JURNAL INFOTEL
VOL. 17, NO. 1, FEBRUARY 2025, PP. 68–80. DOI:10.20895/INFOTEL.V17I1.1266

RESEARCH ARTICLE

Performance of SVM Optimized with PSO as Classification Method for Sentiment Analysis UNNES's Social Media

Miftahul Janaah[1] and Anan Nugroho[2,*]

[1,2]Department of Electrical Engineering, Universitas Negeri Semarang, Semarang 50229, Indonesia

*Corresponding email: anannugroho@main.unnes.ac.id

Received: October 31, 2024; Revised: December 3, 2024; Accepted: February 27, 2025.

Abstract: The reputation of the university is essential to attract prospective students, improve the competitiveness of graduates, and strengthen industry relationships. However, the main challenge in sentiment analysis of a university's reputation is to process and analyze large and diverse social media data effectively to obtain accurate information. This study aims to analyze public sentiment regarding the reputation of Universitas Negeri Semarang (UNNES) on platform X, using a Support Vector Machine (SVM) optimized with Particle Swarm Optimization (PSO). The data comprised 4,606 tweets collected between May 1, 2024, and May 31, 2024. This study tested four SVM kernels: Linear, RBF, Sigmoid, and Polynomial. The evaluation results show that the optimized RBF kernel with PSO achieved the best performance, with an accuracy of 82.05%, recall of 81.21%, Precision of 82.14%, and F-1 score of 81.88%. These results indicate that the proposed approach successfully classifies the sentiment of tweets related to UNNES, providing an effective tool for monitoring the university's reputation on social media.

Keywords: hyperparameter, PSO, sentiment analysis, social media, SVM

1 Introduction

The rapid development of digital technology has impacted various aspects of life, one of which is the use of social media. Social media has become essential for sharing activities, forming individual identities, and expressing opinions [1]. This platform generates text, images, videos, and geographic locations, which can be analyzed to understand public sentiment [2]. This sentiment analysis provides useful insights for monitoring the reputation of institutions such as universities. The reputation of the university plays an important

© Jurnal Infotel ISSN: 2085-3688; e-ISSN: 2460-0997

role in the selection process for prospective students, the learning experience, and the employment prospects of its graduates [3]. A good reputation not only attracts outstanding prospective students but also increases the competitiveness of graduates in the workforce and strengthens relationships with industry. Sentiment analysis classifies opinions into positive, neutral, or negative. However, the main challenge is effectively processing and analyzing large and diverse social media data to obtain accurate information about the university's reputation [4].

Several previous studies have explored sentiment analysis methods for university reputation. One of the relevant studies is the study of [5], which analyzes the sentiment towards Telkom University based on the results of data collection on the LinkedIn networking platform based on posts made by Telkom University accounts. Using the Random Forest method, this study aims to classify sentiment as positive, neutral, and negative. The results of this study provide insight into public perceptions of the university's reputation on professional networks. Another study that examines public sentiment towards Telkom University is [6], which uses data from platform X. This study applies the Long Short-Term Memory (LSTM) method and Word2Vec feature expansion. The results of this study can help Telkom University understand public perception.

In this study, the Support Vector Machine (SVM) is used as a sentiment analysis classification method for university reputation. SVM has been widely applied in various fields, such as text classification, because the basic principle of SVM is to find the best hyperplane to separate data into two classes and maximize the margin between these classes [7]. Using kernel techniques, such as Linear, RBF, Sigmoid, and Polynomial, SVM can handle various types of social media data with large and diverse characteristics that result in accurate classification. One of the advantages of SVM is that it maps data to higher dimensions to separate non-linear classes [8], which are often found in sentiment analysis. However, SVM performance highly depends on proper parameter selection, such as C-value and kernel type [9]. Metaheuristic methods can be effectively used to optimize parameter selection. One of the proven metaheuristic methods is Particle Swarm Optimization (PSO), which shows better performance than the Genetic Algorithm (GA) based on study [10]. PSO can also find optimal solutions efficiently because it is swarm-based, which allows exploration of the search space [11].

The main contribution of this study is the optimization of SVM using PSO for the sentiment analysis of the reputation of Universitas Negeri Semarang (UNNES). The study stages include data collection, preprocessing, labeling, splitting, feature extraction using TF-IDF, oversampling with SMOTE, SVM classification, and optimization with PSO. The results were evaluated by calculating accuracy, F1-score, precision, and recall using a confusion matrix.

