Applied Information System and Management (AISM)
Volume 8, (1) 2025, p. 29–36
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i1.42911

Sentiment Analysis of COVID-19 Booster Vaccines on Twitter Using Multi-Class Support Vector Machine

Andi Nurkholis[1], Styawati[2*], Syahirul Alim[3], Hendi Saputra[4], Andrey Ferriyan[5]

***Abstract*— The Indonesian government's implementation of a booster vaccination program as part of its COVID-19 response has generated diverse public reactions, particularly on social media platforms like Twitter. This study aims to analyze public sentiment regarding booster vaccines by examining Twitter data to understand the prevailing discourse and attitudes toward this policy. The research employs sentiment analysis, a text mining and processing technique, to classify tweets into positive, neutral, and negative categories. The study utilizes the Support Vector Machine (SVM) algorithm, evaluating its performance through a multi-class parameter assessment. Two multi-class strategies, One-against-one (OAO) and One-against-rest (OAR) are combined with various kernels (Sigmoid, Polynomial, and RBF) to identify the most accurate model for sentiment classification. The results show that the OAO method with the RBF kernel achieves the highest accuracy of 96%, outperforming other combinations like OAO with Polynomial (95.2%) and Sigmoid (93.7%) kernels. Similarly, the RBF kernel performs best with 95.5% accuracy in the OAR approach. Using the optimal model, sentiment analysis classifies 49 tweets as positive, 927 as neutral, and 24 as negative, revealing a predominantly neutral public sentiment with limited positive and negative opinions. In conclusion, this study demonstrates the effectiveness of SVM, particularly the OAO method with the RBF kernel, for sentiment analysis of social media data. The findings provide insights into public perceptions of the booster vaccine program, offering policymakers a data-driven basis for designing targeted communication strategies to address concerns and enhance public acceptance.**

***Index Terms*—Booster vaccine, COVID-19, sentiment analysis, support vector machine.**

Received: 09 December 2024; Revised: 25 January 2025; Accepted: 4 February 2025
*Corresponding author
[1]Andi Nurkholis, Department of Informatics, UPN Veteran Yogyakarta, Indonesia (e-mail: andinurkholis@upnyk.ac.id).
[2]Styawati, Department of Information Systems, Universitas Teknokrat Indonesia, Indonesia (e-mail: styawati@teknokrat.ac.id).
[3]Syahirul Alim, Department of Information Systems, Universitas Teknokrat Indonesia, Indonesia (e-mail: syahirul_alim@teknokrat.ac.id).
[4]Hendi Saputra, Department of Information Systems, Universitas Teknokrat Indonesia, Indonesia (e-mail: hendi_saputra.mhs@teknokrat.ac.id).
[5]Andrey Ferriyan, Graduate School of Media and Governance, Keio University, Japan (e-mail: andrey@keio.jp).

I. INTRODUCTION

Corona Virus Disease (COVID-19) is an infectious illness caused by the SARS-CoV-2 virus [1]. The Indonesian government has enacted numerous measures to combat the COVID-19 pandemic, including a primary vaccination program involving two doses. However, research indicates that antibody levels diminish six months after the primary COVID-19 vaccination series is completed. Consequently, booster doses are necessary to enhance individual protection, particularly among vulnerable populations [2]. A study conducted by ITAGI (letter number ITAGI/SR/2/2022) regarding administering additional COVID-19 vaccine doses recommends booster shots to improve vaccine efficacy, which has waned over time [3]. Booster vaccines, introduced by the government, aim to enhance immunity as antibody levels from primary vaccinations decline over time. In early 2022, Indonesia entered the recovery phase caused by the pandemic. The government made an option for the public to do a third dose (booster) vaccine [4]. This is done to increase the body's immunity. Along with implementing the policy with the booster vaccination, several pros and cons emerged, thus inviting many people to give their opinions on social media [5].

The introduction of booster vaccine policies in early 2022 has generated diverse public responses and highlighted several societal challenges. Key public concerns include questions about the safety and effectiveness of additional doses, equitable vaccine access and distribution in remote areas, prioritization of vulnerable populations, and the economic implications of sustained vaccination programs. These multifaceted concerns underscore the importance of understanding public perception and acceptance of booster vaccination policies. Social media platforms, particularly Twitter, have become primary channels through which citizens express these concerns and share their experiences with vaccination programs [6], [7]. Understanding these public sentiments is crucial for policymakers to develop more effective communication strategies and ensure equitable implementation of booster vaccination programs.

Public attitudes toward booster vaccines can be evaluated through Twitter data analysis, specifically through the process known as sentiment analysis [8]. This analytical approach

involves the computational study of textually expressed opinions, feelings, and emotional responses. It is not as easy as the usual classification process because of language that contains ambiguous words, the absence of intonation in a text, and the development of the language itself [9]. Furthermore, sentiment analysis is a sentiment from subjective texts towards analysis, processing, summarizing, and inferential [10], which in this study was carried out on public opinion on booster vaccines by classifying opinions into three classes: positive, neutral, and negative.

