JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. 9, No. 2, August 2026, pp. 164-174 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v9i2.11890 e-ISSN: 2656-1948

PUBLIC SENTIMENT ANALYSIS TOWARDS THE DISCOURSE OF MILITARY SERVICE FOR PROBLEMATIC CHILDREN USING INDOBERT ON SOCIAL MEDIA X

Windu Abdillah[1*], Estu Sinduningrum[2]

[1,2] Department of Informatics Engineering, Faculty of Industrial Technology and Informatics, Prof. Dr. Hamka University (UHAMKA), Jakarta 13460, Indonesia
Email: [1]2003015054@uhamka.ac.id , [2]estu.sinduningrum@uhamka.ac.id

(Received: 12 May 2026, Revised: 31 May 2026, Accepted: 11 June 2026)

Abstract

This study aims to analyze public sentiment toward the discourse of mandatory military service for troubled children using data collected from social media platform X. The research employs a deep learning approach based on the Transformer architecture using the IndoBERT model, which is specifically designed for Indonesian-language text processing. Data were collected through a web scraping process and resulted in 421 valid tweets after preprocessing, cleaning, and manual sentiment labeling stages. The proposed methodology integrates non-standard language preprocessing, including text normalization and cleaning of informal social media expressions, with the IndoBERT model to improve sentiment classification performance. The experimental results show that the model achieved an accuracy of 68.23%, with a precision of 67.39%, recall of 68.23%, and F1-score of 66.85%. Although the accuracy did not exceed 90%, this performance can be considered acceptable because the dataset was characterized by an imbalanced sentiment distribution, where positive sentiment represented only 9.74% of the data, making classification more challenging. The relatively limited dataset size, the presence of informal language, sarcasm, ambiguous expressions, and class imbalance also contributed to the model's performance limitations. Furthermore, the results reveal that negative sentiment dominated public opinion (55.34%), indicating substantial public concern and resistance toward the proposed policy. The findings demonstrate that IndoBERT is capable of capturing sentiment patterns from Indonesian social media discussions while providing valuable insights into public perceptions of controversial policy issues. Practically, these findings may assist policymakers and stakeholders in understanding public concerns, evaluating policy acceptance, and developing more adaptive and evidence-based approaches to addressing youth behavioral issues.

Keywords: *Sentiment Analysis, IndoBERT, Social Media X, Deep Learning, Natural Language Processing.*

**Corresponding Author: Windu Abdillah*

1. INTRODUCTION

In the current digital era, the advancement of internet technology is developing very rapidly and has influenced almost every field of human life [1]. Platforms such as X, Facebook, and Instagram make it easy for someone to share their opinions, experiences, and views quickly and easily with many people [2]. In Indonesia, social media such as X (formerly known as Twitter) serves as an opinion space for the community to express their views on various government policies, including policies that are controversial and often cause public debate [3]. Therefore, sentiment analysis becomes crucial to understanding public views on critical issues such as the discourse of military service for problematic children [4]. This approach allows for the systematic identification and categorization of public opinion, providing deep insight into the acceptance or rejection of a policy [5].

In this context, the use of language models such as IndoBERT becomes highly relevant for analyzing sentiment from Indonesian-language text data sourced from social media, given its ability to classify sentiment accurately [6]. The IndoBERT model has been proven effective in sentiment analysis on various topics on social media X, including public opinion related to government programs such as Free Nutritious Meals [7]. As an example, IndoBERT has been used to analyze public sentiment towards the Constitutional Court's Decision regarding the age limit for presidential and vice-presidential candidates, showing strong performance in identifying community

responses accurately [8]. In line with that, previous research has also utilized the IndoBERT model to analyze public sentiment on platform X regarding the Palestine-Israel conflict, affirming its capability in processing and classifying varied and complex text data [9]. The sentiment classification accuracy of implementing transformer-based models such as BERT and its derivatives is capable of capturing richer linguistic context compared to traditional methods [10].

Recent studies have demonstrated that transformer-based language models consistently outperform conventional machine learning approaches in sentiment classification tasks because they can capture contextual and semantic relationships more effectively [24], [25]. Furthermore, the availability of pre-trained language models has significantly improved the capability of sentiment analysis systems to process large-scale user-generated content on social media platforms. These models have become the state-of-the-art approach for opinion mining because they are capable of understanding contextual meanings, semantic dependencies, and complex language patterns that frequently occur in online discussions [24], [25].

Through this study, an innovative and coordinated approach can be found with the application of BERT in the analysis of public sentiment towards military service for problematic children using data from X (Twitter) regarding how the public views policies concerning military service for children facing problems [11].

Although numerous studies have investigated public sentiment toward political events, government programs, public services, and social issues using IndoBERT, research focusing specifically on public responses to the discourse of military-service-based intervention programs for at-risk youth remains very limited, particularly in the Indonesian context [26]. This issue has attracted substantial public attention and generated extensive debate on social media platforms due to differing opinions regarding its effectiveness, ethical implications, and potential impact on youth development. Consequently, there remains a significant research gap concerning how Indonesian society perceives and responds to this policy discourse in digital spaces [26].

