SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

Sentiment Analysis on Twitter Using Naïve Bayes and Logistic Regression for the 2024 Presidential Election

Alisya Mutia Mantika[1], Agung Triayudi[2]*, Rima Tamara Aldisa[3]

Informatics Study Program, Faculty of Communication and Information Technology, Universitas Nasional, Jakarta, Indonesia

Author Email: alisya.mantika@gmail.com[1], agungtriayudi@civitas.unas.ac.id[2*], rima.tamara@civitas.unas.ac.id[3]

Abstract. In accordance with the notion of democracy which is the basis of the state of Indonesia, general elections will be held in 2024. In the implementation of the General Election there is a campaign to lead the public vote to choose the best candidate according to public opinion. Twitter social media is one of the media to voice opinions as well as share information to become one of the indirect campaigning platforms. Social media also does not escape negative issues, community rumors, and even the digital footprint of presidential candidates which can be a very important consideration in campaigning. This research aims to see the public's response to the 2024 presidential candidates. This research is conducted based on public opinion on presidential candidates, then public opinion data taken from Twitter social media will go through a pre-processing process to clean the data before the data is classified into Naive Bayes and Linear Regression modeling. The two classification models are then sought for the highest performance accuracy value and confusion matrix with 80:20 splitting data. The results showed that the Naive Bayes classification model had a higher accuracy value than the Logistic Regression classification model, which was 63% for Anies Baswedan candidate, 77% for Ganjar Pranowo candidate, and 44% for Prabowo Subianto. The highest accuracy value was obtained by the sentiment data of 2024 presidential candidate Ganjar Pranowo, which was 77%.

Keywords: Sentiment Analysis, 2024 Presidential Election, Logistic Regression, Naive Bayes, Twitter

1 Introduction

The 1945 Constitution's Article 1 paragraph 2—which declares that "Sovereignty is in the hands of the people and shall be exercised according to the Constitution"—designates Indonesia as a democratic nation. One instance of how democracy is implemented by the voice or choice of the people who become the determinants through simultaneous voting is the election of the President and Vice President, which takes place every five years [1].

In organizing general elections, there are campaigning activities to attract the attention of the public so that at the time of the general election the public vote will win or democratically favor 1 candidate with the most votes. The election of presidential candidates will certainly consider opinions and responses from the public on every aspect such as ideas, vision and mission, programs to be implemented, problems to be solved, and others. From this opinion, a survey will be created through the popularity or tendency of the community to create pros and cons related to presidential and vice-presidential candidates which become a reference in choosing the right candidate.

One topic that is currently often discussed is the issue of Presidential elections, both about politics and the activities of its candidates. Twitter social media as a universal opinion data provider and also a means to be actively involved in democracy and support the presidential candidate of choice. [2]. Because Twitter is a media that accommodates the aspirations or opinions of the community, it is not uncommon for tweets or content to be negative or contain diatribes, curses, and harsh words.[3][4][5].

Sentiment analysis is the process of classifying or analyzing the opinions, sentiments and emotions of individuals expressed in the form of text. Sentiment analysis is carried out to determine whether the text tends to have a negative or positive connotation.[6]. Typically, sentiment analysis is used for political, government, education, business, and other purposes. [7].

The algorithms used in this research are Naïve Bayes algorithm and Logistic Regression. The Naïve Bayes algorithm serves to calculate the probability of a text or document in each sentiment category assuming the words in the processed text are independent, while the logistic regression algorithm serves to predict the probability of text against its sentiment class. [8].

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

Naïve Bayes algorithm is a classification model that calculates the probability of the sum of frequencies and combinations of values from a given dataset. [9]. Naïve Bayes algorithm is an algorithm that excels in efficiency, speed, and accuracy which tends to be higher than other algorithms [10].

Logistic regression is one of the classification algorithms in machine learning used to predict the probability of categorical dependent variables. This method is a generalized form of linear regression used to study the relationship of multiple variables with binary or probabilistic variables. [11]. In this study, logistic regression classifies data into 3 classes, namely positive, neutral, and negative using multinomial logistic regression.

2 Methodology

This research uses data from crawling social media twitter with the keyword name of the Indonesian presidential candidate in the 2024 Election. then the tweets data obtained are processed again at the preprocessing stage to clean the data so that it is easy to process. Furthermore, the preprocessed data will be classified with Naive Bayes modeling and Logistic Regression. The accuracy value obtained determines which algorithm is superior in the sentiment data classifier.

2.1 Data Collection

Data collection or data crawling is the process of retrieving data by scrapping social media. This research uses the tweet harvest library in retrieving twitter tweets data with hashtags according to the names of Indonesian presidential candidates in the 2024 election. In retrieving or crawling twitter data, this research uses the tweet harvest library via Google Collaboratory as a cloud computing platform similar to Jupyter Notebook. For crawling data on twitter, an authentication token is also required. Authentication tokens are useful for accessing our twitter so that tweet harvest can crawl data through a personal twitter account.