Numerous study efforts have focused on sentiment analysis, with the Support Vector Machine (SVM) method being a popular choice. As a machine learning technique, SVM is extensively employed for classification tasks and identifying the optimal hyperplane. The fundamental concept behind SVM is determining a separating hyperplane that effectively distinguishes between positive and negative classes. However, a multi-class SVM approach is necessary to handle the increased complexity when dealing with three classes. The SVM algorithm is quite good compared to other algorithms, which were stated from previous studies comparing SVM with KNN by producing successive accuracies of 95% and 80% [11]. A performance evaluation compared Naïve Bayes and SVM algorithms for analyzing sentiments about the COVID-19 vaccine. The results demonstrated that SVM outperformed across multiple performance indicators. The SVM algorithm reached 90.47% accuracy, 90.23% precision, and 90.78% recall. Meanwhile, Naïve Bayes showed comparatively lower results, achieving 88.64% accuracy, 87.32% precision, and 88.13% recall. These findings highlight the effectiveness of the SVM algorithm in accurately classifying sentiments related to the COVID-19 vaccine compared to the Naïve Bayes approach. [12].

Multiple research studies have examined the effectiveness of several classification algorithms, including C4.5, Random Forest, SVM, and Naive Bayes. The analysis consistently showed that SVM emerged as the top performer with a 95% accuracy rate. Both C4.5 and Naive Bayes algorithms achieved identical accuracy levels of 86.67%, while Random Forest demonstrated the least effective performance with an accuracy of 83.33%. These outcomes establish SVM's exceptional performance in classification accuracy when measured against other popular algorithms like C4.5, Random Forest, and Naive Bayes [13]. This means obtaining the best results by conducting a sentiment analysis of booster vaccines using the SVM algorithm is possible. Furthermore, the SVM algorithm depends on parameter configuration for maximum results, including the multi-class and kernel parameter approach. By comparing the combination of multi-class approaches with kernels in SVM, the accuracy of the combination of methods and kernels is expected to be determined.

This study compares the optimal accuracy values achieved by combining the One Against One (OAO) and One Against Rest (OAR) approaches with various kernels, namely Sigmoid, Polynomial, and Gaussian Radial Basis Function (RBF), in the context of Multi-class SVM. This study draws upon Twitter-sourced data divided into training and test datasets. The training dataset contains labeled examples that help identify key data characteristics and establish patterns for model development. The test dataset, meanwhile, consists of independent labeled samples used to assess how well the developed patterns and models perform in categorizing new, previously unexamined data [14]. Performance evaluation of the model employs multiple metrics, including Accuracy, Precision, Recall, and F1-score, to determine which combination of approaches and kernels in the Multi-class SVM method delivers optimal results for analyzing booster vaccine tweets. As an implication, public opinion is more likely to support/neutral/reject through sentiment analysis from the Twitter dataset.

II. Research Method

This study follows a methodical and structured approach to accomplish its objectives. The sequential progression of research phases is illustrated in Fig. 1. Details of the study steps in Fig. 1 are explained as follows.

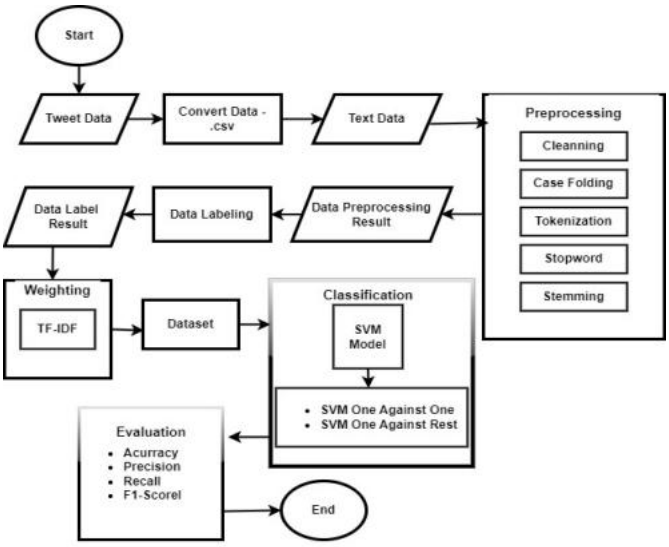

Fig. 1. Research stage

A. Data Collecting

The data collection process involved gathering textual data from Twitter over twelve months, from January 1, 2022, through December 31, 2022. This is based on the emergence of trending public sentiment towards government policies regarding booster vaccines, which will be implemented starting January 2022 as stated in Circular Letter number HK.02.02/II/252/2022 [15]. The crawling technique uses Python programming to collect data on the Twitter social media website. The Steam Twitter API method used by Twitter developers necessitates an API key and an access token for authentication purposes. Data retrieved by searching "Vaksin Booster" is stored in an Excel file with a CSV extension [16]. Figure 2 provides an example of the Twitter data crawling

Applied Information System and Management (AISM)
Volume 8, (1) 2025, p. 29–36
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i1.42911

process.

