International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 6, June - 2022, Pages: 94-101

Performance Comparison of Supporting Vector Machine Method without or with Particle Swarm Optimization Based on Sentiment Analysis WhatsApp Review

Muhammad Syamsul Bahri[1], Agus Hermawan[1], Evlyn Pricilia Kondy[1], Refa Joyce Semida[1], Siswanto[1]*

[1]Department of Statistics, Hasanuddin University, Makassar 90245, Indonesia
*Corresponding author: siswanto@unhas.ac.id

Abstract: Communication technology is undergoing a very rapid development. With this development, it allows humans to communicate remotely using applications available on smartphones. WhatsApp as one of the applications for communicating remotely that offers various advantages to facilitate human communication through smartphones. In addition to its advantages, WhatsApp also has various disadvantages that can be seen in the reviews on the Google Play Store. Reviews on the Google Play Store can be an illustration of the app's eligibility for new users to download. Reviews of the WhatsApp application can be done using sentiment analysis. This study aims to classify positive and negative sentiments on WhatsApp application reviews and compare the performance of the SVM algorithm without and based on PSO. The method used in this study is the method of supporting vector machines without and based on PSO. In the results of the weighting of words using TF-IDF obtained the word "original" with a weight of 0.32. Then the test results, the accuracy of the Support Vector Machine was 79% and the Support Vector Machine based on PSO was 80%. Obtained 1431 positive sentiments and 1069 negative sentiments on the review of the WhatsApp application on the Google Play Store. Based on the results of the paired t-test, it was obtained that there was a significant difference in values between the SVM accuracy values without and based on PSO. The performance of PSO-based SVM algorithms has higher accuracy than the performance of SVM algorithms without PSO.

Keywords: Sentiment analysis, Support Vector Machine, Particle Swarm Optimization, WhatsApp

1. INTRODUCTION

Technology is developing very quickly from time to time. This development brought ease for humans to communicate without being limited by distance [1]. Communication today utilizes smartphones. Indonesia ranks 6th as a user country, which is 73,155,000 or 27.4% of the total Indonesian population [2]. Inside the smartphone there are various applications to communicate. One of them is whatsapp. Whatsapp ranks second on social media, which is a favorite of the Indonesian people with 84% access. Whatsapp is an application used to exchange messages by utilizing a 4G network or wifi for data communication. Whatsapp can be used from various circles because of its conveniences and features. Various features offered by whatsapp are chatting, sharing documents, making phone calls and even videos, sharing location using GPS, and others [3]. In addition to the advantages, there are also disadvantages of this application which are stated in the assessment on the Google Play Store.

Google Play Store is a service provided by Google to provide digital content such as: applications, games, movies, music, and books. The Google Play Store provides an opportunity for users to rate and leave reviews related to the app. Review is an expression in the form of a sentence and contains an assessment or comment to provide feedback on a work. Reviews can be a benchmark for an app's eligibility for new users to download [4]. To process and analyze reviews on the Google Play Store in the form of text, sentiment analysis is needed. Sentiment analysis aims to classify texts by their nature (positive/negative) [5]. Some of the algorithms used in sentiment analysis, such as: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), C 5.0 algorithm, naïve bayes, decision tree, and others [6]. SVM is a sentiment algorithm that can be used for classification and is able to overcome vectors of infinite dimensions on the basis of supervised learning [7]. SVM works on the principle of finding the best hyperplane to separate the two classes in the input space and find the maximum point. The advantage of SVM lies in the fastlearning process and good accuracy in recognizing patterns. However, SVM has a drawback that lies in the difficulty of data with large sample numbers [8]. In general, problems in the world are nonlinear, can solve problems that are nonlinear then can be used kernel functions [9]. Some studies that discuss sentiment analysis using the SVM algorithm include: Wahyuni and Kusumawardhana (2021) using the SVM algorithm in classifying grab application reviews on Google Play Store with an accuracy of 85.54% [10]. Ilmawan and Mude (2020) made a comparison between the naïve bayes algorithm and the Support Vector Machine algorithm and obtained the results that the accuracy of the SVM algorithm was greater than the algorithm naïve bayes was 81.46% [11]. Susanti et al., (2020) used TF-IDF weighting and SVM algorithms in classifying applications in the Google Play Store obtained an accuracy value of 83.3% [12].

