ICIC Express Letters ICIC International ⓒ2020 ISSN 1881-803X
Volume **14**, Number **5**, May **2020** pp. **1–8**

Portrait of Political Map in the 2024 Indonesian Presidential Election Based on Voter Distribution using a Sentiment Analysis Approach

Asno Azzawagama Firdaus [1], Anton Yudhana[2,*], Imam Riadi[3] and Mahsun[4]

[1]Master Program of Informatics
[2]Department of Electrical Engineering
[3]Department of Information System
Universitas Ahmad Dahlan
Jl. Ringroad Selatan, Yogyakarta 55166, Indonesia
2207048008@webmail.uad.ac.id; *Corresponding author: eyudhana@ee.uad.ac.id; imam.riadi@mti.uad.ac.id

[4]Department of Indonesian Language and Literature Education
Universitas Mataram
Jl. Majapahit No. 62, Mataram 83115, Indonesia
mahsun@unram.ac.id

Received October 2019; accepted January 2020

ABSTRACT. *Social media is a solution for politicians as a campaign tool because it can save costs compared to conventional campaigns. The 2024 Indonesian Presidential Election has invited a lot of public attention, especially from social media users. Twitter as one of the social media that is widely used by the people of Indonesia is a solution as the most effective campaign forum. Sentiment analysis is one approach that can be used to get results of public alignment with Indonesian presidential candidates based on Twitter data. The data used was 30,000 data and labeled using the SME (Subject Matter Expert) method. Determination of classification model on sentiment analysis using Naïve Bayes method and Support Vector Machine with TF-IDF feature extraction. Based on the three candidate name datasets studied, the highest accuracy was obtained on Prabowo Subianto's dataset with SVM - RBF and SVM Polynomial methods, which was 87%. While the lowest accuracy is owned in the Ganjar Pranowo's dataset with the Naive Bayes method, which is 70%. The study projects the outcome of the Presidential Election based on vote distribution and sentiment numbers.*
Keywords: Indonesian Presidential, Naïve Bayes, Sentiment Analysis, SVM, Twitter

1. **Introduction.** Social media is an important part of a politician as a place to attract participants. Social media has grown rapidly in recent years and has made it the widest disseminator of information in the world [1,2]. Media control is very important because it has the power to influence, regulate and limit content on social media. Social media has a huge impact on its users [3] thus offering an excellent opportunity for politicians to reach the public directly at minimal cost [4]. Twitter stands as one of the leading social media platforms, enabling users to compose brief messages with a maximum of 140 characters, alongside the capability to share photos and videos. This platform is highly favored for its capacity to convey concise content and multimedia content [5]. Twitter provides a place to express perspectives [6] and makes it the primary online platform for politicians to engage directly with the public through social and political commentary, no wonder it makes political accounts more active than non-political accounts [7]. Approximately 60% of individuals engage in conversations related to political matters on platforms such as Twitter. This indicates that social media has become a widely used platform where people rely on the most recent political updates [8].

Elections are an irreversible and legal process in a democracy where political parties elect the next leader for the political institution that will govern [9]. The April 17, 2019, parliamentary election was the fourth direct presidential election in Indonesia. This is also the first time Indonesia has held a general election at the same time, with 245,000 candidates vying for nearly 20,000 seats in provincial and local councils, councils, and councils across the country. About 190 million people have the right to vote at 800,000 polling stations in one day [10]. Indonesia will experience a change of leadership in 2024. The explanation was released on the official website of the Government through the Ministry of Communication and Information Technology of the Republic of Indonesia. President Joko Widodo explained that there would be no delay for the General Election on February 14, 2024, and the simultaneous Regional Head Election in November 2024. The change of leadership has made several pollsters in Indonesia participate in providing survey results and publications on the upcoming 2024 Presidential Election. The Presidential Election or abbreviated as Presidential Election was attended by several National Figures. The strongest names who have high electability as presidential candidates based on the results of research by pollsters Poltracking Indonesia are Ganjar Pranowo (Governor of Central Java), Prabowo Subianto (Minister of Defense), and Anies Baswedan (Governor of DKI Jakarta 2017-2022) [11].