B. Data Preprocessing

Preprocessing data converts raw data into data suitable for modeling [17], [18]. At this stage, unstructured raw datasets that are not ready to go through the classification stage are

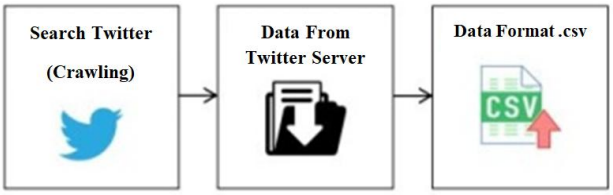

Fig. 2. Data collection

processed so that the data is cleaned first [19], [20].The preprocessing steps in this study are illustrated in Fig. 3.

Following is a description of each stage in Fig. 3.

1) The cleaning process involves eliminating unnecessary elements from the data, including numerical values, extra spaces, punctuation, and irrelevant symbols such as Twitter handles (@username), hashtags (#), and emoji characters.

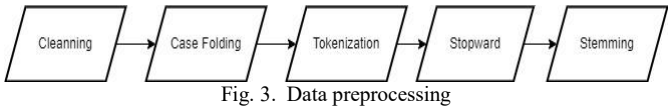

Fig. 3. Data preprocessing

2) Case folding refers to converting all the words in a dataset document to lowercase (standard form), eliminating any uppercase letters. This process transforms characters 'A' to 'Z' into 'a' to 'z'. Non-alphanumeric characters, such as punctuation marks and spaces, are treated as delimiters.
3) Tokenization is breaking text removed with punctuation to become a sentence of words and producing a sentence/word that stands alone. Words, numbers, symbols, punctuation marks, and other items can all be referred to as tokens. In other words, this step seeks to deconstruct answers into individual words.
4) Stopword removal involves filtering high-frequency words with little semantic value, using predefined wordlists or stoplist algorithms to eliminate these non-essential terms. The usage of conjunctions like ‘dan’, ‘yang’, ‘serta’, and ‘setelah’ is an example from this research. These stopwords can be removed to decrease processing time and index size. It also can lower noise levels.
5) Stemming removes all non-standard sentence affixes, which will be converted into essential words in sentences. This stage involves reducing words with suffixes or prefixes to their base form to minimize multiple variations of the same term in the dataset. This process also helps consolidate words that share identical roots and meanings but appear different due to varying affixes.

C. Data Labeling

Following the preprocessing phase, the cleaned data undergoes a labeling process where each entry is classified into one of three sentiment categories: positive, negative, or neutral. These three labels will be the classes used in the SVM classification process in the sentiment analysis process. The labeling process was conducted with expert knowledge [16] provided by Mr. M. Ghufroni An'Ars, S.Pd., M.Pd., an Indonesian language expert and lecturer at Universitas Teknokrat Indonesia.

D. TF-IDF Weighting

This phase involves assigning weights to individual words in the dataset using the TF-IDF methodology. TF-IDF is a weighting technique that combines Term Frequency and Inverse Document Frequency calculations to evaluate word occurrence patterns. Term Frequency (TF) measures how frequently a word appears within a single document, while Inverse Document Frequency (IDF) assesses the word's distribution across the entire document collection [21]. The word weighting process helps determine the relative importance of each term in the documents. The complete

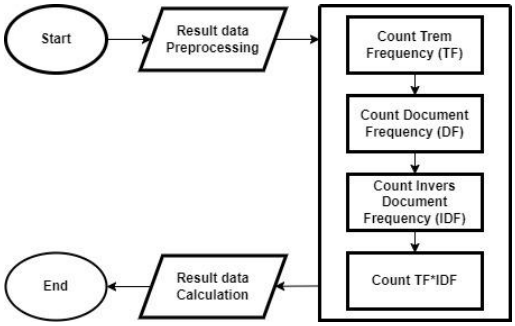

Fig. 4. TF-IDF weighting

TF-IDF weighting process flow is depicted in Fig. 4.

E. Multi-Class Support Vector Machine

SVM, classified as a supervised learning technique in data mining [22], primarily aims to determine the most effective hyperplane among multiple options in a two-class dataset to achieve maximum accuracy. The process involves maximizing the margin, defined as the shortest distance between the hyperplane and the closest data points, known as support vectors [23]. Initially, SVM was applied to classify data into two categories [24]. Later in its development, the SVM was extended to multi-class classification for more complex analysis needs [25]. If the separator is a line in two dimensions, it is called a hyperplane in more than three dimensions.

