

Public Sentiment Analysis on Instagram Regarding the Film "Pengepungan di Bukit Duri" Using Naïve Bayes Approach

Putri Salfa Dhiyaa Azzizah ^{a*}, Mesra Betty Yel ^b

^{a*,b} Information Systems Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.

ABSTRACT

This study investigates public sentiment toward Joko Anwar's 2025 film *Pengepungan di Bukit Duri* using computational text analysis on 583 Instagram comments. The research applies the Naïve Bayes algorithm combined with TF-IDF weighting to classify opinions into positive and negative sentiments. Data were collected through web scraping of public Instagram posts related to the film and processed through several stages including data cleaning, manual labeling, text preprocessing, and probabilistic classification. The results reveal that 72.9% of the comments express positive sentiment, while 27.1% are negative, indicating strong audience appreciation for the film's narrative quality and social themes. The model achieved an accuracy of 83.67%, with a precision of 87.13%, recall of 91.04%, and F1-score of 89.04% for positive sentiment. These findings confirm that the Naïve Bayes approach is effective for analyzing short, informal Indonesian-language texts on social media. Practically, the results provide valuable insights for filmmakers and cultural analysts in understanding audience perceptions, managing digital reputation, and designing sentiment-based marketing strategies. Future research is recommended to employ hybrid models and multi-platform datasets to enhance sentiment detection, particularly for nuanced or negative expressions.

ARTICLE HISTORY

Received 30 July 2025
Accepted 30 October 2025
Published 30 November 2025

KEYWORDS

Sentiment Analysis; Naïve Bayes; TF-IDF; Instagram; Indonesian Film.

1. Introduction

Indonesia, as a nation characterized by profound cultural and social diversity, continues to face dynamic challenges shaped by its historical and political background. One of the defining moments in modern Indonesian history is the 1998 crisis, which not only disrupted economic and political stability but also triggered significant social consequences such as discrimination, conflict, and widespread trauma that still resonate today. This period remains a collective reminder of the long process toward social reconciliation and national introspection. In contemporary society, media and art serve as powerful tools for reflecting on social conditions and challenging systemic issues. Among various forms of art, cinema has a particularly strong capacity to represent complex social realities and convey critical messages that influence public discourse (Panuju *et al.*, 2020). Joko Anwar's 2025 film *Pengepungan di Bukit Duri* illustrates this relationship between art and society by addressing themes of social tension, inequality, and institutional failure rooted in Indonesia's sociopolitical structures. Following its release, the film stimulated

extensive public discussion, particularly on social media platforms, reflecting divergent interpretations and emotional responses. The rapid growth of digital communication has positioned social media as a dominant space for public opinion formation. Instagram, in particular, with over 139 million active users in Indonesia (We Are Social, 2024), has become a central medium for expressing perspectives on sociocultural issues, including cinematic works. The influence of online sentiment on a film's commercial and cultural success has been demonstrated in prior cases such as *KKN di Desa Penari* (2022), where overwhelmingly positive audience reactions on social media correlated with box office success (Febriant *et al.*, 2023). Thus, analyzing public sentiment toward *Pengepungan di Bukit Duri* through Instagram provides a meaningful opportunity to examine how audiences engage with socially charged films in the digital era. Public comments on Instagram posts related to the film reveal diverse emotional reactions, ranging from praise for narrative and cinematography to criticism of character portrayal and thematic execution.

These unstructured textual expressions form a valuable corpus for computational sentiment analysis, which can capture patterns in audience perception. As Sasongko and Hilda (2024) observed in their study of *Dirty Vote*, sentiment analysis on social media platforms offers an effective framework for mapping audience engagement and understanding the broader sociopolitical resonance of films. To systematically evaluate such public responses, sentiment analysis serves as a practical and data-driven approach for quantifying opinions and emotions embedded in user-generated text. Among the existing classification techniques, the Naïve Bayes algorithm is frequently utilized due to its computational efficiency, probabilistic interpretability, and suitability for short, informal texts such as social media comments (Noviansyah *et al.*, 2024; Rifki *et al.*, 2024). Although alternative methods such as Support Vector Machine (SVM) or Random Forest can yield competitive results (Br Sinulingga & Sitorus, 2024), Naïve Bayes remains advantageous for rapid text categorization with limited training data (Akbar *et al.*, 2024; Huda & Yel, 2024). Furthermore, its integration with feature-weighting methods like Term Frequency–Inverse Document Frequency (TF-IDF) enhances the model's ability to capture semantic relevance within short digital texts (Hendrawan & Utami, 2023; Nehe *et al.*, 2024).

