JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. 9, No. 2, August 2026, pp. 190-198 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v9i2.12620 e-ISSN: 2656-1948

# IMPLEMENTATION OF GEOSPATIAL INTELLIGENCE FOR SENTIMENT ANALYSIS ON STUNTING POLICY IN BATANG REGENCY USING INDOBERT

**Dian Fitria Maharani[1*], Bambang Agus Herlambang[2], Nur Latifah Dwi Mutiara Sari[3]**

[1,2,3] Departement of Informatics , Faculty of Engineering and Informatics, Universitas PGRI Semarang, Semarang, 50125, Central Java, Indonesia
Email: [*1]dianmaharanifm@gmail.com, [2]bambangherlambang@upgris.ac.id, [3]nurlatifah@upgris.ac.id

(Received: 30 June 2026, Revised: 15 July 2026, Accepted: 12 August 2026)

**Abstract**

This study develops a Geospatial Artificial Intelligence (GeoAI)-based WebGIS that integrates IndoBERT sentiment classification to evaluate public percetion of stunting-management policy in Batang Regency, Central Java, Indoneia. The Cross-Industri Standart Process for Data Mining (CRISP-DM) framework was applied to the sentiment-analysis pipeline, while Rapid Application Development (RAD) governed the system construction. A total of 478 public-opinion responses were collected through questionnaires from residents of fifteen sub-districts, preprocessed through data cleaning, case folding, tokenizing, stopword removal, and stemming, then labeled and classified into positive, neutral, and negative sentiment using a fine-tuned IndoBERT model. The system was built with Python, Flask, Leaflet.js, and QGIS to visualize sentiment spatially. On a held-out test set of 96 samples, the model achieved 82.29% accuracy, 83.49% weighted precision, 82.29% weighted recall, and an 82.58% weighted F1-score (macro F1-score of 0.77), with class-weighted loss applied during fine-tuning to counter a severe class imbalance in the labeled dataset (Imbalance Ratio = 5.82). Spatial analysis showed that Bandar Sub-district recorded both the highest number of positive (37) and negative (8) responses, indicating the highest level of public engagement, Batang Sub-district recorded the highest number of neutral responses (46). System functionality was further validated through User Acceptance Testing and Black Box Testing, each covering nine functional scenarios spanning authentication, dashoard acces, sentiment-analysis display, spatial map interaction, and page navigation; all eighteen test scenarios were completed successfully (100% valid), confirming that the system operates correctly and satisfies the intended user requirements. The resulting GeoAI-based WebGIS enable policymakers to identify sub-districts requiring closer attention and design more targeted, evidence-based interventions. These findings demonstrate that integrating sentiment classification with spatial visualization provides greater insight into public perception than statistical data.

**Keywords**: *GeoAI, Stunting Policy, Natural Language Processing, Sentiment Analysis, IndoBERT*



**Corresponding Author: Dian Fitria Maharani*

## 1. INTRODUCTION

Stunting remains a major public-health challenge in Indonesia, affecting children's physical growth, cognitive development, and future productivity due to chronic nutritional deficiencies during the first 1.000 days of life[1]. The Indonesian government has implemented various program nutritional interventions, maternal and child collaboration to accelerate stunting reductions [2] However, evaluating the effectiveness of these policies requires more than statistical indicators; it also requires understanding public perception of their implementation.

Recent advances in Natural Language Processing (NLP) with public opinion being analyzed automatically through sentimentn analysis [3], [4]. IndoBERT, a transformer-based language model pre-trained on large Indonesian corpora, has shown strong performance in Indonesian sentiment-classification tasks and has been widely applied to public services, government policy, and social issue [3]. Meanwhile, Geospatial Artificial Intelligence (GeoAI) and WebGIS technologies are increasingly used to visualize spatial patterns and support location-based decision-making [5]. These two research streams are generally developed separately: sentiment analysis focuses on text classification while GeoAI emphasizes spatial analysis, resulting in limited integration between textual sentiment and spatial context [6].

Several studies illustrate this gap. Kaesmetan and Kalengkongan classified public sentiment toward stunting-related social media posts using IndoBERT and SVM, but without any spatial component [7]Siregar et al. combined IndoBERT with geo-tagged data to map sentiment toward a commercial service, not a government health policy or an integrated WebGIS [8]. Marsiska Driptufany et al. mapped the gegraphic distribution of stunting cases in Padang City using GIS-based regression, but relied entirely on quantitative case data without textual sentiment analysis [9]. To date, no study has combined IndoBERT-based sentiment classification with a GeoAI-based WebGIS specifically for evaluating public perception of stunting-management policy at the sub-district level, despite the growing potential of GeoAI for this kind of spatial decision-support [10].

Previous studies have also rarely combined a structured data mining framework for sentiment classification with a rapid system-development method within a single integrated platform [11], [12]. This study therefore proposes a GeoAI-based WebGIS that integrates an IndoBERT sentiment-classification model developed following the CRISP-DM framework with a system built using the RAD method, to visualize public sentiment toward stunting-management policy acroos sub-districts in Batang Regency and support direct policy evaluation from the community.