Multi-class SVM requires specialized approaches when dealing with more than two classes, such as the three sentiment classes in this study (positive, negative, and neutral). This study employs two popular strategies: One-Against-One (OAO) and One-Against-Rest (OAR) [16]. The OAO approach can be understood through a tournament analogy. In a three-class problem, it creates separate "matches" between each pair of classes (positive vs. negative, positive vs. neutral, negative vs. neutral), resulting in $(k\text{-}1)/2$ binary classifiers, where k is the number of classes. For example, OAO builds three separate

classifiers with our three sentiment classes. Each classifier specializes in distinguishing between two specific sentiments, similar to how a sports tournament determines winners through individual matches between pairs of teams.

In contrast, the OAR strategy adopts a "one vs. all others" approach. For our three sentiment classes, it creates k separate binary classifiers, where each classifier distinguishes one class from all others combined. For instance, the first classifier might identify positive sentiments versus all non-positive (neutral and negative combined), the second identifies negative versus non-negative, and so on. This approach is analogous to how a talent show might first identify singers versus non-singers, then dancers versus non-dancers, to categorize participants into distinct groups. The key difference between these approaches lies in their problem-solving strategy: OAO breaks down the multi-class problem into smaller, paired comparisons, while OAR tackles it by isolating each class from the rest.

While SVM fundamentally uses a linear hyperplane for linearly separable data, it employs kernel functions to transform feature spaces when dealing with nonlinear class distributions. Non-linear SVM uses a kernel trick that functions to assist and facilitate classification in a non-linear form. The kernel function transforms input data from its original input space into a feature space, allowing class separation to occur along curved lines or planes in higher dimensions rather than simple straight lines. The approach with the kernel differs from other methods because approaches with other methods usually reduce the initial dimensions to simplify the computation process. Selecting appropriate kernel functions is crucial for SVM performance in sentiment analysis tasks. This study implements three kernel functions that have demonstrated effectiveness in text classification problems: Sigmoid, Polynomial, and RBF. The combination of these three kernels allows for a comprehensive evaluation of different approaches to sentiment classification in the context of vaccine-related social media discussions.

1) The Sigmoid kernel, acting as an activation function for artificial neurons, functions similarly to a neural network's two-layer perceptron architecture. The Sigmoid kernel was included as it functions similarly to neural networks, making it effective for detecting subtle patterns in sentiment expression. Its non-linear transformation capabilities are particularly useful for handling social media posts' informal and diverse language patterns. The sigmoid kernel in SVM is applied using (1) [26].

$$K(x_i, x_j) = \tanh\left(\sigma(x_i, x_j) + c\right) \quad (1)$$

2) Polynomial kernel enables nonlinear model learning by measuring vector similarities in feature space using polynomial functions of the original variables. The Polynomial kernel was selected due to its ability to model feature interactions through the degree parameter, which is valuable for capturing word relationships and contextual patterns in social media text. Its performance in text classification has been well-documented, especially when dealing with varied sentence structures common in Twitter data. The polynomial kernel is applied in SVM using (2) [26].

$$K(x, y) = (x^T . y + 1)^d \quad (2)$$

where d is degree of polynomial

3) RBF is a kernel function that generates output values based on the distance between the input and a fixed point, which can be the origin or any other designated reference point. The RBF kernel was chosen for its versatility in handling non-linear relationships and its proven effectiveness in high-dimensional text classification tasks. It is particularly suitable for sentiment analysis as it can capture complex emotional nuances in text data by mapping them to infinite-dimensional space. The RBF kernel is implemented in SVM using (3) [26].

$$K(x, y) = \exp\left(-\frac{\|x,y\|^2}{2\sigma^2}\right) \quad (3)$$

where $\sigma > 0$ is the constant term

F. Model Evaluation

The evaluation of the Sentiment Analysis model for COVID-19 Booster Vaccines on Twitter, utilizing the Multi-Class SVM, focuses on key performance metrics to assess its effectiveness. Accuracy, precision, recall, and F1-score are employed to accurately measure the model's ability to classify tweets into positive, neutral, and negative sentiments. Accuracy reflects the overall correctness of the model, while precision indicates its reliability in avoiding false positives. Recall evaluates the model's capability to capture all relevant instances of each sentiment class, minimizing false negatives. The F1-score, balancing precision and recall, provides a comprehensive measure of the model's performance, especially in handling imbalanced datasets. These metrics collectively ensure a robust evaluation of the SVM model, highlighting its suitability for analyzing public sentiment on complex topics like booster vaccines.

III. RESULT

A. Preprocessing Data Result

Five thousand records were successfully extracted from Twitter using the crawling approach between January 1 and December 31, 2022. The data preparation phase included various steps such as data cleaning, converting text to lowercase, breaking it into individual tokens, identifying and handling stopwords, and reducing words to their base or root form through stemming. Python programming, which uses the fundamental preprocessing package, is used to implement the data preparation stage. Punctuation, numerical values, and other distinctive symbols are removed during the first step of data preparation, called cleaning. The cleaning procedure depicted in Table 1 is illustrated in the following.

