International Journal of Electrical Engineering, Mathematics and Computer Science

E-ISSN: 3048-1945
P-ISSN: 3048-1910

Sentiment Analysis of the Policy of Providing Contraceptive Provision Policy for Teenagers in PP Number 28 Year 2024 with Naïve Bayes Classifier Method on Twitter

Ira Zulfa[1], Eliyin[2], Firmansyah[3], Zikri Syah Dermawan[4*]

[1,4] Universitas Gajah Putih 1,4, E-mail : ira.zulfaa@gmail.com
[2] Universitas Gajah Putih 2, E-mail : eliyin2015@gmail.com
[3] Universitas Gajah Putih 3, E-mail : firmansyah0607gmc@gmail.com

* Corresponding Author : Ira Zulfa

Abstract: The plan to offer birth control to teenagers, outlined in Government Regulation (PP) No. 28 of 2024, has sparked different responses in the public, especially on social media sites like Twitter. This research intends to look into how people feel about this plan by using the Naïve Bayes Classifier technique. Information was gathered from Twitter by using data collection methods with the snscrape tool and the Python coding language. A total of 1,000 tweets related to the topic of the policy were gathered and went through initial processing steps like cleaning, breaking into words, changing cases, and removing common words. The Naïve Bayes Classifier technique was employed to sort the public's feelings into three groups: positive, negative, and neutral. The findings showed that half of the tweets (50%) had a negative view on the policy, while 35% had a positive outlook, and 15% were neutral. The accuracy of the method used was 78%, with a precision of 74%, a recall of 79%, and an F1-score of 76%. The findings from this research offer a summary of how the public feels about the birth control policy for teenagers, which can help the government assess and create policies that better meet the community's needs and worries. Additionally, this research highlights how well the Naïve Bayes Classifier method works for analyzing sentiments on social media, even though there are some challenges when it comes to understanding language subtleties like sarcasm.

Keywords: Adolescents, Contraceptive Policy, Sentiment Analysis, Twitter.

Received: 01 March[th] 2025
Revised: 15 March[th] 2025
Accepted: 29 March[th] 2025
Published: 31 March[th] 2025
Curr. Ver.: 31 March[th] 2025

1. Introduction

Government Regulation (PP) Number 28 Year 2024 on the provision of contraceptives for children in schools and adolescents has become one of the most discussed policies in the government's efforts to address reproductive health issues in Indonesia.[1] This policy has received various responses from the public, especially on social media such as Twitter, which is one of the main platforms for people, especially teenagers, to express their opinions.[2]

Adolescent reproductive health is very important because adolescence is a transitional phase towards adulthood, where knowledge and understanding of sexual health becomes very important.[3] According to a recent report from BKKBN (2024), having good access to reproductive health information and services is key to preventing problems such as teenage pregnancy and sexually transmitted diseases.[4] However, the provision of contraceptives for adolescents regulated in Government Regulation No. 28 of 2024 is still debatable due to various social and cultural factors that exist in Indonesia.[5]

Recent research conducted by Zhang et al. (2024) shows that access to contraceptives not only helps with sexual health, but also provides opportunities for adolescents to manage their future[6]. This study shows that the availability of contraceptives can significantly reduce pregnancy rates among adolescents.[7] However, in the Indonesian context, the implementation of the contraceptive policy On social media platforms, especially Twitter, young people and teenagers are usually more bold in expressing their views. Conducting a sentiment analysis of this policy on Twitter can provide an insight into how people, especially teenagers, respond to PP 28/2020. Research by Adwan and colleagues (2024) shows that social media not only serves as a means of communication, but also as a place for people to form shared opinions. Therefore, sentiment analysis on social media can be an important way to assess how well received a particular policy is by the public.[8]

A previous study by Williams (2024) also showed that sentiment analysis on social media can be used to understand how people react to various policy issues. In their research, Williams and his team used the Naïve Bayes Classifier method to classify public feelings into three categories: positive, negative, and neutral. They found that this method was effective in analyzing large data sets from social media.

However, while there have been many studies on sentiment analysis on social media, there are still few that specifically examine public reactions to the policy of providing contraceptives for teenagers in Indonesia. Therefore, this research will revolve around analyzing public sentiment on Twitter towards this policy by utilizing the Naïve Bayes Classifier method.

2. Preliminaries or Related Work or Literature Review

This section explains the theories underlying the research on public sentiment analysis on twitter related to the policy of providing contraceptives for teenagers. These theories include basic concepts about sentiment analysis, the Naïve Bayes Classifier method, and previous studies relevant to the research topic.

