Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

Comperative Analysis of Indonesian Political Fake News Detection using IndoBERT-Bi-GRU-Attention Models: Evaluating Performance on Narratives and News Headlines Datasets

Juliana Damayanti Manurung*, Ronsen Purba

Informatics Faculty, Master of Information Technology, Mikroskil University, Medan, Indonesia
Email: [1,*]juliana.manurung@mikroskil.ac.id, [2]ronsen.purba@mikroskil.ac.id
Correspondence Author Email: juliana.manurung@mikroskil.ac.id
Submitted: **06/02/2025**; Accepted: **27/02/2025**; Published**: 07/03/2025**

Abstract−The instant and massive spread of fake news on social media negatively impacts public trust in the media and news agencies. In politics, fake news is often used by politicians to gain support ahead of elections. Detecting fake news in Indonesia poses a significant challenge, especially for communities vulnerable to misinformation. This study aims to develop a new model that combines IndoBERT with Bi-GRU and Attention. Additionally, a comparison is made between the main model and two word embedding models, FastText and GloVe. The tests were conducted on datasets of headlines and news narratives separately. Data was sourced from CNN, Tempo.co, Kompas, and TurnBackHoax.ID. The results show that the IndoBERT-Bi-GRU-Attention model with FastText excelled on the headline dataset with an accuracy of 99.76% and an F1-Score of 99.61%, while the main IndoBERT-Bi-GRU-Attention model excelled on the narrative dataset with an accuracy of 99.08% and an F1-Score of 98.40%. This research demonstrates that IndoBERT can be combined with Bi-GRU, significantly contributing to the development of fake news detection models.

Keywords: Hoax; Politic; IndoBERT; Deep Learning; Word Embedding

1. INTRODUCTION

Fake news refers to the spread of false or misleading information that is deliberately presented as factual news with the intent to deceive, manipulate, or influence public opinion [1]. The distribution of fake news has increased significantly every year. From 2018 to May 2023, KOMINFO identified 11,642 pieces of content as hoaxes, with 1,373 of them classified as political hoaxes[2].Various studies have been conducted to improve fake news detection using deep learning models. One such study compared the performance of CNN and LSTM models enhanced with Word2Vec for word embedding in fake news detection. The results showed that the CNN model achieved higher accuracy than the LSTM model. However, the study also identified an issue of overfitting, and researchers suggested improvements in data preprocessing and word embedding representation, particularly for Indonesian-language datasets[3].

Another study by Junita Amalia et al. (2022) investigated fake news detection using an Indonesian dataset by combining the Bi-LSTM model with Word2Vec. Their model demonstrated good accuracy, which was highly dependent on parameters such as window size, embedding size, and the number of Bi-LSTM units. However, the model performed poorly when tested on datasets consisting only of news headlines. The short sentence structure often resulted in biased predictions compared to narrative-based datasets, making it difficult to distinguish between factual and fake news accurately [4].

To improve word representation beyond Word2Vec, a study by Ryan Adipradana et al. (2021) compared LSTM, GRU, Bi-LSTM, and Bi-GRU models using FastText and GloVe embeddings. The results indicated that the Bi-GRU-FastText model achieved the highest accuracy among all tested models. This study also found that FastText embedding consistently outperformed GloVe in capturing complex word vectors. However, researchers suggested that using original Indonesian news articles instead of translated texts could further enhance the model's effectiveness [5].

Another recent study by Angeline Karen et al. (2023) compared the performance of Bi-LSTM and IndoBERT for fake news detection. Their findings revealed that Bi-LSTM achieved better accuracy than IndoBERT. The study noted that IndoBERT performed well in understanding relationships between short sentences but required fine-tuning for optimal results [6]. Meanwhile, a more recent study by Rifky Maulana et al. (2023) explored the use of IndoBERT combined with LSTM. Their model demonstrated that IndoBERT's ability to process short text relationships could be leveraged effectively when integrated with LSTM. However, their research was limited by a small and less diverse dataset, highlighting the need for further improvements in dataset collection and preprocessing [7].

One of the major challenges faced in previous studies is the heterogeneous nature of textual data. Sentence lengths in news articles vary significantly, which can affect the model's performance. Additionally, detecting fake news based on headlines alone often introduces bias, making it difficult to achieve consistent results across different datasets.

To address these challenges, this study proposes a fake news detection model using IndoBERT-Bi-GRU, while also comparing its performance with IndoBERT-Bi-GRU combined with FastText and GloVe for word embedding. Each model will incorporate an attention mechanism to enhance performance by allowing the model to focus on the most relevant information. This approach aims to achieve higher accuracy and more refined results than previous deep learning models [8], [9].

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

Given the issues and research gaps identified in past studies, this study seeks to contribute to the advancement of fake news detection by leveraging IndoBERT's strong contextual understanding, Bi-GRU's sequence modeling capabilities, and attention mechanisms to improve model performance. The findings from this research could help create more effective tools for combating fake news in the Indonesian language.

2. RESEARCH METHODOLOGY

2.1 Research Stages

This research will focus on six key processes: reading the dataset, preprocessing the data, splitting the data, modeling IndoBERT-Bi-GRU-Attention, and conducting testing. Figure 1 is an illustration of the classification research stages.

