Programming and Algorithm Fundamentals

E-ISSN: 3123-979X

Research Article

Graph Convolutional Network with TF-IDF Embeddings for Public Sentiment Classification on Indonesia's Free Nutritious Meal Program

Zhulfani Faisal Adam

Diponegoro University, Indonesia email: zhulfani@students.undip.ac.id
*Corresponding Author: zhulfani@students.undip.ac.id

Abstract. The Free Nutritious Meal Program, implemented as a national priority policy, has generated a wide array of public responses on social media, reflecting society's perception of the initiative. However, accurately classifying sentiment from short-text content such as tweets remains a challenging task due to limited context, informal language, and high lexical variability. Traditional machine learning models often fall short in capturing the complex semantic and structural information embedded in such data. To address this issue, this study proposes a sentiment classification model based on the Graph Convolutional Network (GCN) architecture, utilizing TF-IDF embeddings as the text representation method. A dataset comprising 5,979 labeled tweets was collected from the social media platform X. A document graph was constructed by computing cosine similarity between TF-IDF vector representations of the tweets, and this graph served as input to the GCN model. The proposed method was evaluated using accuracy, precision, recall, and F1-score. Experimental results show that the TF-IDF-based GCN model achieved an accuracy of 82.71% and an F1-score of 82.65%, outperforming conventional classifiers such as Logistic Regression, Support Vector Machine, and Naive Bayes. These findings demonstrate the effectiveness of integrating semantic-rich embeddings and graph-based learning in classifying public sentiment toward policy-related topics on social media.

Keywords: Free Nutritious Meal Program; Graph Convolutional Network; Indonesian NLP; Sentiment Analysis; Social Media X; TF-IDF.

1. Introduction

Stunting remains a persistent public health and developmental issue in Indonesia, affecting children's growth, cognitive performance, and long-term productivity. Nearly one in three Indonesian children under five suffers from stunted growth, contributing to a cycle of poverty and underdevelopment [1] (Srikandi Alifya et al., 2025). To mitigate this, the Indonesian government introduced the Free Nutritious Meal Program in 2024, targeting school-aged children and pregnant women through daily meal provisions (Risdiansyah et al., 2026; Aqiilah et al., 2026).

The program immediately triggered mixed public reactions, especially on social media platforms such as X (formerly Twitter), where citizens voiced both support and criticism. Social media has become a crucial channel for gauging public perception of government initiatives, but the informal and noisy nature of posts presents unique analytical challenges [2] (Muriyatmoko et al., 2026).

Sentiment analysis is widely used in Natural Language Processing (NLP) to assess opinions in textual data and classify them as positive, negative, or neutral. However, traditional models such as Support Vector Machines (SVM), Random Forests, or Logistic Regression often underperform on short-text social media datasets due to sparse information and lack of relational context [3] (Ifadah et al., 2026). Moreover, imbalanced class distributions frequently bias models toward the dominant sentiment class.

To address these issues, recent works have employed Graph Convolutional Networks (GCN), which extend deep learning to structured graph representations [6] (Darnoto et al., 2026). By modeling tweets as nodes and their similarity as edges, GCNs can incorporate inter-document relationships often missed by conventional methods. It has been shown that

Received: Februari 21, 2026
Revised: March 26, 2026
Accepted: April 27, 2026
Published: April 30, 2026
Curr. Ver.: April 30, 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

combining TF-IDF embeddings with document-level graph structures significantly improves classification performance [7] (Krisnawan et al., 2022).

Building upon this foundation, we propose a sentiment classification model for Indonesian-language tweets using GCN, TF-IDF representation, and undersampling to balance class distributions. The contributions of this work are: (1) constructing a graph from cosine similarity of TF-IDF vectors, (2) integrating a two-layer GCN for sentiment classification, and (3) evaluating the model against five baseline classifiers. Our model achieves 82.71% accuracy and outperforms all baselines, especially in distinguishing neutral sentiments.

The remainder of this paper is organized as follows. Section 2 presents related work. Section 3 describes the proposed methodology. Section 4 presents the results and discussion. Section 5 provides a comparison with state-of-the-art methods. Section 6 concludes the paper.

