Buletin Ilmiah Sarjana Teknik Elektro
Vol. 5, No. 2, June 2023, pp. 303-311
ISSN: 2685-9572, DOI:10.12928/biste.v5i2.7416

Sentiment Analysis of Electronic System Provider (PSE) Method Using Support Vector Machine

Turiyan, Nurirwan Saputra, Ahmad Riyadi, Meilany Nonsi Tentua
Informatics Research Program, Faculty of Science and Technology, PGRI Yogyakarta University

ARTICLE INFORMATION

Article History:

Submitted 20 January 2023
Revised 19 April 2023
Accepted 13 July 2023

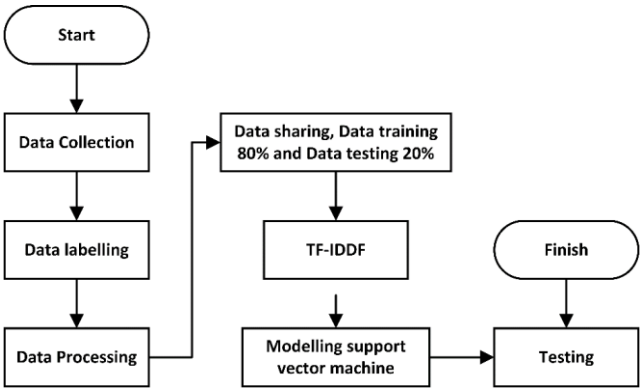
Keywords:

PSE;
Support Vector Machine;
Preprocessing;
Facebook;
Streamlet

Corresponding Author:

Turiyan,
Universitas PGRI Yogyakarta,
Jln ikip pgri 1, Sonosewu,
No.117, Ngestiharjo, Indonesia.
Surel/Email:
turiyan295@gmail.com

ABSTRACT



The study discusses the policy issued by the Indonesian Ministry of Communication and Information on Electronic System Providers (PSE), with a special emphasis on the impact of the policy that has triggered public reactions on social media platforms. Based on the regulations that have been made, it has an impact on the public who believe that the regulation is made to take away access freedom and data privacy on digital systems used by the public. Sentiment analysis can see public opinion on the policy issued by the Ministry of Communication and Information. The data used is taken from social media which is labeled "positive" and "negative". The research method used in sentiment analysis is the Support Vector Machine method. The contribution of this research is to determine the boundary line between the two classes in this research data. The result of the F^1 score can be measured in the model, with a achieved value of 75%.

Citation Document:

Turiyan, N. Saputra, A. Riyadi, M. N. Tentua, "Sentiment Analysis of Electronic System Provider (PSE) Method Using Support Vector Machine," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 5, no. 2, pp. 303-311, 2023, DOI: 10.12928/biste.v5i2.7416.

1. INTRODUCTION

At this time, the Ministry of Communication and Informatics of the Republic of Indonesia issued a policy for Electronic System Operators (PSE), which is currently being widely discussed by the people of Indonesia. The PSE policy that has been implemented by the government has received responses and opinions from the public through social media. Some parties who disagree with the regulations issued by the Ministry of Communication and Informatics consider that blocking sites such as PayPal, Valve Corp, and Yahoo violates privacy and freedom of expression because registering digital platforms endangers user data [1].

The policy of electronic system administrators is very important for all people because electronic system administrators are information as a series of electronic devices and procedures that function to prepare, collect, manage, analyze, store, display, announce, send or disseminate electronic information and also electronic system administrators are very useful to guarantee data collection and maintain digital system security [2]. Electronic system organizers are carried out with the aim that sites and applications that are not registered immediately register on the PSE-Kominfo website that has been implemented, in the Regulation of the Minister of Communication and Informatics Number 5 of 2022 [3].

The PSE policy has received a lot of opinion from the public that it provides the first step in upholding Indonesia's digital sovereignty and can also create a safe digital space for users and entrepreneurs because all the provisions are regulated by law [4]. In article 1 paragraph (4) PP 71/2012 every electronic system operator, state administrator, business entity, and the public who provide, or operate electronic systems to users of electronic systems for the needs of themselves and other parties. So from the article that has been mentioned that electronic system operators are very important for Indonesia to be able to utilize electronic systems by state administrators, individuals, business entities, or the public which can be carried out for public or non-public services [5].

