Sentiment Classification of The Capsule Hotel Guest Reviews using Cross-Industry Standard Process for Data Mining (CRISP-DM)

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

Abstract—Technology advancements empower hotel accommodation service managers to undertake innovative initiatives to enhance guest appeal and ensure safety and comfort. One manifestation of such innovation is exemplified by The Capsule Hotel, which offers novel experiences to both domestic and international tourists. This research seeks to assess the sentiments of guests at The Capsule Malioboro, employing the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology and the Support Vector Machine (SVM) technique with Synthetic Minority Over-sampling Technique (SMOTE) operators. The findings demonstrate that when operated without SMOTE, the SVM algorithm yields a confusion matrix displaying an accuracy of 99.01%, precision of 99.00%, recall of 100%, AUC of 0.944, and an f-measure of 99.49%. With the integration of SMOTE, there is a notable enhancement across all metrics, with accuracy, precision, recall, AUC, and f-measure, all achieving perfect scores of 100%. In addition, an analysis of the top 10 frequently used words in guest reviews, such as "solo," "good," "place," "staff," "comfortable," "room," "clean," "hotel," "capsule," and "Malioboro," provides additional insights. Examining guest profiles within the dataset uncovers a strong inclination among Indonesian individuals to opt for The Capsule Malioboro's services, with solo travelers being the predominant guest type and most stays lasting only a single day. The capsule accommodations cater to various gender preferences, and an examination of overnight data indicates a rising trend, particularly in December 2022 and 2023. These insights enable the hotel to discern guest preferences, offering valuable information for enhancing service ratings and addressing specific needs.

Keywords: CRISP-DM; SMOTE; SVM; Capsule; Guest; Hotel

1. INTRODUCTION

The popularity of capsule hotels has ignited various responses among travelers in Indonesia [1]. The burgeoning trend of these compact and efficiently designed accommodations has captured the attention of a diverse range of individuals seeking unique and cost-effective lodging options [2]. Capsule hotels particularly appeal to those prioritizing functionality and affordability in their travel experiences [3]. Some view these establishments as innovative solutions to the challenges of urban living and limited space, providing travelers with a convenient and novel experience [4]. Nevertheless, capsule hotels in Indonesia are not uniform, as some argue that the confined nature of the accommodations may not suit everyone's preferences [5]. In conclusion, while capsule hotels have gained traction among Indonesian travelers, the varied responses underscore the subjective nature of this emerging trend in the hospitality industry [6].

The evolution of technology has significantly influenced innovative accommodations, providing attractive and secure services that enhance guests’ comfort [7]. Integrating advanced technological solutions into the hospitality sector is the primary driver behind this transformation [8]. For instance, intelligent room systems, biometric access control, and artificial intelligence-driven guest services have become prevalent, ensuring patrons a seamless and secure experience [9].

Implementing such technologies reflects a commitment to modernity and addresses the increasing expectations of contemporary travelers for personalized and technologically enabled services [10]. However, an overreliance on technological features may compromise the personal touch and human element that traditionally characterizes the hospitality industry [11]. In conclusion, while the technological advancements in accommodation services undeniably offer enhanced security and convenience, a balance must be struck to preserve the essential human-centric aspects contributing to a well-rounded guest experience.

The impact of innovative accommodation services on tourists’ travel interests is a subject of considerable academic inquiry [12]. The principal assertion revolves around the premise that cutting-edge services within the accommodation sector significantly influence individuals’ travel preferences and choices [13]. These services encompass a broad spectrum of offerings, from personalized concierge applications to sustainable and eco-friendly lodging options [14].

The deployment of such innovations has been observed to enhance the overall travel experience, captivating the interest of discerning travelers who seek unique and tailored services [15]. Some contend that excessive focus on innovative services may overshadow travel's cultural and authentic aspects [16]. In conclusion, the nexus between accommodation service innovations and tourist interests is intricate and multifaceted, warranting further investigation into the nuanced dynamics shaping contemporary travel behavior.

The tourism sector in Indonesia has undergone notable development, attributable to the introduction of innovative accommodation services, such as capsule hotels and other thematic lodging options [17]. The transformative impact these novel accommodations have had on the landscape of Indonesian tourism. The advent of capsule hotels, characterized by their efficient spatial design and cost-effective approach, has particularly
resonated with contemporary travelers seeking unique experiences [18]. The emergence of thematic hotels, which incorporate cultural or thematic elements into their design and services, has added a layer of diversity to the accommodation offerings [19].

