SENTIMENT ANALYSIS FOR OPINION MINING OF 2024 INDONESIAN ELECTION USING THE SVM AND GENETIC ALGORITHM METHODS

Milyun Ni'ma Shoumi S.Kom, M.Kom[1], Imam Fahrur Rozi, ST., MT[2], and Widiareta Safitri[3]
[1-3]Department of Information Technology, Informatics Engineering Study Program,
Politeknik Negeri Malang, Malang 65163, Indonesia
Email: [1]email_bu_milyun, [2]imam.rozi@polinema.ac.id,
[3]1941720081@student.polinema.ac.id

Abstract *– General Election (Pemilu) is one of the political agendas in Indonesia where members of parliament and the president are elected through the election process. As an important political event, the presidential election of 2024 has been extensively discussed by the Indonesian society, especially through social media. One of the social media platforms widely used by people to express their opinions about the 2024 election is Twitter. With the numerous discussions surrounding the election, it naturally generates various public sentiments, which can be used as data to determine whether the public response to the general election tends to be positive, negative, or neutral. In this research, sentiment analysis is conducted using the SVM (Support Vector Machine) method and Genetic Algorithm as an algorithm to optimize the SVM parameters. The sentiment analysis process starts with data collection through web scraping, followed by preprocessing, TF-IDF feature selection, and then the implementation of GA-SVM. The results of this study show that with the assistance of the genetic algorithm, a higher accuracy of 85.5% is achieved compared to using only SVM with a maximum accuracy of 84.0%..*

Keywords: Sentiment analysis, general election 2024, SVM, GA-SVM

I. INTRODUCTION

The General Election (Pemilu) is one of the political agendas for Indonesia where both parliamentarians and presidents are elected by election [1]. The presidential election of 2024 has been widely discussed among Indonesian people [2] and several candidates and political parties were debated in the general public as well as in Internet-based digital arenas such as online news and social media news [3]. The internet and social media could also be effective tools for the campaign, deliver and promote the candidate's work plan to everyone in the community [4]. This discussion through social media will influence the public's assessment of the opportunities and results as well as their opinion regarding the election in 2024.

One of the social media that used to express their opinion about the election in 2024 is Twitter because it Twitter given its ease in conveying news and opinions. The platform generates diverse public sentiments, providing valuable data to gauge whether the general response to the election is positive, negative, or neutral. To identify the response, sentiment analysis can be use since each individual's opinion is highly subjective which it might present various points of view [5]. Sentiment analysis is a Natural Language Processing method that entails extracting subjective information or views expressed in text data - often negative and positive sentiment [6].

The research utilized sentiment analysis employing the Support Vector Machine (SVM) technique and optimized it using the Genetic Algorithm. SVM is a supervised training method for regression and classification, functioning based on data categorization in the analysis. However, Support Vector Machine has weaknesses in parameter selection or appropriate features that have a substantial impact on classifier performance results [7]. To overcome these deficiencies, Genetic Algorithm (GA) is used as the optimization. GA can be used to synchronously optimize the parameters of SVM and feature selection of in

order to effectively use the appropriate features and parameters [8].

Research conducted by Winda and Haryoko in 2019 [9] compared the accuracy of classification results to positive and negative sentiments of Go-Jek reviews using SVM and classification results using SVM optimized with parameters C and Gamma using GA. The results show that GA-SVM has better accuracy than SVM where GA-SVM achieves 89.5% accuracy and F1 score reaches 91.75% whereas if only using SVM accuracy is 62.1% and F1 Score is 72.1%.

Research with the title "Sentiment Analysis of Thai Online Product Reviews using Genetic Algorithms with Support Vector Machine" by Rawisuda Tesmuang and Nivet Chirawichitchai in 2020 [10] compared several methods including Support Vector Machine, Decision Tree, Naïve-Bayes, and K-Nearest Neighbor techniques to test the model performance and also methods which applies Genetic Algorithm optimization. The genetic algorithm combined with the SVM method provides the highest classification efficiency of 88.64%, followed by the SVM method providing the classification efficiency of 87.17%.

Based on the explanation above, the researcher proposes a study entitled "Sentiment Analysis for Opinion Mining of 2024 Indonesian Election using the SVM and Genetic Algorithm Methods". A sentiment analysis was carried out to see the public's response to the 2024 general election with data taken from community tweets on social media twitter. This study uses the Support Vector Machine method with Genetic Algorithm as optimization. The use of Genetic Algorithm can produce a high accuracy value in determining whether the community's response is negative, neutral or positive.

II. LITERATURE REVIEW

A. Previous Research

Sigit K et al. [11] conducted related research by comparing two methods using PSO optimization and without PSO optimization, which is SVM, SVM+PSO, Naïve Bayes, Naïve Bayes+PSO in a study entitled. The results show that the SVM method with PSO has the highest accuracy and Area Under Curve (AUC) values compared to other methods, namely the accuracy reaches 78.40% and the AUC value is 0.850. The use of PSO optimization in the SVM and Naive Bayes methods increases the accuracy of up to 2.31% in the SVM method and 6.77% in the Naive Bayes method.

