Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 10, No. 2, June 2024, pp. 419-434
ISSN: 2338-3070, DOI: 10.26555/jiteki.v10i2.29185

Genetic Algorithm and GloVe for Information Credibility Detection Using Recurrent Neural Networks on Social Media Twitter (X)

Andi Nailul Izzah Ramadhani, Erwin Budi Setiawan
Telkom University, Jl. Terusan Buah Batu, Bandung 40257, Indonesia

ARTICLE INFO

Article history:

Received June 20, 2024
Revised July 05, 2024
Published July 17, 2024

Keywords:

BERT;
Genetic Algorithm;
GloVe;
Information Credibility;
Recurrent Neural Network;
TF-IDF

ABSTRACT

Social media, especially X, has become a key source of information for many individuals, but the level of trust in the information spread on these platforms is a critical issue. To overcome this problem, this research proposed an information credibility detection system using a Recurrent Neural Network (RNN) with the utilization of TF-IDF feature extraction, GloVe feature expansion, BERT word embedding, and Genetic Algorithm (GA) optimization. This research contributes to assessing the credibility of tweets related to the 2024 Indonesian election by integrating TF-IDF to identify important words, GloVe to enhance word context, BERT for deeper understanding, and GA is present to optimize RNN performance. The main focus is to provide maximum accuracy by integrating these methods. In this research, the dataset used consists of 54,766 tweets relating to the 2024 Indonesia election and includes relatively equal numbers of credible and non-credible labels. The corpus construction utilized source X with a total of 40,466 data, IndoNews with a total of 131,580, and a combination of both with a total of 150,943. This research conducted six experimental scenarios, namely optimal data split, max features, N-grams, Top-N rank similarity corpus, BERT and GA application. Through these scenarios, the model achieved a significant accuracy improvement of 1.81% over the baseline, reaching an accuracy of 90.60%. This result demonstrates the effectiveness of the proposed system by presenting a higher quality of accuracy compared to the baseline model. Moreover, this research underscores the significant contribution of increasing the accuracy of information credibility detection.

Corresponding Author:

Erwin Budi Setiawan, Telkom University, Jl. Terusan Buah Batu, Bandung 40257, Indonesia
Email: erwinbudisetiawan@telkomuniversity.ac.id

1. INTRODUCTION

In today's technological development, social media has become the main platform for accessing information and gathering news from around the world, but its credibility is still questionable [1]. Credibility means the amount of trust and confidence we can gain when reading news from social media sources. The accuracy and trustworthiness of a news story require fact-checking so that it can be used to determine its credibility [2]. Credibility issues become more serious in non-routine and urgent situations [3]. With so much information available, credibility is the most important part of information dissemination in society [4]. Social media has become the main source of news and information for many people, but not all information circulating can be trusted [5]. Twitter (X) is a social media platform founded by Jack Dorsey in 2006. Users can send messages in various formats, such as text, photos, videos, or links. Twitter facilitates its users to communicate by sending and reading messages in the form of text or so-called tweets [6], [7]. The active role of social media platforms in addressing misinformation is an important factor in maintaining the integrity of information in circulation. Social media not only facilitates interaction and communication [8] but also carries the risk of spreading non-credible information among the public. Therefore, information credibility detection is needed to address the problem of spreading non-credible information on social media. Moreover, the use of credibility detection algorithms can help identify fake or misleading news more efficiently.

Research related to information credibility detection has been conducted using either machine learning or deep learning approaches to automatically distinguish false information from credible information [9], [10]. Research conducted by Z. Cao *et al.* [11] showed the effectiveness of deep learning algorithms that work iteratively in detecting the truth of news. The Recurrent Neural Network (RNN) algorithm achieved a high accuracy in determining whether a news story is true or false. This research uses a data set with a total of 44,901 news items containing the full text, title, and release date of each article. The use of a smaller dataset allows the machine to require less training data, thus identifying fake news relatively quickly [11]. Based on the research conducted by Hasan and Setiawan [12] proposed the use of a hybrid RNN-CNN model that combines RNN's ability to handle sequential data with CNN's ability to detect spatial features, using extended features from the top tweet corpus, resulting in the highest accuracy of 76.29%. Besides, research [13] [14] developed an RNN model for text classification. RNN works by processing text sequences one by one while retaining information from previous steps, allowing the model to improve accuracy in text classification tasks that require continuous context understanding. The results showed that the model provided the best performance in terms of accuracy due to the ability to recognize dynamic trends in the text more effectively, improving the accuracy and performance of fake news identification [13].

Research on feature extraction has also been done before, such as in research [15]. Aninditya *et al.* used data on mid and final-semester exam questions. The results show that the use of TF-IDF together with Naive Bayes Classifier and N-grams as indexing terms achieves 85% precision and 80% recall. With TF-IDF, the model can better identify and classify exam questions based on the appropriate cognitive level, resulting in more consistent classification results compared to manual methods. Meanwhile, research related to expansion features conducted by Ramadhanti and Setiawan [16] showed that the RNN-based system with the addition of GloVe for topic classification on the Top 10 features achieved an accuracy of 93.72% and an increase of 0.23%. This research uses data from 55,411 tweets. Since tweets are short text data limited to 280 characters, GloVe was chosen to reduce word mismatches because text data contains many term variations, such as slang. GloVe accelerates parameter training and can significantly improve system performance [17]. Furthermore, research conducted by Poetra *et al.* [18] shows that the use of the GloVe method can produce an accuracy of 88%. In addition, in the research, Kamati [13] GloVe is used as word embedding to provide a vector representation of each word that reflects some secret features of the word. The GloVe was also used in research conducted by Vyas [19], which is an extended version of the Word2Vec model.

Furthermore, research related to word embedding was conducted by Sharma *et al.* [20], which focused on tweet text classification using CNN-RNN and Bidirectional Encoder Representations from Transformers (BERT) hybrid models. In this research, Sharma *et al.* evaluated the performance of several models to determine which model is most effective in handling short and unstructured texts such as tweets. The results showed that the BERT model provided a higher accuracy, which is a significant achievement in the text classification domain [20], [21]. BERT, with its transformer architecture, is able to capture long-range dependencies and nuances in the text, which are often difficult to handle by traditional models [21].

Research related to Genetic Algorithm (GA) as optimization has been done before, such as in research by Fatimatuzzahra *et al.* [22]. Neural network algorithms need to be optimized so that they can perform their duties optimally and produce high accuracy. Based on research [22] shows that the RMSE for neural network parameters optimized with genetic algorithms is 0.044. Thus, GA can be used to define the layer configuration and the number of hidden units, select the appropriate features, set the momentum, and initialize and optimize the weights of the neural network [22], [23]. To improve the efficiency of the model, Maragheh *et al.* in research [23] utilized meta-heuristic algorithms, such as GA, to optimize the model parameters. With this approach, the model not only becomes more accurate in detecting fake news but also more efficient in terms of computation time and resource usage, so it can handle larger and more complex datasets more effectively.

Based on related research, RNN is chosen to analyze sequential text data, utilize temporal relationships between words, and have a simpler structure than LSTM or GRU. TF-IDF is chosen to extract relevant important words in the text, and GloVe is chosen to extend the features with word vector representations that enrich the context and semantic relationships. BERT is chosen to embed words by considering the surrounding context, thus improving the comprehension and accuracy of meaning analysis. Meanwhile, GA was chosen for its ability to find the optimal combination of parameters iteratively. The selection of the 2024 Indonesian election dataset is related to trending issues in social media, especially related to the importance of managing credible information. As far as we know, no one has integrated these methods together with the use of genetic algorithm optimization for RNN to detect information credibility, especially in Indonesian Twitter datasets. This research contributes to the development of an information credibility detection system using RNN on X that uses Genetic Algorithm, TF-IDF feature extraction, GloVe feature expansion, and BERT to fill the gap from previous related research. This novelty approach combines the power of the Genetic Algorithm in finding optimal parameters for RNN, the ability of TF-IDF and GloVe to capture richer word representations and the superiority of BERT in understanding text context in depth. The main focus of this research is to improve accuracy and efficiency in identifying fake news and non-credible information.

