IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 6, December 2025, pp. 4614~4622
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i6.pp4614-4622

Transformer and text augmentation for tourism aspect-based sentiment analysis

Samuel Situmeang, Sarah Rosdiana Tambunan, Jevania, Mastawila Febryanti Simanjuntak, Sandro Sinaga
Information System Study Program, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Toba, Indonesia

Article Info

Article history:

Received Aug 24, 2024
Revised Oct 28, 2025
Accepted Nov 8, 2025

Keywords:

Aspect-based sentiment analysis
BART
BERT
Sentiment analysis
Tourism

ABSTRACT

The 36.98% growth in the quantity of electronic word of mouth (e-WOM) over the past five years presents opportunities for the tourism industry to understand tourists' needs and desires better when analyzed effectively. Aspect-based sentiment analysis (ABSA) is proposed as a solution, as it can identify the sentiment at a more detailed aspect level. Prior research revealed two issues in ABSA: imbalanced datasets and poor performance in representing implicit aspects and opinions. The authors proposed a combination of the bidirectional and auto-regressive transformer (BART) and bidirectional encoder representations from transformers (BERT) models. Leveraging BART capability in modeling context and BERT expertise in modeling text semantics and nuances, the author proposed an ABSA model that combines BART and BERT using the ensemble method. The experimental results reveal that combining these models significantly enhances the performance of the ABSA model, with an F1-score reaching 70%. Furthermore, text augmentation and preprocessing did not bring improvements in ABSA performance.

Corresponding Author:

Samuel Situmeang
Information System Study Program, Faculty of Informatics and Electrical Engineering
Institut Teknologi Del
St. Sisingamangaraja, Sitoluama, Laguboti, Toba, North Sumatra Province, Indonesia
Email: samuel.situmeang@del.ac.id

1. INTRODUCTION

Electronic word of mouth (e-WOM) is an activity where experienced consumers share information and recommendations online regarding certain vendors or products [1], [2]. The Statista survey reports that TripAdvisor's e-WOM witnessed an astounding 36.98% growth from 2018 to 2022. The survey further highlights an annual increase in e-WOM quantity, with projections for continued rising in the upcoming years. The growth undoubtedly presents opportunities for the tourism industry to understand the tourists' needs and desires when analyzed effectively [3].

One form of e-WOM analysis is sentiment analysis, a natural language processing (NLP) task to evaluate the text's sentiment [4]. However, sentiment analysis cannot be used to find out the specific aspects reviewed by visitors. Therefore, aspect-based sentiment analysis (ABSA) is proposed as a solution so that the sentiment of aspects contained can be identified [5]. For the tourism sector, ABSA is applied to understand tourist sentiment towards certain aspects of a tourist attraction.

In ABSA research, the four main focus elements are aspect category (C), aspect term (A), sentiment polarity (S), and opinion term (O) [6]. For instance, considering the sentence "the staff here is very friendly!", the relevant elements are "staff" (A), "service" (C), "quality of service" (O), and "positive" (S). This allows companies to gain more meaningful and specific insight into customer evaluations.

Arianto and Budi [7] revealed that the proposed model performance is suboptimal. However, the study does not specifically elucidate the reasons for this. Potential causes can be found in studies [8], [9], where several issues are identified. The first issue is that the proposed model is sensitive to the imbalanced distribution of data [9]. One of the datasets used is restaurant-ACOS, which has a data distribution heavily skewed towards the quadruple type explicit aspect with explicit opinion (EAEO) at 64%, compared to the other three types: implicit aspect with explicit opinion (IAEO), explicit aspect with implicit opinion (EAIO), and implicit aspect with implicit opinion (IAIO). The second issue is the model's performance on quadruple types containing implicit aspects or implicit opinions, which remains unsatisfactory. The model's performance on the EAIO quadruple type in both datasets shows notably low figures, namely 23.4% for the laptop-ACOS dataset and 20% for the restaurant-ACOS dataset. The IAEO quadruple type shows figures of 52.79% on laptop-ACOS and 43.87% on restaurant-ACOS, while the IAIO quadruple type shows figures of 29.79% on laptop-ACOS and 42.86% on restaurant-ACOS. The results from this study indicate that the proposed model has not yet been able to represent implicit aspects or implicit opinions effectively. Hence, a robust approach to imbalanced data and good characteristics in representing implicit opinions and aspects in reviews is necessary to address these issues.