## 2. RESEARCH METHOD

This study integrates CRISP-DM, which aims to a structure the sentiment classification data processing pipeline, and RAD which governs the development of a GeoAI-based WebGIS system. CRISP-DM was chosen because of its structured, iteractive approach to building validated classification models [13] while RAD was chosen because of its short development cycle and user involvement suitable for system whose map interfaces require iterative adjustments based on stakeholder feedback [14].

### 2.1 Data Mining Process (CRISP-DM)

Bussines Understanding defines the research problem: evaluate public perceptions of stunting management policies in Batang Regency and visualize them spatially at the sub-district level. The data collected by Understanding are primary data through questionnaires distributed to residents in fifteen sub-districts in Batang Regency (Bandar, Banyuputih, Batang, Bawang, Blado, Gringsing, Kandeman, Limpung, Pecalungan, Reban, Subah, Tersono, Tulis, Warungasem, and Wonotunggal); each response consists of a free-text opinion regarding the current stunting management policies in each sub-district where the respondent resides.

Data Preparation applied five sequential preprocessing steps to the 478 collected opinion texts: (1) data cleaning to remove URLs, Number, punctuation, and extra whitespace; (2) case folding to lowercase all text; (3) tokenizing to split text into Sastrawi Indonesian stopword list; and (5) stemming using the Sastrawi stemmer to reduce inflected words to their root form. As an illustration, the raw response "Baik, semoga program nya tepat sasaran dan angka stunting menjadi menurun"; tokenized into eleven word tokens; reduced to ten tokens after stopword removal dropped the conjuction "dan"; and stemmed so that "sasaran" and "menurun" became "sasar" and "turun". Each preprocessed repsonse was then labeled by the authors into positive, neutral, or negative sentiment and tagged with its sub-district for spatial aggregation,

Modeling fine-tuned IndoBERT a transformer-based language model pre-trained on large-scale Indonesian corpora, by converting each opinion text into contextual token embeddings, processing them through multiple self-attention layers, and classifying the result through a Softmax output layer (Eq. 1). Training minimized the Cross Entropy Loss shown in Eq. 2, which measures the discrepancy between the predicted probability distribution and the true sentiment label.

Data Preparation yielded 478 labeled responses distributed unevenly across the three sentiment classes: 262 neutral (54.8%), 172 positive (36.0%), and only 45 negative (9.4%). This skew was quantified using three complementary imbalance metrics (Table 1): an Imbalance Ratio (IR), defined as the ratio between the majority- and minority-class counts, of 5.82; a normalized Shannon entropy of 0.837, where 1.0 denotes a perfectly balanced distribution; and a deviation from the balanced proportion of 47.9%. An IR above 3 is generally considered indicative of severe class imbalance, confirming that the negative class was substantially under-represented relative to the neutral and positive classes.

Table 1. Class Distribution and Imbalance Metrics of the Labeled Dataset

| Metric | Value |
|---|---|
| Neutral class | 262 (54.8%) |
| Positive class | 172 (36.0%) |
| Negatif class | 45 (9.4%) |
| Imbalance Ratio | 5.82 |
| Normalized Shannon entropy | 0.837 |
| Deviation from balanced distributin | 47.9% |

To prevent this imbalance from biasing the model toward the majority neutral class, five candidate strategies, no balancing, class-weighted loss, random oversampling, random undersampling, and SMOTE, were first compared using a TF-IDF and Logistic Regression baseline. In this comparison, balancing was applied only to the training split, while the held-out test split retained its original distribution to avoid data leakage into the evaluation. Class weighting and SMOTE achieved the highest macro F1-scores among the five strategies (0.624 and 0.631, respectively), both clearly outperforming the unweighted baseline (macro F1 = 0.557), for which the negative-class F1-score was 0.00. Because SMOTE generates synthetic feature vectors that cannot be mapped back to natural-language text, class-weighted loss was adopted as the imbalance-handling strategy for the final IndoBERT model, since

it operates directly on the original text without duplicating or synthesizing any response. The weight assigned to each class (Table 2) was computed as inversely proportional to its frequency in the training split, giving the negative class a weight of 3.55, more than five times that of the neutral class, and was incorporated directly into the cross-entropy loss function during fine-tuning (Eq. 2).

Table 2 Class Weights Used in the Weighted Cross-Entropy Loss

| Class | Weight |
|---|---|
| Negative | 3.55 |
| Neutral | 0.61 |
| Positive | 0.93 |

The fine=tuning configuration of the final IndoBERT model is summarized in Table 3. The pre-trained indobenchmark/indobert-base-p1 checkpoint was fine-tuned for a maximum of four epochs using a weighted cross-entropy loss, with model selection based on the highest macro F1-score across epochs; the checkpoint obtained at epoch 2 achieved the best macro F1-score and was retained as the final model.

