JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 96-106

IMPLEMENTATION OF BILSTM AND INDOBERT FOR SENTIMENT ANALYSIS OF TIKTOK REVIEWS

Azziz Fachry Al Farizi*[1)], Yuliant Sibaroni[2)]
1. School of Computing, Telkom University, Bandung, Indonesia
2. School of Computing, Telkom University, Bandung, Indonesia

Article Info
Keywords: BiLSTM; Deep Learning; IndoBERT; Sentiment Analysis; TikTok;

Article history:
Received 15 Oktober 2024
Revised 17 November 2024
Accepted 1 Maret 2025
Available online 1 Maret 2025

DOI :
https://doi.org/10.29100/jipi.v10i1.5815

* Corresponding author.
Azziz Fachry Al Farizi
E-mail address:
kapolda@student.telkomuniversity.ac.id

ABSTRACT
The significant increase in users on TikTok has led to a notable rise in the number of reviews in the form of opinions given to the application. The large number of opinions can be analyzed to identify the prevailing sentiment among the community towards the application. The sentiment analysis method employing machine learning is particularly well-suited to this problem due to its practicality and efficiency. The objective of this research is to develop a model that can be utilized as a sentiment analysis tool with a high degree of accuracy. In this research, the BiLSTM algorithm, combined with IndoBERT, a pre-trained model, is employed. The BiLSTM can comprehend the interrelationships between words within a sentence in a bidirectional manner. IndoBERT is pertinent to this research as it is a model that has been fine-tuned using Indonesian language datasets from various sources on the Internet. To support this research, a scenario was created by considering various aspects when adding methods as an optimization scheme until the optimal model was identified. The outcomes of experimentation demonstrate that sentiment analysis using the BiLSTM+IndoBERT method achieved the highest accuracy, reaching 81% in the classification report test and an average accuracy of 92.03% in cross-validation testing with a total of 10 folds.

.

I. INTRODUCTION

THE internet has become a ubiquitous and indispensable tool in modern society. Social media in particular has become a crucial aspect of people's lives, serving as a primary conduit for communication, entertainment, and business [1]. TikTok, a prominent social media platform, has seen a remarkable surge in user growth over the past six years, with a 109% increase in its user base [2]. This growth rate is significantly higher than that of other social media platforms. Due to its high level of popularity, this application receives a considerable number of reviews in the form of text comments. However, the sheer volume of reviews makes it challenging for readers to discern the true meaning of many of them. To address this issue, this research proposes a sentiment analysis approach, which is considered an effective method for extracting valuable insights from a vast collection of reviews. Sentiment analysis is a stage of data collection in the form of information in the form of text. Its aim is to obtain information from opinion sentences that have a positive or negative tendency [3].

Additionally, research on sentiment analysis of TikTok reviews has been conducted using the long short-term memory (LSTM) and IndoBERTweet methods. The results of this study achieved an accuracy value of 78% for LSTM, while IndoBERTweet, a derivative of the BERT algorithm and trained using Indonesian vocabulary from tweets on Twitter, achieved prediction accuracy up to 80% [4]. Previous research has only discussed the comparison of the two methods without a combination with other methods. The accuracy obtained only reached 80% for IndoBERT. Previous research was considered in this study to discuss and develop variants of LSTM and BERT. However, rather than simply comparing the two methods, they were also combined by considering each of the advantages of these methods to complement the shortcomings between the methods in building a sentiment analysis model on data in the form of Indonesian sentences.

Related research has also demonstrated that the ConvBiLSTM model outperforms other models, achieving an accuracy of 91.13% [5]. This finding is further supported by a similar study that proposed a BiLSTM topic model, which outperformed a sentiment classification task for a benchmark [6]. Collectively, these studies indicate that the BiLSTM's ability to capture sequence, context, and long-term dependency information makes it more effective than LSTM for performing sentiment analysis. BiLSTM is capable of processing input sequences through the first two layers in a forward direction from the beginning to the end, and the other layers in a reverse direction from the

end to the beginning. The output of both layers is then combined. This allows BiLSTM to capture context from both the past and future, thereby improving the overall understanding of context in sequential data [7].

In this study, the bidirectional long short-term memory (BiLSTM) method is combined with IndoBERT, a derivative of the BERT algorithm and one of the deep learning models in natural language processing (NLP) [8]. The algorithm was selected for use in this study because, in a previous study regarding the classification of public opinion on the covid-19 vaccine in Indonesia, the prediction accuracy reached 80% while IndoBERTweet was only 68% [9]. Moreover, user opinion sentences in the form of reviews exhibit similar characteristics with the IndoBERT benchmarking dataset, which is predominantly informal. The IndoBERT model employs a transformer-based architecture for sentiment analysis. This architecture enables bidirectional text processing, allowing the model to comprehend the context of each word in the sentence. This approach offers several advantages over other methods. It is capable of capturing more complex context dependencies in Indonesian sentences, resulting in more accurate text representation and enhanced sentiment classification performance [10]. Consequently, the two methods are employed to analyze the sentiment of the TikTok social media application reviews on the Google Play Store.

The theoretical framework utilized in sentiment analysis encompasses the fundamental principles of NLP, which aims to enable computers to comprehend, interpret, and generate human language. In this context, the BiLSTM model is employed to capture context relationships from both directions within the text, thereby enhancing the model's effectiveness in understanding word order. The IndoBERT model, which is based on the transformer architecture and has been trained on the Indonesian language corpus, processes text in both directions, enabling a more comprehensive and accurate understanding of context. These models, which can capture the meaning of text and are particularly effective in sentiment classification, increase prediction accuracy based on richer context.

