ITEJ June-2026, Volume 11 No 1 Page 154 - 163

ITEJ
Information Technology Engineering Journals
eISSN : 2548-2157

Url : https://itej.uinssc.ac.id
Email : itej@syekhnurjati.ac.id

Sentiment Analysis of Instagram Users Toward the Animated Film *Merah Putih: One For All* Using the Naïve Bayes Method

Nabila Wafa
Undergraduate Program in Informatics Engineering, Faculty of Computer and Education, Ngudi Waluyo University
nabilawafa7461@gmail.com

Yoannes Romando Sipayung
Undergraduate Program in Informatics Engineering, Faculty of Computer and Education, Ngudi Waluyo University
yoannesromando@unw.ac.id

Abstract - Instagram is a widely used social media platform where users express opinions on various forms of entertainment, including animated films. The animated film Merah Putih One For All, as one of Indonesia's local animation works, has received diverse responses from Instagram users that reflect positive, neutral, and negative sentiments. This study aims to analyze and classify the sentiment of Instagram user comments related to the film using the Naïve Bayes algorithm. This research utilized 200 Instagram comments, which were categorized into three sentiment classes. Text preprocessing was applied prior to classification. The dataset was evaluated using the Split Validation method with a 60:40 ratio, where 60% of the data were used for training and 40% for testing. Model performance was assessed using a confusion matrix along with accuracy, precision, and recall metrics. The experimental results show that the Naïve Bayes algorithm achieved an accuracy of 76,67%. The positive sentiment class obtained the highest recall value of 100%, followed by the neutral class with 83,33%, while the negative sentiment class recorded the lowest recall at 46,67%. These findings indicate that the model performs better in identifying positive and neutral sentiments than negative sentiment. Overall, the results demonstrate that the Naïve Bayes algorithm is sufficiently effective for sentiment analysis of Instagram comments, although further improvements are required to enhance the classification of negative sentiment.

Keywords - Sentiment Analysis, Instagram, Naïve Bayes, Animated Film

Article info: Date Submitted: 23/03/2026 | Date Revised: 03/05/2026 | Date Accepted: 13/06/2026
This is an open access article under the CC BY-SA license

I. INTRODUCTION

The development of social media has had a significant impact on the way people express opinions, criticism, and appreciation for various forms of digital works, including animation. Instagram, as one of the platforms with the largest number of active users in Indonesia, provides a broad space for interaction through its comment section. Each comment that appears is not only a direct response from the audience but also a potential data source that can be analyzed to understand public perception comprehensively [1].

In recent years, the local animation industry has shown significant growth. Various Indonesian animation studios have begun producing original works that have attracted widespread public attention. One such work is the animation *Merah Putih: One For All*,

whose release generated diverse opinions from Instagram users. These responses appeared in various forms, including appreciation, criticism, and neutral comments that reflect each user's perception. The diversity of these responses constitutes a valuable data source for analyzing public opinion trends toward the animation.

Public opinion recorded on social media can serve as an indicator for assessing the quality of a work, visual appeal, storyline, and the message conveyed [2]. Positive sentiment indicates good acceptance, while negative sentiment can provide constructive feedback for creators to make improvements. Therefore, sentiment analysis is a relevant approach for understanding opinion trends toward local animated works.

Sentiment analysis itself is the process of identifying and classifying opinions based on certain polarities, such as positive, neutral, or negative [3]. This method is widely applied in the field of Natural Language Processing due to its ability to efficiently process large volumes of text. One commonly used algorithm is Naïve Bayes, which is known for being simple, fast, and effective in text classification tasks [4].

Based on this background, this study was conducted to analyze the sentiment of Instagram users toward the animation *Merah Putih: One For All* using the Naïve Bayes method. The research includes comment collection, sentiment labeling, text preprocessing, model training, and performance evaluation using a confusion matrix. The results of this study are expected to provide an overview of public perception of the animation and serve as input for animation content developers in improving the quality of their work. In addition, this research is expected to contribute to the development of social media–based sentiment analysis studies, particularly in Indonesia.

