E-ISSN: 2962-6110
P-ISSN: 2964-7843
JSRET
jsret.knpub.com
Journal of Scientific Research, Education, and Technology

Sentiment Analysis on Prabowo Subianto Using Naive Bayes on Twitter

Benediktus Rivaldi Toteng, Yuli Asriningtias
University of Technology Yogyakarta

ABSTRACT

General elections are a mechanism to implement the sovereignty of the people with the aim of creating a democratic state government based on Pancasila and the 1945 Constitution of the Republic of Indonesia. Elections are held every five years and involve the election of the President and Vice President, Members of the People's Representative Council, Regional Representative Council, Regional People's Representative Council, as well as regional heads and deputy regional heads. Elections aim to select leaders who can reflect democratic values and represent the aspirations of the people in accordance with the life of the nation and state. To understand public opinion regarding the 2024 elections, a sentiment analysis was conducted. This sentiment analysis aims to determine the public's positive and negative responses towards Presidential Candidate Prabowo Subianto. Sentiment Analysis of Prabowo Subianto Using Naive Bayes On Twitter, data was collected from the social media platform using the hashtag, #Prabowo.

Keywords: *Sentiment Analysis, General election, Prabowo Subianto, Naive Bayes Classifier, Twitter*

Corresponding author
Name: Benediktus Rivaldi Toteng
Email: benediktustoteng2002@gmail.com

INTRODUCTION

Indonesia is a Southeast Asian country with a presidential system of government. The president is an official title held by the head of an organization, association, company, university, or country. In Indonesia, the concept of the president is based on the 1945 Constitution, where the president is the highest position in the government system that holds power in Indonesia (Al Muqsith Prasetyo, 2023). The KPU is socializing the 2024 elections using methods and social media commonly used by the public to raise awareness of participating in the elections among the community (Makmun , 2024).

The president in Indonesia is elected by the people through a democratic process, namely the presidential election (pilpres) which is held every 5 years. Being a president has several requirements in which the requirement is that a person is not allowed to become president if the person has previously been president for 2 consecutive periods, which in this case the current Indonesian president cannot run for President again in the next presidential election which will take place in 2024 (Fais Sya' bani, 2022).

One of the algorithms used for text mining is the Naïve Bayes Classifier. Naïve Bayes Classifier is an algorithm for classification using probability and statistical methods to

predict future probabilities based on past experiences. The Naïve Bayes algorithm uses a simple probabilistic method that calculates a set of probabilities by summing the frequencies and value combinations from the dataset (Mauliza and Sipayung 2024).

Sentiment Analysis On Prabowo Subianto Using Naive Bayes On Twitter was created to observe and gather information in the form of responses from the Indonesian public on Twitter directed at Presidential Candidate Prabowo Subianto, whether these responses from the Indonesian public fall into the positive or negative category. The research "Sentiment Analysis On Prabowo Subianto Using Naive Bayes On Twitter" involves several stages, namely data cleaning, preprocessing, translating, labeling, and visualization. The purpose of Sentiment Analysis On Prabowo Subianto Using Naive Bayes On Twitter is to summarize and conclude the public responses expressed through Twitter regarding the presidential election of Prabowo Subianto. And the benefit of Sentiment Analysis On Prabowo Subianto Using Naive Bayes On Twitter is to help the public understand in a broader context the polarity of the Indonesian people's responses to the 2024 Presidential Election.

LITERATURE REVIEW

The results from the Naïve Bayes method show 515 positive sentiments, 216 negative sentiments, and 122 neutral sentiments. The accuracy value is 77.37%, positive prediction precision is 100.00%, negative prediction precision is 94.01%, and neutral prediction precision is 40.00%. The recall for positive data is 73.98%, recall for negative data is 72.69%, and recall for neutral data is 100.00%. Referring to the accuracy, precision, and recall values, the Naive Bayes method has a high accuracy value. Based on the sentiment results from Naive Bayes, there are 515 positive sentiments, 216 negative sentiments, and 122 neutral sentiments (Supriatna, 2024).