2. Literature Review

Several prior studies have explored sentiment analysis on Indonesian social media platforms, particularly Twitter, using traditional machine learning models. Hermansyah and Hasibuan employed Random Forest and Logistic Regression to classify sentiments regarding the Free Nutritious Meal Program, achieving an accuracy of 85% and 83%, respectively [2]. While their results were promising, the models relied on bag-of-words representations, which limited their ability to capture semantic relationships between documents.

Similarly, Triningsih and Prasetya used Support Vector Machines (SVM) to classify sentiments on tweets related to the same government initiative [3]. Their study achieved 85.74% accuracy, outperforming other traditional models. However, SVM, like many classical algorithms, treats documents independently and does not utilize contextual or relational information between samples.

Recent developments in deep learning have introduced Graph Convolutional Networks (GCN) as a powerful alternative for document classification. GCNs extend neural networks to operate on graph-structured data by propagating node features across graph edges [6]. This makes GCN particularly suitable for short-text sentiment classification, where documents can be modeled as nodes connected by similarity-based edges.

Khemani et al. proposed an enhanced GCN approach that combines Term Frequency-Inverse Document Frequency (TF-IDF) embeddings with document graphs to improve classification accuracy on short-text datasets [7]. Their experiments demonstrated that incorporating TF-IDF into the graph representation helped the model retain important lexical information, outperforming traditional word embedding techniques.

Building on these findings, our work applies a similar graph-based approach but focuses on Indonesian-language sentiment analysis within a policy-driven context. We also address data imbalance—an often overlooked issue in prior works by integrating undersampling to ensure fair representation across sentiment classes.

3. Research Method

This research proposes a sentiment classification model based on Graph Convolutional Network (GCN) using TF-IDF and undersampling. The stages include: data collection, preprocessing, graph construction, model architecture, and training configuration.

Data Collection and Labeling

The dataset consists of tweets related to the Free Nutritious Meal Program, collected from November 1, 2024, to March 31, 2025, using keywords such as "makan siang gratis" and "makan bergizi gratis". A total of 5,979 tweets were obtained.

Manual annotation was carried out to classify each tweet into three sentiment classes: positive (1,526 tweets), negative (2,930 tweets), and neutral (1,523 tweets), based on tone, expressions, and context.

Text Preprocessing

Effective preprocessing techniques, such as stopword removal and text normalization, have been shown to significantly improve tweet classification accuracy [8]. The following preprocessing steps were applied:

a. Case Folding: Lowercasing all characters.
b. Symbol Cleaning: Removing symbols, hashtags, user mentions, links, and emojis.
c. Normalization: Using a lexicon of informal-to-formal Indonesian words.
d. Tokenization: Splitting text into word tokens.

Undersampling

To handle class imbalance, undersampling was applied to the positive and negative classes by selecting tweets with the highest token count (i.e., longest tweets) until they matched the smallest class (neutral: 1,156 tweets). The final balanced dataset consisted of 4,569 tweets total, with equal representation across all sentiment classes.

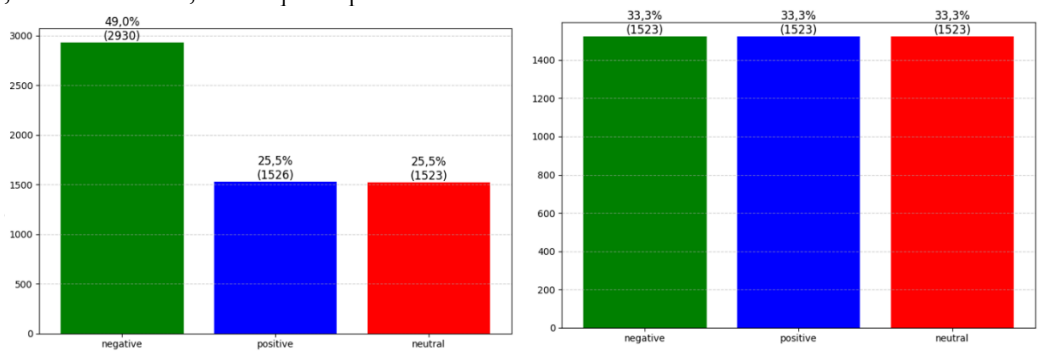


Figure 3. Distribution of sentiment classes (**a**) before undersampling; (**b**) after undersampling.