Data on social media is impossible to select one by one because the number continues to grow. A technique is needed to be able to classify them, an approach that can be used to overcome this is sentiment analysis. Sentiment analysis is a scientific discipline that analyzes and evaluates statements, attitudes, and feelings [12] and allows entrepreneurs to gather information about customer opinions through various online mediums such as surveys, social media, and website reviews [13]. Class determination in sentiment analysis required classification methods as machine learning models. The Naïve Bayes method and Support Vector Machine (SVM) were used as the best analysis for this case study. Naïve Bayes proved to be the most well-known method for sentiment analysis problems [14]. Naïve Bayes are also widely used in many applications [13-16] due to their simplicity, efficiency, and tractability [17].

Overall, this study classifies public opinion toward certain figures related to the upcoming Indonesian elections in 2024. The objectives to be identified in this study:

- The influence of sentiment results on social media on the potential victory in the Presidential Election.
- Political map adjustment in the Presidential Election based on the distribution of voters in each region.
- The best classification method between Naive Bayes and Support Vector Machine to apply to political content.

2. **Methods.** Figure 1 shows the research framework from data collection to sentiment results.

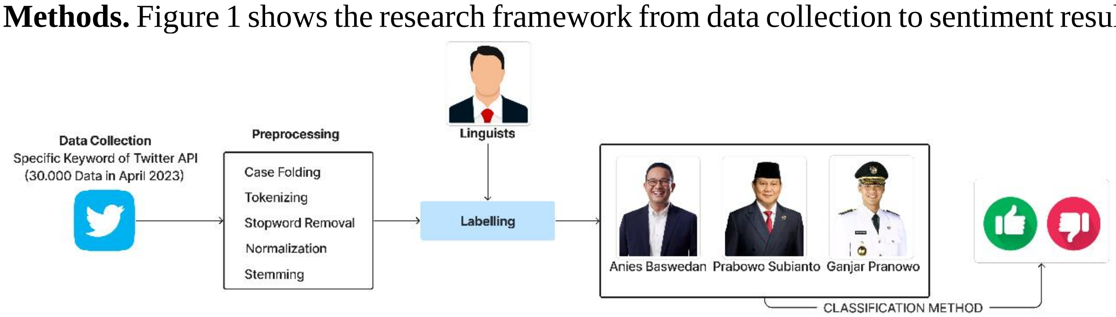

FIGURE 1. Framework for Twitter based sentiment analysis for Indonesia 2024 election

2.1. **Data Collection.** Political Parties and combinations of Political Parties have announced candidates to be carried as candidates for the President of Indonesia in 2024 long before the date of determination of the Presidential and Vice President candidates. This study conducted data collection in a certain period before the official determination of presidential candidates by the KPU-RI. This is to find out the analysis of public opinion about sentiment and the popularity of candidates carried by political parties.

Data collection is done using SNScrape in Python programming. Users are required to have a developer account to retrieve data [18]. The data collected is 30,000 tweets in April 2023. The data collected consists of several Twitter attributes as shown in Table 1.

TABLE 1. Feature Item Description Twitter

Item	Description
Tweet Date	Date the tweet was posted on Twitter
Created Account	The date the user joined Twitter
Username and User ID	Username naming on Twitter and ID
Following and Followers	Number of Accounts Followed and Who Followed
Tweet Count	Number of posts on Twitter
Tweet Location	Name of the location where the tweet was posted
Tweet	The Twitter post

The data used was based on three candidates (Ganjar Pranowo, Prabowo Subianto, and Anies Baswedan).

2.2. **Labelling.** Transformer libraries are used because their architecture achieves prediction-like accuracy with human-performed annotations for text classification at sentimental polarity [19]. In addition to transformers, labeling can also use manual methods [20]. However, the results of the transformer show an accuracy rate of 50.2% when compared to manual labeling. This study compares it to manual labeling because manual labeling produces accurate data [21]. This labeling technique often misinterprets tweets that should have a positive sentiment label but are given a negative label due to several factors, including:
- Words have more negative connotations than positive in one Tweet.
- Contains inductive paragraphs, namely at the beginning of sentences containing negative charges (more words count) and at the end (conclusion) containing positive charges.
- Using religious terms and regional languages.
- Tweets that use the figure of speech/satire (parable).

