Transformer-based Indonesian Language Model for Emotion Classification and Sentiment Analysis

Hendri Ahmadian
Graduate School of Mathematics and Applied Sciences
Universitas Syiah Kuala
Banda Aceh, Indonesia
hendridian@mhs.usk.ac.id

Taufik Fuadi Abidin*
Department of Informatics, Faculty of Mathematics and Natural Sciences
Universitas Syiah Kuala
Banda Aceh, Indonesia
taufik.abidin@usk.ac.id

Hammam Riza
Research Center for Artificial Intelligence and Cyber Security
National Research and Innovation Agency
Jakarta, Indonesia hammam.riza@brin.go.id

Kahlil Muchtar
Department of Electrical and Computer Engineering, Faculty of Engineering
Universitas Syiah Kuala
Banda Aceh, Indonesia
kahlil@usk.ac.id

2023 International Conference on Information Technology and Computing (ICITCOM) | 979-8-3503-5963-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICITCOM60176.2023.10442970

***Abstract*—The rapid development of social networks has made much user-generated data accessible for public evaluation. These data can be used for multiple purposes, such as textual analysis of comments and reviews. This article employs a variant of the bidirectional encoder representations from Transformer (BERT) model designed explicitly for Bahasa Indonesia (called IndoBERT model) to enhance the performance of the Indonesian natural language understanding benchmark in tasks, such as sentiment analysis and emotion classification. The two tasks were tested using a hybrid method, combining the IndoBERT model's last hidden layer summation with a neural network model. The performance of the resulting model was assessed using the F1-score metric. The experimental results show that the proposed model attains an accuracy of 0.92 and 0.76 for sentiment analysis and emotion classification, respectively.**

***Keywords*—sentiment analysis, emotion classification, benchmark IndoNLU, neural networks**

I. Introduction

The development of social media and e-commerce platforms has granted consumers much freedom to express their ideas and opinions. This event has produced a significant volume of interesting data to investigate extensively. Consequently, several scholars [1] have concentrated on devising methodologies for scrutinizing comments and evaluations in text, commonly referred to as sentiment analysis and emotion classification.

Sentiment analysis and emotion classification are crucial in natural language processing (NLP) research. Sentiment analysis involves determining the polarity of opinions conveyed in a specific text. These polarities are mainly categorized into positive, negative, and neutral. Emotion classification involves identifying the expression of emotions expressed in the text. Emotion classification has far-reaching implications and can be used in different fields, such as psychology, e-commerce, customer feedback analysis, e-learning, and human-computer interaction [2].

Deep learning has allowed the utilization of different neural networks to address NLP challenges, including convolutional neural networks (CNNs) [3], recurrent neural networks (RNNs) [4], graph-based neural networks [5], and attention mechanisms [6]. Additionaly, there are several techniques for evaluating emotion classification, including lexicon-based approaches [7], machine learning methods [8], and model unification [9].

Recently, Transformer architecture has significantly advanced NLP research by leveraging pre-trained models. The models provided by Transformers have successfully captured both the syntax [10] and semantics of textual sentences [11]. However, these models are primarily used for languages like English with comprehensive data sources.

Several researchers employed pre-trained Transformers because they can acquire universal language representations and circumvent the need to train new models from the beginning. They are also valuable for evaluating NLP tasks [12]. Embeddings from language models (ELMo) [13] and bidirectional encoder representations from Transformers (BERT) [14] are crucial for evaluating contextual pre-trained language models. Subsequently, broad language models, such as general language understanding evaluation [15] and Chinese language understanding and evaluation [16], were introduced.

The development of BERT architecture models for the Indonesian language datasets started commences with the natural language understanding (IndoNLU) benchmark, encompassing 12 tasks. This study employs the hyperparameters of the BERT and ALBERT models [17]. The accuracy rates for the emotion classification and sentiment analysis, are 79.57% and 92.72%, respectively.

Extensive research has been conducted to address of emo-

*Corresponding Author

tion categorization and sentiment analysis tasks, employing several model techniques and extending to datasets in languages other than English. The BERT model is used for emotion classification and sentiment analysis in various languages such as Chinese [18], Arabic [19], Italian [20], Russian [21], Turkish [22], and Indonesian [23].

