International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 15, No. 2, June 2026, pp. 523~534
ISSN: 2252-8776, DOI: 10.11591/ijict.v15i2.pp523-534 ❒ 523

Enhancing support vector machine performance using particle swarm optimization for sentiment analysis

Christofer Satria[1], Anthony Anggrawan[2], Peter Wijaya Sugijanto[3], Husain Husain[2], I Nyoman Yoga Sumadewa[1], Victoria Cynthia Rebecca[3]
[1]Department of Visual Communication Design, Faculty of Art and Design, Bumigora University, Mataram, Indonesia
[2]Department of Computer Science, Faculty of Engineering, Bumigora University, Mataram, Indonesia
[3]Department of Medical, Faculty of Medicial and Health Sciences, Bumigora University, Mataram, Indonesia

Article Info

Article history:

Received Jul 9, 2025
Revised Dec 24, 2025
Accepted Mar 30, 2026

Keywords:

Data mining
Machine learning
Particle swarm
Sentiment analysis
Social media
SVM

ABSTRACT

Recently, social media has established itself as a leading platform in various sectors. Meanwhile, text extraction and sentiment analysis classification have attracted significant attention in research. Regrettably, traditional sentiment analysis often falls short of accurately capturing sentiment nuances. At the same time, machine learning has enabled more effective sentiment analysis, data mining, and classification, as well as the development of models that incorporate artificial intelligence. Therefore, the purpose of this study is to optimize sentiment analysis of public opinion in social media regarding Grand Prix motorcycle racing (MotoGP) and World Superbike (WSBK) events using machine learning and an optimized machine learning method. This study applies the support vector machine (SVM) machine learning method and enhances its performance through optimization by integrating it with the particle swarm optimization (PSO) algorithm. This study found that the SVM method achieved 80.15% accuracy, 75.63% recall, and 76.89% F1-score. In contrast, the SVM method combined with PSO achieves accuracies of 81.82%, 79.9%, and 79.62% for recall, precision, and F1-score, respectively, in classifying the sentiment of sporting events. The implications suggest that applying Hybrid SVM with PSO significantly enhances classification accuracy in sentiment analysis.

Corresponding Author:

Christofer Satria
Department of Visual Communication Design, Faculty of Art and Design, Bumigora University
Mataram, Indonesia
Email: chris87@universitasbumigora.ac.id

1. INTRODUCTION

In recent eras, numerous users have turned to social media to share their thoughts and express their opinions [1], [2]. Social media has established itself as a dominant public user-generated media platform across various fields [3], including tourism [4]. In essence, public opinion data sets on social media provide valuable insights into users' sentiments through sentiment analysis [5]. Sentiment analysis is the classification of the polarity of emotions, attitudes, or public opinions [6]–[8] regarding an object or subject [7], which classifies the object into positive, negative, and neutral sentiment categories [8]. Sentiment analysis has received considerable attention in research for extracting and classifying sentiment polarities [7]. In short, the primary benefit of sentiment analysis is to determine public reactions and opinions toward a news issue [9]–[11], and the results provide a strong basis for making more accurate and precise decisions [12]. Unfortunately, traditional sentiment analysis often struggles to extract sentiment accurately [4]. As a consequence, sentiment analysis requires a complex classification model [7].

In the meantime, advances in machine learning have enabled the performance of sentiment analysis on large datasets, yielding more effective results [13]. Machine learning is a valuable tool for classifying or predicting class membership, and it plays a significant role in data mining [14]. Moreover, machine learning methods can perform sentiment analysis on complex classification models [7]. It also serves as an effective and accurate method for building models or systems with artificial intelligence [14]. In short, the need to apply complex models for sentiment analysis or to distinguish classes or opinion labels in textual documents leads to the use of machine learning methods [6]. Machine learning is increasingly recognized as a valuable tool for data mining, enabling high-accuracy classification of datasets [15], [16]. That is why this study aims to optimize sentiment analysis on public opinion using the support vector machine (SVM) machine learning method and the SVM optimized with the particle swarm optimization (PSO) method.

SVM is a supervised machine learning [8]. SVM is a machine learning method for classifying or categorizing data [17]. The SVM is a widely used machine learning technique applied in various fields of scientific research [18]. The SVM machine learning method has numerous applications in classifying datasets [19] and is well known as a classification method [19]. The SVM is an advanced algorithm known for its high prediction accuracy [20]. The SVM constitutes the most popularly used machine learning algorithm for sentiment analysis [8]. In the meantime, sporting events have become a rapidly growing trend, as they have a positive impact on societal economic growth [21]. Sporting events not only serve as events but also contribute to Indonesia’s tourism [21]. These events may be why research on support events has garnered significant attention [22], [23]. Unfortunately, research on sporting events is still limited [24]. In the meantime, previous research [25] indicates that using mixed methods can enhance performance. That is why this paper aims to perform sentiment analysis of public opinion on sports events using the SVM method and an SVM optimized with the PSO algorithm.

Table 1 compares our work with previous related studies. Although earlier research includes various studies on sentiment analysis and methodological approaches, our study differs and is not a duplication or a case of plagiarism of existing work. Our study differs from previous studies [6], [8], [12], [21], [26]–[34] (see Table 1). A comparison of this research with related previous studies reveals notable differences in type, focus, methods, results, and subjects. It means that this study offers a novelty that earlier researchers have not explored. In essence, prior research has identified a significant issue. Shami *et al.* [26] and Houssein *et al.* [27] characterize the PSO method as highly popular in the existing literature. Additionally, Shami *et al.* [26] and Houssein *et al.* [27] provide a comprehensive overview of PSO. In contrast, the previous research by Shami *et al.* [26] and Houssein *et al.* [27] used a literature review focused on PSO, which differs from our approach. This study proposes a classification model that uses the SVM method in conjunction with the PSO algorithm. Meanwhile, Sogas *et al.* [21] reviewed the economic and social impacts of sporting events using a cost-benefit analysis. Sogas *et al.* [21] research is purely descriptive and does not propose a sentiment analysis model using machine learning methods, unlike our study. The similarity between Sogas *et al* [21] and our research lies only in the research object: both discuss sporting events.

