Sentiment Analysis Of Free Meal Program For School Students Using Algoritma Naive Bayes

Anggy Dwi Anggreny[1*], Suhardi[2]

[1,2] Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Sumatera Utara, Indonesia.
[*] Corresponding Author:
Email: anggydwianggreny2@gmail.com

Abstract.

The This research aims to analyze public sentiment toward the free meal program by utilizing the Naive Bayes Classifier (NBC) algorithm. The background of this research is based on the high public interaction on social media regarding this policy, thus requiring an analysis method that can effectively classify opinions into positive and negative sentiments. Data was collected from public posts on social media, followed by a text preprocessing stage, including data cleaning, tokenizing, stopword removal, and stemming. After that, the data was analyzed using the NBC algorithm to obtain sentiment classification. The research results show that the NBC model is capable of performing sentiment classification with an accuracy of 72%, precision of 84%, recall of 72%, and f-measure of 72%. These findings indicate that the majority of public opinion tends to be positive towards the free meal program, although there are still some negative opinions highlighting weaknesses in the policy's implementation. Overall, this research contributes by providing an objective picture of public perception through a machine learning approach. Moreover, this study proves that the NBC algorithm can be effectively used to analyze public opinion regarding government policies. Therefore, the results of this research are expected to serve as a reference in decision-making and policy evaluation in the future.

Keywords: Sentiment Analysis; Naive Bayes Classifier and Free Meal Program.

I. INTRODUCTION

The free meal Program for school students is one of the policies that aims to improve nutrition, health, and student achievement. This Program is expected to help ease the economic burden on families while ensuring that every student has access to nutritious food in [1]. In different countries, similar policies have been implemented with varying results. In Indonesia, this program received widespread attention from the public, both in the form of support and criticism delivered through various media, especially social media. Sentiment analysis is one of the effective ways to understand public opinion on the policy. Although in general this program received appreciation, not a few critics also appeared related to its implementation, such as logistical problems, food quality, and program sustainability. Along with the increasing use of social media, people now have easier access to express their opinions regarding government policies. Social Media has become one of the richest sources of data on public opinion, including sentiment towards free-to-eat programs. Sentiment analysis based on social media data can provide an overview of public perception at large, whether the opinions that appear tend to be positive and negative [2].In the context of sentiment analysis, Naive Bayes algorithm is one method that is widely used because of its ability to process text data and provide accurate classification results with high efficiency[2]. The algorithm works on the basis of a probabilistic approach to determining the category of sentiments taking into account the appearance of words in the analyzed text.

Some previous studies have shown that Naive Bayes algorithm performs quite well in sentiment classification compared to other methods such as Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) [3].This study differs from previous research in several aspects. First, this study will use Naive Bayes algorithm for sentiment analysis. The Naive Bayes algorithm was chosen because it has several advantages, among others, it has low computational complexity and can be run quickly even on devices with limited resources, is relatively easy to implement and does not require complex parameter tuning as in deep learning methods, tends to work well on relatively small datasets, making it the right choice when available data is

limited [4]. Secondly, the study will focus on sentiment analysis towards free lunches for school students. Third, this study focused on comparing both data taken from social media X and questionnaires from schools that have been determined. In the digital age, social media is becoming the main platform for people to express their opinions. Thus, sentiment analysis from this study can provide a more accurate and up-to-date picture of public reaction to certain policies.Naive Bayes algorithm is one of the simplest but effective classification methods for sentiment analysis. This method is based on Bayes ' theorem and the assumption that the features in the dataset are independent of each other [5]. Some of the main reasons why the Naive Bayes algorithm was chosen in this study are that it is one of the simplest algorithms to implement and execute, naive bayes has been shown to be effective in a variety of text classification tasks, including sentiment analysis, naive bayes has good resistance to incomplete or imperfect data, and naive bayes often provides highly competitive performance, especially when applied to datasets of a text nature.