Table 3. IndoBERT Fine-Tuning Configuration

| Parameter | Value |
|---|---|
| Pre-trained model | Indobenchmark/indobert-base-p1 |
| Number of classes | 3 (Negative, Neutral, Positive) |
| Max sequence length | 128 tokens |
| Batch size (train/eval) | 16 / 16 |
| Learning rate | $2x10^{-5}$ |
| Epochs (best checkpoint) | 4 (best at epoch 2) |
| Weight decay | 0.01 |
| Optimizer | AdamW (HuggingFace Trainer default) |
| Loss function | Weighted cross-entropy |
| Train / test split | 80% / 20%, stratified, random state = 42 |

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad (1)$$

$$L = -\sum y_i \log(\hat{y}_i) \quad (2)$$

Evaluation used accuracy (Eq. 3), precision (Eq. 4), recall (Eq. 5), and F1-score (Eq. 6), computed per class from the confusion matrix (TP, TN, FP, FN), then aggregated across the positive, neutral, and negative classes using the weighted-average formula in Eq. 7 to account for class imbalance.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1-Score = 2X\frac{precision\ x\ recal}{precision\ x\ recall} \quad (6)$$

$$M_w = \sum_{c=1}^{C} \frac{n_c}{N} M_c \quad (7)$$

In Eq. 7, C denotes the number of sentiment classes (three, in this study), n_c number of test samples belonging to class c, N the total number of test samples, and M_c the metric value (precision, recall, or F1-score) computed for class c; M_w is therefore the support weighted average of the per-class metric across aa classes.

Deployment integrated the best-performing checkpoint into the WebGIS backend so that classification results could automatically be aggregated per sub-distrct and rendered on the spatial dashboard.

### 2.2 System Development

The RAD method proceeded through four stages (Figure 1). Requirements Planning identified the need to visualize the spatial distribution of public sentiment, filter the map by sentiment category, and display a summary dashboard. User Design refined the interface, database structure, and map-layer design; the resulting design is presented as a UML use case diagram in Section 3.1. Construction built the system using Python with Flask for the backend and REST API, Leaflet.js for the interactive map, QGIS for preparing the GeoJSON boundaries of Batang Regency's sub-districts, and Visual Studio Code as the development environment; the fine-tuned IndoBERT model was wrapped as an inference service consumed by the Flask backend. Cutover tested the system's map rendering, filtering, popup details, and classification accuracy before release. A short iterative loop between User Design and Construction allowed the interface to be refined based on stakeholder feedback before finalization.

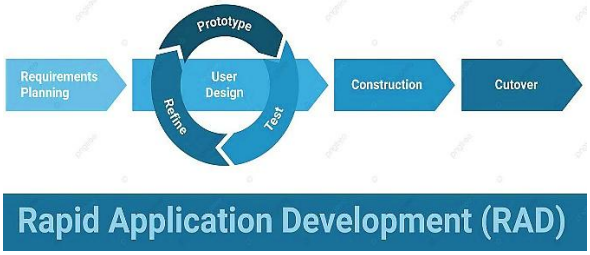



Figure 1. Application Development (RAD) Method

## 3. RESULTS AND DISCUSSION

### 3. 1 System Design

The system was modeled using UML. The use case diagram (Figure 2) shows two actors: Users, who can view the sentiment map, inspect sub-district detail, and filter displayed data; and Admin, who manages survey data, performs sentiment labeling, runs the IndoBERT model, and manages the spatial layer of sub-district boundaries. The end-to-end process flows from questionnaire distributin through the CRISP-DM pipeline to presentation on the Leaflet.js map, with an iterative feedback loop allowing the Admin to request labeling or interface adjustments after reviewing intermediate results.

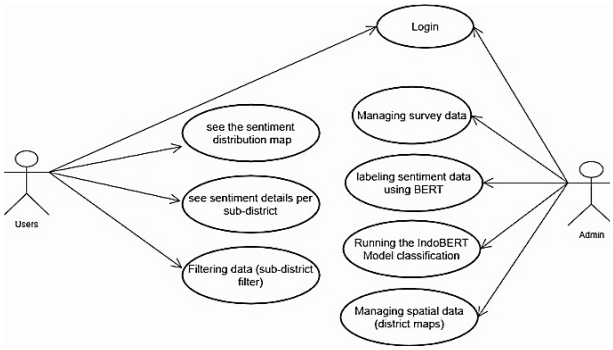



Figure 2. Use Case Diagram of the GeoAI-Based WebGIS

To complement the use case diagram, the activity diagram in Figure 3 illustrates the end-to-end process flow, from questionnaire distribution through the CRISP-DM data-mining pipeline to the presentation of results on the WebGIS map, including the iterative model-evaluation loop characteristic of the RAD approach.

