Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Sentiment Analysis of Public Opinion on the 2024 President-Elect's Administration on Twitter using Naïve Bayes

Puput Rifani[1]*, Bambang Irawan[2], Ahmad Faqih[3], Denni Pratama[4], Dian Ade Kurnia[5]

[1,2,3]Informatics Engineering, STMIK IKMI Cirebon
[4]Accounting Computer, STMIK IKMI Cirebon
[5]Computer Science Management, STMIK IKMI Cirebon
puputrifani2@gmail.com[1]*

Abstract

The increased use of social media, especially Twitter, has created a need for systematic analysis to understand public opinion on political issues, including the performance of the president-elect in 2024. This study analyzes public opinion on these issues using the Naïve Bayes algorithm. Data was collected using scraping techniques and then divided into three sentiment categories positive, negative, and neutral. After the labeling process, the data underwent preprocessing, which included data cleaning, case folding, normalization, tokenization, stop word removal, and stemming. TF-IDF weighting was used to represent features, while the SMOTE technique was applied to balance class distribution. A total of 1,074 tweets were analyzed. The results showed that negative opinions dominated at 59.9%, followed by positive opinions at 29.8% and neutral opinions at 10.3%. Model performance evaluation showed that Naïve Bayes was able to consistently identify sentiment patterns, with 71% accuracy, 74% precision, 71% recall, and an F1 score of 72%. These results prove that the combination of TF-IDF and SMOTE contributes significantly to improving classification effectiveness. This study provides a comprehensive overview of trends in public opinion.

***Keywords**: Sentiment Analysis, Twitter, Naïve Bayes.*

1. Introduction

The advancement of information technology has had a significant impact on various aspects of life, both in the political and governmental sectors, through its ability to provide fast and widespread access to data [1]. Social media has become an important space for conveying information, shaping public opinion, and facilitating two-way interaction between the public and the government [2]. Twitter, which has a very high volume of political conversation, serves as an important indicator in gauging public perception of government issues, especially in the post-2024 election period when public opinion dynamics greatly affect national political stability [3]. This underscores the importance of data-driven analysis to understand public response patterns to governance issues more accurately [4].

A number of previous studies have applied machine learning methods in assessing public opinion on social media, particularly using the Naïve Bayes method, which is known to be stable in handling high-dimensional text data. Research by [5] confirms the effectiveness of Naïve Bayes in mapping public opinion, but does not consider the existence of neutral sentiments that can provide a more objective picture. Another study by [6] produced similar findings, but that study did not integrate supporting techniques such as TF-IDF and SMOTE simultaneously. These findings are strengthened by [7] research showing that the integration of Naïve Bayes and SMOTE provides the highest performance in sentiment classification on political issues on YouTube, with an accuracy of 73%. Thus, a more comprehensive and representative analytical approach is still needed in assessing public opinion on the performance of the president-elect's administration in 2024.

Along with previous research studies, the main focus of this analysis is to understand how Naïve Bayes is used to classify public sentiment towards the performance of the president-elect in 2024 using Twitter data. The main aspects considered include the procedure for applying the method, the classification results, and the evaluation of model performance based on various quantitative metrics [8]. This approach is an important basis for identifying trends in public opinion developing on social media.

As a continuation of this focus, this study also aims to develop a sentiment analysis system capable of identifying public opinion objectively and systematically. The implementation of the Naïve Bayes algorithm focuses on grouping sentiments into positive, negative, and neutral categories, accompanied by performance evaluation using accuracy, precision, recall, and f1-score [9]. This method aims to provide a more comprehensive view of how the public perceives the performance of the president-elect's administration in 2024.

This study is intended to contribute to the development process, particularly in the field of informatics related to the use of machine learning for public opinion analysis. The integration of TF-IDF, SMOTE, and Naïve Bayes is expected to improve the accuracy of the model in understanding sentiment and provide a basis for the government in developing flexible, clear, and responsive communication strategies

[10]. In addition, the findings from this study can be used as a reference for scientists and researchers who want to explore social and political dynamics in the digital world.

2. Research Method

This research is a quantitative study designed to analyze public opinion through large-scale text data processing using data mining and sentiment analysis techniques. A quantitative approach was used because it is capable of converting text into numerical data that can be analyzed systematically, allowing patterns, trends, and relationships between variables to be identified objectively [11]. The research process began with data collection from Twitter as the main source for obtaining text that represents public opinion. The collected data was manually labeled with linguistic expert validation to ensure consistency in determining positive, negative, and neutral sentiments. The reason for choosing the Naïve Bayes algorithm is based on findings [12] showing that the combination of Naïve Bayes with TF-IDF weighting can provide a stable increase in accuracy in text processing. Next is data preprocessing, including data cleaning, case folding, normalization, tokenization, stopword removal, and stemming. Modeling was implemented using Naïve Bayes with TF-IDF weighting, while class imbalance was addressed by applying the Synthetic Minority Oversampling Technique (SMOTE) to generate artificial samples in the minority class, as explained by [13] that this method is effective in improving class distribution so that the model can recognize patterns more proportionally. The results of sentiment analysis and model performance evaluation are then used to describe public opinion trends and provide recommendations for further research development.

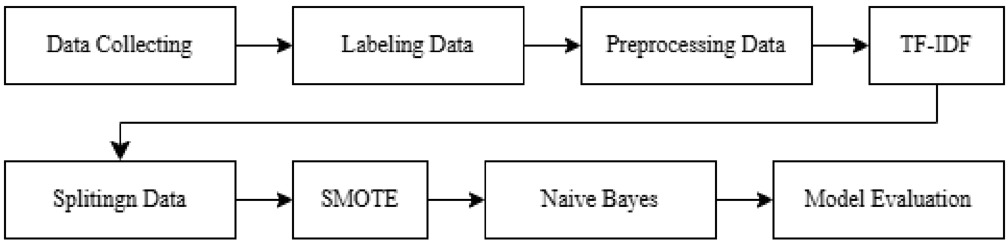

Fig 1: Research method stages

2.1. Data Collection

The data source for this study was secondary data in the form of tweets collected from Twitter using web scraping techniques. The data was gathered through Google Collab using Python libraries such as Tweepy and the tweet-harvest 2.6.1 framework, using the keywords "presidential performance" and "new government." The data collection period ran from January to October 2025. From this process, 1,074 tweets that met the criteria were collected and used as the basis for sentiment analysis.

2.2. Labeling Data

The sentiment categorization process is done manually with language expert validation to ensure the accuracy of the dataset. This approach allows assessors to understand the language context in depth and produce more accurate labels than automated methods. Sentiment labeling is categorized into three classes, namely positive, negative, and neutral, based on consistent annotation guidelines. Expert validation is considered the most effective method for generating labels that accurately reflect the meaning of sentiment and support the creation of more accurate classification model [14].

Table 1: Labeling dataset

Text	label
kinerja pemerintahan presiden prabowo benar benar dirasakan hasilnya secara nyata oleh masyarakat Indonesia	positif
presiden nya wira wiri kemana mana cuma menghasilkan omon omon pepesan kosong wapresnya cuma bisa bagi bagi susu sembunyi. kabinetnya gak kompeten amburadul gak bisa kerja aparat hukumnya sibuk mau nangkepin lawan politik. tapp 81% publik puas atas kinerja mereka. negeri yg aneh..	negatif
menurutku kinerja presiden dan kabinet akan ada plus minus tapi untuk bisa lebih objektif menarik untuk disaksikan Rilis Temuan Survei Nasional Indikator terkait seperti apa dan bagaimana evaluasi publik atas kinerja presiden dan kabinet merah putih.	netral

2.3. Data Preprocessing

Data preprocessing serves to clean, standardize, and prepare text so that it can be optimally processed by machine learning models, ranging from removing non-informative elements through data cleaning to simplifying letters through case folding [15]. Normalization is then carried out to standardize non-standard words, abbreviations, and spelling variations so that the text becomes more consistent and does not reduce the accuracy of the model. The process continues with tokenization, which breaks sentences into word units, and stopword removal to eliminate common words that do not contribute to the sentiment context. Finally, stemming is applied to return words to their base form so that the complex morphological structure of the Indonesian language can be processed more accurately by the sentiment analysis model.

2.4. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF weighting is a method applied to convert text into numerical format by calculating the importance of words based on a combination of local weighting (TF) and global weighting (IDF) [16]. This approach represents each tweet as a numerical vector that highlights the words most relevant to a particular issue, thereby improving the performance of models such as Naïve Bayes in distinguishing sentiment.

2.5. Splitting Data

The data splitting process in this study used an 80:20 ratio, with 80% of the data allocated as training data and 20% as test data. This ratio was chosen because it provides a balance between the model's need to learn patterns from the available data and the need for representative evaluation. The larger proportion of training data allows the model to learn more deeply, while the test data remains sufficient to assess the model's generalization ability. Data division was performed randomly and stratified to ensure that the sentiment class distribution remained consistent in both data subsets. This approach aims to prevent class imbalance that could potentially affect the quality of the training and evaluation processes. In addition, the principle of no data leakage is applied by ensuring that the test data does not affect the training and pre-processing stages. With this approach, model performance evaluation can be carried out objectively and the results better reflect the model's generalization ability to new data [17].

2.6. Synthetic Minority Over-sampling Technique (SMOTE)

The Synthetic Minority Over-sampling Technique (SMOTE) is used to address class imbalance by generating synthetic samples through interpolation between minority class data points. Research by [18] shows that the application of SMOTE, either in combination with TF-IDF or before classification algorithms such as Naïve Bayes, can improve accuracy, precision, recall, f1-score, and model robustness in various text processing contexts. Based on these findings, the use of SMOTE in this study is focused on improving the representation of minority classes so that the model can recognize sentiments more evenly and produce more stable and accurate classification performance.

2.7. Model Evaluation

Model evaluation was conducted to assess the reliability and consistency of the developed sentiment classification system. Model performance was measured using accuracy, precision, recall, and F1-score metrics as key indicators of classification quality. Accuracy was not used as the sole reference because it was not representative of conditions where class distribution was imbalanced. Therefore, precision, recall, and F1-score were used to provide a more comprehensive picture of the model's ability to recognize each sentiment class. The selection of evaluation metrics appropriate to the characteristics of the data and the classification task was an important factor in ensuring that the evaluation results were objective and scientifically credible [19].

3. Result and Discussion

3.1. Labeling Data

The labeling process was implemented by assigning sentiment classes to each tweet in the dataset based on the text content. Each tweet was labeled as positive, negative, or neutral, according to the meaning and tendency of the opinion contained therein. The labeling results were stored in the label column, while the text that had undergone preprocessing was placed in the text_preprocessed column. After the labeling process was completed, a total of 1,074 data points were obtained, ready to be used as training and testing data for the model. The distribution of the labeling results showed that there were 643 negative tweets, 320 positive tweets, and 111 neutral tweets.
The class distribution is visualized through a bar chart, which illustrates the comparison of the number of tweets in each sentiment category. This chart shows that negative sentiment significantly dominates the other two classes, while positive sentiment ranks second, and neutral sentiment is the category with the smallest number.

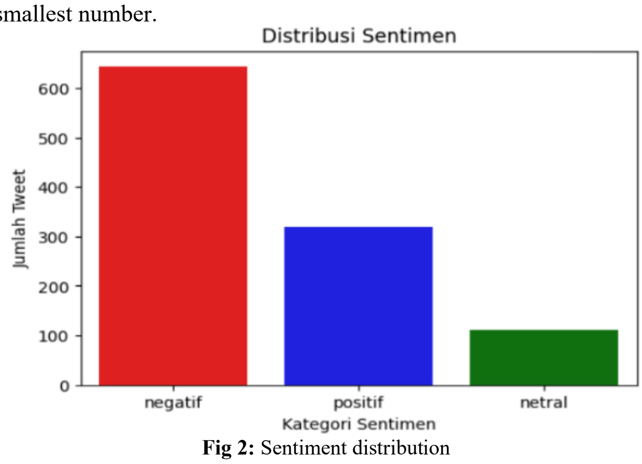

Fig 2: Sentiment distribution

In addition, the pie chart shows the overall percentage proportion of each sentiment class. Based on this visualization, negative sentiment dominates with a proportion of 59.9%, followed by positive sentiment at 29.8%, and neutral sentiment at 10.3%. This display reinforces the finding that public perception on social media regarding the performance of the President-elect's administration in 2024 is more negative than positive or neutral.

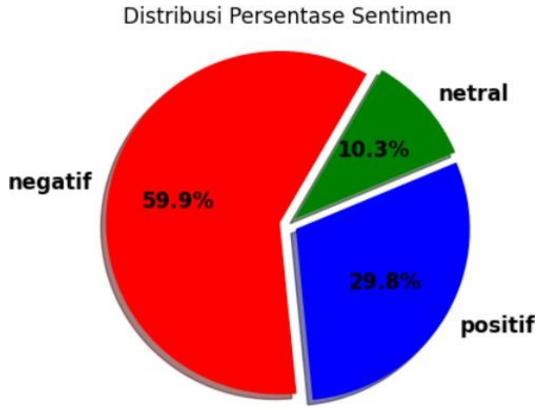

Fig 3: Sentiment percentage distribution

3.2. Data Preprocessing

The data preprocessing phase is used to prepare the text data so that it can be processed effectively by the sentiment classification model. This phase includes data cleaning to remove links, mentions, hashtags, punctuation marks, numbers, and unnecessary spaces. Case folding is used to convert capital letters to lowercase letters to unify all letters, normalization is used to convert non-standard words or abbreviations into standard Indonesian, for example, “gk” is changed to “tidak” and ‘yg’ to “yang”. Next, tokenization is performed to divide the text into words, followed by stopword removal to eliminate common words that do not affect the meaning of the sentiment, and stemming to return derivative words to their basic form using Sastrawi. The result of this entire process is clean, standardized text that is ready for sentiment analysis, which is stored in the “text_preprocessed” column, with a consistent text display that reflects its original meaning.

```
Data setelah Preprocessing:
                                           full_text
0  @Heraloebss @prabowo @KejaksaanRI sejauh ini p...
1  Suara netizen di media sosial mungkin mewakili...
2  @txtdrimedia Artinya sekitar 236 1 juta pendud...
3  Bagaimana ini Pak Presiden @prabowo kinerja Ba...
4  @Bud_Cavalera @hasyimmah Kntol...emang ada pre...

                                   text_preprocessed
0  jauh paling integritas jadi andal utama tuk be...
1  suara netizen media sosial mungkin wakil suara...
2  arti juta duduk indonesia puas kerja pak presi...
3  bagaimana pak presiden kerja bawah mau serius ...
4  kntolemang presiden indo pecintraan lebih namp...
```

Fig 4: Data preprocessing results

3.3. TF-IDF weighting

The implementation of the Term Frequency–Inverse Document Frequency (TF-IDF) method shows that the weighting process produces a matrix measuring (1074, 4047). This size indicates that there are 1,074 tweet data successfully processed with a total of 4,047 unique features (words) formed after the text preprocessing stage. Each tweet is represented in the form of a 4,047-dimensional vector, where each vector element describes the TF-IDF weight value of each word in the tweet data.

The application of Term Frequency–Inverse Document Frequency (TF-IDF) produces a number of words with the highest average weight in the data corpus. The five words with the highest weight values are “work” (0.053820), “president” (0.051157), “command” (0.036683), ‘new’ (0.032328), and “Prabowo” (0.032019). These values indicate the difference in the proportion of each word's appearance in the processed tweet collection. Each weight number describes the contribution level of each word in the numerical representation of the text, which is then used as the basis for forming feature vectors for further analysis.

```
Kata dengan Bobot TF-IDF Rata-Rata Tertinggi:
kerja        0.053820
presiden     0.051157
perintah     0.036683
baru         0.032328
prabowo      0.032019
kamu         0.028459
rakyat       0.028215
pak          0.027330
jadi         0.021520
menteri      0.019498
baik         0.018655
indonesia    0.017306
nya          0.016621
bapak        0.016329
gak          0.016181
dtype: float64
```

Fig 5: TF-IDF word weighting results

3.4. Naïve Bayes Implementation

In the initial stage, the data splitting process is carried out by separating the data set in an 80:20 ratio, where 80% of the data is used for training (training data) and the remaining 20% is used for testing (testing data). This division resulted in 859 samples for training and 215 samples for testing. With this proportional division, the model can learn patterns from most of the data while still having enough data to evaluate its generalization capabilities, allowing the sentiment analysis process to run optimally.
After the data division process, the next step is to perform class balancing using SMOTE to overcome imbalances in the training data. The SMOTE method increases the amount of training data from 859 samples to 1,542 samples through the process of oversampling the minority class. As a result, each class of negative, positive, and neutral sentiment has a balanced amount of data, namely 514 samples each, so that the model has an equal opportunity to learn each sentiment category.

In the training phase, the Multinomial Naïve Bayes model was created using 80% of the training data, which had undergone initial processing and class adjustment using SMOTE. This process allows the model to understand the patterns of each sentiment category proportionally by calculating probabilities and possibilities beforehand. After training was complete, the model was tested with 20% of the test data, which had not been used previously, to assess its generalization ability. At this stage, the model made predictions about sentiment classes based on the highest probabilities and was evaluated using the most relevant metrics. The results of the evaluation show that the accuracy of Naïve Bayes reached 71.16%, precision 74.05%, recall 71.16%, and F1-Score 72.34%, indicating that the model can classify sentiment quite well without showing signs of overfitting when applied to new data.

```
Training Model Naïve Bayes
Model Naïve Bayes berhasil dilatih dengan data hasil SMOTE.
Hasil Evaluasi Model Naïve Bayes
Metrik Kinerja Model
   Metrik  Nilai (%)
 Accuracy      71.16
Precision      74.05
   Recall      71.16
 F1-Score      72.34
Tingkat Akurasi Akhir:71.16%
```

Fig 6: Naive Bayes evaluation tesults

The classification results per class show that the negative category has the highest performance with a precision of 0.81, recall of 0.78, and F1-score of 0.80. The positive category recorded a precision of 0.77 and an F1-score of 0.72, while the neutral category showed the lowest results with a precision of 0.26 and an F1-score of 0.32. The weighted average of all metrics reached a precision of 0.74, a recall of 0.71, and an F1-score of 0.72.

```
Classification Report (Precision, Recall, F1-Score per Kelas)
              precision    recall  f1-score   support

     negatif       0.81      0.78      0.80       129
      netral       0.26      0.41      0.32        22
     positif       0.77      0.67      0.72        64

    accuracy                           0.71       215
   macro avg       0.61      0.62      0.61       215
weighted avg       0.74      0.71      0.72       215
```

Fig 7: Class-based model evaluation

The Confusion Matrix displays the results of the Naïve Bayes model performance evaluation in the sentiment classification process. The model successfully predicted 101 negatively labeled data, 9 positively labeled data, and 43 neutrally labeled data. There were classification errors in 17 negative data identified as positive, 11 negative data detected as neutral, and several positive and neutral data misclassified into other classes. Overall, the best performance was seen in the negative class with the highest number of correct predictions, while the positive class showed the lowest accuracy rate.

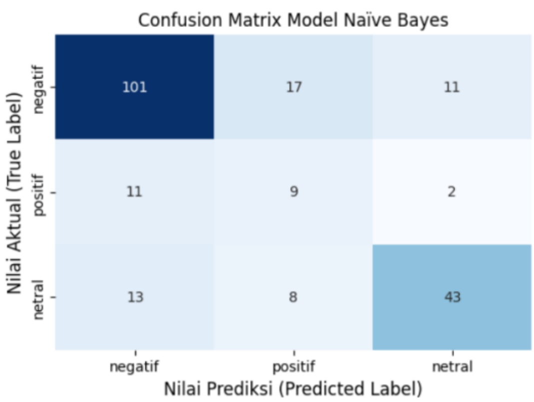

Fig 8: Confusion matrix

3.5. Word Cloud Visualization

The word cloud for each sentiment shows differences in vocabulary focus. The positive word cloud is dominated by words such as “president,” “work,” and “good,” which reflect support and appreciation. The negative word cloud contains words such as “orders,” “people,” and “ministers,” which indicate criticism of the government's performance. Meanwhile, the neutral word cloud contains words such as “work,” “president,” and “program,” which are informative without emotional content.

Fig 9: Word cloud sentiment visualization

4. Conclusion

This study shows that the application of Naïve Bayes with text pre-processing, TF-IDF, and SMOTE produces a systematic analysis flow and is able to improve data quality and classification accuracy. The classification results show that public opinion on the performance of the president-elect in 2024 is dominated by negative sentiment (59.9%), followed by positive sentiment (29.8%) and neutral sentiment (10.3%), indicating a tendency for the public to be more critical. Model evaluation shows fairly good performance with an accuracy of 71.16% and an F1 score of 72.34%, so Naïve Bayes is still considered effective and competitive for analyzing political sentiment on social media.

References

[1] R. Yao and A. Gillen, “Public opinion evaluation on social media platforms: A case study of High Speed 2 (HS2) rail infrastructure project,” *UCL Open Environ.*, vol. 5, 2023, doi: 10.14324/111.444/ucloe.000063.

[2] A. Assaefi, B. Laskarwati, and Zainurrohmah, “The Role of Social Media in Indonesia’s Political Campaigns: A New Era of Electioneering,” *Indones. Discourse*, vol. 1, no. 2, 2024, doi: 10.15294/indi.v1i2.23031.

[3] H. Fardiansyah and E. Komalawati, “Political Communication Strategy and Public Opinion for the Victory of the Prabowo Gibran Pair in the 2024 Presidential and Vice Presidential Elections of the Republic of Indonesia,” *Indones. J. Contemp. Multidiscip. Res.*, vol. 3, no. 4, pp. 675–694, 2024, doi: 10.55927/modern.v3i4.10322.

[4] R. Ackland, F. Gumbert, O. Pütz, B. Gertzel, and M. Orlikowski, “Reciprocal communication and political deliberation on Twitter,” *Soc. Sci.*, vol. 13, no. 1, p. 5, 2024, doi: 10.3390/socsci13010005.

[5] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, and Arief Wibowo, “Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier,” *J. Asiimetrik J. Ilm. Rekayasa Inov.*, vol. 4, pp. 239–248, 2022, doi: 10.35814/asiimetrik.v4i1.3577.

[6] A. Syahrir, H. Harlinda, and F. Umar, “Analisis Sentimen Masyarakat Terhadap Kebijakan Pemerintah Vaksinasi Booster 2 Menggunakan Metode Naïve Bayes Classifier,” *Bul. Sist. Inf. dan Teknol. Islam*, vol. 4, no. 4, pp. 347–359, 2023, doi: 10.33096/busiti.v4i4.1835.

[7] A. A. Putri *et al.*, “OPINI PUBLIK TERHADAP ISU KEASLIAN IJAZAH PADA PLATFORM YOUTUBE DENGAN NAÏVE BAYES , KNN , DAN,” vol. 11, no. 2, pp. 231–242, 2025.

[8] Hardani *et al.*, *Buku Metode Penelitian Kualitatif*, vol. 5, no. 1. 2020.

[9] T. Setiadi, “Analisis Jaringan Opini Publik tentang Pembatasan Sosial Berskala Besar di Twitter,” *J. Media dan Komun. Indones.*, 2020, doi: 10.22146/jmki.82650.

[10] V. D. Setiawan and D. U. Iswavigra, “Sentiment analysis to evaluate public service perception among Surakarta city residents using the BiLSTM model,” *J. Informatics Telecommun. Eng.*, vol. 9, no. 1, pp. 1–12, 2025, doi: 10.31289/jite.v9i1.15498.

[11] J. Cui, “Survey on sentiment analysis: evolution of research methods and topics,” *Artif. Intell. Rev.*, vol. 56, no. 6, pp. 5751–5784, 2023, doi: 10.1007/s10462-022-10386-z.

[12] F. H. Fadel and S. F. Behadili, “A comparative study for supervised learning algorithms to analyse sentiment tweets,” *Iraqi J. Sci.*, vol. 63, no. 6, pp. 2712–2724, 2022, doi: 10.24996/ijs.2022.63.6.36.

[13] M. Altalhan, A. Algarni, and M. T.-H. Alouane, “Imbalanced data problem in machine learning: A review,” *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3531662.

[14] M. Hayaty and A. H. Pratama, “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.

[15] K. P. Gunasekaran *et al.*, “Is text preprocessing still worth the time? A comparative study on twitter datasets,” *Sci. Rep.*, vol. 8, no. 3, p. 111, 2025, doi: https://doi.org/10.1016/j.jclepro.2023.135423.

[16] I. Arroyo-Fernández, C.-F. Méndez-Cruz, G. Sierra, J.-M. Torres-Moreno, and G. Sidorov, “Unsupervised sentence representations as word information series: Revisiting TF-IDF,” *Sci. Rep.*, vol. 13, p. 12345, 2023, doi: 10.1038/s41598-023-12345.

[17] Q. H. Nguyen *et al.*, “Influence of data splitting on performance of machine learning models in prediction of shear strength of soil,” *Math. Probl. Eng.*, vol. 2021, p. 4832864, 2021, doi: 10.1155/2021/4832864.

[18] M. Mujahid and et al., “Data oversampling and imbalanced datasets: An investigation of SMOTE and variants with social-media data,” *J. Big Data*, vol. 11, p. 30, 2024, doi: 10.1186/s40537-024-00943-4.

[19] N. A. Semary, W. Ahmed, K. Amin, P. Pławiak, and M. Hammad, “Enhancing machine-learning-based sentiment analysis through feature-extraction techniques,” *PLoS One*, vol. 19, no. 2, 2024, doi: 10.1371/journal.pone.0294968.