JOIV : Int. J. Inform. Visualization, 9(6) - November 2025 2665-2673

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

COVID-19 Vaccine News Headline: A Sentiment Analysis with Robustly Optimized BERT Pretraining Approach (RoBERTa)

Qurrotul Aini [a,*], Kurnia Yuni Purwati [a], Elvi Fetrina [a], Yunifa Miftachul Arif [b]

[a] *Department of Information Systems, UIN Syarif Hidayatullah Jakarta, Indonesia*
[b] *Department of Electrical Engineering, UIN Maulana Malik Ibrahim Malang, Indonesia*
Corresponding author: [*]qurrotul.aini@uinjkt.ac.id

***Abstract*—The COVID-19 situation and its vaccination management necessitate disseminating information, especially news prevalent in Indonesian society. There is a method for classifying information by feelings. This research analyzes the sentiment of COVID-19 vaccine news headlines in Indonesia using the Robustly Optimized BERT Pretraining Approach. The focus is on news headlines from CNN Indonesia, which contain the word "COVID-19 Vaccine." The methods applied include data acquisition by web scraping and preprocessing. Later on, the RoBERTa model was used to classify sentiments under multiclass and binary conditions. The results showed that the sentiment of public opinion regarding the headlines in news releases was favorable rather than negative between January and December 2021. This research was done under two conditions: multiclass and binary. The training-to-testing data ratio was set to 80:20. In the initial scenario, with a multiclass dataset of 3,680 instances, the accuracy is 78.12%, the precision is 78.90%, the recall is 78.12%, and the F1-score is 78.36%. The second scenario, using a binary dataset of 3,466 entries, yielded an accuracy of 81.99%, precision of 84.42%, recall of 86.98%, and an F1-score of 85.68%. This study contributes to comprehending perceptions of the COVID-19 vaccine as portrayed in Indonesian news media. This study demonstrates the efficacy of the RoBERTa model for sentiment analysis, thereby facilitating the construction of precise and efficient NLP-based public opinion monitors. This study confirms the efficacy of the RoBERTa model in sentiment analysis.**

***Keywords*—Vaccine COVID-19; sentiment analysis; RoBERTa; BERT; news headline.**

Manuscript received 9 Jan. 2025; revised 24 Apr. 2025; accepted 26 Jul. 2025. Date of publication 30 Nov. 2025.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

I. INTRODUCTION

The COVID-19 pandemic has been a world-wide influence, with many nations affected, including Indonesia, with substantial news coverage in relation to the pandemic and related vaccine development. Immunization remains a key factor in influencing transmission in a community, hence protecting not only those who were vaccinated but also those in the community as a whole [1]. However, there are many challenges that are being faced in the distribution of COVID-19 vaccines in Indonesia, including their availability, issues related to quality, lack of substantial uptake among the population, and that this pandemic remains a continuous event. The distribution of news, medications, and vaccines related to a pandemic over the internet, including news, discussion forum, and articles, has been seen to influence global popular behavior. An individual's perspective usually differs as per their consumed news sources [2].

Although the conclusion of the vaccine distribution for COVID-19 has been reached, as well as that of the pandemic, public perception of this vaccine has been dynamic, being influenced by various developments that are taking place in relation to the immunization program, new variants, as well as a wide range of educational and informational programs offered by government and media sources. Systematic evaluation of public attitude towards the COVID-19 vaccine has become relevant for assessing various aspects of communication efficiency, as well as for identifying successes and challenges in relation to implementation of immunization programs. This study remains topical and significant for offering contemporary insights into how Indonesians perceive the COVID-19 vaccine, as well as for developing efficient strategies for communication in relation to the latter stages of this pandemic.

News plays a major role as a vital channel for information dissemination related to local events, especially for their capacities as potential issues for individuals, especially in

periods like the pandemic of COVID-19. On this premise, news organizations are able to systematically identify and examine emerging issues related to news content, especially in understanding issues related to the COVID-19 pandemic in Indonesian society. On the other hand, news platforms like Twitter and Instagram are most popular for performing tasks related to analyzing sentiments, especially due to their subjective nature. News platforms, however, are distinguished by their tendency to maintain objectivity in their contents [3]. This principle has been reinforced by studies that suggest news sources are involved in stringent validation processes for news release, sticking to ethical rules and news values to maintain objectivity of news sources [4]–[6].

News headlines are of great significance in detecting the major message as well as influencing readers in forming their initial perceptions of news information. They are usually framed with objective intentions of capturing readers' attention through a proper reflection of news information, thereby providing a proxy measure for public opinion and moods being analyzed. Previous studies support findings that confirm studies by [7] and [8] that news headlines are common sources of data in analyzing public moods, due to their efficiency in quickly capturing major viewpoints related to certain situations.

Natural Language Processing (NLP), this field of computer science aims for understanding natural languages by integrating Linguistics and Artificial Intelligence (AI) [9]. Some popular techniques used in NLP are named entity recognition, text analysis, machine translation, topic modeling, and text summarization techniques, among others, as classified in [10]–[13].

Classifiers like LogisticRegressionCV, LinearSVC, and Naive Bayes in traditional NLP usually disregard text order (context) and meanings of words in text. A major challenge for most text classification tasks in natural language processing, including Bag-of-words models, Word embeddings, or other traditional models, involves text order and/or text meanings [14]. This challenge arises due to models relying only on patterns of words without consideration for text meanings, hence a lack of a comprehensive understanding of text orders in classification tasks. To tackle this challenge in a text classification task like sentiment analysis, one of the strategies would involve leveraging a set of previously trained models whereby all of these models utilize transformers, with models including Bidirectional Encoder Representations from Transformers (BERT) [15], Robustly Optimized BERT Pretraining Approach (RoBERTa) [11], XLNet models [16], and GPT-2 and GPT-3. RoBERTa, though, can effectively handle larger amounts of data than that of BERT [3]. This follows since it has a different method in model training using dynamic masking. This implies that a different masking strategy will be used with every data point, hence inhibiting this model from being prone to memorizing patterns due to rigid patterns of masking used in model training [17], [18].

