Journal of Information Systems and Informatics
Vol. 6, No. 4, December 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i4.926 Published By DRPM-UBD

Sentiment Analysis and Trend Mapping of Hotel Reviews Using LSTM and GRU

Yerik Afrianto Singgalen[1*]

[1*]Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: [1*]yerik.afrianto@atmajaya.ac.id

Abstract

This study explores applying Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models for sentiment analysis and trend mapping of hotel reviews, specifically focusing on customer feedback from Hotel Vila Ombak in Lombok, Indonesia. The primary objective was to leverage these advanced deep learning models to capture nuanced sentiment patterns in unstructured textual data, enhancing insights into guest satisfaction. The analysis was conducted on a dataset of 326 reviews, achieving an overall model accuracy of 91% (0.91). The results showed that while the models excelled in identifying positive sentiments, with a precision of 0.94, recall of 0.98, and F1-score of 0.96, they struggled with minority classes. Both negative and neutral sentiments exhibited 0% accuracy, primarily due to the dataset's imbalance, where positive reviews constituted 92.3% of the total entries. The macro average metrics (precision 0.31, recall 0.33, F1-score 0.32) highlighted the model's limitations in classifying sentiments less frequently despite high weighted averages driven by the dominant positive class. This research underscores the need to address data imbalance and suggests that future studies incorporate techniques like data augmentation or hybrid models to improve performance across all sentiment categories. By optimizing sentiment analysis models, hospitality businesses can gain deeper insights into customer feedback, ultimately enhancing service quality and customer satisfaction.

Keywords: LSTM; GRU; Sentiment Mapping; Trend; Hotel

1. INTRODUCTION

In the era of digital communication, the proliferation of online reviews has emerged as a pivotal source of consumer feedback, particularly within the hospitality industry. Harnessing the insights embedded in these reviews offers valuable opportunities for businesses to enhance their services [1], [2]. However, this data's vast volume and unstructured nature pose significant challenges in deriving actionable insights. Recent advancements in deep learning, specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models have demonstrated their efficacy in processing sequential data, making them ideal for sentiment mapping and trend analysis of textual reviews [3], [4]. These models

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

capture long-term dependencies within text, allowing for a more nuanced understanding of sentiment shifts over time. Despite traditional approaches in sentiment analysis, such as lexicon-based or rule-based techniques, recurrent neural networks, particularly LSTM and GRU, provide superior capabilities in managing contextual nuances and temporal sequences inherent in customer feedback [5]. Applying these models makes identifying emerging trends and subtle patterns in guest sentiments feasible, which may go unnoticed through conventional analytical methods. Such analytical precision informs strategic decision-making and improves customer satisfaction and competitive positioning within the market. Thus, integrating LSTM and GRU models into sentiment analysis frameworks signifies a progressive step toward a data-driven approach in optimizing service quality within the hospitality sector.

Addressing the intricacies of customer feedback in the hospitality sector is becoming increasingly imperative, especially as businesses seek to maintain competitiveness in an environment driven by digital interactions. The continuous influx of unstructured data from online reviews underscores the urgency for robust analytical models capable of extracting meaningful insights from vast textual content [6]–[8]. Employing advanced techniques such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models facilitates a more precise understanding of customer sentiment [9]–[12]. It allows for detecting trends that may influence strategic decisions. Failing to leverage these sophisticated models' risks overlooking critical patterns in consumer perceptions, which may lead to missed opportunities for service enhancement. By implementing LSTM and GRU architectures, capturing the temporal dynamics and context-specific nuances of guest reviews becomes feasible, thus providing a more accurate assessment of evolving customer preferences. Thus, this research is vital as it contributes to theoretical advancements in sentiment analysis and has practical implications for optimizing service delivery and customer satisfaction in an increasingly competitive market.

The primary objective of this study is to implement Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models to conduct sentiment mapping based on customer reviews, with a particular focus on the feedback from guests at Hotel Vila Ombak in Lombok, Indonesia. This initiative seeks to leverage the advanced capabilities of deep learning algorithms to capture intricate patterns and sentiments embedded in textual data, often overlooked by conventional analytical techniques. Applying these recurrent neural network models is anticipated to provide a more refined understanding of guest experiences, enabling a comprehensive sentiment analysis beyond surface-level interpretations [13]. By focusing on a specific hospitality setting, the research not only aims to demonstrate the effectiveness of LSTM and GRU in extracting meaningful insights from customer feedback but also underscores the potential for these models to enhance strategic decision-making processes within the hotel industry [14]. Such an

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

approach offers practical value, as it informs data-driven strategies that can significantly enhance service quality, customer satisfaction, and overall competitiveness in a rapidly evolving tourism market.

As an advanced computational technique, Sentiment analysis has become a cornerstone in understanding consumer feedback, particularly within the hospitality industry. Its application enables the extraction of valuable insights from unstructured textual data, such as online reviews, facilitating the identification of trends and sentiment patterns critical to improving service quality. Employing models like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) enhances this process by leveraging their ability to capture temporal dependencies and contextual nuances, which traditional methods often fail to address. These deep learning architectures significantly improve the precision of sentiment mapping, especially in handling sequential data, although challenges persist, particularly in classifying less frequent sentiment categories due to dataset imbalances. Addressing these limitations through innovative preprocessing techniques or model enhancements is essential for a more comprehensive analysis. By refining sentiment analysis methodologies, this approach contributes to theoretical advancements in computational linguistics and underscores its practical significance in driving service optimization within a competitive market landscape.

This research offers a significant theoretical contribution by expanding the application of deep learning models, specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), within the domain of sentiment analysis. The study enriches existing literature on using artificial intelligence in natural language processing by demonstrating how these models effectively capture complex patterns in unstructured textual data [15], [16]. Integrating advanced neural network architectures into sentiment mapping highlights the ability to analyze temporal dynamics and contextual subtleties in customer reviews, thus addressing limitations in traditional analytical methods [17]–[19]. From a practical standpoint, the findings provide valuable insights for the hospitality industry, particularly in leveraging data-driven strategies to enhance customer satisfaction [20]–[22]. The successful application of these models to real-world data, such as reviews from Hotel Vila Ombak, establishes a robust framework that can guide businesses in optimizing service delivery and tailoring their offerings to meet customer expectations more effectively. Ultimately, the research bridges the gap between theoretical advancements in machine learning and tangible improvements in business practices, positioning sentiment analysis as a pivotal tool in achieving competitive advantage.

In recent years, studies have increasingly explored using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models for analyzing customer feedback in the hospitality sector, mainly through online hotel reviews. These deep learning techniques have been recognized for their capability to manage sequential

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

data, making them particularly effective in deciphering the nuances of sentiment embedded within unstructured text [23], [24]. Existing research has demonstrated that applying LSTM and GRU models enhances sentiment classification accuracy compared to traditional machine learning algorithms due to their proficiency in capturing long-term dependencies and contextual nuances in customer feedback [25], [26]. Analyzing sentiment trends using these models reveals shifts in customer perceptions and uncovers hidden patterns that might remain unnoticed through conventional approaches. Such studies emphasize that leveraging LSTM and GRU architectures for sentiment analysis can provide actionable insights, ultimately enabling hotels to refine their customer engagement strategies. This body of work highlights the importance of deep learning methodologies in advancing academic understanding and practical applications in customer experience management.

