International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 13, No. 1, April 2024, pp. 42~49
ISSN: 2252-8776, DOI: 10.11591/ijict.v13i1.pp42-49

Traffic accident classification using IndoBERT

Muhammad Alwan Naufal, Abba Suganda Girsang
Department of Computer Science, BINUS Graduate Program, Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Dec 21, 2023
Revised Jan 19, 2024
Accepted Feb 26, 2024

Keywords:

Classification
IndoBERT
Machine learning
Traffic accident
Social media

ABSTRACT

Traffic accidents are a widespread concern globally, causing loss of life, injuries, and economic burdens. Efficiently classifying accident types is crucial for effective accident management and prevention. This study proposes a practical approach for traffic accident classification using IndoBERT, a language model specifically trained for Indonesian. The classification task involves sorting accidents into four classes: car accidents, motorcycle accidents, bus accidents, and others. The proposed model achieves a 94% accuracy in categorizing these accidents. To assess its performance, we compared IndoBERT with traditional methods, random forest (RF) and support vector machine (SVM), which achieved accuracy scores of 85% and 87%, respectively. The IndoBERT-based model demonstrates its effectiveness in handling the complexities of the Indonesian language, providing a useful tool for traffic accident classification and contributing to improved accident management and prevention strategies.

Corresponding Author:

Muhammad Alwan Naufal
Department of Computer Science, BINUS Graduate Program, Master of Computer Science
Bina Nusantara University
Jakarta, 11480, Indonesia
Email: muhammad.naufal@binus.ac.id

1. INTRODUCTION

Accidents on the highway are a serious issue with economic, social, and health impacts. Every year, there are approximately more than 1 million lives lost due to traffic accidents on the road. In addition, as many as 20 to 50 million other individuals suffer nonfatal injuries, with many of them experiencing disabilities as a result of the injuries they have sustained [1].

Studies that are specific to the discipline of microblogging services, especially Twitter, have shown a lot of improvement [2]. Twitter has become a major platform for people to share information and experiences [3], [4]. Twitter's user base has grown significantly, from 30 million users in 2010 to 330 million in 2019 [5]. This includes a wealth of accident-related information, as users regularly share news and personal accounts of accidents [6]. Social media enables the democratization of knowledge and information, turning users into content creators [7]. This vast amount of data, especially regarding traffic accidents, can be easily collected, analyzed, and categorized [8].

Natural language processing (NLP) is essential for analyzing accident-related tweets [9]. Accurate classification of accidents is crucial for effective management and prevention. Our study introduces a novel method for traffic accident classification, contributing significantly to this field.

BERT, a transformative language model, has advanced NLP with its bidirectional context processing [10]. IndoBERT, the Indonesian counterpart, adopts BERT's robust transformer-based approach to process Indonesian text. This forms the basis for our exploration of accident-related tweets, aiming to improve classification precision and efficiency and enhance accident management and prevention in Indonesia [11].

In text mining, a branch of data mining analyzing text documents [12], synergy with models like BERT and IndoBERT is evident. By applying text mining techniques, we extract information patterns from unstructured accident-related tweets, refining our classification approach and contributing insights to enhance accident management and prevention in Indonesia. Leveraging IndoBERT and sophisticated NLP techniques, our model excels in categorizing accidents into four types: car, motorcycle, bus, and others, achieving an impressive 94% accuracy. This outperforms conventional methods like random forest (RF) (85%) support vector machine (SVM) (87%), showcasing IndoBERT's adeptness in the Indonesian language and its potential for substantial contributions to safety strategies in Indonesia.

2. RESEARCH METHOD

The research methodology as shown in Figure 1 follows a systematic seven-step approach. Initially, a comprehensive collection of traffic accident reports and relevant information is carried out through an extensive survey to ensure dataset diversity. The collected data then undergoes a meticulous preprocessing phase to enhance its quality and suitability for analysis. Subsequently, the dataset is strategically divided into subsets for training, validation, and testing purposes, maintaining a balanced representation. To streamline the process, an auto-labeling mechanism is employed to categorize the dataset. The model is then trained using the labeled dataset, and its performance undergoes rigorous testing. Finally, an evaluation phase assesses the model's effectiveness in accurately classifying and predicting traffic accidents based on established criteria.

