IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 4, December 2024, pp. 4688∼4694
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp4688-4694

Indonesian news article authorship attribution multilabel multiclass classification using IndoBERT

Karen Etania Saputra, Riccosan
Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Dec 6, 2023
Revised Mar 14, 2024
Accepted Apr 17, 2024

Keywords:

Authorship attribution
Deep learning
IndoBERT
Indonesian language
Multiclass classification
Multilabel classification

ABSTRACT

Recent developments in technology have made it easier to produce digital content, especially textual articles. But, it has a negative impact in the form of a rising public skepticism of digital data due to plagiarism. Indonesia, one of the world's most populous countries, is not resistant to this problem. To resolve it, the authorship attribution (AA) task must be executed. However, there has been little investigation on AA for Indonesian articles. As a result, this research applies the AA task to an Indonesian digital news articles dataset. Continuing the previous research, dataset modification was carried out to increase data complexity by adding a new class, namely the author's gender, and also by balancing the distribution of data versus labels to minimize potential overfitting, and model hyper-parameter configurations were carried out to enhance the results gained. This research successfully applied the IndoBERT model to the Indonesian AA task, yielding results in the form of precision = 0.92, recall = 0.90, and F1-score = 0.91. These results indicate that the Indonesian AA task has a lot of potential for development since it identifies writing patterns that may benefit the forensic field, detect plagiarism, and analyze Indonesian texts.

Corresponding Author:

Karen Etania Saputra
Department of Computer Science, School of Computer Science, Bina Nusantara University
Bandung Campus, Jakarta 11480, Indonesia
Email: karen.sapurta@binus.ac.id

1. INTRODUCTION

Technological advancements have made it easier to generate content, particularly articles and information, with a rising volume and faster transmission via the internet as a channel with a wide range of topics [1], minimizing the need for physical forms. This causes the issue of a lack of trust from the public in the substance of digital segments, as opposed to traditional news and journalism, which has tight selection and regulation processes [2]. The problem exists because of plagiarism, an act of duplicating an existing idea, method, results, or some data without displaying the original author and its source [3]. Until now, plagiarism has become a root problem in many countries as technology advances and data gets easier to be accessed. Indonesia, as one of the most populated countries, also face this problem and resulting in the creation of many digital articles with unknown authors and the origins, thus having the negative impact of plagiarism, such as deception in searching flow to get genuine information and spreading misinformation to the public [4].

To answer the problems, article classification task using author is critical to implement, especially in Indonesian articles. Article classification is a part of the text analysis task, namely authorship attribution (AA), which has grown into an interesting research topic that has drawn a lot of different sectors, including media, government, economics, and education [5]. AA is a technique for identifying stylometry and linguistic patterns in documents written by humans [6]. AA is part of the text analysis task since it seeks to process vast volumes of textual data to detect plagiarism and find the original article's author. The AA task is supported by the author's personality traits in the form of distinct and unique linguistics in each article as the main data, which is an important factor in determining the originality and source of an article [7]. The AA task in this research is inseparable from its benefits of dynamic interdisciplinary integration in the modern digital world, the ability to obtain important and in-depth information, and its wide contribution to various fields (forensic investigations, digital content analysis, and plagiarism detection) [8].

There are barriers to the AA implementations, including anonymized documents, article lengths, articles with numerous authors [8], the need for an excessive quantity of data [9], articles used by other parties that extend the search flow [10], and the quality of author writing, which continues to improve [11]. These barriers make tracing the origin of digital articles a complicated and time-consuming task. To answer these issues, it is possible to increase the quality of the dataset used in the research by increasing or lowering the amount of data, balancing data distribution, and determining the relevant literary components for the classification process. This research employs a dataset of digital articles from news portal websites from earlier research on automatic Indonesian AA using transformers [12]. For greater complexity, the dataset was modified by including a second feature or class, author gender. Gender elements are included due to the difference in language styles used by men and women, with male authors using more formal language forms and female authors combining the formal, informal, and more familiar figurative language forms [13].