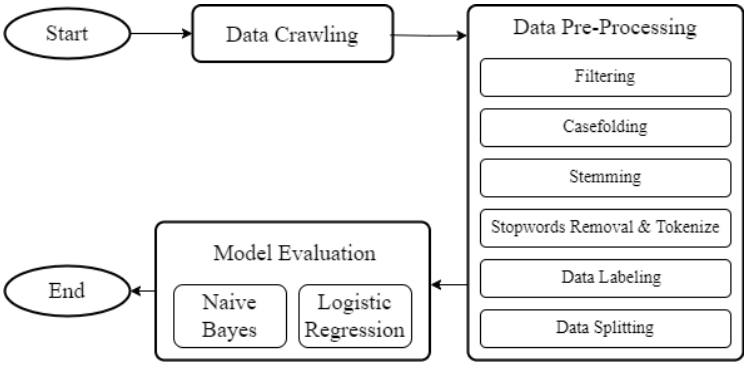

Figure 1. Process Design Flow

2.2 Data Preprocessing

Data preprocessing is the process of cleaning up the initially collected, unprocessed data from the crawl (known as "Raw Data"). The dataset used in this study was initially collected in the original format and underwent the following preprocessing steps:

1. Filtering, is the process of cleaning data from Twitter features such as mentions, urls, emojis, and other characters.
2. Casefolding, is the process of converting capital letters into lowercases, removing punctuation and removing extra whitespace. [10].
3. Stemming, is the process of converting words that have affixes into basic words.
4. Stopword removal, is the process of eliminating words from a sentence that have no sentimental significance. [11].
5. Tokenizing, is converting a text into word fragments. [12].
6. Data Labeling, is the process of categorizing text into sentiment classification categories.
7. Data Splitting, is the process of dividing the dataset for the purposes of training and testing the model.

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

2.3 Naive Bayes

Naïve Bayes is a probabilistic and statistical classification algorithm that assumes each attribute is independent or the characteristics of a class have nothing to do with other classes The following is the equation of Bayes theorem (1) [13].

$$P(x|z) = \frac{P(X)P(Z|X)}{P(Z)} \quad (1)$$

X, Z = events
P(X|Z) = the probability of X if Z is True
P(Z|X) = probability of Z if X is True
P(X), P(Z) = probability of occurrence of X and B, respectively.

2.4 Logistic Regression

Logistic Regression is a statistical model that assesses the relationship between an independent variable and a binary dependent variable. Logistic Regression can be used to classify data into two or more classes, in this case there are 3 sentiment classes, namely Negative, Neutral and Positive. [14]. The multinomial Logistic Regression function is as follows (2)[15].

$$g_j(x) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \cdots + \beta_{jp}x_p \quad (2)$$

Description:
g(x) = Logit function for class j
β_{j0} = intercept for class j
$\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp}$ = coefficient for each predictor variable $(x_1, x_2, \ldots, x_p)$ in class j

2.5 Data Evaluation

A confusion matrix is a table used to assess how well a classification model performs when applied to a given collection of data. In order to ascertain which classification algorithm is better and has the highest accuracy in forecasting public opinion toward Indonesian presidential candidates in the 2024 election, Confusion Matrix is used to compare the performance of the two algorithms. Every algorithm's accuracy, precision, recall, and F1 score values will be taken into account when choosing the best model. The percentage of overall accurate predictions is known as accuracy; the percentage of correct positive predictions is known as precision; the percentage of correctly predicted positive classes is known as recall; and the percentage that strikes a balance between precision and recall is known as the F1 score [16]. The confusion matrix used is a multiclass confusion matrix in accordance with the classification of labeling data which has 3 classes, namely Negative, Neutral, and Positive classes (table 1).

Table 1. Multiclass Confusion Matrix

	Actual Candidate Data with Negative Sentiment	Actual Candidate Data with Neutral Sentiment	Actual Candidate Data with Positive Sentiment
Prediction of Presidential Candidate Data with Negative Sentiment	**TP**	**FP**	**FP**
Prediction of Presidential Candidate Data with Neutral Sentiment	**FN**	**TN**	**TN**
Prediction of Presidential Candidate Data with Positive Sentiment	**FN**	**TN**	**TN**

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

Description:
TP: True Positive
TN: True Negative
FP: False Negative
FN: False Negative
Calculation of accuracy, precision, recall and F1 score values as follows:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1\ Score = \frac{2 \times Presisi \times Recall}{Presisi+Recall} \quad (6)$$

3 Results and Discussion

3.1 Data Crawling

This research uses data from crawling twitter with the hashtags Anies Baswedan, Ganjar Pranowo, and Prabowo Subianto. The data obtained was taken with the tweet harvest library through google colab and was able to retrieve 512 tweets related to Ganjar Pranowo, 509 tweets related to Anies Baswedan and 552 tweets related to Prabowo Subianto. The data was collected in November 2023 with 12 attributes including created_at, id_str, full_text, quote_count, reply_count, retweet_count, favorite_count, lang, user_id, conversation_id_str, username, and tweet_url.

