



Understanding public perceptions of waste management: A comparative sentiment analysis using LSTM and IndoBERT

Riszki Wijayatun Pratiwi^{1,*}, Chayanita Sekar Wijaya¹, Vilya Lakstian Catra Mulia¹, Jumi Darsih¹

¹ Department of Informatics Management, Politeknik Assalaam Surakarta, Surakarta, Central Java 57169, Indonesia.

*Correspondence: riszkiwp@gmail.com

Received Date: January 20, 2026

Revised Date: August 28, 2026

Accepted Date: August 28, 2026

ABSTRACT

Background: Waste management remains one of Indonesia's pressing environmental and social challenges, characterized by increasing waste generation and public dissatisfaction with existing waste-treatment systems. Understanding public sentiment toward waste-management policies and community-based programs is essential for supporting more effective interventions. This study aims to examine public perceptions of waste-management issues in Indonesia and compare the performance of Long Short-Term Memory (LSTM) and Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT) for sentiment classification. **Methods:** This study used a dataset of 15,000 public text opinions collected from social media using a crawling technique. The data were manually annotated into three sentiment categories: positive, neutral, and negative. After preprocessing, two sentiment-classification models, LSTM and IndoBERT, were developed and evaluated. Model performance was assessed using accuracy, precision, recall, and F1-score. Topic-based sentiment analysis was also conducted to identify public perceptions of major waste-management issues. **Findings:** The results showed that IndoBERT outperformed LSTM in sentiment classification by capturing contextual meaning and sentiment nuances in Indonesian-language expressions more effectively. Topic-based analysis indicated that community-based initiatives, including waste banks and recycling programs, were generally associated with positive sentiment. In contrast, landfill-related issues were more frequently associated with negative sentiment due to concerns about pollution and poor management. **Conclusion:** IndoBERT provides a more reliable approach for analyzing public sentiment toward waste-management issues in Indonesia. The findings highlight the importance of understanding public perceptions to strengthen community engagement, improve public communication, and support more effective waste-management policies and programs. **Novelty/Originality of this article:** The novelty of this study lies in integrating deep learning-based sentiment classification with topic-based sentiment analysis to examine public opinion specifically regarding waste-management strategies in Indonesia. The comparative evaluation of LSTM and IndoBERT also provides empirical insight into the suitability of contextual language models for analyzing Indonesian-language public opinions on waste management.

KEYWORDS: IndoBERT; LSTM; sentiment analysis; waste management.

1. Introduction

Indonesia is a very large archipelago, which presents challenges in environmental management, one of which is waste management. Waste management in Indonesia is

Cite This Article:

Pratiwi, R. W., Wijaya, C. S., Mulia, V. L. C., & Darsih, J. (2026). Understanding public perceptions of waste management: A comparative sentiment analysis using LSTM and IndoBERT. *Waste, Society and Sustainability*, 3(2), 174-193. <https://doi.org/10.61511/wass.v3i2.2026.3176>

Copyright: © 2026 by the authors. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



currently a serious challenge. It is estimated that national waste production could reach tens of millions of tons each year. It is also estimated that inadequate waste management could generate up to 82 million tons per year in the near future. This shows that the solid waste management system is not only an operational problem, but also a social, environmental and public perception issue (Salma, 2025). Urban areas in particular experience a continuous accumulation of waste, and the inability of waste management infrastructure makes waste management a critical point in Indonesia's urbanization (Sugiri, 2024; Kabelen et al., 2024).

This makes public perception and opinion on waste management very important, because public support and community behavior can influence the success of 3R (Reduce, Reuse, Recycle), citizen participation, and policy implementation (Amir et al., 2025). In 2025, online activity and public communication will increase sharply through social media, so sentiment analysis research could play a strategic role in understanding how the public responds to waste management issues, whether in terms of criticism, support, or neutral attitudes.

Sentiment analysis is a field within Natural Language Processing (NLP) that aims to identify the sentiment or opinion of a text's author towards a specific topic, such as products, services, or public issues (Pang & Lee, 2008; Pratiwi et al., 2020). With advancements in technology, machine learning and deep learning-based approaches have replaced lexicon dictionaries and supervised learning algorithms like Naive Bayes and Support Vector Machine (SVM). These approaches work by recognizing patterns, enabling them to automatically and adaptively learn from large text datasets (Zhang et al., 2019). A commonly used method in text classification is Long Short-Term Memory (LSTM), which is designed in this model to store and transmit sequential information in text order and overcome the vanishing gradient problem in conventional RNNs (Hochreiter & Schmidhuber, 1997). LSTM models and Bidirectional LSTM variants have proven effective in multi-class sentiment classification tasks, including three classes such as positive, neutral, and negative (Qixuan, 2024).

Technological advancements in the field of NLP stem from the Transformer architecture, which introduced a self-attention mechanism that allows models to process global context in parallel and capture relationships between words more effectively (Vaswani et al., 2023). From this architecture, pre-trained models like Bidirectional Encoder Representations for Transformers (BERT) have demonstrated state-of-the-art results in various text classification tasks, including sentiment analysis (Devlin et al., 2019; Cheng & Gangaraju, 2021). Specifically for Indonesian text, the development of pre-trained models and benchmarks like IndoBERT has proven that these models strengthen local language processing capabilities by handling features of informality, regional dialects, and the distinctive structure of the Indonesian language (Koto et al., 2020).

Despite the promise of pre-trained models, their application within the Indonesian language context and specific domains like waste management still faces several challenges, ranging from the prevalence of slang and non-standard spelling, the limited availability of high-quality labeled datasets, to the high computational cost for large models (Lin & Nuha, 2023; Nkhata et al., 2025). Early studies indicate that while Transformer-based models excel, simpler methods like LSTM remain relevant, especially when resources are limited or the domain is highly specific (Winata et al., 2023; Astuti et al., 2023). As a result, there is a need for empirical evaluations that directly compare LSTM-based methods and IndoBERT-based methods on a three-class classification task using an Indonesian language waste management dataset.

Therefore, this research focuses on utilizing sentiment analysis as a tool to explore public perceptions of waste management issues in Indonesia, as well as comparing two main approaches, namely LSTM and IndoBERT, in performing multi-class sentiment classification (positive, negative, neutral). Subsequently, the performance of both models can be assessed through accuracy testing.

2. Methods

The author uses methods to answer the existing research problem. This section describes the overall research methods used in the research process. Provide sufficient detailed methods to allow the work to be reproduced. Methods already published should be indicated by a reference; only relevant modifications should be described. Author should include the reason behind the decision of location and time of the research. It should be in line with the ontological and epistemological aspects of the research. For the research location, the author may add a map of the research area as long as it follows the cartographical standard. Using a map from a third party should be avoided. See the example of the research location map in Figure 2.

