International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 1, February 2024, pp. 1071~1078
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i1.pp1071-1078

Indonesian multilabel classification using IndoBERT embedding and MBERT classification

Ghinaa Zain Nabiilah, Islam Nur Alam, Eko Setyo Purwanto, Muhammad Fadlan Hidayat
Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta Indonesia

Article Info

Article history:

Received Jun 23, 2023
Revised Jul 13, 2023
Accepted Sep 6, 2023

Keywords:

IndoBERT embedding
Indonesian social media
MBERT
Text classification
Toxic comment

ABSTRACT

The rapid increase in social media activity has triggered various discussion spaces and information exchanges on social media. Social media users can easily tell stories or comment on many things without limits. However, this often triggers open debates that lead to fights on social media. This is because many social media users use toxic comments that contain elements of racism, radicalism, pornography, or slander to argue and corner individuals or groups. These comments can easily spread and trigger users vulnerable to mental disorders due to unhealthy and unfair debates on social media. Thus, a model is needed to classify comments, especially toxic ones, in Indonesian. Transformer-based model development and natural language processing approaches can be applied to create classification models. Some previous research related to the classification of toxic comments has been done, but the classification results of the model still require exploration to get optimal results. So, this research uses the proposed model by using different pre-trained models at the embedding and classification stages, in the embedding stage using Indonesia bidirectional encoder representations from transformers (IndoBERT), and classification using multilingual bidirectional encoder representations from transformers (MBERT). The proposed model provides optimal results with an F1 value of 0.9032.

Corresponding Author:

Ghinaa Zain Nabiilah
Computer Science Department, School of Computer Science, Bina Nusantara University
Kebon Jeruk Street No. 27, West Jakarta 11530, Indonesia
Email: ghinaa.nabiilah@binus.ac.id

1. INTRODUCTION

Social media is internet-based media, which refers to individual interactions and activities in sharing media to exchange information [1]. The rapid advances in information technology have led to increased social media activity, affecting the number of interactions on social media. Social media's openness and speed have unwittingly facilitated information dissemination [2]. The spread of information and the many users who interact are prone to triggering debates due to the ease with which open discussions form between social media users.

Social media users often use toxic comments when debating and cornering an individual or group. So, this often triggers debates that cause great fights on social media. These toxic comments use words that contain elements of hate speech, radicalism, pornography, and or defamation. These comments can easily be spread directly, causing some users to be vulnerable to mental disorders due to unhealthy and unfair debates on social media. With the rapid development of social media activities, keeping online interaction and communication conducive is essential for social media platforms. Making a model that automatically classifies toxic comments, such as hate speech, radicalism, pornography, and or defamation, can help make it easier to keep activities and interactions between users conducive. This can help reduce the negative impact of unhealthy debate on social media. Because of the data from text, the natural language processing (NLP) method can be used to process it. NLP is used in a wide variety of topics involving computational processing and understanding human language [3]. With advances in machine learning and deep learning, various neural networks have been widely used to solve NLP tasks. One example is using artificial neural networks (ANN), long short-term memory (LSTM), and convolution neural networks (CNN) to analyze the sentiments conveyed by Twitter users [4], [5]. Apart from analyzing sentiment, models such as CNN, LSTM, or machine learning models such as ANN and logistic regression can also be used to classify toxic comments from social media [6].

However, neural networks and deep learning models still need to improve in simultaneously analyzing the meaning and semantics of words in sentences. In addition, the resulting model tends to be prone to overfitting on small data [7]. Based on this, pre-trained models can learn language representation as a whole. bidirectional encoder representations from transformers (BERT) are one of the pre-trained models widely used for various NLP tasks. BERT is a model that uses the encoder architecture of the transformer and is designed to use a bidirectional representation [8]. An example of the application of BERT is research conducted by D'Sa *et al.* [9], by classifying toxic comments on Twitter. In this study, a comparison was made using the FastText and BERT word embedding methods classified by the CNN model and bidirectional LSTM (BiLSTM). In addition, a comparison was also made using the entire word embedding model and classification using BERT. The optimal result of this study is using BERT on word embedding and classification with an F1 value of 84% in multi-class classification. Another study by Nabiilah *et al.* [10] also used a pre-trained model with BERT architecture. This study compared several pre-trained models that had been trained with Indonesian language corpus data, namely multilingual BERT (MBERT), Indonesia BERT (IndoBERT), and Indonesia robustly optimized BERT pretraining approach (IndoRoBERTa) small. The optimal result of this study is to use IndoBERT with an F1 value of 0.88978.

