

The Application of Naive Bayes in Analyzing Public Sentiment Toward the Performance of the North Sumatra Regional Government in Handling Flash Floods

Rivaldo Siburian^{1*}, Rikki Josua Tampubolon², Valentino Surbakti³, M. Irvandy Haris⁴, Rizky Rahmansyah⁵

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Prima Indonesia

Corresponding Author: Rivaldo Siburian rivaldosiburian2906@gmail.com

ARTICLE INFO

Keywords: Sentiment Analysis, Naive Bayes, Flash Flood, Government Performance

Received : 16, November

Revised : 18, January

Accepted: 20, March

©2026 Siburian, Tampubolon, Surbakti, Haris, Rahmansyah:
This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

This study analyzes public sentiment towards the performance of the North Sumatra Regional Government in handling flash floods using the Multinomial Naive Bayes algorithm. A total of 1,132 opinion data points were collected from social media and news portals through web crawling from November 2025 to February 2026. Sentiment labeling was performed using a lexicon-based approach with the InSet dictionary. Classification results showed a dominance of negative sentiment at 88.4%, focusing on slow emergency response. Model evaluation with an 80:20 data split yielded 89.43% accuracy and an F1-Score of 0.844 for Naive Bayes, while SVM achieved the highest F1-Score (0.855). This study concludes that AI-based sentiment analysis can serve as an objective instrument for government performance auditing.

INTRODUCTION

Flash floods are one of the most frequent hydrometeorological disasters in Indonesia, including North Sumatra Province (Meidina, 2025). This disaster not only threatens the safety of lives, but also causes significant disruption to the social and economic stability of the community. In an emergency situation like this, the performance of local governments is a crucial factor that determines the speed or slowness of the disaster management process, starting from emergency response, aid distribution, to public communication (DESTA, 2025). Therefore, evaluation of the government's performance in handling floods is very important to be carried out in a sustainable manner.

In today's digital era, public opinion on government performance is no longer only channeled through official channels or conventional media. Online media and social media such as X (Twitter), Facebook, and comment columns on news portals have become virtual public spaces that are very actively used by the public to voice perceptions, criticisms, appreciation, and complaints (Febriana et al., 2025). The textual data generated from these platforms is abundant, but unfortunately it is still stored in an unstructured form that is difficult to analyze manually.

The availability of this large public opinion data actually holds extraordinary potential to be processed into valuable information to evaluate government performance. One of the most relevant techniques for extracting meaning from unstructured text data is sentiment analysis. Sentiment analysis allows us to classify public opinion into positive, negative, or neutral sentiment categories, so that it can be known the tendency of public perception of a policy or the performance of an institution (Safitri et al., 2021).

In the context of handling flash flood disasters in North Sumatra, sentiment analysis is a very potential instrument to measure the extent to which people are satisfied or dissatisfied with the performance of local governments. Unfortunately, research that specifically examines sentiment analysis for government performance evaluation in disaster management is still very limited, especially when compared to studies in the commercial or digital entertainment sectors. In fact, public opinion in times of crisis is crucial for the evaluation of policies and public trust.

To answer this need, this study implements a machine learning approach using a classification algorithm. The algorithm chosen as the main focus is Multinomial Naive Bayes, which is known for its simplicity, computational efficiency, and reliability in handling the classification of medium- to large-scale texts (helmi Setyawan et al., 2025). These algorithms work on the basis of Bayes' theorem assuming independence between features, which, although simple, often provide competitive results in practice.

This study not only uses one algorithm, but also compares with other algorithms such as Support Vector Machine (SVM), Logistic Regression, and Random Forest. This comparative approach aims to understand the advantages and limitations of each method in handling complex public opinion data (Fadhilah et al., 2025). Thus, this study not only produces findings about public sentiment, but also makes a methodological contribution regarding the selection of the most appropriate algorithm for similar contexts.

The data used in this study was collected through web crawling techniques over a certain time frame. The data collection process was focused on opinion texts that discussed the handling of flash floods in North Sumatra, which were sourced from online media, social media, and news reader comments. Specific queries are used to ensure that the data collected is relevant to the topic of local government performance in emergency response, aid distribution, and public communication.

From the initial collection process that resulted in more than 1,500 entries, systematic filtration was then carried out to ensure the quality of the data. Strict inclusion criteria are implemented, including the elimination of duplicate opinions such as retweets or reposts, as well as the cleanup of text that does not contain the substance of relevant information. This process is important to minimize potential bias in algorithmic learning and to ensure the validity of the classification model evaluation.