Previous research has demonstrated the effectiveness of hybrid models that combine neural network models and BERT models for emotion classification and sentiment analysis tasks. These hybrid models have consistently outperformed standalone models in terms of accuracy [24]–[27]. Furthermore, methodologies based on the BERT model have demonstrated substantial proficiency in extracting crucial characteristics, such as word embedding vectors. This method aggregates the features derived from the most recent hidden layers and uses them as input vectors for the neural network. Empirical evidence has shown that implementing this approach enhances the precision of emotion classification and sentiment analysis [28], [29].

Therefore, this study developed a hybrid model that combines the Indonesian BERT and neural network models. The proposed model employs a BERT modification layer with a neural network model. The improved layer technique in the Indonesian BERT model, combined with a neural network model, enhances the accuracy of sentiment analysis results.

The remaining sections of the paper are organized as follows. Section 2 discusses literature review. Section 3 describes the research methods. In Section 4, we present and discuss the experimental results. Finally, Section 5 concludes the paper and provides future research directions.

II. Literature Review

A. Bidirectional Long Short-Term Neural Network

Long short-term memory (LSTM) is a deep learning approach that incorporates three nonlinear gate structures (forget gate, input gate, and output gate) to regulate the information flow. These gates selectively allow or prevent information from entering or leaving, enabling the model to capture temporal dependencies in a unidirectional manner. This study uses the bidirectional LSTM (BiLSTM) architecture to acquire latent characteristics that can process sequential input in both forward and backward directions. The BiLSTM design comprises many hidden layers arranged in a stacked BiLSTM network, which improves the model's capacity to capture and acquire knowledge of its latent properties. Within each concealed layer, two conventional hidden layers handle data in the forward and backward directions, capturing the preceding and subsequent features, respectively.

B. Transformer

Transformer is a neural network design introduced in [30]. Transformer use encoder decoder pairs, however their structure differs from RNN and CNN models. Transformers depend only on a multi-head self-attention mechanism in the encoder (left block) and decoder (right block). This network architecture demonstrates exceptional performance in addressing language tasks such as syntactic parsing and language translation via sequence transduction.

C. Bidirectional Encoder Representations from Transformers

The structure of BERT model has 12 bidirectional Transformers encoder blocks. The model consists of 768 hidden layers and has 110 million parameters. It incorporates the attention mechanism notion. Bidirectionality is a distinguishing feature of this model, which sets it apart from conventional language models trained solely from left to right. The BERT approach encompasses two distinct stages: pre-training and fine-tuning. In the pre-training phase, the BERT model is trained using unlabeled data to perform. various pre-training tasks. The model is initialized with pre-trained parameters in the fine-tuning phase. The parameters are modified using labeled data for various NLP tasks. Despite being initialized with identical pre-training settings, each NLP problem requires its distinct fine-tuning model.

D. Indonesian Natural Language Understanding

The IndoNLU benchmark is a resource used for training, evaluating, and benchmarking the process of understanding the Indonesian language. The IndoNLU benchmark has recently introduced a pre-trained model called IndoBERT for the Indonesian language dataset. The IndoBERT model is further categorized into two versions: IndoBERT Base and IndoBERT Large. The IndoBERT Base and IndoBERT Large comprise has 12 and 24 hidden layers, respectively. Due to restrictions, each model is pre-trained with a maximum sequence length of 128 and 512 in the first and second stages, respectively.

E. Related Studies

Sentiment analysis and emotion classification have recently emerged as prominent areas of study in NLP. Suciati [31] compared machine learning and deep learning techniques for sentiment analysis and emotion classification, using a multi-label classification approach. Ali [32] used sentiment analysis and emotion classification models to promptly identify pandemic outbreaks using machine learning and deep learning techniques.

Recently, several scholars have examined the existing literature on sentiment analysis. For example, Drus [33] conducted a literature review on sentiment analysis on social media, investigating various approaches, platforms, and applications. Xu [34] presented a comprehensive examination of the difficulties and probable obstacles encountered by scholars while conducting sentiment analysis on social media. Furthermore, various studies have been accomplished in the field of emotion classification. These studies include a comparison analysis of different Transformers models. For example, Lin [35] employed a deep learning model to perform sentiment analysis on an Indonesian language dataset. Glenn [36] employed the same deep learning model for emotion classification tasks. Imaduddin [37] performed sentiment analysis tasks using a BERT technique to analyze reviews on health applications.

