Analysis of Public Sentiment Towards President Prabowo's Work Program Using The CNN

Angelina Pramana Thenata [1]*, Dimas Sakti Reka Saputra [2]*
* Informatika, Universitas Bunda Mulia
angelina.pramana@outlook.com [1], s32220142@student.ubm.ac.id [2]

Article Info

Article history:

Received 2025-04-14
Revised 2025-07-16
Accepted 2025-07-30

Keyword:

Sentiment Analysis,
Prabowo's Work Program,
Convolutional Neural Network,
Confusion Matrix.

ABSTRACT

Digital media has now become the primary means for Indonesians to receive and respond to information, including the work programs presented by Prabowo Subianto. One of the programs that is widely discussed by the public is related to efforts to improve the national economy. Public responses to this issue are widespread on social media, reflecting diverse sentiments. Therefore, this study aims to analyze the sentiment of comments from social media users X regarding President Prabowo's work programs in the economic sector, using a deep learning approach based on the Convolutional Neural Network (CNN) architecture. The methods employed include data collection, text preprocessing, and training a CNN model. The dataset used consisted of 2,467 data points, with 1,086 labeled as positive and 1,381 labeled as negative. The test results showed that the model achieved an accuracy of 87.45% and an Area Under the Curve (AUC) score of 0.9373, indicating excellent classification performance in distinguishing between positive and negative sentiments. This study proves that the combination of CNN and FastText is a practical approach to understanding text-based public opinion from social media.

I. INTRODUCTION

The widespread use of digital media in Indonesia means that political communication is no longer limited to conventional media such as television and newspapers. Political information in the digital era is now increasingly accessible through various social media platforms. For example, President Prabowo Subianto utilizes multiple social media platforms, particularly Twitter, to disseminate his work programs and vision to the Indonesian public. One of Prabowo's proposed work programs is to increase economic growth by 6-7% [1]. The dissemination of information about this work program has elicited various responses from the Indonesian public.

Therefore, an in-depth analysis of how the public responds to this program is essential to understand the level of support or opposition and the factors that influence it. The use of text mining technology enables the analysis of public sentiment across various social media platforms. This technique can identify the dominant sentiments contained in public responses [2]. Cholid Fadilah Hasri and Debby Alita have employed this technique. The researchers found that the Naive Bayes Classifier (NBC) and Support Vector Machine (SVM) algorithms were proven to produce classification accuracy rates of 81.07% for NBC and 79.96% for SVM in analyzing Twitter user sentiment regarding the impact of the coronavirus [3].

Furthermore, research conducted by Budi Haryanto, Yova R., Fathur Rohman, et al. analyzed public opinion on the 2019 presidential candidates, Jokowi-Maruf and Prabowo-Sandi, via Facebook using the NBC algorithm. They found that Jokowi-Maruf received a dominant positive sentiment of 56.76% and a dominant negative sentiment of 43.24%. Meanwhile, Prabowo-Sandi received a dominant negative sentiment of 75.79% and a dominant positive sentiment of 24.21% [4].

Furthermore, Nurul Rezki, Sri Astuti Thamrin, and Siswanto Siswanto conducted research on sentiment toward the independent campus policy on Twitter using the SVM algorithm with Word2Vec feature extraction. The study found that the classification model had an accuracy rate of 89.87%, precision of 91.20%, recall of 84.44%, and F-Measure of

87.68% [5]. Most previous studies have proposed various algorithms to analyze public sentiment toward the coronavirus, the 2019 presidential candidates, and the independent campus policy program.

However, these studies have not examined Indonesian public sentiment toward work programs aimed at improving the Indonesian economy. One method that can be used for sentiment analysis is the Convolutional Neural Network (CNN) algorithm. This algorithm is one of the most effective deep learning architectures for text-based sentiment analysis tasks. The primary advantage of CNNs lies in their ability to automatically capture local semantic patterns through convolutional filters, which can identify phrases or n-grams containing significant emotional connotations in text [6]. On the other hand, compared to sequential models like RNN and LSTM, which suffer from training efficiency weaknesses and are susceptible to the vanishing gradient problem, CNNs offer advantages in computational efficiency, robustness to vanishing gradients, and training speed due to their parallelizable architecture, enabling faster training while maintaining high accuracy [7].