X, previously known as Twitter, becomes a digital space for its users to convey their thoughts, feelings, or personal views in the form of short messages or tweets, which can then be read and responded to by other users. In practice, interactions on X often develop into discussions or even debates, especially when discussing social, political, economic issues, or hot topics being widely discussed. This reflects how users utilize X as a means to express themselves while participating in public conversation [12]. In the debate regarding military service for problematic children, discussions on X appear in various variations, including those that are serious, full of emotion, and also contain elements of humor. The existence of diverse expressions of sentiment demands a sophisticated analytical approach to identify and categorize the nuances of public opinion accurately [13]. This research will utilize IndoBERT to analyze public sentiment towards the discourse of military service for problematic children on social media X, with the aim of identifying sentiment distribution (positive, negative, neutral) and understanding the driving factors behind these opinions [14]. Public opinion can be understood as a way for individuals to express their attitudes or views in social spaces without feeling threatened by rejection from their surrounding environment [15].

Sentiment analysis is a method used to understand thoughts, emotions, and evaluations of someone through data such as written opinions, often utilizing NLP techniques to extract and classify sentiment polarity (positive, negative, neutral) from unstructured text [16]. In its development, the Python programming language is often chosen because it is supported by various open-source libraries, its syntax is relatively simple, and it allows the NLP processing to be carried out more efficiently without excessive complexity compared to other programming approaches [17]. Natural Language Processing (NLP) is one of the fields in artificial intelligence that focuses on how computers can understand and interact with human language. This approach has been proven effective in various text processing tasks, including grouping or classifying sentences with a high degree of accuracy [18]. Technological developments make sentiment analysis much more accurate than old methods. One major change comes from deep learning, which allows computers to recognize language patterns without having to be given rules one by one. Neural network models learn these patterns directly from data, so the results are usually more precise [19].

The Bidirectional Encoder Representations from Transformers (BERT) model is one of the important breakthroughs in the field of machine learning introduced by Devlin and team in 2019. The model is built on the Transformer architecture and is designed to learn deep language representations through a pre-training process [20]. IndoBERT is a development of the BERT model specifically adapted for the Indonesian language. This model was trained using a large collection of Indonesian-language text data known as Indo4B. The dataset includes hundreds of millions of words sourced from various platforms, such as Indonesian Wikipedia, news articles from national media such as Kompas, Tempo, and Liputan6, as well as the Indonesian web corpus [21]. The development based on Transformer is an artificial neural network model architecture that has revolutionized the field of NLP, allowing for more efficient and effective sequential data processing compared to previous architectures [22]. Models such as BERT, which is one of the Transformer implementations, are able to understand the context of

words in sentences bidirectionally, thus increasing performance in tasks such as sentiment analysis [23].

Despite the strong performance of IndoBERT in various sentiment analysis tasks, one of the major challenges in analyzing Indonesian social media data is the presence of informal language, slang expressions, abbreviations, misspellings, and non-standard writing styles commonly used by users on platform X [27], [28]. Previous studies have generally focused on sentiment classification performance while providing limited discussion regarding the impact of these linguistic characteristics on model performance. Therefore, effective preprocessing techniques are essential to normalize non-standard text and improve data quality before classification is conducted [27], [28].

The novelty of this study lies not only in examining a contemporary and socially relevant public policy issue, but also in integrating a preprocessing strategy specifically designed to handle non-standard Indonesian social media language prior to sentiment classification using IndoBERT. Furthermore, this research contributes to the literature by providing empirical evidence regarding public sentiment toward a controversial military-service-based intervention discourse that has received limited attention in previous sentiment analysis studies. The findings are expected to provide valuable insights for policymakers, researchers, and social observers in understanding public responses toward the proposed policy while simultaneously enriching the development of sentiment analysis research using Indonesian-language social media data [29], [30].

2. RESEARCH METHOD

This study uses a quantitative approach with a deep learning-based sentiment analysis method. This approach was chosen because it is capable of processing large amounts of data objectively and in a structured manner. The primary focus of the research is to understand the community's views on the discourse of military service for problematic children through data taken from the social media platform X. In this study, the model used is IndoBERT, which is based on the Transformer architecture. This model was chosen for its ability to understand language context bidirectionally, making it more accurate in identifying sentiment compared to conventional methods. The uniqueness of the methodology in this research lies in the integration of a specialized preprocessing process to handle non-standard language on social media with the deep learning model based on IndoBERT [21],. This approach is designed to enhance the model's ability to understand complex linguistic contexts, such as the use of slang, abbreviations, and informal language variations commonly found on the X platform.

To strengthen the preprocessing stage, this research implements a normalization mechanism specifically designed for Indonesian social media text. A customized slang-word dictionary was employed to convert non-standard words, abbreviations, and informal expressions into their standardized Indonesian forms. For example, terms such as “gk”, “ga”, “nggak”, and “tdk” were normalized into “tidak”, while abbreviations such as “yg”, “dr”, and “dgn” were converted into “yang”, “dari”, and “dengan”. Additionally, elongated words expressing emotions, such as “baguuuus” or “jeleeeek”, were reduced to their standard forms. This process aims to improve semantic consistency and reduce vocabulary sparsity before the data are processed by the IndoBERT model [21].

Furthermore, social media content often contains emojis, internet slang, code-mixing, repeated characters, and informal sentence structures that may affect sentiment classification performance. Therefore, preprocessing was not only conducted to remove noise but also to preserve the semantic meaning of public opinions. This preprocessing strategy constitutes one of the methodological contributions of this study because it improves the model’s ability to understand Indonesian social media language more effectively [21].

