Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

Extract Sentiment and Support Vector Machine (SVM) Performance of Hotel Guest Review Classification

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: **28/12/2023**; Accepted: **30/12/2023**; Published**: 30/12/2023**

Abstract−The hotel accommodation business highly depends on consumer preferences regarding products and services. The intensity of hotel guest visits and the level of guest satisfaction with the services provided by hotel management can be seen from various guest reviews on websites used as reservation media. Therefore, this research uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) method to implement the *data mining* process using the webharvy application and *the machine learning* process using the Rapidminer application. Meanwhile, the operators used are Synthetic Minority Over-sampling Technique SMOTE in overcoming data imbalances and *sentiment extract* operators to obtain a total string score before sentiment labels are determined and processed using the Support Vector Machine (SVM) algorithm. The results of this study showed that SVM without using SMOTE operators resulted in an accuracy value of 95.82%, a precision value of 95.80%, a recall value of 100%, and an Area Under Curve (AUC) value of 0.798 (79.8%). Otherwise, SVM performance using SMOTE operators produces an accuracy value of 92.05%, a precision value of 100%, a recall value of 84.08%, and an Area Under Curve (AUC) value of 99.99 (99.9%). Furthermore, based on ten popular words, hotel guests are concerned about breakfast, staff, pool, room, and hotel. Thus, the guests' highlights are the menu served by the hotel, the service provided by employees, room conditions, and hotel brands. Therefore, hotel management needs to improve the quality of products and services to increase satisfaction and intention to stay again.

Keywords: Extract Sentiment; SVM; Hotel; Guest Review; Classification

1. INTRODUCTION

The research background of sentiment analysis in hotel guest reviews is pivotal in computational linguistics and hospitality management studies. Analyzing sentiment in hotel guest reviews involves intricate text-mining techniques to extract valuable insights from unstructured textual data. Natural language processing algorithms enable the identification and categorization of sentiments expressed by guests, ranging from positive commendations to negative critiques, facilitating a comprehensive understanding of customer satisfaction and dissatisfaction factors. This research underscores the significance of sentiment analysis as a powerful tool for hoteliers to enhance service quality, optimize guest experiences, and foster competitive advantage in the hospitality industry.

The urgency of sentiment analysis in hotel guest reviews is paramount within the hospitality industry, given its pivotal role in gauging customer satisfaction and driving operational improvements. By applying advanced computational linguistics techniques, sentiment analysis empowers hoteliers to identify trends swiftly, pinpoint areas for enhancement, and respond promptly to guest feedback. The rapid evolution of online review platforms necessitates real-time analysis of guest sentiments, enabling hotels to adapt and tailor their services to meet evolving consumer preferences. Consequently, integrating sentiment analysis into hospitality management strategies enhances guest experiences and fosters sustainable business growth and competitive advantage in an increasingly dynamic marketplace.

Support Vector Machine (SVM) algorithm exhibits notable advantages in text classification for sentiment analysis. The primary strength lies in its ability to construct an optimal hyperplane that maximally separates different classes within a high-dimensional feature space, making it particularly effective in handling complex relationships within textual data [1]. SVM is adept at discerning subtle patterns and nuances in sentiment expression, enabling it to capture the intricacies of language and context [2]. This capability enhances the algorithm's accuracy in distinguishing between positive and negative sentiments, even when sentiment polarity is less explicit [3]. In conclusion, the robust performance of SVM in navigating intricate feature spaces and discerning nuanced sentiment patterns establishes it as a powerful tool for text classification in sentiment analysis applications.

The performance of the Support Vector Machine (SVM) in text classification for sentiment analysis across diverse domains is noteworthy [4]. SVM excels in handling the intricacies of sentiment expression by constructing an optimal hyperplane within high-dimensional feature spaces, showcasing its adaptability to various fields of study [5]. Its effectiveness is underscored by its capacity to discern subtle patterns in textual data, allowing for accurate sentiment analysis even in contexts where sentiment polarity is less overt [6]. This versatility makes SVM a valuable asset in applications ranging from product reviews and social media sentiments to academic literature and customer feedback [7]. In conclusion, SVM's consistent and robust performance across diverse domains establishes it as a reliable and effective tool for text classification in sentiment analysis across a broad spectrum of research areas.

Examining machine learning applications in sentiment extraction processes is imperative, particularly in identifying sentiments within hotel guest reviews [8]. The primary focus of this study is to understand and enhance the efficacy of machine learning algorithms in extracting nuanced sentiments expressed by hotel guests [9]. With the exponential growth of online reviews, the need for accurate sentiment analysis has become paramount for the

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

hospitality industry [10]. Machine learning algorithms, when applied to hotel guest reviews, have the potential to discern subtle nuances, contextual variations, and user-specific sentiments [11]. This research aims to contribute valuable insights into refining machine learning models, ensuring their applicability and accuracy in discerning sentiment orientations within the intricate realm of hotel guest feedback [12]. In conclusion, a comprehensive investigation into machine learning for sentiment analysis in the context of hotel reviews is essential to advance the precision and relevance of sentiment extraction methods in the hospitality domain.

The sentiment analysis of hotel guests is crucial in generating recommendations for accommodation service providers to enhance guest satisfaction [13]. The primary emphasis lies in understanding and interpreting the sentiments expressed by hotel guests, as this insight holds the key to tailoring services to meet their expectations effectively [14]. By employing sentiment analysis techniques, hotel providers can discern the specific aspects of a guest's experience that contribute to overall satisfaction or dissatisfaction [15]. This nuanced understanding allows for targeted improvements in service delivery, addressing both positive and negative sentiments [16]. Informed by sentiment analysis outcomes, accommodation providers can develop strategic initiatives to elevate the guest experience, fostering loyalty and positive word-of-mouth recommendations [17]. In conclusion, integrating sentiment analysis in evaluating hotel guest experiences provides valuable insights in formulating recommendations for service enhancement and ensuring sustained guest satisfaction.