After modifying the dataset, this research implement multi-label multi-class methods for classification. Multi-label classification is the classification of data that has more than one sort of class designated for it [14], which in this case are the author's name and gender. Multi-class classification is the classification of data utilizing more than one class [15], with every point of data only corresponding to a single class to avoid overlapping. Another modification to the dataset was a reduction in author's name class to balance the distribution of data against the total classes. This research achieved a higher level of data complexity by modifying the dataset, allowing for improved outcomes when implementing AA tasks on Indonesian news articles.

In this research, AA tasks for authorship classification of Indonesian digital news items are implemented using the deep learning approach, as in earlier research [12]. Several prior research utilized deep learning methods for AA tasks and proven their effectiveness. i) the implementation of AA with convolutional neural network (CNN) and POS-Eliding [16] effectively improved the performance of the AA task, resulting in better generalization of textual information of research articles. ii) AA for neural text generation [17] investigated the AA problem with several deep learning models and showed the urgency that machine-generated texts are ever more difficult to detect by humans because their better quality creates the potential for fraud in the form of misleading information generation, demanding the development of AA technologies. iii) the development of the bidirectional encoder representation from transformers for authorship attribution (BertAA) model by Fabien *et al.* [18] utilized the BERT base model [19] and fine-tuned for AA tasks with data in the form of Enron emails, blog posts, and movie reviews, achieving state-of-the-art (SOTA) on model accuracy of 93%. However, most earlier research on AA with deep learning approaches made use of English datasets, hence there is a shortage of local research on AA tasks with Indonesian datasets. As a result, the purpose of this research is to contribute to the field of AA tasks by applying the deep learning method to Indonesian digital articles in particular.

This research utilizes the transformers architecture [20], with Indonesian bidirectional encoder representation from transformers (IndoBERT) [21] as the base model. IndoBERT is a specialized model for natural language processing (NLP) tasks based on the bidirectional encoder representation from transformers (BERT) model [19]. Implementing the IndoBERT model [21] cannot be separated from its ability to focus on Indonesian datasets. This is effectively understood by an in-depth review of previous research that confirms IndoBERT's advantages for specific Indonesian-language datasets [12], [22], [23]. As an outcome, the use of IndoBERT will benefit this research employing an Indonesian news article dataset. Etania *et al.* [12] achieved model evaluation findings in the form of test predict accuracy of 0.70 (70%), and Top-K accuracy of 0.84 (84%). Although exceeding the accuracy values obtained, the model remains short of the BertAA model, which has an accuracy score of 93% [18]. The previous research's model failed to achieve its objective due to an overfit model, which happens when the amount of data is not sufficient for the number of classes. On that basis, this research reapplied the IndoBERT model [21]. As a result, changes were made to the dataset used in this research, which included adding a feature and balancing the data distribution compared to the classes.

This research experiment contributes by showing that AA tasks with deep learning methods for

Indonesian-language articles yield better and more significant results; the implementation of AA tasks can be focused for Indonesian-language digital news articles; and the application of multi-label multi-class methods in AA tasks for Indonesian-language news articles enhances the performance of the IndoBERT model. The research segment is separated into four parts. Section 1, introduction: contains the problem statement, an overview of earlier research, the source of the dataset utilized, and the research's objectives. Section 2, method: covers the research flow used in this article. Section 3, result and discussion: provides an in-depth explanation of the analysis of the study experiment results. Section 4, conclusion: presents the research's conclusions along with recommendations for further research.

2. METHOD

Figure 1 illustrates the research process. Continuing the previous research [12], the dataset was obtained in the data scraping stage, in the form of digital news articles in the Indonesian language retrieved from the author's site on the news website. Once the gathering process is completed, the data is stored in a format that fits this research. In this research, a class category, notably the author's gender, was added to increase the data's complexity as the second label. After data collection, the second stage is data pre-processing. This stage covers sub-processes for cleaning the collected data, such as the removal of special symbols, URLs or hyperlinks, and irrelevant information like picture captions, advertisements, or comments. After the data has been properly cleaned, it is then shuffled to a random representation and avoid bias.

