Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

Improving Model Capability for Sentiment Trend Analysis in Hotel Visitor Reviews with Bi-LSTM Multistage Approach

Bayu Yanuargi[*1], Ema Utami[2], Kusrini[3], Arli Aditya Parikesit[4]

[1,2,3]Program Doktoral, Fakultas Teknologi Indormasi, Universitas Amikom, Indonesia
[4]Department of Bioinformatics, Indonesia International Institute for Life Sciences, Indonesia

Email: [1]bayu.yanuargi@students.amikom.ac.id

Received : Jul 28, 2025; Revised : Aug 4, 2025; Accepted : Aug 4, 2025; Published : Oct 21, 2025

Abstract

This study focuses to improve the sentiment analysis of hotel reviews using Multistage mechanism of two-stage approach based on the Bidirectional Long Short-Term Memory (Bi-LSTM) architecture with 53,000 data from 28 hotels in Yogyakarta that captured from google maps review for hotel in Yogyakarta. Hotel customer reviews often contain mixed sentiment expressions, making it crucial to filter out only sentences with a single dominant sentiment to avoid ambiguity. In the first stage, the model detects sentiment at the token level and counts the number of sentiment expressions in each sentence. Only sentences with a single polarity are passed to the final classification stage. In the second stage, the overall sentiment is classified as positive, negative, or neutral using pooled contextual representations. Experimental results from 30 iterations demonstrate consistently high performance, with precision, recall, and F1-scores above 0.95, and overall accuracy exceeding 96%. The confusion matrix analysis shows strong model performance, although some challenges remain in distinguishing between positive and neutral sentiment. Additionally, sentiment trend analysis of hotel reviews from properties such as Lafayette Boutique Hotel and The Westlake Resort Jogja reveals dynamic shifts in guest perception over time. This multistage mechanism approach proves effectiveness of improving sentiment classification accuracy by avoid the bias on sentiment and also in providing valuable temporal insights for monitoring customer satisfaction.

Keywords: *Bi-LSTM, deep learning, natural language processing, sentiment classification, visitor review.*

1. INTRODUCTION

Currently, sentiment analysis on user reviews has become an increasingly prominent research topic. Websites are often used as data sources to assess hotel service quality, while service providers utilize them for monitoring and evaluation purposes. However, sentiment analysis of unstructured textual data remains time-consuming [1]. Sentiment mining on social media for specific targets has become a major focus in decision-making across various sectors, including services, politics, and entertainment. The focus on sentiment analysis continues to evolve, with many studies delving deeper into subjective texts, going beyond merely extracting general sentiments [2].

For hotel business practitioners, understanding visitor lifestyles is crucial for ensuring business sustainability. One effective effort is analyzing reviews to understand guest sentiments, which supports better-informed decision-making in improving service and business strategies in the hospitality industry [3]. Advances in information technology have driven the rise of online hotel booking, where user reviews serve as critical input. Although many platforms provide review features, not all summarize evaluations of each service aspect [4].

Bi-LSTM is a neural network that utilizes bidirectional LSTM units to capture both past and future context simultaneously. This architecture is capable of understanding long-term dependencies in data without redundantly storing contextual information, making it efficient in processing text sequences

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

[5]. Most current systems adopt LSTM for sentiment analysis, yet few studies have evaluated the impact of various hyperparameters on model performance. The combination of Bi-LSTM with dropout layers has proven to enhance accuracy and deliver improved results in sentiment analysis tasks [6].

Related research has developed a BiLSTM+IndoBERT model for sentiment analysis of Indonesian TikTok user reviews, achieving 81% accuracy in the classification report and 92.03% in cross-validation testing. However, the study had limitations due to ambiguous meanings in the data, highlighting the need to incorporate data from diverse sources to improve prediction accuracy [7]. The tourism industry is rapidly growing, with more travelers booking hotels and sharing their reviews on travel e-commerce platforms. One study employed deep learning techniques to analyze sentiment from hotel reviews crawled from TripAdvisor. A noted limitation was the drop in accuracy when the dataset was changed. The LSTM model achieved 96.42% accuracy on the Padma Hotel dataset and 85.31% on the Hard Rock Hotel dataset. The CNN-LSTM model obtained 85.87% accuracy on the Ayana Hotel dataset, while the BiLSTM model reached 86.09% on the Pullman Hotel dataset [8].

Another study on sentiment trends applied LSTM and GRU models to analyze sentiment and customer review trends at Hotel Vila Ombak, Lombok. Using 326 review records, the model achieved an overall accuracy of 91% (0.91). The model showed excellent performance in classifying positive sentiment, with a precision of 0.94, recall of 0.98, and F1-score of 0.96. However, it failed to recognize negative and neutral sentiments, yielding 0% accuracy due to data imbalance—92.3% of reviews were positive. Consequently, the macro average was also low (precision 0.31, recall 0.33, F1-score 0.32). This study underscores the importance of data augmentation techniques or hybrid models to enhance performance across all sentiment categories [9].

This study proposes a novel Multistage Sentiment Classification approach by modifying the Bi-LSTM architecture into an integrated two-stage model to enhance sentiment trend analysis in hotel reviews. The novelty of this approach lies in the incorporation of a sentiment filtering mechanism at the first stage, where a modified Bi-LSTM model is used to detect and filter sentences containing a single dominant sentiment, thus eliminating noisy data with mixed or ambiguous sentiments. In the second stage, another Bi-LSTM model classifies the filtered sentences into positive, negative, or neutral categories. The classified results are then visualized using a monthly sentiment trend graph based on review timestamps. The key contribution of this research is the development of a robust and accurate multistage sentiment analysis framework that improves upon previous single-stage models by effectively addressing sentiment ambiguity. The main objective is to enhance the detection of sentiment trends over time, thereby providing more reliable insights for decision-making in the hospitality industry.

