Sentiment Analysis of DKI Jakarta 2024 Election (Case Study: Anies Baswedan and Ridwan Kamil)

Muhammad Safrul Safrudin[*1], Syarifah Aini[2]
[1]Politeknik Negeri Ambon Lokasi Banda
[2]Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Muhammadiyah Palembang
Email: [1]safrul34@gmail.com, [2]ainiuzumakiandriyos@gmail.com

Abstrak. Analisis Sentimen Twitter Pilkada DKI Jakarta 2024 (Studi Kasus Anies Baswedan dan Ridwan Kamil). *Penelitian ini menganalisis sentimen publik terhadap dua calon potensial Pilkada DKI Jakarta 2024, Anies Baswedan dan Ridwan Kamil, menggunakan data Twitter. Dengan menerapkan model TextBlob untuk ekstraksi teks dan Naive Bayes untuk klasifikasi sentimen, ditemukan bahwa sentimen terhadap Anies Baswedan sebagian besar positif 52,2%, sementara sentimen netral mendominasi terkait Ridwan Kamil. Akurasi model Naive Bayes mencapai 80% untuk Anies Baswedan dan 72% untuk Ridwan Kamil, dengan precision, recall, dan F1-score masing-masing lebih tinggi pada data Anies. Hasil ini menunjukkan bahwa model lebih akurat dalam mengklasifikasikan sentimen terhadap Anies dibandingkan Ridwan Kamil. Implikasi dari temuan ini penting bagi strategi kampanye politik, di mana Anies dapat memanfaatkan dukungan positif yang ada, sedangkan Ridwan Kamil memiliki peluang untuk memperkuat keterlibatan publik melalui pendekatan yang lebih strategis, sesuai dengan sentimen yang muncul di media sosial.*
Kata Kunci: *analisis sentimen, Anies Baswedan, Jakarta, Pilkada, Ridwan Kamil*

Abstract. *This study analyzes public sentiment toward two potential candidates for the 2024 Jakarta gubernatorial election, Anies Baswedan and Ridwan Kamil, using Twitter data. Applying the TextBlob model for text extraction and Naive Bayes for sentiment classification found that sentiment toward Anies Baswedan is mostly positive, 52.2%, while neutral sentiment dominates for Ridwan Kamil. The accuracy of the Naive Bayes model reached 80% for Anies Baswedan and 72% for Ridwan Kamil, with higher precision, recall, and F1-score for Anies' data. These results indicate that the model is more accurate in classifying sentiment toward Anies compared to Ridwan Kamil. The implications of these findings are important for political campaign strategies, where Anies can leverage the existing positive support, while Ridwan Kamil has an opportunity to strengthen public engagement through a more strategic approach, in line with the sentiment emerging on social media.*
Keywords: *Anies Baswedan, Jakarta, regional head election, Ridwan Kamil, sentiment analysis*

1. Introduction

Regional head elections are one of the main pillars of the democratic process in Indonesia, including in DKI Jakarta, the country's capital city. DKI Jakarta elections are always in the spotlight, not only because of the importance of the governor's role in leading this metropolitan city but also because of its significant influence on national political dynamics. In the 2024 DKI Jakarta elections, two names that are often mentioned as potential candidates are Anies Baswedan and Ridwan Kamil. According to the Kompas survey on June 15-20, 2024, Anies Baswedan has an electability of 29.8% and Ridwan Kamil of 8.6%. Both of these figures have strong track records in government and politics, so they are often the topic of discussion in various media, including social media.

Social media, especially Twitter, has become one of the main platforms for people to express their opinions on various issues, including politics. With hundreds of millions of active users, Twitter provides rich data to analyze how the public views and sentiments toward potential leaders. Twitter-based sentiment analysis can provide valuable insights into people's political preferences, opinions towards certain candidates, and issues that are considered important by the

public. Research [1] shows that sentiment analysis on social media has become an important tool in predicting the outcome of the 2019-2024 Indonesian Presidential election.

In the context of the 2024 DKI Jakarta elections, Twitter sentiment analysis can be used to understand public perceptions of Anies Baswedan and Ridwan Kamil before they officially register as candidates for governor. In addition, this analysis can also provide an overview of the desires of the people of Jakarta for other potential candidates who may not have surfaced.

