Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

Analysis of Static and Contextual Word Embeddings in Capsule Network for Sentiment Analysis of The Free Nutritious Meal Program on Twitter

Virgi Atha Raditya[1], Triando Hamonangan Saragih*[2], Mohammad Reza Faisal[3], Friska Abadi[4], Muliadi[5]

[1,2,3,4,5]Department of Computer Science, Lambung Mangkurat University, Banjarbaru, Indonesia

Email: [2]triando.saragih@ulm.ac.id

Received : Nov 4, 2025; Revised : Nov 14, 2025; Accepted : Nov 17, 2025; Published : Feb 15, 2026

Abstract

Public discourse surrounding Indonesia's Makan Bergizi Gratis (MBG) program reflects diverse opinions that have not yet been systematically examined using computational methods. This study addresses that gap by evaluating the effectiveness of static and contextual word embeddings within a Capsule Network (CapsNet) framework for sentiment analysis of MBG-related tweets on Twitter. A total of 7,133 Indonesian-language tweets were collected through web crawling, preprocessed, and manually labeled into positive, neutral, and negative categories. Four embedding techniques—Word2Vec, FastText, ELMo, and IndoBERT—were tested under two preprocessing settings, raw and stemming. The experimental results show that Word2Vec on raw text achieved the highest accuracy of 96.17%, while FastText obtained the best performance on stemmed data with 94.10%. These findings indicate that morphological normalization benefits static and subword-based embeddings, whereas contextual models maintain stable performance without extensive fine-tuning. Overall, this study demonstrates the potential of combining CapsNet with appropriate embedding strategies for Indonesian-language sentiment analysis and provides evidence that natural language processing can support data-driven evaluation of public programs such as MBG.

Keywords : *Capsule Network, Contextual Embeddings, Makan Bergizi Gratis, Sentiment Analysis, Static Embeddings, Twitter Data*

1. INTRODUCTION

The rapid growth of internet penetration in Indonesia has significantly transformed how people communicate and exchange information. As of early 2024, internet penetration reached 79.5% of the total population [1], illustrating that digital connectivity has become an integral part of daily life, particularly among younger generations. This shift is most evident among Millennials and Generation Z, whose social interactions have been reshaped by digital communication, forming new cultural patterns in information sharing [2]. However, the rapid increase in internet use also presents serious challenges, including greater exposure to disinformation and hoaxes. Therefore, enhancing digital literacy has become an urgent necessity to ensure that users can critically evaluate the authenticity of online information [3][4].

One of the most widely discussed government initiatives in this digital era is the Makan Bergizi Gratis (MBG) or Free Nutritious Meal program. First introduced by Prabowo Subianto in 2006 as a response to Indonesia's high stunting rate [5], the program was finally implemented after his inauguration as the 8th President of Indonesia. With a national budget allocation of Rp71 trillion in the 2025 State Budget Plan (RAPBN 2025) Rp63.356 trillion for nutritional support and Rp7.433 trillion for management [6] MBG represents a large-scale policy with significant socio-economic implications. Recent studies highlight the socio-political dynamics of MBG implementation, showing both strong public support and critical debate on its long-term sustainability [28]. Given its national importance,

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

analyzing public sentiment toward MBG provides valuable, data-driven insight for policy evaluation and communication strategies.

In this context, social media platforms play a key role in capturing public discourse. Twitter (now known as X) remains one of the most influential platforms for expressing opinions and engaging with public figures [9][10][11]. Its short-message format, real-time engagement, and open data ecosystem make it particularly suitable for large-scale sentiment analysis. Recent works have demonstrated how Twitter sentiment reflects socio-political attitudes in Indonesia, providing a real-time pulse of public opinion [29]. With its vast and dynamic data stream, Twitter serves as a valuable resource for understanding how citizens perceive government initiatives like MBG.

Sentiment analysis is a Natural Language Processing (NLP) technique used to identify, extract, and classify opinions or emotions within text [14][15][16]. It enables researchers to determine sentiment polarity positive, negative, or neutral toward an entity such as a product, service, or policy [17]. Earlier studies commonly applied traditional algorithms like Naive Bayes and Support Vector Machine (SVM) for text classification [18]. However, with advancements in deep learning, models such as Bidirectional Long Short-Term Memory (BiLSTM) and Capsule Network (CapsNet) have shown superior performance in capturing complex linguistic contexts [19][20].

Equally important is the text representation technique, or word embedding, which transforms textual data into numerical vectors. Early approaches such as Word2Vec and FastText provided static embeddings, assigning a single vector to each word regardless of context. In contrast, contextual embeddings like ELMo and BERT generate dynamic representations that vary according to sentence structure [21][22]. The emergence of IndoBERT, pretrained on over 220 million Indonesian tokens, offers a more linguistically aligned contextual model for Indonesian text [30]. Nevertheless, few studies have investigated its effectiveness in policy-related sentiment analysis, which involves unique socio-linguistic nuances and domain-specific vocabulary.

Several previous works have explored similar architectures in other contexts. Diviya Prabha and Rathipriya (2022) used CapsNet to analyze COVID-19 tweets, achieving 96% accuracy [23]. Dong et al. (2020) integrated CapsNet with BiLSTM and GloVe, reaching 98.8% accuracy on the IMDB dataset [24]. Toqeer Ehsan et al. (2023) applied contextual embeddings (ELMo) for low-resource languages like Tamil and Tulu, achieving modest results (F1-score 0.51) [25]. Haziq et al. (2024) combined LSTM with FastText for app review sentiment, reaching 89.11% accuracy [26], while Noryasminda et al. (2025) achieved 89.8% using BiLSTM + FastText on monkeypox tweets [27]. Although these studies demonstrate the potential of CapsNet and contextual embeddings, none specifically address Indonesian-language sentiment analysis in socio-political contexts such as MBG.