Despite its contributions, this study encounters several limitations that warrant consideration. One of the primary constraints lies in the reliance on a specific dataset, namely customer reviews from Hotel Vila Ombak, which may limit the generalizability of the findings to other contexts within the hospitality industry. Focusing on a single case study could introduce biases that do not fully capture the diversity of customer sentiments across different regions and hotel categories. Additionally, while LSTM and GRU models are highly effective in processing sequential data, they are computationally intensive and demand substantial resources, which may pose challenges for organizations with limited technological infrastructure. The models' reliance on large datasets for optimal performance also implies that the results are contingent on the quality and volume of available data, potentially leading to skewed insights if the dataset is unbalanced or incomplete. Moreover, the interpretability of deep learning models remains an ongoing challenge, as the complexity of these algorithms often obscures the rationale behind specific predictions. Addressing these limitations in future research may enhance the robustness and applicability of sentiment analysis models in diverse real-world settings.

2. METHODS

2.1 Long Short-Term Memory (LSTM)

Implementing Long Short-Term Memory (LSTM) models involves several systematic stages crucial to achieving accurate and meaningful results in sentiment analysis. Data preprocessing is initially essential, wherein raw textual data undergoes cleaning, tokenization, and transformation into numerical representations, often through word embeddings. This step ensures that the input data is structured in a format suitable for machine learning models, enhancing the model's capacity to understand the underlying patterns in the text. Once the data is prepared, the architecture of the LSTM network is constructed, including defining the number of layers, neurons, activation functions, and optimization

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

algorithms. This stage is critical, as the network's design significantly influences its ability to capture temporal dependencies within sequential data. The training phase follows, wherein the model learns to identify patterns by iteratively adjusting weights based on the error calculated from predictions. Given the complexity of LSTM networks, this process can be computationally intensive and requires fine-tuning of hyperparameters to prevent issues such as overfitting. After training, the model is evaluated using performance metrics like accuracy, precision, recall, and F1-score to assess its effectiveness in sentiment classification. The final stage involves deploying the model to classify new data, allowing for real-time sentiment analysis that can inform strategic business decisions. Each stage is integral to leveraging the full potential of LSTM models in capturing the intricacies of customer sentiment from unstructured textual data.

Selecting specific hyperparameters in machine learning models such as LSTM and GRU is crucial to achieving optimal performance in sentiment analysis tasks. Hyperparameters like the number of layers, learning rate, batch size, and dropout rate were chosen based on their impact on the models' ability to generalize and prevent overfitting. For instance, a learning rate was carefully adjusted to balance convergence speed and accuracy, ensuring the model neither converged prematurely nor diverged. Dropout layers were introduced to mitigate overfitting by randomly deactivating neurons during training, thus enhancing the model's robustness. Batch size was optimized to balance computational efficiency and gradient stability, ensuring effective learning.

Challenges arose during training, particularly regarding overfitting the dominant class due to an imbalanced dataset. This issue was addressed by implementing class-weight adjustments to give more significance to underrepresented classes, encouraging the model to learn their patterns effectively. Additionally, variations in performance across different folds in cross-validation indicated sensitivity to data splits. This was resolved by stratified sampling, ensuring a proportionate representation of all sentiment classes in each fold. Furthermore, computational inefficiencies during training were tackled using early stopping criteria to halt training once the validation loss plateaued, saving resources while maintaining model performance. These strategies collectively ensured a more robust, efficient, and equitable model capable of delivering reliable sentiment analysis results.

Figure 1 illustrates the architecture of an LSTM model, encapsulating the critical stages required to transform raw textual data into meaningful sentiment predictions. The process begins with data preprocessing, where the text undergoes cleaning, tokenization, and conversion into numerical representations to prepare it for analysis. Subsequently, the architecture of the LSTM network is constructed, involving the addition of layers that capture temporal dependencies within sequential data. This phase is pivotal, as the network design, including the selection of activation functions and the incorporation of dropout layers, directly influences

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

the model's ability to generalize from the training data. Following the model construction, a compilation step is necessary to define the loss function, optimization algorithm, and evaluation metrics, ensuring the model learns effectively during training. The training process involves iterative adjustments of model weights to minimize errors, while the evaluation phase assesses the model's performance on unseen data using metrics such as precision, recall, and F1-score. Finally, once the model performs satisfactorily, it is deployed for real-time sentiment analysis, allowing it to classify new data based on patterns learned during training. This architecture enhances the model's capacity to interpret complex sentiment patterns and supports more informed decision-making within the context of customer feedback analysis.

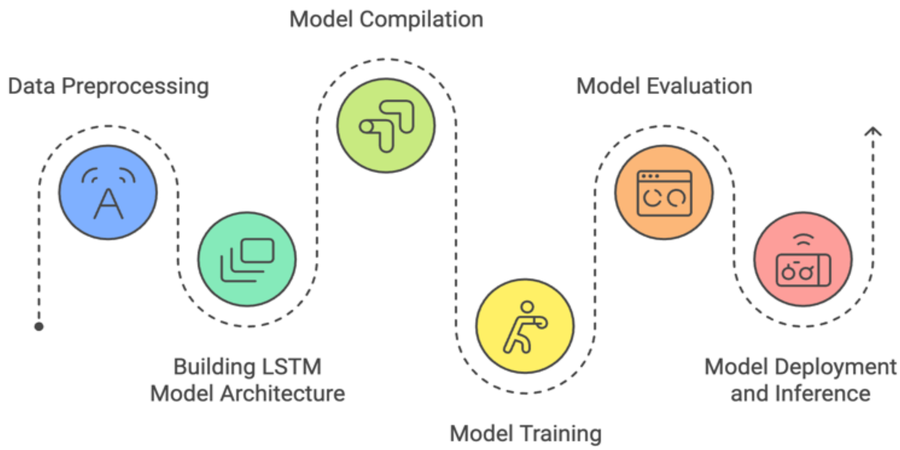

Figure 1. LSTM Model Architecture

The utilization of Long Short-Term Memory (LSTM) networks is particularly relevant to the current research, given its proven capability to handle sequential data with complex temporal dependencies. LSTM's architecture, equipped with forget, input, and output gates, allows it to selectively retain or discard information, thereby addressing challenges such as the vanishing gradient problem that typically hampers standard recurrent neural networks. In analyzing customer reviews, which often involve nuanced sentiment expressions and context-specific language, LSTM is adept at capturing long-term dependencies, making it highly suitable for sentiment mapping. The ability of LSTM models to effectively learn patterns from sequential text data ensures a more accurate sentiment analysis, even when handling extensive datasets containing mixed sentiments. This relevance becomes even more pronounced in hospitality studies, where understanding guest feedback over time is crucial for optimizing service quality. By leveraging LSTM, this research gains a robust methodological foundation to extract deeper insights from

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

textual data, thus enabling more informed strategic decision-making and continuous improvement in customer satisfaction.