Research conducted by Bhakuni, M., Kumar, K., Iwendi, C., & Singh, A. in 2022 [12] entitled "Evolution and Evaluation: Sarcasm Analysis for Twitter Data Using Sentiment Analysis" aims to classify sentiments into positive, negative, and neutral sentiments, as well as identify sarcasm. By using different methods, the accuracy value for the Decision Tree method is 86%, Naive Bayes has an accuracy rate of 83%, KNN has an accuracy rate of 51% and the SVM method has the highest accuracy around 93%.

The research entitled "Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain" conducted by Reza Maulana et al in 2020 [13] compared some of the effects of optimization. The method used is SVM, Naive Bayes, KNN, SVM + Information Gain (IG). Initially, the study compared Kernel Functions in the SVM method, namely Linear, Polynomial, RBF and sigmoid. From the results of the comparison, RBF has the greatest level of accuracy, namely 83.05%. After that, a comparison was also made for Parameters, namely C and Epsilon values where the highest accuracy rate was 83.05% for C 1.0 and Epsilon 0.0 values. After that, a comparison was made for the method where the method that produces the highest results is the SVM method with Information Gain (IG) optimization where the accuracy is 86.65% for the Cornell dataset and 86.62% for the Stanford dataset.

Research entitled "Multi-lingual Twitter sentiment analysis using machine learning" conducted by K. Arun and A. Srinagesh in 2020 [14] conducted a Twitter Sentiment Analysis using different languages. The data training process was carried out using the Naive Bayes, SVM, KNN, Decision Tree, Random Forest and Logistic Regression methods. By using the MLTSA algorithm, the accuracy results show an increase where the best accuracy which best accuracy obtained by the SVM method with accuracy up to 95%.

B. TF-IDF

The TF-IDF method (Term Frequency Inverse Document Frequency) is a method for weighting the link between a term and a document. Term-frequency is a measure of how frequently a term appears in a text, as well as in the corpus as a whole. The algorithm of the ratio of the total number of documents in the corpus to the number of documents that have the word is the inverse document frequency. In calculating the TD-IDF, there are several equations. According to Setyadi et al [15] first is equation 1.1 to calculate the weight on the document.

$$W_{td} = tf_{td} \times IDF_t \quad (1.1)$$

Where:

1. d = d document
2. t = t-th word of keyword
3. W = the weight of the d document against the t-th word
4. tf = the number of words searched for in an IDF document
5. IDF = Inversed Document Frequency

IDF can be calculated by equation 1.2

$$IDF(x) = log\frac{N}{df(t)} \quad (1.2)$$

Where $df(t)$ is the number of documents containing the term t. TF-IDF is a combination of the *TF* method with the *IDF* method. Finally, equation to calculate TF-IDF is in equation 1.3

$$TF - IDF(d,t) = TF(d,t) \times IDF(t) \quad (1.3)$$

C. Genetic Algorithm

The genetic algorithm is a meta-heuristic inspired by evolution that belongs to the broad family of evolutionary algorithms used in informatics and computational mathematics. These algorithms are widely employed to provide high-quality solutions to optimization and search problems by concentrating on bioinspired operators such as selection, convergence, and mutations [16]. In genetic algorithm, there are three operations perform:

a) Selection

A chromosome will be stored in a format that contains information about the solution that it represents. A binary string format is a widely used method of encoding. This is what the chromosome will look like after that. A binary string can be used to represent each chromosome. Each bit in the string is also in charge of containing certain characteristics or standards of the answer [17].

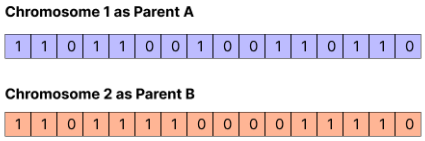

Figure 1. Chromosome

b) Crossover

Crossover operates on a subset of genes from the parent chromosomes, resulting in the formation of a new child. There are several ways to produce a crossover, just as there are numerous crossover spots to choose from. Crossover can be more detailed and sophisticated. It is mostly determined by chromosomal encoding [17].

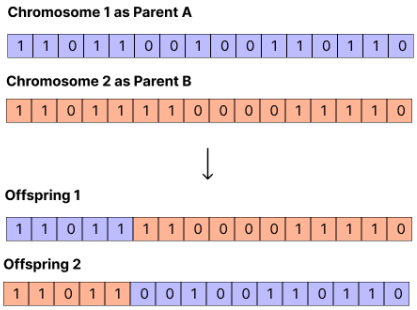

Figure 2. Crossover of Chromosome

c) Mutation

Following the crossing, the next phase is mutation. Mutation is done on purpose to prevent all solutions in the population from falling into a local optimum of the solved issue. The offspring of a crossover are subject to random mutation [17].