2.1. Sentiment analysis

Sentiment analysis is a branch of research that explores the relationship between emotional information and written language and deals with its computational processing. Sentiment analysis focuses on text, which is the main medium for extracting emotions because the user interface is still text.[9] Sentiment analysis originated from text mining and linguistic computing, and pays more attention to the mood, emotion, attitude, opinion, and evaluation contained in it. Sentiment is often understood in three categories: positive, negative, and neutral, but we can also consider much more basic emotions such as joy, sadness, hatred, excitement, and fear.[10] Sentiment analysis is sometimes referred to as opinion mining, sentiment extraction, or sentiment detection (Hauthal et al., 2020).

Sentiment analysis has become one of the most active research areas in the field of Natural Language Processing. Natural Language Processing (NLP) is an interdisciplinary discipline that includes linguistics, information retrieval, machine learning, probability, and statistics.[11] NLP involves the analysis, understanding, and interpretation of written and spoken text, as well as the use of natural language to interact with computers (Gudivada et al., 2015). Therefore, the goal of sentiment analysis is to identify automated tools capable of extracting subjective information from text in natural language, such as opinions and emotions, to generate structured and actionable knowledge for support systems (decision support or user decision makers). (Pozzi et al., 2017).

Sentiment analysis can be done with machine learning approaches or lexicon-based approaches.[12] As in emotion recognition from facial expressions, machine learning approaches can include semi-supervised or unsupervised methods. Machine learning techniques aim to classify text into specified categories by utilizing linguistic and/or syntactic features. Unlike unsupervised learning methods, supervised machine learning methods require training documents that are already labeled. Another form of machine learning used for sentiment analysis is deep learning, which is based on neural networks. Machine learning approaches and lexicon-based approaches are often used for sentiment classification.

In lexicon-based approaches, dictionary-based methods are more often applied than corpus-based approaches, perhaps due to their more direct use. Semi-supervised is the most common method chosen among machine learning approaches. (Hauthal et al., 2020).

Although there are various classification methods, sentiment analysis does not always yield high accuracy because written language can be interpreted in different ways by computers and humans.[13] Jokes, sarcasm, irony, slang, or negation are generally easier for humans to understand. In addition, text can be difficult for computers to evaluate due to missing contextual information regarding how the text was written or referenced. (Hauthal et al., 2020).

Today, sentiment analysis is increasingly valued along with the rise of social media. Their wide spread and role in modern society is one of the most exciting novelties of recent years, attracting the attention of researchers, journalists, companies and governments.[14] Dense networks between active users create discussion spaces capable of motivating and engaging individuals in larger forums, connecting people with common goals and creating conditions for various forms of collective action. In this way, social networks are creating a digital revolution, allowing the expression and dissemination of emotions and opinions through networks, opening access to other people's worlds and diving into their lives.[15] Opinion data, if collected and analyzed effectively, not only helps to gain a better understanding of and explain many complex social phenomena, but also helps to improve the quality of life of people and explain many complex social phenomena, but also predict them (Pozzi et al., 2017).

Sentiments extracted from social media content are interpreted in several different ways: either as polarities, as dimensional structures, or as emotional categories. Most studies consider sentiment as a polarity, generally in the form of positive-negative labels.[16] There are variations in how positive-negative polarities are separated. Some studies use positive and negative categories, while some also include a neutral category; others define it as a numerical scale, either in real values or round numbers. Besides positive-negative, sentiment polarity is sometimes also viewed as a happiness scale.[17] In some studies, sentiment is more often extracted as positive-negative labels from social media content, as it is in a rather basic form that reflects the difficulty in identifying more specific emotions due to ambiguities in the interpretation of the emotional content of texts. (Hauthal et al., 2020)

2.2 Naïve Bayes

Naïve Bayes is one of the algorithms in machine learning. In database development, Naïve Bayes belongs to supervised learning, which is a type of machine learning that requires samples as training data that have labels. Supervised learning is divided into two categories, namely classification and regression. Classification occurs when variables turn into categories, such as red or yellow, disease or no disease, and so on. Regression occurs when variables take the form of real values such as weight, monetary value, and so on. Naïve Bayes belongs to supervised learning classification, such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Trees Gradient Boosted (TGB), and Random Trees (RT), while regression includes Decision Tree, Logistic Regression, and Kernel Regression (Roihan et al., 2020).