Therefore, this study will focus on sentiment analysis of free meal programs on social media X and SMK Multi Karya Medan school. Social Media X was chosen also because it has a very large and diverse user base, so the data collected is more representative also sekolah SMK Multi Karya was chosen because it represents the characteristics of vocational secondary education institutions with diverse student backgrounds, so it is interesting to know how their perception of the program.Based on this explanation, the researchers created a study entitled "sentiment analysis of Free meal programs for school students using Naive Bayes algorithm". By using Naive Bayes algorithm, this study aims to analyze public sentiment towards free meal program for school students and by using Naive Bayes algorithm. This analysis is expected to be an evaluation material for the school and the government so that the program can be more targeted and well received by the community.

II.   METHODS

This study uses a quantitative approach with experimental methods. The goal is to analyze the accuracy of Naïve Bayes Classifier algorithm in performing sentiment classification of user reviews. The quantitative approach was chosen because the research focuses on numerical data processing, model testing, and performance measurement with certain parameters. The research was carried out in 2023 using a computer device that has a minimum specification of an Intel i5 processor, 8GB of RAM, and a Windows operating system. Data processing is done using Python programming language with supporting libraries such as Scikit-learn, Pandas, and NLTK.The research population is a collection of user reviews obtained from certain social media/marketplaces (eg Tokopedia). The study samples were taken randomly from the review dataset with positive and negative labels. The amount of data used is 1,000 reviews, consisting of 500 positive reviews and 500 negative reviews, so that the proportion of data is balanced. The research Model used in the form of systematic stages ranging from data collection, preprocessing, feature extraction, model training, testing, to performance evaluation. This Model is designed to ensure the sentiment classification process can be done in a structured and scalable manner.

III.   RESULTS AND DISCUSSION

In this study, the authors used two main data sources for sentiment analysis of the free meal program for school students, namely:

1.   Data from social Media X (Twitter)

This Data was obtained through the process of web crawling using Google Colab in June 2025. A total of 500 comments were collected from X social media users who gave opinions related to the free meal program. These comments come from the general public, most of whom do not directly experience the program, but provide responses based on news, public opinion, or personal assessment of the implementation of the program.There are opinions from the general public who have not necessarily experienced the program directly, but respond based on the news they see or hear. Examples of comments in Table 1 below :

Table 1. Social Media Data (X)

No.	Full_text
1	therefore, the government is committed to implementing the free nutritious meal (MBG) Program as part of a strategic effort to improve the welfare of the community through targeted nutritional interventions.
2	Support The Free Nutritious Meal Program # Free Nutritious Programming # Benefits of eating bergiziatis # Free Nutritious Programming #Dukung Free Nutritious Programming https://t.co/55ubfBpbXe
3	Free nutritious eating program momentum creates a healthy lifestyle https://t.co/bKqmlpbRXx
4	Babinsa Koramil 1406-01/ Tempe Escort Free Nutritious Programming di SDN 200 Tempe #tniad #dispenad #puspentni #TNIindonesia #kodam14hasanuddin #korem141toddopuli https://t.co/Bbu0yYUlz4
5	Free nutritious meal program brings a new spirit for Indonesian children. Yuk kita dukung bersama! # Free Nutritious Programming # Free Nutritious Programming https://t.co/KuOfTMiGrS
6	Jayapura regency government together with stakeholders are ready to support the Program of Mr. President Prabowo Subianto with the Minister of Social Affairs of the Republic of Indonesia related to the construction of Public Schools and free nutritious meals in Jayapura regency. https://t.co/xpN08gwxNh
7	Free nutritious meal program for children to help reduce stunting and improve Indonesian human resources. #GenerasiEmas #MakanBergizi https://t.co/eqgFA0tJpV
8	Support a free nutritious meal Program to ensure children get balanced nutrition! #MakanBergiziGratis #IndonesiaSehat https://t.co/dUS8S8OO7Y
9	Balanced nutrition for all children! Free nutritious meal Program is here to create a healthy and smart generation. #ProgramMakanBergizi #GiziSeimbang https://t.co/C8ThNQMvoR
10	Support smart and healthy generation through free nutritious meal Program. Make sure they are nutritionally balanced! #DukungGenerasiCerdas #IndonesiaEmas https://t.co/B0AnOOqekX

