JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8037

SENTIMENT ANALYSIS OF TWITTER (X) RESPONSES USING NAÏVE BAYES: A CASE STUDY OF JUMBO

Nurasiah[1*]; Latifah Nur Zam Zami[1]; Yunianto Agung Nugroho[1]

Information System[1]
Insan Pembangunan Indonesia University, Tangerang, Indonesia[1]
https://unipem.ac.id/[1]
nurasiah.sisfo@gmail.com*, latifahnurzam9d@gmail.com, yunianto.agung76@yahoo.com

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *The rapid growth of social media platforms, especially Twitter (now rebranded as X), has made it a rich source for public opinion mining. This study focuses on analyzing public sentiment toward the* animated *"film Jumbo" by applying the Naïve Bayes Classifier algorithm to Twitter data. The research involves collecting tweets related to the film, preprocessing the data by removing noise such as stopwords, user mentions, and special characters, and then classifying sentiments into positive, negative, and neutral categories. The Naïve Bayes algorithm was chosen due to its effectiveness and efficiency in text classification tasks such as sentiment analysis. The classifier is trained on a labeled dataset and evaluated through performance metrics including accuracy, precision, recall, and F1-score. The results demonstrate that the Naïve Bayes Classifier can effectively capture public sentiment on social media, providing insights into audience reception of the "film Jumbo." Moreover, this research highlights the importance of integrating machine learning techniques with social media analytics for real-time sentiment monitoring. The model achieved an accuracy of 77% in classifying sentiments. Analysis results indicate that positive sentiment dominates public opinion, with a precision of 88% and recall of 83%, followed by negative sentiment with precision and recall of 75% and 71%, respectively, and neutral sentiment with precision of 40% and recall of 57%. These findings offer valuable insights for the film industry to better understand audience perceptions and make informed decisions.*

Keywords*: Classifier Algorithm, Jumbo, Naïve Bayes, Sentiment Analysis, Twitter (X).*

Intisari— *Pertumbuhan pesat platform media sosial, khususnya Twitter (X), telah menjadikannya sumber yang kaya untuk penambangan opini publik. Penelitian ini berfokus pada analisis sentimen publik terhadap film "Jumbo" dengan menerapkan algoritma Naïve Bayes Classifier pada data Twitter. Penelitian melibatkan pengumpulan tweet terkait film, pra-pemrosesan data dengan menghilangkan noise seperti stopwords, mention pengguna, dan karakter khusus, kemudian mengklasifikasikan sentimen menjadi kategori positif, negatif, dan netral. Algoritma Naïve Bayes dipilih karena efektivitas dan efisiensinya dalam tugas klasifikasi teks seperti analisis sentimen. Klasifier dilatih pada dataset berlabel dan dievaluasi melalui metrik performa termasuk akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Naïve Bayes Classifier mampu menangkap sentimen publik di media sosial secara efektif, memberikan wawasan tentang penerimaan audiens terhadap film "Jumbo". Selain itu, penelitian ini menyoroti pentingnya integrasi teknik machine learning dengan analitik media sosial untuk pemantauan sentimen secara real-time. Model mencapai akurasi 77% dalam mengklasifikasikan sentimen. Hasil analisis menunjukkan bahwa sentimen positif mendominasi opini publik, dengan presisi 88% dan recall 83%, diikuti sentimen negatif dengan presisi 75% dan recall 71%, serta sentimen netral dengan presisi 40% dan recall 57%.*

Kata Kunci*: Algoritma Klasifier, Jumbo, Naïve Bayes, Analisis Sentimen, Twitter (X).*

INTRODUCTION

Social media platforms have become important sources of public opinion because users can express reactions to products, services, events, and entertainment content in short, informal messages. Twitter (X), in particular, provides a large volume of public posts that can be analyzed to identify sentiment tendencies [1]. The animated film "Jumbo" generated substantial public discussion, including comments about its storyline, characters, moral messages, and viewing experience. This phenomenon provides an opportunity to examine how audiences respond to a local animated film through sentiment analysis.

Sentiment analysis is a Natural Language Processing (NLP) task that identifies and classifies opinions expressed in text.[2] In social-media research, preprocessing is essential because tweets commonly contain URLs, mentions, hashtags, spelling variations, slang, and other noise. Previous studies have shown that Naïve Bayes is a practical classifier for text and Twitter sentiment analysis because it uses the probability of features occurring in each sentiment class and can be implemented efficiently[3], [4], [5]. Recent Indonesian studies also demonstrate the applicability of Naïve Bayes to Twitter sentiment classification after preprocessing and feature extraction[6], [7], [8].

The literature indicates that preprocessing quality and data representation strongly influence sentiment-classification performance. Studies using Indonesian Twitter data commonly apply cleaning, case folding, tokenization, normalization, stopword removal, and stemming before classification [6], [7]. Recent comparative research also shows that classical machine-learning methods remain useful baselines for Indonesian social-media sentiment analysis, although informal language and class imbalance can reduce performance [9], [10]. These findings suggest that a carefully defined preprocessing and labeling pipeline is important when analyzing audience responses to an entertainment topic.

