The Eastasouth Journal of Information System and Computer Science
Vol. 4, No. 01, August 2026, pp. 43-50
ISSN: 3025-566X, DOI: 10.58812/esiscs.v4i01

Improving Sentiment Classification of Indonesia's School Zoning Admission Policy on Twitter

Renaldiansyah Gumay[1], Wiranto Herry Utomo[2]
[1] Master of Science in Information Technology, Faculty of Computing, President University
[2] Faculty of Computing, President University

Article Info

Article history:

Received Jul, 2026
Revised Aug, 2026
Accepted Aug, 2026

Keywords:

Accuracy;
Hybrid Approach;
InSet Lexicon;
LSTM;
Sentiment Analysis;
Zoning Policy

ABSTRACT

Since 2016, the Indonesian government has implemented a zoning-based policy for its online New Student Admission (Penerimaan Peserta Didik Baru, PPDB) system to ensure equitable access to quality education. Despite this intent, the policy has repeatedly drawn mixed public reactions, and in 2024 the debate intensified on social media, particularly on Twitter (X), which has become a primary channel for the public to express opinions on the policy. Understanding this sentiment at scale requires an approach that can cope with the unstructured, informal, and slang-heavy nature of Indonesian social media text. This study proposes a hybrid sentiment analysis approach that combines the InSet (Indonesia Sentiment) Lexicon with a Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) to classify public sentiment toward the 2024 PPDB zoning policy from 1,771 crawled tweets. Two scenarios were tested: Scenario 1 used an LSTM model trained on text alone, while Scenario 2 combined the LSTM with InSet Lexicon polarity scores as an auxiliary input. Scenario 1 achieved a training accuracy of 94.61% but only 66.18% on validation data, indicating overfitting, whereas Scenario 2 achieved a validation accuracy of 95%, with 92% precision, 96% recall, and a 93% F1-score. The results indicate that combining lexicon-based scoring with LSTM substantially improves classification accuracy and generalization on Indonesian-language, informal social media text, and that the overall public sentiment toward the 2024 zoning policy tended to be positive.

Corresponding Author:

Name: Renaldiansyah Gumay
Institution: Faculty of Computing, President University, Jl. Ki Hajar Dewantara, Cikarang, Bekasi, Jawa Barat, Indonesia
Email: renaldiansyah.gumay@president.ac.id

1. INTRODUCTION

The Government of Indonesia has operated an online New Student Admission (PPDB) system since 2016, designed to increase efficiency, accessibility, data integration, and convenience for prospective students and their guardians. Central to this system is a zoning policy intended to ensure equitable access to quality education by prioritizing school placement based on a student's residential proximity to the school. While the policy's objective is widely accepted, its implementation has been controversial: prior studies report conflicts of authority, overlapping jurisdictions among

regions, uneven distribution of school facilities, and dissatisfaction among parents of high-achieving students who feel disadvantaged by the zoning rule [1], [2], [3]. In 2024, this debate intensified on social media platforms, especially Twitter/X, where more than 500 million posts are published daily and where citizens routinely voice grievances and support toward government programs.

Assessing this large volume of informal, unstructured public commentary manually is impractical, which motivates the use of automated sentiment analysis. However, sentiment analysis of Indonesian-language social media text is particularly challenging because of nonstandard spelling, slang, code-mixing, and negation [4], [5], all of which reduce the reliability of purely lexicon-based or purely machine-learning-based classifiers.

This study addresses this gap by proposing a hybrid sentiment analysis approach that combines the InSet Lexicon with an LSTM-based RNN to classify Indonesian-language public sentiment toward the 2024 PPDB zoning policy, using tweets collected directly from Twitter/X. The specific objectives of this research are: (1) to produce an objective, data-driven characterization of public sentiment toward the zoning-based student admission policy; (2) to develop and evaluate a hybrid model that combines InSet Lexicon polarity scores with an LSTM customized for informal Indonesian text; and (3) to assess how much the addition of lexicon information improves classification effectiveness relative to a text-only LSTM baseline.