2 Research Method

The research employs the Knowledge Discovery in Databases (KDD) process to identify significant patterns in data [12]. The KDD process encompasses steps such as selection, transformation, data mining, and interpretation/evaluation, illustrated in Figure 1 [13]. The stages outlined in the KDD methodology are directly adapted and applied in this research to address the identified research problem. The methodology follows a structured flow to ensure comprehensive data analysis, leading to insightful conclusions. The research

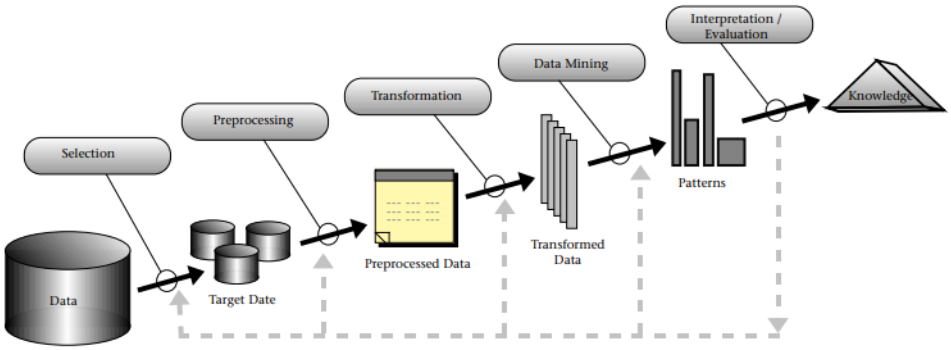

Figure 1: KDD process steps.

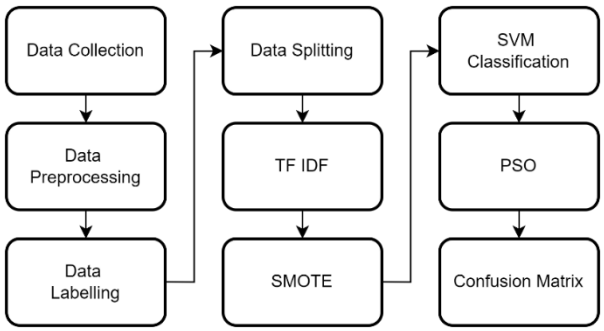

Figure 2: Research steps.

method used in this study is designed to solve the research problem systematically, as depicted in Figure 2, which illustrates the step-by-step process of conducting this research.

2.1 Data Collection

This study used a crawling tool called Tweets-Harvest to automate Twitter's data collection process. Tweets-Harvest leveraged the Twitter API to retrieve tweets containing the keyword "#unnes" posted between May 1, 2024, and May 31, 2024. This targeted keyword selection ensured the dataset's relevance to the study topic centered around Universitas Negeri Semarang (UNNES). The collected tweets were then systematically saved in a Comma-Separated Values (CSV) file, resulting in a dataset of 4,606 tweets.

2.2 Data Preprocessing

Data preprocessing is essential for the development of a classification model [14]. The steps in data pre-processing are shown in Figure 3. The initial step in data preprocessing

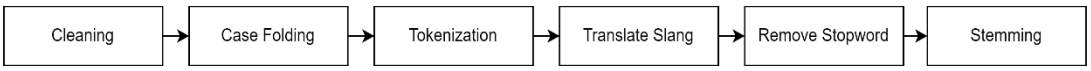

Figure 3: Data preprocessing steps.

is data cleaning, which involves removing extraneous elements such as HTML tags (e.g., `<a>`, `<div>`, `<p>`), URLs (e.g., `www.`, `http://`, `https://`), special characters (e.g., `@#&^`), punctuation marks (e.g., `.`, `,`, `?`), and symbols (e.g., `*$%`) that do not contribute to sentiment analysis. Following this, case folding converts all text to lowercase, ensuring consistency and preventing variations such as `"Manfaat"` and `"manfaat"` from being treated as different words. Next, tokenization splits sentences into individual words, creating a list of tokens for further processing. For example, the sentence `"unnes bersih banget"` would be tokenized into the following list of tokens: `["unnes", "bersih", "banget"]`. Subsequently, the slang translation converts informal words into their more formal equivalents, enhancing the accuracy of the analysis. The removal of stopwords (common words with minimal sentiment value, such as `"yang"`, `"mana"`, `"kalau"`, and `"aku"`) helps to focus on words that carry more significant meaning. Finally, stemming reduces words to their root form, allowing different variations of the same word to be treated as a single entity when using the Sastrawi library, specifically for Indonesian language processing. For example, the words `"membaca"`, `"baca"`, and `"pembaca"` would all stem from the root form `"baca"` using Sastrawi. An example of data pre-processing in the UNNES dataset can be seen in Table 1.