2. METHOD

The research flow illustrated in Figure 1 below outlines the key stages in the development of a sentiment analysis model based on the BiLSTM-SentGate architecture. The process begins with the collection of user review data, which is then subjected to oversampling and data augmentation techniques. These steps are essential for addressing class imbalances within the dataset, ensuring that minority sentiment classes are adequately represented. Once this step is completed, the data undergoes a text preprocessing phase to generate clean and consistent input suitable for model training.

Following preprocessing, the processed data is directed through two distinct paths. The first path is used to train a sentiment count model, which determines how many distinct sentiments are present within a single review. The second path is used to train a sentiment classification model, which categorizes the review into positive, negative, or neutral sentiment—but only if the sentiment count model identifies the review as containing a single sentiment. This multistage approach ensures that only

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

clearly defined sentiment data proceeds to the classification phase, improving model accuracy and robustness.

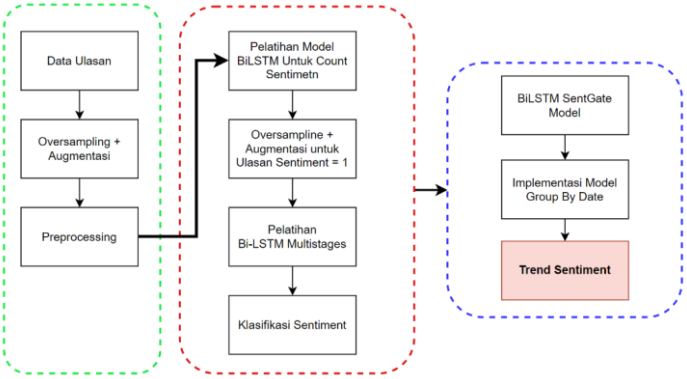

Figure 1. Research Flow

The next step involved the training of two separate models. First, a BiLSTM model was trained to count the number of sentiments within a single review. This model enabled the filtering of reviews containing only one type of sentiment, allowing them to be optimally utilized in the training of the primary model, namely the BiLSTM Multistage. Afterward, the single-sentiment reviews were reprocessed through oversampling and augmentation to balance the sentiment distribution before being used for training the BiLSTM Multistage model. The BiLSTM Multistage model was then trained to classify the sentiment (positive, negative, or neutral) of each review that had passed the earlier filtering stage.

The final phase of the system was the implementation of the trained model to perform sentiment classification on new review data. The classification results were grouped by date to enable sentiment trend analysis over time. Through this approach, the system was not only capable of detecting and classifying sentiments from complex data but also provided temporal insights into user opinion dynamics. This workflow emphasized the importance of review selection based on sentiment count and data balancing to improve the model's accuracy and generalization capabilities.

2.1. Data Preparation

The data collection in this study was carried out through two types of websites, one of which was the official website of the Indonesian Hotel and Restaurant Association (PHRI) at https://phriyogyakarta.com/. This site was used to obtain hotel and restaurant data in the Special Region of Yogyakarta that are affiliated with PHRI. However, not all hotel data were used in the initial model selection phase. At this stage, only data from 28 hotels located in Sleman Regency were utilized to compare and select the best model, resulting in approximately 53,000 review entries. The complete hotel data from the PHRI website were reserved for use in the final stage of business intelligence analysis.

Many real-world classification tasks involved imbalanced data, where the disparity between majority and minority classes could lead to bias toward the majority class, ultimately resulting in less accurate predictions [10]. In this study, data balancing was conducted to ensure that the trained model could perform classification accurately and precisely. Figure 2 below illustrates the steps taken to address data imbalance. As shown in Figure 2, the step following oversampling was the application of augmentation to make the data appear more natural. Augmentation served as an effective technique in machine learning to improve model performance by increasing data variation from existing samples, especially when training data were limited or difficult to obtain directly [11].

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

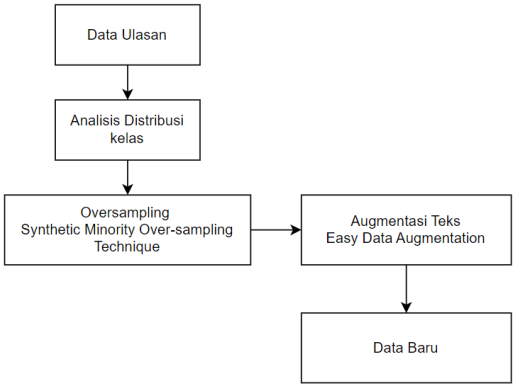

Figure 2. Data Augmentation and Oversampling Flow

This approach proved to be highly beneficial in situations where the process of collecting original data required significant time, cost, or resources. It provided an effective way to address limitations in data availability while maintaining the quality needed for model training. Therefore, data augmentation served as a practical solution to enhance model generalization and reduce the risk of overfitting [12].

2.2. Bi-LSTM Multistages

Figure 3 below illustrated the two main stages in the sentiment analysis process using the BiLSTM architecture. In Stage 1, the input text consisting of a sequence of words was first processed through an embedding layer to convert the words into numerical vectors. These vectors were then processed by a BiLSTM model operating in two directions—forward and backward—enabling it to capture the full context of each word [13]. The output from the BiLSTM was used to identify tokens that carried either positive or negative sentiment. Each token was examined, and the number of sentiment-bearing tokens was counted. If a sentence contained only one explicit type of sentiment, it was considered suitable for further analysis and passed on to the next stage.

In Stage 2, only the sentences that had passed the filtering process in the previous stage were further processed. These sentences were again input into the BiLSTM model, this time with adjusted parameter settings through hyperparameter tuning, to produce the final sentiment classification. The goal of this stage was to accurately determine whether the entire sentence expressed a positive, negative, or neutral sentiment. By separating the sentiment count detection and final classification processes, this approach ensured that only relevant and clean data were classified, thereby improving the accuracy and reliability of the sentiment analysis results.