2. LITERATURE REVIEW

Sentiment analysis is the computational process of identifying and classifying opinions expressed in text as positive, negative, or neutral, and has been applied both in industry (e.g., product reviews) and in public policy evaluation, where it offers a scalable proxy for public opinion. Three broad technique families are commonly used: lexicon-based approaches, machine/deep-learning approaches, and hybrid approaches that combine the two [6], [7], [8].

Lexicon-based approaches classify text using a pre-built opinion dictionary in which each word carries a polarity weight; classification follows from the aggregate score of a text's constituent words. For Indonesian, the InSet Lexicon—comprising 3,609 positive and 6,609 negative words weighted from −5 to +5 by Indonesian-language experts and built from Twitter data—has been reported to outperform translated dictionaries such as SentiWordNet, the Liu Lexicon, and AFINN on Indonesian text [9], [10]. Its main limitation is limited coverage of evolving slang and its inability to model context or negation [11].

RNN-based deep-learning approaches, especially LSTM, are well suited to sentiment analysis because they capture sequential and contextual dependencies between words, which is important given that sentiment often depends on word order and local context rather than isolated keywords [12]. However, RNN/LSTM models typically need large labeled datasets to generalize well and are otherwise prone to overfitting.

Combining a lexicon-derived label with a sequence-based classifier is not itself a new technique family: hybrid lexicon-plus-RNN pipelines have been explored outside Indonesian, for instance on Arabic microblog text [13], and lexicon-derived auto-labeling followed by LSTM training has also been applied within Indonesia to other topics, such as the 2024 presidential election [12]. This study therefore does not claim novelty in the underlying architecture; rather, its contribution lies in applying this pipeline, with an explicit two-scenario ablation, to a topic and data source that have not previously been studied this way.

On the specific topic of the PPDB zoning policy, prior computational studies used K-Means combined with the Levenshtein Distance algorithm on Facebook and YouTube comments [14], K-Nearest Neighbor combined with Levenshtein Distance on YouTube comments (160 comments, 65% accuracy) [15], and Naïve Bayes combined with TensorFlow on Twitter opinion data concerning the same zoning

issue [16]; a separate qualitative study surveyed 102 parents across 13 Indonesian provinces on their perceptions of the policy using a mixed-method approach rather than social-media text mining [17]. Aside from [16], none of these studies used Twitter/X data, and none used a deep-learning sequence model; all operated on datasets at least an order of magnitude smaller than the 1,771 tweets used here.

More broadly, Indonesian sentiment-analysis research has moved rapidly toward transformer-based models: recent studies on education-policy sentiment on platform X report IndoBERT accuracies around 97% when trained on hybrid Lexicon-LLM labels [18], and studies on other government-policy topics (e.g., the capital relocation policy) report IndoBERT and BiLSTM-CNN accuracies in the 66–83% range [19], generally favoring manual or LLM-assisted labeling over purely automated lexicon labeling because of known label-quality concerns with the latter. This positions the present LSTM+InSet Lexicon hybrid model as a lighter-weight, more interpretable alternative to transformer-based pipelines, rather than as a methodologically novel technique; its main contributions are (1) the first application of this specific hybrid pipeline to PPDB zoning-policy sentiment on Twitter, (2) a substantially larger dataset than prior zoning-policy studies, and (3) a controlled two-scenario comparison that isolates the effect of adding lexicon information to an LSTM baseline.

3. METHODS

This research follows an experimental design in which a hybrid sentiment-classification pipeline—combining a rule-based lexicon score with a deep-learning sequence model—is applied to Indonesian-language Twitter data on the PPDB zoning policy. The overall workflow is summarized in Fig. 1.

