

 Open Access

*E-ISSN: 2620 - 4872**Vol.07, No.02**Doi:**https://doi.org/10.21009/j-koma.v7i2.04*

*Received: 28 Agst 2024**Accepted: 09 Nov 2024**Published: 20 Des 2024*

Keywords:*Sentiment Analysis;**LSTM;**Free School Lunch Policy;**Word2Vec;**YouTube Comments.*

Correspondence Email:ari.hendarnoi@unj.ac.id*

Sentiment Analysis of Indonesia's Free School Lunch Policy Using LSTM and Word2Vec on YouTube Comments

Boban Anderson¹, Med Irzal², Ari Hendarno³^{1,2,3}Computer Science, Universitas Negeri Jakarta.**Abstract** (10pt)

This study analyzes public sentiment toward Indonesia's free school lunch policy using sentiment classification on YouTube comments. Data were collected from 5,640 videos, resulting in 485,097 comments, with 392,576 comments used for training and testing. The dataset was preprocessed through cleaning, tokenization, normalization, stopword removal, and stemming. Word2Vec was used for word embedding, and sentiment classification was performed using an LSTM neural network. The model achieved 82.56% accuracy on training data but 57.00% on manually labeled test data. The final sentiment distribution shows that negative sentiment slightly dominates, reflecting public skepticism about budget use and program effectiveness. Frequent keywords such as Indonesia, Prabowo, school, and corruption highlight key concerns. These results provide valuable insights for policymakers to improve communication and address public concerns. Future research should expand data sources, refine labeling, and test hybrid deep learning models to enhance classification performance.

INTRODUCTION

Indonesia's 2024 general election marked a significant political shift, bringing a new administration under President Prabowo Subianto and Vice President Gibran Rakabuming Raka. One of the flagship social welfare programs announced during their campaign is the Free School Lunch policy, aimed at providing nutritious meals for school children nationwide. The initiative seeks to address Indonesia's high stunting rate, which stood at 21.6% in 2022 [1], ranking the country 27th globally and fifth highest in Asia. By targeting children's nutritional needs, the policy aspires to foster a healthier, more productive future generation and reduce household expenses for vulnerable families.

Since its announcement, the free lunch policy has sparked widespread public debate across various social media platforms, particularly on YouTube. Many citizens have voiced support, highlighting its potential to alleviate child malnutrition and strengthen human capital. Others have raised concerns about the policy's fiscal feasibility, arguing that the required budget could be reallocated to other pressing needs, such as education subsidies.

Public sentiment plays a critical role in shaping the implementation and success of government policies. Analyzing such sentiment provides valuable insights into how the public perceives, accepts, or resists new initiatives. Sentiment analysis — as a subfield of Natural Language Processing (NLP) and text mining — enables the extraction of opinions and attitudes from large volumes of unstructured text data [2]. Various machine learning and deep learning algorithms have been applied for sentiment classification, with Long Short-Term Memory (LSTM) networks showing promising results for handling sequential text data [3].

While previous studies have explored public sentiment on diverse policy issues — such as Indonesia's capital city relocation and the implementation of road pricing schemes — limited research has specifically examined reactions to the free lunch program through the lens of sentiment analysis on

YouTube comments. Considering YouTube’s massive user base and active comment sections, it serves as a valuable source for capturing spontaneous public discourse.

Therefore, this study aims to analyze public sentiment towards Indonesia’s free school lunch policy by classifying YouTube comments into positive, negative, or neutral categories using the LSTM algorithm. The findings are expected to offer insights that may assist policymakers in understanding public attitudes and refining strategies for effective policy communication and implementation.

METHOD

2.1 Research Design and Data Collection

This study follows a structured research design that begins with data collection, text cleaning, and preprocessing, followed by feature extraction, modeling, and evaluation. The research flow is summarized in Figure 1.