However, it is crucial to recognize that opinions on these innovative accommodations may vary, as some argue that maintaining a balance between traditional hospitality and avant-garde concepts is essential [20]. In conclusion, the integration of inventive accommodation services has played a pivotal role in shaping the trajectory of tourism in Indonesia, offering a diverse array of options that cater to the evolving preferences of modern travelers.

This research analyzes guest sentiments in capsule hotels based on a case study conducted at The Capsule Malioboro in Yogyakarta. The primary objective is to delve into the multifaceted aspects of guest experiences within the unique setting of capsule accommodations. Through a comprehensive investigation of The Capsule Malioboro, this study seeks to unravel the underlying sentiments of guests, shedding light on factors that contribute to overall satisfaction or dissatisfaction [21].

The research uses CRISP-DM methodologies to discern patterns in guest sentiments, considering personal experience regarding service quality, spatial design, and cost-effectiveness [22]. Furthermore, the study aims to provide valuable insights into the factors influencing the growing popularity of capsule hotels as a distinct lodging choice [23]. In conclusion, the analysis of guest sentiments at The Capsule Malioboro is anticipated to contribute valuable knowledge to the burgeoning field of hospitality research, enhancing our understanding of the dynamics shaping guest preferences in the context of capsule accommodations.

Similar research employing the CRISP-DM methodology has been widely utilized in sentiment analysis of hotel guests using algorithms such as NBC, k-NN, DT, and SVM [24]–[27]. However, this study presents a distinctive advantage by concentrating on the prevalent phenomenon of the Capsule Hotel, providing insights into travelers’ interests regarding innovative services within the accommodation sector [28]. Currently in vogue, the primary focus on The Capsule Hotel allows for a nuanced examination of guest sentiments within this specific lodging trend [29]. By prioritizing this unique setting, the research aims to uncover intricate details of traveler preferences and attitudes toward service innovations, contributing to a more comprehensive understanding of the evolving landscape of accommodation preferences [30]. In conclusion, this study’s emphasis on The Capsule Hotel offers a specialized perspective that contributes valuable insights to the existing body of sentiment analysis research within the hospitality sector.

The limitations of this study are situated within the constraints of the dataset, methodology, and algorithm employed. The research follows the CRISP-DM framework and utilizes the Support Vector Machine (SVM) algorithm to analyze The Capsule Malioboro guest reviews collected from the Agoda platform [31]. While these choices were made to maintain consistency with existing sentiment analysis methodologies, it is crucial to acknowledge the potential bias introduced by the dataset’s reliance on a single platform [32]. The study’s generalizability may be affected by the specific characteristics of the Agoda user base. Additionally, while commonly employed in sentiment analysis, applying the SVM algorithm may not capture specific nuances present in guest sentiments that alternative algorithms could better address [33]. Despite these limitations, the research contributes valuable insights into guest sentiments within the context of The Capsule Malioboro, offering a foundation for further investigations and improvements in sentiment analysis methodologies within the hospitality sector.

The significance of this research lies in examining thematic accommodation services with the capsule concept, a field that has not been extensively studied. The primary impetus for undertaking this study is the relative paucity of academic inquiries into the unique thematic features of capsule accommodations, marking an opportunity to contribute valuable insights to the existing body of knowledge in the hospitality management [34]. Furthermore, adopting data-mining techniques in this research adds intrinsic value, particularly in the contemporary era of big data [33]. By leveraging data-mining methodologies, the study aims to extract meaningful patterns and trends from a large dataset of guest reviews, providing a comprehensive understanding of the factors influencing guest sentiments within the thematic context of the capsule accommodations [35]. In conclusion, this research not only fills a gap in the current literature but also pioneers the application of advanced data-mining techniques, thereby contributing to the evolving landscape of research in the hospitality sector.