The paper is divided into four main parts. The first provides the background and related research. Second, it provides a detailed explanation of the methods used in the research. The third contains the test results, scenarios and discussion. The last part contains the concluding remarks of the research.

2. METHODS

This research builds an information credibility detection system using RNN optimized with a genetic algorithm. The system also utilizes TF-IDF feature extraction, GloVe feature expansion and BERT word embedding to improve accuracy and effectiveness in detecting credible information. This system design consists of several main components that work to produce more accurate and efficient detection. The system architecture design can be seen in Fig. 1. The integration of TF-IDF as an extraction feature makes an important contribution to enriching the document representation, as does GloVe as an expansion feature. While BERT deepens the understanding of text context, the Genetic Algorithm enables the discovery of optimal RNN parameters, allowing this approach to generate models with high accuracy in classifying text for information credibility detection.

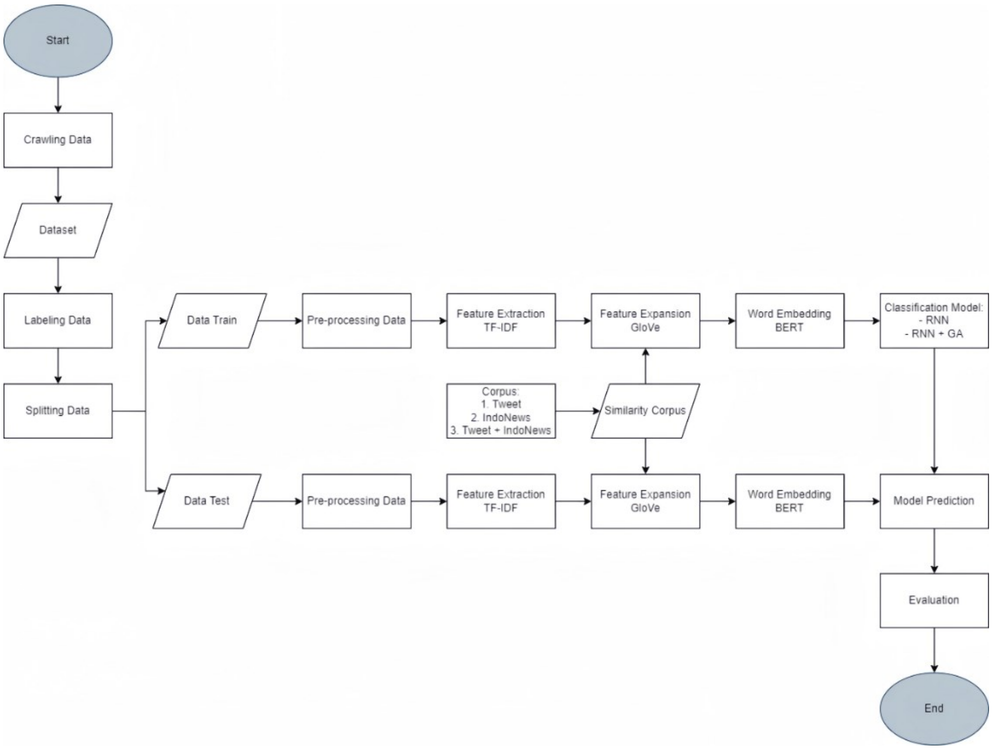

Fig. 1. System Architecture Design

2.1. Crawling Data

The dataset utilized for this research was acquired from crawled data from Twitter, which involved the process of retrieving data from the platform to serve as a dataset. Crawling techniques are used to collect data, which is an effective way to get information from Twitter. Designing research that automates data collection from Twitter, making it easier to get information more quickly and systematically [24], [25]. This crawling process was done using Twitter API to extract the tweets. The data retrieved included a variety of tweets in Bahasa Indonesia, ensuring the relevance and contextuality of the data used for information credibility analysis. The crawling process applies filters to identify and remove entries such as duplicate tweets, spam, or irrelevant content, thus ensuring that only tweets are unique and relevant. The data crawling was carried out with a total of 54.766 tweets with four keywords and two hashtags related to the 2024 Election. The collected data will be automatically saved in Comma Separated Values (CSV) file format. The results of crawling the twitter dataset can be seen in Table 1.

Table 1. Quantity of Crawled Data

Keywords	Quantity
anies baswedan	11.672
pemilu 2024	2.688
pilpres 2024	13.557
prabowo	16.282
#ganjarpranowo	9.345
#debatcapres	1.222

2.2. Labeling Data

Data labeling is the activity of determining the data class of a dataset [26], [27]. The data acquired from the crawling step could not yet indicate whether the data included credible information or not. Therefore, a manual data labeling process was performed, divided into two labels, namely 1 for tweets that were considered credible and 0 for tweets that were considered not credible. The manual labeling process to determine credible and non-credible tweets involves several steps and criteria based on previous research. First, tweets must be identified, whether they come from accounts that are official media or news sources such as Kompas, Metro TV, and CNN, or if they mention these news accounts. Tweets with many retweets, favorites, or that have a URL as a news source are also considered more credible.

In addition, tweets that explain events that have already happened or confirm an activity at a certain time are given priority. Accounts with many followers, many previous tweets, and those that have been created for a long time are also more trusted. The sentiment of the tweets is also considered, where tweets with negative sentiment tend to be more credible. Labeling must also ensure that the information is informative to the public and not a short message between friends, and the topic of the tweet must match the competence of the account that created it. All of these criteria must be met for a tweet to be labeled credible, referring to research conducted by Setiawan *et al.* [4]. Table 2 shows an example of dataset labeling. At the same time, the total of the labeling results can be seen in Table 3. The resulting crawled and labeled dataset shows a relatively balanced number of credible and non-credible tweets, ensuring equal representation for the next process.

Table 2. Example of Labeled Data

Tweets	Label
Calon Presiden Nomor urut 1 Anies Baswedan @aniesbaswedan menyebut perlu adanya debat khusus cawapres agar masyarakat bisa tau bahwa cawapres yang dipilih itu berkualitas. Download #metrotvxtend di app store & play store #aniesbaswedan #muhaiminiskandar #pemilu2024 #pilpres2024	1
sepertinya sudah waktunya pemilu 2024 batal. digantikan oleh rakyat. semua nya harus mundur dan para penghianat bangsa (sampah) harus dibersihkan secara menyeluruh.	0
KPU sudah menyebar desain lima surat suara yang akan dicoblos masyarakat di #Pemilu2024. Kenali perbedaannya dalam #Infografis #CNNIndonesia: https://bit.ly/3T5HbxC	1
Idem Sy juga berpikir untuk golput aja di pilpres 2024 kelak..	0

Table 3. Quantity of Labeled Data

Label	Quantity
Credible	27.489
Non-Credible	27.277

2.3. Preprocessing Data

Data preprocessing was carried out before the data was entered into the classification model. The data obtained from Twitter contained a number of complex and difficult information that had to be processed in the system. Therefore, it is necessary to process the data first so that it becomes appropriate data and can be used effectively. Data preprocessing consists of several steps, including data cleaning, case folding, tokenizing, stopword removal, normalization, and stemming [16], [24]-[27]. Some of these stages serve to eliminate errors and similarities in word forms and reduce the number of words from the data in the dataset that has been collected. Table 4 is an example of data preprocessing.

1. Data Cleaning: Process of cleaning or removing punctuation marks, emojis, numbers, hashtags, and URLs from textual data.
2. Case Folding: Process of converting all textual data from upper-case and capitalization to lower-case.

3. Tokenizing: Process of separating a sentence in the data into individual words.
4. Stopwords Removal: Process of removing common words that are not needed or considered unimportant, such as conjunctions.
5. Normalization: Normalizing text, correcting spelling errors, and normalizing slang words so that words that were originally not standard will become standard.
6. Stemming: The process of changing the words by removing affixes, both prefixes and suffixes, so that only the basic word is left.

