

IndoBERT-Based Sentiment Analysis for Understanding Hotel Guests' Preferences

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

Submitted: 28/01/2025; Accepted: 28/02/2025; Published: 28/02/2025

Abstract—The rapid growth of the hospitality industry and the increasing reliance on online reviews emphasize the need for advanced sentiment analysis tools to understand customer preferences effectively. This study explores the application of IndoBERT, a pre-trained language model tailored for the Indonesian language, in classifying sentiments from hotel guest reviews. Utilizing a dataset of 715 reviews, the study employed the Knowledge Discovery in Databases (KDD) framework for systematic data preprocessing, feature extraction, and machine learning analysis. IndoBERT demonstrated exceptional performance, achieving perfect precision, recall, and F1-scores of 1.00 for both positive (657 reviews) and negative (53 reviews) sentiment classes. The ROC curve analysis also yielded a mean AUC score of 0.86, validating the model's robustness and reliability. The results highlight IndoBERT's capability to accurately capture linguistic nuances and contextual meaning, offering actionable insights into factors influencing guest satisfaction, such as cleanliness, staff behavior, and service quality. This research contributes to advancing natural language processing applications in regional contexts and provides practical implications for enhancing service strategies in the hospitality sector. Future research should expand the model's application to other industries and explore multimodal approaches for a more comprehensive understanding of customer behavior.

Keywords: Sentiment Analysis; IndoBERT; Hotel Guest Preferences; Hospitality Industry; Natural Language Processing

1. INTRODUCTION

The rapid growth of the hospitality industry, coupled with the increasing reliance on online reviews, underscores the need for advanced analytical approaches to comprehend guest preferences effectively. Sentiment analysis, a subset of natural language processing (NLP), has emerged as a critical tool in interpreting the subjective opinions expressed in customer reviews [1], [2]. IndoBERT, a pre-trained language model tailored for the Indonesian language, offers significant potential for accurately capturing nuanced sentiments within this context [3]. The application of IndoBERT enables precise extraction of insights from vast datasets, facilitating an understanding of factors influencing guest satisfaction and dissatisfaction [4]. By systematically analyzing linguistic patterns and sentiment polarity, this approach addresses the complexities of Indonesian textual data that conventional methods often overlook. Employing such a model enhances the accuracy of sentiment interpretation and contributes to the strategic alignment of hospitality services with guest expectations. These advancements highlight the pivotal role of sentiment analysis in driving data-informed decision-making processes within the competitive landscape of the hotel industry.

The increasing reliance on digital platforms for customer feedback in the hospitality sector highlights the critical need for advanced methodologies to interpret and leverage this information effectively. The rapid proliferation of user-generated content, particularly online reviews, necessitates a sophisticated approach to extracting meaningful insights that reflect consumer preferences and expectations [5], [6]. Traditional sentiment analysis methods often struggle to handle regional languages' complexity and linguistic nuances, creating a gap in accurately understanding customer sentiments. Leveraging a model like IndoBERT, specifically designed for the Indonesian language, offers a solution to this limitation by providing enhanced accuracy and contextual relevance in sentiment analysis [7]–[9]. By addressing these challenges, the research fosters an improved alignment of service delivery with guest expectations, enabling businesses to remain competitive in a dynamic market environment. This focus on optimizing customer satisfaction through data-driven insights underscores this study's significance in advancing academic and practical applications within the hospitality industry.

The primary objective of this research is to develop a robust framework for analyzing customer sentiments in the hospitality industry using an advanced natural language processing model tailored to the Indonesian language. Addressing the complexities of regional linguistic structures and the diverse expressions found in online reviews is essential to interpreting guest preferences accurately. IndoBERT enables the extraction of nuanced insights, ensuring a more precise understanding of satisfaction drivers and dissatisfaction factors [10]–[13]. This approach enhances the granularity of sentiment analysis and strengthens its applicability in optimizing service delivery and customer experience management [14]. By integrating advanced computational techniques with practical applications, the study aims to bridge the gap between theoretical advancements and industry needs, ultimately contributing to the broader goal of elevating service standards in the hospitality sector.

The methodology employed in this study utilizes a machine learning approach within the Knowledge Discovery in Databases (KDD) framework to analyze sentiment data systematically. KDD provides a structured

process encompassing key stages such as data selection, preprocessing, transformation, mining, and interpretation, ensuring a comprehensive analysis pipeline [15]. By leveraging this framework, machine learning algorithms extract patterns and insights from large volumes of textual data, enabling a more precise understanding of sentiment trends. This method is particularly effective in handling unstructured data, as it integrates robust techniques for noise reduction, feature extraction, and model optimization. Employing the KDD framework enhances the study's methodological rigor, ensuring that findings are reliable and actionable. This systematic approach underscores the value of combining machine learning with established analytical frameworks, contributing to the development of innovative tools for sentiment analysis in diverse contexts.

Numerous studies have explored sentiment analysis to evaluate customer feedback in the hospitality sector, leveraging various machine learning models and natural language processing techniques. While these efforts have yielded valuable insights, most existing research has focused on widely spoken global languages, often neglecting the linguistic diversity and complexity of regional languages such as Indonesian [16]–[18]. This limitation has resulted in a significant gap in accurately capturing the sentiments expressed by local consumers, particularly in markets where global models fail to address cultural and contextual nuances. IndoBERT, as a language model trained explicitly on Indonesian text, offers a promising avenue for bridging this gap by providing a more contextually relevant analysis of sentiments [19]–[22]. However, its application within the hospitality industry remains underexplored, leaving an opportunity to investigate its potential for delivering more granular and actionable insights. This gap underscores the need for targeted research to advance sentiment analysis capabilities tailored to regional contexts, ultimately enhancing the strategic utility of customer feedback in driving industry innovation.

The novelty of this research lies in its application of IndoBERT, a pre-trained language model tailored for Indonesian, to sentiment analysis within the hospitality industry, addressing a significant gap in localized sentiment evaluation. Unlike conventional approaches that rely on generic language models, this study capitalizes on IndoBERT's capacity to interpret complex linguistic nuances unique to Indonesian, enabling more precise and culturally relevant analysis of customer reviews [23]. Such an approach enhances sentiment classification accuracy and introduces a paradigm shift in understanding localized consumer behavior [24]. This research contributes to natural language processing and hospitality management by integrating advanced computational methods with practical industry challenges. The innovative use of IndoBERT sets a precedent for addressing regional language barriers in sentiment analysis, ultimately fostering greater inclusivity in data-driven decision-making across diverse linguistic contexts.

This research offers a significant theoretical contribution by advancing the application of natural language processing in sentiment analysis, mainly using a model tailored for the Indonesian language. Integrating IndoBERT into the analysis of customer reviews addresses a critical gap in the literature related to the contextual understanding of regional languages, thereby enriching the methodological discourse within computational linguistics. On a practical level, the study provides actionable insights for the hospitality industry, enabling businesses to understand consumer preferences better and align their services accordingly. IndoBERT enhances the precision of sentiment analysis, allowing for a more nuanced interpretation of customer feedback, which can inform strategic decision-making processes [25], [26]. This research expands the academic understanding of sentiment analysis in linguistically diverse contexts by bridging the gap between theoretical advancements and real-world applications. It delivers tangible benefits for improving customer satisfaction and competitive positioning in the market.