Despite the growing number of sentiment-analysis studies on Twitter, research focusing specifically on Indonesian audience responses to local animated films remains limited. Existing studies often address political issues, public policies, products, or general social-media topics. In addition, many studies report classification performance without providing a detailed presentation of the transformation from raw tweets into analysis-ready features. This creates a research gap in documenting a reproducible preprocessing and sentiment-classification workflow for local film discussions.

This study addresses the gap by analyzing Indonesian-language Twitter (X) responses to the animated film "Jumbo" using a structured pipeline consisting of data crawling, preprocessing, lexicon-based labeling, Multinomial Naïve Bayes classification, and model evaluation[11], [12]. The lexicon is used only to generate sentiment labels from word-level weights, whereas Naïve Bayes is used as the supervised classifier. This distinction makes the methodological procedure consistent and separates the labeling mechanism from the machine-learning classification stage.

The study further examines the distribution of positive, negative, and neutral responses and presents representative preprocessing and prediction results in editable tables. The evaluation compares several train–test ratios using accuracy and class-level precision, recall, and F1-score. The research objective is to determine the sentiment tendency toward "Jumbo" and evaluate the capability of Multinomial Naïve Bayes in classifying Indonesian Twitter responses. The contribution is a transparent and reproducible workflow for sentiment analysis of local film discussions using publicly available social-media data.

The research questions are: (1) what sentiment distribution is expressed in Indonesian Twitter responses to "Jumbo"; (2) how does the proposed preprocessing and lexicon-labeling pipeline support Multinomial Naïve Bayes classification; and (3) which train–test ratio provides the best classification accuracy in this dataset?

The study uses the Naïve Bayes classifier because it is computationally efficient and suitable for text classification. Previous research has applied Naïve Bayes to Twitter sentiment analysis and reported useful classification performance after preprocessing and feature selection [4], [6], [7],[13] [14]. These studies provide a methodological basis for applying the same classifier to a film-related dataset while retaining the specific characteristics of Indonesian informal text.

Unlike a purely lexicon-based sentiment study, this research uses the lexicon to create initial sentiment labels and then trains a Multinomial Naïve Bayes model on the labeled data. Therefore, the study does not claim that the final classification itself is lexicon-based. This distinction is important for reproducibility and resolves the methodological inconsistency identified during editorial review.

The research gap is therefore centered on the limited analysis of local animated-film discussions on Twitter (X), the need for a clearly documented

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8037

Indonesian-text preprocessing pipeline, and the need to report class-level model performance rather than only the overall accuracy.

Accordingly, the present study focuses on "Jumbo" as a case study and applies cleaning, case folding, tokenization, normalization, stopword removal, and stemming before lexicon-based labeling and Multinomial Naïve Bayes classification. The analysis also examines sentiment distribution and representative model predictions to provide an interpretable view of audience responses.

The expected contribution is twofold: academically, the study provides an empirical application of Multinomial Naïve Bayes to Indonesian Twitter responses about a local animated film; practically, the findings provide a concise indication of audience reception that can support evaluation of audience preferences and communication strategies in the film industry.

MATERIALS AND METHODS

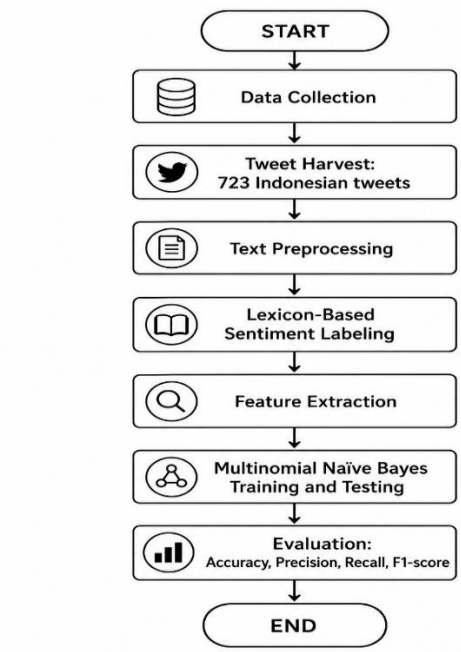
The sampling stage was carried out by collecting data from Twitter. The data collection process employed web scraping, an automated technique for extracting information directly from social media platforms [15], [16], [13], [17]. This approach was specifically applied to gather, a Literature Review was conducted, drawing from previous research studies with similar thematic focus. This comprehensive review incorporated scholarly sources including books, academic journals, and conference proceedings [18].

This study uses a descriptive quantitative approach with text mining and sentiment analysis. The research procedure consists of data collection, data preparation, text preprocessing, lexicon-based labeling, feature extraction, Multinomial Naïve Bayes classification, and model evaluation. [12]. The dataset consists of Indonesian-language public tweets related to the animated film "Jumbo". Tweet text is the primary analytical variable, while available metadata such as date, username, likes, and retweets are retained as supporting information.