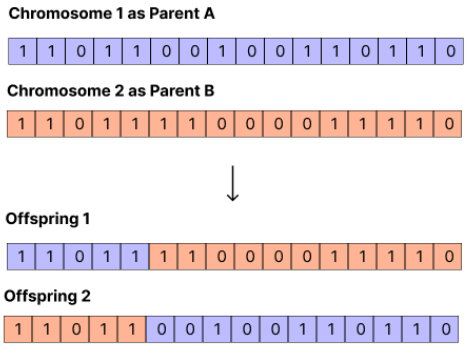

Figure 3. Mutation of Offspring

D. SVM

Support Vector Machine (SVM) is a method for determining optimal boundaries between groups of data that is used to determine the best hyperplane by maximizing the gap between the sets of data [18]. The representation of SVM is shown in Figure 4 as below.

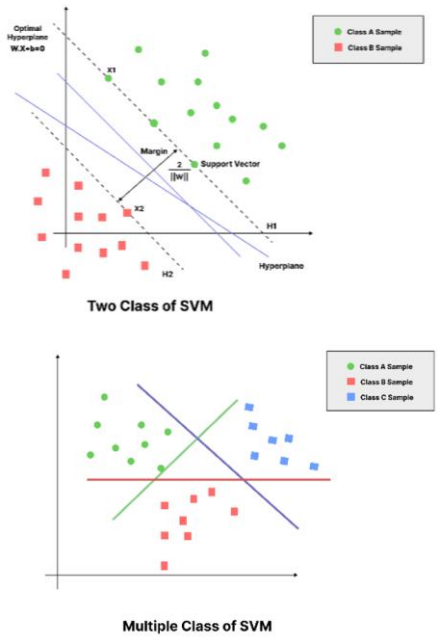

Figure 4. SVM Methods Illustration

On either side of the hyperplane, two parallel hyperplanes are built to segregate the data. The separating hyperplane is the hyperplane that reduces the distance between the two parallel hyperplanes to the smallest possible value. It is assumed that the wider the margin or distance between these parallel hyperplanes, the better the classifier's generalization error will be [19].

The hyperplane is determined by using equation (2.1)

$$w.x + b = 0 \qquad (2.1)$$

Where b is scalar and w is p-dimensional Vector. For the parallel hyperplanes can be described by equation (2.2) and equation (2.3)

$$w.x + b = 1 \qquad (2.2)$$

$$w.x + b = -1 \qquad (2.3)$$

If the training data are linearly separable, we can choose these hyperplanes with no points in between and then strive to maximize their distance. Using geometry, the distance between the hyperplanes can be calculated by

$$\frac{2}{|W|} \qquad (2.4)$$

Then, in order to entice data points, it is necessary to guarantee that for all I either

$$w.x_i - b \geq 1 \; or \; w.x_i - b \leq -1 \qquad (2.5)$$

Or can be written as

$$y_i(w.x_i - b) \geq , \qquad 1 \leq i \leq \text{n} \qquad (2.6)$$

the next step is to manage data labelling mistakes, which can be accomplished through soft margin concept by using equation (2.7)

$$min_{w,b,t} \frac{1}{2}||w||^2 + c\sum_{i=l}^{m} t_i \qquad (2.7)$$

Then, time to obtaining kernel value using polynomial kernel. To calculate the kernel using polynomial, the following formula is used

$$K(x_i, x_j) = (c + X_1^T X_2)^b \qquad (2.8)$$

Where:
c = constant term
b = degree of kernel

The labelling operation can be carried out by utilizing the separator function in equation (2,9). If the outcome is more than 1, the data is

categorized as positive; otherwise, the data is classified as negative.

$$f(x) = y\, a' K(x_{test}, x_{test}) + b \qquad (2.9)$$

II. METHODS

A. *Proposed Method: SVM with Genetic Algorithm for Optimization*

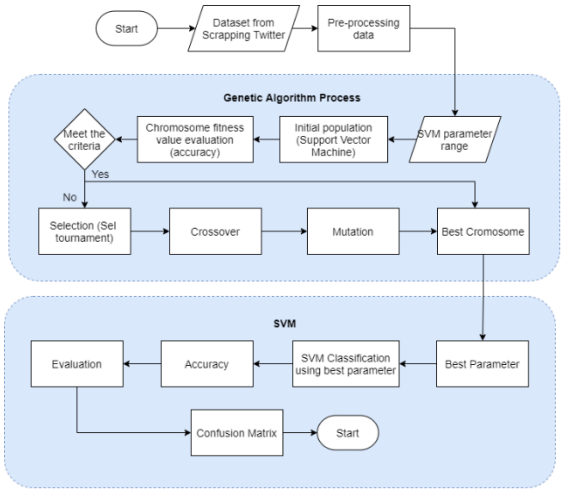

Figure 5. Proposed Method

The classification process involves using SVM and Genetic Algorithm on a Twitter dataset. Genetic Algorithm optimizes SVM parameters (Kernel, C, Gamma, Degree, Maximum Iteration) to find the best configuration for accurate results. This iterative process includes selection, mutation, and crossover, guided by specific parameters. The best configuration is determined by fitness evaluation. The chosen parameters are then applied to the SVM classification process, producing accuracy values that are evaluated using a Confusion Matrix to obtain Precision, Recall, and F1-Score values.