Research [2] conducted sentiment analysis on election-related Twitter data. The results of this study indicate that the Naive Bayes algorithm obtains good accuracy and is suitable for data sentiment analysis on the Twitter platform. Research [3] conducted sentiment analysis related to predicting the results of the 2024 presidential election using Twitter data taken in 2023. The results of this study show that the Naive Bayes algorithm has good performance when used on small amounts of data. The speed and efficiency of Naive Bayes make it possible to process and analyze data quickly, thereby reducing processing time [4]. According to [5], Naive Bayes does not require large training data to function properly. The algorithm can achieve good performance with relatively small datasets, making it suitable for use on limited-label data.

Research [6] conducted sentiment analysis related to tweets about COVID-19 using TextBlob and classification with machine learning. Textblob is a Python library built with the Natural Language Toolkit (NLTK) library and patterns. In sentiment analysis, TextBlob is used for basic modeling in sentiment analysis before further modeling is applied with machine learning classification. TextBlob is used because of its simple and efficient model for handling small data.

In this study, sentiment analysis was conducted on tweets related to the 2024 DKI Jakarta Pilkada, with a focus on Anies Baswedan and Ridwan Kamil. This research uses two approaches: TextBlob for feature extraction and the Naive Bayes algorithm for classification. This study aims to analyze and identify public sentiment towards potential candidates for the 2024 DKI Jakarta Pilkada, especially Anies Baswedan and Ridwan Kamil, based on tweet data taken from Twitter. In addition, this study aims to evaluate voters' preferences towards different candidates by classifying sentiments into positive, negative, and neutral, as well as identifying key issues of public concern.

2. Literature Review

Twitter is unique with its character limit per tweet (280 characters), the use of hashtags, and intensive interaction. These characteristics allow for more specific and real-time sentiment analysis. Data from Twitter is often larger and more volatile due to real-time user activity, allowing for more accurate and quick sentiment analysis. Sentiment on Twitter tends to be more reactive and spontaneous, as users can interact directly and quickly. Research shows that neutral sentiment is most dominant on Twitter, followed by positive and negative sentiment [7].

TextBlob does not require complicated model training, so sentiment analysis can be done quickly. This makes it ideal for real-time analysis or when computing resources are limited. TextBlob can also be easily integrated with other Python tools and libraries, enabling the development of comprehensive analysis pipelines [8]. Research [9] focuses on sentiment analysis of tweets related to PT Telkom Indonesia's products and services. This research uses the TextBlob library to classify tweet sentiment into positive, neutral, and negative categories. The results of sentiment analysis are then visualized in the form of histograms, pie charts, and word clouds to provide an overview of the distribution and sentiment patterns in the tweet data. This research shows that TextBlob can be used effectively for tweet sentiment analysis with an accuracy of 77.2%.

The naive Bayes algorithm is used in various sentiment applications and has proven effective in classifying positive and negative sentiments [10]. The naive Bayes Classifier is used to classify the tweet sentiment into positive, negative, or neutral classes [11]. Research [12] developed a web-based text classification service using the Naive Bayes algorithm for the Indonesian language. This research successfully achieved an accuracy of up to 83.75%, showing that the simple yet effective Naive Bayes method can process text data with fast computation.

3. Method

This research was conducted in several stages, namely: data collection; data pre-processing, which involved two steps: (1) data cleaning and (2) data normalization; feature extraction with three processes: (1) stopwords removal, (2) tokenization, and (3) stemming; followed by modeling and model evaluation as in Figure 1.