Previous research also aligns with recent developments in applying NLP for text classification cases using social media data, where pre-trained models can provide model classification results with better performance in recognizing text patterns. However, the increased access to communication by various groups on social media causes the text and language conveyed to be increasingly unstructured and difficult to analyze the pattern. So, to improve the performance of the model in analyzing the meaning and context of the text conveyed, the development of pre-trained model architecture needs to be further developed. Therefore, this research performs multilabel classification of toxic comments using a BERT-based pre-trained model. The striking difference from this research is using different pre-trained models at several stages. In previous studies, the pre-trained model is usually used simultaneously as a feature extraction and classification model or a feature extraction model (word embedding) such as FastText, or Glove combined with a pre-trained model as a classification model. However, this research needs to do this, where the feature extraction and classification process is carried out using two different pre-trained models, IndoBERT as a feature extraction and MBERT as a classification model. In this study, other pre-processing processes were carried out, namely by translating emoticons and slang words. Because the data comes from social media, words or sentences are often produced in a non-standard form or not following the correct Indonesian spelling. It is necessary to carry out these two different processes.

Toxic comments are a form of spreading or expressing cornering or harassing certain users or groups. The things that are used as comments are usually based on physical form, race, religion, or ethnicity [11]. Research related to the classification of toxic comments conducted by Zhao *et al.* [12] compared several classification models such as BERT, a robustly optimized BERT pretraining approach (RoBERTa), and cross-lingual language model (XLM). The optimal result of this study is using BERT with an F1 value of 0.8824. Other research related to the classification of toxic comments using Indonesian data has been carried out by Azhar and Khodra [13]. The data used in this study are binary by comparison with the CNN, XGBoost, and BERT models. The optimal result is using BERT with an F1 value of 0.9765. In addition, Leite *et al.* [14] also used the BERT model. The types of BERT used in this study were monolingual BERT and multilingual BERT. The optimal result of this study is using multilingual BERT with an F1 value of 0.76.

The study conducted byGuillaume *et al.* [15] used the Reddit dataset to classify toxic comments using several classification models, such as HateBERT, BERTweet, and RoBERTa. The results of this study's classification were with an F1 value of 0.9519 using HateBERT, 0.9603 using BERTweet, and the optimal result was 0.9673 using Roberta. Saraiva *et al.* [16] also used the BR-BERT and MBERT models to classify toxic comments. The results of the model classification are with an F1 value of 0.76 using BR-BERT and 0.75 using M-BERT. Research conducted by Khan *et al.* [17] also uses the MBERT model in analyzing various product sentiments and reviews of user services on social media. The optimal result of this research is the F1 value of 81.49%.

2. RESEARCH METHOD

This section describes the dataset used and the proposed steps in performing multilabel classification of toxic comments of social media users in Indonesia. The proposed method uses a Transformer-based pre-trained model, such as BERT, developed in several languages, especially Bahasa Indonesia. In addition, to process less structured social media data, this research also applies more complete data pre-processing by applying several other processes, such as translating emoticons and translating slang words.

2.1. Dataset

The dataset used in this study is data from users' comments on Indonesian various social media such as Instagram, Twitter, and Kaskus. This data was collected, processed, and labeled manually by Izzan *et al.* [18]. The characteristic of this dataset is multi-label, where each comment can be grouped into one or more than one label. Table 1 contains an example of a multi-label dataset.

- Pornography describes comments that contain obscenity or sexual exploitation that violates the norms of decency in society.
- Hate Speech, describing comments based on identity sentiments concerning ancestry, race, religion, nationality, ethnicity, or specific groups.
- Radicalism describes a comment with an understanding or flow that wants political and social change and renewal with an extreme attitude, even using violence.
- Defamation, describing comments by attacking the honor or good name of a person or a particular group.