In this study, RoBERTa is determined to be the most developed model for sentiment analysis, obtained from improvement in BERT. This model has various advantages, which are additional training with bigger data, removal of the goal of predicting the next sentence, possession of capabilities for longer sequence training, in addition to variation of masking patterns in the training data. These benefits make it possible for RoBERTa to attain a state-of-the-art result in several benchmarks, such as GLUE, RACE, SQuAD, without multi-task tuning and additional data as in [19], [20]. Another point to note about RoBERTa is that it has been proven to be efficient in analyzing emotions with high accuracy in news headlines, with capabilities to understand the subtle context of news, hence suitable for applying it in analyzing the news of COVID-19 vaccines.

The research on analyzing news of the COVID-19 vaccine related to opinion analysis using CNN Indonesia, as part of applying the RoBERTa approach, has pointed out a gap in conventional opinion analysis using traditional classification in NLP, with no direct stress on news sources, and a lack of clarity in opinion interpretation in Indonesian news sources, in this case, CNN Indonesia. This work contributes to connecting a classification algorithm from RoBERTa for opinion analysis in news articles related to the COVID-19 vaccine in Indonesian news sources from CNN Indonesia. The research discusses analysis conditions in terms of multi-class and binary analysis.

II. Materials and Methods

This section describes the proposed methodology, performance, data sources, and design of this study.

A. Sentiment Analysis of News Headlines

Headlines serve as an important source of information for news, setting the tone of the news and influencing readers' opinion interpretation. The news text has been found to contribute to the potential expectations of public opinion, views, and media bias [21].

The analysis of news headlines as a task of natural language processing involves reading, understanding, and classifying news text in relation to the interpretation of emotions connoted in news text, identifying it as either positive, negative, or a group of news with a negative classification [22], [23]. This involves a series of tasks, which include data acquisition, preprocessing of news text, which involves cleansing of news text, developing a model, opinion classification, and evaluation of performance of opinion classification model used [24].

B. RoBERTa

The Robustly Optimised BERT Approach, also known as RoBERTa, was developed in 2019 as a form of improvement of BERT, which was introduced in 2018. BERT stands for "Bidirectional Encoder Representations from Transformers." The BERT model proposed various goals for training, including the prediction of the next sentence in a pair. This approach aimed to improve BERT's efficiency and accuracy in natural language processing. In developing RoBERTa, Liu et al. removed the task of predicting the next sentence in a pair, as experiments showed that this goal does not result in a significant improvement in performance. Another improvement in RoBERTa was that it used a larger and more diverse set of data than BERT, which allowed it to learn from a broader set of sources. The training of RoBERTa also involves using enhanced techniques, including increased batch sizes and dynamic masking, which help improve the model's understanding of the contexts. These features of RoBERTa result in a significant improvement in various

natural language processing tasks, including those related to text classification and sentiment analysis, making it a superior model in this domain as well [17]. RoBERTa further fine-tunes BERT by adjusting various training parameters, which result in performance superior to that of BERT for nearly all tasks that BERT was meant to accomplish. RoBERTa follows a similar structure as BERT. In contrast to BERT, RoBERTa exclusively employs masked language modelling throughout its pretraining phase [25].

The RoBERTa model training pseudocode [17] is presented as follows:

```
FUNCTION TrainRoBERTa(data):
    # 1. Collect data
    dataset=CollectDataset(["CC-NEWS", "OPENWEBTEXT",
      "BOOKCORPUS", "WIKIPEDIA"])
    # 2. Preprocess data
    preprocessed_data = PreprocessData(dataset)
    # 3. Setup model
    model=InitializeModel(architecture="BERT",
     layers=24,hidden_size=1024,  attention_heads=16)
    # 4. Set training parameter
    batch_size = 8192
    training_steps = 500000
    learning_rate = 1e-4
    # 5. Start training
    FOR step IN range(training_steps):
        # a. Get batch data
        batch=GetBatch(preprocessed_data, batch_size)
        # b. Apply dynamic masking
        masked_batch = DynamicMasking(batch)
        #c. Train model with masked language modeling
        loss = model.Train(masked_batch)
        # d. Update the model parameter
        .UpdateParameters(learning_rate)
    # 6. Model evaluation
    evaluation_results= EvaluateModel(model, ["GLUE",
      "SQuAD", "RACE"])
    # 7. Display evaluation result
    PrintResults(evaluation_results)
END FUNCTION

# Execution training function RoBERTa
TrainRoBERTa(data)
```

The mathematical model of the RoBERTa algorithm is as follows:

- Each text input is converted into a vector representation using embedding:

$$X = [x_1, x_2, \ldots, x_n] \tag{1}$$

  x_i is a vector representation of the *i*-th token in the input text.
- RoBERTa applies the Transformer architecture with an attention mechanism. The attention mechanism is expressed as follows:

$$Attention\ (Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

  where Q expresses a query matrix, K is a key matrix, V is a value matrix, and d_k expresses the key dimension.
- During pretraining, RoBERTa uses Masked Language Modeling (MLM). The pseudocode for MLM can be expressed as:

$$P(w_t|w_1, w_2, \ldots, w_n) = \text{softmax}(W_h h_t) \tag{3}$$

  where h_t represents the tokenized token and W_h is the learned weight.
- After pretraining, the model is fine-tuned on specific datasets (such as COVID-19 vaccine news headlines). The classification process is expressed as:

$$y_{\text{pred}} = \text{softmax}(W_{out} h) \tag{4}$$

C. Confusion Matrix

A confusion matrix is a structured table utilized to assess the performance of a classification algorithm [26]. The performance of the algorithm can be measured using a quantity known as accuracy. In a multi-class problem, a confusion matrix has been used, which contains a total of six components: true positive (TP), false positive (FP), true neutral (TNt), false neutral (TNt), true negative (TN), and false negative (FN). True positive describes a statement in which sentences with positive feelings are labeled as such. False positive describes cases in which sentences are labeled as positive when they are positive, but also labeled as either negative or neutral. True negative describes a statement in which sentences with a negative interpretation are labeled as such. False negative describes cases in which sentences with a negative message are labeled as either positive or negative. In the second scenario (binary dataset), the confusion matrix comprises four categories: true positive (TP), false positive (FP), true negative (TN), and false negative (FN) [24], [27]. The six categories (Table I) and calculation of accuracy up to F1-score in the first simulation are shown in (5) to (8).

