Indonesian Journal of Electrical Engineering and Computer Science
Vol. 40, No. 1, October 2025, pp. 271~279
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v40.i1.pp271-279

Unveiling the influence of back-translation on sentiment analysis of Indonesian hotel reviews

Sandy Kurniawan, Retno Kusumaningrum, Priyo Sidik Sasongko
Department of Informatics, Faculty of Sciences and Mathematics, Universitas Diponegoro, Semarang, Indonesia

Article Info

Article history:

Received Nov 14, 2024
Revised Mar 19, 2025
Accepted Jul 3, 2025

Keywords:

Back-translation
Hotel reviews
Machine learning
Sentiment analysis
Translation

ABSTRACT

This study aims to conduct sentiment analysis on hotel reviews in Indonesian using several machine learning classification algorithms, namely multinomial naive bayes (MNB), support vector machine (SVM), and random forest (RF). The back translation method is employed to generate synthetic data variations that are used as additional data variations in building classification models. This research tests three scenarios based on the datasets used: the original dataset, the dataset resulting from back translation, and the combined dataset of both. The experimental results show that the use of combined data yields better results, with the random forest algorithm standing out as the best performer. Back translation significantly improves model evaluation in sentiment analysis for several reasons, including enriching the dataset with new variations, enhancing model robustness, and increasing dataset complexity. However, the differences in the number of word features among scenarios indicate that back translation also significantly influences the dataset's characteristics.

Corresponding Author:

Sandy Kurniawan
Department of Informatics, Faculty of Sciences and Mathematics, Universitas Diponegoro
50275 Semarang, Indonesia
Email: sandy@live.undip.ac.id

1. INTRODUCTION

In the current digital era, society has been empowered to provide assessments and reviews of various services through specific applications. These reviews serve as windows for other users to understand the experiences provided by previous individuals. A prime example is hotel reviews, which serve as valuable feedback for hotel management [1]. From these reviews, they can identify areas where they have excelled, as well as aspects that require improvement. In the context of hotel reviews, sentiment analysis plays a crucial role in understanding the implied meanings behind each comment. It involves not only counting the number of positive and negative reviews but also sensing the emotional nuances contained within them. By understanding the sentiment behind reviews, hotel managers can respond more effectively to the needs and expectations of their customers [2]. However, in the context of sentiment analysis of reviews in Bahasa Indonesia, the challenge becomes more complex. Most existing sentiment analysis models have been trained with data in English, while hotel review data in Bahasa Indonesia is not as readily available. Therefore, the use of data augmentation techniques in sentiment analysis is necessary to increase the number of hotel reviews in Indonesian language [3]. One technique that can be used to augment review data is back translation, which involves translating text into a foreign language and then translating it back into the original language [4], [5]. This method not only increases the number of available reviews but also helps maintain the core message intended by the author.

Sentiment analysis is a technique used to identify, extract, and understand sentiments contained within text [6]-[8]. The main goal of sentiment analysis is to determine whether a piece of text contains positive, negative, or neutral sentiment. The texts analyzed can take various forms. Sentiment analysis can be performed at several levels. The research conducted by [9] falls under document-level sentiment analysis. Sentence-level sentiment analysis has been done in [10]. For aspect-level sentiment analysis, also known as aspect-based sentiment analysis, it is conducted by [11]. Several methods have been used to perform sentiment analysis in Indonesian. Commonly used methods include random forest (RF) [12], support vector machine (SVM) [12]-[14], k-nearest neighbor (K-NN) [15], [16], ensemble classifier [17], and Naïve Bayes [18]. Additionally, sentiment analysis applications have often been used in various cases, such as analyzing satisfaction in teaching [13], [19], public sentiment in elections [20], [21], social media analysis [22]-[24], and sentiment related to COVID-19 [22], [25], [26].

As one of the data augmentation techniques, back translation has been previously used for tasks related to text mining such as text classification, machine translation, and sentiment analysis. This technique involves translating text from its original language into a target language, and then translating it back into the original language. By doing so, back translation introduces diversity into the dataset and enriches it with new linguistic patterns and expressions. Several studies have applied this technique to address natural language

processing (NLP)-related issues. For instance, in sentiment analysis, back translation has been employed to augment data, leading to improved model performance [27]. Cohen *et al.* [28], the study employs back translation in the task of social network hate detection. The use of back translation for text classification in Chinese language is conducted by [4]. Kurniawan and Budi [29] applies translation mechanism as one form of increasing dataset variation in offensive language detection. Luo *et al.* [30], the study combines back translation with transfer learning to address machine translation with low resource data.