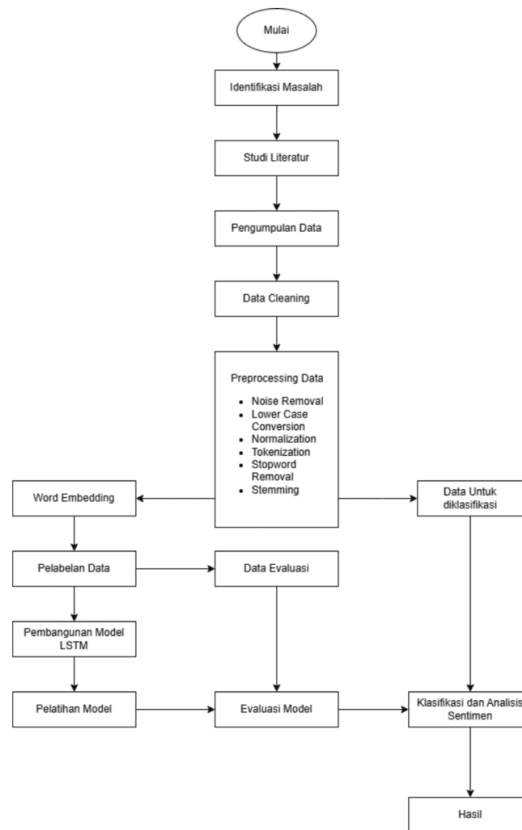


Figure 1. Research flowchart for sentiment analysis using Word2Vec and LSTM.

The dataset was collected using a web scraping technique targeting Twitter posts related to [TOPIK SPESIFIK, misalnya sentiment on online learning] over a defined period. Collected tweets were saved in CSV format and filtered to remove irrelevant or duplicate entries. The final dataset served as the input for the cleaning and preprocessing stages described in the next section.

2.2 Data Cleaning and Preprocessing

The collected text data were first cleaned to remove irrelevant information that could reduce model accuracy. This cleaning process involved eliminating duplicate entries, unwanted characters, HTML tags, URLs, numbers, emojis, and punctuation. Once cleaned, the dataset underwent a series of preprocessing steps to prepare the text for analysis.

The preprocessing stage consisted of six main steps:

- 1) converting all text to lowercase to maintain consistency;
- 2) removing noise such as special symbols and extra whitespace;
- 3) tokenizing the text into individual words;
- 4) normalizing terms by standardizing variations of the same word;
- 5) removing stopwords that do not contribute meaningful context; and
- 6) applying stemming to reduce words to their root form.

This comprehensive cleaning and preprocessing pipeline ensured that the input text was uniform and ready for feature extraction using the Word2Vec model in the next stage [4], [5], [6].

2.3 Word Embedding

To convert tokens into numerical vectors understandable by the LSTM model, the Word2Vec algorithm was used. Word2Vec learns dense vector representations of words by capturing their contextual relationships within a defined window size [7], [8]. This approach enhances the model's ability to recognize semantic similarities and word meanings in the Indonesian language context.

2.4 Data Splitting

The labeled dataset was divided into training (70%), validation (15%), and testing (15%) subsets. This split ensures the model is trained effectively, tuned with validation data, and evaluated fairly on unseen data.

2.5 LSTM Model Development

The sentiment classification task was performed using a Long Short-Term Memory (LSTM) neural network, which is well-suited for handling sequential text data [9], [10]. The model architecture included:

- An input embedding layer initialized with Word2Vec vectors.
- One or more LSTM layers to capture sequential dependencies.
- A dense output layer with softmax activation to classify comments into the three sentiment categories.

The model was trained iteratively, with hyperparameters such as the number of units, epochs, and batch size adjusted to balance performance and computational efficiency.

2.6 Evaluation Metrics

Model performance was evaluated using the multi-class confusion matrix, providing counts of true positives, true negatives, false positives, and false negatives for each sentiment class.

The following standard equations were used:

1) **Accuracy:**

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100$$

2) **Precision:**

$$Precision = \frac{TP}{TP+FP} \times 100$$

3) **Recall:**

$$Recall = \frac{TP}{TP+FN} \times 100$$

4) **F1-Score:**

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where:

- TP (True Positive): Correctly predicted positive comments
- TN (True Negative): Correctly predicted negative or neutral comments
- FP (False Positive): Incorrectly predicted positive comments
- FN (False Negative): Positive comments incorrectly predicted as other classes.

