Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

Detecting Hoax Content on Social Media Using Bi-LSTM and RNN

Hilman Bayu Aji[1*], Erwin Budi Setiawan[2]

Faculty of Informatics, Bachelor's Degree Program in Informatics, Telkom University, Bandung, Indonesia
Email: [1*]hilmanpbayu@student.telkomuniversity.ac.id, [2]erwinbudisetiawan@telkomuniversity.ac.id
Correspondence Author Email: hilmanpbayu@student.telkomuniversity.ac.id
Submitted: **06/06/2023**; Accepted: **27/06/2023**; Published**: 29/06/2023**

Abstract−Online media, such as websites and applications, have become a communication tool available on the internet. Social media is a part of online media that can be used to spread news, opinions, or even hoaxes, such as through Twitter. Although hoaxes are difficult to eliminate, several systems have been built using deep learning approaches that can process text and images to detect the truthfulness of news. In this study, four systems were built using four deep learning methods, namely Bi-directional Long Short-Term Memory (Bi-LSTM), Recurrent Neural Network (RNN), hybrid RNN-Bi-LSTM, and hybrid Bi-LSTM-RNN. Feature extraction was performed using Term Frequency - Inverse Document Frequency (TF-IDF) and feature expansion was performed using Global Vectors (GloVe). The data used has been adjusted according to the keyword of fake news on mainstream news portals. This study attempted several scenarios to compare the various methods that have been built, with the aim of finding the best method that provides the highest accuracy. The results showed that the Bi-LSTM method had the highest accuracy of 96.48%, while the hybrid Bi-LSTM-RNN method ranked second with an accuracy of 96.36%, followed by the RNN method with an accuracy of 95.49%, and the hybrid RNN-Bi-LSTM method with an accuracy of 95.34%.

Keywords: Hoax; Twitter; Bi-LSTM; RNN; Hybrid

1. INTRODUCTION

The advancement of internet technology has made information retrieval easier through online media compared to physical sources such as newspapers, magazines, and books. People can easily search for information through online media such as websites and Twitter due to the rapid technological advancements of today. Social media platforms like Twitter allow users to quickly and freely share the information they have. Indonesia is the fifth-ranked country in terms of Twitter users globally, with a total of 24 million users [1]. On a global scale, Twitter accumulates 500 million tweets worldwide, totaling approximately 200 billion tweets each year [2]. Although Twitter facilitates communication and information sharing, there are also negative aspects, such as the spread of hoaxes on social media [3]. Hoax is a piece of news created and disseminated by anyone, anywhere, and anytime without considering the truth or accuracy of the information conveyed [4]. The purpose of spreading hoaxes is to damage the reputation of individuals, groups, colleagues, or even friends, and it can result in financial losses [5]. As technology advances, society has shifted to social media for communication, inadvertently leading to the spread of hoaxes and fake news among users. In the age of social media, where everyone can be an "information publisher," filter bubbles and algorithms promoting sensational content can amplify misinformation [5]. As a result, there is a widespread dissemination of irresponsible hoaxes related to viral news on social media, leading to misinformation and causing distress among the public.

Several studies have carried out hoax detection systems using various methods, from feature expansion to deep learning approaches In the study [6], various Deep Neural Network (DNN) models such as LSTM, Bi-LSTM, GRU, Bi-GRU, and 1D-CNN, as well as two classifiers SVM and Naïve Bayes, were tested using datasets from GitHub and several Indonesian news websites. The DNN models utilized word embedding features to map each word in the corpus. On the other hand, the classifier models employed Term Frequency – Inverse Document Frequency (TF-IDF) feature extraction to eliminate common terms from the corpus. The research findings indicated that the DNN models outperformed the classifier models in supervised text classification. In the conducted testing, the Bi-LSTM method with dropout demonstrated the most significant increase in accuracy compared to other models, achieving an accuracy rate of 96.60%, which improved by 2.15% from the testing without dropout.

The study [7] demonstrated that the use of GloVe in corpus construction, when compared to the baseline (TF-IDF N-gram), yielded the best accuracy, reaching 88.59%, which improved by 1.25% from the established baseline. In the study [8], two approaches were employed to detect fake news through text classification: machine learning-based and deep learning-based. The vectorization technique used was the bag-of-words with TF-IDF method to calculate the score of each word. In the deep learning approach, a comparison was made among various methods such as LSTM, Bi-LSTM, GRU, RNN, and CNN. Based on the comparison, the CNN and Bi-LSTM methods were considered the most efficient as they achieved remarkably high accuracy, reaching 97%.

In the study conducted by Aini Hanifa et al., RNN architecture along with LSTM and GRU methods were employed to address the vanishing gradient problem. During the performance analysis using an online news portal dataset, it was found that LSTM achieved an accuracy of 73%, while GRU only reached 64% [9]. Another research conducted by Ajao et al. proposed the use of a hybrid RNN model consisting of two variations, namely LSTM and LSTM-RNN. They performed an analysis on the LIAR dataset and obtained an accuracy result of approximately 82% for the LSTM model. Additionally, the hybrid LSTM-CNN model was implemented and achieved an accuracy of 74% in terms of precision and recall. Among the tested models, the LSTM model showed the highest performance in predicting the dataset. These results suggest that the dataset size affects the accuracy of the model [10].

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

Based on previous research, it has been shown that the Bi-LSTM, RNN, TF-IDF N-gram, and GloVe methods yield better accuracy compared to other methods. Some models in those studies used GloVe as word embedding in deep learning models, while TF-IDF N-gram was used in the classification model. This approach aims to achieve optimal accuracy by combining several methods that have been proven effective in previous research. The combinations to be used are Bi-LSTM, RNN, hybrid RNN-Bi-LSTM, and hybrid Bi-LSTM-RNN as classification models, TF-IDF N-gram as a baseline or feature extraction, and GloVe as a feature expansion in corpus construction.

2. RESEARCH METHODOLOGY

2.1 Research Stages

Figure 1 shown the system design of the hoax detection.

