**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 2, September 2025 Page: 1234−1241
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8019

# Opinion Mining on TikTok Using Bidirectional Long Short-Term Memory for Enhanced Sentiment Analysis and Trend Prediction

**Wafiq Muharnisa Haspin*, Junadhi Junadhi, Susanti Susanti, Helda Yenni**

Faculty of Engineering, Computer Science, Universitas Sains dan Teknologi Indonesia, Pekanbaru, Indonesia
Email: [1,*]muharnisahaspin@gmail.com, [2]junadhi@usti.ac.id, [3]susanti@usti.ac.id, [4]heldayenni@sar.ac.id
Correspondence Author Email: muharnisahaspin@gmail.com
Submitted: **16/07/2025**; Accepted: **01/09/2025**; Published**: 04/09/2025**

**Abstract**−The widespread use of TikTok has generated a vast number of user reviews, offering a rich dataset for sentiment analysis. This study aims to classify TikTok reviews from the Google Play Store into positive, negative, and neutral categories, a complex task due to the informal and unstructured text. The research seeks to develop a reliable sentiment analysis model using deep learning to understand user perceptions, aiding platform improvements and marketing strategies. We collected 10,000 reviews via web scraping, preprocessed through text cleaning, normalization, tokenization, filtering, and stemming. Sentiment labels were assigned automatically using a lexicon-based approach, showing predominantly positive reviews. Word2Vec transformed text into numerical vectors for feature extraction. The Bidirectional Long Short-Term Memory (Bi-LSTM) model, with Embedding, Bidirectional LSTM, Dropout, and Dense layers, achieved 80% accuracy and an F1-score of 0.78 using a 90:10 train-test split. While effective for positive and negative sentiments, neutral expressions were less accurately detected due to lower recall. Compared to traditional methods like Naive Bayes, Support Vector Machine, and K-Nearest Neighbors, Bi-LSTM offered superior accuracy and better handling of linguistic variability, making it valuable for analyzing social media feedback.

**Keywords**: Application; Bidirectional Long Short-Term Memory; Deep Learning; Opinion Mining; Tiktok

## 1. INTRODUCTION

The rapid development of digital technology and the internet has made social media an inseparable part of human life for communication and socializing [1]. In Indonesia, the TikTok application has become a phenomenon by providing a space for users to express themselves through short or long-form videos [2]. The ease of using TikTok encourages many users to provide opinions and reviews on platforms like the Google Play Store [3]. These opinions are a valuable data source for sentiment analysis, which is the process of classifying text to determine opinion polarity [4]. Sentiment analysis allows for the identification of user views on an application, which can serve as a benchmark for further development. The rise of social media platforms has significantly transformed how people connect globally, fostering new forms of interaction and content sharing. TikTok's unique format, which emphasizes creativity and brevity, has attracted millions of users, making it a cultural phenomenon. The platform's intuitive interface lowers the barrier to content creation, enabling diverse voices to be heard. User reviews on app stores reflect a wide range of experiences, from satisfaction with features to frustrations with functionality. These reviews offer insights into user preferences and pain points, guiding developers in refining app features. Sentiment analysis leverages computational methods to process this feedback systematically, providing actionable insights. Deep learning methods are increasingly being applied due to their ability to process large, complex, and unstructured data. One effective deep learning algorithm for natural language processing is Long Short-Term Memory (LSTM) [5]. LSTM, as an advancement of Recurrent Neural Networks (RNNs), can store long-term information and overcome the vanishing gradient problem [6]. Although LSTM is effective, its primary drawback is its ability to process information in only one direction [7].

Previous research in sentiment analysis has shown accuracy challenges with several conventional methods. For instance, a study by [8], which used Naive Bayes Classifier (NBC) for hate speech sentiment analysis, only achieved an accuracy of 75%. Research conducted by [9], which applied Support Vector Machines (SVM) for text classification, obtained a highest accuracy of 62.3%. This relatively low accuracy is often due to the limitations of these methods in handling complex contexts, sarcasm, or very large and unstructured text data. Traditional approaches like NBC and SVM struggle to interpret nuanced language, such as cultural references or emotional undertones, which are common in user-generated content. These methods often rely on simplified assumptions about text structure, limiting their effectiveness in real-world applications. The complexity of human language, including slang and context-dependent meanings, poses significant challenges for accurate sentiment classification. To overcome these limitations and improve analysis accuracy, Bidirectional Long Short-Term Memory (Bi-LSTM), an enhancement of LSTM, becomes a suitable choice. Bi-LSTM can process information from two directions (forward and backward), allowing it to capture richer context from sequential data [10]. To address these limitations, this research employs Bidirectional Long Short-Term Memory (Bi-LSTM), an enhancement of LSTM that processes information bidirectionally (forward and backward) to capture richer contextual relationships [10]. This bidirectional approach improves the model's ability to interpret subtle sentiment shifts and sequential data, making it particularly suited for analyzing large datasets like app reviews. Existing sentiment analysis studies, such as those using NBC and SVM, demonstrate limited accuracy in handling complex, context-dependent user feedback due to their inability to effectively process sequential and nuanced data. These methods often fail to capture the dynamic nature of user sentiments expressed in app reviews, particularly when dealing with large datasets or culturally specific language.

