

# Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news

Danny Yongky Yefferson<sup>1</sup>, Viriyaputra Lawijaya<sup>1</sup>, Abba Suganda Girsang<sup>2</sup>

<sup>1</sup>BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

<sup>2</sup>Department Computer Science, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

## Article Info

### Article history:

Received Feb 21, 2023

Revised Oct 11, 2023

Accepted Dec 2, 2023

### Keywords:

Article

Detection system

Hoax

Indobert

Long short-term memory

machine learning

News

## ABSTRACT

The world has entered an era that technology has developed far. Due to rapid technological development, information is easily spread. However, not all information spread through social media is factual information. Responding to this social phenomenon, we initiated to create a hoax detection system using the combined method of Indo bidirectional encoder representations from transformers (IndoBERT) and long short-term memory (LSTM). The dataset used in this study are obtained through the process scraping on the site turnbackhoax.id and cable news network (CNN) Indonesia. We decided to use the IndoBERT-LSTM method to detect hoaxes, using IndoBERT as the feature extractor and LSTM as the classification layer can be an effective method because of its advantages in managing and understanding Indonesian language. The results show that the IndoBERT-LSTM model achieved an accuracy of 93.2%, precision of 92%, recall of 89.7%, and F1-score of 90.8%. From a total of 5876 data composed of a total of 1998 factual news and 3878 hoax data. The hoax detection system using IndoBERT-LSTM is a promising approach for detecting hoaxes accurately and efficiently. This model has the potential to make a significant impact in the fight against the spread of Hoaxes.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Abba Suganda Girsang

Department Computer Science, BINUS Graduate Program-Master of Computer Science

Bina Nusantara University

Jakarta 11480, Indonesia

Email: agirsang@binus.edu

## 1. INTRODUCTION

The world that we now live in has entered at the end of 2022 and entered 2023, which means that technology has developed far and can be said to have developed rapidly. We have entered an era where the dissemination of information is an easy thing to do because it can be done anytime and anywhere. This information can be obtained from all types of media, both from communication media in the form of television to social media in the form of cell phones. According to a survey conducted by the Statista which was held in January 2022, it was stated that there were over 196 million internet users in Indonesia, making it the fourth-largest online market in the world after China, India, and the United States. And it can be assumed, that number has changed a lot in 2023 considering that it has passed more or less than a years ago. Furthermore, According to kemp, there are 48% of the country's overall population who are active on the social media [1].

However, of course not all information spread through social media is factual information. Because information disseminated through social media can come from anyone, not all of this information can be justified. There is a chance that the information submitted is false or incorrect information. In general, this

information is referred to as a hoax. A hoax is defined as fake news or information deliberately spread through the internet with the intention of confusing and misleading readers [2]. Hoax news spreads rapidly through various social media networks, including WhatsApp, LINE, Facebook, Instagram, Twitter, and more, making it accessible to a wide audience. According to the classification by Chris Fleming and John O'Carroll, hoaxes can be categorized into two types: those seeking to deceive and those intending to commit fraud. Some hoaxes are highly performative, while others consist mainly of textual performances [3].

In social media life, no individual can be free from hoaxes. Everyone has the potential to become a victim of a hoax because there are so many of them. Based on one of the studies that have been conducted on hoaxes in Indonesia, one of the triggers for hoaxes to arise is social networking media [2]. Social network media, as a communication tool, transforms social patterns so as to give rise to various types of communication patterns. One form of the new communication pattern is a communication pattern that is considered informative. However, this information does not always convey correct information (indication of hoax) [2]. Hoaxes can be spread due to following several factors and trends that can influence this phenomenon. Among the existing factors, ideology, political affiliation, economy, and popularity are the most common factors that trigger hoaxes. Hoaxes that discuss these four factors/topics tend to lead opinions and offend the interests of their supporters [2].

According to Boese [4] there are two factors that are required for hoax to be present, the first one is public audience. Hoaxes typically need a public audience to be successful because the goal of a hoax is often to deceive or manipulate people on a large scale. Without a wide audience, a hoax would be relatively ineffective because it would not have a significant impact. Hoaxes frequently target a public audience because they aim to exploit people's fears, biases, or desires. By presenting false or misleading narratives that manipulate these emotions, hoaxes become more convincing, leading people to believe in them or act based on them. A hoax directed at a small or limited audience would likely have a reduced impact because it lacks the emotional resonance and the ability to spread through word of mouth or social media [4].

Another reason why hoaxes need a public audience is that they often have a political or social agenda. By creating a false narrative or spreading false information, a hoax can be used to influence public opinion, create controversy, or push a specific agenda. This can be particularly effective when the hoax targets a large or influential group of people, such as the media, government officials, or opinion leaders [4]. Additionally, public audience that lacks of literacy/interest in reading is an easy prey to immediately caught in trap, believing the news and end up being fooled by hoax news [4]. This is in line with Miller and McKenna [5] statement, that the literacy of the Indonesian people was ranked 60<sup>th</sup> compared to other countries.

Another factor is required by hoax to be present, is deception. Hoaxes need deception because they are essentially a form of fraud or trickery that relies on creating a false narrative or presenting false information to a target audience. Without deception, a hoax would not be able to achieve its intended effect [4]. One reason why deception is a necessary component of a hoax is that it allows the hoaxer to control the narrative and manipulate people's beliefs or actions. By presenting false information or creating a fake story, the hoaxer can steer the audience in a specific direction, whether it's to believe a conspiracy theory, support a particular political cause, or purchase a fraudulent product [4]. Another reason why deception is essential to a hoax is that it creates a sense of intrigue or mystery that can capture people's attention and imagination. A hoax that presents itself as a legitimate discovery or revelation, for example, can be more captivating and generate more interest than a straightforward announcement or statement [4].