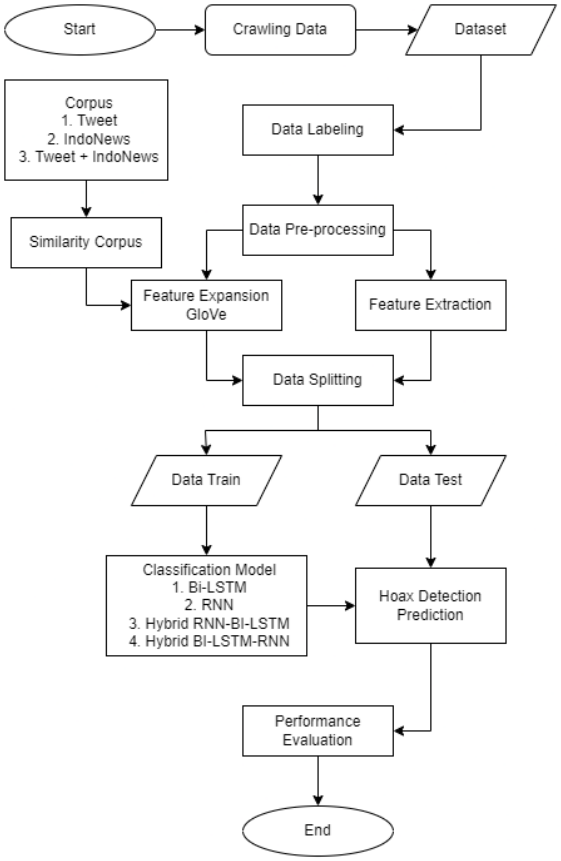

Figure 1. Hoax Detection System

2.2 Crawling Data

For this research, data was collected from Indonesian-language Twitter using the snscrape tool in the Python programming language. The data was collected based on the topics of the Kanjuruhan Tragedy and the Ferdy Sambo case. Before crawling data on Twitter, the researchers ensured that the news spread through social media was indeed hoaxes, using sources such as "Hoax or Not" from Detik.com, "Cek Fakta" from Liputan6.com, "Cek Fakta " from Kompas.com, "Cek Fakta " from Suara.com, and Turnbackhoax.id. The keywords used for data crawling were confirmed to be invalid or hoaxes for each topic, as shown in Tables 1 and 2.

Table 1. Hoax Keywords for the Kanjuruhan Tragedy

No	Keyword
1	Pemukulan jadi penyebab tragedi
2	FIFA bekukan sepak bola Indonesia
3	Komunis uji coba gas beracun
4	Pemain Prancis sindir penggunaan gas air mata
5	Kesaksian penjual dawet

Table 2. Hoax Keywords for the Ferdy Sambo Case

No	Keyword
1	Putri Candrawathi pingsan eksepsi ditolak hakim

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

2	Wasiat Ferdy Sambo sebelum meninggal
3	Bharada Eliezer divonis bebas
4	Pintu rahasia di Rumah Sambo
5	Putri Candrawathi divonis mati
6	Sel mewah Ferdy Sambo
7	Jenazah Ferdy Sambo dikirim ke Magelang
8	Ferdy Sambo akan dieksekusi mati
9	Ferdy Sambo nyaris tewas
10	Ferdy Sambo sujud ke Jokowi mintaampun
11	Ferdy Sambo melarikan diri dari Mako Brimob
12	Putri Candrawathi bunuh diri di rumahnya
13	Kamaruddin Simanjuntak disekap di Bunker
14	Polisi sita puluhan tengkorak dari ruang rahasia
15	Anggota DPR RI disuap Ferdy Sambo
16	Kuat Ma'ruf dibawa kerumah
17	Dua anak Ferdy Sambo dijemput paksa
18	Putri Candrawathi minta ampun
19	Jendral Andika Perkasa panggil tukang di rumah Ferdy Sambo
20	Arwah Brigadir J beri kesaksian
21	Kapolri temukan mayat perempuan tanpa busana
22	Ferdy Sambo satu sel dengan Napoleon Bonaparte
23	Dua organ Brigadir J dijual Ferdy Sambo
24	Sel tahanan Ferdy Sambo kosong
25	Ferdy Sambo divonis bebas

From the above hoax keywords, the topics captured during the crawling process are listed in Table 3. The table includes a total of 25325 tweets, comprising both hoax and non-hoax tweets.

Table 3. Number of Datasets for Each Topic

Topic	Amount
Tragedi Kanjuruhan	2699
Kasus Ferdy Sambo	22626
Total	25325

2.3 Data Labeling

The data labeling process was conducted manually and analyzed by the researcher, considering hoax features. The description of each hoax feature can be found in Table 4.

Table 4. Description for Each Feature [11]

Feature	Description
Username	Whether the usernames used consist of real names or aliases, contain numbers/symbols, include any hateful elements or not.
Display Name	The display name used by Twitter users.
Following > Followers	The number of accounts followed is greater than the number of followers.
Verified Account	Whether the account has been verified or not.
Retweet	The number of retweets of the tweet.
Bio	One of the profile information components on Twitter. It is used to let others know about the user, list interests, or promote a business.
Profile Image	The profile picture identifies the account owner through the displayed photo.
Location	The displayed location in uploaded tweets.

The labeling process is conducted on the dataset before entering the next stage in the hoax detection system. In this system, tweets containing hoax content are labeled as 1, while tweets that do not contain hoax content are labeled as 0. The labeling process is done using indicators of hoaxes such as influencing someone's perspective, tarnishing the reputation of involved parties, inciting conflicts, provocative statements, and hate speech [12]. The total number of labels for the entire data can be seen in Table 5.

Table 5. Number of Labels for All Data

Label	Amount
Hoax	12867
Non-hoax	12458
Total	25325

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

2.4 Data Pre-processing

The data preprocessing process, also known as data preparation, involves a series of steps to analyze raw data and produce quality data. The objective of this phase is to simplify data processing in the classification stage [12]. There are five processes in data preprocessing, namely data cleaning, case folding, stop words, stemming, and tokenizing. Data cleaning is the process of cleaning data by removing non-alphabets, various tags, URLs, punctuation, spaces, and other markup elements [13]. Case folding is the process of converting uppercase letters in the input data into lowercase letters. Stop words involve removing words that are considered irrelevant in determining classifications, such as conjunctions in tweets. In this study, stop words were processed using the Natural Language Toolkit (NLTK), a Python programming language library. Stemming involves removing prefixes or suffixes from a word to obtain its base form. The stemming process in this study was performed using Sastrawi, a Python programming language library. Tokenizing is the process of separating words that are separated by spaces. An example of data preprocessing can be seen in Figure 2.