Guest satisfaction in hotels is determined by various criteria encompassing facilities, cleanliness, service quality, comfortable accommodations, strategic location, and reasonable pricing. The principal determinant of guest contentment is a holistic evaluation of these diverse factors [18]. The availability and quality of amenities, including recreational facilities, dining options, and room features, contribute significantly to overall satisfaction [19]. Cleanliness standards in public spaces and guest rooms are pivotal in shaping positive perceptions [20]. Furthermore, the efficiency and warmth of the service staff, the comfort of accommodations, the strategic positioning of the hotel about key attractions, and the perceived value for money are integral elements that collectively influence guest satisfaction [21]. In essence, synthesizing these multifaceted aspects determines guests' overall impression of their stay [22]. In conclusion, the comprehensive evaluation of facilities, cleanliness, service, comfort, location, and pricing is essential for understanding and optimizing guest satisfaction in the hotel industry.

This research aims to analyze hotel guest sentiments through a machine-learning approach. The primary objective is to employ advanced computational techniques to extract and comprehend sentiments expressed by hotel guests [23]. The utilization of machine learning algorithms facilitates the automatic identification of sentiments, enabling a systematic and efficient analysis of a large volume of textual data from guest reviews [24]. The significance of this approach lies in its potential to uncover nuanced sentiments, identify patterns, and categorize opinions, contributing to a deeper understanding of guest experiences [25]. By employing machine learning in sentiment analysis, this study aspires to provide a robust and automated methodology that enhances the accuracy and efficiency of evaluating hotel guest sentiments, thereby contributing valuable insights to the hospitality industry [26]. In conclusion, integrating machine learning into sentiment analysis offers a promising avenue for comprehensively examining hotel guest sentiments, with implications for improving service quality and guest satisfaction [27].

The limitation of this research lies in the exclusive utilization of the Support Vector Machine (SVM) algorithm for sentiment analysis, overlooking other classification algorithms that could offer insightful performance comparisons based on accuracy, precision, and recall metrics. While SVM is acknowledged for its effectiveness in handling high-dimensional feature spaces and intricate patterns, its singular application may restrict the exploration of alternative algorithms. Various classification algorithms, such as Naive Bayes, Decision Trees, and Random Forests, exhibit diverse strengths and weaknesses that can significantly impact the outcomes of sentiment analysis. A more comprehensive approach involving multiple algorithms could provide a more nuanced understanding of their comparative performance. Therefore, this study acknowledges the limitation of focusing solely on SVM and recognizes the potential for future research to explore a broader spectrum of classification algorithms to enhance the overall robustness and validity of sentiment analysis results.

The recommendation for further research involves conducting a comparative analysis of sentiment classification algorithms in the context of hotel guest reviews to assess performance suitability across diverse datasets. The primary focus is to extend the investigation beyond the singular use of a specific algorithm, such as Support Vector Machine (SVM), by systematically comparing it with alternative classification algorithms. This approach allows for a more comprehensive understanding of algorithmic strengths and weaknesses in handling varied sentiments expressed by hotel guests. By incorporating a range of algorithms, including Naive Bayes, Decision Trees, and Random Forests, researchers can discern the most effective algorithm for sentiment analysis based on dataset-specific characteristics. Such a comparative study can enhance the applicability and generalizability of sentiment analysis findings in hotel guest reviews. In conclusion, recommending further research that embraces algorithmic diversity facilitates a more nuanced evaluation of sentiment classification methods, contributing to the refinement and optimization of sentiment analysis in the context of hotel guest satisfaction.

2. RESEARCH METHODOLOGY

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

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

This research adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM). The primary framework provides a systematic and structured approach to data mining projects, encompassing various stages from understanding business objectives and data collection to model deployment and assessment [28]. Using CRISP-DM allows for a well-defined and transparent workflow, ensuring that each phase of the data mining process is meticulously addressed. The structured nature of CRISP-DM enhances the reproducibility and transparency of the research, contributing to the overall reliability of the findings [30]. In conclusion, the application of CRISP-DM in this study serves as a robust framework, guiding the research through a comprehensive and well-defined process that aligns with industry standards in data mining practices.

This research employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) as its methodological framework for sentiment analysis of guest reviews at The Trans Luxury Hotel. The primary framework facilitates a structured and systematic approach to data mining, ensuring a comprehensive exploration of sentiment patterns within the unique context of a luxury hotel setting. The study further utilizes the Support Vector Machine (SVM) algorithm to conduct sentiment analysis, leveraging SVM's proficiency in handling high-dimensional feature spaces and intricate patterns. Integrating CRISP-DM alongside SVM contributes to a rigorous and well-defined analytical process, enhancing the reliability and interpretability of sentiment analysis outcomes. In conclusion, the combination of CRISP-DM and the SVM algorithm in this research establishes a robust foundation for exploring sentiment dynamics in hotel guest reviews at The Trans Luxury Hotel, offering valuable insights for academic research and practical applications within the hospitality industry.