The implementation of Electronic Systems (PSE) is carried out by the government for citizens by conveying or operating electronic systems to users to obtain security and convenience for the community. Applications and sites used by the public must be registered on the PSE Web so that they can be managed by the Ministry of Communication and Information of the Republic of Indonesia. Applications that are often used by people who have registered with the electronic system operator policy are Facebook, Google, WhatsApp, Tiktok, Netflix, Gojek, and others.

Public complaints against Kominfo's policy which issued Electronic System Operator (PSE) regulations, but the community views PSE policies as being able to take away freedom of access and data privacy in digital systems. This policy also blocks applications and sites that make people afraid because PSE contains threats to privacy and freedom rights on social media, one of which is Facebook [6]. Facebook is one of the popular social media applications among the general public, where users discuss and write about everyday life or talk about various topics that are currently being discussed [7]. Facebook users can provide opinions through comments, so comments can be in the form of opinions, suggestions, or criticisms of the issues being discussed, one of which is the Electronic System Operator (PSE) policy [8].

Public opinion can be processed to analyse sentiment towards a person's opinion by collecting data effectively and efficiently. From the opinion obtained from Facebook comments on one of the Indonesian Ministry of Communication and Informatics accounts which contains information about Electronic System Operators (PSE) which was uploaded on August 2, 2022 and in this information, there are comments used to analyse sentiment on Facebook social media. Sentiment analysis has been widely used in previous studies, most of which used comment data taken from social media such as YouTube, Facebook, Instagram, Twitter, and many more [9]. Sentiment analysis is also called data mining, data mining is a process in the form of a collection of knowledge data that is taken manually which contains opinions, sentiments and emotions expressed in text [10].

Community comments will be taken through a scraping process, scraping by extracting data or information from the intended site in retrieving datasets on the site or websites [11]. With the amount of data taken as many as 1146 data comments that go through the labeling process are given guidelines or labels containing numbers that will facilitate labeling. This number is a label that contains positive information for number 1 and negative for number 2. This results in 366 positive sentiments and 792 negative sentiments related to Electronic System Operator policies. The cleaning process is used to clean unnecessary words or punctuation from comment data taken from Facebook.

This study carried out the data pre-processing process. The pre-processing process has several stages, Remove of punction, Tokenization, Stopword, Stemming, dan Case Folding. The pre-processing process is expected to comment on the data to be processed to produce clean data and can be used to see the results of the accuracy of the data to be processed [12]. In this study, the data was weighted using TF-IDF and the research method used for classification was using the Support Vector Machine which is a suitable method for this calculation. In this study will be tested using a confusion matrix to determine the value of accuracy, precision, recall and F^1 score.

2. METHODS

The method that will be used in this research is to take a dataset about "Public comments on Electronic System Operators (PSE) policies" which predict the accuracy of people's responses using the Support Vector Machine method which is carried out through several stages, starting with data collection through a scraping process. After the data is collected, a labeling process is carried out which gives positive and negative comment values. After the labeling process that assigns a class to each comment, the preprocessing stage is performed, which includes several steps: remove of punctuation, tokenization, stopword, stemming, dan case folding. After the preprocessing stage, the data is weighted using TF-IDF and processed into a Support Vector Machine model Figure 1.

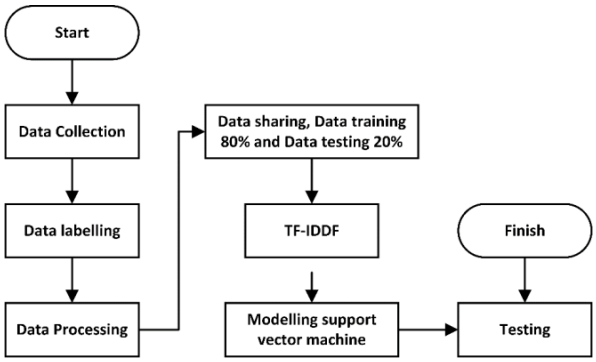


Figure 1. Methodology Flowchart

2.1. Data Collection

This study first collected data by taking comments from the Indonesian Ministry of Communication and Informatics Facebook fan-page which was uploaded on August 2, 2022 and has the most comments until August 2022. The sample data was collected using beautiful soup, with a total of 1446 data collected randomly about public response to the Electronic System Operator Policy.