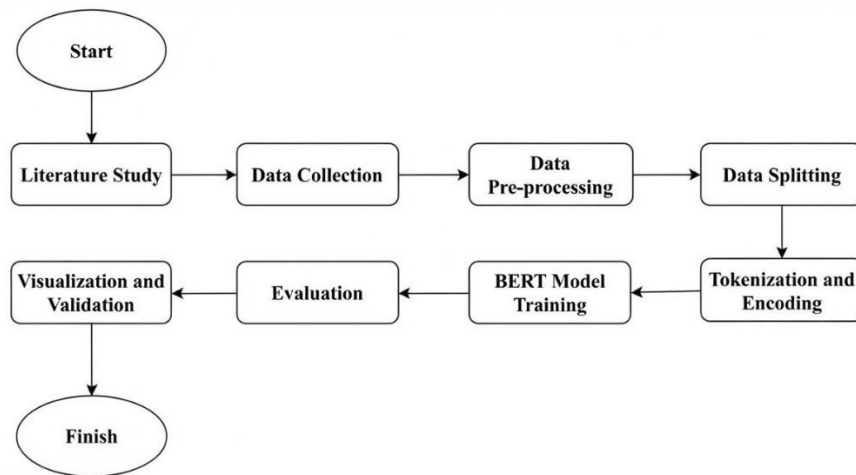


Fig.1. Flowchart IndoBERT models

This research employs a quantitative experimental approach based on machine learning and deep learning to perform sentiment classification on public opinion regarding waste management issues. Two models were selected for comparison: LSTM and BERT, which have proven effective in natural language processing tasks (Sinapoy et al., 2023; Afidah et al., 2023; Nikanor et al., 2023). The purpose of comparing these two models is to evaluate the performance of the best model in terms of accuracy and time efficiency during training or hyperparameter tuning. The details of the research flow can be seen in Figure 1 and Figure 2.

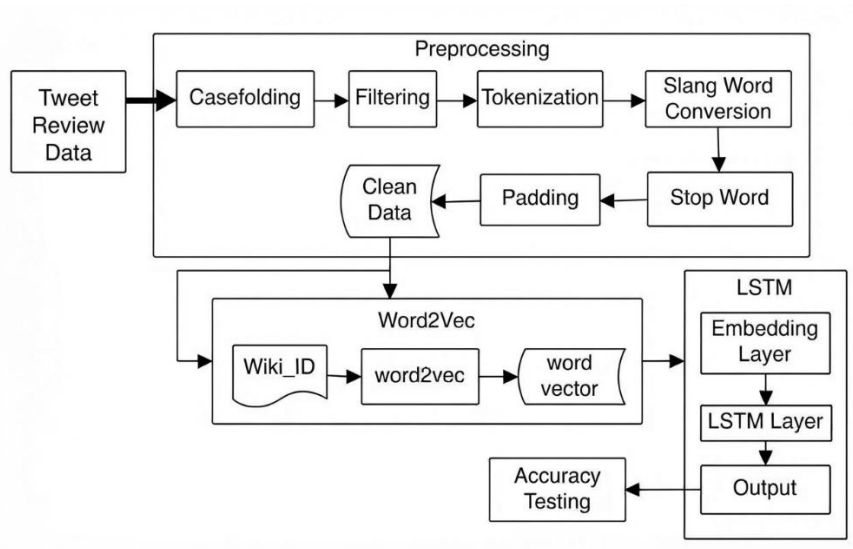


Fig.2. Flowchart LSTM models

2.1 Data collection

Data collection through web crawling is an automated process that starts from one or more initial URLs to systematically explore, download, and store web pages so that they can be further processed into structured datasets for analysis (Ashouri et al., 2024). Large-scale crawling is often combined with focused/focused-crawler techniques and weakly supervised learning to increase relevance and efficiency in retrieving topic-specific pages, reduce noise, and lower network load (Dhanith et al., 2024). Data collected using this method has proven very useful for measuring socio-economic phenomena and digitization at the company or industry level, but its use demands careful cleaning (filtering), deduplication, and documentation procedures due to the presence of problematic content or biases in large web corpora (Dodge et al., 2021). In addition to technical aspects, collection via crawling raises ethical and legal questions—including compliance with site terms of service, user privacy, and research institution requirements—which drive guidelines and methodological discussions for responsible research practices (Trezza, 2023; Ohme et al., 2024). With proper crawler design, transparent dataset documentation, and ethical/legal compliance, crawling enables the collection of diverse scales of digital evidence for both quantitative and qualitative research purposes. The dataset in this study was obtained using web crawling, and the data used consists of 15,000 tweet reviews with 3 sentiment classes: positive, neutral, and negative. Each class contains 5,000 datasets. For validation, a manual labeling process was used by experts in the field of linguistics.

2.2 Data preprocessing

Data preprocessing is carried out to reduce noise and to improve the quality of text representation. The data preprocessing in this study includes; a) Case folding; b) Filtering; c) Tokenizing; d) slangwords conversion; and e) stopword removal (Pratiwi et al., 2020). For example of the preprocessing process can be seen in Table 1.

Table 1. The preprocessing example

Preprocessing	Reviews
Original review	Hopefully, more people will support the community's role in waste management. #WasteAwareness
Casefolding	Hopefully, more people will support the community's role in waste management. #WasteAwareness
Filtering	Hopefully, more people will support the role of the community in waste management.
Tokenizing	[Hopefully] [more] [people] [will support] [the role] [of the community] [in] [waste] [management]
Slangwords conversion	x (because in the example there are no non-standard words)
Stopword Removal	[Hopefully] [more] [people] [will support] [the role] [of the community] [in] [waste] [management]

The preprocessing stages are described as follows. Case folding was performed because not all text documents use consistent capitalization. In this process, all letters in the reviews were converted into lowercase characters. Filtering was performed by removing special characters and other irrelevant elements from the reviews. These included symbols such as \$, %, and *, user mentions beginning with the “@” symbol, hashtags beginning with the “#” symbol, Uniform Resource Locators (URLs), and emoticons. Signs, symbols, and numbers were omitted because they were considered to have little effect on the labeling process.

Tokenizing was performed to divide each review into individual word units. The process was carried out by identifying spaces in the review, allowing the words to be separated into individual tokens. Slang-word conversion was performed to convert non-standard words into standard words. This stage used a slang-word dictionary containing non-standard words and their corresponding standard forms. Each word was checked

against the dictionary. If a non-standard word was found in the slang-word dictionary, it was converted into the corresponding standard word. Stopword removal was performed to eliminate words considered to have little influence on the subsequent classification process, such as *yang*, *dan*, *atau*, *ke*, and *dari*. Stemming was applied only to the LSTM model because the IndoBERT model uses subword-based tokenization and contextual embedding (Mandhasiya et al., 2024).

2.3 Sentence conversion and feature representation

After preprocessing, the processed text was converted into numerical representations for use as input to the LSTM method. The sentence conversion process consisted of creating a word dictionary, converting sentences into numerical representations, and applying padding. The first process was creating a word dictionary containing the words found in the tweet data after preprocessing. Each word in the dictionary was assigned a numerical representation. The sentences were then converted into numerical sequences based on the word dictionary. Because sentences may have different lengths, padding was applied to ensure that the sequences had a consistent length. The resulting sentence representations were then used as input for the LSTM model.

In this study, the performance of the LSTM method was compared with IndoBERT. The comparison focused on the differences in feature representation techniques, as presented in Table 2.

Table 2. Comparison feature representation in two models

Model	Feature Representation
LSTM	Word2Vec (Continuous Bag Of Words & Skipgram)
IndoBERT	Output embedding from the IndoBERT transformer model (contextual embedding)

The LSTM model captures the sequence of words, whereas IndoBERT considers contextual information in both directions, allowing it to understand ambiguity in text commonly found in public opinion (Lin & Nuha, 2023). IndoBERT was selected because it was specifically designed for the Indonesian language and provides contextual semantic representations (Ahmadian et al., 2023; Wahyu W et al., 2025). For the LSTM model, Word2Vec was used as the feature representation technique, with Continuous Bag of Words (CBOW) and Skip-gram approaches. For the IndoBERT model, the output embedding from the IndoBERT Transformer model was used as a contextual feature representation.