The stages of the method used are shown in Figure 2 below:

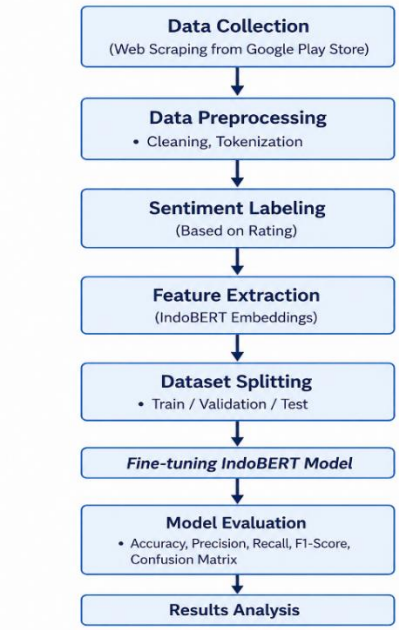

Figure 1. Research Flowchart

The data used in this study originated from tweet posts and user reply comments on social media X discussing the topic of military service for problematic children. Data collection was conducted legally by utilizing official APIs or scraping techniques that only access public data. Data that is duplicated, unrelated to the topic, or contains elements of spam will be eliminated. Furthermore, only Indonesian language data is retained, given that the focus of this research is sentiment analysis in the context of the Indonesian

language. To maintain research ethics, all data used does not include the personal identity of the users. Information such as account names, profile photos, and other sensitive data is removed or anonymized.

The final dataset consisted of 1,529 tweet posts and reply comments that met the inclusion criteria and were subsequently validated for sentiment analysis purposes. Although the dataset size is relatively smaller than those used in large-scale deep learning studies, the utilization of a pre-trained language model such as IndoBERT enables effective transfer learning through the fine-tuning process. Since IndoBERT has been pre-trained on large Indonesian-language corpora, it can achieve competitive performance even when applied to domain-specific datasets with limited sample sizes [21].

Data obtained from the X platform (Twitter) generally still contains various irrelevant elements, such as non-standard abbreviations, symbols, links, and special characters. This condition makes the text not yet ready for immediate analysis. Therefore, data preprocessing is required for cleaning and adjustment so that the information content in the text becomes neater and can be better understood by the model during the data processing stage. The stage of normalizing non-standard language is one of the important contributions of this research, as it plays a role in improving the quality of text representation before being processed by the IndoBERT model [21]. The preprocessing stages carried out in this research include: cleaning data, which is the process of cleaning text from unnecessary elements such as URLs, mentions (@username), hashtags (#topic), emojis, and excessive punctuation. This stage aims to reduce noise that could interfere with the model's process of understanding text content. Case folding is performed to convert all letters in the text to lowercase for consistency. This process is important to standardize word forms so that no differences in meaning occur solely due to variations in capitalization. Tokenization is conducted to break sentences into word units or tokens. By separating words in a sentence, the model can more easily recognize patterns and relationships between words. Word normalization is carried out, particularly to handle non-standard language commonly found on social media; for example, words like "gk", "nggak", or "ga" will be adjusted to standard forms like "tidak". This normalization helps improve data consistency and facilitates the model in understanding the true meaning. Stopword removal is applied to delete common words that do not significantly contribute to sentiment meaning, such as "dan", "yang", "di", and the like. Stemming is performed to convert words into their base form; for instance, "bermasalah", "permasalahan", and "masalahnya" will be reduced to the root word "masalah". Finally, the processed data will go through the labeling stage, which is the assignment of sentiment labels such as positive, negative, or neutral, serving as the primary reference for the deep learning model training process.

In this research, the labeling process was carried out automatically using the content of tweet posts and user reply comments as the basis for determining sentiment categories. Sentiment labeling is grouped into three main categories: Positive Sentiment, which reflects views of support or a tendency to agree with the implementation of military service for problematic children; Negative Sentiment, containing rejection, criticism, or concern regarding the policy; and Neutral Sentiment, which consists of opinions delivered informatively without showing specific bias or emotion.

Prior to model training, the distribution of sentiment classes was analyzed to identify potential class imbalance. The results indicated that the dataset was dominated by neutral sentiment, while positive sentiment represented the smallest proportion. To reduce the risk of model bias toward majority classes, a class weighting strategy was incorporated during the training process. This approach assigns larger weights to minority classes so that classification errors on underrepresented categories receive greater penalties during optimization, thereby improving the model's ability to recognize minority sentiment classes.

After passing the preprocessing and labeling stages, additional steps are required to transform the community opinion text into a more structured form. Feature extraction is performed using IndoBERT [21], a transformer-based model trained on Indonesian language datasets. This method is chosen for its ability to understand deep language meanings, including casual and non-standard language styles. The process begins with tokenization using the WordPiece tokenizer, breaking sentences into subwords to allow the model to understand uncommon words or spelling variations. Each token is then converted into embedding vectors, which are numerical values representing the word's meaning and context. The model then forms a feature representation of the entire sentence, summarized through the special [CLS] token at the beginning of the input, which serves as the basis for sentiment classification.

The dataset obtained is then divided into two parts: the Training Set (80%), used to train the IndoBERT model [21] to understand sentiment patterns, and the Validation Set (20%), used during the training process to evaluate model performance on unseen data, assist in hyperparameter tuning (learning rate, number of epochs), and prevent overfitting. The IndoBERT model is then readjusted through a fine-tuning process. This includes selecting the model architecture variant, generally configurations with 12 hidden layers and 768 units. A classification layer is added on top of the IndoBERT model as the final component, consisting of dropout to reduce overfitting, followed by a dense layer (fully connected) that adjusts the number of outputs to three sentiment categories using the softmax activation

function. The model is prepared for training by determining the optimizer (AdamW), loss function, and evaluation metrics.