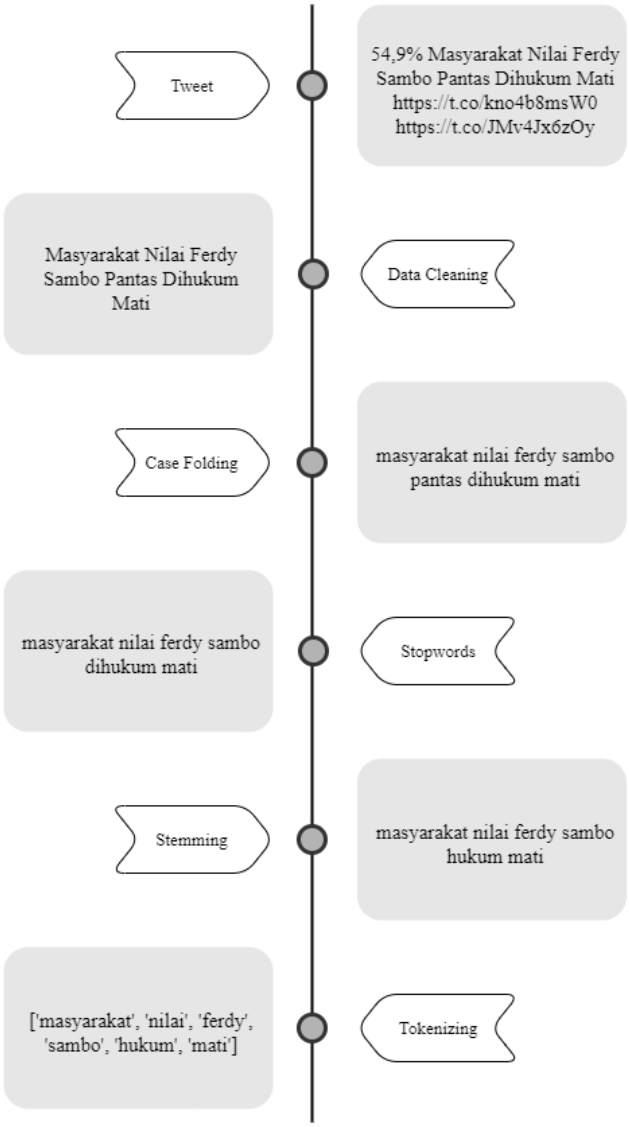

Figure 2. Example of Data Preprocessing Steps

2.5 N-gram

The N-gram model is a probabilistic model that is useful for predicting the next word given a previous word. The N-gram model is also widely used as a feature extraction technique to capture desired word combinations in various tasks, such as predicting correct spellings of limited vocabulary words. In the context of spelling correction, N-gram is used as a collection of N-word sequences. In this study, different types of N-grams, including Unigram, Bigram, and Trigram, as well as combinations of these, are used with the TF-IDF weighting method [7].

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

2.6 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF is a method that is useful for assigning weights to the positions of words in a document. The TF component indicates the frequency of a word in the document, while the IDF component indicates the importance of the word in the document. In TF-IDF, there is a formula for calculating the weight (W) of each document for a specific keyword [14]. In the feature extraction stage, tweets related to the keywords are weighted with TF-IDF scores. After calculating the weight (W) for each document, the W weights are sorted to determine the similarity level between the document and the keyword (higher W values indicate higher similarity levels) [15]. In the calculation of TF-IDF, the weight (W) can be calculated using (1).s

$$W_{ij} = tf_{ij} \times IDF_j, dengan\ IDF_j = log\frac{N}{df_j} \quad (1)$$

The equation represents the formula for TF-IDF, where tf_{ij} represents the term frequency of a word in a document, IDF_j is the inverse document frequency, N is the total number of documents, and df_j is the number of documents containing the searched word.

2.7 Global Vectors (GloVe)

GloVe, short for Global Vectors, is a word embedding technique developed by researchers at Stanford University. GloVe utilizes a global log-bilinear regression model to perform unsupervised learning of word representations in the form of vectors. GloVe consistently outperforms Word2Vec in terms of corpus, vocabulary, window size, and training time. It delivers better and faster results while achieving state-of-the-art performance without sacrificing speed [16]. The goal of using GloVe is to create vector representations of words that capture the semantic differences between words by extracting relationships among words in the corpus. In this study, GloVe is used to construct a corpus from the data and then employed to discover similarities between words in the corpus.

The corpus created using GloVe consists of tweets, news articles, and a combination of tweets and news. The news data used in this study are news articles from various media sources in Indonesia, such as CNN Indonesia, Tempo, Koran Sindo, Kompas, and Republika. Here are examples of words that are similar to "Hukum" (Law) in the word similarity corpus constructed from the tweet dataset, as shown in Table 6.

Table 6. Top 10 Words Similar to "Hukum" (Law)

Rank	Word	Value
1	Layak	0.8920
2	Ringan	0.8758
3	Berat	0.8383
4	Cocok	0.8327
5	Pantas	0.8313
6	Tega	0.8249
7	Negri	0.8224
8	Terap	0.8181
9	Adl	0.8131
10	Dapat	0.8015

Table 6 explains that the rankings are obtained from the similarity values generated by GloVe, starting from the highest rank (Rank-1) to the lowest rank (Rank-10). The number of vocabulary words in each created corpus is presented in Table 7.

