KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA

VOL. 21 (2) (2024), 22-30, p-ISSN: 693-7554, e-ISSN:2654-3990

Fake News Detection in the 2024 Indonesian General Election Using Bidirectional Long Short-Term Memory (BI-LSTM) Algorithm

Shabiq Ghazi Arkaan[1*], **Aldy Rialdy Atmadja**[2], **Muhammad Deden Firdaus**[3]

[1*,2,3] Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University of Sunan Gunung Djati Bandung, Bandung, West Java, 40614, Indonesia

Abstract

The advancement of information technology provides convenience, but it also brings about problems. One area affected by this is the election process in Indonesia, which has seen a rise in fake news often used to discredit political opponents. Fake news misleads the public into believing incorrect information related to the election. To address this issue, a system is needed to detect fake news in the 2024 election to help the public differentiate between true and false information. This system is developed using an artificial intelligence and deep learning approach trained to do text classification on fake news detection. The training data consists of 1999 entries obtained from the Global Fact-Check Database from turnbackhoax.id, detik.com, and cnnindonesia.com. The machine learning model is built using the Bidirectional Long Short-Term Memory (BI-LSTM) algorithm, which is suitable for processing text data. This study compares two types of feature representations: TF-IDF and contextual embeddings with the IndoBERT model. The study results in the best model for text classification with an accuracy of 92% and a loss of 42.92%, achieved by the model using TF-IDF feature rep- resentation. The implementation of this system aims to enhance the integrity of the election process by minimizing the spread of misinformation. Future work will focus on refining the model and expanding the dataset to include more diverse sources for improved accuracy and robustness.

Keywords: *BI-LSTM; Deep Learning; Fake News Detection; Artificial Intelligence; Text Classification*

1. Introduction

The advancement of information technology has significantly transformed human lifestyles [1]. Information technology offers various conveniences for exchanging information both one-way and two-way without the need to be in the same place. However, these offered conveniences have led to various problems [2]. The increasing number of people with access to create, disseminate, and obtain information increases the risk of spreading unverified i

nformation, affecting the credibility of the information received by the public.

The issue of information authenticity is further exacerbated by the growth of the internet, especially in the context of elections in Indonesia [3]. Surveys indicate that the majority of the population has been exposed to fake news ahead of the elections [4]. Ethnographic studies also show that information whose truthfulness is unclear often receives negative responses from social media users, while positive information tends to receive less attention [5]. The circulation of fake news often emerges as an effective way to achieve a political goal, ultimately heating up the political situation and creating unhealthy elections [6].

One of the efforts that can be made to address the authenticity of news issues is by using the Natural Language Processing (NLP) concept. NLP is a part of artificial intelligence that focuses on understanding and processing human language to handle problems with computer assistance [8], [7]. The concept also utilizes deep learning, which is artificial intelligence that has a highly complex computational system based on artificial neural networks capable of understanding and processing human language like humans [9], [10]. By leveraging NLP and deep learning, automated systems can be developed to detect and mitigate the spread of fake news more effectively. This proactive approach not only helps maintain the integrity of information but also fosters a healthier and more transparent electoral process.

Research conducted by Xishuang Dong and Lijun Quan has been undertaken to address similar issues, focusing on the application of Bidirectional Recurrent Neural Network algorithms to detect misinformation,

*Corresponding author. *E-mail address: 1207050118@uinsgd.ac.id*

Received: 04 June 2024, Accepted: 26 July 2024 and available online 31 July 2024

DOI: https://doi.org/10.33751/komputasi.v21i2.5260

achieving an accuracy of 84% [11]. Another study by Bagas Pradipabista Nayoga et al. discusses the analysis of Indonesian-language hoaxes by comparing various deep learning algorithms. This research shows that the highest accuracy comes from the 1D Convolutional Neural Network algorithm, reaching 97%, slightly higher than BI-LSTM, which achieves an accuracy of 96.2% [12]. Furthermore, Jawaher Alghamdi, Yuqing Lin, and Suhuai Luo conducted a study comparing classical machine learning algorithms and deep learning algorithms tested on various datasets. The results indicate that there is no specific algorithm superior for every dataset used [13]. For an Indonesian language dataset, there is a study conducted by Antonius Angga Kurniawan and Metty Mustikasari comparing LSTM and Convolutional Neural Network (CNN) algorithm in their article titled “Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia”. In this study, it is shown that the CNN algorithm produces a better model with an accuracy of 88% compared to the LSTM algorithm, which has an accuracy of 84% [14]. Meanwhile, a study conducted by Xiangyang Li, Yu Xia, Xiang Long, Zheng Li, and Sujian Li in their article titled “Exploring Text-transformers in AAAI 2021 Shared Task: COVID-19 Fake News Detection in English” compares two types of models: BiLSTM- based models and Transformer-based models. The study shows that the Transformer-based model achieved better results [15].