2.2. Data Labelling

Data labelling is done manually by giving a "positive" and "negative" value to each comment with the aim of knowing what opinion is in the dataset that is appropriate or not in accordance with the research theme and this process will be done manually [13]. The labelling given to the dataset used includes:

- Positive word labelling is given a number (1): Explanation of comments containing community approval of the policies provided by the Ministry of Communication and Information and support for policies made as well as discussion outside the topic of electronic system operator policies.
- Labelling of negative words is numbered (2): Explanation of comments that provide negative responses to policies made by the Ministry of Communication and Information and have an adverse impact on readers which will trigger favourable opinion responses to Electronic System Operator (PSE) policies made by Kominfo.

In the labelling process itself, it is necessary to pay attention to the labelling guidelines, so that there are no errors in labelling each comment data shown in Table 1.

Table 1. Labelling Guidelines

Label	Explanation
Positive	Explanation of comments containing public approval of the policies provided by the Ministry of Communication and Informatics and support for the policies made and also discussion outside the topic of electronic system implementation policies
Negative	Explanation of comments that give negative responses to policies made by the Ministry of Communication and Information and have a negative impact on readers which will trigger favourable opinion responses to Electronic System Operator (PSE) policies made by Kominfo

2.3. Data Pre-processing

In this research, the pre-processing stage will be carried out which will make clean sentences with several stages, namely as follows.

Remove of punction, punctuation removal is a process used to remove punctuation marks, numbers, and other characters used in sentences [14]. At this stage of removing punctuation marks to facilitate the next pre-processing stage.

Tokenization, tokenization is the process of breaking sentences into tokens, where each token is a separate word in the dataset based on punctuation marks such as commas (,) or dots (.) and also tokenization will break down the words used for the word weighting process [15]. A unigram token is a token that separates sentences into one word [16]. With the unigram token, the vocabulary used will have a clearer value [17]. The technique process will facilitate the next process or the word weighting process which will improve accuracy for the better.

Stopword removal, this process will reduce the number of words that will not be used in the data [18]. This step is carried out after tokenization which will determine words that are not included in the conjunctions used in the master literature by removing the conjunctions contained in the master literary library [19] . Important words from the token results are retrieved using a stoplist algorithm (eliminating less important words) or wordlist (saving important words) [20], as in Figure 2.

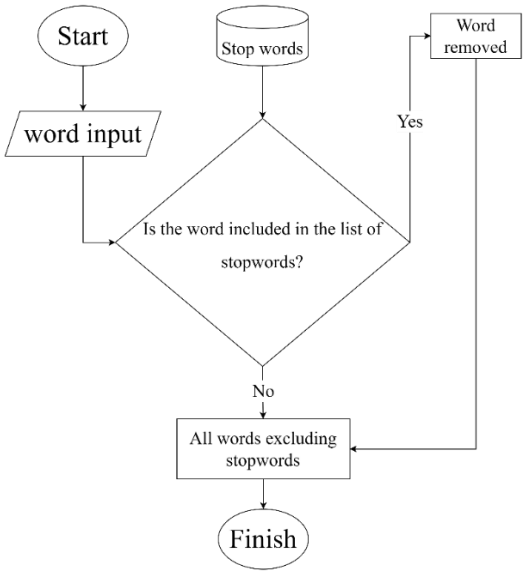


Figure 2. Stopword Removal Flowchart

Stemming is the process of reducing words to their basic form by removing the endings to reduce the number of words and increase the accuracy of the analysis [18]. In Indonesian it is used to separate words into appropriate verb forms in the Big Indonesian Dictionary (KBBI). This process will use the Literature Master library which is a simple python library that provides Indonesian word stemming [21][22]. The purpose of stemming is to remove the inflection of words, this will greatly affect the calculation of the number of words that will be carried out during TF-IDF [23] .

Case folding performed in the case folding process, uppercase is a process of unfirming words in comments, which aims to reduce waste of data storage for use in the classification process [24]. Case folding aims to run the calculation process optimally with word uniformity by changing the data format to uppercase (UPPERCASE) [25].