Based on previous studies, the SVM accuracy value in review sentiment analysis on the Google Play Store is around 81-85% so it is necessary to optimize the SVM algorithm. One of the algorithms that can optimize the SVM algorithm is the Particle Swarm Optimization (PSO) algorithm [13]. Based on research conducted by Indrayuni (2016) an increase in accuracy value of 5.61% was obtained using the PSO-based SVM algorithm compared to the SVM algorithm [14]. In the study conducted by Rizqi, et al (2022) obtained the SVM accuracy value with PSO experienced an increase in the accuracy value by 0.64% using PSO with parameter 1 and experienced decrease in accuracy value when compared to SVM without PSO when the parameter used is 0.5 [15]. Hernawati and Windu (2019) conducted an accuracy comparison between PSO-based SVM algorithms and PSO-based naïve bayes obtained an increase in accuracy values by 3.3% and 3.6% respectively [16].

In this study, we will compare the performance of pso-based and non-PSO-based Vector Machine Support algorithms on the sentiment analysis of whatsapp application reviews on the Google Play Store. The data used in this study were obtained using the google-play-scrapper library. The purpose of this study is to obtain a classification of positive and negative sentiments in the reviews of the whatsapp application on the Google Play Store and compare the performance of supporting vector machines without and PSO-based in the WhatsApp app review classification.

2. Literature Review

2.1 Text Mining

Text mining is the step of automatically analyzing text by a computer to extract quality information from a series of texts collected in a document [17]. The purpose of text mining is to retrieve and extract useful information from a set of documents [18]. Text mining can generate information by collecting processing, extracting, grouping, and analyzing large amounts of unstructured data [19]. Text mining is used to obtain useful information from a series of documents on the data source in an unstructured format [20].

2.2 Sentiment Analysis

Sentiment analysis is a way of assessing oral or written opinions that can be used to determine whether an opinion is neutral, positive, or negative [21]. Sentiment analysis has the task of grouping polarity in text contained in data, sentences, or features/aspect levels and determining an opinion expressed on data, sentences, or features/aspects that are neutral, positive or negative [22]. In addition, sentiment analysis can also describe emotional feelings of joy, anger, or sadness [23].

2.3 Preprocessing Text

In general, the stages of preprocessing text are as follows

a. Case-folding
Case-folding is the process by which all characters become the same, such as changing a word from uppercase to lowercase. For string data, you can use the string.lower() function to make all letters lowercase, or the string.upper() function to make all letters uppercase, you can also specify your own uppercase letters [24].

b. Remove stopwords
Stopwords are words that have no semantics and are not related to information (usually prepositions and subjects) relevant to the case under study. It is also necessary to compile a special dictionary of stop words, which can be updated by correcting the studied data [24].

c. Tokenizing
Tokenization is the process of dividing a line of a word in a sentence, paragraph, or page into a token or a separate part (called a word) that can stand on its own. In addition to eliminating az when tokenizing characters and characters, the separation of sentences and words is also done based on spaces in sentences. This step also removes certain characters, such as punctuation marks, and turns all tokens into lower case [25].

d. Stemming
Stemming is the process of finding the basic form of a sentence by removing affixes. Stemming is a process in an IR system that uses certain rules to convert words contained in a document into basic words. Root Finding is important for improving the efficiency of Indonesian information retrieval, text document search, and document translation [25].