Future research should focus on expanding the application of IndoBERT-based sentiment analysis to a broader range of industries and contexts to further validate its versatility and effectiveness. Investigating its performance in e-commerce, healthcare, and education sectors would provide valuable insights into its adaptability in addressing diverse linguistic nuances. Additionally, exploring the integration of IndoBERT with other advanced machine learning techniques, such as ensemble models or hybrid approaches, could enhance its accuracy and scalability in processing large datasets [27], [28]. Another avenue worth pursuing involves incorporating multimodal data, such as combining textual reviews with visual or audio content, to enrich the analysis and provide a more comprehensive understanding of consumer feedback. By addressing these areas, subsequent studies could strengthen the methodological framework for sentiment analysis and its practical utility, fostering innovation in data-driven decision-making across various domains.

2. RESEARCH METHODOLOGY

2.1 Research Workflow

The research workflow ensures a systematic and efficient approach to achieving the study's objectives. It begins with collecting raw data, primarily customer reviews, which are then subjected to preprocessing to remove noise and standardize the input for analysis. Following this, feature extraction techniques are applied to identify key linguistic and semantic patterns within the textual data, ensuring that relevant information is captured for subsequent analysis. Machine learning algorithms perform sentiment classification and pattern recognition within the Knowledge Discovery in Databases (KDD) framework [29]. This phase is critical for uncovering

insights that reflect customer preferences and satisfaction levels. The results are evaluated for accuracy and relevance using established performance metrics, ensuring the robustness of the analysis. Structuring the workflow into distinct, interdependent stages maximizes the findings' clarity, reliability, and practical value while maintaining methodological rigor throughout the research process.

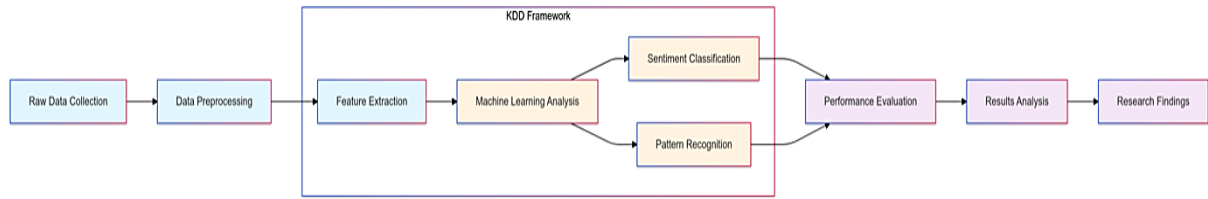


Figure 1. Research Workflow

Figure 1 illustrates a systematic research workflow designed within the Knowledge Discovery in Databases (KDD) framework to achieve an efficient sentiment data analysis. The process begins with raw data collection, where customer reviews are gathered as the primary input for analysis. This is followed by data preprocessing, which eliminates noise, standardizes formats, and ensures the data is suitable for feature extraction. In the next stage, feature extraction isolates key linguistic and semantic elements, preparing the data for machine learning-based analysis. The subsequent step bifurcates into sentiment classification and pattern recognition, where advanced algorithms are employed to categorize sentiments and identify underlying patterns in the data. These findings are then subjected to performance evaluation, utilizing established metrics to ensure the accuracy and reliability of the results. The workflow concludes with results analysis and the presentation of research findings, integrating insights to offer meaningful contributions to both theoretical and practical domains. This structured approach highlights the sequential interdependence of each stage, ensuring methodological rigor and enhancing the validity of the outcomes.

The raw data collection process involves extracting hotel review data from Agoda, explicitly focusing on Qubika Boutique Hotel as the foundation for sentiment analysis. This dataset is acquired through a web scraping procedure using WebHarvy, a sophisticated tool designed for automated data extraction from websites. WebHarvy ensures the efficient gathering of structured and unstructured data while maintaining high accuracy and consistency during scraping. By targeting a specific hotel, the collected reviews offer a well-defined dataset that reflects diverse customer opinions and experiences, providing a comprehensive basis for further analysis. This approach not only streamlines the data collection process but also enhances the relevance of the dataset, as it captures authentic, user-generated content directly associated with the hospitality sector. By employing such a systematic and targeted method, the data collection phase ensures that subsequent analytical stages are based on robust and contextually relevant information.

Data pre-processing is critical in transforming raw data into a clean and structured format suitable for analysis. This process begins with removing irrelevant content, such as advertisements, non-textual elements, and redundant information, to enhance the dataset's quality. Noise reduction techniques, including eliminating special characters, emojis, and stopwords, are applied to ensure the textual data is standardized and meaningful. Tokenization and normalization are then performed to segment the text into analyzable units while maintaining linguistic consistency. Additionally, stemming or lemmatization may reduce words to their base forms, simplifying the data without losing semantic value. Pre-processing ensures that subsequent analytical phases are accurate and efficient by addressing inconsistencies and preparing the data for machine learning models. This systematic refinement underscores the importance of pre-processing as a foundational element in achieving reliable and interpretable results.

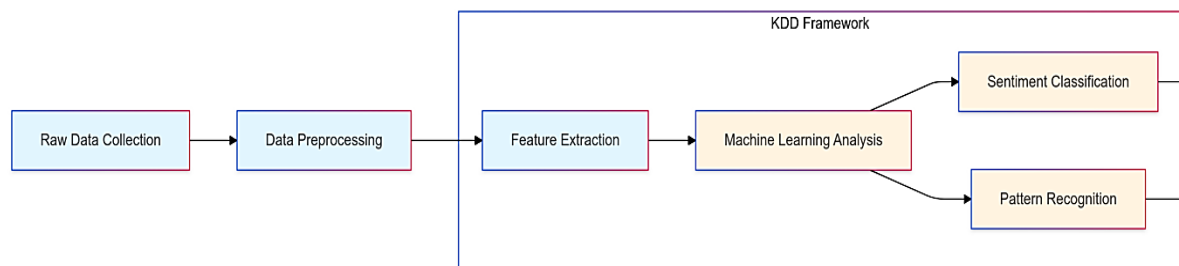


Figure 2. Raw Data Collection, Pre-Processing, and KDD Framework

Figure 2 depicts the systematic process of the Knowledge Discovery in Databases (KDD) framework, encompassing stages such as raw data collection, preprocessing, feature extraction, and machine learning analysis, which ultimately lead to sentiment classification and pattern recognition. The process begins with acquiring raw data from relevant customer reviews, ensuring a comprehensive dataset for analysis. This is followed by data preprocessing, which refines the dataset by removing irrelevant elements, addressing

inconsistencies, and standardizing formats to enhance the quality and usability of the data. Feature extraction then isolates key linguistic and semantic attributes, ensuring that the input for machine learning algorithms is relevant and representative of the underlying patterns. The machine learning phase is divided into sentiment classification and pattern recognition, where advanced algorithms are employed to categorize sentiments and uncover significant trends in the data. This structured and iterative approach ensures methodological rigor and facilitates the generation of actionable insights, highlighting the practical and academic significance of the KDD framework in addressing complex analytical challenges.

The process of feature extraction, machine learning analysis, sentiment classification, and pattern recognition forms the core of a comprehensive analytical framework for understanding textual data. Feature extraction involves identifying and isolating key linguistic and semantic elements from raw data, transforming unstructured text into a structured format suitable for computational analysis. This stage is critical in ensuring that only relevant and meaningful attributes are passed to the subsequent phases, enhancing the efficiency and precision of machine learning algorithms. Machine learning analysis utilizes these features to build predictive models capable of identifying patterns and trends within the data, relying on robust algorithms to ensure accuracy and scalability. Sentiment classification then categorizes textual inputs into predefined sentiment categories, such as positive, negative, or neutral, providing insights into customer attitudes and opinions. Finally, pattern recognition identifies recurring themes or trends within the classified data, offering a deeper understanding of underlying consumer behaviors and preferences. This integrated approach ensures methodological rigor and delivers actionable insights, making it a valuable tool for addressing complex problems across various domains.