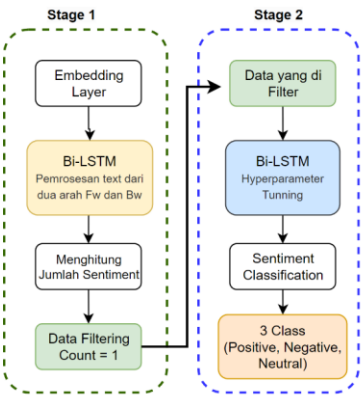

Figure 3. Bi-LSTM Multistage Flow

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

The Bi-LSTM formula in this paper was actually a modified version of the standard Bi-LSTM, specifically applied in the first stage to calculate the number of sentiments in a single sentence. The modification of the Bi-LSTM was reflected in formulas (10) and (9) shown below [14].

$$X = [x_1, x_2, \dots, x_n] \quad (1)$$

$$e_t = Embedding\ (x_t) \epsilon \mathbb{R}^{64} \quad (2)$$

$$h_t = LSTM_{fw}(e_t, h_{t-1}), Dropout = 0.1 \quad (3)$$

$$h_t = LSTM_{bw}(e_t, h_{t+1}) \quad (4)$$

$$h_t = [h_t; h_t] \epsilon \mathbb{R}^{48} \quad (5)$$

$$z_t = Dropout(h_t, p = 0.3) \quad (6)$$

$$d_t = ReLU(W_d\ .\ z_t + b_d), W_d \epsilon R^{16\ x\ 48} \quad (7)$$

$$Loss_{L2} = 0.001\ .\ \|W_d\|_2^2 \quad (8)$$

$$y_t = Softmax\ (W_o\ .\ d_t + b_o\ ) \quad (9)$$

$$S_{t=}\begin{cases} 1, & if \arg\max(y_t) \epsilon \{postive, negative\} \\ 0, & if \arg\max(y_t) = neutral \end{cases} \quad (10)$$

$$Sentiment\ Count = \sum_{t=1}^{T} S_t \quad (11)$$

$$h_{final} = MeanPooling\ (h_1, h_2, \dots, h_t\ ) \epsilon \mathbb{R}^{48} \quad (12)$$

$$z_t = Dropout(h_t, p = 0.3) \quad (13)$$

$$d_t = ReLU(W_d\ .\ z_t + b_d), W_d \epsilon R^{16\ x\ 48} \quad (14)$$

$$Loss_{L2} = 0.001\ .\ \|W_d\|_2^2 \quad (15)$$

$$y_t = Softmax\ (W_o\ .\ d + b_o\ ) \quad (16)$$

The BiLSTM model shown in the Equation (1) represents a sequence of input features where each *xi* is an individual element or token in the sequence. This notation is commonly used in natural language processing to represent tokenized text input. Equation (2) defines the embedding process, where the input token *xt* is transformed into a dense vector representation *et* in a 64-dimensional space, i.e., R^{64}. This allows the model to capture semantic information from discrete input tokens for further processing, which was then passed into a bidirectional LSTM (BiLSTM) [15]. Equation (3) shows that the hidden state ht is computed using a forward LSTM with input et and previous hidden state h_{t-1}, with dropout rate 0.1. The outputs from the forward and backward LSTM were concatenated into ht∈, capturing contextual information from both directions in the sentence. This result enhanced the model's ability to understand the meaning of words based on their surrounding context.

Equation (4) represents the backward LSTM pass, where the hidden state ht is computed using input et and future hidden state h_{t+1}. Equation (5) shows the concatenation of forward and backward hidden states to form a bidirectional representation $h_t \in R^{48}$. Equations (6) to (8) describe key steps in the post-LSTM processing stage of the model. In Equation (6), a dropout operation is applied to the Bi-LSTM hidden state ht, producing zt, with a dropout probability p=0.3. This regularization step helps

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

prevent overfitting by randomly setting a fraction of the activations to zero during training. In Equation (7), a fully connected layer is applied to the output zt, where a linear transformation is followed by a ReLU activation function. The transformation uses weight matrix $W_d \in R^{16\times48}$ and bias bd, resulting in a 16-dimensional output vector dt. This helps map the high-dimensional features into a lower-dimensional space suitable for classification. Equation (8) introduces an L2 regularization loss term on the weights W_d, with a coefficient of 0.001. This encourages smaller weights to reduce model complexity and improve generalization.

The BiLSTM output was then processed through a dropout layer, ReLU activation, and a softmax layer to generate token-level predictions yt. These probabilities were used to determine whether a token carried explicit sentiment ("positive" or "negative") or was neutral [16]. Formula (10) converted the softmax output into binary labels St, and Formula (11) calculated the total number of non-neutral sentiment tokens in a sentence, referred to as the Sentiment Count [17]. This process was crucial for filtering sentences that truly contained a dominant single sentiment, ensuring that only those with explicit single sentiment were passed on to the sentence-level classification stage.

For sentences that contained only one type of sentiment (based on the Sentiment Count value), the sentence classification process was re-executed by repeating the steps from Formula (1) to (9). All hidden states ht from the tokens in the sentence were pooled using mean pooling to produce the final sentence representation hfinal . This representation then passed through dropout, ReLU activation, L2 regularization, and a softmax layer to generate the overall sentiment label (positive, negative, or neutral) [18]. In this way, the model was not only able to detect the presence of sentiment at the token level but also accurately classified sentences with a single explicit sentiment in a selective manner. Equations (11) to (16) outline the final stages of the multistage sentiment classification model, starting from sentiment counting to final sentiment prediction. In Equation (11), the Sentiment Count is computed as the sum of all detected sentiment tokens St across the sequence. This is used to determine whether a sentence has a single dominant sentiment, which is essential for filtering ambiguous sentences.