Table 1: Results of text processing steps

Steps	Result
Original Text	My dream university mixed feeling banget ke unnes pas udah selesai sidang skripsi. rasanya sedih tapi lega juga auk ah https://t.co/n2nMBG8olG
	Syp utbk nya di gedung arsip unnes cung aku ga ada temen https://t.co/SVQ0acRwzC
	CW//HOROR lihatlah apa yg kutemukan di unnes https://t.co/R28ZagyfBy
Cleaning	my dream university mixed feeling banget ke unnes pas udah selesai sidang skripsi rasanya sedih tapi lega juga auk ah
	Syp utbk nya di gedung arsip unnes cung aku ga ada temen
	CW HOROR lihatlah apa yg kutemukan di unnes
Case Folding	my dream university mixed feeling banget ke unnes pas udah selesai sidang skripsi rasanya sedih tapi lega juga auk ah
	syp utbk nya di gedung arsip unnes cung aku ga ada temen
	cw horor lihatlah apa yg kutemukan di unnes

Continued on the next page

Steps	Result
Tokenization	['my', 'dream', 'university', 'mixed', 'feeling', 'banget', 'ke', 'unnes', 'pas', 'udah', 'selesai', 'sidang', 'skripsi', 'rasanya', 'sedih', 'tapi', 'lega', 'juga', 'auk', 'ah']
	['syp', 'utbk', 'nya', 'di', 'gedung', 'arsip', 'unnes', 'cung', 'aku', 'ga', 'ada', 'temen']
	['cw', 'horor', 'lihatlah', 'apa', 'yg', 'kutemukan', 'di', 'unnes']
Translate Slang	['my', 'dream', 'university', 'mixed', 'feeling', 'sekali ', 'ke', 'unnes', 'pas', 'sudah', 'selesai', 'sidang', 'skripsi', 'rasanya', 'sedih', 'tapi', 'lega', 'juga', 'auk', 'ah']
	['siapa', 'utbk', 'nya', 'di', 'gedung', 'arsip', 'unnes', 'cung', 'saya', 'tidak', 'ada', 'teman']
	['cw', 'horor', 'lihatlah', 'apa', 'yang ', 'kutemukan', 'di', 'unnes']
Remove Stopwords	['my', 'dream', 'university', 'mixed', 'feeling', 'pas', 'selesai', 'sidang', 'skripsi', 'sedih', 'lega']
	['utbk', 'gedung', 'arsip', 'teman']
	['horor', 'lihatlah', 'kutemukan']
Stemming	['my', 'dream', 'university', 'mixed', 'feeling', 'pas', 'selesai', 'sidang', 'skripsi', 'sedih', 'lega']
	['utbk', 'gedung', 'arsip', 'teman']
	['horor', 'lihat', 'temu']

2.3 Data Labeling

Sentiment labeling of the collected tweets was performed using InSet (Indonesia Sentiment Lexicon), consisting of 3,609 positive and 6,609 negative words. Each word in a tweet was assigned a sentiment score ranging from -5 (highly negative) to +5 (highly positive) based on its entry in InSet. For example, words such as bahagia (happy, +5), ingkar (denial, -4), and simpel (simple, +3), received specific sentiment scores.

Once individual word scores were assigned, tweet score aggregation summed up these scores to derive a cumulative sentiment score for each tweet. Tweets were then classified into three categories based on their total sentiment score: positive (total score > 0), neutral (total score = 0), and negative (total score < 0), as has been worked on [15]. An example of data labeling after data preprocessing can be seen in Table 2. The data labeling resulted in the following distribution of sentiment classifications among the collected tweets: 2,370 tweets (50.1%) were classified as neutral, 1,239 tweets (26.2%) as positive, and 997 tweets (21.1%) as negative.