Created _at	id_str	full_text	Quote _count	Reply _count	retweet _count	Favorite _count	lang	user_ id_str	Convers ation_id_str	username	tweet_url
Tue Nov 07 05:04:48 +0000 2023	1,72E+18	@Koriandri @aniesbaswedan @mohmahfudmd Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro. https://t.co/UzrEgZVO9v	0	0	0	0	in	1,23E+18	1,72E+18	Bentir16	https://twitter.com/Bentir16/status/1721755558393557487
Tue Nov 07 05:02:38 +0000 2023	1,72E+18	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas @aniesbaswedan di DKI, Buka bahas @OldTraffordNews ,	0	0	0	0	in	1,44E+18	1,72E+18	choymarkochoy	https://twitter.com/choymarkochoy/status/1721755013746471308
Tue Nov 07 05:02:43 +0000 2023	1,72E+18	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarmahfud2024 #GanjarMahfud #jagademokrasi #paktelur #anakpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo https://t.co/Zw4GQSsamu	0	0	0	0	in	1,54E+18	1,72E+18	Ketik_salah24	https://twitter.com/Ketik_salah24/status/1721755034290106622

Figure 2. Twitter Crawling Data Results

3.2 Data Preprocessing

3.2.1 Filtering

In the filtering stage, text data that is not clean from unused characters such as mentions, hashtags, emojis, urls, and punctuation marks is removed so that the text obtained is only the sentiment text. The original data or raw data before preprocessing is as shown in Figure 3.

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

index	full_text
0	@Koriandri @aniesbaswedan @mohmahfudmd Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro. https://t.co/UzrEgZVO9v
1	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas @aniesbaswedan di DKI, Buka bahas @OldTraffordNews ,
2	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarmahfud2024 #GanjarMahfud #jagademokrasi #paktelur #anakpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo https://t.co/Zw4GQSsamu
3	@Tita83079013 Tapi aku ragu ketiga calon bahas ekonomi syariah melulu @ganjarpranowo @aniesbaswedan @prabowo tidak ada ekonomi kerakyatan Pancasila
4	@DPP_PKB @aniesbaswedan @cakimiNOW Bbm gratis jadi ??? Taik

Figure 3. Full_Text Attribute Data Before Preprocessing

Then the data filtering process is carried out to clean the data from hashtags, urls, emojis, mentions and other characters. The result of the filtering process is shown in Figure 4.

index	full_text	clean_text
0	@Koriandri @aniesbaswedan @mohmahfudmd Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro. https://t.co/UzrEgZVO9v	Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro.
1	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas @aniesbaswedan di DKI, Buka bahas @OldTraffordNews ,	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas di DKI, Buka bahas ,
2	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarmahfud2024 #GanjarMahfud #jagademokrasi #paktelur #anakpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo https://t.co/Zw4GQSsamu	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA
3	@Tita83079013 Tapi aku ragu ketiga calon bahas ekonomi syariah melulu @ganjarpranowo @aniesbaswedan @prabowo tidak ada ekonomi kerakyatan Pancasila	Tapi aku ragu ketiga calon bahas ekonomi syariah melulu tidak ada ekonomi kerakyatan Pancasila
4	@DPP_PKB @aniesbaswedan @cakimiNOW Bbm gratis jadi ??? Taik	Bbm gratis jadi ??? Taik

Figure 4. Data After Filtering Process

In Figure 3, the text data still has characters that make the text difficult to process. In Figure 4, the text data is clean from mentions, hashtags, urls and other unused characters.

3.2.2 Casefolding

In Figure 5, the data that has been cleaned is then cleaned from punctuation marks, digits, extra whitespace and capital letters are changed to lowercase letters.

index	full_text	clean_text
0	@Koriandri @aniesbaswedan @mohmahfudmd Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro. https://t.co/UzrEgZVO9v	beneran blum tau kalo anies bukankah petugas partai nasdem siapa presidennya anies apa partainya nasdem ya kaleee kalo mahfud md sih non partai tapi sebelum lahir pun sudah nahdliyin kalo blum tau rajin baca bro
1	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas @aniesbaswedan di DKI, Buka bahas @OldTraffordNews ,	ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas di dki buka bahas
2	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarmahfud2024 #GanjarMahfud #jagademokrasi #paktelur #anakpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo https://t.co/Zw4GQSsamu	hanya ganjar yang konsisten tolak israel datang ke indonesia
3	@Tita83079013 Tapi aku ragu ketiga calon bahas ekonomi syariah melulu @ganjarpranowo @aniesbaswedan @prabowo tidak ada ekonomi kerakyatan Pancasila	tapi aku ragu ketiga calon bahas ekonomi syariah melulu tidak ada ekonomi kerakyatan pancasila
4	@DPP_PKB @aniesbaswedan @cakimiNOW Bbm gratis jadi ??? Taik	bbm gratis jadi taik