Finally, deception is often used in hoaxes to mask the true intentions of the hoaxer. In some cases, a hoax may be used to distract from a more significant issue or to undermine public trust in a specific institution or group. By using deception, the hoaxer can hide their true motivations and avoid being held accountable for their actions [4]. In reality, no individual can be free from hoaxes. Everyone has the potential to become a victim of a hoax because there are so many of them. Everyday there will be a new hoax content produced. According to pew research center found that only 64% fake news/hoax that is detected or found by hoax checking website whereas the remainder of 36% remain undetected and may be still believed as factual news.

Considering that the time leading up to the presidential election in Indonesia is near, namely next year (2024), we speculate that many hoax news about the Indonesian election could spread on social media networks later in the year. From the hoaxes that are spread, it can cause divisions in the family due to different political views and politics pitting one against the other. Responding to this social phenomenon, we took the initiative to create a hoax detection system using the combined method of IndoBERT and Long short-term memory (LSTM). we design the system in order to detect the truth of a news.

The problem addressed in the paper is the detection of fake news or hoaxes in the Indonesian language, which can lead to misleading information and potentially harmful consequences. The proposed solution is to use a machine learning model based on the IndoBERT-LSTM architecture to classify news as

either real or fake. The model's accuracy was compared with two other machine learning models. Hoax is A phenomenon that needs to be addressed. To deal with hoaxes, we propose to create a hoax detection system. Previously, various approaches have been proposed to detect fake news in English and other languages, including using feature-based approaches, rule-based approaches, and machine learning techniques. However, these approaches may not be suitable for detecting fake news in Indonesian, given the unique characteristics of the language. Moreover, there's no research has been conducted before those uses this specific, particular method. The paper's new contribution is the development of an Indonesian fake news detection model using the all new IndoBERT-LSTM architecture. Additionally, the paper highlights the importance of addressing the problem of fake news in the Indonesian language, given its potential impact on society.

Several hoax detection methods have been developed in Indonesia, one of which is hoax detection using the IndoBERT algorithm. Before discussing the hoax detection model that uses the IndoBERT algorithm, it is better if we discuss IndoBERT itself first. IndoBERT is a language model pre-trained BERT language model for Indonesian. This is one of the first monolingual BERT models for Indonesian, trained following best practices in the field [6]. IndoBERT was created because even though Indonesian is spoken by nearly 200 million people and is the 10th most widely spoken language in the world, Indonesian is underrepresented in studies using natural language processing (NLP). Researchers used IndoLEM (Indonesian version of GLUE) to evaluate the performance of IndoBERT. From the experiments and evaluations that have been carried out, the IndoBERT algorithm is able to produce accuracy part-of-speech (POS) tagging from the IndoLEM task namely Morpho-syntax and sequence labeling tasks with a figure of 96.8%, thus showing that Indobert was able to achieve sophisticated and good performance [7].

Back to talking about the hoax detection model, a hoax detection model that uses the IndoBERT algorithm under the title "Indobert for Indonesian fake news detection." The hoax detection model uses datasets collected from turnbackhoax.id website with 3,465 fake news and 766 real news. From the model made, it can be proven that the IndoBERT model is able to detect hoaxes with an accuracy score of 94.66% at precision, recall, and F1-score [8]. F1-Score is a machine learning evaluation metric that measures model accuracy. The numbers can be calculated using (1),

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (1)$$

Besides IndoBERT, another algorithm method that can be applied to detect hoaxes is the LSTM method. LSTM comes from the convolutional neural network (CNN), specifically recurrent neural network (RNN). CNN is one of the deep learning methods that can overcome the shortcomings of classical learning methods. However, CNN is still unable to process data that is sequential (data in sequence). Unlike the RNN. RNNs are specifically designed to handle subsequent data. However, even though it is designed to handle sequential data that works sequentially, RNN has captured limitations and dependencies [9].

The LSTM model was designed to overcome this limitation. The LSTM model is a variant of the RNN method. The LSTM model is able to overcome long-term dependence by remembering long-term information and is good for application in cases of sentiment analysis or classification, such as the detection of hoax news [10].

Furthermore, in terms of our chosen research method, specifically LSTM, there are prior studies that have utilized this method to construct hoax detection models. For instance, there's a hoax detection model that combines the CNN and LSTM algorithms, titled "Fake news classification bimodal using convolutional neural network and long short-term memory." This model was trained on a dataset collected from Kaggle, achieving impressive accuracy rates of 99.7% on the training data and 97.5% on the test data [9].

In searching for methods and references, we found a study that also used the LSTM algorithm in designing a hoax detection model, under the title "A text classification method based on a convolutional and bidirectional long short-term memory model." The hoax detection model uses datasets collected from AGnews, DBPedia, Text\_Class, Yahoo!Answers, Yelp.p, and Amazon.F. Of the models that have been developed, the model is capable of producing accuracy between 72.24% to 92.5% [11].

Subsequent research also involves LSTM using several other deep learning methods, some of which are Bi-LSTM and ID-CNN. Developed in 2021 with the title "Hoax analyzer for Indonesian news using deep learning models." The hoax detection model uses a dataset collected from Indonesian language data.mendeley.com with 372 valid news and 228 hoax news for the initial stage and an additional 128 and 223 fake news taken from github.com, and 49 other hoax news retrieved manually from Kompas.com for the second phase. Of the models that have been developed, the model is able to search for accuracy, precision, recall, and F1-macro with an accuracy of 95.6% LSTM, 96.6% Bi-LSTM and 97% ID-CNN [12].

Similar to the previous case, the researcher also found two studies that also used BERT when looking for methods and references. The first research uses the transformer network, and the second research

uses a BERT-based deep learning approach. The first research model with the title “Indonesia's fake news detection using transformer network.” The hoax detection model uses a dataset also collected from turnbackhoax.id, 1,116 data obtained consisting of valid news and hoax news. The developed model obtains accuracy, precision, recall, F1-score with accuracy of BERT 90%, CNN 74%, Bi-LSTM 85%, Hybrid CNN-BiLSTM 74% [13].