Before the data can be analyzed, the pre-processing stage of data is carried out systematically. These stages include case folding to equalize the text formatting, cleaning to remove unnecessary characters, tokenization to break text into individual words, stopword removal to eliminate common words that don't mean anything, and stemming to change the word to its basic form. This entire process aims to transform raw text into structured data that is ready to be processed.

One of the biggest challenges in sentiment analysis is the subjectivity in data labeling. To overcome this, this study uses a lexicon-based approach with the InSet dictionary as an instrument for standardizing objective assessment. Furthermore, to verify the reliability of the automated method, this study involved more than one annotator at the validation stage. The level of agreement between the annotators is calculated to ensure that the resulting sentiment labels are consistent and reliable.

Once the data has been successfully collected, cleaned, and labeled, the next step is to build and test the classification model. This study used a data splitting method with several ratio scenarios, such as 80:20, 70:30, and 90:10, to find the most stable composition. The feature transformation is carried out by the TF-IDF method which aims to convert text into numerical representations that can be processed by machine learning algorithms, while maintaining data integrity in each experimental scenario.

The model's performance is then comprehensively evaluated. Because of the potential for class imbalance in public opinion data, this study does not rely solely on accuracy metrics. Metrics such as precision, recall, and especially F1-score are used to measure the balance of the model's performance in identifying even minority sentiment. This is important because in public opinion data, often one of the sentiment classes dominates significantly.

The study is also designed to analyze sentiment not only in general, but also aspect-based and temporal. Aspect-based analysis was carried out by grouping sentiment based on three dimensions of government performance, namely emergency response (evacuation speed), aid distribution (logistics), and public communication (clarity of information from regional officials). Meanwhile, temporal analysis was carried out by comparing sentiment across the phases of the disaster, namely before, during, and after the event, to map the shift in public perception over time.

Thus, this research has a fairly broad scope but remains focused. The problem limit is set on sentiment analysis on the performance of the local government of North Sumatra Province in handling flash floods, based on data from certain digital platforms, with a classification of three sentiment classes (positive, negative, neutral), as well as an evaluation of the performance of the Naive Bayes algorithm compared to the other three algorithms. This study does not analyze emotions specifically, but rather on the polarity of opinions.

The novelty of this study lies in the implementation of sentiment analysis to evaluate the performance of local governments in disaster management, which has been rarely done so far. By adapting the comparative approach between algorithms commonly used in non-governmental online opinion into the scope of public policy and disaster management, this study seeks to fill in the existing literature gaps. Ultimately, this research is expected to function as an empirical evaluation instrument for local authorities to map public perception, identify service deficiencies, and formulate optimization of emergency response policies and public communication strategies based on digital data.

LITERATURE REVIEW

Sentimen Analysis

Sentiment analysis is a branch of natural language processing that aims to identify, extract, and classify the tendency of opinions or emotions of a text towards a particular entity. In practice, sentiment analysis allows the grouping of public opinion into three main categories, namely positive, negative, and neutral sentiment (Saputra et al., 2025). This method has been widely adopted in various domains, from commercial product evaluation to public policy, due to its ability to process unstructured text data that is widespread on digital platforms.

The Naive Bayes Algorithm

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem, in which the probability of a text belonging to a particular class of sentiment is calculated based on the occurrence of words in it, assuming that each word is independent of each other (a naïve assumption). Although these assumptions are often not entirely true in natural language, these algorithms have proven to remain effective and efficient for text classification, including sentiment analysis (Kadam & Kaur, 2018). In this study, a Multinomial Naive Bayes variant was used which is suitable for data in the form of frequency of word occurrence (Manning, 2008).

Algoritma Pemanding (SVM, Logistic Regression, Random Forest)

To test the effectiveness of Multinomial Naive Bayes comparatively, the study involved three other classification algorithms. The Support Vector Machine (SVM) works by searching for an optimal hyperplane that separates data between classes in a high-dimensional feature space, making it particularly effective for complex text data (Saputra et al., 2025). Logistic Regression is a regression algorithm used to predict the probability of a data being included in a certain class, with advantages in terms of simplicity and interpretability (Ulya, 2026). Meanwhile, Random Forest is an ensemble method that builds multiple decision trees and combines their predictions to improve accuracy and reduce the risk of overfitting (Nyoto et al., 2026).

TF-IDF Feature Extraction

Before text data can be processed by classification algorithms, text must be converted into numerical representations. One of the most commonly used feature extraction methods is Term Frequency-Inverse Document Frequency (TF-IDF) (Madasu, 2019). This method calculates the weight of each word based on two factors, namely the frequency of the appearance of a word in a document (TF) and the opposite of the frequency of the document that contains the word (IDF). Thus, words that often appear in one document but rarely appear throughout the document will have a higher weight, so that more informative features can be highlighted (Madasu, 2019). Research shows that the use of TF-IDF can improve sentiment analysis performance by 3-4% compared to other feature extraction methods.