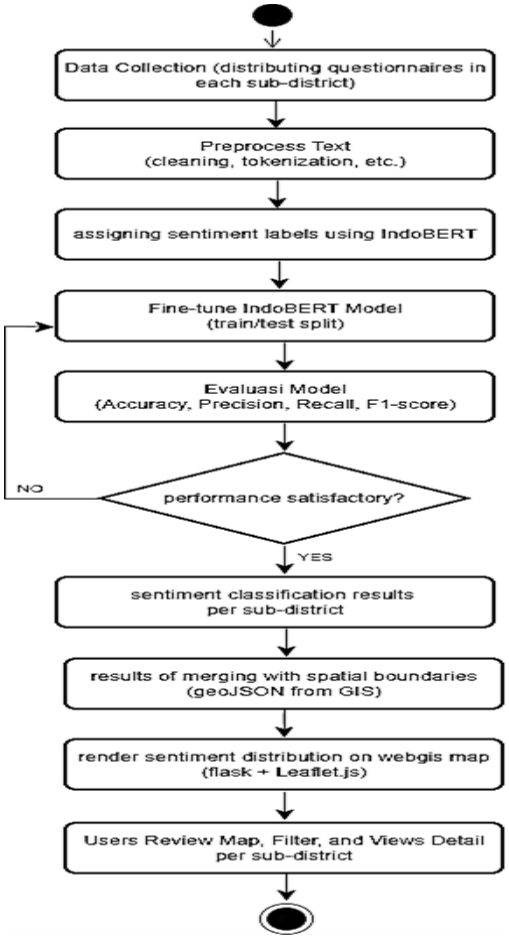



Figure 3. Activity Diagram of the GeoAI-Based WebGIS

### 3. 2 Spatial Integration of Public Sentiment

Before sentiment labeling, the 478 collected responses were first examined by their raw spatial distribution across the fifteen sub-districts of Batang Regency. The WebGIS interface provides an interactive pop-up for each sub-district listing the number of responses collected, the average opinion-text length, a qualitative volume category, and the most frequently occurring root words exracted from the pre-processed text. As illustrated for Blado Sub-district in Figure 4, the pop-up reports 47 responses collected, an average text length of 19.2 words, a volume category of "Very High", and the five most frequent root words "need", "health", "child", "community", and "infomation". These terms recur across most sub-districts, indicating that public discourse converges on similar concerns: the perceived need for intervention, child health and nutrition, and demand for community engagement or information, regardless of each sub-districts response volume.

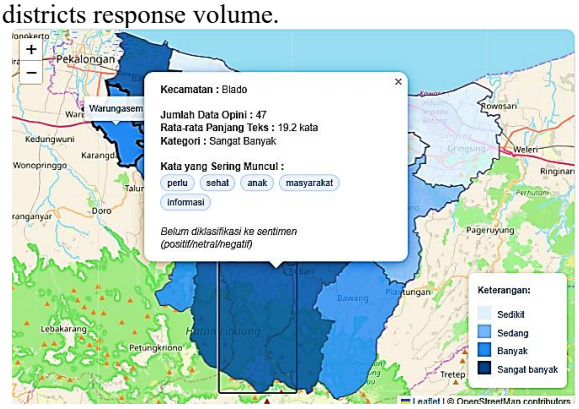



Figure 4. Spatial Integration of Survey Data (Before Labeling)

Each response was also examined by its average opinion text length per sub-district, as presented in Figure 5. Wonotunggal recorded the highest average text length, at approximately 31 words per response, followed by Pecalungan (around 24 words) and Tersono (around 21 words), while Bandar recorded one of the lowest average text lengths, at around 16 words, despite contributing the largest number of both positive and negative responses. This pattern suggest that a sub-districts response volume and the elaborateness of its residents opinions are not necessaliry correlated: some sub-districts submit fewer but more detailed opinions, while others, such as Bandar, submit a larger number of relatively concise responses a distinction relevant to the labeling stage, since longer texts generally provide richer contextual cues for the IndoBERT model to capture sentiment polarity.

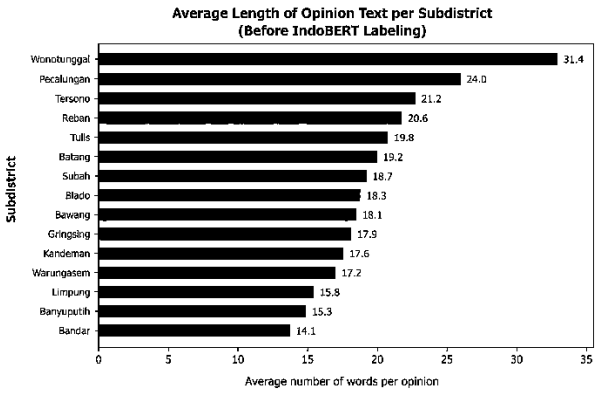



Figure 5. Average Length of Opinion Text per Sub-District

After labeling, a total of 478 labeled responses were obtained: 172 positive (36.0%), 261 neutral (54.6%), and 45 negative (9.4%) (Table 4). Bandar

recorded both the highest number of positive (37) and negative (8) responses. Indicating the highest level of public engagement and diversity of opinion, while Batang the seat of the regency government recorded the highest number of neutral responses (46), suggesting a more cautious, wait and see public stance. Tulis recorded no negative response at all.