Next is case folding, where text with capital letters is changed to lowercase. The procedure of folding a case is illustrated in Table 2 as follows.

Applied Information System and Management (AISM)
Volume 8, (1) 2025, p. 29–36
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i1.42911

Table 1.
Cleaning result

Before	After
@iwansunaryanto: Antri vaksin booster, biar bisa pulang dan mudik, dan lain sebagainya…!	Antri vaksin booster biar bisa pulang dan mudik dan lain sebagainya

Table 2.
Case folding result

Before	After
Antri vaksin booster biar bisa pulang dan mudik dan lain sebagainya	antri vaksin booster biar bisa pulang dan mudik dan lain sebagainya

Tokenizing, which divides words based on spaces, is the following data preparation step. Separation is carried out to facilitate easy word analysis. The tokenizing procedure depicted in Table 3 has an example in the following.

Table 3.
Tokenizing result

Before	After
antri vaksin booster biar bisa pulang dan mudik dan lain sebagainya	“antri” “vaksin” “booster” “biar” “bisa” “pulang” “dan” “mudik” “dan” “lain” “sebagainya”

The following data preparation step is a stopword, which converts irregular words into terms that conform to the Large Dictionary’s of Indonesian Language (KBBI) standards. Table 4 demonstrates an example of the process used for removing stopwords.

Table 4.
Stopword result

Before	After
“antri” “vaksin” “booster” “biar” “bisa” “pulang” “dan” “mudik” “dan” “lain” “sebagainya”	antri vaksin booster pulang mudik sebagainya

The following data preparation step stems from eliminating terms like ‘itu’, ‘nya’, ‘ke’, ‘dari’, and other nondescriptive or ineffective words. Table 5 presents a sample depicting how the stemming technique works.

Table 5.
Stemming result

Before	After
antri vaksin booster pulang mudik sebagainya	antri vaksin booster pulang mudik sebagai

The next stage in data preprocessing involves labeling, where the data is divided into three categories: positive, negative, and neutral. The sentiment classification is carried out by Mr. M. Ghufroni An'Ars, S.Pd., M.Pd., a faculty member at Universitas Teknokrat Indonesia, specializing in the Indonesian language. His expertise states that class categorization can be identified based on the following:

1) Neutral class states that the user does not participate in the discussed topic, which the absence of an attitude statement or discussion of other matters can prove.

2) The positive class agrees on the discussed topic, proven by a sentence of agreement or action that agrees.

3) A negative class expresses disapproval of the topics discussed, which can be proven by vocabulary that expresses disapproval or acts of refusal.

Table 6.
Labeling result

Tweet	Class
antri vaksin booster pulang mudik sebagai	Neutral
segera lakukan vaksinasi booster vaksinasi booster tubuh lindung kali lipat lebih kuat banding vaksinasi dosis	Positive
menteri vaksin kali belum booster kenapa gw selalu ngakak terkait bijak vaksin booster apalagi bijak booster boleh masuk mall tempat umum ngaruh coba penyebar covid strategi habis vaksin ngontrol mobilitas	Negative

An illustration of the data labeling process in this research is presented in Table 6.

TF-IDF weighting, the final step in data preparation, focuses on transforming textual information into numerical values. This enables computations to assess the relevance of individual words, with the assigned weight indicating a word's significance. Additionally, TF-IDF weighting serves as a data filtering mechanism. Words assigned a weight above zero proceed to the subsequent stage for further analysis. In contrast, those with zero weight are filtered out and eliminated from additional processing or visualization. This step is crucial for prioritizing meaningful terms and optimizing the data for the following workflow phases.

B. Multi-Class SVM Model

The dataset comprises 5,000 records created through data preprocessing and is then split into training and test sets in an 80:20 ratio for subsequent modeling. Using Python 3.9 and the Scikit-learn library (version 0.23), the SVM multi-class classification algorithm was implemented on a training dataset consisting of 4,000 instances. Analysis of the dataset distribution reveals several significant patterns in the collected data. From a total of 5,000 analyzed tweets, there is an unbalanced distribution where the majority of tweets have neutral sentiment, totaling 3,856 (77.12%), followed by positive sentiment tweets numbering 812 (16.24%), and negative sentiment tweets only 332 (6.64%) as shown in Fig 5.

The dominance of neutral sentiment tweets may indicate public hesitancy to express strong opinions regarding vaccination policies. Meanwhile, the low proportion of negative sentiment could suggest potential self-censorship on Twitter's content moderation policies and algorithmic content promotion may also influence the visibility and collection of certain opinions. To address potential biases in this analysis,

several steps have been implemented, such as applying stratified sampling in the training-test data split to maintain class proportions.