TF-IDF and Graph Construction

Each tweet was represented using Term Frequency–Inverse Document Frequency (TF-IDF), resulting in a sparse feature matrix X ∈ ℝN×F where N is the number of tweets and F the vocabulary size. To capture inter-document relationships, a pairwise cosine similarity matrix $\boldsymbol{S} \in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$ was computed between TF-IDF vectors, as shown in Eq. (1):

$$s_{ij} = \frac{x_i \cdot x_j}{||x_i|| \cdot ||x_j||} \quad \text{(i)}$$

An undirected graph G = (V, E) was then constructed, where each tweet corresponds to a node $v_i \in V$, and an edge $e_{ij} \in E$ exists if $S_{ij} > \tau$, with τ = 0.05 as the similarity threshold. This graph was encoded into an adjacency matrix A, augmented with self-loops, and normalized to form Â for input into the GCN layers. This approach ensures that semantically similar tweets are directly connected in the graph, allowing the GCN to propagate sentiment information through local neighborhood structures.

GCN Architecture

The proposed model employs a 5-layer Graph Convolutional Network (GCN) to perform node classification on the tweet graph. GCN extends traditional convolution to graph-structured data by aggregating neighborhood features, allowing the model to incorporate both lexical and relational information [6]. Deeper GCN architectures, while uncommon, have been shown to improve performance when combined with appropriate regularization strategies, such as dropout and weight decay [7].

Let $X \in \mathbb{R}^{N\times F}$ be the input feature matrix of N nodes (tweets), each with F-dimensional TF-IDF embeddings. Let Â be the symmetric normalized adjacency matrix with added self-loops. The forward propagation rule is defined as in Eq. (2):

$$H^{(l+1)} = ReLU(\hat{A}\, H^{(l)}\, W^{(l)}) \quad \text{(ii)}$$

where $H^{(0)} = X, W^{(l)}$ is the learnable weight matrix at layer l, and $\hat{A} = D^{-½}(A + I)D^{-½}$ is the normalized adjacency matrix with I as the identity matrix. The final output Z after the last GCN layer is passed through a softmax activation to yield class probabilities, as shown in Eq. (3):

$$Z = softmax(H^{(L)}) \quad \text{(iii)}$$

A total of five GCN layers were used to allow information aggregation from higher-order neighbors in the graph. To prevent overfitting and over-smoothing, dropout with a rate of 0.1 and weight decay of 5×10^{-4} were applied during training.

The use of TF-IDF embeddings retains key lexical features, while the GCN layers enable relational learning across semantically similar tweets. This hybrid approach has shown strong performance in recent studies on short-text sentiment classification [7].

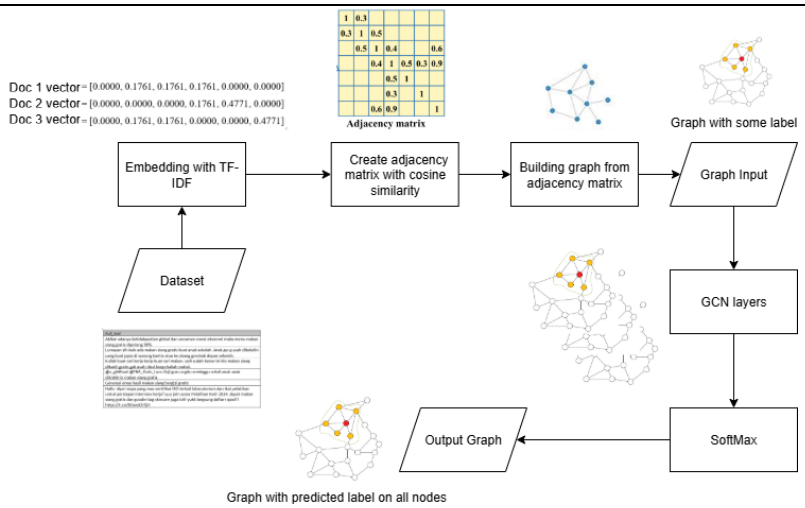


Figure 3. Style menu in MS Word.