Given the importance of reviews in the hospitality industry and the need to understand the sentiment behind them, this research aims to conduct sentiment analysis on Indonesian-language hotel reviews. We will implement back translation techniques as part of the data augmentation process, with the hope of providing a more comprehensive and accurate insight into users' experiences in their own language. This research aims to investigate the influence of using back translation as a data augmentation technique on sentiment analysis. Three machine learning algorithms are used in creating the sentiment analysis model: multinomial naïve bayes (NMB), support vector machine, and random forest. The dataset used contains three sentiments that may be contained within a review: positive, negative, and neutral.

To achieve this goal, the research is conducted in several stages presented in the following sequence. Section 1 explains the background, issues, and related research concerning the use of back translation in sentiment analysis. Section 2 describes the methods employed in conducting this research. The research scenario, experimental results, and discussion of the findings are presented in section 3. Finally, section 4 is used to conclude this research.

2. METHOD

2.1. Data acquisition

This work utilizes a dataset obtained from the study of [10]. The dataset was obtained from the hotel booking service website Traveloka. The dataset consists of hotel review data with positive, negative, and neutral sentiments. The details of the dataset include 430 positive data, 430 negative data, and 860 neutral data.

2.2. Back-translation

After obtaining the data, data augmentation process is conducted using the back translation technique. This technique involves translating the data into a target language, then translating the resulting translated data back into the original language. By performing this process, the dataset is doubled in size compared to the initial dataset. This research translates the data into English as the target language. With this re-translation, it is expected to generate an expansion of words obtained from the translation process.

2.3. Data preprocessing

Data preprocessing stage is carried out to reduce noise in the data used and also to produce a better data representation. This stage is performed after the back translation process so that the resulting back translation dataset also undergoes the same preprocessing process. Some stages performed in preprocessing include case folding, removal of special characters such as numbers, punctuation marks, and white spaces, tokenization, stopwords removal, and ending with stemming.

2.4. Feature extraction

Feature extraction is used to transform text representations into numeric ones so that they can be used in sentiment analysis models. This research uses bag-of-word representation, specifically Unigram type, where each element in the unigram vector represents the occurrence of a single word in the document. Meanwhile, the word weighting used is TF-IDF. This word weighting aims to indicate how important the word is in the document, represented by a numerical weight.

2.5. Model

The development of the sentiment analysis model in this research utilizes three machine learning algorithms, namely multinomial naïve bayes, support vector machine, and random forest. Additionally, to assess the impact of back translation, this study implements the k-fold cross-validation mechanism with a value of k=10. The use of k-fold cross-validation aims to make the performance evaluation generated by the model more objective. Furthermore, the application of this technique can prevent the model from overfitting to the dataset. Model evaluation is conducted using accuracy, precision, recall, and F1 score metrics. This evaluation is performed for each fold for each sentiment analysis model.

3. RESULTS AND DISCUSSION

3.1. Experiment scenario

This study is divided into 3 scenarios based on the datasets used. By applying back translation to the dataset, two types of datasets can be obtained, namely the original dataset without back translation process and the dataset resulting from the back translation process. Thus, the experimental scenarios of this study are based on the datasets used to build its sentiment analysis model. The three experimental scenarios are as follows:

– Scenario 1: development of the sentiment analysis model using only the original dataset. The results of this scenario are used as the baseline performance.
– Scenario 2: development of the sentiment analysis model using only the dataset resulting from back translation.
– Scenario 3: development of the sentiment analysis model using a combination of the original dataset and the dataset resulting from back translation.

By implementing three experimental scenarios, this study aims to analyze the impact of back translation on sentiment analysis. The comparison of results across different datasets provides insights into how back translation influences model performance, dataset diversity, and linguistic complexity. These findings can serve as a foundation for future research on enhancing sentiment analysis using data augmentation techniques.