Furthermore, FastText is a practical word representation method for sentiment analysis due to its ability to capture subword information, thus addressing the problems of spelling variations and informal words that frequently appear in natural language text. With this approach, FastText can produce embeddings that are more robust to spelling errors and language variations, thereby enhancing the performance of sentiment analysis models, particularly when using balanced and diverse datasets. The use of FastText has been proven to provide stable and accurate results in various sentiment analysis studies, particularly in languages with high levels of informality such as Indonesian [8].

Therefore, this study aims to design a CNN architecture model that can analyze the sentiment of social media users X regarding President Prabowo's work program (increasing economic growth). The results of this study provide an overview of public perception of President Prabowo's work program. This information is expected to form the basis for developing more effective communication strategies and provide insight for policymakers to adapt their programs to the needs and expectations of the community.

II. METHOD

This research has a flowchart of the research stage shown in Figure 1.

A. Data Collection

The data collection stage in this study was sourced from tweets on the Indonesian-language social media platform X, using the keywords "Prabowo's economy," "Prabowo's economic policy," and "Prabowo's economic vision." The tweet data was collected from October 20, 2024, to May 31, 2025, comprising a total of 2,467 data points, divided into two classes: a positive class of 1,086 data points and a negative class of 1,381 data points.

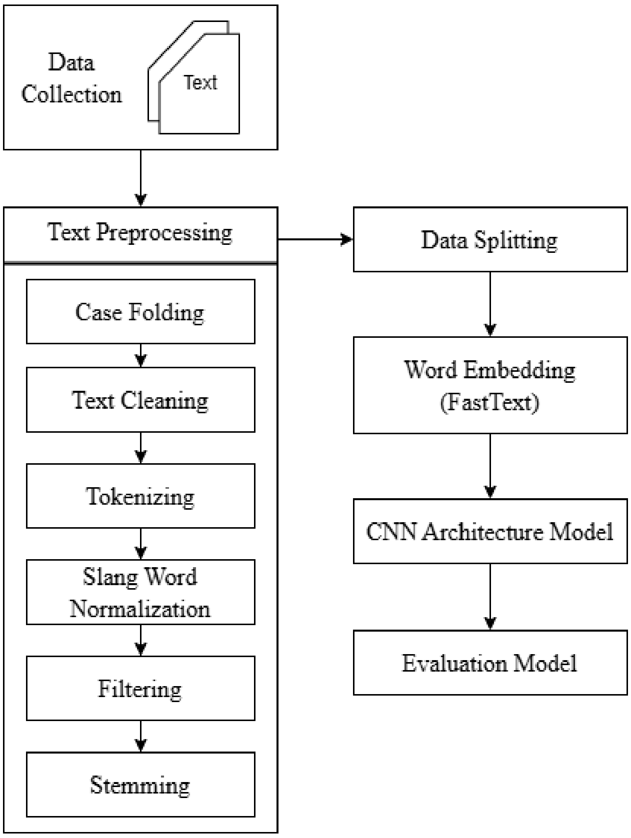

Figure 1 Research Flow

B. Text Preprocessing

The text preprocessing stage prepares ready-to-use data, making it easy to analyze [9]. The stages carried out in text preprocessing are as follows.

1. Case Folding

The first process in this stage is case folding, which converts all letters in the text to lowercase, as shown in Table I. This aims to address the issue of case sensitivity. By standardizing the letters, words like "Prabowo" and "prabowo" will be treated as the same token, thereby reducing redundancy and dimensionality of feature data.

TABLE I
CASE FOLDING

Original Text	Result
@muannas_alaidid Pernyataan yg merendahkan Pak Prabowo ... Padahal beliau sedang berusaha meningkatkan ekonomi masyarakat dengan mengundang banyak investor ... @prabowo @Gerindra Agar rakyatnya gak perlu jual aset untuk hidup ...	@muannas_alaidid pernyataan yg merendahkan pak prabowo ... padahal beliau sedang berusaha meningkatkan ekonomi masyarakat dengan mengundang banyak investor ... @prabowo @gerindra agar rakyatnya gak perlu jual aset untuk hidup ...