For the second research, under the title “FakeBERT: Fake news detection in social media with a BERT-based deep learning approach.” The hoax detection model uses a dataset in the form of a collection of fake and genuine news that was disseminated during the 2016 US Presidential Election. From the developed model it is known that FakeBERT is a classification model that was created and combined with the word embedding BERT model to produce an accuracy of 98.9% and then the same thing was done to several other models such as the BERT embedding model with CNN (92.7%) and LSTM (97.55%), then the GloVe embedding model with CNN (91.5%) and LSTM (97.25%) [14].

In order to broaden our horizons, we are also looking for other methods in developing hoax detection models. The other method in question is a method that does not involve BERT or LSTM. One of the studies that developed the hoax detection model with the title “Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization.” The hoax detection model uses the Naïve Bayes (NB) method, which is optimized using particle swarm optimization (PSO) with datasets originating from google.com. From the developed model, the model is able to produce an average accuracy of 77%, where each news is correctly identified as a hoax in the range of accuracy between 66% and 91% [15].

In the process of searching whether the IndoBERT-LSTM hybrid method had ever been used to create a hoax detection model, we found zero results. However, we found another study that developed a Sentiment Analysis Model by stacking IndoBERT-BiLSTM. This is relevant because we are able to get a glimpse on the performance done by IndoBERT mixed with branch of LSTM method, BiLSTM. Developed by Andry Chowanda and Yohan Muliono with the title “Indonesian sentiment analysis model from social media by stacking BERT and BI-LSTM.” The model uses the IndoBERT-BiLSTM method with datasets originating from various social media. From the developed model the results shows that the model are overfitted proven by their accuracy. The model is able to produce the accuracy with 95.17%, 70.25% and 69.09% for training accuracy, validation accuracy and testing accuracy respectively [16]. From the result we caught a glimpse on how well mixing IndoBERT and LSTM thus, we finally decided to use it as the main method for this research. After combining the indoBERT and LSTM models, the IndoBERT-LSTM models that have been developed will be compared by performance with some other models.

## 2. PROPOSED METHODS

### 2.1. Dataset

The dataset used in this study uses datasets obtained through the process scraping on the site turnbackhoax for hoaxes and CNN Indonesia for factual news. Hoax data was taken starting from April 25, 2021 to February 02, 2023 and January 29, 2023 to February 02, 2023 for factual news. Each hoax datasets are labeled with the number 0 while each factual news datasets are labeled with the number 1. It is intended that the model is able to more easily distinguish which news are hoax and which are factual news.

### 2.2. Methodology

We decided to use the combined IndoBERT-LSTM method in developing an Indonesian hoax detection model. The first reason we use IndoBERT rather than BERT is because the dataset used in making the model uses Indonesian so that it requires the capabilities of the IndoBERT model so that the data can be processed and understood by the model. The second reason we use LSTM is because LSTM is one of the most common RNN methods used to carry out text classification processes. The third reason we combined the two methods was because it was based on research that had been conducted [16] gives us a glimpse that the IndoBERT-LSTM method is a method capable of producing a good accuracy, precision, recall, and F1-score. So according to us, using the IndoBERT and LSTM combined method is not a bad idea and it can be assumed that using the same method, it is capable of producing value precision, recall, and F1-score the highest in detecting hoaxes.

In this study, we were interested and decided to use the combined IndoBERT-LSTM method in developing an Indonesian hoax detection model. A method that combines IndoBERT's pre-trained model with one of the RNN architectures, namely LSTM. Combining IndoBERT and LSTM can be an effective method for designing a hoax detection model because of its advantages in managing and understanding Indonesian language news texts.

Transformers (in this case, IndoBERT), is a type of neural network architecture, which has been widely used for NLP work such as language translation, text classification, and sentiment analysis.

Transformers have shown outstanding performance in this kind of work because of their ability to process and understand the relationships between words in a sentence [17]. When using IndoBERT as the feature extractor and LSTM as the classification layer, the model first sends the tokenized text via IndoBERT to extract the hidden features. These features are then fed into the LSTM, which has been trained to identify patterns and relationships in the data as a classification layer. LSTM then makes predictions about the credibility of the text based on the extracted features and their training.

The architecture of BERT-Base is seen in Figure 1. The layer employed in this study is a transformer encoder with 12 layers in the IndoBERT configuration utilizing the BERT-Base configuration. The processing that takes place in IndoBERT is the provision of sentence vectors that have been processed in IndoBERT's transformer encoder. Vectoring will be performed with an emphasis on extracting features from the sentence-level of the content.

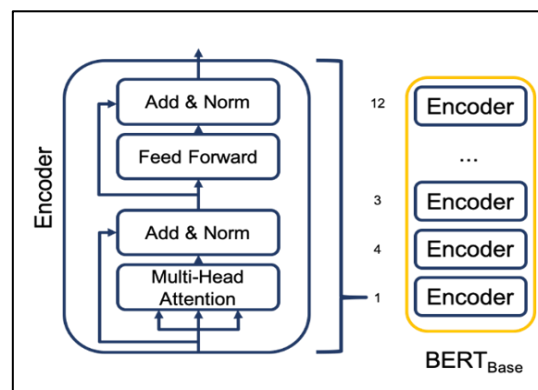


Figure 1. Bert architecture

Classification layers in natural language processing tasks are in charge of transferring learnt representations of input text to particular output classes. The LSTM layer is a prominent form of classification layer in NLP. LSTMs are a sort of recurrent neural network that is built to handle sequential input, making them ideal for natural language text processing. They can detect both short-term and long-term relationships in text, which is critical for many NLP tasks including sentiment analysis and named entity identification. In many NLP tasks, LSTMs have been found to outperform other types of classification layers, making them a popular option among researchers and practitioners [18], [19].