For this study, a multi-class confusion matrix was constructed to handle the three sentiment categories: positive, negative, and neutral. Each class's precision, recall, and F1-score were calculated accordingly.

RESULTS AND DISCUSSION

3.1 Data Overview

A total of 5,640 YouTube videos related to Indonesia's free school lunch policy were retrieved using the YouTube Data API v3. From these videos, 485,097 unique comments were collected within the timeframe from 28 November 2023 to 28 December 2024. To ensure relevance and consistency, only videos containing at least five comments were included in the final dataset.

3.2 Preprocessing Summary

To ensure the raw YouTube comment data were suitable for sentiment classification, several preprocessing steps were applied. These steps aimed to clean, standardize, and optimize the textual data for vector representation and input into the LSTM model. The main preprocessing pipeline consisted of six sequential stages:

- 1) Lowercase Conversion: All characters in the comments were converted to lowercase to prevent duplicate tokens caused by case differences.
- 2) Noise Removal: Unnecessary elements such as numbers, URLs, hashtags, HTML tags, emojis, extra spaces, special symbols, and duplicate comments were removed.
- 3) Tokenization: Each comment was split into individual tokens (words) using the NLTK tokenizer, which separates text based on linguistic rules.
- 4) Normalization: Slang words and non-standard terms were replaced with standard Indonesian words using a manually compiled slang dictionary sourced from Kaggle.
- 5) Stopword Removal: Commonly used words that do not carry significant meaning (stopwords) were filtered out based on an Indonesian stopwords list from the NLTK library.
- 6) Stemming: Words were reduced to their root forms using the Sastrawi library, ensuring morphological consistency within the corpus.

These steps resulted in a cleaner and more semantically consistent dataset, ready for word embedding and subsequent sentiment classification. A sample output of the preprocessing pipeline is shown in Table 1, demonstrating how raw comments were progressively transformed at each stage.

3.3 Sentiment Labeling and Balancing

After the preprocessing phase, sentiment labeling was conducted to classify each cleaned comment into one of three sentiment categories: positive, negative, or neutral.

Labeling was performed automatically using a pretrained Indonesian RoBERTa Base Sentiment Classifier, which assigns a sentiment score based on the text context. This approach enabled efficient annotation of a large dataset while maintaining consistent labeling rules.

The initial labeling results showed an imbalanced class distribution, with a clear dominance of negative sentiments. Out of 392,576 cleaned comments, the model classified approximately 95,032 comments as negative (57.84%), 89,056 as positive (23.50%), and 93,259 as neutral (18.66%). This imbalance is illustrated in Figure 2.

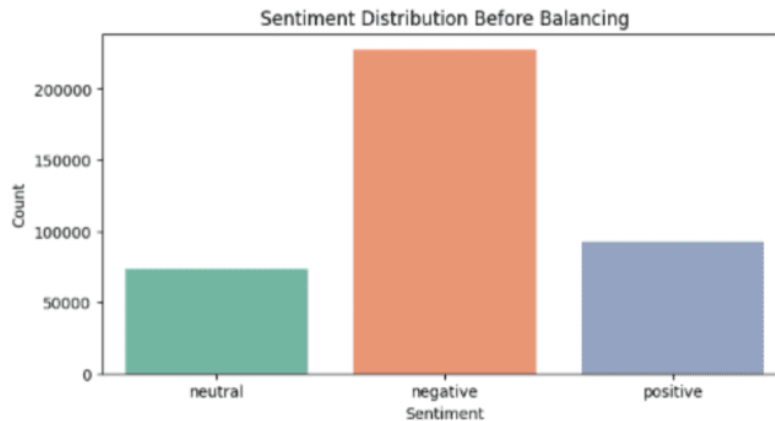


Figure 2. Distribution of Sentiment Labels Before Balancing

Imbalanced datasets pose a challenge for training supervised machine learning models, as the model may become biased toward the majority class, leading to poor generalization for minority classes. To address this issue, a data balancing technique was applied. The study employed a combination of random oversampling for minority classes (positive and neutral) and random undersampling for the majority class (negative). This ensured that each sentiment category contributed equally to model training, improving the classifier's ability to recognize all classes effectively.