3.2. Experiment results

Scenario 1 applies 10-fold cross-validation using the original dataset without any additional data from back translation, with results summarized in Tables 1-2 and Figure 1. Scenario 2 evaluates the dataset generated through back translation using the same validation method, with performance results shown in Tables 3-4 and Figure 2. Scenario 3 combines the original dataset with the back-translated dataset to assess the effectiveness of data augmentation in improving model performance, with results detailed in Tables 5-6 and Figure 3. Comparison of these scenarios provides insights into how different dataset variations influence classification performance, particularly in terms of accuracy and F1 Score.

Table 1 and Table 2 present the performance results for the first scenario, comparing the accuracy and F1 Score of different classification algorithms. Based on these results, the Random Forest algorithm achieved the best performance with an average accuracy of 0.809 and an average F1 Score of 0.806, indicating its ability to balance precision and recall effectively. SVM also demonstrated competitive performance, achieving an average accuracy of 0.805 and an average F1 Score of 0.805. In contrast, multinomial Naïve Bayes recorded the lowest performance, with an average accuracy of 0.766 and an average F1 Score of 0.758, making it the least effective algorithm among the three. To further highlight the performance differences, Figure 1 presents a bar chart comparing accuracy and F1 Score, where Figure 1(a) shows that Random Forest and SVM perform similarly, while Multinomial Naïve Bayes lags behind. Figure 1(b) reinforces this pattern in the F1 Score, visually confirming the trends observed in Table 1 and Table 2.

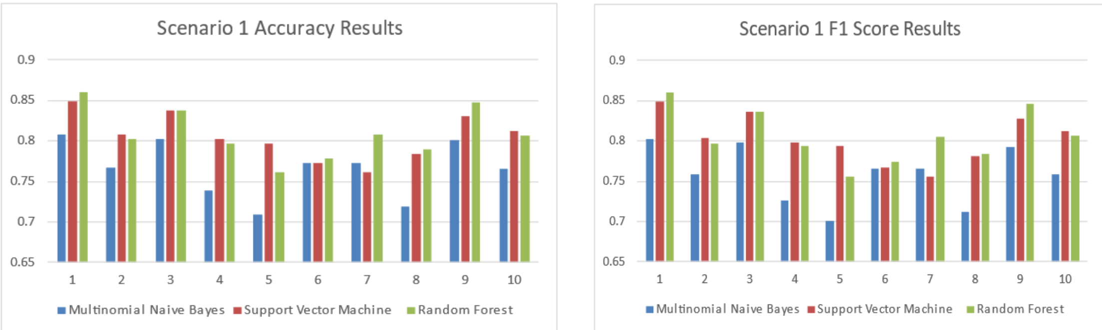
Table 3 and Table 4 present the performance results for the second scenario, showing that the Random Forest algorithm achieved the best accuracy and F1 Score. The Random Forest algorithm recorded the highest average accuracy at 0.8106, followed by SVM with 0.8013, while Multinomial Naïve Bayes had the lowest accuracy at 0.7609. A similar trend is observed in the F1 Score metric, where Random Forest leads with an average of 0.8077, followed by SVM at 0.7987, and Multinomial Naïve Bayes with the lowest score of 0.7538. To enhance the clarity of these comparisons, Figure 2 visualizes the accuracy and F1 Score differences, with Figure 2(a) depicting the accuracy variations across models, while Figure 2(b) illustrates the F1 Score distribution, emphasizing the consistent gap in performance.

Table 1. Scenario 1 accuracy results

Fold	Multinomial naive bayes	Support vector machine	Random forest
1	0.80814	0.848837	0.860465
2	0.767442	0.80814	0.802326
3	0.802326	0.837209	0.837209
4	0.738372	0.802326	0.796512
5	0.709302	0.796512	0.761628
6	0.773256	0.773256	0.77907
7	0.773256	0.761628	0.80814
8	0.719298	0.783626	0.789474
9	0.80117	0.830409	0.847953
10	0.766082	0.812865	0.807018
Average	0.765864	0.805481	0.808979