TABLE I
CONFUSION MATRIX FIRST SCENARIO

		Predicted Label		
		Negative	**Positive**	**Neutral**
True Label	**Negative**	True Negative (TN)	False Positive (FP)	False Neutral (FNt)
	Positive	False Negative (FN)	True Positive (TP)	False Neutral (FNt)
	Neutral	False Negative (FNt)	False Positive (FNt)	True Neutral (TNt)

$$Accuracy = \frac{TP+TNt+TN}{TP+FP+TNt+FNt+TN+FN} \tag{5}$$

$$Precision_{positive} = \frac{TP}{TP+FP} \tag{6}$$

$$Recall_{positive} = \frac{TP}{TP+FN+FNt} \tag{7}$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{8}$$

The four categories (Table II) and calculation of accuracy up to F1-score in the second simulation are shown in (9) to (12).

TABLE III
CONFUSION MATRIX SECOND SCENARIO

		Predicted Label	
		Negative	**Positive**
True Label	**Negative**	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{9}$$

$$Precision_{positive} = \frac{TP}{TP+FP} \tag{10}$$

$$Recall_{positive} = \frac{TP}{TP+FN} \tag{11}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

D. Data Source

This study uses an Indonesian news headline containing the keyword "vaccine COVID-19" from the CNN Indonesia website, covering the period from January to December 2021, via a scraping technique in Node.js. The scraping outcomes are preserved as a dataset. This study solely uses news headlines to obtain 3,680 data points. Additionally, the data is segmented into sentences using a sentence segmentation method. The data has to be separated into sentences; that is, tokenized at the sentence level. After scraping and tokenization, we come up with a dataset of 3,680 sentences, which we save as an Excel file.

E. Research Method

This study proposes a research direction using Madani's framework [28], for predicting sentiment on CNN Indonesia's news portal using RoBERTa. Steps in the study conform to the phases indicated in Fig. 1.

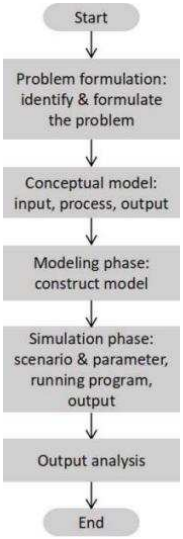


Fig. 1 Study stages

1) Problem formulation: Based on a previous study as a reference to sentiment analysis with RoBERTa, this study identifies the problem. RoBERTa and the results of its performance in the proposed model would come out with sentiment analysis of COVID-19 vaccine headlines. The previous experiments were quite accurate, and that is why RoBERTa is also applied in this study for sentiment analysis. From the invention of RoBERTa to this date, the graph is represented in Fig. 2.

Although this work has already conducted sentiment analysis with RoBERTa on Indonesian COVID-19 vaccine headlines, a number of gaps still remain to be explained. First, prior studies have focused mostly on sentiment analysis and have failed to touch upon how it changes with time-that is, the intensity and polarity of public opinion during the vaccination period. Furthermore, there also lacks a study comparing different categorization models in performing well for news in Indonesian, as the culture and language nuances greatly differ.

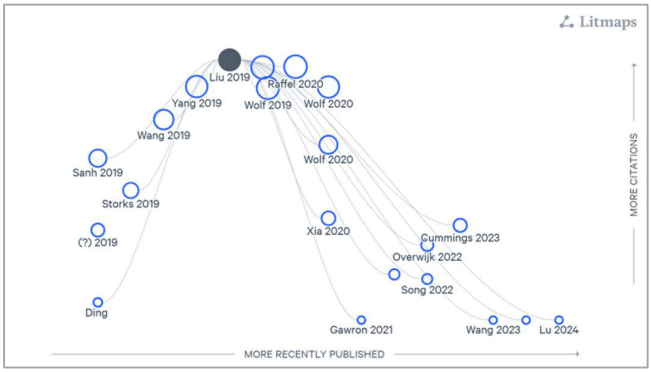


Fig. 2 Research growth of the RoBERTa approach

Therefore, this research needs to further outline these gaps with a view to defining concrete research subjects, for example:

- What is the pattern of change in Indonesian sentiment toward COVID-19 vaccine news from January to December 2021?
- To what extent can the RoBERTa model accurately classify Indonesian news sentiment in multiclass and binary scenarios?
- What are the factors that influence sentiment variation over the period?

This study aims to fill this gap in available literature on news headline analysis for sentiment in an Indonesian context, as well as provide a comprehensive insight into community attitude towards the COVID-19 vaccine through a precise and efficient analysis of news headlines.

2) The conceptual model: This phase describes the key idea of the research, which involves concepts of input, processing, and output. This model involves data scraping, data annotation, and the RoBERTa model. The research adopts data scraping techniques for extracting news headlines from the CNN website. This data will be further annotated for performing sentiment classification. Also, data preprocessing follows classification using RoBERTa. The data flow description of the sentiment analysis task is shown in Fig. 3.

3) Collection input/output data analysis: The news headlines were extracted from the CNN Indonesia website using a keyword “COVID-19 Vaccine” for input data from January 1, 2021, up to December 31, 2021. The data from the website has been classified in terms of date, headlines, and URL. This data set has been used as input data after processing with sentiment analysis.

4) Modeling phase: The input data, which has been derived from the previous step, goes through data preprocessing in which the unstructured data is converted to a structured format to improve the processing of the upcoming research data. Prior to data processing, data representation in form of diagrams has been used to explain the complexities of the research data. The RoBERTa model evaluates data, tokenizes it, and splits it into test, training, and validation sets.

