

A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification

Styawati*¹, Khabib Mustofa²

¹Department of Information System, Universitas Teknokrat Indonesia, Lampung, Indonesia

²Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

e-mail: *styawati@teknokrat.ac.id, khabib@ugm.ac.id

Abstrak

Analisis sentimen yang digunakan dalam penelitian ini merupakan proses klasifikasi teks ke dalam dua kelas yaitu kelas negatif dan positif. Metode klasifikasi yang digunakan adalah Support Vector Machine (SVM). Keberhasilan klasifikasi metode SVM bergantung pada koefisien soft margin C , serta parameter σ dari fungsi kernel. Maka dari itu dibutuhkan kombinasi parameter SVM yang tepat untuk melakukan klasifikasi data opini film. Penelitian ini menggunakan metode Firefly sebagai metode optimasi parameter SVM. Dataset yang digunakan dalam penelitian ini adalah data opini masyarakat terhadap beberapa film. Hasil penelitian ini menunjukkan bahwa Firefly Algorithm (FA) dapat membantu SVM untuk mendapatkan kombinasi parameter yang sesuai berdasarkan akurasi dengan waktu eksekusi lebih singkat. Hal ini dibuktikan dengan hasil pengujian SVM dan FA-SVM dengan rentang nilai $C=1.0-3.0$ dan $\sigma=0.1-1.0$ menghasilkan akurasi tertinggi yaitu 87.84%. Waktu eksekusi SVM adalah 5928 detik, sedangkan FA-SVM membutuhkan waktu 2330 detik. Selisih waktu antara SVM dan FA-SVM adalah 3598 detik. Pengujian berikutnya dengan rentang nilai $C=1.0-3.0$ dan $\sigma=1.0-20$. Berdasarkan rentang tersebut, metode SVM maupun metode FA-SVM menghasilkan akurasi tertinggi yaitu 87.15%. Waktu eksekusi yang dibutuhkan oleh metode SVM adalah 7205 detik, sedangkan waktu yang dibutuhkan oleh metode FA-SVM adalah 2388 detik. Selisih waktu antara metode SVM dengan metode FA-SVM adalah 4817 detik.

Kata kunci— Optimasi, Klasifikasi, SVM, FA-SVM.

Abstract

The sentiment analysis used in this study is the process of classifying text into two classes, namely negative and positive classes. The classification method used is Support Vector Machine (SVM). The success of the SVM method classification depends on the soft margin coefficient C , as well as the σ parameter of the kernel function. Therefore we need the right combination of SVM parameters to classify movie opinion data. This study uses the Firefly method as an SVM parameter optimization method. The dataset used in this study is public opinion data on several movies. The results of this study indicate that the Firefly Algorithm (FA) can help SVM to get the appropriate combination of parameters based on accuracy with shorter execution times. This is evidenced by the results of SVM and FA-SVM testing with a range of values $C = 1.0-3.0$ and $\sigma = 0.1-1.0$ resulting in the highest accuracy of 87.84%. SVM execution time is 5928 seconds, while FA-SVM takes 2330 seconds. The time difference between SVM and FA-SVM is 3598 seconds. The next test with a range of values $C=1.0-3.0$ and $\sigma=1.0-20$. Based on this range, the SVM method and the FA-SVM method produced the highest accuracy of 87.15%. The execution time required by the SVM method is 7205 seconds, while the time required by the FA-SVM method is 2388 seconds. The time difference between the SVM method and the FA-SVM method is 4817 seconds.

Keywords— Optimization, Classification, SVM, FA-SVM.

1. INTRODUCTION

Classification is the process of grouping many data into classes that have been determined and given according to the similarity of the characteristics and patterns contained in these words [1]. Classification is mostly done in various case studies including tone recognition, text categorization, image classification, protein structure prediction, data/document classification, etc[2]. There are many classification methods including Naive Bayes, Support Vector Machine (SVM), Maximum Entropy, J48 [3]. The classification in this study is used to distinguish between the data entered in the positive class and the negative class. Research conducted by [4], comparing three classification methods namely Naive Bayes (NB), maximum entropy and Support Vector Machine (SVM). The best result is SVM.