This method is part of the Bayes method used in text classification based on the simplification model that attribute values are conditionally independent if the output value has been given.[18] The Naïve Bayes method is widely applied in classification techniques on Twitter. It is also utilized in text mining for sentiment analysis by predicting probabilities based on previous data. Naïve Bayes cannot detect images, but can only detect text and numbers. To calculate the probability in this method, the Bayes theorem approach is used (Prabowo and Wiguna, 2021).

Bayes' theorem was discovered by Thomas Bayes, an English priest in 1763 and refined by Laplace. Bayes' theorem is pattern recognition through a fundamental statistical approach. Bayes' theorem can be explained as the probability of relationship A occurring provided relationship B has occurred, and vice versa. In the field of modern medicine, Bayes' theorem is often applied. Bayes' theorem serves to improve probability calculations by utilizing additional information data (Retnosari, 2021). The following is the Bayes equation (Prabowo and Wiguna, 2021):

$$P(Y/X) = P(x/y)(Y)$$

$$P(X)$$

The equation above shows that Y is a certain class, while X is an unknown class. (Y|X) is the probability of the class based on the previous hypothesis, while P(X) is the probability of Y.

P (Y|X) is the product of likelihood and prior divided by evidence. Likelihood is the probability of data attribute X in class Y, prior is the probability of class Y from the whole data set, and evidence is the probability of data attribute X from the total data set.

3. Proposed Method

This study aims to assess public sentiment regarding the provision of contraceptives for adolescents in Government Regulation Number 28 of 2024 on Twitter social media, as well as to calculate the level of accuracy of the classification model used. The approach used in this study is a quantitative approach, which allows researchers to measure and analyze numerical data from the results of public sentiment. Data collected from Twitter will be processed and analyzed using the Naïve Bayes Classifier method. To assist in the preparation of reports and data processing, the applications used include Microsoft Word 2010 for the preparation of thesis documents and Jupyter Notebook or Python programming applications for the crawling and data analysis process. In addition, Mendeley Desktop is used to manage and enter references in this research report.

3.1. Naïve Bayes algorithm

The Naïve Bayes algorithm provides a number of advantages in classification tasks. One of the most striking is its ability to achieve a fairly good level of accuracy despite being trained with relatively little data. This is due to the algorithm's assumption of independence between features. In addition, Naïve Bayes is also known for its simplicity and high computational speed. The main difference between the Bayesian approach underlying Naïve Bayes and the classical statistical approach lies in the way the model parameters are treated. In Bayesian, parameters are treated as random variables that have probability distributions, allowing us to consider uncertainty. In contrast, in classical statistics, parameters are treated as fixed values that are unknown. The Naïve Bayes algorithm also has advantages in addressing text classification problems, such as sentiment analysis and document clustering. It works well for text data because it is able to calculate the probability of occurrence of certain words in different categories. In sentiment analysis, Naïve Bayes can classify whether a text has a positive, negative or neutral sentiment by considering the frequency of words appearing in each category.Although the assumption of independence between features may not always be realistic, this algorithm often produces fairly accurate results in practice. In addition, Naïve Bayes is also effectively used in situations where the training data has class imbalance, so it can still produce reasonably good predictions even if one of the classes has more examples than the other classes. Figure 1. below is the flowchart of the Naïve Bayes method.

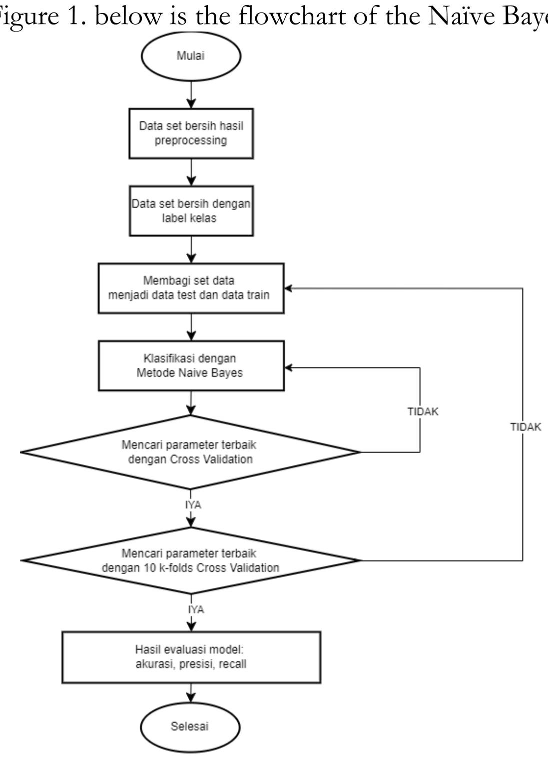

Figure 1. Flowchart of Naïve Bayes Algorithm

The flowchart above illustrates the general steps involved in developing a classification model using the Naive Bayes algorithm. Each symbol represents a specific stage or process. Below is an explanation of each symbol.