Table 4. Example of Data Preprocessing

Preprocessing	Text
Original Tweets	KPU sudah menyebar desain lima surat suara yang akan dicoblos masyarakat di #Pemilu2024. Kenali perbedaannya dalam #Infografis #CNNIndonesia: https://bit.ly/3T5HbxC
Data Cleaning	KPU sudah menyebar desain lima surat suara yang akan dicoblos masyarakat di Kenali perbedaannya dalam
Case Folding	kpu sudah menyebar desain lima surat suara yang akan dicoblos masyarakat di kenali perbedaannya dalam
Tokenizing	[‘kpu’, ‘sudah’, ‘menyebar’, ‘desain’, ‘lima’, ‘surat’, ‘suara’, ‘yang’, ‘akan’, ‘dicoblos’, ‘masyarakat’, ‘di’, ‘kenali’, ‘perbedaannya’, ‘dalam’]
Stopwords Removal	[‘kpu’, ‘menyebar’, ‘desain’, ‘lima’, ‘surat’, ‘suara’, ‘dicoblos’, ‘masyarakat’, ‘kenali’, ‘perbedaannya’]
Normalization	[‘kpu’, ‘menyebar’, ‘desain’, ‘lima’, ‘surat’, ‘suara’, ‘dicoblos’, ‘masyarakat’, ‘kenali’, ‘perbedaannya’]
Stemming	[‘kpu’, ‘sebar’, ‘desain’, ‘lima’, ‘surat’, ‘suara’, ‘coblos’, ‘masyarakat’, ‘kenal’, ‘beda’]

2.4. Feature Extraction TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) is the process of weighting words or grouping the most frequently used words in a document [15], [28], [29]. TF-IDF works by assigning weights to words in a document (TF) and subtracting the weights of words that appear widely across documents (IDF). To be machine-understandable, the data needs to be converted into vectors representing each word used through preprocessing that measures the occurrence of the word in the document. To perform text classification, a feature extraction stage is required so that the system can extract information and represent it based on the input data [29]. TF-IDF was chosen for feature extraction due to its simple computation yet proven ability to extract important words in a document.

In this research, TF-IDF is implemented with adjustments such as n-grams (unigram, bigram, and trigram) to capture a broader context and the max features parameter to limit the number of features used in the model, ensuring only the most informative features are retained. This feature extraction process forms an N-dimensional vector space representing one feature extracted from the dataset. TF-IDF enables the formation of this vector space by emphasizing the most significant words, thus improving accuracy and efficiency in text classification. Weighting using TF-IDF is the most common technique that is often applied. To calculate TF-IDF, you can use the formula below:

$$TFIDF_{(i,j)} = TF_{(i,j)} \times IDF_{(j)} \tag{1}$$

$$TF_{(i,j)} = \frac{Freq\ of\ term\ i\ in\ doc\ j}{Total\ term\ i\ in\ doc\ j} \tag{2}$$

$$IDF_{(j)} = \frac{Total\ doc}{Number\ doc\ with\ term\ i} \tag{3}$$

TF-IDF works with two main components: TF in equation (2) measures how often a term appears in a document and is calculated as the number of occurrences of the term divided by the total number of words in the document. Since not all documents are the same length, TF gives proportional weight to the frequency with which a word appears in a particular document. IDF, on the other hand, in equation (3), measures the importance of a term in the entire document corpus. This is done by reducing the weight of terms that appear in many documents, as they are considered less informative. These two components are combined to give a TF-IDF weight for each term in the document, which is calculated in equation (1).

2.5. Feature Expansion GloVe

Besides feature extraction, this research also conducted feature expansion using Global Vector (GloVe) to expand the features. GloVe is a model that can store global word occurrence statistics, which can later be used to represent words or meanings [30]-[32]. The GloVe model was created from the observations made, and it is known that the ratio of word occurrences has the potential to draw a conclusion. The result obtained from GloVe is the closeness of a word to other words. This value is used to expand the credibility of the information feature.

In this research, GloVe is integrated into the feature expansion process to improve text representation with the NLTK (Natural Language Toolkit) library. We use pre-trained GloVe embeddings on a large corpus to capture richer word representations. The feature expansion process, by replacing the 0-valued vectors in the data with similar words in the GloVe Corpus, is a continuation of the previous feature extraction process. GloVe calculates how often a word co-occurs in a corpus so that the similarity value of the words can be determined. This method is useful for identifying and classifying similarities between words in the data [31]. After obtaining similar words, feature expansion on the representation vector obtained from the feature extraction process is continued.

The utilization of GloVe involves building a corpus from datasets X and IndoNews and a combination of the two (X+IndoNews) to produce accurate and contextual word representations, which can capture semantic relationships between words. The tweet corpus consists of tweets on X related to the 2024 General Election, and the IndoNews corpus consists of news articles from various trusted sources in Indonesia, which are then combined into the X+IndoNews corpus. The feature expansion process is performed by replacing the 0-valued vectors with a vocabulary similar to the GloVe corpus based on the calculated similarity. Table 5 is an example of words in corpus building.

Table 5. Example of Words Similar to "sosialisasi"

Word	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8
sosialisasi	gencar	giat	edukasi	partisipasi	kampanye	masyarakat	pilar	laksana
	Rank 9	**Rank 10**	**Rank 11**	**Rank 12**	**Rank 13**	**Rank 14**	**Rank 15**	**Rank 16**
	harap	masif	aga	imbau	konsolidasi	cegah	teknis	tertib

For example, after getting the most similar vocabulary to the presidential candidate in the X+IndoNews corpus, the next feature expansion will be carried out. On the word "sosialisasi", if it is denoted by the number 0 and the word "gencar" is denoted by the number 1, then the word "sosialisasi" can also be denoted by the number 1 because it is a word similar to "gencar" based on the similarity in Table 5 built by the GloVe model.

2.6. Word Embedding BERT

The pre-processed tweets are feature extracted using the Bidirectional Encoder Representations from Transformers (BERT) method, a word embedding approach that considers the context of word similarity [33]-[36]. BERT provides a contextualized and rich feature representation, taking into account the unique language structure and meaning in each tweet. BERT is a cutting-edge architecture that has been trained on large data for Natural Language Processing (NLP) tasks to gain language understanding. Using this language understanding and transfer learning, BERT is proven to reduce training time and improve performance [33].

In this research, BERT, specifically the IndoBERT variant, which is the base model of BERT, is integrated to perform word embedding using the Transformer library so as to capture a deeper and richer contextual representation of the text. BERT attracts great attention due to its ability to predict words based on left and right contexts and is trained using a widely available corpus of plain text [34]. BERT consists of a stack of coders that use self-attention mechanisms to understand the bidirectional context of text. The model is pre-trained on large datasets for common NLP tasks, which allows BERT to understand language in greater depth and transfer that learning to specialized tasks, thus reducing training time and improving performance. BERT maps each processed token into a word embedding using the pre-trained WordPiece embedding. These embeddings are then processed by a self-attention coding stack [33]. The result is a classification representation that is used as input for the classifier.

2.7. Recurrent Neural Network

Recurrent Neural Networks (RNN) are part of a class of Neural Networks (NN) that have sequential input or output data. RNNs capture the temporal relationship between input/output sequences by introducing feedback to the feed-forward neural network [37]-[44]. RNN is a type of deep learning method that can classify sequence data effectively [40]. RNNs learn from the sequence of items consumed by a user over time, making them ideal for tasks where sequence and context are critical. Thus, many applications with sequential data, such as speech recognition, language translation, and human activity recognition, can benefit from RNNs. Fig. 2 shows the architecture of an RNN as follows:

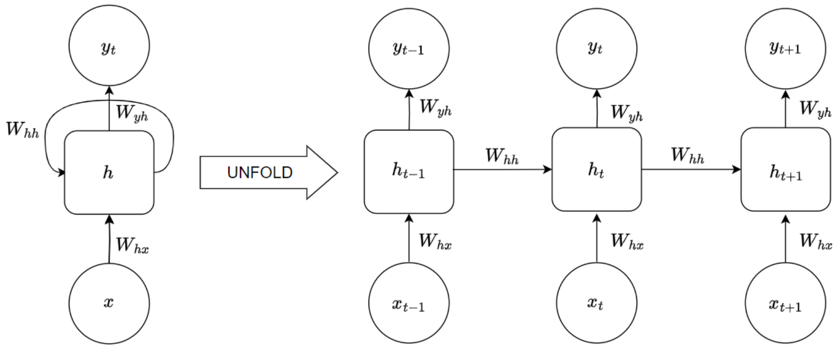

Fig. 2. Recurrent Neural Network Architecture

RNNs have recently attracted significant attention from the research community for sequence modeling and have proven to excel in solving a wide array of natural language processing (NLP) tasks [37]. The advantage of RNNs lies in their ability to process variable-length inputs and outputs, which is particularly beneficial in the processing of text of varying lengths [37]. RNNs operate by maintaining a memory of previous information in a sequence of data, allowing these models to understand the broader context as well as the relationship between words in a sentence or paragraph. These capabilities make RNNs a top choice for various NLP applications such as machine translation, sentiment analysis, and speech recognition [41].