SVM is a machine learning technique that relies on the concept of statistical learning. SVM has good generalization capabilities in cases with small samples [5]. But on the other hand, SVM cannot choose the appropriate parameters so that the use of parameters is not optimal. Using the appropriate parameters is expected to increase the accuracy of SVM [3][5]. The importance of using the appropriate parameter values also explained by [6] is said that the success of the SVM model depends on the soft-margin coefficient C , as well as the parameters of the kernel function. So, choosing optimal parameters for SVM is one important step when using SVM as a classification method. From the SVM shortcomings presented by [5], various SVM methods developed, including Support Vector Machine and Particle Swarm Optimization (SVM-PSO) [5][3], Firefly Algorithm and Support Vector Machine (FA-SVM), Accelerated Particle Swarm Optimization and Support Vector Machine (APSO-SVM) [7] and so on. The research conducted by [7] used heart, diabetes, liver, iris, and cancer datasets using the PSO-SVM, APSO-SVM, and FA-SVM algorithms obtained the highest accuracy results from the combination of the FA-SVM algorithm. Based on the research conducted by [7], the authors are interested in classifying movie opinion data using the FA-SVM method, then the resulting accuracy will be compared with the SVM method without parameter optimization.

FA-SVM is an SVM classification method combined with the Firefly Algorithm (FA). FA is an optimization method based on flashing patterns and firefly behavior [8]. Whereas SVM is a machine learning technique that relies on the concept of statistical learning [5].

2. METHODS

2.1 Research Design

The research design of movie opinion data classification using SVM and FA-SVM method which is built in general consists of four stages, namely data preprocessing, weighting, SVM classification, and FA-SVM classification. Preprocessing is the stage to produce a collection of words that are ready to be processed and used as input at a later stage. Weighting is the process of assigning weight values to each term. The weighting process uses TF-IDF. The SVM classification is the process of classification of Indonesian language movie opinion data into two classes, namely negative class or positive class. The FA-SVM classification is a process of classification of Indonesian opinion movie data using the SVM model, but before the classification process is carried out the search process for the combination of SVM parameter values is done beforehand. The search for SVM parameter combination aims to get the highest accuracy value. The parameters to be optimized are C and σ . Parameter C is a parameter used to control SVM in controlling errors [9]. While the σ parameter is an SVM parameter that is used to find the optimal value in each dataset[10]. The method used in searching SVM parameters is Firefly Algorithm (FA).

Classifier performance evaluation is conducted to find out whether FA-SVM can be used to classify movie opinion data with shorter time. The method used for classifier performance evaluation in this study is K-Fold Cross Validation and Confusion Matrix. The cross validation

process is done by dividing the data into two segments. The first segment is used to train the model while the other segment is used to validate the model. The classifier performance evaluation process uses the Confusion Matrix method, where this method contains the actual classification and predictions made by the classification system. The performance of the system is generally evaluated using data in the matrix. Figure 1 shows the stages of classification of movie opinion data.

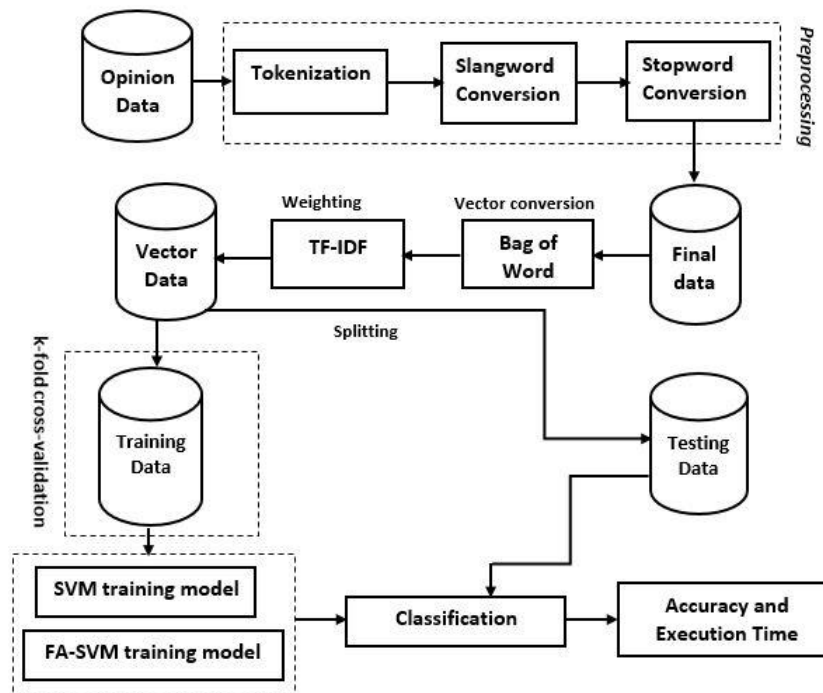


Figure 1 Stages of classification

The first stage of the system architecture is collecting data. Data obtained from the scraping process that has been carried out by previous researchers is [11]. The data produced by the previous researcher was preprocessing using the case folding method and normalization of features. Then in this study coupled with the tokenization process, slang word conversion, and stopword conversion. In this study coupled with the slang words and stopword conversion process because the preprocessing data results in previous studies still found errors in writing words such as "gak jelas" and many words that do not have sentiments such as "yaitu". After the preprocessing process, then the words are mapped into vector models using Bag-of-Word. After the data is in the form of a vector, then weighted by using the term Frequency-Inverse Document Frequency (TF-IDF) method.