2. Text Cleaning

The second process in this stage is to remove all irrelevant characters and elements, or elements considered noise in the text analysis, as shown in Table II. The elements removed include punctuation, numbers, special characters, URL links, mentions (@), and hashtags (#). This removal was done because these elements generally do not contribute significant semantic meaning to the analysis.

TABLE II
TEXT CLEANING

Text	Result
@muannas_alaidid pernyataan yg merendahkan pak prabowo ... padahal beliau sedang berusaha meningkatkan ekonomi masyarakat dengan mengundang banyak investor ... @prabowo @gerindra agar rakyatnya gak perlu jual aset untuk hidup ...	pernyataan yg merendahkan pak prabowo padahal beliau sedang berusaha meningkatkan ekonomi masyarakat dengan mengundang banyak investor agar rakyatnya gak perlu jual aset untuk hidup

3. Tokenizing

The third process in this stage is breaking down text strings (sentences) into smaller units called tokens. Tokens are generally represented as single words. This process converts the data format from a complete sentence into a list of words, forming the foundation for further analysis at the word level, as shown in Table III.

TABLE III
TOKENIZING

Text	Result
pernyataan yg merendahkan pak prabowo padahal beliau sedang berusaha meningkatkan ekonomi masyarakat dengan mengundang banyak investor agar rakyatnya gak perlu jual aset untuk hidup	['pernyataan', 'yg', 'merendahkan', 'pak', 'prabowo', 'padahal', 'beliau', 'sedang', 'berusaha', 'meningkatkan', 'ekonomi', 'masyarakat', 'dengan', 'mengundang', 'banyak', 'investor', 'agar', 'rakyatnya', 'gak', 'perlu', 'jual', 'aset', 'untuk', 'hidup']

4. Slang Word Normalization

The fourth process in this stage aims to convert these words into the standardized forms shown in Table IV. This process requires a previously compiled dictionary. This normalization is crucial to ensuring consistency of meaning and standardization of vocabulary within the corpus.

TABLE IV
TOKENIZING

Text	Result
['pernyataan', 'yg', 'merendahkan', 'pak', 'prabowo', 'padahal', 'beliau', 'sedang', 'berusaha', 'meningkatkan', 'ekonomi', 'masyarakat', 'dengan', 'mengundang', 'banyak', 'investor', 'agar', 'rakyatnya', 'gak', 'perlu', 'jual', 'aset', 'untuk', 'hidup']	['pernyataan', 'yang', 'merendahkan', 'pak', 'prabowo', 'padahal', 'beliau', 'sedang', 'berusaha', 'meningkatkan', 'ekonomi', 'masyarakat', 'dengan', 'mengundang', 'banyak', 'investor', 'agar', 'rakyatnya', 'tidak', 'perlu', 'jual', 'aset', 'untuk', 'hidup']

5. Filtering

The fifth step in this stage involves eliminating common words (stopwords) that frequently appear but make minimal semantic contributions. Words such as conjunctions (e.g., "yang,"dengan”), prepositions, and pronouns (e.g., "beliau”) fall into this category. This process is carried out to reduce the feature dimensionality and focus the analysis on words that convey core meaning. An example of the results of this process is shown in Table V.

TABLE V
FILTERING

Text	Result
['pernyataan', 'yang', 'merendahkan', 'pak', 'prabowo', 'padahal', 'beliau', 'sedang', 'berusaha', 'meningkatkan', 'ekonomi', 'masyarakat', 'dengan', 'mengundang', 'banyak', 'investor', 'agar', 'rakyatnya', 'tidak', 'perlu', 'jual', 'aset', 'untuk', 'hidup']	['pernyataan', 'merendahkan', 'prabowo', 'berusaha', 'meningkatkan', 'ekonomi', 'masyarakat', 'mengundang', 'investor', 'rakyatnya', 'jual', 'aset', 'hidup']

6. Stemming

Finally, this process reduces affixed words to their basic form (stem or root word), as shown in Table VI. For example, the word "mengerak" (increasing) becomes "tingkat" (level), and "jelas" (statement) becomes "benar" (real). This process aims to group the morphological variations of a word into a single representation. The goal is to reduce vocabulary complexity without losing its basic meaning.