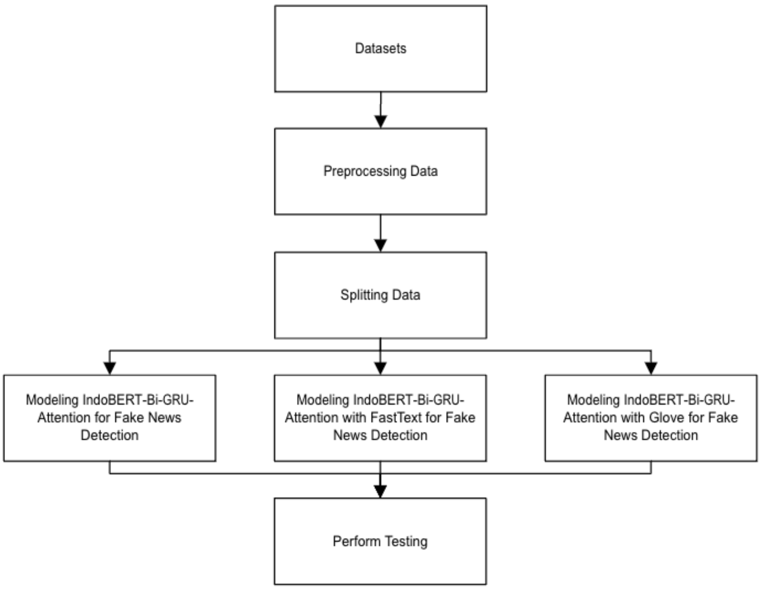


Figure 1. Research Stages

The following are the steps that will be carried out in modeling fake news detection using IndoBERT-Bi-GRU-Attention, namely:

a. First stage, reading the dataset consists of titles, narratives, and fact/hoax labels sourced from CNN, Tempo.co, Kompas, and TurnBackHoax.id between 2020 and 2023. In total, these platforms provide 10,875 narrative records and 17,278 titles, excluding non-political news from TurnBackHoax.id.
b. The preprocessing stages include case folding, punctuation removal, stopword removal, and stemming.
   1. Case folding involves converting all letters in a text to either lowercase or uppercase, eliminating any distinction between uppercase and lowercase letters. This process aims to maintain uniformity in text representation, ensuring that the model treats words the same regardless of their letter casing.
   2. Removing punctuation entails stripping away all non-alphanumeric symbols and punctuation marks, including periods, commas, exclamation points, and more, from the text. This step aims to eliminate unnecessary details and simplify the text, facilitating smoother and more efficient processing.
   3. Stopword removal refers to the process of eliminating common, low-information words like "and," "or," and "from" from the text[10]. These words, known as stopwords, are often removed to retain only the most significant terms. This step helps streamline the text, allowing for a focus on key words and reducing the complexity of the data when constructing a model.
   4. Stemming is the process of removing inflections and affixes from words, leaving only the root or base form. The goal of stemming is to reduce word variations in the text, allowing words with the same meaning but different forms to be treated as identical by the model.
c. On splitting data, the preprocessed dataset is divided into two main parts: 90% of the total data is used for training, while the remaining 10% is split into 5% for validation and 5% for testing. Additionally, experiments are conducted with different splits: 80% training data with 20% validation (10% validation and 10% testing) and 70% training data with 30% validation (15% validation and 15% testing). This division aims to achieve high accuracy, as a sufficiently large training set helps the model learn effectively, while the validation set assists in testing and optimizing the model. The testing set is used for final evaluation, providing insight into how well the model performs on unseen data.
d. The next step is implementing the following hybrid models :
   1. IndoBERT-Bi-GRU-Attention hybrid model

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

2. IndoBERT-Bi-GRU-Attention hybrid model with FastText
3. IndoBERT-Bi-GRU-Attention hybrid model with GloVe

2.2 Word Embedding

Word embedding, also known as word vector representation, is a method that allows words with similar meanings to have similar representations. This technique is essential for creating word vectors that store semantic and syntactic information based on context within a large corpus. In Natural Language Processing studies, word embedding is explored to find effective approaches in language modeling [11], [12], [13].

FastText is a library developed by Facebook that can be used for word embedding. It is an improvement over Word2Vec, which was previously known as a method for word embedding [11]. FastText addresses some of the limitations of Word2Vec by incorporating morphological characteristics through the processing of subwords for each word. A "subword" refers to the character-level n-grams of a given word. Architecture model of FastText is illustrated in Figure 2.

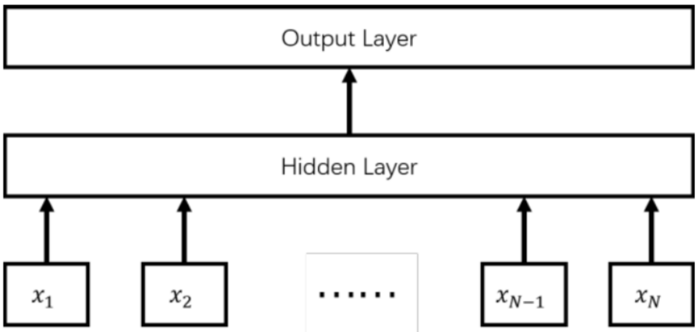


Figure 2. Architecture Model of Fast Text

FastText works by learning the vector representation of each n-gram and then obtaining the word vector by summing the n-gram vectors. The learning method of FastText using n-grams is illustrated in Figure 3.