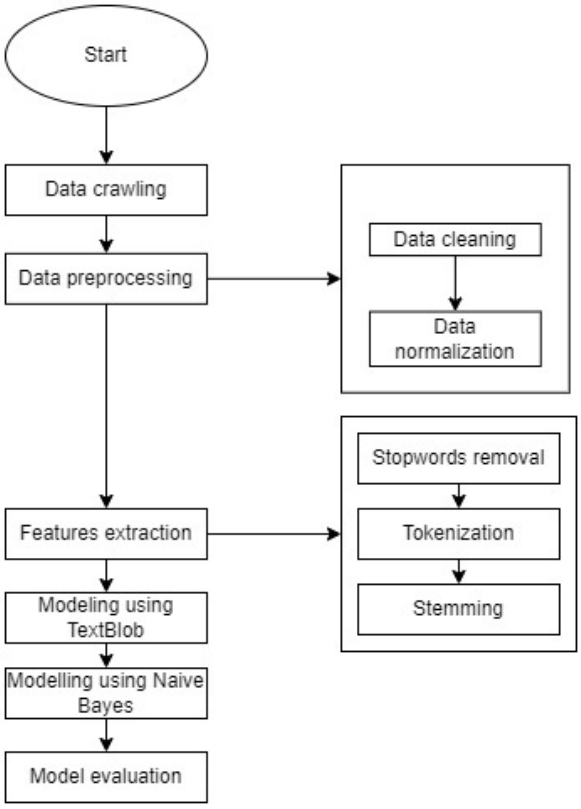

Figure 1. Research Stages

3. 1 Data Crawling

The first step in this process is to collect relevant data from Twitter using Python. In this research, Tweet Harvest is used in the node.js package in Python to crawl data from Twitter. This process requires token authentication from the Twitter API. This token is obtained from a registered Twitter developer account. The data crawling method used in this research cannot ensure the exact amount of data is retrieved, because the system limits it. However, researchers can set a maximum limit for the desired data. In this research, the keywords used are "Anies fav: 100 lang: id", which means that only Indonesian tweets with the word 'Anies' of at least 100 likes and Indonesian language are retrieved. The same thing is applied to the keyword "Ridwan Kamil". The maximum limit set on both data is 500 tweets, but in the crawling process, only 140 tweets were obtained in 10 minutes. The Twitter data retrieved is data from July 26 - August 26, 2024. Data that is successfully obtained at this stage is then processed to the next stage. Table 1 and Table 2 are examples of tweets taken from Twitter for Anies and Ridwan Kamil.

Table 1. Tweet for Anies Baswedan

Tweet Anies
Pemenang sesungguhnya di perpolitikan Indonesia itu Mas @aniesbaswedan. Pak Jokowi bisa ngatur semua lini negara ini bahkan untuk ambisi kepentingan dinastinya wajar dia Presiden. Sedangkan Mas Anies ngga punya partai bahkan ngga masuk partai mana pun karena mungkin alasannya

Table 2. Tweet for Ridwan Kamil

Tweet
@ridwankamil Anda sudah Mengejek Persija Lalu Tetiba Anda Memuji Persija Hanya Karena Jabatan Gurbernur Jakarta Apakah anda sehat ???? https://t.co/K8j2NjqGNG

3.2 Data Cleaning

After the data is collected, the next step is to clean the data by removing duplicate data. Reviews in tweet features still contain sentences, symbols, and characters that are not needed in this study. The data cleansing process, or data cleansing, is performed to remove irrelevant symbols and characters from the text or data [13]. The data cleansing process involves removing elements that are irrelevant or may interfere with the analysis, such as URLs, hashtags (#), mentions (@username), symbols, numbers, and special characters. In addition, tweet data is also converted to lowercase to ensure consistency. Table 3 is an example of data cleaning results. Table 3 shows that after the data cleaning process was performed, the character "@" in the word "@aniesbaswedan" was deleted. After the data is cleaned, it is processed for the next stage.

Table 3. Cleansing data

Tweet Anies	Cleansing text
Pemenang sesungguhnya di perpolitikan Indonesia itu Mas @aniesbaswedan. Pak Jokowi bisa ngatur semua lini negara ini bahkan untuk ambisi kepentingan dinastinya wajar dia Presiden. Sedangkan Mas Anies ngga punya partai bahkan ngga masuk partai mana pun karena mungkin alasannya	Pemenang sesungguhnya di perpolitikan Indonesia itu Mas Pak Jokowi bisa ngatur semua lini negara ini bahkan untuk ambisi kepentingan dinastinya wajar dia Presiden. Sedangkan Mas Anies ngga punya partai bahkan ngga masuk partai mana pun karena mungkin alasannya

3.3 Data Normalization

Normalization is the process of converting all text into a uniform format. In Table 3, words such as “ngga” (no) are used as a short and informal form that reflects spoken language. Using slang in tweets or short messages makes communication more intimate and expressive and shows the identity of a particular group or generation. However, in the context of sentiment analysis or formal research, the use of slang can cause ambiguity or inaccuracies in understanding the text, so it is necessary to perform normalization to remove it. By using normalization, slang words or abbreviations are replaced with their formal form. For example, in Table 3 there is the word “ngga” which is informal language. Table 4 shows that the word “ngga” is normalized to become the formal word “tidak”.