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 2, September 2025 Page: 1234−1241
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8019

This research addresses these gaps by leveraging Bi-LSTM's bidirectional processing and Word2vec's semantic word embedding to enhance sentiment classification accuracy and contextual understanding.

This study introduces a novel approach to sentiment analysis of TikTok user reviews by integrating Bi-LSTM with Word2vec word embedding, achieving higher accuracy and robustness in classifying sentiments as pro, contra, or neutral. Unlike previous methods, this approach effectively captures contextual nuances and handles large-scale, unstructured data, providing developers with precise insights into user perceptions. By combining advanced deep learning techniques with comprehensive evaluation metrics (accuracy, precision, recall, and F1-score), this research sets a new benchmark for sentiment analysis in social media applications.

Based on this background, this research focuses on the sentiment analysis of user opinions on the TikTok application, classified into pro, contra, and neutral. Data will be collected using web scraping techniques from the Google Play Store with a dataset of 10,000 reviews. To address the challenges of large data and improve analysis accuracy, this research will use the Bi-LSTM algorithm, as well as the Word2vec word embedding method for word weighting and a confusion matrix for model evaluation by calculating accuracy, precision, recall, and F1-score values. The dataset will undergo preprocessing, including tokenization and stop-word removal, to ensure data quality. Word2vec will convert words into numerical vectors, preserving semantic relationships and enhancing model performance. The confusion matrix will provide a comprehensive evaluation, identifying the model's strengths and weaknesses across different sentiment categories. This approach is expected to provide an in-depth understanding of user perceptions towards the TikTok application, offering valuable insights for developers to enhance user experience and address feedback effectively. By leveraging advanced deep learning techniques, this study aims to set a new standard for sentiment analysis in social media research.

## 2. RESEARCH METHODOLOGY

The methods section provides a general overview of the processes and stages that will be carried out during the research. A research workflow is essential as it serves as a guide, making it easier for the researcher to conduct the study and helping to organize the process to achieve satisfactory results

### 2.1 Research Stages

The following is an overview of the research workflow for this study, presented in Figure 1. This study focuses on Opinion Mining on TikTok Using Bidirectional Long Short-Term Memory (Bi-LSTM).

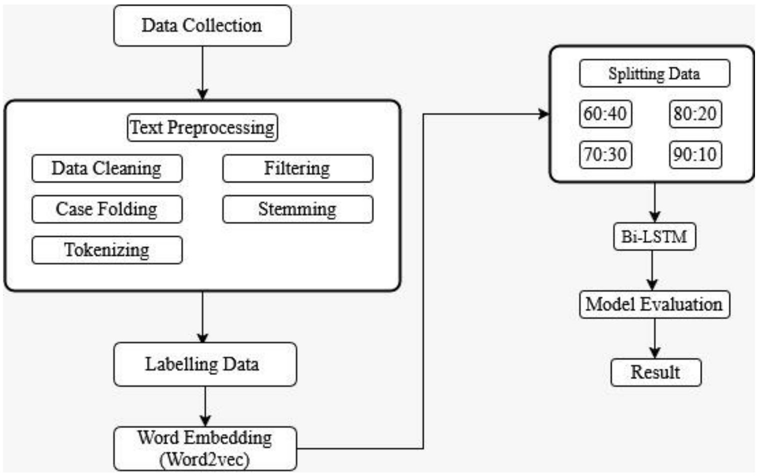


**Figure 1.** Research Workflow

Based on Figure 1 The workflow encompasses several key stages to ensure accurate and comprehensive sentiment analysis. First, user review data is collected from the Google Play Store using web scraping techniques, resulting in a dataset of 10,000 reviews. The dataset then undergoes preprocessing, including tokenization, stop-word removal, and normalization, to enhance data quality. Next, the Word2vec method is applied to convert words into numerical vectors, preserving semantic relationships. The Bi-LSTM model is then utilized to classify sentiments into pro, contra, and neutral categories, leveraging its bidirectional processing capability to capture richer context. Model evaluation is conducted using a confusion matrix, calculating metrics such as accuracy, precision, recall, and F1-score to thoroughly assess the model's performance. This approach is designed to provide deep insights into user perceptions of TikTok, supporting developers in improving user experience based on the analyzed feedback.