Table 7. Number of Words in GloVe Corpus

Corpus	Number of Words
Tweet	20734
IndoNews	79347
Tweet+IndoNews	96359

2.8 Bi-directional Long Short-Term Memory (Bi-LSTM)

BI-LSTM is a model that combines two independent LSTMs, one with normal time order and one with reverse time order, allowing the input to be processed simultaneously. At each time step, the outputs of both LSTMs are concatenated [13], [17]. In many sequence processing tasks, it is important to analyze information from both the future and the past of a point in the sequence [13], [18], [19]. Unlike standard RNNs that only use the previous context, Bi-LSTM is specifically designed to learn long-term dependencies from both sides, and it has been proven to outperform other neural network architectures in phoneme per frame recognition [13], [17], [19]. Several hyperparameters were modified in the Bi-LSTM, including the use of 64 data per batch for 10 epochs, 64 units, and dropout. The created model consists of multiple layers, including Bi-LSTM layers, GlobalAveragePooling1D pooling layer, dense layer, and output layer. As shown in Figure 3, the Bi-LSTM network processes the input sequence in both directions simultaneously.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

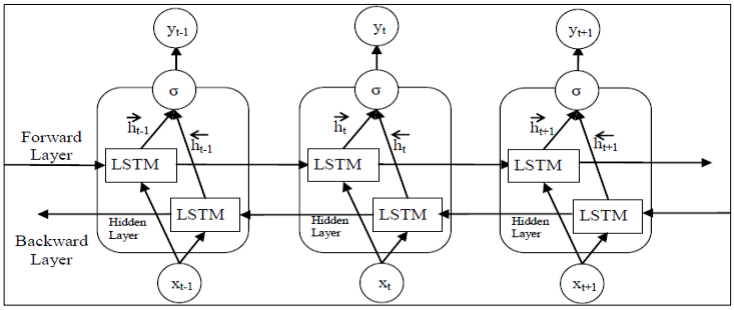

Figure 3. General Architecture of Bi-LSTM [20]

2.9 Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) is a type of feed-forward artificial neural network. RNN can handle variable-length input sequences by using recurrent hidden layers. The activation of these layers at each time step depends on the previous time step, making RNN suitable for capturing long-range contextual information [21]. Long Short-Term Memory (LSTM) is a special type of RNN that can learn long-term dependencies. LSTM is a highly effective solution for addressing the vanishing gradient problem [9]. Several hyperparameters are modified in RNN, including the use of a batch size of 64 for 10 epochs, and dropout. The built model consists of several layers, including an RNN layer that utilizes LSTM, a dense layer, a flatten layer, and an output layer. In LSTM, LSTM cells replace the hidden layer in the basic RNN, as depicted in Figure 4.

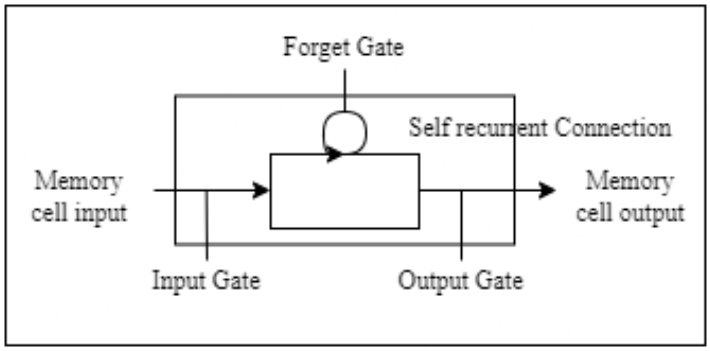

Figure 4. Structure of LSTM Cell

2.10 Hybrid Recurrent Neural Network (RNN) - Bi-directional Long Short-Term Memory (Bi-LSTM)

The hybrid RNN-Bi-LSTM model is the result of combining the RNN and Bi-LSTM models. In the constructed model, there are several components such as the RNN layer with input shape, the Bi-LSTM layer, the dense layer with ReLU activation function, flatten layer, and the output layer with sigmoid activation function. The optimizer used is Adam, and binary crossentropy is used as the loss function. The architecture of the hybrid model built in this research can be seen in Figure 5.

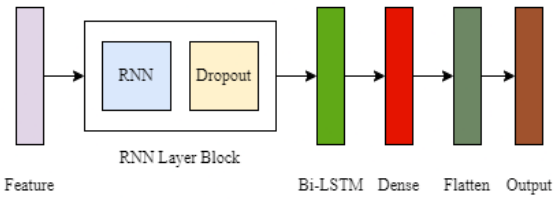

Figure 5. Architecture of the hybrid RNN-Bi-LSTM model.

2.11 Hybrid Bi-directional Long Short-Term Memory (Bi-LSTM) - Recurrent Neural Network (RNN)

This hybrid model is a combination model like the previous models, with the first input shape combination from Bi-LSTM and then RNN. In the constructed model, there are several components such as the Bi-LSTM layer with input shape, the RNN layer, the GlobalAveragePooling1D pooling layer, the dense layer with ReLU activation function, and the output layer with sigmoid activation function. The optimizer used is Adam, and binary crossentropy is used as the loss function. The architecture of the hybrid model built in this research can be seen in Figure 6.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

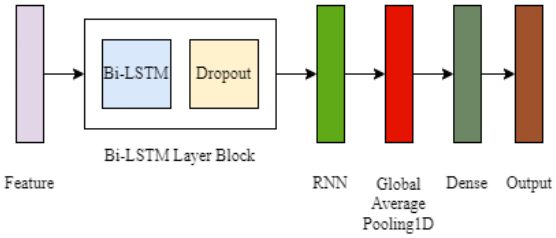

Figure 6. Architecture of the hybrid Bi-LSTM-RNN model.

2.12 System Performance

In this study, the confusion matrix is used as a tool to measure and evaluate the performance of the built model. Table 8 represents the confusion matrix in the presented form.

Table 8. Confussion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

In this context, TP represents True Positive, indicating the correctly predicted positive hoax data. FP represents False Positive, referring to the data that is negative hoax (non-hoax), but predicted as positive hoax. TN represents True Negative, indicating the correctly predicted negative hoax (non-hoax) data. FN represents False Negative, referring to the data that is positive hoax, but predicted as negative hoax (non-hoax). [22]

The system will represent the processed data as predicted values and the desired data as actual data [23]. The results will be indicated by the precision, recall, and accuracy of the dataset processed by the built system [23]. Precision is the number of correctly classified positive samples divided by the total number of positive samples. Recall is the ratio of the total number of positive classifications divided by the total number of positives. F1-Score is the harmonic mean of precision and recall, used to obtain a balanced measure of precision and recall. Accuracy is an evaluation parameter used to measure the accuracy of the built classification system. The values of precision, recall, F1-Score, and accuracy are calculated using (2), (3), (4), and (5).

$$Precision = TP/(TP + FP) \quad (2)$$

$$Recall = TP/(TP + FN) \quad (3)$$

$$F1 - Score = (2 \times Precision \times Recall)/(Precision + Recall) \quad (4)$$

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (5)$$

3. RESULT AND DISCUSSION

In this research, Bi-LSTM, RNN, hybrid RNN-Bi-LSTM, and hybrid Bi-LSTM-RNN were used as classification models to build a hoax detection system. Additionally, the TF-IDF N-gram was used as a baseline or feature extraction, and GloVe was used as a feature expansion. To obtain the best accuracy model, four testing scenarios were conducted. In the first scenario, two tests were conducted on the model to determine the best baseline model by considering the splitting ratio and the type of N-gram used. In the second scenario, testing was conducted on the baseline model by applying a combination of N-gram TF-IDF. In the third scenario, testing was conducted on the baseline model by applying feature expansion using a similarity corpus built by GloVe. In the fourth scenario, testing was conducted on the baseline model by applying several modifications to the dropout hyperparameter. The testing was conducted for 5 (five) iterations in each testing scenario.