Figure 2 depicts on how LSTM architecture works. First, the long-term state  $C_{(t-1)}$  is sent via the forget-gate, which removes certain memories and changes them with those selected by the input gate. After that, the result  $C_{(t-1)} = C_{(t)}$ . Moreover,  $C_{(t-1)}$  is copied and sent through the  $\tanh$  function, which filters the output gate to determine the short-term state  $h_{(t)}$ , which is the cell's output at the  $t$  time step,  $y_t$  [20]. In simpler terms, the LSTM cell has three extra gate controller layers whose output varies from 0 to 1 due to the logistic activation function. The gate closes when the zeros are released; when the ones are released, the gate opens. In the end, the forgotten door determines what should be erased. The input-gate determines how the long-term status is added. The output-gate defines which portions of the long-run state should be read and output in the current time step [21].

Another benefit of employing LSTM as a classification layer in NLP is that it does not necessitate the use of a strong graphics processing unit (GPU). Because of its sequential structure and huge number of parameters, LSTMs may be computationally costly to train. Nevertheless, recent advances in hardware and software have made them more accessible. Smaller LSTM models, in fact, may be trained on a conventional central processing unit (CPU) or GPU, making them a viable option for many NLP applications. This indicates that even academics and practitioners who do not have access to sophisticated hardware can profit from the use of LSTMs in classification tasks [22].

In addition to the increased efficiency and accuracy provided by the combination of IndoBERT and LSTM, this approach also has the advantage of being interpretable. Models can be analyzed to understand how to make predictions, providing insight into the relationships and patterns the model has learned. This interpretability is useful for detecting bias in data or models, and for improving model accuracy over time. In order to be clearer and/or to clarify, the design flow process of the hoax detection model using the combined IndoBERT-LSTM method is described in the following scheme: shown in Figure 3.

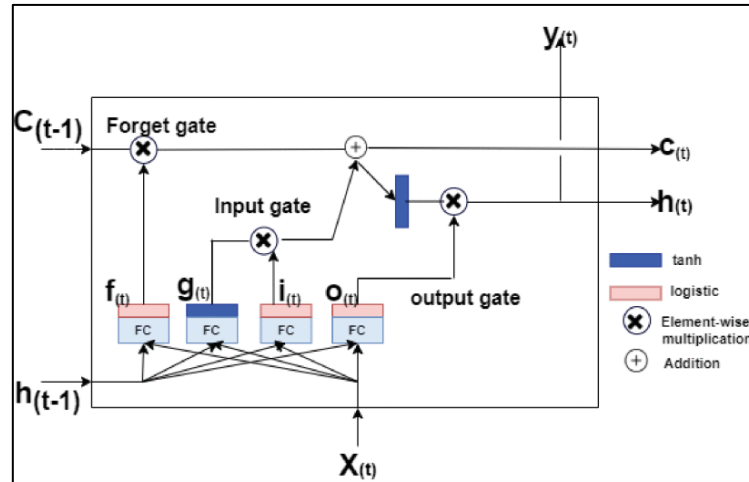


Figure 2. LSTM cell architecture [20]

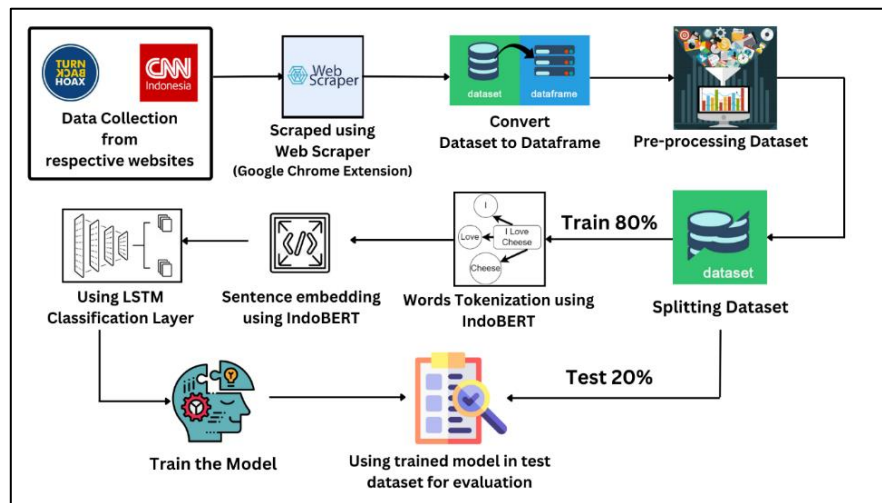


Figure 3. Schematic of design flow of the hoax detection mode

**2.2.1. Data scraping**

In this study, we used two dataset sources and we will discuss the two datasets we used for training and testing. The first dataset was obtained from the CNN Indonesia website, a well-known and well-known news outlet in Indonesia. This data set was obtained via process web scraping using the Google Chrome Web Scraper extension. The scraped dataset contains valuable information for training hoax detection models. CNN Indonesia's fact-checking and editorial processes ensure that the information in the dataset is accurate and up to date.

The second dataset was obtained from turnbackhoax.id, a website that contains a collection of Indonesian hoax news. This dataset was also obtained through web scraping using the Google Chrome Web Scraper extension. The dataset from turnbackhoax.id provides information about the characteristics and patterns of hoaxes, helping the hoax detection model to understand what to look for in determining whether a news item is a hoax or not. This website regularly updates existing news, ensuring that the information is always up to date.

In the end, these datasets provide important information for training the hoax detection model. The first dataset obtained from CNN Indonesia provides a reliable basis for determining factual news. The second dataset obtained from turnbackhoax.id provides complete information about hoaxes as it also provides a large collection of hoax examples as a complement to the training model.