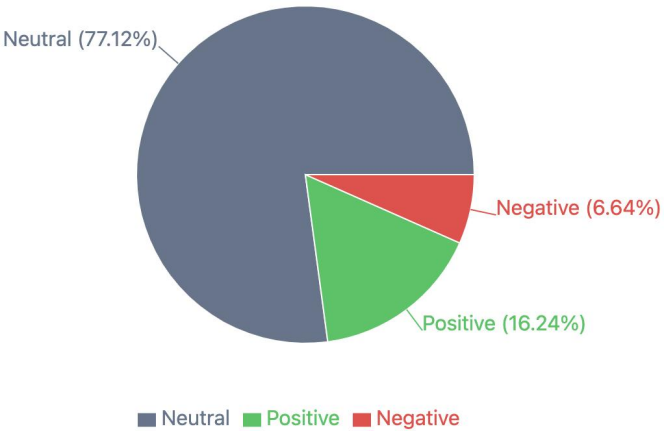

Fig. 5. Class distribution

The current research are using class weights in the SVM model to address class imbalance, and conducting additional manual reviews of tweets from underrepresented regions to ensure diverse geographical representation.

Six distinct model variants were constructed and evaluated to identify the model configuration that yields the best performance, especially regarding accuracy. These variants combine the multi-class SVM approach with two methods: One-Against-One (OAO) and One-Against-Rest (OAR). Each technique was coupled with three kernel functions: sigmoid, polynomial, and RBF. Models 1, 2, and 3 employ the OAO method with the sigmoid, polynomial, and RBF kernels. Similarly, models 4, 5, and 6 utilize the OAR method with the corresponding kernels. This comparative analysis aims to determine the optimal model configuration for the given classification task.

C. Model Evaluation

The six model variations were tested using 1,000 test data points, representing 20% of the total dataset. The evaluation was conducted on this 1,000-test data, which accounts for 20% of the overall data. Table 7 displays the performance metrics, including accuracy, precision, recall, and F1-score, obtained from assessing various combinations of the SVM multi-class approach and kernel.

Regarding accuracy, the RBF kernel outperformed the Sigmoid and Polynomial kernels in the OAO and OAR methods, yielding similar accuracy rates of 96% and 95.5%, respectively. The choice of the OAO or OAR method had minimal impact on the RBF kernel, with the OAO method showing a slight advantage. On the other hand, when combined with either the OAO or OAR method, the Sigmoid and Polynomial kernels achieved the same accuracy. So, applying the OAO and OAR methods does not affect the Sigmoid and Polynomial kernels. The accuracy aspect shows that the difference in the performance of the six resulting model variations is not significant, only a 2.3% difference in the highest and lowest accuracy.

Table 7.
Model evaluation

Method	Kernel	Accuracy	Precision	Recall	F1-Score
OAO	Sigmoid	93.7%	94%	94%	94%
OAO	Polynomial	94.2%	95%	95%	94%
OAO	RBF	96%	96%	96%	96%
OAR	Sigmoid	93.7%	94%	94%	94%
OAR	Polynomial	95.2%	95%	95%	94%
OAR	RBF	95.5%	95%	95%	95%

The model that integrated the OAO method with the RBF kernel demonstrated the highest precision at 96%. This superior performance can be attributed to the more significant proportion of True Positive (TP) predictions among all optimistic predictions compared to the alternative model configurations. Conversely, the models that coupled the Sigmoid kernel with the OAO and OAR methods exhibited the lowest precision scores. This suggests that these models had a smaller ratio of True Positive (TP) predictions relative to the total positive predictions compared to the other model variations under consideration.

Likewise, the model that combined the OAO method with the RBF kernel achieved a recall score of 96%. This suggests that, compared to the other models, this model configuration exhibits a more favorable ratio of True Positive (TP) predictions to the total number of positive instances. On the other hand, the models that incorporated the Sigmoid kernel with the OAO and OAR methods demonstrated the lowest recall scores. This implies that these models have a lower proportion of True Positive (TP) predictions relative to the total number of positive instances compared to the alternative model variations.

Consistent with the performance across other evaluation metrics, the model that integrated the OAO method with the RBF kernel also attained the highest F1 score of 96%. This indicates that this specific model configuration strikes a better balance between precision and recall, as measured by their weighted average, than the other models under consideration. Conversely, the models that coupled the Sigmoid kernel with the OAO and OAR methods exhibited the lowest F1 scores. This suggests that these models have a less favorable weighted average of precision and recall than alternative model variations.