Equation (12) applies Mean Pooling across all hidden states h1,h2,...,ht generated by the Bi-LSTM to produce a single, fixed-size vector $h_{final} \in R^{48}$, representing the entire sentence. Equation (13) applies Dropout with a probability of 0.3 to the hidden representation ht, producing zt, to reduce overfitting by randomly disabling neurons during training. Equation (14) is a fully connected layer with ReLU activation, where $W_d \in R^{16\times48}$ maps the input zt to a lower-dimensional vector dt. Equation (15) adds L2 regularization loss on Wd to penalize large weights and improve generalization. Finally, Equation (16) computes the final sentiment classification output yt using a softmax layer with parameters Wo and bo, producing probability scores for sentiment classes (positive, negative, neutral).

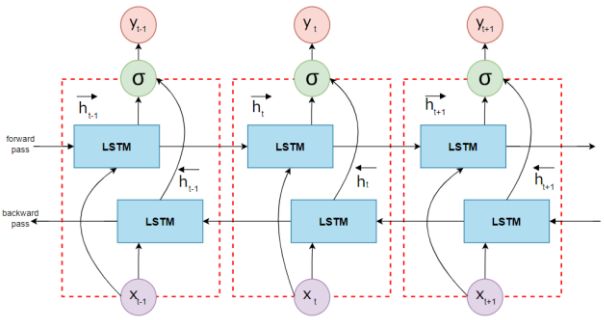

Figure 4. Bi-LSTM Architecture in Each Stage of the Approach [19]

The diagram on figure 4 illustrates the architecture of a Bidirectional LSTM (Bi-LSTM), where each input token xt is processed in both forward and backward directions through separate LSTM layers. The hidden states from both directions, ht and ht , are concatenated to form a comprehensive

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

representation ht, as described in Equation (5). This structure allows the model to capture both past and future context within a sequence. The embedding of each token (Equation 2) feeds into the Bi-LSTM, producing contextualized hidden states (Equations 3 and 4). These are later used in sentiment filtering (Equation 11) and mean pooling (Equation 12) to derive a final sentence representation. Further transformations including dropout (Equation 6), dense layers (Equation 7), and softmax classification (Equation 16) are applied to produce the sentiment label. This Bi-LSTM architecture enhances sentiment analysis by leveraging sequential dependencies in both temporal directions [19].

3. RESULT

The training stages of the model described in the methodology chapter were then carried out over thirty iterations, using a data composition of 80% for training and 20% for validation. Each of the thirty iterations consisted of 100 epochs with an early stopping condition set to 10, with the goal of assessing whether the developed model demonstrated high stability and robustness. Figure 5 below showed the accuracy fluctuations from the experimental results of the Bi-LSTM model using a multistage approach conducted over 30 iterations. Although there were variations between trials, the model generally demonstrated stable performance with accuracy ranging from 94.6% to 96.2%. This indicated that the two-stage strategy—separating the sentiment count detection process from the sentence classification based on filtered data—was capable of delivering consistent classification results. Several sharp drops in the graph were likely caused by random variables such as the split between training and testing data or noise within the input data that contained sentiment ambiguities.

Tabel 1. Accuracy Statistic of 30x Iteration

No	Accuracy Stat	Value
1	Average	95.5%
2	Median	96.2%
3	Maximum Accuracy	96.1%
4	Minimum Accuracy	94.5%
5	Standard Deviation	0.0031

The statistical analysis of 30 trials as shown in table 1 indicates that the model demonstrates consistently high and stable performance in sentiment classification. The average validation accuracy reached 95.53%, with a median of 95.56%, suggesting a balanced distribution of model performance without extreme outliers. The highest accuracy recorded was 96.16%, while the lowest was 94.55%, all within a high-performance range. The relatively low standard deviation of 0.0031 shows that variation across trials was minimal, confirming the model's reliability and stability during repeated evaluations. This level of consistency reinforces the model's effectiveness, making it not only accurate but also dependable for real-world sentiment analysis tasks where precision and reproducibility are crucial.

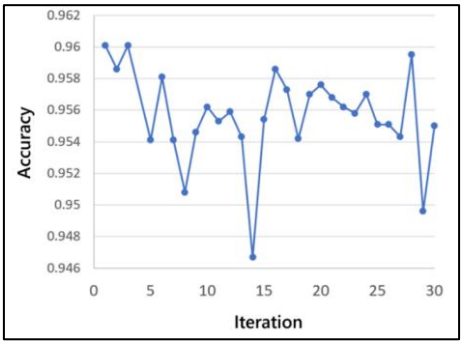

Figure 5. Model Accuracies within 30x Iteration

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

The multistage approach with sentence filtering based on sentiment count proved effective in improving the quality of input data for final classification. By involving only sentences that contained a single explicit sentiment, the model was able to learn more effectively and avoid ambiguity in the training data. In addition, the implementation of Bi-LSTM, which captured bidirectional context, further enhanced the model's ability to understand the full meaning of a sentence. The results obtained from the graph reinforced the argument that this method was suitable for use in integrated token-level and sentence-level sentiment analysis scenarios.

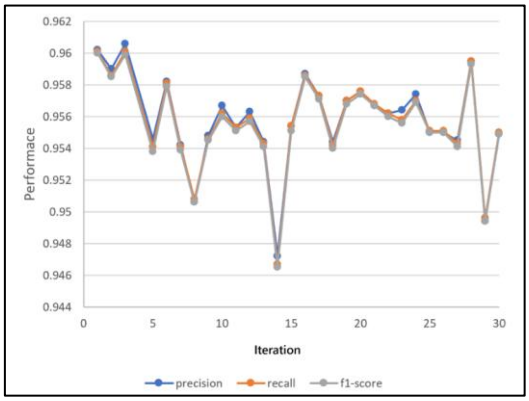

Figure 6 Model Performance Within 30x Iterations

Figure 6 above showed the performance of the Bi-LSTM multistage model in terms of precision, recall, and F1-score over 30 experimental runs. Overall, the three performance metrics remained within a high and relatively stable range, approximately between 0.95 and 0.96. This indicated that the model was not only capable of correctly classifying sentiments (high precision) but also consistently recognized all relevant instances (high recall). The balance between precision and recall was reflected in the F1-score, which consistently fell between the two. This consistency served as an indicator that the two-stage approach effectively minimized classification errors without compromising the model's sensitivity.