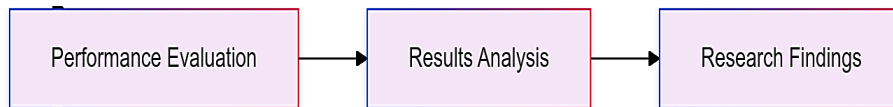


Figure 3. Performance Evaluation, Result Analysis, and Research Findings

Figure 3 illustrates the sequential stages of performance evaluation, result analysis, and research findings, which collectively ensure the validity and relevance of the study's outcomes. Performance evaluation involves assessing the accuracy, reliability, and overall effectiveness of the applied machine learning models using established metrics to determine their ability to classify sentiments and recognize patterns. This phase is essential for identifying potential limitations and optimizing model parameters to enhance analytical precision. Result analysis focuses on interpreting the evaluation outcomes, providing a detailed understanding of the identified sentiment trends and recurring patterns and their implications for the study's objectives. These insights are then synthesized into research findings, presenting actionable conclusions and contributing to theoretical advancement and practical applications. This structured progression ensures that each stage builds upon the previous one, reinforcing the study's rigor and contribution to addressing complex analytical challenges.

The research findings serve as the foundation for the subsequent discussion, offering valuable insights into sentiment trends and behavioral patterns extracted from the analyzed data. These findings clearly understand customer sentiments, preferences, and key drivers of satisfaction and dissatisfaction within the study context. By interpreting these results through a critical lens, the discussion will delve deeper into their implications, exploring how they align with existing theories and contribute to practical advancements in the field. This approach ensures a thorough examination of the results and highlights their relevance in addressing the identified research gap. The transition from findings to discussion enables a comprehensive evaluation of the data, fostering meaningful interpretations that enhance the academic and practical significance of the study. Through this systematic analysis, the research positions itself to deliver robust conclusions and actionable recommendations, emphasizing its value in both theoretical and applied contexts.

2.2 Dataset

The dataset utilized for sentiment classification comprises an extensive range of features extracted from customer reviews, providing a multidimensional view of user experiences and preferences. Key attributes include "Account," identifying the source of feedback; "Country," representing the reviewer's geographical origin; and "Visitor Type," indicating the purpose of their visit. Additional variables, such as "Room Type" and "Length of Stay," offer insights into the context of the accommodation experience, while "Month of Stay" and "Year of Stay" capture temporal dimensions of the data. Furthermore, "Rating" and "Rating Description" provide quantitative and qualitative evaluations of the experience, complemented by textual features like "Title" and "Reviews," which reflect the sentiments and detailed opinions of the customers. This comprehensive dataset enables a nuanced sentiment classification process by combining structured variables with unstructured textual content, offering rich contextual insights. Including diverse features ensures a robust analysis framework, allowing for identifying patterns and sentiment trends that align closely with customer behavior and expectations. Through this structured approach, the dataset supports the development of accurate and contextually relevant models for sentiment classification.

716 reviews were processed using the Knowledge Discovery in Databases (KDD) framework to transform raw textual data into valuable insights for sentiment classification. This volume of data provides a robust foundation for identifying patterns and trends in customer sentiments, ensuring a comprehensive analysis of diverse opinions and experiences. Employing the KDD framework, each review underwent systematic stages, including data preprocessing, feature extraction, and machine learning analysis, to enhance the quality and relevance of the information extracted. Integrating structured and unstructured data elements allowed for a nuanced understanding of sentiment polarity, enabling the classification process to capture subtle variations in customer feedback. This meticulous approach ensures that the resulting insights are accurate and contextually significant, contributing to meaningful interpretations of customer preferences and expectations. The application of this framework demonstrates its capacity to transform raw data into actionable information, supporting informed decision-making and strategic improvements in sentiment-driven analysis.

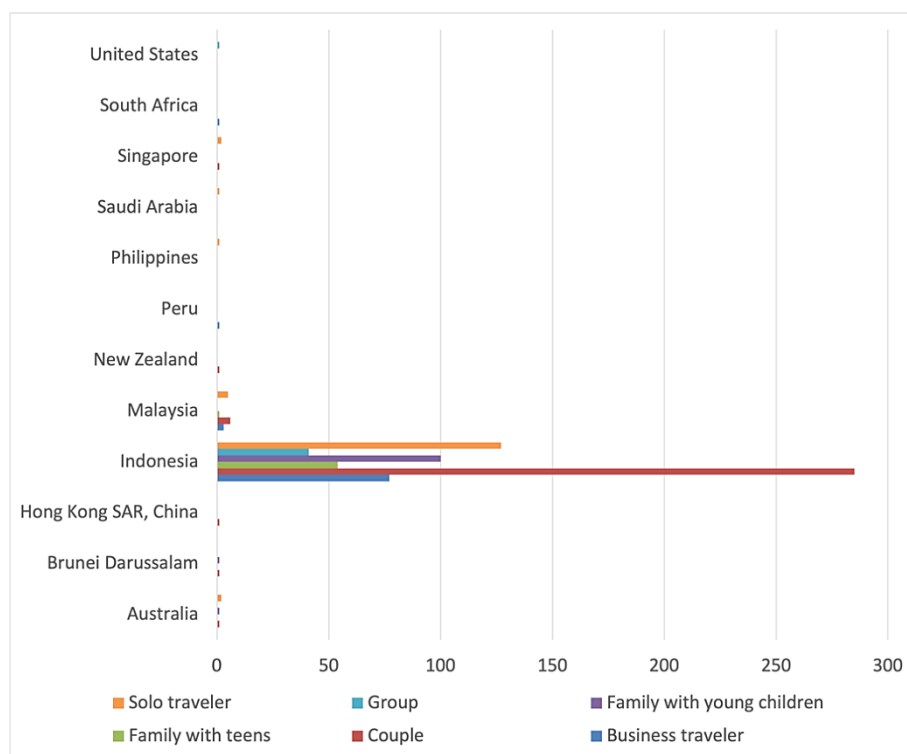


Figure 4. Visitor Type based on Country of Origin (715 Reviews)

Figure 4 illustrates the distribution of visitor types based on their country of origin, encompassing 715 reviews. The data reveals that most visitors originate from Indonesia, with a significant portion being couples and families with young children, highlighting the prominence of domestic tourism in the dataset. Other countries, including Malaysia, Australia, and Singapore, contribute smaller yet notable proportions, varying distributions across solo travelers, business travelers, and family groups. This variation reflects the diverse motivations and travel purposes of international visitors. The predominance of domestic travelers suggests that the hotel caters effectively to local preferences, which may be influenced by accessibility, cultural familiarity, and targeted marketing strategies. Although relatively limited, the presence of international visitors underscores the potential for expanding appeal to a broader audience. This analysis underscores the importance of tailoring services and experiences to meet the specific needs of dominant visitor types while recognizing the opportunities for enhancing global outreach in the hospitality sector.

An analysis of the percentage distribution of visitor types based on country of origin, derived from 715 reviews, reveals insightful patterns in the customer demographic composition. Indonesian visitors constitute the largest segment, accounting for approximately 80% of the data, predominantly represented by couples and families with young children. Visitors from Malaysia follow, contributing around 8%, while smaller proportions are observed from Australia, Singapore, and other regions, collectively comprising the remaining 12%. The significant presence of domestic travelers highlights the hotel's strong resonance within the local market, likely influenced by cultural familiarity, location accessibility, and tailored offerings. The relatively modest international representation, with notable diversity in visitor types such as business travelers and solo tourists, suggests opportunities to enhance the global appeal of the services. This distribution underscores the critical role of understanding market dynamics to optimize customer satisfaction and expand the hotel's outreach. A balanced strategy addressing the dominant domestic market while strategically attracting international clientele could foster sustainable growth and market diversification.