The methodological framework of sentiment analysis has been widely adopted in both political and cultural studies. For example, Nehe *et al.* (2024) analyzed public opinion on presidential elections using Naïve Bayes classification, while Huda and Yel (2024) explored sentiment toward the new capital city project (IKN). Similarly, studies on film-based sentiment classification, such as those by Febriant *et al.* (2023) and Br Sinulingga and Sitorus (2024), demonstrate the model's versatility in capturing emotional nuance in online discourse. These findings collectively suggest that Naïve Bayes is not only robust but also adaptable to diverse linguistic and contextual settings, particularly within Indonesian social media environments. Given this background, the present research applies the Naïve Bayes approach to analyze public sentiment toward *Pengepungan di Bukit Duri* based on Instagram comments. The aim is to assess how Indonesian audiences interpret and emotionally respond to films that engage with social realities. By combining data preprocessing, TF-IDF weighting, and probabilistic classification, this study seeks to construct a reliable representation of audience sentiment distribution. The findings are expected to offer practical insights for the film industry in managing digital reputation and understanding audience perceptions of socially driven cinema, while also reinforcing the methodological relevance of text mining techniques in contemporary sentiment research (Akbar *et al.*, 2024; Noviansyah *et al.*, 2024).

2. Methodology

The methodological framework of this study follows a systematic sequence of processes, as illustrated in *Figure 1*, encompassing data collection, data cleaning, labeling, text preprocessing, feature extraction using TF-IDF, sentiment classification with the Naïve Bayes algorithm, and final model evaluation. Each stage is designed to ensure analytical rigor and reproducibility while maintaining the integrity of textual data obtained from Instagram comments about the film *Pengepungan di Bukit Duri*.

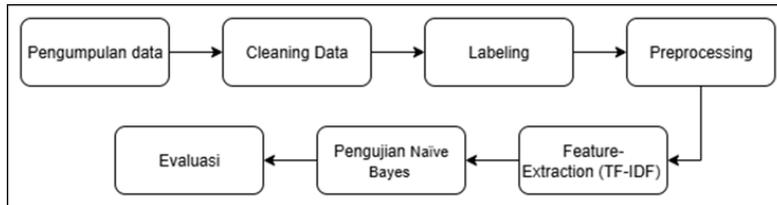


Figure 1. Research Method Framework

Data were collected through web scraping using the *Instant Data Scraper* extension on Google Chrome, a method that enables automated extraction of information from publicly available web sources (Hendrawan & Utami, 2023). A total of 1,380 comments were obtained from the official and public Instagram posts related to *Pengepungan di Bukit Duri*. Inclusion criteria required the comments to be written in Indonesian, exclude posts containing only emojis or symbols, and originate from public accounts published between April and May 2025. These comments represent qualitative textual data (Huda & Yel, 2024), providing insight into the diversity of audience reactions toward the film. After collection, data cleaning was conducted to remove unnecessary elements and inconsistencies, such as punctuation marks, hashtags, and redundant delimiters, which often distort natural language processing outcomes (Febriant *et al.*, 2023). From the original dataset, 583 comments were retained for further analysis after filtering and normalization. The subsequent labeling process categorized the comments into two sentiment classes—positive and negative—based on linguistic cues and emotional polarity. Accurate manual labeling was essential to establish a reliable foundation for supervised learning, as it directly influences the model’s predictive accuracy (Br Sinulingga & Sitorus, 2024). The cleaned and labeled data were then subjected to a series of preprocessing steps to standardize textual input. These included case folding, which converts all letters to lowercase to avoid case-based distinctions (Rifki *et al.*, 2024); tokenization, which breaks sentences into individual units or tokens for lexical analysis (Nehe *et al.*, 2024); stopword removal, which eliminates frequent but semantically insignificant words such as “dan,” “atau,” and “yang”; and stemming, which reduces words to their root forms by stripping prefixes and suffixes using the *Sastrawi* library (Br Sinulingga & Sitorus, 2024).