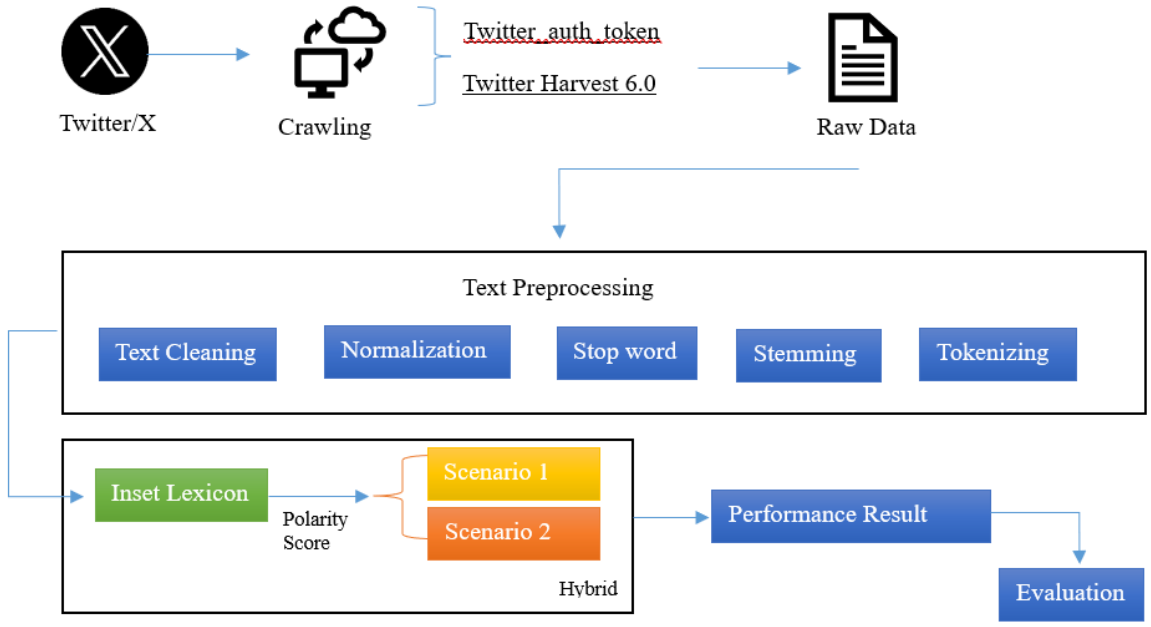

Figure 1. Overall research workflow, from data crawling to model evaluation.

3.1 *Data Collection*

Data were collected using the tweet-harvest 2.6.1 crawler with the keywords "sistem zonasi" (zoning system) and "zonasi PPDB." Three crawling sessions (18 and 21 July, and 24 August 2024) yielded 355, 726, and 690 tweets respectively, for a combined raw dataset of 1,771 tweets. Only the full_text field of each tweet was retained for subsequent processing.

3.2 *Text Pre-Processing*

Because raw tweets are unstructured, five sequential pre-

processing stages were applied, following prior evidence that both the choice and the order of pre-processing steps affect downstream model accuracy [20]: (1) text cleaning—case-folding, and removal of URLs, mentions, hashtags, digits, punctuation, and extraneous whitespace; (2) normalization—mapping non-standard/slang tokens to their formal Indonesian equivalents using a combined dictionary (the Colloquial Indonesian Lexicon, SentiStrength-ID, and Ibrohim's 2018 abusive-language dataset), covering 6,115 standard words; (3) stopword removal using the Sastrawi stopword list; (4) stemming using the Nondeterministic Context (NDETC) stemmer, an extension of Purwarianti's nondeterministic stemmer that resolves root-word ambiguity (e.g., "memalukan" → "malu" or "palu") using contextual cues [21]; and (5) tokenization into word-level units. After cleaning, duplicate removal reduced the corpus from 1,771 tweets (41,904 words) to 1,700 unique tweets (40,354 words).

3.3 *InSet Lexicon Scoring*

Each token in a pre-processed tweet is matched against the InSet Lexicon [9] to obtain a positive score, a negative score, and a neutral score. These are summed across all tokens in a tweet to obtain the overall polarity score. A tweet is labeled Positive if the polarity score is greater than 0, Negative if less than 0, and Neutral if equal to 0. Applying this scheme to the 1,700-tweet corpus produced 804 Positive, 668 Negative, and 228 Neutral tweets; these lexicon-derived labels serve as the ground truth for training and evaluating the LSTM models described next.