Table 4. Total District Sentiment Integration Results

| Sub-district | Positive | Neutral | Negative |
|---|---|---|---|
| Bandar | 37 | 24 | 8 |
| Banyuputih | 5 | 13 | 2 |
| Batang | 18 | 46 | 3 |
| Bawang | 23 | 14 | 4 |
| Blado | 10 | 33 | 4 |
| Gringsing | 7 | 10 | 2 |
| Kandeman | 8 | 4 | 3 |
| Limpung | 4 | 13 | 4 |
| Pecalungan | 8 | 11 | 2 |
| Reban | 12 | 27 | 3 |
| Subah | 9 | 9 | 3 |
| Tersono | 7 | 13 | 2 |
| Tulis | 10 | 15 | 0 |
| Warungasem | 7 | 14 | 1 |
| Wonotunggal | 7 | 15 | 4 |
| **Total** | **172** | **261** | **45** |

A word frequency analysis was also performed on the entire corpus of 478 pre-processed opinion texts to identify which topics dominated public discourse irrespective and stemming, the corpus yielded 6,845 word token; Table 5 lists the fifteen most frequent stemmed words, and Figure 6 presents the corresponding word cloud, in which font size is proportional to word frequency.

Table 5. Most Frequent Words in Public Opinion Text

| Rank | Word (stemmed) | Frequency |
|---|---|---|
| 1 | perlu | 370 |
| 2 | sehat | 204 |
| 3 | anak | 194 |
| 4 | masyarakat | 160 |
| 5 | baik | 138 |
| 6 | bantu | 125 |
| 7 | gizi | 124 |
| 8 | tingkat | 118 |
| 9 | makan | 103 |
| 10 | perintah | 102 |
| 11 | kurang | 97 |
| 12 | informasi | 96 |
| 13 | edukasi | 94 |
| 14 | kena | 85 |
| 15 | tambah | 84 |

Figure 6. Word Claud of the Stunting Policy Opinion Text

Table 5 shows that "perlu" (need) was by far the most frequent term, followed by "sehat" (healthy) and "child" consistent with the dominance of neutral sentiment in Table 1 and indicating that respondents largely framed their opinions around what is still needed, such as additional health services, nutrition, or information, rather than simply approving or rejecting the existing policy. The prominence of "masyarakat" (community), "bantu" (help), "gizi" (nutrition). and related terms further indicates that respondents repeatedly called for stronger community involvement, nutritional support, and public education as part of stunting- management efforts.

### 3.3 Sentiment Classification Model Performance

The IndoBERT model was fine-tuned for a maximum of four epochs, with the checkpoint achieving the highest macro F1-score (obtained at epoch 2) automatically selected as the final model. Table 6 summarizes the training and validation results recorded at each epoch, while Figure 7 and Figure 8 visualize the corresponding training curves, recorded directly from the model-training log.

Table 6. Training and Validation Results per Epoch

| Epoch | Training Loss | Validation Loss | Accuracy | Macro F1 |
|---|---|---|---|---|
| 1 | 0.9842 | 0.7928 | 0.6979 | 0.6211 |
| 2 | 0.6123 | 0.5954 | 0.8229 | 0.7726 |
| 3 | 0.2914 | 0.6617 | 0.8021 | 0.7347 |
| 4 | 0.1716 | 0.6763 | 0.8125 | 0.7531 |

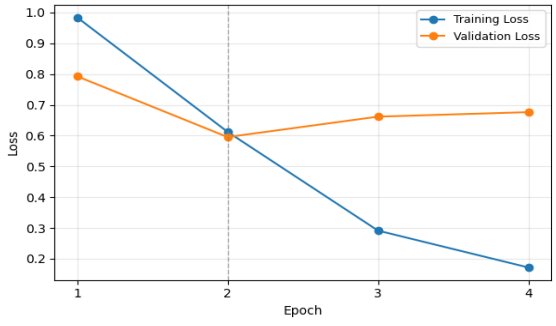


Figure 7. Training Loss vs Validation Loss per Epoch

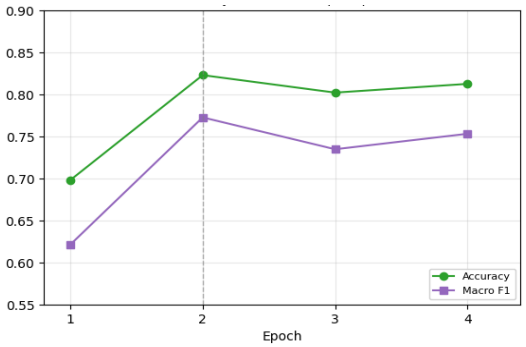


Figure 8. Validation Accuracy per Epoch

As shown in Figure 7, the validation loss and training loss reached their lowest and most favorable balance at epoch 2 (training loss 0.6123, validation loss 0.5954); from epoch 3 onward the training loss continued to fall sharply (donw to 0.1716 at epoch 4) while the validation loss rose again (0.6617-0.6763), a clear signature of overfitting. Figure 8 shows that both accuracy and Macro F1 peaked at epoch 2 (0.229 and 0.7726, respectively, Table 6) and declined there after. The epoch-2 checkpoint, selected automatically as the one with the highest Macro F1, was therefore retained as the final model. It should be noted that, in the current training configuration, the evaluation split monitored during training to select this checkpoint is the same 96 sample split subsequently reported as the test-set result below; consequently, the checkpiont-selection process and the final performance figures are not based on fully independent data, and the reported metrics should be interpreted as an upper=bound estimate of generalization performance. Future work should reserve a separate validation split, distinct from the test set, for checkpoint selection.