3.1 Scenario 1

In the first scenario, two tests were conducted on the model to obtain the optimal baseline model by considering the data splitting ratio and the type of N-gram used. The first test focused on determining the best data splitting ratio, using three options: 90:10, 80:20, and 70:30. Table 9 shows the results of the first test. From the table, it can be observed that using a data separation ratio of 90:10, where 90% of the data is used for training and 10% for testing, resulted in the highest accuracy across all three models. Based on this, it can be concluded that the 90:10 splitting ratio provides the best accuracy results and will be used in the subsequent testing.

Table 9. Comparison Results of Test 1 from Scenario 1

Ratio Splitting	Accuracy (%)			
	RNN	Bi-LSTM	RNN-Bi-LSTM	Bi-LSTM-RNN

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

90:10	**94.67**	**93.48**	**94.51**	**93.80**
80:20	93.97	92.79	94.39	93.24
70:30	93.69	92.19	93.93	92.61

In the second test, the TF-IDF feature extraction method was applied using Unigram, Bigram, and Trigram with a 90:10 splitting ratio. The results of the second test can be seen in Table 10. Based on the table, it can be observed that the second test using TF-IDF feature extraction with Unigram achieved the highest accuracy. The RNN model achieved an accuracy of 94.67%, the Bi-LSTM model achieved an accuracy of 93.48%, the hybrid RNN-Bi-LSTM model achieved an accuracy of 94.51%, and the hybrid Bi-LSTM-RNN model achieved an accuracy of 93.80%.

Table 10. Comparison Results of Test 2 from Scenario 1

TF-IDF	Accuracy (%)			
	RNN	Bi-LSTM	RNN-Bi-LSTM	Bi-LSTM-RNN
Unigram	**94.67**	**93.48**	**94.51**	**93.80**
Bigram	91.82	91.15	91.27	91.47
Trigram	78.95	78.36	78.87	78.76

3.2 Scenario 2

In the second scenario, testing was conducted on the baseline model by applying the N-gram TF-IDF combinations obtained from the previous scenario. Considering that in the previous scenario, TF-IDF Unigram achieved the best accuracy, combinations of N-grams were applied, namely Unigram + Bigram and Unigram + Bigram + Trigram. Table 11 displays the accuracy results of this scenario testing. From the table, using the baseline model with the application of N-gram TF-IDF combinations resulted in improved accuracy. The highest accuracy improvement occurred when using the TF-IDF Unigram + Bigram combination in the RNN model with a relative increase of 0.43%, in the Bi-LSTM model with an increase of 1.10%, in the hybrid RNN Bi-LSTM model with an increase of 0.31%, and in the hybrid Bi-LSTM-RNN model with an increase of 0.39%. Based on the testing results shown in the table, it can be concluded that using the TF-IDF Unigram + Bigram combination provides the best accuracy results for each model and will be used in the next scenario.

Table 11. Comparison Results from Scenario 2

Model	Accuracy (%)		
	Baseline	Unigram + Bigram	Unigram + Bigram + Trigram
RNN	94.67	**95.10 (+0.43)**	94.51 (-0.15)
Bi-LSTM	93.48	**94.59 (+1.10)**	94.35 (+0.86)
RNN-Bi-LSTM	94.51	**94.82 (+0.31)**	94.67 (+0.15)
Bi-LSTM-RNN	93.80	**94.19 (+0.39)**	94.03 (+0.23)

3.3 Scenario 3

In the third scenario, testing was conducted on the baseline model + TF-IDF Unigram + Bigram by applying feature expansion using similarity corpus built by GloVe. Three types of similarity corpus were used, namely Tweet corpus, IndoNews corpus, and Tweet IndoNews corpus, which is a combination of the Tweet corpus and IndoNews corpus. These similarity corpus are ranked based on the highest to lowest similarity level. The more words used, the higher the likelihood of finding similar words that can be utilized.

a. The performance of the baseline RNN model using TF-IDF Unigram + Bigram + GloVe Corpus

The results of applying GloVe to the RNN model are presented in Table 12. From the table, it can be seen that only two tests on the corpus showed an increase in accuracy, while the other tests experienced a decrease in accuracy. The highest accuracy improvement occurred in the test with the top 10 rankings in the similarity of the Tweet IndoNews corpus, with a relative increase of 0.23%. Based on these test results, it can be concluded that using the top 10 rankings in the similarity of the Tweet IndoNews corpus provides the highest accuracy for the RNN model.

Table 12. Comparison Results from Scenario 3 for the RNN model

Rank	Accuracy (%)		
	Baseline + Corpus Tweet	Baseline + Corpus IndoNews	Baseline + Corpus Tweet IndoNews
Top 1	94.11 (-0.55)	93.99 (-0.67)	94.07 (-0.59)
Top 5	94.78 (+0.11)	93.48 (-1.18)	94.51 (-0.15)
Top 10	94.47 (-0.19)	94.31 (-0.35)	**94.90 (+0.23)**
Top 15	94.03 (-0.63)	94.51 (-0.15)	94.55 (-0.11)

b. The performance of the baseline Bi-LSTM model using TF-IDF Unigram + Bigram + GloVe Corpus

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

The results of applying GloVe to the Bi-LSTM model are presented in Table 13. From the table, it can be seen that all tests on the corpus show an increase in accuracy without any decrease in accuracy. The highest accuracy improvement occurs in the test using the top 5 rankings in the similarity of the Tweet corpus, with a relative increase of 2.13%. Based on these test results, it can be concluded that using the top 5 rankings in the similarity of the Tweet corpus provides the highest accuracy for the Bi-LSTM model.