Table 2. Scenario 1 F1-score results

Fold	Multinomial naive bayes	Support vector machine	Random forest
1	0.802665	0.848591	0.860675
2	0.758333	0.804357	0.797321
3	0.797685	0.836139	0.836656
4	0.725604	0.798332	0.793648
5	0.701322	0.793826	0.756218
6	0.765186	0.766968	0.773735
7	0.765856	0.756507	0.805451
8	0.712057	0.781929	0.784048
9	0.793219	0.827398	0.846503
10	0.758022	0.812341	0.806791
Average	0.757995	0.802639	0.806105



(a) (b)

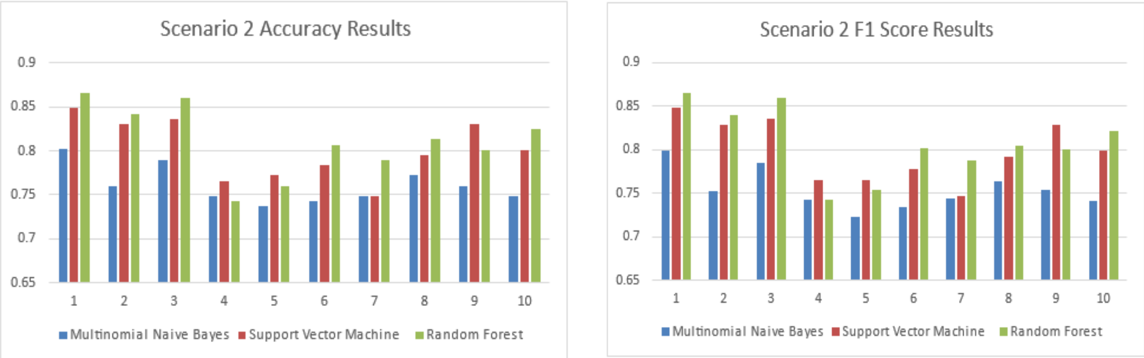
Figure 1. Scenario 1 comparison result (a) accuracy and (b) F1 score

Table 3. Scenario 2 accuracy results

Fold	Multinomial naive bayes	Support vector machine	Random forest
1	0.802326	0.848837	0.866279
2	0.760234	0.830409	0.842105
3	0.789474	0.836257	0.859649
4	0.748538	0.766082	0.74269
5	0.736842	0.77193	0.760234
6	0.74269	0.783626	0.807018
7	0.748538	0.748538	0.789474
8	0.77193	0.795322	0.812865
9	0.760234	0.830409	0.80117
10	0.748538	0.80117	0.824561
Average	0.760934	0.801258	0.810605

Table 4. Scenario 2 F1-score results

Fold	Multinomial naive bayes	Support vector machine	Random forest
1	0.798902	0.848203	0.865972
2	0.751796	0.828475	0.840295
3	0.785175	0.83521	0.859247
4	0.743237	0.764793	0.741902
5	0.722921	0.765692	0.754386
6	0.734183	0.777636	0.801257
7	0.743353	0.747244	0.787283
8	0.763883	0.792596	0.805023
9	0.753536	0.828435	0.80084
10	0.740977	0.798621	0.821117
Average	0.753796	0.798691	0.807732



(a) (b)

Figure 2. Scenario 2 comparison result (a) accuracy and (b) F1 score

Table 5 and Table 6 present the performance results for the third scenario, showing that the Random Forest algorithm once again outperformed the other algorithms. The random forest algorithm achieved the highest average accuracy of 0.9072 and an average F1 Score of 0.9067. Consistent with the trends observed in scenario 1 and scenario 2, SVM ranked second with an average accuracy of 0.8623 and an average F1 Score of 0.8614, while Multinomial Naïve Bayes recorded the lowest performance with an average accuracy of 0.8136 and an F1 Score of 0.8099. To provide a clearer comparison, Figure 3 illustrates the performance differences among the models, with Figure 3(a) depicting accuracy variations and Figure 3(b) highlighting the disparities in F1 Score.