Training Configuration

The final dataset was split into training (3,655 tweets), validation (457 tweets), and test (457 tweets) sets, stratified by class as detailed in Table 1. The model was implemented in PyTorch Geometric using the AdamW optimizer. The training configuration is shown in Table 2.

Table 1. Data Split Configuration.

Dataset	Number of Tweets
Train	3,655
Validation	457
Test	457
Total	4,569

Table 2. Training Configuration of GCN Model.

Hyperparameter	Value
Hidden layer size	3,655
GCN Layers	457
Dropout	457
Learning rate	4,569
Weight decay	5×10^{-4}
Similarity threshold (τ)	0.05
Epochs	300

4. Results and Discussion

The evaluation was conducted on the test set comprising 457 tweets, using accuracy and macro F1-score as performance metrics. The proposed GCN model was compared against five baseline classifiers: Naive Bayes (NB), XGBoost, Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR).

Training Performance

Figure 3 presents the training and validation loss curves (top) along with validation metrics (bottom) for the proposed GCN model. Both training and validation losses consistently decrease and stabilize around epoch 100, with no significant overfitting observed. Despite minor fluctuations, the validation F1-score and accuracy converge to high and stable values. The final test performance, indicated by dashed lines, reaches an accuracy of 82.71% and an F1-score of 82.65%, demonstrating the effectiveness of undersampling in producing a robust and generalizable model.

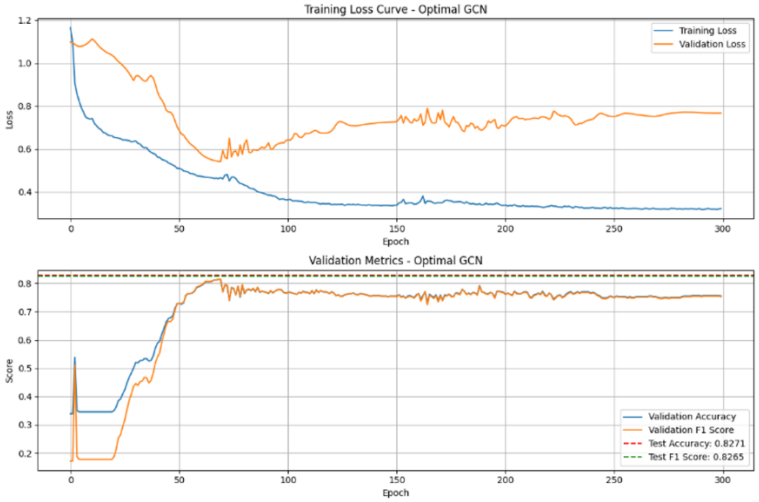


Figure 4. Training and validation loss curves along with final test accuracy and F1-score.

Figure 4 presents the training and validation loss curves (top) along with validation metrics (bottom) for the proposed GCN model. Both training and validation losses consistently decrease and stabilize around epoch 100, with no significant overfitting observed. Despite minor fluctuations, the validation F1-score and accuracy converge to high and stable values. The final test performance, indicated by dashed lines, reaches an accuracy of 82.71% and an F1-score of 82.65%, demonstrating the effectiveness of undersampling in producing a robust and generalizable model.

Confusion Matrix Analysis

Figure 5 shows the confusion matrix of the proposed GCN model. The matrix illustrates the number of correct and incorrect predictions for each sentiment class.