A recommendation for further research involves integrating sentiment classification outcomes with consumer decision-making behavior using various decision support models. The principal proposal stems from the understanding that evaluating sentiment alone may not provide a holistic perspective on consumer preferences and choices. By incorporating sentiment analysis results into decision support models, such as Multi-Criteria Decision Analysis (MCDA) or Analytic Hierarchy Process (AHP), researchers can delve deeper into the intricacies of consumer decision-making processes influenced by sentiment [36]. This approach offers the potential to uncover nuanced relationships between sentiment expressions in reviews and the factors that significantly shape consumers’ choices. Integrating sentiment classification with decision support models extends the research’s analytical depth [37]. It contributes to a more comprehensive comprehension of consumer decision-making dynamics within the studied context. In conclusion, pursuing this avenue of research would augment the scholarly discourse on consumer behavior and sentiment analysis, providing a more nuanced understanding of the intricate interplay between sentiment expressions and decision-making processes.
2. RESEARCH METHODOLOGY

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

The CRISP-DM methodology comprises several stages essential for practical data mining and analysis. The primary stage is Business Understanding, where researchers establish a comprehensive understanding of the objectives and requirements of the analysis, aligning it with the overarching goals of the business or research endeavor. Following this, the Data Understanding stage involves exploring and assessing the available data to gain insights into its characteristics and quality. Subsequently, the Modeling stage focuses on applying various data mining techniques to construct a model that best represents the patterns in the data. The Evaluation stage critically assesses the performance and validity of the model against predefined criteria, ensuring its robustness and reliability. Finally, the Deployment stage involves the integration of the developed model into the operational environment, making it accessible for decision-making processes. This structured approach of CRISP-DM offers a systematic and iterative framework for conducting practical data mining projects, contributing to the overall success and reliability of the analytical process.

![Cross Industry Standard Process for Data Mining (CRISP-DM)](image)

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

Figure 1 shows the implementation of the CRISP-DM and the Support Vector Machine (SVM) algorithm in sentiment classification. Implementing the CRISP-DM methodology in conjunction with the Support Vector Machine (SVM) algorithm represents a robust approach to sentiment classification. The primary thrust of this methodology involves the structured and iterative progression through distinct stages, including Business Understanding, Data Understanding, Modeling, Evaluation, and Deployment. Within this framework, the SVM algorithm, a powerful machine learning technique, is deployed during the Modeling stage to discern patterns and relationships within the dataset. SVM's capacity to handle high-dimensional data and nonlinear relationships effectively enhances sentiment classification accuracy, contributing to a more nuanced understanding of user opinions and attitudes. This integrated approach ensures a systematic and comprehensive analysis. It showcases the synergistic potential of combining a well-established methodology like CRISP-DM with a sophisticated algorithm such as SVM for sentiment classification. In conclusion, this implementation holds promise for advancing the field of sentiment analysis by providing a structured and effective means to decipher nuanced sentiments within a given dataset.

2.2 SMOTE, SVM, and Confusion Matrix

Synthetic Minority Over-sampling Technique (SMOTE) operators are used to overcome data imbalance problems before they are processed to the SVM algorithm performance measurement stage. Data imbalance will occur when the number of objects in a data class is higher than in other classes, where data classes with more objects are called significant classes. In contrast, others are called minor [38]. Processing algorithms that do not consider data imbalance tend to emphasize major classes, not minor ones. Therefore, SMOTE techniques that use oversampling

Yerik Afrianto Singgalen, Copyright © 2024, MIB, Page 634
Submitted: 12/01/2024; Accepted: 30/01/2024; Published: 31/01/2024
methods to multiply random observations by increasing the amount of minor class data (artificial data) to be equivalent to significant classes are needed. Meanwhile, artificial data or synthesis is based on the k-nearest neighbor (k-nearest Neighbor). Artificial numerical data generators are measured for proximity to Euclidean distances, while categorical data based on minor classes whose variables are categorical scales are carried out with the Value Difference Metric (VDM) formula, as shown in the equation below:

$$\Delta(x, y) = w_x w_y \sum_{i=1}^{N} \delta(x, y)_i^r$$  \hspace{1cm} (1)