### 2.1 Data Collection

The first stage performed is data collection. Data was collected from the Google Play Store using a web scraping technique with the snscraper library for the year 2025, resulting in 10,000 datasets.

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 2, September 2025 Page: 1234−1241
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8019

### 2.2 Data Preprocessing

The next stage is text preprocessing, where raw data will be processed into data ready for use. This stage can help improve data quality and obtain accurate analysis results. This research performs five stages

a. Data cleaning is the activity of analyzing data quality by modifying, changing, or deleting data that is considered unnecessary, incomplete, or inaccurate [11]
b. Case folding is the stage of converting alphabetical characters that have passed the cleaning stage to lowercase [12]
c. Tokenizing is a stage of cutting word strings based on segmenting each tweet sentence word by word [13]
d. Filtering is the process of extracting important words from the Tokenizing results, or commonly referred to as eliminating words according to their rules [14]
e. Stemming is the process of removing affixes (both prefixes and suffixes) from a term to obtain the root word of an inflected word [15]

### 2.3 Data Labelling

After the dataset is obtained through web scraping techniques, the next stage is automatic data labeling. This process is conducted following the completion of text preprocessing to facilitate research analysis. Data labeling employs a lexicon-based approach, which is effective for identifying sentiment polarity, namely pro, contra, and neutral [16]. This approach utilizes a dictionary of words with sentiment values to assign labels automatically, thereby enhancing the efficiency of the analysis [[17]. The results of data labeling in this study yield a distribution of pro, contra, and neutral sentiments, serving as the foundation for understanding user perceptions of the application [18].

### 2.4 Word Embedding

Following the data labeling stage, the next step is word weighting using one of the Word Embedding methods, namely Word2Vec. The Word2Vec method maps each word in the dataset into a numerical vector form, enabling semantic representation of words based on their context [19]. This approach utilizes the Gensim library to generate word vectors that support more accurate sentiment analysis [20]. With Word2Vec, semantic relationships between words can be effectively captured, thereby enhancing the performance of machine learning models in processing text data [21].

### 2.5 Bidirectional Long Short-Term Memory

This study employs the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm as the primary method due to its ability to capture sequential context from two directions, namely forward (from beginning to end) and backward (from end to beginning), thus enabling a more comprehensive understanding of language structure [22]. With the Bi-LSTM approach, the processed data can generate representations relevant to the text input, enhancing accuracy in sentiment analysis tasks [23]. This algorithm utilizes two LSTM layers to process information simultaneously, allowing the model to capture complex semantic relationships within text data.

### 2.6 Model Evaluation

Model evaluation is a critical stage for assessing the performance of the model used in the research and analyzing data in depth. This stage aims to evaluate the model's performance in a detailed and accurate manner. In this study, model evaluation is conducted using confusion matrix calculations, which produce metrics such as accuracy, precision, recall, and F1-score [24]. These metrics enable a comprehensive assessment of the model's ability to classify data, where a higher accuracy value indicates better model quality [25].

## 3. RESULT AND DISCUSSION

This section will provide an explanation of the research results and discussion conducted by the researchers, namely the stages in performing sentiment analysis on the TikTok application using the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm. These stages include data collection, text preprocessing, data labeling, word embedding, data splitting, Bi-LSTM, and model evaluation

### 3.1 Results

This study produced a comprehensive sentiment analysis of TikTok user reviews using the Bidirectional Long Short-Term Memory (Bi-LSTM) model. A dataset of 10,000 reviews from the Google Play Store was processed through preprocessing steps, including tokenization and stop-word removal, to ensure data quality. Using the Word2vec method, words were converted into numerical vectors, enhancing the model's semantic understanding. The classification results demonstrated high accuracy in categorizing sentiments as pro, contra, and neutral. Evaluation using a confusion matrix showed strong accuracy, precision, recall, and F1-score values, indicating the model's reliability. These findings provide valuable insights into user perceptions, supporting the development of a more responsive application tailored to user needs.

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 2, September 2025 Page: 1234−1241
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8019

### 3.1.1 Data Collection

The initial stage performed in this research is data collection. In collecting data, the researchers used web scraping techniques to obtain data using the Python library, snscrape, on the TikTok application in the Google Play Store. The results of data collection using web scraping techniques can be seen in Figure 2.