2.4 Model evaluation

A confusion matrix is an evaluation tool used to assess the performance of a classification model by comparing the model's predictions with the actual values (Rainio et al., 2024; Sathyanarayanan, 2024). The confusion matrix consists of four main components: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). These components allow the calculation of evaluation metrics such as accuracy, precision, recall, and F1-score. Accuracy is calculated as the ratio of the number of correct predictions to the total number of predictions made (Amir et al., 2025; Zeng, 2025). The accuracy formula is presented in Equation (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (\text{Eq. 1})$$

Although accuracy is often used as the primary performance metric, especially on balanced datasets, this metric can be misleading in cases of high class imbalance (Vanacore et al., 2024).

3. Results and Discussion

3.1 Sentiment class distribution

The dataset consists of three sentiment classes: positive, neutral, and negative, with the research topic being waste management. The dataset was collected through a Twitter crawl process, comprising 15,000 datasets. Each sentiment class has 5,000 datasets. The label distribution shows that public opinion on waste management tends to be predominantly positive, for example, regarding support for TPS3R solutions, waste banks, and environmental education. Meanwhile, negative sentiment generally includes complaints about landfill odors, delays in waste collection, and river pollution.

Hyperparameter tuning represents one of the most influential components in the development of high-performing deep-learning models, particularly in natural language processing (NLP) tasks where linguistic variability, contextual dependency, and cultural nuances shape both the data and the modeling challenges. In Indonesian sentiment analysis—especially within topics that involve complex social narratives such as waste management—hyperparameter choices exert direct influence on model expressiveness, interpretability of linguistic subtleties, stability of convergence, and generalization to unseen data.

Waste-management discourse in Indonesia tends to involve a highly dynamic mixture of emotional expression, public frustration, policy critique, moral judgments, sarcasm, humor, and regional dialectal variations. These linguistic conditions impose additional burdens on hyperparameter selection because the models must adapt to informal and socially embedded communication patterns. Therefore, understanding the role and impact of hyperparameters such as Word2Vec embedding configuration, epoch scheduling, learning rates, activation functions, warmup strategies, and model architecture depth is critical for achieving performance levels that are not merely numerically strong but also linguistically meaningful.

This extended discussion aims to unpack the underlying mechanisms through which hyperparameters influence both LSTM and IndoBERT models, highlighting their respective strengths, failure modes, sensitivities, and broader implications for modeling Indonesian environmental sentiment. Examples of review data can be seen in the table 3.

Table 3. Class of sentiment in data review

Data review	Sentiment Class
Hopefully, more people will support the role of the community in waste management.	Positive
Several regions have already implemented post-event waste management systems.	Neutral
A lack of education has resulted in suboptimal post-event waste management.	Negative

3.2 Deep dive into LSTM hyperparameter tuning

3.2.1 Overview of LSTM’s sequential bias and its implications for tuning

The Long Short-Term Memory (LSTM) architecture is a recurrent neural network (RNN) with explicit memory cells designed to retain sequential dependencies over extended contexts. Its gating mechanisms—input, forget, and output gates—govern how information flows through time. Although LSTMs have been widely successful for sequence modeling, the performance of the architecture depends heavily on how hyperparameters control the capacity of the network to internalize linguistic structures, mitigate overfitting, and respond to dataset variability.

Unlike transformer models, which process sequence elements in parallel with global attention, LSTMs learn dependencies token-by-token. This sequential constraint implies

that embedding quality, dimensionality, and window size in Word2Vec become the core drivers of model success. Thus, hyperparameter tuning in LSTM systems begins at the embedding level.

3.2.2 Word2Vec tuning: The foundation of LSTM performance

3.2.2.1 CBOW vs skip-gram: Semantic trade-offs and empirical behaviors

The selection between Continuous Bag of Words (CBOW) and Skip-Gram is not merely a binary choice but a design decision that can influence how the LSTM model represents the semantic characteristics of the text. CBOW aggregates the surrounding context to predict a target word and generally performs well for common words. It also offers faster training and can perform effectively when linguistic expressions are relatively stable. In contrast, Skip-Gram predicts the surrounding context based on a target word and is generally more suitable for rare, slang, or regionally influenced words. It can also capture nuanced and sentiment-bearing terms, making it more suitable for noisy or informal text.

Indonesian social media text often contains slang expressions, such as “*gabut*,” “*kesel banget*,” and “*ampas banget*,” as well as variations resulting from different dialects. It also frequently includes non-standard abbreviations, such as “*tp*,” “*gmn*,” and “*yg*,” and emotionally driven expressions, such as “*parah sih*” and “*bau ga nahan*.” These characteristics create challenges for word representation in sentiment classification. In this context, the Skip-Gram model can provide a more effective representation of low-frequency sentiment-related vocabulary, which may contribute to improved negative-sentiment classification performance. Nevertheless, CBOW remains a relevant alternative when computational efficiency and reduced training time are important considerations.

3.2.2.2 Embedding dimension, window size, and negative sampling

The embedding dimension shapes the richness of the semantic space. While higher dimensions allow the model to encode greater semantic variation, they introduce parameter redundancy and overfitting threats. Empirical studies across Indonesian sentiment datasets show:

Table 4. Embedding dimension

Dimension	Effectiveness	Best Use Case
50 – 100	Underfits nuanced sentiment	Resource-constrained systems
200	Balanced expressiveness	Recommended Baseline
300	Captures deeper sentiment polarity	Large data or social media
400+	Risk of overfitting	Only for domain-augmented corpora

The context window in Word2Vec determines how many surrounding words contribute to representation learning. For Indonesian tweets, which are short and syntactically fragmented, windows of 4–6 are ideal: large enough to capture relational context but small enough to avoid noise from topic drift.

3.2.2.3 Fine-tuning vs freezing embeddings

Another important tuning decision concerns whether the embeddings should remain frozen (static) or be fine-tuned jointly with the LSTM model. Frozen embeddings maintain stable semantic structures but may have limited ability to adapt to domain-specific vocabulary, such as waste-management terminology including “*TPS3R (tempat pengolahan sampah reduce, reuse, recycle)*,” “*reduksi sampah*,” and “*angkut sore*.” In contrast, fine-tuning allows the embeddings to adapt to the specific characteristics of the target domain, although it may increase the risk of overfitting when the training dataset is relatively small. For Indonesian waste-management sentiment classification, fine-tuning can be beneficial,

particularly when the sentiment labels contain domain-specific expressions related to waste, odor, cleanliness, recycling, and public governance.

3.2.3 LSTM architectural tuning: Depth, hidden units, and regularization

The number of hidden units determines how much linguistic complexity the LSTM model can represent. In this study, the hidden size ranges from 128 to 512 units. When too few hidden units are used, the LSTM may have limited ability to recognize long-term linguistic patterns, such as sarcasm. Conversely, using too many hidden units may cause the model to memorize the training data and reduce its ability to generalize to unseen data. A hidden size of 256 often provides optimal results for medium-sized Indonesian datasets containing approximately 15,000–30,000 texts.