Equation (1) is the process of generating numerical data. Where $\Delta(x, y)$ is the distance between the intensity of x and y, while $w_x w_y$ is the weight of the warning (negligible). $N$ is a lot clearer, $r$ worth 1 (distance manhattan) or 2 (euclidean distance), also $\delta(x, y)_i^r$ as distance between categories. Meanwhile, the process of generating artificial data (synthesis) for numerical data is carried out by calculating the difference between the primary vector and its nearest k-neighbor, multiplying the difference by a randomized number between 0 and 1, and then adding the difference to the principal value of the original primary vector so that a new main vector is obtained. Furthermore, categorical data generation can be done through equation (2) as follows.

$$\delta(V_1 V_2) = \sum_{i=1}^{n} \left| \frac{C_{1i}}{C_1} - \frac{C_{2i}}{C_2} \right|^k$$  \hspace{1cm} (2)

Where $\delta(V_1 V_2)$ represents the distance between the values $V_1$ and $V_2$ while $C_{1i}$ is the number of $V_1$ which is included in class $i$, and $C_{2i}$ is the number of $V_2$ which is included in class $i$. Meanwhile, $i$ is the multiplicity of $C_1$ classes, as the multiplicity of values of 1, $C_2$ is the multiplicity of value 2, n represents the number of categories, and k is constant. The process of generating artificial data synthesis) for category data is carried out by selecting the majority between the primary vector under consideration and its nearest k-neighbor for face value; if the values are equal, they will be chosen randomly. The Support Vector Machine (SVM) algorithm can classify data using a hyperplane [40]. The SVM concept focuses on risk minimization, which is the estimation of functions by minimizing the limits of generalization errors so that SVM can overcome overfitting [41]. Meanwhile, the regression function of the SVM method is as follows.

$$f(x) = w^T \varphi(x) + b$$  \hspace{1cm} (3)

Where w is a weighting vector, $\varphi(x)$ is a function that maps x into a dimension, and b is a refractive factor. Furthermore, SVM has advantages in high data generalization and can produce good classification models even though it is trained with relatively little data. However, it isn’t easy to apply to datasets with large samples and dimensions [42]. This shows that SVM can perform well even with relatively small data. Furthermore, validation is used to determine the best type of model through the confusion matrix as information about actual classification results that can be predicted by a system through accuracy, precision, and recall values through the following equations.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$  \hspace{1cm} (4)

$$\text{Presisi/Specificity} = \frac{TP}{TP+FP}$$  \hspace{1cm} (5)

$$\text{Recall/Sensitivity} = \frac{TP}{TP+FN}$$  \hspace{1cm} (6)

$$f - \text{measure} = \frac{2x(\text{Presisi x recall})}{\text{presisi} + \text{recall}}$$  \hspace{1cm} (7)

The confusion matrix describes the data classification process's accuracy, precision, and recall. Accuracy is the accuracy of the system in performing the classification process correctly; Precision or Sensitivity is the ratio of the number of relevant documents to the total number of documents found in the classification system; recall or specificity is the ratio of the number of documents recovered by the classification system to the total number of relevant documents; F-measure is a popular evaluation metric for addressing imbalance class problems by combining recall and precision to produce an effective metric for retrieving information in an unbalanced set [43].

**3. RESULT AND DISCUSSION**

The exploration of Capsule Hotels is compelling from the sentiment analysis perspective, thus prompting this research to compute the accuracy, precision, recall, Area Under Curve (AUC), and F-measure metrics for the SVM algorithm when classifying guest review datasets. The primary rationale behind this investigation lies in the unique characteristics of Capsule Hotels that make them an intriguing subject for sentiment analysis. Through applying the SVM algorithm, known for its proficiency in handling intricate relationships within datasets, this research aims to quantitatively assess the effectiveness of sentiment classification within the context of guest reviews. Metrics such as accuracy, precision, recall, AUC, and F-measure will be utilized to evaluate the algorithm's performance comprehensively. By adopting a meticulous quantitative approach, this research contributes empirical evidence to
the discourse on sentiment analysis, specifically in the distinct setting of Capsule Hotels, providing valuable insights into guest sentiments and perceptions. In conclusion, this analytical framework holds promise for advancing our understanding of the intricate interplay between guest reviews and sentiment classification algorithms within hospitality research.