A study by Bordoloi and Biswas [28] reviewed the existing literature on sentiment analysis research conducted by previous scholars. The goal is to describe various systematic techniques and methods for developing sentiment analysis models. A limitation of prior research is that it is a review study and does not propose a sentiment analysis model based on machine learning, unlike the research presented in our article. The differences between previous research and our study lie in the research objectives, categories, and machine learning methods employed. A paper by Rodríguez-Ibánez *et al.* [29] provides a review of sentiment analysis techniques and their applications across various industries. A study by Rodríguez-Ibánez *et al.* [29] is not a study proposing a sentiment analysis model, nor is it an optimization study, as in our article. The similarity between [29] and our research lies solely in the focus on the object of sentiment analysis research. Abiola *et al.* [30] conducted a sentiment analysis on Twitter related to the coronavirus pandemic. Abiola *et al.* [30] utilized TextBlob and a Valence-Aware Dictionary in text mining. In contrast, our study uses term frequency-inverse document frequency (TF-IDF) in text mining. Abiola *et al.* [30] did not optimize the method’s performance for sentiment analysis classification. Denecke and Reichenpfader [31] on individual clinical health sentiment analysis was conducted by reviewing previous studies. The study by Bordoloi and Biswas [28] is a review article and does not propose a sentiment analysis model based on machine learning to improve performance. The research by [31] and our article differ in terms of research methods and objectives

Singgalen [32] analyzed the motor sport event using K-nearest neighbor (KNN) and SVM algorithms. The results demonstrate that the KNN algorithm effectively classifies sentiment analysis. Research conducted by Singgalen [32] is closely related to the sentiment analysis of sports events discussed in our article. However, the difference lies in [32] not incorporating SVM optimization via the PSO approach, unlike our study. The advantage of our research is that it proposes optimizing the SVM method for sentiment analysis of sports events using the PSO algorithm, compared to the study by Singgalen [32]. Research by Pribadi *et al.* [33] employed the naïve Bayes method using data from 1,144 reviews of tourist attractions in Mandalika, comprising 1,033 positive reviews and 111 negative reviews. A paper by Pribadi *et al.* [33] categorizes tourists' opinions of Mandalika tourist attractions into positive or negative sentiments. This study distinguishes itself from our article by utilizing machine learning techniques for sentiment analysis classification [33]. Our research focuses on different sentiments, whereas Pribadi *et al.* [33] research analyzes sentiments related to tourism. Meanwhile, a study by Anggrawan *et al.* [12] proposed a sentiment analysis model that utilizes the KNN and SVM machine learning methods. While both [12] research and our study focus on sentiment analysis using machine learning techniques for sentiment classification and compare the performance of two machine learning methods, Anggrawan *et al.* [12] research does not include a comparison between the SVM and the SVM combined with the PSO method. The research by Anggrawan *et al.* [12] is not research on sentiment classification using machine learning methods that have improved (optimized) their accuracy. On the contrary, the research in this article focuses on sentiment classification using machine learning methods, achieving improved accuracy.

Lamani *et al.* [34] proposes a system for effectively detecting human actions using an SVM machine learning classifier. The focus of Lamani *et al.* [34] research is on different objects that differ from those in this article. Additionally, Lamani *et al.* [34] research does not verify the accuracy of the sentiment analysis results or the system's recognition of human actions. Lamani *et al.* [34] Research is not about optimizing the performance of the SVM method, as in our research article. In the meantime, a paper by Chourasiya *et al.* [6] reviews various approaches to sentiment analysis, including machine learning methods such as Vader, Roberta, Naïve Bayes, and SVM. This prior research is a literature review and does not focus on a system. A paper by Zhao and Yang [8] employed the SVM method for text emotion classification. It proposed a multilevel SVM-based emotion classification model to categorize cultural inheritance tendencies in in-text comments on web networks. Research by Zhao and Yang [8] is similar to our study, as it applies machine learning to propose a sentiment analysis classification model. The difference lies in the research object, as the previous research is not related to MotoGP and WSBK objects. Additionally, the earlier research did not test the performance of SVM optimization with swarm particles, unlike the research presented in this article.

There has been insufficient exploration of integrating the Vader Lexicon method with the SVM algorithm, particularly when optimized using PSO. Without these optimization techniques, feature selection often remains suboptimal, leading to lower accuracy in sentiment classification. The use of the Vader Lexicon and SVM, either separately or without a proper feature optimization strategy, can cause the model to struggle to distinguish complex sentiment polarities, especially with ambiguous or multi-meaning text data. Therefore, the combination of these three approaches has the potential to provide an innovative solution to overcome the limitations of previous research. In essence, this research serves as an example of a performance optimization model for a machine learning method using PSO in science, and the results contribute to the community of observers and organizers regarding public sentiment (positive, neutral, or negative) toward sporting events (MotoGP and WSBK).

Table 1. Comparison of prior related works with our study

Research authors	Research type	Research focus	Using method		Method performance		Sporting event research
			SVM	PSO	SVM	SVM with PSO	
Houssein *et al.* [27]	Literature review	Review PSO	No	Yes	None	None	No
Sogas *et al.* [21]	Descriptive analysis	Cost-benefit	No	No	None	None	Yes
Shami *et al.* [26]	Literature review	Review PSO	No	Yes	None	None	No
Bordoloi *et al.* [28]	Literature review	Sentiment	No	No	None	None	No
Rodríguez-Ibánez *et al.* [29]	Literature review	Sentiment	No	No	None	None	No
Abiola *et al.* [30]	Proposed system	Sentiment	No	No	None	None	No
Denecke *et al.* [31]	Literature review	Sentiment	No	No	None	None	No
Singgalen [32]	Analysis	Sentiment	Yes	No	87.54%	None	Yes
Pribadi *et al.* [33]	Analysis	Sentiment	No	No	None	None	No
Anggrawan *et al.* [12]	Analysis	Sentiment	Yes	No	71%	None	No
Lamani *et al.* [34]	Analysis	Recognition	Yes	No	None	None	No
Liu *et al.* [6]	Literature review	Sentiment	No	No	None	None	No
Zhao and Yang [8]	Analysis	Sentiment	Yes	No	85%	None	No
Our research	Optimizing	Sentiment	Yes	Yes	80.15%	81.82%	Yes

