Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

Sentiment Analysis of Multilingual Customer Reviews in the Hospitality Sector

Titis Puspitarini

Faculty of Business Administration and Communication, the School of Tourism, Atma Jaya Catholic University of Indonesia, Jakarta
Jl. Jend. Sudirman No.51, Karet Semanggi, Setiabudi District, South Jakarta City, Special Capital Region of Jakarta, Indonesia
Email: titis.puspitarini@atmajaya.ac.id
Correspondence Author Email: titis.puspitarini@atmajaya.ac.id
Submitted: **10/01/2025;** Accepted: **31/01/2025;** Published: **31/01/2025**

Abstract−This study investigates the role of sentiment analysis in improving service quality and guest satisfaction in the hospitality industry. Recognizing the increasing importance of customer feedback in shaping operational strategies, the research utilized a dataset of 1,141 original hotel reviews, comprising 1,113 positive and 28 negative sentiments. The methodology employed IndoBERT for sentiment classification, supported by a series of preprocessing steps, including text normalization, stopword removal, and text length filtering, to ensure data integrity. To address the significant class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied, resulting in a balanced dataset of 1,923 reviews, with 1,450 positive and 473 negative sentiments. The model demonstrated strong performance, achieving an Area Under the Curve (AUC) score of 94.7%, highlighting its capability to classify sentiments accurately. Findings reveal that positive sentiments often focus on room quality, staff friendliness, and breakfast service, while negative feedback highlights service delays and cleanliness issues. These insights enable data-driven recommendations for improving guest experiences and addressing critical concerns. The study demonstrates the potential of sentiment analysis as a strategic tool for enhancing service delivery, fostering guest loyalty, and maintaining competitiveness in the hospitality market.

Keywords: Sentiment Analysis; Hospitality Industry; Service Quality; Guest Satisfaction; Data-Driven Approaches

1. INTRODUCTION

The hospitality industry's rapid globalization has intensified the need for an in-depth understanding of customer feedback across diverse linguistic and cultural contexts. Sentiment analysis, a subfield of natural language processing, offers a systematic approach to extracting emotional insights from textual data, making it a valuable tool for evaluating multilingual customer reviews [1], [2]. Customer sentiment is a critical indicator of service quality and customer satisfaction, influencing the sector's decision-making processes and strategic improvements [3], [4]. Despite advancements in sentiment analysis techniques, challenges remain in accurately interpreting nuanced expressions, idiomatic phrases, and cultural variations in multilingual texts [5], [6]. This complexity underscores the importance of adopting advanced models capable of discerning semantic and syntactic patterns to ensure reliable interpretations. A robust analysis of such data enhances the understanding of customer experiences and informs tailored service strategies, driving competitive advantage and fostering customer loyalty in an increasingly interconnected marketplace.

Understanding the sentiments embedded in customer feedback has become increasingly vital in globalized service industries, particularly in the hospitality sector. The dynamic nature of customer preferences and the proliferation of multilingual reviews on digital platforms necessitate a structured approach to analyzing this data [7], [8]. Leveraging sentiment analysis addresses the complexity of linguistic diversity and provides actionable insights that align services with evolving customer expectations [9]. Without systematic exploration, valuable data reflecting consumer satisfaction and areas requiring improvement risk being overlooked, potentially impacting competitive positioning and service quality. This research highlights a strategic pathway for enhancing operational efficiency and cultivating long-term loyalty within a highly competitive marketplace by emphasizing the need for precise and scalable analytical methodologies.

This research systematically analyzes customer sentiments expressed in multilingual reviews within the hospitality industry, addressing the complexities of diverse linguistic and cultural expressions. By focusing on sentiment analysis, the study seeks to uncover patterns and insights that reflect customer satisfaction, preferences, and areas for service enhancement [10], [11]. Such an approach holds the potential to bridge the gap between unstructured textual data and actionable strategies, ensuring a deeper alignment between service offerings and consumer expectations [12], [13]. This research contributes to developing data-driven decision-making frameworks by applying advanced analytical methodologies, empowering the hospitality sector to refine customer experiences, and strengthening market competitiveness. Ultimately, the outcomes are anticipated to offer valuable guidance for stakeholders in optimizing service quality and fostering sustainable growth.

This study offers a dual contribution by advancing theoretical frameworks and providing actionable insights within the hospitality industry's sentiment analysis domain. Theoretically, it enhances understanding of how multilingual sentiment data can be processed and interpreted to capture nuanced customer feedback, addressing gaps in cross-linguistic sentiment modeling [14], [15]. The study also introduces analytical perspectives integrating linguistic diversity and contextual variations, fostering a more holistic approach to evaluating textual data.

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

Practically, the findings are anticipated to inform industry stakeholders in developing data-driven strategies that optimize customer satisfaction and service delivery [16], [17]. By bridging the divide between theory and application, this research equips decision-makers with refined tools to navigate a globalized service environment's complexities while contributing to the academic discourse on natural language processing and consumer behavior analysis. The outcomes are expected to drive scholarly advancement and tangible improvements in customer experience management.

Existing studies in sentiment analysis have extensively explored the application of computational techniques to extract insights from customer reviews, particularly in monolingual contexts or with limited linguistic diversity. These investigations have demonstrated the potential of machine learning and natural language processing methods in identifying customer sentiments and predicting satisfaction levels [18]–[20]. However, a significant research gap persists in addressing the intricacies of multilingual sentiment analysis, especially within sectors where customer feedback spans diverse linguistic and cultural frameworks, such as the hospitality industry [21], [22]. This limitation underscores the need for methodologies integrating cross-linguistic variations and contextual nuances to provide a more accurate and comprehensive understanding of customer experiences. Addressing this gap is critical for advancing both the theoretical knowledge and practical applications of sentiment analysis, paving the way for more inclusive and globally relevant insights in customer-oriented industries.

The uniqueness of this study lies in its integration of multilingual sentiment analysis within the hospitality sector, an area that has been insufficiently addressed in existing literature. The research employs advanced computational techniques to investigate customer reviews spanning diverse languages and cultural contexts, uncovering nuanced insights often overlooked in monolingual or homogeneous analyses [23]. This approach represents a significant departure from conventional studies, which predominantly focus on singular linguistic frameworks, limiting their applicability in globalized industries [24], [25]. The innovative methodology emphasizes the interplay between linguistic diversity and sentiment detection, enabling a deeper understanding of customer feedback dynamics. By addressing these complexities, the study advances methodological innovation. It provides actionable insights tailored to the diverse and interconnected nature of the hospitality industry, positioning it as a critical contribution to academia and industry practice.

Future investigations are encouraged to explore integrating advanced sentiment analysis models with domain-specific data from various industries to enhance the findings' contextual accuracy and applicability. Expanding the scope to include underrepresented languages and dialects in multilingual datasets would provide a more comprehensive understanding of linguistic diversity in customer feedback. Additionally, incorporating real-time sentiment analysis methodologies may offer valuable insights into dynamic consumer behavior, enabling more responsive service strategies. The adoption of interdisciplinary approaches, such as combining sentiment analysis with behavioral economics or cultural studies, can reveal more profound insights into the motivations and expectations of diverse customer groups. Such advancements would refine analytical techniques and contribute to developing more inclusive and adaptive frameworks for understanding global customer experiences. These directions are expected to pave the way for more nuanced and impactful applications of sentiment analysis in addressing complex market demands.