2.4. Classification by method Support Vector Machine

This research classifies using the Support Vector Machine method to determine the value of accuracy, precision, recall and f1score. Support Vector Machine is a method that determines the dividing line between two classes or hyperplanes in this study [26]. The way the Support Vector Machine method works is a concept where this method tries to find the best hyperplane that functions as a dividing line between two classes and also the Support Vector Machine method uses kernel functions to be used on data that is not linearly separated [27]. This study will use testing with validation matrix confusion.

The prediction model that uses the Support Vector Machine method classifies the two classes by finding the best hyperplane that serves as the dividing line between the two classes. Before finding the hyperplane, it is necessary to determine the margin distance that will be used for the distance between vectors so that there is no data noise during the classification process. To do this step, the first step is to determine the margin distance. The following is the equation for finding the width of the margin between vectors to get the classification between the two classes (1). $||w||^2$ is the Vector weight, w_1^2 is the Weight value for positive data class and w_2^2 is the Weight value for negative data class.

$$\frac{1}{2}||w||^2 = \frac{1}{2}(w_1^2 + w_2^2) \quad (1)$$

After obtaining the results of the widest distance for vector classification, the next step is to determine the placement of the hyperplane so that it can determine the centre line between classes so that noise does not occur during the classification process between classes. The following is the equation for determining the hyperplane line (2). yi is the Data class i, W is the Support Vector Machine weight value that is perpendicular to the hyperplane, Xi is the i-th data, b is the Biased value and i is the Data.

$$yi(W.X_i + b) \geq 1, \qquad i = 1,2,3,\ldots..,n \quad (2)$$

After that, get the margin results and get the hyperplane placement results, then do the hyperplane visualization. In this study, a C value of 3 was used to control the margin used to produce the classification of research data with an accuracy of 75%. This research also carried out the process of pre-processing, modelling, and testing as many as 2 labels. In the pre-processing process is done with remove of punctuation, tokenization, stopword, stemming case folding. Then, the Support Vector Machine method will be modelled and the data will be divided into 80% training data and 20% test data.
The Support Vector Machine classification was tested using the confusion matrix method. To calculate accuracy, the equation (3).

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (3)$$

To calculate precision, the equation (4).

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

To calculate recall, the equation (5).

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

F[1]score is a measure of model accuracy that takes into account precision and recall. Calculated using the equation (6).

$$F^1\ score = \frac{2 \times Precision \times Recall}{Precision \times Recall} \quad (6)$$

Note Information is shown in Table 2.

Table 2. Note Equation Confusion Matrix

Symbol	Description
TP (True Positive)	The predicted result is positive and the actual result is also positive
TN (True Negative)	The predicted result is negative and the actual result is also negative
FP (False Positive)	The predicted result is positive but the actual result is negative
FN (False Negative)	The predicted result is negative but the actual result is positive

3. RESULT AND DISCUSSION

3.1. Data Labelling

With the amount of data taken as many as 1158 data comments that go through the labelling process are given guidelines or labels containing numbers that will facilitate labelling. The number is a label that contains negative information for number (2), and positive for number (1). This resulted in 792 negative sentiments and 366 positive sentiments related to Electronic System Operator (PSE) policies. The Clean-up Process is used to clean unnecessary words or punctuation from the comment data retrieved from Facebook.
Then the results of data labelling are shown in Table 3.

Table 3. Data Labelling

No	Comments	Label
1	"Tapi yang digagas kominfo ini gak bermanfaat semua lagi"	Negative
2	"Subhanallah. Semoga kebijakan yang dibuat bermanfaat untuk masyarakat. Semoga PSE ini aman untuk kenyaman dalam sistem elektronik bagi masyarakat. Aamiin"	Positive

This study contains comments that are negative, and positive as shown in Figure 3.

