MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
Volume 17, No. 1(2025), pp 5-11 DOI : 10.18860/mat.v17i1.27995
Received : June 26th 2024; **Accepted** : March 17th 2025 ; **Avalaible Online** : March 28th 2025

Gated Recurrent Unit (GRU) for Sentiment Classification on Imbalanced Data: The COVID-19 Vaccine Program in Twitter

Mukhlis Hadi, Surya Agustian

Abstract— **The initial implementation of the COVID-19 vaccination by the Indonesian government sparked mixed reactions from the public, ranging from strong support to fierce opposition. These differing opinions influenced individuals' decisions to either accept or refuse the vaccination program for themselves or their families. Public sentiment, expressed through posts, comments, or status updates, provides valuable insights into vaccine acceptance or rejection. This study conducts sentiment analysis using deep learning techniques, specifically employing the Gated Recurrent Unit (GRU) method on Twitter data. The dataset consists of three sentiment classes: positive, negative, and neutral. The Word2Vec word embedding model was used as input and trained on a COVID-19 vaccination sentiment dataset collected from Twitter. Since the classes in the existing data tweets are imbalanced, some other steps are required to improve the classification. The best-performing model achieved an F1-score of 66% and an accuracy of 69%. This classification model effectively addresses the class imbalance problem, delivering competitive results compared to other methods.**

Index Terms -- **Gated Reccurent Unit (GRU), COVID-19 vaccine, sentiment classification**

I. Introduction

COVID-19 virus epidemic was first identified in Wuhan, China and has spread around the world, leading to a pandemic at the end of 2019 (December 2019) [1]. The coronavirus spread rapidly across the world, including Indonesia. The first case of covid-19 was found in Indonesia on March 2, 2020.

The impact of the covid-19 epidemic in Indonesia has a major impact in various fields such as the economy, social life and political education [2]. The Indonesian government's efforts for the covid-19 epidemic include implementing health protocol policies, large-scale social restrictions (PSBB) to community vaccinations to suppress the spread of Covid-19. Various types of vaccines have begun to be used in various countries in the world including Indonesia [3], [4]. The covid-19 vaccination is considered controversial and has drawn responses from the public with the emergence of pros and cons expressed on various social media, one of which is Twitter. The response arose because many were concerned about the impact of vaccination and the validity of the clinical feasibility trials of several covid-19 vaccines in circulation.

These responses can be used to conduct sentiment analysis to determine the tendency of public responses to COVID-19 vaccination, whether they tend to comment positively, neutrally, or negatively. Sentiment is a person's opinion, feeling, judgment, or view of a service, event or product [5]. Sentiment analysis is an activity to analyze a person's opinion, assessment, emotions and behavior on a product, organization, service, as well as cases or problems that are happening in the surrounding environment [6].

Sentiment analysis is a method of extracting information about individuals' views on an event or topic by classifying the emotional direction of a text. The classification is done to see whether the text is positive, negative or neutral. Sentiment analysis can be used to determine public opinion on issues such as corruption, and demonstrations based on textual data.

Sentiment analysis of public views on social media can provide valuable insights to understand the dynamics of public opinion regarding COVID-19 vaccination. In that context, this study proposes the application of the Gated Recurrent Unit (GRU) Method for the classification of public sentiment towards the administration of the COVID-19 vaccine on Twitter social media.

GRU method, as a form of recurrent neural network, has proven effective in overcoming the challenges of temporal and contextual text analysis. By utilizing GRU's ability to understand data sequences, this research aims to identify and categorize people's sentiments more

Manuscript was received June 26, 2024. This work was self-supported for B.Eng. graduation requirements from UIN Sultan Syarif Kasim, Riau.

Mukhlis Hadi, is with the Informatic Engineering Departement of UIN Sultan Syarif Kasim, Riau, Indonesia (email 11751102189@students.uin-suska.ac.id)

Surya Agustian, is with the Informatic Engineering Departement of UIN Sultan Syarif Kasim, Riau, Indonesia (corresponding author email surya.agustian@uin-suska.ac.id)

accurately [7]. The superiority of GRU in understanding data sequences is expected to increase the accuracy and accuracy in classifying public sentiment regarding the administration of the COVID-19 vaccine.