Hyperparameters such as learning rate, batch size, and the number of epochs (set to 3) are also determined so the model can learn optimally. During training, performance is monitored via validation data at the end of each epoch to ensure the model generalizes well to new data. To provide a broader perspective on model effectiveness, the performance of IndoBERT was evaluated using multiple metrics, including Accuracy, Precision, Recall, and F1-score. These metrics were selected because accuracy alone may not adequately represent performance when dealing with imbalanced datasets. Precision measures the correctness of positive predictions, Recall evaluates the model's ability to identify all relevant instances, and F1-score provides a balanced assessment of both metrics.

Performance assessment is conducted using several evaluation metrics: Accuracy, Precision, Recall, and F1-score. Accuracy shows the overall correctness rate; Precision describes prediction exactness per sentiment class; Recall shows the ability to find all correct data in each class; and F1-score is the combination of precision and recall providing a balanced view of performance.

3. RESULTS AND DISCUSSION

3. 1 Dataset Description

The data collection process in this research was carried out using web scraping techniques from social media X (formerly Twitter) using tweet-harvest tools. Data retrieval was carried out based on keywords such as "barak anak nakal", "wajib militer anak nakal", and other relevant keywords. The timeframe for data retrieval started from April 1, 2025. Each tweet data obtained has several attributes, including conversation_id_str, full_text, created_at, as well as other metadata that support the analysis process. Initial crawling results produced a number of tweet data which were then processed further using the Python programming language. This process includes data cleaning and merging into the master dataset. At this stage, deletion of duplicate data based on the conversation_id_str attribute was also carried out, so that each tweet used in the research is unique data. Overall, the final dataset amounted to 468 tweets ready to be analyzed further. Table 1 shows some examples of initial data from the dataset:

Table 1. Initial Dataset Data

conversation_id_str	created_at	full_text
1919550721684353096	Tue May 06 00:32:10 +00002025	Satu anak menolak dibawa ke barak militer tapi pemerintah …
1919969262057484532	Wed May 07 04:15:18 +00002025	Menteri Hak Asasi Manusia (HAM) Natalius Pigai mendukung …
1919043071636869414	Mon May 05 03:43:08 +00002025	@ferrykoto Daripada anak anak nakal dimasukin ke barak militer...
1927944748087791771	Thu May 29 04:27:02 +00002025	Ini beneran udah ditahap anak 97 masuk barak…

3. 2 Preprocessing Data

The preprocessing stage is carried out to clean and normalize text data before being used in the sentiment analysis process. This process aims to improve data quality so that it can produce better model performance. The stages of preprocessing carried out in this research are as follows:

1. Case Folding: In this stage, the entire text is converted to lowercase to avoid word writing differences caused by the use of capital letters. Example: "Program Ini Bagus" → "program ini bagus".
2. URL Removal: URL or links contained in the tweet are deleted using regular expression, because they do not contribute to sentiment analysis. Link or URL contained in the tweet are deleted using regular expression. Example: "cek di https://example.com" → "cek di".
3. Mention Removal: User mentions (for example @username) are deleted because they do not have an influence on the meaning of sentiment in the text. Example: "@user ini bagus" → "ini bagus".
4. Hashtag Removal: The hashtag symbol (#) is deleted, but the word following it is still maintained because it still has meaning in the sentence context. Example: "#barak anak nakal" → "barak anak nakal".
5. Excessive Space Removal: Excessive spaces at the beginning, middle, or end of the sentence are deleted to produce a neater and more consistent text.
6. Data Cleaning: After preprocessing, a data cleaning process is carried out to delete data that is not valid. The stages carried out: Deleting data with null values; Deleting data that is empty or only contains spaces. This process is carried out with the aim of: Avoiding errors during the training process; Improving dataset quality.

The results of the preprocessing process are saved in a new column, namely clean_text, which is then used as input on the model training process. As a result, sentences become cleaner without excessive characters. Words are uniform because everything is converted to lowercase. The results of preprocessing are shown in Table 2.

Table 2. Preprocessing Results

full_text	Clean_text
@rainsanna_ @barengwarga Adiknya yang maju? Kesatria sekali. Anak nakal layak dimasukkan ke barak mereka agar belajar jadi kesatria.	adiknya yang maju? kesatria sekali. anak nakal layak dimasukkan ke barak mereka agar belajar jadi kesatria.
@geloraco Gimana masih yakin ngirim anak ke barak merupakan solusi mengatasi anak nakal?	gimana masih yakin ngirim anak ke barak merupakan solusi mengatasi anak nakal?
Anak nakal dikirim ke barak tentara. Polisi kriminal dikirim kemusala. https://t.co/sviBWshXSq	anak nakal dikirim ke barak tentara. polisi kriminal dikirim ke musala.
Anak nakal kirim ke Barak Orang tua maen HP mulu kirim ke Barak https://t.co/59QaEGOinr	anak nakal kirim ke barak orang tua maen hp mulu kirim ke barak
Barak Untuk Anak Nakal: Solusi Semu yang diulang Kembali https://t.co/6K3JPwewIT	barak untuk anak nakal: solusi semu yang diulang kembali

3. 3 Data Labeling

After the data is cleaned, the next step is to prepare the dataset for the manual sentiment labeling process in Table 3. A new column is added to the dataset, namely sentiment. This column will be filled manually with categories: Positive, Negative, and Neutral. The labeling process is carried out manually using the help of a spreadsheet (Google Sheets). Each tweet data is analyzed based on the text content, then given a label according to the sentiment category. The labeling criteria used are:

1. **Positive**: Tweets that contain support, appreciation, or positive opinions regarding the topic.
2. **Negative**: Tweets that contain criticism, rejection, or negative opinions.
3. **Neutral**: Tweets that are informative without containing a clear opinion.