Based on these gaps, this study aims to analyze public sentiment toward Indonesia's Makan Bergizi Gratis (MBG) program using data from Twitter. The research integrates Capsule Network (CapsNet) with both static (Word2Vec, FastText) and contextual (ELMo, IndoBERT) embeddings to evaluate their effectiveness in sentiment classification. The novelty of this study lies in its application of IndoBERT and CapsNet within an Indonesian-language, policy-oriented domain providing both methodological and practical contributions. Methodologically, it compares the effectiveness of embedding types for sentiment classification in a morphologically rich language. Practically, it offers informatics-based insights for policymakers to understand public perception and enhance data-driven decision-making.

2. METHOD

All research processes in this study were conducted using a laptop equipped with an Apple Macbook M2 processor and 8 GB of RAM as the primary development environment. To meet the computational demands required for deep learning model training, the study utilized Google

Colaboratory, a cloud-based Python platform that provides free GPU support. This platform was selected due to its flexibility, seamless integration with popular deep learning libraries, and its ability to accelerate experiments involving complex architectures such as the Capsule Network (CapsNet). The implementation was carried out using the Python programming language, supported by several essential libraries, including scikit-learn, matplotlib, seaborn, gensim, fasttext-wheel, and TensorFlow. Through this computational setup, the research implemented and compared multiple word embedding techniques namely Word2Vec, FastText, BPEmb (ELMo), and IndoBERT which were subsequently integrated into the CapsNet model for sentiment classification.

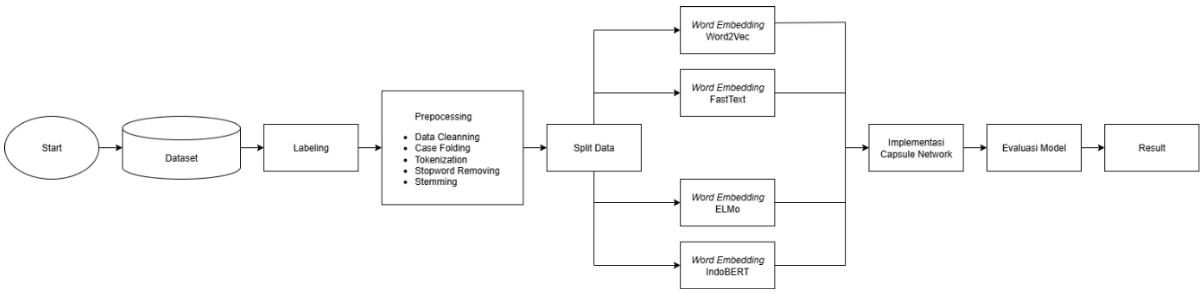

Figure 1. Research Method

2.1. Dataset

The dataset employed in this study was obtained through a web-crawling procedure using the Tweet-Harvest tool, which enables the extraction of public tweets without requiring access to the official Twitter API. The query "Makan Bergizi Gratis OR MBG lang:id" was applied under the Latest tab to ensure that only the most recently posted Indonesian-language tweets were collected. Data acquisition was conducted across five crawling sessions, each yielding approximately 1,500 tweets, resulting in an initial pool of around 7,500 entries. After removing duplicates and non-valid records, a total of 7,133 unique tweets were retained.

The cleaned dataset was stored in CSV format as mbg_dataset.csv and consisted of 15 attributes, including conversation_id_str, created_at, tweet, favorite_count, retweet_count, reply_count, lang, location, and username. All retrieved tweets were written in Bahasa Indonesia and represented a diverse range of public reactions toward the Makan Bergizi Gratis (MBG) program, forming the basis for subsequent exploratory data analysis..

Table 1. Data Collection Parameter

Parameter	Value
Keyword	"makan bergizi gratis" OR "mbg" lang:id
Target Number of Tweets	1500
Search Tab	LATEST
Tool	tweet-harvest versi 2.6.1
Token	Twitter Guest Token

2.2. Data Labeling

All labeling processes were conducted manually on the dataset after completing the cleaning stage. Each tweet was annotated with one of three sentiment labels positive, negative, or neutral based on the contextual meaning of the sentence. In this study, 100% of the data were manually labeled, ensuring higher classification accuracy and enabling the model to better capture the underlying characteristics of each sentiment category [28].

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

2.3. Preprocessing

The preprocessing stage was carried out to ensure that the data were clean, consistent, and suitable for analysis. This process included several steps such as cleaning unwanted elements (URLs, mentions, hashtags, and special characters), converting all text to lowercase to achieve uniformity, tokenizing sentences into individual words, reducing words to their base forms through stemming, and removing common yet semantically insignificant words known as stopwords. These preprocessing steps were essential to eliminate noise and retain meaningful textual components, allowing the subsequent feature extraction and modeling stages to perform effectively [29].

2.4. Word Embedding

Word embedding is a representation technique that transforms words into fixed-dimensional numerical vectors, designed to capture the semantic meaning of words based on their contextual use within sentences. Each word is encoded as a dense vector, allowing semantic relationships among words to be measured mathematically through similarity operations such as cosine similarity [30]. Word embeddings are generally categorized into two main types: static and contextual. Static embeddings, such as Word2Vec and FastText, generate a single, context-independent vector representation for each word, effectively capturing general semantic relationships but struggling to differentiate polysemous words. In contrast, contextual embeddings such as ELMo and BERT produce dynamic vector representations that vary depending on sentence context, enabling a more precise understanding of word meaning [31][32][33]. By employing both static and contextual embeddings, this study aims to compare their effectiveness in improving sentiment classification performance using the Capsule Network (CapsNet) model.