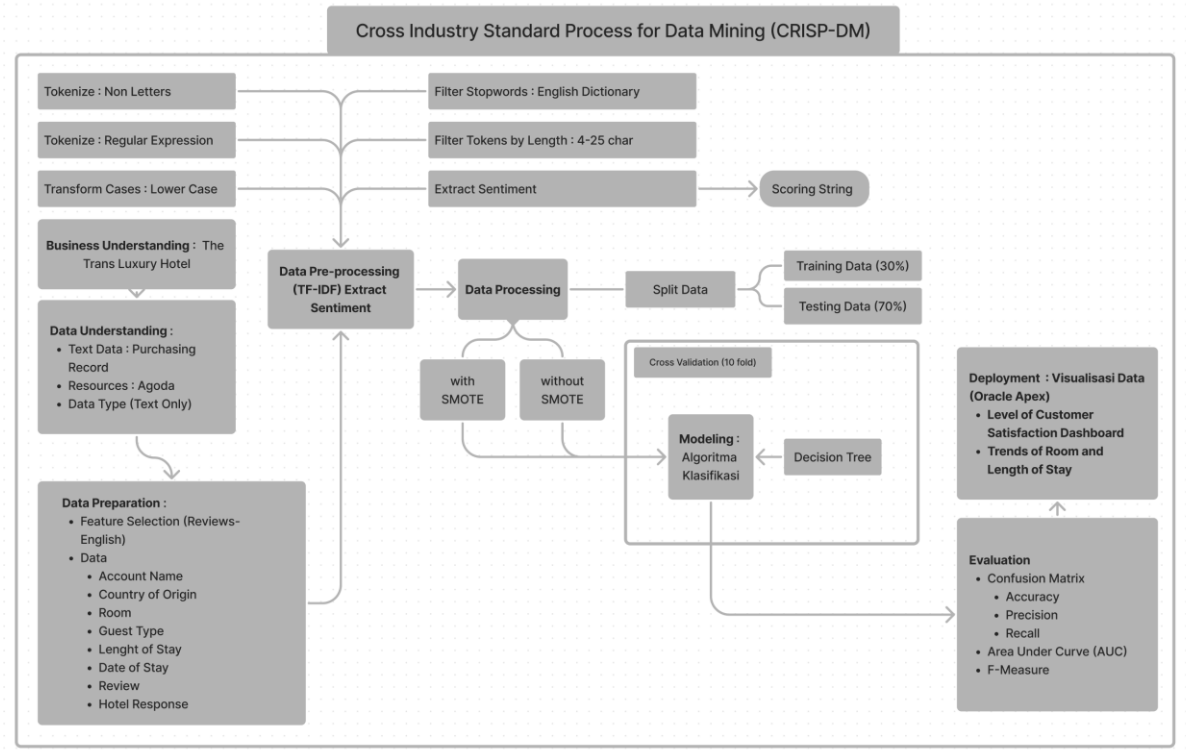

Figure 1. Cross-Industry Standard Process for Data Mining (CRISP-DM) Framework

Figure 1. is a data processing flow using the CRISP-DM framework. The preceding research endeavors employed the Cross-Industry Standard Process for Data Mining (CRISP-DM) as the underlying framework for sentiment analysis of hotel guest reviews. The primary framework provided a structured and comprehensive approach to navigating the complexities of data mining within the context of sentiment analysis [31]. By adhering to CRISP-DM, these studies ensured a systematic progression through critical stages, from data understanding and preprocessing to model evaluation and deployment. The adoption of CRISP-DM facilitated a methodical exploration of sentiment patterns in hotel guest reviews, contributing to the field's understanding of the diverse factors influencing guest satisfaction. In conclusion, integrating CRISP-DM in prior research has played a pivotal role in enhancing the rigor and effectiveness of sentiment analysis methodologies within the hospitality domain.

There are several essential stages in data processing: the business understanding stage, data understanding stage, data preparation stage, modeling stage, evaluation stage, and Deployment stage. This research uses Webharvy to collect data from Agoda applications: account name, country of origin, room type, guest type, length of stay, date of stay, Reviews, date of review, hotel response, and date of response. Next, the dataset is cleaned using the process documents operator from data in which there are operators such as tokenize, transform cases, filter tokens by length, and filter stopwords based on the following configuration: first, tokenize 1 (non-letters), tokenize 2 (regular expression [-!" #$%&'()*+,./:; <=>?@\[\\\]_'{|} ~]); second, transform cases (lower case); third, filter tokens by length (min char

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

= 4 &; max chars = 25); fourth, filter stopwords (dictionary English). In the next stage, the process of extracting sentiment data is carried out based on the review data that has been collected, as shown below.

Result of Scraping Purchasing Record Result of Extract Sentiment

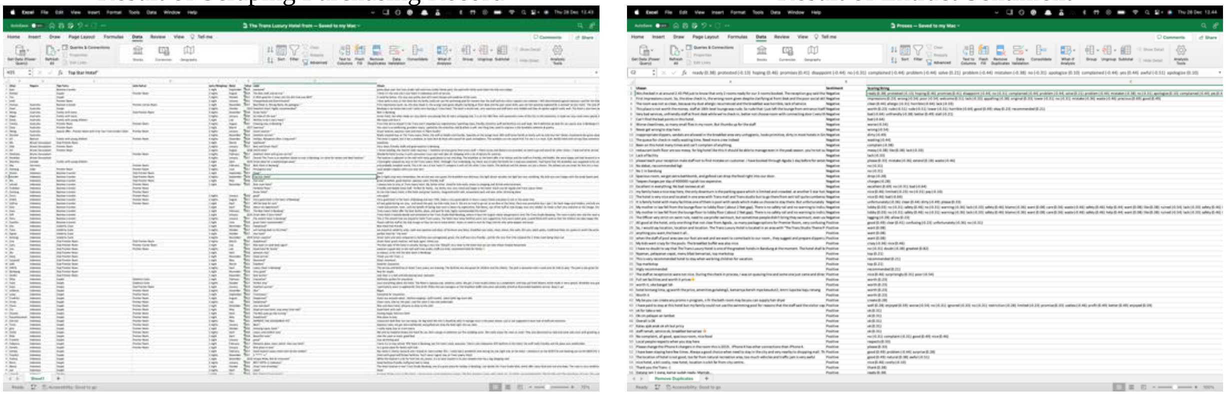

Figure 2. Result of Sraping Purchasing Record Data and Extract Sentiment Operator in Rapidminer

Figure 2 shows the result of scraping purchasing record data, including reviewed data text and extracting the sentiment operator. This research incorporates the Synthetic Minority Over-sampling Technique (SMOTE) operator and the Support Vector Machine (SVM) algorithm, with a data split of 30% for training and 70% for testing. Integrating the SMOTE operator addresses imbalances within the dataset by oversampling the minority class, promoting a more equitable representation of both classes during the training phase. Concurrently, the SVM algorithm is employed for its proficiency in handling high-dimensional feature spaces and intricate patterns, making it particularly suitable for sentiment analysis tasks. The utilization of a 30%-70% split for training and testing data ensures a robust evaluation of the model's generalizability and performance on unseen data. This strategic combination of techniques and parameters contributes to a comprehensive and effective methodology for sentiment analysis within the context of this research. In conclusion, integrating SMOTE, SVM, and the specified data split represents a thoughtfully designed approach to address data imbalance and reliably evaluate sentiment analysis performance, as shown in the figure below.