Key Aspects of the Methodology:

1) Web Scraping Implementation
   Automated data extraction from media platforms and Twitter [1].
2) Literature Review Framework
   Critical examination of existing research on:
   a) Sentiment analysis Methodologies [15].
   b) Machine Learning applications in review analysis [19].
3) Data Analysis Methodology Using Naïve Bayes for Sentiment Classification [20].

The research stages are summarized in Figure 1 and are described below.



Source: (Research Results, 2025)

Figure 1. Research Flow of the Sentiment Analysis Study

In this research, the author conducted data collection using the following methods:

a. Data Collection and Crawling [2]

A total of 723 Indonesian-language tweets related to "Jumbo" were collected using Tweet Harvest. The predefined search parameters were used to improve relevance and reproducibility.

Table 1. Data Collection Parameters

Parameter	Specification
Platform	Twitter (X)
Tool	Tweet Harvest via Google Colaboratory
Keywords	Film Jumbo; Jumbo; #FilmJumbo; Jumbo Akidah
Period	January 1, 2025 – May 31, 2025
Language	Indonesian
Retrieved tweets	723
Output	CSV

Source: (Research Results, 2025)

b. Data Pre-Processing

The author performed stages to prepare the tweet data for proper processing in the classification stage. The pre-processing steps included[21]:

1. Cleaning : Removal of unnecessary characters or symbols, such as: URLs, mentions (@username), hashtags (#) and numbers or punctuation
2. Case Folding : Conversion of all uppercase letters to lowercase.
3. Tokenizing : Breaking sentences into individual words
4. *Normalization* : Converting informal words to standard forms according to KBBI.
5. Stopword Removal : Elimination of common words that do not significantly contribute to sentiment.
6. Stemming : Converting words to their base form by removing affixes

Preprocessing was performed sequentially to reduce noise and standardize informal Indonesian text before sentiment labeling and classification. The representative transformations are summarized in Table 2.

Table 2. Text Preprocessing Stages for Indonesian-Language Dataset

Stage	Representative Transformation
Cleaning	Film Jumbo!!! https://x.com/... → Film Jumbo
Case folding	Film Jumbo → film jumbo
Tokenizing	film jumbo bagus → {film, jumbo, bagus}
Normalization	gk bagus bgt → tidak bagus banget
Stopword removal	film jumbo sangat bagus dan menghibur → film jumbo bagus menghibur
Stemming	menonton film → tonton film

Source: (Research Results, 2025)

c. Feature Extraction and Sentiment Labeling

1. The processed text is converted into features for classification. Sentiment labels are generated using a lexicon containing positive and negative words with weights from −5 to +5. A word with a positive weight contributes to positive sentiment, a negative weight contributes to negative sentiment, and a zero weight is treated as neutral.
   For each tweet, the weights of recognized words are summed to obtain a sentiment score. The resulting score is used to assign one of three labels: positive, neutral, or negative. The score is capped at 10 to prevent unusually long tweets from producing excessively large scores.
2. Multinomial Naïve Bayes Classification
   The lexicon-generated labels are used as the target classes for the Multinomial Naïve Bayes classifier. The processed text is represented as word-frequency features, and the classifier estimates the probability of each sentiment class for an unseen tweet.

d. Model Evaluation

The model is evaluated using accuracy, precision, recall, and F1-score. Three train–test ratios (90:10, 80:20, and 70:30) are compared, and the ratio with the highest reported accuracy is selected for further interpretation

RESULTS AND DISCUSSION

Sentiment Analysis Data Collection

The data source in this research consists of a collection of tweets from Twitter (X) related to the research topic, namely the "film Jumbo", with Google Colaboratory used as the tool. The steps are as follows:

1. Logging into Twitter (X) : Access https://x.com/login using the owned Twitter (X) account and obtain the Twitter (X) auth token
2. Accessing Google Colaboratory : Open Google Colaboratory and navigate to the link https://colab.research.*google*.com/drive/1gXFfewuiJLpHbdi-G7DIYvKomTmCFNWK?usp=sharing, sourced from https://github.com/helmisatria/*tweet*-harvest. This link contains instructions for crawling data using tweet-harvest.
3. Crawling Process : Copy the obtained auth token into Google Colaboratory. Perform the search using keywords "Film Jumbo", "#FILMJumbo", and "Jumbo akidah" within the time range from January 1, 2025, to May 31, 2025, resulting in 723 tweets. The search results are then saved to internal storage in CSV format
4. Data Display : Display the data to verify the crawling process that has been completed.