2.   Data from school (SMK Multi Karya Medan)

In addition to social media data, the author also collected 500 comments through interviews and questionnaires directly to Class X and XI students at SMK Multi Karya Medan. These students are direct recipients of the free meal program, so their comments describe real experiences during the program. This Data represents the internal point of view of users who directly feel the impact of the program.Both types of data are then classified into two categories of sentiment, namely positive and negative. The labeling process is done manually by considering the context of the sentence and the meaning of the opinion. Furthermore, the data is used in modeling using Naïve Bayes algorithm to determine the classification performance of each data source based on evaluation metrics such as accuracy, precision, recall, and F1-score.

Table 2. Comments Data obtained from schools

No.	Komentar
1	This free meal Program is very helpful, so I save pocket money and get a decent healthy and nutritious meal
2	Free meal programs make it easier for me while studying, so my mom doesn't get tired of cooking lunch for me
3	strongly agree
4	Good, because it saves spending on lunch
5	The program is good, save more pocket money because there is a free lunch
6	I think that with the implementation of this free meal, it can be a pleasant and profitable experience because in addition to saving money, you can enjoy food without having to pay.
7	Foods that are healthy and nutritious
8	Good, because with the free meal my snack money can be saved for the tube and the food is delicious nutritious again
9	Very good and helps students pocket money
10	Steady and healthy

Text Cleaning

Text cleaning is the initial stage of pre-processing aimed at cleaning comments from irrelevant elements. Comments from social media X (Twitter) and school interviews were converted to lowercase for consistency. Elements such as URLs, special symbols (@, #, &), numbers, punctuation marks, emoticons, emojis, hashtags, and mentions are removed because they have no informative value in sentiment analysis. This process results in cleaner data and is ready for further processing.

Text Tokens

Text Tokens are a stage of breaking comments into the smallest units called tokens, such as words or phrases, to make them easier to analyze by the Naïve Bayes algorithm [6]. For example, the sentence "this free meal Program is very helpful" is transformed into the tokens "program", "meal", "free", and so on. This process helps in the analysis of the frequency and contribution of each word to the sentiment. Since tokenization also produces common words such as" and"," in", or" or", then this stage is usually followed by stopword removal to remove insignificant words. Tokenization became an important basis in the formation of features for the classification of sentiments.

Text Normalization (Slang Word)

Text normalization is the process of converting non-standard words or slang into forms that are in accordance with the Indonesian language according to KBBI. The purpose of this stage is to equalize the variety of words to be more easily analyzed by the model. In this study, normalization was done by referring to the list of slang words available in the dictionary file.csv Indonesian language from GitHub [7]. Examples of normalized words include "yg "becomes" yang"," ga "becomes" no"," gue "becomes" me", and" lo "becomes"you". This process is important to ensure that the meaning of words in comments remains consistent and can be accurately processed by Bayes ' Naïve algorithm.

Stopwords

Stopwords are common words that often appear in text but have no significance in sentiment analysis [8]. Words such as" and"," which"," for", and" is " do not contribute significantly to the determination of sentiment, so they need to be removed so as not to interfere with the representation of the data. In this context, stopwords are perceived as "noise" in the text that can reduce the accuracy of the model if not eliminated. Therefore, in this study the process of removing stopwords using a list of words from the file stopwordbahasa.csv downloaded from GitHub and contains common words in Indonesian.