The number of datasets used in this study is 7,773 data. Then, the data is divided into training, validation, and test data. The data is divided into 81% train data, 9% validation data, and 10% test data. Table 2 contains the number of data divisions.

Table 1. Example of the dataset

Comment	Translation of Comments in English	Pornography	Hate Speech	Radicalism	Defamation
Hanya orang gila yang tidak tahu aturan yang ngomng seperti kamu dibaca lagi bahasa indonesia kamu dapat brpa sih waktu sekolah miris	Only crazy people who don't know the rules that mumble like you read again Indonesian language you get what the hell time school sadly	0	0	0	1
Ya ampun pahanya kalau difoto kelihatan mulus alus kalau di video tanpa edit gila buyar bagian bokong pas mau berdiri hitam. Anjing ini om fatah	Oh my gosh his thighs if photographed look smooth if in the video without crazy edits the buttocks when he wants to stand black. This dog om fatah	1	0	0	1
Masyarakat harus bersatu dalam satu wilayah negara dengan satu pemimpin baru yang cerdas, dan bendera negara baru untuk mencapai negara maju, sejahtera.	People must unite in one country territory with one smart new leader, and a new country flag to achieve a developed, prosperous country.	0	0	1	0

Information: 0=”no”, 1= “yes”

Table 2. Datasets distribution

Train	Test	Validation
6295	778	700

2.2. Proposed method

This study uses a different approach by using feature extraction and classification with different pre-trained models. In the feature extraction stage, the IndoBERT pre-trained model, while the classification stage uses multilingual BERT (MBERT). The focus of this research is on multi-label datasets in Indonesian. Figure 1 contains the flow of the process carried out in this study.

The process carried out in this study uses a pre-trained model, which is a development of BERT, namely IndoBERT and MBERT. Both models are models that have been trained using Indonesian. IndoBERT is a model specifically trained to use the Indonesian language with a dataset called INDOLEM [19]. Meanwhile, MBERT is a multilingual model of BERT which is trained in many languages, one of which is Indonesian [20]. Both models can be used as a feature extraction or classification model or simultaneously as a feature extraction and classification model. The use of pre-trained models for feature extraction and classification is expected to help models better analyze and study contextual relationships between words. So, the model can better determine the pattern for classifying comments. The evaluation stage is carried out by calculating the model classification in the form of an F1 score derived from the calculation of true positive, true negative, false negative, and false positive [21].

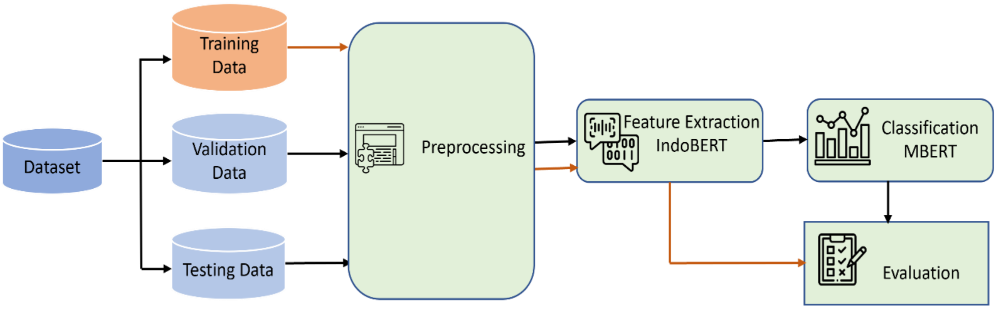


Figure 1. Proposed method stages

2.3. Preprocessing

Preprocessing is the initial stage for carrying out the text classification process. This stage aims to clean data that has a lot of noise. Processed data is usually unstructured and contains many words or characters that are repeated and unnecessary in the classification process [22]. The preprocessing stage consists of several steps, as shown in Figure 2.