3.4 *Hybrid LSTM Architecture*

Tokens are mapped to dense vectors via an embedding layer before being processed by the sequence model. Two scenarios were compared. Scenario 1 (text-only LSTM): tokens are embedded and passed through an LSTM followed by a 256-unit dense layer (311,552 parameters) and a 3-unit softmax output layer (771 parameters) corresponding to the Negative, Neutral, and Positive classes. Scenario 2 (hybrid LSTM + InSet Lexicon): the text branch uses a smaller 32-unit LSTM with a 0.5 dropout rate to curb overfitting; the InSet Lexicon polarity score for each tweet is processed in parallel through a dense layer; the two branches are merged with a Concatenate layer and passed through a 64-unit dense layer with L2 regularization and a further dropout layer before the softmax output. Both models were trained with the Adam optimizer for 10 epochs and a batch size of 32; Scenario 2 additionally used early stopping (patience = 3 epochs, monitoring validation loss) to prevent overfitting. Fig. 2 illustrates the hybrid architecture.

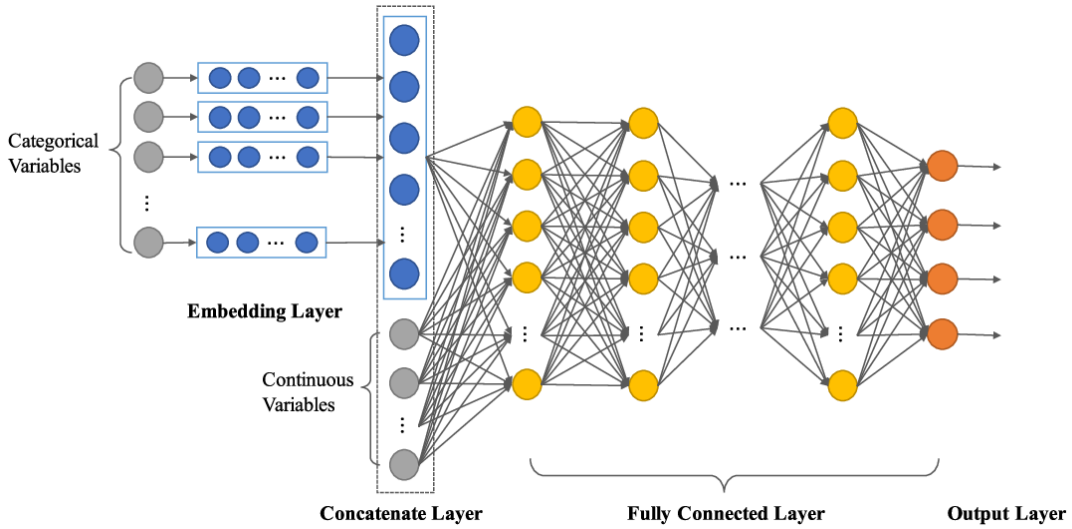

Figure 2. Hybrid RNN architecture combining text embeddings with InSet Lexicon scores

3.5 Model Evaluation

Both scenarios were evaluated on a held-out validation split using a confusion matrix, from which Accuracy, Precision, Recall, and F1-score were computed:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Precision = TP / (TP + FP) \quad (2)$$

$$Recall = TP / (TP + FN) \quad (3)$$

$$F1\text{-}score = (2 \times Precision \times Recall) / (Precision + Recall) \quad (4)$$

4. RESULTS AND DISCUSSION

4.1 Sentiment Distribution from the InSet Lexicon

Applying the InSet Lexicon alone to the 1,700-tweet corpus classified 804 tweets (47.3%) as Positive, 668 (39.3%) as Negative, and 228 (13.4%) as Neutral—an initial indication that public commentary on the 2024 zoning policy skewed positive, while still reflecting a substantial negative minority consistent with the controversy documented in prior studies.

