Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 9, No. 1, February 2024, Pp.101-108

Optimized support vector machine with particle swarm optimization to improve the accuracy amazon sentiment analysis classification

Maylinna Rahayu Ningsih[1], Jumanto Unjung*[2], Dwika Ananda Agustina Pertiwi[3], Budi Prasetiyo[4], Much Aziz Muslim[5]
Department of Computer Science, FMIPA, Universitas Negeri Semarang, Semarang, Indonesia[1,2,4]
Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia [3,5]

Article Info

Keywords:
Particle Swarm Optimization, Classification, Amazon, SVM

Article history:
Received: November 02, 2023
Accepted: January 24, 2024
Published: February 28, 2024

Cite:
M. R. Ningsih, J. Unjung, D. A. A. Pertiwi, B. Prasetiyo, and M. A. Muslim, "Optimized Support Vector Machine with Particle Swarm Optimization to Improve the Accuracy Amazon Sentiment Analysis Classification", KINETIK, vol. 9, no. 1, Feb. 2024. Retrieved from https://kinetik.umm.ac.id/index.php/kinetik/article/view/1888

*Corresponding author.
Jumanto Ujung
E-mail address:
jumanto@mail.unnes.ac.id

Abstract

Text mining is a valuable technique that empowers users to gain a deeper understanding of existing textual data, ultimately allowing them to make more informed decisions. One important application of text mining is in the field of sentiment analysis, which has gained significant traction among companies aiming to understand how customers perceive their products and services. In response to this growing need, various research efforts have been made to improve the accuracy of sentiment analysis classification models. The purpose of this article is to discuss a specific approach using the Support Vector Machine (SVM) algorithm, which is often used in machine learning for text classification tasks and then combined with the application of Particle Swarm Optimization (PSO), which optimizes the SVM model parameters to achieve the best classification results. This dynamic combination not only improves accuracy but also enhances the model's ability to efficiently handle large amounts of text data to achieve better results. The research findings highlight the effectiveness of this approach. The application of the SVM algorithm with PSO resulted in an outstanding accuracy performance of 94.92%. The substantial increase in accuracy compared to previous studies shows the promising potential of this methodology. This proves that the SVM algorithm model approach with Particle Swarm Optimization provides good performance.

1. Introduction

Today's internet infrastructure is increasingly in demand due to its convenience and efficiency. In the advancing digital age, big companies such as Amazon, Google, and many others have pursued digital transformation [1] to meet the needs of their users. The sustainability of this transformation demands a deep understanding of customers' views and desires towards the products or services they offer. Along with this development, sentiment analysis, which is a technique of understanding and analyzing customer reactions, opinions, and feelings towards a service or product, has become an important element in developing business strategies in the modern era.

Websites like Amazon are one of the main platforms customers use to buy and leave reviews on the products or services they use, Amazon is one of the e-Commerce companies [1] with the largest market in the world. Sentiment analysis of buyer reviews is needed to increase company performance and income [2]. In the same way that word of mouth is heard offline, online reviews can help consumers get more accurate information about their products [3]. Specification ratings often reflect user satisfaction with aspects of the product [4], Buyer reviews will generally be a measure for other customers [5] to consider purchasing on the platform. These reviews contain both positive and negative sentiments [6]. For this reason, it is important for a business to understand the customer's view of their product or service by analyzing the sentiments of the reviews provided to gain insight into customer satisfaction and make data-driven decisions. Therefore, knowing these factors is useful to increase sales and income [7].