II. METHOD

The data collection stage is the initial step of this study and aims to obtain Instagram user comments related to the animated film *Merah Putih: One For All*. Data were collected directly from the Instagram platform, as it functions as an active public space where users express opinions and responses to entertainment content. The research process began with problem identification, determination of research objectives, and selection of the research approach, including the use of Instagram as the data source and the Naïve Bayes algorithm as the classification method. Comment data were collected using crawling techniques, either manually or with the assistance of data extraction tools. All relevant comments were stored in a CSV dataset containing only two attributes: comment text and posting time, while personal information was removed to protect user privacy. The collected data were then manually labeled into three sentiment categories positive, negative, and neutral by considering linguistic context, informal language, sarcasm, and social media communication styles to ensure labeling accuracy.

Following labeling, the data underwent preprocessing to clean and prepare the text for classification. This stage included cleansing to remove special characters and symbols, lowercasing to standardize text, tokenizing to split sentences into individual words, and stopword removal to eliminate non-informative terms. The preprocessed data were then used as input for training the Naïve Bayes algorithm, which represents the core of the analysis process [5]. Model training was conducted using the Cross Validation method to enable simultaneous training and testing [6]. The algorithm calculates the probability of each word belonging to a specific sentiment class based on its occurrence patterns in the training data, resulting in a model capable of classifying new comments. Model performance was evaluated using a confusion matrix, accuracy, precision, and recall to assess its effectiveness in identifying sentiment categories and to detect potential data imbalance or bias. The final stage involved drawing conclusions from the analysis results

to address the research objectives and provide an overview of public sentiment toward *Merah Putih: One For All*.

III. Result and Discussion

This section explains the results of each analysis stage that was carried out according to the research workflow, as illustrated in the diagram shown in the previous figure. Each stage produces different outputs that serve as the basis for the subsequent process. The detailed explanation is as follows:

Data Collection

The first stage conducted was the collection of comment data from Instagram. Based on the scraping process, a number of comments relevant to the discussion of the film *Merah Putih: One For All* were obtained.

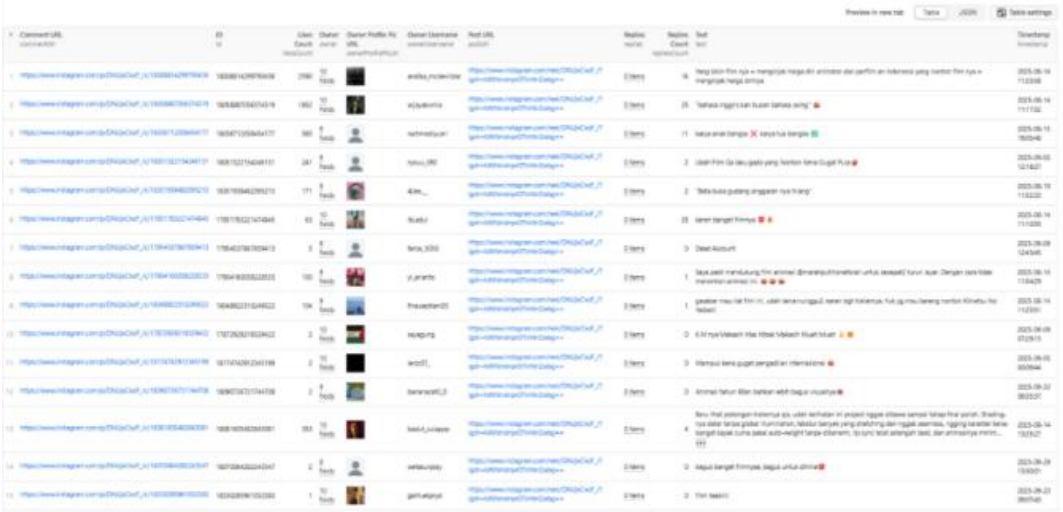

Figure 1. Instagram Comment Data

In the screenshot of the initial dataset (Figure 3.1), the data are still in raw form and contain various types of comments, including long comments, short comments, as well as comments containing emojis or symbols. This raw data were then used as the initial input before proceeding to the preprocessing stage.