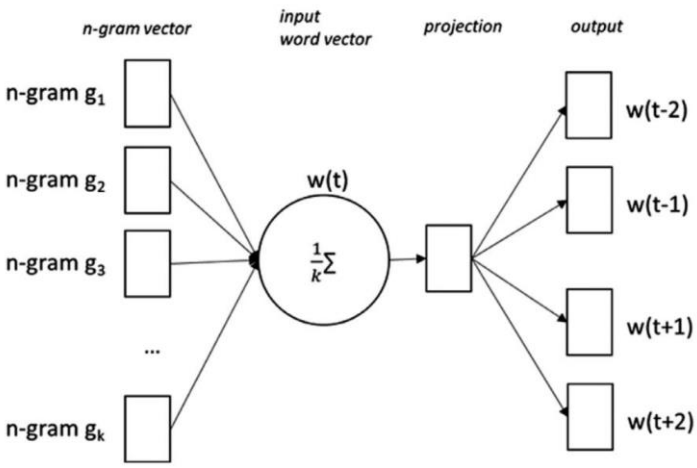


Figure 3. FastText Learning Method (window size = 5)

This method creates word embeddings at the character level and captures morphological features. As a result, FastText can generate word vectors for new words that were not present in the training data. Even for previously unseen words, the sum of n-grams can represent their vector. Additionally, the morphological aspect is preserved, which enhances performance in syntactic evaluations compared to dictionary-based word embedding algorithms [12].

GloVe (Global Vectors for Word Representation) is an unsupervised learning algorithm for obtaining word vector representations. This algorithm combines the global word co-occurrence statistics from a corpus to produce vector representations that capture semantic meaning and relationships between words. GloVe is well known for its computational efficiency and ability to generate high-performing vector representations for various natural language processing tasks. The function examines the co-occurrence probabilities for words w_j and w_k, as well as the correlations between words w_i and w_j to analyze the relationships between words.

$$F(w_i, w_j, w_k) = \frac{P_{ik}}{P_{jk}} \quad (1)$$

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

GloVe is based on a log-bilinear regression model to learn word vector representations. This model minimizes a cost function that measures errors in predicting word co-occurrences.

Compared to Word2Vec, GloVe is more computationally efficient as it does not require negative sampling iterations. GloVe (Global Vectors for Word Representation) is a word embedding algorithm that generates word vector representations based on word co-occurrence within a text corpus. It is well known for its computational efficiency and ability to produce high-performing vector representations for various natural language processing tasks, such as sentiment classification, machine translation, and question answering.

GloVe outperforms Word2Vec in terms of efficiency, handling rare words, leveraging semantic information, and overall performance in NLP tasks.

2.3 IndoBERT

IndoBERT is a transformer-based model similar to BERT [14] but trained exclusively as a masked language model using the Huggingface framework. It follows the default configuration of BERT-Base (uncased).

IndoBERT consists of 12 hidden layers, each with a dimension of 768, 12 attention heads, and a feed-forward hidden layer with a dimension of 3,072 [15]. Fajri Koto et al. (2021) modified the Huggingface framework to process separate text streams for different document blocks and set the training process to use 512 tokens per batch. The IndoBERT model was trained using an Indonesian WordPiece vocabulary of 31,923 tokens.

Overall, IndoBERT was trained on more than 220 million words collected from three main sources: (1) Indonesian Wikipedia (74 million words); (2) news articles from Kompas, Tempo, and Liputan6 (a total of 55 million words); and (3) an Indonesian Web Corpus (90 million words) [15], [16], [17].

The architecture of the IndoBERT model is illustrated in Figure 4, with additional details provided by Ruben Stefanus (2021) in the classification of Indonesian news using IndoBERT, as shown in Figure 5 [18].