4.2 Scenario 1: Text-Only LSTM

Over 10 training epochs, training accuracy rose from 42.06% to 94.61% while training loss fell from 1.6666 to 0.1570. However, validation accuracy fluctuated and ended at 66.18%, with validation loss rising sharply to 2.0588—a clear signature of overfitting: the model memorized patterns in the training data without generalizing to unseen tweets, particularly for the Neutral and Negative classes.

4.3 Scenario 2: Hybrid LSTM + InSet Lexicon

Incorporating the InSet Lexicon score as an auxiliary input, together with dropout, L2 regularization, a reduced LSTM size, a larger batch size, and early stopping, markedly stabilized training. At the final epoch, training accuracy reached 93.67% (loss 0.4003) and validation accuracy reached 89.71% (loss 0.5254), with training and validation curves tracking closely—indicating that the model generalized well rather than overfitting. On a subsequent held-out evaluation, the hybrid model achieved 94.42% accuracy with a loss of 0.3655.

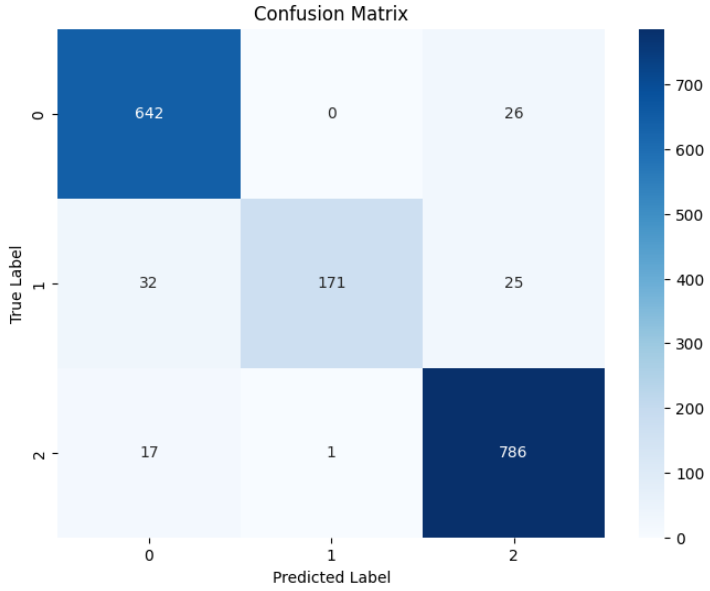

Figure 3. Confusion matrix of the hybrid model (Scenario 2). 0 = Negative, 1 = Neutral, 2 = Positive.

The confusion matrix shows strong performance for the Negative class (642 correctly classified, 26 misclassified as Positive, none as Neutral) and the Positive class (786 correct, 17 misclassified as Negative, 1 as Neutral). The Neutral class was comparatively weaker (171 correct versus 32 misclassified as Negative and 25 as Positive), suggesting that boundary cases with mixed or weak

polarity remain the hardest for the model to resolve—an expected outcome given that Neutral tweets, by construction, carry the least distinctive lexical signal.

4.4 Overall Evaluation Metrics

The InSet Lexicon score was the decisive addition: combining it with the LSTM increased held-out evaluation accuracy from 66.18% to 94.42%, a 28.24 percentage-point improvement, while precision (94%), recall (94%), and F1-score (94%) indicate a well-balanced classifier rather than one biased toward a single class.

4.5 Effect on Label Distribution

Re-scoring the corpus with the hybrid model shifted the label distribution relative to the lexicon-only baseline: Positive tweets increased from 804 to 837, Negative tweets increased from 668 to 691, and Neutral tweets decreased from 228 to 172. This pattern suggests that the LSTM's contextual modeling reassigned a portion of lexically ambiguous (Neutral) tweets to Positive or Negative once sentence-level context was taken into account, consistent with the lexicon-only method's known weakness in handling context-dependent or weak-polarity text.