Labeling Tweets that should have a negative class but be labeled positive are fewer than misinterpretations of previous negative labels. This is due to several factors, such as:
- Tweets containing inductive paragraphs. The beginning of the sentence contains a positive charge (more words) and the end (conclusion) contains a negative charge.
- Tweets that use the figure of speech/satire (parable)
- Tweets that give positive sentiment towards other candidates, but negative towards the candidate in question.

TABLE 2. Comparison of errors on labels using transformers

Candidates	Case	Error Percentage (%)

Ganjar Pranowo	Positive tweets rated negatively	79,22
	Negative tweets rated positively	20,78
Prabowo Subianto	Positive tweets rated negatively	93,66
	Negative tweets rated positively	6,34
Anies Baswedan	Positive tweets rated negatively	96,08
	Negative tweets rated positively	3,92

Table 2 shows the magnitude of errors in labeling techniques using transformers when compared to manual labeling. Based on this, this study proposes the SME (Subject Matter Expert) method as a sentiment labeling technique by research institutions that are experts in their fields, namely the Nusantara Research Institute of the Center for Human Development Studies Foundation. SME eliminates and combines prior knowledge to incorporate controllable features by building better and more practical models [22]. This research uses these insights to develop a new framework to reach a consensus of experts [23].

2.3. **Data Preprocessing.** An important part is to clean raw data [24] and reconstruct text into a more processable form for machine learning algorithms [25]. This research uses stages such as case folding, tokenizing, stop word removal, normalization, and stemming on preprocessing data.

2.4. **Term Frequency – Inverse Document Frequency (TF-IDF).** TF-IDF is a weighting method commonly used in data mining to evaluate the effect of words in accounting for documents and the global corpus [26] with the following formula [27].

$$\mathrm{tf}_i = \frac{n_i}{\sum_k n_k} \quad (1)$$

$$\mathrm{idf}_i = log\frac{|D|}{|\{d_j : t_i \in d_j\}|} \quad (2)$$

$$\mathrm{tfidf}_i = \mathrm{tf}_i \mathrm{idf}_i \quad (3)$$

TF indicates the number of occurrences in the corpus (Equation 1). The IDF is a measure of the importance of an entire corpus term. It consists of the calculation of the logarithm of the inverse relationship of corpus documents (Equation 2). The weight of the TF-IDF is calculated by multiplying the two (Equation 3). The greater the weight indicates the more important words are relevant to the corpus [28].

2.5. **Classification Method.** The Naive Bayes technique, introduced by Thomas Bayes, acts as a point of reference for making predictions about future probabilities using past experiences [51,52]. Theoretically, among various classification algorithms, the Naive Bayes model is believed to exhibit the least error rate [31]. Developed by Thomas Bayes, this technique acts as a guide for forecasting potential outcomes using past experiences [51,52]. In addition to Naive Bayes, another method used is SVM which is an effective algorithm in advanced machine learning [32]. SVM's goal for plotting a hyperplane is ideal and is called a decision limit using distances between nearby samples [33]. In addition, SVM is highly dependent and easier to implement using TF-IDF to calculate weights on documents [34]. The SVM method has a higher classification quality when using nonlinear SVM. However, computational complexity is a barrier to large-scale data sets [35]. The SVM's kernel function involves a linear division within the feature space, particularly for extensive sets of undistinguished data. This aspect has an indirect impact on the efficacy of SVM classification [36]. SVM encompasses four distinct kernels: Linear, Polynomial, Gaussian/Radial Basis Function (RBF), and Sigmoid [37].

2.6. **Model Validation.** Measuring models in machine learning is critical [38]. The classification method used. Precision, accuracy, recall, and F1 scores prove useful for confusion matrices [33].

$$Accuracy = TP + \left(\frac{TN}{\mathrm{TP}}\right) FP + FN + TN \tag{5}$$

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

$$Recall = \frac{TP}{\mathrm{TP}} + FN \tag{7}$$

$$F1 = 2\ x\ precision\ x \frac{recall}{\mathrm{precision}} + recall \tag{8}$$

TP shows a positive result, FP shows a false positive result, TN shows a negative result and FN shows a false negative result.

3. **Result and Discussion.** This study provides projections for the results of the 2024 Presidential Election based on data obtained by giving time keywords from April 2023. This study estimates that the results of the presidential election based on Twitter social media data are determined based on the distribution of voters and public sentiment towards candidates carried by political parties. Another factor is the popularity caused by the increasing number of mentions of the word in Tweets. Figure 2 shows common words that appear frequently.