2.4 Weighting Term Frequency-Inverse Document Frequency (TF-IDF)

Weighting is weighting to see the relationship of a word (term) with a document by giving it a weight [26]. TF-IDF is a combination of term frequency and inverse document frequency. Equation for TF-IDF [27]:

$$TF - IDF(t,d,D) = TF(t,d) \times IDF(t,d,D)$$

In commonality (1) $TF - IDF(t,d,D)$ represents the weight of the word t in document d in corpus D [27]. The greater weight of a word indicates that it appears frequently in documents [6].

2.5 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a method used for prediction in both regression and classification [28]. Support vector machines are included in supervised learning classes and their implementation requires a learning phase using sequential SVM training followed by a testing phase. The support vector machine has a classification concept to find the best hyperplane that serves as a separator between two data classes [29]. The support vector engine can handle multidimensional data sets using kernel tricks [30].

The accuracy of the resulting model uses the SVM transition process and depends largely on the kernel functions and parameters used [31]. Based on its characteristics, the SVM method is divided into two types, namely linear SVM and nonlinear SVM. An SVM is said to be linear if the data can be separated linearly, which then separates the two soft-pigmented classes into hyperplanes. Meanwhile, the non-linear SVM applies kernel trick functions to the high-dimensional space [32].

The Support Vector Machine equation is [33]:

$$f(x) = w.x + b$$

Or

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol.6 Issue 6, Jun – 2022, Pages: 1-9

$$f(x) = \sum_{i=1}^{m} a_i y_i K(x,y) + b$$

Description:

w = the parameter of the hyperplane sought (the line perpendicular between the hyperplane line and the support vector point)

x = Support Vector Machine input data point

a_i = the weight value of each data point

$K(x,y)$ = *kernel* function

2.6 Particle Swarm Optimization (PSO)

The PSO algorithm is a population-based algorithm that utilizes individuals in search. In PSO, populations are called swarms and individuals are called particles. Each particle moves at a speed appropriate for the search area and is kept as the best position ever achieved [34]. In the PSO algorithm, there is the following process [35]:

1. Initialization
   a. Initialize the initial speed. The 0th iteration guarantees that all particles have an initial velocity value of 0.
   b. Initialize the initial position of the particle. At the 0th iteration, the initial position of the particle can be generated by the equation:
   $x = x_{min} + rand[0,1]_x(x_{max} - x_{min})$
   c. Initialize lBest (local best) and gBest (global best). At the 0th iteration, lBest (local best) will be equal to the value of the initial position of the particle. While gBest (global best) is selected from the lBest with the highest fitness.
2. Speed Update

The formula used to perform the speed update is as follows:

$$v_{i,j}^{t+1} = w.v_{i,j}^{t} + c_1.r_1(Pbest_{i,j}^{t} - x_{i,j}^{t}$$
$$c_2.r_2\left(Gbest_{g,j}^{t} - x_{i,j}^{t}\right)$$

Description:

$v_{i,j}$ = i-th individual velocity component on d dimension

w = parameter inertia weight

c_1, c_2 = acceleration constant (learning rate), the value is between 0 to 1

r_1, r_2 = random parameter between 0 to 1

$Gbest_{g,j}$= Gbest (local best) individual g on j dimension

3. Position Update and Fitness Calculation

The formula used in updating the position is as follows:

$$x_{i,j}^{t+1} = x_{i,j}^{t} + v_{i,j}^{t+1}$$

Description:

$x_{i,j}$ = individual position i on j dimensions

4. Update lBest (local best) dan gBest (global best)

Comparing the IBest (local best) in the previous iteration to the results of the position update. Then what will become the new IBest is higher fitness. The latest lBest that has the highest fitness will be the new gBest.

3. Research Methodology

The data is obtained from reviews on the Google Play Store for the WhatsApp application. After the data is received, a cleaning process is carried out so that a total of 1500 sats are obtained which are ready to be processed by sentiment analysis using the Support Vector Machine (SVM) and Particle Swarm Optimization (PSO) methods. processed with R Studio and Python software.