2.4.1. FastText

FastText is an extension of the Word2Vec library that adopts the Skip-Gram with Negative Sampling (SGNS) approach and introduces a word representation model based on character-level n-grams [17]. For instance, the word “sukses” is represented as the sum or average of the vectors of all its character n-grams:

<s, <su, <suk, <suks, suk, suks, sukse, sukses, uks, ukses, ks, kses, se, ses, es, s>.

Here, the symbols ‘<’ and ‘>’ denote the start and end boundaries of a word during the n-gram construction process [32]. By incorporating subword information, FastText is able to better capture morphological variations and semantic relationships within words, particularly in morphologically rich languages such as Indonesian.

2.4.2. Word2Vec

Word2Vec is a neural network–based word embedding algorithm that learns word representations by training a shallow network with a single hidden layer and a fully connected output layer. The weights obtained during the training process in the hidden layer are then used to form dense word vectors that capture semantic and syntactic relationships among words [32]. Word2Vec operates using either the Continuous Bag-of-Words (CBOW) or the Skip-Gram architecture, where CBOW predicts a target word based on its surrounding context, while Skip-Gram predicts surrounding words given a target word. Through this mechanism, Word2Vec effectively captures co-occurrence patterns in large corpora, enabling semantically similar words to be represented with closely related vectors in the embedding space.

Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 1, February 2026, Page. 633-646
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

2.4.3. ELMo

Embeddings from Language Models (ELMo) is a contextual word representation model developed by the Allen Institute for AI, designed to capture the meaning of a word based on its complete sentence context [33]. Unlike static embeddings, ELMo generates different vector representations for the same word depending on its usage within a sentence. For example, in the following sentences:

i. *Kepala sekolah sedang memimpin rapat guru.*
ii. *Ia menundukkan kepala karena malu.*

the word "kepala" carries different meanings. In the first sentence, it refers to a leader (the head of a school), while in the second, it denotes a part of the human body. ELMo captures such polysemy by leveraging a bidirectional LSTM architecture trained on a large text corpus, allowing it to encode both forward and backward contextual dependencies. This capability enables ELMo to provide rich, dynamic embeddings that reflect nuanced word meanings depending on their linguistic context.

2.4.4. IndoBERT

IndoBERT is a variant of the Bidirectional Encoder Representations from Transformers (BERT) model that has been pre-trained on a large-scale Indonesian corpus consisting of more than 220 million words drawn from multiple sources, including Indonesian Wikipedia, news outlets such as Kompas, Tempo, and Liputan6, as well as the Indonesian Web Corpus [34]. By being trained exclusively on Indonesian-language data, IndoBERT achieves a deeper understanding of the morphological structure, syntax, and contextual semantics of the Indonesian language compared to the multilingual BERT model. This allows IndoBERT to generate high-quality contextual representations for Indonesian text, making it particularly effective for natural language processing tasks such as sentiment analysis, named entity recognition, and text classification.

2.5. Capsule Network

The Capsule Network (CapsNet) is a deep neural architecture developed to overcome the limitations of conventional Convolutional Neural Networks (CNNs), particularly in capturing spatial and hierarchical relationships among features. Unlike CNNs, which tend to lose spatial information through pooling operations, CapsNet groups neurons into small clusters called capsules, each capable of representing not only the presence of a feature but also its pose information, including orientation, position, and scale [35].

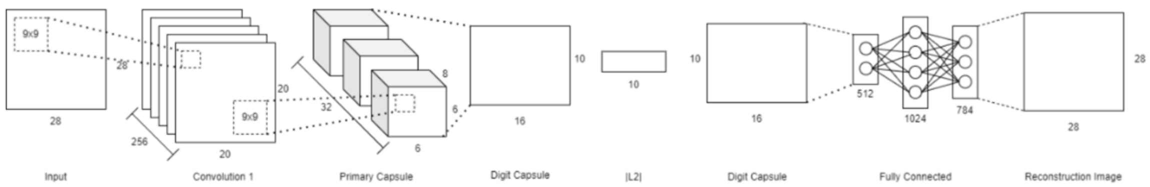

Figure 2. Capsule Network Architecture

The CapsNet architecture consists of two main components: an encoder and a decoder. The encoder is responsible for extracting and encoding textual features through a series of convolutional layers, primary capsules, and digit capsules, which transform low-level features into higher-level representations. The decoder, on the other hand, reconstructs the original input from the capsule outputs to reinforce the model's ability to learn spatial dependencies among features [36]. In this study, CapsNet was chosen for its capability to maintain inter-word and contextual relationships within sentences without requiring extensive data augmentation. However, a key limitation of CapsNet lies in its high

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

computational complexity, as the dynamic routing process used to determine how information flows between capsules requires multiple iterations and involves a large number of parameters, resulting in longer training times and higher computational costs [37].