This research was conducted by , and the initial step taken was crawling data on social media X, resulting in 147 data points. In the next stage, data preprocessing was carried out, which included case folding, cleaning, tokenizing, stopwords removal, and stemming, resulting in 145 data ready to be labeled into three sentiment categories: positive, neutral, and negative. Next, the 145 review data were grouped into two types of data, namely as training data with a total of 43 data and as test data with a total of 102 data, followed by applying the naïve bayes classifier algorithm to perform the classification process. The results of the sentiment prediction show 56 negative sentiments, 15 neutral sentiments, and 31 positive sentiments. This indicates that the negative sentiment is higher, reflecting the public's disagreement with the MK's decision. The high level of negative sentiment affects the public's trust in the Constitutional Court (Izzati 2024).

This research aims to analyze sentiment towards restaurant data in Singapore using the Naive Bayes Classifier algorithm, which has concluded that the Naive Bayes Classifier algorithm can classify a review into two categories: positive, represented by the word "satisfied," and negative, represented by the word "unsatisfied." Based on the results of the tests conducted, sentiment classification using the Naive Bayes Classifier algorithm

yielded a precision value of 73.02%, a recall of 74%, and an accuracy of 73.33% (Permadi, 2020).

Based on the results of the discussed research, the Naive Bayes Method can be used to classify data in the form of text, especially text originating from Twitter. The number of words in each training class greatly affects the classification results on the testing data, therefore the balance of the data needs to be maintained. Non-standard vocabulary can affect the classification results of a testing class if a training class has more non-standard words compared to other training classes (Zuhri, 2020).

Based on the sentiment analysis results of YouTube comments about Anies Baswedan as a potential presidential candidate for 2024 using the Naive Bayes Classifier method with a total of 1009 data points, there are more positive comments totaling 610 compared to negative comments totaling 399. The testing results at the preprocessing stage achieved the best accuracy level when using the preprocessing technique without stopword removal, with a training data ratio of 90% and 10% test data, resulting in an accuracy of 79%. This is because stopword removal was not performed, so the information from the sentences was not lost and obtained higher accuracy (Chely Aulia, 2023).

The results of the classification using the Naïve Bayesian Classifier algorithm yielded 839 positive tweets, 32 negative tweets, and 67 neutral tweets from a total of 938 tweets, or in percentage terms, 90% were positive sentiment, 3% negative sentiment, and 7% neutral sentiment towards Mr. Joko Widodo. And 56 positive tweets, 6 negative tweets, and 8 neutral tweets from a total of 70 tweets, or in percentage form, 80% are positive sentiment, 9% negative sentiment, and 11% neutral sentiment towards Mr. Prabowo. The accuracy level produced by the Naïve Bayesian Classifier algorithm for this research is 77.62% (Silitonga 2019) .

Based on the results of this study, which is a sentiment analysis on the inauguration of the President and Vice President of Indonesia 2024 using the Naive Bayes model and classification. The data used in this study amounted to 1,371, obtained from crawling the comment section of the presidential secretariat's YouTube channel. The data obtained was then processed through a preprocessing stage that included cleansing to clean the data, case folding to convert all letters to lowercase, tokenization to break sentences into words, stopword removal to eliminate less meaningful words, stemming to convert words to their base forms, and TF-IDF to weight the words. After preprocessing, it is followed by labeling to assign labels or categories to the data (Prayoga Siswono et al. 2024).

This study used a Twitter dataset of 15000, with the sentiment classes used being positive and negative. Before classifying, the data obtained will be given weight using TF-IDF. This technique will provide better accuracy when performing classification. Tests are run to get the best model using test scenarios that share training and test data. The best-case scenario consisted of 70% of the training data and 30% of the test data experimented on each dataset. Validation tests are performed with k-fold cross-validation and confusion matrices using these scenarios. K-fold cross-validation was performed using 10 iterations for each dataset, and the accuracy obtained for each dataset was 72.4% for the Anies dataset, 94.46% for Ganjar, and 74.5% for Prabowo. The validation using the

