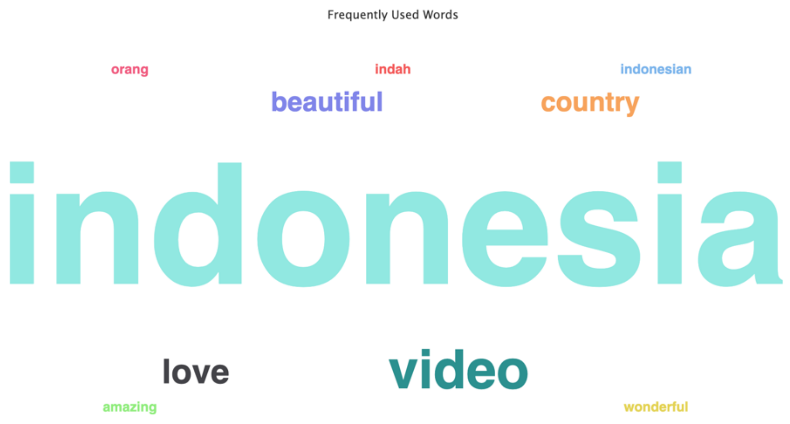

Figure 7. Frequently used Words

Figure 7 shows the frequently used words of the content reviews. From a destination branding perspective, analyzing frequently used words in the review data provides valuable insights into how Indonesia is perceived and portrayed as a tourism destination. The high frequency of terms such as "Indonesia," "video," "love," "beautiful," "country," "wonderful," and "amazing" suggests a positive and visually captivating image of the country. These words evoke admiration, fascination, and appreciation, indicating that Indonesia is seen as a destination with aesthetic appeal and intrinsic charm. The repetition of terms like "beautiful" and "amazing" underscores the visual allure and natural beauty of Indonesia, while words like "wonderful" highlight the overall positive experiences associated with visiting the country. Additionally, the frequent use of "orang" and "Indonesian" signifies a recognition and appreciation of Indonesia's rich cultural heritage and its people. Overall, the analysis reflects Indonesia's solid and favorable brand image as a destination characterized by its natural beauty, cultural richness, and warm hospitality. These positive perceptions and sentiments are essential for destination marketers and managers in crafting effective branding strategies to attract and retain tourists, enhancing Indonesia's competitive position in the global tourism market.

In deploying destination branding strategies based on the insights gathered from the analysis of frequently used words in the review data, several recommendations can be made to enhance the effectiveness of deployment efforts, such as content creation and distribution, capitalize on the positive sentiments expressed toward Indonesia's natural beauty, cultural richness, and hospitality by creating visually captivating and culturally immersive content [38]. Utilize platforms like YouTube to showcase Indonesia's diverse attractions, local experiences, and warm hospitality through high-quality videos and engaging storytelling. Engage with influencers and collaborate with travel influencers, content creators, and local advocates to amplify the reach and impact of the destination branding efforts [39]. Encourage users to share their authentic experiences and positive stories about Indonesia, leveraging their influence to attract and inspire potential visitors [40]. Community Engagement and Participation foster community engagement and participation by encouraging travelers to share their experiences and recommendations about Indonesia on social media platforms [41]. Create interactive campaigns, contests, and challenges to encourage user-generated content and foster a sense of belonging among travelers.

In addition, optimize the targeted marketing and promotion campaigns to specific target audiences based on their interests, preferences, and demographics [42]. Utilize data-driven insights to identify key markets and segments and craft personalized messaging and offers that resonate with their needs and aspirations [43]. Implement a robust monitoring and evaluation framework to track the performance and impact of destination branding efforts over time [44]. Regularly analyze vital metrics such as engagement levels, sentiment trends, and visitor feedback to identify areas for improvement and optimization [45]. By implementing these recommendations, destination marketers and managers can effectively deploy destination branding strategies that leverage Indonesia's positive sentiments and perceptions, ultimately driving visitor engagement, loyalty, and advocacy in the global tourism market.

