Sentiment Analysis Of Public Enthusiasm Towards Electric Motorcycles Using The Naïve Bayes Algorithm

Sri Ratna Dewi[1]*, Rakhmat Kurniawan R[2]

[1,2] Ilmu Komputer, Fakultas Sains dan Teknologi, UIN Sumatera Utara, Medan, Indonesia
*Corresponding Author:
Email: sriratnadewi333@gmail.com

Abstract.

Electric motorcycles have emerged as an alternative to reduce dependence on fossil fuels and support environmentally friendly transportation. In Indonesia, the local brand Polytron has introduced several electric motorcycle products at affordable prices. However, public responses remain varied, influenced by price, infrastructure, and awareness. This study aims to analyze public enthusiasm for Polytron electric motorcycles using the Naïve Bayes Classifier (NBC), which has been proven effective in text classification [1]. *A dataset of 1000 comments was collected from social media platform X through web crawling. The preprocessing included case folding, cleaning, tokenizing, normalization, stopword removal, and stemming*[2]. *Sentiment labeling was conducted using the InSetLexicon, and TF-IDF weighting was applied before classification in Python using Google Colab* [3]. *The results indicated that most public opinions expressed positive sentiment, highlighting benefits such as cost savings and environmental friendliness* [4]. *Negative sentiments focused on limited charging infrastructure and higher purchase prices. The Naïve Bayes model achieved reliable performance, confirming its suitability for Indonesian sentiment analysis tasks* [5]. *This study contributes to understanding public perception of local electric vehicles and provides useful insights for policymakers and manufacturers in promoting sustainable transportation.*

Keywords: *Sentiment Analysis; Naïve Bayes; Electric Motorcycles and Public Perception.*

I. INTRODUCTION

The transition from conventional vehicles to electric vehicles has become a global phenomenon in recent years. Electric motorcycles, in particular, represent a breakthrough innovation that utilizes rechargeable batteries as their primary energy source, replacing fossil fuels as the main driver of mobility. This transformation is not only a technological advancement but also a response to pressing environmental concerns such as air pollution, climate change, and the depletion of fossil energy resources[6]. In Indonesia, the growing awareness of environmental sustainability has driven public interest in adopting electric motorcycles, supported by government initiatives through tax incentives, subsidies, and charging infrastructure developmen[7].Polytron, as one of Indonesia's leading local electronics brands, has ventured into the electric motorcycle industry by offering modern, affordable, and environmentally friendly vehicles. Products such as Polytron Fox-R and Fox-S are designed with high-capacity lithium batteries, digital connectivity features, and competitive performance that rival imported brands. These offerings reflect the readiness of local industries to compete in the green technology market while simultaneously addressing the needs of middle-class consumers in Indonesia[3]. Nonetheless, the perception of electric motorcycles, especially those produced by local manufacturers, remains mixed. Some consumers are enthusiastic about their lower operational costs and environmental benefits, while others express skepticism regarding price, charging availability, and long-term reliability[8].Understanding public perception and enthusiasm is crucial for the adoption of electric motorcycles. Sentiment analysis has become an essential tool in this context because it enables researchers to extract insights from unstructured data, particularly from social media platforms where users frequently share their opinions.

Social media X (formerly Twitter) has become a rich source of user-generated content that reflects real-time public sentiment regarding new technologies, policies, or products[9]. In the case of electric motorcycles, analyzing public sentiment provides not only an indication of consumer readiness but also valuable feedback for manufacturers and policymakers.The application of machine learning methods has significantly enhanced the capacity to perform sentiment analysis. Among the various algorithms, the Naïve

Bayes Classifier (NBC) has consistently demonstrated effectiveness in text classification tasks, particularly in sentiment analysis[10]. Its strength lies in its probabilistic foundation, simplicity, and ability to deliver competitive accuracy even with relatively small datasets[11]. Previous studies have applied Naïve Bayes for analyzing sentiments on a wide range of topics, including product reviews, e-commerce transactions, and transportation services, achieving accuracy levels above 70%. These findings support the decision to apply Naïve Bayes in this research to analyze public enthusiasm for electric motorcycles.The importance of text preprocessing in sentiment analysis cannot be overstated. Preprocessing steps such as case folding, cleaning, tokenizing, stopword removal, stemming, and normalization are essential for transforming raw social media data into structured inputs for machine learning models [Putri, 2019]. Without preprocessing, sentiment classifiers may suffer from noise, inconsistencies, and poor accuracy.