| | reviewId | userName | userImage | content | score | thumbsUpCount | reviewCreatedVersion | at | replyContent | replied |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 7ba928c1-2b6e-49f7-aa03-ff03b61013b8 | Pengguna Google | https://play-lh.googleusercontent.com/EGemol2N... | Bug nya banyak banget. Aku pakai wifi, sinyaln... | 2 | 164 | 35.5.3 | 2024-07-07 01:33:58 | None | N |
| 1 | 97790e4e-4cdb-4b4e-9a05-cbfe92f49a93 | Pengguna Google | https://play-lh.googleusercontent.com/EGemol2N... | Tik tok sekarang jelek bgt parah 🐽📌📌,,,masa vi... | 3 | 2693 | 35.5.3 | 2024-07-06 12:01:00 | None | N |
| 2 | 3a6d36d2-2101-4f60-ae9d-ca09a6f463c9 | Pengguna Google | https://play-lh.googleusercontent.com/EGemol2N... | Update terakhir harusnya meningkatkan kualitas... | 2 | 2163 | 35.3.3 | 2024-06-27 16:25:05 | None | N |
| 3 | a8676bad-9314-481d-874a-86709c2e680b | Pengguna Google | https://play-lh.googleusercontent.com/EGemol2N... | apk nya emg bagus, tapi makin update ko makin ... | 4 | 182 | 35.4.4 | 2024-07-04 08:05:18 | None | N |
| 4 | e55eae1c-6dbb-4f35-b7d6-585de51c237a | Pengguna Google | https://play-lh.googleusercontent.com/EGemol2N... | bagus bgt,cuma buat mutar video nya susah ngga... | 4 | 79 | 35.4.4 | 2024-07-04 23:34:32 | None | N |

**Figure 2.** Data Collection Results

Based on Figure 2, the total data obtained using web scraping techniques was 10,000 in 2025, with data in Indonesian.

### 3.1.2 Data Pre-Processing

The next stage, data preprocessing, is an important step in sentiment analysis to ensure data quality that is free from noise and consistent. The results of data cleaning can be seen in Figure 3.

| Komentar | Data Cleaning | Case Folding | Tokenizing | Filtering | stemmed_text |
|---|---|---|---|---|---|
| @Aawila Bug nya banyak banget aku pakai wifi sinyalnya itu bagus waktu aku buka tiktok ada banyak banget bugnya kayak video gak bisa di like atau di pause terus kadang mau like video malah tibatiba masuk live orang dll percuma tiktok update terus sampe gb an kalo banyak bugnya | Bug nya banyak banget aku pakai wifi sinyalnya itu bagus waktu aku buka tiktok ada banyak banget bugnya kayak video gak bisa di like atau di pause terus kadang mau like video malah tibatiba masuk live orang dll percuma tiktok update terus sampe gb an kalo banyak bugnya | bug nya banyak banget aku pakai wifi sinyalnya itu bagus waktu aku buka tiktok ada banyak banget bugnya kayak video gak bisa di like atau di pause terus kadang mau like video malah tibatiba masuk live orang dll percuma tiktok update terus sampe gb an kalo banyak bugnya | [bug, nya, banyak, banget, aku, pakai, wifi, sinyalnya, itu, bagus, waktu, aku, buka, tiktok, ada, banyak, banget, bugnya, kayak, video, tidak, bisa, di, like, atau, di, pause, terus, kadang, mau, like, video, malah, tibatiba, masuk, live, orang, dan, lain-lain, percuma, tiktok, update, terus, sampe, gb, an, kalo, banyak, bugnya] | [tik, tok, sekarang, jelek, banget, parah, video, draf, lagu, nya, tiba, hilang, videonya, draf, video, mau, posting, nya, malah, nya, tiba, hilang, tik, tok, perbaiki, bug, nya, semakin, di, update, bukan, nya, semakin, bagus, sekarang, malah, semakin, jelek, tok, jelek, sekarang, tik, tok, jga, kualitas, video, di, posting, bures, padahal, awal, posting, video, jernihtpi, beberapa, jam, video, nya, langsung, bures, banget] | tik tok sekarang jelek banget parah masa video draf lagu nya tiba hilang video draf video mau posting nya malah nya tiba hilang tik tok baik bug nya makin di update bukan nya makin bagus sekarang malah makin jelek bgtt tik tok jelek sekarang tik tok juga kualitas video di posting padahal awal posting video jernihtpi beberapa jam video nya langsung banget |

**Figure 3.** Data Pre-Processing Results

Based on Figure 3, the data cleaning stage removes elements such as links, hashtags, emoticons, symbols, and mentions to facilitate data labeling, as shown in the second column. Next, case folding converts all letters to lowercase to reduce data variations, as illustrated in the third column. This process is followed by tokenizing, which separates sentences into single words to facilitate further analysis, as seen in the fourth column. The filtering stage removes unimportant words to focus on relevant information, as shown in the fifth column. Finally, stemming converts inflected words into their base form to simplify the data, as seen in the sixth column. Each of these stages aims to improve the quality of text data, ensuring that the data is ready for sentiment analysis using algorithms such as Bi-LSTM, which requires clean and structured text input.