The objective of this research is to implement sentiment analysis methods in order to create a suitable machine learning model. This will be achieved by utilizing the branch of natural language processing that can classify opinions on sentences into positive, negative, and neutral labels automatically. The results of this research are expected to assist TikTok users in Indonesia in determining the suitability of using the TikTok application in their daily activities, such as finding entertainment and making it a communication tool. Moreover, the findings of this study are anticipated to assist TikTok companies in Indonesia in gaining a deeper comprehension of the characteristics and expectations of application users, based on the insights derived from sentiment analysis of ratings in the form of reviews on the Google Play Store.

II. Research Methods

This research has been conducted in several stages, beginning with the collection of datasets based on the established rules. The datasets then enter the preprocessing stage, during which the results are utilized for model construction using the BiLSTM method, which is integrated with IndoBERTweet. Fig. 1 illustrates the work system employed in this research.

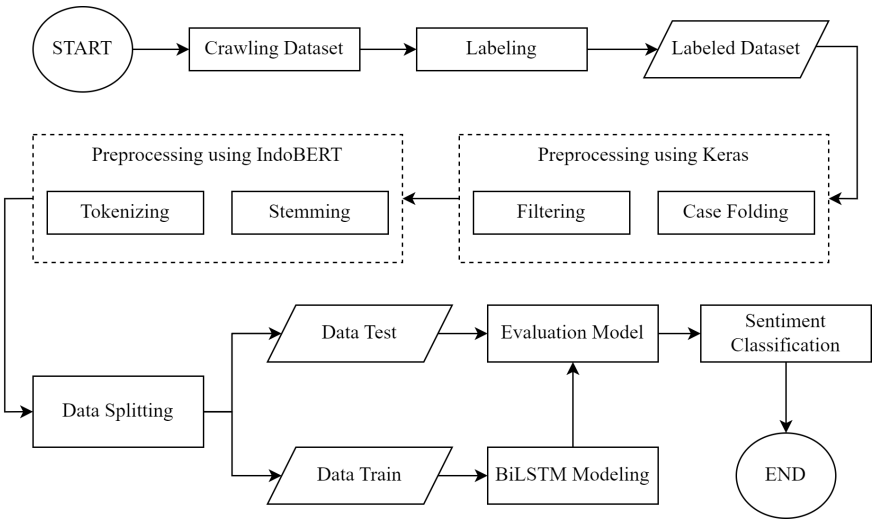

Fig. 1. Workflow of implementation BiLSTM and IndoBERT for sentiment analysis of TikTok reviews

A. Dataset

The dataset was collected through a crawling process using the Python library Google-Play-Scrapper. The downloaded dataset is limited to reviews from Indonesian servers and in Indonesian on the Google Play Store. The

total number of datasets utilized in this research is 25,291, comprising user reviews of the TikTok application. There are various features that contain information in each column in the crawling result dataset.

In this study, feature selection is employed with the objective of obtaining high-quality training data that is pertinent to the research objectives. The features themselves possess value and character that can support an experiment in a study, particularly in the field of machine learning [11]. From the information available on each feature in the dataset, researchers have determined that content features are the most appropriate for supporting this research. The selection of the “content" feature, which contains app reviews in sentiment analysis research on TikTok reviews, as training data is motivated by several factors. Primarily, app reviews directly reflect users' experiences on TikTok, which is the primary focus of sentiment analysis research. Additionally, app reviews encompass a diverse range of sentiments, including positive, negative, and neutral reviews, thereby enabling the model to be trained with representative data from various user perspectives. Consequently, selecting the “content" feature as training data can facilitate the generation of a more accurate and reliable sentiment analysis model for TikTok. Prior to model training, a preprocessing stage is employed to enhance the quality of the content features. The quality of the dataset is influenced by content that contains missing values, links, symbols, or characters other than string variables.

B. Labeling

The objective of labeling the dataset as presented in Table I is to categorize the dataset in order for the model to recognize the content as belonging to one of three sentiment categories: positive, neutral, or negative [12]. This is achieved by converting the labels, which are expressed in words, into numbers and placing them into the Tag column during the reprocessing of the dataset. This is necessary because the model is only able to understand input in the form of integer variables.

TABLE I
SAMPLE DATASET LABELING RESULT

Data	Label	Tag
Aplikasinya sungguh menarik, bisa menghibur kita	Positive	2
Saya bingung mau tulis apa	Neutral	1
Jan di update serius, kotak masuk nya jd aneh	Negative	0

C. Preprocessing using Keras

This preprocessing stage is very important for the purpose of training the model. Natural language processing, which is a branch of machine learning, is very instrumental in the research case in the form of sentiment analysis with text datasets. NLP is a tool that can make computers understand the contents of a text like humans [11]. In this research, the application of NLP in the case folding and filtering process utilizes the Keras and Sastrawi libraries. The case folding process is to convert letters containing capital letters into lowercase letters, with the aim of simplifying each character and only retaining the letters 'a-z'. This goal is also to ensure consistency and avoid discrepancies caused by capitalization. The results of case folding look like in Table II.