However, there were a few sharp fluctuations at certain iterations, such as in trials 14 and 29, which suggested potential sensitivity of the model to variations in training data distribution or noise within the input. This highlighted the importance of model validation, particularly to ensure that performance was not only high on average but also stable under varying conditions. Nevertheless, overall results showed that the Bi-LSTM multistage architecture—focused on sentences with a single explicit sentiment—was able to maintain strong and consistent classification performance in integrated token- and sentence-based sentiment analysis scenarios.

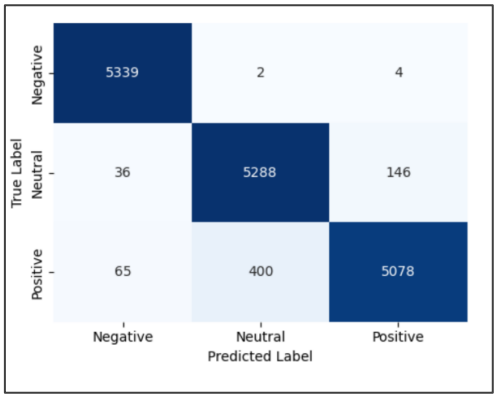

Figure 7. Confusion Matrix of the Model

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

The confusion matrix in Figure 7 above showed differences in classification accuracy across the three sentiment classes (Negative, Neutral, Positive). In the third experiment, the model demonstrated excellent performance, with a high number of correct predictions in all classes: 5,339 for the negative class, 5,288 for neutral, and 5,078 for positive. Misclassifications were relatively low, especially for the negative and positive classes. There were only 400 incorrect predictions for the positive class that were classified as neutral—this was the largest error in the experiment, but still within an acceptable range.

Based on Table 2, the sentiment classification model showed excellent performance with an accuracy value of 96.20%. This meant that 96.20% of the model's total predictions matched the actual labels. In terms of precision, the model achieved high scores for the negative class (98.14%) and the positive class (97.03%), indicating that the model's predictions for these two classes were rarely incorrect. Although the precision for the neutral class was slightly lower (92.95%), the macro average precision remained high at 96.04%. This macro precision demonstrated that the model was able to maintain balanced prediction performance across all classes, rather than relying on performance in only one dominant class.

Tabel 2. Model Performance

No	Performance	Value
1	Accuracy	96.20%
2	Precision Negative	98.14%
3	Precision Neutral	92.95%
4	Precision Positive	97.03%
5	Precision Macro Averages	96.04%
6	Recall Negative	99.89%
7	Recall Neutral	96.67%
8	Recall Positive	91.60%
9	Recall Macro Averages	96.05%
10	F1 Negative	99.01%
11	F1 Neutral	94.77%
12	F1 Positive	94.22%
13	F1 Macro Average	96.00%

Furthermore, the recall metric showed that the model performed very well in identifying actual data for the negative class (99.89%) and the neutral class (96.67%), although it dropped slightly for the positive class (91.60%). This decline suggested that the model still occasionally missed positive instances and misclassified them as neutral. Nevertheless, the macro average recall value remained high at 96.05%, indicating consistent performance. The F1-score, which combined precision and recall, also reflected the model's stability, with an F1-score of 99.01% for the negative class, 94.77% for the neutral class, and 94.22% for the positive class. With a macro average F1-score of 96.00%, the model was considered reliable in classifying the three sentiment categories in a balanced manner.

The model's overall robustness remained intact, as indicated by a stable prediction pattern where the majority of predictions for each class remained on the diagonal of the matrix (correct classifications), and there were no extreme misclassifications where classes were completely confused with one another. This meant the model was able to maintain a consistent prediction structure, particularly in distinguishing negative sentiments accurately. This pattern indicated that the model had a respectable level of generalization and robustness, and its performance could be further improved through fine-tuning without the need to overhaul the entire architecture.

The Bi-LSTM Multistage model was used in the final stage of the sentiment analysis system to classify customer reviews based on a single sentiment polarity—positive, negative, or neutral. After a prior filtering process had been conducted to select only reviews containing one dominant type of

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

sentiment, the final classification process was carried out using the Bi-LSTM architecture and the Multistage mechanism. This model was able to assign accurate sentiment labels to each review. In addition, the classification results were accompanied by time information in the form of review timestamps, which could be used for sentiment trend analysis over specific periods or time ranges.

Each review collected from Google Maps included timestamp information, allowing sentiment data to be associated with specific time periods. This temporal metadata enabled the construction of sentiment trends over time, making it possible to analyze how customer opinions evolved in response to various factors such as service improvements, seasonal changes, or external events. By leveraging the timestamp data, the sentiment analysis could move beyond static classification and support dynamic trend mapping, offering deeper insights into customer behavior patterns and facilitating more informed decision-making for businesses.