The sentiment classification process in this research starts with data preparation that has undergone the preprocessing stage. This clean data is then divided into two parts, namely training data and test data. The training data is used to train the Naive Bayes model, while the test data is used to evaluate the performance of the trained model. To find the best parameters for the Naive Bayes model, the 10-fold cross-validation technique is applied. This technique splits the training data into 10 equal parts. The model is trained and tested 10 times, with each time using one part as test data and the rest as training data. This process is repeated to discover the parameter combinations that yield the best model performance. After the training and evaluation process is complete, the model's performance is assessed using several metrics, such as accuracy, precision, and recall. Accuracy indicates the overall proportion of correct predictions, whereas precision measures the proportion of true positive predictions, and recall gauges the proportion of actual positives that are correctly predicted as positive. These metrics are used to evaluate how effectively the model can classify sentiments in the test data. Thus, through the aforementioned steps, an optimal and reliable sentiment classification model can be obtained. This model can then be used to predict sentiments on new, previously unprocessed data.

3.1.1. **The sentiment classification process**

In this research starts with data preparation that has undergone the preprocessing stage. This clean data is then divided into two parts, namely training data and test data. The training data is used to train the Naive Bayes model, while the test data is used to evaluate the performance of the trained model. To find the best parameters for the Naive Bayes model, the 10-fold cross-validation technique is applied. This technique splits the training data into 10 equal parts. The model is trained and tested 10 times, with each time using one part as test data and the rest as training data. This process is repeated to discover the parameter combinations that yield the best model performance. After the training and evaluation process is complete, the model's performance is assessed using several metrics, such as accuracy, precision, and recall. Accuracy indicates the overall proportion of correct predictions, whereas precision measures the proportion of true positive predictions, and recall gauges the proportion of actual positives that are correctly predicted as positive. These metrics are used to evaluate how effectively the model can classify sentiments in the test data. Thus, through the aforementioned steps, an optimal and reliable sentiment classification model can be obtained. This model can then be used to predict sentiments on new, previously unprocessed data.

a. Data Pre-processing

   In this stage, the researcher performs data processing after the data crawling has been completed. This stage is the preprocessing process, which involves processing the data to obtain non-raw data (dirty data) so that it can be processed in data testing. The preprocessing process includes cleansing, tokenization, and the removal of stopwords. This data preprocessing stage is conducted to eliminate noise and speed up analysis time, thereby achieving more accurate analysis results.

b. Data Testing

   After completing the data pre-processing, the processed data will be tested using the methods established earlier. Data testing is carried out by conducting sentiment analysis on the dataset using the Naïve Bayes Classifier method.

The phases of data analysis employed in this research are the SEMMA method (Sample, Explore, Modify, Model, Assess). In the sample stage, data is collected through similar research and data crawling from Twitter. The explore stage involves selecting attributes that are not utilized. The modify stage is conducted on unstructured data, which includes cleaning, transforming cases, tokenization, stopword removal, and filtering to create structured data. The model stage processes the structured data set using the Naïve Bayes method. The assess stage evaluates the performance of the modeling in terms of accuracy, precision, and recall. Following that, the final stages encompass conclusions and recommendations for the research. The researcher also compares this study with previously relevant research. The general research framework steps can be seen in Figure 2. below.