TABLE VI
STEMMING

Text	Result
['pernyataan', 'merendahkan', 'prabowo', 'berusaha', 'meningkatkan', 'ekonomi', 'masyarakat', 'mengundang', 'investor', 'rakyatnya', 'jual', 'aset', 'hidup']	['nyata', 'rendah', 'prabowo', 'usaha', 'tingkat', 'ekonomi', 'masyarakat', 'undang', 'investor', 'rakyat', 'jual', 'aset', 'hidup']

C. Data Splitting

After going through all pre-processing stages, the cleaned and structured dataset is then divided into two independent data subsets: training data and test data. This process is carried out using an 80:20 split. Eighty percent of the total data is allocated as training data, which serves to train the classification model to recognize relevant patterns. Meanwhile, the remaining 20% of the data is used as test data that is not involved in the training process. The primary purpose of this separation is to evaluate the model's performance on data that has never been "seen" before, thus providing an objective measure of the model's generalization ability and accuracy in classifying new data.

D. Word Embedding (FastText)

In the next step, the categorical word tokens from the preprocessing stage are converted into numeric vector representations using FastText's pretrained word embedding. This method relies on subword-based representation, which represents each word as a vector concatenation of n-gram characters. This approach enables the model to effectively generate vectors for words not in the training vocabulary (Out-of-Vocabulary), making it particularly relevant for analyzing Indonesian text with complex and varied morphology, especially in social media data.

The technique begins by converting each token from the text preprocessing results into a high-dimensional vector based on the subword representation. Because each sentence produces a varying number of tokens, and CNN models require inputs of uniform size, a length standardization process is performed. This technique involves padding, which is the addition of zero vectors to sentences shorter than a predetermined maximum length, and truncating longer sentences. Once the length is standardized, these vectors are then sequentially arranged to form a fixed-size 2D matrix. This matrix serves as the direct input layer for the CNN architecture. By preserving word sequence information through the rows of the matrix and semantic meaning through the columns of the vector, this data structure provides optimal input for CNN convolutional filters to extract essential features relevant for classification tasks.

TABLE VII
HASIL FASTTEXT

Token	Vektor Fitur
pernyataan	[0.15, -0.42, 0.71, ..., 0.88]
merendahkan	[-0.21, 0.59, -0.14, ..., 0.42]
prabowo	[0.75, 0.18, -0.31, ..., -0.69]
berusaha	[0.38, -0.81, 0.52, ..., 0.17]
meningkatkan	[0.93, 0.29, -0.74, ..., -0.31]
ekonomi	[0.47, -0.62, 0.85, ..., 0.59]
masyarakat	[0.61, 0.15, -0.28, ..., -0.92]
mengundang	[-0.84, 0.71, 0.39, ..., 0.75]
investor	[0.19, -0.95, -0.48, ..., -0.27]
rakyatnya	[0.52, 0.38, 0.51, ..., 0.14]
jual	[-0.72, -0.19, 0.73, ..., 0.84]
aset	[0.28, 0.41, -0.93, ..., -0.49]
hidup	[-0.41, 0.68, 0.25, ..., 0.63]
<padding>	[0.0, 0.0, 0.0, ..., 0.0]
…	…
<padding>	[0.0, 0.0, 0.0, ..., 0.0]

For example, the results of the FastText technique, using a fixed matrix size of 50 × 300, are shown in Table VII. The first dimension represents the maximum sequence length (max_length), a hyperparameter determined based on an analysis of the token length distribution in the data corpus. This value was chosen to capture the majority of the data through padding and truncating techniques, thus minimizing the loss of contextual information while maintaining a uniform input size [10]. Meanwhile, the second dimension for the embedding vector was chosen because it is a validated standard and commonly used in various previous studies, particularly in pre-trained models like FastText [11]. This dimension is capable of richly representing the semantic nuances of words without incurring excessive computational burden, making it an optimal configuration for deep learning architectures [12].