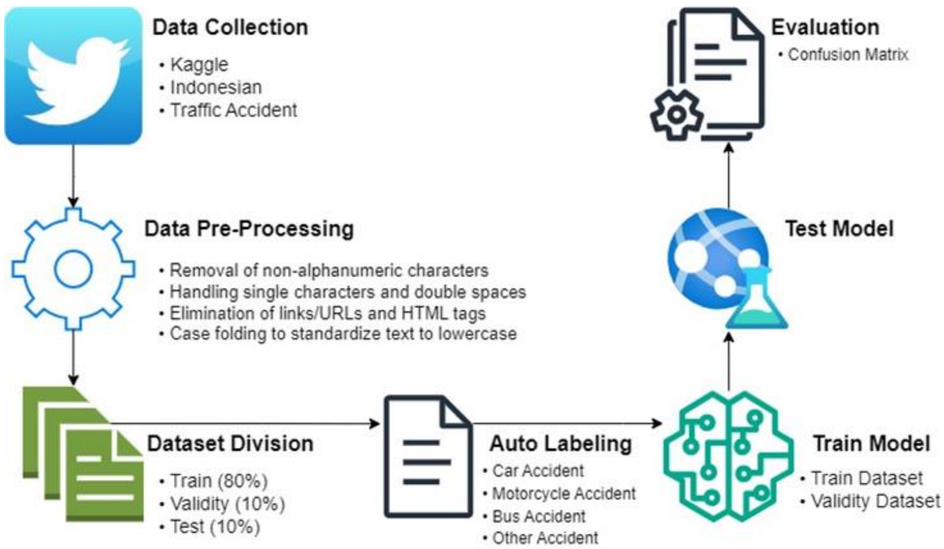

Figure 1. Method

2.1. Data collection

The dataset size is considered a critical property in determining the performance of a machine learning model [13]. Our research follows a pragmatic methodology for building a traffic accident tweet classification model. It commences with data collection from a Kaggle website titled “Traffic Accident in Indonesia” by Dody [14], with a specific focus on curating tweets related to traffic accidents. Subsequent data filtering streamlines the dataset by retaining only tweets directly associated with traffic accidents. In recognizing the significance of dataset size, we aim to empirically evaluate its impact on the classification performance of our machine learning model dedicated to traffic accident tweets.

2.2. Data pre-processing

Data preprocessing is an important step before applying machine learning methods [15]. In data preparation, we conduct essential data preprocessing tasks. These include the removal of non-alphanumeric characters to ensure the text's cleanliness and structure. We also address single characters and double spaces, enhancing the overall consistency of the data. Additionally, we eliminate links/URLs, HTML tags and stopword removal aims to remove words that have no meaning like general words with no or less meaning than other keywords [16], ensuring that the text is free from extraneous elements that could interfere with our analysis. Lastly, we perform case folding to standardize the text to lowercase, promoting uniformity and simplifying the subsequent stages [17] of our text processing pipeline.

2.3. Dataset division and auto labeling

In this study phase, we divide the dataset into three subsets: a training dataset (80% of data), a validation dataset (10%), and a testing dataset (10%). The training dataset helps the model learn and develop the ability to classify data accurately. Model accuracy may be objectively verified using the validation dataset, which provides information on genuine classification results [18]. The dataset for this procedure will be derived via data validation. The testing dataset evaluates the model's real-world performance and reliability. Labeling is necessary because the supervised method can see examples and produce generalizations so that the output will produce the desired label [19]. We label the data into four classes: car accidents, motorcycle accidents, bus accidents, and other accidents. Predefined keywords are used for auto-labeling in the training and validation datasets, resulting in a 91% accuracy rate for recognizing traffic accident-related tweets.