Extract Sentiment Process SVM Classification

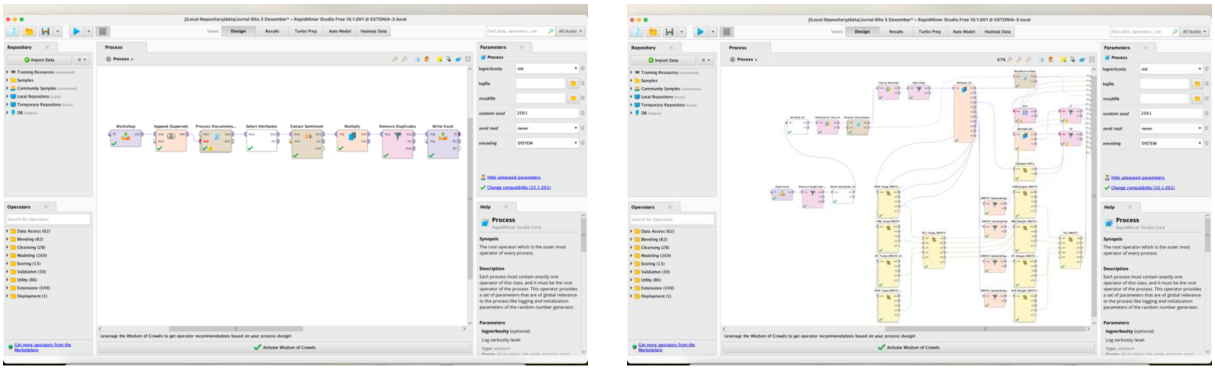

Figure 3. Process of Extract Sentiment and SVM Classification in Rapidminer

Figure 3 shows the process of extracting sentiment and implementing the SVM algorithm in sentiment classification using Rapidminer. The operator employed to address data imbalance in this study is the Synthetic Minority Over-sampling Technique (SMOTE) within the RapidMiner application. The primary purpose of incorporating SMOTE is to mitigate the challenges of imbalanced datasets by oversampling the minority class by generating synthetic instances. SMOTE is a valuable tool in addressing class disparity, ensuring that the learning algorithm is not unduly biased toward the majority class. The utilization of SMOTE within the RapidMiner framework reflects a strategic choice to enhance the model's performance in handling imbalanced data, thereby contributing to the robustness and reliability of the analysis outcomes. In conclusion, integrating the SMOTE operator in RapidMiner constitutes a methodological approach to alleviating data imbalance challenges, fostering more accurate and representative analyses, particularly in skewed class distribution scenarios.

2.2 Support Vector Machine (SVM)

The Support Vector Machine (SVM) process involves constructing an optimal hyperplane to maximize the margin between different classes within a high-dimensional feature space. SVM is particularly adept at handling complex relationships and discerning subtle patterns in data, making it a powerful tool for various machine-learning tasks, including sentiment analysis. The SVM process entails identifying support vectors, which are crucial data points that

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

influence the positioning of the hyperplane. The algorithm aims to minimize classification errors while maximizing class separation, providing a robust framework for effective pattern recognition. In terms of performance, SVM is known for its accuracy and efficiency in high-dimensional spaces, making it suitable for tasks with intricate patterns and complex relationships. However, the performance of SVM is contingent on appropriate parameter tuning and handling imbalanced datasets. While SVM offers a robust process for pattern recognition, its performance is context-dependent, requiring careful consideration of data characteristics and parameter settings for optimal results. The Support Vector Machine (SVM) algorithm can classify data using a hyperplane. The SVM concept focuses on risk minimization, which is the estimation of functions by minimizing the limits of generalization errors so that SVM overcomes overfitting. Meanwhile, the regression function of the SVM method is as follows.

$$f(x) = w^T \varphi(x) + b \quad (1)$$

Where w is the weighting vector, is a function that maps x into a dimension, and b is the refractive factor. Furthermore, SVM has advantages in high data generalization and produces good classification models even though trained with relatively little data.$\varphi(x)$. The supremacy of Support Vector Machine (SVM) in sentiment data classification lies in its inherent capacity to construct an optimal hyperplane within high-dimensional feature spaces, effectively separating different sentiment classes. SVM's ability to identify complex relationships and discern subtle patterns contributes to its superiority in sentiment analysis tasks. The algorithm accurately classifies sentiments by maximizing the margin between different classes, enhancing its robustness in handling diverse and nuanced expressions within textual data. SVM's versatility and effectiveness make it particularly well-suited for sentiment classification applications where precision and discernment of subtle sentiment variations are paramount. In conclusion, the strengths of SVM in sentiment data classification stem from its proficiency in navigating intricate feature spaces, providing a reliable and accurate means for discerning sentiment patterns within textual data.

3. RESULT AND DISCUSSION

3.1 Extract Sentiment of Hotel Guest Reviews

The extraction of sentiment from hotel guest reviews yields significant benefits for various stakeholders within the hospitality industry. The primary advantage lies in gaining valuable insights into customer experiences, preferences, and satisfaction levels. By discerning sentiments expressed in reviews, hotel management can identify areas of strength and weakness in their services, allowing for targeted improvements to enhance overall guest satisfaction. Additionally, the extraction of sentiment provides a basis for benchmarking performance, enabling hotels to compare their guest satisfaction levels with industry standards and competitors. This insightful analysis aids in strategic decision-making, helping hotels tailor their offerings to meet guest expectations and improve their market positioning. Moreover, the sentiment extracted from guest reviews is a valuable resource for online reputation management, influencing potential guests' perceptions and decisions. In conclusion, extracting sentiment from hotel guest reviews offers multifaceted benefits, from internal service enhancement to competitive positioning and effective reputation management.