Based on Fig. 2, for each time t, the hidden layer h_t is maintained and updated based on the input x_{t-1} and hidden state h_{t-1}. Hidden state ht can be calculated with the following equation:

$$h_t = \sigma h \, (W_{hx} X_t + W_{hh} h_{t-1} + b_h \tag{4}$$

W_{hx} is a vector representing the input weights, W_{hh} is a vector representing the recurrent weights, σh is the ReLU activation function, and b_h is the bias vector of the hidden layer. The output of the RNN structure is updated using the equation y_t as follows:

$$y_t = \sigma y \, (W_{yh} h_t + b_y) \tag{5}$$

σy is the activation function of the output layer, W_{yh} is the weight matrix between the hidden layer and the output layer, and b_y is the bias vector of the output layer. By utilizing the results from the previous stage as input, RNNs have the capacity to capture contextual information more effectively, which significantly improves the model's performance in understanding and predicting patterns in sequential data. This capability enables the application of RNNs in various domains, such as product recommendation systems, stock market prediction, and anomaly detection, where temporal patterns and event sequences play a crucial role in analysis and decision-making [38], [39].

In this research, a standard RNN architecture is used for information credibility detection, chosen for its simplicity and ability to process sequences of text data. Classification using RNN is done by utilizing the Keras library to train and test neural network models efficiently. RNNs can be effective in analyzing data sequences to identify patterns that indicate non-credible information. RNNs have the advantage of understanding how the received input affects the prediction of the current output. This ability allows RNNs to develop models capable of capturing temporal relationships and patterns [40].

2.8. Genetic Algorithm

Genetic Algorithm (GA) is an algorithm that belongs to the Evolutionary Algorithm (EA) category and is based on the principle of natural selection [45]-[49]. GA is commonly used to generate high-quality solutions to search and optimization problems, utilizing operators inspired by biological processes such as selection, crossover, and mutation [45]. The basic principle of GA involves a population of candidate solutions that evolves over generations, with the best solutions being selected and combined to create new and better solutions. The mutation is then applied to introduce variation and ensure diversity in the solution population.

In recent years, GA has attracted significant attention due to its ability as an innovative and effective optimization methodology. Factors such as usability, ease of implementation, minimal specification requirements, and simultaneous and cross-national approaches make GAs a popular choice in a wide range of applications. The use of genetic algorithms to optimize the training of neural networks has yielded some important results in various applications [50]. GA has been successfully applied in various domains, ranging from industrial optimization, logistics planning, and engineering design to artificial intelligence and machine learning [46]. This success demonstrates the flexibility and potential of GAs in solving a variety of complex problems that require optimal and efficient solutions. Fig. 3 shows the flowchart of Genetic Algorithm (GA) as follows:

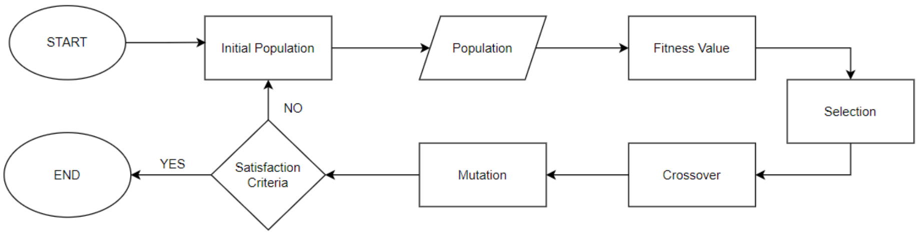

Fig. 3. Flowchart of Genetic Algorithm

In this research, parameter optimization using GA focuses on tuning the number of neurons, epochs, and dropouts to reduce overfitting. GA was chosen for its ability to explore large parameter space, although it faces challenges such as computational complexity that increases with the number of optimized parameters and large population size. Therefore, we limit the value of the GA tuning parameters to still achieve optimal results. The GA process begins with the formation of an initial population of individuals representing potential solutions, evaluated based on fitness values. The best individuals are chosen through selection to become parents in the next generation, followed by crossover to produce new offspring. The mutation is applied to maintain genetic diversity. This process repeats until a stopping criterion is reached, signaling the discovery of an optimal or near-optimal solution.

2.9. Evaluation Performance

The performance measurement of the system accuracy is calculated using the confusion matrix, which serves to measure the performance of the classification model [51], [52]. The use of performance metrics also shows how well the model handles the dataset [53]. The more balanced the class values of the dataset are, the less biased the model will be towards one particular class, and the better it will be to generalize patterns from all classes. This matrix will display the results in the form of actual values and predicted values. For binary classification problems, the confusion matrix can be obtained, as shown in Table 6.

Table 6. Confusion Matrix

	Positive Actual	**Negative Actual**
True Prediction	True Positive (TP)	True Negative (TN)
False Prediction	False Positive (FP)	False Negative (FN)

The performance value of the system can be calculated for accuracy, precision, recall, and F1-Score. The results of confusion matrix calculations will be provided in the form of a percentage, which will then be analyzed as the result of the information credibility detection process. The formula of the calculation can be seen as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

3. RESULTS AND DISCUSSION

This research developed an information credibility detection system model that used RNN as a baseline. Six different scenarios were experimented with to get the maximum increase in accuracy. The model with the highest accuracy in one scenario was used in the next scenario. This research combined RNN with feature extraction, feature expansion, word embedding, and optimization.

3.1. Scenario 1 Result

In the first scenario, the baseline RNN model was tested with 50 units, using sigmoid activation and a loss function using binary cross entropy. The model was tested five times, and the average accuracy was taken. The baseline model becomes a benchmark in comparing the increase in accuracy in the following scenarios. This scenario was conducted to determine the optimal data distribution with the highest accuracy. The results of the first scenario can be seen in Table 7. The results of scenario 1 showed that the model with a ratio of 80:20 data got the highest accuracy of 88.99%. This result will be used in the second scenario.

Table 7. Result of Scenario 1

Split Ratio	Performance Metrics (%)			
	Accuracy	Precision	Recall	F1-Score
90:10	88.90	89.91	88.90	88.90
80:20	**88.99**	89.00	88.99	88.99
70:30	88.64	88.66	88.64	88.63

3.2. Scenario 2 Result

In the second scenario, the baseline RNN applied feature extraction using TF-IDF. This experiment tested the max features parameter value in TF-IDF to determine the optimal value for the model. The max features tested were 2000, 5000, 10000, and 15000. The results of this experiment are seen in Table 8. The results showed that TF-IDF with a max features parameter of 5000 produced the highest accuracy of 89.13%. The model experienced an increase in accuracy of 0.16% from the baseline. Max features with values of 10,000 and 15,000 resulted in slightly lower accuracy than the baseline. So, the best model in this scenario will be used in the next scenario.

Table 8. Result of Scenario 2

Max Features	Performance Metrics (%)			
	Accuracy	Precision	Recall	F1-Score
2000	89.02 (+0.03)	89.06 (+0.06)	89.02 (+0.03)	89.02 (+0.03)
5000	**89.13 (+0.16)**	89.15 (+0.14)	89.13 (+0.16)	89.13 (+0.16)
10000	88.93 (-0.06)	88.94 (-0.06)	88.93 (-0.06)	88.92 (-0.07)
15000	88.89 (-0.11)	88.90 (-0.11)	88.89 (-0.11)	88.89 (-0.11)

3.3. Scenario 3 Result

In the third scenario, the model was tested again by applying TF-IDF to compare n-gram values. An n-gram is a collection of consecutive items in a text document that may contain a word. This experiment was conducted to see the combination of n-gram values that produced the highest accuracy, such as unigram, bigram, and trigram. Table 9 shows the comparison of accuracy values from the third scenario results. It showed that the combination of unigram and bigram gained a 0.25% increase in accuracy, while unigram, bigram, and trigram values decreased in accuracy. This model gets an accuracy of 89.21% and will be used in the next scenario.