confusion matrix obtained accuracy results on the Anies dataset 74.4%, Ganjar 94.93%, and Prabowo 72.93%. However, in the f1-score, the lowest score obtained in the positive class is in the Anies dataset, while in the negative class, namely the Ganjar dataset. Based on previous research that has been carried out using a total of 533 tweet data consisting of Ganjar Pranowo datasets as many as 274, Anies Baswedan as many as 120, Prabowo Subianto as many as 72, and Ridwan Kamil as many as 67. This study got the best accuracy with the Ganjar dataset on the 7th K-Fold, 73.68%. The labeling stage in this study uses the TextBlob library with three classes of sentiment, namely positive, neutral, and negative. Based on this, the updates increase the number of datasets, using two classes of positive and negative sentiments and the scikit-learn library in labeling text (Firdaus, 2023).

METHOD

The research method stages include system research framework, dataset collection, model architecture, and system flowchart.

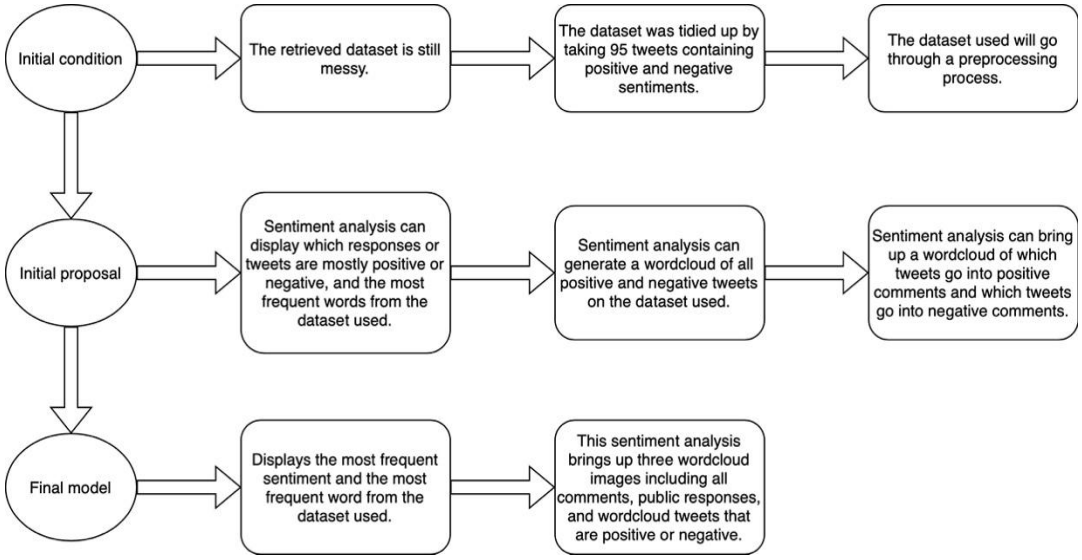

Figure 1 Research Framework System Research Framework

Figure 1 explains the initial condition of the subject matter in this sentiment analysis. From the above problems, the author makes a stage to overcome the sentiment analysis problem. At the stage of the expected final result, the sentiment analysis made by the author using Naive Bayes classifier can find out and determine the responses of the community which are included in the category and the most responses between positive and negative.

Research Data

Dataset X (Twitter)	Description
https://docs.google.com/spreadsheets/d/1Buiw9oHDmk1NwqLtP5TNPN-88WyhOjrNCEJyAdiTpPA/edit?usp=share_link	This data is a dataset taken from social media twitter.

Figure 2 Research Data

Data Source

The data sources in this study were taken by the author from datasets on the kaggle platform. Kaggle is a website that provides a collection of datasets from various fields with the aim of supporting research from researchers around the world (Hadianti, 2022). The dataset used in this study was manually selected because of its relevance to the topic discussed. By using data sources from kaggle, the author can ensure that the data used is well structured so as to facilitate the analysis process.