Based on the aforementioned previous research, to develop a system recognizing the validity of news in the context of artificial intelligence, various algorithms can be utilized, including the BI-LSTM algorithm. BI-LSTM is a deep learning algorithm that is an extension of the LSTM algorithm. BI-LSTM can process sequential data, such as text, bidirectionally [16], [17]. Thus, BI-LSTM is suitable for NLP tasks such as text classification in detecting fake news in this study, aiming to continue exploration in fake news detection, particularly regarding the 2024 general elections in Indonesia. The main contribution of this research is expected to provide a better understanding of the best ways to detect and address the spread of fake news in the context of elections in Indonesia, as well as contribute to the development of more effective and efficient solutions.

2. Methods

2.1. Dataset

2.1.1. Data Preparation

This research utilizes a dataset consisting of Indonesian fake news and facts. These data were collected from various sources using web scraping techniques with several relevant keywords related to the general election, including kampanye, anies, prabowo, ganjar, imin, gibran, mahfud, pemilu, surat suara, kpu, and coblos. Fake news data were collected from a fake news database, the Global Fact-Check Database provided by the turnbackhoax.id website, spanning from February 14, 2023, to April 28, 2024. Additionally, factual news data were collected from trusted national news sites [18], namely Detik.com and CNN Indonesia, within the same timeframe. The data obtained through web scraping included the title, description, date, and source of the news articles. However, the focus is on the news descriptions, extracted from the meta description attribute on Detik.com and CNN Indonesia, and the news content on the GFD site. The collected data from web scraping consisted of 990 fake news articles and 11,683 factual news articles. To address the significant class imbalance, undersampling technique was applied to reduce the number of majority class instances, minimizing bias towards the majority class in the resulting model [19], [20]. Consequently, a more balanced set of factual news data was obtained, totaling 1,009 instances.

2.1.2. Data Labeling

Each data point was labeled 'is_fake' with a value of 0 or 1. A label of 0 indicates that the data is not fake news, while 1 indicates that the news has been identified as fake. Data obtained from the Global Fact-Check Database were automatically labeled as 1 because this website is the source of fake news data used in this research, while data from trusted news portals were automatically labeled as 0 because news from these portals is identified as factual. Table 1 displays a sample of labeled data collected for the study.

Table 1. Data samples with labels

No	Title	Description	Is fake
1	KPU Tak Bisa Tunjukkan Bukti Video Call Verifi. . .	Dalam sidang di DKPP, KPU Sangihe tak bisa men. ..	0
2	PAN Kritik Partai Ummat soal Politik Identitas. . .	Ketum Partai Ummat, Ridho Rahmadi, menyam. . .	0
3	Akun WhatsApp Sekda Temang- gung Hary Ag. ..	Perkenalkan saya Timotius S, S.Sos., M.S sela. . .	1
4	KPU TUNDA PENETAPAN PRABOWO-GIBRAN	GEGER MALAM INI..!!! KPU TUNDA PENETAP. . .	1
5	Rais Aam PBNU: Saya Senang Prabowo Subianto. . .	Miftachul menyebut Prabowo merupakan sos. . .	0

2.1.3. Data Preprocessing

The collected data served as input for the model creation process. The model was built by comparing two feature extraction techniques: TF-IDF and Contextual Representation using the IndoBERT model. Both feature representation techniques were used to generate the best model. Before that, preprocessing was conducted to eliminate noise and irrelevant information during processing.

The preprocessing steps, as illustrated in Figure 1, include case folding, emoji removal, stopword removal, number and symbol removal, whitespace removal, and stemming. All of these steps are used to produce the TF-IDF feature representation. However, for embedding with the IndoBERT model, not all of these preprocessing steps are necessary as they can be handled automatically [21]. The required preprocessing steps are as shown in Figure 2, which include emoji removal, digit and symbol removal, and whitespace removal. The IndoBERT model is used as a pre-trained model to generate features in the form of contextual representation.