The weighting of words in this study was carried out to determine the word weights in the movie opinion data so that we got word weights that could represent the basis of information that would be used to determine the classification of the data. After the weighting process, data is then divided. Data is divided into 2 namely training data and testing data. Training data is divided into 2 namely training data and validation data. The method used in the distribution of training data and data testing is the Splitting method, while the method for dividing training data and validation data is the K-Fold Cross Validation method. Training data is used to model SVM and FA-SVM classification. While the validation data to find the best parameters. Then the model is tested with test data which aims to measure the extent to which the classifier successfully performs the classification process. Table 1 shows examples of training data, validation data, and testing data.

Table 1 Training Data, Validation Data, Testing Data

Document	Opinion	Sentiment	Data Type
D1	membosankan	Negative	Training
D2	film aneh	Negative	
D3	sedikit membosankan	Negative	
D4	saya bosan lihat film ini	Negative	
D5	kamu akan kecewa	Negative	
D6	bagus sekali	Positive	
D7	tidak rugi nonton film ini	Positive	
D8	gak nyangka filmnya bakal seru	Positive	
D9	sumpah keren	Positive	
D10	film yang bagus	Positive	Validation
D11	saya bilang keren	Positive	
D12	alur ceritanya agak bosenin	Negative	
D13	kenapa saya kecewa	Negative	Testing
D14	tempatny keren ceritanya juga keren	Positive	
D15	film dengan tema bagus	Positive	

The opinion data in Table 1 shows that the number of training data is nine data, validation data are three data, and the test data is three data. Then from the training data classification is done to produce a classifier model that will be used to classify the test data.

2.2 Preprocessing

Preprocessing is one of the essential steps in sentiment analysis. Data obtained from previous research only through two preprocessing processes namely case folding and normalization of features, so there are still abbreviations or typos and words that have no sentimental value. Examples of abbreviated words or typos and words that do not have sentiment values are "gak", "membosankan", "yaitu", and so on. The purpose of the preprocessing process is to get clean data so that the process of making word vectors and sentiment classifications becomes more accurate. The preprocessing method added to this study is tokenization, slang word conversion, and stopword removal conversion. Tokenization functions to break comments into units of words. The tokenization process is done by looking at each space in the comment, so based on these spaces comments can be broken down. After the tokenization process a stopword conversion is performed. Stopword is the process of deleting words that are included in the stopword list. Stopword is general words that appear in large numbers that have functions but have no meaning. Examples of stop words are "yang", "yaitu", and so on. Dictionary Stopword in this study was taken from a dictionary made by Tala on Stopwords ID. The last preprocessing is slang word conversion. Slang word conversion is the process of changing to non-standard words to standard words. This stage is done using the help of the slang word dictionary and its equivalent in standard words. This stage will check the words contained in the slang word dictionary or not. If the non-standard word is in the slang word dictionary, the non-standard word will be changed to the default word in the slang word dictionary. The author made the slang word dictionary in this study based on research conducted by[12].

2.3 Weighting

Weighting is the process of assigning weight values to each term in each document [13]. In this study the weighting method used was TF-IDF. TF-IDF is a combination of Term Frequency (TF) and Inverse Document Frequency (IDF) used in calculating the weight of each word (term) in each document. The calculation scheme of TF-IDF is shown in Equation 1.

$$TF - IDF_{t,d} = tf_{t,d} \times \log \frac{N}{df_t} \quad (1)$$

Information:

$tfidf_{t,d}$: Weight of each word from the term t contained in the document d

$tf_{t,d}$: Term frequency which is the weight of the term t of the document d

idf : Many documents containing the term are searched

t : term or word

2.4 Classification with the SVM method

Classification is the process of grouping many data into classes that have been determined and given according to the similarity of the characteristics and patterns contained in these words. In general, the classification process begins with the provision of any data which are used as references to make data classification rules. The data is usually known as training sets. From the training sets, a model is then made to classify the data. The model is then used as a reference for classifying unknown data classes known as test sets [1]. The classification method used in this study is SVM.