The previous approach that can be used as an alternative for imbalanced data problems is data augmentation. Data augmentation helps achieve several goals such as regularization, reducing real-world data, especially in privacy-sensitive domains, minimizing labeling effort, balancing imbalanced datasets, and enhancing resilience to adversarial examples [10]. Shorten and Khoshgoftaar [10] revealed that data augmentation cannot overcome all possible transformations and eliminate all types of bias in the data. As an example of bias in data from research [10], if the dataset in a news classification task does not contain sports-related articles, then the data augmentation method used based on that data set will most likely not produce data related to sports articles, even if it is essential. Data augmentation can introduce supplementary, undesirable biases that result in an inaccurate representation of the entire population. For instance, language models like generative pre-trained transformer (GPT) can create biases that are then transferred into the dataset such as the high correlation between the word “criminal” and male identity in the GPT-2 output, as well as “God” and Christianity. This can cause the model to produce biased decisions [11].

There are various data augmentation techniques with varying complexity. However, complex data augmentation techniques tend to be inefficient. Increased demand on resources, especially when training generative models, is a natural part of data augmentation. To mitigate some of the limitations and maximize the advantages of data augmentation, the authors propose improvements to existing data augmentation approaches based on suggestions proposed in [12]. Longpre *et al.* [12] stated that several data augmentation techniques do not increase the performance of pre-trained transformer models such as bidirectional encoder representations from transformers (BERT), robustly optimized BERT pretraining approach (RoBERTa), and extra-long neural network (XLNet) on simple classification tasks. They hypothesized that data augmentation techniques would only be useful if they could produce novel linguistic patterns that had never been encountered before.

Meanwhile, a previous approach as an alternative to representing implicit opinions for mentioned aspects in reviews is by combining bidirectional and auto-regressive transformer (BART) and BERT using the ensemble method. Lewis *et al.* [13] indicates that combining these two transformer models using ensemble method significantly enhances the performance of transformer models to provide more robust and efficient solutions for various NLP tasks. By integrating BART's strengths in modeling context and BERT's strengths in understanding text' semantics and nuances, this approach can provide more comprehensive performance across various NLP tasks. The ensemble [14] method's use enables the concurrent leveraging of both models' strengths, thus diminishing each model's shortcomings and improving the performance of the resulting model [15].

A factor that may contribute to the model's performance is the dataset quality [16]. In NLP tasks, it's necessary to accurately represent the characters or words in each sentence to achieve high-quality data. Conversely, poor data quality may affect the model's performance. Representing these units presents diverse challenges depending on the language being processed and writing systems [17]. This known as text preprocessing [18], needs to be applied to investigate the dataset characteristics that align with the proposed model in handling the ABSA case and determining its significance.

Hence, this study objective was to address the identified shortcomings by leveraging transformer language models. To the best of our knowledge, this initiative marks a novel approach to integrating BART and BERT using an ensemble method, complemented by text augmentation and preprocessing, to address ABSA challenges. The key contributions of this paper are as follows:

i) Development of an effective transformer model for representing implicit aspects and opinions from review texts.

ii) Significance of the text augmentation impact on the BART+BERT ensemble model.
iii) Significance of the text preprocessing impact on the BART+BERT ensemble model.

The paper is structured in the following way: section 2 describes the proposed models and details the experimental setup. Section 3 analyzes the results and evaluates the models' performance. Finally, section 4 concludes with a discussion of future work.

2. METHOD

This study proposes a novel ABSA model that integrates the strengths of two powerful transformer architectures: BART [13] and BERT [19]. They are pre-trained language models (PLMs) with impressive capabilities that are applicable to various NLP tasks. The proposed ABSA model leverages the large language models (LLMs): BARTLARGE and BERTBASE uncased. Initially, both were designed as general-purpose transformers, applicable to a broad range of NLP tasks. They serve distinct functions. BERT is suitable for NLP tasks that require a deep understanding of semantics and language context, such as text classification [20] and named entity recognition [21]. On the other hand, BART is ideal for NLP tasks that involve text generation, such as text summarization [22], question generation [23], creative text generation [24], and machine translation [13]. These models have been trained on massive datasets, enabling them to capture the human language nuances.