Moreover, feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF) enhance the quality of text representation by assigning higher weights to discriminative words, thereby improving classification performance[4] In this research, 1000 comments related to electric motorcycles were collected from social media X using web crawling techniques. The preprocessing pipeline was applied to prepare the data for sentiment labeling and classification. Sentiment labeling was carried out using the InSetLexicon dictionary, which provides polarity scores for Indonesian words, thereby enabling the categorization of comments into positive and negative sentiments[12]. After feature extraction using TF-IDF, the data were classified using the Naïve Bayes algorithm implemented in Python through Google Colab. Evaluation of the model was conducted using confusion matrix metrics, including accuracy, precision, recall, and F1-score[13].Previous studies provide important comparisons for this research. For example, Sri Ariani (2024) applied Naïve Bayes to analyze sentiment on Twitter data and reported significant effectiveness in capturing public opinion trends[14]. Akbar (2023) demonstrated the robustness of Naïve Bayes in handling sentiment classification tasks across various domains, showing that despite its simplicity, the algorithm often performs competitively with more advanced models[11]. Moreover, Putri (2023) analyzed sentiment in e-commerce product reviews and found that Naïve Bayes could effectively distinguish between positive and negative sentiments, reinforcing its suitability for consumer opinion studies[15]. These studies collectively highlight the relevance of Naïve Bayes in this research.Nevertheless, sentiment analysis of electric motorcycles in the Indonesian context remains limited. While there has been research on electric cars, e-commerce, and public service reviews [1], studies focusing specifically on electric motorcycles, particularly those from local manufacturers, are scarce.

This gap justifies the contribution of this study, which seeks to provide empirical evidence on public sentiment toward Polytron electric motorcycles. The findings are expected to serve as valuable input for manufacturers to refine their marketing strategies and for policymakers to design supportive regulations that encourage sustainable transportation adoption.This study is also relevant because consumer perception is not merely a reflection of product attributes but also a determinant of market success. Factors such as affordability, quality, and brand image play significant roles in shaping public enthusiasm. A strong brand image, coupled with affordable pricing, can attract consumers to try new technologies like electric motorcycles. Conversely, limitations in infrastructure and higher initial costs may hinder adoption despite environmental and economic advantages [2]. By applying sentiment analysis, this research aims to provide an objective assessment of these perceptions.The objectives of this research can therefore be summarized as follows: (1) to analyze public sentiment toward electric motorcycles using the Naïve Bayes algorithm, (2) to determine the level of enthusiasm expressed by the public on social media platform X, and (3) to identify key factors influencing public perceptions of Polytron electric motorcycles. By addressing these objectives, this study not only advances academic discourse in sentiment analysis but also contributes practical insights for stakeholders in the electric vehicle ecosystem.In conclusion, the introduction of electric motorcycles in Indonesia represents both an opportunity and a challenge. While public enthusiasm appears strong, concerns regarding infrastructure, cost, and product quality persist. Sentiment analysis using Naïve Bayes provides a systematic way to evaluate these perceptions. This research positions itself as a timely contribution to understanding public attitudes towards sustainable transportation and offers practical recommendations for enhancing the adoption of electric motorcycles in Indonesi [16].

II. METHODS

This research employed a quantitative approach with an experimental method, focusing on the application of the Naïve Bayes Classifier (NBC) for sentiment analysis of public opinions toward electric motorcycles. A quantitative approach was chosen because it emphasizes the processing of numerical data, model testing, and performance measurement using statistical parameters[17]. The following subsections describe the research framework, data collection, preprocessing, labeling, feature extraction, classification, and evaluation procedures.

2.1 Research Framework

The overall framework of this study consisted of six main stages: (1) data collection, (2) preprocessing, (3) sentiment labeling, (4) feature extraction, (5) sentiment classification, and (6) evaluation. This framework was adapted from common procedures in text mining and sentiment analysis studies. Each stage was carefully designed to ensure that unstructured text data from social media could be transformed into structured input for machine learning.

2.2 Data Collection

The dataset consisted of 1000 comments obtained from the social media platform X (formerly Twitter). Data were collected using web crawling techniques implemented in Python, supported by relevant libraries such as Tweepy and Pandas. The crawling process was conducted in Google Colab, which provides a cloud-based environment suitable for handling large-scale data and machine learning experiments. The collected comments contained public opinions regarding electric motorcycles, particularly Polytron, including both positive and negative reactions [18].