III. Research Methods

This study employs two methodologies for the unification model. The first is the IndoBERT model [38], which combines the BiLSTM model (Fig. 1). The second is the IndoBERT-modified layer model [39], which integrates the BiLSTM model (Fig. 2). The evaluation method was only conducted on the pre-trained IndoBERT base model for each model. Google Collaborative Environment acts as an experimental solution for executing notebooks on Google Cloud Virtual Machines equipped with hardware accelerators, specifically Graphics Processing Units, providing a total of 100 compute units.

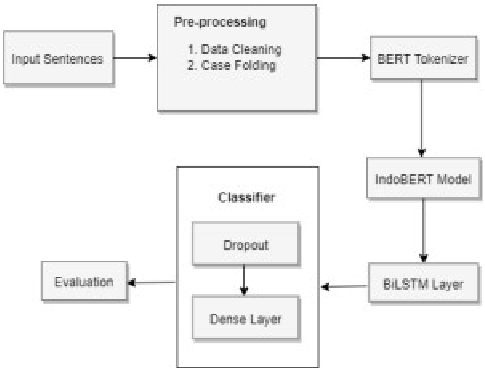

Fig. 1. Representation of the IndoBERT model with the BiLSTM model.

The IndoNLU benchmark dataset was used to train, validate, and test the two models. The IndoNLU benchmark dataset encompass emotion classification (EmoT) and sentiment analysis (SmSA) challenges.The dataset description (EmoT) employs the classifications of anger, fear, happiness, love, and sadness. The training dataset consists of 3,521 samples, the validation dataset comprises 440 records, and the testing dataset encompasses 442 instances. The SmSA dataset, including the positive, negative, and neutral categories, contains 11,000 training data, 1,260 validation data, and 500 tests.

In this study, we employ new models to enhance the accuracy of emotion and sentiment analysis tasks on the IndoNLU dataset. The sentiment analysis tasks conducted in this study are non dependent, where the input is sentence-level sentiment analysis. The initial model integrates the IndoBERT and BiLSTM models. In contrast, the second model employs the IndoBERT model with the sum of the last four hidden layers along with the BiLSTM model.

We performed a preprocessing stage when sentences were inputted. The text was transformed at this stage, changing all capital letters to lowercase letters. Then, the text/sentence was analyzed to identify, fix, and eliminate any flaws or inconsistencies in the data. Subsequently, the input text was tokenized and forwarded to the IndoBERT model. The resulting output of this model is a vector, which is then used as input in the BiLSTM layer. Subsequently, the classification and model evaluation procedures are executed. The second model uses layer summation on the four ultimate hidden layers of the IndoBERT model. Then, the BiLSTM model will receives the

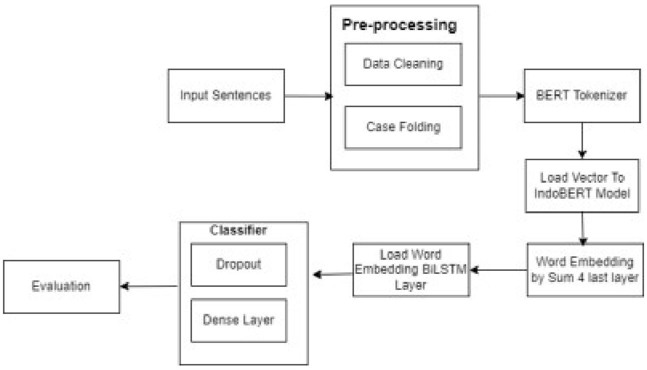

Fig. 2. IndoBERT architecture with the last hidden layer summed to the BiLSTM model. [42]

input, and the feature extraction process is performed on the dense layer. Dropout is then used to mitigate overfitting. The evaluation stage determines the efficiency of the two model approaches using the F1 score calculation. The IndoBERT model sums the last layer on the cumulative output of the previous two, three, and five hidden layers to offer various layer changes.

IV. Results and Discussion

The IndoNLU benchmark dataset was used in emotion classification and sentiment analysis to enhance the model's accuracy in addressing the emotion classification and sentiment analysis tasks. A hybrid model that combines the IndoBERT and BiLSTM models was used as the initial approach. Then, the IndoBERT model employs the summation of the four hidden final layers and combines it with the BiLSTM model. The results demonstrate that the IndoNLU benchmark is a prominent source for evaluating Indonesian language models. The IndoBERT model was also evaluated on various recaps of the last hidden layer, specifically layers 2, 3, and 5. The hyperparameters employed in this study are presented in Table 1.