This study introduced modifications to the model's final (head) layer by appending the linear layer to each pre-trained model. Instead of a single output neuron for sentiment classification (typical in single-label tasks), the head is modified to have multiple output neurons, one for each possible sentiment label per aspect. The next crucial step involves applying fine-tuning to prepare the LLMs for ABSA. In traditional sentiment analysis, the task is typically single-label classification (e.g., positive, negative, and neutral). However, ABSA tasks necessitate the model to discern aspects within the text and subsequently classify the sentiment towards each aspect. Park *et al.* [23] encompasses six aspects, namely attractions, amenities, accessibility, image, price, and human resources, as pivotal aspects of tourism. This multi-label classification nature necessitates adjustments to the model. Fine-tuning entails adapting the pre-trained models to the ABSA needs.

The subsequent step involves calculating the average class probabilities. This entails averaging the probabilities predicted by the model across all labels to determine the overall likelihood of each sentiment class. Computing these averages gains a more stable and reliable estimate of the sentiment for each aspect, mitigating the impact of any single prediction's variability. Subsequently, the prediction employs these averaged probabilities to ascertain the final aspects's sentiment classification. The model assigns the label with the highest average probability, ensuring an accurate sentiment analysis. This method leverages the averaged probabilities to make well-founded predictions, thereby enhancing the model's performance in multi-label sentiment classification tasks. Figure 1 depicts the overall architecture of the model.

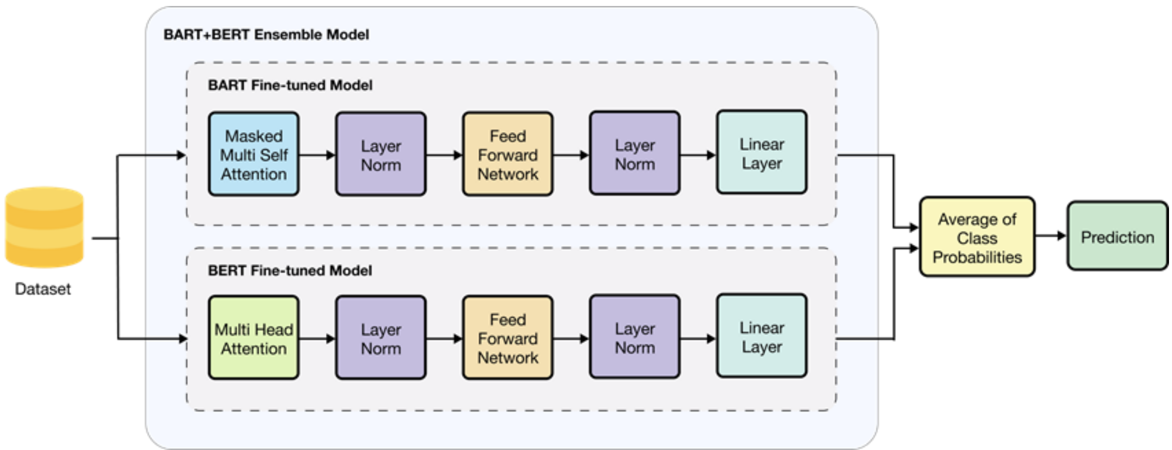

Figure 1. The proposed BART+BERT model architecture

This study employed a dataset of Google Maps user reviews of the Borobudur Temple and Prambanan Temple tourist attractions [7]. Reviews in this dataset use Indonesian, English, and some use both languages simultaneously. Therefore, language standardization was applied to the review text to English which was based on BART's superior performance in text generation in English [13].

As shown in Table 1, the dataset is imbalanced, where the IAEO quadruple type holds more prominence within both groups in contrast to the remaining quadruple types. Adopting an imbalanced dataset can lead to biased model performance, the model's performance tends to be better only for the majority quadruple type. Therefore, this study applies text augmentation to balance the dataset with the dominant type amount, preventing the model from becoming biased towards the majority quadruple types. In this study, the text augmentation techniques to be employed are word replacement, back translation, and random deletion, considering that these are commonly used in text augmentation [25]. Moreover, based on research in [26], [27], it is shown that these have a positive impact on model performance. In the back translation technique, the chosen intermediary language is Chinese due to its established use in prior studies, which are considered mature [28] and leverage the characters' extensive diversity, approximately 47,035, enabling the addition of diverse sentence variations [29].

The exploratory data analysis revealed a substantial degree of character inconsistency within the dataset, necessitating text preprocessing to investigate its impact on model performance. Therefore, this study applies common text preprocessing techniques [18] that have established a beneficial effect on model performance in prior research. These techniques include case folding [30], slang word conversion [31], digit conversion, special character removal [32], stopword removal [33], lemmatization [34], and stemming [34].