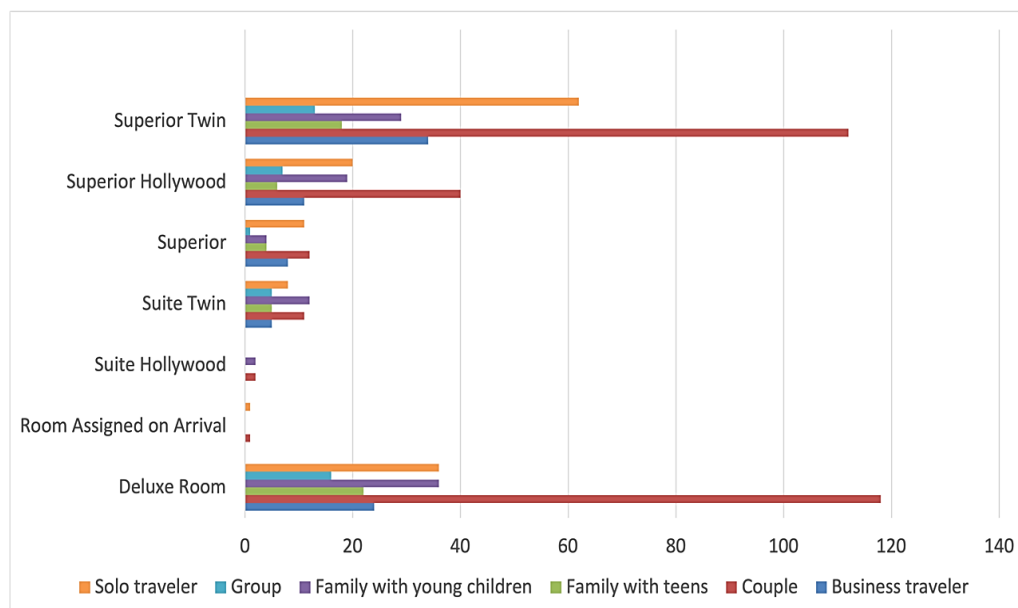


Figure 5. Room Preference-based on visitor Type (715 Reviews)

Figure 5 illustrates the distribution of room preferences among different visitor types based on 715 reviews, highlighting significant variations in accommodation choices. The Superior Twin Room is the most popular option, particularly favored by couples and families with young children, reflecting its suitability for shared occupancy and cost efficiency. The Superior Hollywood and Suite Hollywood rooms show a preference trend among business travelers and solo visitors, likely due to their emphasis on privacy and enhanced amenities. Families with teens and groups exhibit a diverse distribution across room types, indicating their need for flexible arrangements. While less commonly selected, the Deluxe Room and other options cater to niche preferences such as luxury or specific functional requirements. This analysis underscores the interplay between visitor type and room selection, demonstrating the importance of offering a variety of room configurations to accommodate diverse customer needs. Tailoring room features and promotional strategies to align with these patterns could further optimize customer satisfaction and occupancy rates.

An analysis of room preferences based on visitor types, calculated from 715 reviews, reveals distinct patterns highlighting the varying accommodation choices of different traveler categories. Superior Twin rooms account for approximately 45% of total bookings, with the highest demand coming from couples and families with young children, reflecting its practicality and cost-effectiveness for shared occupancy. Superior Hollywood rooms comprise about 25% of preferences, attracting business travelers and solo visitors due to their emphasis on privacy and modern amenities. Suite Hollywood and Deluxe Rooms, collectively representing around 20%, cater to niche segments seeking premium features or spacious configurations, appealing primarily to families with teens and groups. The remaining 10% is distributed across less frequently selected room types, such as Room Assigned on Arrival, often chosen by budget-conscious travelers. This distribution underscores the importance of aligning room offerings with diverse visitor types' specific needs and preferences, suggesting opportunities for targeted marketing strategies and tailored service enhancements. Understanding these patterns facilitates data-driven decision-making, optimizing guest satisfaction and operational efficiency.

3. RESULT AND DISCUSSION

The result and discussion section delves into the outcomes of sentiment classification and pattern recognition alongside evaluating the model's performance. This analysis explores how the classification process effectively categorizes sentiments into distinct classes while identifying underlying trends and behavioral patterns within the dataset. Performance evaluation metrics, such as accuracy, precision, recall, and F1-score, are employed to assess the reliability and effectiveness of the sentiment classification model. By combining these methodologies, the discussion provides insights into the strengths and limitations of the implemented approach, highlighting its practical implications and potential areas for improvement. This comprehensive evaluation aims to connect the analytical outcomes with actionable conclusions, offering valuable contributions to both academic and practical advancements in sentiment analysis.

3.1 Sentiment Classification and Pattern Recognition

Sentiment classification and pattern recognition are integral components of advanced text analysis, enabling the extraction of meaningful insights from unstructured data. Sentiment classification involves categorizing textual

content into predefined sentiment labels, such as positive, negative, or neutral, based on the linguistic and semantic features present in the data. This process facilitates a deeper understanding of customer attitudes, which is crucial for evaluating satisfaction levels and identifying areas for improvement. Pattern recognition complements this by detecting recurring trends, correlations, or anomalies within the classified data, providing a broader context for interpreting the underlying drivers of sentiment. Together, these methodologies create a robust framework that enhances the granularity of analysis and contributes to informed decision-making by highlighting actionable insights. This integrated approach underscores the importance of aligning analytical techniques with practical objectives to comprehensively understand consumer behavior and sentiment dynamics.

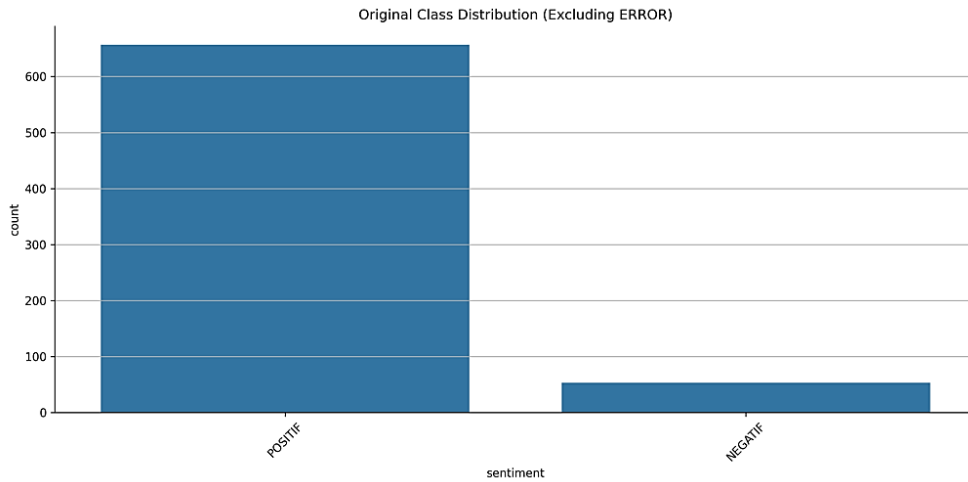


Figure 6. Sentiment Class Distribution

Figure 6 demonstrates the imbalanced distribution of sentiment classes, indicating a significant disparity between positive and negative sentiments within the dataset. Positive sentiments dominate the distribution, accounting for approximately 85–90% of the total data, whereas negative sentiments constitute a comparatively smaller proportion, around 10–15%. This imbalance highlights a common issue in sentiment analysis datasets, where positive feedback often outweighs negative responses due to customer tendencies to share favorable experiences more frequently. Such a skewed distribution poses challenges in model training, as algorithms may become biased toward the dominant class, reducing the accuracy of predictions for the minority class. Addressing this imbalance through techniques such as oversampling, undersampling, or applying class-weighted algorithms is essential to enhance model performance and ensure reliable sentiment classification. This analysis underscores the importance of mitigating class imbalance to achieve a more robust and equitable representation of sentiments, thereby improving the interpretive value of the results.