By comparing previous research [8], For sentiment classification using deep learning techniques with LSTM, using the same data and dividing into 3 classes of positive, negative and neutral. In this study the data was balanced into 2563 training data, 778 validation data, and 400 test data with 1802 neutral data, 1066 negative data, and 566 positive data. The best results in this study provide an F1- Score value of 54% with 66% accuracy. Furthermore, in other related research [9], [10], which uses machine learning techniques with Logistic Regression and SVM, getting F1 - Score values of 60% and 65% with accuracy of 67% and 69%, respectively.

As the problems described in the research above, the authors will use the GRU method using word2vec word embeddings to analyze public sentiment in Indonesia towards the Covid-19 vaccine on Twitter social media.

II. METHOD

The purpose of this research is to classify public sentiment on Twitter social media related to the Covid 19 vaccine. Broadly speaking, this research consists of 9 research steps, namely dataset collection and balancing, preprocessing, word2vec training, tokenization, pad sequences, GRU-classifier training process, best model (optimal) selection, model testing and evaluation as shown in Figure 1.

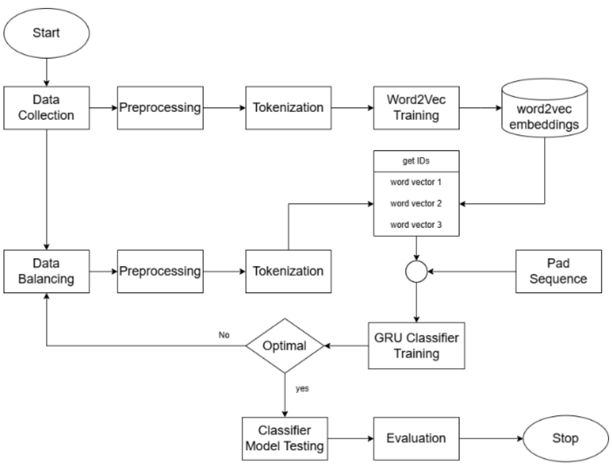


Fig. 1 Methodology

A. Data Collection

The dataset was successfully collected within the time span of March to April 2021, consisting of 12,000 tweets. The same dataset was also utilized in previous research [8]–[17]. The keywords used for data retrieval were selected based on topics related to COVID-19 vaccination activities, including "*Vaksin*", "*Vaksin Covid*", "*Vaksin Corona*", "*Vaksin Sinovac*", "*Vaksin Gratis*", "*Vaksin Gagal*", "*Vaksin Berhasil*", "*Vaksin Aman*", and "*Vaksinisasi Indonesia*".

Then the labeling process is carried out for tweets containing positive, negative sentiments. Labels are annotations given to each tweet, to be classified into positive, negative or neutral classes. Positive labels usually contain words of praise, advice, feedback, and reflections on positive emotions such as joy, pleasure, and happiness. Negative labels may contain complaints, expressions of sarcasm, criticism, and reflections of negative emotions such as frustration, dissatisfaction, and disappointment. Anything beyond that can be categorized as neutral. Of these labels, the classification task to be performed is multilabel classification, which has more difficult challenges than using only 2 labels (binary classification).

The labeling process was conducted through crowdsourcing, with each tweet annotated by three individual annotators [18]. The final label for each tweet is determined by majority vote. The crowdsourcing was conducted by 12 people, divided into 4 groups, with each group consisting of 3 annotators. Each annotator gives label to 3000 data. Each tweet must have a dominant class for label assignment. If a tweet has equal values for positive, negative, and neutral labels (equal to 1 for tweets labeled by 3 annotators), or if no major vote is present, it is removed from the dataset.

Table 1. Dataset Labeling

Sentiment	Tweet
Negative	@501Awani Bodoh. SOP Bodoh, Mentri Bodoh, Vaksin Bodoh
Positive	@kei_aria ga perlu ragu bro,vaksin aztra zeneca itu halal dan aman, MUI udh keluarkan fatwanya kok apalagi vaksinnya sudah lolos uji BPOM
Neutral	6 juta vaksin kembali tiba di Indonesia http://t.co/fEoHkc1r7

From the valid data, we divided the data into 3 datasets: training data, development data and testing data. A total of 400 tweets were set aside for testing, while the remaining data will be used for searching the best classification model. To optimize the model, they were split into 8,000 tweets for training and 746 for validation.