Here are the steps this research was carried out:

1. Sourcing data on the Google Play Store on the WhatsApp application using the text mining method.
2. Data preprocessing with R language to perform Label encoding, Case folding, Cleansing, Tokenization, and filtering for then data ready to be processed sentiment analysis.
3. Sentiment analysis starts with visualization with cloud words, weighting word values, and data mining.
4. Perform sentiment modeling by using the Support Vector Machine (SVM) with python language.
5. Perform sentiment modeling using the Support Vector Machine (SVM) and Particle Swarm Optimization (PSO) with python language.
6. Compare the sentiment results of the two classification models with the significance test and choose the best classification model.

4. Results and Discussion

4.1 Data Collection

The data collection process is carried out by *scraping* using the python programming language at Google Collaboratory. The data obtained is 10,000 data, but data reduction will be carried out to speed up the computational process and eliminate data with inappropriate class values so that data will be used which consists of 2500 Google Play Store Reviews. The data from scraping data is seen in table 1.

Table 1. Whatsapp Application Review Data on Google Play Store

Rating	Review
1	*menurut pengalaman yang ada whatsApp sangat mudah untuk menfasilitas informasi dalam face to face maupun dalam kelompok*
2	*setelah di upgrade, voice note gda suaranya*
3	*Mau nanya min knapa setelah di perbarui klo buat tlfon slalu mati sendiri ada tulisan WhatsApp tidak menangapi*
5	*Apknya baguss bisa buat mengirim pesan ke teman maupun sodara jadi ini apk bagus sih..â¤ï¸*

4.2 Preprocessing Data

4.2.1 Label encoding

In this study, the label class that is expected to consist of two categories of sentiment (Negative and Positive), for ratings with scores of 1 and 2 will be classified as negative sentiment, for rating 4 and 5 will be classified as positive sentiment and for rating a score of 3 will be considered a neutral or unknown sentiment so that for data with a score of 3 will be omitted from data, after encoding process obtained as many as 1431 positive sentiments san 1069 negative sentiments. The encoding label results can be found in table 2.

Table 2. Data from the Encoding Label Process

Rating	Review
Negative	*menurut pengalaman yang ada whatsApp sangat mudah untuk menfasilitas informasi dalam face to face maupun dalam kelompok*
Negative	*setelah di upgrade, voice note gda suaranya*
Positive	*Apknya baguss bisa buat mengirim pesan ke teman maupun sodara jadi ini apk bagus sih..â¤ï¸*

4.2.2 Case Folding

The python programming language has a case-sensitive nature which means that it gives different values between capital and lowercase letters, so it is necessary to apply case folding for the observation of values that has every word in common in uniform observation. The results of case folding can be seen in table 3.

Table 3. Case Folding process result data

Review	**Review (Case Folding)**
menurut pengalaman yang ada whatsApp sangat mudah untuk menfasilitas informasi dalam face to face maupun dalam kelompok	*menurut pengalaman yang ada whatsapp sangat mudah untuk menfasilitas informasi dalam face to face maupun dalam kelompok*
setelah di upgrade, voice note gda suaranya	*setelah di upgrade, voice note gda suaranya*
Apknya baguss bisa buat mengirim pesan ke teman maupun sodara jadi ini apk bagus sih..â¤ï¸	*apknya baguss bisa buat mengirim pesan ke teman maupun sodara jadi ini apk bagus sih*

4.2.3 Cleansing

The text data obtained from the data collection does not fully contain alphabetic letters, but may contain emoticons, symbols, or letters in different languages. In this study, it will be uniformed to use alphabetical letters, so it is necessary to do cleansing on. The results of the cleansing process are found in table 4.

Table 4. Cleansing result data

Review	Review (Cleansing)
menurut pengalaman yang ada whatsapp sangat mudah untuk menfasilitas informasi dalam face to face maupun dalam kelompok	*menurut pengalaman yang ada whatsapp sangat mudah untuk menfasilitas informasi dalam face to face maupun dalam kelompok*
setelah di upgrade, voice note gda suaranya	*setelah di upgrade voice note gda suaranya*
apknya baguss bisa buat mengirim pesan ke teman maupun sodara jadi ini apk bagus sih..â¤ï¸	*apknya baguss bisa buat mengirim pesan ke teman maupun sodara jadi ini apk bagus sih*

4.2.4 Tokenization

The tokenizing process is carried out to break words into units of words. The process is carried out to break down the data so that it can be used in the stemming process, stopword removal and slang word removal.