Text Preprocessing

The quality of the text data greatly determines the accuracy of the classification model. Therefore, the pre-processing stage becomes a crucial foundation in any sentiment analysis project (Jalili et al., 2024). In general, the pre-processing flow includes the following stages: case folding (converting all letters to lowercase), cleaning text (cleaning of irrelevant characters such as punctuation and URLs), tokenization (breaking sentences into individual words), stopword removal (removing common words that have no significant meaning), and stemming (change the adjective to the root word) (Sari et al., 2025). This entire process aims to reduce the data dimension while improving the information signals relevant to the classification model.

Lexicon-based and InSet in Sentiment Labeling

One of the main challenges in sentiment analysis is the subjectivity in data labeling. The lexicon-based approach offers a solution by utilizing dictionaries that contain words and their respective sentimental weights (5). For Indonesian-language data, the Indonesian Sentiment (InSet) Lexicon is one of the most widely used instruments (Munawar et al., 2025). This dictionary contains thousands of words in Indonesian that have been given a polarity score within a certain range, so labeling can be done automatically and objectively. Previous research has shown that the use of InSet Lexicon is capable of producing labeling accuracy of up to 92% (Munawar et al., 2025).

METHODOLOGY

Types of Research

This study applies a quantitative approach based on *text mining* and *machine learning* to classify public sentiment on the performance of local governments in handling flash floods in North Sumatra. The quantitative approach was chosen because it aims to process large-scale text datasets into numerical outputs in the form of sentiment distributions and accurate classification model evaluation metrics (Ramadani, 2024). In addition, this study is comparative because it compares the performance of the Multinomial Naive Bayes algorithm against three other algorithms, namely Support Vector Machine (SVM), Random Forest, and Logistic Regression, to identify the most optimal model for digital opinion data.

Time and Place of Execution

This research was carried out in the period of January 2026 to June 2026. The details of the activity schedule include discussion of topic search, journal search, journal determination, dataset search, proposal making, data analysis, system design, *training/testing*, report writing, journal publication, and dissemination. The place where the research was carried out was not specifically mentioned, but the entire data processing and programming process was carried out virtually using a computer device.

Working Procedure

The working procedure of this study follows the systematic flow described in the flowchart. The stages are as follows:

Data Collection

Data was collected in the form of public opinion texts sourced from online media, comments from news readers, and social media (X/Twitter and Facebook) that discussed the handling of flash floods in North Sumatra Province. The technique used is *web crawling* with the help of third-party APIs in the period November 2025 to February 2026. Specific queries used include "Medan flood", "North Sumatra flood", and "flood government performance". All raw data is stored in CSV format. From a total initial population of more than 1,500 entries, after filtering duplicates and texts without substance, 1,132 valid data were obtained.

Data Pre-processing

This stage aims to clean and prepare the text data before analysis. The procedures performed include:

- a. **Case folding:**
Converts all letters to lowercase.
- b. **Cleaning:**
Removing punctuation, URLs, numbers, and irrelevant characters.
- c. **Tokenization:**
Breaking sentences into individual words.
- d. **Stopword removal:**
Removing common words that have no significant meaning.
- e. **Stemming:**
Converting an affixed word into a root word.

Data Labeling and Subjectivity Control

Sentiment labeling (positive, negative, neutral) is carried out with a *lexicon-based approach* using the InSet dictionary (Indonesian Sentiment Lexicon). To verify the reliability of the automated method, more than one annotator is independently involved at the validation stage. The level of agreement between the annotators (*inter-rater reliability*) is calculated to ensure the consistency and objectivity of the label.

Data Class Distribution Analysis

The distribution of the amount of data in each sentiment class is analyzed and visualized in the form of *pie charts* and *word clouds*. This analysis aims to determine the balance of data and potential *class imbalances* that can affect model performance.

Model Development and Testing

The sentiment classification model is built using the Multinomial Naive Bayes algorithm as the main focus. For comparison, SVM, Random Forest, and Logistic Regression models were also built. The data is divided into training data and test data with three ratio scenarios: 80:20, 70:30, and 90:10. Feature extraction using the TF-IDF method. Model testing is carried out on test data (*unseen data*).

Model Evaluation

Model performance is evaluated using a *confusion matrix* that yields four key metrics:

- a. **Accuracy:** the percentage of prediction correctness of the total test data.
- b. **Precision:** the accuracy of predictions in positive classes.
- c. **Recall:** the success of the model identifies all the data that is truly positive.
- d. **F1-Score:** precision harmonic average and recall to assess the balance of the model, especially on unbalanced data.