Pre-Processing Data

Data preprocessing involves several stages, including cleansing, case folding, tokenizing, stopword removal, normalization, and stemming. This process is crucial because raw text data often contains a significant amount of noise that can negatively affect the quality of sentiment analysis:

1. Cleaning

The data is cleaned from unnecessary attributes for subsequent analysis. In this stage, the data is cleaned by removing unnecessary attributes such as numbers, characters, and symbols that are not relevant for subsequent sentiment analysis

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8037

Table 3. Text Preprocessing Cleaning Result for Indonesian-Language Dataset

No.	Tweet Count	Tweet URL	Cleaning
1	141	x.com/undefined/status/19453268	Nama profilku belum ganti karena kenyataannya
2	0	x.com/undefined/status/...	Cerita tidak akan menjadi cerita kalau tidak...
3	1	x.com/undefined/status/...	Jumbo Full Movie Tonton Dari Tautan Di Bawah ini

Source: (Research Results, 2025)

Table 3 shows the collected data is then saved in a file format, such as CSV, for easy further processing. The next stage is data preprocessing, which removes irrelevant elements such as URLs, mentions (@username), hashtags, and other special characters. Furthermore, the text is converted to lowercase to facilitate further analysis and the data is ready to be used for analysis, to find out the tendency of public opinion towards the film Jumbo, topic analysis to identify frequently discussed issues, displaying data visualization to see the most dominant words that appear.

2. Case Folding
The process of converting uppercase letters to lowercase is essential for text preprocessing in research.

Table 4. Text Preprocessing Case Folding Result for Indonesian-Language Dataset

No.	Tweet Count	Tweet URL	Case Folding
1	141	x.com/undefined/status/19453268	nama profilku belum ganti karena kenyataannya...
2	0	x.com/undefined/status/...	cerita tidak akan menjadi cerita kalau tidak...
3	1	x.com/undefined/status/...	jumbo full movie tonton dari tautan di bawah ini

Source: (Research Results, 2025)

Table 4 shows the test data and training data were processed in the case folding stage where the data in the form of letter characters will be converted to lowercase. The data from the document cleansing process that has been acquired. is cleaned from characters such as html, hashtags, username (@username), punctuation marks such as (. , ? ! [ ] / % : ; < > ( ) *)), numbers (0, 1,2,3,4,5,6,7,8,9) and other characters besides the alphabet. The purpose of the cleaning process is to reduce noise

3. Tokenizing
This stage involves breaking down sentences into individual words from the review data. This process serves as an initial step before further text analysis, such as stemming, feature extraction, or sentiment analysis.

Table 5. Text Preprocessing Tokenizing Result for Indonesian-Language Dataset

No.	Tweet Count	Tweet URL	Tokenizing
1	141	x.com/undefined/status/19453268	[nama, profilku, belum, ganti, karena, kenyataannya, ...]
2	0	x.com/undefined/status/...	[cerita, cerita, menjadi, kalau, tidak, ...]
3	1	x.com/undefined/status/...	[jumbo, full, movie, tonton, taitan, di, bawah, ...]

Source: (Research Results, 2025)

Table 5 shows the collected tweet data related to the film Jumbo was parsed word by word. This process was carried out after the data cleaning stage, so that the processed text was free of irrelevant elements such as URLs, mentions, hashtags, and special characters. Tokenizing helps simplify the text structure, making it easier to analyze in the next stage. For example, the sentence “Film Jumbo sangat bagus dan menghibur” would be converted into tokens such as [“film”, “jumbo”, “sangat”, “bagus”, “dan”, “menghibur”]. The code above uses the NLTK (Natural Language Toolkit) library to perform tokenization on the text contained in the Case Folding column of a DataFrame. Before running the code, ensure that the NLTK library has been downloaded.

4. Stopwords
This stage involves removing words that do not carry meaningful information in data processing. Examples of stopwords in Indonesian include "di," "yang," "dan," and "ke".

Table 6. Text Preprocessing Stopword Result for Indonesian-Language Dataset

No.	Tweet Count	Tweet URL	Stopword
1	141	x.com/undefined/status/19453268	[nama, profilku, ganti, nyata...]
2	0	x.com/undefined/status/...	[nama, profilku, nyata]
3	1	x.com/undefined/status/...	[jumbo, movie, tonton, bawah]

Source: (Research Results, 2025)

Table 6 shows common words that appear frequently but have no significant meaning in the analysis, such as "and," "yang," "di," and so on,

should be removed. This step aims to produce more consistent and representative data. The code utilizes Indonesian stopwords from the NLTK library along with an additional manually defined list. Stopwords are applied to the text in the Normalization column.

5. Normalization

This stage involves standardizing various text forms that have the same meaning. For example, the words "dgn," "dg," and "dengan" can be normalized to "dengan".

Table 7. Text Preprocessing Normalization Result for Indonesian-Language Dataset

No.	Tweet Count	Tweet URL	Normalization
1	141	x.com/undefined/status/19453268	[nama, profilku, ganti, nyata]
2	0	x.com/undefined/status/...	[nama, profilku, ganti, nyata]
3	1	x.com/undefined/status/...	[jumbo, full, movie, tonton, tautan, bawah]

Source: (Research Results, 2025)

Table 7 shows that each tweet has been converted into token form (word by word), then adjusted using a normalization dictionary. For example, in the first data set, the word "profileku" was normalized to "profilku," and the word "kenyatanya" was simplified to "kenyataannya." This process demonstrates that normalization not only improves spelling but also simplifies word forms for more effective analysis. Furthermore, other data sets show corrections for spelling errors, such as "taitan," which was changed to "tautan." This is important because even small spelling errors can impact analysis results, particularly in word frequency calculations or text classification processes. With normalization, words that previously had different spellings but shared meanings can be standardized into a single form.