### 3.1.3 Data Labelling

Data labeling is a crucial stage in sentiment analysis, especially for preparing datasets to be used in training deep learning models. One common approach used for automatic sentiment data labeling is the lexicon-based method. This approach utilizes a pre-defined dictionary of words with their sentiment polarity (positive, negative, or neutral) and their respective weight scores, allowing the system to automatically identify and label sentiment in text based on the words it contains. This process greatly helps in accelerating and standardizing the labeling of large datasets, which would be time and resource-consuming if done manually. The data labeling results are displayed in Figure 4.

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 2, September 2025 Page: 1234−1241
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8019

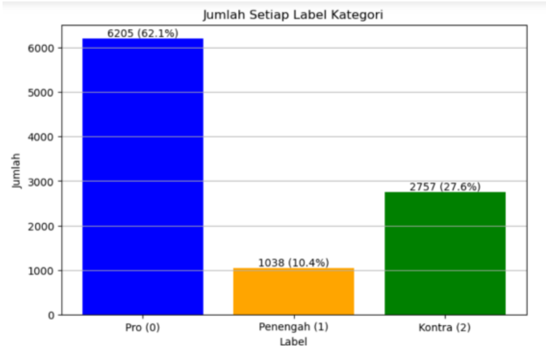


**Figure 4.** Label Distribution

Based on Figure 4, the sentiment distribution of user reviews for the TikTok application. From the total data analyzed, the Pro (0) category dominates with 6205 reviews, or 62.1%, indicating a significant positive sentiment from users. Meanwhile, the Contra (2) category occupies the second position with 2757 reviews, or 27.6%, reflecting a number of dissatisfactions or problems experienced by users. The Neutral (1) category has the smallest number, namely 1038 reviews or 10.4%, indicating that a small portion of reviews are neutral or do not lean towards positive or negative sentiment. This distribution provides a clear picture of the majority of positive sentiment received by the TikTok application, followed by fewer negative and neutral sentiments.

### 3.1.4 Word Embedding

The next stage is Word Embedding, a revolutionary technique that maps words to numerical vectors, where words with similar meanings have adjacent vectors. This allows computers to understand semantic relationships between words, such as synonyms and context, significantly improving model performance in sentiment analysis. The word similarity for "aplikasi" (application) and "tiktok" is shown in Figure 5.

```
# Menghitung kemiripan antara dua kata
word1 = 'aplikasi'
word2 = 'tiktok'
similarity = word2vec_model.wv.similarity(word1, word2)

# Menampilkan hasil kemiripan
print(f"Kemiripan antara '{word1}' dan '{word2}': {similarity}")
```

```
Kemiripan antara 'aplikasi' dan 'tiktok': 0.9590350389480591
```

**Figure 5.** Word Similarity between 'aplikasi' and 'tiktok'

Based on Figure 5, the calculation of similarity between the words 'aplikasi' (application) and 'tiktok' was performed. The result shows a similarity value of 0.959035038948059. This very high value indicates that in the context of the dataset used to train the Word2Vec model, the words 'aplikasi' and 'tiktok' have a very strong semantic similarity. This is very reasonable given that TikTok is a form of application, so these two words often appear in similar or even interchangeable contexts in user reviews. Word2Vec's ability to measure this similarity is an important basis for understanding the relationships between textual entities, which ultimately contributes to the effectiveness of sentiment analysis models in understanding the nuances of user language.

### 3.1.5 Bidirectional Long Short-Term Memory

The next stage is processing the data using the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm. Bi-LSTM consists of two independent LSTM layers that process the input sequence in two directions (forward and backward), then combine the representations from both directions. This capability allows Bi-LSTM to capture richer and more comprehensive context from the entire sentence, thereby significantly improving the accuracy and performance of the model in sentiment analysis tasks. The parameters used in this research are shown in Figure 6.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 100, 100)          1526800

 bidirectional (Bidirection  (None, 100, 128)          84480
 al)

