Jurnal Media Teknik Elektro dan Komputer
Vol. 02, No.02 (July - December, 2025)
E-ISSN: 3063-2927

Sentiment Analysis of Public Opinion on Road Damage in North Sumatra Using the Naive Bayes Method Based on Weak Supervision (Lexicon-Based)

Kaka Davi Dharmawan [1], Nazwa Aliya Muthmainah Hasibuan [2]

[1]Universitas Islam Negeri Sumatera Utara; kaka0701231028@uinsu.ac.id,
[2]Universitas Islam Negeri Sumatera Utara; nazwa0701232104@uinsu.ac.id

ABSTRACT

Road infrastructure is a vital aspect of regional development that often receives public attention in online media, especially in North Sumatra. Manual monitoring of public opinion on this issue is inefficient due to the large volume of data and the imbalance of sentiment, which is dominated by complaints. This study aims to develop an automatic sentiment analysis model using a Weak Supervision approach that combines the Lexicon-Based method for automatic labelling and the Multinomial Naive Bayes algorithm to classify public opinion into three distinct categories: positive, negative, and neutral. Data was collected through web scraping techniques from various online news portals. To overcome data class imbalance, this study applied the Synthetic Minority Over-sampling Technique (SMOTE) to the training data. Test results on the test data showed that the model was able to achieve an accuracy of 70.93%. The model performed very well in detecting negative sentiment with a Precision value of 0.86, and was able to recognize positive sentiment with a Recall of 0.70 thanks to the application of SMOTE. Based on these results, the Naïve Bayes model can be used effectively to classify public sentiment towards road damage. In addition, these findings serve as strategic references and recommendations for stakeholders, such as the Inspectorate, to formulate relevant and data-driven policies in infrastructure improvement and regional development efforts.

Keywords: *Sentiment Analysis, Naïve Bayes, Lexicon-Based, Weak Supervision, Road Infrastructure*

Corresponding Author:
Kaka Davi Dharmawan
Universitas Islam Negeri Sumatera Utara; kaka0701231028@uinsu.ac.id

1. INTRODUCTION

Road infrastructure is a social asset that plays an important role in supporting the movement of goods and services in a region. In North Sumatra Province, road quality is often a public concern because it directly affects economic activity and the welfare of the community. In today's digital era, the flow of aspirations no longer depends solely on formal complaint mechanisms. Various online news portals have become platforms that quickly and massively record public responses, ranging from complaints about road damage to appreciation for improvements made by the government (Manoppo et al., 2025). Thus, online media coverage not only serves as a provider of factual information but also as a reflection of public sentiment towards the performance of local governments.

For internal supervisory agencies such as the Inspectorate, monitoring the dynamics of public opinion in online media is a strategic necessity for assessing the effectiveness of policies and the level of government responsiveness. However, the increasing volume of news coverage and the diversity of narratives make manual monitoring inefficient and prone to bias. This condition encourages the need to utilize a computational approach through Sentiment Analysis to automate the process of identifying positive, negative, or neutral opinion categories objectively and consistently (Rambe et al., 2025).

One of the main challenges in developing machine learning-based sentiment analysis models is the limited availability of labeled data. The manual labeling process by experts requires a lot of time and money, which often becomes an obstacle in system development. To overcome this, the Weak Supervision approach has begun to be widely used (Firda et al., 2025). This approach allows the initial labeling process to be carried out automatically through dictionary-based or rule-based methods, such as Lexicon-Based, so that it can produce large amounts of labeled datasets without the burden of manual annotation costs (Amal & Jayanta, 2023). These datasets then become the basis for training more adaptive classification models, such as the Naive Bayes Classifier (Syah et al., 2023).

In addition to labeling constraints, the distribution of public opinion data on road infrastructure is generally unbalanced. Negative news reports related to damage or public complaints are far more dominant than positive news. This imbalance can cause models to tend to predict the majority class and ignore the minority class, which actually contains important information, such as successful development. To overcome this, data balancing techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) need to be applied so that the model is able to recognize all sentiment categories more proportionally (Budaya & Suniantara, 2024).

Recent studies in the 2023–2025 period have extensively explored optimization techniques to address these challenges. Syah et al. (2023) and Safira & Hasan (2023) compared Lexicon-based approaches with Naive Bayes, consistently finding that machine learning models tend to achieve higher accuracy but struggle with data labelling constraints. To address class imbalance, Hermawan et al. (2025) and Hardiyanti & Fajarlestari (2025) demonstrated that SMOTE significantly improves Naive Bayes performance by generating synthetic samples for minority classes. Furthermore, Ratnaswari et al. (2025) investigated hybrid approaches by utilizing lexicons for initial sentiment weighting, proving that integrating knowledge-based methods with classification algorithms can enhance model robustness in specific domains.