However, it should be recognized that the increasing volume of unstructured data in online reviews is a serious challenge. For this reason, a machine is needed to recognize it, namely text mining [8]. One of the text mining methods is the Support Vector Machine (SVM) algorithm. This Support Vector Machine (SVM) algorithm is a machine learning algorithm that is widely used to classify sentiment data. Classification using SVM is considered one of the strongest algorithms in machine learning [9]. The SVM algorithm [10] includes supervised learning for classification [11], which separates an optimal hyperplane [12] that takes a low-dimensional input vector and maps it to a higher-dimensional feature space. The SVM is also capable of identifying two different classes [13]. In addition, the Particle Swarm Optimization algorithm is an optimization algorithm that applies several parameters [14] and then modifies it. PSO is also defined as the basic concept of the study of bird foraging behavior or is called group-based optimization technology [15]. Techniques that can be used include increasing the attribute weight of attribute selection [16]. Particle Swarm Optimization (PSO) is a method that can help to select the most optimal parameters, that is, produce the highest classification accuracy [17]. The accuracy of a method is also affected by feature weighting [8]. This study used TF-IDF to [18], [19] investigate the relationship between features and text data and with a 10-fold cross-validation optimization.

There are several studies related to sentiment analysis, such as research [8] using the chi square technique to optimize the Naïve Bayes algorithm, and it is has been shown that the use of chi square and TF-IDF weighting improve the performance of Naïve Bayes in predicting Amazon customer sentiment with a final accuracy of 83%. Another study [20] combined Glove and Word2Vec word embedding with CNN-based architecture to analyze sentiment in drug product reviews with an accuracy performance of 84.87%. Further research grouped reviews based on affective characteristics of customers towards products such as happy or disappointed based on reviews on amazon conducted [21] using the development of the clustering algorithm "Cluster-then-label". Another study [22] discussed sentiment prediction on amazon customer reviews using the Naïve Bayes and Support Vector Machine (SVM) algorithms, where the SVM algorithm has better performance.

In this study, we delve into the realm of optimizing the performance of the SVM algorithm for analyzing Amazon customer review data. Although previous research achieved commendable results, there is still untapped potential to improve. In our attempt to improve the accuracy of the analysis, we used the particle swarm optimization (PSO) method to improve the SVM algorithm. We aimed to raise the boundaries of existing performance levels by using this new combination, contributing to the optimization of sentiment analysis in terms of Amazon customer reviews. This integration is a special part of our work, showing a promising path to progress in this area.

2. Research Method

The proposed method used the SVM method optimized with PSO. Before processing with the proposed model, the dataset was pre-processed, starting from Lowercase, Clean text by removing elements that interfere with text analysis such as tags, urls, etc., Stopwords Removal, and Lemmitazion. After the preprocessing stage, we continued with feature extraction with TF IDF and then modeling with SVM algorithm optimized with particle swarm optimization (PSO) and then evaluated with 10-fold cross-validation and confusion matrix. The method used in this research consists of the following steps as presented in Figure 1.

