JHSS (Journal of Humanities and Social Studies)
https://journal.unpak.ac.id/index.php/jhss

Volume 08, Number 02, Page 556-562
e-ISSN: 2598-120X; p-ISSN: 2598-117X

SENTIMENT ANALYSIS ON THE PRESIDENTIAL THRESHOLD POLICY IN ELECTIONS AS A PRINCIPLE OF DEMOCRACY

Jovianus Abel Andreas [a*)], Erick Dazki [a)]

[a)] Pradita University, Tangerang, Indonesia

[*)]Corresponding Author: 1.jovianus.abel@student.pradita.ac.id

Article history: received 17 April 2024; revised 21 May 2024; accepted 20 July 2024

DOI: https://doi.org/10.33751/jhss.v8i2.9444

Abstract. Indonesia is a country that adheres to a presidential system of government. Elections are the basis of a democratic country in its implementation. Democracy is the government of reason by the people, for the people, and of the people, hence the people have the highest position in a democracy. The existence of a threshold is controversial in political dynamics because it is considered as a suppression of democratic values. Sentiment analysis is used to evaluate public assumptions about the application of the threshold by filling out a questionnaire that will be filled by subjects who have attitudes towards politics. Naïve Bayes and SVM are the methods used in solving sentiment analysis classification. Data collected through Twitter crawling is integrated with 1500 data from public assumptions about the presidential threshold. Naïve Bayes and SVM methods will be used to classify comment data. Through testing and classification, 784 comments were obtained which will be used as training data. The accuracy obtained from processing is 75.13% for Naïve Bayes and 83.29% for SVM.

Keywords: Presidential Threshold, Naïve Bayes, Laplacian, Sentiment Analysis, SVM

I. INTRODUCTION

Pemilihan umum merupakan faktor utama yang menjadi The embodiment of democratic life in the governance system of a country[1] is realized through the highest democratic sovereignty vested in the hands of the people as constituents. Article 1, paragraph (2) of the 1945 Constitution of the Republic of Indonesia emphasizes that sovereignty resides in the hands of the people and is exercised according to the Constitution[2]. The transition to a more democratic era occurred after the fall of the New Order regime in the Soeharto era in 1998. Changes in the constitutional amendments of 1945 resulted in a presidential system. The President and Vice President are directly elected by the people, with the role of the People's Consultative Assembly (MPR) as the highest state institution being subordinate. The President and Vice President have a maximum limit of two terms, each lasting five years.

Democracy itself is a value resulting from the struggle that emphasizes freedom, equality, and brotherhood, and is the outcome of consensus among nations and peoples that are part of conflicts or interactions of interests. Democracy is not just about good governance but also about the most practical processes to achieve it. The democratic process poses a challenge for Indonesia due to the high social disparities across various regions. In practice, there is money politics at the grassroots level, and with the existence of a presidential threshold, this practical political practice becomes even more widespread. The establishment of the threshold leads to debates triggering discourses among political elites and experts regarding the effectiveness of strengthening the presidential system in Indonesia[3]. The implementation of the Presidential Threshold is stipulated in Chapter VI, Article 222 of Law Number 7 of 2017 concerning the procedural threshold for presidential candidates, which states: "Presidential and Vice Presidential Candidates nominated by a political party or a coalition of political parties participating in the election that obtains at least 20% of the total seats in the DPR or receives 25% of the valid national votes in the previous DPR election"[4].

The Presidential Threshold is a debated topic every political year before the elections. The election threshold in Indonesia is unique as it is essentially applied as a limit for presidential candidacy, not as a limit for presidential eligibility. The elected president must receive more than 50%+1 of the votes, spread across at least 20% of 50%+1, with no limit on the number of election participants and parties/party coalitions. Substantially, the presidential threshold restricts the number of presidential candidates, as seen in the 2014 and 2019 elections, forcing voters to choose among candidates pre-selected by political figures through closed-room discussions, resulting in a hegemony of power.

Volume 08, Number 02, Page 556-562
e-ISSN: 2598-120X; p-ISSN: 2598-117X

Several previous studies related to sentiment analysis have been conducted by researchers:

Iryanto Saputra (2019) conducted sentiment analysis of user complaints about Indosat SIM cards using the K-Means method, clustering data obtained from crawling results. The analysis focused on clustering words without presenting the overall results of the statements. Unlike this study, sentiment analysis was performed comprehensively using Naïve Bayes classification, providing broader results.