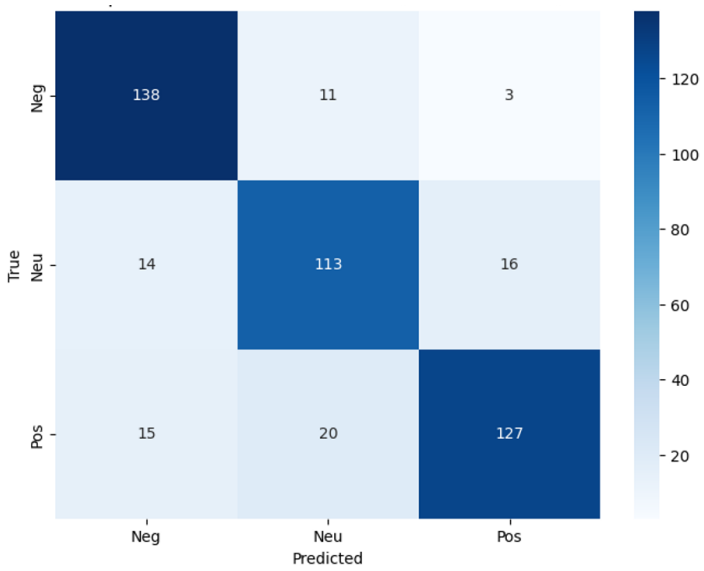


Figure 5. Confusion matrix of GCN model on test set.

Based on the confusion matrix:

a. Negative class: The model correctly predicted 138 negative tweets. However, 11 tweets were incorrectly classified as neutral and 3 as positive;
b. Neutral class: A total of 113 neutral tweets were correctly identified. Misclassifications occurred with 14 tweets predicted as negative and 16 as positive.;
c. Positive class: The model successfully classified 127 positive tweets, while 15 were misclassified as negative and 20 as neutral.

Classification Report:

	precision	recall	f1-score	support
Negative	0.83	0.91	0.87	152
Neutral	0.78	0.79	0.79	143
Positive	0.87	0.78	0.82	162
accuracy			0.83	457
macro avg	0.83	0.83	0.83	457
weighted avg	0.83	0.83	0.83	457

Figure 6. Classification report of GCN model on test set.

The GCN model was especially strong at distinguishing negative sentiments. This is notable because negative tweets often contain more nuanced or sarcastic tones, which are challenging for conventional models.

5. Comparison

Table 3 presents the performance of all models on the test set. The GCN model achieved the highest test accuracy and macro F1-score, outperforming classical machine learning baselines in handling short-text sentiment classification.

Table 3. Model Performance Comparison on Test Set

Model	Accuracy (%)	Macro F1 (%)
Naive Bayes	78	78
XGBoost	78	78
Random Forest	79	79
Logistic Regression	82	82
SVM	82	82
GCN (Proposed)	82.71	82.65

The results show that the GCN model is more effective at capturing relational structure among tweets via graph-based representation. This allowed the model to better generalize in distinguishing between the subtle differences of neutral and negative sentiments, which are often misclassified by traditional models. Compared to the best-performing baseline (SVM and Logistic Regression at 82%), the GCN model achieves an improvement of 0.71% in accuracy and 0.65% in macro F1-score. Although the margin is modest, it is consistent with findings in the literature where graph-based methods show incremental improvements over strong linear baselines on short-text datasets [7].

6. Conclusions

This study proposed a Graph Convolutional Network model with TF-IDF embeddings for sentiment classification of Indonesian tweets regarding the Free Nutritious Meal Program. The model achieved 82.71% accuracy and 82.65% F1-score, outperforming traditional machine learning baselines. The integration of graph-based learning with TF-IDF representations effectively captured both lexical and relational information in short-text social media data.

The use of undersampling to address class imbalance further contributed to a more balanced and generalizable model. The confusion matrix analysis revealed that the model performs particularly well on negative sentiment, which is often the most complex class due to irony and sarcasm in social media language.

Future work will explore attention mechanisms such as Graph Attention Networks (GAT), and pre-trained language models such as IndoBERT, to further improve classification performance on Indonesian-language short texts.