4. CONCLUSION

In conclusion, the analysis of frequently used words in the review data provides valuable insights into the perception and portrayal of Indonesia as a tourism destination. With "Indonesia" mentioned 736 times, "video" 283 times, and "love" 215 times, among other significant terms, it is evident that Indonesia is perceived positively, with terms like "beautiful," "wonderful," and "amazing" emphasizing its aesthetic appeal and allure. Additionally, "orang" and "Indonesian" show appreciation for the country's rich cultural heritage and people. These findings underscore the importance of destination branding efforts in shaping favorable perceptions and emotions toward Indonesia. Moving forward, destination marketers and managers can leverage these insights to develop targeted branding strategies that highlight Indonesia's unique attractions, cultural experiences, and warm hospitality, ultimately enhancing its competitiveness and appeal in the global tourism market. In addition, comparing the two approaches in sentiment classification using the Support Vector Machine (SVM) algorithm reveals exciting insights into their respective performance. The SVM algorithm with Synthetic Minority Over-sampling Technique (SMOTE) showcases promising results with an accuracy of 84.26%, precision of 100.00%, recall of 68.51%, and a remarkable Area Under the Curve (AUC) value of 0.996. These metrics indicate its efficacy in accurately classifying sentiment patterns within analyzed YouTube content, offering valuable practical implications for destination managers and marketers. On the other hand, the SVM algorithm without SMOTE achieves notably high accuracy, precision, and recall scores of 97.08%. Yet, its AUC value of 0.607 suggests potential limitations in discriminating between positive and negative sentiment instances. While this approach demonstrates robustness, further investigation is necessary to enhance its performance in capturing nuanced sentiment patterns. Overall, both approaches highlight the significance of SVM algorithms in sentiment analysis for destination branding strategies, albeit with varying degrees of predictive power and potential areas for improvement.

REFERENCES

[1] V. Kumar and A. K. Kaushik, "Does experience affect engagement? Role of destination brand engagement in developing brand advocacy and revisit intentions," *J. Travel Tour. Mark.*, vol. 37, no. 3, pp. 332–346, 2020, doi: 10.1080/10548408.2020.1757562.
[2] G. G. F. Almeida and P. Almeida, "The influence of destination image within the territorial brand on regional development," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2233260.
[3] B. Suna and M. D. Alvarez, "The role of gastronomy in shaping the destination's brand identity: an empirical analysis based on stakeholders' opinions," *J. Hosp. Mark. Manag.*, vol. 30, no. 6, pp. 738–758, 2021, doi: 10.1080/19368623.2021.1877587.
[4] Ö. Güzel, C. Ünal, and İ. Şahin, "How does the cross-cultural view shape the destination brand personality and brand image? Evidence from the travelers," *J. Qual. Assur. Hosp. Tour.*, 2022, doi: 10.1080/1528008X.2022.2112804.
[5] Q. Jiang, C. S. Chan, S. Eichelberger, H. Ma, and B. Pikkemaat, "Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists," *Curr. Issues Tour.*, vol. 24, no. 17, pp. 2501–2522, 2021, doi: 10.1080/13683500.2021.1874312.
[6] M. N. Islam, "Visitors' Sentiment in Archeological Heritage Sites: A Study Based on TripAdvisor Reviews," *Int. J. Hosp. Tour. Adm.*, vol. 00, no. 00, pp. 1–22, 2023, doi: 10.1080/15256480.2023.2242343.
[7] F. Otay Demir, Ş. Yavuz Görkem, and G. Rafferty, "An inquiry on the potential of computational literary techniques towards successful destination branding and literary tourism," *Curr. Issues Tour.*, vol. 25, no. 5, pp. 764–778, 2022, doi: 10.1080/13683500.2021.1887100.
[8] M. A. M. Au-Yeung, V. W. S. Tung, and S. W. Tsz Tse, "Agile destination management and marketing," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–11, 2022, doi: 10.1080/02508281.2022.2078623.
[9] M. Šerić, Đ. O. Došen, and J. Mikulić, "Antecedents and moderators of positive word of mouth communication among tourist destination residents during the COVID-19 pandemic," *Curr. Issues Tour.*, vol. 26, no. 2, pp. 224–241, 2023, doi: 10.1080/13683500.2022.2132922.
[10] W. Rop, "Gaining competitive advantage through destination branding: a factorial analysis," *Anatolia*, vol. 34, no. 4, pp. 465–475, 2023, doi: 10.1080/13032917.2022.2051056.
[11] D. Amani and E. Chao, "How does destination governance build local residents' behavioural support towards destination branding: An empirical study of the tourism sector in Tanzania," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2192441.
[12] C. Pippirs and G. C. Steckenbauer, "Case study–spa destination branding–a strategic realignment process of five Bavarian thermal spas," *Int. J. Spa Wellness*, vol. 5, no. 3, pp. 308–319, 2022, doi: 10.1080/24721735.2022.2117009.
[13] M. Y. Lai, C. Khoo, and Y. Wang, "Critical realism: a methodological design for destination food image research," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–19, 2022, doi: 10.1080/02508281.2022.2098448.
[14] N. F. Lund and A. N. Kimbu, "Applying the Hollywood scriptwriting formula to destination branding," *Curr. Issues Tour.*, vol. 24, no. 8, pp. 1058–1078, 2021, doi: 10.1080/13683500.2020.1739005.
[15] R. A. Trzonkowski and D. S. Mason, "Sport events, brand extensions, and branding Phoenix, Arizona," *J. Sport Tour.*, vol. 27, no. 1, pp. 33–56, 2023, doi: 10.1080/14775085.2022.2145342.
[16] R. A. Rather, A. H. Najar, and D. Jaziri, "Destination branding in tourism: insights from social identification, attachment and experience theories," *Anatolia*, vol. 31, no. 2, pp. 229–243, 2020, doi: 10.1080/13032917.2020.1747223.
[17] J. Kim and J. Bae, "Influences of symbolic capital on destination brand equity," *J. Glob. Sch. Mark. Sci. Bridg. Asia World*, vol. 33, no. 4, pp. 561–576, 2023, doi: 10.1080/21639159.2023.2233087.
[18] C. H. Yen, H. Y. Teng, and S. T. Chang, "Destination brand identity and emerging market tourists' perceptions," *Asia Pacific J. Tour. Res.*, vol. 25, no. 12, pp. 1310–1327, 2020, doi: 10.1080/10941665.2020.1853578.
[19] D. Amani, "The role of destination brand-oriented leadership in shaping tourism destination brand ambassadorship behavior among local residents in Tanzania," *Cogent Soc. Sci.*, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2266634.
[20] L. Wu, B. Taheri, F. Okumus, and S. Wang, "The effects of host sincerity on tourists' perceived destination image," *Serv. Ind. J.*, vol. 44, no. 1–2, pp. 83–104, 2024, doi: 10.1080/02642069.2021.2011861.
[21] D. Amani, "Mediation Role of Destination Brand Patriotism in Social Media Brand Engagement and Destination Brand Value Co-creation: Local Residents' Perspective in the Tourism Sector in Tanzania," *J. Qual. Assur. Hosp. Tour.*, vol. 00, no. 00, pp. 1–26, 2022, doi: 10.1080/1528008X.2022.2149676.
[22] V. Wilk, S. Sadeque, and G. N. Soutar, "Exploring online destination brand advocacy," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–19, 2021, doi: 10.1080/02508281.2021.1992952.