Table 4. Cleaning up slang words

Tweet Anies
Pemenang sesungguhnya di perpolitikan Indonesia itu Mas Anies Baswedan. Pak Jokowi bisa mengatur semua lini negara ini bahkan untuk ambisi kepentingan dinastinya, wajar dia Presiden. Sedangkan Mas Anies tidak punya partai bahkan tidak masuk partai mana pun karena mungkin alasannya bersih.

3.4 Stopwords Removal

Common words that appear frequently but do not carry significant meaning are removed from the text at this stage. Words such as “dan”, “atau”, “di”, “itu”, “tidak” and the like are known as stopwords. Removing stopwords helps reduce the number of unimportant features, thus improving the performance of the model by making it focus on words that are more relevant for sentiment analysis [14]. Stopwords removal can be done using a predefined list of stopwords. Examples of stopword removal results are presented in Table 5.

Table 5. Stopwords removal

Tweet Anies
Pemenang sesungguhnya perpolitikan Indonesia Mas Anies Baswedan. Pak Jokowi mengatur lini negara ambisi kepentingan dinastinya, wajar Presiden. Sedangkan Mas Anies punya partai masuk partai pun mungkin alasannya bersih

3.5 Tokenization

Tokenization is the process of breaking down a whole string of characters into pieces of words, with the aim of dividing the text, which can be a sentence, paragraph, or document, into small units called tokens [15]. For example, the sentence "Pemenang sesungguhnya perpolitikan Indonesia Mas Anies Baswedan" will be broken down into ["Pemenang", "sesungguhnya", "perpolitikan", "Indonesia", "Mas", "Anies", "Baswedan"]. Tokenization is important because it allows the model to analyze text at the word level, which is the basic unit in text analysis. An example of the results of the tokenization stage can be seen in Table 6.

Table 6. Tokenization

Tweet Anies
['Pemenang', 'sesungguhnya', 'perpolitikan', 'Indonesia', 'Mas', 'Anies', 'Baswedan', 'Pak', 'Jokowi', 'mengatur', 'lini', 'negara', 'ambisi', 'kepentingan', 'dinastinya', 'wajar', 'Presiden', 'Sedangkan', 'Mas', 'Anies', 'punya', 'partai', 'masuk', 'partai', 'pun', 'mungkin', 'alasannya', 'bersih']

3.6 Stemming

Stemming is the process of converting words to their base or root form. For example, words such as "pemenang", "sesungguhnya", and "sedangkan" will be trimmed to "menang", "sungguh", and "sedang". This process helps in reducing the variety of words that have similar meanings. By using stemming, the dimensionality of the data can be reduced, and the model can more easily recognize patterns in the data because word variations have been minimized [16]. Table 7 is an example of the results of stemming. The results of this stemming are then continued to the next stage.

Table 7. Stemming

Tweet Anies
['menang', 'sungguh', 'politik', 'indonesia', 'mas', 'anies', 'baswedan', 'pak', 'jokowi', 'atur', 'lini', 'negara', 'ambisi', 'penting', 'dinasti', 'wajar', 'presiden', 'sedang', 'mas', 'anies', 'punya', 'partai', 'masuk', 'partai', 'pun', 'mungkin', 'alasan', 'bersih']

3.7 TextBlob Model

In this research, the TextBlob model is used as one of the important stages in the sentiment analysis process. After the pre-processing and feature extraction are complete, the text data that has been tokenized and normalized is then entered into the TextBlob model. TextBlob is a deep learning-based model designed to understand the context in text and generate an information-rich vector representation, for each word or phrase. This vector representation is generated by converting each word or phrase in the text into a numeric vector so that the Naive Bayes classification algorithm can further process it. This conversion of words into vectors is called an embedding technique. The embedding techniques used are Word2Vec or GloVe, resulting in vectors representing words' semantic meaning. The TextBlob model has the ability to capture long-term dependencies in text, which is very important in sentiment analysis as context can affect the meaning of words. In this study, the TextBlob model was developed using the library textblob in Phyton.