2.6. Confusion Matrix

Confusion Matrix		**Prediksi**		
		Positive	*Neutral*	*Negative*
Aktual	*Positive*	*True Positive* (TP)	*False Neutral* (FNeu)	*False Negative* (FNeg)
	Neutral	*False Positive* (FP)	*True Neutral* (TNeu)	*False Negative* (FNeg)
	Negative	*False Positive* (FP)	*False Neutral* (FNeu)	*True Negative* (TNeg)

Figure 3. Confusion Matrix

The confusion matrix used in this study categorizes sentiment into three classes: positive, neutral, and negative. In this context, True Positive (TP) represents the number of data instances that genuinely belong to the positive category and are correctly classified by the model. True Neutral (TNeu) refers to data that are actually neutral and correctly identified as such, while True Negative (TNeg) indicates the number of negative data points that are accurately classified. Conversely, False Positive (FP) denotes the number of non-positive data instances (either neutral or negative) that are incorrectly predicted as positive by the system. False Neutral (FNeu) corresponds to data from the positive or negative categories that are misclassified as neutral, and False Negative (FNeg) represents data from the positive or neutral categories that are incorrectly predicted as negative. This classification framework provides a more comprehensive evaluation of model performance by revealing specific areas where misclassification occurs, thereby facilitating targeted improvements to enhance overall accuracy. The accuracy, precision, and recall metrics are calculated using the following standard formulas:

$$Total = TP + FNeu + FNeg + FP + TNeu + FNeg + FP + FNeu + TNeg \quad (1)$$

$$Accuracy = \frac{\mathrm{TP+TNeu+TNeg}}{Total} \quad (2)$$

$$Precision = \frac{\mathrm{TP}}{TP+FP+FP} \quad (3)$$

$$Recall = \frac{\mathrm{TP}}{TP+FNeu+FNeg} \quad (4)$$

2.7. Model Training Configuration

To improve model generalization and evaluate robustness, a randomized hyperparameter search was conducted over 10 independent trials (N = 10) using a fixed random seed (SEED = 42). Each trial used a random combination of key parameters within controlled ranges. The configuration space was defined as follows:

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

Table 2. Parameter Configuration

Parameter	Value
MAX_LEN	128
VAL_SIZE	0.10
TEST_SIZE	0.10
Learning Rate (Lr)	10U (0.10,0.30)
Dropout Rate	U (0.10,0.30)
Batch Size (Bs)	16,24,32
Epochs	10,25
Routing Iteration	2,3

Each trial trained the Capsule Network (CapsNet) model using the Adam optimizer with categorical cross-entropy loss. The best-performing configuration was selected based on validation F1-score. All experiments were implemented in TensorFlow 2.17 and executed on Google Colaboratory (GPU runtime) for computational efficiency. This randomized search approach allows a broader exploration of the parameter space while maintaining reproducibility through a consistent seed value.

3. RESULT

3.1. Exploratory Data Analysis (EDA)

The dataset used in this study was collected through a web crawling process using the tweet-harvest tool, which extracts public tweets without relying on the official Twitter API. The query "Makan Bergizi Gratis OR MBG lang:id" was applied in the “Latest” tab to ensure chronological recency of collected tweets. Data acquisition was performed in five sessions, each retrieving approximately 1,500 tweets, resulting in a total of 7,500 tweets, from which 7,133 unique entries were retained after cleaning.

The final dataset was saved in CSV format as mbg_dataset.csv and contained 15 attributes, including conversation_id_str, created_at, tweet, favorite_count, retweet_count, reply_count, lang, location, and username. All tweets were written in Bahasa Indonesia and reflected various public reactions to the Makan Bergizi Gratis (MBG) program.

Each tweet was manually labeled into one of three sentiment categories: positive, neutral, or negative. As shown in Table 4, positive tweets dominate the dataset (34.38%), followed by neutral (34.68%) and negative (30.63%). This distribution suggests that public opinion on the MBG program tends to be favorable, although the class imbalance may influence the model’s recall performance for minority sentiments such as negative opinions.

Table 3. The dataset has been labeled

No	created_at	full_text	...	Label_sentimen
1	Sat Jun 21 18:19:40 +0000 2025	Dalam satu piring sehat terkandung kekuatan untuk hidup dan belajar. Hari ini mari sukseskan MBG demi generasi unggul Indonesia		positif
2	Sat Jun 21 18:11:44 +0000 2025	Tebak negara mana yang mau belajar soal MBG dari Indonesia wkwk halu banget		negatif
...	...	...	...	...
7268	Sat Jun 21 18:11:18 +0000 2025	Kebijakan MBG tidak pernah menyetujui snack manis atau makanan mentah Lanjutkan MBG.	...	netral

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

Table 4. Distribution of Sentiment Labels

Sentiment	Count	Percentage (%)
Positive	2,474	34.68 %
Neutral	2,474	34.68 %
Negative	2,185	30.63 %
Total	7,133	100 %

3.2. Model Training and Evaluation Setup

Each embedding configuration Word2Vec, FastText, ELMo, and IndoBERT was trained and tested ten times using identical hyperparameters to ensure result stability. The final values represent the best-performing accuracy and mean results across ten runs. The standard deviation remained below 0.8%, confirming model consistency.

3.3. Performance Comparison of Embeddings

Based on the stemmed dataset (Table 5), FastText achieved the highest accuracy at 94.10%, followed by ELMo and IndoBERT (93.67%), while Word2Vec produced 93.48%. On the raw dataset (Table 6), Word2Vec achieved the best performance with 96.17%, followed by FastText (95.25%), ELMo (94.10%), and IndoBERT (93.67%). These results show that Word2Vec performs substantially better on raw text, while FastText benefits more from stemming, likely due to its subword-based architecture.

Table 5. Best parameters for steming data and accuracy results.

Embedding	Learnign Rate	Dropout	Batch _size	Epoch	Routings	Performance Metrics			
						Acuraccy	Precision	Recall	F1-Score
FastText	**0.001070**	**0.100226**	**128**	**37**	**3**	**94.10 %**	**94.11 %**	**94.10 %**	**94.10 %**
Word2Vec	0.001803	0.385502	32	20	2	93.48 %	93.57 %	93.48 %	93.51 %
IndoBERT	0.000023	0.290142	32	16	2	93.67 %	93.75 %	93.75 %	93.67 %
ELMo	0.000270	0.187368	48	17	2	93.67 %	93.67 %	93.67 %	93.66 %

Table 6. Best parameters for raw data and accuracy results.