B. Preprocessing

The text preprocessing stage aims to prepare the data before the classification process [19]. In this research, several preprocessing stages are implemented, such as hyperlink removal (removing hyperlinks, a text started with *http://* or *https://*) and removing mention (for example "@budi0232", "@supermania", and others). We also remove numbers, punctuation, and hashtags from the tweets. Then, we apply case folding (converting all text to lowercase) and space cleaning (removing double spaces resulting from the previous steps).

C. Tokenization

The tokenization process is performed to break down text (tweets) into smaller units called tokens (words). This step is applied after text preprocessing to ensure that each word can be processed individually. Each tokenized word serves as the input for Word2Vec, which generates a set of word embeddings for all recognized words in the dataset.

Tokenization is also essential for preparing text data to train and predict sentiment using the GRU (Gated Recurrent Unit) architecture, as illustrated in Figure 2. By splitting the text into word tokens, the model can process and analyze patterns in sequences more effectively. Before feeding the words into the GRU model, they must first be converted into numerical representations, which allows for computational processing.

To achieve this, each word is transformed into a word vector, enabling mathematical operations necessary for classification. The word vectors are retrieved from the Word2Vec model by searching for their corresponding word IDs. These numerical representations capture semantic relationships between words, improving the model's ability to predict the sentiment classes.

D. Word2Vec Training

Word2Vec in its implementation has two algorithms namely CBOW (continuous bag of words) and Skip-Gram. The main advantage of using Word2Vec is its ability to reduce the dimension of the word vector, thus simplifying and improving computational efficiency. In other words, the Word2Vec model produces a vector representation that has lower dimensions compared to the traditional bag of words model, but retains important information about the semantic relations between words.

In this research, Skip-Gram is used. This model was developed as part of the Word2Vec technique, which aims to produce word embeddings (word representation vectors) that have semantic properties, meaning that words that are similar in context also have similar vectors. The Word2Vec embedding model retains all words seen in the dataset after preprocessing, storing them in tables with their corresponding word IDs and vector embeddings.

The words in a sentence S are then transformed into their respective word embeddings $\vec{v}_k$ (word vectors), and the sentence vector $\vec{S}$ is computed as their average, as equation 1, where n is the number of the words contained in a sentence.

$$\vec{S} = \frac{1}{n}\sum_{k=1}^{n} \vec{v}_k \qquad (1)$$

E. Data Balancing

The training and validation datasets were highly imbalanced, with the neutral class being the most dominant. The training data consisted of 6,664 neutral, 873 negative, and 463 positive tweets, while the validation data had a same distribution. The test data was designed to have a relatively equal distribution between neutral (58.5%) versus sentiment classes (41.5%), that is positive and negative combined (14.5% and 27%). The goal of this strategy is to better evaluate the performance of the classification method on an imbalanced dataset. The statistic of the dataset can be seen in Table 2.

Table 2. Dataset Distribution

Dataset	Positive	Neutral	Negative	Distribution of Class (%)
Train	463	6,664	873	5.8 : 83.3 : 10.9
Val	45	648	85	5.8 : 83.3 : 10.9
Test	58	234	108	14.5 : 58.5 : 27

To improve classification performance, we balance the dataset by oversampling the minority class and undersampling the majority class. The optimal class distribution is determined empirically before training the LSTM model. The model with the highest F1-score on validation data is selected as the final model.

F. Pad Sequences

In deep learning architectures, input dimensions must be consistent across all tweets in the dataset. Due to varying tweet lengths, padding is applied to ensure consistent input dimensions for the GRU model. Pad Sequences is a commonly used method for this purpose.

An array of size n is created to store word IDs representing the tweet. If a tweet has fewer than n words, empty slots are padded with zeros; if it exceeds n, the excess words are truncated. Padding can be applied at the beginning (initial padding) or the end (post-padding). This study adopts the initial padding method. Table 3 illustrates an example of the pad sequences process using this technique.

Table 3. The Example of Pad Sequences Process

Function	Result
Tokenizing	[ga, perlu, ragu, bro, vaksin, aztra, zeneca, itu, halal, dan, aman, mui, udh keluarkan, fatwanya, kok, apa, lagi, vaksinnya, sudah, lolos, uji, bpom]
Gathering word IDs	[7, 109, 371, 463, 1, 572, 573, 712, 493, 95, 557, 49, 29, 467, 202, 1098, 83, 75, 81, 19, 266, 31, 99]
Pad Sequencing	[**0, 0, 0, 0,….0, 0,** 7, 109, 371, 463, 1, 572, 573, 712, 493, 95, 557, 49, 29, 467, 202, 1098, 83, 75, 81, 19, 266, 31, 99]

G. Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network (RNN) architecture introduced by Cho et al. [20] as an alternative to Long Short-Term Memory (LSTM). GRU was developed to address the vanishing gradient problem that commonly occurs in standard RNNs when processing long sequential data.