2.2 Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is an advanced type of recurrent neural network (RNN) designed to handle sequential data with improved efficiency and reduced computational complexity. Unlike traditional RNNs, GRU introduces gating mechanisms, specifically the update gate and reset gate, that control the flow of information, thereby addressing the vanishing gradient problem commonly encountered in sequence modeling. GRUs effectively capture long-term dependencies without the extensive computational demands of Long Short-Term Memory (LSTM) networks by regulating which parts of past information are retained and discarded. This streamlined architecture accelerates training and enables GRUs to perform competitively in text analysis, time-series forecasting, and natural language processing scenarios. Despite the relative simplicity of its structure compared to LSTM, the GRU's ability to maintain critical information over time while efficiently filtering out irrelevant data makes it particularly well-suited for tasks requiring rapid learning and real-time analysis. Thus, the GRU model offers a compelling balance between computational efficiency and accuracy, making it a valuable tool in deep learning for sequential data.

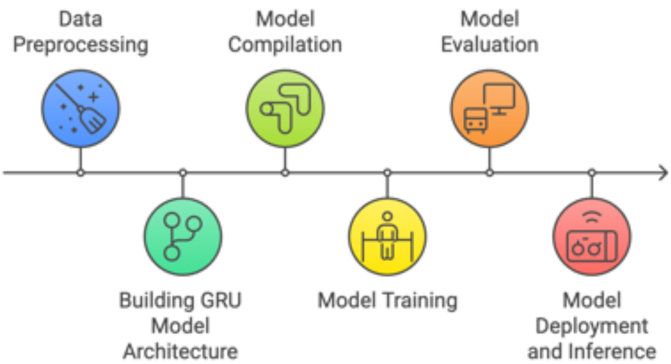

Figure 2. GRU Model Architecture

Figure 2 depicts the architecture of a Gated Recurrent Unit (GRU) model, outlining the essential stages required to transform raw data into actionable insights. The initial phase involves data preprocessing, where unstructured textual content is systematically cleaned, tokenized, and converted into numerical representations. This step is crucial to ensure the data is appropriately structured for subsequent analysis. Following preprocessing, the model architecture is constructed, incorporating GRU layers known for their efficiency in managing sequential dependencies through update and reset gates. These gates optimize the model's ability to retain relevant information while discarding less critical data, enhancing training efficiency and model accuracy. After the architecture is defined, the model undergoes compilation, where critical components such as the loss function, optimization algorithm, and performance metrics are specified. This is followed by the training phase, during which the model iteratively adjusts its parameters to minimize prediction errors. The evaluation stage then assesses the model's performance on unseen data using metrics like precision and recall. Finally, once the model achieves satisfactory accuracy, it is deployed for real-time analysis, allowing for the extraction of insights from new data streams. This systematic approach ensures that the GRU model effectively captures and leverages temporal patterns within the data, making it an invaluable tool in sentiment analysis and other time-dependent applications.

The relevance of Gated Recurrent Unit (GRU) models in this study is underscored by their efficiency in processing sequential data, which is critical for analyzing hotel reviews that exhibit varying sentiment levels over time. GRU models incorporate gating mechanisms, such as the update and reset gates, that enable the model to selectively retain relevant information while efficiently discarding redundant data, thus reducing the risk of information loss. Given the context of sentiment analysis for hotel reviews at Vila Ombak, GRU's simplified architecture, compared to Long Short-Term Memory (LSTM) networks, allows for faster training while maintaining robust performance. This attribute is especially beneficial when dealing with extensive datasets, where computational efficiency without compromising accuracy is crucial. The ability of GRU models to capture temporal dependencies within textual data enhances the precision of sentiment mapping, allowing for deeper insights into customer perceptions and experiences. Therefore, integrating GRU in this research offers a balanced approach to leveraging deep learning for sentiment analysis, aligning well with the study's objective of optimizing customer feedback analysis to enhance service quality.

The preprocessing of the dataset involved several critical steps to prepare the textual data for practical sentiment analysis. Text tokenization was implemented to break down customer reviews into individual words or tokens, allowing the model to process the text as sequential data. This step included cleaning the text by removing special characters, punctuations, and stopwords to focus on meaningful content. Each token was then converted into numerical representations using word embedding techniques, such as Word2Vec or GloVe, which map words to dense vector spaces that capture semantic relationships. Word embedding was chosen for its ability to retain contextual information and semantic similarity between words, which is crucial for understanding nuanced sentiments in hotel reviews. This approach ensured the input data was compact and informative, enhancing the model's ability to learn patterns effectively.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The performance of the LSTM and GRU models was evaluated using metrics such as precision, recall, F1 score, and accuracy. These metrics were selected to assess the models' performance across different sentiment classes comprehensively. Precision and recall were prioritized to understand the models' ability to correctly identify positive, negative, and neutral sentiments, while the F1-score balanced these metrics to evaluate performance under class imbalance. Although a standard metric, accuracy was complemented by macro and weighted averages to provide insights into how well the models performed across both dominant and minority classes.

The use of 5-fold cross-validation was integral to ensuring the robustness of the evaluation process. By dividing the dataset into five subsets, the models were trained and validated on different data splits, reducing the risk of overfitting and ensuring that a particular train-test split did not bias the results. This approach also provided a more reliable estimate of the models' generalization capabilities, particularly given the imbalanced dataset. Cross-validation further highlighted the variability in performance across different folds, offering critical insights into the models' sensitivity to data distribution and guiding the tuning of hyperparameters for optimal results. This rigorous preprocessing and evaluation pipeline ensured the findings were reliable and applicable to real-world sentiment analysis scenarios.

2.3 Vila Ombak Hotel

Vila Ombak Hotel, a distinguished 4-star establishment, offers an inviting sanctuary on the picturesque island of Gili Trawangan, Lombok. Renowned for its serene ambiance, the hotel provides a harmonious blend of comfort and modern luxury. It features spacious rooms with contemporary amenities and a private beach that ensures a tranquil retreat. An indoor pool adds to the allure, enhancing the overall guest experience, particularly for solo travelers seeking relaxation and an immersive getaway. Its strategic location near key landmarks, vibrant restaurants, and nightlife hotspots positions the hotel as an ideal choice for visitors desiring seamless access to local attractions while maintaining a sense of seclusion. Such a balance between accessibility and tranquility creates a compelling value proposition, fostering an environment where guests may unwind while exploring the cultural and natural beauty of the island. Thus, the Vila Ombak Hotel is a quintessential destination for those pursuing adventure and peaceful solitude.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 3. Hotel Facilities (Agoda: Vila Ombak Hotel)

Figure 3 showcases the diverse range of facilities available at Vila Ombak Hotel, reflecting its commitment to providing a luxurious and relaxing guest experience on Gili Trawangan. The aerial view captures the expansive layout of the hotel grounds, featuring lush greenery and a spacious swimming pool area, which serves as a central attraction for guests seeking relaxation under the tropical sun. The image highlighting the beachfront setting, adorned with a heart-shaped arch, also exemplifies the hotel's appeal for romantic getaways, particularly among couples. Such aesthetically pleasing amenities not only enhance the visual allure of the property but also create memorable experiences, aligning with the expectations of travelers seeking both comfort and picturesque scenery. By integrating these elements into its facilities, Vila Ombak caters to various guest preferences, from tranquil retreats to scenic beachfront moments. This thoughtful combination of natural beauty and modern conveniences plays a pivotal role in sustaining high guest satisfaction, reinforcing the hotel's reputation as a top-tier destination on the island.