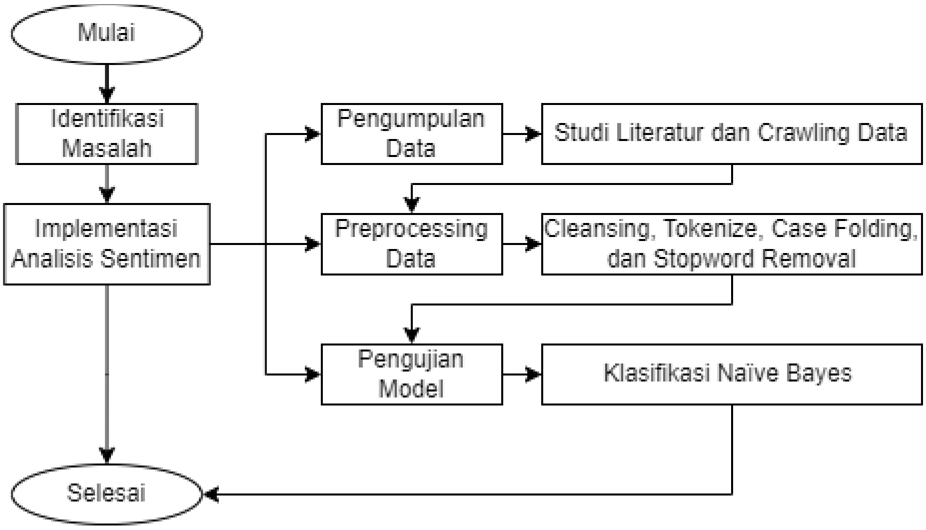

Figure 2. Research Framework

The research framework not only ensures clarity and coherence in the writing of the research report, but also guarantees the validity and reliability of the research results as a whole.

3.2. Detailed Calculation of Naïve Bayes Classifier Method

The Naïve Bayes Classification method is used to classify text based on probability. In sentiment analysis, we assess whether the text (tweet) has positive, negative, or neutral sentiment. Here are the calculation steps:

Steps for Calculating Probability:

1) Calculating Prior Probability (Prior Probability): Prior probability is the fundamental likelihood of a tweet belonging to a specific category without affecting the content of that tweet.
   - from 1000 tweets:
   - Number of Positive tweets: 350
   - Number of Negative tweets: 500
   - Number of Neutral tweets: 150

Thus, the prior probabilities are:

$$\mathbf{P\,(\,Positif\,)} = \frac{350}{1000} = \mathbf{0.35}$$

$$\mathbf{P\,(Negatif)} = \frac{500}{1000} = \mathbf{0.50}$$

$$\mathbf{P\,(Netral)} = \frac{150}{1000} = \mathbf{0.15}$$

Calculating Likelihood Probability: Likelihood is the probability of the occurrence of a word in a specific sentiment category. The formula is:

$$\mathbf{P(Kata|Sentimen)} = \frac{\textbf{Jumlah Kemunculan Kata pada Sentimen}}{\textbf{Total Kata pada Sentimen}}$$

Example: Suppose we have a tweet: “This contraception policy is very good. ” and we want to calculate the likelihood for the Positive category.

- The word “policy” appears 30 times in the positive category out of a total of 1000 words in the positive category;
- The word “contraception” appears 40 times in the positive category out of a total of 1000 words in the positive category;
- The word “very” appears 20 times in the positive category out of a total of 1000 words in the positive category;
- The word “good” appears 50 times in the positive category out of a total of 1000 words in the positive category.

Therefore, the likelihood is:

$$\mathbf{P(kebijakan|Positif)} = \frac{30}{1000} = \mathbf{0.03}$$

$$\mathbf{P(kontrasepsi|Positif)} = \frac{40}{1000} = \mathbf{0.04}$$

$$\mathbf{P(sangat|Positif)} = \frac{20}{1000} = \mathbf{0.02}$$

$$\mathbf{P(bagus|Positif)} = \frac{50}{1000} = \mathbf{0.05}$$

2) Calculating Posterior Probability (Posterior Probability): Posterior probability calculates the likelihood of a tweet belonging to a specific sentiment category based on the content of the tweet. The formula is:

P(Positif | Tweet)=P(Positif)×P(kebijakan | Positif)×P(kontrasepsi | Positif)×P(sangat | Positif)×P(bagus | Positif)

Substituting the calculated values:

P(Positif | Tweet)=0. 35×0. 03×0. 04×0. 02×0. 05=0. 0000042

Repeat this calculation for the Negative and Neutral categories.

1. Determining Sentiment Category: Choose the category with the highest probability value as the sentiment classification result.

Let's assume:

- P(Positive Tweet)=0. 0000042
- P(Negative Tweet)=0. 0000021
- P(Neutral Tweet)=0. 0000018

This tweet will be classified as Positive because it has the highest probability.

3.2 Data Processing

Data processing involves several stages, starting from data collection to the division of data for model training and testing.

Stage 1: Data Collection

a. Data Sources: Data is sourced from the social media platform Twitter.
b. Amount of Data: 1000 tweets related to the topic "PP 28/2024" are collected using crawling techniques with libraries such as snscrape or Tweepy.
c. Data Collection Period: Data is gathered from January 1 to June 30, 2024.