B. *Data Collection*

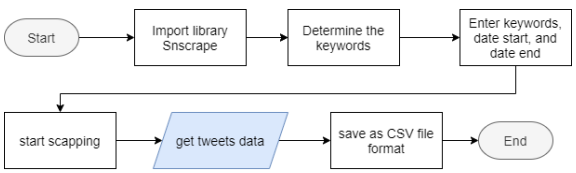

Figure 6. Scrapping Process Flowchart

The first step is to import a library called Snscrape into Python. Then enter the keyword data that will be scrapped. Apart from that, it also includes the start date and end date as the date range of the tweets data to be retrieved. Because the Scrape library can only carry out the scrapping process within one month, it cannot perform the scrapping process within one year, starting from January 2022 - January 2023, the scrapping process is repeated 13 times for each keyword. After that, the scrapped data will appear and can be saved in csv.file format

C. *Data Pre-processing*

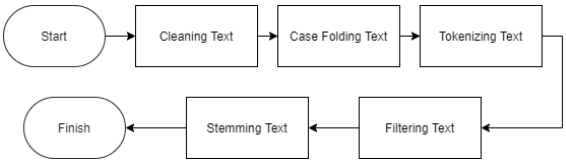

Figure 7. Pre-processing Flowchart

Preprocessing is the first phase in the text mining process, and it prepares the data to use for extraction and clustering. It is useful for removing noise from data and cleaning it up. There are five steps in pre-processing phase including case folding, filtering, tokenization, stemming and stop-word removal. Process flow for data pre-processing can be seen in the Figure 7.

Table 1. Pre-processing Text

Before Preprocessing	***After Preprocessing***
Mari kita bersama-sama memperkuat persatuan untuk melanjutkan berbagai program pembangunan #KoalisiIndonesiaBersatu, menuju Indonesia yang lebih makmur dan lebih maju. #PartaiGolkar #PartaiAmanatNasional	['mari', 'kuat', 'satu', 'lanjut', 'program', 'bangun', 'indonesia', 'makmur', 'maju']

#PartaiPersatuanPembangun
an #Pemilu2024

D. Extraction Features Using TF-IDF

The most important requirement before grouping the sentences into groups is that they be machine understandable. This may be accomplished using the TF-IDF term weight representation approach, which determines the relevance of each phrase in a text. The TF-IDF process considers the frequency of appearance of a term in a document as well as across the entire corpus. In the weighting process using TF-IDF, the input that must be entered is the data from the pre-processing results which will then be carried out in the calculation process of Term Frequency (TF), Weighting Term Frequency (WTF), Inverse Document Frequency (IDF), and Term Frequency-Inverse Document Frequency (TF-IDF).

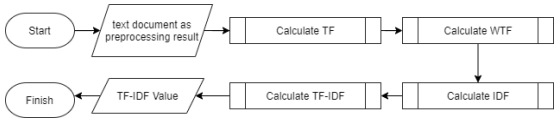

Figure 8. TF-IDF Process

E. Selection Feature Using GA

In this study, the Support Vector Machine (SVM) method is used where there are several parameters needed. The parameter such as Cost (C), Gamma (γ), Degree (λ) dan Maximum Iteration in certain kernel function has a substantial impact on the SVM method's predicting effectiveness. The GA is used to choose the SVM parameters Kernel, Cost (C), Gamma (γ), Degree (λ) dan Maximum Iteration in order to obtain higher accuracy for SVM. The process of optimizing SVM parameters using GA is depicted in the figure below.

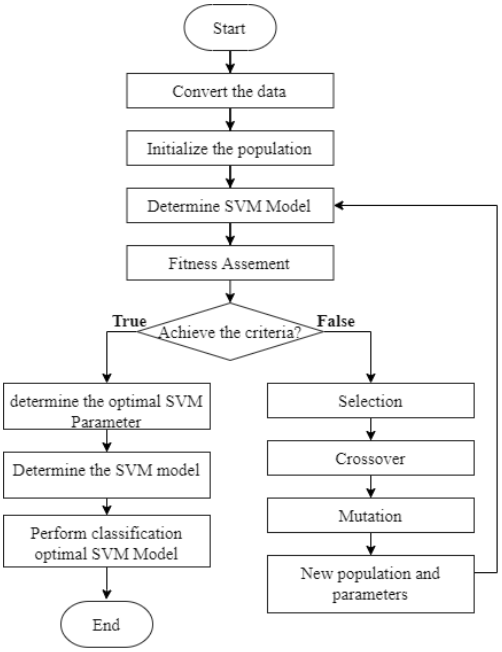

Figure 9. The SVM parameters optimization using the GA method

F. Classification with Support Vector Machine

Text classification using Support Vector Machine is a step that is carried out after text weighting process is complete as well as determine the best parameter for SVM using genetic algorithm. Several processes to do in this text classification. The steps taken are to calculate the hessian matrix, sequential Training, calculate bias, and calculate h(x). Results to be obtained, namely positive, neutral and negative class. The flowchart of SVM process for text classification can be seen in figure x.