3.8 Naive Bayes

The final step in this process is to use the Naive Bayes algorithm for sentiment classification. This model uses word frequencies and assumes that documents are the result of independent trials on random variables that generate words with certain probabilities [17]. In the context of sentiment analysis, Naive Bayes uses features extracted from the TextBlob model to predict whether the sentiment in a tweet is positive, negative, or neutral. The model calculates the probability of each sentiment category based on the distribution of words in the text and selects the category with the highest probability as the prediction. In this study, the Naive Bayes Classifier library in Python is used to develop model classification.

4. Results and Discussion

4.1 TextBlob

From the result of the textblob, we can see the most frequently occurring words in the data. Figure 2 is the word cloud of Ridwan Kamil. From Figure 2 we can see that the words such as “Ridwan”, “maju”, and “pilkada” are some of the most frequently occurring words in the data. This result shows that the discussion on Twitter around the 2024 Jakarta elections is highly centered on candidate Ridwan Kamil, and there is a significant sentiment of support for him. The result also appears the same in Anies Baswedan’s data. Based on the word cloud results in Figure 3, we can see words such as “Anies”, “Jakarta”, “pilkada”, and “dukung” appear as some of the most frequently mentioned words in the data. This indicates that discussions on Twitter regarding the 2024 Jakarta elections are heavily focused on candidate Anies Baswedan, with an indication of considerable.

Figure 2. Word cloud of Ridwan Kamil

Figure 3. Word cloud of Anies Baswedan

4.2 Naive Bayes

The results of sentiment classification using Naive Bayes on Twitter data in the 2024 DKI Jakarta gubernatorial election for two candidates, namely Ridwan Kamil and Anies Baswedan, provide interesting insights into public perceptions of the two figures. The classification results obtained show a significant difference in sentiment distribution between the two candidates. From Figure 4, we can see that for Ridwan Kamil, the sentiment distribution shows that the majority of the analyzed tweets revealed a predominance of neutral sentiments, with 73 neutral tweets, followed by 55 positive tweets and a relatively small number of negative tweets, totaling eight. This finding indicates that while Ridwan Kamil has a positive supporter base, most of the tweets attributed to him are more neutral. This could be due to several factors, such as a less polarizing campaign style on social media or the lack of controversial issues associated with him in the public eye.

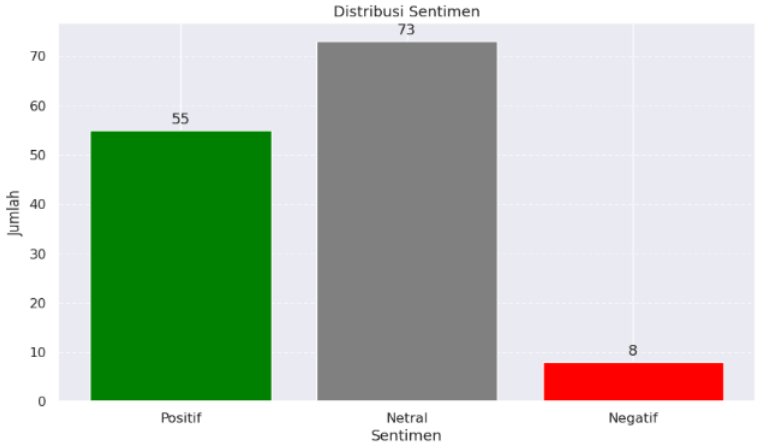

Figure 4. Sentiment Ridwan Kamil

In contrast, the results for Anies Baswedan show a more polarized proportion of sentiment. From Figure 5, we can see the majority of tweets for Anies Baswedan were positive (71 tweets), but there were more negative tweets than Ridwan Kamil (18 tweets) and fewer neutral tweets (47 tweets). This suggests that public perceptions of Anies Baswedan are more divided, while many provide positive support, there is also a significant proportion that express negative sentiments. This may be related to Anies' political track record or the issues he has raised.