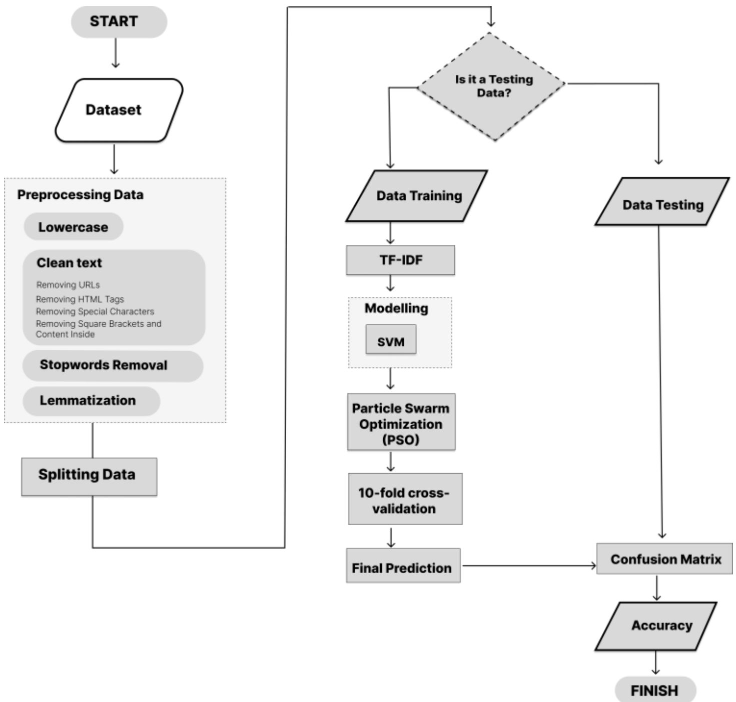

Figure 1. The Model Proposed by the Researcher

2.1 Dataset

The database used in this study were Amazon reviews written in English taken from the kaggle site https://www.kaggle.com/datasets/sid321axn/amazon-alexa-reviews. The data with tsv extension was then uploaded to Google Colab for data processing. This dataset contained over 3000 Amazon customer reviews (input text), star ratings, review date, variant, and comments for various Amazon Alexa goods such as Alexa Echo, Echo dots, Alexa Firesticks, and so on.

2.2 Preprocessing

The pre-processing step aimed to prepare the data [23] so that the pattern can be extracted [24].This stage involved the process of Lowercase, Clean text by removing elements that interfere with text analysis such as tags, urls etc., Stopwords Removal and Lemmitazion. This stage involved several processes, such as ensuring consistency in data analysis, and all text was converted to lowercase. Next, we cleaned the text by removing elements that can interfere with text analysis [25], such as removing elements such as HTML tags, URLs, and special characters from the text. Removing these elements makes the data cleaner and allows more attention to the core text information. Next, the process of removing common words, or Stopwords Removal, from the text was performed. In this case, a list of English stop words was downloaded and used to remove these words from the text [26]. The aim was to reduce the noise in the data and concentrate on words that have greater meaning.This stage plays an important role with the aim of cleaning the data so that they are ready to be processed and it is expected to increase the results for further processing [27]. To see a comparison of text samples before and after pre-processing, see Table 1.

Table 1. Comparison of Text Samples Before and After Preprocessing

Before Preprocessing	After Preprocessing
Sometimes while playing a game, you can answer a question correctly but Alexa says you got it wrong and answers the same as you. I like being able to turn lights on and off while away from home.	sometimes playing game answer question correctly alexa say got wrong answer like able turn light away home
I liked the original Echo. This is the same but shorter and with greater fabric/color choices. I miss the volume ring on top, now it's just the plus/minus buttons. Not a big deal but the ring w as comforting. :) Other than that, well I do like the use of a standard USB charger /port instead of the previous round pin. Other than that, I guess it sounds the same, seems to work the same, still answers to Alexa/Echo/Computer. So what's not to like? :)	liked original echo shorter greater fabriccolor choice miss volume ring top plusminus button big deal ring w comforting well like use standard usb charger port instead previous round pin guess sound seems work still answer alexaechocomputer whats like

2.3 Feature Extraction

An important part of sentiment analysis is feature extension, which includes the process of finding and selecting key words from the text to evaluate sentiment. In sentiment analysis, a method was used to evaluate the significance of words in sentiment text data [28], namely TF-IDF (Term Frequency-Inverse Document Frequency) [18]. The significance of a term in a larger context was then assessed using the Inverse Document Frequency (IDF) component. This was particularly important because simple words such as "is", "a", and "and" were often used but had no special meaning. This method worked by dividing the number of times a term appears in a document [29] by the total number of words in the document to find the term frequency (TF) of the term (t) in the document. TF-IDF formula in a document is formulated by applying Equation 1 and Equation 2.

$$tfidf\ (t,d,D) = tf\ (t,d) \times idf\ (t,D) \tag{1}$$

The word frequency tf (t, d) is the number of occurrences of term t in document d. For a term t in a corpus of documents D, the inverse document frequency idf (t, D) is:

$$idf\ (t,D) = log \frac{N}{|d\ (\in D : t\ \in D)|} \tag{2}$$

where N is the total number of documents in the corpus D and $|d\ (\in D : t\ \in D)|$ is the number of documents containing the phrase t.