- Noise removal: to remove characters such as numbers, punctuation marks, or excess spaces.
- Case folding: to change capital letters to non-capital letters.
- Translated emoticons: to translate emoticons into words.
- Tokenizing, to change sentences word for word.
- Translated slang words to change non-standard words into standard words according to Indonesian spelling.
- Stemming: to remove affix words, so the resulting words are basic words.
- Stopword removal: to remove words that frequently appear in sentences and do not have specific meanings.

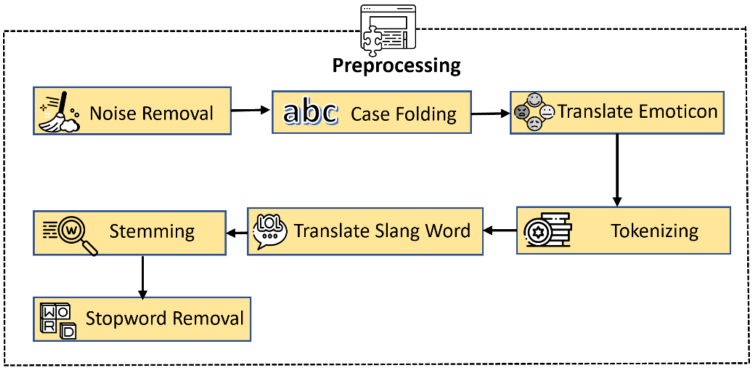


Figure 2. Preprocessing stages

2.4. IndoBERT embedding

Word embedding is the process of converting words into vectors. Traditional word embedding using the bag of words or term frequency-inverse document frequency (TF-IDF) methods. In addition, word embedding can also be done using static and contextual word embedding. Word2vec, glove, or short text for the static embedding method. Contextual embedding uses pre-trained models such as BERT, IndoBERT, and Elmo [23]. IndoBERT is a BERT model trained using masked language modeling with an Indonesian language dataset [24]. IndoBERT has 12 hidden layers and has been trained with words from the Indonesian Wikipedia (74 million), news articles from Kompas, Tempo, and Liputan6 (55 million), as well as the

Indonesian web corpus (90 million). In the embedding process, the first step is to add special tokens at the beginning and end of the sentence, namely the tokens [CLS] and [SEP]. These special tokens are used to separate sentences that are prefixes and suffixes. Then the tokenization process is carried out, which converts sentences into words or tokens. The tokenization process uses word piece tokenization method. This method uses vocabulary owned by the model and vocabulary outside the model.

Each pre-trained model has a fixed dimension in vector representation which is 768. Then the word that has been tokenized will be mapped with the vocabulary contained in the corpus. Figure 3 contains the flow of the embedding process using IndoBERT. The sentence will be converted into a representation of 12 tokens, each comprising 768 embedding tokens. Before the feature extraction process, segment embedding and positional embedding processes are carried out to make the model contextually understand the meaning. In segment embedding, the vector representation that occurs is only index 0 and index 1. If all input consists of only one sentence, then the embedded segment is index zero. In addition, positional embedding applies a lookup table of size (n, 768) where n represents the number of long sentences. The first row is a vector representation of each word in the first position. The second row represents each word in the second position, and so on. The combination of the three embedding processes is called input embedding, a solution to make the pre-trained model adaptable to NLP tasks.