2. RESEARCH METHOD

Figure 1 illustrates that the research process commenced with text preprocessing, specifically tourist reviews related to the implementation of MotoGP and WSBK events in Mandalika. The collected data undergoes further text preprocessing, including case folding, text cleaning, normalization, filtering, tokenization, stop-word removal, and stemming. All of these text preprocessing stages improve data quality, making sentiment analysis more effective and accurate. Or essentially, the preprocessing conducted in this research was the first step towards achieving optimal performance from the applied sentiment classification method. The next step after text preprocessing is to weight the features using the TF-IDF method. TF-IDF converts text into numerical features by emphasizing significant terms and minimizing the impact of less relevant words. The next step is to conduct sentiment analysis using the Vader Lexicon to obtain the sentiment polarity of each review. The classification stage, using the SVM method and PSO to optimize accuracy, is the next step. The final step is to measure or evaluate the model's performance using a confusion matrix, which includes metrics such as accuracy, precision, recall, and F1-score. The next step is to evaluate the best model for classifying tourist sentiment towards events held in Mandalika. This study employs the Python programming language for data preprocessing and implementing TF-IDF.

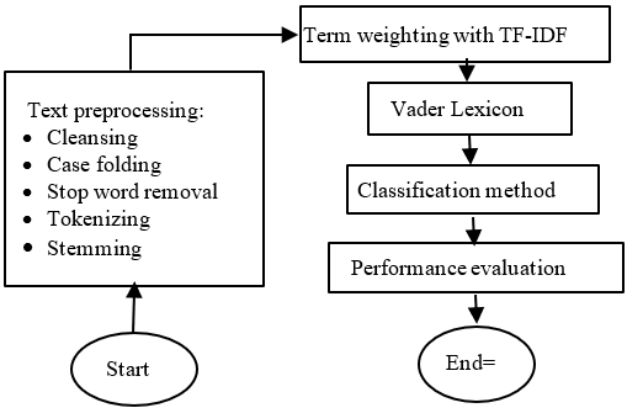

Figure 1. Research methodology flow diagram

2.1. Text preprocessing

Text preprocessing is an essential stage in textual data processing that aims to clean and prepare data for easier analysis. In this study, various text preprocessing techniques were used, including text cleaning (cleansing), case folding, tokenization, stop word removal, and stemming. The entire text preprocessing process involves actions on the public opinion dataset related to the implementation of events in Mandalika. Case folding is the process of changing all letters in a text to lowercase to match the format of the text. Text cleanup involves removing irrelevant characters or elements, such as punctuation marks (e.g., periods, commas, exclamation marks), numbers, special characters (e.g., #, @, %), and HTML elements or URLs.

Furthermore, tokenization breaks text into the smallest units, called tokens, such as words or sentences. The process of removing stop words aims to eliminate common words that appear frequently but have low analytical value, such as "and", "which", or "for", so that the analysis can focus on more meaningful words. Finally, the stemming process is carried out to reduce words to their basic form by removing affixes, thereby bringing variations in word forms together into a consistent representation.

2.2. TF-IDF weighting

The TF-IDF method is a text representation technique widely used in the fields of information retrieval and text mining [35]. The primary purpose of this method is to measure the level of importance of a word (term) in a document relative to the entire collection of documents (corpus). The TF-IDF plays a crucial role in distinguishing words that have significant meaning in a given context by assigning each term a weight based on its frequency of occurrence in individual documents and its rarity in the overall corpus. The TF-IDF method consists of two main components. First, term frequency (TF), which is a measure of how often a term appears in a document. This frequency is calculated by dividing the number of occurrences of a given term by the total number of terms in the document. This TF value indicates the importance of a term within a document's local context. Second, IDF measures the degree of rarity of a term in the entire corpus—the fewer documents that contain the term, the higher the IDF value. The combination of these two components, i.e., the result of multiplication between TF and IDF, results in the weight of TF-IDF. This weight reflects the term's significance in a document relative to the entire corpus, enabling a more accurate and meaningful analysis of the text, as shown in (1)-(3).

$$TF(t,d) = \frac{f_{t,d}}{N_d} \tag{1}$$

The calculation of the IDF value utilized (2). Meanwhile, the (3) functions to obtain the TF-IDF value for term t in document d. In the [2], |D| indicates the total number of documents in the corpus, while |d: t ∈ d| indicates the number of records that contain the term t at least once. The result of multiplication between TF and IDF results in the final weight of TF-IDF, which reflects the level of significance of the term in the document relative to the overall corpus.

$$IDF(t, D) = log\left(\frac{|D|}{1+|d:t\in d|}\right) \quad (2)$$

$$TF - IDF(t, d, D) = TF(t, d)xIDF(t, D) \quad (3)$$

2.3. Vader Lexicon

Vader or Valence Aware Dictionary and Sentiment Reasoner is a lexicon-based sentiment analysis method designed to classify texts into positive, negative, and neutral sentiment categories [36]. This method utilizes a dictionary of words (lexicon) that has been assigned a valence score based on the emotional polarity of each word [37]. One of Vader's strengths is its ability to handle informal language, including emoticons, internet abbreviations, and emphasis on writing styles such as capitalization or exclamation marks [38], [39]. The primary output of Vader is the total score, also known as the compound score, which is calculated by summing the valence scores of all words in the text [37] and then normalized to a range of -1 to +1. A value close to -1 indicates highly negative sentiment, a value close to +1 indicates strong positive sentiment, and a value close to 0 indicates neutral sentiment.