Erwin Yudi Hidayat (2021) used Deep Neural Network (DNN) methodology for sentiment analysis of opinions on public companies, achieving an accuracy of 88.72%. The study compared DNN with Naïve Bayes, showing that Naïve Bayes had slightly easier implications in terms of results and effectiveness.

M. Khairul Anam (2022) presented research on sentiment analysis using Support Vector Machine (SVM) and Social Network Analysis (SNA) methodologies on public opinion about BPJS public policy. The study combined SVM with SNA to enhance accuracy, demonstrating that SVM alone had a relatively low efficiency compared to other methods.

Erwin Yudi Hidayat (2022) explored sentiment analysis using 1D-Convolutional Neural Network (1D-CNN) on female daily review websites, emphasizing the complexity of CNN compared to Naïve Bayes in classifying sentiments.

Habib Hakim Sinaga (2022) defined sentiment analysis results by comparing Decision Tree and XGBoost methodologies, using Random Forest with 75.96% accuracy for sentiment analysis of public opinions on Covid-19 through Twitter data.

Social media serves as a massive platform for interaction and communication, bringing new conflicts and challenges in communicating and conversing with others. Twitter was chosen for this study due to its inclusivity and efficiency in discussing policy issues on a large scale compared to direct interviews[10].

In assessing subjective or opinion-based issues related to the presidential threshold policy in the 2024 election, two methods were used for a higher percentage and accuracy in sentiment analysis. Procedurally, sentiment analysis was measured using the Naïve Bayes and Support Vector Machine (SVM) methodologies. Naïve Bayes assisted in grouping data based on label similarities, and Laplacian was used to avoid zero results in Naïve Bayes classification. Support Vector Machine, a machine learning method, was employed to recognize patterns in data and predict the correct categories in text for analysis[11]..

II. RESEARCH METHODS

The stages occurring in the flow diagram of Flow 1 will serve as the reference for the sentiment data classification process using the Naïve Bayes methodology. Each stage involves sub-processes and explanations defined in Table 1. The research implementation comprises six stages. In the first stage, data will be collected from the use of Google Colab, which employs the Python language. In the second stage, each crawled data will be modified to clean the text from various types other than sentiment text, such as links or unnecessary symbols. In the third stage, the data will be manually labeled with sentiment labels, serving as training data with positive and negative labels for future use. The fourth stage involves cleaning the documents with several sub-processes, including tokenization, case transformation, stopword filtering, and token filtering. In the fifth stage, the data will be applied with sentiment methods, namely Naïve Bayes and SVM. After applying the methods, in the sixth stage, the sentiment results will be analyzed using K-Fold Cross Validation, where the training data will be evaluated into multiple parts (folds) [11].

Rearch Steps	Detail
Data Crawling	*Data Crawling Crawling is performed using the Python language with Google Colab software*
Data Cleansing	Data Cleansing After crawling, the obtained data is cleaned from various irrelevant elements in sentiment data, and duplicate data is removed
Data Labeling	Data Labeling Following the processing and stages mentioned above, sentiment text is labeled as positive and negative.
Proccess Document (TF-IDF)	Process Document (TF-IDF) Comprising sub-processes with four stages: tokenization, breaking the text into more subtle fragments; case transformation, which involves converting uppercase text to lowercase; stopword filtering, removing common and meaningless words such as "and," "or," and "that," or other irrelevant text; token filtering, removing excessively long or short text.
Sentiment Analysis	Sentiment Analysis This stage of sentiment analysis adopts two main methods as the foundational basis and primary parameters: Naïve Bayes and Support Vector Machine (SVM). Sentiment text is polarized into positive and negative
Performance Evaluation	Performance Evaluation The use of the K-Fold Cross Validation method involves training the model using one fold and testing it with another, repeating the process several times by varying the folds used for testing and training. The testing results will be combined to produce a more accurate model performance estimation

Naive Bayes Classification

The primary methodology employed for sentiment classification in the study revolves around the application of the threshold for presidential nomination. The Naïve Bayes classification, conceptualized by Thomas Bayes (1701-1706) and independently discovered by Pierre-Simon Laplace (1749-1827), is a meticulous method for interpreting evidence in the context of experiences or knowledge from the past[12]. It is a classification method based on Bayes' theorem,

Volume 08, Number 02, Page 556-562
e-ISSN: 2598-120X; p-ISSN: 2598-117X

predicting the class of a given data based on its features[13][14]. Naive Bayes operates by calculating the probability of each feature and then multiplying these probabilities to obtain the probability of the desired class[15],[16]. The formula for the Naïve Bayes theorem is as follows::

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \tag{1}$$

In the above formula, it can be defined that P(H|X) is the posterior probability of hypothesis H based on data X, P(X|H) is the likelihood probability of data X based on hypothesis H, P(H) is the prior probability of hypothesis H, and P(X) is the evidence probability of data X[16][17][18]..