Table 1. Extract Sentiment Operator

Classification	Review	Score String
Negative Sentiment	We checked in at around 2.45 PM just to know that only 2 rooms ready for our 3 rooms booked. The reception guy said that the third room will be available soonest at 3.30. We protested and said it is intolerable because we checking in in the noon and the check out time for this hotel is 12.00 PM, how come they cannot prepare the room? Finally, the reception said there is another room available which position not next to each other room and will be available at 3.15. We took that option. We were hoping that the reception can keep his promises to send the third room key to our room. But we have to disappoint. Until 3.35 pm there is no key sent to our room. My husband called the operator and complained. The operator doesn't understand the problem, and my husband asked for duty manager to call our room to solve this problem. The duty manager, if I'm not mistaken named Raisa called to our room and talked to my husband. She has no idea on what happened so, my husband has to explain the situation again. She said that the key will be sent immediately. She also said will come to our room to directly apologize to us. Finally, someone send the key at 3.50 PM. And guess what, she said that the hotel staffs was busy handling another guest. My husband complained to her and asked whether because of handling another guest they delaying to send our room key, and she said yes. What an awful answer from a 5-star hotel staff. The rooms and facilities are 5 stars hotel, but for us, the services are way below the 5-star hotels standard. We were in our room until 6.00 pm, and the hotel duty manager does not come to apologize. When we were back to our room after dinner, there is a slice of cake and letter for a token of apology, but it is too late. We felt not treated in a 5-star hotel standard.	ready (0.38) protested (-0.13) hoping (0.46) promises (0.41) disappoint (-0.44) no (-0.31) complained (-0.44) problem (-0.44) solve (0.21) problem (-0.44) mistaken (-0.38) no (-0.31) apologize (0.10) complained (-0.44) yes (0.44) awful (-0.51) apologize (0.10) apology (0.05)

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

Positive Sentiment	Totally love this hotel from the first time we arrived, friendly customer service. we booked for one night to celebrate my daughter and wife birthday, they assist in room setup which is a very nice setup with birthday cake and teddy bear for my daughter, totally appreciate it. room is spacious, bathroom is cool with bathtub , the amenities is great they even provide bubble bath and bubble bomb, the bed, the pillow which you can request for custom pillow even special pillow for pregnant women, club lounge is great. I recommend for first time traveler to choose club benefit room you pay a little bit more but you get better service, you can have afternoon tea time or evening cocktail, and during breakfast you can have ala carte menu. everything is great. good for family with children, they have swimming pool with some beach theme, that your kids will totally love it they can play with sands, even facilities such as gaming room for kids. and activities for the kids. totally recommended..	love (0.82) friendly (0.56) celebrate (0.69) nice (0.46) appreciate (0.44) cool (0.33) great (0.79) bomb (-0.56) special (0.44) great (0.79) recommend (0.38) benefit (0.51) pay (-0.10) better (0.49) great (0.79) good (0.49) love (0.82) play (0.36) recommended (0.21)

Table 1 shows how the extract sentiment operator works in Rapidminer. The Extract Sentiment operator proves instrumental in evaluating the sentiment of textual data, allowing for the computation of string scores that offer quantitative insights into the opinions expressed. Subsequently, employing visualization techniques such as TF-IDF enables the identification and presentation of the top 10 popular words within hotel guest reviews. This method provides a nuanced understanding of the vital focal points capturing guests' attention. The TF-IDF analysis highlights the significance of terms by considering both their frequency and importance in the corpus, offering a comprehensive perspective on the prevalent themes within the reviews. This approach facilitates a quantitative assessment of sentiment and qualitatively reveals the specific aspects of the hotel experience that guests find noteworthy. In conclusion, the integration of the Extract Sentiment operator and TF-IDF visualization contributes to a comprehensive analysis, allowing for both quantitative and qualitative insights into the concerns and priorities of hotel guests, as shown in the figure below.