Labeling

The labeling stage was performed manually to determine the sentiment category of each comment. In the screenshot of the labeling process (Figure 3.2), the researcher assigns positive, negative, or neutral labels in a specific column labeled “Sentiment.”

text	sentimen
Yang bikin film nya = menginjak harga diri animator dan perfilm an indonesia yang nonton film nya = menginjak harga dirinya	negatif
"bahasa inggris kan bukan bahasa asing" 😂	netral
karya anak bangsa ❌ karya anak tua ☑	negatif
Udah Film Ga Laku gada yang Nonton Kena Gugat Pula 🤣	negatif
"Beta buka gudang anggaran nya hilang"	negatif
keren banget filmnya 😍	positif
Dead Account	negatif
Saya pasti mendukung film animasi @merahputihoneforall untuk secepat2 turun layar. Dengan cara tidak menonton animasi ini. 😂 😂 😂	negatif
gasabar mau liat film ini udah lama nunggu2 keren bgt trailernya. Yuk yg mau bareng nonton Kimetsu No Yaba!!!	positif
6 M nya Makasih Mas Mbak Makasih Muah Muah 👌 🙏	positif
Mampus kena gugat pengadilan internasional 😂	negatif
Animasi tahun 90an bahkan lebih bagus visualnya Ⓜ	negatif
Baru lihat potongan trailernya aja. udah kelihatan ini project nggak dibawa sampai tahap final polish. Shadingnya datar tanpa global Illumination. tekstur banyak yang stretching dan nggak seamless. rigging karakter keras banget kayak cuma pakai auto-weight tanpa dibenerin.	negatif
bagus banget filmnyaa bagus untuk dihina	negatif
film taaaiii	negatif

Figure 2. Labeling of Instagram Comment Data

This figure shows that each comment has been assigned a category according to its underlying meaning. This labeling process serves as an important foundation for model construction in the training stage.

Preprocessing

The next stage is preprocessing, which involves cleaning the collected data using the RapidMiner application to improve analysis accuracy by removing irrelevant or noisy elements. The preprocessing process was carried out using the RapidMiner design shown below:

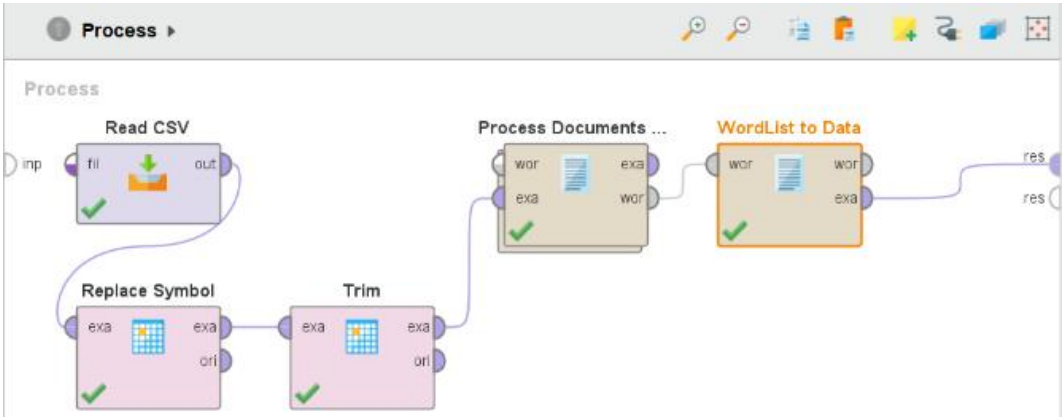

Figure 3. RapidMiner Preprocessing Design

The design can be explained as follows:

1. Read CSV is a tool used to read the animation comment data for *Merah Putih: One For All*, which consist of text data obtained from the Instagram data crawling process in CSV format. At this stage, only the comment text data are selected.
2. Replace Symbol is used to clean the text by removing unnecessary characters or elements. In this stage, the system removes components such as:
   - emojis 😊 🔥 😭
   - punctuation marks (!?,.)
   - symbols (@, #, $, %, *)
   - mentions (@username)
   - hashtags (#topic)
   - links or URLs

text	sentimen
Yang bikin film nya menginjak harga diri animator dan perfilm an indonesia yang nonton film nya menginjak harga dirinya	negatif
bahasa inggris kan bukan bahasa asing	netral
karya anak bangsa karya anak tua	negatif
Udah Film Ga Laku gada yang Nonton Kena Gugat Pula	negatif
Beta buka gudang anggaran nya hilang	negatif
keren banget filmnya	positif
Dead Account	negatif
Saya pasti mendukung film animasi merahputihoneforall untuk secepat2 turun layar Dengan cara tidak menonton animasi ini	negatif
gasabar mau liat film ini udah lama nunggu2 keren bgt trailernya Yuk yg mau bareng nonton Kimetsu No Yaba	positif
6 M nya Makasih Mas Mbak Makasih Muah Muah	positif
Mampus kena gugat pengadilan internasional	negatif
Animasi tahun 90an bahkan lebih bagus visualnya	negatif
Baru lihat potongan trailernya aja udah kelihatan ini project nggak dibawa sampai tahap final polish Shadingnya datar tanpa global Illumination tekstur banyak yang stretching dan nggak seamless rigging karakter keras banget kayak cuma pakai auto weight tanpa dibenerin	negatif
bagus banget filmnyaa bagus untuk dihina	negatif
film taaaiii	negatif

Figure 4. Preprocessing Using Replace Symbol

3. Trim is used to remove leading and trailing spaces in the comments.
4. Next, Process Documents from Data is used to separate words in the comment opinion data related to the animation *Merah Putih: One For All*. The subprocess design in RapidMiner is shown as follows:

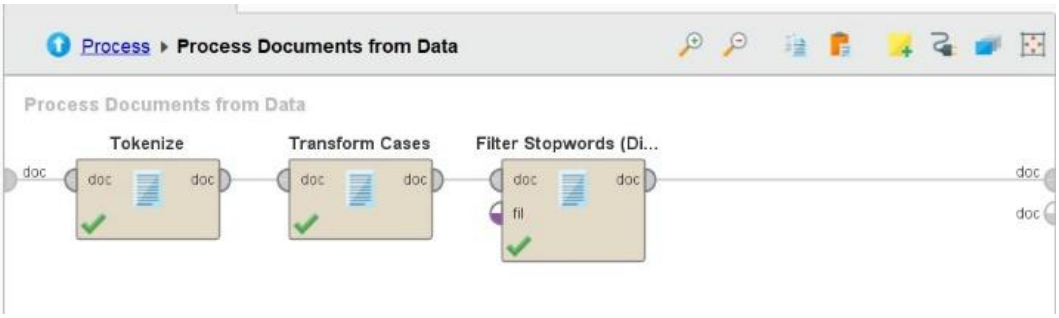

Figure 5. Subprocess of Process Documents from Data

The tools within the Process Documents from Data stage can be explained as follows:

1. Tokenize

   Tokenization is the process of breaking sentences into smaller units called tokens. Classification algorithms do not read sentences as a single long unit but instead analyze the individual words that compose them.