Embedding	Learnign Rate	Dropout	Batch_size	Epoch	Routings	Performance Metrics			
						Acuracy	Precision	Recall	F1-Score
FastText	0.001465	0.222965	32	31	3	95.25%	95.26%	95.25 %	95.26 %
Word2Vec	**0.000117**	**0.212085**	**48**	**26**	**3**	**96.17%**	**96.29 %**	**96.17 %**	**93.18 %**
IndoBERT	0.000053	0.219731	24	12	2	93.67%	93.75%	93.75 %	93.67 %
ELMo	0.000143	0.146798	32	21	2	94.10%	94.10%	94.10 %	94.08 %

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

Figure 4 and 5 illustrates the comparative accuracy across embeddings. Word2Vec shows consistent superiority on the stemmed dataset, whereas FastText benefits more from morphological normalization.

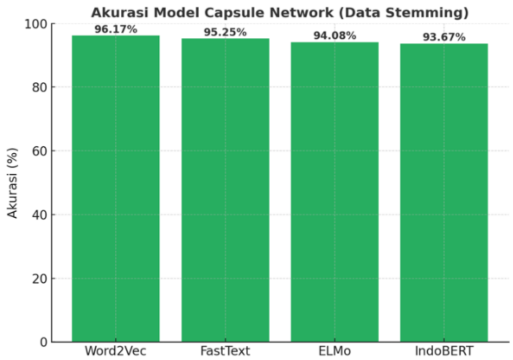

Figure 4. Comparison of Model Accuracy on Stemming Data

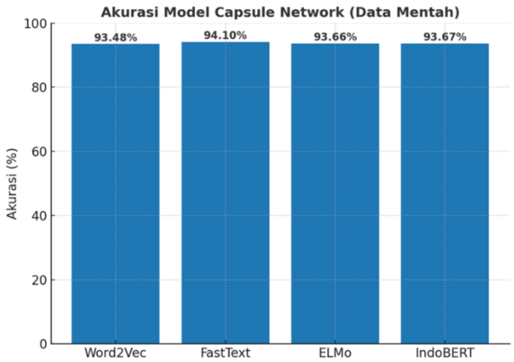

Figure 5. Comparison of Model Accuracy on Raw Data

3.4. Impact of Stemming on Model Performance

f The results demonstrate that stemming has different effects on static and contextual embeddings. On the stemmed dataset, FastText achieved the best performance with an accuracy of 94.10%, showing a slight improvement compared to its raw-text performance (95.25%). This improvement is expected because FastText relies on subword representations; reducing morphological variation through stemming allows its character n-gram units to become more consistent and informative. Word2Vec, on the other hand, showed a slight decrease when applied to stemmed text (93.48%) compared to its raw performance of 96.17%, indicating that Word2Vec captures semantic relationships more effectively when words remain in their original morphological forms.

Contextual embeddings such as ELMo and IndoBERT exhibited minimal performance differences between raw and stemmed datasets (<1%). This aligns with the inherent capability of contextual models to dynamically encode meaning based on sentence structure, making them less dependent on morphological normalization. Overall, stemming primarily benefits subword-based models like FastText, while raw-text representations are more suitable for static models such as Word2Vec.

3.5. Confusion Matrix Analysis

To further evaluate the classification behavior of the best-performing model, a confusion matrix analysis was conducted using the CapsNet + Word2Vec (raw) configuration, which achieved the highest accuracy of 96.17%. Table 7 summarizes classification outcomes across the three sentiment categories.

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

The model demonstrated strong capability in identifying positive and neutral sentiments, reflected by high true positive and true neutral counts. Misclassification was more prominent in the negative class, where some negative tweets were incorrectly predicted as neutral, likely due to subtle sentiment expressions in policy-related discourse. The confusion between neutral and negative categories is consistent with prior sentiment studies on public policy discussions, where linguistic cues can be ambiguous or indirect.

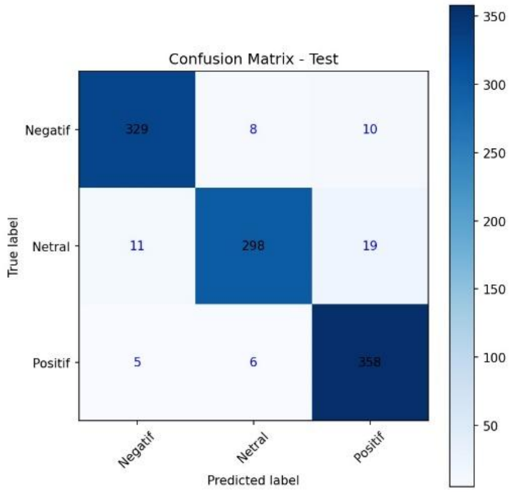

Figure 6. Confusion Matrix Summary for CapsNet + Word2Vec (Raw)

4. DISCUSSIONS

4.1. Comparative Analysis with Previous Studies

Overall, the findings demonstrate that the Capsule Network (CapsNet) model integrated with static embeddings (Word2Vec and FastText) outperforms contextual embeddings (ELMo and IndoBERT) in both stability and accuracy. The best configuration CapsNet + Word2Vec with stemming achieved an accuracy of 96.17%, surpassing results from previous studies employing similar methods.