TABLE II
SAMPLE CASE FOLDING RESULT

Data	Result
Aplikasinya sungguh menarik, bisa menghibur kita	aplikasinya sungguh menarik bisa menghibur kita
Saya bingung mau tulis apa	saya bingung mau tulis apa
Jan di update serius, kotak masuk nya jd aneh	jan di update serius kotak masuk nya jd aneh

The filtering process in this study employs the *stopword* removal method. This process aims to enhance data quality by eliminating superfluous words, including slang terms, in a sentiment sentence while retaining sentences containing pertinent information [13]. The Sastrawi library is also employed in the filtering stage with the objective of eliminating irrelevant or common words in the text, such as conjunctions and interjections. The Sastrawi library is selected due to its extensive Indonesian language corpus, which is a valuable resource in NLP. Additionally, this stage entails transforming some slang terms into standard vocabulary. Table III presents the outcomes of the filtering stage. For a list of *stopwords* utilized in this stage refer to Table IV.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 96-106

TABLE III
SAMPLE FILTERING RESULT

Data	Result
Aplikasinya sungguh menarik, bisa menghibur kita	sungguh menarik menghibur
Saya bingung mau tulis apa	bingung tulis
Jan di update serius, kotak masuk nya jd aneh	serius kotak masuknya jadi aneh

TABLE IV
SAMPLE STOPWORD

Data
aplikasi
bisa
kita
apa
mau
serius
jangan
aku
update
di

D. Preprocessing using IndoBERT

The Bidirectional Encoder Representations from Transformers (BERT) method is a pre-training model based on neural networks in NLP [14]. BERT employs Transformers that can learn the relationship between words through a mechanism called attention. BERT comprises an encoder and a decoder, with the encoder responsible for receiving input and converting it into word vectors, while the decoder is tasked with making predictions. BERT employs three distinct embedding techniques for converting words within sentences into vectors [15]. Fig. 2 illustrates the BERT architecture.

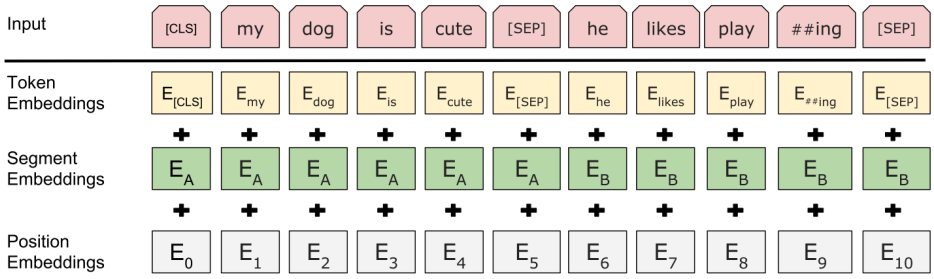

Fig. 2. BERT Architecture [16]

Token embedding in the BERT architecture serves to convert text into vectors in certain dimensions. Segment embedding is specifically used for fine-tuning and distinguishing meaning between sentences. The last part is position embedding, which functions to consider the order of the tokens in the text, while understanding the overall meaning. BERT also has several variants, one of which is IndoBERT, which is one of the pre-trained models designed to handle data in the form of sentences using Indonesian [10]. In this research, IndoBERT is employed as a tokenizer. This algorithm is analogous to BERT and comprises 12 hidden layers, each with a maximum of 786 dimensions, and 12 attention heads.

IndoBERT is also constructed using a vast array of Indonesian vocabulary, encompassing over 220 million words derived from newspapers and the Indonesian Web Corpus. This method has been developed for two months and has demonstrated good performance on several tasks within the NLP domain. Evaluation of the IndoLEM dataset indicates that IndoBERT produces more accurate results than MalayBERT [15]. This research utilizes IndoBERT without further customization, as it is well-suited to the characteristics of the data used.

The stage of processing training data in this study concludes with a tokenizing and stemming process utilizing a pre-trained model, namely IndoBERT. Stemming is a process that eliminates affixes from a word, thereby rendering it a standard word [17]. In the context of this study, the affix word is deemed to add a type of character that lacks significant information in the data training stage. The process of modifying standard words using IndoBERT differs from the traditional stemming approach. Rather than directly modifying words, IndoBERT is capable of understanding word variations and context within text, which enables it to map these words into similar numerical representations. In some cases, these representations can resemble the effects of standard words. This approach

allows the model to process words in text with greater depth, without requiring additional preprocessing steps such as stemming by searching stopword lists manually.

The tokenize process involves the separation of sentences into individual words, referred to as tokens, through a process called tokenization [18]. The token may comprise a word, a sub-word known as (n-gram), or a letter. This research employs IndoBERT as a tokenization tool due to its compatibility with BERT's CLS (classification) and SEP (separation) tokens. The CLS token is utilized as a distinctive token positioned at the outset of every text sequence, serving as a signal for the classification task. SEP tokens are employed to mark the conclusion of each text sequence or pair of separated text sequences within the BERT framework, ensuring that the model can comprehend the interrelationships between texts accurately.

IndoBERT is more effective in the tasks of stemming and tokenizing than other methods. This is due to its utilization of a transformer algorithm, which enables it to comprehend the context of Indonesian language to a greater extent. Additionally, IndoBERT has been specifically trained on a significant corpus of Indonesian texts, thus enabling it to possess a deeper understanding of Indonesian vocabulary, as well as sentence structure and variation. The results of the tokenization and stemming processes are presented in Table V.