2.4. Classification method

The SVM method classifies using training and testing data. In this study, the dataset is split into 80:20, with 80% for training and 20% for testing. The SVM is a powerful and versatile classification algorithm in machine learning. This algorithm works by searching for hyperplanes in a high-dimensional feature space that can separate data classes with maximum margins [40]. The hyperplane serves as a decision boundary that distinguishes the two classes, intending to maximize the distance (margin) between the data points of the two classes closest to the hyperplane. The data points closest to the hyperplane are called support vectors, as shown in (4).

$$f(x) = sign(w.x + b) = \begin{Bmatrix} +1. & w.x + b \geq 0 \\ -1. & w.x + b < 0 \end{Bmatrix} \quad (4)$$

2.5. Performance evaluation

A confusion matrix is a method for evaluating the performance of classification models by comparing the model's predictions to the actual labels. This matrix typically consists of two main classes - Positive (P) and Negative (N) -and contains four key components. True positive (TP) refers to the number of positive data points whose predictions are accurate. At the same time, a false negative (FN) is the amount of positive data whose prediction is wrong as negative. In contrast, a false positive (FP) indicates the number of negative data points that are incorrectly predicted as positive, and a true negative (TN) indicates the number of negative data points that are correctly predicted as negative [15]. Accuracy measures the percentage of total correct predictions from the overall data. Recall, or sensitivity, indicates the proportion of positive data that the model correctly recognizes. Precision measures the extent to which the model's positive predictions are accurate. Meanwhile, the F1-score is the harmonic average of precision and recall, which is useful when there is a class imbalance in the data. Finally, specificity indicates the model's ability to correctly identify negative data, i.e., the true negative proportion of all negative data, as shown in Table 2 (see equations 5-7).

Table 2. Confusion matrix

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+FP +TN+FN}} \quad (5)$$

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \quad (6)$$

$$F1 - Score = \frac{2*(\text{Precision}*\text{Recall})}{(\text{Precision}+\text{Recall})} \quad (7)$$

3. RESULTS AND DISCUSSION

This section presents the results of sentiment analysis of a dataset of tourist reviews on the implementation of sporting events. The analysis process comprises several stages, including text preprocessing, feature weighting, application of the Vader Lexicon method, and optimization using the PSO algorithm. The dataset used in this study comprised 7,184 reviews of the implementation of various events in Mandalika, as shown in Table 3. Each review contains traveler opinions labeled positive, neutral, or negative. In preparing for further analysis, data preparation occurs during the text preprocessing stage. Text preprocessing includes cleaning, case folding, removing unimportant words, tokenization, and stemming.

Table 3. Dataset sentiment

Text	Sentiment
sangat bahagia adik	Positive
Polda antisipasi kemacetan lalu lintas di motogp mandalika	Neutral
Pertunjukan Indonesia akan dimeriahkan oleh penampilan artis nasional	Positive
ditonton oleh ratusan juta orang di seluruh dunia	Neutral
siapa yang akan menjadi pemenangnya	Positive
…………………………….	………..
…………………………….	………..
di negara kita sendiri debunya jelek	Negative

Tables 4-7 present the results of each preprocessing process. Figure 2 shows the source code of text pre-processing. Meanwhile, Table 8 illustrates the subsequent process after preprocessing, specifically the application of the TF-IDF weighting technique.

Table 4. Case folding

PreProcessing	After case folding
Case folding	*bahagia banget dek* *Polda antisipasi kemacetan lalu lintas motogp mandalika* *pertunjukan indonesia akan dimeriahkan penampilan artis nasional* *ditonton ratusan juta orang dunia* *siapa jadi pemenang* .... *negara sendiri debu jelek*

Table 5. Stop word removal

PreProcessing	After stop word removal
Stop word removal	*"bahagia", "banget", "dek"* *"polda", "antisipasi", "kemacetan", "lalu", "lintas", "motogp", "mandalika"* *"pertunjukan", "indonesia", "akan", "dimeriahkan", "penampilan", "artis", "nasional"* *"ditonton", "ratusan", "juta", "orang", "dunia"* *"siapa", "jadi", "pemenang"* .... *"negara", "sendiri", "debu", "jelek"*

Table 6. Tokenizing

PreProcessing	After tokenizing
Tokenizing	*"bahagia", "banget", "dek"* *"polda", "antisipasi", "kemacetan", "lalu", "lintas", "motogp", "mandalika"* *"pertunjukan", "indonesia", "akan", "dimeriahkan", "penampilan", "artis", "nasional"* *"ditonton", "ratusan", "juta", "orang", "dunia"* *"siapa", "jadi", "pemenang"* .... *"negara", "sendiri", "debu", "jelek"*

Figure 3 illustrates a snippet of source code that implements the SVM algorithm combined with the PSO method to enhance classification performance. In this encoding, PSO optimizes the hyperparameters of the SVM, namely C (regularization) and gamma (RBF kernel parameter), which significantly influence the model's accuracy. The process begins by defining objective functions —the functions the PSO will minimize. The function returns a negative value for the average cross-validation accuracy of an SVM model with a specific parameter combination. The PSO then performs a global search over a predefined parameter space to find the C and gamma values that yield the highest classification accuracy, as shown in Table 9.

Table 7. Stemming

Preprocessing	After stemming
Stemming	*"bahagia", "banget", "dek"* *"polda", "antisipasi", "kemacetan", "lalu", "lintas", "motogp", "mandalika"* *"pertunjukan", "indonesia", "akan", "dimeriahkan", "penampilan", "artis", "nasional"* *"ditonton", "ratusan", "juta", "orang", "dunia"* *"siapa", "jadi", "pemenang"* .... *"negara", "sendiri", "debu", "jelek"*

```
Preprocessing: Cleansing
import re
text = re.sub(r'[^\w\s-]', ' ', text)
text = re.sub (r'\s+', ' ', text). Strip ()
Preprocessing: CaseFolding
text = text. lower ()
Preprocessing: Stopword Removal
data['filtered_text'], data['removed_text'] = zip(*data['tokenized_text'].apply(stopword_removal))
for i, row in data.iterrows ():
text = row['Normalized_Text']
filtered_text = ' '.join(row['filtered_text'])
removed_text = ' '.join(row['removed_text'])
Preprocessing: Tokenizing
data['tokenized_text'] = data['Normalized_Text'].apply(tokenizing_text)
new_header_order = ['Normalized_Text', 'tokenized_text']
df = data.reindex(columns=new_header_order)
Preprocessing: Stemming
data['stemmed_text'] = data['filtered_text'].apply(stemming_text)
data = data[data['stemmed_text'].str.len() > 0]
data['stemmed_textjoin'] = data['stemmed_text'].apply(lambda tokens: ' '.join(tokens))
new_header_order = ['Normalized_Text', 'tokenized_text', 'filtered_text'
```