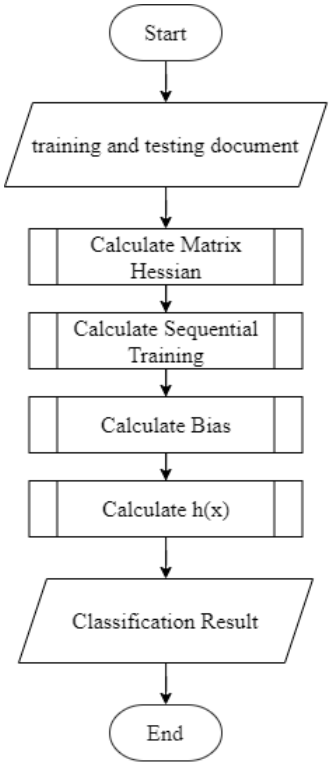

Figure 10. SVM Process for Classification

IV. RESULTS AND DISCUSSION

A. Result

1) Manual Testing of SVM Paramater

Experiment was carried out manually to testing several parameters of SVM method. The maximum iteration effect is different for each kernel. In the RBF, Polynomial, and Sigmoid kernels, the best maximum iteration is 250 which can produce an accuracy of 84.0% in RBF kernels, 81.5% in Polynomial kernel, and 78.5% in Sigmoid kernel. But in Linear kernel, the best maximum iteration is 250. From all the experiments on SVM to find the best parameters manually, the best accuracy was obtained, namely 84.0% by Support Vector Machine with RBF Kernel, C = 10, Degree = 1, Gamma = 0.4 and maximum iteration is 300. This accuracy value will be compared with the value the best accuracy with tuning hype-parameter with Genetic Algorithm.

2) SVM Hyperparameters Tuning Using Genetic Algorithm

Hyperparameters Tuning in SVM using Genetic Algorithm (GA) is intended to find a better accuracy value. The GA-SVM in this study will look for the best parameters, namely Kernel value, C value, Gama, Degree, and Maximum Iteration. The range for parameter values will be taken from the best range from previous manual best parameter search experiments. The following is a table for the range of values used for each parameter.

Table 1. Value Range in Each Parameters

Parameter	Value Range
Kernel	RBF, Linear, Polynomial, Sigmoid
C (Cost)	0.1, 10
Gamma	0.0-1.0
Degree	1-5
Maximum Iteration	100-300

Each parameter in SVM will be represented as a gene, which will be used to form a chromosome. This chromosome will undergo processes such as mutation, crossover, evaluation, and selection to find the best parameter, or in other words, the best chromosome.

In the genetic algorithm, there are four parameters used to execute the genetic algorithm: Mutation Probability (Pm), Crossover Probability (Pc), Number of Generation (G), and Population Size (Pz). In this study, the parameter values used are 50 for Number of Generation, 0.95 for Crossover Probability, and 0.9 for Mutation (Togatorop, P. et al, 2022). However, a value of 10 is used for the Number of Generations.

From the experiment results, an optimal accuracy of 85.5% was obtained, with the following parameters identified as the best for SVM.

Table 2. Best Parameter for SVM

Parameter	Value
Kernel	RBF
C	2.3525701264249683
Gamma	0.8783192364696546
Degree	1
Maximum Iteration	248

Here is the graph representing the results of the genetic algorithm in finding the best parameters for SVM based on accuracy values.

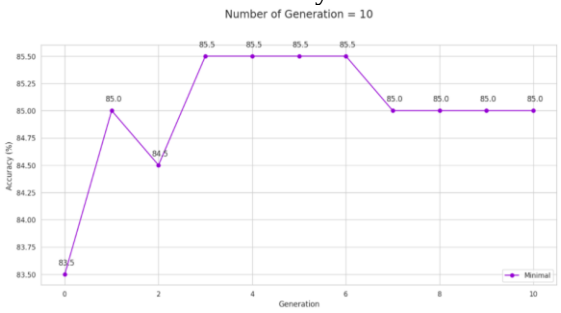

Figure 11. SVM Process for Classification

B. Discussion

The results of previous experiments to find the best parameters for SVM using the manual method and Genetic Algorithm show that the Genetic Algorithm is the best way to search for parameters. The following is a comparison of the maximum accuracy in SVM and GA-SVM implemented on Data Test.

Table 3. Comparison of Precision, Recall, F1-Score, and Accuracy on SVM and GA-SVM on Data Test

	Precision	**Recall**	**F1-Score**	**Accuracy**
SVM	84.0%	84.0%	84.0%	84.0%
GA-SVM	85.5%	85.5%	85.5%	85.5%

After obtaining the accuracy from SVM and GA-SVM, the next process is applied SVM and SVM model to all 2000 datasets. The following is a comparison of the maximum accuracy in SVM and GA-SVM implemented on all 2000 datasets.

Table 4. Comparison of Precision, Recall, F1-Score, and Accuracy on SVM and GA-SVM on all 2000 Datasets

	Precision	**Recall**	**F1-Score**	**Accuracy**
SVM	98.4%	98.4%	98.4%	98.4%
GA-SVM	98.5%	98.5%	98.4%	98.5%

After the GA-SVM model is applied to all data, the classification results are obtained for all data. The results of applying the classification using SVM with the best parameters are as follows.