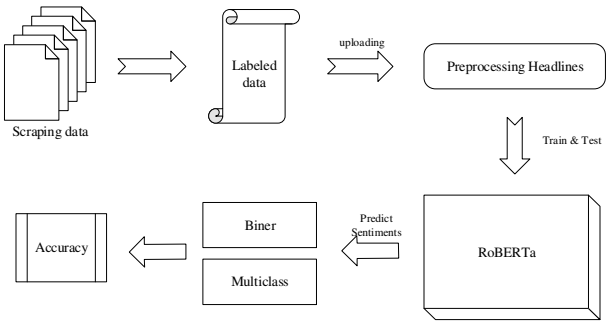


Fig. 3 Sentiment analysis workflow

Data tokenization was performed using the RoBERTa tokenizer from the HuggingFace Transformers Library. The major parameters used were:

- Max length: 512 tokens, as per the standard RoBERTa model.
- Padding and truncation: Data shorter than 512 tokens are padded, and those longer are truncated.
- Label adjustment: Labels are converted to numerical indices (Positive=1, Negative=0, Neutral=2) for classification.

5) Simulation phase: This step simulate the process of running the developed model. This topic will be elaborated in the next point. This step focuses on specifying the simulation parameters and requirements, which are then coded in a computer program.

6) Output analysis: This final stage involves analyzing the performance of the model for accuracy in all scenarios that were run.

III. Results and Discussion

A. Data Preprocessing

This study draws on the data of CNN Indonesia's coverage in 2021 and focuses on the phrase "COVID-19 Vaccine." The data scraped will form a sentiment analysis dataset. News headlines were crawled by Node.js; 3,680 headlines were obtained correctly. A sentence segmentation step splits the data into individual sentences.

The authors perform a manual check of their labeling against the COVID-19 vaccination headlines dataset. After labeling, sorting into positive, negative, and neutral attitudes took place, after which they proceeded with preprocessing. They did URL link removal at this stage to simplify the subsequent processing. The steps involved in the preprocessing phase include data cleansing, case normalization, stemming, and stopword removal.

Preprocessing cleanses and prepares the text data for use in RoBERTa-based NLP applications. Specific activities included are:

- Removing links and unifying data: Removing URLs and links prevent irrelevant text from distorting the sentiment analysis.
- Splitting sentences: Headlines can comprise several sentences; therefore, a splitting process is carried out based on periods (.), exclamation marks (!), or question marks (?) to separate the sentences accurately.
- Tokenizing and lowercasing: Since RoBERTa handles tokenization through its tokenizer, text is uniformly converted to lowercase.
- Removing stopword: Common words without specific sentiment meaning, like “yang,” “dan,” “di,” are removed using a precompiled Indonesian stopword list per NLP standards.
- Stemming: Words are processed by a stemming algorithm, such as SnowballStemmer for Indonesian, to allow the use of root forms. Such preprocessing makes for more consistent processing of words-such as "bervaksin" and "vaksinasi", both of which stem to "vaksin".
- Cleaning special characters and numbers: Special characters, numbers, and extra whitespace are removed in order to keep the text clean and focused on its main content.

It was done using Python with the NLTK library and regular expressions to make sure the input of the model is clean and unified.

The dataset classification presents the aggregate number of datasets. The initial scenario categorized the dataset into negative, positive, and neutral. The histogram's multiclass visualization indicates that neutral sentiment comprises the least data (Fig. 4). Figure 4 illustrates that the multiclass datasets consist of 2,134 positive sentiment labels, 1,332 negative sentiment labels, and 214 neutral sentiment labels.

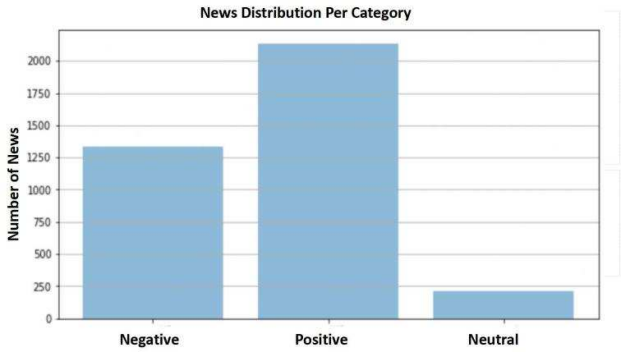


Fig. 4 Visualization of a multiclass dataset

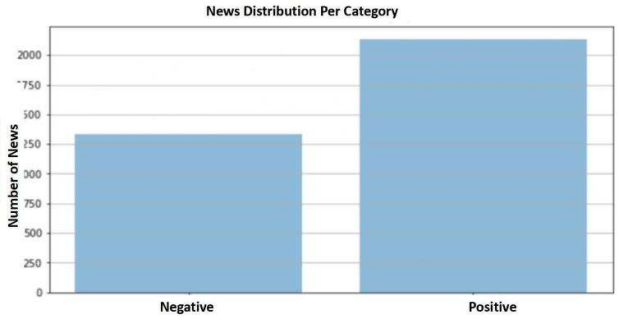


Fig. 5 Visualization of a binary dataset

Figure 5 presents a dataset for the second situation, a binary classification task with 2,134 positive and 1,332 negative sentiment labels. Figure 6 demonstrates that Indonesian news headline sentiment around the COVID-19 vaccine intensified from January 1 to December 1, 2021. In May, adverse sentiment intensified, affecting public perception and

program compliance. Conversely, good mood increased in July, indicating the initiative's heightened vigor. It may help the government achieve herd immunity goals and control the pandemic.

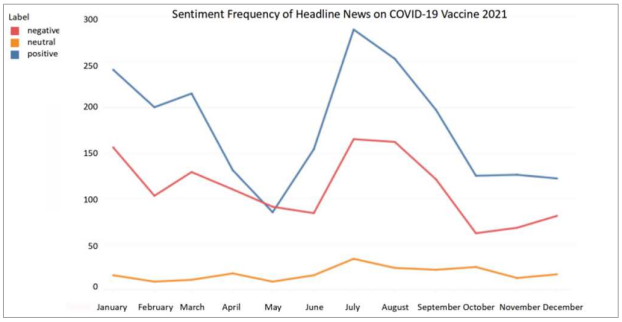


Fig. 6 Sentiment frequency of news headlines

Word clouds are visual depictions of word frequencies, in which more frequently occurring words are shown in larger sizes. The size of a word in the graphic indicates its frequency and prevalence in the document. Tables III–V display the terms in the word clouds for each sentiment category, organized from most to least common by visual prominence.