Figure 5. Data After Casefolding Process

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

3.2.3 Stemming

Stemming converts words that have affixes into their base word.

clean_text	stemword
beneran blum tau kalo anies bukankah petugas partai nasdem siapa presidennya anies apa partainya nasdem ya kaleee kalo mahfud md sih non partai tapi sebelum lahir pun sudah nahdliyin kalo blum tau rajin baca bro	beneran blum tau kalo anies bukankah tugas partai nasdem siapa presiden anies apa partai nasdem ya kaleee kalo mahfud md sih non partai tapi belum lahir pun sudah nahdliyin kalo blum tau rajin baca bro
ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas di dki buka bahas	ini bahas lapang hasil karya anak bangsa yg bangun masa pimpin mas di dki buka bahas
hanya ganjar yang konsisten tolak israel datang ke indonesia	hanya ganjar yang konsisten tolak israel datang ke indonesia
tapi aku ragu ketiga calon bahas ekonomi syariah melulu tidak ada ekonomi kerakyatan pancasila	tapi aku ragu tiga calon bahas ekonomi syariah melulu tidak ada ekonomi rakyat pancasila
bbm gratis jadi taik	bbm gratis jadi taik

Figure 6. Data After Stemming Process

3.2.4 Stopword removal and Tokenizing

Tokenizing separates a sentence into fragments of words. Meanwhile, *stopword* removal is useful for removing words that have no sentimental meaning or basic words such as the words from, to, this, and, in, which, and so on.

stemword	Tokenized_and_No_Stopwords
beneran blum tau kalo anies bukankah tugas par...	[beneran, blum, tau, kalo, anies, tugas, parta...
ini bahas lapang hasil karya anak bangsa yg ba...	[bahas, lapang, hasil, karya, anak, bangsa, yg...
hanya ganjar yang konsisten tolak israel datan...	[ganjar, konsisten, tolak, israel, indonesia]
tapi aku ragu tiga calon bahas ekonomi syariah...	[ragu, calon, bahas, ekonomi, syariah, melulu,...
bbm gratis jadi taik	[bbm, gratis, taik]

Figure 7. Data After Stopwords Removal and Tokenizing Process

3.2.5 Data Labeling

This research uses the BERT model for labeling Indonesian data. The BERT model will process the text data and then label or classify it with 1 - 5 star categories according to the sentiment value of the text, for example in Figure 8.

	clean_text	Sentimen_Label	Sentimen_Score
0	beneran blum tau kalo anies bukankah petugas p...	1 star	0.514378
1	ini membahas lapangan hasil karya anak bangsa ...	5 stars	0.423526
2	hanya ganjar yang konsisten tolak israel datan...	1 star	0.330613
3	tapi aku ragu ketiga calon bahas ekonomi syari...	2 stars	0.350466
4	bbm gratis jadi taik	1 star	0.285715

Figure 8. Labeling Results Using the BERT Model

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

Then, as indicated in Figure 9, modify the labeling with a 1–5 star format to reflect negative, neutral, and positive sentiments, with 1–2 stars denoting negative feeling, 3 stars denoting neutral mood, and 4–5 stars denoting positive sentiment.

	clean_text	Sentimen_value	Labeling	Sentimen_score
0	beneran blum tau kalo anies bukankah petugas p...	Negative	1 star	0.514378
1	ini membahas lapangan hasil karya anak bangsa ...	Positive	5 stars	0.423526
2	hanya ganjar yang konsisten tolak israel datan...	Negative	1 star	0.330613
3	tapi aku ragu ketiga calon bahas ekonomi syari...	Negative	2 stars	0.350466
4	bbm gratis jadi taik	Negative	1 star	0.285715

Figure 9. Changing Labeling to Sentiment_Value

Then add the value of the sentiment. For example, Negative is 0, Neutral is 1, and Positive is 2 as shown in Figure 10.

	clean_text	Sentimen_value	label
0	beneran blum tau kalo anies bukankah petugas p...	Negative	0
1	ini membahas lapangan hasil karya anak bangsa ...	Positive	2
2	hanya ganjar yang konsisten tolak israel datan...	Negative	0
3	tapi aku ragu ketiga calon bahas ekonomi syari...	Negative	0
4	bbm gratis jadi taik	Negative	0

Figure 10. Labeling Data Result

After the labeling process, the data on Indonesian presidential candidates can be known, which is as follows:

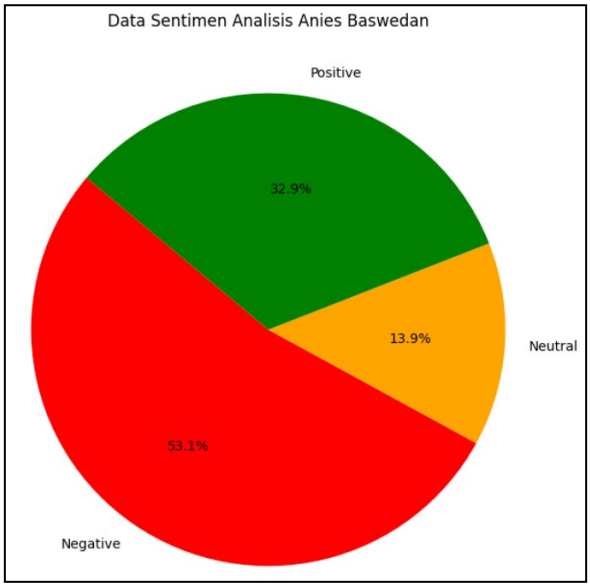

Figure 11. Pie **Chart** of Anies Baswedan's Sentiment

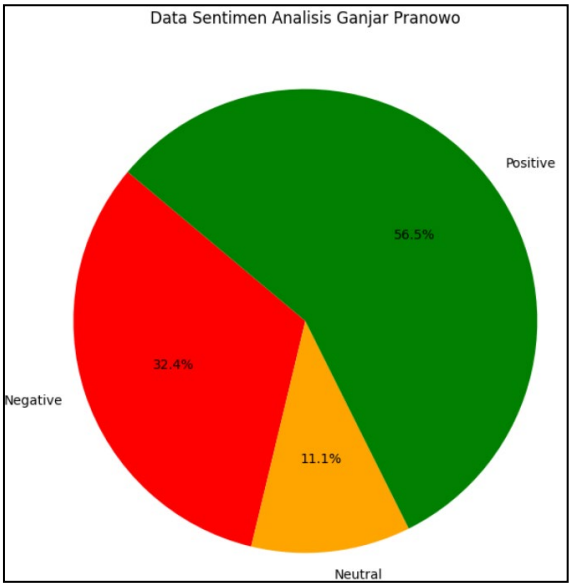

Figure 12. Pie chart of Ganjar Pranowo's Sentiment

```
1 senti_anies['Sentimen_value'].value_counts()

Negative    271
Positive    168
Neutral      71
Name: Sentimen_value, dtype: int64
```

Figure 13. Number of Anies Baswedan Sentiment Data

```
1 senti_Ganjar['Sentimen_value'].value_counts()

Positive    290
Negative    166
Neutral      57
Name: Sentimen_value, dtype: int64
```

Figure 14. Number of Sentiment Data of Ganjar Pranowo

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

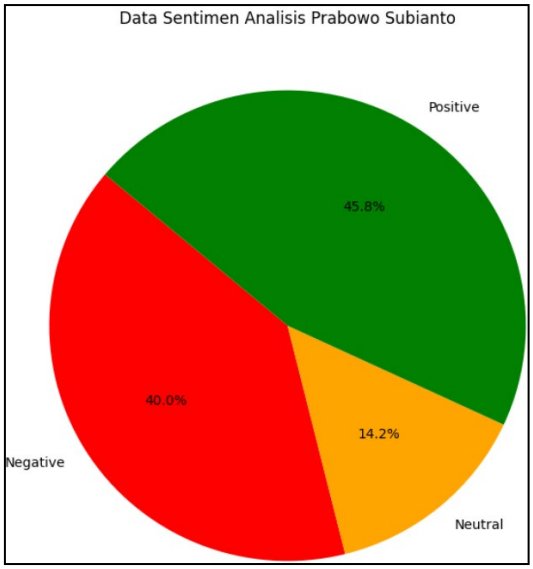

Figure 15. Pie Chart of Prabowo Subianto Sentiment

```
1 senti_Prabowo['Sentimen_value'].value_counts()

Positive    229
Negative    200
Neutral      71
Name: Sentimen_value, dtype: int64
```

Figure 16. Total Sentiment Data of Prabowo Subianto

3.2.6 Data Splitting

In this study, the data is randomly divided into 80:20 with 80% into train data and 20% into test data. Train data is used for training data to train the model which will next be examined using test data to evaluate the model's performance.

```
1 from sklearn.model_selection import train_test_split
2
3 # Misalnya, X adalah fitur dan y adalah label
4 X_train, X_test, y_train, y_test = train_test_split(senti_anies['clean_text'],
5                                   senti_anies['label'],test_size=0.2, random_state=42)
6
```

Figure 17. Data Splitting

3.3 Model Evaluation

This research uses Naive Bayes and Logistic Regression algorithms in classifying sentiment analysis data with data divided into 80% train data and 20% test data. Then each presidential candidate data is processed using Naive Bayes Modeling and Logistic Regression. The results of each model are listed in Table 1 for the classification of sentiment data of presidential candidate Anies Baswedan, Table 2 for the classification of sentiment data of presidential candidate Ganjar Pranowo, and Table 3 for the classification of sentiment data of presidential candidate Prabowo Subianto.