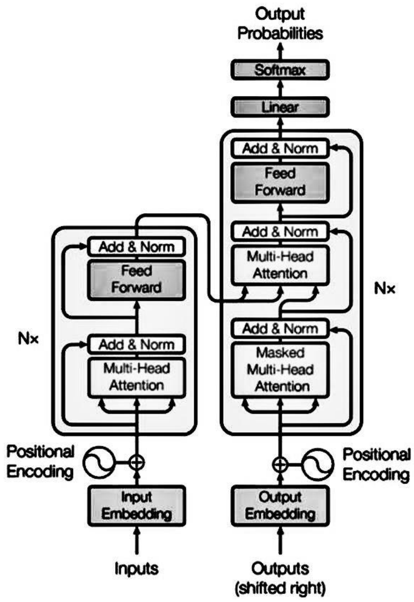


Figure 4. IndoBERT Model Architecture

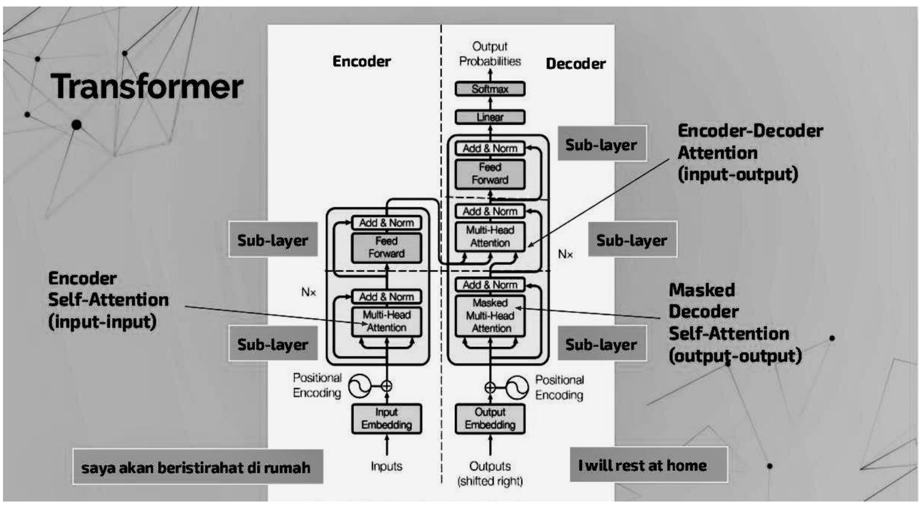


Figure 5. Detail IndoBERT Architecture

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

2.4 Bidirectional Gated Recurrent Unit

Bi-GRU (Bidirectional Gated Recurrent Unit) is a modified Recurrent Neural Network (RNN) architecture designed to process sequential data in both directions. This architecture combines two GRU (Gated Recurrent Units) networks that operate in opposite directions, allowing it to capture contextual information from surrounding words.

Bi-GRU is well known for its ability to capture long-range dependencies and generate rich contextual representations, making it highly effective for natural language processing tasks such as named entity recognition, sentiment analysis, and machine translation.

Bi-GRU consists of two GRUs connected in parallel. The first GRU processes sequential data from the beginning to the end, while the second GRU processes it in reverse, from the end to the beginning. The outputs from both GRUs are then combined to produce a final contextual representation for each word in the sequence.

GRU utilizes a gating mechanism to control the flow of information through the network. This mechanism enables GRU to focus on relevant information while disregarding irrelevant details, which is crucial for natural language processing tasks where word context is highly significant.

Generally, Bi-GRU is more computationally efficient than other RNNs, such as Bi-LSTM, because it has fewer parameters. Additionally, Bi-GRU is better at capturing long-range dependencies in sequential data compared to Bi-LSTM, as it processes contextual information from both directions.

2.5 Attention Mechanism

The Attention Mechanism is a technique that models the dependency between target outputs and input sequences. It enables data-driven models to focus on informative segments while ignoring less relevant ones, thereby enhancing the interpretability of learned representations and improving the overall performance of algorithms.

The attention mechanism applied in this research is based on the work of Bahdanau et al. The following section outlines the steps and formulas used in this study [9]:

a. Alignment Scores

$$e_i = \tanh\,(XiW + b) \tag{2}$$

For each timestep i in the input, the alignment score e is computed by multiplying the input Xi with the weight W, adding the bias b, and then applying the tanh activation function.

b. Calculate Attention Weights (Softmax)

$$\alpha_i = \frac{\exp\,(e_i)}{\sum \exp\,(e_j)} \tag{3}$$

The softmax function converts the alignment scores e into attention weights α. These weights represent probabilities indicating the importance of each timestep in the input.

c. Calculate Context Vector

$$C = \sum_i (X_i . \alpha_i) \tag{4}$$

The context vector C is the result of the weighted sum of the input X based on the attention weights α. It combines information from all timesteps while considering the importance of each timestep.

d. Apply GELU Function

$$GELU(C) = 0.5\,.\,C\,.\left(1 + \text{erf}\left(\frac{C}{\sqrt{2}}\right)\right) \tag{5}$$

The Gaussian Error Linear Unit (GELU) function is applied to the context vector C. This function introduces non-linearity, helping the model capture complex relationships within the data. It provides a smoother mechanism for determining whether a neuron should be activated or not.

e. Final Formula

$$C = GELU\left(\sum_i (X_i . \alpha_i)\right) = 0.5\,.\left(\sum_i (X_i . \alpha_i)\right).\,(1 + \text{erf}\,(\frac{C\,(\sum_i (X_i . \alpha_i))}{\sqrt{2}})\Big) \tag{6}$$

X represents an input tensor with a shape of (batch_size, timesteps, input_dim), where batch_size denotes the number of samples in a batch, timesteps corresponds to the sequence length, and input_dim is the dimensionality of each input feature. The attention mechanism is defined by the attention weight W, which has a shape of (input_dim, 1), and the attention bias b, with a shape of (timesteps, 1). The alignment score e is computed to measure the relevance of each timestep, and the attention weight α is obtained by applying the softmax function to the alignment score e, ensuring that the weights sum to 1. The context vector C is then derived as a weighted sum of the input tensor X using the attention weights α. The exponential function exp and the error function are mathematical operations that may be used in the computation of the alignment score or other components of the attention mechanism, depending on the specific implementation. Together, these elements form the core of an attention-based model, enabling it to focus on the most relevant parts of the input sequence.