How to Get Data

The author obtained this dataset through the social media platform Twitter. Twitter has become a valuable source of information for analyzing public opinion towards institutions and individuals (Hilmi Zain, 2023). by using the keyword #prabowo to ensure that the data is relevant to the research topic. After collecting data from Twitter, the author also conducted manual filtering to ensure that the data collected meets the research needs and does not contain inappropriate or irrelevant information.

Data Collection Time

This data collection was carried out by the author within the time span of December 12, 2023 to December 15, 2023. The period was chosen because it is considered more representative to observe the latest trends related to the research topic. During this time, the author collected data in real time to obtain accurate and up-to-date information.

Model Architecture

At the model architecture stage, the author divides the process into two conditions, namely the condition before the system is created and the condition after it is created. In the condition before the system is created, the author conducts a needs analysis to determine the data, features, and algorithms that best suit the research objectives. After that, in the condition after the system was created, the author designed and implemented a model using the previously planned algorithm. This model is designed by considering various aspects, such as accuracy to measure how good this analysis design model is and the system capability of the model design or scalability.

System Flowchart

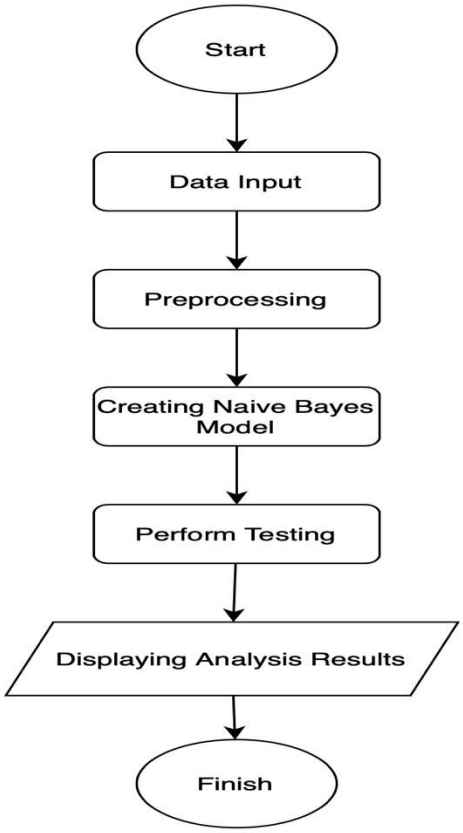

Figure 3 Flowchart

The flowchart describes the data analysis process using Naive Bayes method. The process begins with the Start stage, which signifies the initiation of the system or program. Next, Data Input is performed, which is entering the raw data to be processed. This data then goes through the Preprocessing stage, which aims to clean and prepare the data to make it suitable for use. This stage includes normalization, removal of irrelevant data, and data transformation. Once the data is ready, the next step is to Create a Naive Bayes Model. At this stage, the data that has been processed is used to train the model so that it can recognize certain patterns. After the model is successfully created, testing is carried out using test data to evaluate the performance of the model, such as the level of accuracy and prediction error. The results of the test are then summarized and displayed at the Displaying Analysis Results stage. This process provides an overview of the model performance and relevant information from the data analysis. Finally, the process ends at the Done stage, which indicates that all steps have been completed. Overall, this flowchart provides a systematic overview from the beginning to the end of the process of creating and evaluating Naive Bayes models in data analysis.

FINDING AND DISCUSSION

A system using the Naive Bayes algorithm to determine the positive and negative public responses to presidential candidate Prabowo Subianto requires a dataset about Prabowo Subianto in this study. The dataset containing Prabowo Subianto's tweets will be processed using Naive Bayes to produce a good sentiment analysis in identifying Prabowo Subianto's tweets. The programming language used in this research is Python with the Google Colab platform.

```
data = pd.read_csv(next(iter(uploaded)))  # Memuat file CSV
data = data.dropna()  # Menghapus nilai kosong

# Tampilkan 5 baris pertama untuk memastikan dataset benar
print(data.head())

# Langkah 2: Preprocessing Data
def clean_text(text):
    # Menghapus karakter spesial, URL, angka, dan spasi berlebih
    text = re.sub(r'http\S+', '', text)
    text = re.sub(r'[^a-zA-Z\s]', '', text)
    text = re.sub(r'\s+', ' ', text).strip()
    return text.lower()

data['Cleaned_Komentar'] = data['tweet'].apply(clean_text)
```