For example, Diviya Prabha and Rathipriya (2022) reported 96% accuracy using CapsNet on COVID-19 tweet sentiment data [23], while Dong et al. (2020) achieved 98.8% on the IMDB dataset with a hybrid BiLSTM + CapsNet + GloVe model [24]. Toqeer Ehsan et al. (2023) used ELMo-based embeddings on low-resource Tamil and Tulu texts, yielding modest performance (F1-score 0.51) [25]. In contrast, Haziq et al. (2024) achieved 89.11% accuracy with LSTM + FastText on MyTelkomsel app reviews [26], and Noryasminda et al. (2025) reported 89.8% accuracy with BiLSTM + FastText for sentiment analysis of monkeypox-related tweets [27]. Compared to these works, the present study's 96.17% accuracy confirms a substantial improvement within the same methodological scope, particularly in the context of morphologically rich Indonesian text.

The advantage of the proposed model lies in its ability to maintain semantic coherence through dynamic routing, a hallmark of the CapsNet architecture. This allows the model to retain inter-word spatial relationships that are often lost in CNN or LSTM frameworks. The improvement of +1% accuracy in FastText after stemming further underscores the importance of morphological normalization for static and subword-based embeddings. Meanwhile, contextual embeddings such as ELMo and IndoBERT remained stable across both datasets, showing that they already encode syntactic and semantic context without preprocessing intervention.

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

Table 7. Comparison with Related Works

Ref	Method	Accuracy
[23]	CapsNet	96%
[24]	BiLSTM + CapsNet + GloVe	98.80%
[25]	BiLSTM + ELMo	51.33%
[26]	LSTM + FastText	89.11%
[27]	BiLSTM + FastText	89.80%
Our Proposed Method	CapsNet + Word2Vec (Stemming)	96.17%

4.2. Trade-Off Between Static and Contextual Embeddings

A clear trade-off is observed between static and contextual embeddings. Static embeddings—particularly Word2Vec—perform optimally when applied to raw text, as preserving the original morphological structure enriches semantic co-occurrence signals. In contrast, FastText benefits more from stemming because its subword-based representation becomes more consistent after affix reduction.

Contextual embeddings (ELMo and IndoBERT), which dynamically encode word meaning based on full-sentence context, show limited improvement from stemming and remain stable across both datasets. However, they incur higher computational costs and require large-scale fine-tuning to outperform static embeddings on domain-specific tasks such as MBG sentiment analysis.

In Indonesian, a language with extensive affixation and derivational morphology, static embeddings may struggle when morphological variation is high. Stemming reduces this variation and thus improves FastText performance, but raw input remains more beneficial for Word2Vec. This indicates that embedding choice should be aligned with linguistic properties and preprocessing strategies.

4.3. Implications for Informatics

The combination of CapsNet and word embeddings has meaningful implications for computational informatics and public policy analysis. From a technical standpoint, CapsNet provides a robust feature extraction mechanism capable of capturing nuanced semantic relationships often lost in traditional CNN/LSTM architectures. This makes it suitable for real-time large-scale sentiment monitoring scenarios.

From a policy and governance perspective, the model offers a data-driven mechanism to gauge public perception toward large-scale government initiatives such as the MBG program. Sentiment analytics derived from social media discourse can serve as an early indicator of public support, program bottlenecks, or misinformation trends. This integration demonstrates how informatics can be leveraged to enhance transparency, responsiveness, and evidence-driven policymaking in Indonesia’s social and health sectors.

4.4. Limitations and Future Work

Despite promising results, several limitations remain. First, manual sentiment labeling introduces subjective bias, as interpretations may vary across annotators. Future work should incorporate multiple annotators and inter-rater reliability metrics such as Cohen’s Kappa to improve consistency. Second,

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

although CapsNet offers strong representational capabilities, its dynamic routing mechanism increases computational cost, limiting scalability for very large datasets or production-level systems.

Further research may explore fine-tuning contextual embeddings such as IndoBERT on domain-specific corpora, integrating ensemble architectures (e.g., CapsNet–BiLSTM hybrids), or optimizing routing algorithms to reduce overhead while retaining semantic richness. Expanding the dataset beyond a single trending topic and incorporating multimodal signals (images, user metadata) may also enhance the generalizability of the sentiment classification model.

5. CONCLUSION

This study demonstrated the effectiveness of the Capsule Network (CapsNet) model in classifying public sentiment toward Indonesia's Makan Bergizi Gratis (MBG) program. Among all evaluated configurations, CapsNet combined with Word2Vec on raw text achieved the highest accuracy of 96.17%, highlighting the strong performance of static embeddings when morphological information is preserved. FastText showed the best performance on the stemmed dataset, indicating that subword-based models benefit from reduced morphological variation. Contextual embeddings such as ELMo and IndoBERT achieved stable but lower accuracy, suggesting that additional fine-tuning may be required for optimal performance on Indonesian social-media text.

Overall, these findings confirm the suitability of CapsNet for sentiment analysis in morphologically rich languages and demonstrate its potential for supporting data-driven evaluation of public policies. Future research should investigate domain-specific fine-tuning for IndoBERT, explore ensemble architectures such as CapsNet–BiLSTM, and optimize routing mechanisms to improve scalability while maintaining classification accuracy.

REFERENCES

[1] S. Mufti Prasetiyo, R. Gustiawan, and F. Rizzel Albani, "BIIKMA : Buletin Ilmiah Ilmu Komputer dan Multimedia Analisis Pertumbuhan Pengguna Internet Di Indonesia," *Biikma*, vol. 2, no. 1, pp. 65–71, 2024, [Online]. Available: https://jurnalmahasiswa.com/index.php/biikma

[2] S. F. Zis, N. Effendi, and E. R. Roem, "Perubahan Perilaku Komunikasi Generasi Milenial dan Generasi Z di Era Digital," *Satwika : Kajian Ilmu Budaya dan Perubahan Sosial*, vol. 5, no. 1, pp. 69–87, 2021, doi: 10.22219/satwika.v5i1.15550.