2.6 Confusion Matrix

The evaluation process in machine learning is crucial to determine whether the model used performs well. In text mining, the model evaluation method used is the Confusion Matrix. The Confusion Matrix is a performance measurement for classification problems in machine learning where the output can be two or more classes. The

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

Confusion Matrix is a table with four different combinations of predicted values and actual values. The predicted values are the outputs from the program, which can be Positive and Negative, while the Actual Values are the true values, which can be True and False. The representation of the classification results is divided into four categories: True Positive, True Negative, False Positive, and False Negative[19], [20], [21]. The classification can be seen in Figure 4.

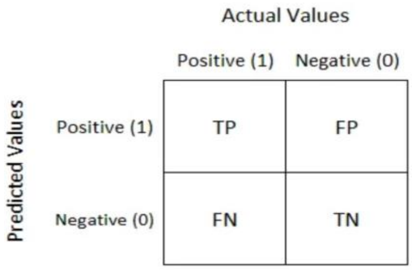


Figure 4. Confusion Matrix

The metrics used in the Confusion Matrix are as follows:
Accuracy measures how accurately the model correctly classifies the data, calculated as:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{7}$$

Precision evaluates the accuracy between the requested data and the results predicted by the model, calculated:

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

Recall or Sensitivity assesses the model's success in retrieving information.

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

F1-Score represents the weighted average of precision and recall.

$$F1 - Score = \frac{2*Recall*Precision}{Recall+Precision} \tag{10}$$

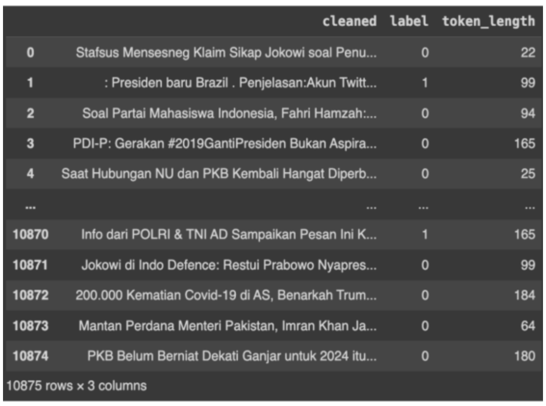
Accuracy is suitable as a performance benchmark for algorithms when the dataset contains nearly equal amounts of False Negative and False Positive data (symmetric). However, if this condition is not met, the F1-Score is a better metric for evaluation.

3. RESULT AND DISCUSSION

3.1 Result

3.1.1 Dataset

This stage involves reading news article dataset from CNN, Kompas, Tempo.co, and TurnBackHoax.id, covering the period from 2020 to April 2023. The dataset is accessible at https://www.kaggle.com/code/linkgish/indobert-training-cleaned-news/input The fields used are fulltext and label, where label 0 (Fact) comes from CNN, Kompas, and Tempo.co, while label 1 (Hoax) is sourced from TurnBackHoax.id. Figure 6 and Figure 7 are representations of dataset that is structured in two forms: one focusing only on news titles and the other based on full narratives,.



	cleaned	label	token_length
0	Stafsus Mensesneg Klaim Sikap Jokowi soal Penu...	0	22
1	: Presiden baru Brazil . Penjelasan:Akun Twitt...	1	99
2	Soal Partai Mahasiswa Indonesia, Fahri Hamzah:...	0	94
3	PDI-P: Gerakan #2019GantiPresiden Bukan Aspira...	0	165
4	Saat Hubungan NU dan PKB Kembali Hangat Diperb...	0	25
...	...	...	...
10870	Info dari POLRI & TNI AD Sampaikan Pesan Ini K...	1	165
10871	Jokowi di Indo Defence: Restui Prabowo Nyapres...	0	99
10872	200.000 Kematian Covid-19 di AS, Benarkah Trum...	0	184
10873	Mantan Perdana Menteri Pakistan, Imran Khan Ja...	0	64
10874	PKB Belum Berniat Dekati Ganjar untuk 2024 itu...	0	180

10875 rows × 3 columns

Figure 6. Narrative Datasets

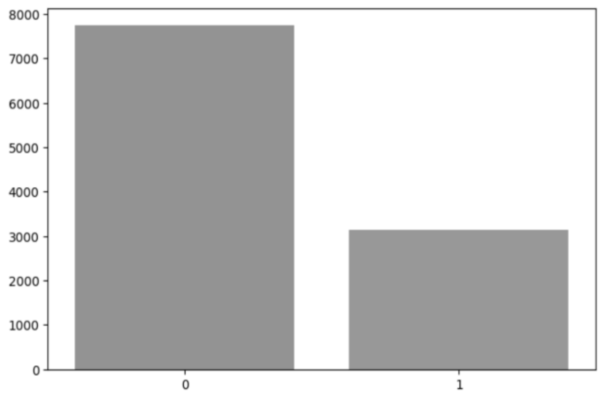


Figure 7. Distribution Narrative Datasets

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

3.1.2 Preprocessing

At this stage, preprocessing will be applied to the data from each news platform. This process includes case folding, punctuation removal, stopword removal, and stemming. The results can be seen in Figure 8.