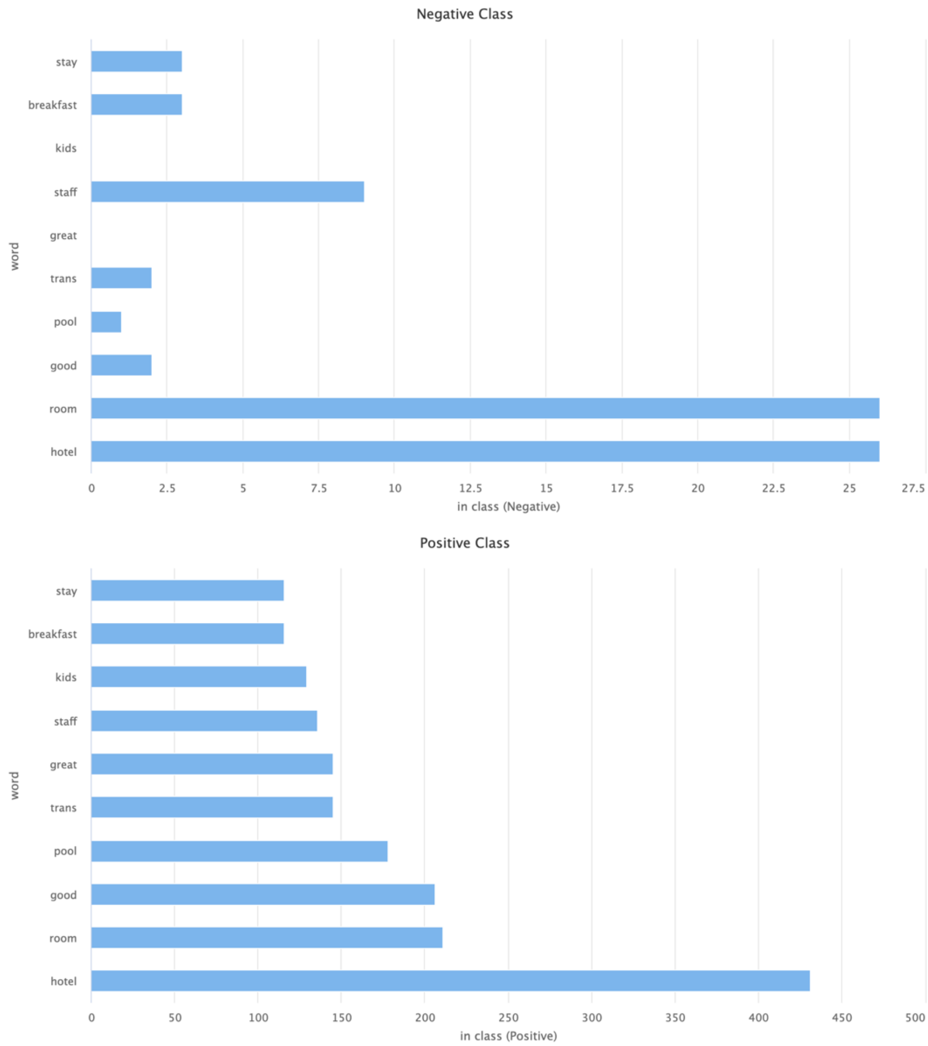

Figure 4. Top 10 Popular Words of Hotel Guest Review

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

Figure 4 shows the result of TF-IDF in counting the ten famous words of hotel guest reviews. Based on the analysis of the top 10 famous words, it is evident that hotel guests place significant emphasis on specific aspects of their experience. The results reveal that the most frequently mentioned terms include "breakfast," "staff," "pool," "room," and the hotel's "brand." This pattern suggests that guests are particularly attentive to fundamental elements such as the quality of breakfast offerings, interactions with staff, the room's condition and amenities, the pool's presence and state, and the overall reputation or identity of the hotel brand. These recurrent themes underscore the importance of these factors in shaping the overall satisfaction and perception of hotel guests. In conclusion, the analysis of the ten famous words provides valuable insights into the focal points of guest attention, offering guidance for hotel management to prioritize and optimize these aspects to enhance the overall guest experience and satisfaction.

3.2 Discussion: Support Vector Machine Performance in Text Classification

Based on the evaluation results of SVM performance without using SMOTE, it is evident that the model exhibits robust accuracy and precision in classifying positive and negative instances. The SVM model achieved an impressive accuracy of 95.82% ± 1.37%, indicating its effectiveness in correctly predicting both positive and negative classes. Furthermore, the 95.80% ± 1.35% precision underscores the model's ability to identify positive instances among the predicted positives accurately. Notably, the 100.00% ± 0.00% recall rate demonstrates the SVM model's exceptional capacity to capture all positive instances without missing any. These evaluation metrics collectively suggest that even without SMOTE, the SVM model showcases remarkable performance in sentiment analysis tasks, highlighting its potential for practical application in real-world scenarios.

Based on the performance evaluation of Support Vector Machine (SVM) utilizing Synthetic Minority Over-sampling Technique (SMOTE), it is evident that the classifier exhibits commendable accuracy and precision in predicting positive sentiment within the dataset. The SVM model achieved an impressive accuracy of 92.05% with a negligible variance of 2.40%, indicating robust performance in classifying hotel guest reviews. Moreover, the model demonstrated perfect precision of 100.00%, underscoring its capability to accurately identify positive sentiments. Although the recall rate of 84.08% suggests slight room for improvement in capturing all positive instances, the overall performance, as indicated by the F-measure of 91.28%, remains highly satisfactory. These findings underscore the efficacy of SVM with SMOTE in sentiment analysis tasks, showcasing its potential utility in enhancing decision-making processes within the hospitality domain.

Thus, the findings of this investigation revealed that employing a Support Vector Machine (SVM) without incorporating SMOTE operators yielded commendable performance metrics. Specifically, the accuracy achieved was 95.82%, indicating the model's proficiency in correctly classifying sentiments. The precision value of 95.80% reflects the model's ability to minimize false positives, showcasing its precision in identifying positive sentiments. Furthermore, the recall value reached 100%, emphasizing the model's capability to capture all instances of positive sentiments. Despite these favorable results, the Area Under Curve (AUC) value was 0.798 (79.8%), indicating the model's moderate discriminatory ability. In conclusion, while SVM without SMOTE demonstrated high accuracy, precision, and recall in sentiment classification, the AUC value suggests potential areas for further optimization to enhance the overall discriminatory power of the model, as shown in the table below.

Table 2. Confusion Matrix of SVM Performance with and without SMOTE

SVM Without SMOTE	SVM With SMOTE
PerformanceVector:	PerformanceVector:
accuracy: 95.82% +/- 1.37% (micro average: 95.82%)	accuracy: 92.05% +/- 2.40% (micro average: 92.05%)
ConfusionMatrix:	ConfusionMatrix:
True: Negative Positive	True: Negative Positive
Negative: 2 0	Negative: 365 58
Positive: 16 365	Positive: 0 307
AUC (optimistic): 0.798 +/- 0.256 (micro average: 0.798) (positive class: Positive)	AUC (optimistic): 0.999 +/- 0.001 (micro average: 0.999) (positive class: Positive)
AUC: 0.798 +/- 0.256 (micro average: 0.798) (positive class: Positive)	AUC: 0.999 +/- 0.001 (micro average: 0.999) (positive class: Positive)
AUC (pessimistic): 0.798 +/- 0.256 (micro average: 0.798) (positive class: Positive)	AUC (pessimistic): 0.999 +/- 0.001 (micro average: 0.999) (positive class: Positive)
precision: 95.80% +/- 1.35% (micro average: 95.80%) (positive class: Positive)	precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positive)
ConfusionMatrix:	ConfusionMatrix:
True: Negative Positive	True: Negative Positive
Negative: 2 0	Negative: 365 58
Positive: 16 365	Positive: 0 307
recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Positive)	recall: 84.08% +/- 4.93% (micro average: 84.11%) (positive class: Positive)
ConfusionMatrix:	ConfusionMatrix:
True: Negative Positive	True: Negative Positive
Negative: 2 0	Negative: 365 58

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

Positive:	16	365	Positive:	0	307
f_measure: 97.85% +/- 0.70% (micro average: 97.86%) (positive class: Positive)			f_measure: 91.28% +/- 2.94% (micro average: 91.37%) (positive class: Positive)		
ConfusionMatrix:			ConfusionMatrix:		
True:	Negative	Positive	True:	Negative	Positive
Negative:	2	0	Negative:	365	58
Positive:	16	365	Positive:	0	307

Table 2 shows the confusion matrix of the SVM algorithm with and without SMOTE. The result of the Support Vector Machine (SVM) with SMOTE operators in this study exhibited distinct performance characteristics. The accuracy was 92.05%, indicating a slightly lower overall correctness in sentiment classification compared to the non-SMOTE SVM model. However, noteworthy improvements were observed in precision, which reached 100%, showcasing the model's capacity to avoid false positives entirely. The recall value of 84.08% signifies the model's ability to correctly identify a substantial portion of positive sentiments. The Area Under Curve (AUC) value was 99.99 (99.9%), implying the model's outstanding discriminatory ability. In conclusion, the SVM model with SMOTE operators demonstrated robust performance, particularly in precision and discriminatory power, suggesting its efficacy in addressing imbalanced datasets for sentiment analysis.