3.4 Word Embedding and LSTM Model

The preprocessed dataset was converted into word vectors using the Word2Vec algorithm implemented with the Gensim library. The purpose of this stage was to transform words into numerical vector representations that could be used as input for the LSTM model. The Word2Vec model was trained using the cleaned text data, with key parameters including a vector size of 120, a context window size of 5, a minimum word count of 3, and multiple CPU cores for parallel processing. The skip-gram (sg=1) approach was used to predict surrounding words within a defined window.

Once the Word2Vec model was built, all tokenized comments were converted into corresponding vector representations and padded to ensure equal sequence length. If a comment contained fewer tokens than the maximum length, zero-padding was applied; if it exceeded the limit, it was truncated to match the defined length. The resulting embeddings were stored in HDF5 format to optimize memory usage and enable batch processing for large volumes of text.

After embedding, the sentiment classification was performed using a Long Short-Term Memory (LSTM) network developed with the PyTorch library. The LSTM architecture consisted of:

- Input-to-Hidden and Hidden-to-Hidden layers, with parameters W (input-to-hidden) and U (hidden-to-hidden) to process sequential data.
- Bias and Batch Normalization, which stabilized activation shifts and improved training speed.
- Four gates: input gate, forget gate, output gate, and candidate gate, controlling the flow of information across time steps.

The model used two stacked LSTM layers to capture deeper sequential patterns in the comment data, with dropout layers (probability 0.2) added after each LSTM layer to reduce overfitting.

Following the LSTM layers, a fully connected linear layer with a ReLU activation was applied to enhance the network's capacity to learn non-linear patterns. The final dense layer used a softmax activation to output probabilities across the three sentiment classes: positive, negative, and neutral. The model was trained using the CrossEntropyLoss function and optimized with the Adam Optimizer at a learning rate of 0.001. Mixed precision training was enabled with `torch.amp.GradScaler()` to improve computational efficiency. Training was conducted over 50 epochs using a balanced dataset of 277,347 comments.

The detailed training log is shown in Table 1, and the convergence of training loss and accuracy across epochs is visualized in Figure 3.

Tabel 1. Training Log of Sentiment Classification Model

Epoch	Loss	Accuracy	Time (seconds)
1	0.9028	63.31%	509.0
5	0.8288	71.47%	887.1
10	0.7881	75.97%	1350.8
20	0.7583	79.13%	2282.3
50	0.7247	82.56%	5083.4

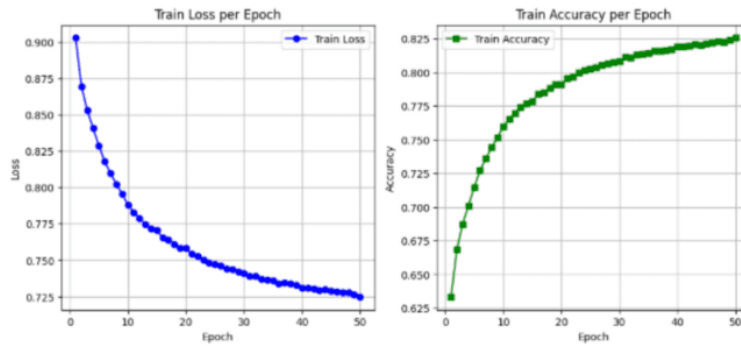


Figure 3. Training Loss and Accuracy per Epoch

The results show that the training loss decreased steadily, while accuracy increased from 63.31% at epoch 1 to 82.56% at epoch 50, indicating that the model successfully learned to classify sentiment in the training data.

3.5 Model Evaluation

To assess the classification performance of the trained LSTM model, an evaluation was conducted using a manually labeled test dataset consisting of 400 YouTube comments. This test set was annotated by human evaluators to ensure the reliability of the ground truth labels. The purpose of this evaluation was to measure how well the trained model could generalize when classifying new, unseen data.