4.6 Comparison with Prior Studies

Prior sentiment-analysis studies on the zoning policy specifically reported 65% accuracy using K-Nearest Neighbor with Levenshtein Distance on YouTube comments [15], and approximately 90% accuracy using K-Means with Levenshtein Distance on 200 Facebook/YouTube comments [14]; broader Indonesian-text hybrid studies (lexicon + machine learning) reported around 84% accuracy [22]. The present hybrid LSTM + InSet Lexicon approach reached 94.42% accuracy on a substantially larger dataset (1,771 raw / 1,700 cleaned tweets), suggesting that combining rule-based lexicon scoring with LSTM's sequential modeling can outperform these earlier distance-based and single-technique approaches on this specific topic, while also handling a larger and more heterogeneous Twitter corpus.

Placed against the wider, more recent Indonesian sentiment-analysis literature, however, this result is competitive but not exceptional. Studies applying IndoBERT to comparable government-policy topics report accuracies ranging from roughly 66% to 97%, depending on the topic, labeling method, and dataset size, with several outperforming the 94.42% achieved here [18], [19]. This suggests that the accuracy gain in this study comes primarily from two factors that are somewhat entangled: (1) the added InSet Lexicon feature, and (2) the regularization techniques (dropout, L2, early stopping, smaller batch/model size) introduced together with it in Scenario 2. Because both changes were introduced simultaneously, this study cannot fully isolate how much of the improvement is attributable to the lexicon feature itself versus the regularization changes—an ablation that future work should perform (see Section 5).

A further caveat is that the 94.42% accuracy is measured against lexicon-derived labels rather than an independently, manually annotated ground truth. Because the InSet Lexicon score is used both to generate the training/validation labels and as an input feature to the Scenario 2 model, part of the reported accuracy gain may reflect this label–feature dependency rather than a genuine improvement in identifying true public sentiment. This limitation, rather than a shortcoming unique to this study, mirrors a broader pattern noted in recent literature, where studies relying on fully automated lexicon-based labeling are generally considered less reliable than those using manual or LLM-assisted annotation. Consequently, the comparison with transformer-based studies above should be read as indicative rather than a strictly like-for-like benchmark.

5. CONCLUSION

This study developed and evaluated a hybrid sentiment-analysis approach that combines the InSet Lexicon with an LSTM-based RNN to analyze Indonesian-language Twitter sentiment toward the 2024 PPDB zoning policy. Using 1,771 crawled tweets (1,700 after de-duplication), a text-only LSTM (Scenario 1) achieved only 66.18% validation accuracy due to overfitting, while incorporating InSet Lexicon polarity scores as an auxiliary input (Scenario 2) raised held-out evaluation accuracy to 94.42%, with 94% precision, 94% recall, and a 94% F1-score. The lexicon-only classification suggests that overall public sentiment toward the 2024 zoning policy leaned positive (47.3% Positive versus 39.3% Negative and 13.4% Neutral), though a sizeable negative minority persisted. These findings support combining lexicon-based and deep-learning methods for sentiment analysis of informal, unstructured Indonesian social media text, and provide policymakers with an evidence-based, large-scale reading of public opinion on the zoning policy. Future work should expand data collection to additional platforms (e.g., Facebook, Instagram), apply cross-validation and additional regularization to further verify generalization, explore transformer-based models such as IndoBERT, and expand the InSet Lexicon's coverage of contemporary slang, while also using an independently, manually annotated test set to validate the lexicon-derived labels used in this study.

ACKNOWLEDGEMENTS

The author thanks the Faculty of Computing, President University, and the thesis advisor for guidance throughout this research.