Table 5. Scenario 3 accuracy results

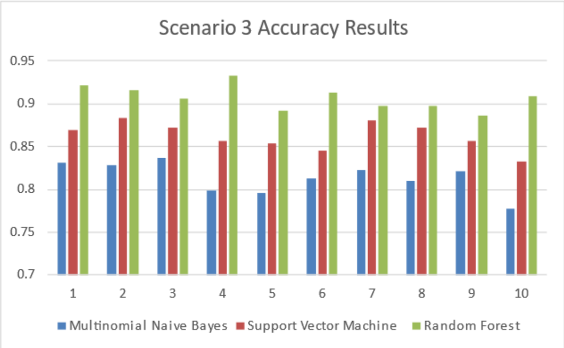
Fold	Multinomial naive bayes	Support vector machine	Random forest
1	0.830904	0.868805	0.921283
2	0.827988	0.883382	0.915452
3	0.836735	0.87172	0.906706
4	0.798834	0.857143	0.932945
5	0.795918	0.854227	0.892128
6	0.813411	0.845481	0.912536
7	0.822157	0.880466	0.897959
8	0.810496	0.87172	0.897959
9	0.821637	0.856725	0.885965
10	0.777778	0.833333	0.909357
Average	0.813586	0.8623	0.907229

3.3. Discussion

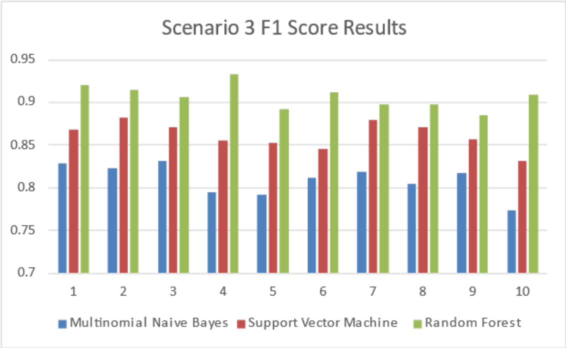
Based on the results obtained from scenarios 1 to 3, the use of a combined dataset, incorporating both original data and back-translated data, produced the highest performance. This is evident from the consistently superior average accuracy and F1 Score values observed in scenario 3 across all algorithms. Additionally, random forest emerged as the best-performing algorithm in every scenario. To provide a clearer comparison of performance across scenarios, Figure 4 summarizes the average accuracy and F1 Score results, where Figure 4(a) highlights the accuracy differences, and Figure 4(b) visualizes the variations in F1 Score, reinforcing the advantage of using the combined dataset.

Table 6. Scenario 3 F1-score results

Fold	Multinomial naive bayes	Support vector machine	Random forest
1	0.828065	0.867799	0.920257
2	0.823663	0.882973	0.915181
3	0.831887	0.870651	0.905857
4	0.795421	0.855944	0.932758
5	0.792271	0.852563	0.891542
6	0.812353	0.845521	0.91237
7	0.819219	0.879481	0.897685
8	0.80485	0.871185	0.897354
9	0.817588	0.856523	0.885338
10	0.773466	0.831572	0.908804
Average	0.809878	0.861421	0.906714

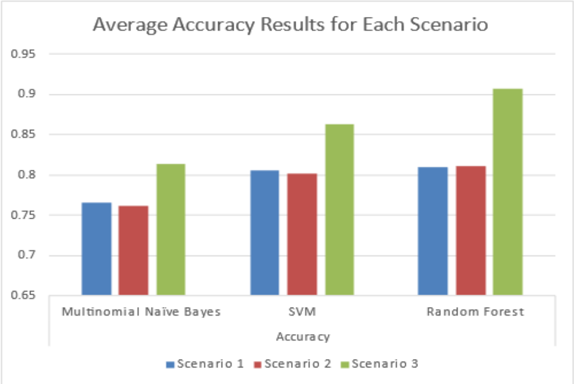


(a)

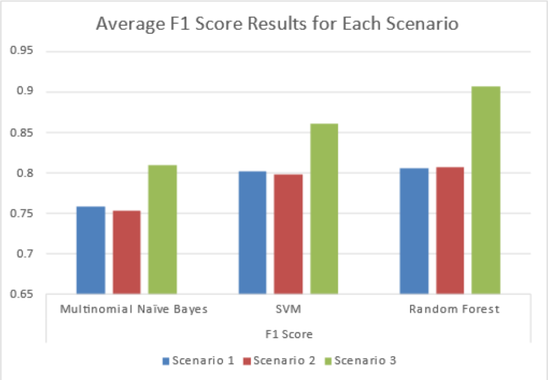


(b)