The model's predictions were compared to the true labels, and the results are summarized in the confusion matrix shown in Figure 4. The confusion matrix illustrates the distribution of correctly and incorrectly classified comments across the three sentiment categories: negative, neutral, and positive.

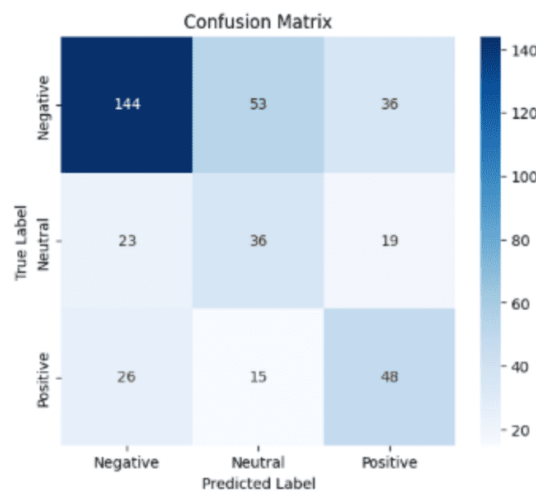


Figure 4. Confusion Matrix of Sentiment Classification on Test Data

Based on the confusion matrix, the following performance metrics were obtained for the test dataset:

- Accuracy: 57.00%

- Precision: 51.94%
- Recall: 53.96%
- F1-Score: 52.39%

Although the training accuracy reached 82.56%, the performance on the manually labeled test data showed a significant drop. This indicates that while the LSTM model performed well during training, its generalization capability on real-world noisy text data was more limited. Contributing factors include the complexity of informal language in YouTube comments, the potential for imbalanced sentiment classes, and differences between the automated labeling used during training and the human annotations used for testing.

These results highlight the importance of addressing class imbalance, noise handling, and labeling consistency to improve sentiment classification models for Indonesian social media text.

3.6 Sentiment Distribution and Wordcloud

After testing and labeling, the trained LSTM model was applied to the full cleaned dataset consisting of 392,576 YouTube comments. This prediction stage generated final sentiment labels distributed across three categories: positive, negative, and neutral. The distribution of the predicted sentiment classes is illustrated in Figure 5.

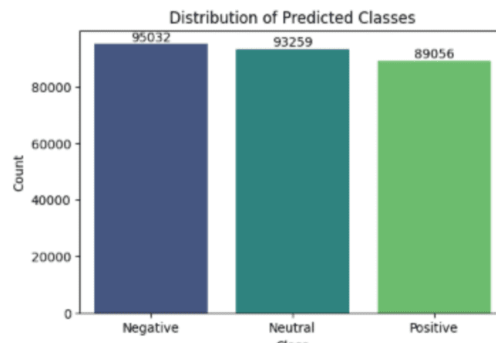


Figure 5. Final Sentiment Class Distribution

The results indicate that negative sentiments were dominant among the comments related to the free school lunch policy, followed by neutral and positive sentiments. This pattern suggests a significant portion of public responses contained skepticism or criticism, while supportive and neutral expressions were relatively balanced.

To gain deeper insights into the content of each sentiment class, a word cloud was generated for each category, highlighting the most frequent terms used in the comments. The word clouds are presented in Figures 6.



Figure 6. (a) Word Cloud for Negative Sentiment, (b) Word Cloud for Neutral Sentiment, (c) Word Cloud for Positive Sentiment.

Frequent words appearing across all sentiment classes include Indonesia, Prabowo, perintah (government order), korupsi (corruption), sekolah (school), and anak (child). These keywords reflect the topics and concerns commonly discussed by the public in response to the free school lunch policy.

The word clouds show that the same core issues are addressed with varying tones — indicating that while many comments convey disapproval or criticism, some also show support and constructive discussion around education and government programs.

3.7 Discussion

The findings indicate that public sentiment towards the free school lunch policy remains mixed, with negative comments slightly dominating over neutral and positive ones. This shows that although the program is welcomed by many, concerns about budget use and implementation still shape public opinion. While the LSTM model performed well on training data (82.56% accuracy), its test performance dropped to 57%, highlighting challenges in handling informal and noisy YouTube comments. Differences between automated and manual labeling, slang variations, and overlapping sentiments contributed to this gap.