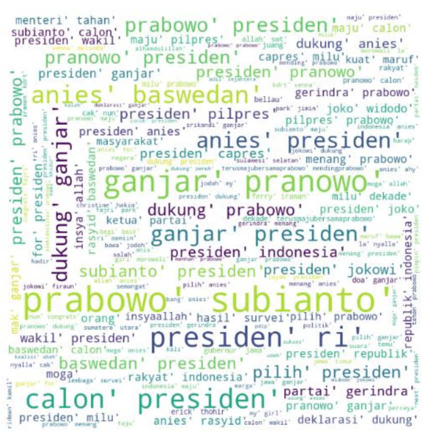

FIGURE 2. Popularity data using word cloud

3.1. **Voter Distribution Prediction.** Analysis of geographical location or distribution of votes in certain regions for each candidate is obtained through positive sentiment owned and spread in each region. The distribution of votes in each of these regions is obtained based on the location field in the Tweet data. Location availability is not owned by every data on the Tweet (NaN) so it cannot be defined as an input on the distribution of voices carried out in this study.

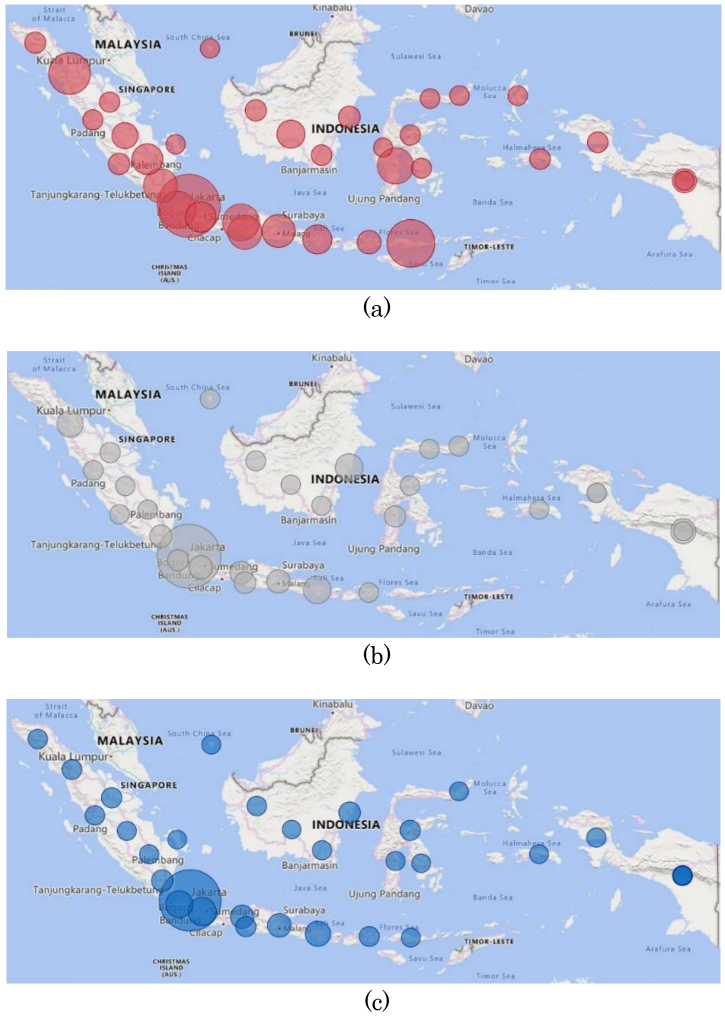

FIGURE 3. Visualization the Political Map of Presidential Electoral Ganjar Pranowo (a), Prabowo Subianto (b) and Anies Baswedan (c)

Figure 3 shows each candidate's voter political map using Power BI. Voter potential is shown like a bubble in the picture for each provincial and/or city area. The choice of bubble color used shows the distinctive colors of the supporting political parties of each candidate. Voters are scattered in each region which is shown the bigger the bubble, the greater the potential voters in that region.

3.2. **Number of Sentiments.** The results of labeling on the dataset of the three candidates are shown in Figure 4.