	cleaned	label	token_length
0	stafsus mensesneg klaim sikap jokowi penundaan...	0	22
1	presiden brazil penjelasan akun twitter plana ...	1	99
2	partai mahasiswa indonesia fahri hamzah berten...	0	94
3	pdi p gerakan aspirasi manuver politik sekjen ...	0	165
4	hubungan nu pkb hangat diperbincangkan indones...	0	25
...	...	...	...
10870	info polri tni ad pesan keluarga kawan kawan p...	1	165
10871	jokowi indo defence restui prabowo nyapres per...	0	99
10872	kematian covid as trump prioritaskan politik a...	0	184
10873	mantan perdana menteri pakistan imran khan kor...	0	64
10874	pkb berniat dekati ganjar dominan ganjar jazil...	0	180

10875 rows × 3 columns

Figure 8. Preprocessed data

3.1.3 Modeling

The modeling process is conducted to develop a model capable of detecting fake news with the highest possible accuracy, precision, recall, and F1-score. The inputs used in this modeling process include narrative data, title data, and labels. The hyperparameters employed in this model are based on previous research, with some modifications to reference the best practices, such as adjustments to the loss function, optimizer, learning rate, batch size, Bi-GRU units, and the number of epochs. Following this, the dataset to be used is selected. The dataset is then divided into two main parts: training data and validation data (where the validation data is further split into validation and testing sets). Two variations of data splitting ratios are applied sequentially: 70:30, 80:20, and 90:10.

The model designed in this research focuses on an IndoBERT-Bi-GRU-Attention architecture, comparing it with combinations of FastText and GloVe word embeddings. As a result, the models produced will differ only in the embedding layer.

In the model development process, the input data is processed in two stages. The first stage involves passing the data through a selected BERT model; this research uses IndoBERT-lite-base-p1. Referring to its development at https://github.com/IndoNLP/indonlu/blob/master/utils/functions.py, the BERT model used is Albert. The Bi-GRU component consists of 128 units, accompanied by four dense layers with 256, 128, and 64 units using the ReLU activation function, and one layer using the Sigmoid function. Layer normalization is applied, and the model is optimized using the Adam Optimizer with a specified loss function. The three models to be developed have the following layer structures:

a. IndoBERT-Bi-GRU-Attention Model: This hybrid model utilizes the Albert Model, a 128-unit Bi-GRU, an attention layer, and layer normalization for the BERT model layer. The embedding layer includes a Dense layer with 128 units, which is expanded to accommodate the output from the embedding layer. A 128-unit Bi-GRU is applied, followed by an attention layer and layer normalization for the embedding model layer. The outputs from both layers are combined and passed through four dense layers with 256, 128, 64, and 1 units, respectively. The model is optimized using the Adam Optimizer, and the loss function used is Binary Crossentropy.
b. IndoBERT-Bi-GRU-Attention with FastText Model: This hybrid model utilizes the Albert Model, a 128-unit Bi-GRU layer, an attention layer, and layer normalization for the BERT model. The embedding layer is based on FastText embedding, followed by another 128-unit Bi-GRU layer, an attention layer, and layer normalization for the embedding model. The two constructed layers are then combined and pass through four dense layers with sizes 256, 128, 64, and 1, concluding with the Adam optimizer. The loss function used is the same as in the previous model, namely Binary Crossentropy.
c. IndoBERT-Bi-GRU-Attention with GloVe Model: This hybrid model utilizes the Albert Model, a 128-unit Bi-GRU layer, an attention layer, and layer normalization for the BERT model. The embedding layer is based on GloVe embedding, followed by another 128-unit Bi-GRU layer, an attention layer, and layer normalization for the embedding model. The two constructed layers are then combined and pass through four dense layers with sizes 256, 128, 64, and 1, concluding with the Adam optimizer. The loss function used is Binary Crossentropy.

3.2 Discussion

This section focuses on testing and evaluating the results of the conducted experiments. The testing is performed to compare the three developed models using the same hyperparameters. The hyperparameters used are in Table 1.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

Table 1. Hyperparameter

Hyperparameter	
Loss function	Binary cross-entropy
Optimizer	Adam Optimizer
Learning rate	1×10^{-5}
Batch size	32
Bi-GRU unit	128
Epoch	20

Each model will use the same hyperparameters and values to ensure the best possible performance based on evaluation metrics. In terms of measurement and testing, the model with the lowest loss function value is considered the best. Additionally, the model with the highest accuracy, precision, recall, and F1-score is selected as the optimal model.

This research focuses solely on evaluating whether IndoBERT-Bi-GRU-Attention can serve as a new model for fake news detection. Therefore, the comparison is conducted between models with the same structure but different word embeddings.

3.2.1 Performance Comparison of Models on the News Title Dataset

The following is a comparison and discussion of each model when tested using the news title dataset.