TABLE I
HYPERPARAMETER SETTING

No	Hyperparameter	Value
1	learning rate	0.000003
2	batch size	16
3	num workers	12
4	max seq length	512
5	epoch	10
6	pre-trained IndoBERT	base phase 1 and 2

A. *Emotion Classification*

Tables 2 and 3 show the F1-score evaluation results for the emotion classification task using the IndoBERT model base phases 1 and 2, respectively. The F1-score accuracy values for the baseline in the IndoNLU benchmark are 0.75 (base phase 1) and 0.76 (base phase 2).

2023 International Conference on Information Technology and Computing (ICITCOM)

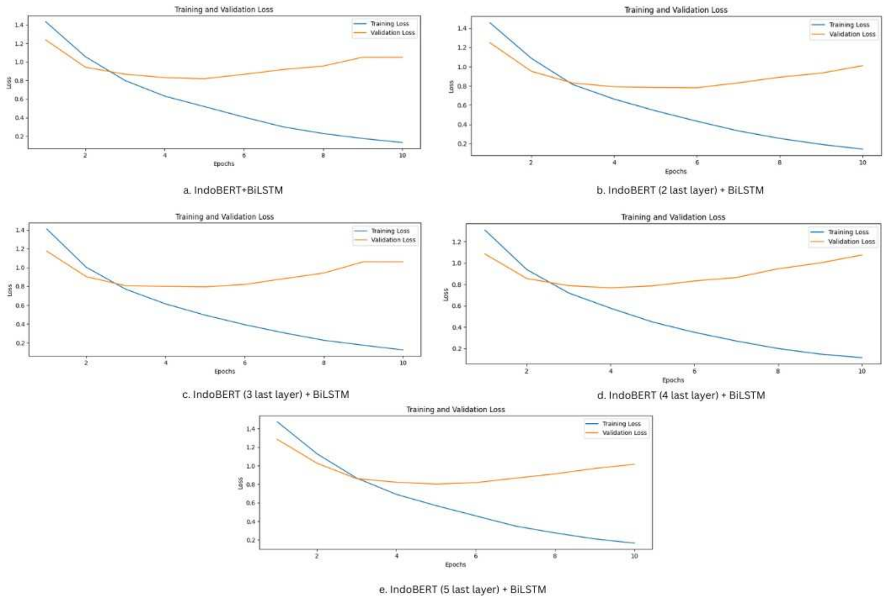

Fig. 3. Training and validation loss for emotion classification.

Table 2 shows that the F1-score value is higher for the IndoBERT model when the last four hidden layers are summed in the pre-trained base phase 1, with an increase of 0.01 by 0.76. However, integrating the BiLSTM model in the pre-trained base phase 2 of the IndoBERT model did not improve the model's performance. Subsequently, we compared the training and validation loss for the task in emotion classification tasks.

The process is demonstrated in Fig. 3, which depicts IndoBERT base phases 1 and 2, which are comparatively similar. As shown in Fig. 3, the loss variable decreases during the data training, whereas the validation data increases during the third and fourth epochs.

TABLE II
EVALUATION RESULTS OF THE HYBRID MODEL FOR EMOTION CLASSIFICATION ON BASE PHASE 1

Hybrid Model	ACC	REC	PRE	F1
IndoBERT+BilLSTM	0.74	0.74	0.76	0.75
IndoBERT(2 last layer)+BiLSTM	0.72	0.72	0.75	0.73
IndoBERT(3 last layer)+BiLSTM	0.74	0.75	0.76	0.75
IndoBERT(4 last layer)+BiLSTM	0.75	0.75	0.76	0.76
IndoBERT(5 last layer)+BiLSTM	0.74	0.75	0.75	0.75

ACC: Accuracy; REC: Recall; PRE: Precision; F1: F1-Score

B. *Sentiment Analysis*

Tables 4 and 5 summarize the F1-score evaluation results for sentiment analysis tasks using IndoBERT pre-trained phases 1 and 2, respectively. For comparison, the F1-score accuracy values of baseline in the IndoNLU benchmark of sentiment analysis tasks are 0.87 (base phase 1) and 0.87 (base phase 2).