On the 96 sample evaluation split, the final model (epoch-2 checkpoint) achieved an overall accuracy of 82%, weighted precision of 83%, weighted recall of 82%, and weighted F1-score of 82%, with a macro F1-score of 0.77 (Table 7).

Table 7. Overall Evaluation Result on the Test Set

| Metric | Value |
|---|---|
| Accuracy | 82.29% |
| Precision | 83.49% |
| Recall | 82.29% |
| F1-score | 82.58% |

The per-class report (Table 8) and confusion Matrix (Figure 9) show that class-weighted fine-tuning improved the models sensitivity to the negative lass relative to an unweighted baseline trained on the same split, which had a negative-class recall of 0.00, correctly identifying none of the 9 negative test responses. With class-weighted loss, 6 of 9 negative responses were correctly classified (recall =0.67), while its precision (0.60) indicates a moderate number of false positive, largely responses that were actually neutral (2 cases) or positive (1 case) but predicted as negative. The positive class achieved the highest recall (0.91; 31 0f 34 correctly classified), while the neutral class, despite the largest support (53 responses), showed a comparatively lower recall of 0.79 (42 of 53 correctly classified), with 3 neutral responses misclassified as negative and 8 as positive. This reduction in neutral recall, relative to the 0.89 achieved by the unweighted baseline, is an expected consequence of the class-weighting strategy rather than a separate weakness of the model: because the loss function penalizes misclassification of the negative class 3.55 times more heavily than the neutral class (Table 2), the models decision boundary was deliberately shifted to become more willing to predict the minority classes for opinions that carry mixed or ambiguous cues, at the cost of occasionally reassigning borderline neutral responses to an adjacent class. Qualitatively, opinion framed as statements of remaining need (for example, opinions built around the word “perlu”/”need”, the most frequent term in the corpus, see Table 5) are inherently ambiguous between a neutral, descriptive reading and a mildly negative or positive one, which likely explains why most of the neutral misclassifications were distributed toward both the negative and positive classes rather than concentrated in one direction. Given the small absolute size of the negative clss in the test set (9 of 96 samples), each individual misclassification also has a disproportionately large effect on its precision and recall percentages, so these per-class figures should be interpreted with that sample-size caveat in mind. Future work should enlarge the negative-labeled portion of the dataset, for example through targeted data collection, to reduce reliance on loss reweighting alone and to obtain more stable per-class estimates.

Table 8. Classification Report per Sentiment Class

| Class | Precison | Recall | F1-score | Support |
|---|---|---|---|---|
| Negative | 0.60 | 0.67 | 0.63 | 9 |
| Neutral | 0.91 | 0.79 | 0.85 | 53 |
| Positive | 0.78 | 0.91 | 0.84 | 34 |
| Macro avg | 0.76 | 0.79 | 0.77 | 96 |
| Weighted avg | 0.83 | 0.82 | 0.82 | 96 |

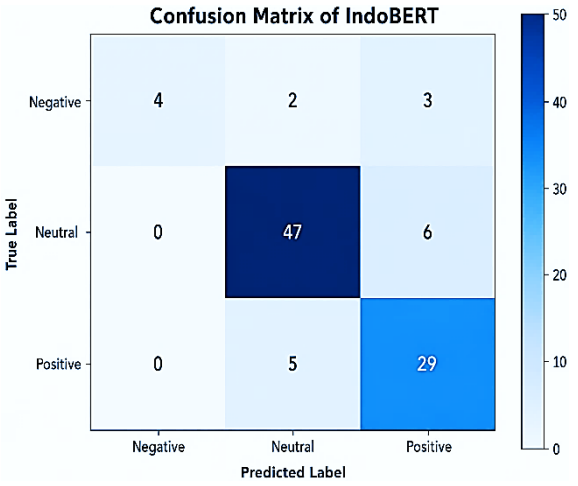
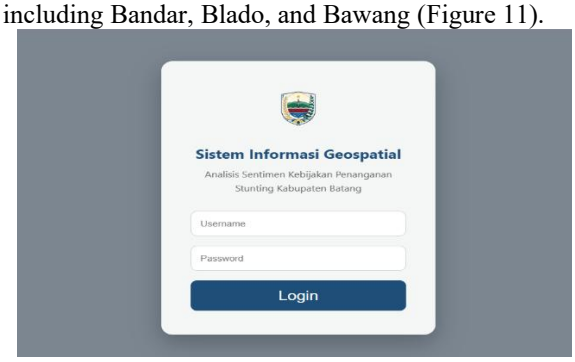