Table 3. Manual Labeling Results

clean_text	sentiment
secaraa pribadi gua setuju banget ama programnya kdm anak nakal di didik di barak militer…	Positive
stuju stuju aja wajib militer utk anak2 nakal . kenapa mrk nakal? krn energi mereka …	Positive
anak nakal kembali berulah setelah pulang dari barak eri cahyadi pilih asramakan mereka	Negative
1. anak nakal dikirim ke barak militer solusi atau jalan pintas? ...	Negative
kak seto tinjau pendidikan karakter siswa di barak militer: tak langgar hak anak ketua lembaga perlindungan anak indonesia …	Neutral

3. 4 Sentiment Distribution

Sentiment distribution is carried out to determine the distribution of data based on the sentiment category that has been given at the labeling stage. This analysis aims to see the data balance before being used in the model training process. Based on the results of manual labeling, the following sentiment distribution is obtained shown in Table 4:

Table 4. Manual Labeling Sentiment Distribution

Sentiment	Total	Percentage
Negative	233	55.34
Neutral	147	34.92
Positive	41	9.74
Total	421	100

Based on the Table 4 above, it is seen that the data distribution is not fully balanced. The negative sentiment category has the largest amount of data compared to other categories. This shows that public opinion regarding the topic being researched tends to reject. Meanwhile, the category with the least amount of data shows that opinions in that category are relatively fewer found on social media. This imbalance in data distribution has the potential to affect model performance in carrying out classification, especially in classes with a smaller amount of data. The model tends to more easily recognize patterns in the majority class compared to the minority class. Even so, this distribution still reflects the real condition of public opinion on social media, so the research results remain relevant to the phenomena that occur. After the preprocessing and data cleaning processes are completed, the dataset is saved in the file dataset_remaja_barak_labeled.csv. This dataset is then used as the basis for the manual labeling process before entering the training stage.

3. 5 Model Training

In this research, the method used is Deep Learning based on Transformer, namely the BERT model with the name indobenchmark/indobert-base-p1. This model was chosen because: BERT was trained on a large-scale Indonesian corpus; Using Transformer architecture (self-attention); More superior compared

to traditional methods (Naive Bayes, SVM). The dataset that has gone through the labeling process is then divided into two parts, namely training data and validation data with a ratio of 80:20. From a total of 421 data, as many as 337 data were used as training data and 84 data as validation data. The data division process was carried out using the train_test_split method with parameters: test_size = 0.2; random_state = 42. The use of the random_state parameter aims to ensure that the data division process is consistent and can be reproduced in subsequent experiments. The purpose of this data division is: Training data is used to train the model in recognizing patterns in text data; Validation data is used to test model performance against data that has never been seen before.

In the tokenization process, text data is converted into numerical representations so that it can be processed by the IndoBERT model. This process is carried out using the built-in tokenizer from the IndoBERT model (indobenchmark/indobert-base-p1). The tokenization used is based on WordPiece, which is a method that breaks text into word units or subwords so that it is able to handle non-standard words or words that rarely appear. The stages in the tokenization process include: Text tokenization, namely breaking sentences into tokens or subwords; Addition of special tokens, such as [CLS] as the initial representation of the sentence and [SEP] as the marker for the end of the sentence; Padding, namely adding zero values to tokens so that the length of each input is uniform; Truncation, namely cutting the text if it exceeds the maximum token length limit. In this research, the maximum token length was determined to be 128 tokens to maintain computational efficiency without significantly reducing sentence context.

The model used in this research is BERT base with the name indobenchmark/bert-base-p1. This model is a pre-trained model which is then reused for sentiment classification tasks through the training process on the research dataset. In implementation, the model is configured using Auto Model For Sequence Classification with the number of classes as 3 classes, namely: Negative (0); Neutral (1); Positive (2). This model automatically adds a classification layer at the final part of the architecture to convert text representation into sentiment label predictions. The training process of the model is carried out using several parameters that have been determined through TrainingArguments. The parameters used in this research are as follows:

Table 5. TrainingArguments Parameters

Parameter	**Score**
Learning Rate	2e-5
Batch Size	8
Epoch	3
Evaluasi	Setiap epoch
Optimizer	AdamW (default transformer)

Learning rate of 2e-5 is used to control the speed of model weight updates during the training process. This value was chosen because it is commonly used in Transformer-based models so that the learning process remains stable. Batch size of 8 shows the amount of data processed in one training iteration. The use of a relatively small batch size is adjusted to the limitations of computational resources. The number of epochs is determined as 3, which means the entire training data will be studied by the model three times. The choice of this number of epochs aims to avoid overfitting while ensuring the model can study data patterns well enough. Evaluation is carried out at every end of an epoch to monitor model performance. In addition, the model is also saved at every epoch, so that the best model can be selected based on the evaluation results.

3. 6 Model Evaluation Results

After the training process is completed, model evaluation is carried out using validation data to measure model performance in classifying sentiment. This evaluation aims to find out the extent to which the model is able to carry out generalization to data that has never been seen before. Evaluation methods used include: Accuracy; Precision; Recall; F1-score. Based on the evaluation results, the following values were obtained:

Table 8. Model Evaluation Results

Matrix	Score
Accuracy	0.682353
Precision	0.673914
Recall	0.682353
F1-score	0.668519

In addition, model performance was also analyzed using a confusion matrix in Figure 2, to see the distribution of prediction results in each class.