Stemming

Stemming is the process of converting words to their basic or root form. In the context of Natural Language Processing (NLP), this process aims to reduce the variety of word forms that have similar meanings, so that the analysis becomes more consistent and efficient[9]. For example, words like "assist", "assisted", and "help" will all be returned to the basic form "assist". In this study, the stemming process was carried out using Indonesian-based tools to ensure the word is returned to the appropriate Form [10]. This stage is important in text preprocessing because it simplifies the process of classification, sentiment analysis, and grouping of documents by the Naïve Bayes algorithm.

Detokenized Text

Detokenized text adalah proses menggabungkan kembali kata-kata hasil tokenisasi menjadi bentuk kalimat yang utuh dan dapat dibaca secara alami [11]. Proses ini bertujuan untuk menyusun kembali token-token individual menjadi teks yang menyerupai bentuk aslinya, dengan menambahkan spasi dan struktur yang sesuai. Dalam konteks penelitian ini, detokenisasi dilakukan setelah tahap stemming dan pembersihan data selesai, sehingga teks yang dihasilkan sudah dalam bentuk bersih dan ringkas. Teks yang telah didetokenisasi inilah yang kemudian digunakan sebagai input akhir pada algoritma Naïve Bayes untuk proses klasifikasi. Dengan komentar yang sudah tersusun rapi dan bermakna secara utuh, proses pengujian dan perhitungan akurasi dapat dilakukan secara optimal sesuai kebutuhan analisis sentimen.

Labeling Using InSetLexicon

The next step is to perform the data labeling process automatically by utilizing the InSetLexicon dictionary. This dictionary is obtained from gitub Indonesian, namely https://github.com/fajri91/InSet. the dictionary also contains a list of words along with their polarity weights. To determine the polarity of a

review, all the weights of the words that appear in the text of the review are counted. Based on the results of such calculations, each review is classified into the category of positive or negative sentiment. A review is declared positive if the total polarity value exceeds zero, and negative if the total value is less than zero.
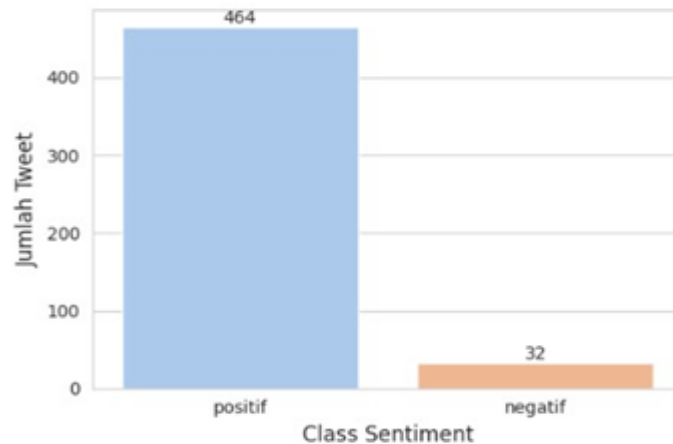
Total sentiment analysis

Fig 1. Visualization of the Number of Sentiment Analysis on Data X

The number of positive and negative comments is shown in Figure 1, where the process is done using python programming in google colaboratory. From the results of this visualization, it can be seen that the image shows the results of visualizing the amount of sentiment data obtained from social media X (formerly Twitter). This graph illustrates the distribution of sentiment in two classes, namely positive and negative. Of the total data, 464 tweets were classified as positive sentiment, while only 32 tweets were included in negative sentiment. This striking difference in numbers indicates that the majority of opinions collected from social media X tend to be positive towards the topic analyzed. Next in Figure 2 is the data obtained from interviews at school.
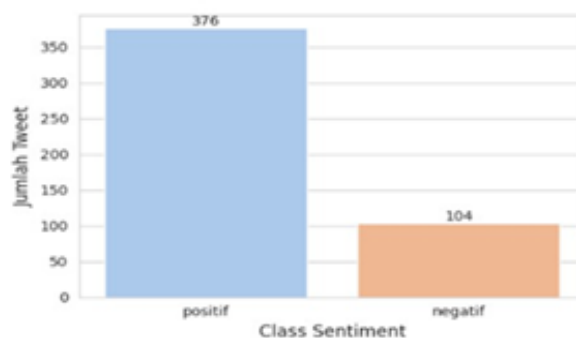