Figure 9. Confusion Matrix of the IndoBERT Sentiment Classification Model

### 3.4 GeoAI-Based WebGIS Visualization

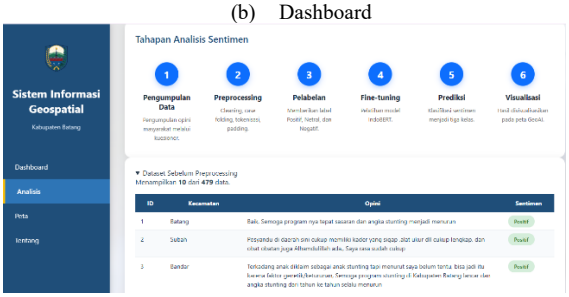
The classification results were integrated into an interactive WebGIS built with Leaflet.js and Flask, where each sub-district polygon prepared in QGIS is color-coded by its dominant sentiment category and linked to a pop-up showing detailed sentiment counts. As shown in Figure 10, Of the, the system consists of four main interface component. Figure 10(a) is the login page, which restricts access to the Admin role responsible for managing survey data, sentiment labeling, and the spatial layer of sub-district boundaries. Figure 10(b) is the dashboard, main page which is designed as a main gateway that simplifies user interaction to quickly direct users into the contents of the system. Figure 10(c) displays the

sentiment analysis results in tabular form, allowing the Users to review the opinion text classified using IndoBERT along with the predicted sentiment labels with the corresponding sub-districts, as well as to filter the results by sentiment category. Figure 10(d) is the mapping-visualization page, the core output of the system, which dispalys the color-coded chloropleth map described above and allows the user to click on any sub-district polygon to view detailed sentiment calculations via a pop-up window. Of the fifteen sub-districts, nine are dominated by positive sentiment, two by neutral sentiment, and four by negative sentiment, which are mostly concentrated in the sub-districts south and southwest of the district capital, including Bandar, Blado, and Bawang (Figure 11).

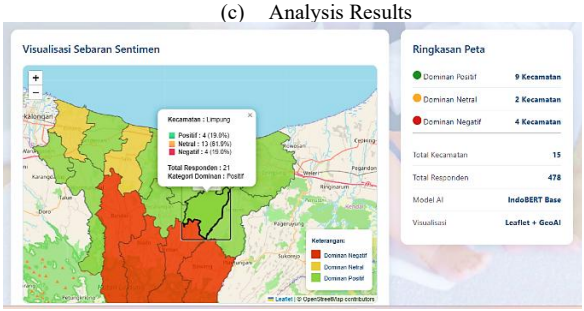



(a) Login Page


(b) Dashboard




(c) Analysis Results




(d) Mapping Visualization

Figure 10. WebGIS Inteface

### 3.5 Discussion

Integrating IndoBERT-based sentiment classification with a GeoAI-based WebGIS reveals spatial variationd in public perception that cannot be identified from aggregate statistics alone. Bandar, Blado, and Bawang show relatively stronger negative sentiment, while Bandar also records the highest positive engagement indicating the greatest diversity of opinion regarding the policys implementation. Batang, by contrast, is dominated neutral sentiment, suggesting more descriptive or balanced public opinion. These geographic differences indicate that public perception of the policy varies across sub-districts and therefore requires differentiated policy responses. By combining sentiment classification with spatial visualization, the proposed WebGIS enables policymakers to identify priority areas, target communication and intervention strategies, and evaluate policy implementation based on geographically distributed public feedback.

As of the current stage of development, the GeoAI-based WebGIS operates as a functioning prototype covering all fifteen sub-districts of Batang Regency, with the Admin and Users interfaces described in Section 3.4 already implemented and connected to the fine-tuned IndoBERT model. The classification component currently deployed is the epoch-2, class-weighted checkpoint reported in Section 3.3, which raised negative-class recall from 0.00 in the unweighted baseline to 0.67, at the cost of a moderate reduction in neutral-class recall (from 0.89 to 0.79) and an overall accuracy of 82.29%. The underlying dataset remains limited to a single data-collected round of 478 questionnaire responses, so the class imbalance identified in Section 2.1 (IR = 5.82) persists in absolute terms even though its effect on the model has been mitigated through loss reweighting rather than eliminated through additional data collection. In its present form, the evaluation split used to select the best training checkpoint also served as the final test split (Section 3.3), so the reported metrics should currently be treated as an upper-bound indication of real-world performance rather than a fully independent estimate. Consequently, at this stage the system is best positioned as a decision-support and screening tool that surfaces spatial sentiment patterns for further review by policymakers, rather than as a fully automated detector of negative public sentiment, until the dataset is enlarged and the evaluation protocol is strengthened with an independent validation split.

## 4. CONCLUSION

The study successfully developed a GeoAI-based WebGIS that integrates sentiment classificationusing IndoBERT, the CRISP-DM framework, and the RAD method to support spatial evaluation of public sentiment toward stunting management policies in Batang Regency. The refined IndoBERT model, fine-tuned with a class-weighted loss to address a severe class imbalance (IR = 5.82), achieved satisfactory

classification performance (82.29% accuracy, weighted F1-score of 82.58%, macro F1-score of 0.77), and spatial analysis revealed that public perception varies acrros sub-districts, with Bandar showing the highest engagement and diversity of opinion and Batang showing the most neutral, balanced responses. These findings demonstrate that integrating sentiment analysis with GeoAI-based spatial visualization provides a more comprehensive understanding of community perceptions than sentiment statisticts alone, and that the resulting WebGIS can serve as an evidence-based decision support tool for local governments to prioritize areas, design targeted interventions, and evaluate stunting management policy more effectively. The study is limited by the small number of negative sentiment samples and its sub-district-level spatial resolution; future work should expand the dataset, improve minority-class representation, extend the analysis to the village level, and integrate real-time public opinion from social media.