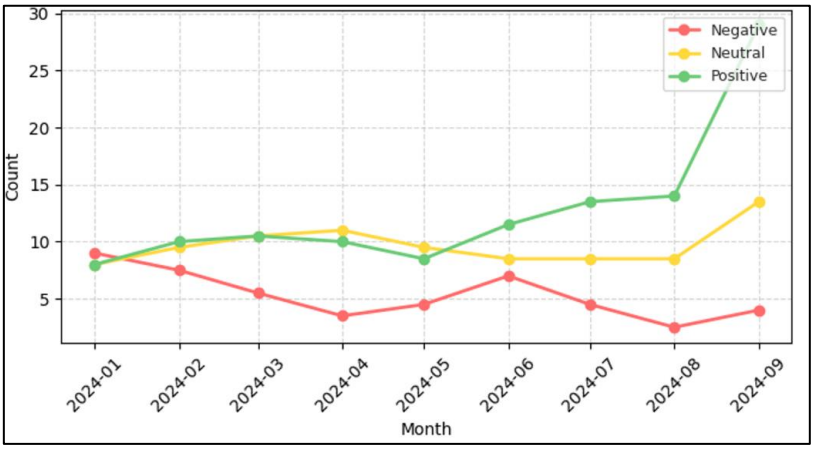

Figure 8. Sentiment Trend of Grand Keisha Hotel Yogyakarta

In Figure 8, the sentiment trend for Hotel Grand Keisha Yogyakarta from January to September 2024 showed an increase in positive sentiment starting in May, peaking in September. Negative sentiment steadily declined and reached its lowest point in August. Neutral sentiment remained stable but rose toward the end of the period, reflecting improvements in service.

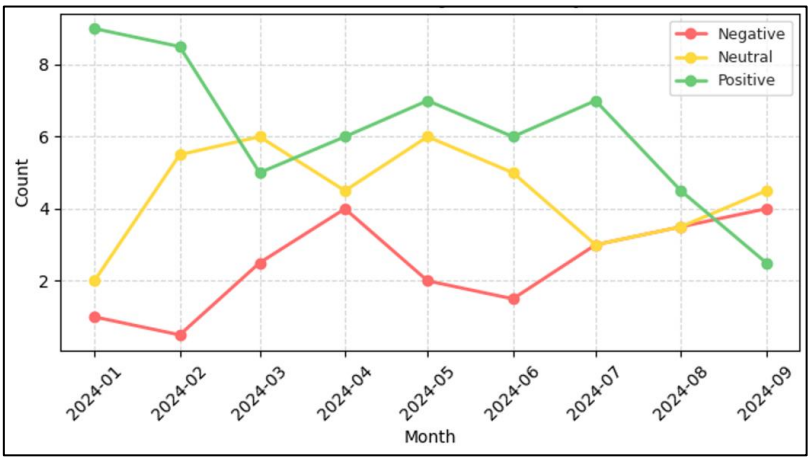

Figure 9. Sentiment Trend of Lafayette Hotel Yogyakarta

Figure 9 showed the sentiment trend graph for Lafayette Boutique Hotel, indicating a decline in positive reviews from January to September 2024, while negative reviews gradually increased. Neutral

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

sentiment remained fluctuating but relatively stable. This pattern reflected a decrease in guest satisfaction and served as an important signal for the management to evaluate their services.

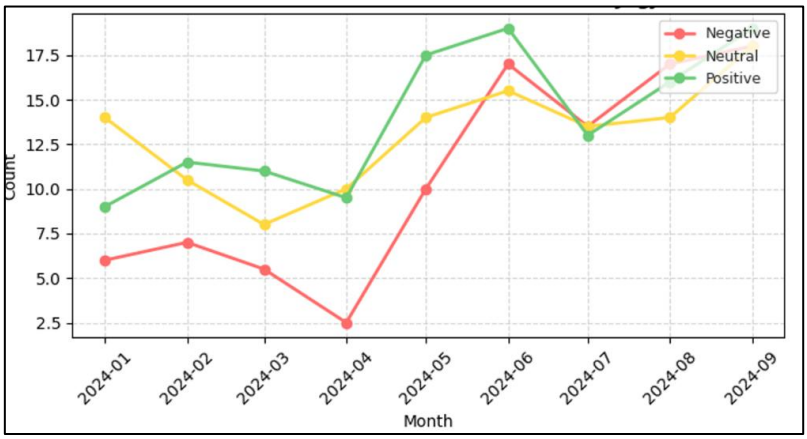

Figure 10. Sentiment Trend Of Westlake Resort Hotel Yogyakarta

Figure 10 displayed the sentiment trend graph for The Westlake Resort Jogja from January to September 2024, showing a fluctuating yet dynamic pattern. Positive reviews rose sharply from April, peaking in June with 19 reviews, then declined briefly before rising again in September. Meanwhile, negative reviews also surged sharply starting in May, reaching 18 reviews in June, indicating an increase in complaints despite the high number of positive reviews. Neutral sentiment remained relatively stable but experienced a slight increase since April. This pattern reflected a surge in activity during the mid-year period, highlighting the need for management to maintain service quality during times of increased demand.

The table 3 above summarized nine model approaches used in sentiment analysis of review texts, each employing a combination of deep learning architectures, machine learning techniques, and varying dataset sizes. The model with the highest accuracy was Bi-LSTM SentGate, which applied a multistage Bi-LSTM approach and achieved an accuracy of 96.20% on a dataset of 53,000 entries. Other high-performing approaches included LSTM-GRU, with 93.12% accuracy, and BiLSTM CF - BiGRU, with 92.56%, both relying on combinations of LSTM and GRU architectures as well as aspect-based techniques to optimize sentiment classification.

Most models in the table utilized hybrid methods to enhance performance, such as the BERT-TCN-BiLSTM-Attention model, which combined dynamic semantic representation and achieved 91% accuracy, despite being trained on a small dataset (3,000 samples). Another notable approach was the use of GloVe embeddings and SMOTE-ENN in the BiLSTM-LSTM model, which aimed to address data imbalance and improve semantic understanding, achieving 89.40% accuracy. The final model combined sentiment dictionary, SVM, Gradient Boosting, and Bi-LSTM with Max Pooling, reaching 89.03% accuracy on a large dataset of 180,379 samples. This demonstrated that integrating classical and modern techniques within the right architecture provided competitive results in sentiment analysis.