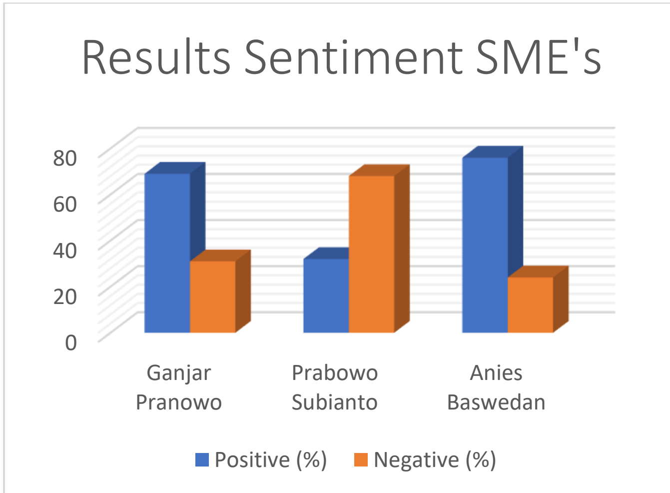

FIGURE 4. Results Sentiment Based on SME's

The testing process on classification methods using the Naïve Bayes and SVM methods is shown in Table 3.

TABLE 3. Experimental Results

Candidate	Method	Data Ratio	Precision		Recall		F1-Score		Accuracy (%)
			Pos	Neg	Pos	Neg	Pos	Neg	
Prabowo Subianto	SVM – RBF	90:10	0,84	0,89	0,83	0,90	0,84	0,89	87
Prabowo Subianto	SVM – Polynomial	90:10	0,87	0,87	0,79	0,92	0,83	0,90	87
Prabowo Subianto	SVM – RBF	70:30	0,86	0,86	0,78	0,91	0,82	0,89	86
…	…	…	…	…	…	…	…	…	…
Ganjar Pranowo	Naïve Bayes	70:30	0,67	0,79	0,91	0,44	0,77	0,57	70
Ganjar Pranowo	Naïve Bayes	80:20	0,68	0,82	0,92	0,46	0,78	0,59	72
Ganjar Pranowo	Naïve Bayes	90:10	0,69	0,75	0,89	0,45	0,77	0,56	70

Table 3 shows the three test models with the highest and lowest accuracy. The highest accuracy was obtained on Prabowo Subianto's dataset with SVM - RBF and SVM Polynomial methods. While the lowest accuracy is owned by the Ganjar Pranowo dataset with the Naive Bayes method.

Among the two methods used, SVM using all four kernels has an average accuracy above that of the Naïve Bayes method as shown in Figure 5. However, the SVM method has a longer computational speed, in contrast to Naïve Bayes which can do it quickly [39]. The accuracy of the Naïve Bayes method is lower than SVM because it requires only small data on training [29]. While the SVM method has good performance when used on large data [40]. The results of high accuracy in the SVM method show that this method is not overfit [41] so it is more suitable for use in big data and sentiment in political discussions.