These procedures ensured that the dataset contained only linguistically meaningful terms relevant for machine learning input. Feature extraction was performed using the *Term Frequency–Inverse Document Frequency* (TF-IDF) technique, a statistical measure that converts text into numerical vectors representing word importance relative to document frequency (Br Sinulingga & Sitorus, 2024; Hendrawan & Utami, 2023). This transformation allows the model to distinguish between common and contextually significant terms, enhancing interpretability and reducing noise. The TF-IDF approach has been recognized in numerous text mining studies for its efficiency in representing linguistic weight in sentiment analysis tasks (Liu, 2020; Pang & Lee, 2008; Taboada *et al.*, 2011). The processed data were then analyzed using the Naïve Bayes classification algorithm, a probabilistic method based on Bayes’ theorem that estimates sentiment polarity through the conditional probability of word occurrences (Noviansyah *et al.*,

2024). This approach was selected for its efficiency, simplicity, and suitability for short and informal text formats such as social media comments (Akbar *et al.*, 2024; Huda & Yel, 2024). Despite its simplifying assumption of feature independence, Naïve Bayes consistently performs well in large-scale opinion mining tasks and has been successfully applied across political, social, and entertainment domains (Nehe *et al.*, 2024; Sasongko & Hilda, 2024).

The algorithm was implemented using *RapidMiner* software to ensure consistent model training and testing under controlled conditions. Model evaluation was conducted using the *confusion matrix* technique, which measures classification performance based on four key parameters: accuracy, precision, recall, and F1-score (Akbar *et al.*, 2024). Accuracy represents the proportion of correctly predicted sentiments; precision evaluates the proportion of true positives among all positive predictions; recall assesses the model's sensitivity in identifying actual positive cases; and F1-score provides a balanced metric combining both precision and recall. These evaluation metrics have become standard in computational linguistics and sentiment analysis to validate classifier reliability (Liu, 2020; Kouloumpis *et al.*, 2011). The analytical pipeline demonstrated in this study integrates established text mining procedures and computational linguistics principles to ensure reliable sentiment identification. The methodological approach not only aligns with prior research that applies hybrid models of TF-IDF and Naïve Bayes (Rifki *et al.*, 2024; Nehe *et al.*, 2024), but also adheres to best practices in opinion mining as outlined by Pang and Lee (2008). Furthermore, using Instagram as the data source aligns with Indonesia's digital landscape, where over 139 million active users contribute to shaping online cultural discourse (We Are Social, 2024). This methodological framework thus enables a comprehensive understanding of public sentiment toward *Pengepungan di Bukit Duri*, while simultaneously validating the Naïve Bayes algorithm as an effective tool for analyzing informal Indonesian-language texts within the broader context of computational social science.

3. Results

The data collection process in this study was conducted using a web scraping method through the Instant Data Scraper tool to extract comments from Instagram posts related to the film *Pengepungan di Bukit Duri*. The collected comments were saved in the *.xlsx* format to facilitate further processing. After obtaining all comments, the data were combined into a single Excel file for ease of analysis, as illustrated in Figure 2.