Thus, it can be discerned that both the SVM algorithm without and with SMOTE demonstrate optimal performance in sentiment data classification. The SVM without SMOTE attains an accuracy level of 95.82%, precision of 95.80%, recall of 100%, and an Area Under the Curve (AUC) of 79.8%, indicating its proficiency in accurately classifying sentiments. Meanwhile, the utilization of SMOTE with the SVM algorithm yields an accuracy rate of 92.05%, precision of 100%, recall of 84.08%, and an AUC of 99.9%, showcasing the model's ability to address data imbalance and enhance precision and discriminative capability. Therefore, the choice between SVM with or without SMOTE can be tailored to the specific needs of sentiment classification tasks, considering the trade-off between accuracy, precision, and the ability to handle imbalanced data.

4. CONCLUSION

The findings of this research indicate that within the context of The Trans Luxury Hotel guest review dataset, the SVM algorithm exhibits the most favorable performance in sentiment classification. The SVM algorithm demonstrates accuracy levels reaching up to 90%, emphasizing its effectiveness in accurately categorizing sentiments expressed by hotel guests. Moreover, the integration of the SMOTE is employed to address the issue of imbalanced datasets, enhancing the model's ability to handle variations in class distribution. While SMOTE improves overall classification performance, the SVM algorithm is the primary driver of success in sentiment analysis for hotel guest reviews. In conclusion, the study highlights the potency of SVM in achieving robust sentiment classification outcomes within the specific domain of The Trans Luxury Hotel guest reviews. In addition, the confusion matrix result shows that the SVM without using SMOTE operators resulted in an accuracy value of 95.82%, a precision value of 95.80%, a recall value of 100%, and an Area Under Curve (AUC) value of 0.798 (79.8%). Otherwise, SVM performance using SMOTE operators produces an accuracy value of 92.05%, a precision value of 100%, a recall value of 84.08%, and an Area Under Curve (AUC) value of 99.99 (99.9%). Based on ten popular words, hotel guests are concerned about breakfast, staff, pool, room, and hotel. Thus, the guests' highlights are the menu served by the hotel, the service provided by employees, room conditions, and hotel brands. Therefore, hotel management needs to improve the quality of products and services to increase satisfaction and intention to stay again.

ACKNOWLEDGMENT

I thank the Tourism Department, Faculty of Business Administration and Communication, and the Atma Jaya Catholic University of Indonesia for their invaluable support in research and publication endeavors.

REFERENCES

[1] H. Chen, Y. Yu, Y. Jia, and L. Zhang, "Safe transductive support vector machine," *Conn. Sci.*, vol. 34, no. 1, pp. 942–959, 2022, doi: 10.1080/09540091.2021.2024511.
[2] S. Wu, X. Chen, C. Shi, J. Fu, Y. Yan, and S. Wang, "Ship detention prediction via feature selection scheme and support vector machine (SVM)," *Marit. Policy Manag.*, vol. 49, no. 1, pp. 140–153, 2022, doi: 10.1080/03088839.2021.1875141.
[3] A. A. Kayode, N. O. Akande, A. A. Adegun, and M. O. Adebiyi, "An automated mammogram classification system using modified support vector machine," *Med. Devices Evid. Res.*, vol. 12, no. 0, pp. 275–284, 2019, doi: 10.2147/MDER.S206973.
[4] J. Farjami, S. Dehyouri, and M. Mohamadi, "Evaluation of waste recycling of fruits based on Support Vector Machine (SVM)," *Cogent Environ. Sci.*, vol. 6, no. 1, pp. 1–120, 2020, doi: 10.1080/23311843.2020.1712146.
[5] T. Wang, G. Li, B. Wu, V. Æsøy, and H. Zhang, "Parameter identification of ship manoeuvring model under disturbance using support vector machine method," *Ships Offshore Struct.*, vol. 16, no. S1, pp. 13–21, 2021, doi: 10.1080/17445302.2021.1927600.
[6] J. A. Carter, A. R. Rivadulla, and E. Preatoni, "A support vector machine algorithm can successfully classify running ability

Building of Informatics, Technology and Science (BITS)
Volume 5, No 3, Desember 2023 Page: 627−635
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i3.4737

when trained with wearable sensor data from anatomical locations typical of consumer technology," *Sport. Biomech.*, vol. 00, no. 00, pp. 1–18, 2022, doi: 10.1080/14763141.2022.2027509.

[7] Y. Liu *et al.*, "Seismic vulnerability and risk assessment at the urban scale using support vector machine and GIScience technology: a case study of the Lixia District in Jinan City, China," *Geomatics, Nat. Hazards Risk*, vol. 14, no. 1, pp. 1–24, 2023, doi: 10.1080/19475705.2023.2173663.

[8] L. Wang, Z. C. Zhao, and Y. C. Weng, "Machine learning in predicting stock indexes: the role of online stock forum sentiment in MIDAS model," *Asia-Pacific J. Account. Econ.*, vol. 00, no. 00, pp. 1–20, 2023, doi: 10.1080/16081625.2023.2215234.

[9] A. Aakash and A. Gupta Aggarwal, "Assessment of Hotel Performance and Guest Satisfaction through eWOM: Big Data for Better Insights," *Int. J. Hosp. Tour. Adm.*, vol. 23, no. 2, pp. 317–346, 2022, doi: 10.1080/15256480.2020.1746218.