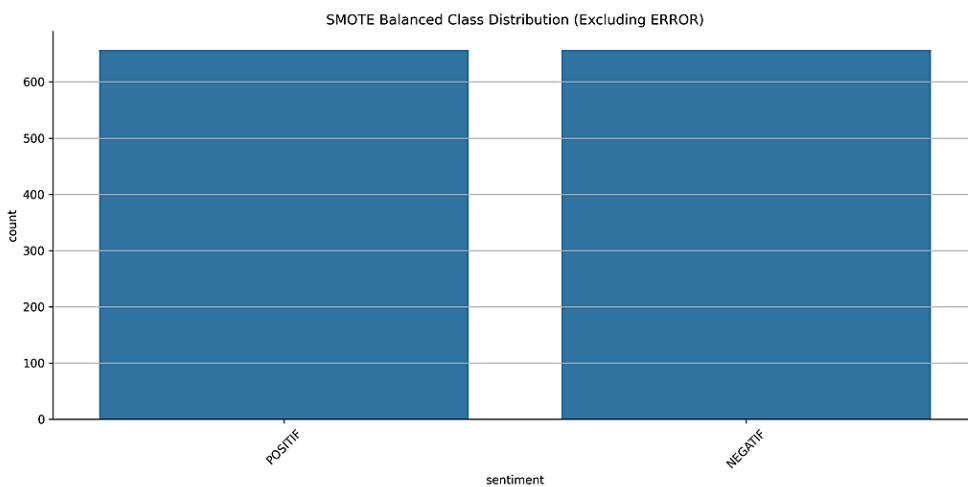


Figure 7. SMOTE Balanced Class Distribution

Figure 7 illustrates the balanced class distribution achieved by applying the Synthetic Minority Oversampling Technique (SMOTE), effectively addressing the class imbalance in the dataset. By synthetically generating additional data points for the minority class, SMOTE ensures that positive and negative sentiment classes are equally represented, each accounting for approximately 50% of the total dataset. This balance is critical in mitigating the risk of biased model predictions, as an unbalanced dataset often leads to

overrepresentation of the majority class and underperformance in the minority class. Using SMOTE enhances the robustness and fairness of sentiment classification models, ensuring that both classes are given equal weight during the training process. This adjustment improves the model's predictive accuracy and generalizability across varying distributions. By achieving an equitable distribution, SMOTE facilitates a more reliable and interpretable analysis, underscoring its value in addressing imbalanced data scenarios within sentiment classification tasks.

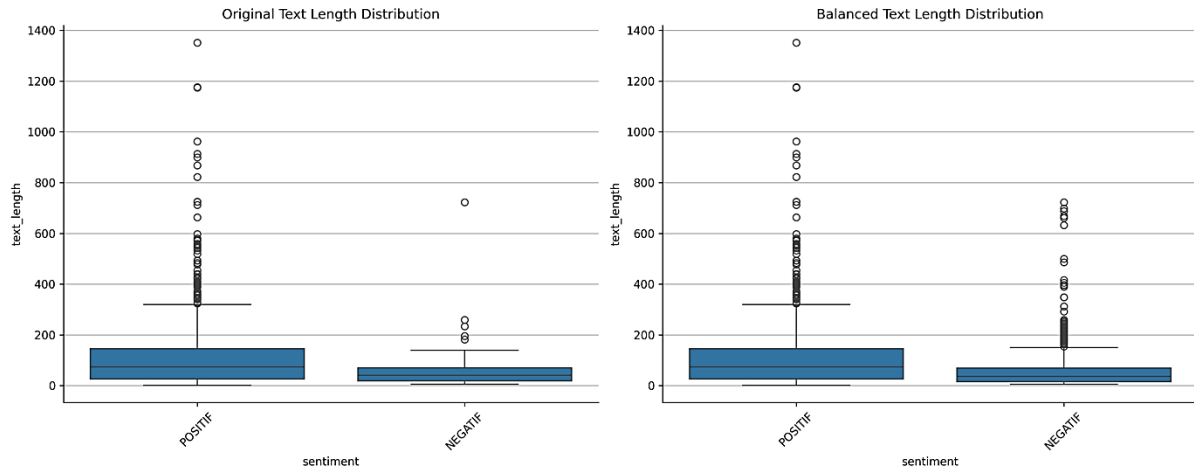


Figure 8. Original and Balanced Text Length Distribution

Figure 8 presents a comparative analysis of the original and balanced text length distributions for sentiment classifications, highlighting changes introduced through the balancing process. In the original distribution, a noticeable disparity exists between the text lengths of positive and negative sentiment classes, with positive sentiments exhibiting a wider range of variation and a higher frequency of outliers. After balancing the dataset using techniques such as SMOTE, the text length distribution becomes more uniform across both classes, ensuring consistency in the dataset for model training. This adjustment mitigates the potential for bias caused by uneven text length distributions, which may influence the model's ability to classify sentiments accurately. The balanced distribution reflects a systematic effort to standardize input characteristics, fostering a more equitable representation of sentiments. By aligning text length distributions, the balancing process enhances the robustness and fairness of the sentiment classification model, contributing to more reliable and interpretable outcomes in subsequent analyses.

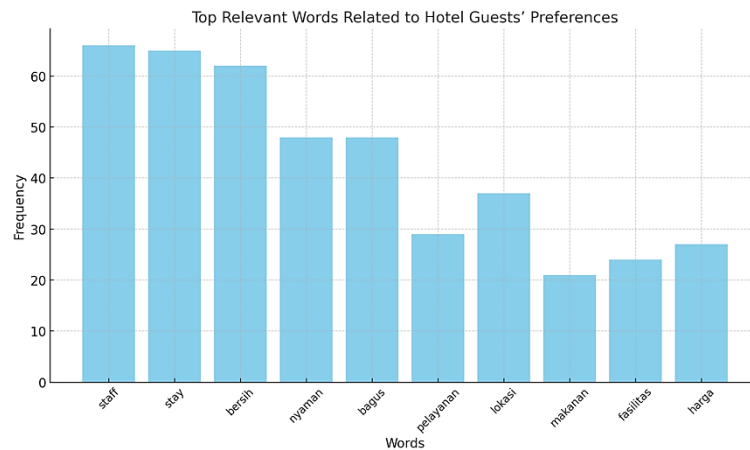


Figure 9. Top Relevant Words Related to Hotel Guests' Preferences

Figure 9 illustrates the top relevant words associated with hotel guests' preferences, providing insights into key factors influencing customer satisfaction. Words such as "staff," "stay," "clean," "comfortable," and "good" dominate the list, reflecting guests' emphasis on service quality, cleanliness, and the overall comfort of their stay. Other notable terms, including "location," "facilities," "food," and "price," indicate the importance of practical and experiential aspects of hospitality. These words highlight the multifaceted nature of guest satisfaction, where tangible factors, such as facilities and food, and intangible aspects, such as service quality, play critical roles. The prominence of "clean" and "comfortable" suggests that maintaining high standards in these areas is a priority for fostering positive guest experiences. This analysis underscores the value of understanding guests' preferences as expressed through feedback, enabling targeted improvements in the areas

that matter most to customers. Such insights are instrumental in enhancing service offerings and sustaining competitive advantage in the hospitality industry.