2.4 Data Splitting

This study divides the dataset into two subsets: training data and testing data. The training data are used to build and train the model, while the testing data are used to evaluate the model's performance. A 70:30 split ratio was applied, with 70% of the data

Table 2: Data labeling

Text	Negative Score	Positive Score	Total Score	Sentiment
['my', 'dream', 'university', 'mixed', 'feeling', 'pas', 'selesai', 'sidang', 'skripsi', 'sedih', 'lega']	-1	2	1	positive
['utbk', 'gedung', 'arsip', 'teman']	0	0	0	neutral
[unnes, horor, lihat, temu]	-5	2	-3	negative

allocated for training and 30% prepared for testing. This approach is consistent with the methodology used in the previous study [16], ensuring a balanced evaluation of the accuracy and generalization ability of the model.

2.5 Feature Extraction

The result of the data separation is followed by the extraction of features using the Term Frequency Inverse Document Frequency (TF-IDF). TF represents the frequency of occurrence of a word in the dataset, while IDF reflects the significance of a word in the dataset, where words that appear frequently throughout the document will have a lower weight; otherwise, if the word appears infrequently, it will have a higher weight [17]. This method ensures that words of greater importance, which occur less frequently in documents, are assigned higher weights in the document representation [18].

2.6 Oversampling

Oversampling is a technique used in machine learning to handle class imbalance in datasets [19]. Class imbalance occurs when the number of samples in one class is significantly more or less than the others. The synthetic minority oversampling technique (SMOTE) is used to address data imbalance by creating synthetic samples for minority classes by interpolating existing samples rather than oversampling by repetition [20]. The difference between the data before SMOTE and after SMOTE can be seen in Figure 4 and Figure 5.

2.7 Classification Method

After the oversampling process, SVM is used as the classification method. This study evaluates the performance of four kernel functions in the SVM models, namely Linear, RBF, Sigmoid, and Polynomial kernels. Each kernel function requires optimization of certain parameters to achieve optimal performance, which is technically calculated using the Scikit-Learn Library. The expressions of the kernel function are presented in Table 3 [21].

The description of each notation in Table 3: C is the cost parameter. γ (gamma) controls the influences of individual data points on the decision boundary. r (coefficient) is a constant that adds to the inner product. d is the degree that controls the complexity of the decision boundary.

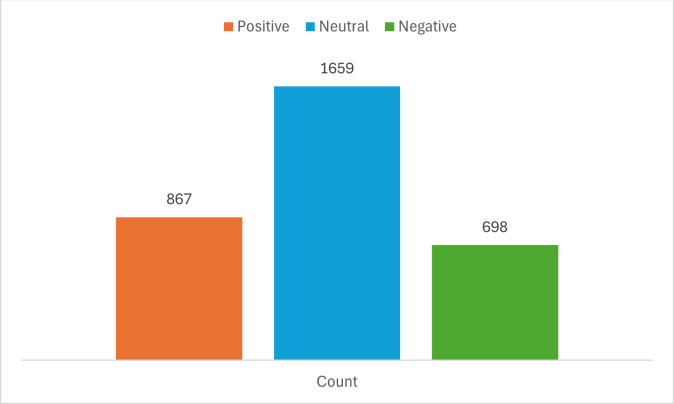

Figure 4: Data before SMOTE.

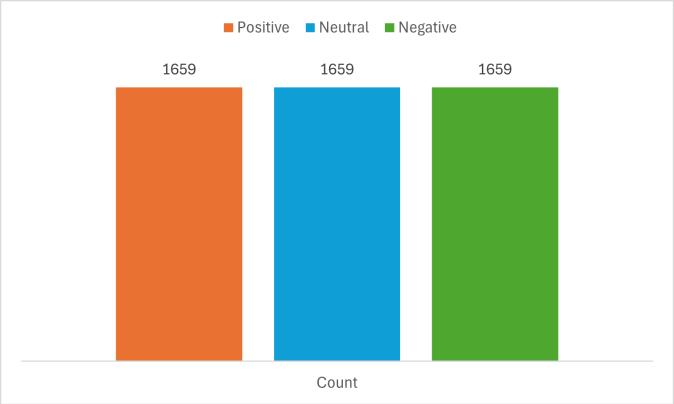

Figure 5: Data before SMOTE.