Table 1. Extract Sentiment Operator Result from Rapidminer

<table>
<thead>
<tr>
<th>Classification</th>
<th>Reviews</th>
<th>Score</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Water supply is often disrupted. Dirty toilet. Limited breakfast.</td>
<td>-0.717948717948718</td>
<td>dirty (-0.49) limited (-0.23)</td>
</tr>
<tr>
<td>Positive</td>
<td>This place is very pleasant, the location is very strategic, the service is good and ordering via Agoda is very profitable... Thank you</td>
<td>1.94871794871795</td>
<td>pleasant (0.59) good (0.49) profitable (0.49) thank (0.38)</td>
</tr>
</tbody>
</table>

Table 1 shows the result of extracting the sentiment operator from Rapidminer. In the data preparation process, the sentiment extraction operator is employed for scoring based on sentiment classification. The primary function of this operator is to extract and evaluate sentiment-related features within the dataset, facilitating the subsequent sentiment-scoring process. This method systematically identifies and categorizes sentiments expressed in textual data, allowing for a nuanced understanding of the sentiment distribution within the dataset. The sentiment extraction operator is pivotal in refining the dataset for sentiment analysis, laying the groundwork for accurate sentiment scoring and subsequent analysis. This meticulous approach ensures the effectiveness of sentiment classification and contributes to the overall reliability and validity of the findings in sentiment analysis research. In conclusion, using the sentiment extraction operator is imperative in preparing the dataset for sentiment scoring, fostering a rigorous and systematic approach to sentiment analysis within the given research context, as shown in the figure below.

Figure 2. Data Pre-processing and Extract Sentiment Operator

Figure 2 shows the data pre-processing and extract sentiment operator to get the score of each word in the reviews dataset. Data pre-processing is crucial for cleansing duplicate entries and eliminating non-meaningful symbols within the dataset. The primary objective of this stage is to enhance the overall quality and coherence of the data. The dataset becomes more refined and amenable to subsequent analysis by removing duplicates and irrelevant symbols. Furthermore, extracting sentiment becomes imperative to gauge the machine's capability in classifying review data into positive and negative classes. This step systematically identifies and categorizes sentiment-related features, contributing to a nuanced understanding of the dataset's sentiment distribution. In essence, combining data pre-processing and sentiment extraction is a fundamental preparatory phase, ensuring the reliability and accuracy of the subsequent sentiment analysis. Consequently, these pre-processing steps are pivotal in fostering a robust and meticulous approach to data analysis within the context of sentiment classification research.

Following the sentiment extraction process, the dataset undergoes configuration by assigning sentiment classification data as labels with a binomial type for subsequent processing in the modeling phase using the SMOTE operator and the SVM algorithm. The primary aim of this stage is to prepare the dataset for effective sentiment classification, where the sentiment labels are defined as positive or negative. Using the binomial type for sentiment labels ensures a precise and binary classification system, facilitating the subsequent modeling process. Incorporating the SMOTE operator is particularly advantageous for handling imbalanced datasets, as it synthesizes minority class instances, enhancing the robustness of the sentiment classification model. Coupled with the SVM algorithm, known for its efficacy in handling non-linear relationships and high-dimensional data, this
configuration ensures a comprehensive and accurate modeling process. In conclusion, the meticulous configuration of the dataset, utilizing SMOTE and SVM, is instrumental in preparing the data for subsequent sentiment analysis, thereby contributing to the overall reliability and effectiveness of the research outcomes, as shown in the figure below.

Figure 3 shows the dataset's total words, positive and negative classes. The most popular words that often appear in guest reviews are “solo” (94), “good” (92), “place” (106), “staff” (108), “comfortable” (110), “room” (120), “clean” (132), “hotel” (160), “capsule” (172), “Malioboro” (164). When visualized within the negative class, certain words stand out with meanings opposite to those typically associated with positivity, including “good” (2), “staff” (4), “comfortable” (2), “room” (2), “hotel” (4), “capsule” (8), and “Malioboro” (4). Despite these contrasting sentiments, it is imperative to emphasize that, overall, guest reviews of Capsule hotels convey a predominantly positive impression. The highlighted words underscore that, although some negative sentiments may be discerned, particularly regarding cleanliness, facilities, and comfort, the prevailing tone of the reviews remains affirmative. This nuanced analysis emphasizes the importance of considering both positive and negative aspects within sentiment visualization to gain a comprehensive understanding of guest perspectives on Capsule hotels, thereby contributing to a more insightful interpretation of the sentiment dynamics within this unique accommodation sector.