Consequently, this study will examine the impact of different text augmentation and preprocessing techniques on ABSA model performance. It aims to identify the most effective methods through independent experimentation, implementing the superior text augmentation technique before evaluating each preprocessing technique compared to the untreated dataset. To ensure the generalizability of the findings and mitigate potential biases, this study utilize the normality test of the performance data using the Shapiro-Wilk test within a 5-fold cross-validation framework. Subsequently, a non-parametric Wilcoxon signed-rank test will be carried out to determine the statistical significance of any observed differences in performance between the models across the folds.

Given the significant role of hyperparameters in model optimization, this study draws upon insights from previous research [35]–[37] to define the hyperparameter settings. Apart from adopting several hyperparameters' settings referred to prior research, the authors also implemented controlled trials and analyzed the experimental results to optimize the hyperparameters' settings. Table 2 lists the optimal hyperparameters.

Table 1. Number of quadruple types per review group

Review group	Quadruple type			
	EAEO	EAIO	IAEO	IAIO
Borobudur temple	1,186	493	2,102	224
Prambanan temple	1,246	679	2,302	309

Table 2. Optimal hyperparameter settings for BART-BERT

Parameter	Values
Epoch	10
Batch size	32
Learning rate	3×10-5

We evaluate model performance using the F1-score, which is the harmonic mean of precision and recall. Precision measures the ratio of accurately predicted aspect or opinion terms out of all the predicted terms. In comparison, recall is defined as the ratio of correctly predicted aspect terms or opinion terms to the total number of aspect terms or opinion terms in the dataset. We also utilize the flesch reading ease (FRE) [38] and the gunning fog index (FOG) [39] to assess the readability and complexity of the textual data. The FRE assigns a score from 0 to 100, where higher numbers mean the text is easier to read. The FOG index, conversely, estimates the years of schooling needed to understand the text on a first reading. A higher FOG score means the text is more complex, with a score above 14 generally considered difficult.

3. RESULTS AND DISCUSSION

The proposed models' effectiveness was assessed through various experiments to evaluate their capabilities. The following subsections show and analyze the model performance evaluation for ABSA.

3.1. Comparative analysis of ABSA model

Table 3 displays the model performance benchmarking. The results show that the proposed model demonstrates a good capability to identify the aspect terms and evaluate sentiment polarity. The experiment results showcased in Table 3 indicate that the BART+BERT ensemble model accomplished superior results compared to recent approaches: BERT [19], k-nearest neighbors (KNN) [40], linear support vector machine

(SVM) [41], radial basis function (RBF)-SVM [42], decision tree [43], random forest [44], multilayer perceptron (MLP) [45], AdaBoost [46], naive Bayes [47], and quadratic discriminant analysis (QDA) [48].

Table 3. Performance of BART+BERT model vs. previous ABSA models

Model	$Precision_{macro}$	$Recall_{macro}$	$F1_{macro}$
BART+BERT (proposed)	0.67	0.78	0.70
BERT	0.24	0.14	0.15
KNN	0.42	0.21	0.25
Linear SVM	0.05	0.06	0.05
RBF-SVM	0.31	0.11	0.13
Decision tree	0.28	0.15	0.17
Random forest	0.08	0.06	0.05
MLP	0.26	0.12	0.13
AdaBoost	0.45	0.25	0.3
Naive Bayes	0.11	0.22	0.12
QDA	0.13	0.87	0.15

However, it's crucial to acknowledge that the data distribution for the BART+BERT combination deviates from normality. Similarly, the BERT model also exhibits a deviation from normal distribution. Consequently, a non-parametric Wilcoxon signed-rank test assessed the statistical significance between the proposed models and recent approaches, such as BERT [19]. This test yielded a p-value of 0.041, indicating the p-value is less than α (0.05), which reveals that the observed difference in performance is statistically significant. The proposed models' impressive capability makes it feasible to address the ABSA challenge to extract the implicit terms, as depicted in Table 4.

