Recursive Journal of Informatics
Vol. 3, No. 2, September 2025
p-ISSN 2963-5551 https://journal.unnes.ac.id/journals/index.php/rji e-ISSN 2986-6588

Implementation of Bidirectional Long-Short Term Memory (Bi-LSTM) and Attention to Detect Political Fake News Using IndoBERT and GloVe Embedding

Adham Satria Firmansyah[1], Anggyi Trisnawan Putra[2]
[1, 2]Computer Science Department, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Semarang, Indonesia

Abstract. Indonesian political news is now increasingly spread through various media, especially social and online media. However, a lot of fake news are spread to bring down political opponents or attract public sympathy in order to find their own supporters. Of course, this news need to be watched out for and preventive measures must be taken so as not to cause misunderstanding in the wider community.
Purpose: This study was conducted to detect the political news whether it's classified as hoax or fact by its narration. Also, understanding how to build the news detector using corresponding architecture and word embeddings.
Methods/Study design/approach: The model architecture of Bi-LSTM and attention mechanism is used to reach the goals from this study's purposes. Many studies have been conducted to detect hoaxes but have not yet paid attention to the context of sentences and the contribution of words in a news text so that this architecture is made to overcome this problem. It uses IndoBERT to optimize the model for Indonesian language and also GloVe to obtain the word weights from pre-trained embedding. Then, the tokenization process is performed with IndoBERT and keras to generate token id and attention mask. After receiving the token id and attention mask as input, the data training process is performed for three architectural scenarios with each configuration of 20 epochs, batch size of 32, and the learning rate is 0.00001.
Result/Findings: The results of this study are defined by a confusion matrix which contains accuracy, recall, precision, and F1-score as the evaluation. The combination of Bi-LSTM + Attention + IndoBERT + GloVe obtains the best result of 97,71% of accuracy, 96,33% of precision, 97,93% of recall, and 97,72% of F1-score.
Novelty/Originality/Value: This study combines two words embeddings in order to make sure the weight of words is completely defined and optimized into the Bi-LSTM and attention mechanism architecture.

Keywords: Natural Language Processing, NLP, political hoax news, IndoBERT, Bi-LSTM, prediction, GloVe embedding, attention mechanism

Received January 2024 / **Revised** September 2025 / **Accepted** September 2025

INTRODUCTION

It is feared that the existence of hoax news will adversely affect the community if they are not careful in sorting and selecting it. The purpose of spreading hoax news is, among others, to shape public perception, lead opinions, and be used for certain interests. In fact, hoax news can cause misinformation and also cause harm to victims such as worry, anxiety, confusion, and various other negative impacts [1]. Negative impacts also come in the political field considering Indonesia as a democratic country. This has also encouraged the media field to make news about politics in Indonesia. Various ways can be done to achieve this goal, one of which is by delivering hoax news to online media. This is certainly detrimental to one party and can certainly create chaos and confusion in the community [2].

To avoid the problems that arise, something is needed that is expected to detect hoax news so that unwanted things do not happen. The utilization of information technology can be used to detect hoax news by implementing Natural Language Processing (NLP). NLP can be used to make a computer know and understand the meaning of natural language/human language [3].

[1]*Corresponding author.
Email addresses: adhsen.firman@students.unnes.ac.id (Firmansyah)
DOI: 10.15294/rji.v3i2.159

There have been many studies conducted to detect hoax news or generally called text classification with NLP [4]. used traditional Machine Learning such as Naïve Bayes, Support Vector Machine, Decision Tree, K-Nearest Neighbor, Logistic Regression, and Random Forest and obtained their highest accuracy of 85%. Also, there is a Deep Learning (DL) method which is one of the machine learning methods in the form of a combination of neural networks in the layer as [5] did a sentiment analysis that compared several DL architectures used for sentiment analysis classification including Neural Network (NN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM). After training, the highest accuracy result is obtained from the Bi-LSTM algorithm which is 90.04%. Bi-LSTM produces better accuracy than RNN and LSTM because Bi-LSTM can overcome exploding and vanishing gradients commonly found in RNN and Bi-LSTM combines two hidden layers, namely forward and backward so that it is better than LSTM which only has 1 hidden layer [6].