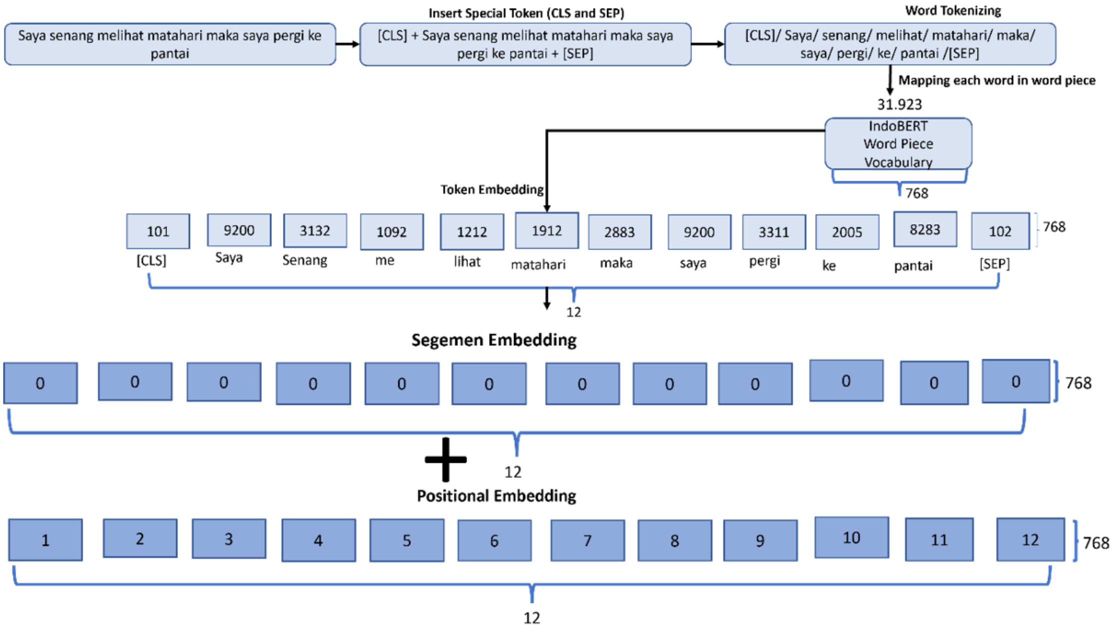

Saya senang melihat matahari maka saya pergi ke pantai
Insert Special Token (CLS and SEP)
[CLS] + Saya senang melihat matahari maka saya pergi ke pantai + [SEP]
Word Tokenizing
[CLS]/ Saya/ senang/ melihat/ matahari/ maka/ saya/ pergi/ ke/ pantai /[SEP]
Mapping each word in word piece
31.923
IndoBERT Word Piece Vocabulary
768
Token Embedding
101 [CLS]
9200 Saya
3132 Senang
1092 me
1212 lihat
1912 matahari
2883 maka
9200 saya
3311 pergi
2005 ke
8283 pantai
102 [SEP]
768
12
Segemen Embedding
0 0 0 0 0 0 0 0 0 0 0 0
768
12
Positional Embedding
1 2 3 4 5 6 7 8 9 10 11 12
768
12