[3] G. Widjaja, "Perubahan pola komunikasi dalam masyarakat akibat penggunaan aplikasi pesan instan," *Prosiding Seminar Nasional Indonesia*, vol. 3, no. 1, pp. 10–16, 2025, [Online]. Available: https://adisampublisher.org/index.php/nasional/

[4] N. M. Malay, "Dampak Positif Dan Negatif Penggunaan Internet Pada Kalangan Mahasiswa Kelas a, B, C, Dan D Angkatan 2021 Prodi Pend. Bahasa Dan Sastra Indonesia, Universitas Nusa Cendana," *Jurnal Lazuardi*, vol. 5, no. 1, pp. 70–88, 2022, doi: 10.53441/jl.vol5.iss1.72.

[5] A. Kiftiyah, F. A. Palestina, F. U. Abshar, and K. Rofiah, "Makan Bergizi Gratis (MBG) Program in the Perspective of Social Justice and Socio – Political Dynamics," *Pancasila: Jurnal Keindonesiaan*, vol. 5, no. 1, pp. 101–112, 2025, doi: 10.52738/pjk.v5i1.726.

[6] A. A. Merlinda and Yusmar Yusuf, "Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan," *Ranah Research : Journal of Multidisciplinary Research and Development*, vol. 7, no. 2, pp. 1364–1373, 2025, doi: 10.38035/rrj.v7i2.1360.

[7] W. Trisno Aji, "Makan Bergizi Gratis di Era Prabowo-Gibran: Solusi untuk Rakyat atau Beban Baru?," *Naafi: Jurnal Ilmiah Mahasiswa*, vol. 2, no. 2, 2025, doi: 10.62387/naafijurnalilmiahmahasiswa.v2i2.134.

[8] D. Herdiana, "Implementasi Kebijakan Makan Bergizi Gratis (MBG): Faktor-faktor Pendorong dan Penghambat," *Jurnal Ilmiah Multidisiplin*, vol. 3, no. 2, pp. 470–477, 2025.

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

[9] A. N. Nurkalyisah, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen pada Twitter Berbahasa Indonesia Terhadap Penurunan Performa Layanan Indihome dan Telkomsel," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 10, no. 4, p. 387, 2022, doi: 10.26418/justin.v10i4.50858.

[10] A. Fatimatuzzahro, A. Larasati, and A. Dwiastuti, "A Sentiment Analysis about The Acquisition and Policy of X (Twitter) by Elon Musk," vol. 18, no. 2, pp. 849–863, 2025.

[11] B. R. Babu, S. Ramakrishna, and S. K. Duvvuri, "Advanced Sentiment and Trend Analysis of Twitter Data Using CNN-LSTM and Word2Vec," *4th International Conference on Sentiment Analysis and Deep Learning, ICSADL 2025 - Proceedings*, no. February, pp. 1536–1543, 2025, doi: 10.1109/ICSADL65848.2025.10933031.

[12] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.

[13] Gadipally Prasanth, "Advanced Feature Extraction and Visualization Techniques for Enhanced Sentiment Analysis on Twitter Data," *Journal of Information Systems Engineering and Management*, vol. 10, no. 3, pp. 1536–1548, 2025, doi: 10.52783/jisem.v10i3.8208.

[14] H. Liu, X. Chen, and X. Liu, "A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis," *IEEE Access*, vol. 10, pp. 32280–32289, 2022, doi: 10.1109/ACCESS.2022.3160172.

[15] S. Chakraborty, P. Das, S. Mahmud Dipto, M. A. Pramanik, and J. Noor, "An Analytical Review of Preprocessing Techniques in Bengali Natural Language Processing," *IEEE Access*, vol. 13, no. July, pp. 112428–112445, 2025, doi: 10.1109/ACCESS.2025.3574234.

[16] M. Iqbal *et al.*, "Journal of Computer Science Artificial Intelligence Implementation of Natural Language Processing for Chatbots in Customer," vol. 1, no. 1, pp. 18–23, 2024.

[17] S. Khomsah, R. D. Ramadhani, and S. Wijayanto, "The Accuracy Comparison Between Word2Vec and FastText On Sentiment Analysis of Hotel Reviews," *Jurnal RESTI*, vol. 6, no. 3, pp. 352–358, 2022, doi: 10.29207/resti.v6i3.3711.

[18] N. Z. B. Jannah and K. Kusnawi, "Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces," *Sinkron*, vol. 8, no. 2, pp. 727–733, 2024, doi: 10.33395/sinkron.v8i2.13559.

[19] Y. Dong, X. Li, M. He, and J. Li, "DC-BiLSTM-CNN Algorithm for Sentiment Analysis of Chinese Product Reviews," *Applied Artificial Intelligence*, vol. 39, no. 1, 2025, doi: 10.1080/08839514.2025.2461809.

[20] A. Rahoof, V. Chaturvedi, and M. Shafique, "FastCaps: A Design Methodology for Accelerating Capsule Network on Field Programmable Gate Arrays," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2023-June, 2023, doi: 10.1109/IJCNN54540.2023.10191653.

[21] M. S. I. Malik, A. Nawaz, M. M. Jamjoom, and D. I. Ignatov, "Effectiveness of ELMo embeddings, and semantic models in predicting review helpfulness," *Intelligent Data Analysis*, vol. 28, no. 4, pp. 1045–1065, 2024, doi: 10.3233/IDA-230349.