[10] H. N. T. Thu, "Measuring guest satisfaction from online reviews: Envidence in Vietnam," *Cogent Soc. Sci.*, vol. 6, no. 1, pp. 1–14, 2020, doi: 10.1080/23311886.2020.1801117.

[11] J. Baek, Y. Choe, and C. M. Ok, "Determinants of hotel guests' service experiences: an examination of differences between lifestyle and traditional hotels," *J. Hosp. Mark. Manag.*, vol. 29, no. 1, pp. 88–105, 2020, doi: 10.1080/19368623.2019.1580173.

[12] T. H. Le, C. Arcodia, M. A. Novais, and A. Kralj, "Proposing a systematic approach for integrating traditional research methods into machine learning in text analytics in tourism and hospitality," *Curr. Issues Tour.*, vol. 24, no. 12, pp. 1640–1655, 2021, doi: 10.1080/13683500.2020.1829568.

[13] J. Luo, S. Huang, and R. Wang, "A fine-grained sentiment analysis of online guest reviews of economy hotels in China," *J. Hosp. Mark. Manag.*, vol. 30, no. 1, pp. 71–95, 2021, doi: 10.1080/19368623.2020.1772163.

[14] P. Manolitzas, N. Glaveli, S. Palamas, M. Talias, and E. Grigoroudis, "Hotel guests' demanding level and importance of attribute satisfaction ratings: an application of MUltiplecriteria Satisfaction Analysis on TripAdvisor's hotel guests ratings," *Curr. Issues Tour.*, vol. 25, no. 8, pp. 1203–1208, 2022, doi: 10.1080/13683500.2021.1915253.

[15] M. Oh and S. Kim, "Role of Emotions in Fine Dining Restaurant Online Reviews: The Applications of Semantic Network Analysis and a Machine Learning Algorithm," *Int. J. Hosp. Tour. Adm.*, vol. 23, no. 5, pp. 875–903, 2021, doi: 10.1080/15256480.2021.1881938.

[16] R. X. Nie, J. H. Hu, H. Y. Zhang, J. Q. Wang, K. S. Chin, and X. Bao, "Classifying Quality Attributes of Hotel Services Considering Review Characteristics and Semantic Consistency: A Review-Driven IPA," *J. Qual. Assur. Hosp. Tour.*, vol. 00, no. 00, pp. 1–30, 2023, doi: 10.1080/1528008X.2023.2259610.

[17] F. Rezaei, I. Raeesi Vanani, A. Jafari, and S. Kakavand, "Identification of Influential Factors and Improvement of Hotel Online User-Generated Scores: A Prescriptive Analytics Approach," *J. Qual. Assur. Hosp. Tour.*, vol. 00, no. 00, pp. 1–40, 2022, doi: 10.1080/1528008X.2022.2146620.

[18] A. Aakash, A. Tandon, and A. Gupta Aggarwal, "How features embedded in eWOM predict hotel guest satisfaction: an application of artificial neural networks," *J. Hosp. Mark. Manag.*, vol. 30, no. 4, pp. 486–507, 2021, doi: 10.1080/19368623.2021.1835597.

[19] F. Xu, L. La, F. Zhen, T. Lobsang, and C. Huang, "A data-driven approach to guest experiences and satisfaction in sharing," *J. Travel Tour. Mark.*, vol. 36, no. 4, pp. 484–496, 2019, doi: 10.1080/10548408.2019.1570420.

[20] A. Lo, P. Yeung, and J. Cronin, "Will 'The spirit of discovery' lead Wharf Hotels to become a preferred international hotel brand?," *Asia Pacific J. Tour. Res.*, vol. 25, no. 10, pp. 1109–1127, 2020, doi: 10.1080/10941665.2020.1745856.

[21] C. Kaveski Peres and E. Pacheco Paladini, "Exploring the attributes of hotel service quality in Florianópolis-SC, Brazil: An analysis of tripAdvisor reviews," *Cogent Bus. Manag.*, vol. 8, no. 1, pp. 1–19, 2021, doi: 10.1080/23311975.2021.1926211.

[22] S. Moro, "Guest satisfaction in East and West: evidence from online reviews of the influence of cultural origin in two major gambling cities, Las Vegas and Macau," *Tour. Recreat. Res.*, vol. 45, no. 4, pp. 539–548, 2020, doi: 10.1080/02508281.2020.1759002.

[23] M. Gharzouli, A. K. Hamama, and Z. Khattabi, "Topic-based sentiment analysis of hotel reviews," *Curr. Issues Tour.*, vol. 25, no. 9, pp. 1368–1375, 2022, doi: 10.1080/13683500.2021.1940107.

[24] M. P. Mehta, G. Kumar, and M. Ramkumar, "Customer expectations in the hotel industry during the COVID-19 pandemic: a global perspective using sentiment analysis," *Tour. Recreat. Res.*, vol. 48, no. 1, pp. 110–127, 2023, doi: 10.1080/02508281.2021.1894692.

[25] N. C. Shereni and M. Chambwe, "Hospitality Big Data Analytics in Developing Countries," *J. Qual. Assur. Hosp. Tour.*, vol. 21, no. 3, pp. 361–369, 2020, doi: 10.1080/1528008X.2019.1672233.

[26] B. Kim, J. Kim, A. Singh, M. Erdem, and A. Hardin, "Factors Predicting Hotel Recommendations: A Comparison of Guest Feedback Before and After the Hotel Closures During the COVID-19 Pandemic," *Int. J. Hosp. Tour. Adm.*, vol. 00, no. 00, pp. 1–24, 2023, doi: 10.1080/15256480.2023.2213218.

[27] M. Cavique, R. Ribeiro, F. Batista, and A. Correia, "Examining Airbnb guest satisfaction tendencies: a text mining approach," *Curr. Issues Tour.*, vol. 25, no. 22, pp. 3607–3622, 2022, doi: 10.1080/13683500.2022.2115877.

[28] 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.

[29] 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.

[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.