E. CNN Architecture Model

The CNN architecture model used in this study consists of the stages depicted in Figure 2.

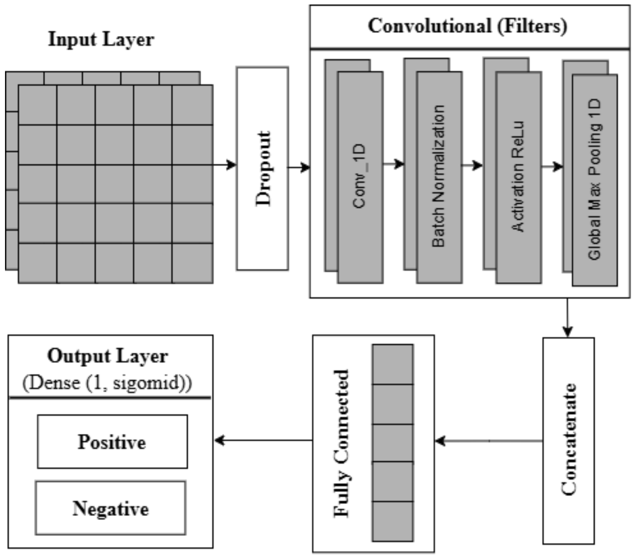

Figure 2 CNN Architecture Model

The designed CNN architecture model begins with the input layer, which in this study receives a feature matrix derived from pretrained FastText word embeddings. The semantically rich input matrix is then passed through a Dropout layer as a regularization mechanism to reduce overfitting. The core component of this architecture utilizes multiple parallel convolutional channels, each using a Conv1D layer with varying kernel sizes to extract linguistic features at varying n-gram scales [13]. After the convolution operation, a Batch Normalization layer is applied to stabilize and accelerate training, followed by a nonlinear ReLU activation. Next, a GlobalMaxPooling1D layer aggregates these features, retaining only the highest response values from each channel. Finally, these most representative features are consolidated into a single, comprehensive feature vector through a Concatenate layer.

A fully-connected block then processes this single feature vector. This block begins with a Dense layer, which models high-level feature interactions, also optimized with Batch Normalization and ReLU activation. After passing through an additional Dropout layer, this architecture is concluded by an output layer with one neuron and a Sigmoid activation function. This layer maps the final feature representation to a probability value between 0 and 1. Then the probability value is used to determine the sentiment class, namely a negative class (0), if the probability is close to zero, and a positive class (1) if the probability is close to one, according to the binary classification framework [10].

F. Evaluation Model

The final stage of this research involved testing the created CNN model using a confusion matrix. The elements tested included the following: [14].

1. Accuracy describes the closeness of the classification result to the actual value.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. Precision refers to the agreement between accurate positive data and predicted data.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

3. Recall describes the classification's success in obtaining information from all available data.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Next, the model was tested using the Receiver Operating Characteristic (ROC) method, which measures the performance of a binary classification model. ROC describes the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) at various decision thresholds, thus providing a comprehensive overview of the model's ability to distinguish between positive and negative sentiments. The Area Under the Curve (AUC) value of the ROC is used as a quantitative indicator of model performance; the closer it is to 1, the better the model is at classifying [15].

III. HASIL DAN PEMBAHASAN

The CNN model used divided the dataset into 80% training and 20% testing. Next, an evaluation was performed using the confusion matrix method, and the results of the model evaluation are presented in Figure 3. Based on the test results presented, the model demonstrated excellent performance, achieving an overall accuracy of 87.45%. Further analysis of the classification report provides in-depth insight into the model's performance for each class. For the positive class (1), the model demonstrated high precision (90.26%), indicating that when it predicted a sentiment as positive, the prediction had a very high degree of accuracy. On the other hand, for the negative class (0), the model excelled in recall (88.02%), demonstrating its ability to successfully identify the majority of all negative samples in the test data. The slightly higher number of False Negatives (36) compared to False Positives (26) explains why the recall for the positive class was somewhat lower than for the negative class.