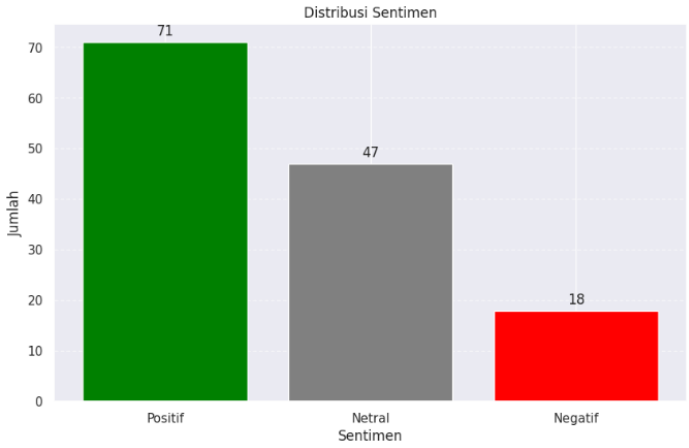

Figure 5. Sentiment Anies Baswedan

4.3. Model Evaluation

From the model evaluation results (Table 8), it can be seen that the performance of the model for Anies Baswedan is better than that of Ridwan Kamil, with an accuracy of 80% compared to 72%. The higher precision and recall rates for Anies Baswedan indicate that the model can more accurately recognize sentiment patterns in tweets related to him. Meanwhile, the performance of Ridwan Kamil shows that the model tends to make slightly more classification errors, especially in distinguishing between positive and neutral sentiments. This may be related to the distribution of data dominated by neutral sentiment, so the model has difficulty distinguishing more specific sentiment categories.

Table 8. Accuracy table

Governor Candidate	Precision	Recall	F1-score	Accuracy
Anies Baswedan	0.75	0.80	0.77	80%
Ridwan Kamil	0.74	0.70	0.72	72%

The difference in model performance may also be due to the characteristics of the data taken from Twitter, where users tend to be more vocal and assertive in expressing sentiments towards controversial or often talked-about figures. In this case, Anies Baswedan may have triggered more strong discussions, both positive and negative, thus providing more information for the Naive Bayes model to learn the sentiment pattern. On the other hand, the neutrality that dominates Ridwan Kamil's sentiment may make the data more homogenous and provide less of a strong signal for the model to predict sentiment correctly.

From the perspective of political implications, these results suggest that Ridwan Kamil may need to further activate campaign strategies that can evoke stronger positive responses from the public, given the high number of neutral sentiments. Meanwhile, while Anies Baswedan already has strong positive support, he must also address significant negative sentiment to minimize any adverse impact on his political campaign.

5. Conclusion

This study provides a comprehensive view of how public sentiment on social media can be mapped and analyzed, and how the results can inform more effective political and communication strategies. The use of the TextBlob model for text extraction and Naive Bayes for sentiment classification in this study shows that the majority of sentiment towards Anies Baswedan is positive (52.2%), while Ridwan Kamil is more dominant in neutral sentiment. The Naive Bayes model achieved classification accuracy of 80% for Anies Baswedan and 72% for Ridwan Kamil. For Anies Baswedan, the precision, recall, and F1-score values were 0.75, 0.80, and 0.77. In comparison, Ridwan Kamil achieved precision, recall, and F1-score values of 0.74, 0.70, and 0.72. The differences in sentiment distribution and model performance between the two candidates highlight the importance of considering data characteristics in sentiment analysis and the limitations of the model in handling unbalanced data distribution. Further evaluation with more complex models or data enrichment approaches can be conducted to improve sentiment prediction accuracy in future research.

References

[1] D. A. Kristiyanti, Normah, and A. H. Umam, "Prediction of Indonesia presidential election results for the 2019–2024 period using Twitter sentiment analysis," in *Proceedings of the 2019 5th International Conference on New Media Studies (CONMEDIA)*, 2019, pp. 36–42, doi: 10.1109/CONMEDIA46929.2019.8981823.