2. RESEARCH METHODOLOGY

2.1 Related Works

The body of literature on sentiment analysis has significantly advanced in recent years, with numerous studies examining its application in various domains, including e-commerce, healthcare, and hospitality. These works have demonstrated the efficacy of machine learning and natural language processing techniques in extracting meaningful insights from textual data, often emphasizing monolingual datasets. While several studies have addressed multilingual sentiment analysis, the focus has frequently been limited to well-resourced languages, leaving a gap in understanding linguistic diversity and its implications. Moreover, the integration of contextual nuances, such as cultural and regional variations, remains underexplored despite its critical role in interpreting customer sentiments accurately. Addressing these limitations is essential for extending the scope and impact of sentiment analysis research. By building on existing methodologies and addressing their shortcomings, future efforts can contribute to more robust and inclusive frameworks capable of handling the complexities of real-world multilingual data. Such advancements hold the potential to broaden the applicability of sentiment analysis across diverse contexts and industries.

Several studies have contributed to developing sentiment analysis, mainly textual data derived from customer feedback and social media platforms. Early investigations predominantly utilized lexicon-based approaches, which relied on predefined dictionaries of sentiment terms to classify sentiments as positive, negative, or neutral. Subsequent advancements introduced machine learning algorithms, including support vector machines and decision trees, demonstrating improved accuracy in capturing sentiment patterns from large datasets. More recent works have emphasized deep learning models, such as convolutional neural networks and recurrent neural networks, which excel in understanding complex linguistic structures and contextual meanings. Despite these advancements, most prior studies have concentrated on monolingual datasets or limited multilingual scenarios,

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

often neglecting the nuanced interplay of cultural and linguistic diversity in customer feedback. Addressing this gap is critical for enhancing the precision and applicability of sentiment analysis, especially in sectors where multilingual data plays a pivotal role in decision-making. These developments provide a strong foundation for expanding analytical methodologies to encompass broader linguistic and contextual dimensions.

Understanding and responding to customer feedback has become a cornerstone for maintaining competitive advantage and enhancing service quality in hotel management. The hospitality industry operates in a customer-centric environment where satisfaction and loyalty are deeply influenced by the ability to effectively address guests' needs and expectations. Analyzing feedback through advanced methodologies such as sentiment analysis provides valuable insights into customer perceptions, preferences, and areas requiring improvement. This approach enables hotel management to tailor services, optimize resource allocation, and refine marketing strategies, fostering a more personalized guest experience. By leveraging data-driven insights, hotels can anticipate emerging trends and adapt proactively, ensuring alignment with dynamic market demands. Such practices improve operational efficiency and build long-term trust and customer retention, which are essential for sustained success in a highly competitive industry.

The significance of this study lies in its potential to address critical challenges in understanding and utilizing multilingual customer feedback within the hospitality sector. As global competition intensifies, businesses must analyze diverse linguistic inputs effectively to capture nuanced sentiments influencing customer satisfaction and loyalty. This study contributes by offering a structured approach to processing and interpreting complex textual data, addressing limitations in existing sentiment analysis methodologies that often overlook linguistic and cultural diversity. This research provides a framework that aligns service improvements with evolving customer expectations by integrating advanced analytical techniques. Such insights are pivotal for enhancing operational strategies, fostering personalized customer interactions, and strengthening competitive positioning. The outcomes of this study hold broader implications for the hospitality industry, where understanding diverse customer perspectives is essential for sustained success and innovation in an increasingly interconnected market.

2.2 Research Workflow

The research workflow for this study is systematically designed to ensure rigor and coherence across all phases, beginning with meticulous planning. The initial stage involves identifying the research problem, conducting a comprehensive literature review, formulating research questions and objectives, and designing a robust methodology tailored to the study's aims. Data collection follows, encompassing the planning and acquisition of raw data, accompanied by preprocessing to prepare it for subsequent analysis. During the analysis phase, sentiment analysis is conducted alongside statistical evaluations and data visualizations to derive meaningful insights. Evaluation of the findings involves interpreting the results, validating the conclusions, and examining their broader implications for the field. The final stage emphasizes documentation, including drafting the research paper, undergoing the peer review process, and ensuring effective publication and dissemination of the findings. This structured approach not only enhances the reliability and validity of the research but also underscores its contribution to academic and practical applications.

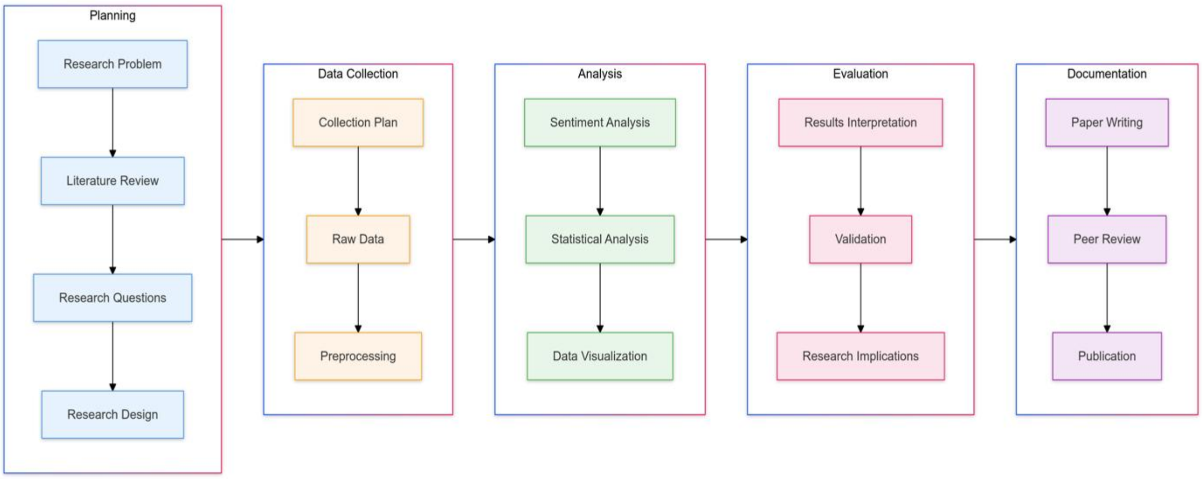

Figure 1. Research Workflow

Figure 1 illustrates the research framework, outlining a structured and sequential study approach. The process begins with the planning phase, which includes identifying the research problem, conducting a literature review, formulating research questions, and designing an appropriate methodology. This is followed by the data collection phase, which emphasizes creating a detailed collection plan, gathering raw data, and preprocessing it to ensure quality and consistency for subsequent analysis. The analysis phase encompasses sentiment analysis, statistical evaluation, and data visualization to derive meaningful insights. The evaluation phase involves

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

interpreting the results, validating the findings, and exploring their broader implications within the research context. Finally, the documentation phase focuses on disseminating the research through paper writing, peer review, and publication. This framework underscores a systematic progression through each stage, ensuring methodological rigor and the effective translation of research outcomes into academic and practical contributions.

The methods employed in this study demonstrate a strong alignment with its objectives, addressing the complexities inherent in multilingual sentiment analysis within the hospitality sector. Integrating advanced sentiment analysis techniques, statistical evaluations, and data visualization ensures a comprehensive approach to capturing and interpreting diverse customer feedback. These methods are particularly relevant given the linguistic and cultural nuances that characterize customer reviews, requiring robust tools capable of extracting meaningful insights from unstructured textual data. The combination of preprocessing techniques and analytical frameworks enhances the precision of sentiment classification and facilitates the identification of trends and patterns across diverse datasets. This methodological alignment ensures that the study effectively bridges theoretical insights and practical applications, offering valuable contributions to understanding and improving customer satisfaction within the global hospitality industry.