The ratings and reviews for a hotel, exceptionally when aggregated from verified guests, provide a crucial indicator of the establishment's service quality and overall guest satisfaction. Vila Ombak Hotel's cumulative rating of 8.2 out of 10, based on over 2,800 reviews from platforms like Agoda and Booking.com, reflects a generally positive reception among its clientele. Guests have commended the hotel's strategic location with a high score of 9.1, alongside favorable ratings for facilities and service, each receiving scores above 8.0. However, areas such as room comfort and value for money, with ratings of 7.9 and 7.8, respectively, indicate potential areas for improvement. The ratings distribution, where a substantial portion of reviews fall into the "Exceptional" and "Excellent" categories, underscores the hotel's capacity to meet and sometimes exceed customer expectations. Nonetheless, a segment of reviews rated below 6 suggests occasional lapses in service quality or unmet expectations. Such reviews, especially when verified, offer valuable feedback for continuous improvement, enabling the hotel to align its offerings more closely with guest preferences and enhance its competitive standing in the hospitality sector.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

The volume of review data and associated metadata intended for processing in this study is substantial and diverse, offering a rich foundation for in-depth analysis. With a dataset encompassing thousands of reviews, complete with ratings, guest profiles, room types, and stay durations, the scope of analysis extends beyond mere sentiment extraction to include insights into guest behavior and preferences. Such a comprehensive dataset allows for a robust application of advanced machine learning models, such as LSTM and GRU, and supports identifying nuanced patterns in customer satisfaction levels over time. Including metadata, such as review sources, stay dates, and guest demographics, enhances the contextual understanding of sentiment trends, enabling a more tailored approach to service enhancement. Leveraging this extensive dataset to uncover actionable insights is expected to significantly contribute to optimizing the hotel's service offerings and strategic decision-making, ultimately driving improvements in guest satisfaction and loyalty.

3. RESULTS AND DISCUSSION

3.1 Vila Ombak Hotel

Vila Ombak Hotel, a distinguished 4-star resort on the idyllic island of Lombok, Indonesia, presents an exquisite retreat for leisure and business travelers. With its breathtaking vistas, luxurious amenities, and renowned hospitality, this establishment offers a harmonious blend of comfort and elegance. Located approximately 30 kilometers from the city center, the hotel serves as a serene escape, enveloped by lush tropical landscapes and pristine beaches, which create an ideal setting for relaxation and adventure. Catering to diverse guest needs, the property features 150 meticulously designed rooms, each furnished with modern comforts to ensure a tranquil and restorative stay. The hotel's commitment to family-friendly service is evident in its generous child policy. It allows children aged 3 to 11 to stay free, making it particularly attractive for families seeking an unforgettable vacation. Built in 1998 and tastefully renovated in 2012, Vila Ombak fuses traditional Indonesian architecture with contemporary design elements, fostering a welcoming and sophisticated ambiance. Whether indulging in a rejuvenating spa experience, lounging by the pool, or exploring the island's natural wonders, guests are enveloped in the tranquility and allure of Lombok. With check-in available at 2:00 PM and check-out until noon, visitors can fully immerse themselves in the experience, leaving with cherished memories and a deep appreciation for this tropical haven.

Figure 4 illustrates the distribution of visitors to Vila Ombak Hotel based on their country of origin, with data derived from 322 guest accounts. The chart reveals that Australia, Hong Kong, and the United Kingdom represent the most significant segments of the hotel's clientele. This pattern suggests a strong presence of travelers from these regions, possibly driven by targeted marketing strategies or

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

established travel preferences for destinations like Gili Trawangan. Meanwhile, a noticeable influx from New Zealand indicates its significance as another key market. In contrast, the relatively lower representation of guests from South Korea, the United States, and other countries suggests untapped potential for expanding outreach efforts. The diversity of guest origins reflected in this chart underscores the hotel's appeal across various regions while highlighting areas where strategic promotional initiatives could increase visitation. Such insights are essential for refining marketing tactics, optimizing service offerings, and enhancing customer experiences tailored to cultural expectations.

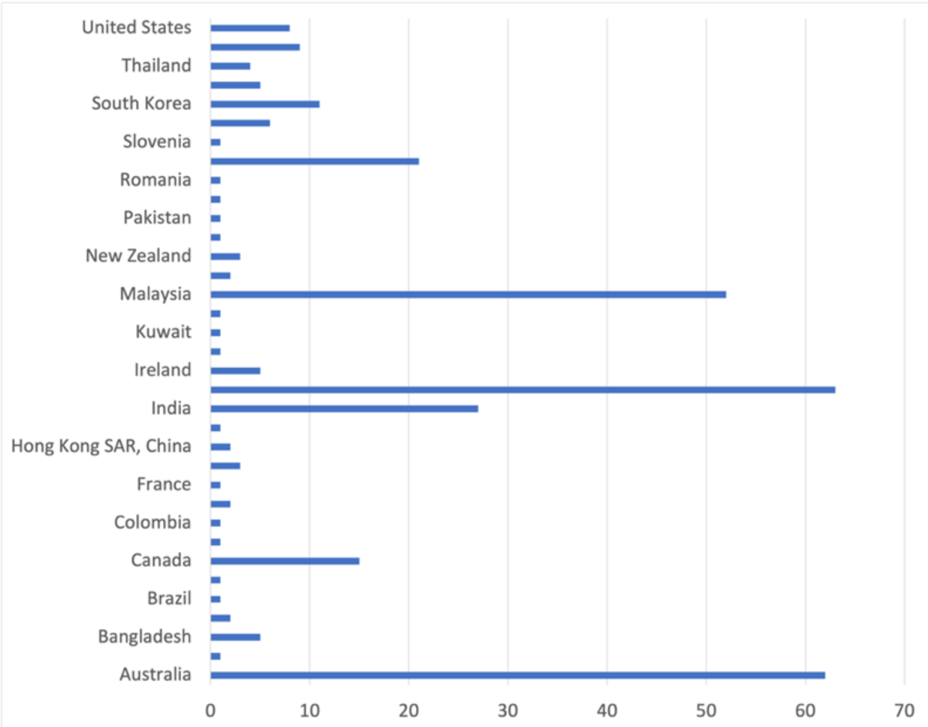

Figure 4. The Visitor Country of Origin (322 Accounts)

Figure 5 provides an overview of the length of stay categorized by different types of guests based on data from 136 accounts. Couples appear to dominate the longer stay durations, particularly for stays of 2 to 3 nights, indicating a preference among this demographic for short getaways, likely motivated by the romantic appeal of Gili Trawangan as a travel destination. Families with teenagers also show a significant presence for 2 to 3 nights, suggesting that the hotel's amenities cater effectively to this group's needs. While fewer in number, Solo travelers display a tendency for shorter stays, typically just one or two nights, possibly reflecting a

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

pattern of quick visits as part of broader travel itineraries. The distribution highlights a sharp decline in guest numbers for stays exceeding 4 nights, regardless of the guest type, which may imply either a preference for shorter trips or limitations in the current service offerings for extended vacations. The hotel can better align its marketing efforts and tailor its packages to optimize occupancy rates across different guest segments by analyzing these patterns.