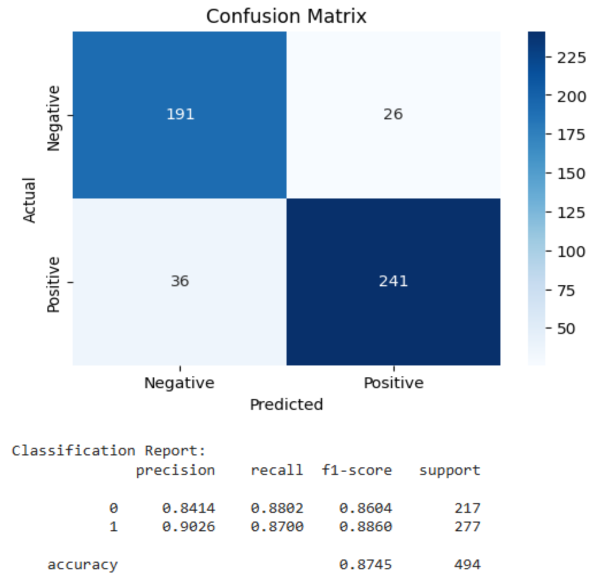

Figure 3 Evaluation Results Confusion Matrix

This study evaluated performance using a Receiver Operating Characteristic (ROC) curve, as shown in Figure 4.

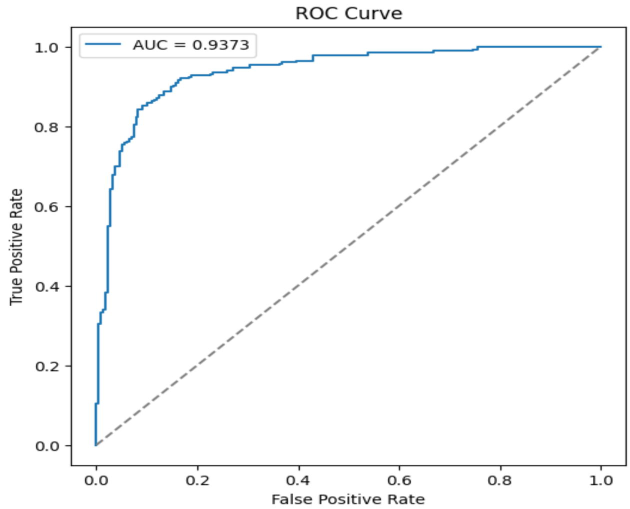

Figure 4 Evaluation Results ROC

This evaluation was conducted to assess the model's ability to discriminate between positive and negative classes. The curve's sharp curve toward the upper left corner indicates that the model achieved a high True Positive Rate while maintaining a low False Positive Rate.

This performance was quantified by the Area Under the Curve (AUC) value, with the model achieving an AUC score of 0.9373. This AUC value, very close to 1.0, indicates the model's excellent capability to separate classes. These results demonstrate that the proposed architecture is highly effective and reliable in identifying sentiment in the analyzed text data.

IV. KESIMPULAN

This study successfully applied a Convolutional Neural Network (CNN) architecture combined with FastText word embeddings to conduct sentiment analysis of public opinion regarding Prabowo's work program related to Indonesia's economic growth, utilizing social media data. The dataset used consisted of 2,467 data points, consisting of two sentiment classes: 1,086 positive data points and 1,381 negative data points. Based on the test results, the proposed model demonstrated excellent and reliable performance. This was shown by an overall accuracy of 87.45% and an AUC (Area Under the Curve) score of 0.9373, indicating the model's excellent ability to distinguish between positive and negative sentiment.