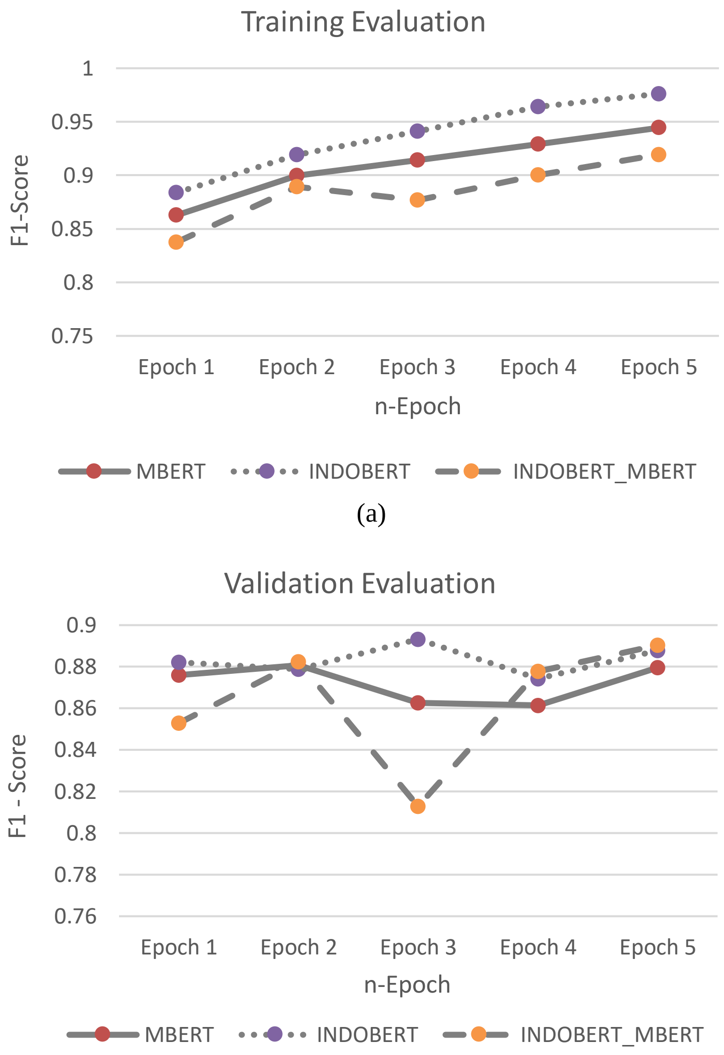
Figure 3. IndoBERT embedding process

2.5. MBERT classification

Multilingual BERT (MBERT) is the development of the BERT single language model, which has been trained using a monolingual corpus in 104 languages. MBERT is refined using specific training data from one language, and an evaluation process is carried out in a different language, making it possible to use it across languages, and even MBERT can carry out cross-language generalization tasks well [25]. The MBERT model has also been trained using the Indonesian language so that it is possible to use it for tasks in Indonesian.

3. RESULT AND DISCUSSION

The experiment was carried out using the PyTorch library with the Python programming language. Because the model is pre-trained, the experimental process requires to use the Google Colab Pro platform to meet large resource requirements and memory allocations. The number of epochs used in this study was 5, the batch size was 16, and the learning rate was 5e-5. All models trained in this study use the same number of an epoch, batch sizes, and learning rates. The model classification results are shown in Figure 4, which consists of two different types of model classification results. Figure 4(a) contains the F1-score of the training data and Figure 4(b) contains the F1-score of the validation data.



(a)

(b)

Figure 4. Comparing model classification results on (a) training data (b) validation

Experiments on training data show that the entire training model can continue to learn at each epoch so that the resulting classification results continue to increase, even the IndoBERT model tends to have a high F1 value of 0.976. Meanwhile, if tested using validation data, the resulting F1 value tends to experience an increase and decrease, but in a range of values that are not too high. In addition, all models also tend to have an F1 value below 0.90, where IndoBERT_MBERT has the highest F1 value of 0.89. From the training and validation processes, the entire model tends to have a stable F1 value so that the possibility of overfitting can be avoided. The final results of the three models are shown in Table 3 which contains the results of model evaluation using data testing.

Table 3. Evaluation results on testing data

Architecture	Training Accuracy	Validation Accuracy	Testing Accuracy
Multilingual BERT	0.94432	0.87964	0.89492
IndoBERT	0.97613	0.88785	0.89717
Proposed Model (IndoBERT_MultilingualBERT)	**0.91930**	**0.89035**	**0.90327**

From the results of the classification model on data testing, the proposed model in this study, namely by using a combination of feature extraction from IndoBERT and the MBERT classification model, obtained optimal results with an F1 value of 0.9032. The proposed model also tends to have a stable F1 value on data testing, validation, and training. From the resulting values, the proposed model is more stable and does not experience overfitting because the F1 value between the training and testing data has a very small difference of around 0.01. Whereas in the MBERT and IndoBERT models, the difference in the value of f1 generated on the testing and training data is 0.04 and 0.07, so the model tends to experience mild overfitting. The proposed model in this study also has a better f1 value than the previous study conducted by [10], with an F1 value of 0.889 on data testing. This study also uses the same dataset as the research conducted by [10].

4. CONCLUSION

Based on the experiments that have been carried out, the proposed model in this study can improve the model's ability to classify toxic comments from social media users in Indonesia. It is possible to use a combination of pre-trained models that have been trained in Indonesian, such as IndoBERT for feature extraction and MBERT for classification models. The proposed model also tends to have a stable F1 value and does not experience overfitting, compared to using one pre-trained model in the feature extraction and classification models. For further exploration, further research can apply a combination of pre-trained models such as Indo Electra and Indo Roberta as feature extraction with machine learning models such as k-nearest neighbors (KNN) and support vector machine (SVM) or deep learning models such as CNN and LSTM.