The final result of the normalization process, shown in the image, shows that the data has become cleaner, more structured, and easier to understand. Each tweet is now represented in a standardized word list, ready for use in subsequent analysis stages, such as sentiment analysis or topic clustering. Thus, normalization plays an important role in improving data quality and the accuracy of analysis results in research related to public opinion on the "film Jumbo". This code reads the file normalisasi.csv, which contains a list of non-standard words (such as abbreviations or slang) and their standard forms. It then converts non-standard words in the text found in the tokenize column into their standard forms according to the mappings in the normalisasi.csv file. If a word is not found in the list, it remains unchanged.

6. *Stemming*

This stage involves removing affixes from words to reduce them to their root forms. The code performs stemming using the Sastrawi library along with Swifter to efficiently handle large datasets. Sastrawi is a specialized library designed for stemming Indonesian language text. The text subjected to stemming is taken from the stopwords column.

Table 8. Text Preprocessing Stemming Result for Indonesian-Language Dataset

Index	Stemming Result
0	['nama', 'profilku', 'ganti', 'nyata', 'pamit...']
1	['cerita', 'cerita', 'dengar']
2	['jumbo', 'full', 'movie', 'tonton', 'taut']
3	['bang', 'acil']
4	['iseng', 'gambar', 'atta', 'film', 'jumbo', ...]
...	...
718	['op', 'tidak', 'nonton', 'yang', 'seperti', ...]
719	['sudah', 'laku', 'cinta', 'kabib', 'jumbo', ...]
720	['tolol', 'nah', 'muncul', 'dari', 'orang', ...]
721	['iya', 'saya', 'tuju', 'urus', 'akidah', 'mua...']
722	['hahaha', 'insyaallah', 'wan', 'awek', 'comel...']

Source: (Research Results, 2025)

After completing all preprocessing steps, the processed data is saved by retaining only the *stemming* column for use in subsequent processes. The data is stored in a CSV file named Pre Processing.

Data Visualisation

This stage aims to display a list of frequently occurring words in the processed data. The visualization method used is a Wordcloud. A visual representation of the most frequently occurring words in the text data. The size of each word indicates its frequency, with more frequently appearing words displayed larger.

Source: (Research Results, 2025)

Figure 2. Word Cloud of Frequently Occurring Words in the Indonesian-Language Dataset

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8037

Data Labelling

Data labelling stages play an important role in the classification process because this research uses an approach at the word level where the data processed is the word to obtain sentiment scores, The next stage is the process of weighting each word in the previously processed data. This can be seen in the table below.

Table 9 Sentiment Lexicon Weight Categories

Sentiment Analysis	Amount *Tweets*	Weight Range	Description
Positive	*402*	*> 0 (1 to 5)*	*Words with positive meaning*
Neutral	*203*	*= 0*	*Words with no sentiment tendency*
Negative	*118*	*< 0 (-1 to -5)*	*Words with negative meaning*

Source: (Research Results, 2025)

Table 10. Sentiment Classification Performance

No	Sentiment Type	Precision	Recall
1	Positive	88%	83%
2	Negative	75%	71%
3	Neutral	40%	57%

Source: (Research Results, 2025)

Table 11. Interpretation of Results

Aspect	Description
Dominant Sentiment	Positive sentiment dominates public opinion
Model Strength	High precision and recall in positive sentiment classification
Model Weakness	Lower performance in neutral sentiment classification
Insight	Useful for understanding audience perception and supporting decision-making in the film industry

Source: (Research Results, 2025)

The Table 11 shows In the feature extraction stage, data labeling is performed based on a lexicon-based dictionary consisting of positive and negative data stored in CSV files. Each file contains words and their corresponding weights, ranging from -5 to 5. Positive data have weights greater than 0, negative data have weights less than 0, and words with a weight of 0 are categorized as neutral. After assigning weights to each word in the data, the weights of all words within a single review are summed to produce a sentiment score. This sentiment score is then used to determine the overall sentiment of each review as positive, negative, or neutral. The maximum sentiment score is capped at 10 to ensure that the score does not become excessively large, even if the text contains many words

Table 12 The Result Data Labelling for Indonesian-Language Dataset

No	*Stemming*	Sentiment Analysis
1	terima kasih menciptakan karya ditonton anak bioskop pengalaman pertamanya semoga nilainilai kebaikan disuarakan menancap inti ingatannya	*positive*
2	jujur nonton jumbo *film* hantu ya bund sebenarnya wakakakak	*negative*
3	aspek mustahil *film* jumbo lihat pejabat ditangkap hantu saya percaya hantu pejabat ditangkap	*negative*