Table 4 presents an overview of several findings obtained from the results of BART+BERT and BERT. In the first instance, BERT identifies both "attraction" and "image" aspects with positive sentiments, which does not align with the expected target. BERT fails to recognize that "must visit" implies a recommendation without expressing a strong positivity. However, the BART+BERT model successfully focuses on implicit opinion consistent with the overall context. On the other hand, in the second instance, the BERT model inaccurately predicts the "attraction" and "image" aspects implied in the text, capturing only the explicitly mentioned "price" aspect. Meanwhile, the BART+BERT model successfully predicts the implied aspects despite the low readability and high complexity of the text, as evidenced by its high FOG score (27.02) and low FRE score (18.02). The BART+BERT model effectively captures the "Attraction" and "Image" aspects implied by the strong dissatisfaction with the high entrance ticket prices to Borobudur. The results highlight the proposed model's excellence in addressing sentiment analysis at the aspect level, showcasing its ability to discern and interpret various tourism aspects contained within textual content, even in the context of low readability and complex sentences. The results highlight the proposed model's excellence in addressing sentiment analysis at the aspect level, showcasing its ability to discern and interpret various tourism aspects contained within textual content, even in the context of low readability and complex sentences.

Table 4. Comparative ABSA task example

FRE	FOG	Input text	Predictions		Target
			BART+BERT	BERT	
82.81	5.84	"A must visit in Yogyakarta! Go for the sunrise for sure"	(Attraction, 0)	(Attraction, 1), (Image, 1)	(Attraction, 0)
18.02	27.02	"The entrance ticket for foreigners is very expensive, 350 thousand, local 50 thousand, very expensive, it's no wonder tourists rarely come to Borobudur, in Bali alone there are many interesting and great tourist attractions to visit but the entrance ticket price is not even 5x that for foreigners, very disappointed came all the way from Bali to Magelang bringing foreign guests"	(Price, -1), (Image, -1), (Attraction, -1)	(Price, -1)	(Price, -1), (Image, -1), (Attraction, -1)

3.2. Performance BART+BERT with text augmentation

Table 5 presents a comparative analysis of model performance across various text augmentation techniques. The results imply that the original dataset without augmentation, demonstrates good performance. The Wilcoxon signed-rank test obtained a p-value of 0.074 for word replacement, 0.078 for back translation, and 0.074 for random deletion. Notably, all p-values are more than 0.05, indicating no statistically significant difference in performance between the BART+BERT and untreated model. The study found that augmenting the text often altered or even removed the specific tourism context crucial to the reviews. This change in context resulted in the modification of the quadruple type, which consists of the aspect term and opinion term. Consequently, this introduced noise rather than enhancing the dataset, leading to no improvement in the model's ability.

Table 5. Performance of BART+BERT with text augmentation

Model	$\text{Precision}_{\text{macro}}$	$\text{Recall}_{\text{macro}}$	F1_{macro}
Original	0.67	0.78	0.70
Back translation	0.44	0.55	0.40
Random deletion	0.43	0.52	0.38
Word replacement	0.43	0.52	0.38

3.3. Performance BART+BERT with text preprocessing

Table 6 showcases a comparative analysis of model performance employing different text preprocessing techniques. The result points to the original dataset without preprocessing, demonstrating effective performance. The Wilcoxon signed-rank test produced are as follows: case folding at 0.074, slang word processing at 0.074, digit conversion at 0.074, stop word removal at 0.074, special character removal at 0.078, lemmatization at 0.078, and stemming at 0.078. Since all p-values are more than 0.05, thus there is no statistically significant difference in performance between the BART+BERT and untreated model.

Table 6. Performance of BART+BERT with text preprocessing

Model	$\text{Precision}_{\text{macro}}$	$\text{Recall}_{\text{macro}}$	F1_{macro}
Original	0.67	0.78	0.70
Case folding (lowercasing)	0.56	0.35	0.40
Slang words conversion	0.56	0.33	0.39
Digit conversion	0.56	0.34	0.40
Special character removal	0.57	0.31	0.36
Stopword removal	0.47	0.23	0.27
Lemmatization	0.57	0.31	0.36

4. CONCLUSION

Our study reveals the superior performance of the BART+BERT ensemble model in analyzing aspect-level sentiment. The Wilcoxon signed-rank test reveals that BART+BERT ensemble models demonstrate significantly superior performance than non-ensemble models, with a p-value of 0.041, under the significance threshold of α (0.05). Furthermore, achieving an F1-score of 70%, the proposed model exhibited robust capabilities in recognizing and interpreting tourism-related aspects within reviews, even amidst low readability and complex sentences. Furthermore, we investigated the impact of two supplementary techniques, namely text augmentation and text preprocessing, on the capability of the BART+BERT model. Surprisingly, neither technique brought improvements to the proposed model performance. In concluding our study, we have demonstrated the effectiveness of the BART+BERT ensemble model in ABSA, even in low readability and complex sentences. This underscores its potential as an efficient solution for sentiment analysis in tourism reviews. However, for future work endeavors, we recommend exploring domain-specific augmentation techniques for tourism-related data that can preserve the quadruple type. Furthermore, future studies may also consider automating the hyperparameters configuration for the BART+BERT ensemble model to maintain its performance edge in ABSA.