Table 2. Performance Comparison of Models on the 70:30 News Title Dataset

Model	Loss	Accuracy	Precision	Recall	F1-Score
IndoBERT-Bi-GRU-Attention	0.0415	99.24%	**99.79%**	97.90%	98.84%
IndoBERT-Bi-GRU-Attention with FastText	0.0486	99.14%	99.33%	98.03%	98.67%
IndoBERT-Bi-GRU-Attention with GloVe	**0.0357**	**99.34%**	99.61%	**98.19%**	**98.89%**

Table 3. Performance Comparison of Models on the 80:20 News Title Dataset

Model	Loss	Accuracy	Precision	Recall	F1-Score
IndoBERT-Bi-GRU-Attention	0.0457	99.18%	99.40%	97.86%	98.63%
IndoBERT-Bi-GRU-Attention with FastText	**0.0096**	**99.76%**	**100%**	99.22%	**99.61%**
IndoBERT-Bi-GRU-Attention with GloVe	0.0169	99.53%	99.03%	**99.42%**	99.22%

Table 4. Performance Comparison of Models on the 90:10 News Title Dataset

Model	Loss	Accuracy	Precision	Recall	F1-Score
IndoBERT-Bi-GRU-Attention	0.0391	99.35%	99.40%	98.62%	99.01%
IndoBERT-Bi-GRU-Attention with FastText	**0.0328**	**99.53%**	**99.60%**	98.83%	**99.22%**
IndoBERT-Bi-GRU-Attention with GloVe	0.0646	98.84%	96.97%	**99.22%**	98.08%

Based on Tables 2, 3, and 4, this study examines whether the generated model is effective in detecting news when each available dataset contains fewer than 30 words. Referring back to the research by Junita Amalia [4], the model tended to make incorrect predictions due to bias caused by excessively short titles. Similarly, in the study by Angeline Karen [6], the Bi-LSTM model struggled to detect fake news based on titles.

Thus, this study successfully demonstrates that IndoBERT can be combined with Bi-GRU, and after incorporating the attention mechanism, the results surpass previous research. The model is able to focus more effectively on important information, leading to improved performance. Even when the dataset consists of single-sentence titles, detection remains accurate. This research also confirms that IndoBERT's advantage in handling short words can be leveraged in the latest model, delivering optimal performance.

3.2.1 Performance Comparison of Models on the News Narrative Dataset

The following is a comparison and discussion of each model when tested using the news narrative dataset.

Table 5. Performance Comparison of Models on the 70:30 News Narrative Dataset

Model	Loss	Accuracy	Precision	Recall	F1-Score
IndoBERT-Bi-GRU-Attention	**0.0538**	**98.77%**	**98.91%**	96.81%	**97.85%**
IndoBERT-Bi-GRU-Attention with FastText	0.0964	98.28%	96.42%	**97.65%**	97.04%
IndoBERT-Bi-GRU-Attention with GloVe	0.0619	98.71%	98.07%	97.44%	97.75%

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

Table 6. Performance Comparison of Models on the 80:20 News Narrative Dataset

Model	Loss	Accuracy	Precision	Recall	F1-Score
IndoBERT-Bi-GRU-Attention	0.1138	**98.53%**	97.14%	**97.76%**	**97.45%**
IndoBERT-Bi-GRU-Attention with FastText	**0.0964**	98.52%	**97.44%**	97.44%	97.44%
IndoBERT-Bi-GRU-Attention with GloVe	0.1208	98.06%	96.20%	97.12%	96.66%

Table 7. Performance Comparison of Models on the 90:10 News Narrative Dataset

Model	Loss	Accuracy	Precision	Recall	F1-Score
IndoBERT-Bi-GRU-Attention	0.0651	**99.08%**	**98.08%**	**98.71%**	**98.40%**
IndoBERT-Bi-GRU-Attention with FastText	0.0964	98.52%	97.44%	97.44%	97.44%
IndoBERT-Bi-GRU-Attention with GloVe	**0.0604**	98.71%	98.06%	97.43%	97.74%

Based on Tables 5, 6, and 7, the IndoBERT-Bi-GRU-Attention model effectively captures word meanings and classifies data more accurately, positively impacting the evaluation metrics.

The results show that IndoBERT-Bi-GRU-Attention outperforms in nearly all tests on the narrative dataset, while IndoBERT-Bi-GRU-Attention with FastText excels in most tests on the news title dataset. However, even without additional word embedding models, the main model already performs well. This indicates that adding specific word embeddings is not necessarily required to achieve high evaluation metrics. Compared to previous studies, all three models successfully surpass prior research results.

4. CONCLUSION

This study demonstrates that the IndoBERT-Bi-GRU-Attention model effectively detects fake news in both news titles and narratives, achieving 99.08% accuracy on narratives and 99.76% on titles when combined with FastText. By utilizing different hyperparameters from previous research, the model improved accuracy and successfully addressed bias in short-text classification, a challenge in earlier studies. While integrating external word embeddings like FastText and GloVe enhanced performance, it also doubled the number of parameters, increasing training time. However, the attention mechanism significantly boosted the model's effectiveness by focusing on critical information, even with a simpler deep learning structure. These results confirm IndoBERT-Bi-GRU-Attention as a reliable approach for Indonesian fake news detection. For future research, exploring different hyperparameter values, incorporating additional evaluation metrics such as AUC-ROC, AUC-PR, F1-Score, and MCC, and experimenting with alternative deep learning models could further enhance performance and provide a more comprehensive assessment of fake news detection systems.