TABLE IIIII
NEGATIVE SENTIMENT KEYWORDS

No.	Keyword (ID)	Translation (EN)
1	batal	canceled
2	tolak	reject
3	hoaks	hoax
4	efek	effect
5	meninggal	died
6	takut	afraid
7	gagal	failed
8	bahaya	danger
9	isu	issue
10	risiko	risk

TABLE IVV
POSITIVE SENTIMENT KEYWORDS

No.	Keyword (ID)	Translation (EN)
1	vaksin	vaccine
2	aman	safe
3	efektif	effective
4	lindungi	protect
5	sehat	healthy
6	dukung	support
7	berhasil	successful
8	gratis	Free (of charge)
9	distribusi	distribution
10	heard	Herd (in immunity)

TABLE V
NEUTRAL SENTIMENT KEYWORDS

No.	Keyword (ID)	Translation (EN)
1	covid	COVID
2	indonesia	Indonesia
3	presiden	president
4	pemerintah	government
5	jadwal	schedule
6	data	data
7	berita	news
8	update	update
9	jumlah	number
10	kasus	case

B. *Data Splitting*

This study encompasses three sentiment categories—positive, negative, and neutral—in the first simulation situation and two—positive and negative—in the second. The dataset was subsequently divided into an 80:20 ratio, as in a previous study [29], meaning 80% of the data for training and 20% for testing. Another study [30] indicated that an 80% training and 20% testing ratio produced optimal performance, while another study [31] further asserted that the 80:20 ratio is regarded as a more precise division.

Prior to the start of model classification, data splitting into training, validation, and test data takes place. Data validation is used in this step to validate all model training processes. Overfitting takes place when complications exist in data training, marked by a situation where the value of loss does not decrease and the accuracy fails to increase. Data validation eliminates overfitting, which in most cases takes place when complications exist in data training due to differences between data in training and actual data. Data testing, on the other hand, provides a rigorous validation of model correctness that has been trained using training data.

C. *Applying RoBERTa*

Using RoBERTa through the transformers library, which is offered by HuggingFace and provides a range of pre-trained models. After pre-training on existing datasets, the training process is divided into three components: training, validation, and testing.

Concerning the fine-tuning procedure utilizing hyperparameters, the subsequent recommendations are offered for RoBERTa [32].

- Batch size: 8 and 16;
- Epoch: 2 to 10;
- Learning rate: 1e-5 Adam (0.005).

Various factors influence hyperparameter selection. Batch sizes of 8 and 16 were selected because of the correlation between greater batch sizes and extended completion times for each batch [31]. The use of a learning rate 1e-5 Adam was chosen, as in previous studies [3], [17], [33], [34]. Subsequently, several simulations were done over epochs 1 to 10, revealing that 8 epochs with a batch size 16 attained the highest accuracy.

The example of simulated training results of the multiclass dataset with 8 epochs and 16 batch sizes is shown below:

```
Training...
Epoch 1/8
184/184 [==============================] - 184s
951ms/step - loss: 0.8549 - accuracy: 0.6393 -
val_loss: 0.6979 - val_accuracy: 0.7079
Epoch 2/8
184/184 [==============================] - 181s
986ms/step - loss: 0.5655 - accuracy: 0.7554 -
val_loss: 0.5809 - val_accuracy: 0.7486
⋮
Epoch 8/8
184/184 [==============================] - 181s
986ms/step - loss: 0.1624 - accuracy: 0.9429 -
val_loss: 0.8525 - val_accuracy: 0.7812
```

After several iterations per epoch and batch, the model results with the highest accuracy are retained, and the accuracy is documented. Figures 7 and 8 show a graphical comparison of accuracy during training and validation across

two experimental conditions with batch sizes of 16 and 8, respectively.

Figure 7 above shows how accurately a multiclass classification model performed in the course of training. The results indicate that there was a substantial improvement in how accurately the model performed on both the training and validation sets as it progressed from epoch to epoch, with a near 95% rate for the training set. However, as it progressed, it seemed to settle on a 75% validation rate. There seems to be a discrepancy between how well it performed on both sets of data, with a tendency to overfit. This can also be seen in a binary classification set in Fig. 8.

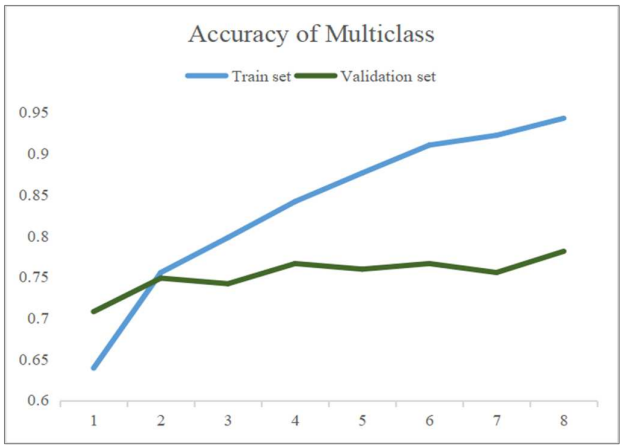


Fig. 7 Training and validation performance on multiclass datasets

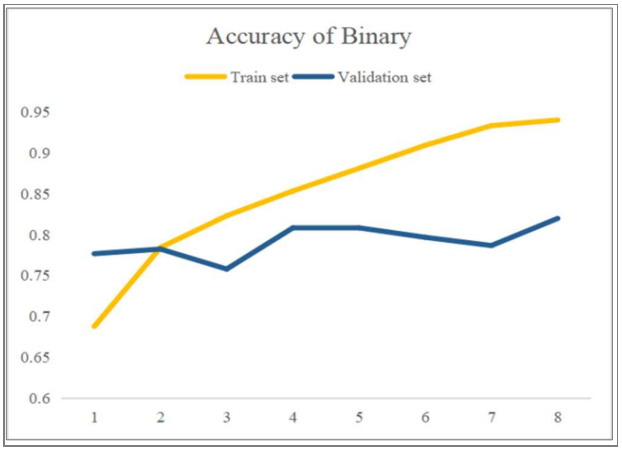


Fig. 8 Training and validation performance on binary datasets

D. Output Analysis

The categorization results demonstrate that positive emotion has predominated in COVID-19 vaccine news on the CNN Indonesia National channel website for one year. This favorable feeling signals substantial positive news and widespread public excitement about vaccines in Indonesia's efforts to combat the pandemic.