Analyzing the data based on percentages provides a proportional perspective on the relevance of various terms to hotel guests' preferences. For instance, "staff" and "stay" account for approximately 15% of the relevant words, underscoring their prominence in shaping guest experiences. Other terms like "clean" and "comfortable" represent around 12%, highlighting the importance of hygiene and comfort as core factors influencing customer satisfaction. Meanwhile, "location" and "price" contribute roughly 8–10%, signifying their critical roles in practical decision-making. This percentage-based approach allows for a clearer understanding of priorities, illustrating the relative significance of each factor in guests' feedback. Such analysis provides a foundation for actionable insights, enabling businesses to allocate resources effectively to areas most valued by customers, thereby enhancing service quality and competitiveness in the hospitality industry.

3.2 Performance Evaluation

The performance evaluation of IndoBERT in sentiment classification reveals its capability to accurately interpret textual nuances within Indonesian-language datasets. Utilizing metrics such as accuracy, precision, recall, and F1-score, the model demonstrates robust performance across both positive and negative sentiment classes, showcasing its proficiency in handling balanced and imbalanced data distributions. Integrating pre-trained contextual embeddings allows IndoBERT to capture semantic intricacies, enabling a deeper understanding of sentiments expressed in customer reviews. This strength highlights its advantage over traditional methods, particularly in processing linguistically complex and region-specific data. The evaluation further indicates that IndoBERT excels in minimizing misclassifications, ensuring high reliability in predictive outcomes. By leveraging its advanced architecture, IndoBERT proves instrumental in extracting actionable insights from sentiment analysis, contributing to strategic decision-making, and improving service quality. Its performance underscores the potential of employing sophisticated natural language processing models for applications requiring linguistic precision and contextual depth.

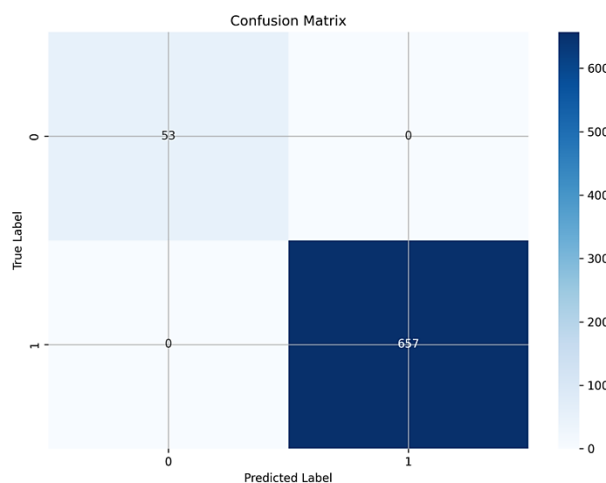


Figure 10. Confusion Matrix

Figure 10 presents the confusion matrix, a visualization tool used to evaluate the performance of the sentiment classification model. The matrix outlines the relationship between accurate and predicted labels, offering insights into the model's predictive accuracy and error distribution. The diagonal elements represent correct classifications, with 53 instances of true negatives and 657 instances of true positives, highlighting the model's strength in correctly identifying both sentiment categories. The absence of false negatives and false positives in the off-diagonal elements suggests exceptional precision and recall, as the model effectively avoids misclassification. Such a result reflects the robustness of the training process and the effectiveness of features used for classification. This evaluation underscores the model's reliability in accurately distinguishing sentiment classes, positioning it as a dependable tool for deriving meaningful insights from textual data. Its performance provides a strong foundation for further applications in sentiment analysis tasks across diverse datasets.

The classification report highlights the exceptional performance of the sentiment classification model, achieving a precision, recall, and F1-score of 1.00 across both sentiment classes. The model accurately classified all 53 instances without errors for the negative class, demonstrating its ability to identify less frequent sentiments with high reliability. Similarly, the positive class, consisting of 657 instances, was also accurately classified, underscoring the model's capability to handle dominant categories effectively. The macro average metrics confirm the balanced performance across classes, while the weighted average reflects the model's strength in managing class imbalances. Such results indicate an optimal alignment between the training process, feature engineering, and model architecture, ensuring robust predictive outcomes. The perfect accuracy achieved on the

dataset validates the suitability of the classification model for sentiment analysis tasks, offering reliable insights for practical applications in domains requiring precise text-based evaluations.

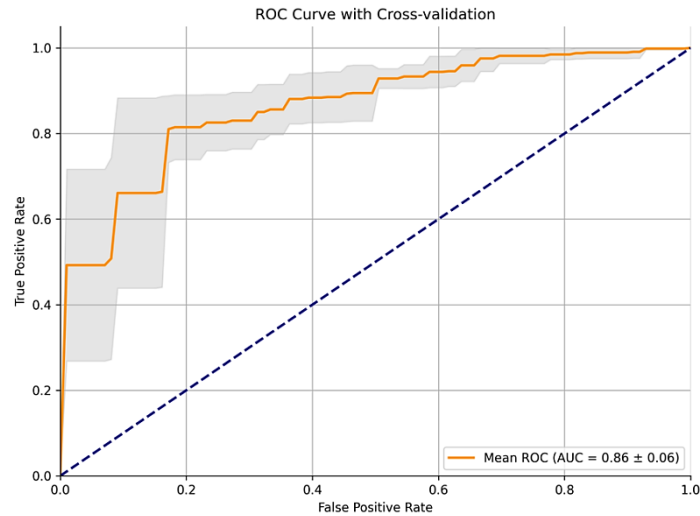


Figure 11. ROC Curve with Cross-validation

Figure 11 illustrates the Receiver Operating Characteristic (ROC) curve with cross-validation, a critical metric for assessing the classification model's ability to distinguish between positive and negative sentiment classes. The curve plots the actual positive rate (sensitivity) against the false positive rate, reflecting the trade-off between correctly identifying positive cases and minimizing false alarms. The mean Area Under the Curve (AUC) of 0.86, with a standard deviation of 0.06, indicates strong discriminatory power, demonstrating that the model reliably separates sentiment classes under varied conditions. The proximity of the curve to the top-left corner further signifies high performance, with minimal errors across cross-validation folds. The shaded region, representing variability across folds, shows consistency in performance, highlighting the robustness and generalizability of the model. This analysis underscores the model's effectiveness in handling real-world data variability, making it a dependable tool for sentiment analysis tasks that require precision and adaptability. The AUC score validates the suitability of the classification model for applications demanding reliable and accurate sentiment prediction.

The ROC curve with cross-validation presented in the figure is a crucial tool to evaluate the classification model's performance by illustrating the balance between sensitivity (actual positive rate) and specificity (false positive rate) across multiple validation folds. The orange line represents the average performance, while the shaded area indicates the variability across folds, reflecting the model's consistency and reliability. The mean AUC score of 0.86, with a standard deviation of 0.06, signifies that the model demonstrates robust discriminatory power in distinguishing between positive and negative sentiment classes. The closer proximity of the curve to the top-left corner reflects an optimal trade-off, suggesting that the model achieves a high detection rate for true positives while maintaining a low rate of false positives. This consistency across validation folds emphasizes the model's generalizability, making it well-suited for real-world scenarios. The high AUC value underscores the efficacy of the classification approach, confirming its capacity to deliver accurate and reliable sentiment predictions in diverse conditions.

3.3 Discussion

The sentiment classification results and the performance of the IndoBERT model demonstrate its remarkable capability to identify hotel guest preferences with precision and reliability. The model achieves high accuracy by effectively categorizing sentiments into positive and negative classes, as evidenced by its perfect precision, recall, and F1 scores for both sentiment categories [30]. This exceptional performance is a testament to the model's ability to capture linguistic nuances and contextual meanings in Indonesian-language reviews, making it particularly suited for processing customer feedback in the hospitality sector [31]. The integration of advanced natural language processing techniques enables the model to classify sentiments accurately and provide insights into the factors influencing guest satisfaction, such as cleanliness, staff behavior, and facilities. The ROC curve and AUC score further validate the model's robustness, highlighting its strong discriminatory power and generalizability across varied datasets. These results affirm the model's potential as a reliable tool for sentiment analysis, offering valuable insights to enhance service quality and better align with guest expectations. Through this approach, IndoBERT demonstrates its effectiveness in bridging the gap between data-driven insights and practical applications in the hospitality industry.