2.4. Model train

For model training, we harness IndoBERT, as a language model trained on Indonesian language data, offers several advantages in processing the Indonesian language. Through pretraining and fine-tuning, this model can learn patterns and structures in Indonesian, including word meanings, syntax, and word dependencies [20]. As a result, IndoBERT can analyze text more effectively, recognize word relationships, and comprehensively understand context. Subsequently, the model is subjected to testing using the dedicated testing dataset to evaluate its performance in categorizing traffic accident-related tweets.

2.5. Model test

In the model testing phase, we rigorously assess our IndoBERT-based classification model using a separate dataset, ensuring an unbiased evaluation of its real-world effectiveness. Emphasizing objectivity, this dataset remains distinct from the training and validation phases. Our focus is on evaluating the model's ability to accurately categorize traffic accident-related tweets and handle previously unseen data.

Using standard metrics like accuracy, precision, recall, and the F1 score, we comprehensively measure the model's performance. The results provide valuable insights into the model's practical utility and reliability in accident management and prevention. These findings contribute significantly to the overall evaluation detailed in section 3.7 Model Evaluation, employing a pragmatic and methodical approach to ensure a balanced and objective perspective, avoiding exaggerated claims.

3. RESULTS AND DISCUSSION

3.1. Data collection

The data collection process was initiated by accessing the Kaggle website, which is dedicated to traffic accident incidents. A total of 4,245 data points were meticulously gathered from the website, forming a substantial and comprehensive dataset. This dataset serves as a solid and extensive foundation for conducting in-depth analyses and research pertaining to various aspects of traffic accidents, enabling researchers and analysts to derive valuable insights and make informed decisions. Additionally, Figure 2 provides an illustrative example of the extracted data, offering a visual representation of the dataset in practice.

	id_str	full_text
0	1113812743138697216	Pelajar SMP Tewas Kecelakaan di Jalinsum Banda...
1	1113848173166977024	#TrukTangki\n\n#Kecelakaan\n#Evakuasi\n#NaganR...
2	1113859406666530817	Mobil Dinas Kepala BPN Prabumulih Mengalami Ke...
3	1113860178883964928	MASUK JALAN TOL GRATESSS EMG ENAK MULUSSHHH LI...
4	1113870948237758465	Laka Lantas di Tanah Hitam, SPM Beat Vs SPM Ka...
...	...	...
4240	1247149864376385537	19:58 Wib. Interchange Pluit arah Gedong Panja...
4241	1247152094034841601	5. Bis Hantu-Nusantara\n\nDari bus yang kecela...
4242	1247167265776103425	Kajati Sulut Andi Muh Iqbal Arief, SH. MH pimp...
4243	1247252793624158208	T'lah terjadi kembali pagi tadi kecelakaan men...
4244	1247324276325265408	.\n.\nIni orang tolol apa begok? 😟\n.\nKecelak...

Figure 2. Data collection result

3.2. Data preprocessing and dataset division

Data processing stands as a crucial element in the data analysis procedure, often demanding additional effort and time [21]. The data has undergone a series of transformations, including the removal of non-alphanumeric characters replaced with spaces, handling of single characters, elimination of double spaces, removal of links/URLs, removal of HTML tags, stopwords, and application of case folding. This ensures that the data is well-prepared for further analysis, making the preprocessing stage a crucial initial step in data processing [22]. The following as shown in Figure 3 is an example of data that has been preprocessed.

```
id_str                                         1113812743138697216
created_at                                     2019-04-04 21:37:06
screen_name                                           vtvindonesia
processed      pelajar smp tewas kecelakaan di jalinsum banda...
Name: 0, dtype: object
```

Figure 3. Example of data preprocessing result

3.3. Dataset division

The data is divided into three essential subsets: the training data, the validation data, and the test data. This division is crucial for machine learning and statistical analysis, where the training data is used to build and train the models, the validation data helps fine-tune the model's parameters, and the test data is used to assess the model's performance on unseen data. Such partitioning ensures the integrity and accuracy of the subsequent analysis and modeling processes, ultimately leading to reliable results and insights. As shown in Table 1, the dataset is categorized into three subsets: “Train Data” (3,396 data), “Validity Data” (424 data), and “Test Data” (424 data). This division ensures a balanced representation for model training, validation, and evaluation, contributing to the robustness of the research findings and the model's predictive capabilities.