Table 9. Result of Scenario 3

N-gram	Performance Metrics (%)			
	Accuracy	Precision	Recall	F1-Score
Unigram	89.13 (+0.15)	89.15 (+0.14)	89.13 (+0.15)	89.13 (+0.15)
Bigram	86.37 (-2.94)	86.38 (-2.94)	86.37 (-2.94)	86.37 (-2.94)
Trigram	83.26 (-6.44)	83.49 (-6.19)	83.26 (-6.44)	83.23 (-6.47)
Unigram + Bigram	**89.21 (+0.25)**	89.22 (+0.24)	89.21 (+0.25)	89.21 (+0.25)
Unigram + Bigram + Trigram	89.17 (+0.20)	89.20 (+0.22)	89.17 (+0.20)	89.17 (+0.20)

3.4. Scenario 4 Result

In the fourth scenario, the model implemented feature expansion using GloVe. Before testing the model, it was necessary to build a corpus using the X, IndoNews, and combined X + IndoNews datasets. The corpus was then examined for the top n rankings to implement GloVe. GloVe provides a rich and contextual representation of words, allowing the model to capture deeper nuances, although it requires higher computation during model training. Table 10 shows the number of vocabulary words generated from the corpus building. Additionally, Table 11 shows the accuracy value of the model applying each corpus in feature expansion. It could be seen that the highest accuracy was obtained from the model that implemented feature expansion for the Tweet + IndoNews corpus with the Top 15. The accuracy obtained was 89.57%, outperforming other models. The best model in the scenario has an increase of 0.65% from the baseline and will be used in the next scenario.

Table 10. Quantity of Corpus Vocabulary

Corpus	Quantity
X	40.466
IndoNews	131.580
X + IndoNews	150.943

Table 11. Result of Scenario 4

Corpus	Rank	Performance Metrics (%)			
		Accuracy	Precision	Recall	F1-Score
X	Top 1	89.23 (+0.26)	89.25 (+0.28)	89.23 (+0.26)	89.23 (+0.26)
	Top 5	89.26 (+0.30)	89.24 (+0.26)	89.26 (+0.30)	89.26 (+0.30)
	Top 10	89.33 (+0.38)	89.34 (+0.40)	89.33 (+0.38)	89.33 (+0.38)
	Top 15	89.45 (+0.51)	89.46 (+0.51)	89.45 (+0.51)	89.45 (+0.51)
	Top 16	89.40 (+0.46)	89.40 (+0.44)	89.40 (+0.46)	89.40 (+0.46)
IndoNews	Top 1	89.13 (+0.15)	89.13 (+0.14)	89.13 (+0.15)	89.13 (+0.15)
	Top 5	89.15 (+0.17)	89.15 (+0.16)	89.15 (+0.17)	89.14 (+0.16)
	Top 10	89.18 (+0.21)	89.20 (+0.22)	89.18 (+0.21)	89.18 (+0.21)
	Top 15	89.34 (+0.39)	89.34 (+0.40)	89.34 (+0.39)	89.34 (+0.39)
	Top 16	89.26 (+0.30)	89.26 (+0.29)	89.26 (+0.30)	89.25 (+0.29)
X + IndoNews	Top 1	89.14 (+0.16)	89.14 (+0.15)	89.14 (+0.16)	89.14 (+0.16)
	Top 5	89.26 (+0.30)	89.26 (+0.29)	89.26 (+0.30)	89.25 (+0.29)
	Top 10	89.42 (+0.48)	89.43 (+0.48)	89.42 (+0.48)	89.42 (+0.48)
	Top 15	**89.57 (+0.65)**	89.58 (+0.65)	89.57 (+0.65)	89.57 (+0.65)
	Top 16	89.47 (+0.53)	89.47 (+0.52)	89.47 (+0.53)	89.47 (+0.53)

3.5. Scenario 5 Result

In the fifth scenario, the model implemented word embedding using BERT. After going through the previous experiments, specifically integrating feature extraction and expanding features, the model was tested again using BERT to process word semantics with the aim of achieving an increase in accuracy. BERT contributes significantly to the model's performance with its ability to deeply understand text context, which is enhanced through interaction with RNN to integrate temporal information. Table 12 shows the increase in accuracy of the model after being tested using BERT, increased by 1.10%, achieving an accuracy of 89.97%.

Table 12. Result of Scenario 5

Model	Performance Metrics (%)			
	Accuracy	Precision	Recall	F1-Score
RNN + TF-IDF + GloVe + BERT	89.97 (+1.10)	89.99 (+1.12)	89.97 (+1.10)	89.96 (+1.09)

3.6. Scenario 6 Result

In the sixth scenario, the model was tested with the last experiment using Genetic Algorithm (GA) optimization. The goal was for the model to work optimally with maximum accuracy. In this research, GA tunes specific parameters to search for the best combination in the model, with a randomly initialized population to represent potential variations in units, dropout rates, and number of epochs. The chosen population size (pop size) and number of generations provide a balance between exploration and exploitation in the parameter search space, while the mutation rate introduces new variations to prevent convergence.

Although it requires a large computational complexity, we have limited the values of these parameters, as seen in Table 13, to find the best individual in each generation. Table 14 shows the best individual in the GA process with the best parameters and scores. The results showed that the best individual was obtained with parameters: units of 57, dropout of 0.220, and epochs of 19. This best individual achieved the highest score, outperforming other individuals in the previous generation, with a score of 90.60%. This demonstrated a 1.81% increase in accuracy from the baseline model.

Table 13. Genetic Algorithm Parameters

Parameters	Values	Description
pop_size	5	The number of individuals in the population is to be evaluated and updated every generation.
generations	5	Number of evaluation cycles and population updates performed.
mutation_rate	0.4	The probability that each individual will undergo mutation in the process of evolution.

Table 14. Result of Scenario 6

Parameters	Values	Performance Metrics (%)			
		Accuracy	Precision	Recall	F1-Score
Units	57				
Dropout	0.22056077228483756	90.60 (+1.81)	90.55 (+1.74)	90.55 (+1.73)	90.55 (+1.73)
Epochs	19				

3.7. Discussion

This research designed an information credibility detection system with a Recurrent Neural Network (RNN) that implemented feature extraction, feature expansion, word embedding, and optimization. The dataset used is Indonesian language tweet data related to the 2024 Election. Data pre-processing for tweet classification involves a series of steps to create clean data without duplicates and address potential biases. These steps include data cleaning to remove unwanted characters, case folding to bring consistency in text representation, tokenization to divide text into smaller tokens, stopword removal to remove common words that do not provide additional meaning, normalization to improve text formatting, and stemming to reduce words to their base form. This process also includes manual labeling to distinguish between credible (label 1) and non-credible (label 0) tweets so that they can be used by the model. This research conducted six sets of scenarios in experiments with the aim of obtaining maximum accuracy.

In the first scenario, tests were conducted with varying distributions of training and test data, which resulted in the finding that a ratio of 20% test data and 80% training data gave the highest accuracy of 88.88% compared to other ratios. Testing in the first scenario with sigmoid activation and binary cross-entropy loss functions for binary classification showed that this combination gave good results with fairly high accuracy. These functions played an important role in the formation of an effective RNN model for classification tasks. This showed the importance of choosing the right data distribution to improve model performance.

In the second scenario, the use of feature extraction with TF-IDF showed an accuracy improvement of 0.16% from the baseline. Experiments with the maximum feature parameter value in TF-IDF showed that the value of 5000 gave the best result with an accuracy of 89.13%. TF-IDF was used to measure the importance of a word in a document, giving higher weights to words that appeared infrequently but were contextually important. The advantage of TF-IDF lies in its ability to capture relevant and significant information, thus improving the accuracy of the model in identifying the credibility of information.

The third scenario introduced n-gram value testing in TF-IDF, which is a combination of n consecutive items in a text document. The results showed that using a combination of unigram and bigram values resulted in an accuracy improvement of 0.25%. Proper selection of n-grams improved the understanding of the context of the words in the document. N-grams allowed the model to capture the relationship between adjacent words, revealing deeper meaning than single-word analysis. Thus, the use of n-grams helped identify more complex and relevant linguistic patterns for classification tasks.