2.8 Optimization Method

In this study, PSO is used as the optimization method to tune the parameters of SVM. PSO optimizes a problem iteratively, starting with a swarm of particles, each representing a candidate solution. Each particle knows its personal best position and value, as well as the global best position within the swarm. At each iteration, the velocity and position of each particle are updated based on both its knowledge and the swarm's collective knowledge,

Table 3: Kernel functions expression

Kernel Function	Function Expression	Parameter
Linear	$K(x_i, x_j) = x_i \cdot x_j$	C
RBF	$K(x_i, x_j) = \exp\left(-\gamma \|\|x_i - x_j\|\|^2 + C\right)$	C and γ
Sigmoid	$K(x_i, x_j) = \tanh(\gamma(x_i \cdot x_j) + r)$	C, γ, and r
Polynomial	$K(x_i, x_j) = (\gamma(x_i \cdot x_j) + r)^d$	C, γ, r, and d

guiding the particles toward the optimal solution until a stopping criterion is met [22]. PSO runs by applying a population of particles with their respective velocities and positions. The algorithm iterates over all the updated particles until it reaches the optimal value. Each particle gets a solution from the process, individually or globally, to update all velocities and positions [23].

2.9 Model Evaluation (Evaluation)

A standard model evaluation method is the confusion matrix, which describes the model prediction results in matrix form to further evaluate the model's performance. The confusion matrix shows the number of correct and incorrect predictions in specific categories, namely true positive (TP), false positive (FP), true negative (TN), and false negative (FN). From these values, various evaluation metrics such as accuracy, precision, recall, and F1-score can be calculated to assess how well the model works [24]. The corresponding formulas for these metrics are as follows: (1)-(4) [24].

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}} + \text{FN} \tag{2}$$

$$\text{Precision } = \frac{\text{TP}}{\text{TP}} + \text{FP} \tag{3}$$

$$\text{F1-score } = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

3 Results

In this section, the results of two experiments carried out to determine the optimization accuracy between the SVM method and the SVM optimized with PSO will be presented. The results of each experiment were analyzed to compare the performance of the two methods in achieving the best accuracy on the SVM model.

3.1 Experiment A

In the first experiment, the SVM model was tested by changing the value of the parameter C to analyze its impact on the accuracy of the model. The tested C values include 0.01, 0.1, 1, 10, and 100. In addition, the parameter gamma (γ) was set to 10, the parameter r was set to 0, and the parameter degree (d) was set to 3. Table 4 presents the accuracy obtained for each variation value of the parameter C [25].

From Table 4, each kernel achieves the highest accuracy at a certain C parameter value. The Linear kernel achieves the highest accuracy of 80.39% at a value of C = 5, while the RBF kernel also provides the highest accuracy of 80.96% at the same value of C = 5. For the Sigmoid kernel, the highest accuracy of 79.73% is obtained at a value of C = 1. Meanwhile, the Polynomial kernel shows the highest accuracy of 57.52% at a value of C = 1.

Table 4: Experiment A result

Parameter C	Accuracy (%)			
	Linear	RBF	Sigmoid	Polynomial
0.01	66.28	67.29	65.19	56.72
0.1	68.08	69.46	68.01	55.64
1	79.95	79.45	79.73	57.52
5	80.39	80.96	78.36	57.38

3.2 Experiment B

In the second experiment, SVM parameter optimization was performed using PSO to find the combination of parameters that maximized accuracy. The result of the optimization is presented in Table 5.

Table 5: Experiment B result

Kernel Function	Parameter				Accuracy (%)
	c	γ	r	d	
Linear	32.50	-	-	-	80.53
RBF	53.37	0.073	-	-	82.05
Sigmoid	71.27	0.096	-1	-	80.96
Polynomial	17.22	0.18	1	3	80.82

Based on the results presented in Table 5, the RBF kernel achieved the highest accuracy of 82.05%, followed by the Polynomial kernel at 80.82%, the Sigmoid kernel at 80.96%, and the Linear kernel with the lowest accuracy of 80.53%.

3.3 Comparison of SVM and SVM Optimized PSO

Table 6 compares the performance of SVM and SVM-optimized PSO on four kernel functions: Linear, RBF, Sigmoid, and Polynomial. The results include important evaluation metrics such as Accuracy, Recall, Precision, and F-1 Score.