FUNDING INFORMATION

The authors would like to express their gratitude to the Research and Community Service Unit (LPPM) of Institut Teknologi Del for-funding research under contract number 021.33/ITDel/LPPM/Penelitian/IV/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Samuel Situmeang	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓
Sarah Rosdiana Tambunan	✓	✓		✓	✓	✓				✓		✓		
Jevania		✓	✓		✓	✓		✓	✓		✓			
Mastawila Febryanti Simanjuntak			✓		✓			✓	✓		✓			
Sandro Sinaga			✓		✓			✓	✓		✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [SS], upon reasonable request.

REFERENCES

[1] N. Donthu, S. Kumar, N. Pandey, N. Pandey, and A. Mishra, “Mapping the electronic word-of-mouth (eWOM) research: a systematic review and bibliometric analysis,” *Journal of Business Research*, vol. 135, pp. 758–773, Oct. 2021, doi: 10.1016/j.jbusres.2021.07.015.

[2] H. Simanjuntak, D. Tarigan, I. Sibarani, C. J. Hutapea, R. Lumbantoruan, and M. Sigiro, “Weighted hybrid recommendation system for Toba tourism based on Google Review data,” in *2022 IEEE International Conference of Computer Science and Information Technology*, IEEE, Oct. 2022, pp. 1–8, doi: 10.1109/ICOSNIKOM56551.2022.10034911.

[3] J. Li, L. Xu, L. Tang, S. Wang, and L. Li, “Big data in tourism research: a literature review,” *Tourism Management*, vol. 68, pp. 301–323, Oct. 2018, doi: 10.1016/j.tourman.2018.03.009.

[4] M. D. Devika, C. Sunitha, and A. Ganesh, “Sentiment analysis: a comparative study on different approaches,” *Procedia Computer Science*, vol. 87, pp. 44–49, 2016, doi: 10.1016/j.procs.2016.05.124.

[5] E. Martini, F. Tisadinda, M. A. Sultan, and R. Hurriyati, “Impact of e-WOM and WOM on destination image in shopping tourism business,” *Jurnal Dinamika Manajemen*, vol. 13, no. 1, pp. 66–77, Mar. 2022, doi: 10.15294/jdm.v13i1.31819.

[6] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, “A survey on aspect-based sentiment analysis: tasks, methods, and challenges,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 11, pp. 11019–11038, Nov. 2023, doi: 10.1109/TKDE.2022.3230975.

[7] D. Arianto and I. Budi, “Aspect-based sentiment analysis on Indonesia’s tourism Destinations based on Google Maps User Code-mixed reviews (study case: Borobudur and Prambanan Temples),” in *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, Hanoi, Vietnam: Association for Computational Linguistics, 2020, pp. 359–367.

[8] C. D. Hoang, Q. V. Dinh, and N. H. Tran, “Aspect-category-opinion-sentiment extraction using generative transformer model,” in *2022 RIVF International Conference on Computing and Communication Technologies*, IEEE, Dec. 2022, pp. 1–6. doi: 10.1109/RIVF55975.2022.10013820.

[9] X. Bao, W. Zhongqing, X. Jiang, R. Xiao, and S. Li, “Aspect-based sentiment analysis with opinion tree generation,” in *The Thirty-First International Joint Conference on Artificial Intelligence*, California: International Joint Conferences on Artificial Intelligence Organization, Jul. 2022, pp. 4044–4050, doi: 10.24963/ijcai.2022/561.

[10] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, pp. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[11] I. Solaiman *et al.*, “Release strategies and the social impacts of language models,” *arXiv:1908.09203*, Nov. 2019.

[12] S. Longpre, Y. Wang, and C. DuBois, “How effective is task-agnostic data augmentation for pretrained transformers?,” in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 4401–4411, doi: 10.18653/v1/2020.findings-emnlp.394.

[13] M. Lewis *et al.*, “BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 7871–7880, doi: 10.18653/v1/2020.acl-main.703.