The results of traditional ML and DL discussed above show that the DL method certainly brings more optimal results compared to conventional methods. The results of DL are better because of the advantages of DL which is able to learn data through its own computation like the human brain in making decisions so that it can be used to analyze data dynamically [7]. As of [6], DL can also generate context information from a learning sequence. However, the drawback of DL is that it still cannot focus on the most contributing words. To overcome that issue, [6] combined DL with a layer called attention layer and word embedding too.

The advantage of attention layer is that it can focus on the most contributing words by weighting a word contained in a sequence that has been represented by DL from word embedding. In practice, the addition of this attention layer succeeded in surpassing the accuracy of [5]'s research.

The [6]'s research also shows that word embedding is an important element in text classification. The embedding is in the form of a vector containing real numbers which represent a text that represents the context in which it appears. The word is converted into a number or numeric form that is inserted and mapped each word into the vector. There are many types of word embedding in circulation, including Word2Vec, GloVe, FastText, BERT, ELMo, and many more. Each word embedding has its own advantages and disadvantages [8].

Pre-trained word embedding is more commonly used in text classification, one of which is Global Vector (GloVe). The GloVe representation as a word embedding is obtained and trained from matrix factorization on large word collections such as Wikipedia. Due to its high quality as a word embedding, GloVe is widely used in text mining and NLP research with good performance [9]. GloVe is trained with a collection of words from Wikipedia & Gigaword with 6 billion tokens, 400 thousand words, with dimensions varying from 50 to 300 so that it will better understand words [10].

There is another pre-trained word embedding named Bidirectional Encoder Representation from Transformers (BERT). BERT is a method that belongs to transformers introduced in 2018 by [11]. Transformers use an encoder-decoder structure but do not rely on Recurrent and Convolutional methods to generate output. BERT uses transformers to learn contextual relationships between words in text with a self-attention mechanism [12]. This method also utilizes an encoder which means BERT reads the text as input, rather than as a sequential sequence so that it can properly understand the relationship and context of each token [13]. There are several examples of BERT such as DistilBERT, RoBERTa, and so on. BERT is commonly used for English, but there is also IndoBERT for Indonesian.

[14] stated that combining BERT and attention mechanism can improve the performance of the model in understanding the context between words and sentences. BERT plays a role in training text in several contexts and forming embedding that contains context information in it. Then, the attention mechanism plays a role in deepening the understanding of the context of each word and giving weight to the words.

METHODS

Research approach and design are needed so that the implementation of research is more focused and well organized. In this research, Bi-LSTM with an additional attention layer and IndoBERT and GloVe embedding are used to detect hoax news in Indonesian. In the initial stage, the dataset for research was

prepared in Microsoft Excel Spreadsheet (XLSX) format. Then, the dataset goes through a preprocessing stage so that it can improve the performance of the model. Furthermore, the data is separated into 3 parts, namely training as much as 70%, validation and testing as much as 15% each. The next process is to train and evaluate the model with confusion matrix to get the results of accuracy, precision, recall, and F1-score. The research flow is shown in Figure 1.