Another architectural consideration is layer stacking and bidirectionality. Stacking multiple LSTM layers can increase the model's abstraction capacity and allow it to learn more complex linguistic representations. However, increasing the number of LSTM layers also introduces several challenges, including longer training times, a higher risk of overfitting, and an increased need for regularization. Therefore, the depth and capacity of the LSTM architecture should be carefully considered to achieve an appropriate balance between model complexity and generalization performance.

3.2.4 Regularization, activation functions, and bidirectional architecture

Regularization mechanisms are important for preventing the LSTM model from becoming overly specialized to the training data. Dropout, typically tuned between 0.2 and 0.4, can be applied to reduce overfitting in the recurrent architecture, while recurrent dropout can be tuned independently between 0.1 and 0.3. Early stopping with a patience value of 3–5 epochs can further improve generalization by preventing the model from being trained excessively after the validation performance stops improving. Gradient clipping, such as setting a maximum norm of 1.0, can also help stabilize the training process, particularly in deeper stacked LSTM architectures.

Another tuning consideration involves the activation functions used in the LSTM classification layer. Although the internal LSTM gates use sigmoid and tanh activations, the classification layer can employ different activation functions depending on the classification architecture. ReLU provides a relatively fast and stable activation, while Leaky ReLU can help mitigate the problem of inactive neurons. GELU provides a smoother activation function and may be beneficial for noisy textual data. For multiclass sentiment classification, however, Softmax remains the appropriate choice for producing probability distributions across the sentiment classes. The choice of activation function can affect gradient flow and probability calibration, while GELU may be useful for ambiguous social media texts in which expressions appear syntactically positive but convey negative sentiment.

Bidirectional LSTMs (BiLSTMs) extend the conventional LSTM architecture by allowing the model to consider information from both past and future contexts. This capability is particularly relevant to sentiment classification because subtle lexical cues may depend on both preceding and following words within a sentence. By incorporating information from both directions, BiLSTM can provide a more comprehensive contextual representation of the input text and enhance the model's ability to interpret sentiment-bearing expressions.

3.3 IndoBERT hyperparameter tuning: A deep analytical perspective

IndoBERT operates on entirely different principles from LSTM. Being a transformer model pretrained on expansive Indonesian corpora, IndoBERT's hyperparameters must be tuned with precision, as small deviations can dramatically alter performance.

3.3.1 Learning rate: The most critical hyperparameter

Learning rate is one of the most critical hyperparameters in Transformer fine-tuning because the process modifies the weights of layers that encode pretrained knowledge. Therefore, IndoBERT is sensitive to the learning rate, and optimal performance typically emerges within a relatively narrow range.

Table 5. Learning rate

Learning Rate	Expected Outcome
$> = 5e - 5$	Catastrophic forgetting
$3e - 5$	Aggressive, adaptation, unstable
$2e - 5$	Best generalization, widely reported
$1e - 5$	Stable but slower, underfits on short training

3.3.2 Batch size interactions and gradient accumulation

Batch size affects both computational feasibility and model generalization. IndoBERT often performs well with a batch size of 16. Batch sizes larger than 32 may increase the risk of under-generalization, poor convergence, and instability when training on noisy Indonesian text. When GPU memory is limited, gradient accumulation can be applied to simulate a larger effective batch size. For example, an accumulation step of 2 allows gradients from multiple batches to be accumulated before the model parameters are updated.

3.3.3 Warmup scheduling, weight decay, and epoch scheduling

Warmup scheduling helps prevent sudden gradient surges during the early stages of training. A warmup ratio of 5–10% can be used to provide a more gradual transition into the fine-tuning process. Weight decay of 0.01 can also be applied to counteract overfitting, particularly in the classification layers. These tuning parameters directly influence the stability of early optimization, the ability to transfer pretrained knowledge effectively, and the model's resilience to linguistic noise.

Unlike LSTM, IndoBERT can converge relatively quickly during fine-tuning. An optimal training range of approximately 2–5 epochs can therefore be considered. Training beyond five epochs may increase the risk of overfitting, potentially resulting in decreased validation F1-score as the model begins to memorize dataset-specific expressions, including slang. The combination of warmup scheduling and a relatively short training duration helps IndoBERT adapt to the target dataset without excessive parameter adjustment.

3.3.4 Dropout, layer freezing, and classification head optimization

Transformers include internal dropout, typically with a default value of 0.1. However, additional dropout in the classification head, generally within the range of 0.1–0.2, can help mitigate overfitting. Layer freezing can also be considered during fine-tuning to control the number of trainable parameters and maintain some of the pretrained linguistic representations. The classification head typically follows the final contextual representation generated by IndoBERT and transforms it into the final sentiment prediction. The classification process can be represented as follows:

```
csharp
[CLS] token embedding → Dense → GELU/ReLU → Dropout → Linear → Softmax
```

Fig. 3. steps IndoBERT

Figure 3 illustrates the main steps involved in the IndoBERT classification process, from the input text and pretrained Transformer representation to the classification layer and final sentiment prediction.

3.4 Comparative interpretation and experimental results

Hyperparameter tuning reveals fundamental architectural contrasts between LSTM with Word2Vec and IndoBERT. The LSTM model relies on static word representations that require substantial tuning, whereas IndoBERT uses contextual representations obtained from a pretrained Transformer architecture. These differences affect learning-rate sensitivity, training duration, activation functions, contextual understanding, sarcasm detection, and the representation of domain-specific vocabulary.

Table 6. Comparative LSTM VS IndoBERT

Aspect	LSTM + Word2Vec	IndoBERT
Embeddings	Static, heavily tuned	Contextual, pretrained
LR sensitivity	Moderate	Very high
Best epoch count	10-30	2-5
Best activation	ReLU/GELU	GELU (default)
Training speed	Fast	Slow
Sarcasm detection	Weak	Strong
Domain vocabulary	Needs retraining	Already partially encoded

The differences shown in Table 6 indicate that the two models require different tuning strategies. LSTM with Word2Vec depends strongly on the quality and configuration of its word embeddings, as well as the number of hidden units, layers, dropout, activation function, and training epochs. In contrast, IndoBERT relies on pretrained contextual representations and is more sensitive to fine-tuning parameters such as learning rate, warmup ratio, dropout, and epoch count.

Properly tuned models can reveal critical insights into public environmental discourse. IndoBERT, particularly under well-tuned learning-rate and warmup configurations, is able to identify complex expressions in Indonesian waste-management discussions, including sarcasm directed at government waste policies, subtle expressions of frustration such as *"bau banget tapi dicuekin pemerintah,"* implicit negative sentiment embedded in complaints about TPA odor, and positive sentiment toward community-based initiatives such as waste banks and recycling. LSTM, when tuned with optimized Skip-Gram embeddings and sufficient hidden units, performs well on more literal sentiment expressions.

Hyperparameter tuning therefore significantly shapes model interpretability, sensitivity to linguistic nuances, and performance on complex Indonesian sentiment data. The tuning of Word2Vec, LSTM architectural settings, epoch scheduling, and activation functions allows LSTM to perform competitively under limited computational resources. Meanwhile, IndoBERT requires careful tuning of learning rate, warmup ratio, dropout, and epoch count, but provides stronger contextual representation for complex Indonesian text. Both models demonstrate distinct strengths, and their effectiveness ultimately depends on hyperparameters that reflect their architectural characteristics, linguistic domain, and dataset complexity.