Table 6: Comparison of SVM and SVM-optimized PSO

Kernel Function	**Method**	**Accuracy (%)**	**Recall (%)**	**Precision (%)**	**F-1 Score (%)**
Linear	SVM	80.39	80	80	80
	SVM + PSO	80.53	81.39	80	80.07
RBF	SVM	80.96	81	81	81
	SVM + PSO	82.05	81.21	82.14	81.88
Sigmoid	SVM	79.73	80	80	80
	SVM + PSO	80.96	81.21	80.86	80.87
Polynomial	SVM	57.67	57.98	67	49
	SVM + PSO	80.82	80.57	80.57	80.35

Based on Table 6, the SVM optimized with PSO shows a significant performance improvement in all kernels, especially in RBF and polynomial kernels. For the RBF kernel, optimization with PSO results in accuracy improvement from 80.96% to 82.05%, precision improvement from 81% to 82.14%, and F-1 score improvement from 81% to 81.88%. Meanwhile, the Polynomial kernel, which initially performed poorly on SVM with an accuracy of 57.67%, shows drastic improvement with an accuracy increase to 80.82%. Although the Sigmoid and Linear kernels also improved, the difference in accuracy and recall was relatively small after optimization with PSO, with the precision of the Linear kernel increasing from 80.39% to 80.53% and the recall of the Sigmoid kernel increasing from 80% to 81.21%. These results indicate that optimization with PSO is very effective in improving model performance.

4 Discussion

Based on the results of Experiment A, the SVM model was tested with various parameter values to analyze their effect on accuracy. The test results show that each kernel has different optimal parameters, and some kernels perform well with certain parameter choices. In Experiment B, hyperparameter optimization was performed using PSO. PSO was used to optimize the SVM results because it allows for a more comprehensive and efficient exploration of the hyperparameter space than manual tuning. PSO improves the effectiveness of optimization by organizing a group of particles, each representing a possible solution. Each particle updates its speed and position based on the best results obtained by itself and the best results found by the group to help find the best solution. This tuning allows PSO to avoid suboptimal choices and find the best combination of parameters more efficiently, thus significantly improving the model's accuracy. The results show a significant performance improvement. This indicates that PSO optimizes SVM models by ensuring better hyperparameter selection and maximizing model performance. However, the drawback of using PSO is the computational complexity when applied to problems with very large search spaces, where the optimal search can take considerable time.

5 Conclusion

This study has applied sentiment analysis to evaluate the reputation of UNNES using an optimized SVM model using PSO. The study stages include data collection, preprocessing, labeling, splitting, feature extraction using TF-IDF, SMOTE oversampling, SVM classification, and PSO optimization. The proposed approach successfully classifies the public sentiment of UNNES's reputation into positive, neutral, or negative categories. This study tested four SVM kernels: Linear, RBF, Sigmoid, and Polynomial. The evaluation results showed that the optimized RBF kernel with PSO achieved the best performance, with 82.05% accuracy, 81.21% recall, 82.14% precision, and 81.88% F1-score. These results indicate that the proposed approach successfully classifies the sentiment of tweets related to UNNES, providing an effective tool for monitoring the university's reputation on social media.

Although this study has provided good results in sentiment analysis of UNNES' reputation, some limitations must be considered. One of them is the use of data only in Bahasa Indonesia. For future study, it would be beneficial to expand the scope of sentiment anal-

ysis to include data in other languages, such as English or other regional languages, to obtain a more holistic understanding of public sentiment towards universities.

Acknowledgments

This work was supported by an inter-institutional research collaboration of the Postgraduate School of Universitas Negeri Semarang No: DPA 023.17.2.690645/2024.09. Special thanks to the Artificial Intelligence, Robotics & Internet of Things (AIRIoT) Laboratory, Digital Center, Universitas Negeri Semarang, for the facilities provided during the study process.

References

[1] A. M. A. Ausat, "The role of social media in shaping public opinion and its influence on economic decisions," *Technology and Society Perspectives (TACIT)*, vol. 1, no. 1, pp. 35–44, 2023.

[2] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," *Procedia Computer Science*, vol. 161, pp. 707–714, 2019.

[3] M. A. Mateus, A. G. Rincón, F. J. Acosta, I. R. Soler, and D. R. Valero, "Keys to managing university reputation from the students' perspective," *Heliyon*, vol. 10, no. 21, 2024.

[4] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "Social media analytics–challenges in topic discovery, data collection, and data preparation," *International journal of information management*, vol. 39, pp. 156–168, 2018.

[5] I. B. Prakoso, D. Richasdy, and M. D. Purbolaksono, "Sentiment analysis of telkom university as the best bpu in indonesia using the random forest method," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, p. 2050, 2022.

[6] A. Alfarel, H. Hasmawati, B. Bunyamin, *et al.*, "Sentiment analysis of telkom university using the long short-term memory and word2vec feature expansion," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 17, no. 2, pp. 468–481, 2024.

