



Sentiment Analysis of Public Response to the Free Nutritious Meal Program on Instagram Using IndoBERT

Fikry Destryanto^{1*}, Putri Rizqiyah², Petrus Sokibi³

¹Informatic engineering, Faculty of Information Technology, Universitas Catur Insan Cendekia, Cirebon, Indonesia
fikry.destryanto.ti.21@cic.ac.id^{1*}, putri.rizqiyah@cic.ac.id², petrus.sokibi@cic.ac.id³

Abstract

The Free Nutritious Meal (MBG) Program, as one of the government's social initiatives, has attracted diverse discussions on social media, particularly Instagram. This study aims to analyze public sentiment toward the program by applying a machine learning approach based on IndoBERT. The research data were collected from user comments on Instagram posts related to the MBG program. The methodological framework includes data collection, text preprocessing, sentiment labeling, as well as the training and evaluation of the IndoBERT model. The results indicate that IndoBERT successfully classifies public sentiment into three categories: positive, negative, and neutral, with a high level of accuracy. Furthermore, the analysis shows that most public responses reflect positive sentiment, although neutral opinions and criticism are also present. In conclusion, IndoBERT is effective in capturing patterns of public opinion on social media and provides valuable insights for policymakers in evaluating and improving the implementation of the MBG program.

Keywords: *IndoBERT, Instagram, MBG Program, Sentiment Analysis, Social Media*

1. Introduction

The increasing use of social media has significantly transformed how people communicate, share their views, and engage with public policies. Instagram, for instance, is not only a platform for sharing daily moments but also a space where discussions about government initiatives and programs take place. Understanding public opinion expressed on these platforms offers valuable insights that can be used to evaluate the level of acceptance and the real-time impact of policies [1]

One of the government initiatives that has recently drawn considerable public attention is the Free Nutritious Meal (MBG) Program. As a policy closely tied to fundamental aspects of social welfare, it has elicited a wide range of public reactions, from strong support to critical assessments. Analyzing these sentiments is crucial, as the insights gained can assist policymakers in refining implementation strategies while also strengthening public trust [2]

Sentiment analysis, which seeks to categorize opinions expressed in text into positive, negative, or neutral classes, has become a widely applied method in the field of natural language processing (NLP). The rapid progress of machine learning, particularly with the advent of transformer-based language models, has greatly enhanced the performance and accuracy of sentiment classification [3]. In the Indonesian context, IndoBERT has emerged as one of the most reliable models for processing and analyzing Indonesian-language text, owing to its strong contextual representation and pretraining on extensive language corpora [4].

This study focuses on analyzing public sentiment toward the MBG program by utilizing the IndoBERT model on Instagram comments. The primary objectives are (i) to examine the distribution of opinions across positive, negative, and neutral categories, and (ii) to assess the capability of IndoBERT in effectively capturing sentiment expressed in the Indonesian language on social media platforms. The outcomes of this research are expected to provide contributions both to academic advancements in sentiment analysis and to practical recommendations for policymakers in enhancing the implementation of the MBG program [5].

2. Methodology

This research employs a sentiment analysis framework with the IndoBERT model to examine public opinions regarding the Free Nutritious Meal (MBG) Program on Instagram. The methodological approach comprises several key stages, including data collection, text preprocessing, annotation, model training, and performance evaluation.

2.1. Data Collection

The dataset used in this study was derived from Instagram comments on posts associated with the MBG program. Data collection was conducted using an Instagram comment scraping tool within a defined observation period. The final dataset was refined to include only comments that explicitly referred to or discussed the MBG program, thereby ensuring its relevance for sentiment analysis [1].

2.2. Data Preprocessing

Prior to the analysis, the text data were subjected to several preprocessing procedures, including case folding, punctuation removal, stopword elimination, and normalization of informal expressions. These steps were carried out to enhance the quality of the input data and to minimize the noise that is typically present in social media text [2].

2.3. Sentimen Anotation

The comments collected from Instagram were manually annotated into three sentiment categories, namely positive, negative, and neutral. This annotation process was established as the ground truth, forming the basis for both model training and subsequent evaluation. In addition, careful attention was given to the distribution of labels to ensure the dataset remained balanced and representative.

2.4. Model Training IndoBERT

IndoBERT, which is a pre-trained transformer-based model designed specifically for the Indonesian language, was utilized to perform the sentiment classification task. The model was fine-tuned with the annotated dataset in order to adapt its performance to the domain of MBG-related discussions on Instagram. The classification process produced a three-class output that represented the sentiment categories of positive, negative, and neutral [4].

2.5. Evaluation

The performance of the IndoBERT model was assessed through commonly applied metrics in sentiment analysis, namely accuracy, precision, recall, and F1-score. The use of these metrics offered a comprehensive assessment of the model's capability in accurately classifying sentiments within the dataset [5].

3. Research Method

This research applied a sentiment analysis approach utilizing the IndoBERT model to classify Instagram comments related to the Free Nutritious Meal (MBG) program. The methodological framework comprises several key stages, as presented in Figure 1.