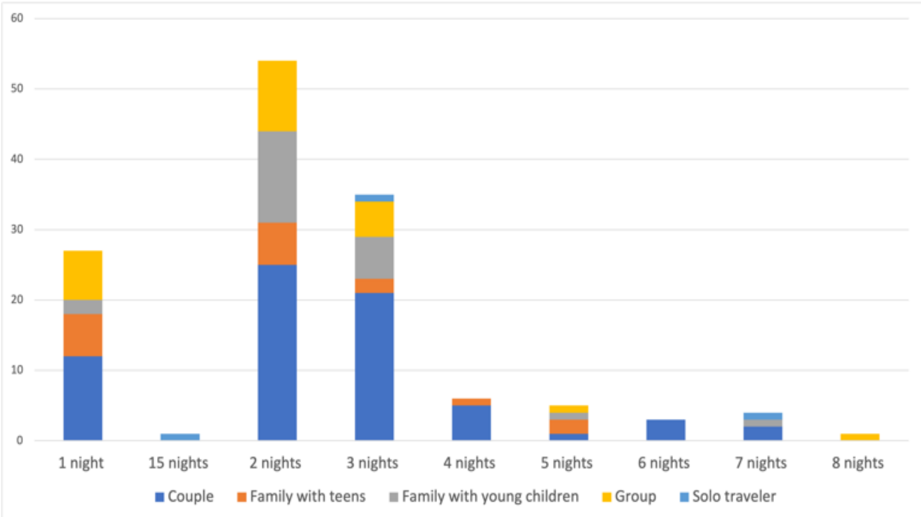

Figure 5. Length of Stay based on Guest Type (136 Accounts)

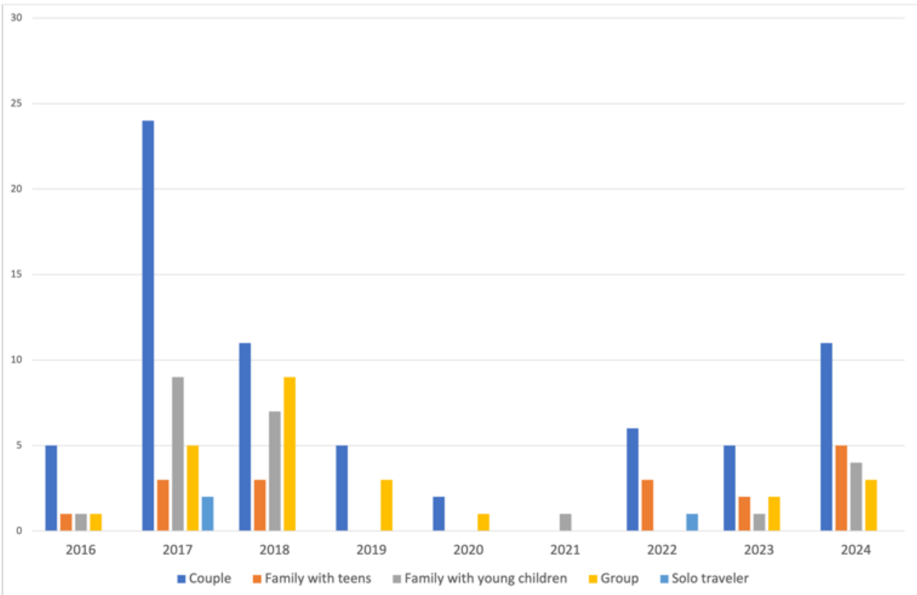

Figure 6. Year of Stay based on Guest Type (136 Accounts)

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Figure 6 provides an overview of the distribution of guest stays at Vila Ombak Hotel segmented by year and guest type, based on data from 136 accounts. The chart reveals a noticeable peak in 2017, particularly among couples, indicating a surge in visits during that period, possibly driven by effective marketing campaigns or favorable travel conditions. Families with teenagers and groups also showed significant activity this year, suggesting a broad appeal across different demographics. The subsequent years depict a gradual decline, with a notable dip around 2020, which aligns with global travel disruptions. However, a resurgence is visible in 2023, led primarily by couples, followed by families and solo travelers, reflecting a potential recovery in the tourism sector. The data indicates that couples consistently represent the largest segment across most years, underscoring the hotel's strong positioning as a romantic getaway destination. The fluctuations in guest demographics over the years emphasize the impact of external factors on travel trends, highlighting opportunities for targeted strategies to attract a diverse clientele in the future.

Based on the observed trends in tourist visits, it becomes apparent that both internal and external factors influence the preferences and behaviors of guests. The data indicates that couples consistently make up the most significant proportion of visitors, especially during peak travel years, highlighting the hotel's attractiveness as a romantic retreat. Families, particularly those with teenagers, also constitute a significant segment, which suggests that the property's amenities and location align well with the needs of multigenerational travelers. A noticeable decline in guest numbers around 2020 reflects the impact of global travel restrictions. Yet, the gradual recovery in subsequent years, especially among solo travelers and couples, signals a renewed interest in leisure travel as conditions improve. This resurgence implies that destinations like Gili Trawangan, known for their natural beauty and relaxed atmosphere, are becoming increasingly sought after in the post-pandemic era. Recognizing these patterns is crucial for devising targeted marketing strategies and optimizing service offerings to meet evolving guest expectations, ultimately enhancing customer satisfaction and market competitiveness.

3.2 LSTM and GRU in Sentiment Mapping

Leveraging Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models for sentiment mapping offers significant advancements in accurately analyzing unstructured textual data. Both models are specifically designed to capture temporal dependencies and contextual nuances, making them highly effective in processing sequences of customer reviews where sentiments are often implied rather than explicitly stated. LSTM networks utilize memory cells to retain essential information over longer sequences, thereby mitigating issues related to vanishing gradients. Meanwhile, with their streamlined architecture, GRU models achieve similar performance but with enhanced computational efficiency, making them ideal for scenarios requiring faster processing without sacrificing accuracy.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In the context of sentiment analysis, these models excel at discerning patterns in customer feedback, enabling a more profound understanding of guest satisfaction and emerging trends. By implementing LSTM and GRU for sentiment mapping, businesses are better equipped to refine their strategies, optimize customer experiences, and respond proactively to market demands, ultimately driving competitive advantage.

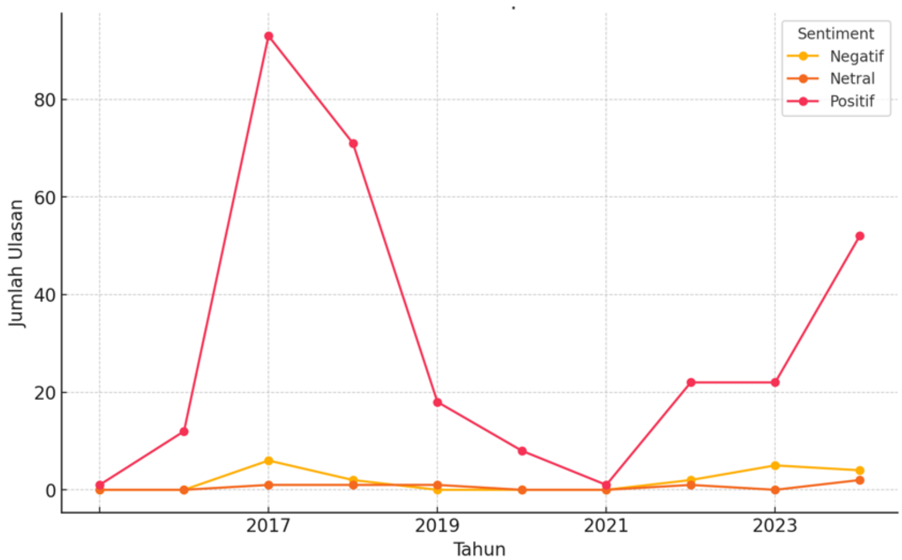

Figure 7. Sentiment Trend based on Year (326 Accounts)