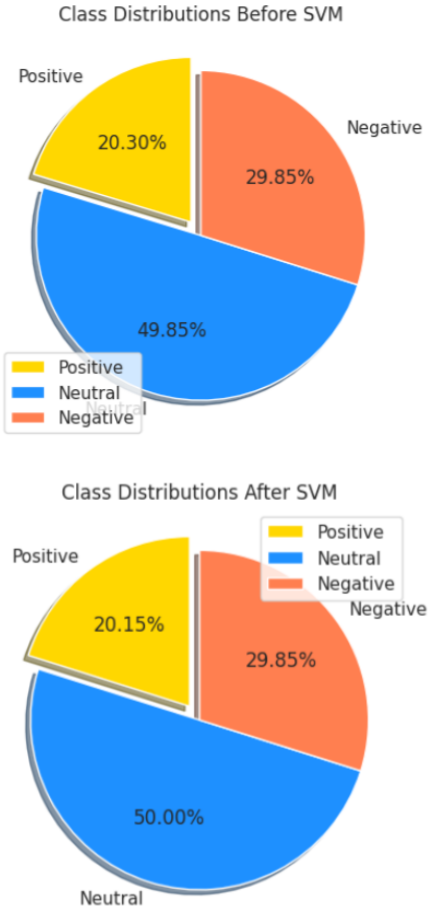

Figure 11. Comparison of Data Class before and After Classifications

The positive class experienced a decrease which should have been 20.30% to 20.15%, then the negative class increased to 50%. And for the neutral class, there was no increase or decrease and remained 29.85%.

In this study, it was previously explained that dataset retrieval uses keyword hashtags, usernames, and certain locations. The hashtags used are "#Pemilu2024" and "#Serentak Election2024" while the usernames used are candidates that have been mentioned quite a lot in research, namely "@ganjarpranowo", "aniesbaswedan", and "@prabowo".

a) Sentiment Classification Based on Keyword “#Pemilu2024”

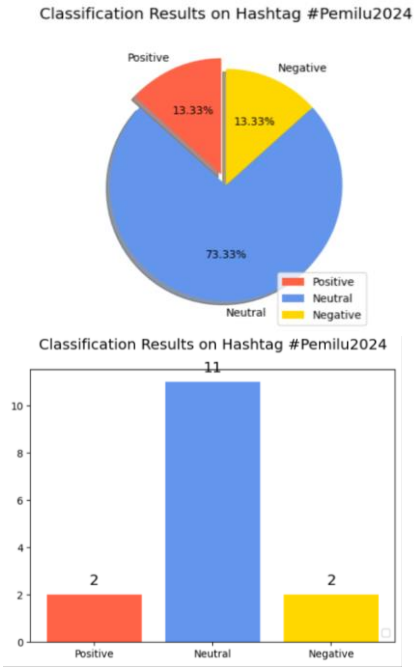

Figure 12. Classification Results on the Keyword "#Pemilu2024" From 15

Based on the classification results, keywords with the hashtags "Pemilu2024" have the most sentiment from the Neutral sentiment. Of the 15 data, 73.33% or 11 data is neutral, for negative and positive are 13.33% or only 2.

b) Sentiment Classification Based on Keyword “#PemiluSerentak2024”

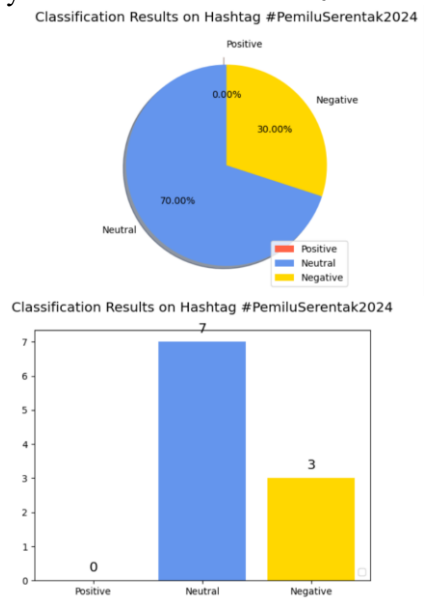

Figure 13. Classification Results on the Keyword "#PemiluSerentak2024" From 10 Data

Based on the classification results on the keyword "#Serentak2024 Election", as much as 70% of the 10 data or 7 data are neutral sentiments. Then, for sentiment with a negative class, that is, only 3 data with negative sentiment.

c) Sentiment Classification Based on Keyword “@ganjarpranowo”

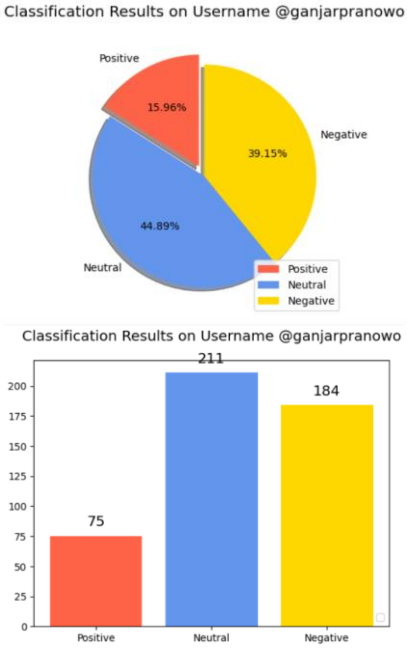

Figure 14. Classification Results on the Keyword "@ganjarpranowo" From 470 Data

Sentiment Classification Based on Keyword "@ganjarpranowo" shows that out of 470 sentiment data, the most sentiment is neutral with percentage is 44.89% or 211 data. For negative sentiment have total 184 data or 39.15% and for positive sentiment are 75 data or 15.96%.