Tools and Materials

Tools

- a. Computer/laptop devices with the Windows operating system.
- b. Python programming language.
- c. Supporting libraries: Pandas, Scikit-learn, NLTK.

Ingredients

- a. Text data on public opinion from online media, reader comments, and social media (1,132 valid entries).
- b. Datasets that have gone through a sentiment pre-processing and labeling process.

Data Analysis Techniques

Data analysis was carried out in three approaches:

- a. General sentiment classification: grouping opinions into positive, negative, and neutral using a pre-built algorithm.
- b. Aspect-based analysis: sentiment is grouped based on three dimensions of government performance, namely emergency response (evacuation speed), aid distribution (logistics), and public communication (clarity of information from local officials).
- c. Temporal analysis: sentiment compared across three phases of a disaster (before, during, and after the flood) to map shifts in public perception.

The results of the classification were then evaluated based on class distribution and model performance, thus providing a comprehensive picture of public opinion trends and the effectiveness of the methods applied.

RESEARCH RESULT

Overview of Research Data

Data collection was carried out in the period from November 2025 to February 2026 using *web crawling techniques* and third-party APIs on platform X (Twitter), Facebook, and online news portals. Specific queries used include "Medan flood", "North Sumatra flood", and "flood government performance" to capture the dynamics of opinion across disaster phases. All raw data is integrated in CSV format. From the total initial population exceeding 1,500 entries, 1,132 valid data were obtained after going through a systematic filtration process. There was a 24.5% reduction in volume due to the implementation of strict inclusion criteria, including the elimination of duplicate opinions (*retweets/reposts*) and the cleanup of text without the substance of relevant information.

This stage transforms the raw text into structured data through five systematic procedures. Here is an illustration of the transformation of one sample data at each stage:

Table 1. Preprocessing Results

Stages	Results
Teks Asli	Sumatra Floods: Government rejects foreign aid, recovery of affected areas predicted to take 30 years
Case Folding	Sumatra Floods: Government Rejects Foreign Aid, Recovery of Affected Areas Predicted to Take 30 Years
Cleaning	Sumatra Floods Government Rejects Foreign Aid Recovery of Affected Areas Predicted to Take Year
Tokenizing	['flood', 'sumatra', 'government', 'reject', 'aid', 'foreign', 'recovery', 'region', 'affected', 'predicted', 'need', 'year']
Stopword Removal	['flood', 'sumatra', 'government', 'reject', 'aid', 'foreign', 'recovery', 'region', 'affected', 'predicted']
Stemming	['flood', 'sumatra', 'order', 'reject', 'help', 'foreign', 'recover', 'region', 'impact', 'prediction']

Data Distribution and Labeling Analysis

The results of *data preprocessing* are converted into training *data*, labeled sentiment, and statistically depicted. The labeling of 1,132 data uses the main approach of *Lexicon-based Sentiment Analysis* with the InSet dictionary as an objective assessment standardization instrument. To verify the reliability of the automated method, the *role of the annotator* is applied at the validation stage through agreement testing (*inter-rater reliability*) of the data sample.

Table 2. Pelabelan Sentimen

Text	Score	Sentiment
East Kalimantan Cares for Disasters, Governor Harum Issues Official Appeal to Raise Aid for Sumatra	4	Positive
Sumatra Floods: Government rejects foreign aid, recovery of affected areas predicted to take 30 years	1	Negatives
There is a Fund of Rp 500 Billion, the Government Has Not Yet Opened International Aid for Sumatra	3	Neutral

The labeling results showed an unbalanced distribution with a significant dominance of negative sentiment of 88.4% (1,001 data), followed by positive sentiment of 9.3% (105 data) and neutral sentiment of 2.3% (26 data).

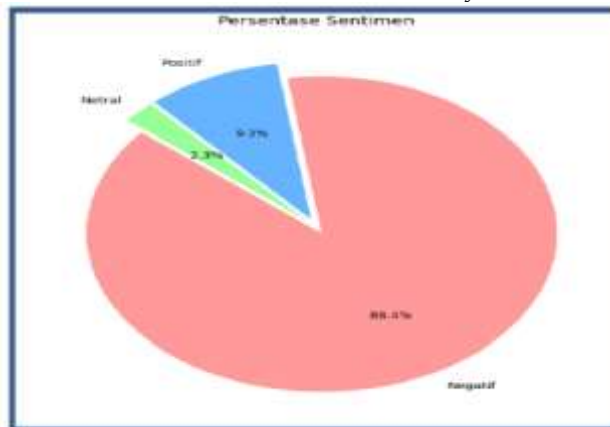


Figure 1. Sentiment Percentage Pie Chart

Classification Model Test Results

Multinomial Naive Bayes model testing on 1,132 valid data using the TF-IDF method for feature transformation. Model performance is evaluated through *Accuracy, Precision, Recall, and *F1-Score** metrics based on the *Confusion Matrix*.