2.2. BI-LSTM

This research employs text classification using the BI-LSTM algorithm. The BI-LSTM algorithm is an advancement of the LSTM algorithm, which itself is an improvement of the Recurrent Neural Network (RNN) algorithm designed to address the vanishing gradient problem [22]. In the LSTM algorithm, there exists a memory cell block consisting of forget gates, input gates, and output gates.

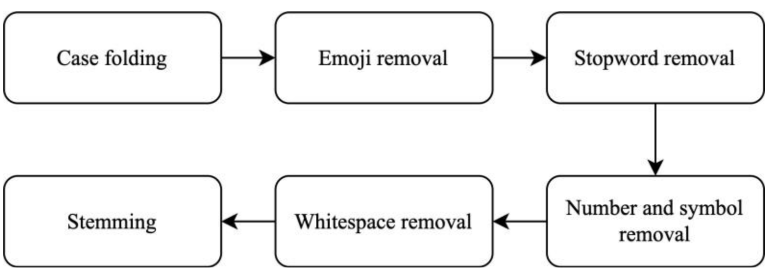

Figure 1. The preprocessing steps for TF-IDF representation

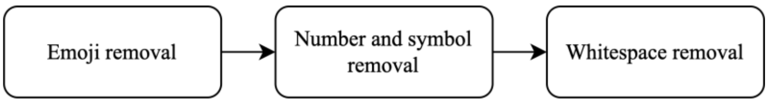

Figure 2. The preprocessing steps for contextual embeddings with IndoBERT model representation

Forget gate is a gate that decides which information to discard or forget from the cell state, Input gate is a gate that determines how much information to let into the cell state. Output gate controls how much information to pass from the cell state to the next layer or to the outside of the model.

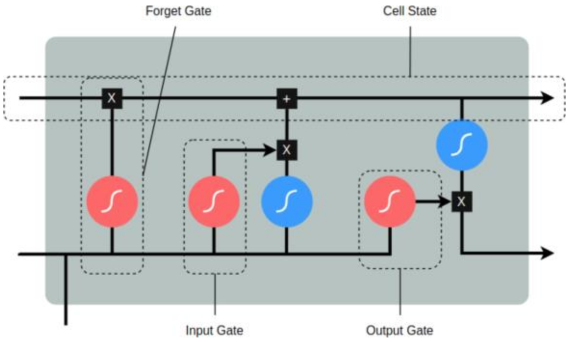

Figure 3. LSTM Architecture with forget gate, input gate, and output gate [23]

Unlike the LSTM algorithm, in the BI-LSTM algorithm, processing is conducted bidirectionally, from both the beginning to the end and vice versa, allowing access to information from both directions in the sequence of data. In this research, this is considered because with BI-LSTM, contextual understanding is enhanced through bidirectional processing.

2.3. Evaluation Method

The evaluation stage is necessary to measure how well the built model performs. The evaluation stage is conducted with four scenarios, where each scenario comprises a combination of feature rep- resentation techniques and the number of epochs used in training. These four testing scenarios are divided into two categories: the first category tests the model with TF-IDF feature representation, and the second category tests the model with contextual embeddings feature representation using the IndoBERT model. Each category of scenarios undergoes two rounds of testing with different numbers of epochs, namely 10 and 20 epochs.

The method for measuring the model's quality in this study is using a confusion matrix. A confusion matrix is a matrix used to evaluate the performance of a classification model [24]. This matrix lists the number of correct and incorrect predictions made by the model for each target class. The metrics present in the confusion matrix are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). To calculate the quality of the resulting model, measurements of Accuracy, Precision, Recall (Sensitivity), and F1 Score are used.

3. Result and Discussion

The results of this research focus on the model created using the BI-LSTM algorithm. From the four testing scenarios, one optimal model was obtained with the most optimal accuracy and loss. The testing results also include the values of its confusion matrix. The model created with the BI-LSTM algorithm utilizes several layers and hyperparameters as listed in Table 2.

