Jurnal Krisnadana (Jurnal Komputer, Sistem Kendali & Jaringan)
Vol 4, Issue 3, Mei 2025 , pp. 131-141
https://doi.org/10.58982/krisnadana.v4i2.718

E-ISSN: 2808-3563
P-ISSN: 2808-3520

Sentiment Analysis of Indonesian Public Response to The 12% Value Added Tax Increase Using The Naïve Bayes and Multinomial Naïve Bayes Method

I Made Ary Widiantara Putra [1], Ketut Jaya Atmaja[2*], Putu Surya Lesmana[3] ,Desak Made Dwi Utami Putra[4], I Gede Andika[5]

[1,2*,3,4] Informatika, Institut Bisnis Dan Teknologi Indonesia , Denpasar, Bali, Indonesia
[5] Rekayasa Sistem Komputer, Institut Bisnis Dan Teknologi Indonesia , Denpasar, Bali, Indonesia

[1]arywidiantara86@gmail.com, [2*]ketutjayaatmaja@instiki.ac.id, [3]suryawedra@gmail.com, [4]desak.utami@instiki.ac.id, [5]gdandika@instiki.ac.id

ARTICLE INFO

Article history:
Received 15 April 2025
Revised 17 Mei 2025
Accepted 13 Juni 2025

***Keywords**:*
Multinomial Naïve Bayes;
Naïve Bayes;
Sentiment Analysis;
Value Added Tax

ABSTRACT

The policy of increasing the Value Added Tax (VAT) to 12% implemented by the government has sparked various reactions from the public, both supportive and opposing. This study aims to analyze public sentiment toward this policy using two classification methods: Naïve Bayes and Multinomial Naïve Bayes. The data was obtained through crawling from social media platform X with a total of 3,024 comments that had undergone preprocessing stages. The analysis was conducted using TF-IDF weighting, and model evaluation was performed using a Confusion Matrix. The evaluation results showed that the Multinomial Naïve Bayes algorithm performed better with an accuracy of 75%, compared to Naïve Bayes which only reached 71%. The majority of public sentiment towards the 12% VAT increase was negative, driven by concerns over economic impacts such as rising prices of goods. This study is expected to provide input for the government in formulating more targeted public communication strategies related to fiscal policies.

1. Introduction

The taxation system in Indonesia is clearly regulated under Law Number 28 of 2007 concerning General Provisions and Taxation Procedures, which serves as the foundation for facilitating national tax administration [1]. Among the various types of taxes implemented in Indonesia, the Value Added Tax (VAT) stands out as a fundamental component. VAT is imposed on the sale or importation of goods and services and plays a pivotal role in generating state revenue. The legal basis for the implementation of VAT was initially established through Law Number 8 of 1983, which was subsequently revised and updated by Law Number 42 of 2009 [2]. As one of the primary sources of public income, VAT functions as a crucial financial instrument to support the state budget and contribute significantly to national development initiatives.

In recent years, the Indonesian government's policy to raise the VAT rate to 12% has sparked widespread public discourse. The proposed increase has led to mixed reactions from society, ranging from expressions of support to sharp criticism. These reactions have been particularly prominent on social media platform X, which has emerged as a dynamic and valuable source for capturing public opinion. The platform hosts a large volume of user-generated content, including comments, opinions, and sentiments related to the tax policy, thereby offering an extensive dataset for deeper exploration.

To systematically understand the spectrum of public responses and to determine the prevailing sentiment whether supportive or oppositional it is essential to conduct sentiment analysis. Sentiment analysis refers to the management of sentiments, opinions, and subjective text [3]. Sentiment analysis provides comprehension information related to public views, as it analyzes different tweets and reviews. This technique enables researchers to identify patterns and trends in public opinion, offering insights that can inform policymakers about the level of acceptance or resistance to the VAT policy. The outcomes of such analysis are expected to provide a comprehensive overview of how the general population perceives the tax increase, thus serving as meaningful input for future policy decisions.

This study compares two classification methods Naïve Bayes and Multinomial Naïve Bayes for sentiment analysis on Indonesian-language social media data regarding the 12% VAT increase. Naïve Bayes is a simple and efficient probabilistic classifier based on Bayes' Theorem with an independence assumption [4], while Multinomial Naïve Bayes is tailored for text classification by considering word frequency [5]. Both methods are computationally efficient and suitable for high-dimensional data, with the Multinomial variant often performing better in capturing sentiment nuances. Despite prior use of Naïve Bayes in sentiment analysis, few studies have compared both models in socio-economic contexts. This study fills that gap and contributes to selecting appropriate methods for analyzing public opinion on policy changes.