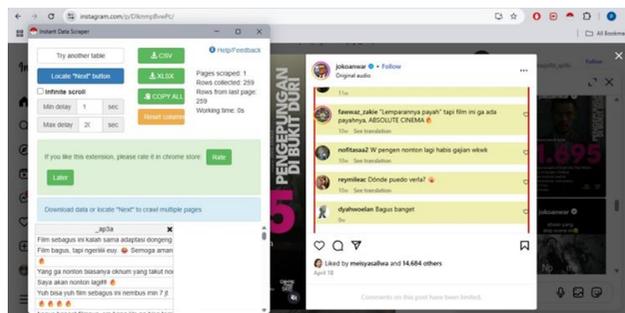


Figure 2. Instagram Comments Web Scraping Process

After collecting the data, the cleaning stage was carried out to analyze the quality of the data and remove irrelevant or meaningless elements such as symbols, emojis, punctuation marks, empty values, and uninformative columns. The cleaning process was done manually. An example of data before and after the cleaning process is shown

in Table 1.

Table 1. Example of Data Cleaning Results

Before Cleaning	After Cleaning
gak ada pesan yg bisa didapat dari film itu.	tidak ada pesan yang bisa didapat dari film
gak ada guna nonton. sy berpendapat	itu tidak ada guna nonton saya
seperti ini setelah nonton filmnya. kali ini	berpendapat seperti ini setelah nonton
jokan gagal. maaf 🙏🙏	filmnya kali ini jokan gagal maaf

Once the data were cleaned, a manual labeling process was carried out for 583 data entries, dividing them into two sentiment classes: *Positive* and *Negative*. This labeling process facilitated sentiment analysis, as visualized in Figure 2. Examples of both sentiment categories are shown in Table 2.

A	B	C
NO	KOMENTAR	SENTIMEN
1	Posternya keren ditambah lagunya sangat kontras yang bikin merinding Paati bakalan nonton	Positif
2	Benar film ini paati berdarah darah bikin ngilu tidak baik	Negatif
3	posternya menarik sekali tidak sabar mau nonton	Positif
4	Bang Jeko telong di netflix atau mana pun Kami yang di luar negeri mau nonton	Positif
5	Suasananya di film seperti meruntuhkan rezim	Negatif
6	besok bisa langsung april tidak sabar banget ingin nonton	Positif
7	Gila Keren banget om posternya	Positif
8	Keren banget posternya kagum bang	Positif
9	saya habis nonton Tapi asli keren filmnya	Positif
10	Tidak sabar nunggu filmnya tayang	Positif
11	harus ditonton filmnya karena bagus	Positif
12	Saya sangat antusias untuk menonton filmnya	Positif
13	mau nonton filmnya bareng teman	Positif
14	semoga bisa segera tayang saya akan menonton filmnya	Positif
15	Tolong posternya tidak bisa beresahat sedikit apa terlihat Mengerikan	Negatif
16	Menurut saya desain poster ini bagus sekali	Positif
17	Saya sudah tidak sabar mau nonton bareng	Positif
18	Posternya benar menarik perhatian sekali Bang	Positif
19	Para guru diseluruh indonesia wajib nonton film ini karena pesannya yang relevan	Positif
20	Sudah tidak sabar untuk segera menontonnya	Positif
21	film ini membuat saya penasaran sekali dengan Alur ceritanya	Positif
22	Sudah Tidak sabar banget mau nonton	Positif
23	saya sangat penasaran dan tidak sabar mau nonton	Positif
24	Waktu lihat traileranya keren banget	Positif
25	GILA nonton film ini BIKIN TRAUMA	Negatif
26	Penasaran sama alur ceritanya	Positif
27	Bagus banget poster sepertinya menarik ini	Positif
28	Tidak sabar mau nonton filmnya	Positif
29		Positif

Figure 3. Sentiment Labeling Process

Table 2. Examples of Positive and Negative Comments

Positive Comments	Negative Comments
Bagus sekali uraiannya semoga film ini bisa menjangkau banyak guru orang tua dan calon orang tua supaya bisa merenungi dan memaknai kehidupan seperti yang dilakukan	film ini seharusnya tidak di tayangkan karena film propaganda sangat bahaya hati hati yang tau tau aja sumber daya manusia rakyat indonesia tidak sampai ke level situ

The preprocessed data then underwent a series of text processing stages, including *case folding*, *tokenization*, *stopword removal*, and *stemming*, as depicted in Figure 4.