### 2.2.2. Datasets to dataframes

The scraped datasets only provide low-level functionality for training data. Therefore, it is needed to convert dataset objects to Pandas DataFrames for read and manipulate the data. DataFrames also provides easy and flexible indexing, merging, and grouping operations, which makes it a great choice for making it easier to process, save time, and make code processing simpler.

### 2.2.3. Data pre-processing

After converting the dataset to DataFrames, the dataset (or DataFrames) will be pre-processed before being converted into a token. The pre-processing stage is the stage for cleaning the dataset. The cleaning action in question is cleaning the dataset from unimportant words from scrape results and removing news headlines. In this study, we used the natural language toolkit (NLTK) library to assist in the dataset cleaning process.

The first step in dataset pre-processing using the NLTK library is removing irrelevant information such as stop words, punctuation marks, and numbers. The NLTK library provides several functions for removing stop words, punctuation, and numbers from data, such as the `stopwords.words()` function, the `string.punctuation` constant, and the `isdigit()` method. Another important step in pre-processing a data set is to remove white spaces from data. The `strip()` method is used to remove white spaces from the beginning and end of text data. As well as removing irrelevant information, it's also important to lowercase text data. This is important because the model should not be able to distinguish between uppercase and lowercase characters, because they have the same meaning. The `lower()` method is used to convert text data to lowercase. The last step is to remove the URL contained in the text data because the URL (`http/`) has no relevance to the text data to be processed.

### 2.2.4. Splitting dataset

After cleaning the dataset, the dataset is divided into two, namely datasets for training and testing. Dividing the data set into two parts is a common and important step. This is because splitting the dataset into two parts makes it possible to evaluate the performance of the model on a given data.

The process of separating the datasets was carried out with 80% data used for training and the remainder 20% used for testing. We split the data into training and testing sets because first, it helps prevent overfitting and improves the generalization ability of the model. Second, it allows us to tune the hyperparameters of the model based on the performance on the testing dataset. Third, it helps us compare the performance of different models on the same dataset and finally, it provides an estimate of the expected performance of the model on new, unseen data. overall, by evaluating the performance on the testing dataset, we can estimate the generalization ability of the model and compare the performance of different models on the same dataset.

The training dataset is used to train the model by adapting the model to the data. Dataset testing, on the other hand, is used to evaluate model performance. This process involves using a trained model to make predictions on the test data and comparing these predictions with the actual target values in the test data. The comparison between the predicted and actual values is used to calculate evaluation metrics such as precision, recall, and F1-score.

### 2.2.5. Word's tokenization

Tokenization is an important step in the pre-processing of data in the form of text before it can be included in a machine learning model. The goal of tokenization is to break down text input into individual units, called tokens, which can then be processed by the model [17]. In this study, we used IndoBERT, where IndoBERT used the WordPiece tokenization method for its tokenization process. WordPiece tokenization is a type of Sub-Word Tokenization, which breaks the input text into word pieces. This method is used in IndoBERT to mark input text [17], because it helps in dealing with out-of-vocabulary (OOV) words, i.e., words that are not in the model vocabulary [23].

In traditional word tokenization, if a word is not in the vocabulary, it is treated as OOV and is usually discarded or replaced with a special token. This can cause loss of information and/or change of meaning [17]. On the other hand, sub-word tokenization methods such as WordPiece can handle OOV words by breaking them into chunks that are in a vocabulary set. With this in mind, WordPiece tokenization ensures that the model can still process and store information from OOV words [24].

The WordPiece tokenization method uses a vocabulary consisting of sub-words (Smaller unit words), which are the most frequently occurring word chunks in the training data. During the tokenization process, the input text is broken into sub-words and each sub-word is assigned a unique numerical representation, which is called embedding. This embedding is then entered into the model as input, not raw text [24].

### 2.2.6. Token embedding

After the tokenization process, the generated token is then converted into an embedding token. Token embedding is one of the important components in the modern NLP model as is the case with IndoBERT. The embedding token provides a representation of the input token and can receive and read the information contained in the original text [17].

Embedding tokens are created by mapping each token of input text to a fixed-length vector representation. These vector representations are learned through the process of training NLP models. During training, the model is exposed to a large corpus of text data, and the token embedding is fine-tuned to best capture the relationships and patterns present in the data [25], [26].

Once the embedding token is created, the embedding token can be used as input to the IndoBERT encoder block layer. The encoder block layer is responsible for processing the token embedding and generating the hidden state representation for each token. The hidden state representation captures the context present in the original text, enabling NLP models to make more informed decisions about the meaning and structure of the text [17].

The encoder block layer consists of several self-attention mechanisms and a feed-forward neural network. The self-attention mechanism allows the model to pay attention to different parts of the input text, giving it the ability to focus on the most relevant information [27]. This is very important because it allows the model to process input text in a more sophisticated way, taking into account local and global contexts.

Feed-forward neural networks are used to study complex interactions between token embeddings. Feed-forward neural networks take the output of the self-attention mechanism as input and apply a series of non-linear transformations to it [28]. This allows the model to capture more complex relationships between tokens, beyond what self-attention mechanisms alone can capture [25].

The hidden state is a mathematical representation of the token's meaning and context in the input text [17]. The hidden state generated by IndoBERT is the result of a series of calculations and transformations that take into account the relationship between each token in the input text. In short, embedding tokens is like a compact version of the input text, capturing the most important information about each token in a concise and efficient form. The hidden state generated by the IndoBERT encoder then builds on top of this embedding token, creating a more sophisticated representation of the input text, taking into account the local and global context. After hidden states have been created for all input tokens, these states can be used as input to other models [17], such as LSTM, for further processing. In the case of using IndoBERT as a feature extractor for the hoax detection model, the hidden state will be used to capture the meaning and context of the input text, which can then be entered into the classification layer to make predictions about whether the input news text is hoax news or not.