[7] F. Y. Sari, M. sukma Kuntari, H. Khaulasari, and W. A. Yati, "Comparison of support vector machine performance with oversampling and outlier handling in diabetic disease detection classification," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 3, pp. 539–552, 2023.

[8] D. M. Abdullah and A. M. Abdulazeez, "Machine learning applications based on svm classification a review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 81–90, 2021.

[9] J. Yang, Z. Wu, K. Peng, P. N. Okolo, W. Zhang, H. Zhao, and J. Sun, "Parameter selection of gaussian kernel svm based on local density of training set," *Inverse Problems in Science and Engineering*, vol. 29, no. 4, pp. 536–548, 2021.

[10] A. Sa'adah, A. Sasmito, A. A. Pasaribu, *et al.*, "Comparison of genetic algorithm (ga) and particle swarm optimization (pso) for estimating the susceptible-exposed-infected-recovered (seir) model parameter values," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 2, pp. 290–301, 2024.

[11] S. Cheng, H. Lu, X. Lei, and Y. Shi, "A quarter century of particle swarm optimization," *Complex & Intelligent Systems*, vol. 4, no. 3, pp. 227–239, 2018.

[12] X. Shu and Y. Ye, "Knowledge discovery: Methods from data mining and machine learning," *Social Science Research*, vol. 110, p. 102817, 2023.

[13] T. Husain and N. Hidayati, "The optimize of association rule method for the best book placement patterns in library: A monthly trial," *Teknokom*, vol. 4, no. 2, pp. 53–59, 2021.

[14] S. S. Berutu, H. Budiati, J. Jatmika, and F. Gulo, "Data preprocessing approach for machine learning-based sentiment classification," *Jurnal Infotel*, vol. 15, no. 4, pp. 317–325, 2023.

[15] S. R. Aisy and B. Prasetiyo, "Sentiment analysist of the tpks law on twitter using inset lexicon with multinomial naïve bayes and support vector machine based on soft voting," *Recursive Journal of Informatics*, vol. 1, no. 2, pp. 93–101, 2023.

[16] A. Larasati, R. P. Editya, Y.-W. Chen, and V. E. Darmawan, "The effect of data splitting ratio and vectorizer method on the accuracy of the support vector machine and naïve bayes model to perform sentiment analysis," in *2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, pp. 1–6, IEEE, 2023.

[17] C. N. Agustina, R. Novita, N. E. Rozanda, *et al.*, "The implementation of tf-idf and word2vec on booster vaccine sentiment analysis using support vector machine algorithm," *Procedia Computer Science*, vol. 234, pp. 156–163, 2024.

[18] M. A. Toçoğlu and A. Onan, "Sentiment analysis on students' evaluation of higher educational institutions," in *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions: Proceedings of the INFUS 2020 Conference, Istanbul, Turkey, July 21-23, 2020*, pp. 1693–1700, Springer, 2021.

[19] M. Achirul Nanda, K. Boro Seminar, D. Nandika, and A. Maddu, "A comparison study of kernel functions in the support vector machine and its application for termite detection," *Information*, vol. 9, no. 1, p. 5, 2018.

[20] D. Elreedy and A. F. Atiya, "A comprehensive analysis of synthetic minority oversampling technique (smote) for handling class imbalance," *Information Sciences*, vol. 505, pp. 32–64, 2019.

[21] M. Achirul Nanda, K. Boro Seminar, D. Nandika, and A. Maddu, "A comparison study of kernel functions in the support vector machine and its application for termite detection," *Information*, vol. 9, no. 1, p. 5, 2018.

[22] D. Freitas, L. G. Lopes, and F. Morgado-Dias, "Particle swarm optimisation: a historical review up to the current developments," *Entropy*, vol. 22, no. 3, p. 362, 2020.

[23] A. B. P. Utama, A. P. Wibawa, M. Muladi, and A. Nafalski, "Pso based hyperparameter tuning of cnn multivariate time-series analysis," *Jurnal Online Informatika*, vol. 7, no. 2, pp. 193–202, 2022.

[24] Ž. Vujović *et al.*, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.

[25] M. Achirul Nanda, K. Boro Seminar, D. Nandika, and A. Maddu, "A comparison study of kernel functions in the support vector machine and its application for termite detection," *Information*, vol. 9, no. 1, p. 5, 2018.