2.3 Laplacian Smoothing

The use of the Laplacian method involves adding a value of 1 to each data set present in the training data. This is done to avoid zero probabilities in the training data. The variable ν represents the number of pseudo-counts added to each possible value of the random variable to avoid zero probability values..

$$P(H|X) = \frac{P(X|H).P(H)+1}{P(X)+v} \tag{2}$$

2.4 Support Vector Machine (SVM)

It is one of the methods and machine learning techniques for identifying patterns in data and predicting precise categories for each text. Conceptually, SVM divides the data into a higher-dimensional space and seeks the most accurate hyperplane to separate different polarized categories. SVM analyzes the hyperplane that best divides the data with the largest distance between two categories. It then predicts the precise category for each text based on its position relative to the hyperplane[19]..

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b\right) \tag{3}$$

2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

A crucial component serving as a technique to evaluate the importance of words in the structure of a document is TF-IDF. TF-IDF calculates points for each word in a document, indicating how substantially important the word is overall. This is achieved by multiplying the frequency of each occurrence of the data in the document (TF), on the other hand, represents the frequency of the word in the set of documents[20]. The higher the TF-IDF score of a word, the more important its role in the document overall. The TF-IDF formula is as follows:

$$tf - idf(t, d) = tf(t, d) * idf(t) \tag{4}$$

Where tf(t, d) represents the frequency of the occurrence of the term t in document d, and idf(t, d) is the inverse of the frequency of the term t across all documents in the document set. The formula is as follows:.

$$idf(t) = log\left(\frac{N}{df(t)}\right) \tag{5}$$

Where tf(t, d) is the frequency of the term t in document d, and idf(t, d) is the inverse of the frequency of the term t across all documents in the document set.

III. RESULTS AND DISCUSSION

Data Crawling

Data acquisition was performed through Google Colab tools using the Python language via the Twitter social media platform. The search was conducted with the keywords "presidential threshold lang:id," where lang:id helps filter the Indonesian language, and non-Indonesian languages will not be crawled in the resulting output. A total of 1500 data points were obtained. Duplicate and irrelevant data were filtered using the replace and remove duplicates operators. Below, in Table 2, examples of the crawled data are presented, illustrating the process for sentiment analysis.

Creation Date	Text
Wed Nov 22 11:55:23	I actually agree with changing the age requirements for presidential and vice-presidential candidates as long as the root of the problem is addressed first, namely the 20% Presidential Threshold (PT) that limits choices and prevents the people from fully exercising their political rights. It should be abolished.
Thu Oct 26 09:08:41	RT @Fahrihamzah Just want to say #thickface talking about the presidential threshold, it's you who are in the parliament and approved it. People in Papua are starving, and dozens are dying. #rejectignorance
Tue Oct 17 04:38:48	The public urges a presidential threshold of zero (0) percent so that democracy benefits the people, and oligarchy cannot control the government. https://t.co/JOaJzkdZDO @RamliRizal https://t.co/8CraOdxYRl
Mon Nov 13 13:14:48	@primawansatrio @ainunnajib We need to consider the constitutional mechanisms in place, that the presidential threshold provision is constitutionally valid and cannot be challenged, as it has undergone judicial review multiple times.

In this stage, data that is not significant and unrelated to the sentiment analysis process will be nullified. The obtained data after the cleansing process amounts to a total of 784 records out of the initial 1500 crawled data from Twitter. The sequence of operators used to process the data is as follows:

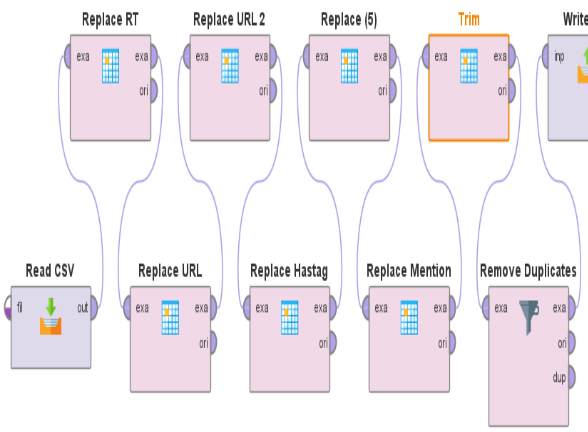

Figure 1. : Data Cleaning Process

The CSV file is read, where the input file has been crawled. The replace operator performs substitutions as instructed, the trim operator removes leading and trailing white spaces from numerical values, and the remove duplicate operator ensures there is no duplicate data. Subsequently, the results are written back to a CSV file through the Write CSV operator. The following represents the outcome after the data cleansing process.

Text
I genuinely support the changes in the age requirements for presidential and vice-presidential candidates, as long as the root issue is addressed first, namely the Presidential Threshold set by PT. This threshold restricts choices, limiting the people's ability to fully exercise their political rights. Its elimination is essential.
I'd like to point out that discussions about the presidential threshold involve those present in the parliament who also approve it. Meanwhile, people in Papua are suffering from hunger and numerous deaths. The public insists on reducing the presidential threshold to 0 percent to ensure that democracy serves the people, preventing oligarchy from dominating the government.
We need to consider the existing constitutional mechanisms, noting that the presidential threshold is a legitimate constitutional provision that should not be challenged. It has undergone repeated judicial reviews.
We need to consider the existing constitutional mechanisms, noting that the presidential threshold is a legitimate constitutional provision that should not be challenged. It has undergone repeated judicial reviews
If the presidential threshold were challenged, many individuals would likely register as presidential candidates, potentially turning the situation into a double-edged sword. Those with non-political-related popularity and knee-jerk reactions might easily be elected, considering the tendency of Indonesian citizens to be drawn to viral trends.

Data Labelling

The results of the data cleansing will then be labeled as polarized sentiments into positive and negative. The positive benchmark is if the text sentiment agrees with the presidential threshold policy and vice versa is negative if the text sentiment disagrees. Labeling is carried out manually, with the division of a total of 784 data obtained, 500 data will be taken as training data and 284 data will be taken as test data. Here's an example of text that has been labeled sentiment

Text	Sentiment
"I actually agree with the change in the requirements for the age of presidential and vice-presidential candidates, as long as the root of the problem is addressed first, namely the Presidential Threshold PT, which limits choices and prevents the people from fully exercising their political rights. It should be eliminated."	Negative
Just want to say, discussing the presidential threshold, you, who are in the parliament, approve while people in Papua are starving and dozens of them are dying. Shame on you	Negative
The public urges a presidential threshold of zero percent so that democracy benefits the people and oligarchs cannot control the government.	Negative
We need to look at the existing constitutional mechanisms, that the provision of the presidential threshold is a constitutionally valid provision that should not be contested because it has been subjected to judicial review several times	Positive
If the presidential threshold is challenged, there will likely be many who register as presidential candidates. It has the potential to backfire, as those with popular but non-political-related knee-jerk popularity will easily be elected. You know the quality of Indonesian citizens, they love things that go viral	Positive

Process Document (TF-IDF)

Volume 08, Number 02, Page 556-562
e-ISSN: 2598-120X; p-ISSN: 2598-117X

The result of the next data cleansing step will be labeling sentiment as polarized into positive and negative. The positivity benchmark is if the sentiment text agrees with the presidential threshold policy, and conversely, negative if the sentiment text disagrees. Labeling is done manually, with a distribution of the total 784 obtained data; 500 data will be taken as training data, and 284 data will be taken as test data. The following is an example of text that has been labeled with sentiment..

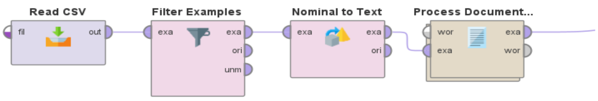

Figure 2: Pre-processing Steps before TF-IDF

In Figure 2, several stages can be seen. The first is the CSV read operator obtained from the previous process in Figure 1, namely, write CSV; the second is the filter examples, which is the retrieval of labeled sentiment data; the third is nominal to text to transform parameters that are still in polynomial form into text so that they can be processed in the TF-IDF stage, then the fourth is document processing which has 4 sub-processes in it. In Figure 3, it can be seen that there are 4 operator stages of sub-processes in the document process, namely:

Tokenize: Fragmenting text, audience words, or phrases. Tokenization is done using spaces, punctuation, or other special characters as separators. Example: "Just want to talk about the presidential threshold, you in Senayan and agree with the people in Papua who are starving and dying, dozens of them."