[14] G. Wang, J. Hao, J. Ma, and H. Jiang, “A comparative assessment of ensemble learning for credit scoring,” *Expert Systems with Applications*, vol. 38, no. 1, pp. 223–230, Jan. 2011, doi: 10.1016/j.eswa.2010.06.048.

[15] D. Opitz and R. Maclin, “Popular ensemble methods: an empirical study,” *Journal of Artificial Intelligence Research*, vol. 11, no. 1, pp. 169–198, 1999.

[16] J. M. Moreira, A. C. P. L. F. de Carvalho, and T. Horváth, “Data quality and preprocessing,” in *A General Introduction to Data Analytics*, Wiley, 2018, pp. 71–97, doi: 10.1002/9781119296294.ch4.

[17] N. Indurkhya and F. J. Damerau, *Handbook of natural language processing*, 2nd ed. New York: Chapman and Hall/CRC, 2010, doi: 10.1201/9781420085938.

[18] M. J. Denny and A. Spirling, “Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it,” *Political Analysis*, vol. 26, no. 2, pp. 168–189, Apr. 2018, doi: 10.1017/pan.2017.44.

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of North American Chapter of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
[20] E. C. G.-Merchan, R. G.-Brizuela, and S. G.-Carvajal, "Comparing BERT against traditional machine learning models in text classification," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 4, pp. 352–356, Apr. 2023, doi: 10.47852/bonviewJCCE3202838.
[21] M. Arkhipov, M. Trofimova, Y. Kuratov, and A. Sorokin, "Tuning multilingual transformers for language-specific named entity recognition," in *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 89–93, doi: 10.18653/v1/W19-3712.
[22] V. Goyal, R. Singh, M. Dhawley, A. Kumar, and S. Sharma, "Aerial object detection using deep learning: a review," in *Computational Intelligence*, Springer, Singapore, 2023, pp. 81–92, doi: 10.1007/978-981-19-7346-8_8.
[23] G.-M. Park, S.-E. Hong, and S.-B. Park, "Post-training with interrogative sentences for enhancing BART-based Korean question generator," in *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, Association for Computational Linguistics, 2022, pp. 202–209.
[24] Z. Lin *et al.*, "Text generation with diffusion language models: a pre-training approach with continuous paragraph denoise," in *Proceedings of the 40th International Conference on Machine Learning*, ACM Digital Library, 2023, pp. 21051–21064.
[25] B. Li, Y. Hou, and W. Che, "Data augmentation approaches in natural language processing: a survey," *AI Open*, vol. 3, pp. 71–90, 2022, doi: 10.1016/j.aiopen.2022.03.001.
[26] J. Chen, D. Tam, C. Raffel, M. Bansal, and D. Yang, "An empirical survey of data augmentation for limited data learning in NLP," *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 191–211, 2023, doi: 10.1162/tacl_a_00542.
[27] M. Queudot, L. Marceau, R. M. Belbahar, E. Charton, and M.-J. Meurs, "Dataset augmentation using back-translation to improve early stage dialog systems," in *Proceedings of the Canadian Conference on Artificial Intelligence*, Canadian Artificial Intelligence Association, May 2022, doi: 10.21428/594757db.3986fa37.
[28] S. Xiao and J. W. Oller, "Can relatively perfect translation between English and Chinese be achieved?," *Language Testing*, vol. 11, no. 3, pp. 267–289, Nov. 1994, doi: 10.1177/026553229401100303.
[29] S. Zhao and D. Zhang, "The totality of Chinese characters - a digital perspective," *Journal of Chinese Language and Computing*, vol. 17, no. 2, pp. 107–125, 2007.
[30] M. Munikar, S. Shakya, and A. Shrestha, "Fine-grained sentiment classification using BERT," in *2019 Artificial Intelligence for Transforming Business and Society*, IEEE, Nov. 2019, pp. 1–5, doi: 10.1109/AITB48515.2019.8947435.
[31] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis," *Expert Systems with Applications*, vol. 110, pp. 298–310, Nov. 2018, doi: 10.1016/j.eswa.2018.06.022.
[32] H. Fang, G. Xu, Y. Long, and W. Tang, "An Effective ELECTRA-based pipeline for sentiment analysis of tourist attraction reviews," *Applied Sciences*, vol. 12, no. 21, Oct. 2022, doi: 10.3390/app122110881.
[33] M. M. Samia, A. Rajee, M. R. Hasan, M. O. Faruq, and P. C. Paul, "Aspect-based sentiment analysis for Bengali text using bidirectional encoder representations from transformers (BERT)," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 12, 2022, doi: 10.14569/IJACSA.2022.01312112.
[34] J. Tao, L. Zhou, and C. Feeney, "I understand what you are saying: leveraging deep learning techniques for aspect based sentiment analysis," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, Grand Wailea, Hawaii: HICSS, 2019, pp. 470-479.
[35] A. Murtadha, B. Wen, S. Pan, J. Su, L. Ao, and Y. Liu, "BERT-ASC: auxiliary-sentence construction for implicit aspect learning in sentiment analysis," *Expert Systems with Applications*, vol. 258, Dec. 2024, doi: 10.1016/j.eswa.2024.125195.
[36] C. Sun, L. Huang, and X. Qiu, "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 380–385. doi: 10.18653/v1/N19-1035.
[37] A. Das and W. E. Zhang, "ABSA-Bench: towards the unified evaluation of aspect-based sentiment analysis research," in *Proceedings of the 18th Annual Workshop of the Australasian Language Technology Association*, Australasian Language Technology Association, 2020, pp. 65–71.
[38] R. Flesch, "'Simplification of Flesch reading ease formula': reply.," *Journal of Applied Psychology*, vol. 36, no. 1, pp. 54–55, 1952, doi: 10.1037/h0051965.
[39] R. Gunning, "The fog index after twenty years," *Journal of Business Communication*, vol. 6, no. 2, pp. 3–13, Jan. 1969, doi: 10.1177/002194366900600202.
[40] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: consistency properties," *International Statistical Review/Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, Dec. 1989, doi: 10.2307/1403797.
[41] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, New York, United States: ACM, Jul. 1992, pp. 144–152. doi: 10.1145/130385.130401.
[42] B.-C. Kuo, H.-H. Ho, C.-H. Li, C.-C. Hung, and J.-S. Taur, "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 317–326, Jan. 2014, doi: 10.1109/JSTARS.2013.2262926.
[43] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1023/A:1022643204877.
[44] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[45] M.-C. Popescu, V. E. Balas, L. P.-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Transactions on Circuits and Systems*, vol. 8, no. 7, pp. 579–588, 2009.
[46] Y. Freund and R. E. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, no. 5, pp. 771–780, 1999.
[47] G. I. Webb, "Naïve Bayes," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 713–714. doi: 10.1007/978-0-387-30164-8_576.
[48] Y. Wu, Y. Qin, and M. Zhu, "Quadratic Discriminant analysis for high-dimensional data," *Statistica Sinica*, vol. 29, no. 2, pp. 939–960, 2019, doi: 10.5705/ss.202016.0034.