In conclusion, the model configuration that emerges as the top performer across all evaluation metrics combines the OAO method and the RBF kernel. Conversely, the models incorporating the RBF kernel with the OAO and OAR methods consistently underperformed, achieving the lowest scores in all aspects. A confusion matrix was employed to examine the class distribution of the best-performing model (OAO with RBF kernel). Table 8 presents the results obtained from the 3×3 confusion matrix analysis. Table 8 shows that the model made 960 correct predictions across the neutral, positive, and

Applied Information System and Management (AISM)
Volume 8, (1) 2025, p. 29–36
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i1.42911

negative classes. Specifically, it correctly predicted 927 neutral cases, 22 positive cases, and 11 negative cases. However, there were also 40 incorrect predictions, representing instances where the model's predicted class label did not match the actual label, whether neutral, positive, or negative. These errors may arise from the training data needing to represent the test data,

Table 8.
Confusion matrix result

Actual	Prediction		
	Neutral	Positive	Negative
Neutral	927	0	0
Positive	26	22	1
Negative	0	13	11

Table 9.
Sentiment analysis result

Class	Total
Neutral	927
Positive	49
Negative	24

leading to fully mismatched predictions. Nevertheless, with an accuracy of 96%, the resulting model variations are highly effective and accurately capture most of the data. The final analysis of public sentiment towards the booster vaccine policy is shown in Table 9.

The sentiment analysis of COVID-19 Booster Vaccines on Twitter revealed a predominantly neutral sentiment, which accounted for 92.7%. This indicates that most tweets expressed neither strong support nor opposition to the booster vaccine policies, reflecting a cautious or indifferent attitude among the public. In contrast, positive sentiment represented 4.9%, suggesting a small but notable portion of tweets expressed approval or optimism toward the booster vaccines. Meanwhile, negative sentiment constituted 2.4%, highlighting limited but significant opposition or concerns about the policy. Neutral sentiment shows that people tend not to have strong views or are not too one-sided. This response may indicate uncertainty, confusion, or inadequate information about the Booster Vaccine policy. As a suggestion for development, the government needs to improve the communication and information provided to the public about the vaccine booster policy. In addition to providing information, the government can involve the public in educational campaigns focusing on booster vaccines' benefits and safety. With these development suggestions, the government can better understand the feelings and needs of the community regarding the booster vaccine policy and strive to ensure that the information provided is clear and accurate to support the handling of COVID-19 more effectively.

IV. CONCLUSION

The study demonstrated that the OAO method combined with the RBF kernel achieved the highest accuracy of 96% in classifying public sentiment on Twitter regarding booster vaccine policies. The sentiment distribution revealed a predominantly neutral sentiment, with 927 neutral, 49 positive, and 24 negative tweets. The OAO method consistently outperformed the OAR approach, particularly when paired with the RBF kernel, while the Sigmoid kernel yielded the lowest accuracy (93.7%) across both methods. These findings provide valuable insights into public sentiment toward booster vaccines, highlighting Indonesians' largely neutral but cautiously accepting attitude. The results can guide policymakers in designing targeted communication strategies to address concerns, improve vaccine acceptance, and ensure equitable implementation of booster programs. Additionally, the study underscores the effectiveness of SVM, particularly the OAO method with the RBF kernel, for sentiment analysis on social media data. However, the research has limitations, including reliance on Twitter data, which may not fully represent the broader population due to demographic biases, and a dataset limited to a specific time frame, potentially missing evolving public sentiment. Future studies could expand the dataset to include other platforms or survey data, conduct longitudinal analyses to track sentiment changes, and explore advanced machine learning techniques, such as deep learning models (e.g., BERT or LSTM), to enhance accuracy and robustness further.

REFERENCES

[1] N. W. P. Y. Praditya, A. K. Syaka, and R. Anggraini, "Correlation Analysis and Prediction of Confirmed Cases of Covid 19 and Meteorological Factor," *Applied Information System and Management (AISM)*, vol. 7, no. 1, pp. 9–16, 2024.

[2] H. Harapan *et al.*, "Drivers of and Barriers to COVID-19 Vaccine Booster Dose Acceptance in Indonesia," *Vaccines (Basel)*, vol. 10, no. 12 (1981), pp. 1–20, 2022.

[3] R. Rahmadyanti and M. Masruloh, "Community Knowledge and Attitude to Conduct Covid-19 Booster Vaccination," *Jurnal Keperawatan Komprehensif (Comprehensive Nursing Journal)*, vol. 8, no. 3, pp. 362–367, 2022.

[4] H. Harapan *et al.*, "Willingness to Pay (WTP) for COVID-19 Vaccine Booster Dose and Its Determinants in Indonesia," *Infect Dis Rep*, vol. 14, no. 6, pp. 1017–1032, 2022.

[5] S. Styawati, A. Nurkholis, E. Winarko, Y. Rahmanto, M. A. Reza, and I. Ismail, "Sentiment Analysis of Indonesian Government Policy using Support Vector Machine-Word2Vec," in *2022 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, Dec. 2022.

[6] T. Aichner, M. Grünfelder, O. Maurer, and D. Jegeni, "Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019," *Cyberpsychol Behav Soc Netw*, vol. 24, no. 4, pp. 215–222, 2021.

[7] E. Chen, K. Lerman, and E. Ferrara, "Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set," *JMIR Public Health Surveill*, vol. 6, no. 2, Art.. no. e19273, 2020.