Table 1. Total dataset

Dataset	Count
Train data	3396
Validity data	424
Test data	424

3.4. Auto labeling

In the autolabeling process, data is systematically labeled using relevant keywords to categorize it effectively. Keywords like “car accident” lead to labeling as “Car Accidents,” “motorcycle accident” to “Motorcycle Accidents,” and “bus accident” to “Bus Accidents.” When no specific keywords apply, data is automatically labeled as “Other Accidents.” This streamlines data organization, making analysis and management based on accident types more straightforward. The autolabeling process boasts an impressive 91% accuracy, as indicated in Figure 4, demonstrating its effectiveness in accurate data categorization.

```
Auto Labeling Accuracy: 91.00%


Confusion Matrix:
[[ 3  0  0  0]
 [ 0  5  0  6]
 [ 0  0  2  3]
 [ 0  0  0 81]]
Precision: 0.919
Recall: 0.91
F1-Score: 0.8946898496240603
```

Figure 4. Auto labeling performance

3.5. Classification training using IndoBERT

In the training process, tokenization is crucial for breaking down text into meaningful components, a necessary step in text preprocessing for analysis [23]. The AutoTokenizer from the Transformers library is used to transform text into a format understandable by the model [24]. This ensures effective analysis by converting text into a sequence of tokens. Label encoding is also employed, using the LabelEncoder, to convert label categories into numerical values, essential for deep learning models that require numeric data.

Data loading, the initial step in preparation, involves extracting both training and validation data from CSV files. Loading input text and labels separately helps distinguish between training and validation datasets. The Indobert model is initialized with an Indonesian language-specific tokenizer, and tokenization and padding are applied to text data to ensure effective processing.

Preparing input tensors is crucial, as deep learning models require data in tensor format. Model compilation involves defining the optimizer, loss function, and accuracy metric, guiding model updates during training. The model is then trained on the training dataset and evaluated on the validation data to assess performance. Finally, the trained model is saved in a “saved_model” directory for future use, allowing application in production or further testing without retraining. These steps collectively contribute to the development of an efficient and accurate text classification model.

3.6. Classification testing using test dataset

In this study, we share the outcomes of our traffic accident classification model powered by IndoBERT, a language model customized for Indonesian. The model effectively categorizes accidents into four types: car, motorcycle, bus, and others, achieving an impressive 94.34% overall accuracy. This accuracy is crucial for robust accident management and prevention strategies.

The model's classification performance is further detailed in Table 2, presenting a confusion matrix. The matrix reveals the model's predictions aligning with actual instances for each category. Notably, the model correctly predicted 10 car accidents out of 25, 8 motorcycle accidents out of 13, and 6 bus accidents out of 10. The most notable performance was in the “other accidents” category, where 400 predictions matched the 376 actual cases. This breakdown allows for a comprehensive evaluation of the model's accuracy across various traffic accident categories.

Table 2. Comparsion of predicted and actual categories

Label	Predicted	Actual
Car Accident	10	25
Motorcycle Accident	8	13
Bus Accident	6	10
Other Accident	400	376

Table 3 showcases an example result, demonstrating the model's classification output for a specific text. The provided text describes a motorbike accident that occurred on February 25th, and the model accurately predicts the category as “Motorcycle Accident.” This example highlights the model's effectiveness in correctly categorizing real-world instances based on the information provided.