REFERENCES

[1] D. S. M and Hairunnisa, “The Phenomenon of Fake News (Hoax) in Mass Communication: Causes, Impacts, and Solutions,” *Open Access Indonesia Journal of Social Sciences*, vol. 6, no. 3, pp. 980–988, May 2023, doi: 10.37275/oaijss.v6i3.161.

[2] “Sampai Mei 2023, Kominfo Identifikasi 11.642 Konten Hoaks – Ditjen Aptika.” Accessed: May 02, 2024. [Online]. Available: https://aptika.kominfo.go.id/2023/06/sampai-mei-2023-kominfo-identifikasi-11-642-konten-hoaks/

[3] A. A. Kurniawan and M. Mustikasari, “Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia,” Jurnal Informatika Universitas Pamulang, vol. 5, no. 4, pp. 2622–4615, 2020, doi: 10.32493/informatika.v5i4.7760.

[4] J. Amalia, J. Pakpahan, M. Pakpahan, Y. Panjaitan, F. Informatika dan Teknik Elektro, and I. Teknologi Del, “Model Klasifikasi Berita Palsu Menggunakan Bidirectional LSTM Dan Word2Vec Sebagai Vektorisasi,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 4, 2022

[5] R. Adipradana, B. P. Nayoga, R. Suryadi, and D. Suhartono, “Hoax analyzer for indonesian news using rnns with fasttext and glove embeddings,” *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 2130–2136, Aug. 2021, doi: 10.11591/eei.v10i4.2956.

[6] A. Karen, M. Christopher, N. N. Qomariyah, I. B. K. Manuaba, and A. K. Anom, “Clarifact-AI: Detecting Fake News in Indonesian Language with Natural Language Processing Using BiLSTM and IndoBERT Models,” in *10th International Conference on ICT for Smart Society, ICISS 2023 - Proceeding*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICISS59129.2023.10291714.

[7] M. Rifky Maulana *et al.*, “Penerapan Metode Lstm Dalam Pembuatan Sistem Pendeteksi Berita Palsu Berbahasa Indonesia,” *Syntax : Jurnal Informatika*, vol. 12, no. 02, pp. 43–51, Nov. 2023, doi: 10.35706/SYJI.V12I02.9924.

[8] T. E. Trueman, A. K. J., P. Narayanasamy, and J. Vidya, “Attention-based C-BiLSTM for fake news detection,” *Appl Soft Comput*, vol. 110, Oct. 2021, doi: 10.1016/j.asoc.2021.107600.

[9] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.0473

[10] S. Sarica and J. Luo, “Stopwords in Technical Language Processing,” *PLoS One*, vol. 16 8, no. 8 August, Aug. 2021, doi: 10.1371/JOURNAL.PONE.0254937.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2459−2468
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6938

[11] "Word Embedding dengan FastText – Master of Computer Science." Accessed: May 06, 2024. [Online]. Available: https://mti.binus.ac.id/2021/12/31/word-embedding-dengan-fasttext/

[12] J. Choi and S. W. Lee, "Improving FastText with inverse document frequency of subwords," *Pattern Recognit Lett*, vol. 133, pp. 165–172, May 2020, doi: 10.1016/j.patrec.2020.03.003.

[13] J. Xu and Q. Du, "A deep investigation into fasttext," in *Proceedings - 21st IEEE International Conference on High Performance Computing and Communications, 17th IEEE International Conference on Smart City and 5th IEEE International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2019*, Institute of Electrical and Electronics Engineers Inc., Aug. 2019, pp. 1714–1719. doi: 10.1109/HPCC/SmartCity/DSS.2019.00234.

[14] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Online]. Available: https://github.com/tensorflow/tensor2tensor

[15] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," Online. [Online]. Available: https://huggingface.co/

[16] F. Koto Jey Han Lau Timothy Baldwin, "INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization." [Online]. Available: https://huggingface.co/huseinzol05/

[17] A. B. Y. A. Putra, Y. Sibaroni, and A. F. Ihsan, "Disinformation Detection on 2024 Indonesia Presidential Election using IndoBERT," in *2023 International Conference on Data Science and Its Applications, ICoDSA 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 350–355. doi: 10.1109/ICoDSA58501.2023.10277572.

[18] "indo-bert-news-classification/transformer.png at master · rubentea16/indo-bert-news-classification · GitHub." Accessed: May 07, 2024. [Online]. Available: https://github.com/rubentea16/indo-bert-news-classification/blob/master/transformer.png

[19] B. Juarto and Yulianto, "Indonesian News Classification Using IndoBert," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 454–460, Feb. 2023, Accessed: Feb. 13, 2025. [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/2654

[20] S. Narkhede, "Understanding Confusion Matrix." Accessed: May 08, 2024. [Online]. Available: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

[21] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci.*, vol. 507, pp. 772–794, Jan. 2020, doi: 10.1016/J.INS.2019.06.064.