2.4 Classification Method

The method employed in sentiment analysis used the SVM algorithm approach and feature selection with PSO optimization to improve accuracy performance. According to the authors, the SVM algorithm is used to identify hyperplanes, which are points of interest or patterns in data, and to classify texts into positive, negative, or neutral categories [30]. We applied a 10-fold cross-validation process to test and validate the performance of the techniques used. This method has great benefits: the risk of overfitting is reduced [31], which makes sentiment analysis results more reliable. In addition, the feature selection method with PSO an optimization algorithm based on the idea of a colony of particles moving simultaneously towards an optimal solution is a smart and effective strategy for selecting the most relevant key words in the text for sentiment evaluation. Particle Swarm Optimization with a simple concept is easy to implement and efficient when compared with other optimization algorithms [32]. This experiment is expected to provide more accurate and reliable sentiment analysis results for identifying sentiment in text that is evaluated using a combination of the PSO and SVM feature selection algorithms.

2.5 Model Evaluation

Following the analysis of the experiments, the optimized SVM learning model with PSO should be evaluated. The purpose of this test is to acquire a better understanding of how the model works and its ability to characterize sentiment in text data. The Confusion Matrix is a very informative evaluation tool to be used in this context. The confusion matrix is a method that allows us to measure how well our model succeeds in categorizing text into various sentiment categories, such as positive, negative, or neutral. Using this matrix, we can determine how many predictions are correct for the positive category, how many predictions are incorrect for the negative category, and how many predictions are correct for the negative category. We can calculate various relevant evaluation metrics, such as accuracy, precision, recall, and F1 score, with the information generated from the confusion matrix. This will help us thoroughly assess the ability of PSO-optimized SVM models to perform sentiment analysis. The evaluation metrics calculation can be seen in the following Equations 3 to 6.

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Precission = \frac{TP}{TP + FP} \quad (5)$$

$$F1 - Score = \frac{2\ x\ precision\ x\ recall}{precision +\ recall} \quad (6)$$

3. Results and Discussion

The stages carried out in this research were dataset processing, preprocessing, SVM model, testing with SVM, and PSO optimization. Starting from preparing the dataset to be processed, then pre-processing was carried out including the Tokenize, Transform Case, Stopword Removal, and Stemming processes.

The processing was then carried out with feature weighting with TF-IDF with 10-fold cross-validation optimization, where [33] the data set was optimized into 10 segments of the same size in each testing and training process. Then, further feature selection was carried out, this study used Particle Swarm Optimization feature selection. The level of accuracy produced by the Support Vector Machine (SVM) algorithm is 94.92%.

Then after preprocessing, with the condition that positive reviews have a rating of more than 4, while negative reviews have a rating of less than 2. Wordcloud is used for visual representation of text data [34] to describe positive and negative word reviews, as shown in Figure 2.

(a) (b)

Figure 2. (a) Wordcloud Reviews Postive, (b) Wordcloud Reviews Negative

Terms that appear in larger sizes indicate that the word is more likely to be used than other words. This can be seen from the results of this Wordcloud review when comparing the positive and negative wordcloud reviews. The phrases "love", "Alexa", "use", "great", "music", "echo", and "easy" have a larger font size in the like word cloud, indicating these words are used more often in positive reviews. Whereas in negative Wordcloud reviews, words like “device”, “work”, “amazon”, “alexa”, echo, time, one Wordcloud can provide an overview of words that often appear in positive reviews and negative reviews, which can help understand review sentiment quickly and effectively.

This experiment applied the particle swarm optimization (PSO) method to optimize the SVM model for sentiment analysis. In applying the optimization, the optimal parameters found by PSO were C 10.0 and gamma=0.186. This parameter describes the SVM model that has optimal performance based on accuracy metrics on the training data. Next, the SVM model was trained with these optimal parameters. The results show that the SVM model that has been optimized and evaluated using 10-Fold Cross-Validation provides high accuracy performance. This shows that the SVM model optimized with PSO indicates good classification ability. Performance can be classified and evaluated with the confusion matrix [11]. The confusion matrix is an evaluation of model performance in predicting targets which can then be used to calculate precision, recall, and f1-score values. To see the results of the Confusion Matrix, refer to Table 2.