Table 13. Comparison Results from Scenario 3 for the Bi-LSTM model

Rank	Accuracy (%)		
	Baseline + Corpus Tweet	Baseline + Corpus IndoNews	Baseline + Corpus Tweet IndoNews
Top 1	93.88 (+0.39)	93.92 (+0.43)	94.39 (+0.90)
Top 5	**95.61 (+2.13)**	94.63 (+1.14)	95.26 (+1.77)
Top 10	95.22 (+1.73)	94.19 (+0.71)	95.10 (+1.61)

c. The performance of the baseline hybrid RNN-Bi-LSTM model using TF-IDF Unigram + Bigram + GloVe Corpus

The results of applying GloVe to the hybrid RNN-Bi-LSTM model are presented in Table 14. From the table, it can be seen that many of the tests on the Tweet corpus experienced a decrease in accuracy. The highest accuracy improvement occurred in the test with the top 10 rankings in the similarity of the Tweet IndoNews corpus, with a relative increase of 0.43%. Based on these test results, it can be concluded that using the top 10 rankings in the similarity of the Tweet IndoNews corpus provides the highest accuracy for the hybrid RNN-Bi-LSTM model.

Table 14. Comparison Results from Scenario 3 for the hybrid RNN-Bi-LSTM model

Rank	Accuracy (%)		
	Baseline + Corpus Tweet	Baseline + Corpus IndoNews	Baseline + Corpus Tweet IndoNews
Top 1	94.11 (-0.39)	94.74 (+0.23)	93.60 (-0.90)
Top 5	94.63 (+0.11)	94.82 (+0.31)	94.70 (+0.19)
Top 10	92.93 (-1.57)	93.56 (-0.93)	**94.94 (+0.43)**
Top 15	94.35 (-0.15)	93.92 (-0.59)	94.15 (-0.35)

d. The performance of the baseline hybrid Bi-LSTM-RNN model using TF-IDF Unigram + Bigram + GloVe Corpus

The results of applying GloVe to the hybrid Bi-LSTM-RNN model are presented in Table 15. From the table, it can be seen that all tests on the corpus show an increase in accuracy without any decrease in accuracy. The highest accuracy improvement occurs in the test using the top 5 rankings in the similarity of the Tweet corpus, with a relative increase of 1.85%. Based on these test results, it can be concluded that using the top 5 rankings in the similarity of the Tweet corpus provides the highest accuracy for the hybrid Bi-LSTM-RNN model.

Table 15. Comparison Results from Scenario 3 for the hybrid Bi-LSTM-RNN model

Rank	Accuracy (%)		
	Baseline + Corpus Tweet	Baseline + Corpus IndoNews	Baseline + Corpus Tweet IndoNews
Top 1	93.84 (+0.03)	94.07 (+0.27)	95.02 (+1.22)
Top 5	**95.65 (+1.85)**	94.98 (+1.18)	95.18 (+1.38)
Top 10	95.57 (+1.77)	94.90 (+1.10)	95.53 (+1.73)

3.4 Scenario 4

In the fourth scenario, testing was conducted on the baseline model TF-IDF Unigram + Bigram + GloVe Corpus obtained from the previous scenario by making several changes to the dropout hyperparameter. This testing was performed using dropout variations of 20%, 30%, 40%, 50%, 60%, 70%, and 80%. The highest accuracy improvement occurred in the Bi-LSTM model with a relative increase of 3.00%. The results of applying dropout to each model can be found in Table 16.

Table 16. Comparison Results from Scenario 4 with Dropout

Model	Accuracy (%)							
	Baseline	20%	30%	40%	50%	60%	70%	80%
RNN	94.67	94.23 (-0.43)	**95.49 (+0.82)**	94.94 (+0.27)	95.18 (+0.51)	94.51 (-0.15)	92.73 (-1.93)	90.76 (-3.90)
Bi-LSTM	93.48	95.34 (+1.85)	95.73 (+2.25)	96.01 (+2.52)	96.28 (+2.80)	95.93 (+2.44)	96.05 (+2.56)	**96.48 (+3.00)**

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

RNN-Bi-LSTM	94.51	**95.34 (+0.82)**	95.22 (+0.71)	95.26 (+0.75)	94.35 (+0.15)	93.68 (-0.82)	92.45 (-2.05)	89.14 (-5.36)
Bi-LSTM-RNN	93.80	95.77 (+1.97)	95.73 (+1.93)	95.97 (+2.17)	95.85 (+2.05)	96.05 (+2.25)	96.24 (+2.44)	**96.36 (+2.56)**

3.5 Discussion

In the test results, there was an improvement in accuracy across all testing scenarios. The implementation of N-gram combinations extends the N-gram and provides more context to predict possible words in the data. Specifically, the combination of TF-IDF Unigram + Bigram showed the most optimal results compared to other combinations in the test. The use of similarity corpus can transform words that initially have zero values into TF-IDF values of similar words, allowing previously zero-valued words to have an impact on the test results. In the RNN model, the application of similarity corpus showed the greatest improvement when using the Tweet IndoNews similarity corpus. In the Bi-LSTM and hybrid Bi-LSTM-RNN models, the application of similarity corpus showed the greatest improvement when using the Tweet similarity corpus.

Various dropout variations were applied to enable the models to make predictions with higher accuracy. In the RNN model, the highest accuracy improvement occurs when using a dropout rate of 30%, while in the hybrid RNN-Bi-LSTM model, the highest accuracy improvement occurs when using a dropout rate of 20%. The accuracy improvement continued until a dropout rate of 50%, but consistently decreased when the dropout rate was increased to 80%. The application of an 80% dropout rate yielded optimal testing results in the Bi-LSTM model and the hybrid Bi-LSTM-RNN model. Based on these results, the confusion matrix for the best model can be seen in Figure 7. In these figures, the X-axis represents the predicted labels, while the Y-axis represents the actual labels.