[22] Z. Zhan, "Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis: A Case of Food Reviews," *ITM Web of Conferences*, vol. 70, p. 02013, 2025, doi: 10.1051/itmconf/20257002013.

[23] V. Diviya Prabha and R. Rathipriya, "Competitive Capsule Network Based Sentiment Analysis on Twitter COVID'19 Vaccines," *Journal of Web Engineering*, vol. 21, no. 5, pp. 1583–1602, 2022, doi: 10.13052/jwe1540-9589.2159.

[24] Y. Dong, Y. Fu, L. Wang, Y. Chen, Y. Dong, and J. Li, "A sentiment analysis method of capsule network based on BiLSTM," *IEEE Access*, vol. 8, pp. 37014–37020, 2020, doi: 10.1109/ACCESS.2020.2973711.

[25] T. Ehsan, A. Tehseen, K. Sarveswaran, and A. Ali, "Sentiment Analysis of Code-Mixed Tamil and Tulu by Training Contextualized ELMo Representations," *DravidianLangTech 2023 - 3rd Workshop on Speech and Language Technologies for Dravidian Languages, associated with 14th International Conference on Recent Advances in Natural Language Processing, RANLP 2023 - Proceedings*, no. September, pp. 152–159, 2023, doi: 10.26615/978-954-452-085-4_020.

Vol. 7, No. 1, February 2026, Page. 633-646
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5424

[26] M. R. Haziq, Y. Sibaroni, and S. S. Prasetyowati, "Word Embedding Optimization in Sentiment Analysis of Reviews on Mytelkomsel App Using Long Short-Term Memory and Synthetic Minority Over-Sampling Technique," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 6, pp. 1581–1589, 2024, doi: 10.52436/1.jutif.2024.5.6.2498.

[27] Noryasminda, T. H. Saragih, R. Herteno, M. R. Faisal, and A. Farmadi, "Analysis of the Effect of Feature Extraction on Sentiment Analysis using BiLSTM: Monkeypox Case Study on X/Twitter," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, pp. 344–357, 2025, doi: 10.35882/ijeeemi.v7i2.73.

[28] The Indonesian Institute, Policy Assessment 2025 – Analisis Implementasi Program Makan Bergizi Gratis sebagai Strategi Penurunan Stunting di Indonesia, The Indonesian Institute – Center for Public Policy Research, Policy Assessment 2025, 2 July 2025. [Online]. Available: https://www.theindonesianinstitute.com/policy-assessment-2025-analisis-implementasi-program-makan-bergizi-gratis-sebagai-strategi-penurunan-stunting-di-indonesia

[29] F. M. S. Putra, S. Rakasiwi, and N. Ariyanto, "Twitter Sentiment Classification towards Telecommunication Provider Users in Indonesia," J. Artif. Intell. Comput., vol. , 2025. [Online]. Available: https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/9143

[30] R. I. Perwira, et al., "Domain-Specific Fine-Tuning of IndoBERT for Aspect-Level Sentiment Analysis," J. Inf. Syst. Eng. Bus. Intell., vol. 11, no. 1, pp. 30–40, 2025. [Online]. Available: https://e-journal.unair.ac.id/JISEBI/article/download/63660/32187

[31] H. Alkaabi, A. K. Jasim, and A. Darroudi, "From Static to Contextual: A Survey of Embedding Advances in NLP," *PERFECT: Journal of Smart Algorithms*, vol. 2, no. 2, pp. 57–66, 2025, doi: 10.62671/perfect.v2i2.77.

[32] D. Kurniasari and F. Aulia Pradana, "Deep Learning-Driven Comparative Study of Word Embedding Techniques: Word2Vec, GloVe, and FastText in Health Condition Reviews," *ISRG Journal of Multidisciplinary Studies (ISRGJMS)*, vol. III, no. II, pp. 51–62, 2025, doi: 10.5281/zenodo.14890438.

[33] M. Kurniawan, K. Kusrini, and M. R. Arief, "Part of Speech Tagging Pada Teks Bahasa Indonesia dengan BiLSTM + CNN + CRF dan ELMo," *Jurnal Eksplora Informatika*, vol. 11, no. 1, pp. 29–37, 2022, doi: 10.30864/eksplora.v11i1.506.

[34] P. Sayarizki and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," *Journal on Computing*, vol. 9, no. 2, pp. 61–72, 2024, doi: 10.34818/indojc.2024.9.2.934.

[35] B. Thiyagarajan, M. Thenmozhi, and K. Revathy, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING An Improved Routing based Capsule Network for Hyperspectral Image Classification," *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 2024, no. 2, pp. 79–89, 2024, [Online]. Available: www.ijisae.org

[36] M. FATONI and E. ERNASTUTI, "Ornamental Plants Classification Using Integration of Convolution With Capsule Network," *MIND (Multimedia Artificial Intelligent …*, vol. 8, no. 2, pp. 158–172, 2023, [Online]. Available: https://ejurnal.itenas.ac.id/index.php/mindjournal/article/view/9157%0Ahttps://ejurnal.itenas.ac.id/index.php/mindjournal/article/download/9157/3486

[37] A. E. Regi and G. J. Leelipushpam, "Enhancing Aspect Based Sentiment Analysis Using BERT and Capsule Networks," *Proceedings - 2nd IEEE International Conference on Device Intelligence, Computing and Communication Technologies, DICCT 2024*, pp. 364–369, 2024, doi: 10.1109/DICCT61038.2024.10532838.