Transform Cases: Transforming all characters to lowercase or uppercase, which in the contextualization of the process helps text analysis. Example: "just want to say talking about the presidential threshold, you in Senayan and agree with the people in Papua starving and dying, dozens of them."

Filter Stopwords: A process of removing common and meaningless words contained in text analysis. Example: "Just want to say talking about the presidential threshold, you in Senayan and agree with the people in Papua starving and dying, dozens of them."

Filter Tokens: A process of removing irrelevant or unwanted tokens from the text. Includes numbers, special characters, or irrelevant words.

The process carried out at this stage is to process sentiment data totaling 500 training data and 284 test data. Some stages that have been carried out before are to ensure that the classification process can provide accurate and precise calculations for this model with the naive Bayes and SVM methods. The following are the operators used in the sentiment classification process

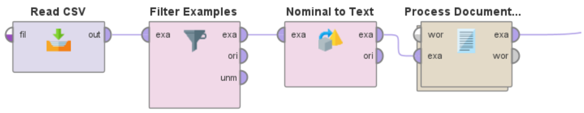

Figure 1 Sub-process Stages in a Process Document

Sentiment Classification

The process carried out at this stage is to process sentiment data which amounts to 500 training data and 284 test data. Some of the steps that have been carried out previously are so that the classification process can provide accuracy and precise calculations for this model using the naïve Bayes and SVM methods. Here are the operators used in the sentiment classification process.

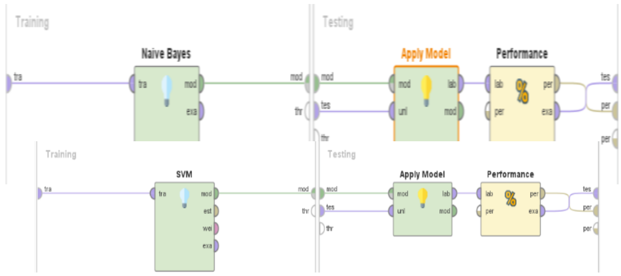

Image 3: Sub-process Stages within the Document Processing

Performance Evaluation

Statistics generated from the machine learning model in Figure 4 and Figure 5 require the use of the K-Fold Cross-Validation method. The parameter used, where the value of K is the number of data segments separated into training and testing, is 10-fold cross-validation in this case. Sub-processes that can be seen in Figure 4 and Figure 5. On the left, which is training, the operator methods used to create an algorithm for classification can be seen. On the right, which is part of the testing of unlabeled data from the training section to test the model's performance. Apply the model used to test unlabeled sentiment data. Afterward, data from processing will enter the performance evaluation stage to provide results from the tested data. The following are the results from the confusion matrix table of the processing using the K-Fold Cross-Validation method

	True Positive	True Negative	Class Precision
Pred. Positive	**262(TP)**	65 (FP)	80.12%
Pred. Negative	130 (FN)	**327 (TN)**	71.55%
Class Recall	66.84%	83.42%	

Results of calculations using Cross-Validation operator in the form of Confusion Matrix using Naïve Bayes calculations

	True Positive	True Negative	Class Precision
Pred. Positive	**318 (TP)**	60 (FP)	83.72%
Pred. Negative	74 (FN)	**332 (TN)**	82.87%
Class Recall	82.65%	83.93%	

Volume 08, Number 02, Page 556-562
e-ISSN: 2598-120X; p-ISSN: 2598-117X

shows the results of the calculation using the Cross-Validation operator in the form of a Confusion Matrix using SVM calculation.

In Tables 5 and 6, True Positive (TP) is the number of positive data correctly predicted by the model. False Positive (FP) is the number of negative data incorrectly predicted by the model. False Negative (FN) is the number of positive data incorrectly predicted by the model. True Negative (TN) is the number of negative data correctly predicted by the model. The results of confusion matrix Table 5 show that the number of True Positive is 262, False Positive is 65, True Negative is 327, and False Negative is 130. In Table 6, it is shown that the number of True Positive is 318, False Positive is 60, True Negative is 332, and False Negative is 74. Class Precision measures how accurately the model calculates the positive class, while Class Recall identifies all positive classes in the data set. The following are the formula calculations for Class Precision and Class Recall..

$$\textit{Class Precision} = \frac{TP}{TP+FP} \tag{6}$$

$$\textit{Class Recall} = \frac{TP}{FP+FN} \tag{7}$$

The calculations performed by the operator employ K-Fold Cross Validation, entailing training and testing the model k times. In order to identify the optimal accuracy yielded by this method, data will be tested using K-2 through K-10, and the results will be presented in the table below.