In comparison to previous studies with the same objective of analyzing sentiment trends, this research offers a significant advantage in terms of model robustness testing. The model was evaluated 30 times, ensuring the consistency and reliability of the results. In contrast, earlier studies [9] relied on a single test run, making it difficult to confirm whether the reported accuracy was genuinely representative. Moreover, previous research did not apply single sentiment filtering, which could lead to sentiment bias, as a single sentence might contain multiple conflicting sentiments. This study addressed both concerns, enhancing the validity and precision of sentiment trend analysis.

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

Tabel 3. Comparison with Other Research

No	Model	Dataset	Approach	Accuracies
1	Bi-LSTM SentGate*	53,000	Multistage Bi-LSTM	96.20%
2	LSTM-GRU	326	The use of deep learning with LSTM and GRU for sentiment analysis of hotel review texts. [9]	91%
3	D-RNN	162,980	A combination of deep learning (BERT and D-RNN), traditional NLP techniques (BoW, Word2Vec), and aspect-priority based sentiment modeling to improve accuracy [20]	91%
4	LSTM-GRU	20,491	The use of deep learning with LSTM and GRU for sentiment analysis of hotel review texts. [21]	93.12%
5	BERT-TCN-BiLSTM-Attention	3,000	A hybrid approach based on deep learning and dynamic semantic representation, combining BERT, TCN, Bi-LSTM, and Attention. [22]	91%
6	LSTM	50,000	The approach combines deep learning (LSTM), a separated frontend-backend architectural design, and structured data formats (JSON, Pickle). [23]	87.53%
7	BiLSTM-LSTM GloVe	266	A combination of BiLSTM-LSTM, GloVe embeddings, and SMOTE-ENN in a single hybrid deep learning architecture. [24]	89.40%
8	BiLSTM CF - BiGRU	23,485	A hybrid model of BiLSTM-CF and Bi-GRU in aspect-based sentiment analysis. [25]	92.56%
9	BiLSTM	180,379	An approach using a combination of sentiment dictionary, machine learning (SVM, Gradient Boosting), and a deep learning model Bi-LSTM with Max Pooling. [26]	89.03%

4. CONCLUSSION

This study proposed a two-stage sentiment analysis approach using a Bi-LSTM architecture combined with a SentGate mechanism. The process began with token-level sentiment identification to count the number of sentiment expressions in a sentence. Only sentences containing a single type of sentiment were processed further in the sentence classification stage. This strategy proved effective in improving the quality of model input and minimizing data ambiguity. The model implementation was also enhanced with timestamp information, allowing sentiment trend analysis over time.

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

The model's performance evaluation results demonstrated high and stable precision, recall, and F1-scores over 30 trials, with overall accuracy exceeding 96%. Although the performance remained high, there were fluctuations in the model's ability to distinguish between neutral and positive sentiments. This indicated that the quality and distribution of training data influenced classification outcomes. These metrics served as indicators that the Bi-LSTM Multistage model was reliable for handling real-world data.

Sentiment trend analysis of several hotels, such as Lafayette Boutique Hotel and The Westlake Resort Jogja, provided valuable insights. Lafayette showed a gradual decline in positive sentiment and an increase in negative sentiment throughout the year, indicating a potential decline in service quality. In contrast, trends at The Westlake were more dynamic, with a surge in both positive and negative reviews during the mid-year period. These patterns served as important signals for hotel management to maintain service consistency, especially during peak seasons. This combined approach of sentiment classification and temporal analysis proved useful for data-driven decision-making in the service sector.

This research contributes to the advancement of sentiment trend modeling in the hospitality domain using deep learning approaches, which can be adapted in recommender systems or customer satisfaction analytics. But the research has limitation that conducted in all topic of review, the future research expecting can be conducted to improve the model by accurately identifying each topic that impacting the hospitality business such as services quality, location satisfaction and operations hour feedback.

REFERENCES

[1] T. Arwila Utami and I. Wayan Ordiyasa, "Sentiment Analysis of Hotel User Review Using RNN Algorithm," *International Journal of Informatics and Computation (IJICOM)*, vol. 3, no. 1, doi: 10.35842/ijicom.

[2] B. Yanuargi, Ema Utami, Kusrini, and A. A. Parikesit, "Data Clustering for Sentiment Classification with Naïve Bayes and Support Vector Machine," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 6, pp. 819–827, Dec. 2024, doi: 10.29207/resti.v8i6.6139.

[3] V. Chandradev, I. Made, A. Dwi Suarjaya, I. Putu, and A. Bayupati, "Analisis Sentimen Review Hotel menggunakan Metode Deep Learning BERT 107 Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," vol. 14, no. 2, pp. 107–115, Oct. 2023, doi: https://doi.org/10.24002/jbi.v14i02.7244.

[4] R. Jayanto, R. Kusumaningrum, and A. Wibowo, "Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory," *International Journal of Advances in Intelligent Informatics*, vol. 8, no. 3, pp. 391–403, Nov. 2022, doi: 10.26555/ijain.v8i3.691.

[5] 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," *Applied Sciences (Switzerland)*, vol. 10, no. 17, Sep. 2020, doi: 10.3390/app10175841.

[6] A. Singh, R. Thapliyal, R. Vanave, R. Shedge, and S. Mumbaikar, "Analysis of hyperparameters in Sentiment Analysis of Movie Reviews using Bi-LSTM," *ITM Web of Conferences*, vol. 44, p. 03012, 2022, doi: 10.1051/itmconf/20224403012.

[7] A. F. Al Farizi and Y. Sibaroni, "Implementation of BiLSTM and IndoBERT for Sentiment Analysis of TikTok Reviews," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, pp. 96–106, Jan. 2025, doi: 10.29100/jipi.v10i1.5815.