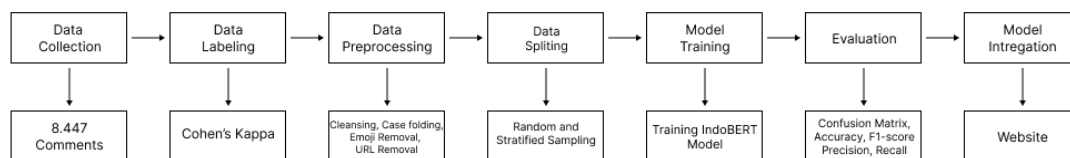


Fig. 1: Reaserch Framework

The research framework, as depicted in Figure 1, is structured into several stages, namely data collection, labeling, preprocessing, data splitting, model training, and evaluation. The process begins with the collection of Instagram comments related to the MBG program, which are then manually labeled based on sentiment categories and refined through text preprocessing. The prepared dataset is subsequently divided into training and testing sets, forming the basis for fine-tuning the IndoBERT model. In the final stage, the model's performance is systematically evaluated using standard classification metrics to assess its effectiveness in sentiment analysis.

3.1. Data Collection

The data for this study were obtained using the IG Comment Exporter tool, which facilitated the extraction of Instagram comments into CSV format. In total, 8,447 comments were collected from 95 Instagram posts related to the MBG program within the observation period spanning from January 6 to May 5, 2025. Following a filtering process to remove irrelevant and empty entries, a refined dataset consisting of 3,721 comments was retained for the annotation stage.

3.2. Data Collection

The annotation process was carried out manually to establish ground truth labels for model training. Two annotators were tasked with assigning each comment to one of three sentiment categories: positive, neutral, or negative. To ensure consistency, both annotators adhered to predefined labeling guidelines, and any discrepancies were resolved through discussion. Out of 3,721 comments that passed preprocessing, 3,136 were successfully annotated and deemed valid for further analysis. The reliability of the annotations was evaluated using Cohen's Kappa, yielding a score of 0.71, which reflects a substantial level of agreement between the two annotators.

3.3. Data Preprocessing

The preprocessing stage was carried out to improve the quality of text data before model training. This process included case folding to standardize text into lowercase, noise removal to eliminate punctuation, emojis, URLs, and extra spaces, stopword removal to filter out non-informative words, and text normalization to replace informal expressions with standard forms.

3.4. Data Splitting

The annotated dataset was partitioned into training and testing sets through a stratified random sampling technique, applying a 70:30 ratio to maintain proportional representation across sentiment classes. To further address potential class imbalance, an undersampling strategy was employed, resulting in a more balanced distribution of the dataset and ensuring fairness in the training process.

3.5. Model Training

The IndoBERT model was fine-tuned using a stratified sampling strategy with a learning rate of $2e-5$ and six epochs, which yielded the best performance among the experimental configurations. The training process was conducted on Google Colaboratory with GPU acceleration using PyTorch, where the Adam optimizer and categorical cross-entropy loss function were applied to optimize the model.

3.6. Evaluation

The final stage of this research involved integrating the trained IndoBERT model into a web-based application. This integration was carried out by connecting the model with a web interface, enabling users to interact directly with the system. The application was designed to accept user input in the form of comments, process them through the IndoBERT model, and present the classification results into three sentiment categories: positive, neutral, and negative. Through this step, the study not only produced a sentiment analysis model but also provided a practical implementation that can be utilized to assess public opinion regarding the Free Nutritious Meal (MBG) Program.

3.7. Integration

The final stage of this study involved integrating the trained IndoBERT model into a web-based application. This integration connected the model with a user-friendly web interface, enabling direct interaction with the system. The application was designed to accept user input in the form of comments, process them through IndoBERT, and present the classification results into three sentiment categories: positive, neutral, and negative. Through this step, the research not only produced a sentiment analysis model but also delivered a practical implementation that can be utilized to assess public opinion on the Free Nutritious Meal Program.

4. Result and Discussion

This section outlines the performance of the IndoBERT model in sentiment analysis, along with the distribution of sentiments reflected in Instagram comments concerning the MBG program.

4.1. Data Preprocessing Result

Following the preprocessing stage, the dataset was reduced from 8,447 raw comments to 3,721 cleaned entries after eliminating irrelevant, duplicate, and empty data. Of these, 3,136 comments were successfully annotated and retained as valid data for subsequent analysis.

Table 1: Data Reduction during Preprocessing

Stage	Number of Content
Raw Data	8.447
After Cleaning	3.721
After Anotation	3.136

4.2. Sentimen Distribution

Of the 3,136 annotated comments, 40% were identified as positive, 33.3% as neutral, and 26.7% as negative. These results suggest that the majority of users conveyed supportive views toward the MBG program, although neutral perspectives and critical remarks were also evident.