Fig 2. Visualization of the Number of Sentiment Analysis on School Data

The second image shows the results of sentiment analysis of comment data collected through direct interviews at the school. The graph shows that there were 376 positive comments and 104 negative comments. When compared with data from social media X, where there were 464 positive comments and only 32 negative comments, it was seen that although both sources showed a predominance of positive sentiment, comments from schools had a higher proportion of negative sentiment. This suggests that responses from students in schools are more diverse and tend to be more critical compared to opinions shared publicly on social media X, which tend to be more supportive or enthusiastic about the topics analyzed.

3.   Naïve Bayes Classification

In this study, sentiment analysis of free meal program was conducted using Naïve Bayes algorithm. The Data used comes from two main sources, namely:

1.   480 comments from SMK Multi Karya Medan students as a representation of the direct recipients of the free meal program.

2.   496 comments from social media X (Twitter) as a representation of the general public who did not experience the program firsthand.

https://ijsenet.com

After the duplicate removal and data cleaning process, the total data used was 976 comments. This Data is then divided into training data and test data with a ratio of 80:20. However, for the following manual example, only a small part of the data is used as an illustration.

Fig 3. Sample Train Data

Train Label Comments	Label
Good Program save money snack Free Lunch positive	Positive
Reject milk free order personal condition compulsory school milk drink eat negative	Negative

From the practice comments, word extraction was carried out and the frequency (term frequency) was calculated based on the Class (positive or negative).

Fig 4. Word Weighting In Train Data

No.	Vocabulary	Tf (positif)	Tf(Negatif)
1.	Programs	1	0
2.	Nice	1	0
3.	Save	1	0
4.	Money	1	
5.	Cake	1	
6.	Eat	1	1
7.	Lunch	1	0
8.	Free	1	1
9.	Decline	0	1
10.	Milk	0	
11.	Command	0	1
12.	Conditions	0	1
15.	Personal	0	1
16.	Schools	0	
17.	Mandatory	0	1
18.	Drink	0	
Total *Term*		**8**	

A Total of 16 unique words were obtained from the combination of all words that appeared in two comments of the training data, namely positive comments and negative comments. These words include: program, good, save, money, snack, eat, lunch, free, refuse, milk, order, condition, personal, school, compulsory, and drink. Although some words such as" eat"," milk", and" free " appear more than once, in the unique word count, each word is counted only once. This unique word count is important in the Naïve Bayes algorithm because it is used in the smoothing process (Laplace smoothing), that is, adding the value +1 in the numerator and the number of unique words in the denominator when calculating conditional probabilities. This is to avoid a zero probability value if a word does not appear in one of the sentiment classes.It can be concluded, the positive value = 8 and negative = 13 with a total of = 21 words. The class classification process begins by calculating prior probabilities, conditional probabilities, and posterior probabilities.