2.2 Dataset

The dataset utilized in this study consists of customer reviews collected from online platforms, encompassing 12 key attributes relevant to sentiment analysis in the hospitality sector. These columns include "Nama," "Negara," "Tipe Tamu," "Tipe Kamar," "Lama Menginap," "Bulan Menginap," "Tahun Menginap," "Rating," "Deskripsi Rating," "Judul," "Ulasan," and "Tanggal Mengulas." Each attribute provides a comprehensive understanding of customer feedback, ranging from demographic details to qualitative and quantitative assessments of their experiences. This dataset reflects diverse customer perspectives, capturing linguistic and contextual variations across multiple entries. Including detailed qualitative reviews alongside structured quantitative ratings offers a robust foundation for performing sentiment analysis, ensuring the extraction of numerical insights and nuanced textual data. Such a dataset is highly appropriate for this research, as it enables a multi-dimensional evaluation of customer sentiments, contributing to the reliability and depth of the analytical outcomes.

The composition of the dataset undergoes significant transformation following the implementation of preprocessing and Synthetic Minority Oversampling Technique (SMOTE). Initially, the original dataset comprised 1,141 rows, with a highly imbalanced distribution of 1,113 positive and only 28 negative reviews. This imbalance limits the model's ability to generalize effectively, as it disproportionately favors the majority class. During preprocessing, steps such as text normalization, stopword removal, and text length filtering ensure the dataset's quality and consistency, preparing it for more robust analysis. Subsequently, SMOTE is applied to address the imbalance by synthetically generating additional entries for the minority class, resulting in a balanced dataset comprising 1,923 rows—1,450 positive reviews and 473 negative reviews. This adjustment ensures equitable representation of sentiment classes, enhancing the model's ability to learn patterns across both classes effectively. The transformation underscores the importance of preprocessing and balancing techniques in ensuring the dataset's suitability for accurate and unbiased sentiment classification.

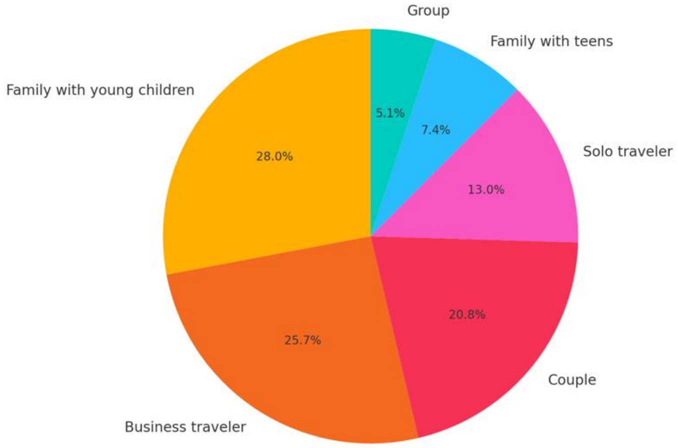

Figure 2. Percentage Distribution of Guest Type

Figure 2 illustrates the percentage distribution of guest types, revealing significant insights into the demographic composition of the hotel's visitors. "Family with young children" represents the most significant proportion, accounting for 28.0% of the total guests, followed closely by "Business travelers" at 25.7% and "Couples" at 20.8%. "Solo travelers" constitute 13.0% of the guests, while "Families with teens" represent 7.4%, and "Group" travelers make up the smallest segment at 5.1%. This distribution highlights the hotel's primary

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

appeal to family-oriented and professional guests, reflecting the suitability of its facilities and services for these groups. The presence of smaller categories, such as "Group" and "Families with Teens," suggests potential opportunities for targeted marketing or service enhancement to attract a broader audience. These insights are crucial for aligning the hotel's offerings with its customer base's diverse needs and preferences, ensuring both satisfaction and competitive positioning.

Guest preferences for room types are critical in understanding customer inclinations and optimizing hospitality services. Variations in room selection often reflect diverse needs, including budget constraints, group size, and specific expectations related to comfort and amenities. Superior categories, such as "Superior Twin Bed" and "Superior King Bed," which dominate booking percentages, suggest a preference for versatile and cost-effective accommodations. While less frequently chosen, Deluxe options highlight the segment of guests willing to invest in premium features for enhanced experiences. Analyzing these trends provides valuable insights into the demand dynamics of different room categories, enabling targeted adjustments in marketing, pricing, and resource allocation. Recognizing these patterns ensures a customer-centric approach to service delivery, fostering greater satisfaction and loyalty while aligning operational strategies with market demands.

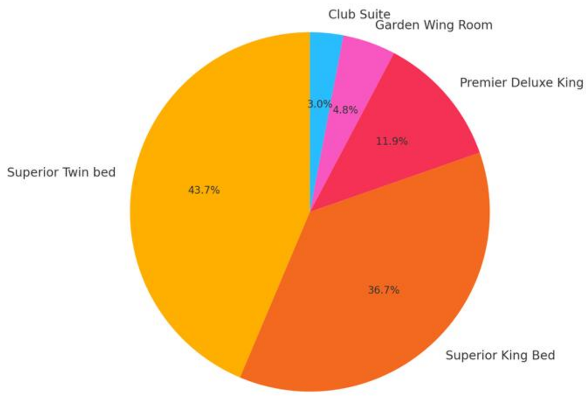

Figure 3. Top Five Percentage Distribution of Room Types

Figure 3 highlights the top five percentage distributions of room types, shedding light on customer preferences within the hotel. The "Superior Twin Bed" emerges as the most favored room type, accounting for 43.7% of bookings, followed closely by "Superior King Bed" at 36.7%. The "Premier Deluxe King" comprises 11.9% of the total bookings, while the "Garden Wing Room" and "Club Suite" contribute 4.8% and 3.0%, respectively. These figures reflect a clear preference for standard room categories, which likely meet the needs of budget-conscious guests and those seeking practical accommodations. The lower percentages associated with more premium room types suggest their appeal to niche segments or specific guest profiles. Understanding these preferences is critical for optimizing room allocation, tailoring promotional strategies, and effectively aligning service offerings to enhance guest satisfaction and occupancy rates. This distribution is a vital tool for aligning operational strategies with customer demand.

Room type preferences are driven by a combination of factors related to individual guest needs, purpose of travel, and perceived value for money. For instance, family travelers may prioritize larger rooms or suites that offer additional space and amenities suitable for children. In contrast, business travelers might prefer rooms with work-friendly features such as desks, high-speed internet, and proximity to meeting facilities. Budget constraints often play a pivotal role, as guests tend to weigh the cost of a room against the level of comfort, amenities, and services provided. Location within the hotel also influences preferences, with some guests favoring rooms closer to elevators, dining areas, or recreational facilities. Furthermore, cultural factors and personal lifestyle choices, such as a preference for modern versus traditional aesthetics or interest in luxury experiences, significantly shape decision-making. Seasonal and promotional factors, including discounts or value-added services, may also sway guest preferences toward specific room types. Understanding these diverse motivations allows hotels to tailor their offerings and marketing strategies to align with customer expectations and maximize satisfaction.

3. RESULT AND DISCUSSION

3.1 Model Performance Evaluation of IndoBERT-base-p1 from Hugging Face

Text cleaning is a crucial preprocessing step in natural language processing to standardize and optimize raw textual data for analysis. Transforming all characters into lowercase, such as converting "Hotel" to "hotel," ensures uniformity and eliminates discrepancies caused by case sensitivity during model interpretation. Removing website links and email addresses eliminates extraneous information that does not contribute to the semantic understanding

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

of the text. Deleting irrelevant numerical data prevents noise in the dataset, while removing symbols such as @, #, $, and % ensures that the input is free from distractions that may distort the analysis. These steps collectively enhance the clarity and quality of the data, enabling more accurate and reliable sentiment analysis or other linguistic tasks. Text cleaning is essential for extracting meaningful insights from unstructured textual information by systematically addressing inconsistencies and irrelevant elements.