Table 5. Tokenization result data

Review	Review (Tokenization)
menurut pengalaman yang ada whatsapp sangat mudah untuk menfasilitas informasi dalam face to face maupun dalam kelompok	*[menurut, pengalaman, yang, ada, whatsapp, sangat, mudah, untuk, menfasilitas, informasi, dalam, face, to, face, maupun, dalam, kelompok]*
setelah di upgrade voice note gda suaranya	*[setelah, di, upgrade, voice, note, gda, suaranya]*
apknya baguss bisa buat mengirim pesan ke teman maupun sodara jadi ini apk bagus sih	*[apknya, baguss, bisa, buat, mengirim, pesan, ke, teman, maupun, sodara, jadi, ini, apk, bagus, sih]*

4.2.5 Filtering

Filtering is removing Slang Word (slang), Stop Word and do Stemming. Reviews obtained from the Google Play Store are not all arranged according to good and correct Indonesian rules and not all of them are standard Indonesian words, the influence of words it is in many cases not very significant because it is not a commonly used word. Package Sastrawi in python programming language provides a feature to eliminate stopwords and slang words from the text data owned. The results of processing can be seen in the table.

Table 6. Filtering result data

Review	Review (removing slang word, stopword and stemming)
[menurut, pengalaman, yang, ada, whatsapp, sangat, mudah, untuk, menfasilitas, informasi, dalam, face, to, face, maupun, dalam, kelompok]	*alam mudah menfasilitas informasi face face kelompok*
[setelah, di, upgrade, voice, note, gda, suaranya]	*upgrade voice note suara*

4.3 Sentiment Analysis

4.3.1 Visualization

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol.6 Issue 6, Jun – 2022, Pages: 1-9

Figure 1. Combined Sentiment (Positive and Negative).

Based on the visualization with the cloud word above, it can be seen that for some words that are found in many whatsapp reviews on the Google Play Store are the words "good", "update", "send ", "video", "message" and "status", in order a better conclusion can be obtained, so for cloud work it will be distinguished into two categories of sentiment as shown below.

Positive Sentiment Negative Sentiment

Figure 2. Positive and negative sentiments

Based on the results of word cloud visualizations for different sentiment categories, it can be seen in the positive sentiment of many reviews that contain the words "good", "communication", "love" and "cent". On the negative sentiment, many reviews contain the words "Update", "Video", "Video" and "Block". It can be concluded that some whatsapp users feel they have a positive sentiment towards whatsapp because it is a good communication tool while most of the negative sentiment has to do with update on WhatsApp application.

4.3.2 Word Value Weighting

Word weighting is a score value that is based on the appearance of a word or term in a particular document (TF) and the number of documents containing that word (IDF). It shows how important a word is and also how much it is in the whole document. The TF-IDF values can be seen in table 7.

Table 7. TF-IDF value

Words	**TF-IDF**
Original	0.32
Updated	0.19
photograph	0.18
result	0.27
How	0.30
return	0.29
send	0.15
quality	0.23
Person	0.20
step	0.20
message	0.17
Synchronization	0.31
appropriate	0.31
video	0.16
web	0.24

The results above are the result of calculating the TF-IDF value of several words that are not zero, which means that in addition to the words in the table of results above have a value of TF-IDF = 0 in this study. Words that have a high TF-IDF value are considered to be features that contribute greatly to the classification of models. In this study, the word with a weight value of 0 is still stored because if it is deleted, the model will lose very many features, allowing observations that are not has a predictor value.