Figure 4 Text Cleanup

In Figure 4 above, the purpose of the code is to clean the text in the tweet column so that there are no elements that can interfere with the analysis process, such as URLs, numbers, or special characters. In performing text cleaning, it is necessary to perform several stages of steps such as removing URLs using the regular expression r “http\S+” to remove all URLs from the text, removing non-alphabetic characters from characters such as numbers or symbols (@, #, etc.) are removed to focus on pure text, converting text to lowercase to avoid differences between capital and non-capital letters. Next is a new dataset with a cleaned_tweet column, containing the cleaned text.

```
# Langkah 3: Tokenisasi dan Ekstraksi Fitur
vectorizer = CountVectorizer(stop_words='english', max_features=5000)
X = vectorizer.fit_transform(data['Cleaned_Komentar']).toarray()
y = data['label']  # Label

# Langkah 4: Split Data ke Training dan Testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Langkah 5: Membuat Model Naive Bayes
model = MultinomialNB()
model.fit(X_train, y_train)
```

Figure 5 Training Model

In Figure 5 the model is trained with Naive Bayes Classifier to classify the sentiment of tweets. To train the model, several training steps are required, such as data division with the dataset divided into 80% training data and 20% test data using the train_test_split function. Next, it uses feature extraction such as text converted into numerical representation using the Count Vectorizer function. Each word is considered a feature, and its frequency is counted. Model training is trained using the training data that has been converted into numeric vectors. Prediction of the model predicts the sentiment label on the dataset for the test data and the result is that the Naive Bayes model is ready to be used to analyze the sentiment on the test data.

```
Confusion Matrix:
 [[ 4  1]
 [ 2 12]]

Classification Report:
              precision    recall  f1-score   support

     NEGATIF       0.67      0.80      0.73         5
    POSITIVE       0.92      0.86      0.89        14

    accuracy                           0.84        19
   macro avg       0.79      0.83      0.81        19
weighted avg       0.86      0.84      0.85        19
```

Figure 6 Model Evaluation

Model evaluation is done to see how well the model classifies the test data. The first thing to do is to perform classification reports such as precision, recall, F1-score, and accuracy metrics calculated for each positive and negative category. Confusion Matrix shows the number of true positive and false positive/negative predictions for each category. The final results show that the model has a high accuracy of 84%, with the most prevalent sentiment being negative.

Figure 7 Wordscloud Visualization

The wordcloud visualization section uses the wordcloud library to display all comments from the tweet dataset used. In addition, wordcloud visualization displays the words that appear most often in the positive tweet wordcloud and negative tweet wordcloud. To use this function, a text merging process is performed first where all texts in one category are merged into one string.

```
#Demo Program
import pandas as pd
from textblob import TextBlob

# Data opini contoh
data = {
    "opini": [
        "Gemira and Semeton Bali Muslims declare support for Prabowo as president, news upda
        "The presidential palace is waiting for the general, ready for the general's orders.
        "I am sure that more than half of the dps will choose Pak Prabowo as president of th
        "Prabowo Subianto, a serious human rights violator, does not deserve to be president
        "Again, various comments came from netizens who prayed for Prabowo to become the nex
    ]
}

# Membuat DataFrame
df = pd.DataFrame(data)

# Fungsi untuk analisis sentimen (hanya positif atau negatif)
def analisis_sentimen(teks):
    analisis = TextBlob(teks)
    if analisis.sentiment.polarity > 0:
        return "Positif"
    else:
        return "Negatif"

# Menambahkan kolom sentimen ke DataFrame
df["sentimen"] = df["opini"].apply(analisis_sentimen)

# Menampilkan hasil
print(df)
```

Figure 8 Program Demo

In the final stage, the program demo aims to show how the sentiment analysis program works. Starting from the input data that is manually selected by the author in the form of a collection of opinions or texts from this research dataset, the author also translates the opinion into English because textblob by default is more effective for English texts. Because textblob is basically designed to work well with English text. The author has

also tried to use the text in Indonesian but the output results become errors in the sense that all the results become negative.