Figure 9 illustrates six categories inside the initial scenario confusion matrix: TP, FP, TNt, FNt, TN, and FN. In the second case, as seen in Fig. 10, the confusion matrix comprises four categories: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

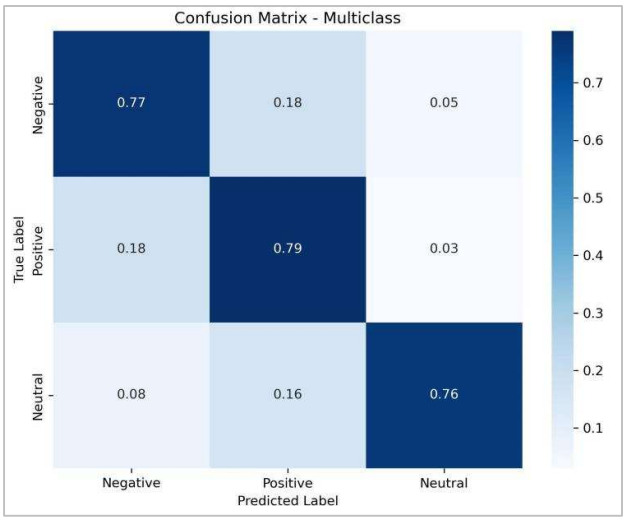


Fig. 9 Confusion Matrix of RoBERTa with Multiclass Dataset

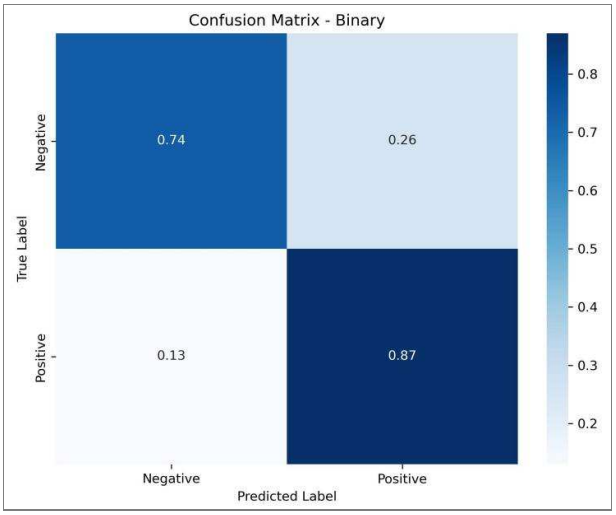


Fig. 10 Confusion Matrix of RoBERTa with Binary Dataset

The confusion matrices in both instances reveal that RoBERTa accurately classifies sentiment, as shown in Fig. 10. In Fig. 9, RoBERTa struggles to classify neutral sentiment. This results from a significant imbalance in the data, with the quantity of neutral-class data considerably less than that of the positive and negative classes. Furthermore, the ambiguous nature of words in a neutral class, along with a lack of specific features, makes it even tougher for the model to pick out distinct patterns. The heatmap shows that there is a substantial amount of data in the neutral class that has been misclassified either as a positive class or as a negative class, which further proves that this class requires a different approach, including data balancing and/or data augmentation.

Precision, recall, and F1-score are evaluation metrics that are often employed in order to test models for sentiment analysis, able to measure a model's performance in a manner that goes beyond accuracy. The F1-score can be seen as a superior measure of performance in a model as opposed to precision and recall. Table VI shows the performance of both cases with RoBERTa.

TABLE VI
PERFORMANCE OF ROBERTA WITH A MULTICLASS DATASET

Batch size	Epoch	Accuracy	Precision	Recall	F1-score
8	2	71.88	72.06	71.88	70.36
8	3	74.46	74.96	74.46	73.83
8	4	73.37	78.02	73.37	73.59
8	5	75.41	76.10	75.41	75.50
8	6	77.17	77.50	77.17	76.40
8	7	76.49	76.30	76.49	76.20
8	8	75.54	75.41	75.54	74.89
8	9	75.00	78.04	75.00	75.19
8	10	76.09	76.43	76.09	76.18
16	2	73.91	74.06	73.91	72.88
16	3	72.69	73.56	72.69	72.56
16	4	75.41	75.04	75.41	74.32
16	5	75.14	74.67	75.14	74.73
16	6	75.95	76.29	75.95	76.07
16	7	76.36	76.69	76.36	75.67
16	**8**	**78.12**	**78.90**	**78.12**	**78.36**
16	9	75.68	77.00	75.68	75.92
16	10	75.27	75.41	75.27	73.79

Table VI illustrates a gradual improvement in performance metrics with increasing epochs as well as batch sizes, including accuracy, precision, recall, and F1 measurement. The classifier reaches a point of maximized accuracy of 78.12% in the 10th epoch with a batch size of 16, which suggests that increased training periods as well as bigger batch sizes can help improve the classification capabilities of the model. Other performance metrics, including precision and recall, also report a consistent improvement.

Table VII also shows a gradual enhancement in performance metrics such as accuracy, precision, recall, and F1-score with an increase in epochs.