2. Method

This diagram illustrates a common machine learning workflow, starting with Data Crawling to collect raw text, followed by Preprocessing for cleaning, Data Labeling for categorization, and Data Splitting for training/testing. TF-IDF is then used for feature extraction, which feeds into Naïve Bayes and Multinomial Naïve Bayes classifiers, with performance ultimately assessed using a Confusion Matrix.

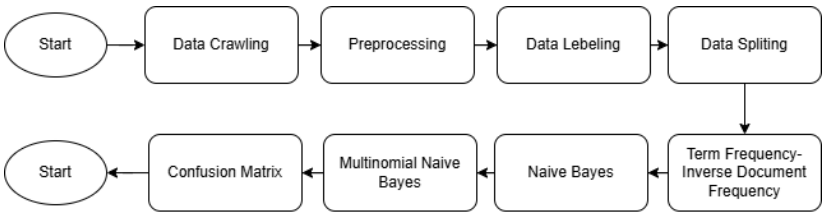

Figure 1. Research Stage

This study involves a series of essential stages that have been systematically and comprehensively designed to produce findings and conclusions with a high degree of

accuracy. Each stage—from the initial data collection process, followed by data processing, to the final stage of in-depth analysis—has been carefully planned and considered based on scientific methodological principles. This meticulous planning aims to ensure that the collected data is truly relevant to the research topic, that the methods employed are appropriate for the characteristics of the data and the objectives of the study, and that the results obtained are reliable and academically accountable. The entire research process plays a crucial role in fully and comprehensively understanding the public's perceptions, views, and responses toward the government policy regarding the increase in Value Added Tax (VAT).

Data collection for this study was conducted over a relatively extended period, starting from March 2024 to January 2025. This long timeframe was deliberately chosen to capture the evolving dynamics of public opinion and sentiment in response to the 12% VAT increase policy. By collecting data over a broad period, the researchers were able to observe how public perception was formed, changed, and influenced by various events, announcements, and circulating information throughout the implementation of the policy. Accordingly, the research process can be described as follows:

Data Crawling

This study specifically focuses on sentiment analysis using Indonesian-language data, with the aim of understanding how the public responds to the policy of increasing the Value Added Tax (VAT). The initial step in the research process involved data collection through a crawling method on one of the most popular social media platforms, X. This technique allowed the researchers to automatically retrieve a large volume of data relevant to the research topic. The collected data consisted of various forms of content, such as posts, comments, opinions, and responses from netizens that directly or indirectly referred to the VAT increase issue [6].

Through this approach, the study seeks to explore in greater depth the public's perceptions, attitudes, and reactions toward the government policy, as reflected in their social media activity. During implementation, the crawling process successfully gathered a total of 3,004 data entries, which were then used as the foundation for further sentiment analysis..

Text Preprocessing

Preprocessing is a crucial process carried out to prepare text data for subsequent analysis [7]. This stage is crucial in data processing because it ensures that the data is in the proper format for further handling. The preprocessing phase involves several procedures, including data cleansing, case folding, normalization, tokenization, removing meaningless words and irrelevant characters while showcasing only important words so that the model can train better , stopwords removal, and stemming [8]. These steps result in cleaner and more structured data, making it easier to analyze. Moreover, preprocessing helps reduce noise in the data, which can otherwise negatively impact the analysis results. It has an important function of data cleaning, returning more accurate results of unstructured social media data sources. It is used to remove any noise information, to decrease data content, and to enhance the accuracy of classification [9]. Overall, preprocessing is a vital phase in the data processing pipeline to achieve optimal outcomes.

Data Labeling

Labeling is one of the processes in text mining, where it is used to categorize sentences into positive and negative categories. This process aims to assign labels or markers to each sentence based on the sentiment contained within it. With labeling, sentences that have positive or negative meanings can be grouped more easily[10].

Vader Lexicon is one of the analysis methods within the category of lexicon-based algorithms used to evaluate sentiment in text. In this algorithm, data is analyzed and assessed based on a predefined dictionary of words, where each word has a polarity score reflecting its emotional intensity. The output of this method includes sentiment classification into polarity categories such as positive, neutral, and negative, as well as a composite or total score that reflects the overall sentiment of the text [11].