Figure 2. Source code of text preprocessing

Table. 8 TF-IDF

Tf			Df	IDF (log D/df)	W		
D1	D2	D3			D1	D2	D3
1	0	0	1	0.85	0,85	0	0
0	1	0	1	0.85	0	0.28	0
0	1	0	1	0.85	0	0.28	0
0	1	0	1	0.85	0	0.28	0
0	0	1	1	0.85	0	0	0.28
0	0	1	1	0.37	0	0	0.12
0	0	1	1	0.37	0	0	0.12

```
svm_model3sentfit (X_train_vectorized, y_train_3sent)
predictions_3sent = svm_model_3sent.predict(X_test_vectorized)
accuracy_3sent = accuracy score (y_test_3sent, predictions_3sent)
results_df = pdDataFrame ({'Sentimen': ['3 Sentimen'])
print(results_df)
```

Figure 3. Source code SVM PSO

Table 9. hyper parameter SVM PSO

Best hyperparameter (c)	Accuracy
1,4347501497014852	0,818282

Figure 4 shows the classification results using the original SVM model (without optimization). Based on these results, the model correctly predicted 50 instances for the negative class, 191 cases for the meutral class, and 286 instances for the Positive class. These results indicate that classification accuracy remains limited, particularly in the negative class, which has the lowest true prediction rate. Conversely, Figure 5 shows the classification results obtained with an SVM model optimized using PSO. In this model, the number of correct predictions has increased significantly: 105 for the negative class, 410 for the neutral class, and 531 for the positive class. The improved performance across the class demonstrated that implementing the PSO algorithm improved the quality of hyperparameter selection in SVM, making the model more adaptable to data patterns and resulting in higher overall classification accuracy.

Figure 6 illustrates that applying the PSO method as an optimization algorithm to the SVM model yields a significant improvement in classification performance. PSO plays a role in finding the optimal combination of hyperparameters, including C (cost) and other kernel parameters that affect the margin of separation between classes. The experimental results showed that PSO successfully improved the model's accuracy compared to SVMs run with default parameters or without optimization. The optimal parameter value found is 1.4347, representing the optimal configuration obtained by PSO after an iterative process to identify the best solution in the parameter space. This increase in accuracy suggests that parameter search via PSO can better tailor the model to the data's characteristics, resulting in more precise classification. In addition, it indicates that SVMs rely heavily on parameter selection and that metaheuristic-based approaches, such as PSO, can be an efficient strategy for improving their performance. Thus, integrating PSO into the SVM model training process not only improves numerical accuracy but also demonstrates that an optimization-based approach can enhance the model's ability to capture essential patterns in the data, particularly in high-complexity sentiment classification tasks.

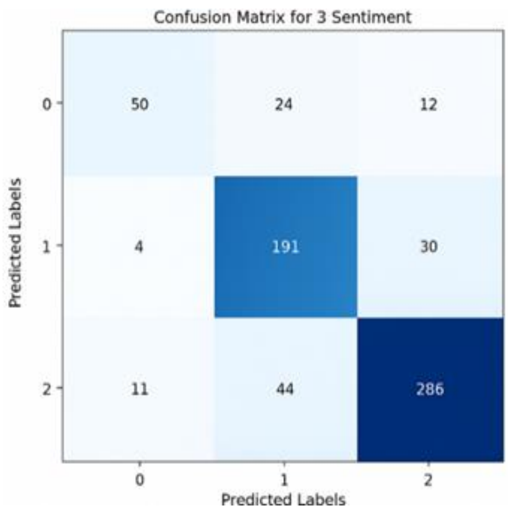

Figure 4. Confusion matrix of SVM original

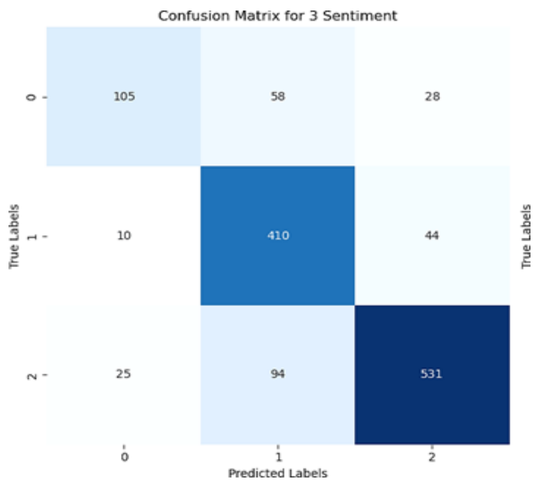

Figure 5. Confusion matrix of SVM PSO

The results of this study demonstrate that using the SVM method to classify sentiment in sports event text data, when optimized with PSO, yields improved performance metrics. Specifically, the accuracy, recall, and F1-score increased to 81.82%, 79.9%, and 79.62%, respectively. In contrast, the performance without PSO yielded lower metrics: 80.15% accuracy, 75.63% recall, and 76.89% F1-score. The findings of this study align with the opinions of previous research [27], [41] that PSO is a crucial aspect of optimization research in science and technology, as demonstrated by this study, which shows that SVM with PSO improves SVM classification performance. In addition, this study strengthens the literature review of [26], which suggests that the use of optimization techniques, such as PSO, can improve the performance of classification models, particularly for the SVM method in sentiment analysis tasks. This research is unique and distinct from previous related studies [6], [8], [12], [21], [26]–[34]; in other words, the results of this research are novel.