The stop words operator is applied using both the Indonesian and English languages to enhance accuracy in the sentiment extraction process. The primary objective of employing this operator is to refine the dataset by removing common and non-informative words that may not contribute significantly to sentiment analysis. By incorporating stop words in Indonesian and English, the operator aims to encompass a broader spectrum of linguistic nuances in the dataset, ensuring a more comprehensive identification of sentiment-related features. This meticulous approach acknowledges the multilingual nature of the dataset and seeks to optimize the accuracy of sentiment extraction by accounting for language-specific stop words. In conclusion, integrating stop word operators in both languages represents a strategic step in bolstering the accuracy of sentiment extraction, contributing to the reliability and effectiveness of subsequent sentiment analysis processes.
Upon completing the visualization of famous words, the testing phase involves evaluating the SVM algorithm by comparing the confusion matrix values with and without the Synthetic Minority Over-sampling Technique (SMOTE) operator. The primary focus is assessing the impact of SMOTE on the algorithm's performance in handling imbalanced datasets. The supplementary step of applying SMOTE addresses potential biases caused by imbalances between positive and negative sentiment classes. By comparing the confusion matrix outcomes with and without SMOTE, this analysis seeks to ascertain the effectiveness of the oversampling technique in enhancing the SVM algorithm's ability to classify sentiments accurately. Integrating SMOTE represents a strategic consideration to optimize the algorithm's performance in sentiment analysis, offering insights into the potential benefits of addressing class imbalances. In conclusion, this comparative evaluation serves as a critical component in gauging the efficacy of the SVM algorithm, providing valuable insights into the importance of SMOTE in handling imbalanced sentiment datasets.

<table>
<thead>
<tr>
<th>Table 2. Confusion Matrix of SVM (with and without SMOTE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Without SMOTE</strong></td>
</tr>
<tr>
<td>PerformanceVector:</td>
</tr>
<tr>
<td>Accuracy: 99.01% +/- 1.41% (micro average: 99.01%)</td>
</tr>
<tr>
<td>ConfusionMatrix:</td>
</tr>
<tr>
<td>True: Negative Positive</td>
</tr>
<tr>
<td>Negative: 8 0</td>
</tr>
<tr>
<td>Positive: 5 491</td>
</tr>
<tr>
<td>AUC (optimistic): 0.944 +/- 0.105 (micro average: 0.944)</td>
</tr>
<tr>
<td>(positive class: Positive)</td>
</tr>
<tr>
<td>AUC: 0.944 +/- 0.105 (micro average: 0.944)</td>
</tr>
<tr>
<td>(positive class: Positive)</td>
</tr>
<tr>
<td>AUC (pessimistic): 0.944 +/- 0.105 (micro average: 0.944)</td>
</tr>
<tr>
<td>(positive class: Positive)</td>
</tr>
<tr>
<td>precision: 99.00% +/- 1.41% (micro average: 98.99%)</td>
</tr>
<tr>
<td>(positive class: Positive)</td>
</tr>
<tr>
<td>ConfusionMatrix:</td>
</tr>
<tr>
<td>True: Negative Positive</td>
</tr>
<tr>
<td>Negative: 8 0</td>
</tr>
<tr>
<td>Positive: 5 491</td>
</tr>
<tr>
<td>recall: 100.00% +/- 0.00% (micro average: 100.00%)</td>
</tr>
<tr>
<td>(positive class: Positive)</td>
</tr>
<tr>
<td>ConfusionMatrix:</td>
</tr>
<tr>
<td>True: Negative Positive</td>
</tr>
<tr>
<td>Negative: 8 0</td>
</tr>
<tr>
<td>Positive: 5 491</td>
</tr>
<tr>
<td>f_measure: 99.49% +/- 0.72% (micro average: 99.49%)</td>
</tr>
<tr>
<td>(positive class: Positive)</td>
</tr>
<tr>
<td>ConfusionMatrix:</td>
</tr>
<tr>
<td>True: Negative Positive</td>
</tr>
<tr>
<td>Negative: 8 0</td>
</tr>
<tr>
<td>Positive: 5 491</td>
</tr>
</tbody>
</table>