BIOGRAPHIES OF AUTHORS

Samuel Situmeang received his bachelor's degree in the Information Technology Study Program, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Indonesia, in 2014. He received his master's degree from the Department of Computer Science and Information Engineering, College of Engineering, National Chung Cheng University, Taiwan, in 2017. He currently lectures in the Information System Study Program, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del. His main research interest is AI-driven solutions for education and cultural heritage, which intersect areas of data science, natural language processing, and computer vision. He can be contacted at email: samuel.situmeang@del.ac.id.

Sarah Rosdiana Tambunan received her bachelor's degree in the Information System Study Program, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Indonesia, in 2019. She received her master's degree from the Computer Sciences Master Program, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Indonesia, in 2023. She currently lectures in the Information System Study Program, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del. Her research focuses on software engineering and data security. She can be contacted at email: sarah.tambunan@del.ac.id.

Jevania received her bachelor's degree in the Information System Study Program, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Indonesia, in 2024. Her curiosity over technology earned her a scholarship to the Apple Developer Academy in her second year. With a profound interest in iOS development and machine learning, she was involved in several projects such as an application that leverages machine learning to generate public speaking feedback. Her passion lies in exploring research domains that include machine learning, artificial intelligence, and natural language processing. She can be contacted at email: jevaddicted@gmail.com.

Mastawila Febryanti Simanjuntak received her bachelor's degree in the Information System Study Program, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Indonesia, in 2024. Her research interests are natural language processing, and data mining. She can be contacted at email: mastawilasimanjuntak82171@gmail.com.

Sandro Sinaga received his bachelor's degree in the Information System Study Program, Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Indonesia, in 2024. His research interests are natural language processing and data mining. He can be contacted at email: sanndrosinaga89@gmail.com.