The removal of stopwords represents a pivotal step in text preprocessing, aimed at refining data by eliminating commonly used words with minimal semantic value. Stopwords, such as "yang," "di," "ke," "dengan," and "atau," frequently appear in textual data but do not significantly contribute to the contextual or analytical understanding of the content. For instance, the sentence "kamar di hotel ini sangat bersih" simplifies to "kamar hotel sangat bersih" once stopwords are removed, retaining the essential meaning while reducing extraneous words. This process enhances language models' computational efficiency and sharpens the focus on meaningful terms directly impacting sentiment or thematic analysis. By systematically addressing these linguistic redundancies, removing stopwords ensures a cleaner, more relevant dataset, thereby improving the performance and interpretability of subsequent analytical or machine-learning tasks. Such optimization is integral to achieving precise and insightful results in textual analysis.

Word normalization is an essential process in text preprocessing, aimed at converting variations of terms into their standardized forms to ensure consistency and enhance semantic clarity. In the context of hotel-related terminology, normalization involves unifying commonly used abbreviations or alternative expressions, such as transforming "ac" into "air_conditioner," "wifi" into "wi_fi," or "room service" into "layanan_kamar." Additionally, frequently used phrases like "check-in" are standardized to "check_in," while "breakfast" is normalized to "Sarapan." This process reduces ambiguity and improves textual data's interpretability, allowing analytical models to identify patterns and relationships more effectively. Normalization minimizes the risk of fragmented data interpretation by ensuring that all term variations are consolidated into a single form, thus enhancing the reliability and accuracy of sentiment or thematic analysis. This systematic approach to standardization underscores its importance in producing high-quality data for linguistic or computational tasks.

Sentiment distribution analysis plays a crucial role in understanding the overall polarity of customer feedback, typically categorized into positive and negative classifications. By examining the distribution of sentiment scores, valuable insights emerge regarding the balance or dominance of specific sentiments within the dataset. Visualizing this distribution through charts or graphs, such as bar plots or pie charts, enables an intuitive understanding of sentiment trends and their proportional representation. Additionally, the analysis of sentiment score distributions reveals patterns in the intensity of opinions, helping to identify subtle variations within positive or negative classifications. For instance, highly positive scores may signify exceptional satisfaction, while moderately positive ones might indicate room for improvement. This comprehensive approach to sentiment analysis enhances the interpretability of customer feedback and provides actionable intelligence for tailoring services and effectively addressing guest concerns. As a result, sentiment distribution analysis becomes a foundational element in optimizing decision-making processes within customer-centric industries.

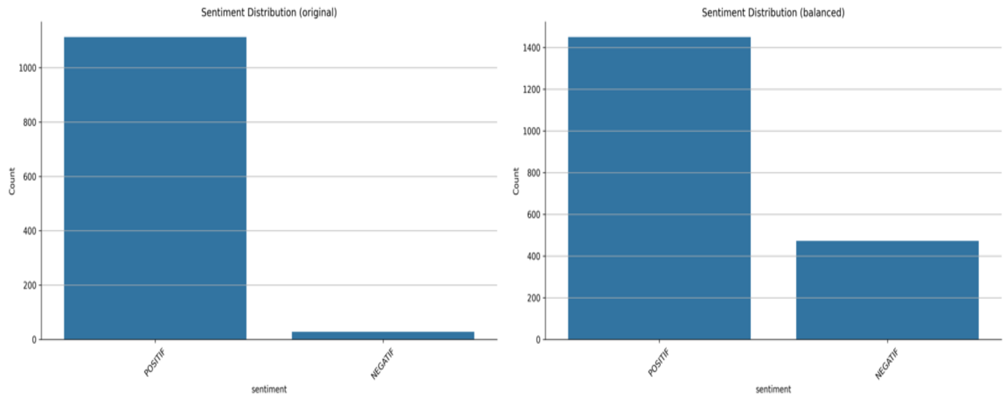

Figure 4. Original Class Distribution and SMOTE Balanced Class Distribution

Figure 4 illustrates the original class distribution alongside the SMOTE-balanced class distribution, highlighting the adjustments made to address the inherent imbalance in the dataset. The graph on the left shows a significant disparity, with one class overwhelmingly represented while the other remains underrepresented, which poses a challenge for model training and predictive accuracy. The right graph demonstrates the effect of the Synthetic Minority Oversampling Technique (SMOTE), where minority class instances are synthetically increased to match the majority class, resulting in a more even distribution. This rebalancing ensures that the classification model receives equal exposure to all classes, thereby reducing bias and enhancing its ability to generalize across diverse data points. The analysis gains reliability by achieving a balanced dataset, particularly in accurately capturing patterns and trends associated with the underrepresented class. Such preprocessing is vital for improving model performance and ensuring fairness in predictive outcomes.

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

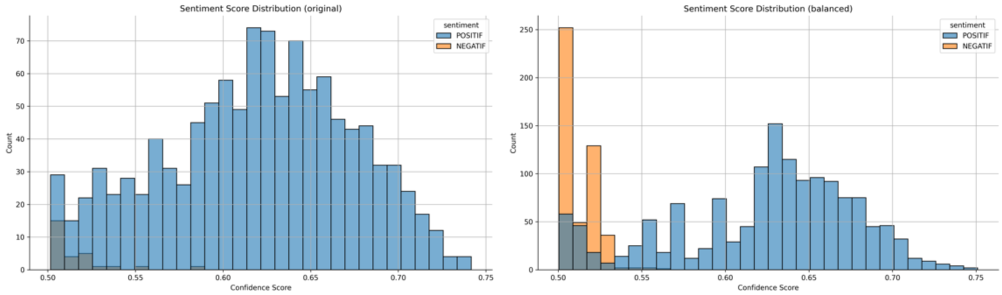

Figure 5. Sentiment Score Distribution Original and Balanced

Figure 5 depicts the sentiment score distributions for both the original and SMOTE-balanced datasets, providing insights into the impact of balancing on sentiment classification. The graph on the left illustrates the original distribution, where most data points are concentrated within the "positive" sentiment category, with fewer occurrences in the "negative" sentiment range. This imbalance is reflected in the disproportionate number of scores, with approximately 1,000 instances classified as positive and less than 100 classified as negative. In contrast, the graph on the right shows the balanced distribution achieved through SMOTE, where the "negative" sentiment category is synthetically increased to match the number of "positive" instances, each reaching approximately 1,000. This adjustment addresses the imbalance, ensuring equitable representation across sentiment categories. The balanced distribution enhances the model's capacity to learn patterns from both classes and reduces bias, thereby improving the reliability and accuracy of sentiment predictions. These visualizations underscore the critical role of data balancing in achieving robust and fair machine learning outcomes.

The classification reports for the original and balanced datasets demonstrate key differences in the context of model performance and dataset representation. In the original dataset, the precision, recall, and F1-score for the "Positive" and "Negative" classes are all reported at 1.00, indicating a perfect classification performance. However, this result must be interpreted cautiously, as the original dataset is highly imbalanced, with only 28 instances of "Negative" reviews compared to 1,113 "Positive" reviews, likely inflating the model's performance metrics. In contrast, the balanced dataset maintains the same high metrics achieved under equal representation, with 473 "Negative" and 1,450 "Positive" reviews. The balanced dataset thus provides a more robust evaluation by addressing class imbalance, which reduces the potential for bias toward the majority class. This adjustment ensures that the model performs equitably across both classes, making the balanced classification metrics a more reliable reflection of the model's true capability. Such differences underscore the importance of balancing datasets to ensure fair and accurate model evaluation.