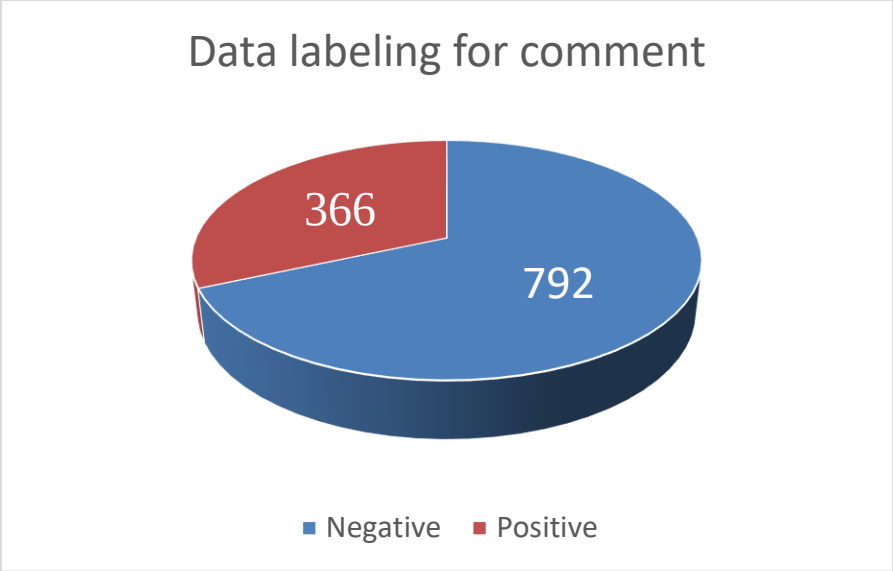


Figure 3. Graph for Data Labelling for Comment

3.2. Data Pre-processing

The Clean-up process is used to clean unnecessary words or punctuation from the comment data retrieved from Facebook. The above process will produce data in Table 4.

Table 4. Example Data Pre-processing

Stages	Result
Before pre-processing	tapi yang digagas kominfo ini gak bermanfaatnya semua anjer
Remove of punctuation	tapi yang digagas kominfo ini gak bermanfaatnya semua anjer
Tokenization	'tapi', 'yang', 'digagas', 'kominfo', 'ini', 'gak', 'bermanfaatnya', 'semua', 'anjer'
Stopword removal	'digagas', 'kominfo', 'gak', 'bermanfaatnya', 'semua', 'anjer'
Stemming	'gagas', 'kominfo', 'gak', 'manfaat', 'semua', 'anjer'
Case folding	'GAGAS', 'KOMINFO', 'GAK', 'MANFAAT', 'SEMUA', 'ANJER'

3.3. Testing

After testing, accuracy, precision, recall, and F^1 score values will be generated from the confusion matrix shown in Table 5.

Table 5. Result Confusion Matrix

	POSITIVE	NEGATIVE
POSITIVE	41	42
NEGATIVE	15	148

The percentage of accuracy produced by the Support Vector Machine method is 75%. Percentage of precision, recall, and F^1 scores for each label can be seen in Table 6. These results indicate that the classifier is able to correctly classify 75% of the data based on the labels given in the study. The values for precision, recovery and F^1 scores in Table 6 provide more detailed information about the performance of the classifier for each label. Precision measures the proportion of true positive predictions among all the positive predictions made by the classifier. Remember to measure the proportion of predicted true positives among all actual positive events. The F^1 score is a measure that combines precision and memory, this is the harmonic average of precision and memory shown in Table 6.

Table 6. Percentage Result of Each Label

Label	Precision	Recall	F^1score	Accuracy
Positive	58%	33%	42%	
Negative	78%	91%	84%	
Support Vector Machine				75%

The prediction model of this study chose Unigram tokenization with an accuracy of 70% and the largest percentage value was obtained for negative labels due to the large amount of data labelled negatively during the data labelling process, and resulted in 78% precision, 91% recall, and 84% F^1 score. The accuracy results obtained were 75%, while using the Random Forest Classifier method it was obtained 60%, because this study used 2 labels, namely positive and negative, while the Random Forest Classifier used 3 labels, namely negative, positive and neutral. This study also tested the 2×2 confusion matrix, because it uses 2 labels, while the Random Forest Classifier method uses 3 labels and uses the 3x3 confusion matrix test.

4. CONCLUSIONS

The classification results using the Support Vector Machine method on comments posted on August 2, 2022, on the Facebook fan page of the Indonesian Ministry of Communication and Information Technology that received public opinions or feedback on the policy of electronic system providers that blocked some unregistered applications and performed scraping using Python programming language and pre-processing regarding Electronic System Policy (PSE) comments. This study produced a classification with a total of 1128 pre-processed data, and the majority of the classification was negative. This study has good accuracy, precision, recall, and F^1 score values. The classification results using several pre-processing steps with the Support Vector Machine method had an accuracy of 75%, with the highest percentage value obtained in the negative label, namely precision of 78%, recall of 91%, and F1 score of 84%, which will be implemented on a website to be used as an evaluation material for the government. This study shows that the use of pre-processing steps and the Support Vector Machine method is effective in classifying comments on the Electronic System Provider (PSE) policy in Indonesia.