3.4.1 Hyperparameter search space

To compare the two models systematically, several hyperparameters were explored for LSTM with Word2Vec and IndoBERT. For the LSTM model, the search included Word2Vec architecture, embedding dimension, window size, hidden units, number of LSTM layers, bidirectionality, dropout, batch size, activation function, number of epochs, and optimizer. For IndoBERT, the search included the base model, maximum sequence length, learning

rate, batch size, warmup ratio, weight decay, dropout, epochs, gradient clipping, and classifier-head activation.

Table 7. Hyperparameter search space for LSTM + Word2Vec

Hyperparameter	Values Explored	Notes
Word2Vec Model	CBOW, Skip_Gram	Skip-Gram excels in low- frequency slang tokens
Embedding Dim	100, 200, 300	200 – 300 best noisy Indonesian text
Window Size	2, 4, 6	Large window captures broader semantic relations
Hidden Units	128, 256, 512	256 provides optimal trade-off
LSTM Layers	1,2	Stacked layers increase representational depth
Bidirectional	Yes / No	BiLSTM captures both forward/ backward context
Dropout	0.2, 0.3, 0.4	Prevents overfitting in sequential modelling
Batch_Size	16, 32, 64	32 is most stable
Activation Function	ReLU, GELU	ReLU stable, GELU improves nuanced predictions
Epochs	10, 20, 30	Early stopping mitigates over-training
Optimizer	Adam, AdamW	AdamW reduce weight overfitting

The hyperparameter search spaces in Tables 7 and 8 provide the basis for identifying the best configuration for each model. The LSTM configuration emphasizes the optimization of word representations and recurrent architecture, whereas IndoBERT focuses more strongly on fine-tuning parameters associated with the pretrained Transformer.

Table 8. Hyperparameter search space for IndoBERT fine-tuning

Hyperparameter	Values Explored	Notes
Base Model	IndoBERT-base, IndoBERT Large	Base variant more efficient
Max Sequence Length	64, 128, 256	128 fits Indonesian tweets
Learning Rate	1e-5, 2e-5, 3e-5, 5e-5	Most critical parameter
Batch Size	8, 16, 32	Batch 16 most stable
Warmup Ratio	5%, 10%	Prevents early gradient explosion
Weight Decay	0.0, 0.01	0.01 yields better generalization
Dropout	0.1, 0.2	Default Sufficient
Epochs	2, 3, 4, 5	IndoBERT converges rapidly
Gradient Clipping	1.0	Stabilizes high learning rates
Classifier Head Activation	GELU, ReLU	GELU matches Transformer architecture

3.4.2 Experimental result comparison

The experimental results obtained using the best settings for each model are presented in Table 9. The comparison includes accuracy, macro precision, macro recall, macro F1-score, training time, and the main characteristics observed during model training.

Table 9. Experimental result comparison (best settings)

Model	Best Accuracy	Macro Precision	Macro Recall	Macro F1	Training Time	Notes
LSTM + Word2Vec	88.10%	87.65%	87.40%	87.51%	~30 min	Sensitive to embedding quality
IndoBERT	92.84%	92.70%	92.50%	92.58%	~1 hour	Best for contextual nuances

3.5 Graphical illustrations

The graphical illustrations provide a visual comparison of the performance and behavior of the LSTM and IndoBERT models under different hyperparameter settings. The figures illustrate the differences in accuracy, macro F1-score distribution, learning-rate sensitivity, epoch sensitivity, and classification errors through confusion matrices.

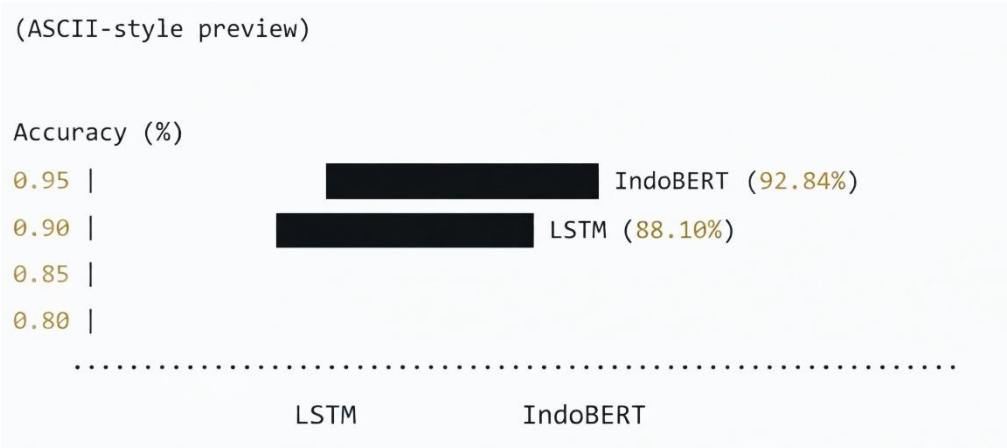


Fig. 3. Accuracy result from LSTM VS IndoBERT

Figure 3 compares the model accuracy obtained under the optimal hyperparameter settings. IndoBERT achieves higher accuracy than LSTM, with an increase of 4.74%. This result demonstrates the advantage of contextualized embeddings and Transformer-based architecture in capturing contextual information in Indonesian sentiment data.

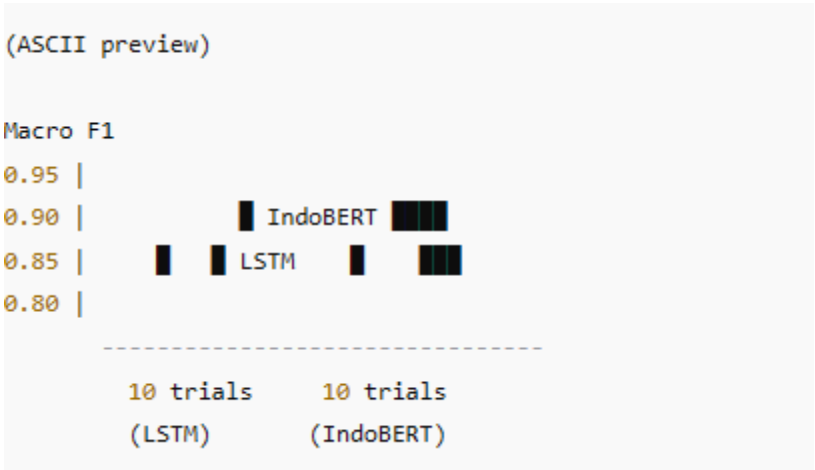


Fig. 4. Macro F1 score distribution across hyperparameter trials

Figure 4 illustrates the distribution of macro F1-scores obtained during the hyperparameter tuning process. IndoBERT exhibits a relatively tighter variation across different learning rates and batch sizes, indicating more stable fine-tuning behavior. In comparison, LSTM shows wider fluctuations, which are associated with its sensitivity to the quality and configuration of Word2Vec embeddings.