[8] S. Styawati, A. Nurkholis, A. A. Aldino, S. Samsugi, E. Suryati, and R. P. Cahyono, "Sentiment Analysis on Online Transportation Reviews Using Word2Vec Text Embedding Model Feature Extraction and Support Vector Machine (SVM) Algorithm," in *2021 International Seminar on Machine Learning, Optimization, and Data Science, ISMODE 2021*, 2022. doi: 10.1109/ISMODE53584.2022.9742906.

[9] A. Kumar and G. Garg, "Systematic literature review on context-based sentiment analysis in social multimedia," *Multimed Tools Appl*, vol. 79, pp. 15349–15380, 2020.

[10] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc Netw Anal Min*, vol. 11, Art. No. 81, 2021, doi: 10.1007/s13278-021-00776-6.

[11] A. Budianto, R. Ariyuana, and D. Maryono, "Comparison of K-Nearest Neighbor (Knn) and Support Vector Machine (Svm) in the Recognition of Motor Vehicle Plate Characters," *Jurnal Ilmiah Pendidikan Teknik dan Kejuruan*, vol. 11, no. 1, pp. 27–35, 2019.

[12] F. Fitriana, E. Utami, and H. Al Fatta, "Opinion Sentiment Analysis of the Covid-19 Vaccine on Twitter Social Media Using Support Vector Machine and Naive Bayes," *Jurnal Komtika (Komputasi Dan Informatika)*, vol. 5, no. 1, pp. 19–25, 2021.

[13] M. Azhari, Z. Situmorang, and R. Rosnelly, "Comparison of Accuracy, Recall, and Classification Precision on C4.5 Algorithm, Random Forest, SVM and Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 640–651, 2021.

[14] R. R. N. Bhactiar and D. Hartanti, "Hybrid Decision Tree Method and C4. 5 Algorithm for a Recommendation System in Determining Recipients of Direct Cash Assistance (BLT)," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 5, no. 2, pp. 368–377, 2023.

[15] A. A. A. Mas'amah, F. E. Jelahut, and A. Mallongi, "The Influence of Mass Media Content on the Effectiveness of Covid-19 Vaccination Achievements in East Nusa Tenggara, Indonesia," *Journal of Namibian Studies: History Politics Culture*, vol. 34, pp. 2594–2608, 2023.

[16] A. Nurkholis, D. Alita, and A. Munandar, "Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 227–233, Apr. 2022.

[17] A. Nurkholis and I. S. Sitanggang, "A spatial analysis of soybean land suitability using spatial decision tree algorithm," in *Proceedings of SPIE - The International Society for Optical Engineering*, 2019. doi: 10.1117/12.2541555.

[18] A. Nurkholis, Styawati, D. Alita, A. Sucipto, M. Chanafy, and Z. Amalia, "Hotspot Classification for Forest Fire Prediction using C5.0 Algorithm," in *2021 International Conference on Intelligent Cybernetics Technology and Applications (ICICyTA)*, 2021, doi: 10.1109/ICICyTA53712.2021.9689085.

[19] Q.-T. Phan, Y.-K. Wu, and Q.-D. Phan, "An overview of data preprocessing for short-term wind power forecasting," in *2021 7th International Conference on Applied System Innovation (ICASI)*, 2021, pp. 121–125.

[20] A. Nurkholis, I. S. Sitanggang, Annisa, and Sobir, "Spatial decision tree model for garlic land suitability evaluation," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 666–675, 2021, doi: 10.11591/ijai.v10.i3.pp666-675.

[21] M. I. Alfarizi, L. Syafaah, and M. Lestandy, "Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory)," *JUITA: Jurnal Informatika*, vol. 10, no. 2, pp. 225–232, 2022.

[22] D. A. Otchere, T. O. A. Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *J Pet Sci Eng*, vol. 200, pp. 1–20, 2021.

[23] A. A. Ajhari, "The Comparison of Sentiment Analysis of Moon Knight Movie Reviews between Multinomial Naive Bayes and Support Vector Machine," *Applied Information System and Management (AISM)*, vol. 6, no. 1, pp. 13–20, 2023, doi: 10.15408/aism.v6i1.26045.

[24] S. Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 3, pp. 219–230, 2019.

[25] A. Nurkholis, Styawati, I. S. Sitanggang, Jupriyadi, A. Matin, and P. Maulana, "SVM Multi-Class Algorithm for Soybean Land Suitability Evaluation," in *2022 International Conference on Information Technology Research and Innovation, ICITRI 2022*, 2022. doi: 10.1109/ICITRI56423.2022.9970216.

[26] X. He, Z. Wang, C. Jin, Y. Zheng, and X. Xue, "A simplified multi-class support vector machine with reduced dual optimization," *Pattern Recognit Lett*, vol. 33, no. 1, pp. 71–82, 2012.