Table 3. Example result

Text	Label
25 feb terjadi kecelakaan motor roda tiga masuk ke sungai di jl pulau misol belum diketahui secara detail kronologis kejadian info dari upiksa_honda_mobil_bali citizenjournalist infodenpasar	Motorcycle Accident

The provided classification metrics outline the performance of a model in categorizing accidents into different classes. Notably, the model achieves high precision across most classes, indicating a low rate of false positives. For instance, the “Car Accident,” “Motorcycle Accident,” and “Bus Accident” classes all demonstrate a precision of 1.00, implying that when the model predicts an accident in these categories, it is almost always correct. However, the recall values vary across classes. The “Motorcycle Accident” class exhibits a relatively low recall of 0.40, suggesting that the model struggles to capture all instances of actual motorcycle accidents, potentially leading to underreporting. Similarly, the “Car Accident” class has a recall of 0.60, indicating that there is room for improvement in identifying all occurrences of car accidents. On the contrary, the “Other Accident” class displays a high recall of 1.00, implying that the model effectively captures all instances of accidents categorized as “Other.”

The overall accuracy of 94% indicates the model's proficiency in making correct predictions, but it's crucial to address the variations in recall, especially for specific accident types, to enhance the model's comprehensiveness in identifying diverse accident scenarios. The macro-averaged metrics provide a comprehensive view of the model's performance across all categories. The macro-averaged precision was 0.98, recall stood at 0.65, and the F1-Score reached 0.76. These metrics offer a balanced assessment of the model's overall capabilities, considering all classes equally without bias. The weighted averages for precision (0.95), recall (0.94), and F1-Score (0.93) account for the contribution of each category to the overall dataset. They highlight the model's strong performance in a more representative context, indicating its efficacy in real-world applications as shown in Table 4.

Table 4. Performance metrics for traffic accident classification

Label	Precision	Recall	F1-Score	Support
Car accident	1.00	0.60	0.75	10
Motorcycle accident	1.00	0.40	0.57	25
Bus accident	1.00	0.62	0.76	13
Other accident	0.94	1.00	0.97	376
Accuracy			0.94	424
Macro Avg	0.98	0.65	0.76	424
Weighted Avg	0.95	0.94	0.93	424

The confusion matrix in Table 5 displays the classification outcomes of a model across four accident categories: Car Accident, Motorcycle Accident, Bus Accident, and Other Accident. The diagonal entries represent accurate predictions, while off-diagonal entries indicate misclassifications. For example, in the "Car Accident" row, the model accurately predicted 6 instances but misclassified 4 as "Other Accident." Similarly, for "Motorcycle Accident," the model made 10 correct predictions but misclassified 15 as "Other Accident." Notably, the model achieved perfect accuracy in predicting "Other Accident" with 376 correct classifications. This matrix sheds light on the model's performance nuances, emphasizing areas for refinement, particularly in accurately identifying Motorcycle and Car Accidents.

Table 5. Confusion matrix label

		Actual			
		Car accident	Motorcycle accident	Bus accident	Other accident
Predicted	Car accident	6	0	0	4
	Motorcycle accident	0	10	0	15
	Bus accident	0	0	8	5
	Other accident	0	0	0	376

3.7. Evaluation

Evaluation metrics are crucial for achieving an optimal classifier during classification training [25]. In our evaluation, we compared the performance of the IndoBERT model with SVM and RandomForest models for traffic accident classification using the same dataset. The comparative results reveal that IndoBERT achieved a superior accuracy rate of 94.34%, surpassing the SVM model (87.97%) and the RandomForest model (85.85%), as shown in Table 6.

Table 6. Comparsion with another model

Classification model	Accuracy
IndoBERT	94.34%
SVM	87.97%
RandomForest	85.85%

This underscores IndoBERT's effectiveness in classifying accident categories, especially in the Indonesian language context. While SVM and RF have their merits in specific situations, IndoBERT excels in handling the intricacies and nuances of the Indonesian language, making it the preferred choice for this accident classification task. Its deep learning capabilities enable it to capture context and subtleties in text data, resulting in improved accuracy and performance.

4. CONCLUSION

Our study introduces a novel traffic accident classification approach that leverages IndoBERT, a language model customized for the Indonesian language. The notable achievement of a 94% accuracy rate in categorizing accidents into four distinct classes, surpassing traditional methods such as RF and SVM, highlights the model's remarkable proficiency in capturing the intricacies of the Indonesian language. This breakthrough signifies a significant contribution to accident management and prevention, with the potential to reduce the loss of life, injuries, and economic costs associated with traffic accidents in Indonesia and beyond.