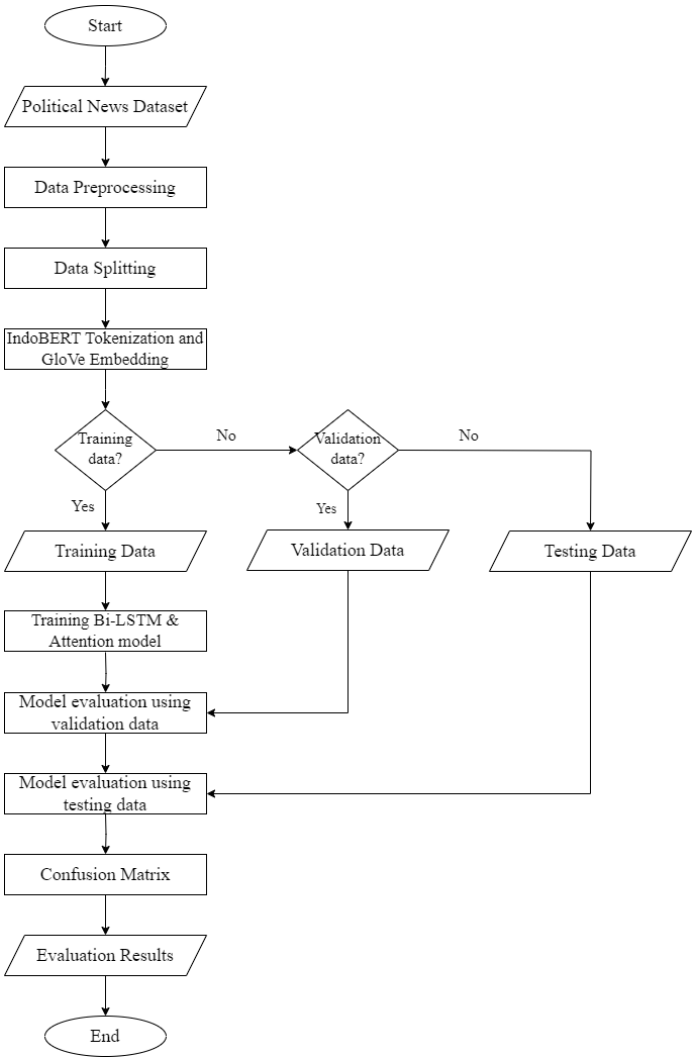

Figure 1. Flowchart of research

Dataset

The data used in this study are political hoax news texts in Indonesian. The dataset is titled 'Indonesian Fact and Hoax Political News' by [15]. The dataset contains news from national credible media sources called Tempo (years 2022-2023) and TurnBackHoax (years 2015-2023). The dataset is supervised, where the data training uses labels. The dataset detail can be seen in Table 1 and Table 2.

Table 1. Column on Tempo dataset

Column	Description
Title	News title
Timestamp	Time when the news was created
FullText	Original content of the news
Tags	Keywords of the news
Author	News author
Url	News link
text_new	Cleaned narration text
hoax	Type of the news (hoax or non-hoax)

Table 2. Column on TurnBackHoax dataset

Column	Description
Title	News title
Timestamp	Time when the news was created
FullText	Original content of the news
Tags	Keywords of the news
Author	News author
Url	News link
Narasi	Quoted narration from 'FullText' column
Clean Narasi	Cleaned narration text
hoax	Type of the news (hoax or non-hoax)

Data Preprocessing

Data preprocessing is done to adjust the raw data to what is needed by the model so that the training process will run more focused and efficient. In this research, data preprocessing included data cleaning, data filtering, and text preprocessing. Data cleaning is a process to clean the dataset to avoid errors, anomalies, and redundancies. In this phase, removing missing values or empty data in a dataset. Secondly, data filtering will be applied to filter the data by providing a threshold on the number of tokens in a news narrative [16]. And then, text preprocessing on this research uses case folding which converts text into lowercase, removing emoji, removing URLs, removing punctuation marks, removing numbers, converting text to lowercase, removing stop words, and removing white space [17].

Data Splitting

This data splitting is done randomly without recognizing hoax or non-hoax labels. Training, validation, and testing data will be divided by 70%, 15%, and 15% respectively refers to [18].