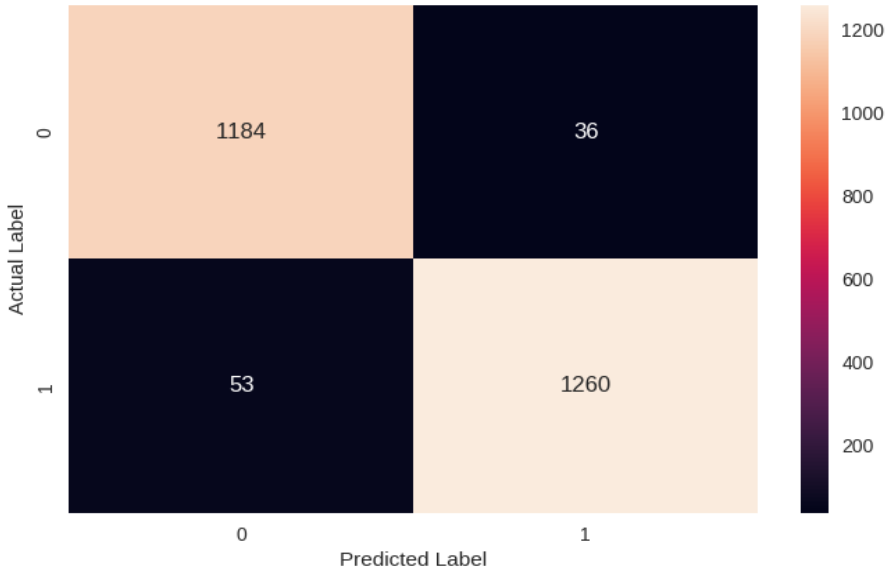

Figure 7. Confusion Matrix

In the figure, there are 1184 correct positive predictions (True Positive/TP) and 1260 correct negative predictions (True Negative/TN). Additionally, there are 53 false positive predictions (False Positive/FP) and 36 false negative predictions (False Negative/FN). In the above confusion matrix calculations, the predictions that result in assigning the hoax label to TP and the non-hoax label to TN are correct. With proportional values of TP and TN, these calculation results indicate an optimal level of accuracy.

The graph in Figure 8 shows the increase in accuracy in each scenario that has been conducted. The values taken are the most optimal accuracy values in each scenario. The following is the comparison of the highest relative increase for each model compared to the baseline: (1) In the RNN model, the highest relative increase compared to the baseline is 0.82%, which occurred in the test using the combination of baseline + TF-IDF Unigram + Bigram + GloVe Top 10 similarity of the Tweet IndoNews corpus + dropout 30%, with an accuracy of 95.49%. (2) In the Bi-LSTM model, the highest relative increase compared to the baseline is 3.00%, which occurred in the test using the combination of baseline + TF-IDF Unigram + Bigram + GloVe Top 5 similarity of the Tweet corpus + dropout 80%, with an accuracy of 96.48%. (3) In the hybrid RNN-Bi-LSTM model, the highest relative increase compared to the baseline is 0.82%, which occurred in the test using the combination of baseline + TF-IDF Unigram + Bigram + GloVe Top 10 similarity of the Tweet IndoNews corpus + dropout 20%, with an accuracy of 95.34%. (4) In the hybrid Bi-LSTM-RNN model, the highest relative increase compared to the baseline is 2.56%, which occurred in the test using the combination of baseline + TF-IDF Unigram + Bigram + GloVe Top 5 similarity of the Tweet corpus + dropout 80%, with an accuracy of 96.36%.

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

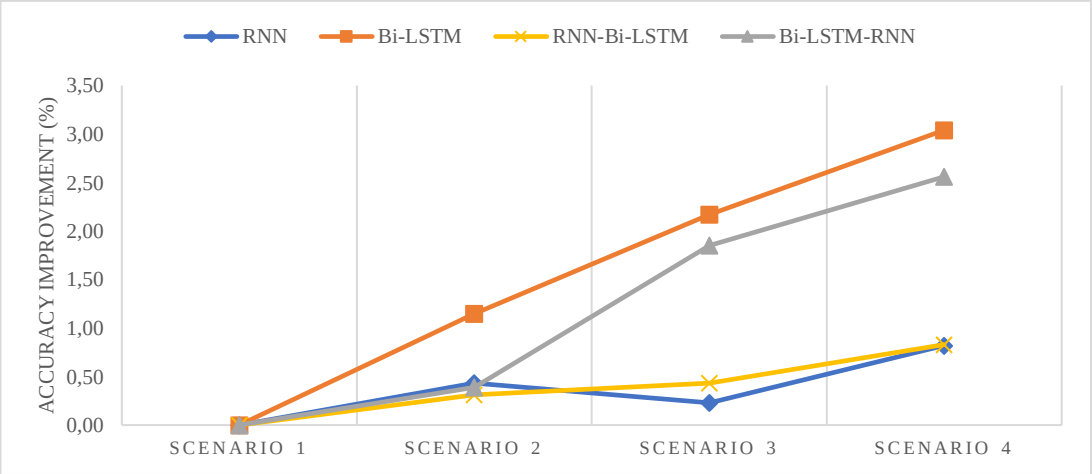

Figure 8. Accuracy Improvement Graph

4. CONCLUSION

In this study, predictions were made on fake news or hoaxes spread on social media, especially on Twitter. The methods used included RNN, Bi-LSTM, hybrid RNN-Bi-LSTM, and hybrid Bi-LSTM-RNN. The dataset consisted of 25325 tweets with a balanced distribution between hoax and non-hoax labels. TF-IDF feature extraction method was used to assign weights to words, and GloVe was used to create the corpus for feature expansion. Testing was conducted with four scenarios, which demonstrated that data splitting ratio, types, and combinations of TF-IDF, corpus in feature expansion, and dropout influenced the performance of each model. The results showed that the Bi-LSTM model achieved the highest accuracy of 96.48%, with the most significant improvement in accuracy compared to the other three methods. The hybrid Bi-LSTM-RNN method achieved an accuracy of 96.36%, followed by the RNN method with an accuracy of 95.49%, and the RNN-Bi-LSTM method with an accuracy of 95.34%. Based on the test results, it can be concluded that the Bi-LSTM model remained superior to the combination of Bi-LSTM and RNN models. Additionally, the use of dropouts proved to enhance accuracy. As a suggestion for future research, other methods such as Restricted Boltzmann Machines (RBMs) could be explored, or adjustments to other hyperparameters could be made to achieve even higher levels of accuracy.