REFERENCES

[1] Anzaldi and Sukmana, "SWOT analysis of the zoning system policy in new student admission," 2022.
[2] Y. Sulistyosari, A. Wardana, and S. A. Dwiningrum, "School zoning and equal education access in Indonesia," Int. J. Eval. Res. Educ. (IJERE), pp. 586–593, 2023.
[3] T. Widiyati and A. Sudrajat, "Conflict and overlapping authorities in the newly implemented school zoning policy in Indonesia: The case in the urban–rural regency of Magelang," in Proc. 2nd Int. Conf. Soc. Sci. Character Educ. (ICoSSCE), Atlantis Press, 2019.
[4] Christina and Sherly, "Sarcasm in sentiment analysis of Indonesian text: A literature review," J. Teknol. Inf., pp. 54–59, 2019.
[5] R. I. Terecha, F. Wahyudi, and U. M. Jannah, "Penanganan negasi dalam analisa sentimen Bahasa Indonesia," JUSIFOR, vol. 1, no. 1, pp. 51–58, 2022.
[6] Ariyanto and Chamidah, "Comparison of SVM and Naive Bayes methods for Indonesian sentiment analysis," 2021.
[7] E. Daniati and H. Utama, "Analisis sentimen dengan pendekatan ensemble learning dan word embedding pada Twitter," J. Inf. Syst. Manag. (JOISM), pp. 125–231, 2023.
[8] S. Kausar, X. Huahu, W. Ahmad, and M. Y. Shabir, "A sentiment polarity categorization technique for online product reviews," IEEE Access, vol. 8, pp. 3594–3605, 2020.
[9] F. Koto and Rahmaningtyas, "InSet lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in Int. Conf. Asian Lang. Process. (IALP), 2017, pp. 391–394.
[10] B. Liu, Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
[11] A. Nurkasanah and M. Hayaty, "Feature extraction using lexicon on the emotion recognition dataset of Indonesian text," Ultimatics J. Tek. Inform., vol. 14, no. 1, pp. 20–27, 2022.
[12] D. A. Firdlous, "Analisis sentimen publik Twitter terhadap Pemilu 2024 menggunakan model Long Short-Term Memory (LSTM)," Undergraduate thesis, Universitas Pendidikan Indonesia, 2023.
[13] K. Elshakankery et al., "Hybrid deep learning for sentiment polarity determination of Arabic microblogs," in Neural Information Processing (ICONIP), Part II, 2019.
[14] A. Al Farisi, Arini, Wardhani, Matin, and Y. Durachman, "K-Means algorithm and Levenshtein Distance algorithm for sentiment analysis of school zonation system policy," in 2021 6th Int. Conf. Informatics Comput. (ICIC), 2021, pp. 1–6.
[15] D. Anggraini and Tursina, "Sentiment analysis of new student admission zoning system policy using K-Nearest Neighbor and Levenshtein Distance," 2019.
[16] A. A. Dwisanny and S. Suptami, "Twitter opinion sentiment analysis based on new student admission zoning issues using the Naïve Bayes and TensorFlow methods," in 9th Int. Conf. Signal Process. Intell. Syst. (ICSPIS), 2023, pp. 1–8.
[17] M. N. Chozin, L. W. Savitri, R. E. Apriyanto, and M. Marzuki, "Public perception in evaluating Indonesia's school zoning admission policy: Insight into the ongoing debate," Al-Ishlah J. Pendidikan, vol. 17, no. 1, pp. 751–770, 2025.

[18] G. F. S. Medantoro and M. Muljono, "Comparative analysis of IndoBERT and classic machine learning models for sentiment classification of education policy on social media X," J. Appl. Inform. Comput. (JAIC), vol. 10, no. 1, pp. 548–557, 2026.
[19] W. Nugraha, M. T. Julianto, M. K. Najib, and E. Khatizah, "Public sentiment toward the Indonesian capital relocation policy on X using a BiLSTM-CNN model," J. Telematika, vol. 20, no. 2, pp. 114–126, 2025.
[20] S. Shevira, I. D. Suarjaya, and P. W. Buana, "Pengaruh kombinasi dan urutan pre-processing pada tweets Bahasa Indonesia," J. Ilm. Teknol. Komput., pp. 1–8, 2022.
[21] Bunyamin, F. A. Huda, and A. Ardiyanti, "Indonesian stemmer for ambiguous word based on context," in Int. Conf. Data Sci. Its Appl. (ICoDSA), 2021, pp. 1–9.
[22] S. J. Putra, I. Khalil, M. N. Gunawan, R. Amin, and T. Sutabari, "A hybrid model for social media sentiment analysis for Indonesian text," in Int. Conf. Inf. Integr. Web-based Appl. Serv., 2018, pp. 297–301.