Data Splitting

Data splitting is the process of dividing a dataset into two or more distinct subsets, usually for the purpose of training and testing a machine learning model [12]. The main goal of data splitting is to evaluate the model's performance on data that has not been seen during the training process. In this study, 80% of the total dataset is used as training data, while the remaining 20% is used as testing data. The data splitting process is carried out using the sklearn library in the Python programming language. This step is important to ensure that the resulting model can be accurately tested on unseen data, allowing its performance to be evaluated more objectively.

Term frequency inverse document frequency(TF-IDF)

The word weighting stage is the process of assigning scores to words in a text to determine the extent to which those words reflect positive, negative, or neutral sentiment. A commonly used technique is TF-IDF (Term Frequency - Inverse Document Frequency), which calculates how often a word appears in a text and how rarely that word appears across the entire collection of documents. This process aims to highlight important and relevant words in the analysis [13].

TF-IDF is used as a feature extraction method to transform text from documents into numerical vectors that can be utilized by classification algorithms. By applying TF-IDF, the importance level of words within a document can be measured, producing feature representations that help distinguish between different classes in text classification. Feature extraction in data mining involves steps to reduce the amount of data available while effectively describing large datasets. In analyzing complex text sentiment, one major challenge is the high number of variables involved. Therefore, applying feature extraction techniques to input data before feeding it into classification algorithms can significantly improve the accuracy of the classification model [14]. The following is an example of TF-IDF weighting results.

Tabel 1. Example TF-IDF Weighting

Words	Weight
Adil	0,309075
Bangun	0,307698
bijak	0,2415320
Dorong	0,4049659
Ekonomi	0,249174
Rata	0,452209
Ppn	0,0930492

Stabilitas	0,4522089

3. Result and Discussions

This study uses data sourced from social media platform X, totaling 3,004 entries. After data collection, a preprocessing step was carried out to tidy and clean the data by removing irrelevant elements. Subsequently, the data were labeled with two sentiment classes, namely positive and negative, while neutral sentiment data were excluded. Following the labeling process, the dataset was reduced to 2,048 entries.

The study applies two classification methods, namely Naïve Bayes and Multinomial Naïve Bayes, with the aim of comparing the accuracy levels of both methods in analyzing sentiment on the processed data. Thus, the evaluation results are expected to provide insights into which method is more effective for sentiment classification in the context of this research.

Model Training and Testing Using Confusion Matrix

This study compares two methods, Naïve Bayes and Multinomial Naïve Bayes, to determine which is better suited for the research. The confusion matrix evaluates classification model performance by showing correct and incorrect predictions [15]. It is a table where rows represent predicted classes and columns show actual classes, helping to identify true and false predictions for easier model evaluation. The matrix also highlights errors like False Positives and False Negatives. From it, metrics such as accuracy, precision, and recall can be calculated to give a comprehensive view of the model's performance.

Naïve Bayes

After the classification process, the first evaluation was carried out using the Naïve Bayes method with the aim of measuring the model's performance in classifying the data. The evaluation results, presented using a Confusion Matrix, are shown in Figure 3 (a) to provide a clear overview of the model's performance

Here is the confusion matrix showing the results of true positive, false positive, true negative, and false negative :

a. True Positive (TP) 154 data data Points.
b. False Positive (FP) 32 data points.
c. True Negative (TN) 137 data points.
d. False Negative (FN) 87 data points.

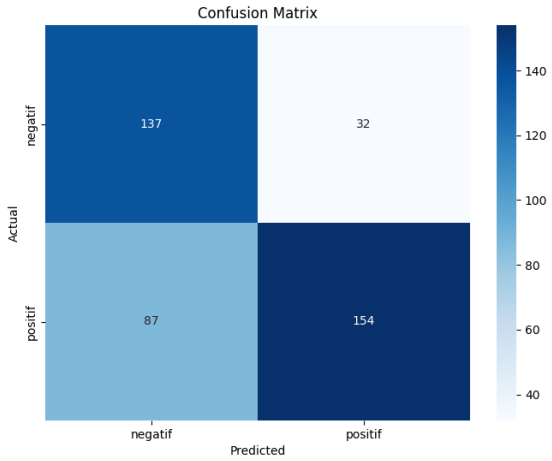

Figure 2. Confusion Matrix

Using evaluation with a confusion matrix, the model achieved an accuracy of 71%, meaning it correctly predicted 71% of the test data. The precision for the negative class is 0.61 and for the positive class is 0.83, indicating the percentage of correct predictions for each class. Recall for the negative class is 0.81 and for the positive class is 0.64, representing the percentage of actual data that was successfully detected. The F1-score of 0.70 (negative) and 0.72 (positive) reflects the balance between precision and recall. Support for the negative class is 169 and for the positive class is 241, which are the actual numbers of data points in each class.

Tabel 2. Confusion Matrix Results

	Precision	**Recall**	**F1-Score**	**Support**
Negative	0,61	0,81	0,70	169
Positive	0,83	0,64	0,72	241
Accuracy				0,71

Multinomial Naïve Bayes

The second evaluation was conducted after the classification process using the Multinomial Naïve Bayes method, aiming to measure the model's performance in classifying the data. The evaluation results, presented using a Confusion Matrix, are shown in Figure 3 (b) to provide a clear overview of the model's performance.

Here is the confusion matrix showing the results of true positive, false positive, true negative, and false negative:

a. True Positive (TP) 227 data points.
b. False Positive (FP) 37 data points.
c. True Negative (TN) 123 data points.
d. False Negative (FN) 40 data poits.

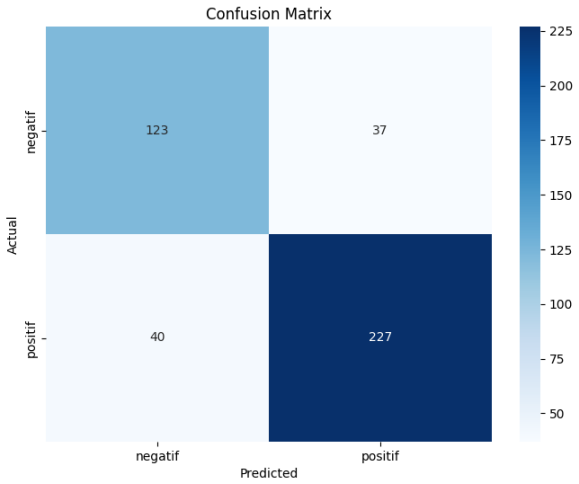

Figure 3. Confusion Matrix

Using evaluation with a confusion matrix, the model achieved an accuracy of 75%, meaning it correctly predicted 75% of the test data. The precision for the negative class is 0.73 and for the positive class is 0.76, indicating the percentage of correct predictions for each class. Recall for the negative class is 0.62 and for the positive class is 0.84, representing the percentage of actual data that was successfully detected. The F1-score of 0.67 (negative) and 0.80 (positive) reflects the balance between precision and recall. Support for the negative class is 169 and for the positive class is 241, which are the actual numbers of data points in each class.

Tabel 3. Confusion Matrix Results

	Precision	Recall	F1-Score	Support
Negative	0,73	0,62	0,67	169
Positive	0,76	0,84	0,80	241
Accuracy				0,75

Visualization of Analysis Results

Data visualization facilitates the presentation of research findings, making complex information easier to understand and more engaging. Patterns within the data become clearer through images, thereby accelerating comprehension.

Figure 4. Positive Wordcloud

The positive word cloud shows the dominance of words such as “ppn”, “pajak”, and “naik”, indicating support for the VAT increase. Words like “dampak”, “bijak”, “masyarakat”, and “negara” reflect a positive perspective on this policy. The appearance of words such as “layak”, “program”, and “dukung” reinforces that many comments view it as a proper step to boost the economy and public services.

Figure 5. Negative Wordcloud

The negative word cloud highlights words such as “ppn”, “naik”, “pajak”, “tolak”, and “harga”, reflecting concerns over a policy perceived as burdensome. Words like “rakyat”, “kena”, “barang”, and “susah” express complaints from the public, especially the lower-middle class. The appearance of words such as “salah”, “perintah”, and “buruk” indicates dissatisfaction with the government, while emotional terms like “gila”, “gak”, and “bodoh” suggest a strong negative reaction.

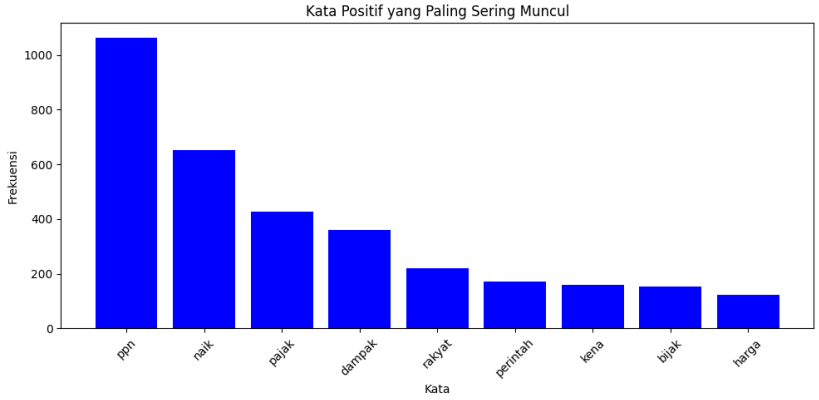

Figure 6. Positive Graph

The graph shows that the words “ppn”, “naik”, and “pajak” frequently appear in positive sentiments, indicating support and positive information related to the policy. The words “dampak”, “rakyat”, “pemerintah”, and “bijak” reflect the view that this policy is beneficial and in favor of the public.

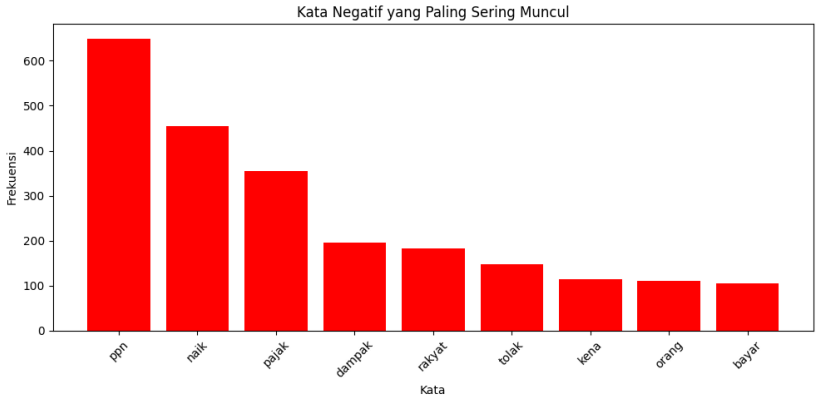

Figure 7. Negative Graph

The graph shows the dominance of the words “ppn”, “naik”, and “pajak” in negative sentiments, reflecting concerns and the perception that this policy is burdensome. The words “dampak”, “rakyat”, and “tolak” indicate rejection, while “kena”, “orang”, and “bayar” suggest economic pressure and a decline in people's purchasing power.

Model Evaluation and Findings

The following presents a comprehensive summary of the model evaluation table, which is based on an 80:20 data split between training and testing sets. This division allows for a balanced assessment of each model’s performance by training the model on 80% of the data and evaluating its accuracy and effectiveness using the remaining 20%. The results provide insights into how well each algorithm generalizes to unseen data and highlight the comparative strengths of the models used in the sentiment analysis.

Tabel 4. Model Evaluation

Model	Sentiment	Model Evaluation			
		Accuracy	Precision	Recall	F1-Score
Naïve Bayes	Positif	71%	83%	64%	72%
	Negatif		61%	81%	70%
Multinomial	Positif	75%	76%	84%	80%
Naïve Bayes	Nrgatif		73%	62%	67%

The evaluation compared the performance of Naïve Bayes and Multinomial Naïve Bayes using a confusion matrix based on an 80:20 data split. The results showed that Multinomial Naïve Bayes achieved higher accuracy (75%) than Naïve Bayes (71%). This indicates that the frequency-based approach of Multinomial Naïve Bayes is more effective for text classification in this context. A key finding is the dominance of negative sentiment in public

response, mainly driven by concerns over price increases and economic pressure. The combination of TF-IDF and Multinomial Naïve Bayes proved effective in capturing sentiment patterns in Indonesian language social media data.

4. Conclusion

This study analyzes the sentiments of Indonesian society toward the 12% VAT increase using the Naïve Bayes and Multinomial Naïve Bayes methods. Evaluation results show that Multinomial Naïve Bayes has a higher accuracy of 75%, compared to Naïve Bayes which only achieved 71%. The majority of public sentiment tends to be negative, as indicated by the dominance of words such as "naik", "harga", and "susah" in the data. This reflects public concern about the economic impact of the policy. The use of TF-IDF and an 80:20 data split has proven to support the effectiveness of the analysis. These findings provide important insights for the government in understanding public opinion on fiscal policies. In addition, the results of this study can serve as a reference in selecting text-based sentiment analysis methods in the future.

References

[1] G. Samuel, "Analisis Yuridis Tingkat Kepatuhan Membayar Pajak Masyarakat Indonesia," 2022.

[2] D. Bisnis *dkk.*, "ASSET: JURNAL MANAJEMEN Analisis Penerapan Pajak Pertambahan Nilai (PPN) Atas Transaksi E-Commerce Pada Platform Marketplace PT. Bukalapak," *Jurnal Ilmiah Bidang Manajemen dan Bisnis Vol. 4*, 2021, [Daring]. Tersedia pada: http://journal.umpo.ac.id/index.php/ASSEThttp://journal.umpo.ac.id/index.php/asset

[3] Q. T. Ain *dkk.*, "Sentiment Analysis Using Deep Learning Techniques: A Review," 2017. [Daring]. Tersedia pada: www.ijacsa.thesai.org

[4] S. Hajar, A. Samsudin, N. M. Sabri, N. Isa, U. Fatihah, dan M. Bahrin, "Sentiment Analysis on Acceptance of New Normal in COVID-19 Pandemic using Naïve Bayes Algorithm," 2024. [Daring]. Tersedia pada: www.ijacsa.thesai.org

[5] Yuyun, Nurul Hidayah, dan Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, hlm. 820–826, Agu 2021, doi: 10.29207/resti.v5i4.3146.

[6] O. Dini, F. Sari, A. Kusjani, D. Kurniawati, dan I. Setiawan, "PENCARIAN DATA QUICK COUNT PILPRES DENGAN TEKNIK WEB SCRAPING," *JIRK Journal of Innovation Research and Knowledge*, vol. 3, no. 5, 2022.

[7] D. Cahyo Ramadhan dan F. Irwiensyah, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Pengguna Terhadap Aplikasi Bing Chat di Google Play Store dengan Metode Naïve Bayes," *Media Online*, vol. 4, no. 5, hlm. 2410–2418, 2024, doi: 10.30865/klik.v4i5.1769.

[8] S. Jeetendra Vadher, "Performance Analysis of Naïve Bayes and Stochastic Gradient Descent-based SVM for Sentiment Analysis," *SSRG International Journal of Computer*

Science and Engineering, vol. 12, hlm. 10–18, 2025, doi: 10.14445/23488387/IJCSE-V12I15P102.

[9] M. Kumar, L. Khan, dan H. T. Chang, "Evolving techniques in sentiment analysis: a comprehensive review," 2025, *PeerJ Inc.* doi: 10.7717/PEERJ-CS.2592.

[10] L. Legito *dkk.*, "Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Terhadap Isu Khilafah dan Radikalisme di Indonesia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, hlm. 324–330, Nov 2023, doi: 10.57152/malcom.v3i2.893.

[11] M. Angga Gumilang, A. Sirojudin, F. Abdilllah, M. Yusril Amin, dan W. Budi Lestari Politeknik Negeri Jember, "SNESTIK Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika Analisis Sentimen terhadap Kecurangan Pemilu dan SIREKAP di Twitter menggunakan Metode Vader Lexicon dan Naïve Bayes," *e jurnal. itats*, 2024, doi: 10.31284/p.snestik.2024.5877.

[12] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, dan L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, hlm. 273–281, Jan 2024, doi: 10.57152/malcom.v4i1.1085.

[13] M. Hafizh Mahendra, D. Triantoro Murdiansyah, dan K. Muslim Lhaksmana, "Dike : Jurnal Ilmu Multidisiplin Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer," *Dike : Jurnal Ilmu Multidisiplin*, 2023.

[14] L. Afuan, M. Khanza, dan A. Zahira Hasyati, "ENHANCING SENTIMENT ANALYSIS OF THE 2024 INDONESIAN PRESIDENTIAL INAUGURATION ON X USING SMOTE-OPTIMIZED NAIVE BAYES CLASSIFIER," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 1, hlm. 325–333, Feb 2025, doi: 10.52436/1.jutif.2025.6.1.4290.

[15] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, hlm. 4023–4031, Nov 2024, doi: 10.53555/AJBR.v27i4S.4345.