REFERENCES

[1] L. Wu, F. Morstatter, K. M. Carley, and H. Liu, “Misinformation in social media,” *ACM SIGKDD Explorations Newsletter*, vol. 21, no. 2, pp. 80–90, Nov. 2019, doi: 10.1145/3373464.3373475.

[2] Y. A. Ahmed, M. N. Ahmad, N. Ahmad, and N. H. Zakaria, “Social media for knowledge-sharing: a systematic literature review,” *Telematics and Informatics*, vol. 37, pp. 72–112, Apr. 2019, doi: 10.1016/j.tele.2018.01.015.

[3] D. W. Otter, J. R. Medina, and J. K. Kalita, “A survey of the usages of deep learning for natural language processing,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 604–624, Feb. 2021, doi: 10.1109/TNNLS.2020.2979670.

[4] M. Mansoor, K. Gurumurthy, A. R. U, and V. R. B. Prasad, “Global sentiment analysis of COVID-19 tweets over time,” *arxiv.org/abs/2010.14234*, Oct. 2020.

[5] S. Rani and P. Kumar, “Deep learning based sentiment analysis using convolution neural network,” *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3305–3314, Aug. 2019, doi: 10.1007/s13369-018-3500-z.

[6] B. van Aken, J. Risch, R. Krestel, and A. Löser, “Challenges for toxic comment classification: An in-depth error analysis,” in *2nd Workshop on Abusive Language Online - Proceedings of the Workshop, co-located with EMNLP 2018*, 2018, pp. 33–42, doi: 10.18653/v1/w18-5105.

[7] X. P. Qiu, T. X. Sun, Y. G. Xu, Y. F. Shao, N. Dai, and X. J. Huang, “Pre-trained models for natural language processing: a survey,” *Science China Technological Sciences*, vol. 63, no. 10, pp. 1872–1897, Sep. 2020, doi: 10.1007/s11431-020-1647-3.

[8] 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*, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[9] A. G. D’Sa, I. Illina, and D. Fohr, “BERT and FastText embeddings for automatic detection of toxic speech,” *2020 International Multi-Conference on: “Organization of Knowledge and Advanced Technologies” (OCTA)*, Tunis, Tunisia, 2020, pp. 1-5, doi: 10.1109/OCTA49274.2020.9151853.

[10] G. Z. Nabiilah, S. Y. Prasetyo, Z. N. Izdihar, and A. S. Girsang, “BERT base model for toxic comment analysis on Indonesian social media,” *Procedia Computer Science*, vol. 216, pp. 714–721, 2022, doi: 10.1016/j.procs.2022.12.188.

[11] N. Lashkarashvili and M. Tsintsadze, “Toxicity detection in online Georgian discussions,” *International Journal of Information Management Data Insights*, vol. 2, no. 1, Art. no. 100062, Apr. 2022, doi: 10.1016/j.jjimei.2022.100062.

[12] Z. Zhao, Z. Zhang, and F. Hopfgartner, “A comparative study of using pre-trained language models for toxic comment classification,” in *The Web Conference 2021 - Companion of the World Wide Web Conference, WWW 2021*, Apr. 2021, pp. 500–507, doi: 10.1145/3442442.3452313.

[13] A. N. Azhar and M. L. Khodra, “Fine-tuning pretrained multilingual BERT model for Indonesian aspect-based sentiment analysis,” *2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA)*, Tokoname, Japan, 2020, pp. 1-6, doi: 10.1109/ICAICTA49861.2020.9428882.

[14] J. A. Leite, D. F. Silva, K. Bontcheva, and C. Scarton, “Toxic language detection in social media for Brazilian Portuguese: new dataset and multilingual analysis,” *arxiv.org/abs/2010.04543*, Oct. 2020.

[15] P. Guillaume, C. Duchêne, and R. Dehak, “Hate speech and toxic comment detection using transformers,” *EPITA Speech and Language Recognition Group (ESLR)*, 2022.