TABLE VII
PERFORMANCE OF ROBERTA WITH BINARY DATASET

Batch size	Epoch	Accuracy	Precision	Recall	F1-score
8	2	78.39	77.56	91.63	84.01
8	3	81.41	84.60	85.58	85.09
8	4	81.12	85.18	84.19	84.68
8	5	80.55	83.60	85.35	84.46
8	6	81.56	**87.01**	82.56	84.73
8	7	81.12	83.45	86.74	85.06
8	8	79.68	79.67	90.23	84.62
8	9	81.27	80.36	**92.33**	**85.93**
8	10	81.70	86.68	83.26	84.93
16	2	80.69	82.17	87.91	84.94
16	3	80.98	84.33	85.12	84.72
16	4	80.98	80.41	91.63	85.65
16	5	78.96	86.60	78.14	82.15
16	6	81.70	84.20	86.74	85.45
16	7	81.41	81.55	90.47	85.78
16	**8**	**81.99**	84.42	86.98	85.68
16	9	80.55	85.37	82.79	84.06
16	10	78.67	83.41	81.86	82.63

The highest accuracy of 81.99% and recall of 92.33% were obtained by this model in epoch 9 with a batch size of 16, which clearly reflects how effective this model was in identifying instances from the positive class. These results indicate that this model, named RoBERTa, works well in performing binary classification, as it has a good accuracy as well as a good recall value for the positive class.

E. Discussion

There is a population gap when compared to other studies done in the past. This is due to the amount of data used. In study [3], a total of 10,727 news headlines were used for analysis, while in study [33], 1,000 FEVER (Fact Extraction and Verification) datasets were used. This study also used a total of 3,680 news headlines in Indonesian. In a previous study, RoBERTa obtained a 90% accuracy validation with a set of hyperparameters that consisted of 3 epochs, a learning rate of 1e-5, and a batch size of 16 for analysis of news related to COVID-19, particularly in classifying news headlines into positive and negative ones. In this study, there is an approximation of 82% in binary classification. Another study, in [35], used FEVER in determining BERT, RoBERTa, and Electra models in terms of their accuracy and F1-Score. The result showed that RoBERTa obtained a 95.4% and 95.3% value for accuracy and F1-Score, respectively, using a set of hyperparameters that consisted of 5 epochs and a batch size of 32. Different hyperparameters are usually used in different studies for better result performance.

IV. CONCLUSION

This research shows that public sentiment towards COVID-19 vaccine coverage in CNN Indonesia news was mostly positive, with a higher number of positive news than negative news from January to December 2021. The classification result of news headlines using RoBERTa model showed a high accuracy of 81.99% in classifying news as either positive, negative, and/or neutral. These findings suggest that public attitude towards the COVID-19 vaccine becomes progressively optimistic, which can result in a growing interest in vaccination. The research, however, admits that it has certain limitations. Firstly, it only used news from a sole news source, which may not be a true reflection of public attitude in general. Secondly, it only analyzed news headlines, which lack comprehensive information that news articles can offer. Lastly, it only analyzed news data for a period of one year, which can only provide a limited perspective of data.

Further studies are needed in order to examine news data in a broader way, which involves articles as well as comments, in order to gain further insights into the trends of opinion sentiment as the news situation with regards to the pandemic unfolds in Indonesia. Further application of deep learning techniques might help improve upon or provide a comparison of the findings derived from using RoBERTa. Other models like XLNet and DistilBERT can provide different insights for opinion analysis. The project aims to further investigate ensemble methods of applying various deep learning models. Different models put together might provide a better solution for opinion analysis, which can provide a robust model for different types of data.

REFERENCES

[1] B. N. Harapan, T. Harapan, L. Theodora, and N. A. Anantama, "From archipelago to pandemic battleground: Unveiling Indonesia's COVID-19 crisis," *J. Epidemiol. Glob. Health*, vol. 13, pp. 591–603, 2023, doi:10.1007/s44197-023-00148-7.

[2] S. Vakhare, A. Yadav, A. Kannaujiya, and A. Pandey, "A framework for sentiment analysis of online news articles," *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)*, vol. 12, no. IX, pp. 1035–1041, Nov. 2024, doi: 10.22214/ijraset.2024.65276.

[3] P. Ghasiya and K. Okamura, "Investigating COVID-19 news across four nations: A topic modeling and sentiment analysis approach," *IEEE Access*, vol. 9, pp. 36645–36656, 2021, doi:10.1109/access.2021.3062875.

[4] M. G. Lopez *et al.*, "Making newsworthy news: The integral role of creativity and verification in the human information behavior that drives news story creation," *J. Assoc. Inf. Sci. Technol.*, vol. 73, no. 10, pp. 1445–1460, Mar. 2022, doi: 10.1002/asi.24647.

[5] M.-F. De-Lima-Santos, W. N. Yeung, and T. Dodds, "Guiding the way: A comprehensive examination of AI guidelines in global media," *AI & Soc.*, Jul. 2024, doi: 10.1007/s00146-024-01973-5.

[6] L. Dierickx *et al.*, "A data-centric approach for ethical and trustworthy AI in journalism," *Ethics Inf. Technol.*, vol. 26, no. 4, Sep. 2024, doi:10.1007/s10676-024-09801-6.

[7] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, May 2012, doi:10.2200/s00416ed1v01y201204hlt016.

[8] M. Osama, A. Dey, K. Ahmed, and M. A. Kabir, "BeliN: A novel corpus for Bengali religious news headline generation using contextual feature fusion," *Nat. Lang. Process. J.*, vol. 8, p. 100138, Mar. 2025, doi: 10.1016/j.nlp.2025.100138.

[9] T. Huynh-The *et al.*, "Artificial intelligence for the metaverse: A survey," *Eng. Appl. Artif. Intell.*, vol. 117, p. 105581, Jan. 2023, doi:10.1016/j.engappai.2022.105581.

[10] G. W. Wicaksono *et al.*, "Text summarization on verdicts of industrial relations disputes using the cross-latent semantic analysis and long short-term memory," *JOIV: Int. J. Informat. Vis.*, vol. 7, no. 3, pp. 847–853, Sep. 2023, doi: 10.30630/joiv.7.3.2052.

[11] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2020, doi:10.1017/9781108639286.

[12] Y. Zhang and G. Xiao, "Named entity recognition datasets: A classification framework," *Int. J. Comput. Intell. Syst.*, vol. 17, p. 71, Mar. 2024, doi: 10.1007/s44196-024-00456-1.

[13] S. Rouhani and F. Mozaffari, "Sentiment analysis researches story," *Soc. Sci. Humanit. Open*, vol. 6, p. 100309, Jan. 2022, doi:10.1016/j.ssaho.2022.100309.