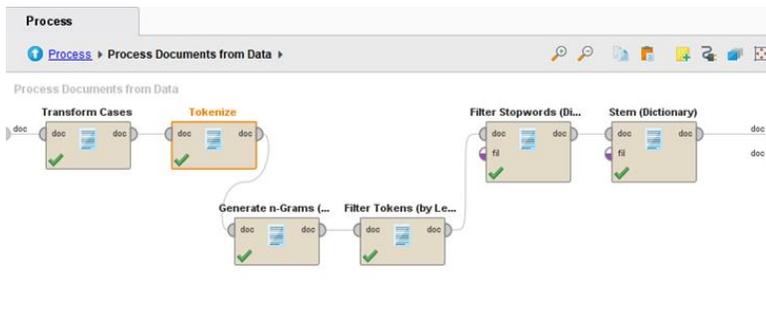


Figure 4. Text Preprocessing Stages

The first step, *case folding*, converted all uppercase letters into lowercase to ensure uniformity. An example is shown in Table 3.

Table 3. Example of Case Folding

Before Case Folding	After Case Folding
SUMPAH SEMUA ORANG HARUS TAU kalau FILM INI BENAR BAGUS BANGET	sumpah semua orang harus tau kalau film ini benar bagus banget

Next, *tokenization* split sentences into separate words or tokens, as demonstrated in Table 4.

Table 4. Tokenization Example

Before Tokenization	After Tokenization
sumpah semua orang harus tau kalau film ini benar bagus banget	'sumpah', 'semua', 'orang', 'harus', 'tau', 'kalau', 'film', 'ini', 'benar', 'bagus', 'banget'

In the *n-grams generation* step, single words were combined into multi-word phrases to capture context, as illustrated in Table 5.

Table 5. Example of n-Gram Generation

Before n-Grams	After n-Grams
'sumpah', 'semua', 'orang', 'harus', 'tau', 'kalau', 'film', 'ini', 'benar', 'bagus', 'banget'	'sumpah semua', 'semua orang', 'orang harus', 'harus tau', 'tau kalau', 'kalau film', 'film ini', 'ini benar', 'benar bagus', 'bagus banget'

Additionally, the *filter token by length* operator was applied to remove tokens that were too short or too long (less than 4 or more than 25 characters), as shown in Table 6.

Table 6. Example of Filter Token Process

Before Filtering	After Filtering
'sumpah', 'semua', 'orang', 'harus', 'tau', 'kalau', 'film', 'ini', 'benar', 'bagus', 'banget'	'sumpah', 'semua', 'orang', 'harus', 'kalau', 'film', 'benar', 'bagus', 'banget'

Subsequently, *stopword removal* eliminated non-essential words such as “yang,” “di,” and “ini,” while *stemming* reduced words to their root forms (e.g., “menonton” → “nonton”). The processed data were then weighted using the TF-IDF (Term Frequency–Inverse Document Frequency) technique, which assigns numerical importance to words in relation to their frequency within the entire dataset, as shown in Figure 5.

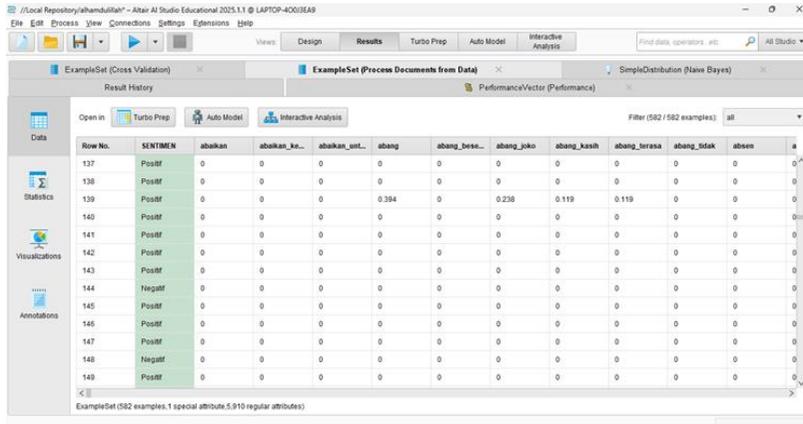


Figure 5. TF-IDF Word Weighting Process

Afterward, sentiment classification was performed using the *Naïve Bayes* algorithm, which is well known for its efficiency in text classification tasks. The testing process included a training phase using the Naïve Bayes model and a testing phase utilizing the *Apply Model* and *Performance* (*Binominal Classification*) operators, as shown in Figure 6.