GRU has a gating unit mechanism that controls the flow of information within the network. Unlike LSTM, which has three main gates (input gate, forget gate, and output gate), GRU has only two main gates:

- *Update Gate* (z_t): Controls how much information from the previous step is carried forward to the current step.
- *Reset Gate* (r_t): Controls how much information from the previous step is discarded.

As in Fig. 2, the operational of GRU block is formulated as follow:

Update Gate: $z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$

Reset Gate: $r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$

Candidate Activation: $\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t])$

Hidden State: $h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$

Where z is update gate, r is reset gate, h is hidden state, and W is weight matrix, σ is the sigmoid function, tanh is a hyperbolic tangent function, and t represents the time step.

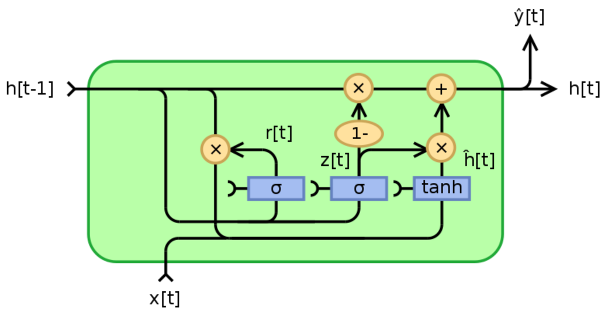


Fig. 2. GRU architecture[1]

In this research, we implement a Bidirectional Gated Recurrent Unit (Bi-GRU), as illustrated in Figure 3. Unlike a standard GRU, which processes sequential data in only one direction, Bi-GRU enhances contextual understanding by incorporating two separate hidden layers:

1. *Forward Layer* – Processes the sequence from the beginning to the end, capturing dependencies from past information.
2. *Backward Layer* – Processes the sequence in reverse, capturing dependencies from future information.

By combining information from both directions, Bi-GRU provides a more comprehensive representation of sequential data, improving performance in tasks that require a deeper understanding of context, such as sentiment analysis and natural language processing (NLP).

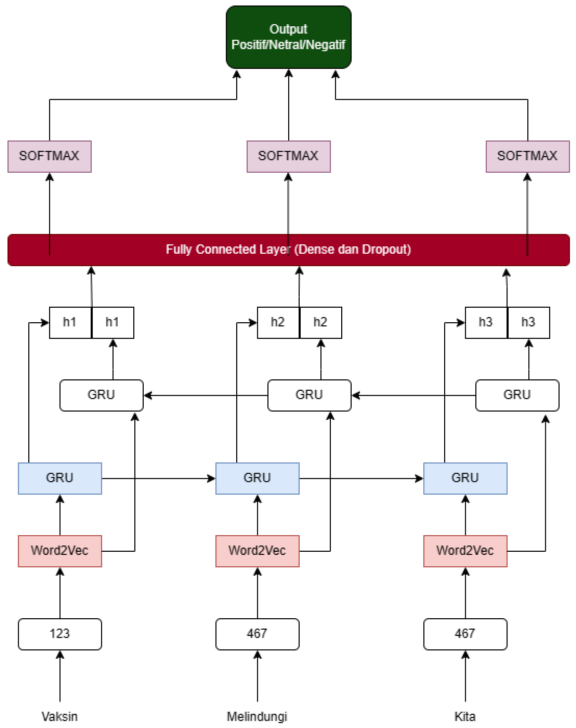


Fig. 3. Bi-GRU architecture for classification

The training process was conducted using the training dataset, while the validation dataset was used to evaluate model performance. Each tweet is limited to 100 words. Tweets exceeding this limit are truncated, while those with fewer than 100 words undergo initial padding. Training was performed over 50 epochs with a batch size of 64. Hyperparameter tuning was carried out to identify the most optimal classification model. In this study, the F1-score metric was used to determine the best-performing GRU-based classification model.