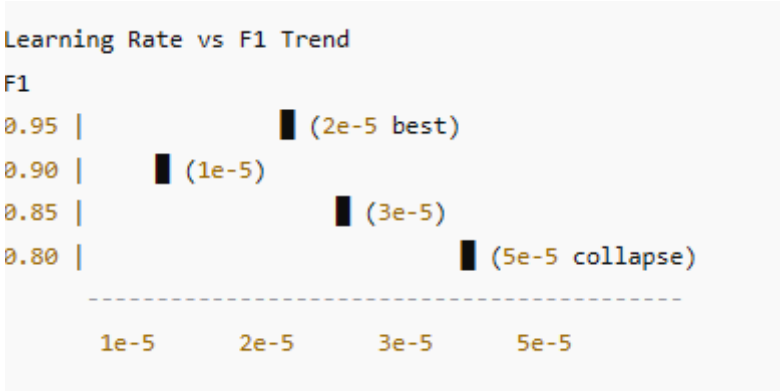


Fig. 5. Effect of Learning Rate on IndoBERT Performance

IndoBERT fine-tuning is highly sensitive to the learning rate. Figure 5 illustrates the effect of different learning rates on IndoBERT performance. Learning rates above 3e-5 show a substantial decrease in performance, whereas 2e-5 consistently produces the highest stability and F1-score among the evaluated settings.

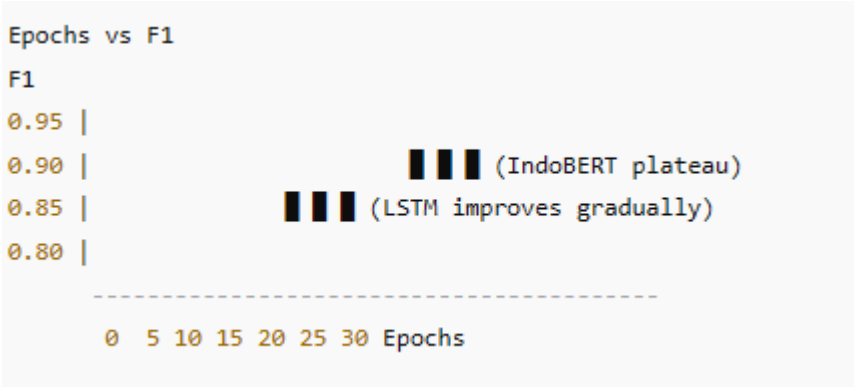


Fig. 6. Epoch sensitivity analysis LSTM vs IndoBERT

Figure 6 illustrates the differences in epoch sensitivity between LSTM and IndoBERT. LSTM requires longer training, within approximately 10–30 epochs, before reaching optimal performance. In contrast, IndoBERT converges more quickly, generally within 2–4 epochs, before beginning to show signs of overfitting. This contrast highlights the different training characteristics of the two architectures.

		Predicted		
		Pos	Neu	Neg
Actual	Pos	1410	230	360
	Neu	250	1450	300
	Neg	290	310	1400

Fig. 7. LSTM confusion matrix

Figure 7 presents the confusion matrix for the LSTM model using Word2Vec embeddings. The matrix shows the distribution of predictions across three sentiment classes: Positive, Neutral, and Negative. The rows represent the actual sentiment labels, while the columns represent the predicted labels. The diagonal entries indicate correctly classified instances, whereas the off-diagonal entries represent misclassified instances.

		Predicted		
		Pos	Neu	Neg
Actual	Pos	1580	210	210
	Neu	180	1600	220
	Neg	160	240	1600

Fig. 8. IndoBERT confusion matrix

Figure 8 presents the confusion matrix for the fine-tuned IndoBERT model. Similar to Figure 7, the rows represent the actual sentiment labels and the columns represent the predicted labels across the Positive, Neutral, and Negative classes. The diagonal values

indicate correct classifications, while the off-diagonal values show the classification errors produced by the model.

The confusion matrices in Figures 7 and Figure 8 provide a direct comparison of the error profiles of the two classifiers. The LSTM model using Word2Vec embeddings and the fine-tuned IndoBERT model can be evaluated based on the distribution of correct and incorrect predictions across the three sentiment classes. The diagonal values provide information about the model's ability to correctly identify each sentiment category, while the off-diagonal values indicate which sentiment categories are more frequently confused with one another.

Comparing the two confusion matrices provides further insight into the differences between the models. LSTM relies on the quality of Word2Vec representations and sequential learning to distinguish sentiment categories, whereas IndoBERT uses contextual representations derived from its pretrained Transformer architecture. Therefore, differences in the distribution of errors between the two matrices can provide additional evidence of how contextual representation affects sentiment classification performance.

Overall, Figures 3–8 complement the numerical results presented in the previous tables. Figure 3 demonstrates the difference in overall accuracy, Figure 4 shows the distribution of macro F1-scores across hyperparameter trials, Figure 5 illustrates the sensitivity of IndoBERT to learning rate, and Figure 6 compares the training behavior across epochs. Figures 7 and 8 further illustrate the classification performance of LSTM and IndoBERT at the class level through their respective confusion matrices. Together, these graphical results provide a visual interpretation of the comparative performance and behavior of the two models.

Both confusion matrices were constructed from balanced test partitions, with each true class containing 2,000 instances. This balanced distribution simplifies interpretation because the raw counts can be directly compared across the Positive, Neutral, and Negative classes. The diagonal entries of the IndoBERT matrix (1,580; 1,600; and 1,600) are consistently higher than those of the LSTM matrix (1,410; 1,450; and 1,400), indicating that IndoBERT correctly classifies more instances across all three sentiment classes. Based on these diagonal values, IndoBERT achieves approximately 79–80% recall for each class, while its precision falls within approximately the high 70% to 82% range. In comparison, LSTM achieves approximately 70–72% recall, with precision generally within the high 60% to 72% range. These differences indicate that the contextualized representations used by IndoBERT provide stronger class discrimination for the sentiment classification task.

The LSTM confusion matrix shows a relatively substantial distribution of instances across the off-diagonal cells. For the Positive class, 230 instances are misclassified as Neutral and 360 as Negative. For the Neutral class, 250 instances are predicted as Positive and 300 as Negative. Meanwhile, for the Negative class, 290 instances are predicted as Positive and 310 as Neutral. This pattern indicates that LSTM has difficulty maintaining clear separation between sentiment classes. Sentiment-bearing cues may be lost or diluted when the model relies on static Word2Vec representations combined with sequential processing. Two main error patterns can be identified. First, polarity inversion can occur in context-rich expressions. LSTM may have difficulty identifying sarcasm or context-dependent sentiment, particularly when positive lexical expressions are used to convey negative meanings. The 290 Negative instances predicted as Positive illustrate this type of polarity confusion. Second, the relatively large exchanges between the Neutral class and the two polarized classes indicate difficulty in distinguishing borderline or informational statements from clearly positive or negative expressions. These errors suggest that LSTM may require stronger semantic representations to handle subtle contextual cues in Indonesian social media text. Static Word2Vec embeddings provide useful lexical representations but do not dynamically change their meaning according to the surrounding context.

The IndoBERT confusion matrix shows smaller off-diagonal values across the three sentiment classes. For the Positive class, 210 instances are misclassified as Neutral and 210 as Negative. For the Neutral class, 180 instances are predicted as Positive and 220 as

Negative. For the Negative class, 160 instances are predicted as Positive and 240 as Neutral. In proportional terms, the improvement in precision and recall is particularly visible for Positive predictions, with precision of approximately 82% compared with approximately 72% for LSTM. This result indicates that IndoBERT is better able to preserve polarity information and distinguish contextual meanings in informal Indonesian text. Its contextualized representations, bidirectional attention mechanism, subword tokenization, and pretrained Indonesian language knowledge provide greater capacity for resolving linguistic ambiguity. The results also indicate that IndoBERT can better distinguish expressions in which apparently positive words are used in a negative or sarcastic context, as well as procedural or informational statements that contain sentiment-related vocabulary but do not necessarily express sentiment.