4.3.3 Data Mining

The stage of data mining carried out in this study is to model the Support Vector Machine using a radial base function or RBF kernel, the reason for using the RBF kernel type in the study is because this function in general, it is widely used in modeling problems that are non-linear kernel learning. The division of data trains and data sets was carried out with a value of 20% of test data and 80% of data trains, in line with 5 iterations of K-Folds Cross validation used in this study.

4.3.4 Support Vector Machine (SVM)

Modeling using a support vector machine algorithm with K-Fold Cross Validation of 5 iterations resulted in confusion matrix values and accuracy as detailed in table 8.

Table 8. SVM model accuracy results without PSO

K-Folds Cross validation					
	Iteration				Average
	I	II	III	IV	
True Positive	210	214	214	193	207
False Positive	59	62	62	55	59
True Negative	280	297	281	293	287
False Negative	76	52	68	84	70
Accuracy	0.78	0.82	0.79	0.77	0.79
Precision	0.83	0.83	0.82	0.84	0.83
Recall	0.79	0.85	0.80	0.78	0.8

Based on the table values above, an average accuracy value of 79% was obtained in the test data so that it can be said that the model has a fairly good estimation. The precision value of the model is 83% or the accuracy of the information provided by the model is quite good, and the model's ability to find information can be said to be good at 80%.

4.3.5 *Support Vector Machine - Particle Swarm Optimization* (SVM-PSO)

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol.6 Issue 6, Jun – 2022, Pages: 1-9

Modeling using the Support Vector Machine – Particle Swarm Optimization (SVM-PSO) algorithm with K-Fold Cross Validation of 4 iterations resulted in confusion matrix values and accuracy as detailed in tables 9 and 10, differences between the two tables lies in the particle values and the iterations attempted on the PSO method.

4.3.6 Model SVM-PSO 1

Table 9. SVM-PSO with particle = 5 and iteration = 50

K-Folds Cross validation					
	Iteration				Average
	I	II	III	IV	
True Positive	208	219	216	194	209
False Positive	61	57	60	54	58
True Negative	290	297	280	294	290
False Negative	66	52	69	83	67
Accuracy	0.79	0.83	0.80	0.79	0.80
Precision	0.83	0.84	0.83	0.85	0.84
Recall	0.82	0.86	0.80	0.80	0.82

4.3.7 Model SVM-PSO 2

Table 10. SVM-PSO with particle = 5 and iteration = 100

K-Folds Cross validation					
	Iteration				Average
	I	II	III	IV	
True Positive	208	218	217	197	210
False Positive	61	58	59	51	57
True Negative	290	298	280	292	290
False Negative	66	51	52	85	63
Accuracy	0.80	0.83	0.80	0.79	0.80
Precision	0.83	0.84	0.83	0.85	0.84
Recall	0.81	0.85	0.80	0.77	0.81

In the table above the values of accuracy, precision and recall that the two models have are almost the same, in the iteration above, an output value is also obtained in the form of maximum values for the parameters C and gamma (γ) generated by the iteration of the PSO method, where the value of C is a regularization parameter to determine how much optimization should avoid misclassification in each *training* data and gamma determines how much influence each has observation of data train. For the first model and the second model have different values on the results of the best parameters issued, for PSO with 50 iterations have a value of C = 2.96 and gamma = 0.16 then for PSO optimization with 100 iterations having a value C = 3.06 and gamma = 0.19. PSO optimization models for models with 50 and 100 iterations have almost the same accuracy, precision and recall values.