Table 2. Model Parameter

Layer	Parameter	Value
Input Layer	input_shape	1000
Reshape Layer	target_shape	(1, 1000)
BI-LSTM 1	units	64
	return_sequences	True
BI-LSTM 2	units	32
	return_sequences	False
Dense Layer	units	1
	activation	sigmoid
Compilation	optimizer	adam
	loss	binary_crossentropy
	metrics	accuracy
Training	batch_size	64

3.1. The results of testing the model with TF-IDF feature representation

The testing of the model with TF-IDF feature representation is conducted twice with different numbers of epochs. This category includes scenarios 1 and 2.

Scenario 1 is an experiment with 10 epochs, resulting in a model with a training accuracy of 100% and a training loss of 0.16%, and a validation accuracy of 92% and a validation loss of 42.92%. The training graph can be seen in Figure 4.

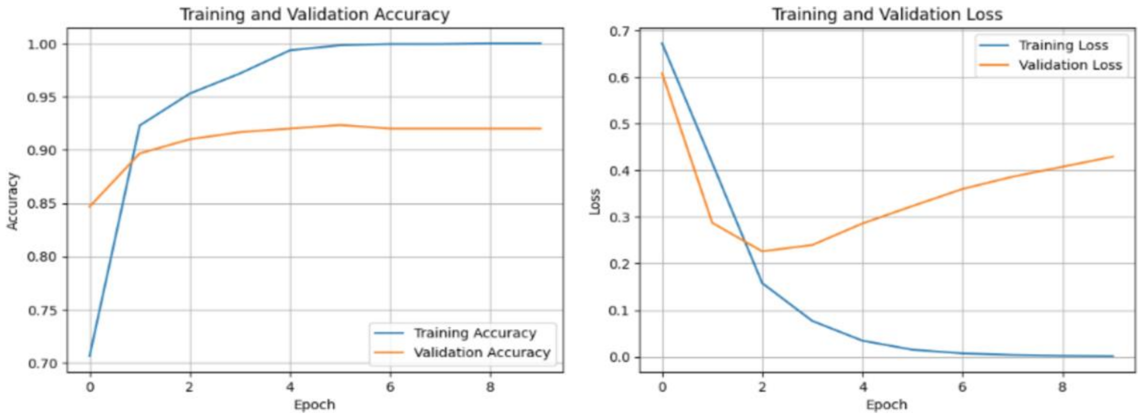

Figure 4. The training result graph for scenario 1

The graph generated in scenario 1 shows a training accuracy reaching 1.00, while the validation accuracy remains relatively stable at around 0.92. Meanwhile, the validation loss curve increases as the training progresses.

Scenario 2 is an experiment with 20 epochs, resulting in a model with a training accuracy of 100% and a training loss of 0.02%. The validation accuracy is 91.33%, and the validation loss is 55.83%. The training graph can be seen in Figure 5.

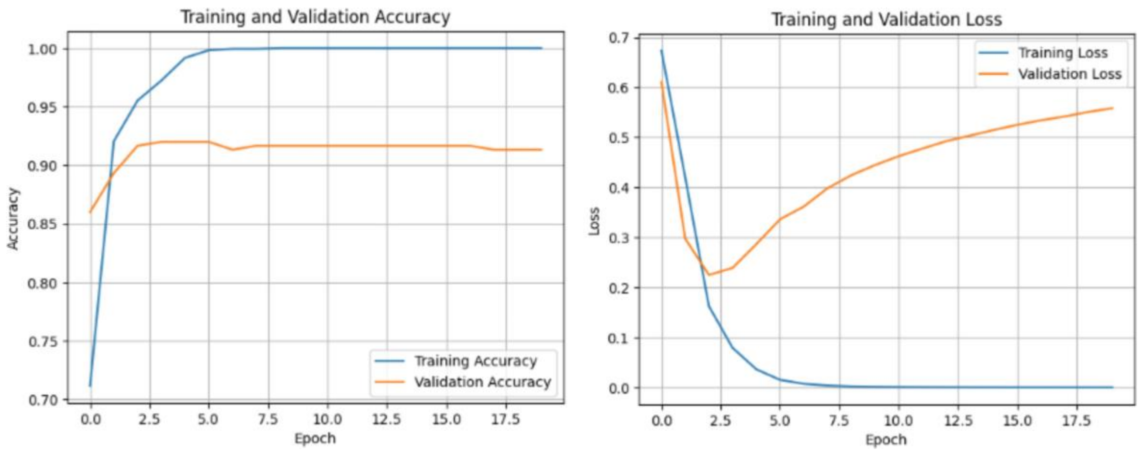

Figure 5. The training result graph for scenario 2

The graph generated in scenario 2 shows a stable training accuracy at 1.00 and a decreasing validation accuracy compared to scenario 1. Meanwhile, the validation loss increases over time.