In the fourth scenario, implementing feature expansion with GloVe resulted in a significant improvement in accuracy. By building the corpus and implementing GloVe for feature expansion on the Tweet + IndoNews similarity corpus with the Top 15 ranks, the model reached its highest accuracy of 89.57%. GloVe helped generate semantically richer word representations, improving the model's ability to understand information content. The advantage of GloVe lies in its ability to capture relationships between words based on the global context in the corpus, in contrast to other methods that might only consider local context. Testing the Top N

Vol. 10, No. 2, June 2024, pp. 419-434

values showed that the Top 15 provided the optimal balance between the number of features and accuracy compared to the other Top N tested.

The fifth scenario introduced the use of word embedding with BERT, resulting in a 1.10% increase in accuracy. BERT assisted the model in understanding the semantic relationships between words in a sentence, providing a deeper understanding of the content of the information being analyzed. The advantage of BERT lies in its ability to consider word context bidirectionally, both from left to right and from right to left. As such, BERT could capture more accurate and contextualized meaning from text, improving the model's performance in classification tasks.

In the last scenario, GA optimization was utilized to find the best parameters for the model. The results demonstrated a 1.81% improvement in accuracy over the base model, with the best individual in the last generation achieving the highest score of 90.60%. The use of GA in model parameter optimization offers advantages over other optimization methods. GA can automatically adjust model parameters to achieve the best results, with the ability to explore the parameter space thoroughly and efficiently. This allows GA to find more optimal solutions in a relatively short time, increasing the effectiveness and efficiency in significantly improving model performance.

Overall, the experiments in these six scenarios demonstrate that a combination of feature extraction, feature expansion, word embedding, and parameter optimization can lead to significant accuracy improvements in the RNN-based information credibility detection system model. Each feature and technique applied has a specific role in enhancing the model's understanding and analysis of information content, highlighting the importance of a holistic approach in developing an effective and accurate information credibility detection system. Although it requires a large amount of computation, the integration of these methods results in maximum accuracy. This research showcases significant advantages through the accuracy increases observed in each scenario. Starting from the first scenario, which achieved an accuracy increase of 88.99% through the selection of the right data distribution, to the sixth scenario, which resulted in an accuracy increase of up to 1.81% from the base model with the application of the GA optimization.

Next, a statistical significance test is conducted to validate whether there is a statistically significant change in accuracy. The P-value statistical measure indicates the likelihood of no change in accuracy if it is smaller than 0.05. The Z-value statistical measure assesses whether the observed difference between the two scenarios is large enough to be considered significant if it is greater than 1.96, indicating that the difference is statistically significant at the 95% confidence level [26]. Based on Table 15, the overall accuracy during scenario testing tends to increase significantly, except from the first scenario to the second scenario (S1→S2) and the second scenario to the third scenario (S2→S3). The change in accuracy across scenarios (S1→S6) shows significance, which underlines that the proposed model provides the best accuracy compared to the baseline. The incremental accuracy improvements in each scenario, as depicted in Fig. 4, also validate that a combination of features and techniques can yield superior results in the development of information credibility detection systems.

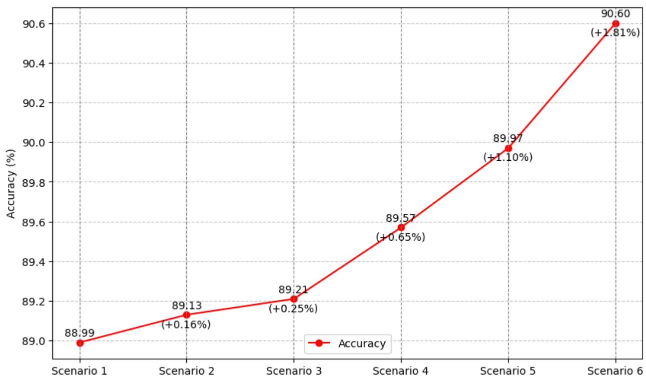

Fig. 4. Relative Increase of All Scenarios

The final part of this research compares the results with similar methods from previous research to improve the contribution of this research. Table 16 shows the differences in the performance results from some previous studies. Based on Table 16, the RNN model integrated with TF-IDF feature expression, GloVe feature expansion, BERT word embedding, and GA optimization shows the highest performance compared to the methods proposed in previous studies that only use the baseline or only integrate some features.

Table 15. Significant Test Result

Parameters	Scenarios					
	S1→S2	S2→S3	S3→S4	S4→S5	S5→S6	S1→S6
Z-Value	1.282	1.164	9.996	1.992	3.960	16.095
P-Value	0.199	0.244	0.0	0.046	0.0	0.0
Significant?	**False**	**False**	**True**	**True**	**True**	**True**

Table 16. Comparative Analysis of Related Research

Ref	Proposed Method	Feature Extraction	Feature Expansion	Word Embedding	Optimizer	Performance Metrics (%)			
						Acc	Prec	Rec	F1
Hasan *et al.* [12]	Hybrid RNN-CNN	TF-IDF	GloVe	-	-	76.29	-	-	-
David *et al.* [14]	RNN	-	FastText	-	-	83.33	-	-	-
Ni *et al.* [17]	LSTM-GRU	-	GloVe	-	Adam	87.16	-	-	86.08
Wang *et al.* [21]	Bi-LSTM	-	-	BERT	-	90.04	-	-	90.04
Maragheh *et al.* [23]	LSTM	-	-	-	GA	82.32	83.17	82.57	82.87
Wijaya *et al.* [26]	CNN	TF-IDF	FastText	-	-	88.79	-	-	86.43
This Research	**RNN**	**TF-IDF**	**GloVe**	**BERT**	**GA**	**90.60**	**90.55**	**90.55**	**90.55**

4. CONCLUSION

This research presents a significant development in information credibility detection systems using an RNN-based approach with the utilization of TF-IDF optimization, GloVe, BERT, and Genetic Algorithm. The dataset used is Indonesian language tweet data related to the 2024 General Election. The selection of datasets is closely related to the topics that are being discussed on social media, especially regarding the importance of handling credible information. Data crawling was performed on a total of 54,766 tweets with four keywords and two hashtags. The dataset was cleaned for proper model use, including data cleaning from unwanted characters, letter folding, tokenization, stopword removal, normalization, and stemming. This step aims to ensure the data is clean, consistent, and free from potential bias. Manual labeling is also performed to distinguish between credible (label 1) and non-credible (label 0) tweets that will be used in the classification model.

Through six rigorous experimental scenarios, with steps ranging from data distribution adjustment in the first scenario, the use of TF-IDF in the second and third scenarios, corpus development for the use of GloVe in the fourth scenario, integration of BERT as word embedding in the fifth scenario, to the use of GA in the last scenario, there was an accuracy improvement of 1.81% from the base model with an initial accuracy of 90.60%. In addition, the proposed method also obtained 90.55% for precision, recall, and F1-score metrics. Such scenarios involve various methods, from data processing to the use of various techniques, embedding words and optimizing parameters for the model, effectively improving the quality and accuracy of the model in real-time. This reflects the success of the holistic approach applied in significantly improving the performance of the model.

The accuracy improvements observed in the scenario experiments highlight the advantages of the approach used in this research. The holistic combination of features and techniques leads to a deeper understanding of the information content and improves the model's ability to recognize and classify information credibility more accurately. The combination of TF-IDF, GloVe, BERT, and GA synergistically improves the overall performance by focusing on accurate word representation, deep context understanding, and optimal optimization of model parameters, enabling the system to tackle complex challenges in text classification more effectively. The use of GA as an optimization tool also offers the advantage of efficiently adjusting model parameters; although it requires a large amount of computation, GA produces maximum accuracy due to optimal parameter tuning. This research contributes to addressing the challenges of dealing with misinformation on social media with the proposed method, thereby improving the system's ability to identify and distinguish credible and non-credible information. The use of algorithms and the tuning of experimental parameters can be suggested through a literature review of previous research. The limitation of this research is the Indonesian language dataset that can be obtained from Twitter (X). So, suggestions for future research, exploring more diverse datasets and utilizing more complex hybrid neural networks, can help expand the understanding of information credibility dynamics in a broader context. By continuing to develop and apply innovative approaches, it is hoped that a more effective and adaptive information credibility detection system can be created to support more accurate and credible information analysis.