5. FUTURE WORKS

The future of this research suggests several recommendations for future studies. The research is expected to implement automatic scraping that collects public comments on government policies and processes them with several pre-processing stages. The research is expected to be improved with better methodology to achieve higher levels of accuracy. Additionally, the research is expected to be continued with a comparison of other related sentiment analysis methods.

REFERENCES

[1] A. Syaifuddin *et al.*, “Analisis Sentimen Pada Sosial Media Tentang Implementasi Kebijakan Pse Kominfo Menggunakan Algoritme Lexicon,” *Pros. SEMASTEK 2022*, vol. 1, no. 1, pp. 8–15, 2022, https://doi.org/10.36815/semastek.v1i1.2.

[2] G. Kane, "The technology fallacy: people are the real key to digital transformation," *Research-Technology Management*, vol. 62, no. 6, pp. 44-49, 2019, https://doi.org/10.1080/08956308.2019.1661079.

[3] Martin, T., Karopoulos, G., Hernández-Ramos, J. L., Kambourakis, G., & Nai Fovino, I. (2020). Demystifying COVID-19 digital contact tracing: A survey on frameworks and mobile apps. *Wireless Communications and Mobile Computing*, *2020*, 1-29, 2020, https://doi.org/10.1155/2020/8851429.

[4] A. Itten and N. Mouter, "When digital mass participation meets citizen deliberation: combining mini-and maxi-publics in climate policy-making," *Sustainability*, vol. 14, no. 8, p. 4656, 2022, https://doi.org/10.3390/su14084656.

[5] I. Veza *et al*., "Electric vehicles in malaysia and indonesia: opportunities and challenges," *Energies*, vol. 15, no. 7, p. 2564, 2022, https://doi.org/10.3390/en15072564.

[6] J. Kittur, "Measuring the Programming Self-Efficacy of Electrical and Electronics Engineering Students," in *IEEE Transactions on Education*, vol. 63, no. 3, pp. 216-223, 2020, https://doi.org/10.1109/TE.2020.2975342.

[7] S. Pudaruth, S. Moheeputh, N. Permessur, and A. Chamroo, “Sentiment Analysis from Facebook Comments using Automatic Coding in NVivo 11,” *ADCAIJ Adv. Distrib. Comput. Artif. Intell. J.*, vol. 7, no. 1, pp. 41–48, 2018, https://doi.org/10.14201/ADCAIJ2018714148.

[8] N. Kewsuwun and S. Kajornkasirat, “A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2829–2838, 2022, https://doi.org/10.11591/ijece.v12i3.pp2829-2838.

[9] M. Singh, A. K. Jakhar and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 33, 2021, https://doi.org/10.1007/s13278-021-00737-z.

[10] M. Ben Hajhmida and O. Oueslati, “Predicting mobile application breakout using sentiment analysis of Facebook posts,” *J. Inf. Sci.*, vol. 47, no. 4, pp. 502–516, 2021, https://doi.org/10.1177/0165551520917099.