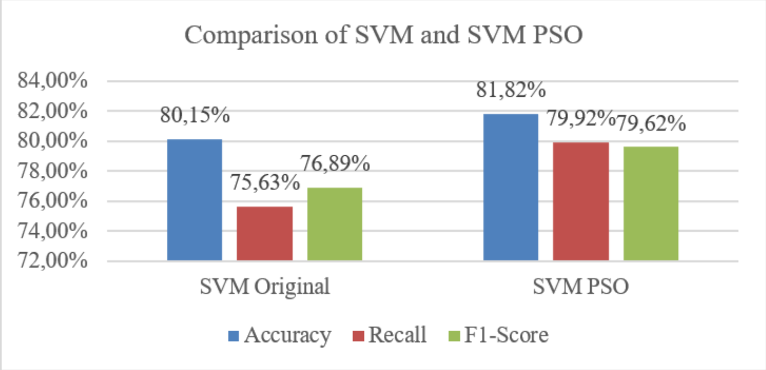

Figure 6. Comparison of SVM and SVM PSO

4. CONCLUSION AND FUTURE WORK

The results showed that classifying text data related to sentiment towards the sporting event into three classes - positive, neutral, and negative - produced accuracy, recall, and F1-score of 81.82%, 79.9%, and 79.62%, respectively, when using the SVM method optimized with PSO. Meanwhile, the SVM method without optimization (original) yields accuracy, recall, and F1-scores of 80.15%, 75.63%, and 76.89%, respectively. This research represents a new study in terms of its topics, research objects, objectives, and results, which have not been explored by researchers before. The main contribution of this study is to reveal the sentiment of user X towards the event in Mandalika in more depth by classifying opinions into three main categories. Additionally, this study demonstrates that applying PSO can significantly improve the accuracy of SVM for sentiment classification.

This research confirms that SVM relies heavily on appropriate parameter selection and that metaheuristic-based approaches such as PSO can improve its performance. The findings of this study support the views of previous researchers who have argued that PSO is a crucial element in optimization research within science and technology. This study demonstrates that using PSO with SVM can improve SVM classification performance. Additionally, these results are consistent with earlier literature indicating that PSO can improve the performance of machine learning classification models. The weakness of this study is that it only compares the original SVM with the SVM optimized using PSO. Further analysis with diverse object data sets is needed to clarify the accuracy differences between the SVM method without PSO and the SVM method with PSO. Besides that, further research proposals should compare other machine learning methods, such as Naïve Bayes and random forest, both without and with PSO treatment. To determine the extent of optimization in accuracy, recall, and F1-score using PSO on various machine learning methods. In addition, it is necessary to conduct further research by applying other optimization methods, such as genetic algorithm, grid search, or bayesian optimization, to compare and determine the best optimization method and to determine which machine learning method to use.

FUNDING INFORMATION

Authors state no funding involved.

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
Christofer Satria	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	
Anthony Anggrawan	✓	✓	✓		✓				✓	✓		✓		
Peter Wijaya Sugijanto	✓					✓		✓		✓	✓	✓		
Husain Husain		✓	✓	✓			✓			✓	✓			
I Nyoman Yoga Sumadewa						✓		✓		✓	✓		✓	
Victoria Cynthia Rebecca					✓	✓				✓	✓			

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

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The data used in this study can be accessed at the following link: https://shorturl.at/dal3F.

REFERENCES

[1] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[2] M. Z. Asghar, A. Khan, A. Bibi, F. M. Kundi, and H. Ahmad, "Sentence-level emotion detection framework using rule-based classification," *Cognitive Computation*, vol. 9, no. 6, pp. 868–894, Dec. 2017, doi: 10.1007/s12559-017-9503-3.

[3] A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," *Journal of Big Data*, vol. 10, no. 1, p. 110, Jun. 2023, doi: 10.1186/s40537-023-00781-w.

[4] M. Água, N. António, P. Carrasco, and C. Rassal, "Large language models powered aspect-based sentiment analysis for enhanced customer insights," *Tourism and Management Studies*, vol. 21, no. 1, pp. 1–19, 2025, doi: 10.18089/tms.202501011.

[5] O. Alsemaree, A. S. Alam, S. S. Gill, and S. Uhlig, "Sentiment analysis of Arabic social media texts: a machine learning approach to deciphering customer perceptions," *Heliyon*, vol. 10, no. 9, p. e27863, May 2024, doi: 10.1016/j.heliyon.2024.e27863.

[6] A. Chourasiya, A. Khan, K. Bajaj, M. Tomar, T. Kohli, and D. Chauhan, "A review of sentiment analysis and emotion detection from text using different models," *International Journal of Engineering Applied Science and Management ISSN*, vol. 6, no. 1, pp. 2582–6948, 2025.

[7] X. Liu, R. Li, S. Ye, G. Zhang, and X. Wang, "Multimodal aspect-based sentiment analysis under conditional relation," in *Proceedings - International Conference on Computational Linguistics, COLING*, 2025, pp. 313–323.

[8] Z. Zhao and S. Yang, "Research on red cultural inheritance and application of SVM support vector," *Applied Mathematics and Nonlinear Sciences*, vol. 10, no. 1, pp. 1–18, 2025.

[9] M. A. J. Eljatin, R. W. S. Sumadinata, and D. S. Sari, "Sports and cultural diplomacy integration in post-pandemic COVID-19 international events: evidence from Indonesia's MotoGP Mandalika 2022," *Daengku: Journal of Humanities and Social Sciences Innovation*, vol. 5, no. 1, pp. 113–120, Feb. 2025, doi: 10.35877/454RI.daengku3745.

[10] H. Q. Low, P. Keikhosrokiani, and M. P. Asl, "Decoding violence against women: analysing harassment in middle eastern literature with machine learning and sentiment analysis," *Humanities and Social Sciences Communications*, vol. 11, no. 1, p. 497, Apr. 2024, doi: 10.1057/s41599-024-02908-7.

[11] N. A. Semary, W. Ahmed, K. Amin, P. Pławiak, and M. Hammad, "Enhancing machine learning-based sentiment analysis through feature extraction techniques," *PLOS ONE*, vol. 19, no. 2, p. e0294968, Feb. 2024, doi: 10.1371/journal.pone.0294968.