	opini	sentimen
0	Gemira and Semeton Bali Muslims declare suppor...	Negatif
1	The presidential palace is waiting for the gen...	Positif
2	I am sure that more than half of the dps will ...	Positif
3	Prabowo Subianto, a serious human rights viola...	Negatif
4	Again, various comments came from netizens who...	Negatif

Figure 9 Output Results

Based on the output results shown in Figure 9, sentiment analysis using the Naïve Bayes classifier produces sentiment classification for several opinions taken from this research dataset. Opinions that receive negative sentiment may contain words that disagree with or criticize the topic being discussed. Meanwhile, opinions that receive positive sentiment show confidence or support and hope regarding the topic.

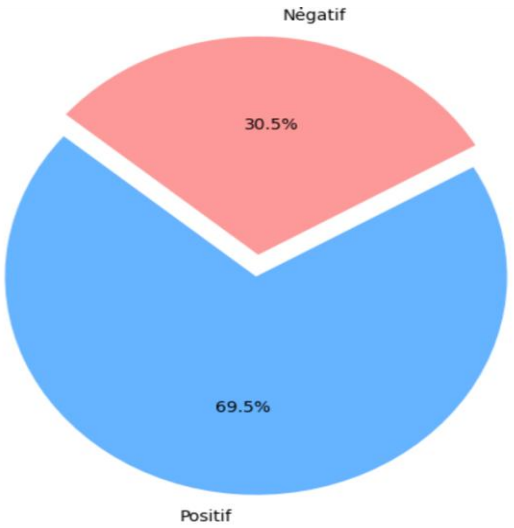

Figure 10 The Matplotlib Piechart

The matplotlib pie chart that will be used as an interface shows the proportion of positive sentiment responses at 69.5% and negative sentiment responses at 30.5%.

DISCUSSION

From the system development results, the Naive Bayes Classifier model was able to achieve a fairly high accuracy, indicating that this model is suitable for text-based sentiment analysis. The analysis results show that the positive sentiment towards presidential candidate Prabowo Subianto is greater than the negative sentiment. WordCloud reveals the words frequently used by the public in their tweets, such as Prabowo, president, and Subianto, providing additional insights into the discussion topic about Prabowo Subianto's candidacy in the 2024 presidential election. Similar research

using the Naive Bayes Classifier model was also able to achieve quite high accuracy, conducted by (Salsabila, 2023).

In the Sentiment Analysis Research on Prabowo Subianto Using Naive Bayes on Twitter, the results have not yet achieved perfect accuracy due to a significant lack of dataset. For further development, the author will try using other algorithms such as Support Vector Machine (SVM) and Random Forest to compare with the Naive Bayes Classifier, so that it can be identified which algorithm is superior in terms of accuracy or performance.

CONCLUSION

This research successfully applied the Naive Bayes Classifier method to analyze sentiment towards Presidential Candidate Prabowo Subianto based on positive and negative tweets. The resulting model has a good accuracy of 84%. Analysis shows that most of the sentiment is positive compared to negative sentiment, and for the development of future research, it is hoped that the authors will use other algorithms for comparison such as Support Vector Machine (SVM) and Random Forest to see the comparison of accuracy levels and performance obtained. Thus, the algorithm with the highest accuracy becomes the best choice for conducting sentiment analysis based on the topic of presidential candidate Prabowo Subianto.

REFERENCES

Chely Aulia Misrun, Elin Haerani, Muhammad Fikry, and Elvia Budianita. 2023. "Analisis Sentimen Komentar Youtube Terhadap Anies Baswedan Sebagai Bakal Calon Presiden 2024 Menggunakan Metode Naive Bayes Classifier." *Jurnal CoSciTech (Computer Science and Information Technology)* 4(1):207–15. doi: 10.37859/coscitech.v4i1.4790.