text	sentimen
[Yang, bikin, film, nya, menginjak, harga, diri, animator, dan, perfilm, an, indonesia, yang, nonton, film, nya, menginjak, harga, dirinya]	negatif
[bahasa, inggris, kan, bukan, bahasa, asing]	netral
[karya, anak, bangsa, karya, anak, tua]	negatif
[Udah, Film, Ga, Laku, gada, yang, Nonton, Kena, Gugat, Pula]	negatif
[Beta, buka, gudang, anggaran, nya]	negatif
[keren, banget, filmnya]	positif
[Dead, Account]	negatif
[Saya, pasti, mendukung, film, animasi, merahputihoneforall, untuk, secepat2, turun, layar, Dengan, cara, tidak, menonton, animasi, ini]	negatif
[gasabar, mau, liat, film, ini, udah, lama, nunggu2, keren, bgt, trailernya, Yuk, yg, mau, bareng, nonton, Kimetsu, No, Yaba]	negatif
[6, M, nya, Makasih, Mas, Mbak, Makasih, Muah, Muah]	positif
[Mampus, kena, gugat, pengadilan, internasional]	positif
[Animasi, tahun, 90an, bahkan, lebih, bagus, visualnya]	negatif
[Baru, lihat, potongan, trailernya, aja, udah, kelihatan, ini, project, nggak, dibawa, sampai, tahap, final, polish, Shadingnya, datar, tanpa, global, Illumination, tekstur, banyak, yang, stretching, dan, nggak, seamless, rigging, karakter, keras, banget, kayak, cuma, pakai, autoweight, tanpa, dibenerin]	negatif
[bagus, banget, filmnyaa, bagus, untuk, dihina]	negatif
[film, taaaiii]	negatif

Figure 6. Tokenization Process

   In the screenshot of the tokenization process, each comment is shown to be split into a list of words. This stage helps the system analyze the frequency, occurrence, and relevance of each word during the training process.

2. Transform Cases

   The Transform Cases stage converts all letters into lowercase. Its purpose is to standardize text formatting so that words with different capitalization are not treated as different entities by the system.

text	sentimen
[yang, bikin, film, nya, menginjak, harga, diri, animator, dan, perfilm, an, indonesia, yang, nonton, film, nya, menginjak, harga, dirinya]	negatif
[bahasa, inggris, kan, bukan, bahasa, asing]	netral
[karya, anak, bangsa, karya, anak, tua]	negatif
[udah, film, ga, laku, gada, yang, nonton, kena, gugat, pula]	negatif
[beta, buka, gudang, anggaran, nya]	negatif
[keren, banget, filmnya]	positif
[dead, account]	negatif
[saya, pasti, mendukung, film, animasi, merahputihoneforall, untuk, secepat2, turun, layar, dengan, cara, tidak, menonton, animasi, ini]	negatif
[gasabar, mau, liat, film, ini, udah, lama, nunggu2, keren, bgt, trailernya, yuk, yg, mau, bareng, nonton, kimetsu, no, yaba]	negatif
[6, m, nya, makasih, mas, mbak, makasih, muah, muah]	positif
[mampus, kena, gugat, pengadilan, internasional]	positif
[animasi, tahun, 90an, bahkan, lebih, bagus, visualnya]	negatif
[baru, lihat, potongan, trailernya, aja, udah, kelihatan, ini, project, nggak, dibawa, sampai, tahap, final, polish, shadingnya, datar, tanpa, global, Illumination, tekstur, banyak, yang, stretching, dan, nggak, seamless, rigging, karakter, keras, banget, kayak, cuma, pakai, autoweight, tanpa, dibenerin]	negatif
[bagus, banget, filmnyaa, bagus, untuk, dihina]	negatif
[film, taaaiii]	negatif

Figure 7. Transform Cases Process

3. Filter Stopword

   Filter Stopword is the process of removing common words that do not have significant meaning in sentiment determination. Words such as:

   - *"yang"*
   - *"dan"*

- *"saya"*
- *"itu"*
- *"di", "ke", "dari"*
- *"udah", "aja"*

do not contribute significantly to classification and therefore need to be removed.

text	sentimen
[bikin, film, menginjak, harga, diri, animator, perfilm, indonesia, nonton, film, menginjak, harga, diri]	negatif
[bahasa, inggris, bukan, bahasa, asing]	netral
[karya, anak, bangsa, karya, anak, tua]	negatif
[film, ga, laku, gada, nonton, kena, gugat]	negatif
[beta, buka, gudang, anggaran, hilang]	negatif
[keren, banget, filmnya]	positif
[dead, account]	negatif
[pasti, mendukung, film, animasi, merahputihoneforall, secepat2, turun, layar, cara, menonton, animasi]	negatif
[gasabar, liat, film, lama, nunggu2, keren, bgt, trailernya, nonton, kimetsu, yaba]	positif
[6, m, makasih, mas, mbak, makasih, muah, muah]	positif
[mampus, kena, gugat, pengadilan, internasional]	negatif
[animasi, tahun, 90an, bagus, visualnya]	negatif
[lihat, potongan, trailernya, kelihatan, project, nggak, dibawa, sampai, final, polish, shadingnya, datar, tanpa, global, illumination, tekstur, stretching, nggak, seamless, rigging, karakter, keras, banget, kayak, cuma, pakai, auto, weight, tanpa, dibenerin]	negatif
[bagus, banget, filmnyaa, bagus, dihina]	negatif
[film, taaaiii]	negatif