Figure 3. Scenario 3 comparison result (a) accuracy and (b) F1 score



(a)



(b)

Figure 4. Comparison of average (a) accuracy and (b) F1 score for each scenario

The impact of using back translation in sentiment analysis can improve the performance of model evaluation. There are several reasons influencing the improvement in model performance. These reasons include:

– Larger data variation: back translation generates new variations in the data used. These new variations can help the model learn sentences or words that do not appear in the original dataset. Back translation in this study provided 321 new word features not found in the original dataset.
– Increased robustness: related to sentences or words that do not appear in the original dataset, this can increase the model's resistance to overfitting. The variation in the dataset being learned helps prevent the

model from memorizing small details from each training data used. Thus, the model will be more focused on the general patterns of a class and reduce the possibility of overfitting.

– Higher complexity: the combination of original and back-translated datasets provides sentences with more complex grammar and writing styles. This can make the model adapt to the complexity of the learned sentences.

The number of features obtained using different datasets in each scenario conducted results in quite different amounts. In scenario 1, the generated word features amount to 2,176; scenario 2 produces 1,696-word features, while scenario 3 produces 2,497-word features. The difference in the number of features is caused by several factors, including:

– Limitations in back translation: the back translation process does not always generate word variations that match the words in the original dataset. Some words or phrases may not be translated correctly or have fewer variations in other languages.
– Reduction of information in the translation process: in the translation process, some words may be lost or not translated word by word if they have direct equivalents in other languages.
– Differences in vocabulary: different languages have different vocabularies. Some words or phrases in the original language may not have direct equivalents in the language used in the back translation process, thereby reducing the number of words features generated.

4. CONCLUSION

This study analyzes sentiment in Indonesian hotel reviews using classic machine learning algorithms: multinomial naive bayes, support vector machine, and random forest. To enhance model performance, back translation is applied to generate synthetic data, leading to three research scenarios based on different datasets: the original dataset, the back-translated dataset, and a combination of both. Experimental results show that the combined dataset consistently outperforms the other two scenarios, with the random forest algorithm achieving the best performance. Back translation significantly improves model evaluation by enriching the dataset with diverse patterns, enhancing model robustness, and improving generalization. It introduces higher linguistic complexity, helping the model adapt better. Additionally, variations in word features among the scenarios-2,176 in the original dataset, 1,696 in the back-translated dataset, and 2497 in the combined dataset-highlight its substantial impact on dataset structure and diversity. Thus, back translation is proven to enhance sentiment analysis model performance while significantly altering dataset characteristics.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Department of Informatics and the Faculty of Science and Mathematics, Universitas Diponegoro, for their support under the 2023 fiscal year.

FUNDING INFORMATION

This research was funded by the Faculty of Science and Mathematics, Universitas Diponegoro, under funding reference number 25.E/UN7.F8/PP/II/2023, fiscal year 2023.

AUTHOR CONTRIBUTIONS STATEMENT

The specific contributions of each author are summarized below, following the CRediT taxonomy:

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sandy Kurniawan	✓		✓	✓		✓	✓	✓	✓	✓	✓			✓
Retno Kusumaningrum	✓	✓			✓	✓	✓			✓		✓	✓	
Priyo Sidik Sasongko		✓		✓	✓					✓	✓	✓	✓	

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

All authors have read and approved the final version of the manuscript.

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The original hotel review dataset used in this study was obtained from [10]. As part of this research, we produced a back-translated version of the dataset to facilitate further sentiment analysis tasks. The back-translated dataset is publicly available at the following Github repository : [https://github.com/sandykurniawan/Backtranslation-Hotel-Review].

REFERENCES

[1] B. J. Ali *et al.*, "Hotel service quality: the impact of service quality on customer satisfaction in hospitality," *International Journal of Engineering, Business and Management*, vol. 5, no. 3, pp. 14–28, 2021, doi: 10.22161/ijebm.5.3.2.

[2] R. Jayanto, R. Kusumaningrum, and A. Wibowo, "Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory," *International Journal of Advances in Intelligent Informatics*, vol. 8, no. 3, pp. 391–403, Nov. 2022, doi: 10.26555/ijain.v8i3.691.