Acknowledgments

The author would like to thank those who always support us and the guiding professors who always help with guidance and expertise to finish this research with careful preparation.

REFERENCES

[1] F. Ahmad and S. A. M. Rizvi, "Information credibility on Twitter using machine learning techniques," *in Communications in computer and information science*, pp. 371–381, 2020, https://doi.org/10.1007/978-981-15-4451-4_29.

[2] G. Pennycook and D. G. Rand, "Fighting misinformation on social media using crowdsourced judgments of news source quality," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 7, pp. 2521–2526, Jan. 2019, https://doi.org/10.1073/pnas.1806781116.

[3] J. Son, J. Lee, O. Oh, H. K. Lee, and J. Woo, "Using a Heuristic-Systematic Model to assess the Twitter user profile's impact on disaster tweet credibility," *International Journal of Information Management*, vol. 54, p. 102176, Oct. 2020, https://doi.org/10.1016/j.ijinfomgt.2020.102176.

[4] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, "Measuring information credibility in social media using combination of user profile and message content dimensions," *International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, p. 3537, Aug. 2020, https://doi.org/10.11591/ijece.v10i4.pp3537-3549.

[5] N. Sitaula, C. K. Mohan, J. Grygiel, X. Zhou, and R. Zafarani, "Credibility-Based fake news detection," *in Lecture notes in social networks*, pp. 163–182, 2020, https://doi.org/10.1007/978-3-030-42699-6_9.

[6] S. Sharma, K. Rathore, S. Mittal, and S. Srivastava, "An Analytical Study of Rumoured Tweets by Using Twitter Data," *Journal of Web Development and Web Designing*, vol. 4, no. 2, pp. 1–9, 2019, https://doi.org/10.5281/zenodo.3066521.

[7] S. N. S. Rajini, K. Anuradha, S. Umadevi, and E. M. Beulah, "Twitter Sentiment Analysis on big data in SPARK Framework," *IOP Conference Series. Materials Science and Engineering*, vol. 925, no. 1, p. 012015, Sep. 2020, https://doi.org/10.1088/1757-899x/925/1/012015.

[8] R. Yunanto, A. P. Purfini, and A. Prabuwisesa, "Survei literatur: Deteksi berita palsu menggunakan pendekatan Deep Learning," *Jurnal Manajemen Informatika/Jamika*, vol. 11, no. 2, pp. 118–130, Sep. 2021, https://doi.org/10.34010/jamika.v11i2.5362.

[9] M. Azer, M. Taha, H. H. Zayed, and M. Gadallah, "Credibility Detection on Twitter News Using Machine Learning Approach," *International Journal of Intelligent Systems and Applications*, vol. 13, no. 3, pp. 1–10, Jun. 2021, https://doi.org/10.5815/ijisa.2021.03.01.

[10] M. A. Kaufhold, M. Bayer, D. Hartung, and C. Reuter, "Design and Evaluation of Deep Learning Models for Real-Time Credibility Assessment in Twitter," *Artificial Neural Networks and Machine Learning – ICANN*, pp. 396–408, 2021, https://doi.org/10.1007/978-3-030-86383-8_32.

[11] Z. Cao, X. Ding, and Z. Tao, "News Detection for Recurrent Neural Network Approach," *in 2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), IEEE*, pp. 794–797, Sep. 2022, https://doi.org/10.1109/CEI57409.2022.9950209.

[12] M. K. Hasan and E. B. Setiawan, "Sentiment Analysis of Twitter Data on Bank Central Asia Stocks (BBCA) Using RNN and CNN Model with GloVe Feature Expansion," *2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, pp. 195-200, Nov. 2023, https://doi.org/10.1109/comnetsat59769.2023.10420731.

[13] S. P. Akula and N. Kamati, "Credibility of Social-Media Content Using Bidirectional Long Short-Term Memory-Recurrent Neural Networks," *in 2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), IEEE*, pp. 170–175, Aug. 2021, https://doi.org/10.1109/ICETCI51973.2021.9574061.

[14] M. S. David and S. Renjith, "Comparison of word embeddings in text classification based on RNN and CNN," *IOP Conference Series. Materials Science and Engineering,* vol. 1187, no. 1, p. 012029, Sep. 2021, https://doi.org/10.1088/1757-899x/1187/1/012029.

[15] A. Aninditya, M. A. Hasibuan, and E. Sutoyo, "Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom's Taxonomy," *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, pp. 112-117, Nov. 2019, https://doi.org/10.1109/iotais47347.2019.8980428.

[16] W. Ramadhanti and E. B. Setiawan, "Topic Detection on Twitter Using Deep Learning Method with Feature Expansion GloVe," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 780–792, 2023, https://doi.org/10.26555/jiteki.v9i3.26736.

[17] R. Ni and H. Cao, "Sentiment Analysis based on GloVe and LSTM-GRU," *2020 39th Chinese Control Conference (CCC)*, pp. 7492-7497,, Jul. 2020, https://doi.org/10.23919/ccc50068.2020.9188578.