Author Contributions**:** Conceptualization: F.L. and F.L.; Methodology: F.L.; Software: F.L.; Validation: F.L., F.L., and F.L.; Formal analysis: F.L.; Investigation: F.L.; Resources: F.L.; Data curation: F.L.; Writing—original draft preparation: F.L.; Writing—review and editing: F.L.; Visualization: F.L.; Supervision: F.L. All authors have read and agreed to the published version of the manuscript. **Funding:** This research received no external funding. Data Availability Statement: The dataset used in this study consists of publicly available tweets collected from the social media platform X. The processed dataset and implementation code are available upon request from the corresponding author.

Acknowledgments

The authors would like to thank the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for supporting this research under the Independent Campus Program. We also express gratitude to the annotators involved in the manual labeling process and to Diponegoro University for providing research infrastructure and computational resources. Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] UNICEF Indonesia. (2019). *Stunting in Indonesia: A public health crisis*. https://www.unicef.org/indonesia/stunting-crisis

[2] Hermansyah, A., & Hasibuan, H. (2023). Analisis sentimen Twitter terhadap program makan bergizi gratis menggunakan *Random Forest* dan *Linear Regression*. *Jurnal Informatika Indonesia, 12*(3).

[3] Triningsih, Y., & Prasetya, D. (2023). Klasifikasi sentimen pada Twitter tentang program makan gratis menggunakan SVM. In *Proceedings of Seminar Nasional Teknologi Informasi dan Komputer (SENTIKA)*.

[4] Elissa, K. (n.d.). *Deep learning sentiment classification for short texts in Bahasa Indonesia* (Unpublished manuscript).

[5] Nicole, R. (n.d.). Improved representation learning using GCN with TF-IDF integration. *IEEE Transactions on Knowledge and Data Engineering*. Advance online publication.

[6] Zhou, J., Cui, G., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open, 1*, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

[7] Khemani, R., Banerjee, S., & Ahmad, F. (2024). Enhanced GCN for short-text sentiment classification using TF-IDF embedding. *International Journal of Data Science, 9*(1).

[8] Abdullah, M., et al. (2025). An integration of corpus applied linguistics with deep learning for Arabic spam and ham tweets detection model. *Journal of Radiation Research and Applied Sciences, 18*(2), 101463. https://doi.org/10.1016/j.jrras.2025.101463

[9] Aqiilah, et al. (2026). Sentence-level sentiment analysis of Indonesian app reviews using IndoBERTweet. *Journal of Computing Theories and Applications, 4*(1). https://doi.org/10.62411/jcta.16240

[10] Darnoto, et al. (2026). Language-similarity-guided transfer fine-tuning of pre-trained transformer models for sentiment analysis across 12 Indonesian regional languages. *Journal of Computing Theories and Applications, 3*(4). https://doi.org/10.62411/jcta.15975

[11] Krisnawan, et al. (2022). Analisis sentimen program makan bergizi gratis menggunakan *bidirectional gated recurrent unit*. *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi, 4*(3). https://doi.org/10.24246/itexplore.v4i3.2025.pp282-294

[12] Risdiansyah, et al. (2026). Aspect-based sentiment analysis (ABSA) pada aplikasi JMO menggunakan IndoBERT, *neural network*, dan SMOTE. *Teknik: Jurnal Ilmu Teknik dan Informatika, 6*(1). https://doi.org/10.51903/teknik.v6i1.1259

[13] Srikandi Alifya, et al. (2025). Penerapan NLP menggunakan algoritma *Naive Bayes*, C4.5, XGBoost untuk analisis sentimen ulasan produk kecantikan di Tokopedia dan Shopee. *Prosiding Seminar Nasional Ilmu Teknik, 2*(2). https://doi.org/10.61132/prosemnasproit.v2i2.71

[14] Ifadah, M., et al. (2026). Analisis kinerja metode *long short-term memory* (LSTM) dalam klasifikasi sentimen ulasan pengguna Shopee. *Jurnal Elektronika dan Komputer, 18*(2). https://doi.org/10.51903/elkom.v18i2.3407

[15] Muriyatmoko, D., et al. (2026). Sentiment analysis of YouTube comments on Indonesia's performance in 2026 World Cup qualifiers using *Naive Bayes*. *Prosiding Seminar Nasional Ilmu Teknik, 2*(2). https://doi.org/10.61132/prosemnasproit.v2i2.196