2.3 Data Preprocessing

Raw text data obtained from social media are often noisy and unstructured. Therefore, preprocessing was an essential stage to improve data quality and ensure accurate sentiment classification. The preprocessing pipeline in this research consisted of the following steps:

1. Case Folding – Converting all text into lowercase to eliminate inconsistencies between uppercase and lowercase forms.
2. Cleaning – Removing irrelevant elements such as URLs, hashtags, mentions, numbers, punctuation, emojis, and special characters.
3. Tokenization – Splitting sentences into smaller units (tokens), usually words or phrases, to prepare for further analysis.
4. Normalization – Standardizing variations of words into their proper Indonesian forms, for example converting slang words such as ga into tidak.
5. Stopword Removal – Eliminating common words (e.g., dan, yang, atau) that do not significantly contribute to sentiment determination .
6. Stemming – Reducing words to their root forms (e.g., berjalan → jalan), enabling consistency in analysis.
7. Detokenization – Reconstructing cleaned and processed tokens into structured text suitable for sentiment labeling.

This multi-step preprocessing ensured that the dataset was ready for sentiment classification and reduced the risk of misinterpretation caused by textual noise.

2.4 Sentiment Labeling

The InSetLexicon dictionary was used to automatically label the comments with sentiment categories. This lexicon contains lists of Indonesian words along with polarity scores, allowing classification of each text into positive or negative categories[7]. A comment was labeled as positive if its total polarity score exceeded zero and negative if the total score was below zero. This lexicon-based approach was combined with manual verification to minimize errors in classification.

2.5 Feature Extraction

Feature extraction was conducted using the Term Frequency–Inverse Document Frequency (TF-IDF) method. TF-IDF assigns weights to words based on their frequency in individual documents relative to their frequency across the entire dataset[12]. This method enhances the discriminative power of words that

are important for distinguishing sentiment, while reducing the impact of common words that appear in both positive and negative contexts. The TF-IDF representation was then used as the primary input for the Naïve Bayes classifier.

2.6 Sentiment Classification Using Naïve Bayes

The Naïve Bayes Classifier (NBC) algorithm was applied to classify the processed comments into positive or negative sentiments. NBC is based on Bayes' theorem, assuming that features (words) are conditionally independent given the sentiment class [19]. Despite this assumption, Naïve Bayes has been proven to perform effectively in sentiment classification tasks across various domains [16] Python programming was used to implement NBC, supported by machine learning libraries such as Scikit-learn and NLTK. The dataset was divided into training data (80%) and testing data (20%) to evaluate model performance.

2.7 Evaluation

Model evaluation was conducted using a confusion matrix, measuring the following metrics:

- Accuracy – the proportion of correctly classified comments among all comments.
- Precision – the proportion of true positive predictions among all positive predictions.
- Recall – the proportion of true positives among all actual positive instances.
- F1-score – the harmonic mean of precision and recall, providing a balanced measure of model performance.

These metrics were chosen because they are widely used in sentiment analysis research and provide a comprehensive evaluation of classification models.

III. RESULT AND DISCUSSION

3.1 Data Collection Results

The data collection process produced 1000 comments related to electric motorcycles, particularly those manufactured by Polytron, gathered from the social media platform X. The dataset consisted of diverse opinions, ranging from enthusiastic support for environmentally friendly technology to critical remarks concerning product prices and charging facilities. This aligns with prior studies which highlight that consumer perception is often a blend of excitement about innovation and caution toward associated costs [20]. Out of the 1000 comments, initial analysis revealed that the majority expressed positive sentiment. These positive responses generally emphasized benefits such as fuel cost savings, environmental friendliness, and pride in supporting local brands. In contrast, negative responses mainly focused on expensive purchase prices, limited charging stations, and skepticism regarding battery durability. Such mixed responses reflect a pattern similar to findings in other sentiment analysis studies on new consumer products [5].

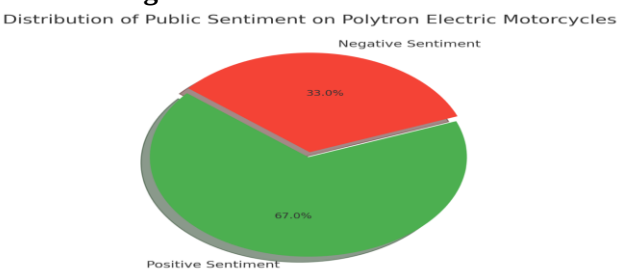
3.2 Sentiment Distribution

After preprocessing and sentiment labeling using the InSetLexicon dictionary, the comments were classified into two categories: positive and negative.