Testing Scenarios

1. Data Sharing

Testing three comparative ratios between Data Training and Data Testing:

- a. Scenario A (80:20):
 80% data (905 lines) for training and 20% data (227 lines) for testing.
- b. Scenario B (70:30):
 70% data (792 lines) for training and 30% data (340 lines) for testing.
- c. Scenario C (90:10):
 90% data for training and 10% for testing.

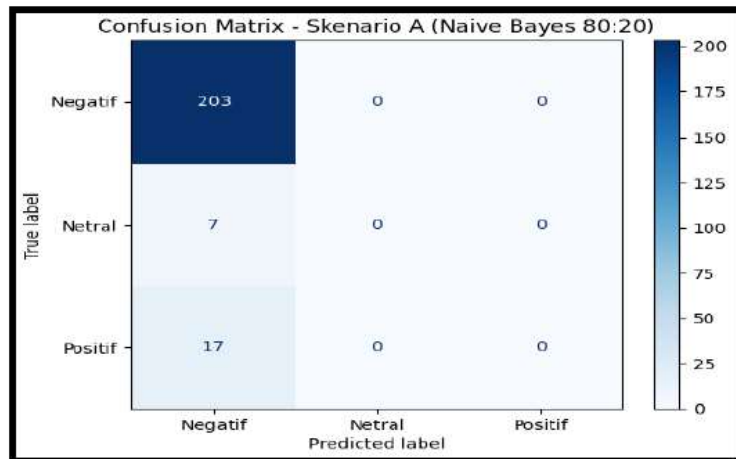


Figure 5. Confusion Matrix Multinomial Naive Bayes Skenario A (80:20)

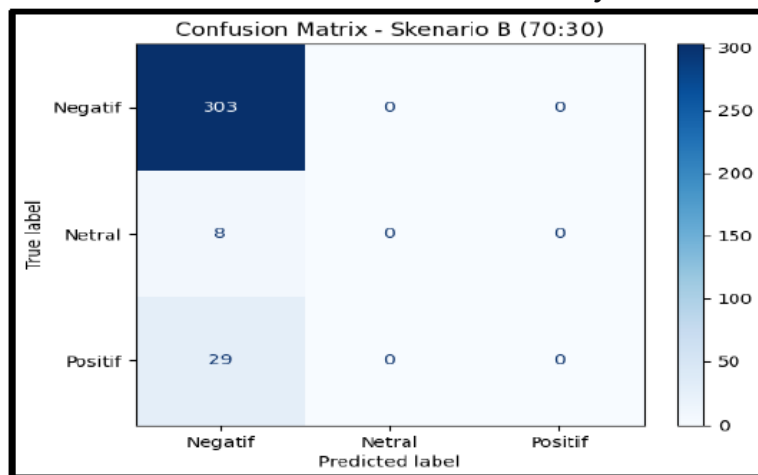


Figure 6. Confusion Matrix Multinomial Naive Bayes Skenario B (70:30)

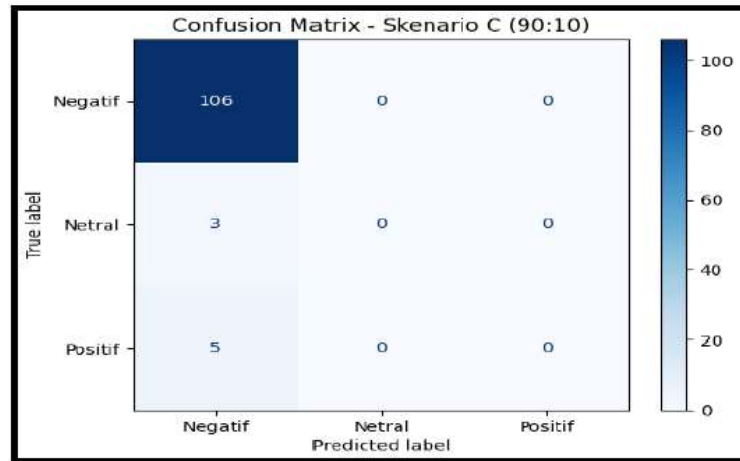


Figure 7. Confusion Matrix Multinomial Naive Bayes Skenario C (90:10)

Algorithm Comparison Results

To validate the effectiveness of Multinomial Naive Bayes, comparative tests were conducted on SVM, Random Forest, and Logistic Regression.