Support Vector Machine (SVM) is one of the supervised machine learning algorithms that have outstanding performance[9]. Support Vector Machine (SVM) is linear classifier based on the principle of maximizing margins. SVM uses the hyperplane optimally to classify data into two groups of data in a higher dimensional space [13]. The margin is the distance between the hyperplane and the closest data from each class. The closest data is called support vector[9]. The hyperplane is the best separator between two predetermined classes[9]. The basic principle of SVM is a linear classifier and then developed to work on non-linear problems. By incorporating the kernel trick concept in a high-dimensional workspace[14]. The SVM kernel used in this study is the RBF kernel for the transformation process from the input space into feature space.

The SVM method has the central concept in classifying data, namely finding the best hyperplane to separate between two predetermined classes[9]. The best hyperplane is obtained by maximizing the margin support vector. The process of maximizing support vector margins can be done by minimizing lagrangian and being reduced to w and b found in equation 1 with terms 1 and 2.

$$Lp = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i + b) - 1 \quad (2)$$

Terms 1:

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (3)$$

Terms 2:

$$b = y_i - w \cdot x_i \quad (4)$$

Because the value of α is unknown, the values w and b cannot be determined. The value of α is sought by maximizing the Lagrangian multiplier with the optimal conditions for its duality using the Karush-Kuhn-Tucker (KKT) constraint. The use of KKT constraints makes the Lagrange multiplier (α) value equal to the amount of training data.. The process of maximizing the Lagrangian multiplier still has many possible values of w , b , and α . Based on these problems, the maximization process of Lagrange multiplier must be transformed into Lagrange multiplier duality in equation 5 with constraints 1 and 2.

$$\text{Maks } Ld = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (5)$$

Terms 1:

$$\sum_{i=1}^N \alpha_i y_i = 0, \quad i = 1, 2, \dots, N \quad (6)$$

Terms 2:

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N \quad (7)$$

After the values of w , b , and α are obtained, then determine the label using the SVM model.

$$F(\Phi(x)) = \text{sign}(w \cdot \Phi(x) + b) \quad (8)$$

If the value of $f(x)$ produced is $f(x) > 0$ then the data is classified into a positive class (+1), if $f(x) < 0$ then the data is classified into a negative class (-1).

2.5 Classification with the FA-SVM Method

The process of data classification with the FA-SVM method begins with initializing the parameters needed for the search process with the firefly algorithm, which is determining the number of firefly population (`number_of_fireflies`), generation (`maximun_generation`), initial attractiveness coefficient (β_0), light absorption coefficient (γ), and random parameter coefficient (α). After initializing the parameters needed, then optimize the parameters C and σ using the Firefly Algorithm method. The optimization uses several steps, namely [15]:

1. Determine the objective function $f(x)$
2. Initialize the population of firefly noodles ($i = 1, 2, \dots, N$)
3. Define the light absorption coefficient γ
4. Calculate the distance

The distance between two fireflies i and j at the position of the x_i and x_j coordinates is the cartesian distance which is formulated as

$$r_{ij} = \|m_i - m_j\| = \sqrt{\sum_{k=1}^d (m_{i,k} - m_{j,k})^2} \quad (9)$$

If an equation uses dimensions ($d = 2$) then the above equation becomes [16]

$$r_{ij} = \sqrt{(m_i - m_j)^2 + (n_i - n_j)^2} \quad (10)$$

5. Calculate Attractiveness

Attractiveness of a firefly is proportional to the intensity of light seen by other fireflies. Attractiveness is formulated by:

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (11)$$

Information:

$\beta(r)$ = Attractiveness of fireflies at distance r

β_0 = Attractiveness at distance 0

γ = coefficient of light absorption

r = distance between source fireflies and fireflies

6. Calculate the movements of fireflies

Firefly movements that are attracted to fireflies j (which are brighter or have higher attractiveness) are formulated as

$$m_i = m_i + \beta_0 e^{-\gamma r_{ij}^2} (m_j - m_i) + \alpha \left(\text{rand} - \frac{1}{2} \right) \quad (12)$$

The parameter values used in general are $\beta_0 = 1$ and $\alpha \in [0,1]$. The randomization process can be done using a normal distribution of $N(0,1)$ or another distribution.

After the values C and σ are obtained from each firefly then the value is used to train the data using the SVM method. After the data is trained using the SVM model, then calculate the accuracy of each C and σ produced by fireflies. Then ranking is done to determine the most optimal parameter values C and σ . Then the values of C and σ are used to model the SVM classifier then the model is tested using test data. From the test results, accuracy will be obtained based on the test data.