Source: (Research Results, 2025)

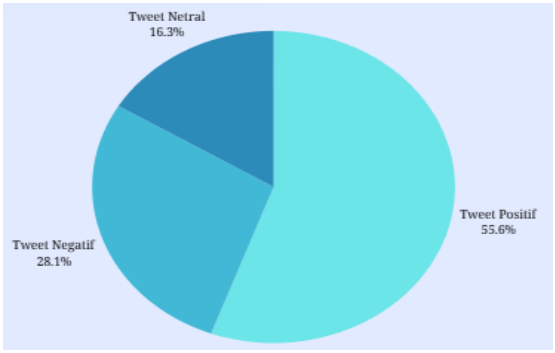
Based on table 12, the results of text data processing that has undergone stemming and sentiment analysis, variations in public opinion regarding the “film Jumbo" are evident. The analyzed data shows three examples of reviews with different sentiment categories: positive and negative. In the first data set, the stemmed text shows expressions of appreciation for the film "Jumbo," such as "thank you for creating this work" and "your first experience." This review carries a positive connotation because viewers felt the film provided a memorable experience, especially for children who were watching it in a cinema for the first time. Furthermore, there is hope that the positive values in the film will be embedded in the audience's memory. This reinforces the reason why this review is categorized as positive sentiment. In the second data set, despite the presence of humorous elements such as "wakakakak," the text demonstrates a less serious perception of the film, calling it a "film hantu."

This expression can be interpreted as criticism or a discrepancy between audience expectations and the film's content. Therefore, this review is classified as negative sentiment, despite its humorous tone. Furthermore, in the third data set, reviews contained ironic and unrealistic statements, such as "tidak mungkin" and comparisons between ghosts and officials. These statements indicate disbelief or criticism of the storyline of the film "Jumbo." The use of these words reflects dissatisfaction or a negative view of the film, thus categorizing it as negative sentiment. Overall, the results of these reviews indicate that despite positive appreciation for the educational and experiential value of "Jumbo," there was still criticism from viewers regarding aspects of the story and perceptions deemed unrealistic. This illustrates that public opinion towards the film is diverse and influenced by each viewer's perspective.

Sentiment Data Visualisation

After labeling and weighting the data,

sentiment visualization is performed to show the percentage distribution of each sentiment category. The visualization method used is a pie chart created with the Matplotlib library. Matplotlib is a widely used Python data visualization library that enables the creation of various types of graphs and plots. Its flexibility allows users to produce simple to complex visualizations, making it ideal for effectively presenting sentiment analysis results



Source: (Research Results, 2025)
Figure 3. Percentage of Tweet Labels

Figure above shows that positive sentiment dominates compared to other sentiments. With a percentage of 55.6% or 402 positive labels, 28.1% or 203 negative labels, and 16.3% or 118 neutral labels. The results are then saved in a CSV file. The algorithm used in this study is Multinomial Naïve Bayes. Naïve Bayes is a machine learning algorithm based on Bayes' Theorem with the "naive" assumption that each feature in the dataset is independent of the others. Although this assumption rarely holds true in reality, the algorithm is highly effective for various tasks, especially in text classification such as sentiment analysis, spam detection, and others.

Model Development

The review data that has been processed up to the feature extraction stage is divided into training and testing datasets. The training data is first used to build the classification model, which then serves as a reference for classifying the testing data. The classification results on the testing data are subsequently evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score

Model evaluation

Model evaluation in machine learning is the process of measuring the performance of a model in predicting or solving a specific problem. The purpose of evaluation is to ensure that the model not only performs well on the training data but also generalizes effectively to new, unseen data (testing data).

Table 13 Rasio, Data Train and Test Data

No	*Rasio*	*Data Train*	*Test Data*
1	90:10	650	72
2	80:20	578	144
3	70:30	506	216

Source: (Research Results, 2025)

After determining the ratios to be used, the accuracy values for all three data ratios above were then calculated. The data with the highest accuracy will be selected for the confusion matrix computation.

Table 14. Accuracy Comparison Based on Data Splitting

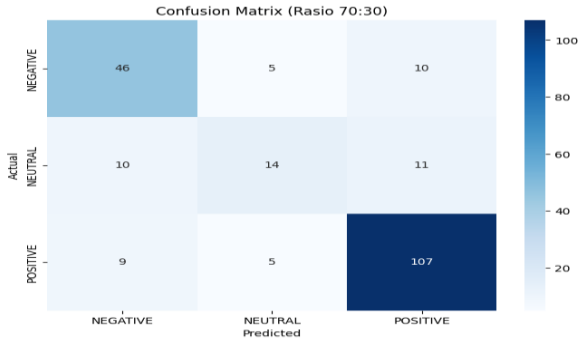
Training–Testing Split	**Accuracy**
90:10	0.6849
80:20	0.7586
70:30	**0.7788**

Source: (Research Results, 2025)

From the results above, it can be concluded that among the three ratios, the 70:30 ratio has the highest accuracy value of 0.77 or 77%.