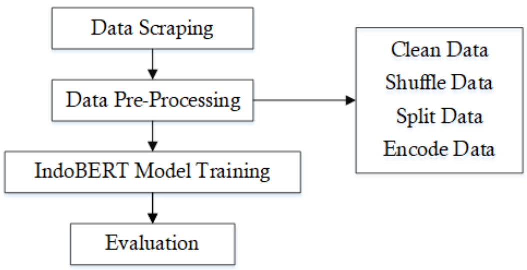


Figure 1. Research flow

Additionally, the utilized dataset is separated into 3 parts: training, valid, and test data. The first one is training data for the model training phase; the second is valid data used for model parameter tuning; the last is test data used to evaluate the utilized model's performance. Because IndoBERT can read more than one pattern from the given data, data splitting is employed to generate more diverse training data. Out of the total 2000 collected data, the primary dataset is divided into 1300 training data, 455 validation data, and 245 test data. Furthermore, each of the content and label columns are merged to organize the data in a suitable form for training the multilabel multiclass model. The final step involves encoding the data using the one-hot encoding method. One-hot encoding is used to transform the text into binary numeric vectors that can be processed by the model [24], [25]. Each unique word in the dataset is mapped to a binary vector of the same size, enabling the model to understand and process the data in a numeric format. With the help of this data preprocessing pipeline, the model will be able to learn and accompanying labels to predict the author and gender with accuracy.

For the training phase, the IndoBERT model was used, which has proven to be effective in processing the Indonesian language. Before beginning training, the data is tokenized with the Bert tokenizer, which separates the text into fragments such as words or sub-words. The tokenization steps prepare the data to ensure the model can utilize it. The hyperparameters that employed in the training process are: 10 epochs, 8 batch size, 3e-5 learning rate, 0.01 weight decay, and 100 logging steps. After the training process, the model is evaluated using classification report in the form of accuracy and F1-score. The F1-score calculate using precision and recall values to offer a comprehensive evaluation of the model's performance. Accuracy measures the percentage of model's accurate predictions based on total predictions, meanwhile the F1-score measures the percentage of model's correct predictions based on total predictions. Following these steps, this research could provide an effective and precise multi-label multi-class classification model for Indonesian news items, allowing for improved information structure and analysis in the context of Indonesian news article.

3. RESULTS AND DISCUSSION

The dataset for this research is from earlier research [12], that contained 2,000 Indonesian digital news articles in the form of comma separated value (CSV) files. For each author in the dataset contributed 100 unique articles to gain a more wide range of writing styles and patterns. The number of authors in the dataset has been reduced from 80 in the previous research [12] to 20 authors in this research to minimize the potential of model overfitting. To increase the complexity of the data used in this research, a new class was added, namely author gender, to make the research multi-labeled. Due to its multi-label approach, the utilized model can determine relevant authors with high accuracy, even when assigning many different author classes to each digital news article. In this research, the model's hyper-parameters have been adapted at learning rate (LR) = 3e-5, batch size = 8, weight decay = 0.01, logging steps = 100, and training epoch = 10, to improve the model's performance in minimizing overfitting and obtaining more optimal results than previous research. For the training stage, the sigmoid activation function was used for the multi-label classification task.

The model's training results yield impressive performance metrics as seen in Table 1. The results are in the form of training accuracy = 0.8879, the F1-score = 0.9360, the training loss =0.0187, and the validation loss = 0.0449. These evaluation metrics indicate the model's robustness and effectiveness in classifying the news articles and attributing them to their respective authors. The model demonstrates high accuracy and F1-score, implying that the model utilized can correctly identify authors' distinct writing styles and characteristics. Additionally, a comparison of model training results, shown in Table 2, indicates this research model succeeded in overcoming the overfitting problem in previous research [12]. This can be seen through the validation loss value which is decreased to 0.0449 and the training accuracy which is slightly increased to 0.8879. These results were supported by modifications to the dataset utilized and adding additional training epochs.