REFERENCES

[1] We Are Social, "Jumlah Pengguna Media Sosial di Dunia Capai 4,2 Miliar," *Databoks*, p. 2021, 2021, [Online]. Available: https://databoks.katadata.co.id/datapublish/2021/02/18/jumlah-pengguna-media-sosial-di-dunia-capai-42-miliar
[2] L. Samaras, E. García-barriocanal, and M. Sicilia, "Sentiment analysis of COVID-19 cases in Greece using Twitter data," *Expert Syst. Appl.*, p. 120577, 2023, doi: 10.1016/j.eswa.2023.120577.
[3] I. Muslim, K. Karo, S. Dewi, and P. M. Fadilah, "Hoax Detection on Indonesian Tweets using Naïve Bayes Classifier with TF-IDF," vol. 4, no. 3, pp. 914–919, 2023, doi: 10.47065/josh.v4i3.3317.
[4] M. Sargsyan, "' ALL TRUTH IS RELATIVE ' OR HOW TO NOT BE FOOLED IN THE POST-TRUTH AGE ?," vol. 19, no. 1, pp. 69–82, 2023, doi: 10.46991/AFA/2023.19.1.069.
[5] D. S. M. and Hairunnisa, "The Phenomenon of Fake News (Hoax) in Mass Communication: Causes, Impacts, and Solutions Deddy," *Open Access Indones. J. Soc. Sci.*, vol. 4, no. 1, pp. 132–142, 2021, doi: 10.37275/oaijss.v6i3.161.
[6] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 704–712, 2021, doi: 10.1016/j.procs.2021.01.059.
[7] F. Anistya and E. B. Setiawan, "Hate Speech Detection on Twitter in Indonesia with Feature Expansion Using GloVe," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1044–1051, 2021, doi: 10.29207/resti.v5i6.3521.
[8] W. H. Bangyal *et al.*, "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," *Comput. Math. Methods Med.*, vol. 2021, 2021, doi: 10.1155/2021/5514220.
[9] A. Hanifa, S. A. Fauzan, M. Hikal, and ..., "Perbandingan Metode LSTM dan GRU (RNN) untuk Klasifikasi Berita Palsu Berbahasa Indonesia," *Din. Rekayasa*, vol. 17, no. 1, pp. 33–39, 2021, [Online]. Available: http://dinarek.unsoed.ac.id/jurnal/index.php/dinarek/article/view/436
[10] O. Ajao, D. Bhowmik, and S. Zargari, "Fake news identification on Twitter with hybrid CNN and RNN models," *ACM Int. Conf. Proceeding Ser.*, no. July, pp. 226–230, 2018, doi: 10.1145/3217804.3217917.
[11] C. W. Kencana, E. B. Setiawan, and I. Kurniawan, "Hoax Detection System on Twitter using Feed-Forward and Back-Propagation Neural Networks Classification Method," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 4, pp. 655–663, 2020, doi: 10.29207/resti.v4i4.2038.
[12] W. Pamungkas and S. Suryani, "Deteksi Hoax Untuk Berita Hoax Covid 19 Indonesia Menggunakan CNN," vol. 8, no. 5. pp. 10264–10276, 2021.
[13] V. D. Derbentsev, V. S. Bezkorovainyi, A. V. Matviychuk, and ..., "Sentiment Analysis of Electronic Social Media Based on Deep Learning," no. M3e2 2022, pp. 163–175, 2023, doi: 10.5220/0011932300003432.
[14] A. Fauzi, E. B. Setiawan, and Z. K. A. Baizal, "Hoax News Detection on Twitter using Term Frequency Inverse Document

Building of Informatics, Technology and Science (BITS)
Volume 5, No 1, June 2023 Page: 114−125
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v5i1.3585

Frequency and Support Vector Machine Method," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012025.

[15] V. Amrizal, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.

[16] P. M. Brennan, J. J. M. Loan, N. Watson, P. M. Bhatt, and P. A. Bodkin, "GloVe: Global Vectors for Word Representation," *Br. J. Neurosurg.*, vol. 31, no. 6, pp. 682–687, 2017, doi: 10.1080/02688697.2017.1354122.

[17] A. Kurniawati, E. Mulyanto, Y. Kusnendar, and ..., "Automatic note generator for Javanese gamelan music accompaniment using deep learning," vol. 9, no. 2, pp. 231–248, 2023, doi: 10.26555/ijain.v9i2.1031.

[18] X. Zhao, K. Sun, S. Gong, and X. Wu, "RF-BiLSTM Neural Network Incorporating Attention Mechanism for Online Ride-Hailing Demand Forecasting," *Symmetry (Basel).*, vol. 15, no. 3, p. 670, 2023, doi: 10.3390/sym15030670.

[19] H. Li, Y. Lu, H. Zhu, and Y. Ma, "A Novel AB-CNN Model for Multi-Classification Sentiment Analysis of e-Commerce Comments," *Electron.*, vol. 12, no. 8, 2023, doi: 10.3390/electronics12081880.

[20] P. Bahad, P. Saxena, and R. Kamal, "Fake News Detection using Bi-directional LSTM-Recurrent Neural Network," *Procedia Comput. Sci.*, vol. 165, no. 2019, pp. 74–82, 2019, doi: 10.1016/j.procs.2020.01.072.

[21] B. Roy *et al.*, "Hybrid Deep Learning Approach for Stress Detection Using Decomposed EEG Signals," pp. 1–19, 2023, doi: 10.3390/diagnostics13111936.

[22] I. M. Mubaroq and E. B. Setiawan, "The Effect of Information Gain Feature Selection for Hoax Identification in Twitter Using Classification Method Support Vector Machine," *Indones. J. ...*, vol. 5, no. September, pp. 107–118, 2020, doi: 10.21108/indojc.2020.5.2.499.

[23] F. Ismayanti and E. B. Setiawan, "Deteksi Konten Hoax Berbahasa Indonesia Di Twitter Menggunakan Fitur Ekspansi Dengan Word2vec," *eProceedings ...*, vol. 8, no. 5, pp. 10288–10300, 2021, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15697%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15697/15410