[23] N. Pagano and R. Sharpley, “Destination image / branding of Sicily and the Mafia phenomenon : a corpus-based analysis of keywords in English guidebooks,” *J. Tour. Cult. Chang.*, pp. 1–19, 2024, doi: 10.1080/14766825.2024.2311422.

[24] H. Chauhan, V. K. Jain, and H. Verma, “Destination branding as new tool for economic development: a qualitative approach with reference to Jaunsar Bawar Region, Uttarakhand,” *Int. J. Spa Wellness*, vol. 6, no. 1, pp. 109–130, 2023, doi: 10.1080/24721735.2022.2115331.

[25] J. Hemmonsbey and T. M. Tichaawa, “Brand messages that influence the sport tourism experience: the case of South Africa,” *J. Sport Tour.*, vol. 24, no. 3, pp. 177–194, 2020, doi: 10.1080/14775085.2020.1822200.

[26] M. Aktan, U. Zaman, and S. Nawaz, “Examining destinations’ personality and brand equity through the lens of expats: moderating role of expat’s cultural intelligence,” *Asia Pacific J. Tour. Res.*, vol. 26, no. 8, pp. 849–865, 2021, doi: 10.1080/10941665.2021.1925314.

[27] Y. A. Singgalen, “Analisis Sentimen Konsumen terhadap Food , Services , and Value di Restoran dan Rumah Makan Populer Kota Makassar Berdasarkan Rekomendasi Tripadvisor Menggunakan Metode CRISP-DM dan,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1899–1914, 2023, doi: 10.47065/bits.v4i4.3231.

[28] Y. A. Singgalen, “Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT),” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1551–1562, 2023, doi: 10.30865/mib.v7i3.6461.