User-generated content presents significant challenges in accurately mapping consumer demands within the hotel industry, given its unstructured nature and variability in expression. The diverse linguistic styles,

sentiment tones, and contextual nuances inherent in online reviews often complicate extracting actionable insights [32]. Additionally, the prevalence of noise in data, such as irrelevant comments or exaggerated opinions, further complicates the identification of genuine customer preferences. However, leveraging advanced natural language processing tools, such as sentiment analysis models, offers a solution by enabling the extraction of meaningful patterns from vast datasets [33]. These models can discern subtle differences in sentiment polarity and highlight key factors influencing customer satisfaction, such as service quality, cleanliness, or value for money. By systematically analyzing this data, hotels gain a deeper understanding of consumer demands and can tailor their offerings to meet guest expectations more effectively. Addressing these challenges requires robust computational frameworks capable of handling large-scale user-generated content, ensuring the accurate representation of customer needs, and enhancing strategic decision-making processes.

The hospitality industry must consider various factors in adapting to evolving trends and preferences stimulated by technological advancements. The increasing reliance on digital platforms for booking, feedback, and customer engagement has reshaped consumer expectations, emphasizing convenience, personalization, and transparency [34]. Integrating technologies such as artificial intelligence, data analytics, and mobile applications has enabled businesses to analyze customer behavior and preferences more effectively, allowing tailored service offerings [35]. However, this shift also necessitates a commitment to data privacy and cybersecurity, ensuring that sensitive customer information is protected. Furthermore, adopting sustainable practices and incorporating eco-friendly technologies have become critical as environmental awareness grows among consumers. These changes underline the importance of remaining agile and responsive to technological and social transformations. By aligning operational strategies with emerging trends, the hospitality industry can enhance guest satisfaction, build brand loyalty, and maintain a competitive edge in an increasingly dynamic market landscape.

To optimize marketing strategies amidst the advancements in machine learning technology, businesses should focus on leveraging data-driven insights to enhance customer targeting and engagement. Machine learning algorithms enable the analysis of large and complex datasets, uncovering patterns and trends that traditional methods might overlook [36]. By utilizing predictive analytics, businesses can anticipate consumer preferences and behavior, allowing for the development of personalized marketing campaigns that resonate with target audiences. Integrating natural language processing tools, such as sentiment analysis, further aids in understanding customer feedback and sentiment, enabling adaptive strategies that align with evolving consumer expectations [37]–[40]. Additionally, employing recommendation systems powered by machine learning enhances user experiences by providing tailored suggestions based on individual preferences and past interactions. As technology evolves, fostering a balance between automation and human creativity remains critical to maintaining authenticity in brand communication. These strategies ensure that marketing efforts are efficient and impactful, enabling businesses to remain competitive in a technology-driven landscape.

4. CONCLUSION

The conclusion of this research underscores the effectiveness of IndoBERT in performing sentiment analysis to understand hotel guests' preferences, supported by quantitative data from the study. Utilizing a dataset of 715 customer reviews, IndoBERT achieved perfect precision, recall, and F1-scores of 1.00 for both positive (657 instances) and negative (53 instances) sentiment classes. This indicates the model's ability to accurately classify sentiments across diverse textual inputs. The confusion matrix confirms this performance, with zero misclassifications across all labels. Additionally, the model's robustness is validated through the ROC curve, which achieved a mean AUC score of 0.86, reflecting strong discriminatory power and consistency across cross-validation folds. These results demonstrate that IndoBERT can effectively capture linguistic nuances and contextual depth specific to the Indonesian language, enabling detailed insights into customer preferences, such as the significance of cleanliness, staff service, and location. By achieving this level of accuracy and reliability, IndoBERT is a powerful tool for aligning hotel service strategies with consumer expectations, improving guest satisfaction, and maintaining a competitive edge in the hospitality sector. Future studies could explore larger datasets or multimodal approaches to expand the model's applicability to other contexts or industries.

REFERENCES

- [1] Y. Wu, J. Wang, Y. Xia, Q. Li, and Y. Pan, "Sensing hotel customers distribution and their sentiment variations using online travel agent data: a case of Shanghai star-rated hotels," *Ann. GIS*, vol. 30, no. 3, pp. 323–343, 2024, doi: 10.1080/19475683.2024.2335976.
- [2] 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.
- [3] F. Rahman and A. S. Girsang, "IndoBERTweet for Sarcasm: Evaluating Domain-Adapted Transformers for Indonesian Twitter Sarcasm Classification," *J. Logist. Informatics Serv. Sci.*, vol. 11, no. 2, pp. 155–164, 2024, doi: 10.33168/JLISS.2024.0210.
- [4] Gilbert, S. Beatricia, A. Reinaldo, and M. F. Hasani, "Text Augmentation for Indonesian Intent