## 5. REFERENCE


[1] UNICEF, World Health Organization, and World Bank Group, "Levels and trends in child malnutrition: UNICEF/WHO/World Bank Group joint child malnutrition estimates: Key findings of the 2020 edition," 2020. Accessed: Aug. 07, 2026. [Online]. Available: https://www.who.int/publications/i/item/9789240003576

[2] Musmarlinda, Yuanita Windusari, and Haerawati Idris, "KEBIJAKAN PENANGGULANGAN STUNTING DI INDONESIA BERDASARKAN PROGRAM SANITASI TOTAL BERBASIS MASYARAKAT," *Jurnal Kesehatan*, vol. 13, no. Supplementary 1, pp. 109–114, Sep. 2022, doi: 10.35730/jk.v12i0.808.

[3] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," Nov. 2020. [Online]. Available: http://arxiv.org/abs/2011.00677

[4] D. Jurafsky and J. H. Martin, *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Third Edition draft. Stanford University, 2026. Accessed: Aug. 07, 2026. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/

[5] D. Lunga *et al.*, "GeoAI at ACM SIGSPATIAL: The New Frontier of Geospatial Artificial Intelligence Research," *SIGSPATIAL Special*, vol. 13, no. 1–3, pp. 21–32, Dec. 2022, doi: 10.1145/3578484.3578491.

[6] M. N. Kamel Boulos, G. Peng, and T. Vopham, "An overview of GeoAI applications in health and healthcare," *Int. J. Health Geogr.*, vol. 18, no. 1, p. article 7, May 2019, doi: 10.1186/s12942-019-0171-2.

[7] Y. R. Kaesmetan and W. W. Kalengkongan, "Classification of Public Sentiment Related to Stunting in Indonesia Using BERT and SVM," *Journal of Business and Audit Information System (JBASE)*, vol. 8, no. 2, pp. 11–33, 2025, doi: 10.23965/jbase.v8i2.8960.

[8] R. A. Siregar, F. Fitriyani, and L. S. Darfiansa, "Sentiment Analysis Of Indihome Service Based On Geo Location Using The Bert Model On Platform X," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 2975–2990, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.4993.

[9] D. Marsiska Driptufany, I. Armi, and D. Arini, "Pemetaan dan Analisis Spasial Kasus Stunting di Kota Padang," *Jurnal Sains dan Ilmu Terapan*, vol. 8, no. 2, pp. 41–49, Dec. 2025, doi: 10.59061/jsit.v8i2.1163.

[10] G. Mai *et al.*, "Towards the next generation of Geospatial Artificial Intelligence," *International Journal of Applied Earth Observation and Geoinformation*, vol. 136, p. Article 104368, Feb. 2025, doi: 10.1016/j.jag.2025.104368.

[11] G. W. Sasmito, D. S. Wibowo, and D. Dairoh, "Implementation of Rapid Application Development Method in the Development of Geographic Information Systems of Industrial Centers," *Journal of Information and Communication Convergence Engineering*, vol. 18, no. 3, pp. 194–200, Sep. 2020, doi: 10.6109/jicce.2020.18.3.194.

[12] E. N. Hayati, "Artificial Intelligence in Web-Based Geographic Information Systems: A Cross-Disciplinary Review," *International Journal of Research and Applied Technology*, vol. 4, no. 2, pp. 309–316, Dec. 2024, doi: 10.34010/injuratech.v4i2.16433.

[13] D. Ruswanti, D. Susilo, and R. Riani, "Implementasi CRISP-DM pada Data Mining untuk Melakukan Prediksi Pendapatan dengan Algoritma C.45," *Go Infotech: Jurnal Ilmiah STMIK AUB*, vol. 30, no. 1, pp. 111–121, Jun. 2024, doi: 10.36309/goi.v30i1.266.

[14] E. Sofiati, "IMPLEMENTASI METODE RAPID APPLICATION DEVELOPMENT (RAD) DALAM PERANCANGAN SISTEM INFORMASI BIMBINGAN KONSELING DI SMKN 1 SIJUNJUNG," *JIKO (Jurnal Informatika dan Komputer)*, vol. 8, no. 2, p. 437, Sep. 2024, doi: 10.26798/jiko.v8i2.1318.

[15] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT 2019*, Jill Burstein, Christy Doran, and Thamar Solorio, Eds., Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[16] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using

fine-tuning IndoBERT and R-CNN," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.

[17] M. A. Nur, N. Umar, Z. Feng, and H. Gani, "EVALUATION OF INDOBERT AND ROBERTA: PERFORMANCE OF INDONESIAN LANGUAGE TRANSFORMER MODELS IN SENTIMENT CLASSIFICATION," *JIKO (Jurnal Informatika dan Komputer)*, vol. 8, no. 2, pp. 121–127, Jul. 2025, doi: 10.33387/jiko.v8i2.9988.