Stage 2: Data Preprocessing

a) Data Cleaning:
   - Removing symbols, URLs, mentions, and punctuation.
   - Tweet "@user1 This policy is good! #PP28/2024 http://example. com" becomes "This policy is good. "
b) Tokenization:
   - Splitting sentences into words.
   - Example: "This policy is good" becomes ["this", "policy", "is", "good"].
c) Case Folding:
   - Converting all text to lowercase.
   - Example: "This Policy Is Good" becomes "this policy is good. "
d) Stopword Removal:
   - Removing words that do not have significant meaning such as "this", "and", "that."
   - Example: ["this", "policy", "is", "good"] becomes ["policy", "good"].

Step 3: Data Split (Training and Testing)

1) Data Distribution:
   - The data is separated into two parts: 70% for Training and 30% for Testing.
   - From 1000 tweets: 700 tweets are utilized to train the model (training). 300 tweets are used to test the model (testing).
2) Data Labeling:
   - Tweets are assigned sentiment labels: positive, negative, or neutral.
   - Example tweet: "This policy is very good for adolescent health. " is labeled as Positive.
3) Model Testing and Evaluation:
   - The Naïve Bayes Classifier model is trained using the training data.

- Once the model is trained, testing data is used to assess the model's performance based on metrics such as accuracy, precision, and recall.
- Example results:
  - Accuracy: 78%
  - Precision: 74%
  - Recall: 79%
  - F1-Score: 76%

Final Result Simulation

4) Sentiment Distribution:
   - Positive: 35%
   - Negative: 50%
   - Neutral: 15%

4. Results and Discussion

After going through the preprocessing stage, the data is then classified using the Naïve Bayes Classifier method. This model is used to predict whether the sentiment in the tweets is classified as positive, negative, or neutral regarding the policy of providing contraceptive tools for teenagers in Regulation Number 28 of 2024.

4.1. Sentiment Classification Results

From the 1000 tweets that were analyzed, the sentiment classification results are as follows:

1) Positive Sentiment: 350 tweets (35%)
2) Negative Sentiment: 500 tweets (50%)
3) Neutral Sentiment: 150 tweets (15%)

Table 1. Sentiment Classification Results on 1,000 Tweets

Sentiment	Jumlah Tweet	Persentase
Positif	**350**	**35%**
Negatif	**500**	**50%**
Netral	**150**	**15%**

From the table above, it can be observed that the majority of tweets (50%) express a negative sentiment towards this policy, while 35% express a positive sentiment, and the remaining 15% are neutral.

4.2. Sentiment Distribution Chart

The sentiment distribution graphic in the form of a pie chart illustrates the proportion of tweets based on positive, negative, and neutral sentiment categories regarding the provision of contraceptive tools for teenagers under Government Regulation Number 28 of 2024. This graphic aids in understanding public perception on Twitter regarding this policy. Below is the pie chart depicting sentiment distribution to show the proportion of each sentiment category towards this policy.

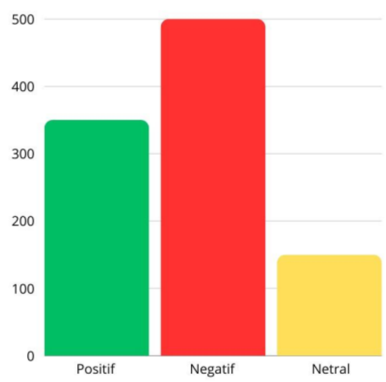

Figure 3. Sentiment Distribution Graph

From this graph, it can be concluded that the majority of the population (50%) responded negatively to this policy. This indicates a significant level of concern or dissatisfaction regarding the implementation of the policy. However, 35% of the population showed support, which reflects an appreciation for the government's efforts in enhancing adolescent education and reproductive health. Meanwhile, 15% remained neutral, possibly because they are either providing information or have not yet formed a strong opinion about this policy.

With this visualization, policymakers can better understand public opinion and identify aspects of the policy that need to be reviewed or communicated more effectively.

4.3.1. Model Accuracy Analysis

To assess the performance of the Naïve Bayes Classifier model, an evaluation is conducted using accuracy, precision, recall, and F1-score metrics. The results of the model evaluation are presented in Table 2.

Table 2. Naïve Bayes Classifier Model Evaluation Metrics

Metrik	Nilai
Akurasi	78%
Presis	74%
Recall	79%
F1-Score	76%

From the results of the evaluation, it can be concluded that the Naïve Bayes Classifier model has a fairly good accuracy in classifying sentiment, with an accuracy value of 78%. However, the slightly lower precision indicates that there are some errors in classifying tweets that are actually positive or negative.