[29] Y. A. Singgalen, “Analisis Sentimen Top 10 Traveler Ranked Hotel di Kota Makassar Menggunakan Algoritma Decision Tree dan Support Vector Machine,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 323–332, 2023, doi: 10.30865/klik.v4i1.1153.

[30] Y. A. Singgalen, “Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1870–1882, 2023, doi: 10.47065/bits.v4i4.3199.

[31] Y. A. Singgalen, “Analisis Sentimen Wisatawan terhadap Taman Nasional Bunaken dan Top 10 Hotel Rekomendasi Tripadvisor Menggunakan Algoritma SVM dan DT berbasis CRISP-DM,” *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 367–379, 2023, doi: 10.47065/josyc.v4i2.3092.

[32] Y. A. Singgalen, “Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM,” *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 614–625, 2023, doi: 10.47065/josh.v4i2.2999.

[33] Y. A. Singgalen, “Analisis Perilaku Wisatawan Berdasarkan Data Ulasan di Website Tripadvisor Menggunakan CRISP-DM : Wisata Minat Khusus Pendakian Gunung Rinjani dan Gunung Bromo,” *J. Comput. Syst. Informatics*, vol. 4, no. 2, pp. 326–338, 2023, doi: 10.47065/josyc.v4i2.3042.

[34] Y. A. Singgalen, “Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW,” *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1343–1353, 2023, doi: 10.47065/josh.v4i4.3917.

[35] Z. Ghaderi, M. J. Mahdavizadeh, M. Rajabi, and C. M. Hall, “Does storytelling affect destination image, destination personality, and tourists’ behavioural intention?,” *Anatolia*, vol. 00, no. 00, pp. 1–13, 2023, doi: 10.1080/13032917.2023.2191250.

[36] G. Manimont, S. Pike, A. Beatson, and S. Tuzovic, “Culinary destination consumer-based brand equity: exploring the influence of tourist gaze in relation to FoodPorn on social media,” *Tour. Recreat. Res.*, vol. 47, no. 3, pp. 221–240, 2022, doi: 10.1080/02508281.2021.1969623.

[37] M. A. Saleem, J. Li, and H. Afzal, “Protect for affinity? The role of destination brand engagement in forming environmentally responsible tourist behaviours,” *J. Sustain. Tour.*, vol. 29, no. 8, pp. 1344–1364, 2021, doi: 10.1080/09669582.2020.1835932.

[38] M. S. Satar, R. A. Rather, S. H. Parrey, H. Khan, and T. Rasul, “Eliciting consumer-engagement and experience to foster consumer-based-brand-equity: moderation of perceived-health-beliefs,” *Serv. Ind. J.*, vol. 0, no. 0, pp. 1–26, 2023, doi: 10.1080/02642069.2023.2191953.

[39] S. Zakiah, A. Winarno, and D. Hermana, “Examination of consumer engagement for loyalty in sustainable destination image,” *Cogent Soc. Sci.*, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2269680.

[40] D. Amani, “It’s mine, that’s why I speak: Destination brand psychological ownership as a mediator of the link between social media brand engagement and local residents’ voice behavior,” *Cogent Bus. Manag.*, vol. 9, no. 1, 2022, doi: 10.1080/23311975.2022.2031432.

[41] C. Dolezal and M. Novelli, “Power in community-based tourism: empowerment and partnership in Bali,” *J. Sustain. Tour.*, vol. 30, no. 10, pp. 2352–2370, 2022, doi: 10.1080/09669582.2020.1838527.

[42] M. W. Lever, S. Elliot, and M. Joppe, “Exploring destination advocacy behavior in a virtual travel community,” *J. Travel Tour. Mark.*, vol. 38, no. 5, pp. 431–443, 2021, doi: 10.1080/10548408.2021.1940421.

[43] D. Y. Reindrawati, “Challenges of community participation in tourism planning in developing countries,” *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2022.2164240.

[44] S. Moghavvemi, K. M. Woosnam, A. Hamzah, and A. Hassani, “Considering Residents’ Personality and Community Factors in Explaining Satisfaction with Tourism and Support for Tourism Development,” *Tour. Plan. Dev.*, vol. 18, no. 3, pp. 267–293, 2021, doi: 10.1080/21568316.2020.1768140.

[45] S. Kim and E. Park, “Identifying film tourism impacts on grassroots communities: implications for film tourism development and community planning,” *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–5, 2022, doi: 10.1080/02508281.2022.2133810.