 dropout (Dropout)           (None, 100, 128)          0

 lstm_1 (LSTM)               (None, 32)                20608

 dense (Dense)               (None, 3)                 99

=================================================================
Total params: 1631987 (6.23 MB)
Trainable params: 1631987 (6.23 MB)
Non-trainable params: 0 (0.00 Byte)
```

**Figure 6.** Bi-LSTM Parameters

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 2, September 2025 Page: 1234−1241
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8019

Based on Figure 6, the first layer is Embedding, which is responsible for transforming discrete word representations into dense vectors with an output shape of (None, 100, 100) and 1,526,800 parameters. This layer acts as a bridge between raw text and the feature space that can be learned by the neural network. Next, the Bidirectional (Bi-LSTM) layer with an output shape of (None, 100, 128) and 84,480 parameters processes word embeddings from two directions to capture global context. A Dropout layer is applied to prevent overfitting by randomly omitting a portion of neurons during training. Then, the LSTM layer with an output shape of (None, 32) and 20,608 parameters functions to condense the extracted features. Finally, the Dense layer (output shape (None, 3) and 99 parameters) acts as the output layer, generating probabilities for three sentiment categories (pro, contra, neutral), with a total of 1,631,987 trainable parameters, indicating the model's complexity which is capable of learning intricate sentiment patterns.

### 3.1.6 Model Evaluation

Model evaluation is a crucial stage for measuring the performance of the algorithm in classifying sentiment. Confusion Matrix is an effective tool for visualizing performance, showing the number of correct and incorrect predictions for each class. Metrics such as accuracy, precision, recall, and F1-score are derived from this matrix, providing a comprehensive overview of the model's ability to recognize pro, neutral, and contra sentiments in review data. The confusion matrix results in this study are shown in Figure 7.

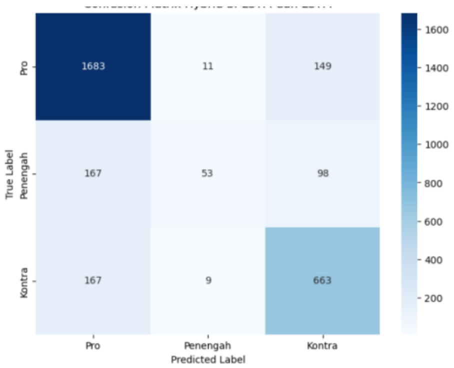


**Figure 7.** Confusion Matrix

Based on Figure 7, the model's ability to classify sentiment is visible. For the "Pro" label, the model correctly predicted 1683 reviews, but incorrectly classified 167 "Neutral" and 167 "Contra" reviews as "Pro". In the "Neutral" class, only 53 reviews were correctly classified, with dominant errors in predicting as "Pro" (11) and "Contra" (98). Meanwhile, for the "Contra" class, the model correctly identified 663 reviews, but incorrectly predicted 149 "Pro" reviews as "Contra" and 9 "Neutral" reviews as "Contra". This matrix highlights that the model performs best in identifying "Pro" sentiment, while classifying "Neutral" remains a major challenge. The classification report containing accuracy, precision, recall, and F1-score is shown in Figure 8.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Pro | 0.83 | 0.91 | 0.87 | 1843 |
| Penengah | 0.73 | 0.17 | 0.27 | 318 |
| Kontra | 0.73 | 0.79 | 0.76 | 839 |
| accuracy | | | 0.80 | 3000 |
| macro avg | 0.76 | 0.62 | 0.63 | 3000 |
| weighted avg | 0.79 | 0.80 | 0.78 | 3000 |

**Figure 8.** Classification Report

Based on Figure 8, this classification report presents comprehensive model performance evaluation metrics. It provides an overview of the model's overall performance, considering the amount of data (support) in each class. For a data split of 90% training data and 10% testing data, the precision of 0.79, recall of 0.80, and F1-score of 0.78 indicate that the model has good capability in sentiment classification on average, with weights given based on the proportion of each class. The overall accuracy of the model reaches 0.80 (80%), indicating that the Bi-LSTM model has effectively learned sentiment patterns on the TikTok review dataset.

### 3.2 Discussion

This analysis includes interpreting the model's performance and comparing it with previous research to contextualize this study's contribution, as well as affirming the superiority of the proposed approach. Comparison with previous research is shown in Table 1.

**Table 1.** Comparison with previous research

| Researcher | Algorithm | Accuracy |
|---|---|---|
| [26] | Naïve bayes classifier | 71.38% |

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 2, September 2025 Page: 1234−1241
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8019

| Researcher | Algorithm | Accuracy |
|---|---|---|
| [27] | Support Vector Machine | 70.75% |
| [28] | K-Nearest Neighbors | 73.6% |
| This study | Bi-LSTM | 80% |