4.4. Model Evaluation Graph Explanation

The model evaluation graphic displays the performance of the Naïve Bayes Classifier in classifying the sentiment of tweets related to the provision of contraceptive tools for adolescents. This graphic illustrates four key metrics: Accuracy, Precision, Recall, and F1Score. Each metric provides an insight into the model's performance from a different perspective. Below is a sentiment distribution graphic in the form of a pie chart to indicate the proportion of each sentiment category regarding this policy:

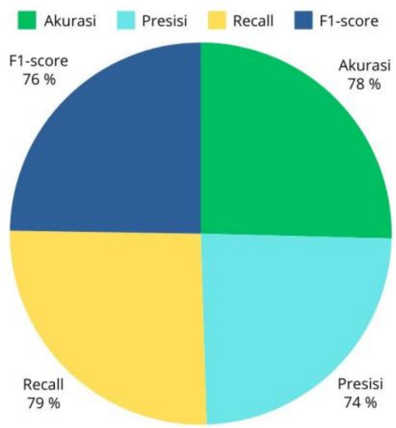

Figure 4. Sentiment Distribution Graph

By comprehending the explanation of each metric, this chart provides a clear perspective on the model's performance and the areas that need improvement for more accurate and effective sentiment analysis.

4.4.1. Interpretation of Public Sentiment

Based on the classification results, the majority of the public responded negatively to this policy. Several common reasons identified in tweets with negative sentiment include concerns about the impact of this policy on the morality of teenagers and the potential increase in premarital sexual behavior.

Example of a tweet with negative sentiment:

a. Why should teenagers be provided with contraceptives? This is like supporting free sex. The government needs to rethink this policy! ”
b. “PP 28/2024 concerning contraceptives for teenagers is actually confusing. Has the government considered the negative impacts on children? ”

Meanwhile, tweets with positive sentiment tend to support this policy due to the importance of reproductive health education and the prevention of unwanted pregnancies.

Example of a tweet with positive sentiment:

1) “Great! Finally, there’s a policy that supports contraceptive education for teenagers. It’s important to avoid misconceptions about sex. ”
2) “The provision of contraceptives is essential for teenagers so they better understand how to maintain reproductive health. ”

Neutral tweets tend to not explicitly express support or rejection but instead focus more on academic discussion or the dissemination of information.

5. Comparison

This research utilizes data in the form of tweets sourced from the social media platform Twitter. The data was collected using web scraping techniques with keywords such as “PP 28/2024”, “teen contraception”, and “contraceptive tools”. The data collection period occurred from January 1 to June 30, 2024. After the preprocessing phase to eliminate noise such as symbols, URLs, and stopwords, the number of tweets utilized in this analysis totals 1,000 tweets. The preprocessing steps undertaken include:

a) Removal of symbols and URLs: Cleaning the tweets of unnecessary symbols such as @, #, and URL links.
b) Case folding: Converting all letters in the tweets to lowercase to minimize differences in similar words.
c) Stopword removal: Eliminating common words that do not significantly impact sentiment analysis, such as “and”, “that”, “in”..

6. Conclusions

The results of this analysis indicate that the government needs to pay more attention to the socialization aspect of this policy. The numerous negative reactions may serve as an indication that this policy has not been well understood by the public. More in-depth education and open dialogue regarding the objectives and benefits of this policy could assist in reducing public resistance.

Dominance of Negative Sentiment: The majority of the analyzed tweets show negative sentiment (50%) towards this policy. This indicates the public's concern regarding the impact of this policy on adolescent morality and the increase in premarital sexual behavior. Positive and Neutral Responses: Despite the many negative tweets, there are also 35% of tweets that exhibit positive sentiment, supporting the importance of reproductive health education and access to contraceptive methods. Meanwhile, the remaining 15% are neutral, neither clearly supporting nor opposing. Model Accuracy: The Naïve Bayes Classifier model used in this analysis has an accuracy of 78%, indicating a fairly good ability to classify sentiment despite some errors.