K-Fold	SVM	Naïve Bayes
K-2	79.46%	**75.13%**
K-3	79.08%	74.36%
K-4	81.63%	74.62%
K-5	81.38%	73.85%
K-6	**83.29%**	72.96%
K-7	82.53%	74.23%
K-8	82.14%	72.96%
K-9	82.91%	73.85%
K-10	82.28%	72.72%

Comparison of results from K-Fold calculations with two different methods

In Table, the results of K-Fold for the Naïve Bayes and SVM methods are presented. The highest accuracy among these methods is observed at K-6, reaching 83.29% using the SVM method, and at K-2, achieving 75.13% using the Naïve Bayes method. Tables 5 and 6 illustrate the representations of each method with the highest accuracy. It can be concluded that the method demonstrating the overall best accuracy in sentiment classification is the Support Vector Machine (SVM).

II. CONCLUSIONS

This study adapts sentiment analysis regarding the presidential threshold policy issue in elections as a fundamental aspect of polarized democracy, manifesting sentiments as positive and negative. The testing results using Naïve Bayes and SVM methods show an accuracy of 83.29% with the SVM method and 75.13% with the Naïve Bayes method. Notably, SVM attains the highest accuracy, even with the limitations of available data. Sentiment analysis on the presidential threshold contributes to understanding the polarization dynamics, especially concerning this policy. This method is crucial for analyzing sentiments. However, limitations exist in crawling data using Python language on Twitter, as Twitter's policy restricts the maximum data extraction. Obtaining more data requires a higher-priced subscription The researcher hopes that the analysis of the presidential threshold policy can be further developed in the future, especially aiming for higher accuracy. This involves seeking higher-quality data to vividly portray the genuine polarization regarding this policy. With the ongoing evolution of policies, particularly entering the political year, this policy holds substantial significance in Indonesian political dynamics and will persist until the majority's desired outcome is achieved: a 0% presidential threshold without any imposed limits.

REFERENCES

[1] A. M. Al Mas'udah, "The Presidential Threshold As An Open Legal Policy In General Elections In Indonesia," *Prophetic Law Review*, Vol. 2, No. 1, Jun. 2020, Doi: 10.20885/Plr.Vol2.Iss1.Art3.

[2] A. Fitri And W. Setiadi, "Presidential Threshold Dalam Pemilihan Umum Serentak: Kemunduran Demokrasi Konstitusional?," 2022.

[3] S. D. Ambarwati, M. R. Saifulloh, And S. M. S. Aritonang, "Rekonstruksi Sistem Presidential Threshold Dalam Sistem Pemilu Di Indonesia (Studi Perbandingan Sistem Presidential Threshold Indonesia Dan Brazil)," Bulan Kedelapan, 2020. [Online]. Available: Https://Jhlg.Rewangrencang.Com/

[4] V. Anggara, "Dinamika Presidential Threshold Dalam Sistem Presidensial Di Indonesia," 2019.

[5] E. Y. Hidayat, R. W. Hardiansyah, And A. Affandy, "Analisis Sentimen Twitter Untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network," *Jurnal Nasional Teknologi Dan Sistem Informasi*, Vol. 7, No. 2, Pp. 108–118, Sep. 2021, Doi: 10.25077/Teknosi.V7i2.2021.108-118.

[6] M. K. Anam, M. I. Mahendra, W. Agustin, R. Rahmaddeni, And N. Nurjayadi, "Framework For Analyzing Netizen Opinions On Bpjs Using Sentiment Analysis And Social Network Analysis (Sna)," *Intensif: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, Vol. 6, No. 1, Pp. 11–28, Feb. 2022, Doi: 10.29407/Intensif.V6i1.15870.