Table 1. Model training result

Epoch	Training loss	Validation loss	F1 score	Training accuracy
1 - 5	...	...	...	...
6	0.0318	0.0502	0.9319	0.8747
7	0.0248	0.0461	0.9354	0.8901
8	0.0215	0.0449	0.9371	0.8923
9	0.0198	0.0449	0.9360	0.8879
10	0.0187	0.0449	0.9360	0.8879

Table 2. Model training result comparison

	Total training epoch	Training loss	Validation loss	Training accuracy
Earlier research [12]	8	0.062	0.992	0.84
This research	10	0.0187	0.0449	0.8879

After the completion of the training process, an evaluation is conducted to assess the model's performance in the form of classification report that provides valuable insights into key metrics, including precision, recall, and F1-score, which play significant roles in gauging the model's effectiveness in classifying news articles accurately. Precision measures the proportion of true positive predictions (correctly predicted positive cases) out of all positive forecasts made by the model. Recall represents the proportion of positive instances in the dataset that the classifier correctly identified among all the actual positive cases. It showcases the model's capability to capture all relevant instances belonging to a specific class. The F1-score serves as an overall measure of a model's performance, balancing both the precision and recall scores. F1-score combines these two metrics to provide a comprehensive evaluation of the model's accuracy in predicting both positive and negative cases. The formula that serves as the foundation for the evaluation calculation is calculated as follows.

$$precision = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1score = \frac{TP}{\frac{1}{2} + (FP + FN)} \tag{3}$$

When evaluating the model in the context of AA, the assessment measures indicated in the previous formulations are crucial. The counts of true positive (TP), false positive (FP), and false negative (FN) instances are used to calculate these metrics. Where TP denotes the number of accurate positive predictions made by the model, FP denotes the number of inaccurate positive predictions, and FN denotes the number of positive instances that were mistakenly classified as negative. Table 3 presents a detailed breakdown of the evaluation findings for each label (article's author), revealing distinct performance levels across different authors. Notably, Diva Lufiana Putri exhibits the lowest evaluation values among all labels, with a precision of 0.50, recall of 0.40, and F1 score of 0.44. These results suggest that the author's language choices might be less diverse, leading to comparatively less clarity in the AA task. On the other hand, author Abdul Haris Maulana and author Jawahir Gustav Rizal receive the highest ratings for their labels, achieving precision, recall, and F1-scores of 1. These exceptional scores indicate a flawless performance in the classification and attribution of articles to those authors.

Table 3. Classification report result

Label	Precision	Recall	F1 Score
Aisyah Sekar Ayu Maharani	0.80	0.92	0.86
Diva Lufiana Putri	0.50	0.40	0.44
Fika Nurul Ulya	0.75	0.75	0.75
Kiki Safitri	0.69	0.82	0.75
Lely Maulida	1.00	1.00	1.00
Nadia Faradiba	0.79	1.00	0.88
Nur Fitriatus Shalihah	0.83	0.59	0.69
Rintan Puspita Sari	0.94	1.00	0.97
Silmi Nurul Utami	0.88	0.70	0.78
Vanya Karunia Mulia Putri	1.00	0.86	0.92
Abdul Haris Maulana	1.00	1.00	1.00
Jawahir Gustav Rizal	1.00	1.00	1.00
Labib Zamani	0.93	1.00	0.96
M Adika Faris Ihsan	1.00	1.00	1.00
Muhammad Idris	1.00	0.91	0.95
Nicholas Ryan Aditya	1.00	0.86	0.92
Nur Jamal Shaid	0.90	1.00	0.95
Reza Agustian	1.00	0.85	0.92
Rully R Ramli	1.00	0.57	0.73
Reza Kurnia Darmawan	0.93	0.88	0.90
L	0.97	0.93	0.95
P	0.94	0.97	0.95
Micro average	0.92	0.90	0.91

The differences in assessment metrics across authors underline each author's unique writing style and language. While linguistic patterns may be more difficult to discover in certain authors, resulting in lower evaluation scores, for others, the model may be able to generate accurate predictions with high precision and recall. This performance variance shows the difficulties of AA task, which requires a deep examination of each author's unique linguistic preferences and writing characteristics. The analysis of gender labels shows remarkable gender classification skills, with an average accuracy, recall, and F1-score of 0.95. This high accuracy in identifying gender is crucial as it shows the model's ability to differentiate gender-based writing patterns in news articles. Table 3 shows the results of model evaluation using a classification report, containing precision, recall, and F1-score values, as well as the average of each category calculated with the micro-averaging approach. The model's average precision score of 0.92 indicates that the model accurately detects and labels the data that belong to a certain category. The model's average recall score of 0.90 indicates that the model effectively retrieved events from different classes. The classification report in Tabel 3 shows a micro-average F1-score of 0.91, showing a balanced performance in terms of recall and precision.