Table 1 shows that the accuracy value of Naive Bayes modeling is 63% while the precision value is 64%, 75% and 54%. Meanwhile, the accuracy value of Logistic Regression modeling is 61% while the precision value is 74%, 50%, and 44%. So that based on the assessment of the accuracy level of modeling on Anies Baswedan sentiment data, the algorithm whose accuracy value is superior is the Naive Bayes accuracy value of 63% while the Logistic Regression accuracy value is 61%. Confusion matrix sentiment data of 2024 presidential candidate Anies Baswedan is depicted in Figure 18 and Figure 19.

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

Table 2. Anies Baswedan Sentiment Data Classification Report

Evaluation Results of Anies Baswedan Sentiment Data Model					
Algorithm	Sentiment	Precision	Recall	F1-socre	Accuracy
Naive Bayes	Negative	64%	92%	75%	63%
	Neutral	75%	19%	30%	
	Positive	54%	26%	35%	
Logistic Regression	Negative	74%	71%	72%	61%
	Neutral	50%	6%	11%	
	Positive	44%	70%	54%	

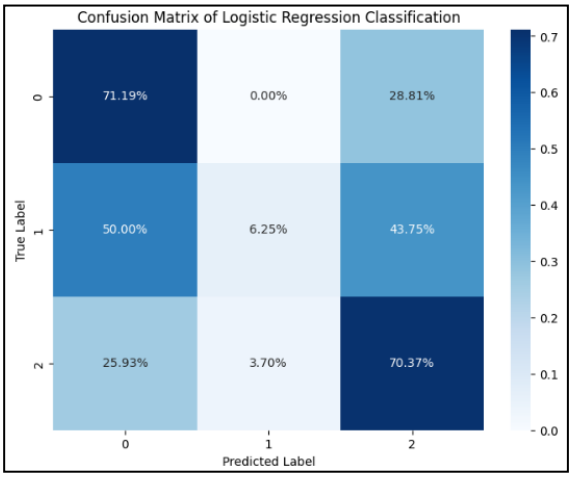

Figure 18. Confusion Matrix Naive Bayes Anies B.

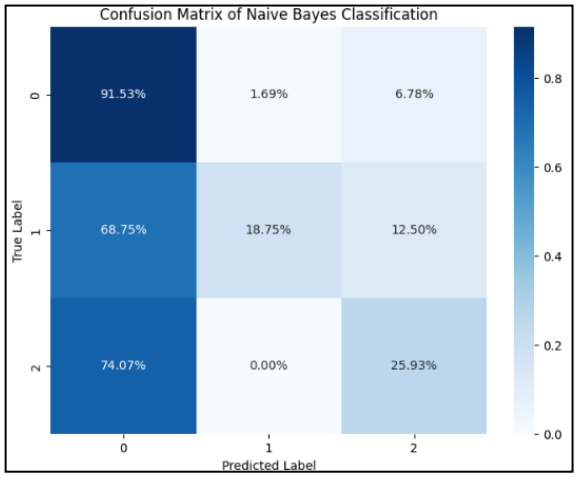

Figure 19. Confusion Matrix of Logistic Regression Anies B.

Table 2 shows that the accuracy value of Naive Bayes modeling is 77% while the precision value is 72%, 33% and 81%. Meanwhile, the accuracy value of Logistic Regression modeling is 77% while the precision value is 84%, 0%, and 76%. So based on the assessment of the modeling accuracy level on Ganjar Pranowo sentiment data, both algorithms have the same accuracy value of 77%, which makes both modeling balanced. Confusion matrix of sentiment data of 2024 presidential candidate Ganjar Pranowo is depicted in Figure 20 and Figure 21.

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

Table 3. Ganjar Pranowo Sentiment Data Classification Report

Model Evaluation Result of Ganjar Pranowo Sentiment Data					
Algorithm	Sentiment	Precision	Recall	F1-score	Accuracy
Naive Bayes	Negative	72%	77%	74%	77%
	Neutral	33%	8%	13%	
	Positive	81%	90%	85%	
Logistic Regression	Negative	84%	70%	76%	77%
	Neutral	0%	0%	0%	
	Positive	76%	95%	85%	