As illustrated in Fig. 3, the GRU process begins with tokenization, where each sentence is broken down into individual words and assigned an index. These tokenized words are then transformed into vector representations using Word2Vec, serving as input for the GRU model. The Bidirectional GRU (Bi-GRU) architecture is then applied, processing the word vectors through forward and backward layers to capture both past and future contextual dependencies. This dual-layer processing enhances the model's ability to understand the relationships between words, ultimately improving classification accuracy.

III. Result And Discusion

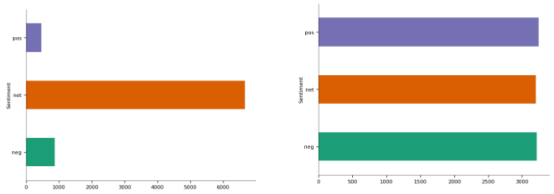
A. *Baseline and Data Balancing*

As explained above, this train and val data are extremely imlabance, where neutral class have more than 80% tweets compare to others. Machine learning systems often struggle with imbalanced data, favoring the dominant class while failing to detect positive and negative sentiments. This issue was also observed in our experiment. After training the GRU architecture with 8,000 tweets, this baseline model was tested on the validation dataset. However, the model's performance was poor, as it failed to classify sentiment correctly, as shown in Table 4. The precision, recall, and F1-score for both the negative and positive classes were 0, while the macro-average F1-score was only 30%.

Table 4. The baseline model performance on val data

Class	Recall	Precision	F1 Score
Negative	0	0	0
Neutral	1	0,84	0,91
Positive	0	0	0
Macro Average	0.33	0.28	0.30

The first step in optimizing the model is balancing the training data by reducing the size of the neutral class and the expanding the size of positve and negative classes. The reduction and expansion are determined empirically by experiment. Figures 4(a) and (b) illustrate the distribution of the unbalanced and balanced training datasets.



(a) (b)

Fig. 4 (a) Unbalanced composition of training data (b) Balanced training dataset composition (balanced dataset)

[1] https://paperswithcode.com/method/gru

After balancing the training data, the GRU model shows significant improvements in F1-score, particularly for the positive and negative classes, as shown in Table 5. The system is now better at recognizing and detecting sentiment. The precision, recall and F1-score increased by 33%, 40% and 37% respectively. The classification report of the model trained on the balanced dataset can be seen in Table 4.

Table 5. Balanced dataset performance on val data

Class	Recall	Precision	F1 Score
Negative	0.66	0.70	0.68
Neutral	0.76	0.67	0.71
Positive	0.56	0.68	0.61
Macro Average	0.66	0.68	0.67

B. *Parameter Tuning*

The next step in optimization is parameter tuning, which involves finding the optimal combination of hyperparameters that control the model training process. Unlike model parameters such as neural network weights that are updated during training, hyperparameters are predefined and remain unchanged. Key hyperparameters in deep learning include *learning rate*, *batch size*, *epochs*, *number of layers and units*, *dropout rate*, and *activation function*.

In this study, parameter optimization was performed as a second test to determine the optimal model configuration. The tuning process focused on two key parameters: *dropout rate* and the *number of nodes*. The *dropout rate* was selected from {0.1, 0.2, 0.3, 0.4}, representing the proportion of nodes randomly dropped during training. The *number of nodes* in fully connected layer (dense layer) was chosen from {32, 64, 128, 160, 256}, influencing model complexity. Table 6 presents the parameter tuning results for dropout and GRU node count on the validation data.

Table 6. Prediction Result on Val Data after Parameter Tuning

Number of Nodes	Drop Out	F1 Score	Accuracy
32	0,1	50%	63%
	0,2	53%	64%
	0,3	56%	66%
	0,4	57%	65%
64	0,1	55%	65%
	0,2	57%	59%
	0,3	59%	65%
	0,4	64%	**68%**
128	0,1	60%	66%
	0,2	59%	62%
	0,3	61%	64%
	0,4	61%	63%
160	0,1	52%	62%
	0,2	61%	65%
	0,3	**67%**	**68%**
	0,4	64%	**68%**
256	0,1	59%	66%
	0,2	58%	64%
	0,3	59%	64%
	0,4	62%	63%

The training process incorporates an early stopping mechanism to detect overfitting. When overfitting occurs, the system reverts to the model from the previous epoch, where validation accuracy was higher.