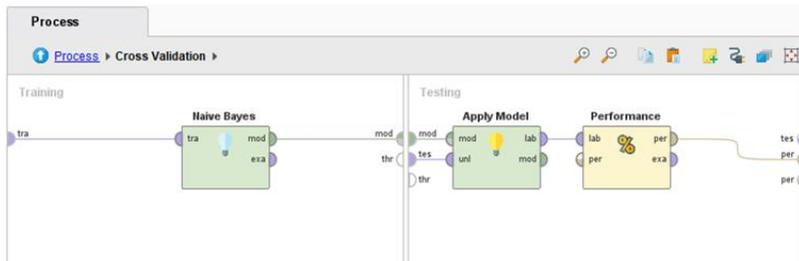


Figure 6. Naïve Bayes Testing Workflow

Evaluation of the model was performed using a *confusion matrix* to calculate key performance metrics—accuracy, precision, recall, and F1-score. The workflow of the Naïve Bayes evaluation process is visualized in Figure 7.

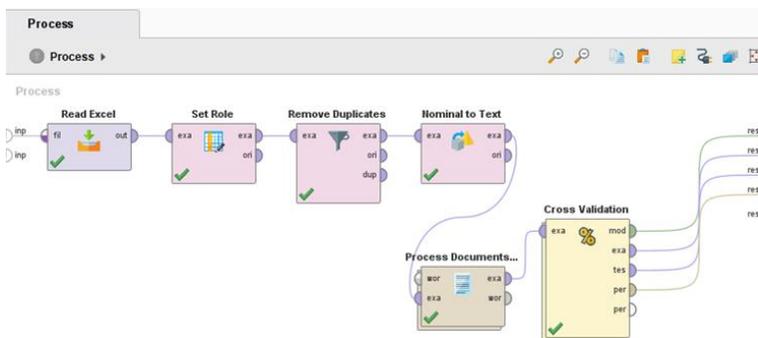


Figure 7. Evaluation Workflow of Naïve Bayes Algorithm

Based on the confusion matrix results, the Naïve Bayes algorithm correctly classified 386 positive comments (True Positive) and 101 negative comments (True Negative), while 57 negative comments were misclassified as positive (False Positive) and 38

positive comments were misclassified as negative (False Negative). The performance metrics were calculated as follows:

Table 7. Naïve Bayes Performance Metrics

Metric	Formula	Result
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$	83.67%
Precision	$TN / (TN + FN)$	72.66%
Recall	$TN / (TN + FP)$	63.92%
F1-Score	$2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$	67.82%