[16] G. D. Saraiva, R. T. Anchiêta, F. A. R. Neto, and R. S. Moura, “A semi-supervised approach to detect toxic comments,” in *International Conference Recent Advances in Natural Language Processing, RANLP*, 2021, pp. 1261–1267, doi: 10.26615/978-954-452-072-4_142.

[17] L. Khan, A. Amjad, N. Ashraf, and H. T. Chang, “Multi-class sentiment analysis of Urdu text using multilingual BERT,” *Scientific Reports*, vol. 12, no. 1, Mar. 2022, doi: 10.1038/s41598-022-09381-9.

[18] A. Izzan, C. Wibisono, and I. F. Putra, “Indonesian social media post toxicity classification,” *GitHub*, 2018. https://github.com/ahmadizzan/netifier (accessed Feb. 21, 2023).

[19] 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.

[20] T. Pires, E. Schlinger, and D. Garrette, “How multilingual is multilingual BERT?,” in *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 2020, pp. 4996–5001, doi: 10.18653/v1/p19-1493.

[21] R. Yacouby and D. Axman, “Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models,” in *Proceedings of the First Workshop on Evaluation and Comparison of {NLP} Systems*, 2020, pp. 79–91, doi: 10.18653/v1/2020.eval4nlp-1.9.

[22] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, “Effective text data preprocessing technique for sentiment analysis in social media data,” Oct. 2019. doi: 10.1109/KSE.2019.8919368.

[23] E. T. Luthfi, Z. I. M. Yusoh, and B. M. Aboobaider, “Enhancing the Takhrij Al-Hadith based on contextual similarity using BERT embeddings,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 286–293, 2021, doi: 10.14569/IJACSA.2021.0121133.

[24] S. Saadah, K. M. Auditama, A. A. Fattahila, F. I. Amorokhman, A. Aditsania, and A. A. Rohmawati, “Implementation of BERT, IndoBERT, and CNN-LSTM in classifying public opinion about COVID-19 vaccine in Indonesia,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 648–655, Aug. 2022, doi: 10.29207/resti.v6i4.4215.

[25] S. Wu and M. Dredze, “Are all languages created equal in multilingual BERT?,” in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 120–130. doi: 10.18653/v1/2020.repl4nlp-1.16.

BIOGRAPHIES OF AUTHORS

Ghinaa Zain Nabiilah is a lecturer from Bina Nusantara University (BINUS). She graduated from BINUS Computer Science Department in 2023. Since 2020, her research has been related to natural language processing, especially in analyzing human personality and emotions. In addition, she is also active in research on the management and investigation of toxic sentences, hoaxes, and hate speech for decision support. She can be contacted in email: ghinaa.nabiilah@binus.ac.id.

Islam Nur Alam is a lecturer from Bina Nusantara University (BINUS). He has 2-year experience data science researcher with proven success in building successful algorithms and predictive models for image classification with convolutional neural network, highly adept clustering and classification, content-based filtering, data analysis and visualization and he currently training data science skills all about natural language processing and machine translation task. He can be contacted in email: islam.alam@binus.ac.id.

Eko Setyo Purwanto is a lecturer from BINUS University who has experience teaching and guiding students in developing research in computer science. He has conducted research focusing on the internet of things. His study aims to understand and address challenges related to the internet of things. He has also published several articles and scientific publications focusing on the internet of things, augmented reality, and UI/UX in leading journals in computer science. Currently, Eko is trying to explore research in natural language processing. With his commitment to computer science research and development, Eko strives to contribute to advancing knowledge and technology in this field. He believes that the combination of theoretical study and practical applications can have a significant positive impact on the development of society and industry in this digital era. He can be contacted in email: eko.purwanto@binus.ac.id.

Muhammad Fadlan Hidayat is a lecturer at Bina Nusantara University (BINUS). Finished his Master Degree from Bina Nusantara University majoring computer science in 2022 with research topic deep learning in speech recognition. Fadlan has research interest in deep learning especially speech recognition and natural language processing and try to contribute to advancing technology in computer science field. He can be contacted in email: muhammad.hidayat003@binus.ac.id.