Although the grades are very small and almost balanced, the highest grade is for the negative grade, so the comment "Free meal program helps the family economy" is classified as positive.After sentiment testing using Naïve Bayes algorithm, sentiment classification results obtained in the form of labels, which are then compared with the actual label to calculate the value of accuracy, precision, recall, and F1-score. The evaluation was carried out on two types of data, namely data from school students who directly experienced and received the free meal program, and data from crawling results from social media X (Twitter) which contains the opinion of the general public who did not directly experience the program. The results of this test aim to compare the effectiveness of the model in recognizing sentiment from two different points of view. In other words, the comparison is done to find out whether the naïve Bayes model gives more accurate classification results on data from direct experience (school data) or on indirect opinions (crawling data). The final evaluation value will show which data source is more suitable and stable in the sentiment analysis process towards the free meal program.
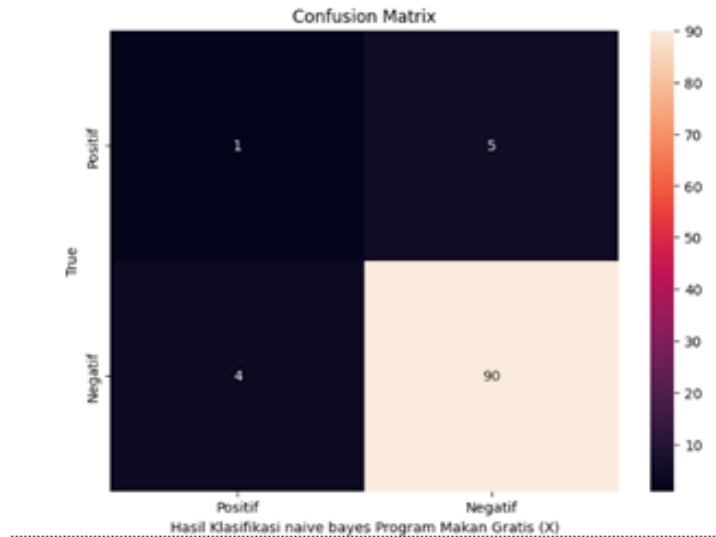
Fig 3. Confusion Matrix Data X (Twitter)

In Figure 3 the classification results to find accuracy, precision, recall and F1-score follow the following formula :

Fig 4. Confusion Matrix Results (X)

Based on confusion matrix, Naive Bayes model in classifying sentiment towards free meal Program (X) has excellent accuracy, with overall accuracy of 91%, precision of 94.74%, recall 95.74%, and F1-score of 95.23%. This shows that the model is very effective in identifying comments or data with negative sentiment towards the Free Meal Program (X), with a very low error rate in classifying negative data. Then the following results confusion matrix for data obtained from the school.
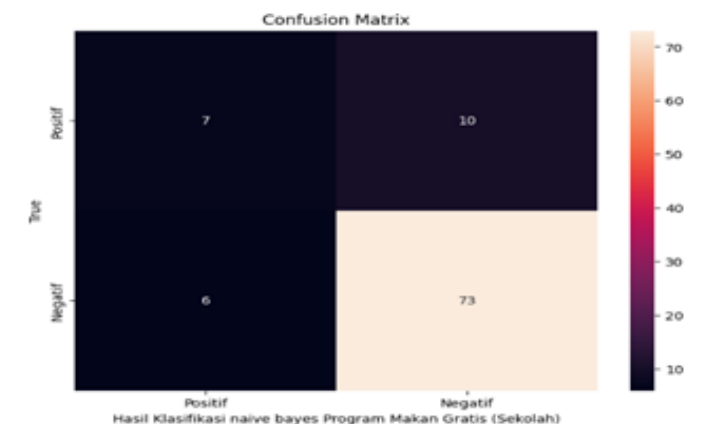

Fig 5. Confusion Matrix Data (School)

In Figure 5 the classification results to find accuracy, precision, recall and F1-score follow the following formula :

$$accuracy = \frac{73 + 70}{73 + 7 + 10 + 6} = \frac{80}{96} = 0.8333 \; atau \; 83.33\%$$

```
Accuracy: 0.8333333333333334
              precision    recall  f1-score   support

     negatif       0.54      0.41      0.47        17
     positif       0.88      0.92      0.90        79

    accuracy                           0.83        96
   macro avg       0.71      0.67      0.68        96
weighted avg       0.82      0.83      0.82        96
```

Fig 6. Confusion Matrix Results (School)

Based on the results of sentiment classification evaluation using Naïve Bayes on school data (students who experienced the free meal program firsthand), the performance of the positive Class model showed excellent results, with 88% precision, 92% recall, F1-score 90%, and overall accuracy of 83.33%. This means that the model was able to accurately recognize and classify the positive sentiments of the students, where most of the really positive comments were correctly predicted. Although the crawling data from platform X showed higher overall accuracy (91%) as well as excellent performance in the negative class (94.74% precision, 95.74% recall, and 95.23% F1-score), it is worth highlighting that this data came from people who did not directly experience the free meal program, but only provided comments based on outside observations, opinions, or perceptions. In other words, these comments do not necessarily represent reality on the ground, but rather are assumed or reactive.