Tokenization and Embedding Matrix Assembling

Tokenization of text data is needed for the embedding matrix creation process in GloVe. The embedding matrix is obtained from the number of vocabularies in the entire text data [6]. The tokenization process begins by retrieving data that has been preprocessed and split. Then, the keras library will perform tokenization for each news narrative data and then do padding as long as the threshold is. In the other hand, the other tokenization comes from IndoBERT using WordPiece algorithm. The next step, this will also do padding and then make an attention mask to each data. So, the IndoBERT tokenization create token id and attention mask as input to the model.

Prediction Modeling Algorithm

The modeling framework to be used is based on the RNN architecture and also the application of Bi-LSTM and attention layer in it. The first stage is call the embedding as the input. Then, Bi-LSTM layer processes the output of each IndoBERT and/or GloVe embedding result. The parameters used are units of 128 and return_sequences is true. The next step, attention layer will simplify the output of the Bi-LSTM layer with the weighting in it. Layer Normalization is also used to calculate the average and standard deviation at each point for normalization. It will be concatenated if two embeddings are used, otherwise, the input will directly through several dense layers with batch normalization are applied for each dense layer. Lastly, the output comes from last dense layer with sigmoid activation function in it.

Model Evaluation

In this stage, the results of the training model will be calculated which represents whether the model is good or not. In general, text classification uses confusion matrix evaluation to calculate accuracy, precision, recall, and F1-score. Table 3 represents the confusion matrix.

Table 3. Confusion Matrix

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

Accuracy is a number that represents the amount of data that is correctly predicted by the system against the entire data. Precision is the ratio of correctly predicted positive data results to all positive data. Next, a recall calculation is performed which represents the ratio of positive and correct results to the total of all TPs and FNs. Then, the F1-score is calculated as the average of the precision and recall values. The formula for each evaluation metrics will be shown on Equation 1, 2, 3, and 4.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1} - score = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (4)$$

RESULT AND DISCUSSION

This research aims to find out how the results of the implementation of Bi-LSTM and attention mechanism in detecting hoax news with IndoBERT and GloVe embedding. The detection process with the above algorithms will go through various processes as follows.

The dataset used is "Indonesian Hoax and Fact Political News" which contains political news narratives labeled as facts and hoaxes in Indonesian. The selected dataset is a news narrative in the cleaned folder from Tempo and Turnbackhoax. News from Tempo has a fact label, while Turnbackhoax has a hoax label. The amount of data to be used is shown in Table 4.

Table 4. Data amount

Label	Non-hoax	Hoax
Summary	6.592	10.381
Total	16.973	

After collecting the data, the data is processed to make it easier to train the model. At this stage, 3 preprocessing steps are carried out, namely data cleaning, data filtering, and text preprocessing. The results of each stage is as follows. The dataset has the possibility of having missing values which cause the data to be a lot but actually empty so it must be checked. The isna() and sum() functions will be used to check for missing values so that the columns with missing values and the number of missing values will appear. The result of missing data cleaning is shown in Table 5.

Table 5. Data amount after deleting missing value

Dataset	Initial data amount	Missing values	Data amount
Tempo	6.592	0	6.592
Turnbackhoax	10.381	3.879	6.502

After data cleaning, the next step is data filtering. In this research, data filtering will be carried out by providing a threshold based on the token_length. The limit set is only between 20 and 200 tokens. The results of the filtering data that has been filtered is shown in Table 6.

Table 6. Data amount after filtering

Dataset	Initial data amount	Wasted data	Data amount
Tempo	6.592	1.570	5.022
Turnbackhoax	6.502	3.370	3.132
TOTAL	13.094	4.940	8.154

The next step is text preprocessing to make the text easier to process by the model. This process is shown in Table 7 and continued to Table 8.