[12] A. Anggrawan, C. Satria, H. Wardhana, P. W. Sugijanto, A. D. Dayani, and A. S. Abdi, "Comparison of sentiment analysis evaluation regarding international mobile equipment identity blocking between the K-nearest neighbor and support vector machine methods," in *2024 5th International Conference on Computational Science & Information Management (ICoCSIM)*, IEEE, Oct. 2024, pp. 193–199. doi: 10.1109/ICoCSIM65098.2024.00003.

[13] V. Echambadi and I. High, "Financial market sentiment analysis using LLM and RAG," *Journal SSRN*, pp. 1–7, 2025.

[14] A. Anggrawan, C. Satria, C. K. Nuraini, L. -, N. G. A. Dasriani, and M. -, "Machine learning for diagnosing drug users and types of drugs used," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 111–118, 2021, doi: 10.14569/IJACSA.2021.0121113.

[15] D. Rifaldi *et al.*, "Machine learning 5.0: in-depth analysis trends in classification," *Scientific Journal of Computer Science*, vol. 1, no. 1, pp. 1–15, 2025, doi: 10.64539/sjcs.v1i1.2025.18.

[16] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical study on classifiers for earlier prediction of COVID-19 infection cure and death rate in the Indian States," *Healthcare*, vol. 10, no. 1, p. 85, Jan. 2022, doi: 10.3390/healthcare10010085.

[17] A. Alqarni, "A support vector machine (SVM) model for privacy recommending data processing model (PRDPM) in internet of vehicles," *Computer, Materials & Continua*, vol. 82, no. 1, pp. 389–406, 2024, doi: 10.48084/etasr.7743.

[18] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," in *ACRS 2020 - 41st Asian Conference on Remote Sensing*, 2020.

[19] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6308–6325, 2020, doi: 10.1109/JSTARS.2020.3026724.

[20] S. Xu, H. Wu, J. Luo, J. Chen, and H. Jia, "The application of support vector machine (SVM) in industrial carbon accounting prediction and green electricity control strategies," *E3S Web of Conferences*, vol. 615, p. 01012, Feb. 2025, doi: 10.1051/e3sconf/202561501012.

[21] P. C. Sogas, I. F. Molina, À. A. Batlle, and J. M. R. Vílchez, "Economic and social yield of investing in a sporting event: sustainable value creation in a territory," *Sustainability*, vol. 13, no. 13, p. 7033, 2021, doi: 10.3390/su13137033.

[22] D. Parra-Camacho, M. H. González-Serrano, M. Alguacil Jiménez, and P. Jiménez-Jiménez, "Analysis of the contribution of sport events to sustainable development: Impacts, support and resident's perception," *Heliyon*, vol. 9, no. 11, p. e22033, Nov. 2023, doi: 10.1016/j.heliyon.2023.e22033.
[23] S. Owen and D. Chambers, "Volunteers' sense of (Dis)connection at a sport event," *Leisure Sciences*, vol. 46, no. 2, pp. 105–122, Feb. 2024, doi: 10.1080/01490400.2021.1916660.
[24] K. Kogoya, T. S. Guntoro, and M. F. P. Putra, "Sports event image, satisfaction, motivation, stadium atmosphere, environment, and perception: a study on the biggest multi-sport event in Indonesia during the pandemic," *Social Sciences*, vol. 11, no. 6, p. 241, May 2022, doi: 10.3390/socsci11060241.
[25] H. Giroh, V. Kumar, and G. Singh, "Improving the performance of hybrid models using machine learning and optimization techniques," *International Journal of Membrane Science and Technology*, vol. 10, no. 2, pp. 3396–3409, 2023, doi: 10.15379/ijmst.v10i2.3138.
[26] T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, "Particle swarm optimization: a comprehensive survey," *IEEE Access*, vol. 10, pp. 10031–10061, 2022, doi: 10.1109/ACCESS.2022.3142859.
[27] E. H. Houssein, A. G. Gad, K. Hussain, and P. N. Suganthan, "Major advances in particle swarm optimization: theory, analysis, and application," *Swarm and Evolutionary Computation*, vol. 63, p. 100868, Jun. 2021, doi: 10.1016/j.swevo.2021.100868.
[28] M. Bordoloi and S. K. Biswas, "Sentiment analysis: A survey on design framework, applications and future scopes," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 12505–12560, Nov. 2023, doi: 10.1007/s10462-023-10442-2.
[29] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Systems with Applications*, vol. 223, p. 119862, Aug. 2023, doi: 10.1016/j.eswa.2023.119862.
[30] O. Abiola, A. Abayomi-Alli, O. A. Tale, S. Misra, and O. Abayomi-Alli, "Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 5, Jan. 2023, doi: 10.1186/s43067-023-00070-9.
[31] K. Denecke and D. Reichenpfader, "Sentiment analysis of clinical narratives: A scoping review," *Journal of Biomedical Informatics*, vol. 140, p. 104336, Apr. 2023, doi: 10.1016/j.jbi.2023.104336.
[32] Y. A. Singgalen, "Sentiment and toxicity analysis of sport event MotoGP Mandalika Circuit using cross-industry standard process for data-mining," *Journal of Information System Research (JOSH)*, vol. 5, no. 3, pp. 731–741, Apr. 2024, doi: 10.47065/josh.v5i3.5056.
[33] T. I. Pribadi, F. Fahry, M. Muharis, and E. D. P. Marswandi, "Analysis of tourist sentiment towards tourist attractions in the mandalika special economic zone using the Naïve Bayes method," *Jurnal Bumigora Information Technology (BITe)*, vol. 6, no. 1, pp. 105–114, Aug. 2024, doi: 10.30812/bite.v6i1.4081.
[34] D. Lamani *et al.*, "SVM directed machine learning classifier for human action recognition network," *Scientific Reports*, vol. 15, no. 1, p. 672, Jan. 2025, doi: 10.1038/s41598-024-83529-7.
[35] C. N. Agustina, R. Novita, Mustakim, and N. E. Rozanda, "The implementation of TF-IDF and Word2Vec on booster vaccine sentiment analysis using support vector machine algorithm," *Procedia Computer Science*, vol. 234, pp. 156–163, 2024, doi: 10.1016/j.procs.2024.02.162.
[36] V. Nurcahyawati, Z. Mustaffa, and M. Khalaf, "Exceeding manual labeling: Vader Lexicon as an accurate alternative to automatic sentiment classification," *The International Arab Journal of Information Technology*, vol. 22, no. 2, pp. 225–235, 2025, doi: 10.34028/iajit/22/2/2.
[37] K. Barik and S. Misra, "Analysis of customer reviews with an improved Vader lexicon classifier," *Journal of Big Data*, vol. 11, no. 1, pp. 1–19, 2024, doi: 10.1186/s40537-023-00861-x.
[38] A. Mahmoudi, D. Jemielniak, and L. Ciechanowski, "Assessing accuracy: a study of Lexicon and rule-based packages in R and Python for sentiment analysis," *IEEE Access*, vol. 12, pp. 20169–20180, 2024, doi: 10.1109/ACCESS.2024.3353692.
[39] N. W. S. Saraswati, I. Ketut Gede Darma Putra, M. Sudarma, and I. Made Sukarsa, "Enhance sentiment analysis in big data tourism using hybrid lexicon and active learning support vector machine," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 13, no. 5, pp. 3663–3674, Oct. 2024, doi: 10.11591/eei.v13i5.7807.
[40] T. A. Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. M. Su'ud, "Sentiment analysis using support vector machine and random forest," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 67–75, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.5.
[41] Y. Li, J. Zhou, F. Chen, and M. Sun, "An improved particle swarm optimization for wind resistance performance design of high-rise buildings," *Advances in Wind Engineering*, vol. 2, no. 2, p. 100053, Jun. 2025, doi: 10.1016/j.awe.2025.100053.