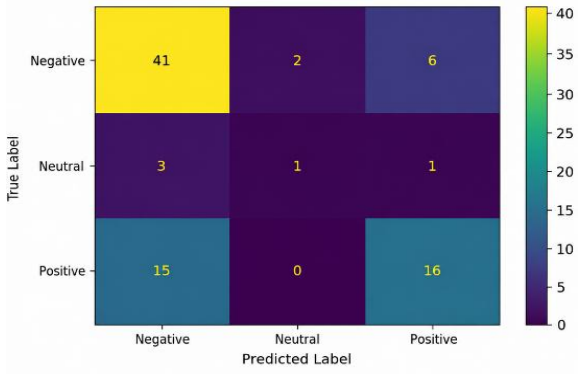

Figure 2. Confusion Matrix

Based on the confusion matrix displayed in Figure 4, several analyses can be carried out: In the Negative class, the model is able to classify quite well, where as many as 41 data were successfully predicted correctly. However, there are classification errors to the Neutral class (2 data) and Positive class (6 data). In the Neutral class, model performance is still low. From the total neutral data, only 1 data was successfully classified correctly, while most were incorrectly classified as Negative (3 data) and Positive (1 data). In the Positive class, the model was able to classify 16 data correctly,

but there was a significant error to the Negative class (15 data).

Based on the evaluation results, the BERT model that has been trained is able to classify sentiment with an accuracy level of 68.23%. This shows that most of the data in the validation set was successfully predicted correctly by the model. The precision value of 67.39% shows that the predictions produced by the model are quite precise in classifying data into each sentiment category. Meanwhile, the recall value of 68.23% shows the model's ability to identify relevant data in each class. The F1-score value of 66.85% is a result of the balance between precision and recall. This value indicates that model performance is in the fairly good category, although there is still room for improvement, especially in handling data with complex or ambiguous context.

The accuracy value of 68.23% can be considered acceptable given the characteristics of the dataset used in this study. As shown in Table 6, the sentiment distribution is imbalanced, where negative sentiment dominates the dataset (55.34%), while positive sentiment accounts for only 9.74% of the total data. Under such conditions, sentiment classification becomes more challenging because the model tends to learn patterns from the majority class more effectively than from minority classes. Therefore, evaluation should not rely solely on accuracy but also consider precision, recall, and F1-score, which provide a more comprehensive assessment of model performance across all sentiment categories.

The obtained precision score of 67.39%, recall score of 68.23%, and F1-score of 66.85% indicate that the model maintains relatively balanced performance in identifying sentiment classes. The similarity among these metric values suggests that the model does not exhibit extreme bias toward a particular class and is capable of generalizing sentiment patterns reasonably well despite the limited dataset size and class imbalance.

The confusion matrix further reveals that the model performs better in identifying negative sentiment compared to neutral and positive sentiments. This result is consistent with the class distribution presented in Table 6, where negative sentiment constitutes the majority class. The lower performance observed in the positive class can be attributed to the relatively small number of positive samples available during training, which limits the model's ability to learn representative sentiment patterns from that category. In addition, the model training process was also analyzed using training loss and validation loss graphs.

The graph shows that the training loss and validation loss values tend to decrease as the epoch increases. This indicates that the model successfully studied the pattern from the data well. There is no significant difference between training loss and validation loss, so it can be concluded that the model did not experience significant overfitting. Overall, the evaluation results show that the BERT model is able to understand sentiment patterns in text data quite well. However, model performance can still be improved through increasing the amount of data, improving labeling quality, as well as handling more complex informal language. After the training and evaluation process is completed, the trained model is then saved for reuse at the next stage.

Figure 3. Training Loss vs Validation Loss Graph

3. 7 Model Discussion

Based on the training and evaluation results that have been carried out, the BERT model shows quite good performance in carrying out sentiment classification against tweet data related to the topic "barak anak nakal". This is shown by the evaluation values obtained, specifically on the accuracy and F1-score metrics which reflect the model's ability to classify data as a whole. There are several factors that influence model performance in this research, including: Manual labeling quality; Dataset size; Class distribution, namely the balance of the amount of data in each sentiment category (positive, neutral, negative).

The performance achieved by IndoBERT in this study is closely related to its ability to capture contextual information through the Transformer architecture. Unlike conventional machine learning approaches that rely heavily on manually engineered features, IndoBERT learns contextual relationships between words bidirectionally, enabling it to better understand the meaning of social media texts that often contain informal language and diverse linguistic structures. This capability contributes to the model's ability to achieve satisfactory classification performance despite the complexity of public opinion data collected from platform X.

However, the dataset characteristics also significantly influence model performance. The relatively limited dataset size and the imbalance among sentiment classes create additional challenges during the training process. The dominance of negative sentiment data causes the model to learn negative sentiment patterns more effectively than positive sentiment patterns. Consequently, prediction performance for minority classes may be lower compared to the majority class. This finding is consistent with previous studies indicating that class

imbalance can affect the ability of deep learning models to recognize underrepresented categories.

Another important factor contributing to model performance is the preprocessing stage. The cleaning and normalization processes help reduce noise originating from URLs, mentions, hashtags, and inconsistent writing styles commonly found on social media. By transforming text into a more structured representation, preprocessing improves the quality of input data and facilitates the model in learning sentiment-related patterns more effectively. Therefore, the preprocessing strategy implemented in this study plays an important role in supporting the performance of the IndoBERT model.

The use of the BERT model in this research provides several advantages compared to traditional methods, including: Able to understand sentence context more deeply because it is based on Transformer architecture; More effective in handling informal language commonly used on social media; Has a better accuracy level in classification tasks of Indonesian language texts. However, the model used still has several limitations, namely: Difficulty in understanding sentences that contain sarcasm or irony; Less optimal in handling slang language that is very non-standard; Potential errors in interpreting sentences that are ambiguous or have double meanings.