In contrast, although the accuracy of the school data was slightly lower (83.33%), the model's performance against the positive class was very strong, with 88% precision, 92% recall, and 90% F1-score. This suggests that comments from students who experience the program firsthand are much more clear, consistent, and easily recognized as positive sentiment by the model. That is, the quality and depth of students ' real experiences provide more valuable and representative data, compared to outside opinions that tend to be less accurate.It can be concluded that the data from the school is superior in representing real and positive sentiment towards the free meal Program. Although the overall accuracy was higher in the crawling data, because the data from the school came from respondents who actually experienced the program firsthand, the model was better able to recognize authentic patterns of positive sentiment. Therefore, in the context of assessing the quality of perception of programs, school data is stronger, more valid, and more relevant than data from X.

IV.   CONCLUSION

Based on the results of research on sentiment analysis of the free meal Program using the Naïve Bayes algorithm, it was found that although the accuracy of the X data was higher, the school data actually provided a more realistic and authentic representation of sentiment, because it came from respondents who experienced the free meal program firsthand. This suggests that data based on real experience is more easily recognized by the model as consistent positive sentiment. Therefore, it can be concluded that the school data is superior in terms of quality and relevance to assess the perception of the program, while data X is stronger in the aspect of quantity and generalization of public opinion at large.

REFERENCES

[1] A. Sitanggang, Y. Umaidah, and R. I. Adam, "Analisis sentimen masyarakat terhadap program makan siang gratis pada media sosial X menggunakan algoritma Naïve Bayes," *J. Inform. Dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024.

[2] P. P. O. Mahawardana, G. A. Sasmita, and I. P. A. E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," *J. Ilm. Teknol. dan Komput.*, vol. 3, no. 1, pp. 810–820, 2022.

[3] R. Yunita and M. Kamayani, "Perbandingan algoritma SVM dan Naïve Bayes pada analisis sentimen penghapusan kewajiban skripsi," *Indones. J. Comput. Sci.*, vol. 12, no. 5, 2023.

[4] S. H. Ramadhani and M. I. Wahyudin, "Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN," *J. JTIK (Jurnal Teknol. Inf. Dan Komunikasi)*, vol. 6, no. 4, pp. 526–534, 2022.

[5] A. H. Hasugian, R. A. Putri, and M. A. Simatupang, "Penerapan Algoritma Klasifikasi Naïve Bayes Untuk Analisis Sentimen Tentang Pemindahan Ibu Kota Negara," *J. Sci. Soc. Res.*, vol. 7, no. 2, pp. 635–644, 2024.

[6] A. Hidayat, "Aplikasi Teks Mining Untuk Mendeteksi Spam Pada Email Berbasis Naive Bayes," *J. Teknol. Pint.*, 2022.

[7] N. Khoirunnisaa, K. N. N. Kesuma, S. Setiawan, and A. Y. P. Yusuf, "Klasifikasi Teks Ulasan Aplikasi Netflix Pada Google Play Store Menggunakan Algoritma Naive Bayes dan SVM," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 7, no. 1, pp. 64–73, 2024.

[8] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *J. SIMETRIS*, vol. 11, no. 2, pp. 626–636, 2020.

[9] A. Nisa, E. Darwiyanto, and I. Asror, "Analisis Sentimen Menggunakan Naive Bayes Classifier dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi," *eProceedings Eng.*, vol. 6, no. 2, 2019.

[10] E. A. Lisangan, A. Gormantara, and R. Y. Carolus, "Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 1, 2022.

[11] H. Z. Muflih, A. R. Abdillah, and F. N. Hasan, "Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1613–1621, 2023.