References

[1] G. Ruz, P. Henríquez, and A. Mascareño, “Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers,” Futur. Gener. Comput. Syst., vol. 106, pp. 92–104, 2020, doi: 10.1016/j.future.2020.01.005.
[2] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks pada Twitter,” J. SAKTI, vol. 5, pp. 697–711, 2021, doi: 10.30645/J-SAKTI.V5I2.369.
[3] M. Wongkar and A. Angdresey, “Sentiment analysis using Naive Bayes algorithm of the data crawler: Twitter,” 2019 4th Int. Conf. Informatics Comput., pp. 1–5, 2019, doi: 10.1109/ICIC47613.2019.8985884.
[4] S. Sindhuja, K. Sai, R. Kotharu, and S. Devi, “Twitter sentiment analysis using enhanced TF-DIF Naive Bayes classifier approach,” 2023 7th Int. Conf. Comput. Methodol. Commun., pp. 547–551, 2023, doi: 10.1109/ICCMC56507.2023.10084106.
[5] 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,” J. Tek. Inform., 2021, doi: 10.20884/1.JUTIF.2021.2.2.51.

[6] A. A. Dwisanny and S. Supatmi, "Twitter opinion sentiment analysis based on new student admission zoning issues using the Naïve Bayes and TensorFlow methods," 2023 9th Int. Conf. Signal Process. Intell. Syst., pp. 1–8, 2023, doi: 10.1109/ICSPIS59665.2023.10402745.

[7] P. Wahyuni and M. A. Romli, "Comparison of Naïve Bayes Classifier and Decision Tree algorithms for sentiment analysis on the House of Representatives' Right of Inquiry on Twitter," J. Appl. Informatics Comput., 2024, doi: 10.30871/jaic.v8i2.8670.

[8] F. Millennianita, U. Athiyah, and A. W. Muhammad, "Comparison of Naïve Bayes Classifier and Support Vector Machine methods for sentiment classification of responses to bullying cases on Twitter," J. Mechatronics Artif. Intell., 2024, doi: 10.17509/jmai.v1i1.69959.

[9] F. Astiko and A. Khodar, "The sentiment analysis reviewing Indosat services from Twitter using the Naive Bayes Classifier," J. Appl. Comput. Sci. Technol., 2020, doi: 10.52158/JACOST.V1I2.79.

[10] C. Wicaksana, M. Fatkhurrokhman, H. P. Pratama, R. Tryawan, Alimuddin, and R. Febriani, "Twitter sentiment analysis in Indonesian language using Naive Bayes classification method," 2022 Int. Conf. Informatics Electr. Electron., pp. 1–6, 2022, doi: 10.1109/ICIEE55596.2022.10010002.

[11] Z. Zhang, "Sentiment analysis of Twitter comments using Naive Bayes classifier," Commun. Humanit. Res., 2023, doi: 10.54254/2753-7064/10/20231338.

[12] A. Irwanto and L. Goeirmanto, "Sentiment analysis from Twitter about COVID-19 vaccination in Indonesia using Naive Bayes and XGBoost classifier algorithm," SINERGI, 2023, doi: 10.22441/sinergi.2023.2.001.

[13] M. R. Ramli, H. Sulastri, and R. Rianto, "Sentiment analysis of student opinion related to online learning using Naïve Bayes classifier algorithm and SVM with AdaBoost on Twitter social media," Telematika, 2023, doi: 10.31315/telematika.v20i2.8827.

[14] I. Wardhani, Y. Chandra, and F. Yusra, "Application of the Naïve Bayes classifier algorithm to analyze sentiment for the COVID-19 vaccine on Twitter in Jakarta," Int. J. Innov. Enterp. Syst., 2023, doi: 10.25124/ijies.v7i01.171.

[15] Y. Mayasari and Y. R. Nasution, "Post-election sentiment analysis 2024 via Twitter (X) using the Naive Bayes classifier algorithm," J. Dinda Data Sci. Inf. Technol. Data Anal., 2024, doi: 10.20895/dinda.v4i2.1582.

[16] I. Oktavia and A. Isnain, "Analisis sentimen opini terhadap tools Artificial Intelligence (AI) berdasarkan Twitter menggunakan algoritma Naïve Bayes," J. MEDIA Inform. BUDIDARMA, 2024, doi: 10.30865/mib.v8i2.7524.

[17] H. Fakhrurroja, T. M. Chiqamara, F. Hamami, and D. Pramesti, "Sentiment analysis of local water company customer using Naive Bayes algorithm," 2024 IEEE Int. Conf. Ind. 4.0, Artif. Intell. Commun. Technol., pp. 168–173, 2024, doi: 10.1109/IAICT62357.2024.10617472.

[18] A. S. Wicaksono, "Twitter sentiment analysis of electric vehicle subsidy policy using Naïve Bayes algorithm," J. Stat. Methods Data Sci., 2023, doi: 10.31258/jsmds.v1i1.3.