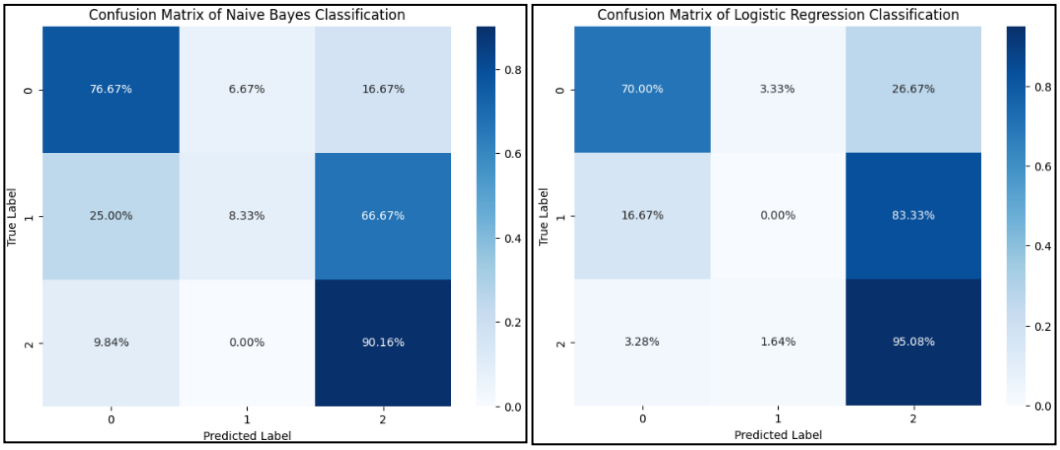

Figure 20. Confusion Matrix Naive Bayes Ganjar P. **Figure 21.** Confusion Matrix Logistic Regression Ganjar P.

Table 4 shows that the accuracy value of Naive Bayes modeling is 44% while the precision value is 33%, 50% and 52%. Meanwhile, the accuracy value of Logistic Regression modeling is 41% while the precision values are 33%, 50%, and 46%. Therefore, based on the assessment of the accuracy level of modeling on Prabowo Subianto sentiment data, the algorithm whose accuracy value is superior is the Naive Bayes accuracy value, which is 44% while the Logistic Regression accuracy value is 41%. The confusion matrix of the sentiment data of 2024 presidential candidate Prabowo Subianto is illustrated in Figure 22 and Figure 23.

Table 4. Prabowo Subianto Sentiment Data Classification Report

Model Evaluation Result of Prabowo Subianto Sentiment Data					
Algorithm	Sentiment	Precision	Recall	F1-score	Accuracy
Naive Bayes	Negative	33%	42%	37%	44%
	Neutral	50%	14%	22%	

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 1, February 2023, Pages 44-55
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i1.267

	Positive	52%	53%	52%	
Logistic Regression	Negative	33%	45%	38%	41%
	Neutral	50%	7%	12%	
	Positive	46%	47%	47%	

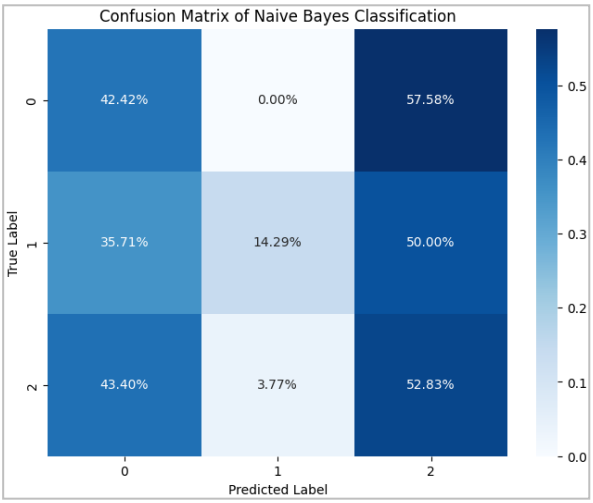

Figure 22. Confusion Matrix of Naive Bayes Prabowo S.

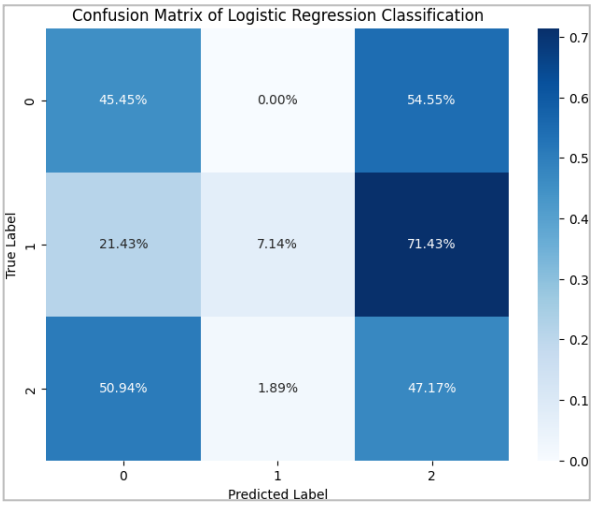

Figure 23. Confusion Matrix of Logistic Regression Prabowo S.