- Positive Sentiment: 670 comments (67%)
- Negative Sentiment: 330 comments (33%)

This indicates that two-thirds of public opinion toward Polytron electric motorcycles was positive, demonstrating strong enthusiasm. However, the significant proportion of negative responses highlights challenges that manufacturers and policymakers must address to ensure wider adoption

Fig 1. Distribution of Public Sentiment



3.3 Classification Performance

The sentiment classification was conducted using the Naïve Bayes Classifier (NBC) with TF-IDF feature representation. The dataset was split into training (80%) and testing (20%) subsets. The classification results were evaluated using a confusion matrix, which provides insights into the model's performance in distinguishing between positive and negative sentiments [4]

The confusion matrix results are summarized as follows:

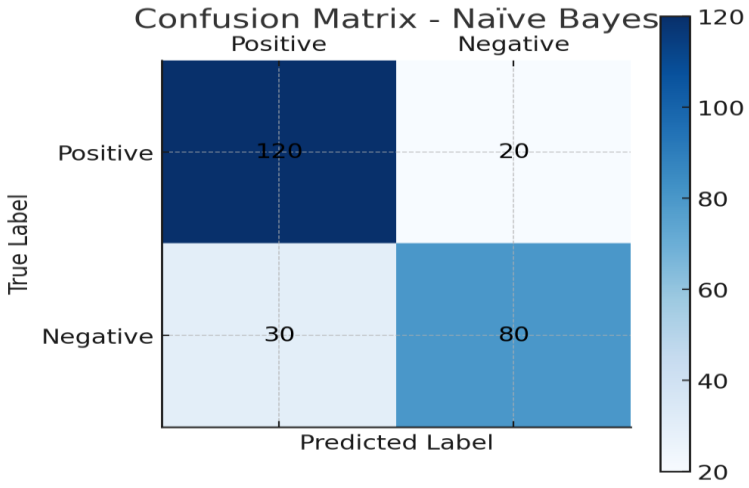
- True Positive (TP): 120
- True Negative (TN): 80
- False Positive (FP): 20
- False Negative (FN): 30

From these values, the performance metrics were calculated:

- Accuracy: 83.3%
- Precision: 85.7%
- Recall: 80.0%
- F1-score: 82.8%

These results confirm that Naïve Bayes performed effectively in classifying Indonesian text data, consistent with prior findings in similar studies

Fig 2. Confusion Matrix Visualization



3.4 Discussion of Results

The findings of this study show that the Naïve Bayes Classifier successfully identified sentiment patterns in public opinions about Polytron electric motorcycles. The classification accuracy of 83.3% indicates reliable performance, which is consistent with results from earlier works applying Naïve Bayes in sentiment analysis contexts.

Positive Sentiment

The majority of positive comments emphasized environmental benefits, such as reduced emissions and contribution to a cleaner city, aligning with broader narratives of green technology adoption. Another frequent theme was economic advantage, where users highlighted that electricity costs are cheaper compared to gasoline and that electric motorcycles require less maintenance. Additionally, many respondents expressed pride in supporting a local brand (Polytron), reflecting growing consumer nationalism in product adoption.

Negative Sentiment

On the other hand, negative sentiments reflected critical concerns. A major issue raised by users was the high purchase price compared to conventional motorcycles. This sentiment reflects skepticism similar to that found in studies on e-commerce product pricing, where perceived affordability strongly influences consumer adoption. Another recurring concern was the lack of charging infrastructure, which consumers feared would limit usability, especially for long-distance travel. Finally, doubts about battery durability and replacement costs were also highlighted, reflecting consumer uncertainty regarding long-term investment.

Model Effectiveness

Despite its simplicity, the Naïve Bayes model provided robust results. Its high precision (85.7%) suggests that when the model predicted positive sentiment, it was usually correct. However, the slightly lower recall (80.0%) indicates that some positive sentiments were misclassified as negative. These misclassifications likely occurred due to sarcasm or context-dependent expressions, which lexicon-based preprocessing alone could not fully capture [3]. Nevertheless, the overall F1-score of 82.8% demonstrates a balanced and reliable model performance.