2.6 Testing

Classification testing is carried out using the SVM and FA-SVM methods. The dataset used in this test is the movie opinion dataset. The purpose of this test is to analyze whether the classification with the FA-SVM method produces good accuracy with a faster processing time compared to the SVM method without parameter optimization. Tests are carried out with several combinations of SVM parameters and several combinations of populations and generations.

3. RESULTS AND DISCUSSION

3.1 Testing Scenarios

The first test is done by the SVM method without parameter optimization. Tests are carried out with a range of SVM parameter values $C=1.0-3.0$ and $\sigma=0.1-1.0$. The next test uses the range $C = 1.0-3.0$ and $\sigma = 1.0-2.0$. The difference used is 0.01. Experiment with some of these combinations to find out the value of a better combination based on the highest accuracy. The second test was carried out using the FA-SVM method using several combinations of population and generation with the same goal as SVM testing without optimization. The last test is to compare the processing time required by the SVM and FA-SVM methods. The comparison is done to find out whether the FA-SVM method can be used to classify movie opinion data with good accuracy results and have a faster processing time compared to SVM without parameter optimization.

3.2 Testing the SVM method without parameter optimization

Testing the SVM method is done by using the k-fold cross validation approach. The k fold cross validation process in this study uses 10 fold ($k = 10$) with the same partition data size and is done randomly. The data in the fold will be divided into 10 subsets so that each subset has the same size and has different data. Using 10 fold of 2179 data, the distribution of 436 data as test data and 1743 data were divided into training data. The process will be repeated as many as 10 fold with different data distribution. This test uses a combination of SVM parameters with a range of values $C=1.0-3.0$ and $\sigma=0.1-1.0$. The test produced 18000 combinations of C and σ with the highest accuracy of 87.84%. The processing time needed is 5928 seconds. Figure 2 shows the results of SVM testing with a range of values $C = 1.0-3.0$ and $\sigma = 0.1-1.0$.

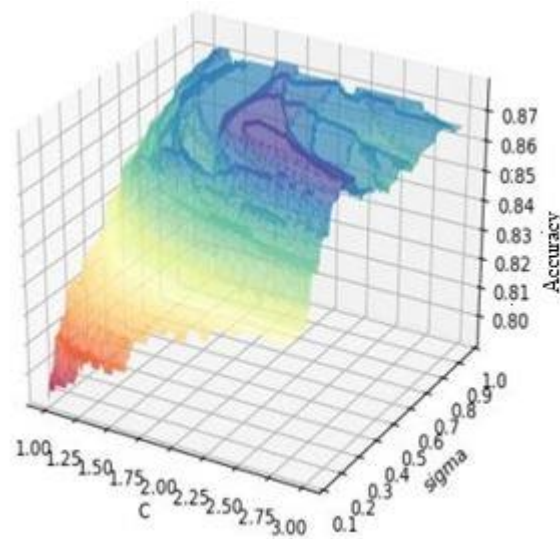


Figure 2 SVM Test Results $C = 1.0-3.0$ and $\sigma = 0.1-1.0$

Figure 2 shows that the higher the value of C and the value of σ then the accuracy obtained increases, but at a certain point when the values of C and σ are high, the accuracy obtained decreases. The example at $C = 2.33$ and $\sigma = 0.45$ produces an accuracy of 87.84%, but when the SVM parameter values are elevated to $C = 2.84$ and $\sigma = 0.81$ the resulting accuracy is 86.69%.

Testing the next SVM method uses a range of parameter values $C = 1.0-3.0$ and $\sigma = 1.0-2.0$. The test produced 20000 combinations of C and σ with the highest accuracy of 87.15%. This test requires execution time of 7205 seconds. Figure 3 shows the results of SVM testing with a range of parameter values $C = 1.0-3.0$ and $\sigma = 1.0-2.0$.