TABLE V
SAMPLE TOKENIZING AND STEMMING RESULT

Data	Stemmed	Tokenized
sungguh menarik menghibur	sungguh tarik hibur	[‘[CLS]’, ‘sungguh’, ‘tarik’, ‘hibur’, ‘[SEP]’]
bingung tulis	bingung tulis	[‘[CLS]’, ‘bingung’, ‘tulis’, ‘[SEP]’]
serius kotak masuknya jadi aneh	kotak masuk jadi aneh	[‘[CLS]’, ‘kotak’, ‘masuk’, ‘jadi’, ‘aneh’, ‘[SEP]’]

The data is then divided into two distinct categories: training data and test data [19]. The modeling process utilizes the training data for learning purposes. A small percentage of the test data is retained for validation purposes. This validation data is employed to identify any errors that may arise during the data training process. Meanwhile, the test data is utilized to evaluate the model that has been constructed [20].

E. BiLSTM Modeling

One of the most prevalent modifications of the recurrent neural network algorithm is long short-term memory (LSTM). This LSTM method incorporates a feedback connection, which distinguishes it from standard neural networks [21]. The LSTM architecture comprises three gates, namely the forget gate (f_t), input gate (i_t), and output gate (o_t).

The gate represents a decision-making point at which important information is determined based on previously obtained information (h_{t-1}) and the value entered at that time (x_t). The result of the decision is then processed at the input gate (i_t) stage. At this stage, the process of updating information in the previous state (c_{t-1}) is carried out and then replaced by the latest information (c_t). The output gate (o_t) is used to output information that has been processed. This is achieved through the use of a sigmoid layer, which determines the cell state as a place to exit information [22]. Fig. 3 illustrates the architecture of the LSTM algorithm.

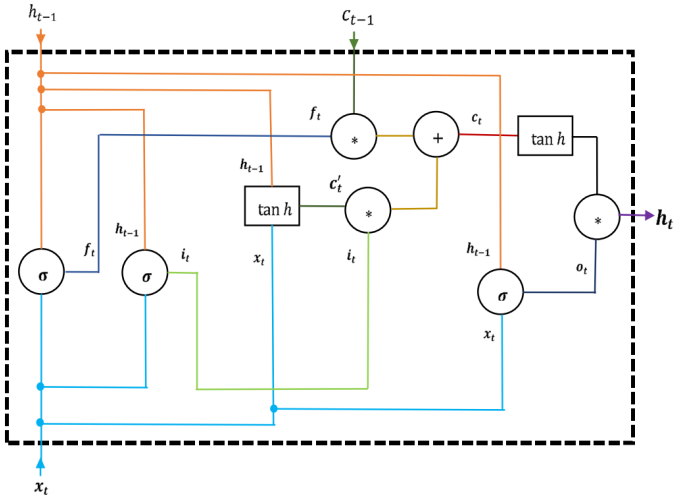

Fig. 3. LSTM Architecture [22]

The computational process of the LSTM, which is based on the three gates, is expressed in a mathematical formula. This formula represents the letters *W*, which correspond to the weight of the input value. *U* which

represents the output value. b which is a bias in a particular cell. The calculation formula shows in (1)-(6).

$$forget\ gate\ (f_t) = \sigma_g(W_f \times x_t \times U_f \times h_{t-1} + b_f) \quad (1)$$

$$input\ gate\ (i_t) = \sigma_g(W_i \times x_t \times U_i \times h_{t-1} + b_i) \quad (2)$$

$$output\ gate\ (o_t) = \sigma_g(W_o \times x_t \times U_o \times h_{t-1} + b_o) \quad (3)$$

$$cell\ memory\ (c'_t) = \sigma_g(W_c \times x_t \times U_c \times h_{t-1} + b_c) \quad (4)$$

$$cell\ state\ (c_t) = f_t \times c_{t-1} + i_t \times c'_t) \quad (5)$$

$$hidden\ state\ (h_t) = o_t \times \sigma_c(c_t) \quad (6)$$

Bidirectional long short-term memory (BiLSTM) is an RNN algorithm that can be utilized to analyze sequential data. This algorithm is a development of the LSTM model, through the modification of the algorithm's capabilities, enabling it to perform forward and backward context understanding. Additionally, BiLSTM can understand the relationship between words that are situated at disparate locations within a complex sentence [7].

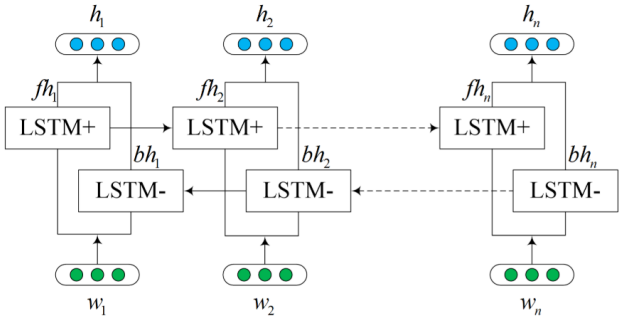

Fig. 4. BiLSTM Architecture

Fig. 4 shows the architecture of BiLSTM, which comprises a forward LSTM and a backward LSTM. The forward LSTM computes the hidden vector fh_t based on the previous vector fh_{t-1} and the input of word embedding x_t. In contrast, the backward LSTM computes the hidden vector bh_t based on the previous vector bh_{t-1} and the input of the word embedding x_t. The subsequent stage involves the combination of the two vectors, thereby forming the BiLSTM method. The illustration in Fig. 4 serves as the foundation for the BiLSTM model. The notation $\{w_1, w_2, ..., w_n\}$ represents the word in vector form, with n denoting the length of the sentence. Similarly, $\{fh_1, fh_2, ..., fh_n\}$ and $\{bh_1, bh_2, ..., bh_n\}$ represent the forward and backward hidden vectors, respectively. The relationship between fh_n and bh_n is denoted by h_n. Thus, the final hidden vector is designated as h_t and expressed in (7).