Table 3. Comparison of Accuracy, Precision, Recall, and F1-Score

Algoritma	Accuracy (%)	Precision	Recall	F1-Score
Naïve Bayes	89.43	0.800	0.894	0.844
SVM	89.87	0.878	0.899	0.855
Random Forest	87.22	0.801	0.872	0.835
Logistic Regression	89.43	0.800	0.894	0.844

Based on the table, key analysis points are obtained:

- a. SVM Advantages:
The SVM algorithm shows the highest performance on all metrics, especially the F1-Score (0.855).
- b. Model Consistency:
Multinomial Naive Bayes and Logistic Regression resulted in identical performance values (89.43% accuracy and F1-Score 0.844).
- c. Random Forest Stability:
Despite the lowest accuracy (87.22%), this model remains competitive.

Performance Analysis and Algorithm Characteristics

- a. SVM and Random Forest advantages:
Both models have a higher F1-Score on 227 test data. SVM excels at handling complex text through optimal hyperplane search in high-dimensional space (TF-IDF). Random Forest reduces the possibility of overfitting.
- b. Naive Bayes Characteristics:
Although it is difficult to overcome non-linear relationships with massive text features, Naive Bayes was chosen for its computational efficiency and reliability for the classification of medium-scale texts with short response times.

Aspect-Based and Temporal Sentiment Analysis

The mapping of sentiment movements is divided into three main phases:

- a. Pre-Disaster / Warning Phase:

Dominated by neutral sentiment due to the high volume of factual information dissemination from BMKG.

- b. Emergency Response Phase (At the Time of Incident):
 A sharp spike in negative sentiment due to public panic, slow evacuation, and damage to infrastructure.
- c. Recovery Phase (Post-Disaster):
 Mixed sentiment (hybrid), with appreciation for volunteers and logistical assistance (positive) as well as bureaucratic complaints and slow recovery (negative).

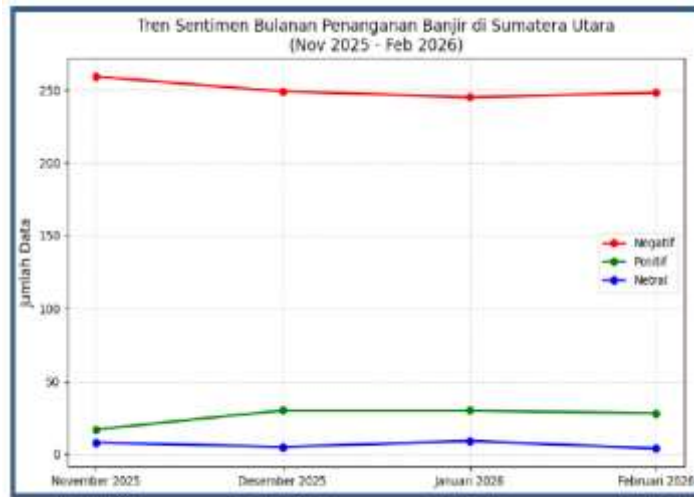


Figure 8. Line Chart Sentimen

Temporal analysis by month:

- a. November - December 2025:
 Neutral and negative sentiment began to be detected along with the early warning of the weather.
- b. January 2026 (Emergency Response Phase):
 Community response to the effectiveness of assistance in critical times.
- c. February 2026 (Recovery Phase):
 Positive sentiment (105 data) began to emerge in response to relief.

Analysis by Aspect

Table 4. Sentiment Based on Government Performance Aspects

Aspects	Negatives	Neutral	Positive
Emergency Response	199	4	2
Distribution of Assistance	260	4	25
Public Communication	49	0	2
Miscellaneous	493	18	51
Total	1.001	26	105

Aspect Description:

- a. Emergency Response:
Assessment of the evacuation speed of the SAR/Regional Government Team.
- b. Distribution of Aid:
Criticism or praise of the distribution of logistics (basic necessities/public kitchens).
- c. Public Communication:
Assessment of the clarity of information from regional officials (Governor/Regent).

DISCUSSION

The dominance of negative sentiment of 88.4% in this study reflects deep public dissatisfaction with the performance of the North Sumatra Regional Government in handling flash floods. This finding is in line with the Kompas R&D survey (December 2025) which recorded that 37.0% of the public stated that the government's response was "Slow" and 10.3% "Very Slow", as well as the MEDIAN survey (January 2026) which revealed that 59.0% of respondents were dissatisfied with the central government's handling of disasters. Thus, social media has proven to be a valid barometer to measure public perception in crisis situations.