Frequent terms like Indonesia, Prabowo, school, and corruption illustrate recurring public topics and worries. These insights can guide policymakers to address budget transparency and communication strategies more clearly. This research is limited to YouTube data and a single deep learning approach. Future work should expand data sources, explore hybrid models, and include more refined labeling to improve sentiment classification performance.

CONCLUSION

This study analyzed public sentiment towards Indonesia's free school lunch policy by classifying YouTube comments using a Long Short-Term Memory (LSTM) model combined with Word2Vec embeddings. The data collection covered 5,640 videos and 485,097 comments, resulting in 392,576 cleaned comments for sentiment analysis. The LSTM model achieved a training accuracy of 82.56%, but its test performance was lower at 57.00%, showing challenges in handling informal, noisy text data.

The final sentiment prediction indicated that negative sentiment slightly dominated over neutral and positive sentiments. Frequent keywords revealed concerns about budget allocation, corruption, and the overall implementation of the program. These findings suggest that while the policy is generally supported, public skepticism remains significant. The results highlight the need for transparent communication strategies to address concerns and improve public trust.

Future research should consider expanding data sources beyond YouTube, refining labeling methods, and testing hybrid deep learning models to enhance classification performance for Indonesian social media sentiment analysis.

REFERENCES

- [1] A. Iskandar, I. Tanzaha, M. N. Asri, and A. R. Wulandari, "Analysis of Dietary Patterns and Food Security in Relation to Children's Nutritional Status in Central Sulawesi, 2021–2023," *Journal of Global Nutrition*, vol. 5, no. 1, pp. 430–442, Apr. 2025, doi: 10.53823/jgn.v5i1.107.
- [2] J. Jayasudha and M. Thilagu, "A Survey on Sentimental Analysis of Student Reviews Using Natural Language Processing (NLP) and Text Mining," 2022, pp. 365–378. doi: 10.1007/978-3-031-23233-6_27.
- [3] N. Shi, Z. Chen, L. Chen, and R. S. T. Lee, "CNO-LSTM: A Chaotic Neural Oscillatory Long Short-Term Memory Model for Text Classification," *IEEE Access*, vol. 10, pp. 129564–129579, 2022, doi: 10.1109/ACCESS.2022.3228600.
- [4] S. Hong, J. Kim, H.-G. Woo, Y.-C. Kim, and C. Lee, "Screening ideas in the early stages of technology development: A word2vec and convolutional neural network approach," *Technovation*, vol. 112, p. 102407, Apr. 2022, doi: 10.1016/j.technovation.2021.102407.
- [5] J. Ma, L. Wang, Y.-R. Zhang, W. Yuan, and W. Guo, "An integrated latent Dirichlet allocation and Word2vec method for generating the topic evolution of mental models from global to local," *Expert Syst Appl*, vol. 212, p. 118695, Feb. 2023, doi: 10.1016/j.eswa.2022.118695.
- [6] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec And Long Short-Term Memory (LSTM) For Indonesian Hotel Reviews," *Procedia Comput Sci*, vol. 179, pp. 728–735, 2021, doi: 10.1016/j.procs.2021.01.061.

-
- [7] H. D. Abubakar and M. Umar, "Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec," *SLU Journal of Science and Technology*, vol. 4, no. 1 & 2, pp. 27–33, Aug. 2022, doi: 10.56471/slujst.v4i.266.
- [8] D. Držík and J. Kapusta, "Effect of dimension size and window size on word embedding in classification tasks," Jul. 08, 2024. doi: 10.21203/rs.3.rs-4532901/v1.
- [9] U. D. Gandhi, P. Malarvizhi Kumar, G. Chandra Babu, and G. Karthick, "Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)," *Wirel Pers Commun*, May 2021, doi: 10.1007/s11277-021-08580-3.
- [10] M. Umer, I. Ashraf, A. Mehmood, S. Kumari, S. Ullah, and G. Sang Choi, "Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model," *Comput Intell*, vol. 37, no. 1, pp. 409–434, Feb. 2021, doi: 10.1111/coin.12415.