$$h_t = [fh_t, bh_t] \quad (7)$$

The BiLSTM model was created in this study by applying the SoftMax parameter as an activation function. In the context of sentiment analysis using BiLSTM, the Softmax activation function is employed in the output layer to generate a probability distribution of sentiment classes based on the outputs produced by the model [23]. Using the Adam optimizer as the model parameters, including weights and biases, during the training process by adjusting them based on the gradients of the loss function [21]. A SpatialDropout1D value of 0.4 was also applied to the modeling. The utilization of SpatialDropout1D serves to mitigate the issue of overfitting by randomly disregarding a portion of the input features, such as word dimensions, at each training iteration. This prevents the model from becoming overly reliant on specific features, thereby enhancing its generalizability.

F. Model Evaluation Method

1) Classification Reports

The outcomes of the constructed model are then assessed using the classification report method, which employs the values of the confusion matrix. This method calculates by considering the precision, recall, f1-score, and accuracy matrix of the model in making predictions and is formulated in (8)-(11). This confusion matrix comprises true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

$$precision = TP/(TP + FP) \quad (8)$$

$$recall = TP/(TP + FN) \quad (9)$$

$$f1 = (2\ x\ recall\ x\ precision)/(recall + precision) \quad (10)$$

$$accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (11)$$

Precision is an evaluation metric used to assess the model's ability to accurately predict positive classes, expressed as a proportion of the total number of positive predictions. Sensitivity (recall) is a metric used to evaluate the model's performance in identifying positive classes correctly. F1-Score is an evaluation metric that describes the balance between precision and recall. Accuracy is a calculation used to assess the model's ability to make accurate predictions [24].

2) Cross Validation

Another method employed to assess the model developed in this study is k-fold cross-validation. This statistical technique partitions data into subsets for training and validation purposes [25]. The objective of this method is to generalize the data effectively and avoid overfitting, thereby facilitating the identification of the optimal model in each scenario within this study.

The fundamental principle of cross-validation is the partitioning of the data set *D* into several subsets or *k-folds*. The subset data *i* is designated by D_i, representing the totality of all data *D*. In this study, the *k* value employed is 10 folds, with each fold subjected to 10 *epochs*. The mean accuracy value and f1-score are calculated from the individual accuracy values of each fold to assess the performance of the model in each evaluation scenario. The mathematical formula for data division and the calculation of the mean value of each fold are shown in (12) and (13), respectively.

$$D = \{D_1, D_2, \dots, D_k\} \quad (12)$$

$$Mean\ Evaluation\ Metric = \frac{1}{k}\sum_{i=1}^{k} Evaluation\ Metric_i \quad (13)$$

III. Result And Discussion

A. Model Evaluation

In this test, three scenarios were created as outlined in Table VI. These scenarios are employed to ascertain the optimal method based on the highest accuracy. In each scenario, several tests were conducted under specified conditions to obtain the optimal dataset, which was then utilized for model training.

TABLE VI
Model Scenario

Scenario	Model
1	LSTM
2	Bidirectional + LSTM
3	Bidirectional + LSTM + IndoBERT

Table VI in the first scenario will utilize the LSTM method to determine the accuracy and initial ratio. In scenario two, the LSTM method is integrated with the bidirectional method, with the aim of enabling the model to process data in both directions. In the third scenario, IndoBERT is employed to enhance the data quality and accuracy of the preceding two scenarios.

TABLE VII
Implementation of Indonesian Text Preprocessing Methods

Dataset	Preprocessing Method
1	Sastrawi + Keras Toenizer *(Traditional)*
2	IndoBERT

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 96-106

In accordance with the scenarios that have been prepared, this research employs the same dataset, albeit with a slightly different preprocessing stage, as detailed in Table VII. The first dataset is a dataset that has not been processed using a pre-trained model. The dataset employs Sastrawi as a stemming tool and the Keras as a tokenizer. The second dataset has been processed using IndoBERT and will be utilized in the third scenario of this study.

1) Scenario 1

In scenario 1, the objective is to determine the optimal dataset ratio for modeling this research. Additionally, the efficacy of the standard LSTM method in modeling with these datasets across a range of predetermined ratios will be evaluated. The dataset is divided into three ratios, comprising training data and test data, with a ratio of 70:30, 80:20, and 90:10. The determination of the optimal dataset ratio is based on the scenario design, utilizing the standard LSTM method with test results presented in Table VIII.

TABLE VIII
CLASSIFICATION REPORT TEST RESULTS BASED ON DATASET RATIO

Model	Ratio	Metrics			
		Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
1	70:30	75.35	75.13	75.23	75.13
2	80:20	**75.45**	**75.35**	**75.36**	**75.35**
3	90:10	74.38	74.02	74.17	74.02

The cross-validation method is employed to validate the model by utilizing training data as test data. The objective is to obtain precise results in determining the optimal ratio. The outcomes of the cross-validation implementation for the first scenario are presented in Table IX.