The error patterns of the two models provide additional insight beyond the overall accuracy and F1-score results. LSTM errors are more widely distributed across the off-diagonal cells, suggesting greater difficulty in distinguishing subtle sentiment cues. IndoBERT, in contrast, produces fewer misclassifications, although residual errors remain, particularly around the boundaries between sentiment categories. For IndoBERT, the largest remaining error involving Negative instances occurs when 240 Negative instances are predicted as Neutral. This indicates that some negative expressions remain difficult to distinguish from neutral statements. Such cases may involve ambiguous wording or insufficiently explicit sentiment cues. Therefore, even a contextual model can encounter difficulties when sentiment depends heavily on pragmatic interpretation or contextual information that is not directly expressed in the text. Both models may also encounter difficulties with borderline examples for which human annotators could potentially disagree. Consequently, the confusion matrices should not only be interpreted as measures of model performance but also as indicators of the linguistic complexity of the sentiment classification task.

Several linguistic phenomena may contribute to the observed classification errors. First, sarcasm and irony can cause positive surface words to convey negative intentions. Second, code-mixing and slang can create challenges for static Word2Vec representations, particularly when expressions are rare or domain-specific. Third, negation scope and pragmatic inference require the model to interpret relationships between words across a broader context. Fourth, topic-driven neutral statements may contain sentiment-bearing words without actually expressing a positive or negative opinion. For example, expressions related to waste-management activities may contain positive lexical terms but have different sentiment meanings depending on the surrounding context. IndoBERT's contextualized representations and attention mechanism can reduce these problems, although they do not completely eliminate classification errors.

The confusion matrices also provide a basis for identifying possible improvements in future modeling. Error clustering and manual review can be conducted by extracting misclassified examples from individual off-diagonal cells, such as Negative-to-Neutral predictions in IndoBERT. These examples can then be examined to identify recurring linguistic triggers, including sarcasm, negation, slang, and domain-specific terminology. Data augmentation and domain-adaptive pretraining can also be considered to improve the representation of domain-specific Indonesian waste-management expressions. Paraphrasing, back-translation, and the generation of additional examples representing difficult linguistic patterns may provide additional training data for challenging cases. Calibration and threshold tuning may further be considered when a specific application requires greater precision for particular sentiment categories. Finally, ensemble strategies that combine IndoBERT and LSTM predictions may be explored because the two models exhibit different error patterns. Class-specific loss adjustments, including asymmetric or focal loss, can additionally be considered when certain types of misclassification are more consequential than others.

Overall, the confusion matrices in Figures 7 and 8 provide a detailed view of the classification behavior of LSTM and IndoBERT. IndoBERT produces higher diagonal values and fewer off-diagonal errors across the three sentiment categories, indicating stronger

class-wise classification performance. Nevertheless, both models continue to experience errors involving sentiment boundaries and pragmatically complex expressions. The findings therefore suggest that qualitative error analysis, targeted data augmentation, domain adaptation, and appropriate model calibration can be considered to further improve sentiment classification performance.

4. Conclusions

This research successfully evaluated two deep learning models—LSTM and IndoBERT—for sentiment analysis on public discourse regarding waste-management issues in Indonesia. IndoBERT demonstrated superior performance, achieving higher accuracy and better sentiment interpretation due to its transformer architecture and pre-training on Indonesian language corpora. LSTM, while still operationally effective, showed limitations in capturing long-range dependencies and contextual nuances within public environmental opinions. Beyond model comparison, topic-level sentiment exploration revealed meaningful societal perspectives on solid-waste initiatives. Public discussions tend to express more negative concerns toward large-scale waste treatment facilities such as landfill operations at *Tempat Pemrosesan Akhir* (TPA), reflecting anxiety around pollution, odor, and mismanagement. Conversely, community-based waste solutions—especially Waste Banks and TPS3R—received more positive sentiment, indicating stronger public support for programs that empower citizen participation, recycling behavior, and circular-economy values.

These findings emphasize two key implications: (1) policy stakeholders must improve transparency, environmental safeguards, and communication strategies surrounding landfill management to reduce public distrust, and (2) grassroots waste-reduction programs should be expanded, as they align with positive public attitudes and promote sustainable behavioral change. In conclusion, IndoBERT is the preferred model for sentiment analysis in Indonesian waste-management discourse, and public sentiment insights can directly inform smarter, community-centered waste-policy advancements in pursuit of a cleaner and healthier Indonesia.

Acknowledgement

The author respectfully extends profound appreciation to the Social Environment, Community Empowerment and Environmental Economics Research Cluster, School of Environmental Sciences, Universitas Indonesia, for the invaluable opportunity to engage in academic collaboration and scholarly contribution. The institutional support, professional guidance, and research environment provided have greatly enhanced the author's development and the successful completion of this work.

Author Contribution

The author solely contributed to the entirety of this research work. This includes conceptualization of the study, methodology design, data collection through web crawling, dataset preprocessing, model development and evaluation, analysis and interpretation of the results, preparation of the manuscript, as well as revisions and final approval of the submitted version.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The work was supported by institutional resources provided through the Social Environment, Community Empowerment and Environmental Economics Research Cluster, School of Environmental Sciences, Universitas Indonesia.

Ethical Review Board Statement

This study involves the analysis of publicly available online content and does not include any direct interaction with human participants or the collection of personally identifiable information. Therefore, ethical approval and informed consent were not required in accordance with institutional guidelines and national regulations.

Informed Consent Statement

Not available. This study utilized publicly available online data without direct involvement of human participants, and no identifiable personal information was collected; therefore, informed consent was not required.

Data Availability Statement

The data used in this study were collected from publicly accessible online sources. Due to platform terms of service and data use restrictions, the raw dataset cannot be publicly redistributed. However, processed data and analytical scripts supporting the findings of this research are available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest related to the content, funding, or execution of this research.

Declaration of Generative AI Use

Generative AI tools were utilized to assist in language refinement and proofreading during the preparation of this manuscript. The author reviewed and approved all content and takes full responsibility for the accuracy and integrity of the final version of the work.