Compared with previous studies employing IndoBERT for sentiment analysis on social media, the performance obtained in this study is relatively comparable considering the limited dataset size and the complexity of the issue being analyzed. The discourse surrounding military service programs for problematic children is characterized by diverse opinions, emotional expressions, and nuanced arguments, making sentiment classification more challenging than topics with more explicit sentiment expressions. Nevertheless, the results demonstrate that IndoBERT remains effective for capturing public sentiment patterns within controversial policy discussions.

Research results show that the approach used is able to identify public sentiment patterns towards policy issues in a quite representative manner. This shows that special preprocessing integration and the BERT model are effective in understanding the context of digital public opinion in Indonesia, especially on social issues that are complex and controversial.

From a practical perspective, the findings of this study provide valuable insights for policymakers and stakeholders. The dominance of negative sentiment indicates that a considerable portion of the public expresses concerns regarding the proposed policy. Such information can serve as an initial reference for evaluating public acceptance, identifying major concerns, and formulating communication strategies before implementing similar policies. Additionally, social media sentiment analysis can function as an efficient tool for monitoring public opinion in real time, enabling policymakers to better understand societal responses to emerging policy issues.

Despite its promising results, this study has several limitations. The relatively small dataset size and imbalanced sentiment distribution may restrict the model's ability to achieve higher classification performance. Future research is therefore recommended to utilize larger datasets, apply data balancing techniques, and compare IndoBERT with other machine learning and transformer-based models to obtain a more comprehensive evaluation of sentiment classification performance.

3. 8 Public Opinion Pattern Analysis

Analysis of public opinion patterns is carried out to understand the community's sentiment tendency towards the issue of "barak anak nakal" based on the classification results that have been carried out using the BERT model. Based on the results of sentiment classification, public opinion is divided into three main categories, namely positive, negative, and neutral. Each category reflects a different perspective on the policy or phenomenon discussed. Positive sentiment shows the existence of community support for the military barracks policy as a solution in handling delinquent children. Opinions in this category generally contain hope that a disciplinary approach can form a better character. Negative sentiment reflects rejection or criticism of the policy. Some opinions in this category highlight the aspects of children's rights, policy effectiveness, as well as the potential negative impacts of an approach that is militaristic. Neutral sentiment contains information or statements that do not show a clear opinion tendency.

In general, the public opinion pattern formed shows that this topic is still a debate in society. This is seen from the existence of significant sentiment variations in each category. The difference in opinion shows that the related policy still requires further study from various perspectives. In addition, the analysis results also indicate that social media becomes an important means in expressing public opinion. Therefore, sentiment analysis of social media data can provide an initial overview that is quite representative regarding public perception of a policy. The dominance of negative sentiment found in this research becomes an indication of a community resistance tendency towards a militaristic approach in handling problematic children. This finding provides an additional contribution in understanding public perception of social policies that are currently developing.

4. CONCLUSION

Based on the results of this study, it can be concluded that public sentiment toward the discourse of mandatory military service for troubled children on social media platform X tends to be negative, indicating the dominance of critical responses and rejection of the policy because it is considered to have

the potential to violate children's rights, create psychological impacts, and fail to address the root causes of social problems. On the other hand, some people expressed support for the military discipline approach as an effort to build children's character, while neutral sentiments generally consisted of informational statements without taking a particular stance. This study also proves that IndoBERT has fairly good performance in classifying informal Indonesian-language sentiment on social media, achieving an accuracy of 68.23%, precision of 67.39%, recall of 68.23%, and F1-score of 66.85%, although it still has limitations in understanding sarcasm, irony, and imbalanced data distributions. The main contribution of this study lies in the integration of a preprocessing strategy for handling non-standard Indonesian social media language with the IndoBERT model, as well as providing empirical evidence regarding public perceptions of military-service-based intervention programs for troubled children, an issue that has received limited attention in previous studies.

From a practical perspective, the findings of this study can serve as a reference for policymakers, educators, and child protection stakeholders in understanding public responses to the proposed policy. The dominance of negative sentiment indicates the need for careful consideration of public concerns before implementing similar programs. Nevertheless, this study has several limitations, including the relatively small dataset size, imbalanced sentiment distribution, and the use of data collected only from platform X. Therefore, future studies are recommended to utilize larger and more diverse datasets, incorporate data balancing techniques, compare IndoBERT with other classification methods, and develop more detailed sentiment and emotion classification frameworks to generate deeper insights and support more adaptive, evidence-based policymaking.

5. REFERENCES

[1] W. Amananti, "Analisis Sentimen Pengguna Media Sosial X Terhadap Isu Kesehatan Mental Mahasiswa Menggunakan Metode Naïve Bayes Dengan Pembobotan Weighted Inverse Document Frequency (WIDF)," vol. 4, no. 02, pp. 7823–7830, 2024.

[2] A. Sitanggang, Y. Umaidah, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, 2024.

[3] N. A. Mansyur, F. Wajidi, and M. R. Rasyid, "Klasifikasi Sentimen terhadap Program Barak Militer Anak Dedi Mulyadi Menggunakan Support Vector Machine," vol. 25, no. 1, pp. 153–169, 2026.

[4] S. A. Ihalauw, N. T. Lapatta, D. W. Nugraha, Wirdayanti, and C. A. Lamasitudju, "Analisis Sentimen Terhadap Kinerja Awal Pemerintahan Menggunakan," pp. 803–815, 2025.