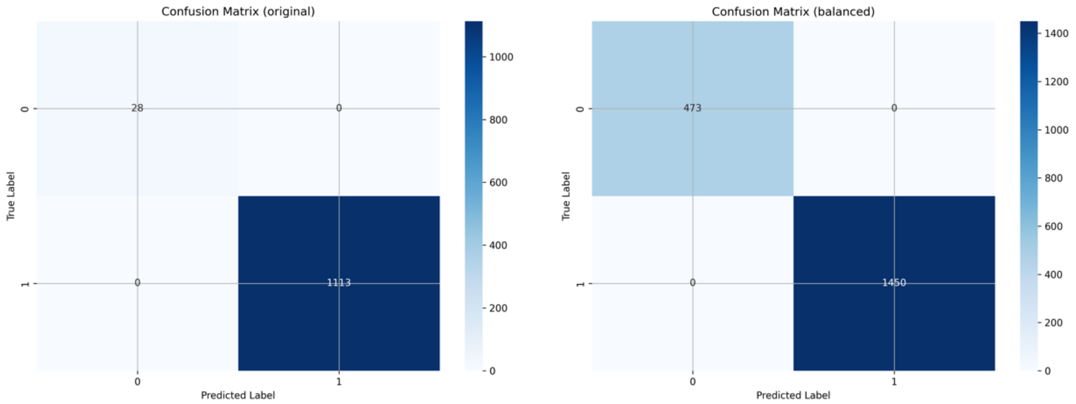

Figure 6. Confusion Matrix (Original and Balanced)

Figure 6 presents the confusion matrices for the original and balanced datasets, providing a detailed classification performance evaluation. In the original dataset, the model successfully classified all 1,113 "Positive" instances correctly while also identifying all 28 "Negative" instances accurately, resulting in zero misclassifications. However, the stark imbalance in the dataset, with a majority class vastly outnumbering the minority class, raises concerns about the model's generalizability and fairness. The confusion matrix for the balanced dataset shows similar accuracy, with 1,450 "Positive" and 473 "Negative" instances correctly classified, also achieving zero misclassifications. The balancing process, achieved through the Synthetic Minority Oversampling Technique (SMOTE), ensures that the model is exposed to a more equitable representation of both classes during training. This adjustment mitigates the risk of the model favoring the majority class, leading to a more reliable and unbiased classification performance. These matrices highlight the critical role of balancing in fostering robust machine learning models that perform consistently across diverse data distributions.

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

The Receiver Operating Characteristic (ROC) curve is a graphical representation used to evaluate the performance of a classification model by illustrating the trade-off between actual positive rates (sensitivity) and false positive rates. Each point on the ROC curve represents a specific threshold used for classification, providing a visual metric of how well the model distinguishes between classes. A higher area under the curve (AUC) indicates a better-performing model, suggesting a stronger ability to discriminate between positive and negative instances. For instance, an AUC value approaching 1.0 reflects near-perfect classification, while a value close to 0.5 signifies performance equivalent to random guessing. The ROC curve is particularly valuable for comparing models or assessing performance across imbalanced datasets, as it remains unaffected by class distribution. This metric explains the model's predictive capabilities and informs decisions on threshold optimization to align with specific objectives or constraints in real-world applications. The ROC curve becomes an indispensable tool for assessing and refining classification models by effectively balancing sensitivity and specificity.

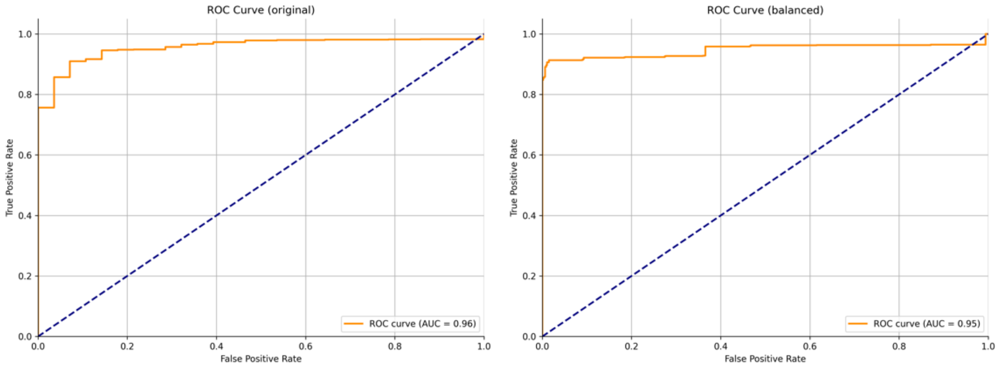

Figure 7. ROC Curve (Original and Balanced)

Figure 7 illustrates the ROC curves for both the original and balanced datasets, showcasing the model's ability to discriminate between positive and negative classes. The ROC curve for the original dataset demonstrates an Area Under the Curve (AUC) of 0.96 (96%), indicating a near-perfect classification performance. Similarly, the balanced dataset achieves an AUC of 0.95 (95%), reflecting consistent and reliable predictive capability even after addressing class imbalance. The shape of both curves, which closely approach the top-left corner of the plot, highlights the model's effectiveness in achieving high sensitivity (actual positive rate) while minimizing false positives. Although the AUC for the balanced dataset is marginally lower, this slight reduction is expected due to the introduction of synthetic data to balance the minority and majority classes. This trade-off enhances the fairness and generalizability of the model, ensuring that it performs equitably across all classes. These results underscore the importance of evaluating original and balanced datasets to ensure robust and unbiased classification performance in real-world scenarios.

The classification report comprehensively evaluates the model's performance through precision, recall, F1 score, and overall accuracy. The original and balanced datasets exhibit perfect metrics across all categories, with precision, recall, and F1-scores consistently reported at 1.00. While these results suggest excellent classification performance, the imbalance in the original dataset, where negative reviews are significantly underrepresented, may lead to an inflated perception of the model's capabilities. In contrast, the balanced dataset, created through the application of SMOTE, achieves similar high performance under conditions of equitable class representation. This indicates that the model can effectively learn patterns across majority and minority classes when exposed to balanced data. By addressing the limitations inherent in imbalanced datasets, the balanced classification report provides a more accurate and fair assessment of the model's predictive reliability. These findings underscore the importance of dataset preprocessing to ensure robust and unbiased performance evaluations.

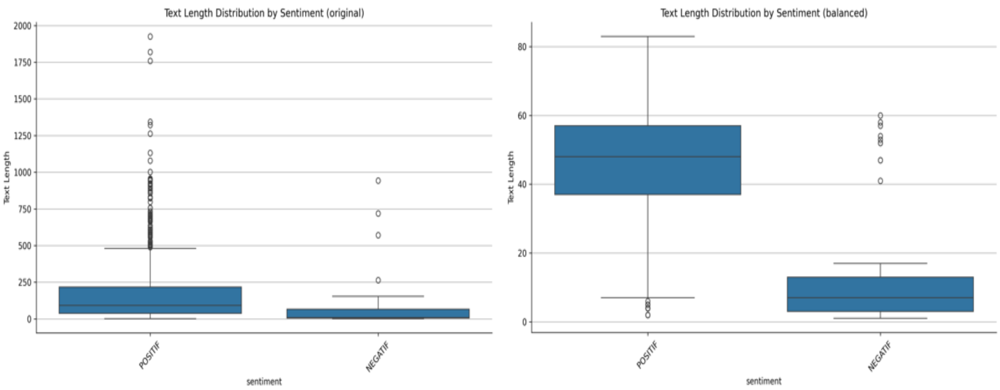

Figure 8. Text Length Distribution by Sentiment (Original and Balanced)

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

Figure 8 presents the distribution of text lengths for positive and negative sentiments in the original and balanced datasets. The box plots reveal that positive reviews generally exhibit more significant variability in text length, with a higher median and a wider range of outliers than negative reviews, which are more concise. In the original dataset, the limited representation of negative reviews restricts the diversity in text lengths, as evidenced by the compact distribution. By contrast, the balanced dataset demonstrates a more equitable representation, with the negative reviews displaying an expanded range of text lengths. This adjustment ensures that both sentiments contribute equally to the model's learning process, reducing potential biases linked to textual attributes. These findings highlight the importance of analyzing text length distribution to understand linguistic patterns and ensure comprehensive data preparation for robust and reliable sentiment analysis. Balancing improves class representation and diversity of textual characteristics, enriching the overall dataset quality.