[3] Z. Jiang, M. Y. R. Yang, M. Tsirlin, R. Tang, Y. Dai, and J. Lin, "'Low-resource' text classification: a parameter-free classification method with compressors," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 6810–6828. doi: 10.18653/v1/2023.findings-acl.426.

[4] J. Ma and L. Li, "Data augmentation for chinese text classification using back-translation," *Journal of Physics: Conference Series*, vol. 1651, no. 1, p. 012039, Nov. 2020, doi: 10.1088/1742-6596/1651/1/012039.

[5] D. R. Beddiar, M. S. Jahan, and M. Oussalah, "Data expansion using back translation and paraphrasing for hate speech detection," *Online Social Networks and Media*, vol. 24, p. 100153, Jul. 2021, doi: 10.1016/j.osnem.2021.100153.

[6] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, "Sentiment analysis on social media," in *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012*, IEEE, Aug. 2012, pp. 919–926. doi: 10.1109/ASONAM.2012.164.

[7] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[8] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems*, vol. 60, no. 2, pp. 617–663, Aug. 2019, doi: 10.1007/s10115-018-1236-4.

[9] M. Rahardi, A. Aminuddin, F. F. Abdulloh, and R. A. Nugroho, "Sentiment analysis of COVID-19 vaccination using support vector machine in Indonesia," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, pp. 534–539, 2022, doi: 10.14569/IJACSA.2022.0130665.

[10] S. Kurniawan, R. Kusumaningrum, and M. E. Timu, "Hierarchical sentence sentiment analysis of hotel reviews using the naïve bayes classifier," in *2018 2nd International Conference on Informatics and Computational Sciences, ICICoS 2018*, IEEE, Oct. 2018, pp. 104–108. doi: 10.1109/ICICOS.2018.8621748.

[11] S. Ounacer, D. Mhamdi, S. Ardchir, A. Daif, and M. Azzouazi, "Customer sentiment analysis in hotel reviews through natural language processing techniques," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, pp. 569-579, 2023, doi: 10.14569/IJACSA.2023.0140162.

[12] V. Bidve *et al.*, "Use of explainable AI to interpret the results of NLP models for sentimental analysis," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 35, no. 1, pp. 511–519, Jul. 2024, doi: 10.11591/ijeecs.v35.i1.pp511-519.

[13] O. Chamorro-Atalaya *et al.*, "Supervised learning using support vector machine applied to sentiment analysis of teacher performance satisfaction," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 28, no. 1, pp. 516–524, Oct. 2022, doi: 10.11591/ijeecs.v28.i1.pp516-524.

[14] C. N. Permatasari and A. A. R. Fernandes, "Sentiment analysis and support vector machine one versus one for collectibility classification of bank's house ownership loan," *International Journal of Applied Decision Sciences*, vol. 17, no. 3, pp. 293–312, 2024, doi: 10.1504/IJADS.2024.138193.

[15] F. Firmansyah *et al.*, "Comparing sentiment analysis of indonesian presidential election 2019 with support vector machine and k-nearest neighbor algorithm," in *6th International Conference on Computing, Engineering, and Design, ICCED 2020*, IEEE, Oct. 2020, pp. 1–6. doi: 10.1109/ICCED51276.2020.9415767.

[16] D. A. Kristiyanti, S. A. Sanjaya, V. C. Tjokro, and J. Suhali, "Dealing imbalance dataset problem in sentiment analysis of recession in Indonesia," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 2, pp. 2058–2070, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp2060-2072.

[17] S. Khomsah, A. F. Hidayatullah, and A. S. Aribowo, "Comparison of the effects of feature selection and tree-based ensemble machine learning for sentiment analysis on Indonesian YouTube comments," in *Lecture Notes in Electrical Engineering*, vol. 746 LNEE, 2021, pp. 161–172. doi: 10.1007/978-981-33-6926-9_15.

[18] M. L. F. Martanto and W. Istiono, "Sentiment analysis of M-Paspor app reviews using multinomial naive bayes," *Journal of Logistics, Informatics and Service Science*, vol. 11, no. 10, pp. 311–326, Sep. 2024, doi: 10.33168/JLISS.2024.1017.