Confusion Matrix

The next step is to calculate the confusion matrix, which helps evaluate how well the predictions performed.



Source: (Research Results, 2025)
Figure 4. Confusion Matrix of the Sentiment Classification Model

From the results obtained, the correct predictions for positive sentiment (true positive) were 107, for negative sentiment (true negative) were 46, and for neutral sentiment (true neutral) were 14 Then, the calculation of precision, recall, and F1-Score was performed.

Table 15. Accuracy, Precision, Recall, and F1- Score.

Accuracy	*Precision*	*Recall*	*F1-Score*
77%	80%	83%	86%

Source: (Research Results, 2025)

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8037

Interpretation of Results

The next stage to be conducted is sentiment prediction and interpretation of the prediction results. This stage is performed by creating a program to input words, which will then be processed and produce sentiment results from the inputted words.

Table 16. Sentiment Analysis Application Results for Indonesian-Language Dataset

No.	Input Text	Sentiment	Negative	Neutral	Positive	Interpretation
1	*film jumbo bagus banget tt*	Positive	17.55%	6.43 %	76.02 %	Positive words detected. (*senang, puas, bagus*)
2	*film ini kerjasama dengan hantu*	Positive	19.14%	5.86 %	75.00 %	Positive words detected. (*puas, bagus*)

Source: (Research Results, 2025)

From the table 16, when testing with 2 different texts, the Naïve Bayes prediction shows positive results. This occurs because the analyzed data between positive and negative data is imbalanced, causing the inputted negative words to be biased and produce positive results. The results of sentiment analysis on tweet data for the film Jumbo show that public opinion tends to be dominated by positive sentiment, indicating that the audience appreciates the film, especially in terms of moral values, viewing experience, and the entertaining impression for the audience, especially children and families. However, there are also negative sentiments that emerge as a form of criticism of several aspects of the film, such as the storyline that is considered less realistic or does not match the audience's expectations, including comments with sarcastic or humorous nuances that are still classified as negative. Meanwhile, neutral sentiment is relatively less and indicates that some opinions do not have a strong emotional tendency.

Inter-Class Sentiment Analysis.

Inter-Class Sentiment Analysis. The reported class-level results show that positive sentiment achieved 88% precision and 83% recall, corresponding to an F1-score of 85.4%. Negative sentiment achieved 75% precision and 71% recall, with an F1-score of 72.9%. Neutral sentiment obtained 40% precision and 57% recall, with an F1-score of 47.5%. The lower neutral-class performance indicates that neutral expressions are more difficult to distinguish from positive or negative language in informal Twitter text. The best train–test ratio was 70:30, with a reported accuracy of 77.9%

CONCLUSION

Based on the analysis of 723 Indonesian-language Twitter (X) posts about the animated film "Jumbo", the proposed pipeline successfully transformed raw social-media text through cleaning, case folding, tokenization, normalization, stopword removal, and stemming, followed by lexicon-based labeling and Multinomial Naïve Bayes classification. Positive sentiment was the dominant class, accounting for 55.6% of the dataset, followed by negative (28.1%) and neutral (16.3%) sentiment. The best reported train–test ratio was 70:30 with an accuracy of 77.9%. Class-level evaluation indicates that the model performs best for positive sentiment and less effectively for neutral sentiment. These findings suggest that audience responses to "Jumbo" were generally positive, while criticism remained present. The main limitation is the reliance on lexicon-generated labels and the difficulty of handling sarcasm, irony, and contextual language in Indonesian social-media text. Future work should validate labels through manual annotation and compare Multinomial Naïve Bayes with stronger contextual models such as IndoBERT.

REFERENCE

[1] P. S. Zalukhu, T. Handhayani, M. Sitorus, T. Informatika, and U. Tarumanagara, "Kenaikan Bbm Di Indonesia Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes," vol. 8, no. 1, 2023, doi: 10.51876/simtek.v8i1.177.

[2] A. Gosai, S. De, and K. Thankachan, "Sentiment Analysis on Movie Reviews : A Deep Dive into Modern Techniques and Open Challenges," pp. 1–31, 2026, doi: 10.48550/arXiv.2601.07235.

[3] M. A. Al-Garadi, Y. C. Yang, and A. Sarker, "The Role of Natural Language Processing during the COVID-19 Pandemic: Health Applications, Opportunities, and Challenges," *Healthc.*, vol. 10, no. 11, pp. 1–19, 2022, doi: 10.3390/healthcare10112270.

[4] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 148–155, 2022, [Online]. Available:

https://djournals.com/klik

[5] A. R. Abdillah, F. N. Hasan, T. Informatika, F. N. Hasan, D. Mining, and A. Sentimen, "Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier Sentiment Analysis of Presidential Candidates Based on Tweets on," vol. 13, no. 1, pp. 117–130, 2024, doi: 10.47065/bits.v6i2.5766.