Fig. 4 shows that the loss decreases during training, whereas the validation data increases during the second epoch. The

TABLE III
EVALUATION RESULTS OF THE HYBRID MODEL FOR EMOTION CLASSIFICATION ON BASE PHASE 2

Hybrid Model	ACC	REC	PRE	F1
IndoBERT+BilLSTM	0.72	0.73	0.74	0.73
IndoBERT(2 last layer)+BiLSTM	0.72	0.73	0.76	0.74
IndoBERT(3 last layer)+BiLSTM	0.75	0.75	0.76	0.75
IndoBERT(4 last layer)+BiLSTM	0.74	0.75	0.76	0.75
IndoBERT(5 last layer)+BiLSTM	0.74	0.75	0.75	0.75

ACC: Accuracy; REC: Recall; PRE: Precision; F1: F1-Score

TABLE IV
EVALUATION RESULTS OF THE HYBRID MODEL FOR SENTIMENT ANALYSIS ON BASE PHASE 1

Hybrid Model	ACC	REC	PRE	F1
IndoBERT+BilLSTM	0.90	0.85	0.91	0.87
IndoBERT(2 last layer)+BiLSTM	0.93	0.91	0.92	0.92
IndoBERT(3 last layer)+BiLSTM	0.91	0.87	0.92	0.89
IndoBERT(4 last layer)+BiLSTM	0.92	0.88	0.93	0.90
IndoBERT(5 last layer)+BiLSTM	0.92	0.88	0.92	0.90

ACC: Accuracy; REC: Recall; PRE: Precision; F1: F1-Score

TABLE V
EVALUATION RESULTS OF THE HYBRID MODEL FOR SENTIMENT ANALYSIS ON BASE PHASE 2

Hybrid Model	ACC	REC	PRE	F1
IndoBERT+BilLSTM	0.92	0.89	0.92	0.90
IndoBERT(2 last layer)+BiLSTM	0.90	0.84	0.91	0.86
IndoBERT(3 last layer)+BiLSTM	0.91	0.86	0.91	0.88
IndoBERT(4 last layer)+BiLSTM	0.92	0.88	0.92	0.90
IndoBERT(5 last layer)+BiLSTM	0.92	0.89	0.92	0.90

ACC: Accuracy; REC: Recall; PRE: Precision; F1: F1-Score

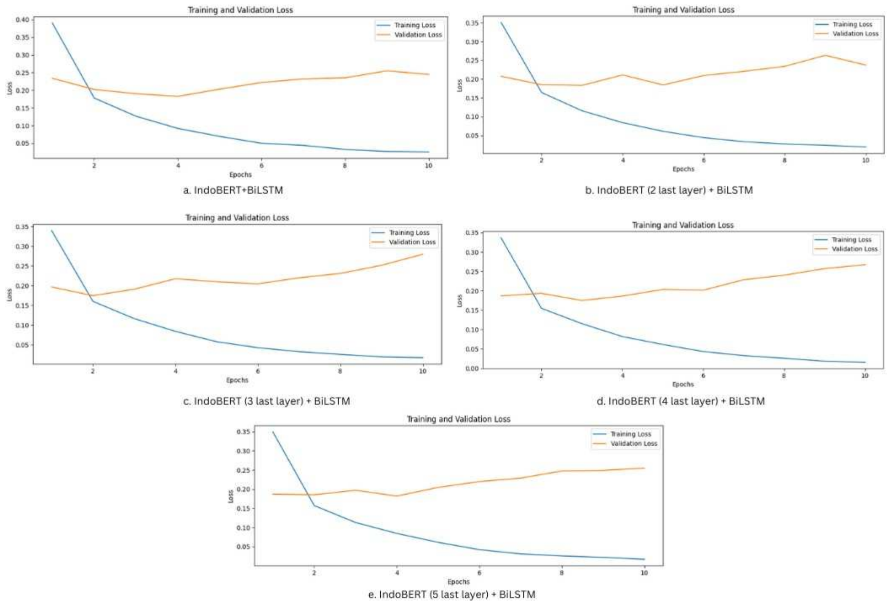

Fig. 4. Training and validation loss for sentiment analysis

IndoBERT model was evaluated using the BiLSTM model by summing the last 2, 3, 4, and 5 hidden layers. The results show an increase in the F1-score by 0.02, 0.03, and 0.005, respectively. The highest F1-score of 0.92 was achieved when summing the last two hidden layers in the pre-trained IndoBERT base phase 1 model. The pre-trained IndoBERT base phase 2 experienced a marginal rise of 0.01 and 0.03, with the maximum increase recorded at 0.90.