Table 2. Confusion Matrix of SVM with PSO

		Predicted	
		0	**1**
Real	**0**	29	29
	1	3	569

This table has four different cells that show the various results of the classification process. The two main categories of sentiment analysis consist of category "0" that indicates negative sentiment and category "1" that indicates positive sentiment. In the top left corner of the table, we have 29 correct predictions (True Negatives, TN) for category “0,” indicating that the model correctly classified 29 texts as negative sentiment. In the top right corner of the table, we have 29 false predictions (False Positives, FP) for category “0,” indicating that the model classified 29 texts as positive sentiment even though they were actually negative sentiment. In the lower left corner of the table, there are three false estimates (False Negatives, FN) for category "1", which indicates that the model considers three texts to be negative sentiments even though they are actually positive sentiments. In the bottom right corner of the table, there are 569 correct estimates (True Positives, TP) for category "1", indicating that the model correctly considers 569 texts as positive sentiment. These results show an evaluation report on the performance of the SVM model in classification. The SVM model with PSO optimization shows model classification performance with an accuracy of 94.92% which is better than previous research.

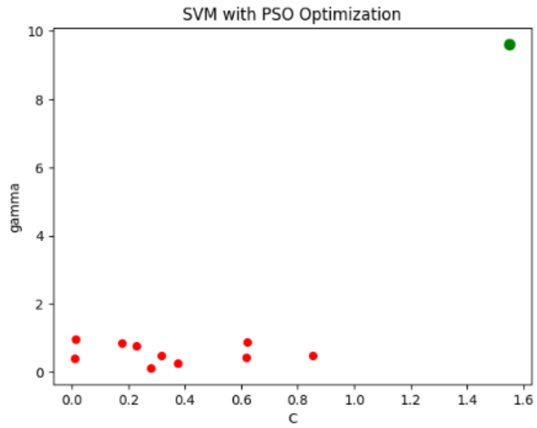

Figure 3. SVM with PSO Optimazion

The plot displayed in Figure 3 is the results of SVM optimization using the Particle Swarm Optimization (PSO) algorithm. The aim was to provide a visual summary of the SVM model optimization results, by describing the position of the cluster and the best position (global best) found by the PSO algorithm. The red color shows the position of the cluster produced by the PSO algorithm, while the green color shows the best position (global best) found by the algorithm.

In this research, efforts have been made to improve classification accuracy in previous studies. By using Particle Swarm Optimization (PSO) algorithm to optimize the SVM model, the optimal parameters C=10.0 and gamma=0.186 were found. To see the comparison between the performance of this research and other studies, it is shown in Table 3.

Table 3. Model Performance Comparison

Comparison	Algorithm	Accuracy (%)
In [8]	Naive Bayes + Chi Square + TF-IDF	83
In [20]	Glove and Word2Vec word embedding with CNN	84,87
In [35]	Multilayer Perceptual (MLP) classifier	88
In [36]	VADER Sentiment + SVM	80
In [37]	SVM	59
Proposed method	**SVM+PSO**	**94,92**

This research successfully showed superior performance compared to previous studies, as evidenced by the highest accuracy reaching 94.92% with the proposed SVM+PSO approach.