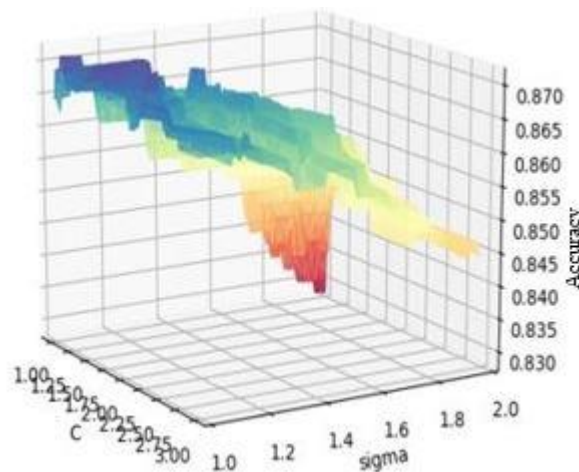


Figure 3 SVM Test Results $C = 1.0-3.0$ and $\sigma = 1.0-2.0$

Figure 3 shows that when the value $C = 1.0$ and $\sigma = 1.0$ the accuracy is 86.46%. When the values C and σ are elevated to $C = 1.63$ and $\sigma = 1.08$ the resulting accuracy increases to 87.15%, but when the values C and σ are increasingly elevated, the accuracy obtained decreases. Accuracy starts according to when the value of $C = 1.68$, $\sigma = 1.06$ up to the value of $C=2.99$, $\sigma=1.99$, even though in that range there is an increase in accuracy, but it does not reach the highest accuracy of 87.15%.

Based on experiments with values $C = 1.0-3.0$, $\sigma = 0.1-1.0$ and $C = 1.0-3.0$, $\sigma = 1.0-2.0$ it can be concluded that the high value of accuracy is not always due to the high values of C and

σ , so to obtain the value of C and σ best based on the highest accuracy requires a trial and error process.

3.3 Testing the FA-SVM method

Evaluation of the performance of the FA-SVM method is done by using the k-fold cross-validation approach. The k fold cross-validation process in this study uses 10 fold ($k = 10$) with the same partition data size and is done randomly. The data in the fold will be divided into 10 subsets so that each subset has the same size and has different data. Using 10 fold of 2179 data, the distribution of 436 data as test data and 1743 data were divided into training data. The process will be repeated as many as 10 folds with different data distribution. Before evaluating the FA-SVM method, we first look for SVM parameters using the Firefly Algorithm method. The SVM parameters are searched in the range $C = 1.0-3.0$, $\sigma = 0.1-1.0$ and $C = 1.0-3.0$, $\sigma = 1.0-2.0$. The number of fireflies used is 30, the number of generations used is 50. The random value is $\alpha = 0.2$ and the value of $\gamma = 1.0$.

The evaluation results of the FA-SVM method used movie opinion data with a range of parameter values $C = 1.0-3.0$ and $\sigma = 0.1-1.0$ giving the highest accuracy results of 87.84%. Figure 4 shows the results of FA-SVM testing with a range of parameter values $C = 1.0-3.0$ and $\sigma = 0.1-1.0$.

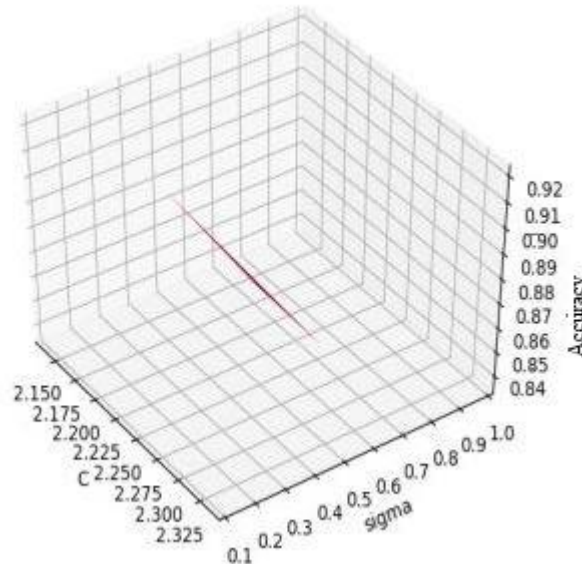


Figure 4 FA-SVM Test Results $C = 1.0-3.0$ and $\sigma = 0.1-1.0$

The FA-SVM method is tested by finding the best C and σ values based on the highest accuracy value obtained from each Firefly. Searching with the Firefly Algorithm method produces 50 combinations of C and σ . Based on the 50 combinations there are three best combinations seen from the highest accuracy which is 87.84%. This test requires execution time of 2330 seconds. Table 2 shows the values of C , σ , accuracy and execution time of FA-SVM with a range of parameter values $C = 1.0-3.0$ and $\sigma = 0.1-1.0$.

Table 2 Results of FA-SVM evaluation $C = 1.0-3.0$ and $\sigma = 0.1-1.0$

C	σ	Accuracy %	Time/s
2.33	0.45	87.84	2330
2.25	0.46	87.84	
2.13	0.50	87.84	

Based on Table 2, it can be seen that the Firefly Algorithm method produces three combinations of the best C and σ values based on the highest accuracy, which is 87.84%.