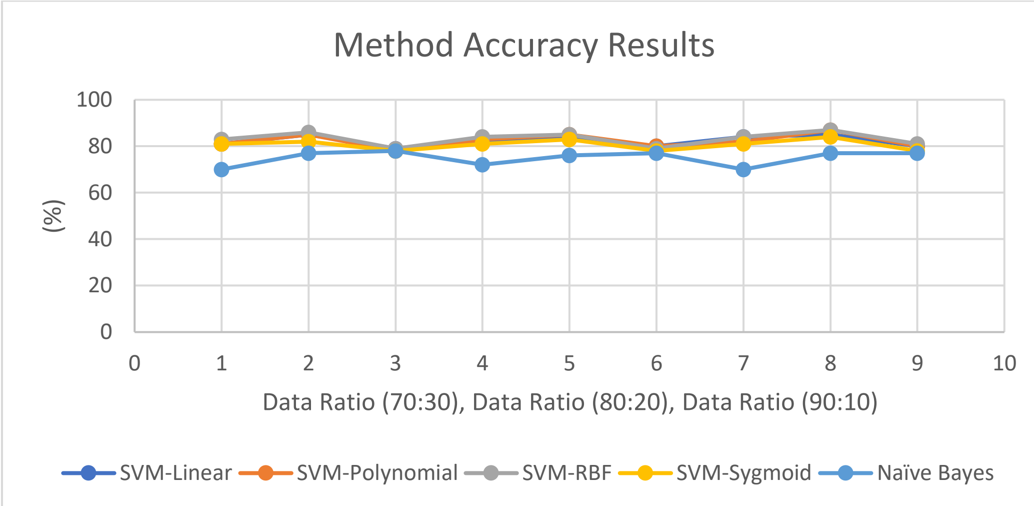

FIGURE 5. Method Accuracy Results Curve

4. **Conclusion.** This study utilized Twitter data from April 2023, encompassing 30,000 tweets with each candidate contributing 10,000 tweets. The observed candidates were Ganjar Pranowo, Prabowo Subianto, and Anies Baswedan, in the context of a presidential election. Various steps were employed to process the data, including data cleansing techniques such as lowercase conversion, tokenization, stop word removal, normalization, and stemming. The collected tweets were categorized into positive and negative classes with the assistance of linguistic experts. The TF-IDF extraction feature was used to gauge word frequency within documents and applied in classification methods using Support Vector Machine and Naïve Bayes. The study explored three data ratios (70:30, 80:20, and 90:10) to identify the best models in each method. The evaluation was conducted using precision, recall, f1 score, and accuracy metrics. Optimal results included accuracy in the SVM method with a 90:10 ratio and RBF/Polynomial kernels on the Prabowo Subianto dataset at 87%. Meanwhile, the lowest accuracy results were obtained in the Naive Bayes method with ratios of 70:30 and 90:10 on the Ganjar Pranowo dataset, which was 70%.

This study provides an interpretation of the results obtained based on voter turnout and sentiment. However, these factors can be misinterpreted because they have not been able to control bot/computer accounts and paid users/fake accounts and the data used has not been adjusted for population size or voter turnout.

Acknowledgment. This work is supported by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia based on Letter Number 0557/E5.5/AL.04/2023. The author also gratefully for the cooperation as a partner of the Nusantara Research Institute, the Center for Human Development Studies Foundation.

REFERENCES

[1] H. Yas *et al.*, "The Negative Role of Social Media During the COVID-19 Outbreak," *Int. J. Sustain. Dev. Plan.*, vol. 16, no. 2, pp. 219–228, Apr. 2021, doi: 10.18280/ijsdp.160202.

[2] B. Al sari *et al.*, "Sentiment analysis for cruises in Saudi Arabia on social media platforms using machine learning algorithms," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00568-5.

[3] A. Zahri, R. Adam, and E. B. Setiawan, "Social Media Sentiment Analysis using Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU)," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 1, pp. 119–131, 2023, doi: 10.26555/jiteki.v9i1.25813.

[4] E. S. Prihatini, “Women and social media during legislative elections in Indonesia,” *Womens. Stud. Int. Forum*, vol. 83, p. 102417, Nov. 2020, doi: 10.1016/j.wsif.2020.102417.

[5] S. Hinduja, M. Afrin, S. Mistry, and A. Krishna, “Machine learning-based proactive social-sensor service for mental health monitoring using twitter data,” *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100113, 2022, doi: 10.1016/j.jjimei.2022.100113.

[6] N. M. Azahra and E. B. Setiawan, “Sentence-Level Granularity Oriented Sentiment Analysis of Social Media Using Long Short-Term Memory (LSTM) and IndoBERTweet Method,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 1, pp. 85–95, 2023, doi: 10.26555/jiteki.v9i1.25765.

[7] D. Antypas, A. Preece, and J. Camacho-Collados, “Negativity spreads faster: A large-scale multilingual twitter analysis on the role of sentiment in political communication,” *Online Soc. Networks Media*, vol. 33, no. January, 2023, doi: 10.1016/j.osnem.2023.100242.

[8] R. H. Ali, G. Pinto, E. Lawrie, and E. J. Linstead, “A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election,” *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00633-z.

[9] K. K. Saha, “Election Violence,” in *Encyclopedia of Violence, Peace, & Conflict*, Elsevier, 2022, pp. 357–366. doi: 10.1016/B978-0-12-820195-4.00298-3.