4. Conclusion

The dataset used in this research was taken from Kaggle. This research proposed SVM algorithm modeling using PSO feature optimization. After processing starting from dataset preparation, pre-processing, classification, and model evaluation, SVM with the PSO feature obtained an accuracy of 94.92%. The solution to the problem of the SVM algorithm not being able to work well on large data, in addition to that a large number of attributes will result in low accuracy due to decreased performance, can be solved in this research by applying PSO (Particle Swarm Optimization) optimization to increase accuracy. Apart from these problems, further research is needed on how the results of this research can be applied to selling products on Amazon Alexa. Future research could try to classify by topic, not just by sentiment. Furthermore, considering the potential implications for future research, evaluating the adaptability of the proposed SVM+PSO model in various real-world datasets and scenarios can enhance its applicability. Moreover, recognizing the limitations of the study, such as dataset-specific constraints and potential problems in wider applications, will promote better understanding, paving the way for future improvements.

References

[1] I. Castell-uroz, P. Barlet-ros, dan U. Polit, “Amazon Alexa traffic traces,” vol. 205, 2022. https://doi.org/10.1016/j.comnet.2022.108782

[2] M. Shaden, A. Fadel, S. Achmad, dan R. Sutoyo, “Sentiment analysis for customer review : Case study of Traveloka Sentiment analysis for customer review : Case study of Traveloka,” *Procedia Comput. Sci.*, vol. 216, no. 2022, hal. 682–690, 2023. https://doi.org/10.1016/j.procs.2022.12.184

[3] C. Kim dan Y. Na, “Consumer reviews analysis on cycling pants in online shopping malls using text mining,” *Fash. Text.*, 2021. https://doi.org/10.1186/s40691-021-00264-7

[4] T. Nguyen, T. Ngoc, H. Nguyen, T. Thu, dan V. A. Nguyen, “Mining aspects of customer ’ s review on the social network,” *J. Big Data*, hal. 1–21, 2019. https://doi.org/10.1186/s40537-019-0184-5

[5] L. Kong, C. Li, J. Ge, V. Ng, dan B. Luo, “Predicting Product Review Helpfulness – A Hybrid Method,” hal. 1–14, 2020. https://doi.org/10.1109/TSC.2020.3041095

[6] V. A. Fitri, R. Andreswari, M. A. Hasibuan, V. A. Fitri, R. Andreswari, dan M. A. Hasibuan, “Sentiment Analysis of Social Media Twitter with Case of Anti- Sentiment Analysis of Social Media Twitter with Case of Anti- LGBT Campaign in Indonesia using Naïve Bayes , Decision Tree , LGBT Campaign in Indonesia using Naïve Bayes , Decision Tree , and R,” *Procedia Comput. Sci.*, vol. 161, hal. 765–772, 2019. https://doi.org/10.1016/j.procs.2019.11.181

[7] Y. Zhou dan S. Yang, “Roles of Review Numerical and Textual Characteristics on Review Helpfulness Across Three Different Types of Reviews,” *IEEE Access*, vol. 7, hal. 27769–27780, 2019. https://doi.org/10.1109/ACCESS.2019.2901472

[8] A. Falasari dan M. A. Muslim, “Optimize Naïve Bayes Classifier Using Chi Square and Term Frequency Inverse Document Frequency For Amazon Review Sentiment Analysis,” *J. Soft Comput. Explor.*, vol. 3, no. 1, hal. 31–36, 2022. https://doi.org/10.52465/joscex.v3i1.68

[9] B. Navaneeth dan M. Suchetha, “PSO optimized 1-D CNN-SVM architecture for real-time detection and classi fi cation applications,” *Comput. Biol. Med.*, vol. 108, no. September 2018, hal. 85–92, 2019. https://doi.org/10.1016/j.compbiomed.2019.03.017

[10] J. Unjung dan M. R. Ningsih, “Optimized Handwriting-based Parkinson ’ s Disease Classification Using Ensemble Modeling and VGG19 Feature Extraction,” *Sci. J. Informatics*, vol. 10, no. 4, hal. 489–498, 2023. https://doi.org/10.15294/sji.v10i4.47108

[11] D. Valero-carreras, J. Alcaraz, dan M. Landete, “Computers and Operations Research Comparing two SVM models through different metrics based on the confusion matrix,” *Comput. Oper. Res.*, vol. 152, no. April 2022, hal. 106131, 2023.https://doi.org/10.1016/j.cor.2022.106131