TABLE IX
CROSS VALIDATION TEST RESULTS

Model	Fold (%)										Avg	Avg
	1	2	3	4	5	6	7	8	9	10	Accuracy (%)	F1-Score (%)
1	76.63	84.08	88.71	90.57	91.53	91.81	93.51	93.56	93.28	93.06	89.67	89.69
2	79.5	84.14	87.75	89.82	91.75	91.6	92.39	93.23	93.48	93.63	**89.73**	**89.78**
3	80.5	83.31	86.82	88.76	89.46	91.35	91.39	92.05	92.89	93.5	89	89.05

The test results, which were generated using the optimal parameters for the standard LSTM method, indicate that the dataset ratio of 80:20, where 80% is train data and 20% is test data, yields the most favorable metric values compared to other ratios. This is evident in Table VIII. The results of k-fold cross-validation testing also indicate that model 2, with a total of 10 folds, achieved an average accuracy value of 89.73% and an f1-score of 89.78%. These results are reflected in the test tables, where the 80:20 ratio was used to test comparisons in the second scenario.

2) Scenario 2

In the second scenario, the research employs a bidirectional layer to enhance the standard LSTM method. The incorporation of the layer facilitates the model's capacity to learn the training data in two directions, thereby enhancing its ability to extract information from the dataset. The bidirectional long short-term memory model constructed in this sentiment analysis research employs the SoftMax activation function in the output layer. This function enables the model to provide an appropriate class probability based on each input, given that the research encompasses datasets with positive, neutral, and negative sentiment classes.

Related research has demonstrated the efficacy of the SoftMax function in the BERT-BiGRU model for sentiment analysis of e-commerce product reviews, achieving notable success in classifying negative and positive sentiments [23]. This research serves as a valuable reference for the utilization of the SoftMax activation function, with the incorporation of a dropout value to mitigate the risk of overfitting in the model. The results of the BiLSTM model training are presented in Table X. These results demonstrate an increase of 0.19% in the precision value and an increase of 0.06% in the f1-score value when compared to the standard LSTM method.

TABLE X
CLASSIFICATION REPORT RESULTS IN THE SECOND SCENARIO

Method	Metrics			
	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
LSTM	75.45	75.35	75.36	75.35
Bidirectional + LSTM	**75.64 (+0.19%)**	75.35	**75.42 (+0.06%)**	75.35

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 96-106

The *k-fold* cross-validation method was also employed in the second scenario with the objective of conducting a more rigorous validation process. This was because the discrepancy in test outcomes with regard to classification report metrics was not particularly pronounced between the two models that had been developed. Table XI illustrates the outcomes of cross-validation testing, indicating that the average accuracy and f1-score of the BiLSTM model are superior to those of the LSTM method. This improvement occurs because bidirectional layers help the model understand more complex contexts through bidirectionality.

TABLE XI
CROSS VALIDATION TEST RESULTS IN THE SECOND SCENARIO

Method	Fold (%)										Avg Accuracy (%)	Avg F1-Score (%)
	1	2	3	4	5	6	7	8	9	10		
LSTM	79.5	84.14	87.75	89.82	91.75	91.6	92.39	93.23	93.48	93.63	89.73	89.78
Bidirectional + LSTM	78.66	84.63	90.32	92.49	93.63	93.77	95.01	95.21	94.96	95.21	**91.39**	**91.46**

3) Scenario 3

The final scenario in this research is to implement the dataset that has been processed using IndoBERT. This research applies IndoBERT to the construction of LSTM and BiLSTM models to ascertain the impact of one of the pre-trained models. Table XII demonstrates that the method integrated with IndoBERT has increased in each metric value. The greatest increase was observed in the integration between BiLSTM and IndoBERT, with an accuracy of 81.03%.

TABLE XII
CLASSIFICATION REPORT RESULTS IN THE THIRD SCENARIO

Method	Metrics			
	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
LSTM	75.45	75.35	75.36	75.35
LSTM + IndoBERT	79.09	79.25	78.83	79.25
Bidirectional + LSTM	75.64	75.35	75.42	75.35
Bidirectional + LSTM + IndoBERT	**81**	**81.03**	**80.88**	**81.03**

The validation of the models that have been constructed is also evaluated using the cross-validation method to ascertain the performance of these models. As illustrated in Table XIII, the BiLSTM method integrated with IndoBERT not only excels at evaluating metrics using the classification report, but also exhibits an average accuracy of 92.03% and an average f1-score of 92.01% in cross-validation testing.

TABLE XIII
CROSS VALIDATION TEST RESULTS IN THE THIRD SCENARIO

Method	Fold (%)										Avg Accuracy (%)	Avg F1-Score (%)
	1	2	3	4	5	6	7	8	9	10		
LSTM	79.5	84.14	87.75	89.82	91.75	91.6	92.39	93.23	93.48	93.63	89.73	89.78
LSTM + IndoBERT	79.25	85.87	87.84	88.98	91.25	92.49	93.43	94.07	95.4	95.3	90.39	90.37
Bidirectional + LSTM	78.66	84.63	90.32	92.49	93.63	93.77	95.01	95.21	94.96	95.21	91.39	91.46
Bidirectional + LSTM + IndoBERT	77.82	86.96	90.21	92.14	93.18	94.56	95.55	95.95	96.84	97.13	**92.03**	**92.01**

The integration of BiLSTM and IndoBERT in sentiment analysis of TikTok reviews provides significant advantages compared to using a single LSTM. BiLSTM, with bidirectional processing capabilities, can better capture text context by considering the order of words from the past and future. On the other hand, IndoBERT has been trained on Indonesian texts, provides a richer representation of words by understanding the meaning and structure of the language. By combining the strengths of both, models can produce more accurate sentiment predictions and generally show improvements in evaluation metrics such as accuracy and f1-score. This proves that this integrated approach is more effective in capturing the complexity of sentiment in TikTok reviews compared to using a single LSTM.