Table 7. Text preprocessing steps

Original text	URL removal	Punctuation removal
Sementara itu, di Arab Saudi Salju turun untuk pertama kali dalam 100 tahun . Penjelasan:Akun Twitter @Strawberygirli mengunggah video yang memperlihatkan dua ekor unta di gurun yang tengah ditutupi salju. https://www.youtube.com/ watch?v de5jM9-7raM	*Sementara itu, di Arab Saudi Salju turun untuk pertama kali dalam 100 tahun . Penjelasan:Akun Twitter @Strawberygirli mengunggah video yang memperlihatkan dua ekor unta di gurun yang tengah ditutupi salju.*	*Sementara itu di Arab Saudi Salju turun untuk pertama kali dalam 100 tahun Penjelasan Akun Twitter Strawberygirli mengunggah video yang memperlihatkan dua ekor unta di gurun yang tengah ditutupi salju*

Table 8. Continued text preprocessing steps

Digit removal	Case folding	Stop words removal
Sementara itu di Arab Saudi Salju turun untuk pertama kali dalam tahun Penjelasan Akun Twitter Strawberygirli mengunggah video yang memperlihatkan dua ekor unta di gurun yang tengah ditutupi salju	*sementara itu di arab saudi salju turun untuk pertama kali dalam tahun penjelasan akun twitter strawberygirli mengunggah video yang memperlihatkan dua ekor unta di gurun*	*arab saudi salju turun pertama tahun penjelasan akun twitter strawberygirli mengunggah video memperlihatkan dua*

After all data that has been processed, they will be split into three parts namely train, test, and validation with a ratio of 70:15:15. A total of 8,154 data is taken randomly so as to fulfill the data division ratio.. Table 9 presents the details of the data splitting.

Table 9. Data splitting detail

Data type	Data amount
Training (70%)	5.707
Validation (15%)	1.223
Testing (15%)	1.224
TOTAL	8.154

After data division, the next step is to perform tokenization. In this process, tokenization is performed using the library from Keras. Tokenization will break down a sentence into word fragments. The vocabulary is updated to match the vocabulary in the entire narrative using the fit_on_texts(X) function of the k_tokenizer with the variable X representing the entire narrative. After the vocabulary has been updated, tokenization can be done on the narrative. The results of the tokenization process with the Keras library are presented in Table 10.

Table 10. Keras tokenization process

Original text	Tokenized	Convert to token id	Padding
forum ijtima ulama sumut dukung sandiaga uno capres sah hak said perwakilan pemuda islam indonesia sumatera utara wizdan fauran lubis menganggap sandiaga tokoh menyelesaikan permasalahan ekonomi	[*'forum', 'ijtima', 'ulama', 'sumut', 'dukung', 'sandiaga', 'uno', 'capres', 'sah', 'hak', 'said', 'perwakilan', 'pemuda', 'islam', 'indonesia', 'sumatera', 'utara', 'wizdan', 'fauran', 'lubis', 'menganggap', 'sandiaga', 'tokoh', 'me nyelesaikan', 'permasalahan', 'ekonomi'*]	[613, 1857, 441, 2159, 145, 205, 398, 17, 702, 270, 882, 683, 766, 155, 5, 1074, 175, 14305, 20423, 9445, 1619, 205, 212, 1058, 1287, 167, 945]	[613, 1857, 441, 2159, 145, 205, 398, 17, 702, 270, 882, 683, 766, 155, 5, 1074, 175, 14305, 20423, 9445, 1619, 205, 212, 1058, 1287, 167, 945, 0, 0, 0, 0, 0, 0, ..., 0, 0, 0]

The next tokenization is tokenization for IndoBERT. In this process, tokenization is carried out using the library from transformers. Tokenization with transformers will use WordPiece tokenization. After the tokenizer variable is initiated, tokenization can be performed on the narrative. The results of the WordPiece tokenization process are presented in Table 11.