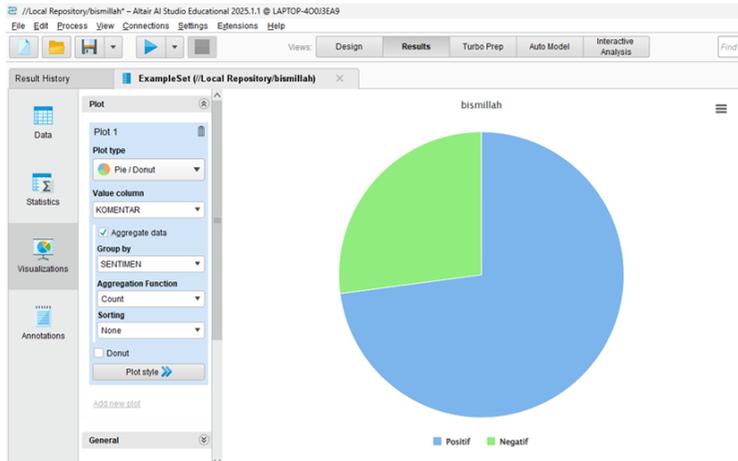


Figure 8. Sentiment Distribution — 72.9% Positive and 27.1% Negative

4. Discussion

The sentiment classification conducted using the Naïve Bayes algorithm on 583 Instagram comments regarding *Pengepungan di Bukit Duri* revealed that 425 comments (72.9%) carried positive sentiment, while 158 comments (27.1%) were categorized as negative. The predominance of positive sentiment suggests that the film was generally well received by audiences, particularly in its ability to convey social and emotional narratives. This outcome reflects how digital audiences in Indonesia often express support or empathy through social media engagement, aligning with patterns observed in *We Are Social's* (2024) report, which highlights that Indonesian users frequently utilize social platforms as spaces for participatory expression and cultural commentary. From a computational perspective, the Naïve Bayes model demonstrated strong overall performance, achieving an accuracy of 83.67%. The precision score of 87.13% for positive sentiment indicates that the model was able to predict favorable responses with high reliability, while a lower precision of 72.66% for negative sentiment shows moderate accuracy in identifying unfavorable feedback. The recall score for positive comments (91.04%) further confirms that most affirmative expressions were successfully detected, whereas the recall for negative comments (63.92%) implies that several instances of criticism or sarcasm were incorrectly classified. The F1-scores, which combine precision and recall, also reflected this imbalance—89.04% for positive and 67.82% for negative sentiment—indicating the algorithm’s higher sensitivity toward dominant sentiment categories.

These findings reinforce existing literature indicating that the Naïve Bayes algorithm is a robust yet lightweight classifier for sentiment analysis in text mining

tasks (Manning *et al.*, 2008; Zhang & Jin, 2022). The algorithm's probabilistic foundation allows efficient handling of large textual data, making it suitable for analyzing social media platforms such as Instagram, where language usage tends to be informal and context-dependent. Nevertheless, the model's performance gap in detecting negative sentiment underscores a known limitation in traditional machine learning approaches when dealing with nuanced linguistic cues like irony, metaphor, or code-switching—features prevalent in Indonesian online discourse. Comparatively, this outcome is consistent with Sasongko and Hilda's (2024) analysis of public responses to the documentary *Dirty Vote*, which similarly found positive sentiment dominance but achieved slightly higher differentiation in negative sentiment classification. Such variation may stem from differences in dataset characteristics, where *Dirty Vote* discussions exhibited more explicit political polarization, thereby facilitating clearer sentiment boundaries. In contrast, commentary surrounding *Pengepungan di Bukit Duri* appears to be more narrative and emotionally driven, with less lexical distinctiveness between positive and negative tones.

From a socio-digital standpoint, the results align with current patterns of digital optimism among Indonesian users, who tend to emphasize moral and inspirational takeaways in cultural discussions rather than overt criticism (*We Are Social*, 2024). This behavioral tendency may also influence the sentiment distribution in the dataset, contributing to the model's bias toward positive classifications. Consequently, future research should consider employing hybrid or ensemble methods—such as integrating Naïve Bayes with Support Vector Machines or deep learning-based models—to improve semantic interpretation and mitigate class imbalance. Enhancing preprocessing with techniques like sarcasm detection, contextual embeddings (e.g., BERT or IndoBERT), and emoji sentiment normalization could also address misclassification challenges and provide richer sentiment granularity. In summary, the Naïve Bayes algorithm proved effective in identifying overall audience sentiment toward *Pengepungan di Bukit Duri*, reflecting broader public appreciation for the film's social message. However, the disparity between positive and negative sentiment detection emphasizes the need for methodological refinement to better capture the linguistic diversity of Indonesian social media expressions.