Accuracy results of 87.84% were obtained from the evaluation process using Confusion Matrix. The evaluation produced 266 data true positive, 117 data true negative, 39 false positive data, and 14 false negative data.

The next test still uses 436 test data but uses a different range from the previous test. The FA-SVM parameter range used is $C = 1.0-3.0$ and $\sigma = 1.0-2.0$. Figure 5 shows the results of the FA-SVM test with a range of parameter values $C = 1.0-3.0$ and $\sigma = 1.0-2.0$.

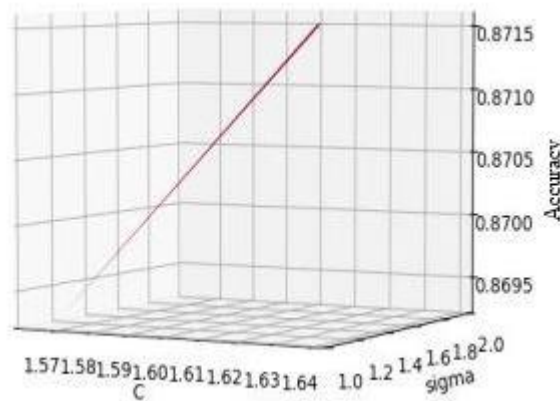


Figure 5 FA-SVM Test Results $C=1.0-3.0$ and $\sigma=0.1-2.0$

The second test with a different range of values produces 50 combinations of C and σ . Based on the combination the highest accuracy is 87.15% with a value of $C = 1.63$ and $\sigma = 1.08$. 87.15% accuracy is obtained from the evaluation process using the Confusion Matrix. The evaluation produced 265 true positive data, 115 true negative data, 41 false positive data, and 15 false negative data. This test requires execution time of 2388 seconds.

3.4 Comparison of SVM and FA-SVM methods

The evaluation results of SVM and FA-SVM methods on 436 test data with ranges $C = 1.0-3.0$ and $\sigma = 0.1-1.0$ giving the highest accuracy of 87.84%. Table 3 shows the results of evaluating the SVM and FA-SVM methods.

Table 3 SVM and FA-SVM results with $C = 1.0-3.0$ and $\sigma = 0.1-1.0$

	C	σ	Accuracy	Time/s
SVM	2.33	0.45	87.84	5928
	2.25	0.46	87.84	
	2.13	0.50	87.84	
	2.84	0.81	86.69	
FA-SVM	2.33	0.45	87.84	2330
	2.25	0.46	87.84	
	2.13	0.50	87.84	

Based on the evaluation results of the SVM and FA-SVM methods in Table 3, that with the range $C = 1.0-3.0$ and $\sigma = 0.1-1.0$ the SVM method produces the highest accuracy of 87.84%. The FA-SVM method with a range $C=1.0-3.0$ and $\sigma=0.1-1.0$ also produces the best combination of C and σ with the highest accuracy of 87.84%. The process of evaluating the SVM method takes 5928 seconds. While the evaluation process using the FA-SVM method takes 2330 seconds. The SVM method takes longer than the FA-SVM method. The SVM method requires a longer execution time because the method tries every combination of C and σ with the range $C=1.0-3.0$ and $\sigma=0.1-1.0$. The process produces 18000 combinations of C and σ .

The FA-SVM method requires a shorter time because the method does not try every combination of C and σ with a certain range. The best combination search process based on accuracy is done by using the objective function as a reference for moving towards a better point.

The next evaluation uses the same amount of data but uses a different range of C and σ values. The range used is $C=1.0-3.0$ and $\sigma=1.0-2.0$. Tests performed on the SVM and FA-SVM methods using these ranges provide the highest accuracy results of 87.15%. Table 4 shows the evaluation results of SVM and FA-SVM with a range of parameter values $C=1.0-3.0$ and $\sigma=1.0-2.0$.

Table 4 SVM and FA-SVM results with $C = 1.0-3.0$ and $\sigma = 1.0-2.0$

	C	σ	Accuracy	Time/s
SVM	1.63	1.08	87.15	7205
	1.68	1.06	86.92	
	2.99	1.99	84.40	
FA-SVM	1.63	1.08	87.15	2388

Table 4 shows that the SVM method and the FA-SVM method provide the highest accuracy results of 87.15%. The execution time required by the SVM method is 7205 seconds, while the FA-SVM method requires execution time of 2388 seconds. The time required by SVM is greater because SVM tries all combinations of C and σ with ranges $C = 1.0-3.0$ and $\sigma = 1.0-2.0$. The difference used is 0.01. Based on that range the combinations of C and σ are tried as many as 20000 combinations. While the time required by the FA-SVM method is smaller because the method does not try every combination of C and σ , but the method uses the objective function as a reference to find the best combination.