Figure 7 illustrates the sentiment trends over several years based on an analysis of 326 accounts. The graph reveals that positive sentiments dominate the timeline, with a noticeable spike in 2017, followed by a gradual decline in subsequent years. This initial surge suggests a period of heightened guest satisfaction, potentially driven by improvements in service quality or successful marketing initiatives during that time. In contrast, negative and neutral sentiments remain relatively stable, with only minor fluctuations, indicating a generally consistent level of guest expectations being met. However, a slight uptick in positive sentiments from 2022 onwards signals a recovery phase, possibly correlating with the easing global travel restrictions and a renewed focus on enhancing customer experiences. The stability of neutral and negative feedback suggests that while specific challenges persist, the overall guest perception leans positively. These insights can guide strategic efforts to sustain and elevate customer satisfaction by addressing recurring concerns identified in the neutral or negative reviews.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Based on the implementation results of LSTM and GRU models, the provided visualizations reveal a nuanced performance across different evaluation metrics. The graph depicting model accuracy across folds illustrates significant fluctuations during the initial epochs, gradually stabilizing as the training progresses, with some folds achieving accuracy close to 1.0. This suggests that while the models can effectively learn patterns in the data, their convergence rate varies depending on the specific fold, indicating sensitivity to data splits. A general downward trend is observed in model loss, reflecting the models' improved performance over epochs; however, specific folds exhibit occasional spikes, which may indicate overfitting or challenges in generalizing across validation data. The final graph of mean validation accuracy highlights a consistent upward trend during the early epochs, followed by a plateau, suggesting that further training beyond this point yields diminishing returns in model performance. These results underscore the robustness of LSTM and GRU models in capturing complex temporal dependencies in the dataset while highlighting the need for careful hyperparameter tuning to optimize consistency across different data partitions.

The initial class distribution of the dataset reveals a significant imbalance, with positive sentiments comprising 92.3% of the entries, while damaging and neutral sentiments account for only 5.8% and 1.8%, respectively. This pronounced skewness toward the positive class suggests potential challenges in model training, as the overwhelming dominance of one class may hinder the model's ability to learn and predict the minority classes accurately. Imbalanced datasets often result in models biased toward the majority class, potentially overlooking or misclassifying instances of the less frequent categories. This discrepancy can compromise the model's generalization capabilities, particularly in real-world applications where detecting negative or neutral sentiments is crucial for nuanced sentiment analysis. Addressing this imbalance through techniques such as oversampling, undersampling, or employing specialized algorithms is essential to enhance the model's robustness and ensure a more equitable representation of all sentiment classes, ultimately leading to more accurate and actionable insights.

Implementing a 5-fold cross-validation process reveals considerable variability in model performance across the different folds. The recorded accuracy rates demonstrate substantial fluctuations, with Fold 2 achieving the highest accuracy at 92.31%, while Fold 3 exhibits a significant drop to 43.08%. Such disparities suggest that the model's performance is susceptible to the specific data subset used for training and validation in each fold. The overall mean accuracy of 71.38%, accompanied by a relatively high standard deviation of ±17.96%, highlights the inconsistent predictive capability of the model when exposed to different partitions of the dataset. This variation could be attributed to the inherent imbalance in the dataset or the presence of outliers affecting specific folds. Addressing these inconsistencies through stratified cross-validation or data augmentation may be

Journal of Information Systems and Informatics
Vol. 6, No. 4, December 2024

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

necessary to achieve more stable and reliable results, ultimately enhancing the model's generalizability.

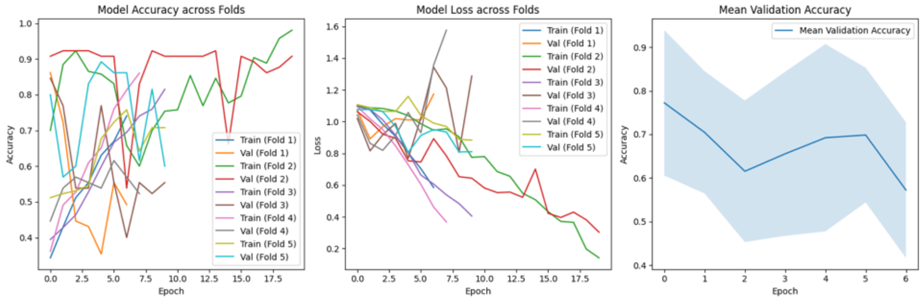

Figure 7. Training History

Figure 7 illustrates the training history of the implemented model, displaying performance metrics across multiple folds. The graph on the left shows model accuracy over epochs, where specific folds exhibit rapid convergence, achieving high accuracy early on, while others demonstrate more fluctuating progress before stabilizing. This variability indicates that model learning is influenced by the distinct characteristics of each fold's data subset. The middle graph tracks model loss, revealing a general downward trend as training progresses, suggesting that the model minimizes errors. However, occasional spikes in loss indicate potential challenges with overfitting in specific folds, which could be attributed to the uneven distribution of class labels. The graph on the right, depicting mean validation accuracy, shows an initial increase followed by a plateau and slight decline, suggesting that additional training does not necessarily translate to better performance on unseen data after a certain point. These visualizations collectively highlight the model's capacity to learn from the data while emphasizing the need for fine-tuning to mitigate overfitting and enhance generalization across diverse subsets.

The evaluation of the final model reveals a stark disparity in classification performance across different sentiment classes. The positive class achieves an impressive accuracy of 98.33%, supported by a substantial sample size of 60 and an average confidence level of 85.04%. This suggests that the model is proficient at identifying positive sentiments, likely due to the overwhelming representation of positive samples in the training data. However, the model's performance drastically declines for the negative and neutral classes, where the accuracy drops to 0%. All four negative samples were misclassified as positive, while the single neutral sample was erroneously classified as negative. This misclassification pattern indicates a clear bias towards the positive class, a common issue in models trained on imbalanced datasets. The lack of accurate predictions for minority classes highlights a significant limitation in the model's ability to generalize beyond the

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

dominant class. Addressing this imbalance through techniques such as data augmentation or adjusted class weights is crucial to enhance the model's performance in recognizing underrepresented sentiments, thereby improving its overall robustness and applicability in real-world scenarios.