4.4 Significance Test

To test whether the model with the PSO has significant differences in accuracy values. T-tests of paired samples were carried out at the average value of the accuracy of the model that had been modeled in the previous section. With the hypothesis as follows:

$H_0 : \mu_1 - \mu_2 = 0$ (no difference between the average accuracy value of models without PSO and models with PSO)

$H_1 : \mu_1 - \mu_2 \neq 0$ (there is a significant difference between the average accuracy value of models without PSO and models with PSO)

Table 11. The significance value of the t-test of paired samples

	P-value
SVM-SVM PSO model 1	0.015
SVM-SVM PSO model 2	0.013

With a degree of significance$\alpha = 5\%$ obtained the value $\mathrm{p-value}$ for the first model pair of 0.015 and the second model pair of 0.013, which means that it H_0 is rejected or there is a significant difference in the accuracy value of the model with or without PSO, where the average for SVM without PSO is 79% and SVM with PSO is 80%, meaning that the accuracy value of SVM with PSO is 80%, meaning that the accuracy value of SVM with PSO better than SVM without PSO. Based on the test results, it was concluded that it is better to use the SVM-PSO model in classifying the problem of reviewing the WhatsApp application because it has a better accuracy value.

5. Conclusion

The results of the study on the review data of 2500 data from the Google Play Store for the WhatsApp application obtained two of the best sentiment models, namely, the Support Vector Machine (SVM) model and Particle Swarm Optimization (PSO), with 1439 observations have positive sentiments and 1069 observations have negative sentiments.

A comparison of the SVM method with or without PSO results in a table of accuracy, precision and recall values as in table 12.

Table 12. Summary of sentiment results with SVM and/or without PSO

No	Model	Accuracy	Precision	Recall
1	SVM	79%	83%	80%
2	SVM-PSO	80%	84%	82%
3		80%	84%	81%

After being tested using a paired t-test. A conclusion was reached to reject H_0 which means that there is a significant difference in the accuracy value of the model with or without pso, so it is concluded that the SVM-PSO model is better used because it has better accuracy.

References

[1] Habibie, A., & Rachmawati, I. (2020). Analisis Preferensi Konsumen Dalam Memilih Smartphone di Indonesia. *eProceedings of Management*, *7*(**1**), 114-124.

[2] Saepulrohman, A., Saepudin, S., & Gustian, D. (2021). Analisis Sentimen Kepuasan Pengguna Aplikasi Whatsapp Menggunakan Algoritma NaÃ¯ve Bayes Dan Support Vector Machine. *Accounting Information Systems and Information Technology Business Enterprise*, *6*(**2**), 91-105.

[3] Rusni, A., & Lubis, E. E. (2017). Penggunaan Media Online Whatsapp dalam Aktivitas Komunitas One Day One Juz (ODOJ) dalam Meningkatkan Minat Tilawah ODOJER di Kota Pekanbaru. *JOM FISIP*. *2*(**1**), 1-15

[4] N. Herlinawati *et al*. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine. *Journal of Computer Engineering System and Science*, *5*(**2**), 502–714.

[5] Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm for Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(**2**), 121-130.

[6] Mulyani, S. (2022). Perbandingan Kinerja Metode Mesin Pendukung Tanpa dan Dengan Fitur Berbasis Leksikon Pada Analisis Sentimen Vaksinasi Covid-19 di Indonesia. *Skripsi*. Fakultas MIPA Universitas Hasanuddin, Makassar.

[7] l Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, *127*(**1**), 511-520.

[8] Budianto, A., Ariyuana, R., & Maryono, D. (2018). Perbandingan K-Nearest Neighbor (KNN) Dan Support Vector Machine (SVM) Dalam Pengenalan Karakter Plat Kendaraan Bermotor. *Jurnal Ilmiah Pendidikan Teknik dan Kejuruan*, *11*(**1**), 27-35.

[9] Suhardjono, S., Wijaya, G., & Hamid, A. (2019). Prediksi Waktu Kelulusan Mahasiswa Menggunakan Svm Berbasis Pso. *Bianglala Informatika*, *7*(**2**), 97-101.

[10] Wahyudi, R., & Kusumawardana, G. (2021). Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, *8*(**2**), 200-207.

[11] Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. *ILKOM Jurnal Ilmiah*, *12*(**2**), 154-161.

[12] Fransiska, S., Rianto, R., & Gufroni, A. I. (2020). Sentiment Analysis Provider by. U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method. *Scientific Journal of Informatics*, *7*(**2**), 203-212.

[13] Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi|* *9*(**2**), 162-170.