Based on these issues, this study aims to develop a sentiment analysis model that can map public opinion on road infrastructure in North Sumatra through a hybrid approach (Fatharani & Syahrul, 2025). This method integrates Lexicon-Based automatic labeling with a domain-specific dictionary and a Naive Bayes algorithm enhanced using the SMOTE technique (Haikal et al., 2025). This approach is expected to efficiently process large-scale data without the need for manual annotation, while producing accurate and balanced classification models (Hermawan et al., 2025). Therefore, the results of this study serve as a strategic reference and recommendation for stakeholders to formulate relevant and data-driven policies, particularly in determining infrastructure monitoring priorities in a more responsive manner.

2. LITERATURE REVIEW

2.1 Sentiment Analysis

Sentiment analysis is part of text mining that focuses on extracting opinions, assessments, emotions, and public reactions to a particular issue or entity (Albab et al., 2023). Its main purpose is to identify the polarity of a text, whether it is positive, negative, or neutral. This method has developed rapidly alongside the increasing use of digital media as a space for public expression (Gibran et al., 2024). In the government ecosystem, sentiment analysis plays an important role as a tool for monitoring public perception quickly and based on data, enabling public service providers to follow up on criticism or complaints in a more measured manner.

2.2 Naïve Bayes Classifier Method

The Naïve Bayes Classifier (NBC) is known as a supervised learning algorithm that relies on Bayes' Theorem with the assumption of independence between features (Jelita et al., 2025). In the realm of text classification, the Multinomial Naive Bayes variant is the primary choice due to its ability to process features based on word frequency such as TF-IDF (Oktavia et al., 2025). The advantages of this algorithm lie in its computational efficiency and competitive performance even when faced with high-dimensional data. However, NBC is highly dependent on the quality of the training data; the existence of an unbalanced dataset can cause the model to be biased towards certain classes and reduce overall accuracy.

2.3 Lexicon-Based Method

The Lexicon-Based approach is a sentiment analysis method that does not rely on labeled data (Sabrina et al., 2025). The classification process is carried out by matching words in the document with a list of words in a sentiment dictionary that has been given a specific polarity weight. The polarity values are then summed to determine the sentiment tendency of the document. This model is relatively easy to implement and efficient because it does not require model training. However, its performance is greatly influenced by the completeness and suitability of the dictionary to the domain context. The presence of local terms, irony, or ambiguous words often reduces the accuracy of this method.

2.4 Previous Studies and Research Position

Recent research in the period 2023–2025 has focused intensively on optimizing the Naive Bayes algorithm using data balancing techniques. Kurnia et al. (2023) analyzed public sentiment towards Bank BTN on Twitter using a combination of Naive Bayes, SMOTE, and Adaboost. Their study found that this combination produced the best modelling results, with an accuracy of 87.05%, precision of 90.63%, and recall of 83.00%. This proves that handling noise through pre-processing and data balancing is very important in banking service sentiment analysis.

Additionally, Afuan et al. (2025) applied the SMOTE-Optimized Naive Bayes Classifier to analyse the 2024 Indonesian presidential inauguration on the X platform. Their research results show that handling class imbalance is a critical step, where the application of SMOTE increases the accuracy of NBC from 98% to 99%, with precision increasing significantly from 0.98 to 1.0. In the field of e-commerce, Kurniawan et al. (2025) tested the Naive Bayes Classifier with SMOTE on 2,500 Blibli app reviews. Their research proved that this method is very robust, achieving 90% accuracy, with 92% recall and an F1-score of 91%.

Referring to these studies, this research aims to apply the Naive Bayes and SMOTE methods in a different domain, namely public infrastructure (road damage), which has the unique characteristic of negative sentiment dominance (complaints). The main difference between this study and previous studies lies in the labelling approach. While Kurnia et al. (2023) and Kurniawan et al. (2025) relied on manual labelling or standard datasets, and Afuan et al. (2025) used TextBlob for labelling, this study proposes a Weak Supervision approach. This study integrates lexicon-based automatic tagging with a domain-specific dictionary to efficiently generate a large-scale dataset, which is then used to train a Naive Bayes model optimized with SMOTE. This approach is designed to offer a more efficient solution for government oversight without the high cost of manual tagging.