### 2.2.7. Classification layer

The LSTM component of the model plays a crucial role in handling sequential data, particularly significant in the context of hoax detection [29]. This is because the meaning of a sentence can vary depending on its context, highlighting the importance of comprehending the relationships between words in a sentence. LSTM excels at learning patterns and identifying relationships within the features extracted by IndoBERT. This capability empowers the models to make more precise and accurate classifications.

Using LSTM as a classification layer, the LSTM layer receives feature vector sequences extracted from the input text, as already obtained from IndoBERT. The LSTM layer then processes these feature vectors one at a time, maintaining an internal state that captures information about the previous time step. The internal state is updated at each timestep, allowing the LSTM to incorporate information from previous timesteps into its predictions for the current timestep.

After the LSTM processes the entire feature vector sequence, it issues a final hidden state which summarizes the information contained in a sequence. This final hidden state is then used as input to the fully connected layer which performs the final classification. The LSTM's ability to retain information in memory over a longer period of time allows it to capture the relationships between words in a sentence.

## 3. ANALYSIS PERFORMANCE

In this study, we used two data sources with a total of 5,876 data composed of a total of 1,998 factual news collected from the official CNN Indonesia website which were scraped using a web scraper and 3,878 hoax data. The hoax data is also collected through scraping techniques using a web scraper. The dataset is then divided by a ratio of 80% : 20% for training data and test data respectively. So that the total training data used is 3,102 and 741 for hoax test data, and 1,598 and 435 for factual news test data.

In this experiment, we created two other models which were only used as benchmarks and/or comparisons, which were compared with the main model, namely IndoBERT-LSTM. The two models consist



of IndoBERT-BiLSTM and sole LSTM model. Since all the model, both the main model and the experimental model producing only two outcomes (hoax or non-hoax) we can consider all the models as binary classifications model, and from which, the confusion matrix can be used to describe the discrimination assessment of the best (optimal) solution during classification training [30]. From the experimental results that have been carried out, the researcher gets the results of the confusion matrix as shown in Table 1.

Table 1. IndoBERT-LSTM confusion matrix

	Hoax	Factual
Hoax	707	34
Factual	45	390

In the initial phase of testing with the IndoBERT-LSTM model, a comprehensive analysis revealed a total of 707 hoax news articles successfully identified, along with 390 factual news articles out of a dataset totaling 435. Following this, the subsequent experiment utilizing the IndoBERT-BiLSTM model culminated in the compilation of a detailed confusion matrix, depicted in Table 2. This careful study effort provided insights into the models' individual capabilities and performance indicators, showing their effectiveness in differentiating hoax news from reliable sources. The results quantified the models' capabilities and gave a comprehensive knowledge of their usefulness in practice for the crucial task of distinguishing between accurate and false information.

Table 2. IndoBERT-BiLSTM confusion matrix

	Hoax	Factual
Hoax	707	34
Factual	46	389

In addition to the quantitative data presented in the table, it's worth noting the meticulous curation process undertaken. Out of a total of 435 real news articles, 389 were accurately identified, reflecting the precision of the selection procedure. Similarly, in terms of fake news, the model adeptly pinpointed 707 out of 741 articles, indicating a high degree of discernment. Our approach to detailing the outcomes of the final experiment, specifically centered on the exclusive use of LSTM, is exemplified in the comprehensive confusion matrix provided in Table 3. This visual representation not only encapsulates the model's performance but also offers a nuanced understanding of its effectiveness in the critical task of distinguishing between factual and hoax news.

Table 3. LSTM confusion matrix

	Hoax	Factual
Hoax	658	83
Factual	108	327

The data in the table shows that, in the first experiment, 327 true news pieces were correctly detected out of 435 factual news data, whereas 658 fake news articles were properly identified out of a total of 707 hoax news data. This initial result paves the way for a thorough assessment of the models' efficacy. A variety of assessment metrics may then be computed using information from three different confusion matrix tables that reflect different model trials. These measures, indicated as (1), (2), (3), and (4), provide a useful way to evaluate the efficiency and dependability of each model, allowing for a more detailed knowledge of their propensity to discern between fake and real news stories. Researchers may make educated judgments and develop improvements in the quest of more accurate news classification by analyzing these metrics in order to acquire deeper insights into the benefits and drawbacks of the models used in the trials.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

$$Recall = \frac{True\ Positive}{True\ Positive + True\ Negative} \quad (3)$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (4)$$

And the results that have been calculated are listed in the following Table 4. The IndoBERT-LSTM model emerges as the most effective hoax detection method, demonstrating superior performance in terms of evaluation metrics compared to the alternative models. Following closely is the IndoBERT-BiLSTM model, which exhibits commendable results, albeit with a marginal variance of approximately 0.01% to 0.03% in evaluation metric scores when compared to the IndoBERT-LSTM. Additionally, a visual representation of the model's loss can be observed in Figure 4, providing further insights into the model's training process and performance trends. This comprehensive analysis affirms the IndoBERT-LSTM model's prominence in hoax detection while acknowledging the competitive performance of the IndoBERT-BiLSTM model, underscoring the significance of these findings for future research and application in the realm of news classification.