Figure 8. Filter Stopword Process

As shown in Figure 9, these common words are removed from the token list. This stage makes the text representation more concise and focused on words that truly convey sentiment meaning.

After the preprocessing stage was completed, a word cloud was generated as follows:

Figure 9. Word Cloud of the Animated Film *Merah Putih: One For All*

The word cloud shows that the most frequently appearing words, represented by larger sizes, include *film*, *AI*, and *Indonesia*.

Naïve Bayes Training

The next step involves using the Naïve Bayes algorithm for classification. The training and testing processes of the Naïve Bayes model were carried out using the RapidMiner application by applying the Split Validation method. The design of the model training process is shown in Figure 10.

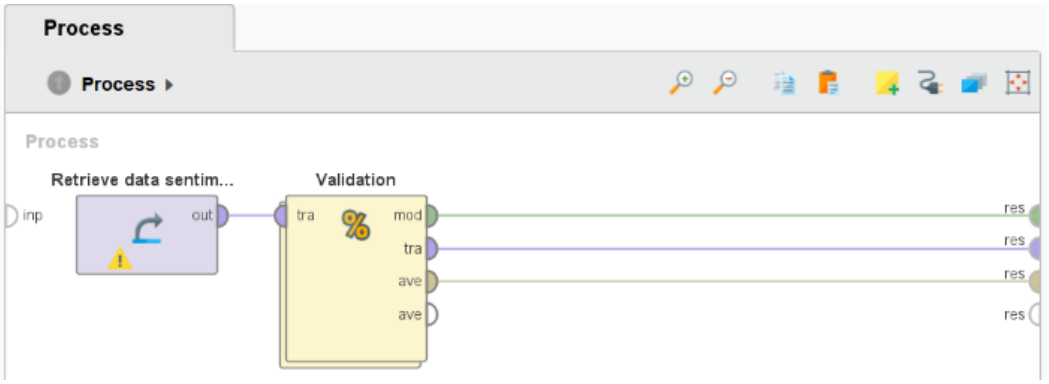

Figure 10. Naïve Bayes Algorithm Model Design

After the model training process, the data were divided into training and testing sets with a ratio of 60:40 to evaluate model performance. The data splitting workflow and evaluation process are shown in Figure 11.