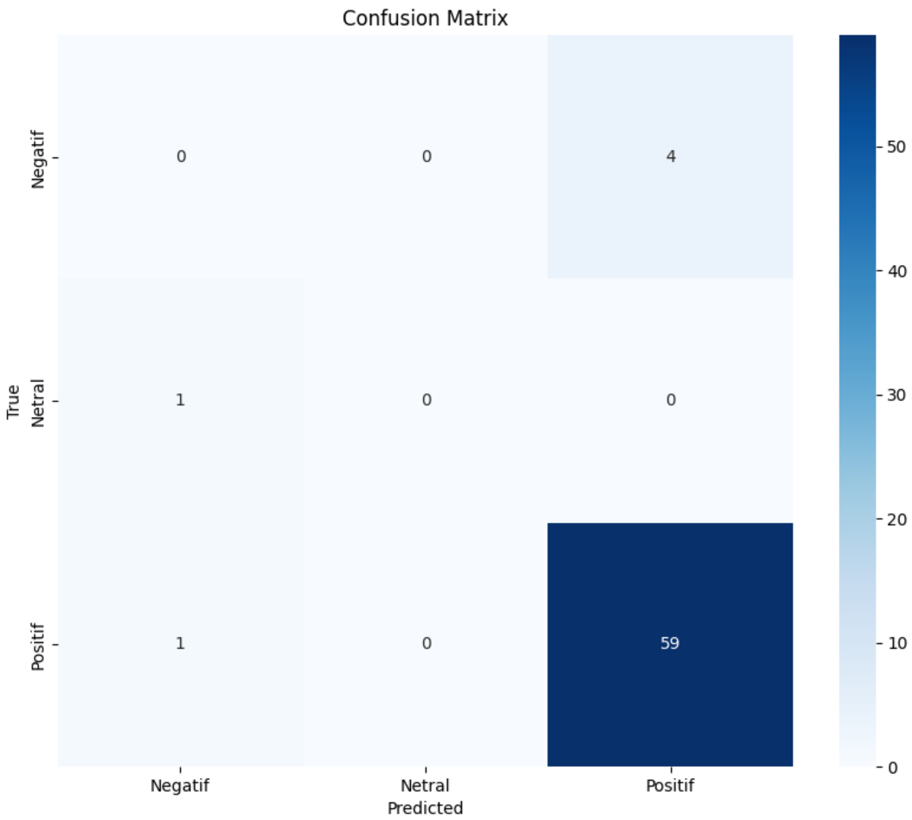

Figure 8. Confusion Matrix

Figure 8 presents the confusion matrix for the final evaluation of the sentiment classification model, providing a visual representation of its performance across different sentiment classes. The matrix indicates a strong bias towards the positive class, as evidenced by the high concentration of correct classifications within this category. Specifically, almost all predictions for positive sentiments align accurately with their proper labels, reflecting the model's proficiency in handling the dominant class. However, the negative and neutral classes suffer from severe misclassification, with all negative instances incorrectly identified as positive and the single neutral instance wrongly classified as negative. This imbalance suggests that the model struggles to differentiate minority classes, likely due to their

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

underrepresentation in the training dataset. Such a skewed distribution significantly hampers the model's generalization ability, especially in cases where nuanced sentiment distinctions are critical. Addressing this issue may require resampling techniques or leveraging class-weight adjustments to improve the model's sensitivity to less frequent categories, thereby fostering a more balanced and accurate classification outcome.

The results of the prediction tests indicate a clear tendency for the model to classify nearly all input cases as positive, regardless of the underlying sentiment. For instance, positive phrases like "The hotel was amazing..." and "Best hotel ever!" were correctly identified with high confidence levels of 89% and 85%, respectively. However, the model erroneously assigned a positive sentiment to statements such as "Terrible experience..." and "Room was dirty..." with 73% and 67% confidence scores, even though these clearly conveyed negative sentiments. Additionally, a neutral review, "It was okay...", was inaccurately classified as positive with a 52% confidence level. This pattern suggests that the model's training on an imbalanced dataset has resulted in a significant bias toward predicting positive sentiments. Such misclassification, particularly in cases where negative feedback is crucial for service improvement, highlights the model's limitation in discerning less frequent sentiment classes. Enhancing model performance may necessitate addressing class imbalance through data augmentation or rebalancing techniques, thereby fostering a more accurate representation of diverse sentiments in future predictions.

3.3 Discussion

The poor performance of the models with minority sentiment classes, such as negative and neutral categories, is primarily attributed to the severe class imbalance in the dataset, where positive sentiments accounted for 92.3% of the entries. This imbalance disproportionately influenced the optimization of model weights, leading to a bias towards the dominant class. Consequently, the models failed to generalize effectively to the underrepresented classes, resulting in 0% classification accuracy for negative and neutral sentiments. The lack of sufficient examples further restricted the models' ability to learn nuanced features of minority classes, exacerbating their classification challenges.

To overcome these limitations, several strategies can be implemented. Data augmentation, such as generating synthetic samples for minority sentiments through text paraphrasing or neural text generation, can improve class balance. Adjusting class weights during training to assign higher penalties for misclassifying minority classes ensures that these categories receive adequate attention in the learning process. Hybrid approaches, such as ensemble learning that combines multiple models with complementary strengths, could further enhance the accuracy of classifying minority sentiments. Transfer learning, leveraging pre-

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

trained models on larger, balanced datasets, offers an additional avenue to address data scarcity and improve generalization.

Unlike traditional sentiment analysis techniques, such as lexicon-based or rule-based methods, LSTM and GRU models demonstrate superior capabilities in capturing contextual nuances and temporal dependencies in sequential textual data. However, these deep learning models require balanced datasets or sophisticated handling of imbalances to ensure consistent performance across all sentiment categories. The study highlights the significant potential of LSTM and GRU in achieving high accuracy for dominant sentiments while offering a framework for addressing limitations with minority classes. These models can be adapted by applying advanced preprocessing techniques and leveraging ensemble or transfer learning to provide more comprehensive sentiment analysis.

The application of LSTM and GRU models in sentiment mapping, as demonstrated in the analysis of Hotel Vila Ombak reviews, achieved an impressive overall accuracy of 91%, excelling in detecting positive sentiments with precision, recall, and F1-scores of 0.94, 0.98, and 0.96, respectively. Despite challenges with minority sentiments, these models provide actionable insights that businesses can use to identify trends, refine service strategies, and enhance customer satisfaction. Positive sentiment analysis allows businesses to emphasize their strengths in marketing campaigns and loyalty programs, fostering customer retention and competitive advantage. By addressing the challenges of minority class representation through advanced techniques, these models offer significant potential for more robust sentiment analysis frameworks in the hospitality industry.

4. CONCLUSION

In the era of digital communication, customer feedback through online reviews has become an essential source of information for businesses, especially within the hospitality industry. Often unstructured and vast in volume, these reviews contain valuable insights that, if properly analyzed, can significantly enhance service quality and customer satisfaction. However, extracting actionable insights from such data is challenging due to its complex, sequential nature. Deep learning models like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) have emerged as powerful tools for addressing these challenges, offering superior capabilities in capturing temporal dependencies and contextual nuances in text. By leveraging these models for sentiment mapping and trend analysis, it becomes feasible to identify shifts in customer satisfaction over time, thus informing data-driven decision-making for service optimization. This research specifically focused on applying LSTM and GRU to analyze customer reviews for Hotel Vila Ombak, aiming to uncover sentiment patterns that could guide strategic improvements in guest experiences. The implementation of LSTM and GRU models on a dataset of

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

326 hotel reviews from Hotel Vila Ombak demonstrated their capability in sentiment classification, achieving an overall accuracy of 91% (0.91). The models were highly influential in identifying positive sentiments, with a precision of 0.94, a recall of 0.98, and an F1-score of 0.96 for the positive class, which dominated the dataset at 92.3% (300 out of 325 entries). However, the models' performance significantly declined for minority classes, with 0% accuracy for both negative and neutral categories. The classification report revealed a clear bias towards predicting positive sentiments, as all negative (5.8% of the data) and neutral (1.8% of the data) samples were misclassified. Low macro averages (precision 0.31, recall 0.33, F1-score 0.32) further highlighted this imbalance, despite the high weighted averages driven by the positive class (precision 0.86, recall 0.91, F1-score 0.89). These findings indicate that while LSTM and GRU models are proficient in handling dominant classes, they struggle with minority classes due to the dataset's skewed distribution. Future studies should prioritize addressing data imbalance by employing techniques such as synthetic data generation, oversampling of minority classes, or hybrid modeling approaches to enhance the generalizability of sentiment analysis tools. Additionally, exploring interpretability frameworks for these deep learning models could improve their adoption in practical business by offering clear explanations for sentiment predictions.