Table 4. Overall comparison of all evaluation metric

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LSTM	83.7	79.8	75.2	77.4
IndoBERT-BiLSTM	93.1	92	89.4	90.7
IndoBERT-LSTM	93.2	92	89.7	90.8

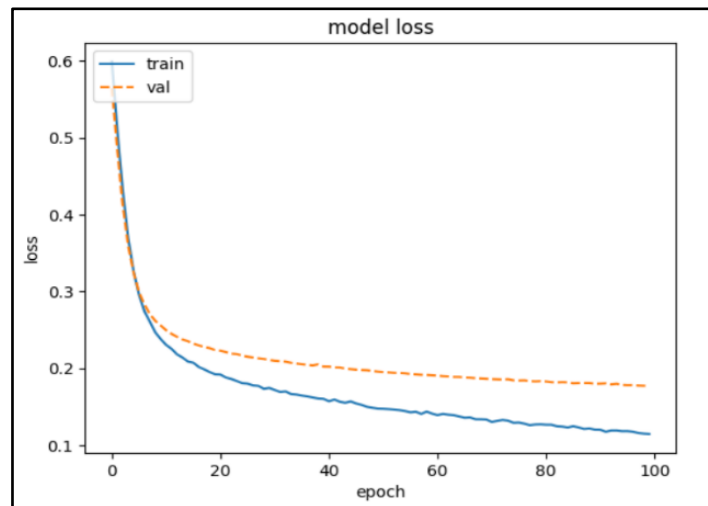


Figure 4. Training/validation loss

The loss graph exhibits a steady lower trend, demonstrating the model's improvement over time and learning from the supplied data. This accomplishment shows that the IndoBERT-LSTM model has achieved its main goal of becoming capable of making exact predictions for the training dataset. This achievement may be credited to the model's use of LSTM network and pre-trained IndoBERT embeddings, which let it understand the contextual nuances within the text data. Furthermore, this effective combination of pre-trained embeddings and sequential learning enables the model to capture intricate patterns and semantic relationships within the text, contributing to its overall success in hoax detection.

The findings of this investigation confirm the IndoBERT-LSTM model's effectiveness in hoax detection, which is a very positive outlook. This effective technology has the potential to considerably help businesses, media outlets, and people spot bogus information quickly and stop its spread. The model's strength rests in its ability to handle the nuances unique to text data, which is shown in the model's high accuracy and recall values. These scores highlight the model's ability to distinguish true positive and true negative situations with accuracy, highlighting its resilience in classification tasks.

The IndoBERT-LSTM model performs with balance, establishing a pleasing mix between precision and recall, as seen by the F1-score of 90.8%. Any classification model must have this equilibrium since it guarantees the fairness and accuracy of the model's predictions. Such a comprehensive performance reinforces the IndoBERT-LSTM model's position as a robust tool for hoax identification, with potential applications in other fields where separating fact from fiction is crucial. Moreover, this balanced performance signifies that the model is not only accurate in pinpointing hoaxes but also minimizes the chances of false

alarms, providing a reliable system for discerning between genuine and deceptive news. This further highlights the model's versatility and reliability in real-world applications beyond the realm of hoax detection.

#### 4. CONCLUSION

In today's digital age, the spread of fake news is a growing concern, especially in Indonesia where it can lead to social unrest and damage the credibility of media sources. To address this issue, we proposed the IndoBERT-LSTM model as a solution for detecting fake news. The dataset used in this study consisted of 3,878 fake news and 1,998 real news articles in the Indonesian language. The results of the experiment showed that the proposed model achieved an impressive accuracy rate of 93.2% in detecting fake news. This demonstrates its effectiveness in reducing the spread of fake news and promoting the credibility of media sources. Additionally, it's critical to recognize that given the possibility of significant changes in news trends and patterns, the model's ability to distinguish fake news from true news may become less effective after February 02, 2023. This timeframe highlights the dynamic nature of news reporting and the changing information landscape. Future research may focus on building a more adaptable model that has the capacity to automatically ingest and learn from current news data in order to address this problem.




#### REFERENCES

- [1] We Are Social, "Digital 2019 Indonesia," *We Are Social-Hootsuite*. p. 77, 2019. Accessed: Feb. 02, 2023. [Online]. Available: [https://es.slideshare.net/DataReportal/digital-2019-indonesia-january-2019-v01?from\\_action=save](https://es.slideshare.net/DataReportal/digital-2019-indonesia-january-2019-v01?from_action=save)
- [2] N. P. S. Meinarni and I. B. A. I. Iswara, "Hoax and its Mechanism in Indonesia," in *Proceedings of the International Conference of Communication Science Research (ICCSR 2018)*, 2018, pp. 183–186, doi: 10.2991/iccsr-18.2018.39.
- [3] C. Fleming and J. O'Carroll, "The Art of the Hoax," *Parallax*, vol. 16, no. 4, pp. 45–59, Nov. 2010, doi: 10.1080/13534645.2010.508648.
- [4] A. Boese, *The museum of hoaxes: a collection of pranks, stunts, deceptions, and other wonderful stories contrived for the public from the Middle Ages to the new millennium*. Dutton, 2002.
- [5] J. W. Miller and M. C. McKenna, *World Literacy*. USA: Routledge, 2016, doi: 10.4324/9781315693934.
- [6] Y. Hao, L. Dong, F. Wei, and K. Xu, "Visualizing and Understanding the Effectiveness of BERT," Aug. 2019, doi: <https://doi.org/10.48550/arXiv.1908.05620>.
- [7] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *COLING 2020-28th International Conference on Computational Linguistics, Proceedings of the Conference*, 2020, pp. 757–770, doi: 10.18653/v1/2020.coling-main.66.
- [8] S. M. Isa, G. Nico, and M. Permana, "Indobert for Indonesian Fake News Detection," *ICIC Express Letters*, vol. 16, no. 3, pp. 289–297, 2022, doi: 10.24507/icicel.16.03.289.
- [9] M. Awan, M. Shehzad, and M. Ashraf, "Fake News Classification Bimodal using Convolutional Neural Network and Long Short-Term Memory," *International Journal on Emerging Technologies*, vol. 11, no. 5, pp. 209–212, 2020.
- [10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [11] H. Huan, Z. Guo, T. Cai, and Z. He, "A text classification method based on a convolutional and bidirectional long short-term memory model," *Connection Science*, vol. 34, no. 1, pp. 2108–2124, Jul. 2022, doi: 10.1080/09540091.2022.2098926.
- [12] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," in *Procedia Computer Science*, 2021, vol. 179, pp. 704–712, doi: 10.1016/j.procs.2021.01.059.
- [13] J. Fawaid, A. Awalina, R. Y. Krisnabayu, and N. Yudistira, "Indonesia's Fake News Detection using Transformer Network," in *6th International Conference on Sustainable Information Engineering and Technology 2021*, Sep. 2021, pp. 247–251, doi: 10.1145/3479645.3479666.
- [14] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11765–11788, Jan. 2021, doi: 10.1007/s11042-020-10183-2.
- [15] H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. Rosal Ignatius Moses Setiadi, and R. S. Basuki, "Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 799–806, Apr. 2020, doi: 10.12928/telkomnika.v18i2.14744.
- [16] A. Chowanda and Y. Muliono, "Indonesian Sentiment Analysis Model from Social Media by Stacking BERT and BI-LSTM," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, Sep. 2022, pp. 278–282, doi: 10.1109/aidas56890.2022.9918717.
- [17] L. Tunstall, L. Von Werra, and T. Wolf, *Natural language processing with Transformers*. Canada: O'Reilly Media, Inc., 2022.
- [18] P. M. Lavanya and E. Sasikala, "Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network: A Comprehensive Survey," in *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, May 2021, pp. 603–609, doi: 10.1109/icpsc51351.2021.9451752.
- [19] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, May 2020, doi: 10.1007/s10462-020-09838-1.
- [20] T. Jiang, J. P. Li, A. U. Haq, A. Saboor, and A. Ali, "A Novel Stacking Approach for Accurate Detection of Fake News," *IEEE Access*, vol. 9, pp. 22626–22639, 2021, doi: 10.1109/access.2021.3056079.
- [21] S. Senhadji and R. A. S. Ahmed, "Fake news detection using naïve Bayes and long short term memory algorithms," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 2, pp. 746–752, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp746-752.
- [22] Z. Sun, L. Di, and H. Fang, "Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series," *International Journal of Remote Sensing*, vol. 40, no. 2, pp. 593–614, Oct. 2018, doi: 10.1080/01431161.2018.1516313.
- [23] A. Nayak, H. Timmapathini, K. Ponnalagu, and V. Gopalan Venkoparao, "Domain adaptation challenges of BERT in