Fais Sya' bani, Muhammad Raihan, Ultach Enri, and Tesa Nur Padilah. 2022. "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes. *JURIKOM (Jurnal Riset Komputer)* 9(2):265. doi: 10.30865/jurikom.v9i2.3989.

Firdaus, Asno Azzawagama, Anton Yudhana, and Imam Riadi. 2023. "Public Opinion Analysis of Presidential Candidate Using Naïve Bayes Method." *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*. doi: 10.22219/kinetik.v8i2.1686.

Hadianti, Sri, Firman Yosep Tember, Nusa Mandiri, Jl Raya, Jatiwaringin No, Cipinang Melayu, and Jakarta Timur. n.d. "ANALISIS SENTIMENT COVID-19 DI TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SVM." *Jurnal Teknologi Informasi)* 6(1).

Hilmi Zain, Haekal, Rolly Maulana Awannga, and Woro Isti Rahayu. n.d. "Perbandingan Model Svm, Knn Dan Naïve Bayes Untuk Analisis Sentiment Pada Data Twitter: Studi Kasus Calon Presiden 2024." *JIMPS: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah* 8(3):2083–93. doi: 10.24815/jimps.v8i3.25342.

Izzati, Afifah Nurul. 2024. “Analisis Sentimen Hasil … ANALISIS SENTIMEN HASIL PUTUSAN MK TERKAIT SENGKETA PILPRES 2024 PADA MEDIA SOSIAL X DENGAN METODE NAÏVE BAYES.” *JTII* 09(01).

Makmun, Muhammad, Ahmad Turmudi Zy, and Asep Arwan. 2024. “Analisis Sentimen Media Sosial Twitter Terhadap Calon Presiden RI Tahun 2024 Menggunakan Klasifikasi Algoritma Naïve Bayes.” doi: 10.47065/josyc.v5i3.5210.

Mauliza, Risha Nur, and Yoannes Romando Sipayung. 2024. “Penerapan Text Mining Dalam Menganalisis Pendapat Masyarakat Terhadap Pemilu 2024 Pada Media Sosial X Menggunakan Metode Naive Bayes.” *Technomedia Journal* 9(1):1–16. doi: 10.33050/tmj.v9i1.2212.

Al Muqsith Prasetyo, Panji, and Arief Hermawan. 2023. “Analisis Sentimen Twitter Terhadap Pemilihan Presiden Menggunakan Algoritma Naïve Bayes.” *INFOTECH : Jurnal Informatika & Teknologi* 4(2):224–33. doi: 10.37373/infotech.v4i2.863.

Permadi, Vynska Amalia. n.d. *Analisis Sentimen Menggunakan Algritma Naïve Bayes Terhadap Review Restoran Di Singapura 141*.

Prayoga Siswono, Andika, Syahrul Fauzi, Aliefino Zalva Surya Hermawan, Arif Riyandi, Jl DI Panjaitan No, Purwokerto Kidul, Kec Purwokerto Selatan, and Kab Banyumas. 2024. *Analisis Sentimen Pelantikan Presiden Indonesia 2024 Menggunakan Model Klasifikasi Dan Algoritma Naive Bayes*. Vol. 4.

Salsabila, Fadila, and Utomo Budiyanto. 2023. *Implementasi Naïve Bayes Classifier Terkait Pencalonan Ganjar Pranowo Sebagai Calon Presiden 2024 Di Twitter*. Vol. 2.

Silitonga, Wiranto Horsen, and Jay Idoan Sihotang. n.d. *Analisis Sentimen Pemilihan Presiden Indonesia Tahun 2019 Di Twitter Berdasarkan Geolocation Menggunakan Metode Naïve Bayesian Classification*.

Supriatna, Reza, and Dede Rohman. 2024. *PENERAPAN NATURAL LANGUAGE PROCESSING DALAM ANALISIS SENTIMEN CAWAPRES 2024 MENGGUNAKAN ALGORITMA NAIVE BAYES*. Vol. 8.

Zuhri, Khoirul, Nurul Adha, and Oktarini Saputri. 2020. *Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier*. Vol. 1.