Looking ahead, future research should prioritize expanding the model's multilingual capabilities to accommodate the linguistic diversity of the region. Additionally, exploring real-time implementation for immediate accident response and refining the classification system for more granular accident categorization are essential objectives. Furthermore, enriching the dataset, establishing collaboration with traffic authorities, and integrating the model into decision support systems and accident reporting platforms are crucial steps to maximize its real-world impact and contribute to safer road environments.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to Mr. Dody, who generously provided the valuable traffic accident dataset from Indonesia via Kaggle. His contribution played a pivotal role in the success of our research, enabling us to carry out our study and achieve meaningful results in the field of traffic accident classification. We deeply appreciate his support and willingness to share this essential dataset, which has significantly contributed to advancing our understanding of traffic accidents in Indonesia.

Additionally, we extend our appreciation to the authors and organizations whose research and publications have been integral to our study. The contributions who have enriched our understanding and influenced the methodologies applied in our research. We also acknowledge the diverse range of sources that have been instrumental in providing the necessary context and insights for our investigation.

REFERENCES

[1] World Health Organization, "Road traffic injuries," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. [Accessed: July 2023].
[2] L. Sheela, "A review of sentiment analysis in Twitter data using hadoop," *International Journal of Database Theory and Application*, vol. 9, no. 1, pp. 77-86, 2016.
[3] A. S. Cahyono, "Influence of social media on social changes in Indonesian society." *Publiciana*, vol. 9, no. 1, pp. 140-157, 2016.
[4] A. Gal-Tzur, S. M. Grant-Muller, T. Kuflik, E. Minkov, S. Nocera, and I. Shoor, "The potential of social media in delivering transport policy goals," *Transport Policy,* vol. 32, pp. 115–123, Mar. 2014, doi: 10.1016/j.tranpol.2014.01.007.
[5] "Twitter: monthly active users worldwide | Statista", [Online]. Available: https://www.statista.com/statistics/282087/number-of-monthly-activetwitter-users/. [Accessed: Jul. 03, 2020].
[6] A. Z. Arifin and D. S. Prayoga, "Mining Indonesian tweets to identify road traffic accident: A case study in Jakarta", 2020.
[7] N. F. Astuti, "Understanding the functions of social media, along with their definitions and types." [Online]. Available: https://www.merdeka.com/jabar/mengenal-fungsi-media-sosial-berikut-pengertian-dan-jenisnya-kln.html?page=3, 2021
[8] D. A. Saputro and A. S. Girsang, "Classification of traffic accident information using machine learning from social media," *International Journal of Emerging Trends in Engineering Research,* vol. 8, no. 3, pp. 630–637, Mar. 2020, doi: 10.30534/ijeter/2020/04832020, 2020.
[9] S. W. Sakti, A. Purwarianti, and S. Hartati, "Named entity recognition on Indonesian language using IndoBERT," *Procedia Computer Science,* vol. 135, pp. 425-432, 2021.
[10] D. Abbot, "Introduction to text mining: virtual data intensive summer school," *Abbot Analytics, Inc* 2013.
[11] K. R. Prilianti and H. Wijaya, ""Text mining application for automating the determination of trending thesis topics using the K-Means clustering method (in Indonesian: Aplikasi text mining untuk automasi penentuan tren topik skripsi dengan metode K-Means clustering,)" *Jurnal Cybermatika,* vol. 2, no. 1, pp. 1–6, 2014.
[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
[13] A. Althnian *et al.* "Impact of dataset size on classification performance: An empirical evaluation in the medical domain," *Applied Sciences*, vol. 11, p. 796, 2021, [Online]. Available: https://doi.org/10.3390/app11020796.
[14] Dody, "Traffic accident in Indonesia," [Online]. Available: https://www.kaggle.com/datasets/dodyagung/accident/data. [Accessed: August 2023].
[15] Z. Xiao *et al.,* "Impacts of data preprocessing and selection on energy consumption prediction model of HVAC systems based on deep learning," *Energy and Buildings,* vol. 258, 111832, 2022, doi: 10.1016/j.enbuild.2022.111832.
[16] A. Kulkarni and A. Shivananda, "Natural language processing recipes," *Apress,* 2019.
[17] W. Wijaya, K. Rianto, and R. Manurung, "IndoBERT: A pretrained language model for Indonesian language processing".
[18] M. Rafało, "Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis," *ICT Express,* vol. 8, no. 2, pp. 138-188, May 2021, doi: 10.1016/j.icte.2021.05.001.
[19] Y. Goldberg and G. Hirst "Neural network methods in natural language processing," *Morgan and Claypool Publishers,* p. 69, 2017.