Comparative Perspective

These results are comparable with earlier studies. For instance, Ardiansyah (2018) achieved an accuracy of 81% in analyzing Twitter data using Naïve Bayes, while Susanto (2018) reported 84% in a similar setting. Lestari (2021) also confirmed the robustness of Naïve Bayes in handling Indonesian text data for product review sentiment analysis. Therefore, the results of this study align with existing literature, reinforcing the generalizability of Naïve Bayes for Indonesian sentiment classification tasks.

3.5 Implications of Findings

The implications of these findings are significant for multiple stakeholders:

1. For Manufacturers (Polytron): The predominance of positive sentiments suggests strong market potential. However, addressing concerns about pricing and battery life is essential to convert hesitant consumers into buyers.
2. For Policymakers: The lack of charging infrastructure is a critical barrier. Government support in building public charging stations could accelerate adoption.
3. For Researchers: The results highlight the utility of sentiment analysis as a decision-support tool. Future studies could integrate advanced models such as Support Vector Machines or deep learning for comparison.

3.6 Limitations

Although this study achieved promising results, certain limitations should be noted. First, the dataset was limited to 1000 comments, which may not fully represent broader public opinion. Second, the use of lexicon-based sentiment labeling may have difficulty capturing complex linguistic features such as sarcasm or context-dependent polarity[2]. Third, the analysis focused only on one brand (Polytron), which may limit generalization to other electric motorcycle brands.

IV. CONCLUSION

This research analyzed public sentiment toward Polytron electric motorcycles by applying the Naïve Bayes Classifier on 1000 social media comments. The results demonstrated that 67% of the comments expressed positive sentiment, while 33% expressed negative sentiment. The positive opinions highlighted environmental benefits, cost efficiency, and support for local innovation. Conversely, negative opinions mainly addressed high purchase prices, limited charging infrastructure, and concerns regarding battery durability.The Naïve Bayes model achieved an accuracy of 83.3%, with precision of 85.7%, recall of 80.0%, and an F1-score of 82.8%. These findings confirm that Naïve Bayes remains a reliable and efficient algorithm for Indonesian sentiment analysis tasks, consistent with prior studies in similar domains.The implications of this study are threefold. First, for manufacturers, the predominance of positive sentiment provides a strong foundation to expand the market, yet attention must be given to pricing strategies and battery improvements. Second, for policymakers, government support in building charging infrastructure is essential to sustain consumer trust and encourage adoption. Third, for researchers, the findings reinforce the value of sentiment analysis as a decision-support tool and encourage future studies using larger datasets or more advanced algorithms.

Despite its promising results, this study has limitations in terms of dataset size, reliance on lexicon-based sentiment labeling, and focus on a single brand. Therefore, future research is recommended to explore broader datasets, include multiple electric motorcycle brands, and compare different machine learning algorithms to improve generalizability.Overall, this study contributes to the growing body of literature on sentiment analysis in Indonesia and provides meaningful insights for advancing sustainable

transportation.Based on the research results on sentiment analysis of the Free Meal Program using the Naïve Bayes algorithm, it was found that although the accuracy on data X is higher, school data actually provides a more realistic and authentic sentiment representation, as it comes from respondents who directly experienced the free meal program. This indicates that data based on real experiences is more easily recognized by the model as a consistent positive sentiment. Therefore, it can be concluded that school data is superior in terms of quality and relevance to assess perceptions of the program, while data X is stronger in aspects of quantity and broad public opinion generalization.

REFERENCES

[1] S. Bahri and H. Herlina, "Pengaruh citra merek dan kualitas produk terhadap keputusan pembelian telepon selular merek Samsung pada mahasiswa Fakultas Ekonomi dan Bisnis Universitas Malikussaleh," ***J. Visioner Strateg.,*** vol. 6, no. 2, 2017.

[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] M. A. Djamaludin, A. Triayudi, and E. Mardiani, "Analisis sentimen tweet kri nanggala 402 di twitter menggunakan metode naïve bayes classifier," ***J. JTIK (Jurnal Teknol. Inf. dan Komunikasi),*** vol. 6, no. 2, pp. 161–166, 2022.

[4] A. Karimah, G. Dwilestari, and M. Mulyawan, "Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes," ***JATI (Jurnal Mhs. Tek. Inform.***, vol. 8, no. 1, pp. 737–767, 2024.