[18] C. K. Poetra, S. F. Pane and N. S. Fatonah, "Meningkatkan akurasi long-short term memory (lstm) pada analisis sentimen vaksin covid-19 di twitter dengan glove," *Jurnal Telematika*, vol. 16, no. 2, pp. 85-90, 2021, https://doi.org/10.61769/telematika.v16i2.400.
[19] P. Vyas and O. F. El-Gayar, "Credibility Analysis of News on Twitter using LSTM: An exploratory study," *AMCIS 2020 Proceedings*, Jan. 2020, https://scholar.dsu.edu/bispapers/107/.
[20] S. Sharma, M. Saraswat, and A. K. Dubey, "Fake news detection on Twitter," *International Journal of Web Information Systems*, vol. 18, no. 5/6, pp. 388– 412, Dec. 2022, https://doi.org/10.1108/IJWIS-02-2022-0044.
[21] C. Wang, P. Nulty, and D. Lillis, "A Comparative Study on Word Embeddings in Deep Learning for Text Classification," *In Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval (NLPIR '20). Association for Computing Machinery*, pp. 37-46, Dec. 2020, https://doi.org/10.1145/3443279.3443304.
[22] F. Fatimatuzzahra, R. Hammad, A. Z. Amrullah, and P. Irfan, "Optimasi Neural Network Dengan Menggunakan Algoritma Genetika Untuk Prediksi Jumlah Kunjungan Wisatawan," *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 3, no. 4, pp. 227–235, Feb. 2022, https://doi.org/10.35746/jtim.v3i4.190.
[23] H. K. Maragheh, F. S. Gharehchopogh, K. Majidzadeh, and A. B. Sangar, "A New Hybrid Based on Long Short-Term Memory Network with Spotted Hyena Optimization Algorithm for Multi-Label Text Classification," *Mathematics*, vol. 10, no. 3, p. 488, Feb. 2022, https://doi.org/10.3390/math10030488.
[24] C. Wintang Kencana, E. B. Setiawan, and I. Kurniawan, "Hoax Detection System on Twitter using Feed-forward and Back-propagation Neural Networks Classification Method," *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 4, no. 4, pp. 655–663, 2020, https://doi.org/10.29207/resti.v4i4.2038.
[25] A. Z. R. Adam, and E. B. Setiawan, "Social Media Sentiment Analysis Using Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU)," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 1, pp. 119-131, 2024, https://doi.org/10.26555/jiteki.v9i1.25813.
[26] K. U. Wijaya, and E. B. Setiawan, "Hate Speech Detection Using Convolutional Neural Network and Gated Recurrent Unit with FastText Feature Expansion on Twitter," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 619-631, 2024, https://doi.org/10.26555/jiteki.v9i3.26532.
[27] J. vol. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," *International Journal of Research in Marketing*, vol. 36, no. 1, pp. 20–38, 2019, https://doi.org/10.1016/j.ijresmar.2018.09.009.
[28] H. Syahputra, and A. Wibowo, "Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, vol. 9, no. 1, pp. 165-173, 2023, https://doi.org/10.26555/jiteki.v9i1.25861.
[29] N. M. Azahra, and E. B. Setiawan, "Sentence-Level Granularity Oriented Sentiment Analysis of Social Media Using Long Short-Term Memory (LSTM) and IndoBERTweet Method," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, vol. 9, no. 1, pp. 85-95, 2023. https://doi.org/10.26555/jiteki.v9i1.25765.
[30] S. M. Mohammed, K. Jacksi, and S. R. M. Zeebaree, "A state-of-the-art survey on semantic similarity for document clustering using GloVe and density-based algorithms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 1, p. 552, Apr. 2021, https://doi.org/10.11591/ijeecs.v22.i1.pp552-562.
[31] N. F. Anistya and E. B. Setiawan, "Hate Speech Detection on Twitter in Indonesia with Feature Expansion Using GloVe," *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1044–1051, Dec. 2021, https://doi.org/10.29207/resti.v5i6.3521.
[32] M. Kamyab, G. Liu, and M. Adjeisah, "Attention-Based CNN and Bi-LSTM model based on TF-IDF and GLOVE word embedding for sentiment analysis," *Applied Sciences*, vol. 11, no. 23, p. 11255, Nov. 2021, https://doi.org/10.3390/app112311255.
[33] W. Harly and A. S. Girsang, "CNN-BERT for measuring agreement between argument in online discussion," *International Journal of Web Information Systems*, vol. 18, no. 5/6, pp. 356–368, Dec. 2022, https://doi.org/10.1108/IJWIS-12- 2021-0141.
[34] N. D. T. Putra and N. E. B. Setiawan, "Sentiment Analysis on Social Media with Glove Using Combination CNN and RoBERTa," *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 7, no. 3, pp. 457–563, Jun. 2023, https://doi.org/10.29207/resti.v7i3.4892.
[35] N. B. A. C. Martani and E. B. Setiawan, "Naïve Bayes-Support vector machine combined BERT to classified Big Five personality on Twitter," *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, vol. 6, no. 6, pp. 1072–1078, Dec. 2022, https://doi.org/10.29207/resti.v6i6.4378.
[36] A. S. Alammary, "BERT Models for Arabic Text Classification: A Systematic Review," *Applied Sciences*, vol. 12, no. 11, p. 5720, Jun. 2022, https://doi.org/10.3390/app12115720.
[37] I. Banerjee *et al.*, "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification," *Artificial Intelligence in Medicine*, vol. 97, pp. 79–88, Jun. 2019, https://doi.org/10.1016/j.artmed.2018.11.004.
[38] S. Bansal and N. Baliyan, "Remembering past and predicting future: a hybrid recurrent neural network based recommender system," *J Ambient Intell Humaniz Comput*, vol. 14, no. 12, pp. 16025-16036 2022, https://doi.org/10.1007/s12652-022-04375-x.
[39] B. Choe, T. Kang, and K. Jung, "Recommendation System with Hierarchical Recurrent Neural Network for Long-Term Time Series," *IEEE Access*, vol. 9, pp. 72033– 72039, 2021, https://doi.org/10.1109/ACCESS.2021.3079922.

[40] B. M. G. A. Awienoor, and E. B. Setiawan, "Movie Recommendation System Based on Tweets Using Switching Hybrid Filtering with Recurrent Neural Network," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 2, pp. 277–293, Apr. 2024, https://doi.org/10.22266/ijies2024.0430.24.

[41] M. A. Saputra and E. B. Setiawan, "Aspect Based Sentiment Analysis Using Recurrent Neural Networks (RNN) on Social Media Twitter," *2023 International Conference on Data Science and Its Applications (ICoDSA)*, Aug. 2023, https://doi.org/10.1109/icodsa58501.2023.10276768.

[42] W. K. Sari, D. P. Rini, and R. F. Malik, "Text classification using Long Short-Term memory with GloVE features," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, vol. 5, no. 2, p. 85, Feb. 2020, https://doi.org/10.26555/jiteki.v5i2.15021.

[43] M. S. Islam, M. S. Sultana, U. Kumar, J. A. Mahmud, and S. J. Islam, "HARC-New Hybrid Method with Hierarchical Attention Based Bidirectional Recurrent Neural Network with Dilated Convolutional Neural Network to Recognize Multilabel Emotions from Text," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, vol. 7, no. 1, p. 142, Apr. 2021, https://doi.org/10.26555/jiteki.v7i1.20550.

[44] J. Du, C.-M. Vong, and C. L. P. Chen, "Novel Efficient RNN and LSTM-Like Architectures: recurrent and gated broad learning systems and their applications for text classification," *IEEE Transactions on Cybernetics*, vol. 51, no. 3, pp. 1586–1597, Mar. 2021, https://doi.org/10.1109/tcyb.2020.2969705.

[45] M. A. Albadr, S. Tiun, M. Ayob, and F. AL-Dhief, "Genetic Algorithm Based on Natural Selection Theory for Optimization Problems," *Symmetry (Basel)*, vol. 12, no. 11, p. 1758, Oct. 2020, https://doi.org/10.3390/sym12111758.

[46] T. Alam, S. Qamar, A. Dixit, and M. Benaida, "Genetic Algorithm: reviews, implementations, and applications," *International Journal of Engineering Pedadogy*, vol. 10, no. 6, p. 57, Dec. 2020, https://doi.org/10.3991/ijep.v10i6.14567.

[47] A. G. Putrada, N. Alamsyah, I. D. Oktaviani, and M. N. Fauzan, "A Hybrid Genetic Algorithm-Random Forest Regression Method for Optimum Driver Selection in Online Food Delivery," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, vol. 9, no. 4, pp. 1060-1079. 2023. https://doi.org/10.26555/jiteki.v9i4.27014.

[48] S. Nikbakht, C. Anitescu, and T. Rabczuk, "Optimizing the neural network hyperparameters utilizing genetic algorithm," *Journal of Zhejiang University. Science A*, vol. 22, no. 6, pp. 407–426, Jun. 2021, https://doi.org/10.1631/jzus.a2000384.

[49] M. A. A. Albadr, S. Tiun, M. Ayob, and F. T. Al-Dhief, "Genetic algorithm based on natural selection theory for optimization problems," *Symmetry*, vol. 12, no. 11, p. 1758, Oct. 2020, https://doi.org/10.3390/sym12111758.

[50] J. Chai, "Optimizing neural network training with Genetic Algorithms," *Applied and Computational Engineering*, vol. 42, no. 1, pp. 220–224, Feb. 2024, https://doi.org/10.54254/2755-2721/42/20230780.

[51] B. Gaye and A. Wulamu, "Sentiment analysis of text classification algorithms using Confusion Matrix," *in Communications in computer and information science*, pp. 231–241, 2019, https://doi.org/10.1007/978-981-15-1922-2_16.

[52] M. M. Dakwah, A. A. Firdaus, Furizal, and R. A. Faresta, "Sentiment Analysis on Marketplace in Indonesia using Support Vector Machine and Naïve Bayes Method," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, vol. 10, no. 1, pp. 39-53, 2024. https://dx.doi.org/10.26555/jiteki.v10i1.28070.

[53] M. S. Islam, S. Sultana, U. K. Roy, and J. A. Mahmud, "A review on Video Classification with Methods, Findings, Performance, Challenges, Limitations and Future Work," *JITEKI: Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, vol. 6, no. 2, pp. 47-57, 2020. https://dx.doi.org/10.26555/jiteki.v6i2.18978.

BIOGRAPHY OF AUTHORS

Andi Nailul Izzah Ramadhani, is a final-year student in the Faculty of Informatics at Telkom University, Bandung, Indonesia, currently pursuing a bachelor's degree in computer science. Email: andinidhaa@student.telkomuniversity.ac.id.

Erwin Budi Setiawan, is a senior lecturer at the School of Computing at Telkom University, Bandung, Indonesia, with more than ten years of experience in research and teaching within the field of Informatics. Currently, he is an associate professor with research interests in machine learning, people analytics, modeling and simulation, and social media analysis. Email: erwinbudisetiawan@telkomuniversity.ac.id.