3.2. The results of testing the model with contextual embeddings feature representation using the IndoBERT model

Testing the model with contextual embeddings feature representation generated by the IndoBERT model is conducted twice with different numbers of epochs. This category includes scenarios 3 and 4.

Scenario 3 is an experiment with 10 epochs, resulting in a model with a training accuracy of 100% and a training loss of 0.24%, and a validation accuracy of 87.63% and a validation loss of 60.27%. The training graph can be seen in Figure 6.

Figure 6. The training result graph for scenario 3

Figure 6 depicts that scenario 3 produces unstable training accuracy. Meanwhile, the validation loss values continue to increase over time.

Scenario 4 is an experiment with 20 epochs, resulting in a model with a training accuracy of 100% and a training loss of 0.02%. The validation accuracy is 85.28%, and the validation loss is 75.45%. The training graph can be seen in Figure 7.

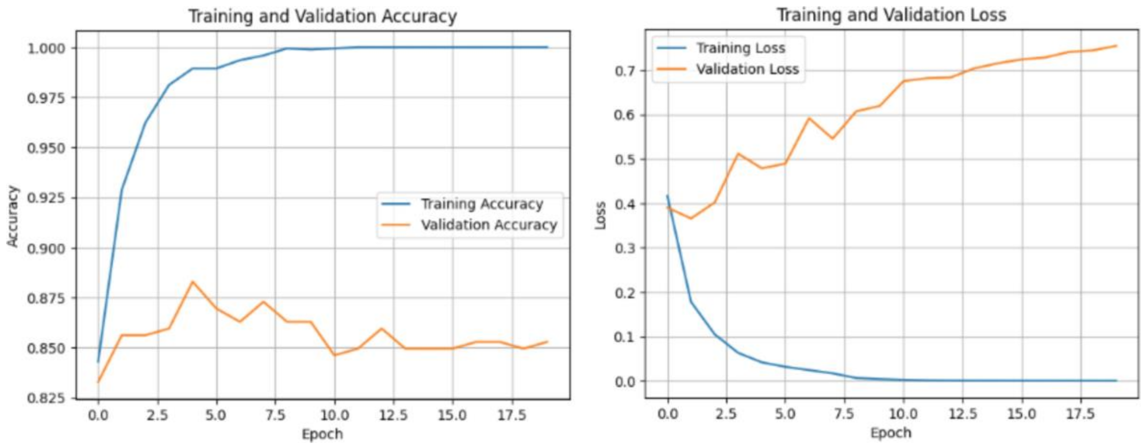

Figure 7. The training result graph for scenario 3

Figure 7 illustrates stable training accuracy and loss. However, the validation accuracy tends to decrease. Meanwhile, the validation loss values continue to increase, resulting in higher numbers over time.

The testing results from scenarios 1-4 all exhibit signs of overfitting, indicated by several factors. These include a training accuracy of 100%, which differs significantly by approximately 8-15% from the validation accuracy [25]. Additionally, there is a notable difference between the training loss and validation loss, ranging from 42-76%.

The best model among scenarios 1-4 is the one generated in scenario 1, specifically the model with TF-IDF feature representation and 10 training epochs. This model yields the highest accuracy among the others, at 92%. Furthermore, it also has the lowest loss, at 42.92%.

The accuracy of 92% is derived from the sum of True Positive and True Negative values from a total of 300 validation data points. The results are listed in Table 3.

Table 3. Confusion Matrix Results for Scenario 1 Model

Prediction	Value
True Positive (TP)	138
True Negative (TN)	138
False Positive (FP)	9
False Negative (FN)	15

The prediction results shown in Table 3 are used to calculate Accuracy, Precision, F1 Score, and Recall to indicate the model's quality. The calculation results are shown in Table 4.