d) Sentiment Classification Based on Keyword “@aniesbaswedan”

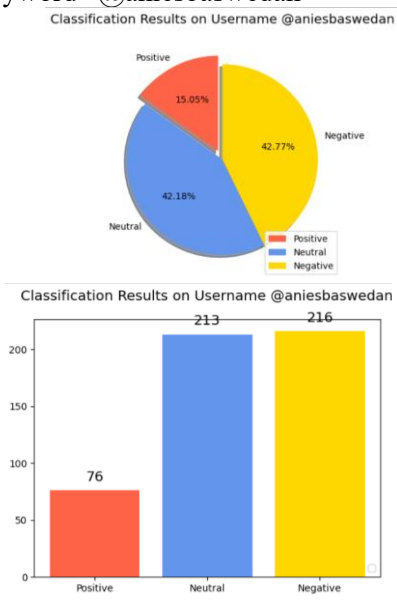

Figure 15. Classification Results on the Keyword "@aniesbaswedan" From 505 Data

Based on the results of the Sentiment Classification Based on the Keyword "@aniesbaswedan", there are the most negative class sentiments. Of the 505 data, negative sentiment are 216 data or 42.77% of the total data. The second most sentiment is neutral sentiment with 213 data and positive is with 76 data.

e) Sentiment Classification Based on Keyword “@prabowo”

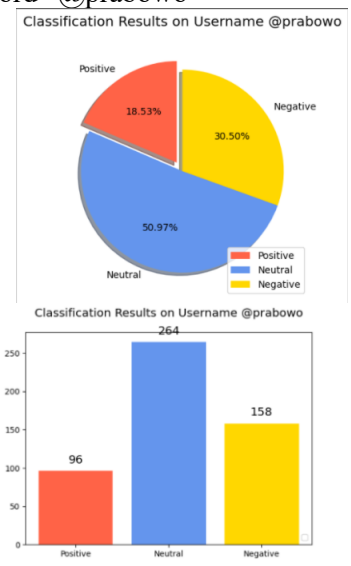

Figure 16. Classification Results on the Keyword "@prabowo" From 518 Data

Based on the results of the Sentiment Classification Based on Keyword "@prabowo" it shows that the most are sentiments with a neutral class, namely 264 data out of 518 data or 50.97%. Then, followed by sentiment with a negative class of 158 data and positive sentiment are 96 data.

f) Comparison of Sentiments of the 2024 Presidential Candidates: @ganjarpranowo, @aniesbaswedan, @prabowo

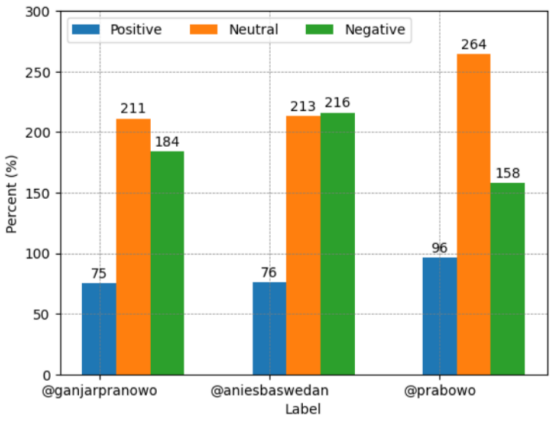

Figure 17. Comparison of Sentiments of the 2024 Presidential Candidates

A comparison of the sentiments of the candidate candidates: @ganjarpranowo, @aniesbaswedan, @prabowo in the 2024 presidential election shows that for @ganjarpranowo and @prabowo, the most common sentiment is neutral sentiment. However, for @aniesbaswedan, the most common sentiment is negative sentiment but it is close to neutral sentiment with total data is 216 for negative sentiment an 213 for neutral sentiment.

V. CONCLUSION

The validation process involving a linguist aims to minimize label errors, enhancing the effectiveness of the SVM model. Utilizing the Support Vector Machine and Genetic Algorithm methods for sentiment analysis in the 2024 General Election includes optimizing SVM parameters like Kernel value, C value, Degree, Gamma, and Maximum Iteration. Through the Genetic Algorithm, optimal parameters enhance SVM performance, resulting in improved Accuracy, Precision, Recall, and F1-Score values compared to the

SVM method alone. For instance, the optimized SVM method exhibits higher Precision (85.5%), Recall (85.5%), F1-Score (85.5%), and accuracy (85.5%) compared to the unoptimized SVM method (Precision, Recall, F1-Score, and accuracy all at 84.0%).