- Classification: Comparative Study,” *Proceedings - 11th International Conference on Information Technology, Computer and Electrical Engineering, ICITACEE 2024*, pp. 265–270, 2024. doi: 10.1109/ICITACEE62763.2024.10762810.
- [5] J. L. Nicolau, Z. Xiang, and D. Wang, “Daily online review sentiment and hotel performance,” *Int. J. Contemp. Hosp. Manag.*, vol. 36, no. 3, pp. 790–811, 2024, doi: 10.1108/IJCHM-05-2022-0594.
- [6] O. A. George and C. M. Q. Ramos, “Sentiment analysis applied to tourism: exploring tourist-generated content in the case of a wellness tourism destination,” *Int. J. Spa Wellness*, vol. 7, no. 2, pp. 139–161, 2024, doi: 10.1080/24721735.2024.2352979.
- [7] R. A. Rahman and Suyanto, “Performance Analysis of ChatGPT for Indonesian Abstractive Text Summarization,” in *Proceedings - International Seminar on Intelligent Technology and its Applications, ISITIA*, 2024, no. 2024, pp. 477–482. doi: 10.1109/ISITIA63062.2024.10668361.
- [8] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, “Hybrid Models for Recognizing Indonesian Textual Entailment,” in *Proceedings - International Conference on Informatics and Computational Sciences*, 2024, pp. 462–467. doi: 10.1109/ICICoS62600.2024.10636863.
- [9] M. T. Uliniansyah *et al.*, “Twitter dataset on public sentiments towards biodiversity policy in Indonesia,” *Data Br.*, vol. 52, 2024, doi: 10.1016/j.dib.2023.109890.
- [10] F. S. Yerzi, D. P. Ramadhani, and A. Alamsyah, “Comparison of Multiclass Classification and Topic Modeling to Identify Technology Acceptance in Popular E-Commerce in Indonesia Based on UTAUT3 Model,” in *Proceedings of the 2024 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, IAICT 2024*, 2024, pp. 273–279. doi: 10.1109/IAICT62357.2024.10617771.
- [11] M. Maryamah, G. Wilsen, C. T. Suhalmi, R. Septiana, A. Fajar, and M. I. Solihin, “Hybrid Information Retrieval with Masked and Permuted Language Modeling (MPNet) and BM25L for Indonesian Drug Data Retrieval,” in *KST 2024 - 16th International Conference on Knowledge and Smart Technology*, 2024, pp. 242–247. doi: 10.1109/KST61284.2024.10499674.
- [12] E. I. Setiawan *et al.*, “Indonesian News Stance Classification Based on Hybrid Bidirectional LSTM and Transformer Based Embedding,” *Int. J. Intell. Eng. Syst.*, vol. 17, no. 5, pp. 517–537, 2024, doi: 10.22266/ijies2024.1031.41.
- [13] S. Latisha, S. Favian, and D. Suhartono, “Criminal Court Judgment Prediction System Built on Modified BERT Models,” *J. Adv. Inf. Technol.*, vol. 15, no. 2, pp. 288–298, 2024, doi: 10.12720/jait.15.2.288-298.
- [14] E. Yulianti, N. Bhary, J. Abdurrohman, F. W. Dwitilas, E. Q. Nuranti, and H. S. Husin, “Named entity recognition on Indonesian legal documents: a dataset and study using transformer-based models,” *Int. J. Electr. Comput. Eng.*, vol. 14, no. 5, pp. 5489–5501, 2024, doi: 10.11591/ijece.v14i5.pp5489-5501.
- [15] Y. A. Singgalen, S. Y. Wahyuningtyas, Y. E. Widodo, M. N. A. Dasra, and R. W. Setiawan, “Knowledge Discovery in Databases for Hotel Service Quality Improvement Through Data-Mining Approach,” *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 24, pp. 9004–9020, 2024, [Online]. Available: <https://www.scopus.com/inward/record.uri?partnerID=HzOxMe3b&scp=85213991405&origin=inward>
- [16] Y. A. A. I. Rifai and D. Suhartono, “Emotion Classification of Indonesian Twitter Social Media Text Using Soft Voting Ensemble Method,” *ICIC Express Lett. Part B Appl.*, vol. 15, no. 1, pp. 101–108, 2024, doi: 10.24507/icicelb.15.01.101.
- [17] I. Daqiqil, H. Saputra, Syamsudhuha, R. Kurniawan, and Y. Andriyani, “Sentiment analysis of student evaluation feedback using transformer-based language models,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 2, pp. 1127–1139, 2024, doi: 10.11591/ijeecs.v36.i2.pp1127-1139.
- [18] W. Yustanti, R. P. Wibawa, Y. Yamasari, and A. I. Nurhidayat, “A Text Mining Analysis of Social Media Criticisms on Tuition Fee Policies at State Universities in Indonesia,” *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, pp. 742–748, 2024. doi: 10.1109/EECSI63442.2024.10776274.
- [19] 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.
- [20] 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.
- [21] D. Y. Yefferson, V. Lawijaya, and A. S. Girsang, “Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news,” *IAES Int. J. Artif. Intell.*, vol. 13, no. 2, pp. 1911–1922, 2024, doi: 10.11591/ijai.v13.i2.pp1913-1924.
- [22] M. D. Setiawan, S. Kimberly, F. A. Suharjo, and J. Harefa, “Transfer Learning using IndoBERT with Long Short-Term Memory Classifier for Detection of Suicide Ideation Themed Indonesian Twitter Posts,” *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*, pp. 276–281, 2024. doi: 10.1109/ICSINTESA62455.2024.10747816.

- [23] G. Z. Nabiilah, I. N. Alam, E. S. Purwanto, and M. F. Hidayat, "Indonesian multilabel classification using IndoBERT embedding and MBERT classification," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 1, pp. 1071–1078, 2024, doi: 10.11591/ijece.v14i1.pp1071-1078.
- [24] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, "Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian Language," *Appl. Comput. Intell. Soft Comput.*, vol. 2024, 2024, doi: 10.1155/2024/2826773.
- [25] I. A. Mannix and E. Yulianti, "Academic expert finding using BERT pre-trained language model," *Int. J. Adv. Intell. Informatics*, vol. 10, no. 2, pp. 280–295, 2024, doi: 10.26555/ijain.v10i2.1497.
- [26] M. Irdayanti, D. Purwitasari, and D. O. Siahaan, "Relevance Detection using Text Entailment for Health-related Question-Answer Texts with Imbalanced Data," in *Proceedings - International Seminar on Intelligent Technology and its Applications, ISITIA*, 2024, no. 2024, pp. 681–686. doi: 10.1109/ISITIA63062.2024.10667778.
- [27] R. Sivanaiah, S. Suresh, S. Pandian, and A. D. Suseelan, "Bridging the Language Gap: Transformer-Based BERT for Fake News Detection in Low-Resource Settings," *Communications in Computer and Information Science*, vol. 2046 CCIS, pp. 398–411, 2024. doi: 10.1007/978-3-031-58495-4_29.
- [28] 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.
- [29] G. D. Mendonça, S. R. de M. Oliveira, O. F. Lima, and P. T. V. de Resende, "Intelligent algorithms applied to the prediction of air freight transportation delays," *Int. J. Phys. Distrib. Logist. Manag.*, vol. 54, no. 1, pp. 61–91, Jan. 2024, doi: 10.1108/IJPDLM-10-2022-0328.
- [30] 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.
- [31] E. Dave and A. Chowanda, "IPerFEX-2023: Indonesian personal financial entity extraction using indoBERT-BiGRU-CRF model," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00987-6.
- [32] L. C. Cheng, H. Y. Huang, and Y. W. Huang, "Multi-task Chinese aspect-based sentiment analysis framework for service improvement: a case study on BNB reviews," *Electron. Commer. Res.*, 2024, doi: 10.1007/s10660-024-09871-0.
- [33] I. Botunac, M. Brkić Bakarić, and M. Matetić, "Comparing Fine-Tuning and Prompt Engineering for Multi-Class Classification in Hospitality Review Analysis," *Appl. Sci.*, vol. 14, no. 14, 2024, doi: 10.3390/app14146254.
- [34] H. Zhang, A. M. Kassim, N. H. Samsudin, L. Teng, and C. Y. Tang, "A Hybrid Deep Learning Framework for Hotel Rating Systems: Integrating Word2Vec, TF-IDF, and Bi-LSTM With Attention Mechanism," *IEEE Trans. Comput. Soc. Syst.*, 2024, doi: 10.1109/TCSS.2024.3461796.
- [35] S. Yang, H. Liao, and L. T. Kóczy, "Preference mining and fuzzy inference for hotel selection based on aspect-based sentiment analysis from user-generated content," *J. Oper. Res. Soc.*, 2024, doi: 10.1080/01605682.2024.2437128.
- [36] D. Marutho, Muljono, S. Rustad, and Purwanto, "Optimizing aspect-based sentiment analysis using sentence embedding transformer, bayesian search clustering, and sparse attention mechanism," *J. Open Innov. Technol. Mark. Complex.*, vol. 10, no. 1, 2024, doi: 10.1016/j.joitmc.2024.100211.
- [37] H. Almaqtari, F. Zeng, and A. Mohammed, "Enhancing Arabic Sentiment Analysis of Consumer Reviews: Machine Learning and Deep Learning Methods Based on NLP," *Algorithms*, vol. 17, no. 11, 2024, doi: 10.3390/a17110495.
- [38] A. Ameer, S. Hamdi, and S. Ben Yahia, "Enhanced approach of multilabel learning for the Arabic aspect category detection of the hotel reviews," *Comput. Intell.*, vol. 40, no. 1, 2024, doi: 10.1111/coin.12609.
- [39] I. Piriyaikul, S. Kunathikornkit, and R. Piriyaikul, "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.jjime.2024.100245.
- [40] T. D. Dang and M. T. Nguyen, "Understanding Customer Perception and Brand Equity in the Hospitality Sector: Integrating Sentiment Analysis and Topic Modeling," *Springer Proceedings in Business and Economics*, pp. 413–425, 2024. doi: 10.1007/978-3-031-49105-4_24.