[10] R. J. Toha, D. D. Gueorguiev, and A. Sinpeng, “The normalization of intolerance: The 2019 presidential election in Indonesia,” *Elect. Stud.*, vol. 74, p. 102391, Dec. 2021, doi: 10.1016/j.electstud.2021.102391.

[11] A. A. Firdaus, A. Yudhana, and I. Riadi, “Public Opinion Analysis of Presidential Candidate Using Naïve Bayes Method,” *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 2, pp. 563–570, May 2023, doi: 10.22219/kinetik.v8i2.1686.

[12] B. Haryanto, Y. Ruldeviyani, F. Rohman, T. N. Julius Dimas, R. Magdalena, and F. Muhamad Yasil, “Facebook analysis of community sentiment on 2019 Indonesian presidential candidates from Facebook opinion data,” *Procedia Comput. Sci.*, vol. 161, pp. 715–722, 2019, doi: 10.1016/j.procs.2019.11.175.

[13] K. Gupta, N. Jiwani, and N. Afreen, “A Combined Approach of Sentimental Analysis Using Machine Learning Techniques,” vol. 37, no. 1, pp. 1–6, 2023, doi: 10.18280/ria.370101.

[14] Aldinata, A. M. Soesanto, V. C. Chandra, and D. Suhartono, “Sentiments comparison on Twitter about LGBT,” *Procedia Comput. Sci.*, vol. 216, pp. 765–773, 2023, doi: 10.1016/j.procs.2022.12.194.

[15] Z. Geng, Q. Meng, J. Bai, J. Chen, and Y. Han, “A model-free Bayesian classifier,” vol. 482, pp. 171–188, 2019, doi: https://doi.org/10.1016/j.ins.2019.01.026.

[16] S. Wang, J. Ren, and R. Bai, “A semi-supervised adaptive discriminative discretization method improving discrimination power of regularized naive Bayes,” *Expert Syst. Appl.*, vol. 225, no. November 2022, 2023, doi: 10.1016/j.eswa.2023.120094.

[17] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, “Variable selection for Naïve Bayes classification,” *Comput. Oper. Res.*, vol. 135, 2021, doi: 10.1016/j.cor.2021.105456.

[18] K. K. Agustiningsih, E. Utami, and O. M. A. Alsyaibani, “Sentiment Analysis and Topic Modelling of The COVID-19 Vaccine in Indonesia on Twitter Social Media Using Word Embedding,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 8, no. 1, p. 64, 2022, doi: 10.26555/jiteki.v8i1.23009.

[19] D. Rozado, R. Hughes, and J. Halberstadt, “Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with Transformer language models,” *PLoS One*, vol. 17, no. 10 October, pp. 1–15, 2022, doi: 10.1371/journal.pone.0276367.

[20] A. H. Pratama, M. Hayaty, U. A. Yogyakarta, and R. R. Utara, “Performance of Lexical Resource and Manual Labeling on Long Short-Term Memory Model for Text Classification,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 1, pp. 74–84, 2023, doi: 10.26555/jiteki.v9i1.25375.

[21] R. Azizah Arilya, Y. Azhar, and D. Rizki Chandranegara, “Sentiment Analysis on Work from Home Policy Using Naïve Bayes Method and Particle Swarm Optimization,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 7, no. 3, p. 433, 2021, doi: 10.26555/jiteki.v7i3.22080.

[22] K. K. G. Lee, H. Kasim, W. J. Zhou, R. P. Sirigina, and G. G. T. Hung, “Feature redundancy assessment framework for subject matter experts,” *Eng. Appl. Artif. Intell.*, vol. 117, p. 105456, Jan. 2023, doi: 10.1016/j.engappai.2022.105456.

[23] H. Monkman, S. Mir, E. M. Borycki, K. L. Courtney, J. Bond, and A. W. Kushniruk, “Updating professional competencies in health informatics: A scoping review and consultation with subject matter experts,” *Int. J. Med. Inform.*, vol. 170, p. 104969, Feb. 2023, doi: 10.1016/j.ijmedinf.2022.104969.

[24] D. Arifah, T. H. Saragih, D. Kartini, and M. I. Mazdadi, “Application of SMOTE to Handle Imbalance Class in Deposit Classification Using the Extreme Gradient Boosting Algorithm,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 2, pp. 396–410, 2023, doi: 10.26555/jiteki.v9i2.26155.