[12] A. F. Limas, R. Rosnelly, dan A. Nursie, “A Comparative Analysis on the Evaluation of KNN and SVM Algorithms in the Classification of Diabetes,” *Sci. J. Informatics*, vol. 10, no. 3, hal. 251–260, 2023. https://doi.org/10.15294/sji.v10i3.44269

[13] D. Asante, T. Omar, A. Ganat, R. Gholami, dan S. Ridha, “Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties : Comparative analysis of ANN and SVM models,” *J. Pet. Sci. Eng.*, vol. 200, no. November 2020, hal. 108182, 2021. https://doi.org/10.1016/j.petrol.2020.108182

[14] R. Zheng, Y. Bao, L. Zhao, dan L. Xing, “Method to predict alloy yield based on multiple raw material conditions and a PSO-LSTM network,” *J. Mater. Res. Technol.*, 2023. https://doi.org/10.1016/j.jmrt.2023.10.046

[15] W. Zhou, M. Chen, Z. Yang, dan X. Song, “Socio-Economic Planning Sciences Real estate risk measurement and early warning based on PSO-SVM,” *Socioecon. Plann. Sci.*, no. November, hal. 101001, 2020. https://doi.org/10.1016/j.seps.2020.101001

[16] A. Dongoran, S. Rahmadani, Zakarias, M. Zarlis, dan Zakarias, “Feature weighting using particle swarm optimization for learning vector quantization classifier Feature weighting using particle swarm optimization for learning vector quantization classifier,” *IOP Publ.*, 2018. https://doi.org/10.1088/1742-6596/978/1/012032

[17] A. Nazir, A. Akhyar, dan E. Budianita, “Toddler Nutritional Status Classification Using C4 . 5 and Particle Swarm Optimization,” vol. 9, no. 1, hal. 32–41, 2022. https://doi.org/10.15294/sji.v9i1.33158

[18] A. Mee, E. Homapour, F. Chiclana, dan O. Engel, "Sentiment analysis using TF – IDF weighting of UK MPs ' tweets on Brexit," *Knowledge-Based Syst.*, vol. 228, hal. 107238, 2021. https://doi.org/10.1016/j.knosys.2021.107238

[19] J. Jumanto, M. A. Muslim, Y. Dasril, dan T. Mustaqim, "Accuracy of Malaysia Public Response to Economic Factors During the Covid-19 Pandemic Using Vader and Random," *J. Inf. Syst. Explor. Res.*, vol. 1, no. 1, hal. 49–70, 2023. https://doi.org/10.52465/joiser.v1i1.104

[20] N. H. Jeremy, D. Suhartono, dan S. Philip, "Deep neural networks and weighted word embeddings for sentiment analysis of drug product reviews analysis," *Procedia Comput. Sci.*, vol. 216, no. 2022, hal. 664–671, 2023. https://doi.org/10.1016/j.procs.2022.12.182

[21] W. Kim, K. Nam, dan Y. Son, "Electronic Commerce Research and Applications Categorizing affective response of customer with novel explainable clustering algorithm : The case study of Amazon reviews," *Electron. Commer. Res. Appl.*, vol. 58, no. February, hal. 101250, 2023. https://doi.org/10.1016/j.elerap.2023.101250

[22] P. Pandey dan N. Soni, "Sentiment Analysis on Customer Feedback Data : Amazon Product Reviews," *2019 Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput.*, hal. 320–322, 2019. https://doi.org/10.1109/COMITCon.2019.8862258

[23] M. R. Ningsih, K. Aalifian, H. Wibowo, dan A. U. Dullah, "Global recession sentiment analysis utilizing VADER and ensemble learning method with word embedding," *J. Soft Comput. Explor.*, vol. 4, no. 3, hal. 142–151, 2023. https://doi.org/10.52465/joscex.v4i3.193