BIOGRAPHIES OF AUTHORS

Christofer Satria is dean of the Faculty of Design and Arts at Bumigora University in Mataram, Indonesia. He received a bachelor's degree (S.Sn.) in Visual Communication Design from Petra Christian University, Surabaya, Indonesia, and a master's degree (M.Sn.) in Visual Communication Design from the Indonesian Art Institute (ISI), Denpasar, Bali, Indonesia. He also holds a master's degree in computer science. His highest level of education is a doctorate (Dr.) in the same field of science as his expertise. He is currently a senior lecturer at Bumigora University in Indonesia and the Head of the Photography, Animation, and Video Laboratory. His research interests include machine learning, animation learning media, video learning media, design systems, educational methods, and experimental design. He can be contacted at email: chris87@universitasbumigora.ac.id.

Anthony Anggrawan currently works as a professor in the Department of Computer Science as a lecturer, university Rector, and State Civil Apparatus at Bumigora University, Indonesia. In addition, he serves as a reviewer for several reputable international scientific journals. He received his Master of Computer Science (M.T) from the 10 November Institute of Technology, Surabaya, Indonesia. After that, he earned his first Doctoral degree (Ph.D.) in Accounting Information Systems from Universiti Utara Malaysia. Then, he received his second Doctoral degree (Dr.) from Hasanuddin University, Makassar, Indonesia, in linguistics. Finally, he earned his third Doctorate in Educational Technology from the State University of Jakarta. His research interests include educational technology, online learning, data mining, text mining, and the internet of things. He has been the advisor of the International Conference on Computational Science and Information Management (ICoCSIM) Association; and advisor of Cooperation Research Inter University (CORIS), Indonesia. He can be contacted at email: anthony.anggrawan@universitasbumigora.ac.id.

Peter Wijaya Sugijanto is a lecturer and researcher in the Faculty of Medicine, Bumigora University, Mataram, Indonesia. He holds a general practitioner degree from Maranatha University and is currently pursuing a specialist degree in Obstetrics and Gynecology at Andalas University in Padang. His research interests include obstetrics and gynecology, prenatal care, reproductive health, maternal-fetal medicine, and women's health education. He actively contributes to interdisciplinary research and community-based technology development in the fields of obstetrics, gynecology, and health education. His research applies machine learning methods, test cases, and field practice, and employs statistical methods. He can be contacted at email: peter.ws@universitasbumigora.ac.id.

Husain Husain currently serves as head of the Information Technology Undergraduate Study Program in the Faculty of Engineering—his academic Functional Position is senior lecturer (associate professor). He was born in the province of West Nusa Tenggara, Sumbawa Regency, Lopok Beru Village, on 22 February 1986. Completed his undergraduate studies and obtained a Bachelor of Computer Science (S.Kom) at the College of Informatics and Computer Management (STMIK Bumigora Mataram) in 2009 and in 2017 obtained a Master of Computer Science (M.Kom.) at Dian Nuswantoro University, Semarang and completed the Doctoral Program in the Computer Science study program with a Doctoral Degree (Dr.) at the Sumatera Utara University in 2021. He is currently working at Bumigora University (UBG) in Mataram as a Homebase lecturer in the Information Technology (IT) Study Program. He can be contacted at email: husain@universitasbumigora.ac.id.

I Nyoman Yoga Sumadewa is a lecturer and head of the Visual Communication Design Program at Bumigora University, Mataram, Indonesia. He holds master's and doctoral degrees in visual communication design. His academic interests include visual communication, user experience, digital content development, and educational media design. He is also actively involved in institutional digital transformation initiatives and community-based design outreach. His highest level of education is a doctorate in the same field of science as his expertise. He can be contacted at email: yoga@universitasbumigora.ac.id.

Victoria Cynthia Rebecca is a researcher at the Faculty of Medicine, Bumigora University, Mataram, Indonesia. She holds a general practitioner degree from Maranatha University and is currently pursuing a master's degree in Medical and Health Professional Education at Hasanuddin University in Makassar. Her research interests include clinical medicine, medical education, health professional education, evidence-based practice, and curriculum development. She actively contributes to interdisciplinary research and community-driven technology innovations as a medical doctor and health professional educator in the fields of health informatics and education systems. Her research applies machine learning methods, test cases, or field practice, and applies statistical methods. She can be contacted at email: victoria.cynthia@universitasbumigora.ac.id.