[20] C. D. Manning, "Speech and language processing: an introduction to natural language processing," SPEECH Lang. Process. An Introd. to Nat. Lang. Process. Comput. Linguist. Speech Recognit., pp. 1–18, 2021, [Online]. Available: http://www.cs.colorado.edu/~martin/slp.html.

[21] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, and F. Herrera, "A survey on data preprocessing for data stream mining: Current status and future directions," *Neurocomputing*, vol. 239, pp. 39–57, May 2017, doi: 10.1016/j.neucom.2017.01.078, 2017.

[22] H. Putra, "Classification of sentiment in online motorcycle taxi services using the Naive Bayes classifier method (in Indonesian: Klasifikasi sentimen layanan ojek online menggunakan metode Naive Bayes classifier)," Skripsi Thesis, Universitas Islam Negeri Sultan Syarif Kasim Riau, 2018.

[23] A. Kulkarni and A. Shivananda, "Natural language processing recipes," *Apress*, 2019.

[24] T. Wolf *et al.*, "Transformers: State-of-the-art natural language processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, 2020.

[25] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining and Knowledge Management Process*, vol. 5, no. 2, pp. 1–11, 2015, doi: 10.5121/ijdkp.2015.5201, 2015.

BIOGRAPHIES OF AUTHORS

Muhammad Alwan Naufal is currently student at master information techonology at Bina Nusantara University, Jakarta. He completed his undergraduate studies majoring in Computer Science at BINUS University from 2015 to 2019, where he gained a strong foundation in various aspects of computer science and technology. After graduating, he embarked on his professional journey, joining PT QPRO Sukses Mandiri as a Software Developer Intern from March 2018 to February 2019. During this period, he honed his software development skills and gained practical experience in creating applications. Following this internship, he continued his tenure at PT QPRO Sukses Mandiri as a Developer from March 2019 to March 2020. This role allowed him to work on various projects, contributing to the development of software solutions and enhancing his expertise in the field. Subsequently, he took on a new challenge by joining PT Mitra Integrasi Informatika as an Associate Application Developer from July 2020 to July 2022. In this role, he actively participated in designing and developing applications, further expanding his knowledge and proficiency in the realm of application development. In August 2022, he embarked on a new chapter by becoming a Full Stack Developer at PT Siloam International Hospitals Tbk. His responsibilities in this role encompass a wide range of tasks related to web and application development. He continues to excel in this position, bringing his skills and expertise to contribute to the healthcare industry. He can be contacted at email: muhammad.naufal@binus.ac.id.

Abba Suganda Girsang is currently lecturer at master information technology at Bina Nusantara University, Jakarta. He obtained Ph.D. degree in the Institute of Computer and Communication Engineering, Department of Electrical Engineering and National Cheng Kung University, Tainan, Taiwan, in 2014. He graduated bachelor from the Department of Electrical Engineering, Gadjah Mada University (UGM), Yogyakarta Indonesia, in 2000. He then continued his master degree in the Department of Computer Science in the same university in 2006–2008. He was a staff consultant programmer in Bethesda Hospital, Yogyakarta, in 2001 and also worked as a web developer in 2002–2003. He then joined the faculty of Department of Informatics Engineering in Janabadra University as a lecturer in 2003-2015. He also taught some subjects at some universities in 2006–2008. His research interests include swarm intelligence, combinatorial optimization, and decision support system. He can be contacted at email: agirsang@binus.edu.