[11] S. Chaudhari, R. Aparna, V. G. Tekkur, G. L. Pavan and S. R. Karki, "Ingredient/Recipe Algorithm using Web Mining and Web Scraping for Smart Chef," *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pp. 1-4, 2020, https://doi.org/10.1109/CONECCT50063.2020.9198450.
[12] H. Samy, A. Helmy and N. Ramadan, "Aspect-based Sentiment Analysis of Mobile Apps Reviews using Class Association Rules and LDA," *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 183-189, 2021, https://doi.org/10.1109/ICICIS52592.2021.9694242.
[13] Y. Wang, Q. Chen, J. Shen, B. Hou, M. Ahmed, and Z. Li, "Aspect-level sentiment analysis based on gradual machine learning," *Knowledge-Based Syst.*, vol. 212, p. 106509, 2021, https://doi.org/10.1016/j.knosys.2020.106509.
[14] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini and D. Risdiansyah, "Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization," *2020 Fifth International Conference on Informatics and Computing (ICIC)*, pp. 1-7, 2020, https://doi.org/10.1109/ICIC50835.2020.9288655.
[15] W. Bourequat and H. Mourad, "Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine," *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 1, pp. 36–44, 2021, https://doi.org/10.25008/ijadis.v2i1.1216.
[16] J. Sangeetha and U. Kumaran, "A hybrid optimization algorithm using BiLSTM structure for sentiment analysis," *Meas. Sensors*, vol. 25, no. 100619, p. 100619, 2023, https://doi.org/10.1016/j.measen.2022.100619.
[17] M. K. Sohrabi and F. Hemmatian, "An efficient preprocessing method for supervised sentiment analysis by converting sentences to numerical vectors: a twitter case study," *Multimedia tools and applications*, *78*(17), 24863-24882, 2019, https://doi.org/10.1007/s11042-019-7586-4.
[18] A. W. Pradana and M. Hayaty, "The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, pp. 375–380, 2019, https://doi.org/10.22219/kinetik.v4i4.912.
[19] L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in *IEEE Access*, vol. 8, pp. 23522-23530, 2020, https://doi.org/10.1109/ACCESS.2020.2969854.
[20] S. Ferilli, "Automatic multilingual stopwords identification from very small corpora," *Electronics*, vol. 10, no. 17, p. 2169, 2021, https://doi.org/10.3390/electronics10172169.
[21] N. Saputra, K. Nurbagja, and T. Turiyan, "Sentiment Analysis of Presidential Candidates Anies Baswedan and Ganjar Pranowo Using Naïve Bayes Method," *J. Sisfotek Glob.*, vol. 12, no. 2, p. 114, 2022, https://doi.org/10.38101/sisfotek.v12i2.552.
[22] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pp. 1-5, 2019, https://doi.org/10.1109/ICIC47613.2019.8985884.
[23] A. Septiana Sani and I. Budi, "Sentiment Analysis on Jurassic Park Development in the Komodo Conservation Area," *Int. Conf. Inf. Sci. Technol. Innov.*, vol. 1, no. 1, pp. 105–108, 2022, https://doi.org/10.35842/icostec.v1i1.18.
[24] N. Kankanamge, T. Yigitcanlar, A. Goonetilleke M. Kamruzzaman, "Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets," *International journal of disaster risk reduction*, vol. 42, p. 101360, 2020, https://doi.org/10.1016/j.ijdrr.2019.101360.
[25] R. Kosasih and A. Alberto, "Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier," *Ilk. J. Ilm.*, vol. 13, no. 2, pp. 101–109, 2021, https://doi.org/10.33096/ilkom.v13i2.721.101-109.
[26] V. Ramanathan and T. Meyyappan, "Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism," *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*, pp. 1-5, 2019, https://doi.org/10.1109/ICBDSC.2019.8645596.
[27] N. S. Mohd Nafis and S. Awang, "An Enhanced Hybrid Feature Selection Technique Using Term Frequency-Inverse Document Frequency and Support Vector Machine-Recursive Feature Elimination for Sentiment Classification," in *IEEE Access*, vol. 9, pp. 52177-52192, 2021, https://doi.org/10.1109/ACCESS.2021.3069001.

AUTHOR BIOGRAPHY

Turiyan is currently completing his undergraduate education in the informatics study program at PGRI Yogyakarta University in 2023.

Nurirwan Saputra completed his undergraduate education in the informatics engineering study program at the Islamic University of Indonesia and completed his master's education in the electrical engineering study program at Gajah Mada University. Currently author 2 is a permanent lecturer in the informatics study program at the Yogyakarta PGRI University. His research area is sentiment analysis.

Ahmad Riyadi completed his undergraduate education in the mathematics study program at Gajah Mada University and completed his master's education in the computer science study program at Gajah Mada University. Currently author 3 is a permanent lecturer in the informatics study program at PGRI Yogyakarta University. His research area is computer science.

Meilany Nonsi Tentua completed his undergraduate education at Gajah Mada University's mathematics study program and completed his master's degree at Gajah Mada University's electrical engineering study program. Currently author 2 is a permanent lecturer in the informatics study program at the Yogyakarta PGRI University. His research areas are Artificial Intelligence, Machine Learning, Data Science and Natural Language Processing.