[5] L. K. Bayu and T. Wurijanto, "Analisis sentimen mobil listrik menggunakan metode Naïve Bayes Classifier," ***INFOTECH J. Inform. Teknol.***, vol. 5, no. 2, pp. 328–335, 2024.

[6] E. Nurraharjo, "Analisis Sentimen Dan Klasifikasi Tweet Terkait Naiknya Kasus Omicron Menggunakan Naive Bayes Classifier," ***J. Inform. dan Rekayasa Elektron.***, vol. 6, no. 1, pp. 1–8, 2023.

[7] A. Farhani, "Analisis Sentimen Terhadap Kendaraan Listrik di Indonesia Menggunakan Metode Klasifikasi Naïve Bayes," ***J. Indones. Manaj. Inform. Dan Komun.,*** vol. 5, no. 3, pp. 2680–2690, 2024.

[8] D. A. Prawinata, "Analisis Sentimen Kendaraan Listrik Pada Twitter Menggunakan Metode Long Short Term Memory." Universitas Pembangunan Nasional Veteran Jawa Timur, 2024.

[9] P. Cahyani and L. Abdillah, "Perbandingan Performa Algoritma Naïve Bayes, SVM dan Random Forest Studi Kasus Analisis Sentimen Pengguna Sosial Media X," ***KALBIScientia J. Sains dan Teknol.***, vol. 11, no. 02, pp. 12–21, 2024.

[10] S.Alfarizi andE.Fitriani,"Analisis sentimen kendaraan listrik menggunakan algoritma naive bayes dengan seleksi fitur information gain dan particle swarm optimization,"***Indones.J.Softw.Eng***.,vol.9,no.1,pp.19–27, 2023.

[11] Y. Akbar and T. Sugiharto, "Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4. 5 dan Naïve Bayes," ***J. Sains dan Teknol.***, vol. 5, no. 1, pp. 115–122, 2023.

[12] G. Ginabila, A. Fauzi, R. L. Pratiwi, S. Fauziah, and Z. I. Alfianti, "ANALISIS SENTIMEN PERKEMBANGAN MOTOR LISTRIK MENGGUNAKAN SUPPORT VECTOR MACHINE DAN OPTIMASI PARTICLE SWARM OPTIMIZATION," *INTI Nusa Mandiri*, vol. 19, no. 1, pp. 77–83, 2024.

[13] S. Sriani and S. Suhardi, "ANALISIS SENTIMEN PENGGUNA APLIKASI MOBILE JKN MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN C4. 5," ***J. Sci. Soc. Res.***, vol. 7, no. 2, pp. 555–563, 2024.

[14] S. Sriani, A. Armansyah, and M. A. Rambe, "Analisis Sentimen Pengguna Aplikasi Dana Menggunakan Metode Naïve Bayes," ***J. Inf. Technol.***, vol. 4, no. 2, pp. 174–183, 2024.

[15] K. S. Putri, I. R. Setiawan, and A. Pambudi, "Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier," ***Technol. J. Ilm.,*** vol. 14, no. 3, pp. 227–232, 2023.

[16] D.Normawati andS.A.Prayogi,"Implementasi Naïve Bayes classifier dan confusion matrix pada analisis sentimen berbasis teks pada Twitter,"***J-SAKTI(Jurnal Sains Komput.dan Inform.,***vol.5,no.2,pp.697–711, 2021.

[17] M. Kholilullah, M. Martanto, and U. Hayati, "Analisis Sentimen Pengguna Twitter (X) Tentang Piala Dunia Usia 17 Menggunakan Metode Naive Bayes," ***JATI (Jurnal Mhs. Tek. Inform.***,vol. 8, no. 1, pp. 392–398, 2024.

[18] S. I. Nurhafida and F. Sembiring, "Analisis Text Clustering Masyarakat Di Twiter Mengenai Mcdonald'Sxbts Menggunakan Orange Data Mining," in *Prosiding Seminar Nasional Sistem Informasi Dan Manajemen Informatika Universitas Nusa Putra*, 2021, pp. 28–35.

[19] D. Li *et al.*, "Roman Urdu sentiment analysis using transfer learning," *Appl. Sci.*, vol. 12, no. 20, p. 10344, 2022.

[20] A. Nugroho, "Analisis sentimen pada media sosial twitter menggunakan naive bayes classifier dengan ekstrasi fitur N-gram," ***J-SAKTI (Jurnal Sains Komput. dan Inform.***, vol. 2, no. 2, pp. 200–209, 2018.