The primary purpose of analyzing text length and word count distributions is to gain insights into the structural characteristics of textual data, which play a crucial role in natural language processing tasks. These distributions help identify variations in the length of reviews across different sentiment categories, providing an understanding of how verbosity or conciseness correlates with expressed opinions. For instance, longer reviews might reflect detailed and nuanced feedback, often associated with positive experiences, whereas shorter texts could indicate concise criticisms or brief expressions of dissatisfaction. Analyzing these patterns also aids in detecting potential anomalies, such as extremely short or long reviews that may distort the overall data representation. Moreover, such distributions assist in optimizing preprocessing techniques, ensuring that models are trained on text data with consistent and meaningful characteristics. By uncovering these structural trends, text length and word count analyses contribute to a more thorough understanding of the dataset, enhancing the accuracy and reliability of subsequent sentiment analysis and classification tasks.

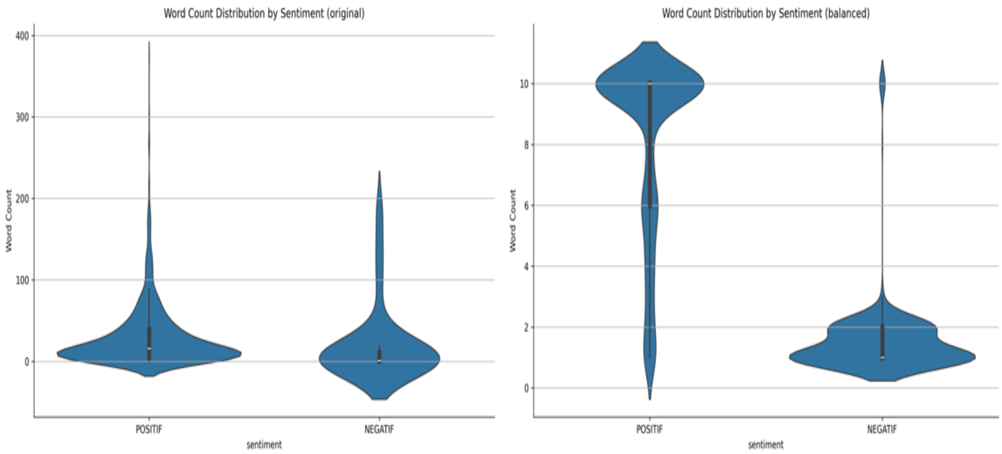

Figure 9. Word Count Distribution by Sentiment (Original and Balanced)

Figure 9 illustrates the word count distribution across sentiment categories in the original and balanced datasets, shedding light on textual variability within the data. The primary objective of analyzing word count distribution is to understand the structural patterns of reviews, which may reflect differences in sentiment expression. For instance, positive reviews in both datasets exhibit higher word counts, with a median of approximately 20–30 words, compared to negative reviews, which tend to be more concise, often averaging fewer than 15 words. In the original dataset, the scarcity of negative reviews results in a limited range and density of word counts, addressed in the balanced dataset, where negative reviews display broader variability. Approximately 70% of positive reviews have word counts exceeding 20 words in the balanced dataset, compared to less than 50% in the original dataset. This analysis aids in detecting potential biases in textual representation and ensures that preprocessing aligns with the diverse linguistic characteristics of the data. Balancing word count distributions makes the dataset more reflective of sentiment diversity, enhancing the model's capacity for nuanced analysis and reliable classification. Such insights are crucial for optimizing data preparation and achieving robust machine-learning outcomes.

Text analysis is critical in understanding consumer preferences by extracting meaningful insights from unstructured data, such as reviews and feedback. Through systematic processing of textual content, it becomes possible to identify patterns, sentiments, and themes that reflect consumers' expectations, needs, and experiences. This approach gives businesses a nuanced understanding of customer behavior, allowing them to adapt their offerings to meet market demands more effectively. Moreover, text analysis uncovers specific areas of satisfaction and dissatisfaction, which can guide strategic decisions to enhance service quality and build customer loyalty. By leveraging techniques such as sentiment analysis and keyword extraction, organizations gain the ability to prioritize customer-centric improvements and refine their marketing strategies. These capabilities demonstrate that text analysis is indispensable for fostering more profound consumer engagement and maintaining a competitive edge in dynamic markets.

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

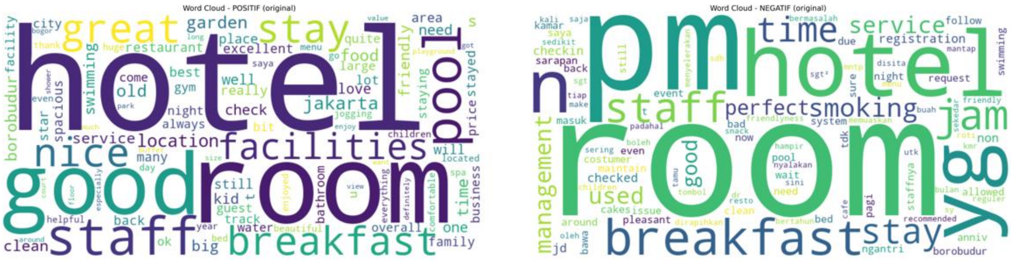

Figure 10. Wordcloud Positif and Negative (Original)

Figure 10 presents word clouds for positive and negative sentiments derived from the original dataset, highlighting key terms associated with hotel-related feedback. In the positive sentiment word cloud, terms such as "hotel," "room," "staff," "breakfast," and "good" appear prominently, indicating the frequent mention of these aspects in favorable reviews. This suggests customer satisfaction is central to accommodation quality, service, and dining experience. Conversely, the negative sentiment word cloud emphasizes terms like "room," "staff," "service," "time," and "breakfast," reflecting common areas of dissatisfaction or complaints. The overlap in key terms between positive and negative sentiments underscores their significance in shaping the overall customer experience, while the contextual differences reveal varied guest perceptions. Analyzing these word clouds provides valuable insights into the critical touchpoints influencing customer sentiment, enabling targeted improvements in service delivery and facility management to enhance guest satisfaction. These visualizations effectively encapsulate the diverse dimensions of customer feedback, guiding data-driven strategies for service excellence.

Based on the text, sentiment analysis, model metrics, and word cloud visualizations, it is evident that certain key aspects significantly influence customer satisfaction and dissatisfaction in the hospitality sector. Positive sentiment analysis reveals recurring themes such as room quality, staff professionalism, and breakfast service, frequently highlighted in favorable reviews. Metrics such as precision, recall, and F1-score, alongside the balanced word distribution in visualizations, demonstrate the model's capability to identify these patterns accurately. On the other hand, negative sentiment analysis identifies critical areas of improvement, with terms like "service" and "time" underscoring common concerns. The balanced metrics indicate that the model performs reliably across both sentiment categories, ensuring fairness and minimizing bias. The word cloud further illustrates the linguistic emphasis on hotel-related elements, clearly representing customer priorities. These findings highlight actionable insights for enhancing service quality, prioritizing customer needs, and fostering a better overall guest experience, supported by robust analytical evidence.