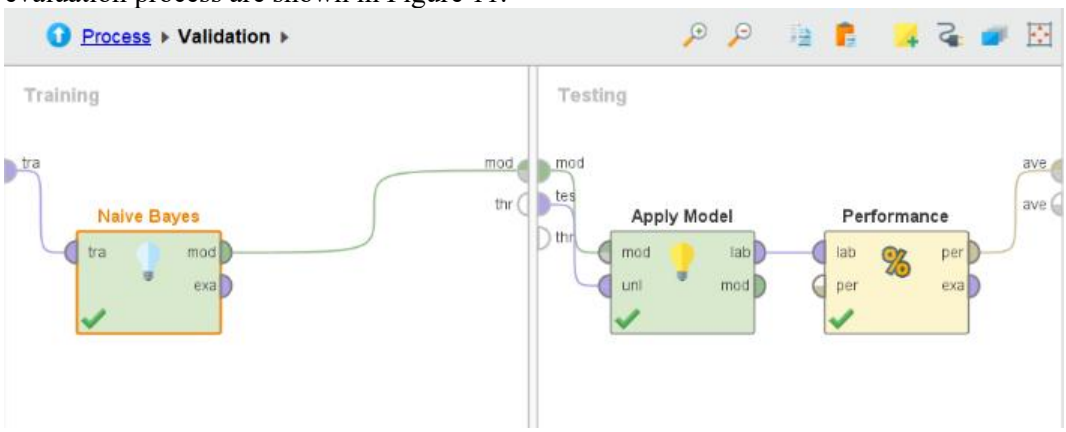

Figure 11. Split Validation

Evaluation of Classification Results

After completing the training and testing of the Naïve Bayes model, the next stage was model performance evaluation. Evaluation was conducted using a confusion matrix along with evaluation metrics such as accuracy, precision, and recall to assess the model's ability to classify the sentiment of Instagram user comments.

accuracy: 76.67%

	true negatif	true positif	true netral	class precision
pred. negatif	14	0	2	87.50%
pred. positif	12	30	3	66.67%
pred. netral	4	0	25	86.21%
class recall	46.67%	100.00%	83.33%	

Figure 12. Results of Naïve Bayes Algorithm Testing

The Naïve Bayes testing results produced an accuracy of 76.67%, indicating that the model performs fairly well in classifying sentiment. The positive class achieved the best performance with a recall of 100%, followed by the neutral class with a recall of 83.33%. In contrast, the negative class obtained a recall of 46.67%, indicating that misclassification still occurs for negative sentiment.

V. CONCLUSION

This study applies the Naïve Bayes algorithm to analyze the sentiment of Instagram user comments on the animated film Merah Putih One For All using 200 comment data collected from the Instagram platform. The data represent a wide range of public opinions, which were classified into three sentiment classes positive, neutral, and negative through manual labeling and text preprocessing stages. The model training and testing processes were conducted using the Split Validation method, with evaluation based on 90 test data generated by the system. The testing results show that the Naïve Bayes model achieved an accuracy of 76.67%, indicating fairly good classification performance. The positive sentiment class demonstrated the best performance with a recall of 100%, followed by the neutral class with a recall of 83.33%, while the negative class achieved a recall of 46.67%. The recall value for the negative sentiment class indicates that the model still encounters difficulties in identifying patterns in negatively toned comments. This limitation is caused by class imbalance, the use of informal and sarcastic language by Instagram users, and the inherent limitations of the Naïve Bayes algorithm, which is not yet capable of capturing deeper semantic context. Therefore, further improvements are required, such as increasing the amount of training data, balancing class distributions, or employing more complex classification methods. This study demonstrates that the Naïve Bayes algorithm has reasonably good capability in performing sentiment analysis on Instagram user comments related to the animated film Merah Putih One For All. Although the classification results show adequate performance particularly for the positive and neutral sentiment classes limitations remain in optimally identifying negative sentiment. Consequently, future research is recommended to enrich the dataset, balance class distributions, and explore more advanced classification methods to improve the accuracy and reliability of sentiment analysis. Thus, the findings of this study are expected to contribute academically to the development of sentiment analysis research and serve as a reference for further studies in the fields of natural language processing and social media opinion analysis.

REFERENCES

[1] Y. Li and W. Zhuge, "Application of animation control technology based on internet technology in digital media art," *Mobile Information Systems,* vol. 1, no. 4009053, p. 4009053, 2022. https://doi.org/10.1155/2022/4009053

[2] W. Qin, "Digital Narrative and Social Change: The Expressive Power of Visual Design in Public Issues," in *2025 4th International Conference on Social Sciences and Humanities and Arts (SSHA 2025).*, Atlantis Press, 2025.