[14] Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. *EVOLUSI: Jurnal Sains dan Manajemen*, *4*(**2**), 20-27.

[15] Darmawan, R., Indra, I., & Surahmat, A. (2022). Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru di Twitter. *Jurnal Kajian Ilmiah*, *22*(**2**), 143-152.

[16] Hernawati, H. (2019). Sentimen Analisis Operasi Tangkap Tangan KPK Menurut Masyarakat Menggunakan Algoritma Support Vector Machine, Naïve Bayes Berbasis Particle Swarm Optimization. *Faktor Exacta*, *12*(**3**), 230-243.

[17] Julianto, Y., Setiabudi, D. H., & Rostianingsih, S. (2022). Analisis Sentimen Ulasan Restoran Menggunakan Metode Support Vector Machine. *Jurnal Infra*, *10*(**1**), 1-7.

[18] Wijaya, A. P., & Santoso, H. A. (2016). Naïve Bayes Classification pada klasifikasi dokumen untuk identifikasi konten E-Government. *Journal of Applied Intelligent System*, *1*(**1**), 48-55.

[19] Fu, W., Choi, E. K., & Kim, H. S. (2022). Text mining with network analysis of online reviews and consumers' satisfaction: A case study in Busan wine bars. *Information*, *13*(**3**), 127.

[20] Samsir, S., Ambiyar, A., Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, *5*(**1**), 157-163.

[21] Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, *9*(**3**), 483.

[22] Kurniawan, I. (2022). Analisis Sentimen Vaksin Covid-19 Di Indonesia Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization. *Skripsi*. Fakultas MIPA Universitas Hasanuddin, Makassar.

[23] Taufik, I. (2018). Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (SVM). *Thesis*. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah, Jakarta.

[24] Astuti, S. P. (2020). Analisis sentimen berbasis aspek pada aplikasi tokopedia menggunakan lda dan naïve bayes. *Thesis*. Fakultas Sains dan Teknologi UIN Syarif Hidayatullah, Jakarta.

[25] Alfanzar, A. I. (2019). Topic Modelling Skripsi menggunakan metode Latent Dirichlet Allocation. *Disertasi*. UIN Sunan Ampel, Surabaya.

[26] Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan

Pembobotan TF-IDF dan Cosine Similarity. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.*, *2*(**1**), 306-312.

[27] Jiang, Z., Gao, B., He, Y., Han, Y., Doyle, P., & Zhu, Q. (2021). Text classification using novel term weighting scheme-based improved TF-IDF for Internet media reports. *Mathematical Problems in Engineering*, *2021*.

[28] Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. (2021). Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter. *Journal of Information System Management (JOISM)*, *3*(**1**), 16-21.

[29] Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer genomics & proteomics*, *15*(**1**), 41-51.

[30] Ly, T. V., & Son, D. V. T. (2022). Churn prediction in telecommunication industry using kernel Support Vector Machines. *PloS one*, *17*(**5**), 1-18.

[31] Nurkholis, A., Alita, D., & Munandar, A. (2022). Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *6*(**2**), 227-233.

[32] Gupta, A. K., Singh, V., Mathur, P., & Travieso-Gonzalez, C. M. (2021). Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. *Journal of Interdisciplinary Mathematics*, *24*(**1**), 89-108.

[33] Filemon, B., Mawardi, V. C., & Perdana, N. J. (2022). Penggunaan Metode Support Vector Machine Untuk Klasifikasi Sentimen E-Wallet. *Jurnal Ilmu Komputer dan Sistem Informasi*, *10*(**1**), 1-6.

[34] Zhu, P., Saadati, H., & Khayatnezhad, M. (2021). Application of probability decision system and particle swarm optimization for improving soil moisture content. *Water Supply*, *21*(**8**), 4145-4152.

[35] Darmawan, R., Indra, I., & Surahmat, A. (2022). Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru di Twitter. *Jurnal Kajian Ilmiah*, *22*(**2**), 143-152.