An aspect-based analysis showed that Emergency Response (199 negative opinions) and Aid Distribution (260 negative) were the two aspects most criticized by the public. This reinforces the findings (Akbar & Qoiriah, 2025) that the performance of the model in aspect-based sentiment analysis varies widely, with the disaster infrastructure aspect tending to receive the lowest rating. Sebastian, (Meidina, 2025) also emphasized that the Sumatra flash flood has become a test of administrative credibility, where the slowness of response reflects weak coordination and resource mobilization at the provincial level.

In terms of algorithm comparison, SVM showed the highest performance with an F1-Score of 0.855, followed by Naive Bayes and Logistic Regression (0.844), and Random Forest (0.835). The advantages of SVM are consistent with the findings (Saputra et al., 2025) that hyperplane-based algorithms are more precise in handling the complexity of digital opinion texts. Nevertheless, the study retains Multinomial Naive Bayes as the primary focus due to its superior computational efficiency for real-time large-scale data processing, as recommended by (Prasetyo et al., 2023) in the context of social media-based disaster relief identification.

The phenomenon of class imbalance (88.4% negative, 9.3% positive, 2.3% neutral) is a serious challenge in model evaluation. Therefore, this study focuses more on the F1-Score metric of accuracy alone, in line with the recommendation (Amalia, 2023) that uses F1-score weighted to address class imbalances in public stance analysis. This approach ensures that the model's sensitivity to minority classes remains accurately measured.

Temporally, negative sentiment jumped sharply in the emergency response phase (January 2026), while positive sentiment only began to emerge in the recovery phase (February 2026). This pattern shows that public perception is dynamic and is greatly influenced by the speed of government intervention in each phase of the crisis. These findings reinforce the argument that AI-based sentiment analysis can serve as an objective and responsive government performance audit instrument to the dynamics of public opinion (Salsa, 2025).

This research has practical implications for local governments to transform emergency response systems, increase evacuation speeds, and improve public communication strategies. Further studies are recommended to explore class imbalance handling techniques such as SMOTE and implement the IndoBERT architecture to improve the classification precision of texts with high linguistic complexity.

CONCLUSIONS AND RECOMMENDATIONS

This study concludes that public sentiment towards the performance of the North Sumatra Regional Government in handling flash floods is dominated by negative sentiment of 88.4%, which is concentrated on the aspects of emergency response and aid distribution. Temporally, negative sentiment jumped sharply in the emergency response phase (January 2026) and began to subside in the recovery phase. In terms of algorithms, SVM shows the highest performance (F1-Score 0.855), but Multinomial Naive Bayes remains the most efficient choice for real-time data processing due to its simplicity and speed. Local governments are advised to increase the speed of emergency response and improve the aid distribution chain, as well as optimize digital data-based public communication strategies. The government also needs to utilize sentiment analysis as a tool for continuous performance monitoring.

ADVANCED RESEARCH

Further research is recommended to apply class imbalance handling techniques such as SMOTE to increase the sensitivity of the model to minority classes. In addition, implementations of IndoBERT or Transformer architectures can be explored to capture more complex linguistic contexts. The use of multimodal data from platforms such as TikTok and Instagram, as well as the integration of emojis and images, is also recommended to enrich sentiment analysis. Finally, the development of a real-time sentiment monitoring system based on an interactive dashboard can help local governments in faster and more responsive decision-making.

ACKNOWLEDGMENT

Praise be to the author for the presence of God Almighty for the completion of this research report. The author would like to express his deepest gratitude to the supervisor who has provided direction and guidance during the research process. Thank you are also expressed to Universitas Prima Indonesia for supporting this research, as well as to all parties who have assisted in data collection, including the annotators involved in the validation of sentiment labeling. Not to forget, thank you to family and friends who have provided motivation and moral support. Hopefully the results of this research will be useful for the development of science, especially in the application of sentiment analysis for public policy evaluation in Indonesia.