[5] L. Rohmatun and A. Baita, "Machine Learning-Based Sentiment Analysis on Twitter (X): A Case Study of the 'Kabur Aja Dulu' Issue Using SVM," vol. 9, no. 4, pp. 1972–1983, 2025.

[6] R. Merdiansah, Siska, and A. A. Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," vol. 7, pp. 221–228, 2024.

[7] M. Haikal and A. Erfina, "JURNAL LOCUS: Penelitian & Pengabdian Analisis Sentimen Warganet terhadap Program Makan Bergizi Gratis (MBG) Menggunakan Model Indobert," vol. 5, no. 3, 2026.

[8] L. Septian, T. Aljauza, and C. Juliane, "Analisis Sentimen Putusan Mahkamah Konstitusi Terhadap Batas Usia Capres Dan Cawapres Menggunakan IndoBERT," vol. 12, no. 1, pp. 4428–4439, 2024.

[9] M. G. Al-kadzim, Rasim, and Herbert, "Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik Palestina-Israel Menggunakan Model IndoBERT," vol. 4, no. 2, pp. 1167–1174, 2024.

[10] A. Z. Muhabbab, Bunyamin, and Hasmawati, "Sentiment Analysis of Indonesia's Free Nutritious Meal Program on Platform X (Formerly Twitter) Using IndoBERT," vol. 9, no. 3, pp. 884–893, 2025.

[11] H. A. Jiddan and R. E. Putra, "Analisis Sentimen Sebagai Deteksi Sarkasme di Media Sosial X Berbahasa Indonesia Menggunakan Transformer dengan Model XLNet," vol. 07, pp. 588–593, 2026.

[12] K. Pramudya, Sujarwo, and D. Safitri, "Media Sosial X dalam Membentuk Identitas Sosial Digital Generasi Z," *Jurnal Ilmiah Penelitian Mahasiswa*, vol. 3, no. 2, pp. 178–183, 2025.

[13] P. Subarkah, A. N. Ikhsan, E. Anggraeni, and A. P. Sabaniyah, "Sentiment Perspective of Government's Free Nutritious Meal Policy on Social Media X using Indo-BERT and Bi-LTSM," vol. 7, no. 2, pp. 201–210, 2025.

[14] J. D. Yuswira and J. A. Ginting, "ANALISIS SENTIMEN DAN SOCIAL NETWORK ANALYSIS PADA ISU MAKANAN BERGIZI GRATIS DI X / TWITTER MENGGUNAKAN INDOBERT," vol. 10, no. 1, pp. 1259–1265, 2026.

[15] Z. Putri Ajie and A. Wardhana, "Analisis Sentimen Negative Publik pada Tagar #KaburAjaDulu di Media Sosial 'X'," *Jurnal Audiens*, vol. 6, no. 3, pp. 430–442, 2025.

[16] S. P. Mahardika, R. Mardhiyyah, and F. I. Sanjaya, "ANALISIS SENTIMEN PUBLIK TERHADAP BADAN INVESTASI DANANTARA PADA MEDIA SOSIAL X MENGGUNAKAN MODEL INDOBERT," vol. 7, no. 4, pp. 1711–1721, 2025.

[17] V. Chandradev, I. M. A. D. Suarjaya, and I. P. A. Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *Jurnal Buana Informatika*, vol. 14, no. 02, pp. 107–116, 2023.

[18] M. Fazri and A. Voutama, "ANALISIS SENTIMEN PUBLIK TERHADAP DANANTARA DI MEDIA SOSIAL X," vol. 9, no. 1, pp. 197–206, 2025.

[19] A. Abdullah and S. Yahya, "Kajian Pemanfaatan Deep Learning Dalam Pembelajaran Pada Lembaga Pelatihan," *Journal of Management, Administration, Education, and Religious Affairs*, vol. 7, no. 1, pp. 25–41, 2025.

[20] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019.

[21] N. Herlinawati, I. Ariyati, Samudi, and A. Y. Asih, "ANALISIS SENTIMEN ZOOM CLOUD MEETINGS DI PLAY STORE MENGGUNAKAN METODE INDOBERT Nuraeni," pp. 243–251, 2026.

[22] A. S. Rizkia, Wufron, and F. F. R. Roji, "Sentiment Analysis of Coretax: A Comparison of Manual, Transformers-Based, and Lexicon-Based Data Labeling on IndoBERT Performance," vol. 5, pp. 1037–1048, 2025.

[23] A. Jaya, "Analisis Sentimen pandangan public terhadap profesi PNS (Pegawai Negeri Sipil) dari Twiter menerapkan Indonesian," vol. 4, no. 1, pp. 38–44, 2023.

[24] T. B. Brown et al., "Language Models are Few-Shot Learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[25] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019, doi: 10.18653/v1/N19-1423.

[26] A. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *Proceedings of COLING 2020*, pp. 757–770, doi: 10.18653/v1/2020.coling-main.66.

[27] A. Ziems, A. Jha, and D. Kumar, "Racism and Bias in Social Media Language Processing: A Survey," *Computational Linguistics*, vol. 49, no. 2, pp. 1–40, 2023, doi: 10.1162/coli_a_00491.

[28] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool Publishers, 2012.

[29] F. Barbieri, J. Camacho-Collados, L. E. Anke, and M. Neves, "TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification," *Findings of EMNLP 2020*, pp. 1644–1650, doi: 10.18653/v1/2020.findings-emnlp.148.

[30] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," *Advances in Neural Information Processing Systems*, vol. 32, 2019.