The application of comprehensive text preprocessing techniques and the use of FastText word embedding proved capable of producing semantically rich feature representations, even for diverse and non-standard vocabulary commonly found on social media. The CNN architecture with multiple parallel convolutional channels was also effective in capturing linguistic features (n-grams) of varying lengths, contributing to the model's high performance. These findings demonstrate that a deep learning approach combining CNN and FastText architectures is a robust and reliable method for analyzing and measuring public sentiment on specific issues, such as economic policy.

Future research could explore hybrid models, such as CNN-LSTM, utilize the Transformer architecture, or employ pretrained Indonesian-language models, like IndoBERT. This approach could also involve applying multiclass sentiment analysis and evaluation to real-time data to enhance model generalization and relevance.

DAFTAR PUSTAKA

[1] P. Subianto and G. R. Raka, "Visi, Misi dan Program Calon Presiden dan Wakil Presiden 2024-2029," *Medcom.id*, 2024.

[2] B. Hakim, "Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 2, pp. 16–22, 2021, doi: 10.30813/jbase.v4i2.3000.

[3] C. F. Hasri and D. Alita, "Penerapan Metode Naive Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, doi: 10.33365/jatika.v3i2.2026.

[4] B. Haryanto, Y. Ruldeviyani, F. Rohman, T. N. Julius Dimas, R. Magdalena, and F. Muhamad Yasil, "Facebook Analysis of Community Sentiment on 2019 Indonesian Presidential Candidates From Facebook Opinion Data," *Procedia Comput Sci*, vol. 161, pp. 715–722, 2019, doi: 10.1016/j.procs.2019.11.175.

[5] N. Rezki, S. A. Thamrin, and S. Siswanto, "Sentiment Analysis of Merdeka Belajar Kampus Merdeka Policy Using Support Vector Machine With Word2Vec," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 1, pp. 0481–0486, 2023, doi: 10.30598/barekengvol17iss1pp0481-0486.

[6] S. N. Listyarini and D. A. Anggoro, "Analisis Sentimen Pilkada di Tengah Pandemi Covid-19 Menggunakan Convolution Neural Network (CNN)," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 1, no. 7, pp. 261–268, 2021, doi: 10.52436/1.jpti.60.

[7] F. M. Shiri, T. Perumal, N. Mustapha, and R. Mohamed, "A Comprehensive Overview and Comparative Analysis on Deep Learning Models."

[8] R. N. Razzak, "Ekspansi Fitur dengan FastText untuk Analisis Sentimen di Media Sosial X Menggunakan Recurrent Neural Network dan Covlutional Neural Network," 2025.

[9] M. Alfonso and D. B. Rarasati, "Sentiment Analysis of 2024 Presidential Candidates Election Using SVM Algorithm," *JISA(Jurnal Informatika dan Sains)*, vol. 6, no. 2, pp. 110–115, 2023, doi: 10.31326/jisa.v6i2.1714.

[10] H. Kim and Y. S. Jeong, "Sentiment classification using Convolutional Neural Networks," *Applied Sciences (Switzerland)*, vol. 9, no. 11, Jun. 2019, doi: 10.3390/app9112347.

[11] M. Umer *et al.*, "Impact of convolutional neural network and FastText embedding on text classification," *Multimed Tools Appl*, vol. 82, no. 4, pp. 5569–5585, Feb. 2023, doi: 10.1007/s11042-022-13459-x.

[12] I. N. Khasanah, "Sentiment Classification Using fastText Embedding and Deep Learning Model," in *Procedia CIRP*, Elsevier B.V., 2021, pp. 343–350. doi: 10.1016/j.procs.2021.05.103.

[13] A. Rajesh and T. Hiwarkar, "Sentiment analysis from textual data using multiple channels deep learning models," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, Nov. 2023, doi: 10.1186/s43067-023-00125-x.

[14] A. Hagi and D. B. Rarasati, "Sentiment Analysis of Sirekap Application Review Using Logistic Regression Algorithm," *Jurnal Informatika*, vol. 11, no. 2, pp. 55–64, 2024.

[15] J. Li, "Area under the ROC Curve has the most consistent evaluation for binary classification," *PLoS One*, vol. 19, no. 12 December, Dec. 2024, doi: 10.1371/journal.pone.0316019.