Table 11. WordPiece tokenization process

Original text	Tokenized	Convert to token id	Padding	Attention mask
forum ijtima ulama sumut dukung sandiaga uno capres sah hak said perwakilan pemuda islam indonesia sumatera utara wizdan fauran lubis menganggap sandiaga tokoh menyelesaikan permasalahan ekonomi	[‘CLS’, *'forum', 'ij', '##tim', '##a', 'ulama', 'sumut', 'dukung', 'sandiaga', 'uno', 'capres', 'sah', 'hak', 'said', 'perwakilan', 'pemuda', 'islam', 'indonesia', 'sumatera', 'utara', 'wiz', '##dan', 'fau', '##ran', 'lubis', 'menganggap', 'sandiaga', 'tokoh', 'menyelesaikan', 'permasalahan', 'ekonomi'*, ‘SEP’]	[2, 4226, 4475, 858, 30354, 3716, 11109, 9162, 27216, 24056, 15994, 1456, 1319, 9219, 5130, 3624, 868, 300, 4049, 2024, 26840, 3703, 12680, 3823, 20616, 4480, 27216, 2546, 3426, 3472, 1447, 3]	[2, 4226, 4475, 858, 30354, 3716, 11109, 9162, 27216, 24056, 15994, 1456, 1319, 9219, 5130, 3624, 868, 300, 4049, 2024, 26840, 3703, 12680, 3823, 20616, 4480, 27216, 2546, 3426, 3472, 1447, 3, 0, 0, 0, 0, ... ,0, 0]	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, ... ,0, 0, 0, 0, 0, 0]

After all those processes are done, the next step is define the architecture scenarios for training the model. The model architecture formed is a combination of Bi-LSTM and attention layer with GELU activation function as the main layer and IndoBERT and GloVe embedding as word embedding. Layer normalization,

batch normalization, and dense layer as layers for normalization and output. In this study, three scenarios of model architecture formation were carried out, which are presented in Table 12.

Table 12. Model architecture scenario detail

Scenario	Bi-LSTM	*Attention*	GloVe	IndoBERT
1			√	-
2	√	√	-	√
3			√	√

All of the scenarios using the same hyperparameter such as 20 epochs, 32 batch size, 0,00001 learning rate, and AdamW optimizer.

Model training is done by creating three architecture scenarios, namely Bi-LSTM + attention + GloVe architecture; Bi-LSTM + attention + IndoBERT; and Bi-LSTM + attention + GloVe + IndoBERT. The results of the model evaluation comparison are shown in Table 13.

Table 13. Model result comparison

Scenario	Accuracy	Precision	Recall	F1-score
(Bi-LSTM + *Attention* + GloVe)	90,28%	88,05%	86,07%	87,05%
(Bi-LSTM + *Attention* + IndoBERT)	96,90%	98,26%	93,77%	95,96%
(Bi-LSTM + *Attention* + GloVe + IndoBERT)	97,71%	96,33%	97,93%	97,12%

Referring to Table 13, it is obtained that high evaluation values are obtained in scenario 2 and scenario 3 where both scenarios use IndoBERT as word embedding so that IndoBERT is clearly better when compared to GloVe. However, when comparing the results in scenarios 2 and 3, it is found that scenario 3 has better results due to the combination of IndoBERT and GloVe which strengthens the weight on the words in the news narrative. This proves that the model evaluation results will be better if IndoBERT and GloVe are combined.

The highest accuracy results in this research were obtained in scenario 3 with 97.71% and 96.33% precision, 97.93% recall, and 97.12% F1-score. These results will be compared with the results of previous studies presented in Table 14.