[6] A. Prajela, D. Permana, and D. Fitria, "Twitter Data Sentimen Analysis for 2024 Presidential Candidate using Algorithm Naïve Bayes Classifier by K-Fold Cross Validation Methods," vol. 2, pp. 99–104, 2024, doi: 10.24036/ujsds/vol2-iss1/149.

[7] A. Hasnining, "Text Mining Untuk Klasifikasi Emosi Pengguna Media Sosial Degan Algoritma Naïve Bayes," vol. 7, no. 1, pp. 57–67, 2023, doi: 10.33857/patj.v7i1.671.

[8] A. Sitanggang, Y. Umaidah, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes," Jurnal Informatika dan Teknik Elektro Terapan, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4902.

[9] M. D. Lestari and I. Komputer, "Analisis Data Mining Untuk Pemantauan Kesehatan," *Duniadata.org*, vol. 1, no. 6, pp. 1–18, 2024.

[10] W. Astriani, O. S. Bachri, and B. Irawan, "Classification of Sentiment of Emina Product Reviews Using the Naive Bayes Algorithm," *J. Bin. Digit. - Technol.*, vol. 8, no. 2, 2025, doi: 10.32877/bt.v8i2.3554.

[11] M. K. Anam, T. A. Fitri, A. Agustin, L. Lusiana, M. B. Firdaus, and A. T. Nurhuda, "Sentiment Analysis for Online Learning using The Lexicon-Based Method and The Support Vector Machine Algorithm," *Ilk. J. Ilm.*, vol. 15, no. 2, pp. 290–302, 2023, doi: 10.33096/ilkom.v15i2.1590.290-302.

[12] I. A. Ghofur and A. D. Putri, "Perbandingan Kinerja Svm, Knn , Dan Naïve Bayes Pada Analisis Sentimen Tweet MBG," Computer Based Information System Journal, vol. 14, no. 1, pp. 100–117, Mar. 2026, doi: 10.33884/cbis.v14i1.10994.

[13] J. J. A. Limbong, I. Sembiring, and K. D. Hartomo, "Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naive Bayes dan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, p. 347, 2022, doi: 10.25126/jtiik.2022924960.

[14] U. Ayman, T. A. Akash, T. Akter, S. Z. Mridul, and Y. Sutradhar, "BanglaEcomReviewCorpus: A Dataset for E-Commerce Product Review Sentiment Analysis," *Data Br.*, p. 112663, 2026, doi: 10.1016/j.dib.2026.112663.

[15] W. Gumilang and A. Riyandi, "Sentimen Analisis Pengguna Twitter Terhadap Sea Games 2023 Dengan Metode Naive Bayes," Jurnal Akademika, vol. 16, no. 1, pp. 95–100, Nov. 2023, doi: 10.53564/akademika.v16i1.1125.

[16] D. P. Wilie, "Analisis Sentimen Opini Publik Terhadap Chatgpt di Twitter Menggunakan Metode Naive Bayes," Jurnal Nasional Ilmu Komputer, vol. 4, no. 4, pp. 35–44, Nov. 2023, doi: 10.47747/jurnalnik.v4i4.1417.

[17] S. B. Ade Bagus Ferdiyawan, Nurasiah, Imam Fauzy, Winanti, "Analisis Sentimen Pada Media Sosial Twitter(X) Terhadap Pemain Diaspora Di Tim Nasional Sepak Bola Indonesia Studi Kasus Pesepakbola Mees Hilgers Dan Eliano Reinjders Menggunakan Metode Naïve Bayes Dan Support Vector Machine (SVM)," *J. Ipsikom Vol 13 no. 1 - Juni 2025*, vol. 13, no. 1, pp. 63–70, 2025, doi: https://doi.org/10.58217/ipsikom.v13i1.418.

[18] B. W. Andrian, F. A. T. Tobing, I. Z. Pane, and A. Kusnadi, "Implementation of Naïve Bayes Algorithm in Sentiment Analysis of Twitter Social Media Users Regarding Their Interest to Pay the Tax," International Journal of Science, Technology & Management, vol. 4, no. 6, pp. 1733–1742, Nov. 2023, doi: 10.46729/ijstm.v4i6.1015.

[19] S. Riyadi, M. D. Mubarok, C. Damarjati, and A. J. Ishak, "Improving Sentiment Analysis Accuracy Using CRNN on Imbalanced Data: a Case Study of Indonesian National Football Coach," *JUITA J. Inform.*, vol. 12, no. 2, p. 159, 2024, doi: 10.30595/juita.v12i2.21847.

[20] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," Jurnal Sains Komputasi dan Informatika (J-SAKTI), vol. 5, no. 2, pp. 697–711, Sep. 2021, doi: 10.30645/j-sakti.v5i2.368.

[21] M. A. Palomino, "applied sciences Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *MDPI Stay.*, pp. 1–21, 2022, doi: https://doi.org/10.3390/app12178765.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri