

# Sentiment Analysis of Social Media X in the 2024 Indonesian Presidential Election Using the Naive Bayes Algorithm: Candidates' Backgrounds and Political Promises

Santi Prayudani<sup>1\*</sup>, Dita Rouli Basa Situmorang<sup>2\*</sup>, Rizki Hidayah<sup>3\*</sup>, Heri Sanjaya Ginting<sup>4\*</sup>  
\* Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan  
[santiprayudani.polmed.ac.id](mailto:santiprayudani.polmed.ac.id)<sup>1</sup>, [ditaroulibasa@students.polmed.ac.id](mailto:ditaroulibasa@students.polmed.ac.id)<sup>2</sup>, [rizkihidayah@students.polmed.ac.id](mailto:rizkihidayah@students.polmed.ac.id)<sup>3</sup>  
[herisanjayaginting@students.polmed.ac.id](mailto:herisanjayaginting@students.polmed.ac.id)<sup>4</sup>

## Article Info

### Article history:

Received 2024-05-15  
Revised 2024-08-09  
Accepted 2024-10-04

### Keyword:

*Sentiment analysis,  
Naive Bayes,  
Social Media,  
2024 presidential election.*

## ABSTRACT

In 2024, Indonesia holds a presidential election, and the candidates are making promises to each other to attract voters. Many people gave their opinions on X. This study uses the Naïve Bayes algorithm to analyze the sentiment of these tweets, with the aim of understanding the background of the candidates and their campaign promises. Data is collected from X by crawling technique, then data is pre-processed, trained using Naïve Bayes model, and evaluated for accuracy. Sentiments in tweets were classified as positive, negative, or neutral. The results showed that the Prabowo Subianto - Gibran Rakabuming Raka pair was the most talked about with 1005 tweets, followed by Anis Rasyid Baswedan - Muhaimin Iskandar with 707 tweets, and Ganjar Pranowo - Mohammad Mahfud M.D. with 572 tweets. The Prabowo Subianto - Gibran Rakabuming Raka pair received the most positive sentiment, which was 446 more than the other candidates.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. INTRODUCTION

In modern politics, understanding public opinion and sentiment is becoming increasingly important for the success of political campaigns. In this digital era, social media provides a vast and dynamic platform for the public to express their views, including on presidential candidates. Research on public sentiment on social media not only provides insights into how the public responds to political promises but also has the potential to predict election trends. Therefore, sentiment analysis of presidential candidates and their political promises holds significant importance, both academically for understanding political dynamics and practically for developing more effective and data-driven campaign strategies.

Political campaigns are a primary means for political parties to communicate their political messages to the public, aiming to influence their thoughts, beliefs, sentiments, and perspectives toward specific political actors [1]. In the continuously evolving digital era, social media, particularly X (formerly known as Twitter), has become a key platform for disseminating opinions, support, and criticism regarding

presidential candidate pairs and their political promises [2][3]. Social media's role in political campaigns is increasingly important, as it allows candidates and political parties to reach a broader audience at a lower cost compared to traditional media, though there is a risk of bias in recommendation systems that can influence user perceptions [1][3].

Previous research has shown that the Naïve Bayes algorithm is highly effective in analyzing sentiments on social media, particularly on platforms like X, which are widely used in political contexts, including presidential elections [4]. Sentiment analysis techniques, such as decision tree methods, have been applied in Indonesian presidential elections to effectively map public opinion, providing valuable insights into how the public perceives presidential candidate pairs [5]. With its ability to classify sentiment into positive, negative, and neutral categories, Naïve Bayes is well-suited for analyzing the large volumes of text data generated during political campaigns [7][8].

This research aims to investigate how the backgrounds of presidential candidate pairs and their political promises are

perceived by the public through sentiment analysis on X. This analysis not only contributes academically to understanding how political dynamics are shaped in the digital era but also provides practical insights for candidates in developing more effective campaign strategies [9]. By focusing on public conversations on X, this study is expected to provide a clearer view of public preferences and sentiments toward presidential candidates in the 2024 election and reveal how these sentiments may influence the election outcome in Indonesia.

## II. METHOD

In this research, the methodology employed involves collecting data based on public opinions regarding presidential and vice-presidential candidates using Google Colab with Python language. The data collection process begins with authentication using the X API to gather data from X, specifying search parameters such as keywords "Anies", "Prabowo", "Ganjar", and a specific date range. The collected data is then processed using cleaning and preprocessing techniques to remove and modify irrelevant words. Subsequently, a translation process is conducted to ensure consistency in the same language so that opinion data can be classified with labeling into three sentiment categories: negative, positive, and neutral. Labeling is done based on the context of the collected opinions. Visualization is performed to display frequently occurring words visually. Following that, classification is carried out using Naive Bayes [9-18] with these three sentiment categories, aiming to classify the collected opinions into the appropriate sentiment categories.

### A. Crawling

Crawling method to gather sentiment data from the X platform using the API (Application Program Interface) Tweet Harvest version 2.5.3. We performed crawling oriented towards mentions of the presidential candidates' names in tweets. We conducted 6 crawling sessions using the respective nicknames of each presidential candidate. After the data collection, empty tweets and duplicates were removed from the dataset. As a result, we obtained a dataset consisting of 2284 rows of data. Your paper should be written in a two-column format with a space of 4.22 mm between columns.

TABEL I  
DATA SENTIMENT

No	Sentiment	Text
1	Positive	anies once turn half turn clean dead skin cells and dirt tar turn on face prabowo gibran
2	Negative	I think many do not feel that's why many claim anies sweet promise.
3	Positive	may ganjar mahfud win
4	Positive	go forward ganjar mahfud
.....	.....	.....
2089	Positive	congratulations on the spirit of the nation, mas gibran.

2123	Positive	prabowo gibran always protects his people
2189	Positive	the most correct is only prabowo gibran
2204	Positive	I believe in mas gibran, he can't lie.
2225	Positive	obviously I choose prabowo gibran so tough amazingly great

### B. Preprocessing

After collecting the sentiment dataset, we performed preprocessing on the dataset before using it as training data for the model to be created. The steps we took are as follows:

- **Cleaning:** This involves removing links, hashtags, and mentions from each sentence in the sentiment dataset.
- **Normalization:** This includes converting abbreviations that are difficult to interpret by automatic translators into full words, and combining different mentions that have the same meaning into one common word.
- **Tokenizing, Stemming, Casefolding:** This process involves manipulating the data to make it clearer and more understandable for the analysis model.
- **Translation:** We translated the dataset into English due to the unavailability of programming libraries that can be used to analyze sentiment in the Indonesian language.

### C. Training Naive Bayes Model

After preprocessing the dataset, we utilized the TextBlob library in the Python programming language to classify each data row into three categories: positive, negative, and neutral. Subsequently, we used this classified dataset as the training dataset (train set) to train the sentiment analysis model we created. We employed the Naive Bayes Classification (NBC) [19-24] classifier from the TextBlob library to be applied to the model. Next, this model will be utilized to perform sentiment analysis on the dataset of each candidate pair that we have prepared.

### D. Evaluation

Before performing classification using the Naive Bayes Classification (NBC) model that has been created, we first utilized this model to classify the dataset previously used as the training dataset. We then compared the classification results of this model with the previous classification results to calculate the accuracy of this model. The result showed that this model has an accuracy of 77%.

### E. Sentiment Analysis

Election Candidates 2024 After successfully creating the model, we then divided the dataset into 3. Each dataset contains sentiments discussing each pair of presidential candidates. If there is data discussing more than one candidate pair, then that data is included in all relevant candidate pair datasets. Subsequently, we utilized the analysis model we



Figure 5 displays the results of sentiment categorized into three categories: Positive, Negative and Neutral. The data shows that most tweets give positive responses to the presidential and vice presidential candidates Anies Rasyid Baswedan - Muhaimin Iskandar.

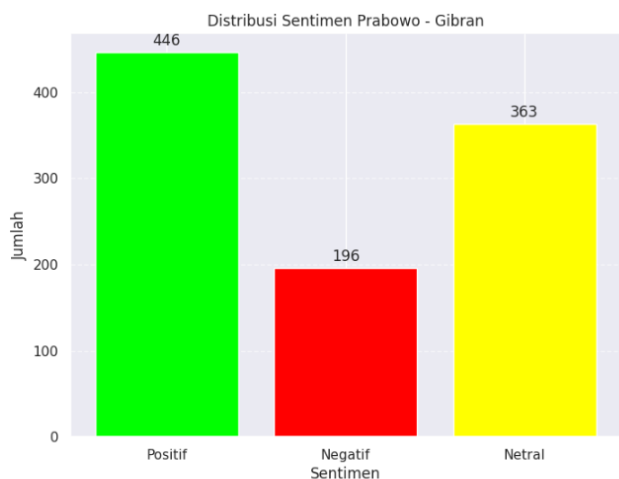


Figure 5. Diagram Sentiment Prabowo-Gibran

Figure 5 displays the results of sentiment categorized into three categories: Positive, Negative and Neutral. The data shows that most of the tweets give positive responses to the presidential and vice presidential candidates PrabowoSubianto-Gibran Rakabuming Raka.

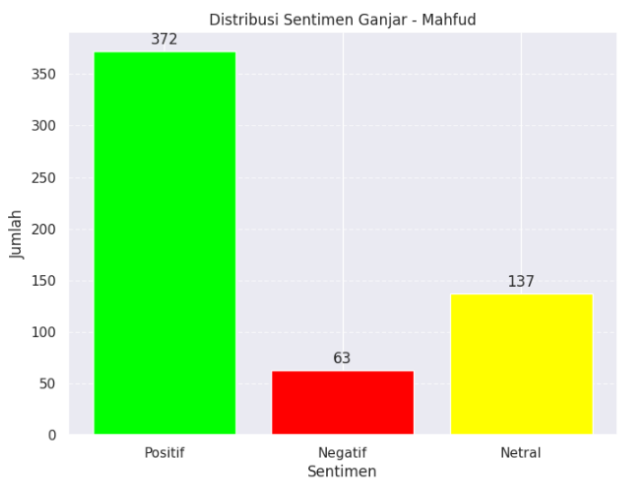


Figure 6. Diagram Sentiment Prabowo-Gibran

Figure 6 displays the sentiment results categorized into three categories: Positive, Negative and Neutral. The data shows that most of the tweets give positive responses to the presidential and vice presidential candidates Ganjar-Mahfud.

Based on the data, the presidential and vice presidential candidate pair Prabowo Subianto - Gibran Rakabuming Raka received the highest number of tweets with 1005 tweets, followed by Anies Rasyid Baswedan - Muhaimin Iskandar

with 707 tweets, and Ganjar Pranowo - Mohammad Mahfud M.D. with 572 tweets. Sentiment analysis showed that the Prabowo-Gibran pair had the highest number of positive tweets with 446 positive tweets, 196 negative tweets, and 363 neutral tweets. This was followed by the Ganjar-Mahfud pair who received 372 positive tweets, 63 negative tweets, and 137 neutral tweets, and the Anies-Muhaimin pair who received 335 positive tweets, 146 negative tweets, and 226 neutral tweets. This indicates that although Prabowo-Gibran received the most interactions and the highest number of positive tweets, the three pairs still received more positive sentiments than negative ones, reflecting their appeal and support from the public.

## V. CONCLUSION

In this study, researchers used crawling data from X social media by utilizing the Tweet Harvest library and obtained 2444 sentiment data samples during the 6-month time period before the election. This study found facts that corroborate previous research that sentiment classification using the Text Blob library in Python has higher accuracy than sentiment in Indonesian, both with and without additional classifiers.

The sentiment data samples obtained by the crawling method related to political topics are also influenced by the popularity of the figures. The significant difference in popularity between figures makes the number of data samples for each figure unbalanced so that the accuracy of the classification model is not very high. In addition, each sentiment obtained sometimes does not only discuss one character. There are also sentiments that discuss two or more figures in one tweet. These tweets contain positive sentiment towards one character and negative sentiment for another character, making it difficult for the model to classify whether this tweet includes positive or negative sentiment.

This research uses the classification results from Text Blob with a built-in classifier that distinguishes sentiment based on the word content in the sentence to label each line (entity). The labeled sentences were used as a train set as well as a validation dataset to train the model that utilizes the Naïve Bayes Classification method. We found that the accuracy of this model only ranges from 70% to 80%.

For future research, it is recommended that any lines of data (tweets) containing two or more opposing figures be removed from the dataset as it results in ambiguity. In addition, the number of positive, negative, or neutral sentiments should be balanced for each candidate pair in the training dataset to improve the accuracy of the model. Before training the model, the dataset needs to be separated into train set and validation set with the appropriate portion of the data.

## DAFTAR PUSTAKA

- [1] O. Papakyriakopoulos, J. Carlos, M. Serrano, and S. Hegelich, "Political communication on social media: A tale of hyperactive users and bias in recommender systems," 2019, doi: 10.1016/j.osnem.2019.10.
- [2] P. Al Muqsith Prasetyo and A. Hermawan, "Analisis sentimen X terhadap pemilihan presiden menggunakan algoritma Naive Bayes,"

- INFOTECH : Jurnal Informatika & Teknologi, vol. 4, no. 2, pp. 224–233, Dec. 2023, doi: 10.37373/infotech.v4i2.863.
- [3] N. G. Ramadhan, M. Wibowo, N. F. L. Mohd Rosely, and C. Quix, "Opinion mining indonesian presidential election on X data based on decision tree method," JURNAL INFOTEL, vol. 14, no. 4, pp. 243–248, Nov. 2022, doi: 10.20895/infotel.v14i4.832.
- [4] A. Hijratul Rakhmah and T. Allita Putri, "Analisis Sentimen Terhadap Pasangan Calon Presiden 2019 Pada Media Sosial X," Jakarta Gedung Sentra Kramat Jl. Kramat Raya.
- [5] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler : X."
- [6] E. Kouloumpis, T. Wilson, and J. Moore, "X Sentiment Analysis: The Good the Bad and the OMG!" [Online]. Available: [www.aaai.org](http://www.aaai.org)
- [7] T. Rosyida, H. P. Putro, and H. Wahyono, "Analisis Sentimen Terhadap Pilpres 2024 Berdasarkan Opini Dari X Menggunakan Naive Bayes dan SVM", [Online]. Available: [www.apjii.or.id](http://www.apjii.or.id)
- [8] S. Chohan, A. Nugroho, A. Maezar Bayu Aji, W. Gata, and S. Nusa Mandiri, "Analisis Sentimen Aplikasi Duolingo Menggunakan Metode Naive Bayes dan Synthetic Minority Over Sampling Technique," vol. 22, no. 2, 2020, doi: 10.31294/p.v2i2.
- [9] R. V. B. Vangara\*, K. Thirupathur, and S. P. Vangara, "Opinion Mining Classification using Naive Bayes Algorithm," International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 5, pp. 495–498, Mar. 2020, doi: 10.35940/ijitee.E2402.039520.
- [10] Y. Y. Lase, A. R. Lubis, F. Elyza, and S. A. Syafla, "Mental Health Sentiment Analysis on Social Media TikTok with the Naive Bayes Algorithm," in Proceedings - 2023 6th International Conference on Computer and Informatics Engineering: AI Trust, Risk and Security Management (AI Trism), IC2IE 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 186–191. doi: 10.1109/IC2IE60547.2023.10331126.
- [11] S. Inkoom, J. Sobanjo, A. Barbu, and X. Niu, "Pavement Crack Rating Using Machine Learning Frameworks: Partitioning, Bootstrap Forest, Boosted Trees, Naive Bayes, and K -Nearest Neighbors," Journal of Transportation Engineering, Part B: Pavements, vol. 145, no. 3, p. 04019031, Sep. 2019, doi: 10.1061/jpeodx.0000126.
- [12] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media X with case of Anti-LGBT campaign in Indonesia using Naive Bayes, decision tree, and random forest algorithm," in Procedia Computer Science, Elsevier B.V., 2019, pp. 765–772. doi: 10.1016/j.procs.2019.11.181.
- [13] A. F. Hidayatullah, S. Cahyaningtyas, and A. M. Hakim, "Sentiment Analysis on X using Neural Network: Indonesian Presidential Election 2019 Dataset," IOP Conf Ser Mater Sci Eng, vol. 1077, no. 1, p. 012001, Feb. 2021, doi: 10.1088/1757-899x/1077/1/012001.
- [14] J. Hartmann, M. Heitmann, C. Siebert, and C. Schamp, "More than a Feeling: Accuracy and Application of Sentiment Analysis," International Journal of Research in Marketing, vol. 40, no. 1, pp. 75–87, Mar. 2023, doi: 10.1016/j.ijresmar.2022.05.005.
- [15] K. Chakraborty, S. Bhattacharyya, and R. Bag, "A Survey of Sentiment Analysis from Social Media Data," IEEE Trans Comput Soc Syst, vol. 7, no. 2, pp. 450–464, Apr. 2020, doi: 10.1109/TCSS.2019.2956957.
- [16] A. L. Hananto, A. P. Nardilasari, A. Fauzi, A. Hananto, B. Priyatna, and A. Y. Rahman, "International Journal of Intelligent Systems And Applications In Engineering Best Algorithm in Sentiment Analysis of Presidential Election in Indonesia on X." [Online]. Available: [www.ijisae.org](http://www.ijisae.org)
- [17] R. G. Bhati, "A Survey On Sentiment Analysis Algorithms And Datasets," Review of Computer Engineering Research, vol. 6, no. 2, pp. 84–91, Sep. 2019, doi: 10.18488/journal.76.2019.62.84.91.
- [18] K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," Applied Sciences (Switzerland), vol. 13, no. 7, MDPI, Apr. 01, 2023. doi: 10.3390/app13074550.
- [19] E. Prabowo, D. Rahmat Hidayat, D. Sugiana, and B. Aly, "Political Socialisation and Political Communication in Delivering Political Education and Being Community Aspiration Absorber in Indonesia." [Online]. Available: [www.ijicc.net](http://www.ijicc.net)
- [20] O. Papakyriakopoulos, J. Carlos, M. Serrano, and S. Hegelich, "Political communication on social media: A tale of hyperactive users and bias in recommender systems," 2019, doi: 10.1016/j.osnem.2019.10.
- [21] H. Fitra, S. Dosen, J. Jurnalistik, F. Dakwah, and D. Komunikasi, "Pengaruh Dan Efektivitas Penggunaan Media Sosial Sebagai Bentuk Saluran Komunikasi (Haidir Fitra Siagian) Pengaruh Dan Efektivitas Penggunaan Media Sosial Sebagai Saluran Komunikasi Politik Dalam Membentuk Opini Publik."
- [22] P. Basile, V. Basile, M. Nissim, N. Novielli, and V. Patti, "Sentiment Analysis of Microblogging Data," in Encyclopedia of Social Network Analysis and Mining, Springer New York, 2017, pp. 1–17. doi: 10.1007/978-1-4614-7163-9\_110168-1.
- [23] K. Orkphol and W. Yang, "Sentiment Analysis on Microblogging with K-Means Clustering and Artificial Bee Colony," Int J Comput Intell Appl, vol. 18, no. 3, Sep. 2019, doi: 10.1142/S1469026819500172.
- [24] M. Iqbal, "Social Network Analysis: Public Trust And Digital Movement Of The Covid-19 Era In Indonesia," 2022.