Table 14. Comparison of accuracy from previous researches

Reference	Dataset	Method	Best result
[19]	English (3843 data) and Indian (4594 data) data	*Attention layer* embedded in: · BERT + GRU · BERT + LSTM · FastText + GRU · FastText + LSTM · (BERT+FastText)+LSTM · (BERT+FastText)+GRU	(BERT + FastText) +GRU with 76,32% of accuracy
[20]	3000 English-Indonesian translation data from Github	DL architecture (LSTM, Bi-LSTM, GRU, Bi-GRU) with GloVe and FastText embedding respectively	87,825% from Bi-LSTM + *FastText*
[21]	6930 data from turnbackhoax.id	· SVM and NB with TF-IDF · fine-tuned IndoBERT	Fine-tuned IndoBERT with accuracy of 94,66%
[22]	Scrapping data from Twitter as much as 5163 data	Traditional ML and Deep Learning each with: - Single classifier · - Two-stage classifier	Single classifier with Bi-LSTM + IndoBERT 91.15% accuracy
[23]	Indonesian Fact and Hoax Political News	Using LSTM architecture to compare Word2Vec and GloVe	LSTM + *Word2Vec* 95% accuracy
Proposed Method	Indonesian Fact and Hoax Political News	Bi-LSTM, Attention mechanism, IndoBERT, dan GloVe	Accuracy 97,71%; Precision 96,33%; Recall 97,93%; F1-score 97,12%

These results are higher than the results of the studies conducted in Table 14. This is because the method used in this study is newer and better at predicting hoax news, especially in the research of [23] which has

the same dataset as this study. The prominent difference in the method is the use of Bi-LSTM which is better than the LSTM used by [23] because Bi-LSTM has the ability to read input from two directions so that connections between words can be better understood when using Bi-LSTM. Then the use of IndoBERT also affects the accuracy results because IndoBERT is specialized for Indonesian text so that it is in line with the language used in news data and the use of GloVe to give weight to each word.

CONCLUSION

After the stages of research on the implementation of Bi-LSTM and attention mechanism with IndoBERT and GloVe embedding to detect hoax news and the results of the research that has been carried out, implementation of Bi-LSTM architecture and attention mechanism through stages including data retrieval, data processing, model training stage, and system implementation. Then tokenization is carried out with the results in the form of token_id and attention mask which will be used as model input. Furthermore, model training is carried out to obtain hoax news prediction results with three model architecture scenarios and model evaluation is carried out using confusion matrix. The model training process is carried out with 20 epochs and 32 batch sizes for each scenario. The best model training evaluation results are obtained in the third scenario using the Bi-LSTM attention GloVe IndoBERT architecture. The evaluation results of this study increased by 2.71% from previous research with the same dataset [23].

REFERENCES

[1] G. Brotosaputro, W. Windihastuty, and R. A. Mutiarawan, "Penentuan hoax pada artikel politik berbahasa indonesia di sosial media dengan similarity jaccard dan algoritma stemming," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 1, pp. 79–86, Apr. 2022, doi: https://doi.org/10.32736/sisfokom.v11i1.1358.

[2] I. M. Kartika and I. Mustika, "Peran generasi muda dalam menangkal hoax di media sosial untuk membangun budaya demokrasi indonesia," *JOCER Journal of Civic Education Research*, vol. 1, no. 2, pp. 29–40, Dec. 2023, doi: https://doi.org/10.60153/jocer.v1i2.26.

[3] D. Marta, G. L. Ginting, and A. H. Sihite, "Deteksi berita palsu tentang vaksinasi covid-19 dengan menggunakan text mining dan algoritma cosine similarity," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 6, no. 1, pp. 129–139, 2022, doi: https://doi.org/10.30865/komik.v6i1.5738.

[4] L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, "Komparasi ekstraksi fitur dalam klasifikasi teks multilabel menggunakan algoritma machine learning," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 653–666, Jul. 2022, doi: https://doi.org/10.30812/matrik.v21i3.1851.

[5] F. P. Rachman, "Perbandingan model deep learning untuk klasifikasi sentiment analysis dengan teknik natural languange processing," *Jurnal Teknologi dan Manajemen Informatika*, vol. 7, no. 2, pp. 113–121, Dec. 2021, doi: https://doi.org/10.26905/jtmi.v7i2.6506.

[6] Moh. A. Rohman, S. Suhartono, and T. Chamidy, "Bidirectional gru dengan attention mechanism pada analisis sentimen pln mobile," *Techno.COM Jurnal*, vol. 22, no. 2, pp. 358–372, May 2023, doi: https://doi.org/10.33633/tc.v22i2.7876.