- tokenization and sub-word representations of Out-of-Vocabulary words,” in *Proceedings of the First Workshop on Insights from Negative Results in NLP*, 2020, pp. 1–5, doi: 10.18653/v1/2020.insights-1.1.
- [24] M. Schuster and K. Nakajima, “Japanese and Korean voice search,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012, pp. 5149–5152, doi: 10.1109/icassp.2012.6289079.
- [25] K. Ue, A. Chatterjee, and Tushti, *Problems on Array: For Interviews and Competitive Programming*. Independently published, 2021. [Online]. Available: <https://amzn.to/3n4OHrJ>
- [26] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *NAACL HLT 2019-2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies-Proceedings of the Conference*, 2019, vol. 1, pp. 4171–4186.
- [27] Vasvani A. *et al.*, “Attention is all you need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [28] ChrisMcCormickAI, “BERT Research-Ep. 7-Inner Workings IV-FFN and Positional Encoding,” *Youtube*. 2018. Accessed: Feb. 05, 2023. [Online]. Available: <https://www.youtube.com/watch?v=YIEe7d7YqaU>
- [29] A. Graves, “Long Short-Term Memory,” in *Supervised Sequence Labelling with Recurrent Neural Networks*, Germany: Springer Berlin Heidelberg, 2012, pp. 37–45, doi: 10.1007/978-3-642-24797-2\_4.
- [30] H. M and S. M.N, “A Review on Evaluation Metrics for Data Classification Evaluations,” *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 1–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.

## BIOGRAPHIES OF AUTHORS






**Danny Yongky Yefferson**    is a dedicated college student in the Bina Nusantara University Graduate Program (BGP). Pursuing a degree in Computer Science, Danny has actively contributed to this paper through writing this paper. showcasing a strong foundation in natural language processing. With a passion for research and a well-rounded approach to education, Danny aspires to make significant future contributions to the field of computer science. He is currently in his 7<sup>th</sup> semester of his college year at Bina Nusantara University, Jakarta 11480, Indonesia. He can be contacted at his email: [danny.yefferson@binus.ac.id](mailto:danny.yefferson@binus.ac.id)



**Viriyaputra Lawjiaya**    is a dedicated college student in the Bina Nusantara University Graduate Program (BGP). Pursuing a degree in Computer Science, Viriyaputra has actively contributed to this paper through writing this paper. showcasing a strong foundation in natural language processing. With a passion for research and a well-rounded approach to education, Viriyaputra aspires to make significant future contributions to the field of computer science. He is currently in his 7<sup>th</sup> semester of his college year at Bina Nusantara University, Jakarta 11480, Indonesia. He can be contacted at his email: [viriyaputra@binus.ac.id](mailto:viriyaputra@binus.ac.id)



**Abba Suganda Girsang**    obtained Ph.D. degree in the Institute of Computer and Communication Engineering, Department of Electrical Engineering and National Cheng Kung University, Tainan, Taiwan, in 2014. He graduated bachelor from the Department of Electrical Engineering, GadjahMada University (UGM), Yogyakarta Indonesia, in 2000. He then continued his master’s degree in the Department of Computer Science in the same university in 2006–2008. He was a staff consultant programmer in Bethesda Hospital, Yogyakarta, in 2001 and worked as a web developer in 2002–2003. He then joined the faculty of Department of Informatics Engineering in Janabadra University as a lecturer in 2003-2015. He also taught some subjects at some universities in 2006–2008. His research interests include swarm intelligence, business intelligence, machine learning and media social text mining. He can be contacted at email: [agirsang@binus.edu](mailto:agirsang@binus.edu).