The hybrid model that combines the IndoBERT model by summing the last 2, 3, 4, and 5 hidden layers with the BiLSTM model consistently achieves a higher F1 score than the IndoNLU benchmark model in the emotion classification and sentiment analysis tasks.

V. Conclusions

This study aims to improve the accuracy of the IndoBERT model in emotion classification and sentiment analysis using the IndoNLU benchmark dataset. This accuracy was achieved using a model fusion technique using neural networks. The proposed IndoBERT model combines the summation of the last 2, 3, 4, and 5 hidden layers with the BiLSTM model. This combination achieves better performance than the IndoNLU benchmark. In future studies, we will modify the IndoBERT model layer using alternative bidirectional RNN models, such as BiGRU [40] [41], and evaluate the model's effectiveness with various NLP tasks like named entity recognition on the IndoNLU benchmark.

Acknowledgment

We sincerely thank the Ministry of Education, Culture, Research, and Technology, Directorate for Research, Technology, and Community Services, who has funded this research under the Doctoral Dissertation Research Grant (PDD) 672/UN11.2.1/PT.01.03/DPRM/2023. Also, thank the Institute for Research and Community Services (LPPM), Universitas Syiah Kuala, for monitoring and evaluating the research process. Additionally, we would like to thank the technical team who assisted us in handling experiments. All the contributions and assistance provided have been invaluable to the progress of this research.

References

[1] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," Mathematical and computational applications, vol. 23, no. 1, p. 11, 2018.

[2] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," Knowl Inf Syst, vol. 62, pp. 2937–2987, 2020.

[3] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional Sequence to Sequence Learning," 2017.

[4] P. Liu, X. Qiu, and X. Huang, "Recurrent Neural Network for Text Classification with Multi-Task Learning," 2016.

[5] D. Marcheggiani, J. Bastings, and I. Titov, "Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks," Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.08313

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

[7] Diamantini, C., Mircoli, A., Potena, D., and Storti, E. (2015, June). Semantic disambiguation in a social information discovery system. In 2015 International Conference on Collaboration Technologies and Systems (CTS) (pp. 326-333). IEEE.

[8] A. Glenn, P. LaCasse, and B. Cox, "Emotion classification of Indonesian Tweets using Bidirectional LSTM," Neural Comput Appl, vol. 35, no. 13, pp. 9567–9578, 2023.

[9] C. Karras, A. Karras, D. Tsolis, M. Avlonitis, and S. Sioutas, "A Hybrid Ensemble Deep Learning Approach for Emotion Classification," in 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 3881–3890.

[10] G. Jawahar, B. Sagot, and D. Seddah, "What does BERT learn about the structure of language?," in ACL 2019-57th Annual Meeting of the Association for Computational Linguistics, 2019.

[11] C. Sun, L. Huang, and X. Qiu, "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence," arXiv preprint arXiv:1903.09588, 2019.

2023 International Conference on Information Technology and Computing (ICITCOM)

[12] X. Han et al., “Pre-trained models: Past, present and future,” AI Open, vol. 2, pp. 225–250, 2021, doi: https://doi.org/10.1016/j.aiopen.2021.08.002.

[13] M. E. Peters et al., “Deep contextualized word representations,” Feb. 2018, [Online]. Available: http://arxiv.org/abs/1802.05365

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[15] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, “GLUE: A multi-task benchmark and analysis platform for natural language understanding,” arXiv preprint arXiv:1804.07461, 2018.

[16] L. Xu et al., “CLUE: A Chinese Language Understanding Evaluation Benchmark,” Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.05986

[17] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R.. (2019). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. abs/1909.11942. http://arxiv.org/abs/1909.11942

[18] F. Li, Y. Huo, and L. Wang, “Research on Chinese Emotion Classification using BERT-RCNN-ATT,” WSEAS Transactions on Communications, vol. 22, pp. 17–31, 2023.

[19] M. M. Abdelgwad, “Arabic aspect based sentiment classification using BERT,” arXiv preprint arXiv:2107.13290, 2021.

[20] A. Chiorrini, C. Diamantini, A. Mircoli, D. Potena, and E. Storti, “EmotionAlBERTo: Emotion Recognition of Italian Social Media Texts Through BERT,” in 2022 26th International Conference on Pattern Recognition (ICPR), IEEE, 2022, pp. 1706–1711.