Table 4. Evaluation Metrics Table

Metric	Value
Accuracy	92%
Precision	93.88%
Recall	90.20%
F1 Score	0.9485

Based on the research findings, the BI-LSTM algorithm achieves good results in text classifi- cation for detecting fake news. The results show an accuracy of 92%, precision of 93.88%, recall of 90.20%, and F1 Score of 0.9485. This means that the system can effectively detect fake news. However, there are still weaknesses in classifying news with different variations.

The comparison of results with other studies highlights that the BI-LSTM algorithm used in this research achieved an accuracy of 92%, which is notable. Specifically, when compared to the study "A proposed bi-lstm method to fake news detection" by Islam et al. [26], which also utilized a BI-LSTM algorithm with TF-IDF as the feature representation, the accuracy obtained was 84% . This indicates that the BI-LSTM method implemented in our study demonstrates a slightly higher accuracy. This comparison underscores the effectiveness of the BI-LSTM algorithm with different feature representations and across different datasets and contexts, demonstrating its competitive performance in fake news detection for the Indonesian language.

4. Conclusion

The comparison between models trained with TF-IDF feature representation and models trained with contextual embeddings feature representation using the IndoBERT model, each tested with epochs 10 and 20, results in models that tend to experience overfitting, indicated by significant differences in accuracy and loss between training data and validation data.

The best model is produced with TF-IDF feature representation trained with 10 epochs, achieving an accuracy of 92% and a loss of 42.92%. These figures indicate that the resulting model has good accuracy. However, the high loss also suggests a high prediction bias for certain data points. This research is expected to be continued to reduce the loss in the resulting model by increasing the number of varied training data, thus producing a better model for detecting increasingly diverse types of fake news in the future.

References

[1] M. Ngafifi, "Kemajuan Teknologi Dan Pola Hidup Manusia Dalam Perspektif Sosial Bu- daya," *J. Pembang. Pendidik. Fondasi dan Apl.*, vol. 2, no. 1, pp. 33–47, 2014, doi: 10.21831/jppfa.v2i1.2616.

[2] A. S. Cahyono, "Pengaruh Media Sosial terhadap Penyebaran Hoax oleh Digital Native," Pengaruh media Sos. terhadap perubahan Sos. Masy. di Indones., vol. 1, no. 1, pp. 140–157, 2016.

[3] C. Juditha, "Hoax Communication Interactivity in Social Media and Anticipation (Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya)," *J. Pekommas*, vol. 3, no. 1 SE-Communication, pp. 31–44, Sep. 2018, doi: 10.30818/jpkm.2018.2030104.

[4] MASTEL, "Hasil Survey Wabah HOAX Nasional 2017 | MASTEL Living Enabler." Accessed: May 23, 2024. [Online]. Available: https://mastel.id/hasil-survey-wabah-hoax-nasional-2017/

[5] R. Anisa and R. Rachmaniar, "Hoax Politik Pada Media Sosial Instagram (Studi Etnografi Virtual Tentang Keberadaan Instagram dan Hoax Politik)," *Pros. Komun.*, vol. 1, no. 2, 2017.

[6] M. Rama, D. Sulistyo, & Fatma, and U. Najicha, "PENGARUH BERITA HOAX TER-HADAP KESATUAN DAN PERSATUAN BANGSA INDONESIA," *J. Kewarganegaraan*, vol. 6, no. 1, 2022.

[7] O. N. Julianti et al., "PENERAPAN NATURAL LANGUAGE PROCESSING PADA ANAL- ISIS SENTIMEN JUDI ONLINE DI MEDIA SOSIAL TWITTER," *JATI J. Mhs. Tek. In-form.*, vol. 8, no. 3, pp. 2936–2941, 2024.

[8] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, 2023, doi: 10.1007/s11042-022-13428-4.

[9] D. Sanjaya and S. Budi, "Prediksi Pencapaian Target Kerja Menggunakan Metode Deep Learning dan Data Envelopment Analysis," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 288–300, 2020, doi: 10.28932/jutisi.v6i2.2678.

[10] M. Soori, B. Arezoo, and R. Dastres, "Artificial intelligence, machine learning and deep learning in advanced robotics, a review," *Cogn. Robot.*, vol. 3, pp. 54–70, 2023, doi: https://doi.org/10.1016/j.cogr.2023.04.001.