Volume 08, Number 02, Page 556-562
e-ISSN: 2598-120X; p-ISSN: 2598-117X

[7] E. Y. Hidayat And D. Handayani, "Penerapan 1d-Cnn Untuk Analisis Sentimen Ulasan Produk Kosmetik Berdasar Female Daily Review," *Jurnal Nasional Teknologi Dan Sistem Informasi*, Vol. 8, No. 3, Pp. 153–163, Jan. 2023, Doi: 10.25077/Teknosi.V8i3.2022.153-163.

[8] H. H. Sinaga And S. Agustian, "Pebandingan Metode Decision Tree Dan Xgboost Untuk Klasifikasi Sentimen Vaksin Covid-19 Di Twitter," *Jurnal Nasional Teknologi Dan Sistem Informasi*, Vol. 8, No. 3, Pp. 107–114, Dec. 2022, Doi: 10.25077/Teknosi.V8i3.2022.107-114.

[9] M. O. Odim, A. O. Ogunde, B. O. Oguntunde, And S. A. Phillips, "Exploring The Performance Characteristics Of The Naïve Bayes Classifier In The Sentiment Analysis Of An Airline's Social Media Data," *Advances In Science, Technology And Engineering Systems*, Vol. 5, No. 4, Pp. 266–272, Jul. 2020, Doi: 10.25046/Aj050433.

[10] W. Budiharto And M. Meiliana, "Prediction And Analysis Of Indonesia Presidential Election From Twitter Using Sentiment Analysis," *J Big Data*, Vol. 5, No. 1, Dec. 2018, Doi: 10.1186/S40537-018-0164-1.

[11] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, And J. G. Hsieh, "Twitter Sentiment Analysis Towards Covid-19 Vaccines In The Philippines Using Naïve Bayes," *Information (Switzerland)*, Vol. 12, No. 5, May 2021, Doi: 10.3390/Info12050204.

[12] A. A. Farisi, Y. Sibaroni, And S. Al Faraby, "Sentiment Analysis On Hotel Reviews Using Multinomial Naïve Bayes Classifier," In *Journal Of Physics: Conference Series*, Institute Of Physics Publishing, May 2019. Doi: 10.1088/1742-6596/1192/1/012024.

[13] M. Z. Nafan And A. E. Amalia, "Kecenderungan Tanggapan Masyarakat Terhadap Ekonomi Indonesia Berbasis Lexicon Based Sentiment Analysis," *Jurnal Media Informatika Budidarma*, Vol. 3, No. 4, P. 268, Oct. 2019, Doi: 10.30865/Mib.V3i4.1283n. Ramadhani And N. Fajarianto, "Sistem Informasi Evaluasi Perkuliahan Dengan Sentimen Analisis Menggunakan Naïve Bayes Dan Smoothing Laplace," *Jurnal Sistem Informasi Bisnis*, Vol. 10, No. 2, Pp. 228–234, Dec. 2020, Doi: 10.21456/Vol10iss2pp228-234.

[15] D. Berrar, "Bayes' Theorem And Naive Bayes Classifier," In *Encyclopedia Of Bioinformatics And Computational Biology: Abc Of Bioinformatics*, Vol. 1–3, Elsevier, 2018, Pp. 403–412. Doi: 10.1016/B978-0-12-809633-8.20473-1.

[16] B. Liu, "Sentiment Analysis And Opinion Mining," Morgan & Claypool Publishers, 2012.

[17] M. Nababan *Et Al.*, "The Diagnose Of Oil Palm Disease Using Naive Bayes Method Based On Expert System Technology," In *Journal Of Physics: Conference Series*, Institute Of Physics Publishing, Apr. 2018. Doi: 10.1088/1742-6596/1007/1/012015.

[18] S. Samsir, K. Kusmanto, A. H. Dalimunthe, R. Aditiya, And R. Watrianthos, "Implementation Naïve Bayes Classification For Sentiment Analysis On Internet Movie Database," *Building Of Informatics, Technology And Science (Bits)*, Vol. 4, No. 1, Pp. 1–6, Jun. 2022, Doi: 10.47065/Bits.V4i1.1468.

[19] Y. Surya, S. Al Faraby, And M. Dwifebri, "Analisis Sentimen Terhadap Ulasan Film Menggunakan Word2vec Dan Svm."

[20] T. I. Saputra And R. Arianty, "Implementasi Algoritma K-Means Clustering Pada Analisis Sentimen Keluhan Pengguna Indosat," *Jurnal Ilmiah Informatika Komputer*, Vol. 24, No. 3, Pp. 191–198, 2019, Doi: 10.35760/Ik.2019.V24i3.2361.