These noteworthy micro-average results indicate the model utilized, namely IndoBERT, has performed well in classifying news articles and assigning them to their authors, especially in this research with Indonesian digital news article dataset [12]. The model's ability to accurately project positive occurrences while limiting FP and FN is shown through the model's high precision, recall, and F1-score, proving its robustness and dependability in the Indonesian AA task. Table 3 about evaluation details not only provides valuable insight into the model's performance but also acts as a critical reference point for further research and study in the field of Indonesian AA. The metrics used in this research can be implemented by other researchers to evaluate the performance of different models, find areas for advancement, and make data-driven decisions that optimize the model's performance.

4. CONCLUSION

This research successfully overcomes the problems faced in earlier research by modifying the dataset utilized in two ways: i) adding a new class in the form of the author's gender and ii) reducing the number of author labels, which equalizes the amount of data versus class. In addition, the result of the research is tightly associated with the proper model's hyper-parameters tuning. The evaluation results demonstrates the model's success in classifying Indonesian digital news articles and identifying their authors, with outstanding performance, the precision 0.92, recall 0.90, and F1-score 0.91. These results will act as a benchmark for upcoming research and development in Indonesian AA as the field continues to evolve. However, despite its excellent results, the model's accuracy is still below. This research still has potential for more advancement, for example utilizing Indonesian datasets with higher complexity and larger amounts and implementing better model parameters configuration to obtain more optimal results. Researchers can use the implemented metrics in this research to compare models, enhance performance by making data-driven decisions, or identify areas for improvements in Indonesian AA.

REFERENCES

[1] H. S. A. -Ash, M. F. Putri, P. Mursanto, and A. Bustamam, "Ensemble learning approach on Indonesian fake news classification," in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, 2019, pp. 1–6, doi: 10.1109/ICICoS48119.2019.8982409.

[2] R. Baly *et al.*, "What was written vs. who read it: news media profiling using text analysis and social media context," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 3364–3374, doi: 10.18653/v1/2020.acl-main.308.

[3] P. Abbasi, J. Yoosefi-Lebni, A. Jalali, A. Ziapour, and P. Nouri, "Causes of the plagiarism: a grounded theory study," *Nursing Ethics*, vol. 28, no. 2, pp. 282–296, 2021, doi: 10.1177/0969733020945753.

[4] T. Foltýnek, N. Meuschke, and B. Gipp, "Academic plagiarism detection: a systematic literature review," *ACM Computing Surveys*, vol. 52, no. 6, pp. 1–42, 2020, doi: 10.1145/3345317.

[5] P. Barberá, A. E. Boydstun, S. Linn, R. McMahon, and J. Nagler, "Automated text classification of news articles: a practical guide," *Political Analysis*, vol. 29, no. 1, pp. 19–42, 2021, doi: 10.1017/pan.2020.8.

[6] T. Schuster, R. Schuster, D. J. Shah, and R. Barzilay, "The limitations of stylometry for detecting machine-generated fake news," *Computational Linguistics*, vol. 46, no. 2, pp. 499–510, 2020, doi: 10.1162/COLI_a_00380.

[7] D. D. Maharani, "A sociolinguistics analysis of language style in wild child," *Bachelor Thesis*, Department of English Education, Universitas Muhammadiyah Mataram, Mataram, Indonesia, 2019.

[8] F. Iqbal, M. Debbabi, and B. C. M. Fung, *Machine learning for authorship attribution and cyber forensics*. Cham: Springer, 2020, doi: 10.1007/978-3-030-61675-5.

[9] U. Mahor and A. Kumar, "A comparative study of stylometric characteristics in authorship attribution," in *Information and Communication Technology for Competitive Strategies (ICTCS 2021)*, Singapore: Springer, 2023, pp. 71–81, doi: 10.1007/978-981-19-0095-2_8.