Table 6 presents the prediction results after the parameter tuning process on the validation data. The highest F1-score of 67% is achieved with 160 nodes, a dropout rate of 0.3, and word embeddings derived from the Covid dataset. This configuration is selected as the final model for testing.

C. *Test Data Testing*

The final optimal model is applied to the unseen test data, which was not used during GRU training or word2vec language model formation. The classification results for positive, negative, and neutral classes on 400 test tweets are presented in Figure 5 as a confusion matrix.

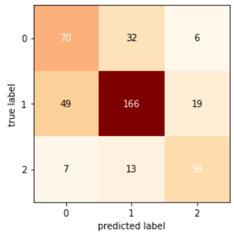


Fig. 3. Confusion Matrix of Test Data

These results demonstrate that the proposed GRU method achieves highly competitive performance in terms of F1-score, accuracy, precision, and recall compared to previous studies. It outperforms all other methods in term of F1-score, which serves as the official evaluation metric in this research, and ranks first in accuracy alongside SVM with FastText [17], as shown in Table 7.

Table 7. Comparison to Other Methods

Methods	Accuracy	F1 Score
Naïve Bayes [16]	61%	57%
SVM + TF.IDF[15]	65%	56%
SVM + Word2Vec[10]	68%	59%
LSTM + Word2Vec[8]	66%	54%
SVM + FastText[17]	**69%**	65%
XGBoost[12]	66%	57%
Logistic Regression[9]	67%	60%
KNN + FastText[11]	66%	57%
GRU + Word2Vec (This Research)	**69%**	**66%**

Despite its high computational complexity, GRU has demonstrated strong performance, even with a relatively small training dataset. This suggests that deep learning approaches, particularly GRU, can still be effective in sentiment classification tasks with limited data. However, one of the main challenges observed in this study is the model's difficulty in accurately detecting positive and negative sentiments. This issue arises due to an imbalance in the training data, where the neutral class dominates, and there is insufficient training data for tweets containing sentiment.

A lack of representative samples for positive and negative sentiment means the model is less exposed to sentiment-specific linguistic patterns, making it harder to generalize well. Consequently, the model tends to predict neutral sentiment more frequently, as it has learned from a larger proportion of neutral data. To address this limitation, expanding the dataset with more sentiment-rich tweets is crucial. By increasing the diversity and quantity of positive and negative samples, the model will gain a better understanding of sentiment-related features, leading to improved classification performance.

IV. Conclusion

the Gated Recurrent Unit (GRU) method demonstrates an effective performance for sentiment classification on Twitter. After balancing the training data and optimizing hyperparameters, the model can improve detection rate of positive and negative sentiments significantly compare to baseline. Testing on unseen data achieved an F1-score of 66% and accuracy of 69%, highlighting the model's strong generalization ability.

Future work with this method could explore various strategies to enhance sentiment detection, such as data augmentation techniques, more refined keyword selection for dataset collection, or leveraging pre-trained language models that have been trained on extensive sentiment-labeled corpora. These approaches should be able to mitigate the data imbalance issue and further optimize the model's performance in classifying tweets with positive and negative sentiments.

References

[1] A. Syauqi, "Jalan panjang covid19 (sebuah refleksi dikala wabah merajalela berdampak pada perekonomian)," *Jurnal Keuangan dan Perbankan Syariah.*, vol. 1, no. 1, pp. 1–19, 2020.

[2] J. Adhani, I. N. Yulita, A. Sholahuddin, M. N. Ardisasmita, and D. Agustian, "COVID-19 Social Safety Nets Sentiment Analysis On Twitter Using Gated Recurrent Unit (GRU) Method," in *2021 International Conference on Artificial Intelligence and Big Data Analytics*, Oct. 2021, pp. 100–103. doi: 10.1109/ICAIBDA53487.2021.9689737.

[3] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Indonesian Health and Management Journal vol 8 (2)*, 2020.

[4] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Information*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.

[5] V. N. T. Le, S. Ahderom, and K. Alameh, "Performances of the lbp based algorithm over cnn models for detecting crops and weeds with similar morphologies," *Sensors (Switzerland)*, vol. 20, no. 8, pp. 1–18, 2020, doi: 10.3390/s20082193.

[6] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *J. Dinda Data Science, Information Technology and Data Analysis*, vol. 1, no. 1, pp. 42–51, 2021, doi: 10.20895/dinda.v1i1.180.