Based on Table 1, the superiority of Bi-LSTM compared to machine learning methods such as SVM, NBC, and KNN lies in its ability to understand long-term dependencies and sequential context in text data. Unlike SVM and NBC which tend to treat text as a collection of words and ignore order, Bi-LSTM can process information from the beginning to the end of a sentence and vice versa, thus capturing more complex semantic relationships. In addition, as part of deep learning, Bi-LSTM can automatically extract relevant features from raw data without requiring extensive manual feature engineering, a significant advantage over KNN which heavily relies on high-quality feature representation. Bi-LSTM's adaptive capability to unstructured and large data makes it a more robust and accurate choice for sentiment analysis tasks involving the complexity of natural language in user reviews.

# 4. CONCLUSION

This research successfully implemented a sentiment analysis model based on Bidirectional Long Short-Term Memory (Bi-LSTM) to classify TikTok application user reviews into pro, neutral, and contra categories. With an architecture that includes Embedding, Bidirectional, Dropout, LSTM, and Dense layers, the model is capable of extracting rich contextual features from textual data. Evaluation results show the model achieved an overall accuracy of 80%, with an F1-score of 0.78, indicating good performance in general sentiment classification. Compared to previous research using conventional machine learning methods such as NBC, SVM, and KNN with accuracies below 80%, the Bi-LSTM algorithm proved superior in handling long-term dependencies and sequential context, making it a more robust and accurate choice for sentiment analysis on complex natural language data. Despite these strengths, a primary limitation encountered was the lower recall in identifying "Neutral" sentiment, suggesting that the model sometimes struggles to accurately classify inherently ambiguous neutral expressions. For future work, it is recommended to focus on developing more sophisticated techniques for neutral sentiment classification (e.g., alternative labeling strategies or advanced attention mechanisms), expanding and diversifying the dataset, exploring other state-of-the-art deep learning architectures like Transformers, incorporating multi-modal data (e.g., emojis), and delving into aspect-based sentiment analysis for more granular insights into user opinions on specific TikTok features.

# REFERENCES

[1] M. Ali, A. W. Budyastomo, and M. H. Harun, “The impact of social media for the development of da’wah in Indonesia,” *Religia*, vol. 24, no. 1, pp. 22–33, Apr. 2021, doi: 10.28918/religia.v24i1.2224.

[2] G. A. Szalkowski, I. M. Windekilde, and C. Johansen, “Towards sustainable short-form video: Modelling solutions for social and environmental challenges,” *F1000Res*, vol. 14, p. 265, Mar. 2025, doi: 10.12688/f1000research.161812.1.

[3] X. Wang, Y. Yu, Z. Zhu, and J. Zheng, “Visiting Intentions toward Theme Parks: Do Short Video Content and Tourists’ Perceived Playfulness on TikTok Matter?,” *Sustainability (Switzerland)*, vol. 14, no. 19, Oct. 2022, doi: 10.3390/su141912206.

[4] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, “Sentiment analysis and opinion mining on educational data: A survey,” *Natural Language Processing Journal*, vol. 2, p. 100003, Mar. 2023, doi: 10.1016/j.nlp.2022.100003.

[5] J. Hartmann, M. Heitmann, C. Siebert, and C. Schamp, “More than a Feeling: Accuracy and Application of Sentiment Analysis,” *International Journal of Research in Marketing*, vol. 40, no. 1, pp. 75–87, Mar. 2023, doi: 10.1016/j.ijresmar.2022.05.005.

[6] L. Bai, J. Guo, T. Xu, and M. Yang, “Emotional monitoring of learners based on EEG signal recognition,” *Procedia Computer Science*, vol. 176, pp. 364–368, 2020, doi: 10.1016/j.procs.2020.06.100.

[7] J. Oruh, S. Viriri, and A. Adegun, “Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition,” *IEEE Access*, vol. 10, pp. 30069–30079, 2022, doi: 10.1109/ACCESS.2022.3159339.

[8] G. Airlangga, “Comparative Analysis of NLP Techniques for Hate Speech Classification in Online Communications,” *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 1, pp. 674–683, Jan. 2024, doi: 10.33379/gtech.v8i1.3959.

[9] A. K. Maulaya and Junadhi, “Analisis sentimen menggunakan Support Vector Machine masyarakat Indonesia di Twitter terkait Bjorka,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 495–500, Dec. 2022, doi: 10.37859/coscitech.v3i3.4358.

[10] R. M. Al-Khatib, L. Heilat, W. Qudah, S. Alhatamleh, and A. Al-Khateeb, “A novel improved deep learning model based on Bi-LSTM algorithm for intrusion detection in WSN,” *Networks and Heterogeneous Media*, vol. 20, no. 2, pp. 532–565, 2025, doi: 10.3934/nhm.2025024.