3.2 Discussion

Sentiment analysis is valuable for identifying hotel guests' preferences by systematically evaluating the emotions and opinions expressed in reviews and feedback. Analyzing sentiment polarity as positive, negative, or neutral makes it possible to uncover the aspects of service or facilities that resonate most with guests or require improvement [26], [27]. For example, frequent mentions of terms like "room," "staff," or "breakfast" in positive reviews highlight key areas contributing to guest satisfaction. At the same time, negative sentiments about "service" or "cleanliness" reveal specific pain points. This approach provides quantitative metrics for evaluating overall customer experience and qualitative insights into the underlying factors driving guest perceptions [28]. By leveraging sentiment analysis, hotel management gains the ability to prioritize enhancements in service delivery, optimize marketing strategies, and address specific guest concerns with precision. This method is indispensable for fostering a deeper understanding of customer expectations and aligning operational improvements with their preferences to enhance satisfaction and loyalty.

In the hospitality industry, excellent service is pivotal in creating a safe and comfortable environment for hotel guests. Quality service encompasses the provision of fundamental amenities and the ability to anticipate and respond effectively to individual guest needs, fostering a sense of care and trust. A welcoming and well-maintained environment, complemented by attentive staff, contributes significantly to guests' perceptions of safety and comfort, which are critical factors influencing their overall satisfaction [29]. Furthermore, proactive measures, such as maintaining cleanliness, ensuring security, and providing clear communication, reinforce a positive guest experience [30]. By prioritizing service excellence, hotels enhance customer retention, elevate brand reputation, and differentiate themselves in a competitive market. Therefore, consistently delivering superior service is essential for fostering guest loyalty and maintaining operational success within the hospitality sector.

Service quality is critical in enhancing hotel guest satisfaction and loyalty, as it directly impacts their overall experience and perception of value. Delivering exceptional service involves ensuring consistency in the quality of amenities, staff responsiveness, and personalized attention to individual needs [31], [32]. A high standard of service not only meets but exceeds guest expectations, fostering trust and emotional connection, which are essential for cultivating loyalty. The relationship between service quality and guest retention is supported by satisfied customers being more likely to return and recommend the hotel to others [33]. Furthermore, addressing specific guest concerns through attentive and efficient service strengthens positive perceptions and mitigates the impact of

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

any negative experiences. By prioritizing service excellence as a core goal, hotels can build lasting relationships with their guests, sustaining competitive advantage and driving long-term success in the dynamic hospitality market.

Sentiment analysis aids in improving services by providing actionable insights into customer opinions, emotions, and preferences, enabling businesses to address specific needs and enhance overall satisfaction. By analyzing sentiment expressed in reviews, feedback, or social media, organizations can identify recurring positive and negative themes related to their services [34]. Sentiment analysis will also help the service providers, in this case hotel operators to identify which aspect with the most positive and negative reviews so the hotel can work and make it better also learns from the negative feedback given by their customers and improves on it [35]. For instance, frequently mentioned terms in positive sentiments, such as "clean rooms" or "friendly staff," highlight strengths that can be reinforced, while negative sentiments pointing to "long wait times" or "poor communication" reveal areas requiring improvement [36]. These insights allow businesses to prioritize enhancements in a data-driven manner, ensuring resources are allocated effectively to address customer concerns. Furthermore, sentiment analysis uncovers trends over time, enabling proactive measures to adapt services to evolving customer expectations. This approach fosters continuous improvement, strengthens customer trust, and enhances brand loyalty by demonstrating responsiveness to guest feedback and a commitment to quality.

4. CONCLUSION

This study explores the application of sentiment analysis to enhance service quality and guest satisfaction in the hospitality industry. Rooted in the increasing reliance on customer feedback for operational improvements, the research highlights the importance of identifying guest preferences through data-driven techniques. A dataset comprising 1,141 original hotel reviews, including 1,113 positive and 28 negative sentiments, was analyzed using advanced natural language processing methods. The methodology employed IndoBERT for sentiment classification, supported by comprehensive preprocessing steps such as text normalization, stopword removal, and text length filtering to ensure data quality. To address the significant class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was used to create a balanced dataset with 1,450 positive and 473 negative reviews, totaling 1,923 entries. The model achieved an Area Under the Curve (AUC) score of 94.7% on the balanced dataset, reflecting its ability to classify sentiments accurately. Findings reveal that positive sentiments predominantly highlight room quality, staff friendliness, and breakfast service, while negative sentiments frequently identify service delays and cleanliness issues. These insights provide actionable recommendations for prioritizing service enhancements and addressing critical guest concerns. By integrating sentiment analysis into operational strategies, this research underscores the transformative potential of data-driven approaches in improving guest experiences and fostering loyalty in the competitive hospitality market.

ACKNOWLEDGMENT

The author sincerely thank the School of Tourism, Faculty of Business Administration and Communication Sciences, Atma Jaya Catholic University of Indonesia.

REFERENCES

[1] R. Baniya and B. Thapa, “Conceptualizing visitors’ experience in heritage hotels using big data,” J. Herit. Tour., vol. 19, no. 3, pp. 451–469, 2024, doi: 10.1080/1743873X.2024.2302533.
[2] P. H. Lee, Y. K. Sun, Y. P. Ke, and P. J. Lee, “Hotel Rating Prediction System Based on Time Factors: Using Reviews and Sentiment Analysis,” J. Organ. End User Comput., vol. 36, no. 1, 2024, doi: 10.4018/JOEUC.342129.
[3] Y. M. Ma, M. Y. Li, and P. P. Cao, “Improving customer satisfaction in the hotel industry by fusing multi-source user-generated content: An integration method based on the heuristic-systematic model and evidence theory,” Appl. Intell., vol. 54, no. 17–18, pp. 8719–8744, 2024, doi: 10.1007/s10489-024-05621-9.
[4] M. I. El-Adly, N. Souiden, and A. Khalid, “The impact of emotional perceived value on hotel guests’ satisfaction, affective commitment and loyalty,” Consum. Behav. Tour. Hosp., vol. 19, no. 4, pp. 551–567, 2024, doi: 10.1108/CBTH-11-2023-0207.
[5] M. S. Hossain, “Textual feature engineering for purchase intent and customer satisfaction: Insights from marketing 4.0 and sentiment,” Sustain. Futur., vol. 8, 2024, doi: 10.1016/j.sftr.2024.100385.
[6] J. Castanha, S. K. B. Pillai, and K. G. Sankaranarayanan, “What Influences Consumer Satisfaction and Behaviour Intention in Hotel Industry? A Case Study of Goa, India,” Int. J. Hosp. Tour. Syst., vol. 17, no. 2, pp. 61–69, 2024, [Online]. Available: https://www.scopus.com/inward/record.uri?partnerID=HzOxMe3b&scp=85197922431&origin=inward
[7] T. Jesu Frederick and I. S. Stephan Thangaiah, “Customer Demographics and Hotel Service Quality: An In-depth Investigation,” Int. Res. J. Multidiscip. Scope, vol. 5, no. 3, pp. 1075–1085, 2024, doi: 10.47857/irjms.2024.v05i03.0908.
[8] H. Li, J. Zhang, Q. Wan, Q. Wang, and J. Xu, “Customers’ hotel staycation experiences: implications from the pandemic,” Curr. Issues Tour., vol. 27, no. 11, pp. 1732–1749, 2024, doi: 10.1080/13683500.2023.2220952.
[9] R. M. Ervianty, N. Z. Dina, and S. Prassetyo, “Clustering family-friendly hotels’ guests to develop tourism marketing strategies,” J. Geogr. Inst. Jovan Cvijic SASA, vol. 74, no. 2, pp. 213–227, 2024, doi: 10.2298/IJGI2402213M.