REFERENCES

[1] F. El Jamiy and R. Marsh, "Distance estimation in virtual reality and augmented reality: A survey," in *2019 IEEE International Conference on Electro Information Technology (EIT)*. Brookings, SD, USA: IEEE, May 20–22, 2019, pp. 063–068.

[2] J. Mittelstaedt, J. Wacker, and D. Stelling, "Effects of display type and motion control on cybersickness in a virtual bike simulator," *Displays*, vol. 51, pp. 43–50, 2018

[3] Rajagukguk, K. J., Aripin, S., & Wahyudi, H. (2021). Simultaneous General Election: It Is Fair for Democracy in Indonesia. JIP (Jurnal Ilmu Pemerintahan): Kajian Ilmu Pemerintahan dan Politik Daerah, 6(1), 56-64.

[4] Syamsurrijal, M., Nurmandi, A., Jubba, H., Hidayati, M., Baharuddin, T., & Qodir, Z. (2021). Prediction Candidates and Political Parties in the Presidential Election 2024 in Indonesia Based on Twitter.

[5] Baharuddin, T., Qodir, Z., Jubba, H., & Nurmandi, A. (2022). Prediction of Indonesian presidential candidates in 2024 using sentiment analysis and text search on Twitter. ijcs, 4, 512.

[6] S. Adi, M. Wulandari, A. K. Mardiana, and A. Muzakki, (2018). Survei: topik dan tren analisis sentimen pada media online". Seminar Nasional Teknologi Informasi dan Multimedia 2018, pp. 55 – 60.

[7] Lee, Vivian & Gan, Keng Hoon & Tan, Tien-Ping & Abdullah, Rosni. (2019). Semi-supervised Learning for Sentiment Classification using Small Number of Labeled Data. Procedia Computer Science. 161. 577-584. 10.1016/j.procs.2019.11.159.

[8] Rahardi, M., Aminuddin, A., Abdulloh, F. F., & Nugroho, R. A. (2022). Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia. Int. J. Adv. Comput. Sci. Appl., 13(6), 2022.

[9] Kristiyanti, Dinar & Normah, & Hairul Umam, Akhmad. (2019). Prediction of Indonesia Presidential Election Results

[10] Tao, Zhou; Huiling, Lu; Wenwen, Wang; Xia, Yong (2018). GA-SVM based feature selection and parameter optimization in hospitalization expense modeling. Applied Soft Computing, (), S1568494618306264–. doi:10.1016/j.asoc.2018.11.001

[11] Winda M. P.D. and Haryoko. (2019). Optimization Of Parameter Support Vector Machine (SVM) using Genetic Algorithm to Review Go-Jek's Services. 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, pp. 301-304, doi: 10.1109/ICITISEE48480.2019.9003894.

[12] Tesmuang, R., & Chirawichitchai, N. (2020). Sentiment Analysis of Thai Online Product Reviews using Genetic Algorithms with Support Vector Machine. Progress in Applied Science and Technology, 10(2), 7–13. https://doi.org/10.14456/past.2020.8

[13] Sigit, K., Dewi, A. P., Windu, G., Muhamad, T., & Kadinar, N. (2019, November). Comparison Of Classification Methods on Sentiment Analysis of Political Figure Electability Based on Public Comments on Online News Media Sites. In IOP Conference Series: Materials Science and Engineering (Vol. 662, No. 4, p. 042003). IOP Publishing.

[14] Bhakuni, M., Kumar, K., Iwendi, C., & Singh, A. (2022). Evolution and Evaluation: Sarcasm Analysis for Twitter Data Using Sentiment Analysis. Journal of Sensors, 2022.

[15] Maulana, R., Rahayuningsih, P. A., Irmayani, W., Saputra, D., & Jayanti, W. E. (2020). Improved accuracy of sentiment analysis movie review using support vector machine-based

information gain. In Journal of Physics: Conference Series (Vol. 1641, No. 1, p. 012060). IOP Publishing.

[16] Arun, K. & Srinagesh, Ayyagari. (2020). Multi-lingual Twitter sentiment analysis using machine learning. International Journal of Electrical and Computer Engineering (IJECE). 10. 5992. 10.11591/ijece.v10i6.pp5992-6000.

[17] Setyadi, I., Khrisne, D., & Suyadnya, I. (2018). Automatic Text Summarization Menggunakan Metode Graph dan Metode Ant Colony Optimization. Majalah Ilmiah Teknologi Elektro, 17(1), 124-130. doi:10.24843/MITE.2018.v17i01.P17.

[18] Whitley D,A (1994). Genetic algorithm tutorial, Statistics and computing, Jun 14 (2):65-85

[19] Lambora, Annu; Gupta, Kunal; Chopra, Kriti (2019). IEEE 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) - Faridabad, India (2019.2.14-2019.2.16)] 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) - Genetic Algorithm- A Literature Review 380–384. doi:10.1109/COMITCon.2019.8862255

[20] Hidayat, T. H. J., Ruldeviyani, Y., Aditama, A. R., Madya, G. R., Nugraha, A. W., & Adisaputra, M. W. (2022). Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier. Procedia Computer Science, 197, 660-667.

[21] V. Vapnik (1995) The Nature of Statistical LearningTheory. NY: Springer-Verlag.