The results of the scenarios applied in this research, along with suggestions from previous research aimed at improving model performance, indicate that a BiLSTM combined with IndoBERT is a more effective approach than a single LSTM method, which achieves an accuracy of 78% and IndoBERT, which achieves 80% prediction accuracy without integration with other methods [4]. This is evidenced by the findings of this research, which demonstrate that the accuracy of the devaluation using the classification report reaches a predicted accuracy value of 81.03%, while the evaluation using k-fold cross validation reaches 92.03%.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 96-106

B. Result Analysis

This research employs two of the most effective models derived from a series of simulations to assess the precision of the predictions. Additionally, it utilizes the confusion matrix in Fig. 5 to illustrate the accuracy of the model in anticipating user sentiment towards the TikTok application. The results demonstrate that the model has a low error rate in predicting user sentiments. Furthermore, the analysis reveals that users in Indonesia tend to express more positive sentiments regarding the TikTok application. Negative sentiments are observed to occur at a frequency of no more than 50%. However, further analysis is required to gain insight into the specific reasons behind user dissatisfaction with one of these social networking applications.

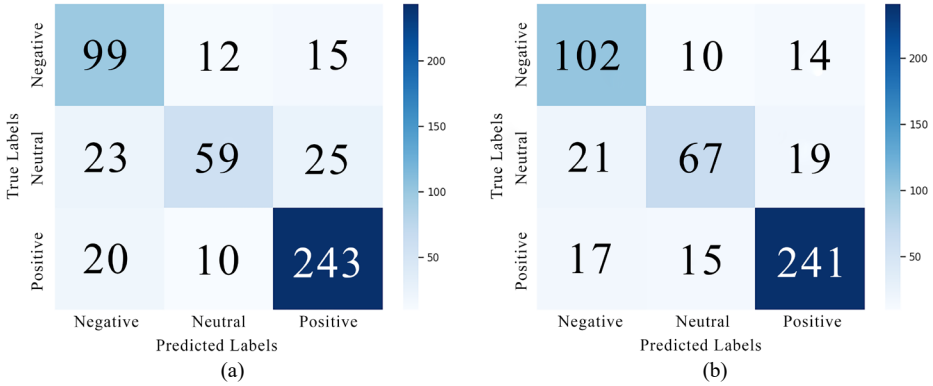

Fig. 5. Confusion Matrix Model Evaluation Results (a) LSTM+IndoBERT and (b) BiLSTM+IndoBERT

A review of the results of the sentiment analysis conducted on the TikTok review using the two most effective models reveals that model 5(a) correctly identified 99 instances of negative sentiment, incorrectly identified 23 instances that should have been neutral, and incorrectly identified 20 instances of positive sentiment. The model was then tested on neutral sentiment, correctly predicting 59 data points, incorrectly predicting 12 data points as negative, and incorrectly predicting 10 data points as positive. This model is effective in that it correctly predicts positive sentiment on 243 data points, incorrectly predicts negative sentiment on 15 data points, and incorrectly predicts neutral sentiment on 25 data points. Model 5(b) demonstrated an ability to correctly predict negative sentiment on 102 data points, while incorrectly predicting neutral sentiment on 21 data points and positive sentiment on 17 data points. The prediction results for negative sentiment towards true labels were 67 data points, with 10 data points incorrectly predicted as neutral and 15 data points incorrectly predicted as positive. In the results of positive sentiment predictions, the model shows 241 positive data results from correct predictions, 19 data should be neutral, and 14 data should be negative. When comparing the prediction results of the two models, it can be seen that TikTok users still have a positive sentiment towards this application.

IV. CONCLUSION

This research discusses the creation of a machine learning model that can be used to perform classification in the form of sentiment analysis of TikTok application users in Indonesia. The classification is based on positive, neutral, and negative sentiments. The dataset used for training the model consisted of 25,291 entries. The model was developed using a combination of the bidirectional long short-term memory method and the IndoBERT method. To validate the research, three scenarios were compared. In each scenario, the addition of appropriate methods is carried out with the aim of determining the performance of the model with the optimal level of prediction accuracy. The base in this research employs the Long Short-Term Memory (LSTM) method because it has the advantage of being able to remember sequential information over an extended period and can replace information that is no longer relevant with new information. The LSTM method is combined with a bidirectional layer to produce more accurate information when training data. The role of IndoBERT in this research is to serve as a tool for pre-processing datasets. This method is one of the pre-trained models that has been trained with Indonesian language benchmarking datasets. Based on the test results, the integration of other methods to the LSTM base in modeling sentiment analysis is quite appropriate. The BiLSTM+IndoBERT model exhibited a notable enhancement in performance, with an accuracy value of 81% on the classification report and an average accuracy of 92.03% on the cross-validation test results.

The author's suggestion for further research is that preprocessing needs to pay more attention to the contents of the dataset, because some reviews contain double meanings. For this reason, further handling needs to be done to improve the quality of the data to be trained. The author also hopes that future research will implement optimization

and addition of datasets from other sources. With the aim that the model can be more accurate in performing prediction tasks on sentiment analysis of various kinds of user reviews of the TikTok application.

REFERENCES

[1] P. A. Permatasari, L. Linawati, and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, pp. 177–186, Dec. 2021, doi: 10.24843/mite.2021.v20i02.p01.