[3] L. Bharadwaj, "Sentiment analysis in online product reviews: mining customer opinions for sentiment classification," *Int J Multidiscip Res,* vol. 5, no. 5, 2023. https://doi.org/10.36948/ijfmr.2023.v05i05.6090

[4] M. Veziroğlu, E. Veziroğlu and İ. Ömür , "Performance comparison between Naive Bayes and machine learning algorithms for news classification," *Bayesian Inference-Recent Trends. IntechOpen,* 2024. 10.5772/intechopen.1002778

[5] Y. R. Sipayung, "Penerapan Metode Klasifikasi Naive Bayes untuk Analisis Sentimen terhadap Undang Undang ITE di Media Sosial Twitter," *Elementary,* 2025. https://doi.org/10.30651/else.v9i2.24343

[6] A. Seraj, M. M.-. Khanaposhtani , R. Daneshfar and M. Naseri , "Chapter 5 - Cross-validation," *Handbook of hydroinformatics ,* vol. 1, no. 1, pp. 89-105, 2023. https://doi.org/10.1016/B978-0-12-821285-1.00021-X

[7] B. I. O. N. Irvandi, "Naive Bayes dan Wordcloud untuk Analisis Sentimen Wisata," *INFOTECH Journal,* pp. 236-242, 2023. https://doi.org/10.31949/infotech.v9i1.5322

[8] Y. A. Audina Tazkia, "Implementasi Naive Bayes Classifier Dalam Menganalisis," *Jurnal Teknika,* p. 139 – 149, 2024. https://doi.org/10.5281/zenodo.10851075

[9] F. A. Artanto, "Analisis Sentimen Opini Publik terhadap Fenomena Bunuh Diri," *Jurnal Sains Teknologi dan Sistem Informasi,* pp. 70-77, 2024. https://doi.org/10.54259/satesi.v4i1.2908

[10] Y. Nurtikasari, " Analisis sentimen opini masyarakat terhadap film pada platform Twitter menggunakan algoritma Naïve Bayes" *Jurnal Sains dan Teknologi,* pp. 411-423, 2022. https://doi.org/10.55123/insologi.v1i4.770

[11] D. Berliansyah, "Analisis Sentimen Pada Media Sosial Instagram Terhadap Akun Presiden Joko," *Journal of Digital Innovation & Information Technology ,* pp. 1-11, 2024. https://doi.org/10.54066/jptis.v2i2.1895

[12] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review," *KLIK: Kajian Ilmiah Informatika dan Komputer,* pp. 148-155, 2022. https://djournals.com/klik

[13] Y. R. Sipayung, "Penerapan Data Mining Pada Persediaan Barang Mentah dan Barang Jadi dengan Metode Naive Bayes untuk Optimasi Kerjasama antara Masyarakat dan Badan Usaha Desa," *Jamastika,* 2022. https://doi.org/10.35473/.v1i2.1868

[14] A. C. M. R. F. Oktavia Putri Zusrotun, "Sentimen Analisis Belajar Online di Twitter Menggunakan Naive Bayes," *Janapati,* 2022. https://doi.org/10.23887/janapati.v11i3.49160

[15] I. S. Putri Gea Aryanti, "Analisis Sentimen Pada Twitter Terhadap Mobil Listrik Menggunakan Algoritma Naive Bayes," *Jurnal IKRAITH-INFORMATIKA,* 2023.. https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/2821

[16] S. Lestari, "Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes," *SISMATIK,* 2021. https://sismatik.nusaputra.ac.id/index.php/sismatik/id/article/view/23/20

[17] Y. R. S. Arga Aditia Purnama, "Sentiment Analysis of Public Service Using Naïve Bayes Classifier," *Journal of Information Systems and Informatics,* 2025. https://doi.org/10.51519/journalisi.v7i3.1207

[18] I. H. Kusuma, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Informatika: Jurnal pengembangan IT (JPIT),* 2023. https://doi.org/10.30591/jpit.v8i3.5734

[19] Syafrizal, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science,* 2023. https://doi.org/10.57152/malcom.v4i1.983

[20] K. S. Putri, "Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naive Bayes Classifier," *Technologia,* 2023.

[21] Z. S. R. R. Dede Ardian Tarigan, "Analisis Sentimen Aplikasi Playstore Sirekap 2024 Pasca Pilpres Dengan Perbandingan Metode Support Vector Machine (SVM), Naïve Bayes Classifier Dan Random Forest," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* 2025. https://doi.org/10.25126/jtiik.2025129608