[19] O. Chantamuang, J. Polpinij, V. Vorakitphan, and B. Luaphol, "Sentence-level sentiment analysis for student feedback relevant to teaching process assessment," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13651 LNAI, 2022, pp. 156-168. doi: 10.1007/978-3-031-20992-5_14.

[20] P. Chauhan, N. Sharma, and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2601-2627, Feb. 2021, doi: 10.1007/s12652-020-02423-y.

[21] O. Oyebode and R. Orji, "Social media and sentiment analysis: the Nigeria presidential election 2019," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2019*, IEEE, Oct. 2019, pp. 140–146. doi: 10.1109/IEMCON.2019.8936139.

[22] N. Braig, A. Benz, S. Voth, J. Breitenbach, and R. Buettner, "Machine learning techniques for sentiment analysis of COVID-19-related Twitter data," *IEEE Access*, vol. 11, pp. 14778–14803, 2023, doi: 10.1109/ACCESS.2023.3242234.

[23] B. Gaind, V. Syal, and S. Padgalwar, "Emotion detection and analysis on social media." 2019. [Online]. Available: https://arxiv.org/abs/1901.08458

[24] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14–46, Nov. 2015, doi: 10.1016/j.knosys.2015.06.015.

[25] K. H. Manguri, R. N. Ramadhan, and P. R. M. Amin, "Twitter sentiment analysis on worldwide COVID-19 outbreaks," *Kurdistan Journal of Applied Research*, pp. 54–65, 2020.

[26] L. Nemes and A. Kiss, "Social media sentiment analysis based on COVID-19," *Journal of Information and Telecommunication*, vol. 5, no. 1, pp. 1–15, Jan. 2021, doi: 10.1080/24751839.2020.1790793.

[27] T. Liesting, F. Frasincar, and M. M. Trusc, "Data augmentation in a hybrid approach for aspect-based sentiment analysis," in *Proceedings of the ACM Symposium on Applied Computing*, New York, NY, USA: ACM, Mar. 2021, pp. 828–835. doi: 10.1145/3412841.3441958.

[28] S. Cohen, D. Presil, O. Katz, O. Arbili, S. Messica, and L. Rokach, "Enhancing social network hate detection using back translation and GPT-3 augmentations during training and test-time," *Information Fusion*, vol. 99, p. 101887, Nov. 2023, doi: 10.1016/j.inffus.2023.101887.

[29] S. Kurniawan and I. Budi, "Utilizing translation to enhance NLP models in offensive language and hate speech identification," *Jurnal Improsci*, vol. 1, no. 4, pp. 182–197, Feb. 2024, doi: 10.62885/improsci.v1i4.187.

[30] G. X. Luo, Y. T. Yang, R. Dong, Y. H. Chen, and W. B. Zhang, "A joint back-translation and transfer learning method for low-resource neural machine translation," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–11, May 2020, doi: 10.1155/2020/6140153.

BIOGRAPHIES OF AUTHORS

Sandy Kurniawan earned his bachelor's degree in computer science from Universitas Diponegoro and his master's degree in computer science from Universitas Indonesia. He is currently a lecturer in the Department of Informatics, Faculty of Science and Mathematics at Universitas Diponegoro. His research focuses on machine learning, artificial intelligence, and natural language processing. He can be contacted at email: sandy@live.undip.ac.id.

Retno Kusumaningrum earned her bachelor's degree in computer science from Universitas Diponegoro and her master's and doctoral degrees in computer science from Universitas Indonesia. She is currently a lecturer in the Department of Informatics, Faculty of Science and Mathematics at Universitas Diponegoro. As a researcher, she has expertise in computer vision, pattern recognition, machine learning, topic modeling, and natural language processing. She can be contacted at email: retno@live.undip.ac.id.

Priyo Sidik Sasongko received his bachelor's degree in mathematics from Universitas Diponegoro and a master's degree in computer science from Universitas Gadjah Mada. He is currently a lecturer in the Department of Informatics, Faculty of Science and Mathematics at Universitas Diponegoro. His research focuses on machine learning, fuzzy logic, and artificial intelligence, Japan. He can be contacted at email: priyoss_undip@yahoo.co.id.