[25] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, “Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset,” *Expert Syst. Appl.*, vol. 212, no. September 2022, p. 118715, 2023, doi: 10.1016/j.eswa.2022.118715.

[26] S. Xu, Y. Leng, G. Feng, C. Zhang, and M. Chen, “A gene pathway enrichment method based on improved TF-IDF algorithm,” *Biochem. Biophys. Reports*, vol. 34, no. December 2022, p. 101421, 2023, doi: 10.1016/j.bbrep.2023.101421.

[27] M. Liang and T. Niu, “Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs,” *Procedia Comput. Sci.*, vol. 208, pp. 460–470, 2022, doi: 10.1016/j.procs.2022.10.064.

[28] M. Chiny, M. Chihab, Y. Chihab, and O. Bencharef, “LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 7, pp. 265–275, 2021, doi: 10.14569/IJACSA.2021.0120730.

[29] A. Yudhana and A. Dwi, “Spatial distribution of soil nutrient content for sustainable rice agriculture using geographic information system and Naïve Bayes classifier,” vol. 16, no. 1, 2023, doi: 10.2478/ijssis-2023-0001.

[30] A. Yudhana, D. Sulistyo, and I. Mufandi, “GIS-based and Naïve Bayes for nitrogen soil mapping in Lendah, Indonesia,” *Sens. Bio-Sensing Res.*, vol. 33, p. 100435, 2021, doi: 10.1016/j.sbsr.2021.100435.

[31] Z. Ye, P. Song, D. Zheng, X. Zhang, and J. Wu, “A Naive Bayes model on lung adenocarcinoma projection based on tumor microenvironment and weighted gene co-expression network analysis,” *Infect. Dis. Model.*, vol. 7, no. 3, pp. 498–509, 2022, doi: 10.1016/j.idm.2022.07.009.

[32] A. Tariq, J. Yan, A. S. Gagnon, M. Riaz Khan, and F. Mumtaz, “Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest,” *Geo-Spatial Inf. Sci.*, vol. 00, no. 00, pp. 1–19, 2022, doi: 10.1080/10095020.2022.2100287.

[33] A. Tariq *et al.*, “Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive bayes tree models with optical remote sensing data,” *Heliyon*, vol. 9, no. 2, 2023, doi: 10.1016/j.heliyon.2023.e13212.

[34] R. Khan, Y. Qian, and S. Naeem, “Extractive based Text Summarization Using KMeans and TF-IDF,” *Int. J. Inf. Eng. Electron. Bus.*, vol. 11, no. 3, pp. 33–44, 2019, doi: 10.5815/ijieeb.2019.03.05.

[35] S. Cipolla and J. Gondzio, “Training very large scale nonlinear SVMs using Alternating Direction Method of Multipliers coupled with the Hierarchically Semi-Separable kernel approximations,” *EURO J. Comput. Optim.*, vol. 10, 2022, doi: 10.1016/j.ejco.2022.100046.

[36] W. Liu, S. Liang, and X. Qin, “Weighted p-norm distance t kernel SVM classification algorithm based on improved polarization,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–17, 2022, doi: 10.1038/s41598-022-09766-w.

[37] N. H. Putri, M. Fatekurohman, and I. M. Tirta, “Credit risk analysis using support vector machines algorithm,” *J. Phys. Conf. Ser.*, vol. 1836, no. 1, pp. 1–11, 2021, doi: 10.1088/1742-6596/1836/1/012039.

[38] A. Peryanto, A. Yudhana, and R. Umar, “Convolutional Neural Network and Support Vector Machine in Classification of Flower Images,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 8, no. 1, pp. 1–7, 2022, doi: 10.23917/khif.v8i1.15531.

[39] A. Basuki, “Sentiment Analysis of Customers ’ Review on Delivery Service Provider on Twitter Using Naive Bayes Classification,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 2, pp. 420–428, 2023, doi: 10.26555/jiteki.v9i2.26327.

[40] H. Syahputra and A. Wibowo, “Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 1, pp. 165–173, 2023, doi: 10.26555/jiteki.v9i1.25861.

[41] A. D. Cahyani, “Aspect-Based Sentiment Analysis from User-Generated Content in Shopee Marketplace Platform,” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 2, pp. 444–454, 2023, doi: 10.26555/jiteki.v9i2.26367.