[7] A. Asrafil, M. R. D. Septian, M. Cahyanti, and E. R. Swedia, "Klasifikasi penyakit tanaman apel dari citra daun dengan convolutional neural network," *Sebatik*, vol. 24, no. 2, pp. 207–212, Dec. 2020.

[8] F. K. Khattak, S. Jeblee, C. Pou-Prom, M. Abdalla, C. Meaney, and F. Rudzicz, "A survey of word embeddings for clinical text," *Journal of Biomedical Informatics: X*, vol. 4, p. 100057, Dec. 2019, doi: https://doi.org/10.1016/j.yjbinx.2019.100057.

[9] F. Sakketou and N. Ampazis, "A constrained optimization algorithm for learning glove embeddings with semantic lexicons," *Knowledge-Based Systems*, vol. 195, p. 105628, May 2020, doi: https://doi.org/10.1016/j.knosys.2020.105628.

[10] J. Pennington, R. Socher, and C. D. Manning, "Glove: global vectors for word representation," in *2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[11] J. Devlin, M.-W. . Chang, K. Lee, and K. Toutanova, "Bert: pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018, doi: https://doi.org/10.48550/arXiv.1810.04805.

[12] A. Vaswani et al., "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[13] F. A. Pratama and A. Romadhony, "Identifikasi komentar toksik dengan bert," *eProceedings of Engineering*, vol. 7, no. 2, Aug. 2020.

[14] J. Meng, Y. Zhu, S. Sun, and D. Zhao, "Sarcasm detection based on bert and attention mechanism," *Multimedia Tools and Applications*, Sep. 2023, doi: https://doi.org/10.1007/s11042-023-16797-6.
[15] M. R. Rizqullah and R. H. Ghaida, "Indonesian fact and hoax political news [Dataset] ," 2023. https://www.kaggle.com/datasets/linkgish/indonesian-fact-and-hoax-political-news (accessed April 23rd, 2023)
[16] M. Ostendorff, P. Bourgonje, M. Berger, J. M. Schneider, G. Rehm, and B. Gipp, "Enriching bert with knowledge graph embeddings for document classification.," *KONVENS*, Jan. 2019.
[17] S. R. Aisy and B. Prasetiyo, "Sentiment analysist of the tpks law on twitter using inset lexicon with multinomial naïve bayes and support vector machine based on soft voting," *Recursive Journal of Informatics*, vol. 1, no. 2, pp. 93–101, Sep. 2023, doi: https://doi.org/10.15294/rji.v1i2.68324.
[18] N. Toiganbayeva et al., "Kohtd: kazakh offline handwritten text dataset," *Signal Processing: Image Communication*, vol. 108, pp. 116827–116827, Oct. 2022, doi: https://doi.org/10.1016/j.image.2022.116827.
[19] K. Maity, A. Kumar, and S. Salia, "Attention based bert-fasttext model for hate speech and offensive content identification in english and hindi languages," *Forum for Information Retrieval Evaluation (Working Notes)(FIRE)*, 2021.
[20] 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: https://doi.org/10.11591/eei.v10i4.2956.
[21] S. M. Isa, E. I. Setiawan, and J. Santoso, "Indobert for indonesian fake news detection," *ICIC Express Letters*, vol. 16, no. 3, pp. 289–297, 2022, doi: https://doi.org/10.24507/icicel.16.03.289.
[22] D. R. Faisal and R. Mahendra, "Two-stage classifier for covid-19 misinformation detection using bert: a study on indonesian tweets," *arXiv (Cornell University)*, Jun. 2022, doi: https://doi.org/10.48550/arxiv.2206.15359.
[23] M. G. Adrian, S. S. Prasetyowati, and Y. Sibaroni, "Effectiveness of word embedding glove and word2vec within news detection of indonesian using lstm," *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1180–1188, 2023, doi: https://doi.org/10.30865/mib.v7i3.6411.