[11] A. P. Joshi and B. V. Patel, “Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process,” *Oriental journal of computer science and technology*, vol. 13, no. 0203, pp. 78–81, Jan. 2021, doi: 10.13005/ojcst13.0203.03.

[12] E. Mulyani, F. P. B. Muhamad, M. Yani, and M. Alfarizi, “A comparative analysis: Spelling checker methods for syntactic ambiguity detection in software requirements statements using SMART rules between TextBlob and CyHunspell,” in *Proc. 17th Int. Conf. on Evaluation of Novel Approaches to Software Engineering (ENASE)*, Dec. 2023, pp. 120–126, doi: 10.5220/0011723600003575.

**Building of Informatics, Technology and Science (BITS)**
Volume 7, No 2, September 2025 Page: 1234−1241
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8019

[13] W. Bourequat and H. Mourad, "Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine," *International Journal of Advances in Data and Information Systems*, vol. 2, no. 1, pp. 36–44, Apr. 2021, doi: 10.25008/ijadis.v2i1.1216.

[14] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences (Switzerland)*, vol. 12, no. 17, Sep. 2022, doi: 10.3390/app12178765.

[15] A. Jabbar, S. Iqbal, M. I. Tamimy, A. Rehman, S. A. Bahaj, and T. Saba, "An Analytical Analysis of Text Stemming Methodologies in Information Retrieval and Natural Language Processing Systems," *IEEE Access*, vol. 11, pp. 133681–133702, 2023, doi: 10.1109/ACCESS.2023.3332710.

[16] R. Catelli, S. Pelosi, C. Comito, C. Pizzuti, and M. Esposito, "Lexicon-based sentiment analysis to detect opinions and attitude towards COVID-19 vaccines on Twitter in Italy," *Comput Biol Med*, vol. 158, May 2023, doi: 10.1016/j.compbiomed.2023.106876.

[17] V. Nurcahyawati, Z. Mustaffa, and M. Khalaf, "Exceeding Manual Labeling: VADER Lexicon as an Accurate Alternative to Automatic Sentiment Classification," *International Arab Journal of Information Technology*, vol. 22, no. 2, pp. 225–235, Mar. 2025, doi: 10.34028/iajit/22/2/2.

[18] Z. Kastrati, A. S. Imran, S. M. Daudpota, M. A. Memon, and M. Kastrati, "Soaring Energy Prices: Understanding Public Engagement on Twitter Using Sentiment Analysis and Topic Modeling with Transformers," *IEEE Access*, vol. 11, pp. 26541–26553, 2023, doi: 10.1109/ACCESS.2023.3257283.

[19] M. Chiny, M. Chihab, A. A. Lahcen, O. Bencharef, and Y. Chihab, "Effect of word embedding vector dimensionality on sentiment analysis through short and long texts," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, pp. 823–830, Jun. 2023, doi: 10.11591/ijai.v12.i2.pp823-830.

[20] A. Qostal, A. Moumen, and Y. Lakhrissi, "CVs Classification Using Neural Network Approaches Combined with BERT and Gensim: CVs of Moroccan Engineering Students," *Data (Basel)*, vol. 9, no. 6, Jun. 2024, doi: 10.3390/data9060074.

[21] A. L. Lezama-Sánchez, M. Tovar Vidal, and J. A. Reyes-Ortiz, "An Approach Based on Semantic Relationship Embeddings for Text Classification," *Mathematics*, vol. 10, no. 21, Nov. 2022, doi: 10.3390/math10214161.

[22] B. Omarov and Z. Zhumanov, "Bidirectional Long-Short-Term Memory with Attention Mechanism for Emotion Analysis in Textual Content," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, pp. 129–136, 2023, doi: 10.14569/IJACSA.2023.0140615.

[23] K. Zhu and N. H. Samsudin, "Attention-based Spatialized Word Embedding Bi-LSTM Model for Sentiment Analysis," *Pertanika J Sci Technol*, vol. 32, no. 1, pp. 79–98, Jan. 2024, doi: 10.47836/pjst.32.1.05.

[24] M. Anand, A. Velu, and P. Whig, "Prediction of Loan Behaviour with Machine Learning Models for Secure Banking," *Journal of Computer Science and Engineering (JCSE)*, vol. 3, no. 1, pp. 1–13, Feb. 2022, doi: 10.36596/jcse.v3i1.237.

[25] A. Bandi, P. V. S. R. Adapa, and Y. E. V. P. K. Kuchi, "The power of generative AI: A review of requirements, models, input–output formats, evaluation metrics, and challenges," *Future Internet*, vol. 15, no. 8, Aug. 2023, doi: 10.3390/fi15080260.

[26] D. Marlina, T. B. Kurniawan, M. Z. Zakaria, and S. F. Abdullah, "Sentiment analysis on natural skincare products using SVM," Journal of Data Science, vol. 4, no. 1, pp. 1–7, 2022. [Online]. Available: http://eprints.intimal.edu.my/1667/1/jods2022_12.pdf

[27] A. Witanti, "Analisis sentimen masyarakat terhadap vaksinasi COVID-19 pada media sosial Twitter menggunakan algoritma Support Vector Machine (SVM)," J. Sist. Inf. dan Inform. (Simika), vol. 5, no. 2, pp. 107–112, 2022, doi: 10.47080/simika.v5i2.1716. [Online]. Available: https://ejournal.akakom.ac.id/index.php/JISS/article/view/1330/351

[28] A. Salma and W. Silfianti, "Sentiment Analysis of User Review on COVID-19 Information Applications Using Naïve Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors," *International Research Journal of Advanced Engineering and Science*, vol. 6, no. 4, pp. 158–162, 2021.