[8] F. Amali, H. Yigit, and Z. H. Kilimci, "Sentiment Analysis of Hotel Reviews using Deep Learning Approaches," *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*, pp. 1–8, Apr. 2024, doi: 10.1109/eStream61684.2024.10542593.

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

[9] Y. A. Singgalen, "Sentiment Analysis and Trend Mapping of Hotel Reviews Using LSTM and GRU," *Journal of Information Systems and Informatics*, vol. 6, no. 4, pp. 2814–2836, Dec. 2024, doi: 10.51519/journalisi.v6i4.926.

[10] S. U. Sabha, A. Assad, N. M. U. Din, and M. R. Bhat, "Comparative Analysis of Oversampling Techniques on Small and Imbalanced Datasets Using Deep Learning," *2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)*, pp. 1–5, Mar. 2023, doi: 10.1109/AISP57993.2023.10134981.

[11] Gilbert, S. Beatricia, A. Reinaldo, and M. F. Hasani, "Text Augmentation for Indonesian Intent Classification: Comparative Study," *2024 11th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, pp. 265–270, Aug. 2024, doi: 10.1109/ICITACEE62763.2024.10762810.

[12] S. Wada and N. Morimoto, "Investigating Relationship between Data Augmentation Intensity and Model Performance in Natural Language Processing," *2024 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, pp. 445–446, Jul. 2024, doi: 10.1109/ICCE-Taiwan62264.2024.10674562.

[13] N. Kapali, T. Tuhin, A. Pramanik, Md. S. Rahman, and S. R. H. Noori, "Sentiment Analysis of Facebook and YouTube Bengali Comments Using LSTM and Bi-LSTM," *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, Oct. 2022, doi: 10.1109/ICCCNT54827.2022.9984395.

[14] T. Pang, J. Liu, L. Han, H. Liu, and D. Yan, "A Deep Learning-Based Analysis of Customer Concerns and Satisfaction: Enhancing Sustainable Practices in Luxury Hotels," *Sustainability*, vol. 17, no. 8, p. 3603, Apr. 2025, doi: 10.3390/su17083603.

[15] A. S. Mirdan, S. Buyrukoğlu, and M. R. Baker, "Advanced deep learning techniques for sentiment analysis: combining Bi-LSTM, CNN, and attention layers," *International Journal of Advances in Intelligent Informatics*, vol. 11, no. 1, pp. 55–71, Feb. 2025, doi: 10.26555/ijain.v11i1.1848.

[16] H. Yuan and A. A. Hernandez, "Sentiment Analysis of Student Evaluation of Teaching Based on Bi-LSTM Algorithm," *2023 IEEE 15th International Conference on Advanced Infocomm Technology (ICAIT)*, pp. 417–422, Oct. 2023, doi: 10.1109/ICAIT59485.2023.10367267.

[17] J. Amalia, D. F. Matondang, G. E. M. Hutajulu, and A. Hasibuan, "Impact Of Sarcasm Detection on Sentiment Analysis Using Bi-LSTM and FastText," *Jurnal Sistem Informasi Bisnis*, vol. 14, no. 4, pp. 353–362, Oct. 2024, doi: 10.21456/vol14iss4pp353-362.

[18] Md. M. Rahman and Y. Watanobe, "Multilingual Program Code Classification Using n-Layered Bi-LSTM Model With Optimized Hyperparameters," *IEEE Trans Emerg Top Comput Intell*, vol. 8, no. 2, pp. 1452–1468, Apr. 2024, doi: 10.1109/TETCI.2023.3336920.

[19] D. Naik and C. D. Jaidhar, "A novel Multi-Layer Attention Framework for visual description prediction using bidirectional LSTM," *J Big Data*, vol. 9, no. 1, p. 104, Nov. 2022, doi: 10.1186/s40537-022-00664-6.

[20] P. Durga and D. Godavarthi, "Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN)," *IEEE Access*, vol. 11, pp. 108433–108447, 2023, doi: 10.1109/ACCESS.2023.3320738.

[21] M. M. Mathew, P. Nanjundan, and S. Bashir, "Sentiment Analysis of Online Hotel Reviews Employing Bidirectional GRU with Attention Mechanism," in *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, 2024, pp. 237–242. doi: 10.1109/ESIC60604.2024.10481617.

[22] D. Chi, T. Huang, Z. Jia, and S. Zhang, "Research on sentiment analysis of hotel review text based on BERT-TCN-BiLSTM-attention model," *Array*, vol. 25, p. 100378, Mar. 2025, doi: 10.1016/j.array.2025.100378.

[23] S. J. Hegde, H. M. Madhunandana, and Mohana, "Sentiment Analysis with LSTM Recurrent Neural Network Approach for Movie Reviews using Deep Learning," *2023 3rd International*

Vol. 6, No. 5, October 2025, Page. 3652-3666
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5185

Conference on Innovative Mechanisms for Industry Applications (ICIMIA), pp. 463–468, Dec. 2023, doi: 10.1109/ICIMIA60377.2023.10426266.

[24] Y. A. Singgalen, "BiLSTM-LSTM Hybrid Model with Glove Embeddings for Hotel Review Sentiment Analysis," *Journal of Information System Research*, vol. 6, no. 2, pp. 857–868, 2025, doi: 10.47065/josh.v6i2.6420.

[25] M. R. R. Rana, A. Nawaz, T. Ali, A. M. El-Sherbeeny, and W. Ali, "A BiLSTM-CF and BiGRU-based Deep Sentiment Analysis Model to Explore Customer Reviews for Effective Recommendations," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11739–11746, Oct. 2023, doi: 10.48084/etasr.6278.

[26] Z. Tao and Z. Wu, "Sentiment Analysis of Product Reviews Based on Bi-LSTM and Max Pooling," *Artificial Intelligence Technologies and Applications*, vol. 382, pp. 1046–1053, Feb. 2024, doi: 10.3233/FAIA231407.