Table 2 shows the difference in performance of the SVM algorithm while using SMOTE and without using SMOTE. Without using SMOTE, the performance evaluation of the sentiment classification model yields auspicious results. The model's overall accuracy is reported at 99.01% with a narrow variance of +/- 1.41%, demonstrating its robustness in accurately classifying sentiments. The confusion matrix further elucidates the model's efficacy, revealing no instances of false negatives and a minimal count of false positives. The Area Under the Curve (AUC) values indicate a consistently high level of performance, with a micro-average of 0.944, reaffirming the model's reliability. The precision, recall, and F-measure metrics consistently demonstrate the model's proficiency in correctly identifying positive sentiments, each boasting impressive scores above 98%. These comprehensive performance metrics collectively affirm the model's accuracy, precision, and reliability in sentiment classification, making it a robust tool for analyzing guest sentiments in the context of Capsule hotels. In conclusion, the meticulous evaluation of the model's performance underscores its effectiveness in capturing and classifying sentiments within the analyzed dataset.

However, after using SMOTE, the performance assessment of the sentiment classification model reveals exceptional results, attaining a flawless accuracy score of 100.00% with a negligible variance of +/- 0.00%. The confusion matrix underscores the model's impeccable precision in correctly classifying negative and positive sentiments, yielding no instances of false positives or false negatives. The Area Under the Curve (AUC) values
consistently register a perfect score 1.000, affirming the model's impeccable discriminatory power in distinguishing between sentiment classes. Precision, recall, and F-measure metrics further validate the model's flawless performance, each reflecting a micro-average of 100.00%. These outstanding metrics collectively establish the sentiment classification model as highly accurate and reliable, showcasing its proficiency in discerning sentiments within the context of Capsule hotels. In conclusion, the unparalleled performance metrics underscore the model's effectiveness in achieving perfect sentiment classification, providing robust insights into guest sentiments and opinions with commendable precision.

**Figure 4. Area Under Curve (AUC) of SVM Performance (with and without SMOTE)**

Figure 4 shows SVM's difference in AUC value while using SMOTE or without SMOTE. A notable discrepancy in the Area Under the Curve (AUC) values is observed between the SVM algorithm implemented with Synthetic Minority Over-sampling Technique (SMOTE) and the one without SMOTE, with a score of 1.000 and 0.944, respectively. The primary observation underscores the substantial impact of employing SMOTE on the discriminatory power of the SVM algorithm. The AUC value of 1.000 in the SMOTE-enhanced model suggests a perfect ability to distinguish between positive and negative sentiment classes. In contrast, the AUC value of 0.944 in the non-SMOTE model, while still indicative of good discriminatory performance, falls short of the perfection achieved with the SMOTE implementation. This discrepancy underscores the importance of addressing class imbalances in sentiment datasets, as demonstrated by the notable enhancement in model performance when employing SMOTE. In conclusion, the comparative AUC values emphasize the efficacy of SMOTE in augmenting the SVM algorithm's capacity to discern sentiments accurately, substantiating its relevance in sentiment analysis within the context of Capsule hotels.

Moreover, it is noteworthy that the f-measure values exhibit a marginal difference between the SVM algorithm without Synthetic Minority Over-sampling Technique (SMOTE) at 99.49% and the SVM algorithm utilizing SMOTE at 100%. This nuanced contrast in f-measure values signifies a slight improvement in precision and recall achieved with the SMOTE-enhanced model, highlighting the effectiveness of addressing class imbalances in the sentiment dataset. While both values indicate high accuracy and reliability in sentiment classification, the 100% f-measure with SMOTE underscores the optimization achieved in the model's ability to balance precision and recall, affirming the significance of SMOTE in enhancing the overall performance of the SVM algorithm within the specific context of sentiment analysis, particularly in the domain of Capsule hotels. In conclusion, the minute variation in f-measure values accentuates SMOTE's subtle yet meaningful impact on refining the model's precision-recall balance, offering valuable insights into its utility in handling imbalanced sentiment datasets.