[21] Smetanin S. 2022. RuSentiTweet: a sentiment analysis dataset of general domain tweets in Russian. PeerJ Computer Science 8:e1039 https://doi.org/10.7717/peerj-cs.1039

[22] M. M. Mutlu and A. Özgür, “A dataset and BERT-based models for targeted sentiment analysis on Turkish texts,” arXiv preprint arXiv:2205.04185, 2022.

[23] E. I. Setiawan, L. Kristianto, A. T. Hermawan, J. Santoso, K. Fujisawa, and M. H. Purnomo, “Social Media Emotion Analysis in Indonesian Using Fine-Tuning BERT Model,” in 2021 3rd East Indonesia Conference on Computer and Information.

[24] Z. Gou and Y. Li, “Integrating BERT Embeddings and BiLSTM for Emotion Analysis of Dialogue,” Comput Intell Neurosci, vol. 2023, 2023.

[25] X. Li, Y. Lei, and S. Ji, “BERT-and BiLSTM-Based Sentiment Analysis of Online Chinese Buzzwords,” Future Internet, vol. 14, no. 11, p. 332, 2022.

[26] R. Cai et al., “Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM,” IEEE access, vol. 8, pp. 171408–171415, 2020.

[27] C.-H. Lin and U. Nuha, “Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy,” J Big Data, vol. 10, no. 1, p. 88, 2023, doi: 10.1186/s40537-023-00782-9.

[28] A. Neelima and S. Mehrotra, “A Comprehensive Review on Word Embedding Techniques,” in 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), IEEE, 2023, pp. 538–543.

[29] S. Selva Birunda and R. Kanniga Devi, “A review on word embedding techniques for text classification,” Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2020, pp. 267–281, 2021.

[30] A. Vaswani et al., “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[31] A. Suciati and I. Budi, “Aspect-Based Sentiment Analysis and Emotion Detection for Code-Mixed Review,” International Journal of Advanced

[32] M. Z. Ali, K. Javed, and A. Tariq, “Sentiment and emotion classification of epidemic related bilingual data from social media,” arXiv preprint arXiv:2105.01468, 2021.

[33] Z. Drus and H. Khalid, “Sentiment analysis in social media and its application: Systematic literature review,” Procedia Comput Sci, vol. 161, pp. 707–714, 2019

[34] Q. A. Xu, V. Chang, and C. Jayne, “A systematic review of social media-based sentiment analysis: Emerging trends and challenges,” Decision Analytics Journal, vol. 3, p. 100073, 2022, doi: https://doi.org/10.1016/j.dajour.2022.100073.

[35] C.-H. Lin and U. Nuha, “Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy,” J Big Data, vol. 10, no. 1, p. 88, 2023.

[36] A. Glenn, P. LaCasse, and B. Cox, “Emotion classification of Indonesian Tweets using Bidirectional LSTM,” Neural Comput Appl, vol. 35, no. 13, pp. 9567–9578, 2023.

[37] H. Imaduddin, F. Y. A’la, and Y. S. Nugroho, “Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach,” International Journal of Advanced Computer Science and Applications, vol. 14, no. 8, 2023.

[38] Wilie, Bryan, et al. ”IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding.” arXiv preprint arXiv:2009.05387 (2020).

[39] Song, M., Feng, Y., and Jing, L. (2022, December). Utilizing BERT intermediate layers for unsupervised keyphrase extraction. In Proceedings of the 5th International Conference on Natural Language and Speech Processing (ICNLSP 2022) (pp. 277-281).

[40] Tan, K. L., Lee, C. P., Lim, K. M., and Anbananthen, K. S. M. (2022). Sentiment analysis with ensemble hybrid deep learning model. IEEE Access, 10, 103694-103704.

[41] Islam, M. S., Sultana, S., Roy, U. K., Al Mahmud, J., and Jahidul, S. (2021). HARC-new hybrid method with hierarchical attention based bidirectional recurrent neural network with dilated convolutional neural network to recognize multilabel emotions from text. J. Ilm. Tek. Elektro Komput. Dan Inform.(JITEKI), 7, 142-153.

[42] N. Rai, D. Kumar, N. Kaushik, C. Raj, A. Ali, “Fake News Classification using Transformer Based Enhanced LSTM and BERT,” International Journal of Cognitive Computing in Engineering, vol. 3, pp. 98–105, June 2022, doi: 10.1016/j.ijcce.2022.03.003.