[11] X. Dong and L. Qian, "Semi-supervised bidirectional RNN for misinformation de- tection," *Mach. Learn. with Appl.*, vol. 10, no. September, p. 100428, 2022, doi: 10.1016/j.mlwa.2022.100428.

[12] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, "Hoax Analyzer for Indonesian News Using Deep Learning Models," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 704–712, 2021, doi: 10.1016/j.procs.2021.01.059.

[13] J. Alghamdi, Y. Lin, and S. Luo, "A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection," *Inf.*, vol. 13, no. 12, pp. 1–28, 2022, doi: 10.3390/info13120576.

[14] A. A. Kurniawan and M. Mustikasari, "Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 2622–4615, 2020, doi: 10.32493/informatika.v5i4.7760.

[15] X. Li, Y. Xia, X. Long, Z. Li, and S. Li, "Exploring Text-Transformers in AAAI 2021 Shared Task: COVID-19 Fake News Detection in English," *Commun. Comput. Inf. Sci.*, vol. 1402 CCIS, pp. 106–115, 2021, doi: 10.1007/978-3-030-73696-5_11.

[16] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Anali- sis Sentimen Terhadap Layanan Grab Indonesia," *J. Manaj. Inform.*, vol. 12, no. 2, pp. 89–99, 2022, doi: 10.34010/jamika.v12i2.7764.

[17] R. Siringoringo, J. Jamaluddin, R. Perangin-angin, E. J. G. Harianja, G. Lumbantoruan, and E. N. Purba, "Model Bidirectional Lstm Untuk Pemrosesan Sekuensial Data Teks Spam," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 7, no. 2, pp. 265–271, 2023, doi: 10.46880/jmika.vol7no2.pp265-271.

[18] Dewan Pers, "Data Perusahaan Pers." Accessed: May 23, 2024. [Online]. Available: https://dewanpers.or.id/data/perusahaanpers

[19] M. C. Untoro and M. A. N. M. Yusuf, "Evaluate of Random Undersampling Method and Majority Weighted Minority Oversampling Technique in Resolve Imabalanced Dataset," *IT J. Res. Dev.*, vol. 8, no. 1, pp. 1–13, 2023, doi: 10.25299/itjrd.2023.12412.

[20] M. Zheng et al., "UFFDFR: Undersampling framework with denoising, fuzzy c-means cluster- ing, and representative sample selection for imbalanced data classification," *Inf. Sci. (Ny).*, vol. 576, pp. 658–680, 2021, doi: https://doi.org/10.1016/j.ins.2021.07.053.

[21] V. Lendave, "A Guide to Text Preprocessing Using BERT," analyticsindiamag. Accessed: May 23, 2024. [Online]. Available: https://analyticsindiamag.com/a-guide-to-text-preprocessing- using-bert/

[22] M. N. A. Putera Khano, D. R. S. Saputro, S. Sutanto, and A. Wibowo, "Sentiment Analysis With Long-Short Term Memory (Lstm) and Gated Recurrent Unit (Gru) Algo- rithms," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 4, pp. 2235–2242, 2023, doi: 10.30598/barekengvol17iss4pp2235-2242.

[23] D. Rengasamy, M. Jafari, B. Rothwell, X. Chen, and G. Figueredo, "Deep Learning With Dynamically Weighted Loss Function for Sensor-based Prognostics and Health Management," *Sensors*, Jan. 2020, doi: 10.3390/s20030723.

[24] M. Fahmy Amin, "Confusion Matrix in Three-class Classification Problems: A Step-by-Step Tutorial," *J. Eng. Res.*, vol. 7, no. 1, pp. 0–0, 2023, doi: 10.21608/erjeng.2023.296718.

[25] X. Ying, "An Overview of Overfitting and its Solutions," *J. Phys. Conf. Ser.*, vol. 1168, no. 2, 2019, doi: 10.1088/1742-6596/1168/2/022022.

[26] T. Islam, Hosen, Md Alamin, A. Mony, M. T. Hasan, I. Jahan, and A. Kundu, "A Proposed BI- LSTM Method to Fake News Detection," *2022 International Conference for Advancement in Technology, ICONAT 2022*, 2022, doi: https://doi.org/10.1109/ICONAT53423.2022.9725937.