[2] L. Stappen, A. Baird, E. Cambria, B. W. Schuller, and E. Cambria, "Sentiment Analysis and Topic Recognition in Video Transcriptions," *IEEE Intell Syst*, vol. 36, no. 2, pp. 88–95, Apr. 2021, doi: 10.1109/MIS.2021.3062200.

[3] O. Somantri and D. Apriliani, "Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 5, pp. 537–548, Oct. 2018, doi: 10.25126/jtiik.201855867.

[4] J. C. Setiawan, K. M. Lhaksmana, and B. Bunyamin, "Sentiment Analysis of Indonesian TikTok Review Using LSTM and IndoBERTweet Algorithm," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 774–780, 2023, doi: 10.29100/jipi.v8i3.3911.

[5] S. Tam, R. Ben Said, and Ö. Tanriöver, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification," *IEEE Access*, vol. 9, pp. 41283–41293, 2021, doi: 10.1109/ACCESS.2021.3064830.

[6] Y. Huang, Y. Jiang, T. Hasan, Q. Jiang, and C. Li, "Topic BiLSTM model for sentiment classification," *ACM International Conference Proceeding Series*, vol. Part F1376, pp. 143–147, 2018, doi: 10.1145/3194206.3194240.

[7] J. Xie, B. Chen, X. Gu, F. Liang, and X. Xu, "Self-Attention-Based BiLSTM Model for Short Text Fine-Grained Sentiment Classification," *IEEE Access*, vol. 7, pp. 180558–180570, 2019, doi: 10.1109/ACCESS.2019.2957510.

[8] R. Mas, R. W. Panca, K. Atmaja, and W. Yustanti, "Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *JEISBI*, vol. 2, no. 3, p. 2021, Jul. 2021, [Online]. Available: ejournal.unesa.ac.id/index.php/JEISBI/article/view/41567

[9] S. Saadah, Kaenova Mahendra Auditama, Ananda Affan Fattahila, Fendi Irfan Amorokhman, Annisa Aditsania, and Aniq Atiqi Rohmawati, "Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 648–655, 2022, doi: 10.29207/resti.v6i4.4215.

[10] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.

[11] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed Tools Appl*, pp. 1–32, Jul. 2022, doi: 10.1007/s11042-022-13428-4.

[12] Q. Bi, K. E. Goodman, J. Kaminsky, and J. Lessler, "What Is Machine Learning: a Primer for the Epidemiologist Qifang," *Am J Epidemiol*, vol. 188, no. 12, pp. 2222–2239, Dec. 2019, doi: https://doi.org/10.1093/aje/kwz189.

[13] A. A. V. A. Jayaweera, Y. N. Senanayake, and P. S. Haddela, "Dynamic Stopword Removal for Sinhala Language," in *2019 National Information Technology Conference (NITC)*, Oct. 2019, pp. 1–6. doi: 10.1109/NITC48475.2019.9114476.

[14] K. S. Nugroho, A. Y. Sukmadewa, H. Wuswilahaken Dw, F. A. Bachtiar, and N. Yudistira, "BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews," *ACM International Conference Proceeding Series*, pp. 258–264, 2021, doi: 10.1145/3479645.3479679.

[15] B. Juarto and Yulianto, "Indonesian News Classification Using IndoBert," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 454–460, 2023.

[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[17] D. Khyani, S. B. S, N. N. M, and D. B. M, "An Interpretation of Lemmatization and Stemming in Natural Language Processing," *Journal of University of Shanghai for Science and Technology*, vol. 22, no. 10, pp. 350–357, Oct. 2021, [Online]. Available: https://www.researchgate.net/publication/348306833

[18] S. Ahmadi, "A Tokenization System for the Kurdish Language," *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*, pp. 114–127, 2020, [Online]. Available: https://aclanthology.org/2020.vardial-1.11

[19] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons.B*, vol. 4, no. December 2017, pp. 51–62, 2017, doi: 10.20544/horizons.b.04.1.17.p05.

[20] E. Breck, N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "Data Validation for Machine Learning," *Proceedings of Machine Learning and Systems 1 (MLSys 2019)*, pp. 334–347, 2019, [Online]. Available: https://proceedings.mlsys.org/paper/2019/file/5878a7ab84fb43402106c575658472fa-Paper.pdf

[21] Dr. G. S. N. Murthy, S. R. Allu, B. Andhavarapu, M. Bagadi, and M. Belusonti, "Text based Sentiment Analysis using LSTM," *International Journal of Engineering Research and*, vol. V9, no. 5, pp. 299–303, May 2020, doi: 10.17577/ijertv9is050290.

[22] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artif Intell Rev*, vol. 53, no. 6, pp. 4335–4385, 2020, doi: 10.1007/s10462-019-09794-5.

[23] Y. Liu, J. Lu, J. Yang, and F. Mao, "Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax," *Mathematical Biosciences and Engineering*, vol. 17, no. 6, pp. 7819–7837, 2020, doi: 10.3934/MBE.2020398.

[24] E. Beauxis-aussalet and L. Hardman, "Visualization of Confusion Matrix for Non-Expert Users," in *IEEE Information Visualization (InfoVis 2014)*, 2014.

[25] J. M. Gorriz, F. Segovia, J. Ramirez, A. Ortiz, and J. Suckling, "Is K-fold cross validation the best model selection method for Machine Learning?," no. Ml, 2024, [Online]. Available: http://arxiv.org/abs/2401.16407