[2] P. Gaur, S. Vashistha, and P. Jha, "Twitter sentiment analysis using Naive Bayes-based machine learning technique," in *Sentiment Analysis and Deep Learning*, Singapore, 2023, pp. 367–376.

[3] A. A. Firdaus, A. Yudhana, and I. Riadi, "Prediction of presidential election results using sentiment analysis with pre and post candidate registration data," *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 10, no. 1, pp. 36–46, 2024.

[4] E. Hasibuan and E. A. Heriyanto, "Analisis sentimen pada ulasan aplikasi Amazon Shopping di Google Play Store menggunakan Naive Bayes classifier," *Jurnal Teknik dan Science*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:255367773

[5] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment analysis of review datasets using Naive Bayes and K-NN classifier," *International Journal of Information Engineering and Electronic Business*, vol. 8, no. 4, pp. 54–62, Jul. 2016, doi: 10.5815/ijieeb.2016.04.07.

[6] P. Kathiravan, R. Saranya, and S. Sekar, "Sentiment analysis of COVID-19 tweets using TextBlob and machine learning classifiers," in *Proceedings of the International Conference on Data Science and Applications*, Singapore, 2023, pp. 89–106.

[7] F. F. Mailoa, "Analisis sentimen data Twitter menggunakan metode text mining tentang masalah obesitas di Indonesia," *Journal of Information Systems for Public Health*, vol. 6, no. 1, p. 44, 2021, doi: 10.22146/jisph.44455.

[8] B. S. Shekhawat, "Sentiment Classification of Current Public Opinion on BREXIT: Naïve Bayes Classifier Model vs Python's TextBlob Approach," MSc thesis, School of Computing, National College of Ireland, Dublin, Ireland, 2019.

[9] I. G. S. Mas Diyasa, N. M. I. M. Mandenni, M. I. Fachrurrozi, S. I. Pradika, K. R. N. Manab, and N. R. Sasmita, "Twitter sentiment analysis as an evaluation and service base on Python TextBlob," *IOP Conference Series: Materials Science and Engineering*, vol. 1125, no. 1, p. 012034, May 2021, doi: 10.1088/1757-899X/1125/1/012034.

[10] S. Paudel, P. W. C. Prasad, A. Alsadoon, M. D. R. Islam, and A. Elchouemi, "Feature selection approach for Twitter sentiment analysis and text classification based on chi-square and Naïve Bayes," in *International Conference on Applications and Techniques in Cyber Security and Intelligence (ATCI 2018)*, J. Abawajy, K.-K. R. Choo, R. Islam, Z. Xu, and M. Atiquzzaman, Eds. Cham: Springer International Publishing, 2019, pp. 281–298.

[11] A. Goel, J. Gautam, and S. Kumar, "Real-time sentiment analysis of tweets using Naive Bayes," *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pp. 257–261, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:17961386.

[12] E. Setiani and W. Ce, "Text classification services using Naive Bayes for Bahasa Indonesia," in *Proceedings of the 2018 International Conference on Information Management and Technology (ICIMTech)*, 2018, doi: 10.1109/ICIMTech.2018.8528258.

[13] S. Zainudin, "Sentiment analysis techniques in recent works," *Science and Innovation Advances Conference*, 2015, doi: 10.1109/SAI.2015.7237157.

[14] S. Sarica and J. Luo, "Stopwords in technical language processing," *Public Library of Science ONE (PLOS ONE)*, vol. 16, no. 8, Aug. 2021, doi: 10.1371/journal.pone.0254937.

[15] A. Q. Surbakti, R. Hayami, and J. Al Amien, "Analisa tanggapan terhadap PSBB di Indonesia dengan algoritma decision tree pada Twitter," *Jurnal Computer Science and Information Technology)*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:246335901.

[16] A. A. Siswandi, Y. Permana, and A. Emarilis, "Stemming analysis Indonesian language news text with Porter algorithm," *Journal of Physics: Conference Series*, vol. 1845, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233802494.

[17] K.-M. Schneider, "Techniques for improving the performance of Naive Bayes for text classification," in *Computational Linguistics and Intelligent Text Processing*, A. Gelbukh, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 682–693.