3. METHODS

The research flow was designed using a quantitative approach with the Knowledge Discovery in Database (KDD) stages. This process includes data collection, automatic labeling, pre-processing, data balancing, and model training and evaluation.

3.1 Data Collections

Research data was collected from various credible national and local online news portals (such as Detik, Kompas, Tribun Medan, Waspada, etc.) with a focus on the topic of "Road Damage in North Sumatra". The data collection process was carried out using the Web Scraping technique by utilizing the Python newspaper3k and requests libraries. The data collected was in the form of article text, which was then broken down into sentences (sentence tokenization) to obtain more specific sentiment granularity. A total of 859 sentences were collected (Hafiz & Sudarmilah, 2023).

3.2 Knowledge-Based Data Labeling (Lexicon-Based Labeling)

Due to the absence of labeled data (ground truth), this study applied the Weak Supervision technique for data labeling. Labeling was performed automatically using the Lexicon-Based method with a specially curated sentiment dictionary (domain-specific dictionary) (Saputra et al., 2025)

1. Dictionary Curation: The dictionary is compiled based on infrastructure terminology, covering negative words such as “genangan” (puddle), “terbengkalai” (derelict), and “hancur” (destroyed), as well as positive words such as “mulus” (smooth), “aspal” (asphalt), and “selesai” (completed).
2. Weighting Rules: Each word is given a score weight (-5 to +5). Sentiment is determined based on the accumulation of sentence scores, taking into account negation rules (e.g., “tidak rusak” or "not damaged" becomes positive) and the context of infrastructure keywords (Fikri et al., 2022).

3.3 Data Preprocessing (Text Preprocessing)

Raw text data obtained from web scraping often contains noise and unstructured formats. In this study, preprocessing was designed to clean the data while explicitly maintaining the integrity of word affixes and functional markers. This approach was chosen because the Weak Supervision (Lexicon-Based) method relies on a specific lexicon where sentiment values are tied to unmodified whole words. Therefore, standard procedures such as stemming and stopword removal were deliberately omitted from the main pipeline to prevent the loss of semantic context and sentiment nuances during the tagging process.

However, to ensure clear and meaningful thematic mapping in the results section, a separate filtering step is applied exclusively for Word Cloud visualization. The specific steps taken for the main dataset are as follows:

1. Case Folding

This step aims to standardize the text format by converting all capital letters to lowercase. This ensures that words such as 'Jalan' and 'jalan' are treated as identical features.

a. Process *:* text.lower()
b. Example : "*Jalan Rusak PARAH*" to "*jalan rusak parah*"

2. Cleansing
This process removes non-alphanumeric characters that do not contain sentiment information, such as punctuation marks, numbers, URL links, and mentions, using Regular Expressions (Regex).
a. Process *:* re.sub(r'[^a-zA-Z\s]', '', text)
b. Example : "*Lapor @pemprovsu jalan berlubang!! #viral*" → "*lapor jalan berlubang viral*"

3. Tokenizing
Sentences are broken down into individual words (tokens) to facilitate the analysis of the frequency and weight of each word during the TF-IDF process.
a. Process: Splitting sentences based on whitespace.
b. Example: "*jalan rusak*" → ['*jalan*', '*rusak*']

To ensure transparency, Table 1 shows the transformation of the top three data samples from the dataset before and after the preprocessing stage.

Table 1. Comparison of data before and after preprocessing

No	Raw Data (Before Processing)	Preprocessed Data (After Processing)
1	*"Satu video memperlihatkan sejumlah warga MEMANCING di jalan yang RUSAK di wilayah Deli Serdang VIRAL!!"*	*"satu video memperlihatkan sejumlah warga memancing di jalan yang rusak di wilayah deli serdang viral"*
2	*"Aksi warga itu untuk memprotes kondisi jalan yang rusak hingga dipenuhi air (27/07/2025)."*	*"aksi warga itu untuk memprotes kondisi jalan yang rusak hingga dipenuhi air"*
3	*"Dilihat detikSumut dari video Minggu 27/7/2025, dalam video terlihat warga berada di jalan rusak yang penuh air. http://detik.com/sumut"*	*"dilihat detiksumut dari video minggu dalam video terlihat warga berada di jalan rusak yang penuh air"*

3.4 Feature Extraction and Data Division

After the pre-processing stage, the text data is converted into numerical vectors using Term Frequency-Inverse Document Frequency (TF-IDF) with a maximum feature limit of 2,000 words. This process converts raw text into a format that can be processed by machine learning algorithms (Ulgasesa et al., 2022).