REFERENCES

- Akbar, A. M., & Qoiriah, A. (2025). Analisis Sentimen Berbasis Aspek Kinerja Pemerintah Kota Surabaya Menggunakan Algoritma Naïve Bayes. *Journal Of Informatics And Computer Science (Jinacs)*, 7(02), 539–545.
- Amalia, N. L. (2023). Analisis Sentimen Pada Perpindahan Ibukota Indonesia Dengan Algoritma Support Vector Machine: Evaluasi Leksikon, Metode Ekstraksi Fitur, Dan Kernel Trick. *Fakultas Sains Dan Teknologi Universitas Islam Negeri Syarif Hidayatullah*.
- Desta, P. R. (2025). Responsivitas Kepala Pelaksana Badan Penanggulangan Bencana Daerah Dalam Menanggulangi Bencana Alam (Studi Kasus Pada Kota Bandar Lampung).
- Fadhillah, O. S. D., Jaman, J. H., & Carudin, C. (2025). Perbandingan Naive Bayes, Support Vector Machine, Logistic Regression Dan Random Forest Dalam Menganalisis Sentimen Mengenai Tiktokshop. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1).
- Febriana, K. A., Watie, E. D. S., Fanani, F., & Setiawan, Y. B. (2025). *Media Sosial Dan Transformasi Komunikasi Publik: Dari Opini Hingga Mitigasi Bencana*. Penerbit Mitra Cendekia Media.
- Helmi Setyawan, M. Y., Herwanto, P., & Wiharko, T. (2025). *Machine Learning: Memahami Multinomial, Distribusi Probabilitas & Multinomial Naive Bayes*. Tangguh Denara Jaya Publisher.
- Jalili, A., Tabrizchi, H., Mosavi, A., & Varkonyi-Koczy, A. R. (2024). Enhancing Language Model Performance With A Novel Text Preprocessing Method. *Acta Physica Polonica: A*, 146(4).

- Kadam, T., & Kaur, P. (2018). Enhanced Approach Of Using Custom Heuristic Rules With Stanford And Naive Bayes Classification Technique. 2018 International Conference On Advances In Communication And Computing Technology (Icacct), 601-605.
- Madasu, A. (2019). A Study Of Feature Extraction Techniques For Sentiment Analysis. Arxiv Preprint Arxiv:1906.01573.
- Manning, C. D. (2008). Introduction To Information Retrieval. Syngress Publishing.
- Meidina, S. C. M. S. C. (2025). Akuntabilitas Pemerintah Daerah Dalam Penanggulangan Bencana Hidrometeorologi Di Sumatera. Equality Before The Law, 5(2).
- Munawar, M., Sartika, D., & Husnah, F. (2025). Lexicon-Based Comparison For Suicide Sentiment Analysis On Twitter (X). Telkomnika (Telecommunication Computing Electronics And Control), 23(5), 1314-1322.
- Nyoto, V. J., Riti, Y. F., & Tantokusumo, R. V. P. (2026). Analisis Perbandingan Algoritma Random Forest, Decision Tree Dan Naive Bayes Dalam Mendeteksi Spam Sms. Jurnal Sains Informatika Terapan, 5(1), 279-286.
- Prasetyo, V. R., Erlangga, G., & Prima, D. A. (2023). Analisis Sentimen Untuk Identifikasi Bantuan Korban Bencana Alam Berdasarkan Data Di Twitter Menggunakan Metode Kmeans Dan Naïve Bayes. Jurnal Teknologi Informasi Dan Ilmu Komputer (Jtiik), 10(5), 1055-1062.
- Ramadani, N. C. (2024). Analisis Sentimen Untuk Mengukur Ulasan Pengguna Aplikasi Mobile Legend Menggunakan Algoritma Naive Bayes, Svm, Random Fores, Decision Tree, Dan Logistic Regression. J. Sist. Inf, 16(1), 123-138.
- Safitri, R., Alfira, N., Tamitiadini, D., Dewi, W. W. A., & Febriani, N. (2021). Analisis Sentimen: Metode Alternatif Penelitian Big Data. Universitas Brawijaya Press.

- Salsa, A. (2025). Analisis Sentimen Tentang Respon Masyarakat Terhadap Pelayanan Digital Administrasi Pertanahan (Studi: Di Kantor Wilayah Badan Pertanahan Nasional Provinsi Lampung).
- Saputra, J., Maryani, L., Wulandari, D., & Eka, W. (2025). Analisis Performa Naive Bayes Dan Svm Terhadap Sentimen Teks Media Sosial Dengan Word2vec Dan Smote. *Jurnal Instek (Informatika Sains Dan Teknologi)*, 10(1), 143-155.
- Sari, A., Irawan, B., Faqih, A., Dikananda, A. R., & Fathurrohman, F. (2025). Analisis Sentimen Ulasan Pengguna Aplikasi Kredivo Menggunakan Algoritma Support Vector Machine (Svm) Dengan Metode Tf-Idf. *Literatur Informatika Dan Komputer*, 2(4), 593-601.
- Ulya, D. M. (2026). Evaluasi Kinerja Algoritma Regresi Logistik Menggunakan Oversampling Smote Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Mobile Jkn. Universitas Islam Negeri Maulana Malik Ibrahim.