[7] J. Patihullah and E. Winarko, "Hate Speech Detection for Indonesia Tweets Using Word Embedding And Gated Recurrent Unit," *IJCCS (Indonesian Journal of Computer Cybernetics System*, vol. 13, no. 1, p. 43, Jan. 2019, doi: 10.22146/ijccs.40125.

[8] M. Ihsan, B. S. Negara, and S. Agustian, "LSTM (Long Short Term Memory) for Sentiment COVID-19 Vaccine Classification on Twitter," *Digital Zone: J. Teknologi Informasi dan Komunikasi*, vol. 13, no. 1, pp. 79–89, 2022, doi: 10.31849/digitalzone.v13i1.9950.

[9] Ash Shiddicky and S. Agustian, "Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 pada Media Sosial Twitter menggunakan Metode Logistic Regression," *J. CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 99–106, 2022, doi: 10.37859/coscitech.v3i2.3836.

[10] M. Sahbuddin and S. Agustian, "Support Vector Machine Method with Word2vec for Covid-19 Vaccine Sentiment Classification on Twitter," *J. Informatics Telecommunication and Engineering.*, vol. 6, no. 1, pp. 288–297, 2022, doi: 10.31289/jite.v6i1.7534.

[11] A. Naldi and S. Agustian, "Klasifikasi Sentimen Vaksin Covid-19 Menggunakan K-Nearest Neighbor Berdasarkan Word Embeddings Fasttext pada Twitter," *Zonasi: J. Sistem. Informasi*, vol. 5, pp. 323–333, Jun. 2023, doi: 10.31849/zn.v5i2.12548.

[12] H. H. Sinaga and S. Agustian, "Pebandingan Metode Decision Tree dan XGBoost untuk Klasifikasi Sentimen Vaksin Covid-19 di Twitter," *J. Nasional. Teknololgi dan Sistem Informasi*, vol. 8, no. 3, pp. 107–114, 2022, doi: 10.25077/teknosi.v8i3.2022.107-114.

[13] Roihan, "Metode Decision Tree Dengan Fitur Fasttext Untuk Klasifikasi Sentimen Vaksin Covid-19 Pada Twitter," *Universitas Islam Negeri Sultan Syarif Kasim Riau*, 2023. [Online]. Available: http://repository.uin-suska.ac.id/74348/

[14] Ardi Mursyidi, "Penerapan Bidirectional Encoder Representations From Transformers (BERT) untuk Analisis Sentimen Vaksin Covid-19 pada Twitter," *Universitas Islam Negeri Sultan Syarif Kasim Riau*, 2023. [Online]. Available: http://repository.uin-suska.ac.id/74313/

[15] M. Rizki, "Analisis Sentimen Masyarakat Terhadap Vaksin Covid-19 Menggunakan Metode Support Vector Machine Pada Media Sosial Twitter," *Universitas Islam Negeri Sultan Syarif Kasim Riau*. [Online]. Available: http://repository.uin-suska.ac.id/id/eprint/58497

[16] P. Yohana, S. Agustian, and S. Kurnia Gusti, "Klasifikasi Sentimen Masyarakat terhadap Kebijakan Vaksin Covid-19 pada *Twitter* dengan Imbalance Classes Menggunakan Naive Bayes," in *SNTIKI (Seminar Nasional Teknologi Informasi, Komunikasi dan Industri)*, 2022.

[17] M. M. Kusairi and S. Agustian, "SVM Method with FastText Representation Feature for Classification

of Twitter Sentiments Regarding the Covid-19 Vaccination Program", *Digital Zone: J. Teknologi Informasi dan Komunikasi*, vol. 13 (1) 2022, pp. 140–150.

[18] R. Wardoyo, A. Musdholifah, G. Pradipta, and H. Sanjaya, "Weighted Majority Voting by Statistical Performance Analysis on Ensemble Multiclassifier". in *Fifth International Conference on Informatics and Computing (ICIC)* 2020. doi: 10.1109/ICIC50835.2020.9288552.

[19] C. J. E. Munthe, N. A. Hasibuan, and H. Hutabarat, "Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace," *Rekayasa Teknik Informatika dan Informasi*, vol. 2, no. 3, pp. 110–115, 2022, [Online]. Available: https://djournals.com/resolusi

[20] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, (2014), "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation", ArXiv: 1406.1078, doi: 10.48550/arXiv.1406.1078