Table 2: Sentiment Distribution of Annotated Data

Sentimen	Number of Comments	Number of Content
Positive	1.254	40.0%
Neutral	1.0d3	33.3%
Negative	839	26.7%
Total	3.136	100%

The findings reveal that the majority of comments conveyed positive sentiment (40%), followed by neutral (33.3%) and negative (26.7%). This suggests that most Instagram users responded favorably to the MBG program, while a significant share of responses were either neutral or critical. Negative remarks primarily raised concerns regarding the program's implementation and long-term sustainability, whereas neutral comments were generally descriptive or factual in nature, lacking strong emotional evaluation.

4.3. Model Performance

The performance of the IndoBERT model was assessed using standard evaluation metrics, namely accuracy, precision, recall, and F1-score. A summary of the results for both the training and testing datasets is presented in Table 3.

Table 3: Evaluation Metrics of IndoBERT

Dataset	Accuracy	Precision	Recall	F1-score
Training	0.9989	0.9989	0.9989	0.9989
Testing	0.9395	0.9396	0.9395	0.9393

The findings reveal that IndoBERT achieved near-perfect performance on the training set, with accuracy approaching 100%. On the testing set, the model sustained strong performance, recording accuracy above 93%. These results demonstrate the model's ability to generalize effectively to unseen data while avoiding overfitting.

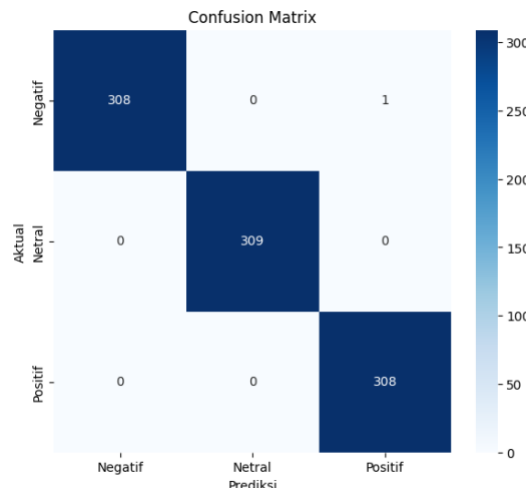


Fig. 2: Confusion Matrix on Training Data

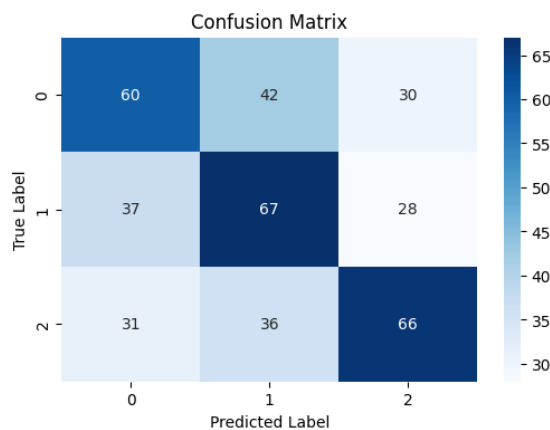


Fig. 3: Confusion Matrix on Testing Data

5. Conclusion

This study implemented IndoBERT to analyze public sentiment toward the Free Nutritious Meal (MBG) program based on Instagram comments. Through preprocessing, the dataset was refined from 8,447 raw comments to 3,136 valid annotated entries. The sentiment distribution revealed that 40% of the comments were positive, 33.3% were neutral, and 26.7% were negative, suggesting that most users responded favorably to the program while a portion remained neutral or expressed concerns. In terms of performance, IndoBERT achieved an accuracy of 99.89% on the training set and 93.95% on the testing set, with precision, recall, and F1-scores demonstrating consistent reliability. The confusion matrix further confirmed that only minor misclassifications occurred, particularly between neutral and negative classes. These findings highlight IndoBERT's effectiveness in sentiment classification for the Indonesian language, particularly in processing informal social media text. Future work could explore the use of larger and more diverse datasets, as well as alternative transformer-based architectures, to enhance performance and strengthen the role of sentiment analysis in supporting government decision-making and policy evaluation.

Reference

- [1] A. Dwijayanti, "Policy Spillover: Analisis Jaringan Dampak Kebijakan Makan Siang Bergizi Gratis terhadap Sektor Pertanian," *Jurnal Administrasi Publik*, vol. 20, no. 2, hlm. 281–308, Des 2024, doi: 10.52316/jap.v20i2.435.
- [2] F. V. Sari dan A. Wibowo, "ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI," *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.
- [3] D. Alita dan A. Rahman, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," 2020.
- [4] M. Hamka dan D. Ratna Sari, "ANALISIS SENTIMEN DAN INFORMATION EXTRACTION PEMBELAJARAN DARING MENGGUNAKAN PENDEKATAN LEXICON," 2022.
- [5] P. Rizqiyah, E. Yulianti, dan M. A. Jiwangi, "Analyzing public perception toward COVID-19 vaccines in Indonesia," *Int J Publ Health Sci*, vol. 13, no. 1, hlm. 428–437, Mar 2024, doi: 10.11591/ijphs.v13i1.23134