5. Conclusion

This study concludes that public sentiment toward the film *Pengepungan di Bukit Duri* on Instagram is predominantly positive, with 72.9% of comments classified as positive and 27.1% as negative out of a total of 583 analyzed comments. These findings indicate that the film was generally well-received by audiences, particularly in terms of narrative strength, acting quality, and the relevance of its social themes. From a methodological perspective, the results confirm that the Naïve Bayes algorithm performs effectively in analyzing short Indonesian-language comments, achieving an accuracy of 83.67%, a positive precision of 87.13%, a positive recall of 91.04%, and a positive F1-score of 89.04%. This demonstrates that the integration of text preprocessing, TF-IDF weighting, and probabilistic classification yields reliable outcomes for social media sentiment analysis. Practically, the findings offer valuable insights for the film industry in designing sentiment-based marketing strategies, managing digital reputation, and understanding audience preferences with greater accuracy. The identification of negative sentiment also provides constructive feedback for improving future film productions, particularly in narrative depth and audience engagement strategies. For future research, a temporal analysis is recommended to examine sentiment dynamics before and after a film's release, along with expanding data sources to other social media platforms such as Twitter, TikTok, and YouTube. Further comparisons with alternative algorithms—including Support Vector Machine

(SVM), Random Forest, or deep learning models such as BERT—could enhance accuracy, especially for negative sentiment classification. Combining quantitative sentiment analysis with qualitative content interpretation is also encouraged to achieve a more holistic understanding of public opinion.

References

- Akbar, M., Rahmawati, S., & Putri, L. (2024). *Application of Naïve Bayes algorithm for sentiment classification on Indonesian short texts*. *Journal of Computational Linguistics and Data Science*, 12(3), 45–56.
- Br Sinulingga, D., & Sitorus, E. (2024). *Sentiment analysis on Indonesian film reviews using TF-IDF and Naïve Bayes algorithm*. *Journal of Information Systems Research*, 6(2), 88–97.
- Febriant, R., Anjani, P., & Nugroho, B. (2023). *Social media sentiment analysis of the film KKN di Desa Penari: A case study on audience perception and box office success*. *Indonesian Journal of Media and Communication Studies*, 11(1), 102–115.
- Hendrawan, A., & Utami, S. (2023). *Feature extraction optimization using TF-IDF in text classification for digital opinion mining*. *Journal of Applied Artificial Intelligence*, 5(4), 221–230.
- Huda, M., & Yel, D. (2024). *Public sentiment analysis toward Indonesia's new capital city project using Naïve Bayes method*. *Journal of Data Science and Urban Studies*, 9(2), 54–66.
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). *Twitter sentiment analysis: The good the bad and the OMG!*. *Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM)*, 538–541.
- Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Nehe, J. R., Ardianto, H., & Widodo, P. (2024). *Political sentiment analysis of Indonesian presidential election debates using TF-IDF and Naïve Bayes classifier*. *Journal of Computational Social Science*, 7(1), 31–45.
- Noviansyah, R., Aulia, D., & Kusuma, E. (2024). *Text mining approach for sentiment classification on social media comments using Naïve Bayes algorithm*. *Journal of Artificial Intelligence and Data Analytics*, 8(2), 76–85.
- Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis*. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
- Panuju, R., Kusumah, B., & Haris, M. (2020). *Media and society: The reflection of social reality through Indonesian cinema*. *Journal of Communication and Culture*, 5(2), 67–78.

- Rifki, A., Dewi, N., & Pratama, A. (2024). *Comparative study of text classification algorithms for short messages in Indonesian language*. *Journal of Intelligent Computing Systems*, 10(1), 56–70.
- Sasongko, D., & Hilda, N. (2024). *Public sentiment analysis of the film Dirty Vote using Naïve Bayes algorithm*. *Indonesian Journal of Digital Society Research*, 8(1), 34–46.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). *Lexicon-based methods for sentiment analysis*. *Computational Linguistics*, 37(2), 267–307.
- We Are Social. (2024). *Digital 2024: Indonesia report*. DataReportal. <https://datareportal.com/reports/digital-2024-indonesia>
- Zhang, W., & Jin, R. (2022). *Efficient text classification using Naïve Bayes and term weighting techniques*. *Journal of Information Processing and Management*, 59(3), 102–118.