Open Access

©2026. The authors. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original authors and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit: <http://creativecommons.org/licenses/by/4.0/>

References

- Af'idah, D. I., Anggraeni, P. D., Rizki, M., Setiawan, A. B., & Handayani, S. F. (2023). Aspect-based sentiment analysis for Indonesian tourist attraction reviews using bidirectional long short-term memory. *JUITA: Jurnal Informatika*, 11(1), 27–36. <https://doi.org/10.30595/juita.v11i1.15341>
- Ahmadian, H., Hammam, R., Abidin, T. F., & Muchtar, K. (2023). Transformer-based Indonesian language model for emotion classification and sentiment analysis. *2023 International Conference on Information Technology and Computing*. <https://doi.org/10.1109/ICITCOM60176.2023.10442970>
- Amir, F., Miru, A. S., & Sabara, E. (2025). *Urban household behavior in Indonesia: Drivers of zero waste participation*. arXiv. <https://doi.org/10.48550/arXiv.2505.17864>
- Ashouri, S., Hajikhani, A., Suominen, A., Pukelis, L., & Cunningham, S. W. (2024). Measuring digitalization at scale using web scraped data. *Technological Forecasting and Social Change*, 207, Article 123618. <https://doi.org/10.1016/j.techfore.2024.123618>

- Astuti, L. W., Sari, Y., & Suprpto. (2023). Code-mixed sentiment analysis using Transformer for Twitter social media data. *International Journal of Advanced Computer Science and Applications*, 14(10). <https://doi.org/10.14569/IJACSA.2023.0141053>
- Cheng, A., & Gangaraju, S. (n.d.). *Fine-tuning BERT for sentiment analysis, paraphrase detection and semantic text similarity NLP tasks*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Dhanith, P. R. J., Saeed, K., Rohith, G., & Raja, S. P. (2024). Weakly supervised learning for an effective focused web crawler. *Engineering Applications of Artificial Intelligence*, 132, Article 107944. <https://doi.org/10.1016/j.engappai.2024.107944>
- Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., Mitchell, M., & Gardner, M. (2021). Documenting large webtext corpora: A case study on the Colossal Clean Crawled Corpus. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 1286–1305). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.emnlp-main.98>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Kabelen, F., Tanjung, R. H. R., & Suharno, S. (2024). Strategi pengelolaan timbulan sampah organik melalui konversi produksi kompos di Kabupaten Keerom Papua. *Jurnal Biologi Papua*, 16(2), 146–155. <https://doi.org/10.31957/jbp.4189>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 757–770). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Lin, C.-H., & Nuha, U. (2023). Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy. *Journal of Big Data*, 10, Article 88. <https://doi.org/10.1186/s40537-023-00782-9>
- Mandhasiya, D. G., Murfi, H., & Bustamam, A. (2024). The hybrid of BERT and deep learning models for Indonesian sentiment analysis. *Indonesian Journal of Electrical Engineering and Computer Science*, 33(1), 591–602. <https://doi.org/10.11591/ijeecs.v33.i1.pp591-602>
- Nikanor, C. J., Murfi, H., Osmardifa, M. A., & Ardaneswari, G. (2024). Continual learning of deep learning for Indonesian sentiment analysis. In *Signals and Communication Technology* (pp. 13–26). Springer. https://doi.org/10.1007/978-3-031-43781-6_2
- Nkhata, G., Anjum, U., & Zhan, J. (2025). *Sentiment analysis of movie reviews using BERT*. arXiv. <https://doi.org/10.48550/arXiv.2502.18841>
- Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2024). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Communication Methods and Measures*, 18(2), 124–141. <https://doi.org/10.1080/19312458.2023.2181319>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/15000000011>
- Pratiwi, R. W., Sari, Y., & Suyanto, Y. (2020). Attention-based BiLSTM for negation handling in Sentimen Analysis. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(4), 397–406. <https://doi.org/10.22146/ijccs.60733>
- Qixuan, Y. (2024). *Three-class text sentiment analysis based on LSTM*. arXiv. <https://doi.org/10.48550/arXiv.2412.17347>
- Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1), Article 6086. <https://doi.org/10.1038/s41598-024-56706-x>
- Salma. (2025, February 26). *Indonesia faces waste crisis, projected to generate 82 million tons by 2045*. Universitas Gadjah Mada. [Universitas Gadjah Mada](https://doi.org/10.31957/jbp.4189)

- Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, 27(4S), 4023–4031. <https://doi.org/10.53555/AJBR.v27i4S.4345>
- Sinapoy, M. I. K., Sibaroni, Y., & Prasetyowati, S. S. (2023). Comparison of LSTM and IndoBERT method in identifying hoax on Twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(3), 657–662. <https://doi.org/10.29207/resti.v7i3.4830>
- Sugiri, D. (2024). Upaya pengelolaan sampah melalui pelebagaan unit dalam Badan Layanan Umum Daerah (BLUD) di Indonesia. *Jurnal Ilmiah Akuntansi Manajemen*, 7(2), 177–185. <https://doi.org/10.35326/jiam.v7i2.6669>
- Trezza, D. (2023). To scrape or not to scrape, this is dilemma. The post-API scenario and implications on digital research. *Frontiers in Sociology*, 8, Article 1145038. <https://doi.org/10.3389/fsoc.2023.1145038>
- Vanacore, A., Pellegrino, M. S., & Ciardiello, A. (2024). Fair evaluation of classifier predictive performance based on binary confusion matrix. *Computational Statistics*, 39(1), 363–383. <https://doi.org/10.1007/s00180-022-01301-9>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *arXiv*. <https://doi.org/10.48550/arXiv.1706.03762>
- Wahyu W, A., Andana, H. H., Zeniarja, J., & Febriyanto, A. (2025). Sentiment analysis of the 2024 general election through Twitter using long-short-term memory algorithm. *Journal of Informatics and Web Engineering*, 4(2), 387–400. <https://doi.org/10.33093/jiwe.2025.4.2.25>
- Winata, G. I., Aji, A. F., Cahyawijaya, S., Mahendra, R., Koto, F., Romadhony, A., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Fung, P., Baldwin, T., Lau, J. H., Sennrich, R., & Ruder, S. (2023). *NusaX: Multilingual parallel sentiment dataset for 10 Indonesian local languages*. arXiv. <https://doi.org/10.48550/arXiv.2205.15960>
- Zeng, G. (2025). Invariance properties and evaluation metrics derived from the confusion matrix in multiclass classification. *Mathematics*, 13(16), Article 2609. <https://doi.org/10.3390/math13162609>
- Zhang, X., Liang, X., Zhiyuli, A., Zhang, S., Xu, R., & Wu, B. (2019). AT-LSTM: An attention-based LSTM model for financial time series prediction. *IOP Conference Series: Materials Science and Engineering*, 569(5), Article 052037. <https://doi.org/10.1088/1757-899X/569/5/052037>

Biographies of Authors

Riszki Wijayatun Pratiwi, Department of Informatics Management, Politeknik Assalaam Surakarta, Surakarta, Central Java 57169, Indonesia.

- Email: riszkiwp@gmail.com
- ORCID: 0000-0003-4381-9318
- Web of Science ResearcherID: N/A
- Scopus Author ID: 58988133700
- Homepage: <https://sinta.kemdiktisaintek.go.id/authors/profile/6776015>

Chayanita Sekar Wijaya, Department of Informatics Management, Politeknik Assalaam Surakarta, Surakarta, Central Java 57169, Indonesia.

- Email: chayanitasekarw@politeknikassalaam.ac.id
- ORCID: 0009-0000-0199-9889
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: <https://sinta.kemdiktisaintek.go.id/authors/profile/6899881>

Vilya Lakstian Catra Mulia, Department of Informatics Management, Politeknik Assalaam Surakarta, Surakarta, Central Java 57169, Indonesia.

- Email: v_lakstian@politeknikassalaam.ac.id
- ORCID: 0000-0002-3254-4059
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: <https://sinta.kemdiktisaintek.go.id/authors/profile/6116744>

Jumi Darsih, Department of Informatics Management, Politeknik Assalaam Surakarta, Surakarta, Central Java 57169, Indonesia.

- Email: jumidarsih@gmail.com
- ORCID: N/A
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A