Table 5 shows the accuracy value of each presidential candidate and also the modeling algorithm used. From the table we can conclude that Naive Bayes is a modeling that has a superior accuracy value with an accuracy value of 63%, 77% and 44%, compared to Logistic Regression whose accuracy value is 61%, 77%, and 41%. The highest accuracy value is achieved by Ganjar pranowo sentiment data with a serial accuracy value of 77%.

Table 5. Accuracy values of Naive Bayes modeling and Logistic Regression

Modeling accuracy value on Presidential Candidate Sentiment Data		
	Naive Bayes	Logistic Regression
Anies Baswedan	63%	61%
Ganjar Pranowo	77%	77%
Prabowo Subianto	44%	41%

4 Conclusions

Along with the organization of the 2024 General Election, many rumors have begun to emerge in the community. One of the factors for the rise of rumors ahead of the 2024 General Election is Social media, which is a place for people to exchange information and also channel opinions. Sentiment analysis is the right research to see public sentiment towards Indonesian presidential candidates. From the above analysis, it can be concluded that the accuracy value of Naive Bayes modeling is superior to Logistic Regression in classifying sentiment data of 2024 presidential candidates with an accuracy value of 63% for candidate Anies Baswedan, 77% for candidate Ganjar Pranowo, and 44% for Prabowo Subianto.

In this study, only 1,573 data were processed and obtained before the campaign period began. It is hoped that further research can obtain more relevant data, add word weighting, and also text lebeling and model classification can be more accurate.

References

[1] B. Delvika, Apriana, N. Abror, and U. R. Gurning, "Perbandingan Algoritma NBC dan C4.5 DalamAnalisa Sentimen Pemilihan Presiden 2024 Pada Twitter," *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, vol. 1, no. 1, pp. 41–48, Aug. 2023.

[2] H. H. Zain, R. M. Awannga, and W. I. Rahayu, "Perbandingan Model Svm, Knn Dan Naïve Bayes Untuk Analisis Sentiment Pada Data Twitter: Studi Kasus Calon Presiden 2024," *JIMPS: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, vol. 8, no. 3, Jun. 2023.

[3] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 479, Apr. 2023, doi: 10.30865/jurikom.v10i2.5945.

[4] Fathir, M. A. Hariyadi, and Y. Miftachul A, "ANALISIS SENTIMEN ARTIKEL BERITA PEMILU BERBASIS METODE KLASIFIKASI," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 2, pp. 485–493, May 2023, doi: 10.35870/jimik.v4i2.220.

[5] S. Juanita, "Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 3, p. 552, Jul. 2020, doi: 10.30865/mib.v4i3.2140.

[6] R. Asmara, M. F. Ardiansyah, and M. Anshori, "Analisa Sentiment Masyarakat terhadap Pemilu 2019 berdasarkan Opini di Twitter menggunakan Metode Naive Bayes Classifier," *INOVTEK Polbeng - Seri Informatika*, vol. 5, no. 2, p. 193, Nov. 2020, doi: 10.35314/isi.v5i2.1095.

[7] A. Muzaki and A. Witanti, "SENTIMENT ANALYSIS OF THE COMMUNITY IN THE TWITTER TO THE 2020 ELECTION IN PANDEMIC COVID-19 BY METHOD NAIVE BAYES CLASSIFIER," *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 2, pp. 101–107, Mar. 2021, doi: 10.20884/1.jutif.2021.2.2.51.

[8] A. Averina, H. Hadi, and J. Siswantoro, "Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model Regresi Logistik," *Teknika*, vol. 11, no. 2, pp. 123–128, Jun. 2022, doi: 10.34148/teknika.v11i2.461.

[9] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *Journal of Information Systems and Informatics*, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.

[10] K. Kelvin, J. Banjarnahor, E. I. -, and M. NK Nababan, "Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)," *Jurnal Sistem Informasi dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 47–52, Feb. 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365.

[11] Ash Shiddicky and Surya Agustian, "Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 pada Media Sosial Twitter menggunakan Metode Logistic Regression," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 99–106, Aug. 2022, doi: 10.37859/coscitech.v3i2.3836.

[12] I. Iwandini, A. Triayudi, and G. Soepriyono, "Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor," *Journal of Information System Research (JOSH)*, vol. 4, no. 2, pp. 543–550, Jan. 2023, doi: 10.47065/josh.v4i2.2937.

[13] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA JURNAL*, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455.

[14] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, Apr. 2021, doi: 10.28932/jutisi.v7i1.3216.

[15] R. Prabowo, H. Sujaini, and T. Rismawan, "Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode Regresi Logistik Multinomial," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 1, p. 85, Jan. 2023, doi: 10.26418/justin.v11i1.57450.

[16] M. R. Romadhon and F. Kurniawan, "A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia," in *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, IEEE, Apr. 2021, pp. 41–44. doi: 10.1109/EIConCIT50028.2021.9431845.