[10] V. Kalgutkar, R. Kaur, H. Gonzalez, N. Stakhanova, and A. Matyukhina, "Code authorship attribution: Methods and challenges," *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–36, 2019, doi: 10.1145/3292577.

[11] P. Rizvi, "The problem of microattribution," *Digital Scholarship in the Humanities*, vol. 34, no. 3, pp. 606–615, 2018, doi: 10.1093/digitalsh/fqy066.

[12] S. K. Etania, Riccosan, and A. Chowanda, "Automatic Indonesian authorship attribution recognition using transformer," *ICIC Express Letters*, vol. 17, no. 5, pp. 497–503, 2023, doi: 10.24507/icicel.17.05.497.

[13] F. Togatorop and D. Y. Saragih, "The difference between male and female language style of students grade viii," *Journal on Education*, vol. 5, no. 3, pp. 9471–9479, 2023, doi: 10.31004/joe.v5i3.1762.

[14] M. L. Zhang, Y. K. Li, H. Yang, and X. Y. Liu, "Towards class-imbalance aware multi-label learning," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 4459–4471, 2022, doi: 10.1109/TCYB.2020.3027509.

[15] R. M. Mathew and R. Gunasundari, "A review on handling multiclass imbalanced data classification in education domain," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2021, pp. 752–755, doi: 10.1109/ICACITE51222.2021.9404626.

[16] J. Hitschler, E. V. D. Berg, and I. Rehbein, "Authorship attribution with convolutional neural networks and POS-eliding," in *Proceedings of the Workshop on Stylistic Variation*, 2017, pp. 53–58, doi: 10.18653/v1/W17-4907.

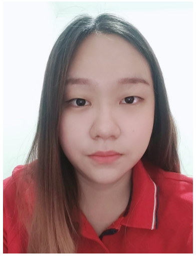
[17] A. Uchendu, T. Le, K. Shu, and D. Lee, "Authorship attribution for neural text generation," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 8384–8395, doi: 10.18653/v1/2020.emnlp-main.673.

[18] M. Fabien, E. Villatoro-Tello, P. Motlicek, and S. Parida, "BertAA: BERT fine-tuning for Authorship Attribution," in *Proceedings of the 17th International Conference on Natural Language Processing (ICON)*, 2020, pp. 127–137.

[19] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

Technologies, 2019, vol. 1, pp. 4171–4186.
[20] A. Vaswani *et al.*, "Attention is all you need," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017, pp. 1–15.
[21] 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 *28th International Conference on Computational Linguistics*, 2020, pp. 757–770, doi: 10.18653/v1/2020.coling-main.66.
[22] B. Juarto and Yulianto, "Indonesian news classification using IndoBert," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 454–460, 2023.
[23] D. U. K. Putri and D. N. Pratomo, "Clickbait detection of Indonesian news headlines using fine-tune bidirectional encoder representations from transformers (BERT)," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 7, no. 2, pp. 162-168, 2022, doi: 10.25139/inform.v7i2.4686.
[24] S. H. Kong, L. M. Tan, K. H. Gan, and N. H. Samsudin, "Fake news detection using deep learning," in *2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 2020, pp. 102–107, doi: 10.1109/ISCAIE47305.2020.9108841.
[25] M. K. Dahouda and I. Joe, "A deep-learned embedding technique for categorical features encoding," *IEEE Access*, vol. 9, pp. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.

BIOGRAPHIES OF AUTHORS



Karen Etania Saputra received her master's degree in Computer Science from Bina Nusantara University. She is a full-time lecturer in the computer science program at Bina Nusantara University's Bandung Campus. Her primary research focus lies in the field of deep learning. She can be contacted at email: karen.saputra@binus.ac.id.

Riccosan obtained his master's degree in Computer Science from Bina Nusantara University. He currently serves as a full-time lecturer in the computer science program at Bina Nusantara University's Bandung Campus. His primary research interest revolves around emotions and sentiment analysis, delving into the fascinating world of understanding human emotions through technology. He can be contacted at email: riccosan@binus.ac.id.