REFERENCES

[1] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews," *Procedia Computer Science*, vol. 179. pp. 728–735, 2021. doi: 10.1016/j.procs.2021.01.061.

[2] M. S. Hossen, A. H. Jony, T. Tabassum, M. T. Islam, M. M. Rahman, and T. Khatun, "Hotel review analysis for the prediction of business using deep learning approach," *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*. pp. 1489–1494, 2021. doi: 10.1109/ICAIS50930.2021.9395757.

[3] R. A. Hameed, W. J. Abed, and A. T. Sadiq, "Evaluation of Hotel Performance with Sentiment Analysis by Deep Learning Techniques," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 9, pp. 70–87, 2023, doi: 10.3991/ijim.v17i09.38755.

[4] N. Zhao, H. Gao, X. Wen, and H. Li, "Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis," *IEEE Access*, vol. 9, pp. 15561–15569, 2021, doi: 10.1109/ACCESS.2021.3052937.

[5] N. Khamphakdee and P. Seresangtakul, "An Efficient Deep Learning for Thai Sentiment Analysis," *Data*, vol. 8, no. 5, 2023, doi: 10.3390/data8050090.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[6] P. B. Weerakody, K. W. Wong, G. Wang, and W. Ela, "A review of irregular time series data handling with gated recurrent neural networks," *Neurocomputing*, vol. 441, pp. 161–178, 2021, doi: 10.1016/j.neucom.2021.02.046.
[7] F. Ali *et al.*, "An intelligent healthcare monitoring framework using wearable sensors and social networking data," *Futur. Gener. Comput. Syst.*, vol. 114, pp. 23–43, 2020, doi: 10.1016/j.future.2020.07.047.
[8] B. Jang, M. Kim, G. Harerimana, S. U. Kang, and J. W. Kim, "Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism," *Appl. Sci.*, vol. 10, no. 17, 2020, doi: 10.3390/app10175841.
[9] A. Al Wazrah and S. Alhumoud, "Sentiment Analysis Using Stacked Gated Recurrent Unit for Arabic Tweets," *IEEE Access*, vol. 9, pp. 137176–137187, 2021, doi: 10.1109/ACCESS.2021.3114313.
[10] A. S. Fadel, M. E. Saleh, and O. A. Abulnaja, "Arabic Aspect Extraction Based on Stacked Contextualized Embedding With Deep Learning," *IEEE Access*, vol. 10, pp. 30526–30535, 2022, doi: 10.1109/ACCESS.2022.3159252.
[11] H. J. Alshahrani, A. Q. A. Hassan, N. S. Almalki, M. M. Alnfiai, A. S. Salama, and M. A. Hamza, "Applied Linguistics With Red-Tailed Hawk Optimizer-Based Ensemble Learning Strategy in Natural Language Processing," *IEEE Access*, vol. 11, pp. 132448–132456, 2023, doi: 10.1109/ACCESS.2023.3331061.
[12] M. Jamshidi *et al.*, "Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment," *IEEE Access*, vol. 8, pp. 109581–109595, 2020, doi: 10.1109/ACCESS.2020.3001973.
[13] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model," *IEEE Access*, vol. 10, pp. 39313–39324, 2022, doi: 10.1109/ACCESS.2022.3165621.
[14] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9713–9729, 2020, doi: 10.1007/s00521-019-04504-2.
[15] S. Mifrah and E. H. Benlahmar, "Sentence-Level Sentiment Classification A Comparative Study between Deep Learning Models," *J. ICT Stand.*, vol. 10, no. 2, pp. 339–352, 2022, doi: 10.13052/jicts2245-800X.10213.
[16] Mamta, A. Ekbal, P. Bhattacharyya, S. Srivastava, A. Kumar, and T. Saha, "Multi-domain tweet corpora for sentiment analysis: Resource creation and evaluation," *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*. pp. 5046–5054, 2020.
[17] Q. Jiang, C. S. Chan, S. Eichelberger, H. Ma, and B. Pikkemaat, "Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists," *Curr. Issues Tour.*, vol. 24, no. 17, pp. 2501–2522, 2021, doi: 10.1080/13683500.2021.1874312.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[18] U. Chakraborty and S. K. Biswal, "Impact of Online Reviews on Consumer's Hotel Booking Intentions: Does Brand Image Mediate?," *J. Promot. Manag.*, vol. 26, no. 7, pp. 943–963, 2020, doi: 10.1080/10496491.2020.1746465.
[19] M. Gharzouli, A. K. Hamama, and Z. Khattabi, "Topic-based sentiment analysis of hotel reviews," *Curr. Issues Tour.*, vol. 25, no. 9, pp. 1368–1375, 2022, doi: 10.1080/13683500.2021.1940107.
[20] Y. H. Hsiao and Y. T. Hsiao, "Online review analytics for hotel quality at macro and micro levels," *Ind. Manag. Data Syst.*, vol. 121, no. 2, pp. 268–289, Jan. 2021, doi: 10.1108/IMDS-04-2020-0214.
[21] M. Jeong, H. H. Shin, M. Lee, and J. Lee, "Assessing brand performance consistency from consumer-generated media: the US hotel industry," *Int. J. Contemp. Hosp. Manag.*, vol. 35, no. 6, pp. 2056–2083, Jan. 2023, doi: 10.1108/IJCHM-12-2021-1516.
[22] J. Wang, Y. Xia, and Y. Wu, "Emotional characteristics and theme mining of star-rated hotels from the perspective of social sensing: a case study of Nanchang City, China," *Comput. Urban Sci.*, vol. 2, no. 1, 2022, doi: 10.1007/s43762-022-00042-1.
[23] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," *IEEE Access*, vol. 10, pp. 21517–21525, 2022, doi: 10.1109/ACCESS.2022.3152828.
[24] J. Wang, L. C. Yu, K. R. Lai, and X. Zhang, "Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 581–591, 2020, doi: 10.1109/TASLP.2019.2959251.
[25] Y. Liu, L. Wang, T. Shi, and J. Li, "Detection of spam reviews through a hierarchical attention architecture with N-gram CNN and Bi-LSTM," *Inf. Syst.*, vol. 103, 2022, doi: 10.1016/j.is.2021.101865.
[26] A. Parasyris, G. Alexandrakis, G. V. Kozyrakis, K. Spanoudaki, and N. A. Kampanis, "Predicting Meteorological Variables on Local Level with SARIMA, LSTM and Hybrid Techniques," *Atmosphere (Basel).*, vol. 13, no. 6, 2022, doi: 10.3390/atmos13060878.