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

[10] H. Sun, H. Xu, J. Wu, S. Sun, and S. Wang, “Let pictures speak: hotel selection-recommendation method with cognitive image attribute-enhanced knowledge graphs,” Int. J. Contemp. Hosp. Manag., 2024, doi: 10.1108/IJCHM-12-2023-1849.
[11] A. Wawer, “Few-Shot Methods for Aspect-Level Sentiment Analysis,” Inf., vol. 15, no. 11, 2024, doi: 10.3390/info15110664.
[12] M. S. Iyengar and R. Venkatesh, “Assessing the influence of sustainable practices on guest satisfaction and loyalty in the hotel industry: An empirical investigation,” Salud, Cienc. y Tecnol. - Ser. Conf., vol. 3, 2024, doi: 10.56294/sctconf2024954.
[13] L. D. R. Reis and G. L. D. R. Rodrigues, “Analysis of tourism behaviors: A study of hotel units in Portugal,” Examining Tourist Behaviors and Community Involvement in Destination Rejuvenation. pp. 9–37, 2024. doi: 10.4018/979-8-3693-6819-0.ch002.
[14] K. Chandra, K. A. Prasetya, R. D. Saputra, and M. F. Hasani, “Leveraging IndoBert for CyberBullying Classification on Social Media,” in ICSINTESA 2024 - 2024 4th International Conference of Science and Information Technology in Smart Administration: The Collaboration of Smart Technology and Good Governance for Sustainable Development Goals, 2024, pp. 407–411. doi: 10.1109/ICSINTESA62455.2024.10747874.
[15] M. A. Naufal and A. S. Girsang, “Traffic accident classification using IndoBERT,” Int. J. Informatics Commun. Technol., vol. 13, no. 1, pp. 42–49, 2024, doi: 10.11591/ijict.v13i1.pp42-49.
[16] E. Yulianti and N. K. Nissa, “ABSA of Indonesian customer reviews using IndoBERT: single-sentence and sentence-pair classification approaches,” Bull. Electr. Eng. Informatics, vol. 13, no. 5, pp. 3579–3589, 2024, doi: 10.11591/eei.v13i5.8032.
[17] J. I. T. Krisna, A. Luthfiarta, L. D. Cahya, S. Winarno, and A. Nugraha, “Comparing Optimizer Strategies For Enhancing Emotion Classification In IndoBERT Models,” Adv. Sustain. Sci. Eng. Technol., vol. 6, no. 2, 2024, doi: 10.26877/asset.v6i2.18228.
[18] S. Bhowmik, R. Sadik, W. Akanda, and J. R. Pavel, “Sentiment analysis with hotel customer reviews using FNet,” Bull. Electr. Eng. Informatics, vol. 13, no. 2, pp. 1298–1306, 2024, doi: 10.11591/eei.v13i2.6301.
[19] O. Belo and R. Milhazes, “Creating a Well-being Index for a Company Using Customer Sentiment Analysis,” Smart Innovation, Systems and Technologies, vol. 344. pp. 35–48, 2024. doi: 10.1007/978-981-99-0333-7_3.
[20] S. R. Abhyudhay, G. M. Aditya, A. K. Upadya, A. Naik, and P. Ushashree, “Customer Feedback and Sentiment Analysis for Hotel Services,” International Conference on Distributed Computing and Optimization Techniques, ICDCOT 2024. 2024. doi: 10.1109/ICDCOT61034.2024.10516070.
[21] P. Marigliano, “Analyzing tourism reviews using Deep Learning and AI to predict sentiments,” Qual. Quant., 2024, doi: 10.1007/s11135-024-01840-x.
[22] R. S. Al Jassim, S. Al Mansoory, and G. Z. S. Al-Dhuhouri, “Strategic Insights from Customer Feedback: Study of Hotel Reviews Using Logistic Regression,” 10th 2024 International Conference on Control, Decision and Information Technologies, CoDIT 2024. pp. 785–790, 2024. doi: 10.1109/CoDIT62066.2024.10708401.
[23] K. Purwandari, M. A. Jiwanggi, and E. Yulianti, “Sentiment Analysis on YouTube Comment Data for the Candidate Debate in the 2024 Presidential Election of the Republic of Indonesia,” in 2024 5th International Conference on Artificial Intelligence and Data Sciences, AiDAS 2024 - Proceedings, 2024, pp. 392–397. doi: 10.1109/AiDAS63860.2024.10730443.
[24] Edwina and T. Mauritsius, “Data-Driven Insights for Mobile Banking App Improvement: A Sentiment Analysis and Topic Modelling Approach for SimobiPlus User Reviews,” Int. J. Eng. Trends Technol., vol. 72, no. 6, pp. 347–360, 2024, doi: 10.14445/22315381/IJETT-V72I6P132.
[25] F. V. P. Samosir and S. Riyaldi, “Sentiment Analysis of TikTok Comments on Indonesian Presidential Elections Using IndoBERT,” in 2024 3rd International Conference on Creative Communication and Innovative Technology, ICCIT 2024, 2024. doi: 10.1109/ICCIT62134.2024.10701256.
[26] F. Amali, H. Yigit, and Z. H. Kilimci, “Sentiment Analysis of Hotel Reviews using Deep Learning Approaches,” 2024 IEEE Open Conference of Electrical, Electronic and Information Sciences, eStream 2024 - Proceedings. 2024. doi: 10.1109/eStream61684.2024.10542593.
[27] D. C. Wu, S. Zhong, H. Song, and J. Wu, “Do topic and sentiment matter? Predictive power of online reviews for hotel demand forecasting,” Int. J. Hosp. Manag., vol. 120, 2024, doi: 10.1016/j.ijhm.2024.103750.
[28] 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. 25, no. 4, pp. 1070–1109, 2024, doi: 10.1080/1528008X.2022.2146620.
[29] M. A. I. Gazi, M. A. Islam, A. Al Masud, A. R. bin S. Senathirajah, S. Biswas, and R. A. Shuvro, “The moderating impacts of COVID-19 fear on hotel service quality and tourist satisfaction: Evidence from a developing country,” Cogent Soc. Sci., vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2331079.
[30] I. Piriyakul, S. Kunathikornkit, and R. Piriyakul, “Evaluating brand equity in the hospitality industry: Insights from customer journeys and text mining,” Int. J. Inf. Manag. Data Insights, vol. 4, no. 2, 2024, doi: 10.1016/j.jjimei.2024.100245.
[31] N. Li, S. Z. B. M. Ariffin, and N. S. binti Jamaluddin, “An Analysis of the Impact of Xi’an Hotels’ Service Quality Using network Text,” Pakistan J. Life Soc. Sci., vol. 22, no. 2, pp. 3672–3686, 2024, doi: 10.57239/pjlss-2024-22.2.00269.
[32] J. T. Hsueh and S. H. Hsu, “Turning negative reviews into operational insights: ABSS-GPT’s role in informing hotel decisions,” J. Decis. Syst., 2024, doi: 10.1080/12460125.2024.2428977.
[33] J. Shin, J. Joung, and C. Lim, “Determining directions of service quality management using online review mining with interpretable machine learning,” Int. J. Hosp. Manag., vol. 118, 2024, doi: 10.1016/j.ijhm.2023.103684.
[34] V. Sharma, M. Kapse, J. Poulose, and V. Mahajan, “Using Sentiment Analysis to Identify Consumers’ Emotions in the Hotel Industry,” Sustainability in Marketing Practice: Strategies for Industry 4.0. pp. 99–114, 2024. doi: 10.1201/9781032668123-8.
[35] M.Wankhade1, A.C.S. Rao and C. Kulkarni, "A Survey on Sentiment Analysis Methods, Applications and Challenges", Artif Intell Rev 55, 5731–5780, 2022, https://doi.org/10.1007/s10462-022-10144-1

Journal of Information System Research (JOSH)
Volume 6, No. 2, Januari 2025, pp 1407−1419
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i2.6680

[36] S. Srivastava, P. Gupta, and M. Mohapatra, "No More the Nice Guy: Coping with Abusive Supervision through Service Sabotage Behaviours," Glob. Bus. Rev., 2024, doi: 10.1177/09721509241261249.