Based on the values derived from the confusion matrix, it is evident that there exists a discrepancy in accuracy, precision, recall, Area Under the Curve (AUC), and f-measure between the SVM algorithm implementation with Synthetic Minority Over-sampling Technique (SMOTE) and without SMOTE. The primary observation indicates that the inclusion of SMOTE as an oversampling technique significantly impacts the performance metrics of the SVM algorithm. The differences in accuracy, precision, recall, AUC, and f-measure highlight the substantial influence of SMOTE in addressing imbalances within the sentiment dataset. This nuanced contrast emphasizes the importance of carefully considering and managing class imbalances, particularly in sentiment analysis scenarios. In conclusion, the varied metrics extracted from the confusion matrix underscore the noteworthy impact of SMOTE on enhancing the SVM algorithm's performance in accurately classifying sentiments, contributing to a more robust and balanced sentiment analysis within the domain of Capsule hotels.

Hence, this study recommends the SVM algorithm utilizing the Synthetic Minority Over-sampling Technique (SMOTE) operator as the ideal model with the best performance in classifying both negative and positive classes within the guest review dataset of Capsule Malioboro Hotel. The primary assertion is grounded in...
the superior outcomes observed in various performance metrics, including accuracy, precision, recall, Area Under the Curve (AUC), and f-measure, when SMOTE is integrated into the SVM algorithm. The supplementary step of employing SMOTE significantly enhances the algorithm's capability to handle imbalances in sentiment classes, resulting in a more comprehensive and accurate classification. Therefore, adopting SVM with SMOTE is recommended as a robust and effective approach for sentiment analysis within the unique context of Capsule hotels, offering valuable insights into guest sentiments and preferences with improved precision and reliability. In conclusion, the commendable performance metrics support the endorsement of SVM with SMOTE as the preferred model for sentiment classification in the Capsule Malioboro hotel guest reviews domain.

Based on the sentiment analysis results, it is evident that most guests at Capsule Malioboro express positive sentiments toward the accommodation services they received. However, analyzing this finding within the contextual framework of guest backgrounds and data related to their stay history at The Capsule Malioboro is essential. The primary observation underscores the overall favorable sentiments, indicating a positive guest experience. Nonetheless, the need for a contextual examination arises from the understanding that various factors, including individual preferences, expectations, or specific events during the stay, may influence sentiment expressions. By scrutinizing contextual elements and stay-related data, a more nuanced interpretation of sentiment dynamics can be achieved, contributing to a comprehensive understanding of the intricacies in guest sentiments toward the accommodation services provided by The Capsule Malioboro. In conclusion, while positive sentiments dominate, a contextual analysis is imperative to thoroughly comprehend the diverse factors influencing guest sentiments, as shown in the figure below.

![Figure 5](image-url)
shedding light on the nuanced preferences of guests towards the capsule hotel concept and facilitating informed decisions for those within the hospitality sector.

4. CONCLUSION

The outcomes of this study reveal that within the context of The Capsule Malioboro, the predominant visitor profile is characterized by solo travelers, primarily originating from Indonesia, with a dominant length of stay being one day. Additionally, the analysis indicates that the highest number of guests occurs in December, escalating from 2022 to 2023. Furthermore, sentiment analysis results demonstrate that the SVM algorithm, coupled with the SMOTE, exhibits optimal performance, achieving 100% accuracy and an AUC value of 100%. These detailed findings contribute to a comprehensive understanding of the guest demographic and temporal patterns at The Capsule Malioboro while highlighting the efficacy of the sentiment analysis methodology in capturing and classifying sentiments with remarkable precision. In conclusion, the combination of visitor profiles, temporal patterns, and sentiment analysis outcomes provides a holistic perspective that informs hotel strategic decisions and contributes to the broader understanding of guest preferences in the context of capsule hotels.

ACKNOWLEDGMENT

I sincerely thank Lembaga Penelitian dan Pengabdian Masyarakatat (Institute for Research and Community Service), Faculty of Business Administration and Communication, Tourism Department, Atma Jaya Catholic University of Indonesia. The institution's unwavering support, guidance, and resources have been instrumental in completing my research endeavors. The collaborative and nurturing environment fostered by the faculty and staff has significantly contributed to the depth and quality of my academic pursuits. I extend my heartfelt thanks to all institution members for their dedication to fostering a conducive research atmosphere. This acknowledgment is a testament to the invaluable role played by Lembaga Penelitian dan Pengabdian Masyarakat in shaping my academic journey and enhancing my scholarly pursuits within the field of tourism.

REFERENCES