[24] R. A. Sinoara, "Text mining and semantics : a systematic mapping study," 2017. https://doi.org/10.1186/s13173-017-0058-7

[25] T. Singh dan M. Kumari, "Role of Text Pre-Processing in Twitter Sentiment Analysis," *Procedia - Procedia Comput. Sci.*, vol. 89, hal. 549–554, 2016. https://doi.org/10.1016/j.procs.2016.06.095

[26] R. Rani dan D. K. Lobiyal, "Performance evaluation of text-mining models with Hindi stopwords lists," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, hal. 2771–2786, 2022. https://doi.org/10.1016/j.jksuci.2020.03.003

[27] T. Hema *et al.*, "Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as development using classifier," *Procedia Comput. Sci.*, vol. 197, no. 2021, hal. 660–667, 2022. https://doi.org/10.1016/j.procs.2021.12.187

[28] R. Ahuja, A. Chug, S. Kohli, S. Gupta, dan P. Ahuja, "The Impact of Features Extraction on the Sentiment Analysis," *Procedia Comput. Sci.*, vol. 152, hal. 341–348, 2019. https://doi.org/10.1016/j.procs.2019.05.008

[29] G. N. H, R. Siautama, A. C. I. A, dan D. Suhartono, "Extractive Hotel Review Summarization based on TF / IDF and Adjective-Noun Pairing by Considering Annual Sentiment Trends," *Procedia Comput. Sci.*, vol. 179, no. 2020, hal. 558–565, 2021. https://doi.org/10.1016/j.procs.2021.01.040

[30] E. Gul, N. Alpaslan, dan M. E. Emiroglu, "Robust optimization of SVM hyper-parameters for spillway type selection," *Ain Shams Eng. J.*, vol. 12, no. 3, hal. 2413–2423, 2021. https://doi.org/10.1016/j.asej.2020.10.022

[31] S. M. Malakouti, "Case Studies in Chemical and Environmental Engineering Improving the prediction of wind speed and power production of SCADA system with ensemble method and 10-fold cross-validation," *Case Stud. Chem. Environ. Eng.*, vol. 8, no. April, hal. 100351, 2023. https://doi.org/10.1016/j.cscee.2023.100351

[32] K. F. Irnanda, A. P. Windarto, dan I. S. Damanik, "Optimasi Particle Swarm Optimization Pada Peningkatan Prediksi dengan Metode Backpropagation Menggunakan Software RapidMiner," *J. Ris. Komput.*, vol. 9, no. 1, hal. 122–130, 2022. http://dx.doi.org/10.30865/jurikom.v9i1.3836

[33] N. Hussain, H. T. Mirza, F. Iqbal, dan I. Memon, "Spam Review Detection Using the Linguistic and Spammer Behavioral Methods," *IEEE Access*, vol. 8, hal. 53801–53816, 2020. https://doi.org/10.1109/ACCESS.2020.2979226

[34] V. Ahuja dan M. Shakeel, "Twitter Presence of Jet Airways-Deriving Customer Insights Using Netnography and Wordclouds," *Procedia Comput. Sci.*, vol. 122, hal. 17–24, 2017. https://doi.org/10.1016/j.procs.2017.11.336

[35] H. Cam, A. Veli, U. Demirel, dan S. Ahmed, "Heliyon Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers," *Heliyon*, vol. 10, no. 1, hal. e23784, 2024. https://doi.org/10.1016/j.heliyon.2023.e23784

[36] M. Isnan, G. Natanael, dan B. Pardamean, "Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model," *Procedia Comput. Sci.*, vol. 227, hal. 168–175, 2023. https://doi.org/10.1016/j.procs.2023.10.514

[37] D. Wang dan Y. Zhao, "Using News to Predict on Investor Sentiment : on SVM and Knowledge Internet of Using News to Predict Investor Sentiment : Based on SVM Model," *Procedia Comput. Sci.*, vol. 174, no. 2019, hal. 191–199, 2020. https://doi.org/10.1016/j.procs.2020.06.074