4. CONCLUSION

The results of this study indicate that the FA can help SVM to get the appropriate combination of parameters based on accuracy, so there is no need for trial and error to get the parameter values. This conclusion is proved by the results of the evaluation of the SVM method with a range of values $C = 1.0-3.0$ and $\sigma = 0.1-1.0$ giving the highest accuracy of 87.84%. The execution time needed is 5928 seconds. While the results of the evaluation of the FA-SVM method with the same range produce the same accuracy as the SVM method, which is 87.84%, but the execution time needed is shorter which is 2330 seconds. The difference in execution time of the SVM method with the FA-SVM method with a value range of $C = 1.0-3.0$ and $\sigma = 0.1-1.0$ is 3598 seconds. The next evaluation is carried out with a range of values $C = 1.0-3.0$ and $\sigma = 1.0-2.0$. Based on this range, the SVM method and the FA-SVM method produced the highest accuracy of 87.15%. SVM takes 7205 seconds, while the FA-SVM requires 2388 seconds to classify data. The time difference between the SVM method and the FA-SVM method is 4817 seconds.

5. FUTURE WORKS

This research still has shortcomings which can be further improved in future studies. The suggestion for the next research is that this study only tests the FA algorithm as an optimization method on SVM, then can be tested using other metaheuristic algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Gravitational Search Algorithm (GSA), etc.

REFERENCES

- [1] M. H. WAHYUDI, "Klasifikasi Spermatozoa Sapi Pembawa Kromosom X atau Y Dengan Menggunakan The Classification Of OX'S Spermatozoa Carries X or Y Chromosomes Using Support Vector Machine," Institut Teknologi Sepuluh Nopember, 2015.
- [2] Y. Fei, "Simultaneous Support Vector Selection and Parameter Optimization Using Support Vector Machines for Sentiment Classification," *IEEE*, pp. 59–62, 2016.
- [3] K. N. J. P. Devi, "Sentiment Classification Using SVM and PSO," pp. 1–3, 2016.
- [4] B. Pang, L. Lee, H. Rd, and S. Jose, "Thumbs up ? Sentiment Classification using Machine Learning Techniques," 2002.
- [5] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion Mining of Movie Review Using Hybrid Method of Support Vector Machine and Particle Swarm Optimization," *Procedia Eng.*, vol. 53, pp. 453–462, 2013.
- [6] E. Tuba, L. Mrkela, and M. Tuba, "Support Vector Machine Parameter Tuning using Firefly," *Conf. Radioelektronika*, 2016.
- [7] A. Sharma, A. Zaidi, R. Singh, S. Jain, and A. Sahoo, "Optimization of SVM Classifier Using Firefly Algorithm," pp. 198–202, 2013.
- [8] X. Yang, *Cuckoo Search and Firefly Algorithm Theory and Applications*. Springer, 2014.
- [9] A. Kowalczyk, "Support Vector Machines Succinctly," *Book*, vol. Alexandre, 2017.
- [10] R. Diani, U. N. Wisesty, and A. Aditsania, "Analisis Pengaruh Kernel Support Vector Machine (SVM) pada Klasifikasi Data Microarray untuk Deteksi Kanker," *Ind. J. Comput.*, vol. 2, no. 1, NaN pp. 109–118, pp. 109–118, 2017.
- [11] F. Ratnawati, "Analisis Sentimen Opini Film Pada Twitter Menggunakan Algoritme Dynamic Convolutional Neural Network," Universitas Gadjah Mada, 2017.
- [12] Suryadi, "Analisis Sentimen Review Hotel Menggunakan Algoritme Naive Bayes Classifier Dan Pendekatan Lexicon Based," Universitas Gadjah Mada, 2017.
- [13] R. Wijayanti and A. Arisal, "Ensemble Approach for Sentiment Polarity Analysis in User-Generated Indonesian Text," *IEEE*, pp. 158–163, 2017.
- [14] S. Patnaik and X. Li, *Proceedings of International Conference on Computer Science and Information Technology (Advances in Intelligent Systems and Computing)*. 2013.
- [15] X. S. Yang, Z. Cui, R. Xiao, A. H. Gandomi, and M. Karamanoglu, *Swarm Intelligence and Bio-inspired Computation Theory and Applications*. 2013.
- [16] B. Xing and Wen-Jing Gao, *Innovative Computational Intelligence: A Rough Guide to 134 Clever Algorithms*. 2014.