[14] T. T. Khoei and N. Kaabouch, "Machine learning: Models, challenges, and research directions," *Future Internet*, vol. 15, no. 10, p. 332, Oct. 2023, doi: 10.3390/fi15100332.

[15] B. Juarto and A. S. Girsang, "Neural collaborative with sentence BERT for news recommender system," *JOIV: Int. J. Informat. Vis.*, vol. 5, no. 4, pp. 448–455, Dec. 2021, doi: 10.30630/joiv.5.4.678.

[16] X. Han *et al.*, "Pre-trained models: Past, present and future," *AI Open*, vol. 2, pp. 225–250, Jan. 2021, doi: 10.1016/j.aiopen.2021.08.002.

[17] A. J. Lak *et al.*, "RoBERTa, ResNeXt and BiLSTM with self-attention: The ultimate trio for customer sentiment analysis," *Appl. Soft Comput.*, vol. 164, p. 112018, Oct. 2024, doi: 10.1016/j.asoc.2024.112018.

[18] A. Jose and S. Harikumar, "Predicting IMDB movie ratings using ROBERTA embeddings and neural networks," in *Lect. Notes Electr. Eng.*, vol. 900, 2022, pp. 181–189, doi: 10.1007/978-981-19-4453-6_13.

[19] Y. Liu *et al.*, "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv*, Jul. 26, 2019. [Online]. Available: https://arxiv.org/abs/1907.11692.

[20] M. Mosbach *et al.*, "A closer look at linguistic knowledge in masked language models: The case of relative clauses in American English," in *Proc. 28th Int. Conf. Comput. Linguistics (COLING)*, 2020, pp. 771–787, doi: 10.18653/v1/2020.coling-main.67.

[21] D. Rozado, R. Hughes, and J. Halberstadt, "Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with transformer language models," *PLoS ONE*, vol. 17, no. 10, p. e0276367, Oct. 2022, doi: 10.1371/journal.pone.0276367.

[22] M. Vicari and M. Gaspari, "Analysis of news sentiments using natural language processing and deep learning," *AI & Soc.*, vol. 36, no. 3, pp. 931–937, Nov. 2020, doi: 10.1007/s00146-020-01111-x.

[23] S. Supriyono *et al.*, "Advancements in natural language processing: Implications, challenges, and future directions," *Telemat. Informat. Rep.*, vol. 16, p. 100173, Dec. 2024, doi: 10.1016/j.teler.2024.100173.

[24] Q. Aini, R. R. Fauzi, and E. Khudzaeva, "Economic impact due COVID-19 pandemic: Sentiment analysis on Twitter using naive Bayes classifier and support vector machine," *JOIV: Int. J. Informat. Vis.*, vol. 7, no. 3, pp. 733–741, Sep. 2023, doi: 10.30630/joiv.7.3.1474.

[25] P. López-Úbeda *et al.*, "Role of natural language processing in automatic detection of unexpected findings in radiology reports: A comparative study of RoBERTa, CNN, and ChatGPT," *Acad. Radiol.*, vol. 31, no. 12, pp. 4833–4842, Dec. 2024, doi:10.1016/j.acra.2024.07.057.

[26] E. Wahyudi *et al.*, "The implementation of the K-Medoid clustering for grouping hearing loss function on excessive smartphone use," *JOIV: Int. J. Informat. Vis.*, vol. 7, no. 4, pp. 2523–2530, Dec. 2023, doi: 10.62527/joiv.7.4.1873.

[27] A. A. Ajhari, "The comparison of sentiment analysis of moon knight movie reviews between multinomial naive Bayes and support vector machine," *Appl. Inf. Syst. Manag.*, vol. 6, no. 1, pp. 13–20, Apr. 2023, doi: 10.15408/aism.v6i1.26045.

[28] A. Gandomi, V. A. Tafti, and H. Ghasemzadeh, "Wireless sensor networks modeling and simulation in Visualsense," in *Proc. 2nd Int. Conf. Comput. Res. Develop. (ICCRD)*, Kuala Lumpur, Malaysia, 2010, pp. 251–254, doi: 10.1109/ICCRD.2010.72.

[29] V. R. Joseph, "Optimal ratio for data splitting," *Stat. Anal. Data Mining: ASA Data Sci. J.*, vol. 15, no. 4, pp. 531–538, Apr. 2022, doi:10.1002/sam.11583.

[30] F. Demir, "DeepCoroNet: A deep LSTM approach for automated detection of COVID-19 cases from chest X-ray images," *Appl. Soft Comput.*, vol. 103, p. 107160, May 2021, doi:10.1016/j.asoc.2021.107160.

[31] R. B. Hani *et al.*, "Malicious URL detection using machine learning," in *Proc. 15th Int. Conf. Inf. Commun. Syst. (ICICS)*, Irbid, Jordan, 2024, pp. 1–5, doi: 10.1109/ICICS63486.2024.10638299.

[32] J. A. Ilemobayo *et al.*, "Hyperparameter tuning in machine learning: A comprehensive review," *J. Eng. Res. Rep.*, vol. 26, no. 6, pp. 388–395, Jun. 2024, doi: 10.9734/jerr/2024/v26i61188.

[33] A. Murarka, B. Radhakrishnan, and S. Ravichandran, "Detection and classification of mental illnesses on social media using RoBERTa," *arXiv*, Jan. 2020. [Online]. Available: https://arxiv.org/abs/2011.11226.

[34] M. Zhu, Y. Song, G. Jin, and K. Jiang, "Identifying personal experience tweets of medication effects using pre-trained RoBERTa language model and its updating," in *Proc. 11th Int. Workshop Health Text Mining Inf. Anal. (LOUHI)*, 2020, pp. 127–137, doi:10.18653/v1/2020.louhi-1.14.

[35] M. Naseer, M. Asvial, and R. F. Sari, "An empirical comparison of BERT, RoBERTa, and Electra for fact verification," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Jeju Island, South Korea, Apr. 2021, pp. 241–246, doi: 10.1109/ICAIIC51459.2021.9415192.