Next, the dataset consisting of 859 documents is divided into two subsets using the Hold-Out method with a ratio of 80:20. The detailed data distribution is as follows:

1. Training Set (80%): Consists of 687 data points. This subset is used to train the Naive Bayes model. In this study, the SMOTE process is applied only to this training set to address class imbalance without affecting the validity of the test data.
2. Test Set (20%): Consists of 172 data points. This subset is kept separate and used exclusively for final evaluation. The separation was carried out strictly before the data balancing process (SMOTE) to prevent data leakage. This ensures that the evaluation metrics (Accuracy,

Precision, Recall) reflect the model's performance on real-world data, thereby guaranteeing the validity of the research results.

The weight of each word is calculated using the following formula:
Term Frequency (TF): Calculates the frequency of occurrence of a term (t) relative to the total number of words in a document (d). This ensures that longer documents do not unfairly dominate the weighting. The formula is defined as in Equation (1):

$$TF(t,d) = \frac{n_{t,d}}{\sum_k n_{k,d}} \quad (1)$$

Description :
$n_{k,d}$: The number of times term t appears in document d.
$\sum_k n_{k,d}$: The total number of all terms (words) in document d.

Inverse Document Frequency (IDF): Measures how much information a word provides based on how common or rare it is across the entire document corpus (D). The formula is defined as in Equation (2):

$$IDF(t,D) = log\left(\frac{N}{|\{d \in D: t \in d\}|}\right) \quad (2)$$

Description :
N : The total number of documents in the corpus.
$|\{d \in D: t \in d\}|$: The number of documents containing the term t (document frequency).
log : The logarithm function used to dampen the effect of high frequency.

TF-IDF weighting: The final weight is the product of TF and IDF, which assigns a high weight to words that appear frequently in a specific document but are rare in the entire corpus. The formula is defined as in Equation (3):

$$W_{t,d} = TF(t,d) \times IDF(t,D) \quad (3)$$

Description :
$W_{t,d}$: The specific weight of term t in document d.
$TF(t,d)$: The Term Frequency score from Equation (1)
$IDF(t,D)$: The Inverse Document Frequency score from Equation (2).

3.5 *Data Imbalance Handling (Resampling)*

Given the dominance of negative sentiment in the training data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the class distribution (Badriyah et al., 2025). Unlike random over-sampling, which simply duplicates existing data, SMOTE works by generating new synthetic samples based on the similarity of the minority class feature space (Putra & Salam, 2025).

The SMOTE algorithm operates based on the concept of K-Nearest Neighbors (K-NN) concept. The process begins by selecting samples from the minority class x_i and identifying its K-Nearest Neighbors. A neighbor x_{zi} is then selected at random, and a synthetic sample is created along the line segment connecting the two points. The mathematical formula for generating a new synthetic sample is defined as follows:

$$x_{new} = x_i + \delta \times (x_{zi} - x_i) \quad (4)$$

Description :

x_{new} : The generated synthetic sample.

x_i : The original sample vector from the minority class.

x_{zi} : The selected nearest neighbor vector.

δ : A random number between 0 and 1 (scaling factor).

By applying this formula, the model learns the decision boundary of the minority class more effectively without over-fitting, as the new samples introduce variations rather than mere replications (Rahmatullah et al., 2025). In this study, SMOTE is applied only to the training set to prevent data leakage and ensure the integrity of the evaluation results.

3.6 Naïve Bayes Classification Process

As shown in the research flowchart (Figure 1), the classification stage is the core of this study. After the training data was balanced using SMOTE, the Multinomial Naive Bayes (MNB) algorithm was used to build a sentiment classification model. MNB was specifically chosen for its effectiveness in text classification tasks involving discrete data counts, such as word frequencies generated by TF-IDF.

The modelling process begins by calculating the prior probability of each sentiment class (Positive, Negative, Neutral) based on the balanced training data set. The algorithm then calculates the likelihood of each word feature appearing in these classes. To predict the sentiment of a new document (d), the model applies Bayes' theorem to calculate the posterior probability, as defined in Equation (5) (Ratnaswari et al., 2025):

$$P(c|d) = \frac{P(d|c) \times P(c)}{P(d)} \quad (5)$$

Description :

$P(c|d)$: Posterior Probability, the probability of document d belonging to class c.

$P(d|c)$: Likelihood, the probability of document d appearing given class c.

$P(c)$: Prior Probability, the probability of class c occurring in the training data.

$P(d)$: Predictor Prior, the probability of document d occurring (evidence).

In the classification phase, the model classifies documents into the class (c) that has the highest posterior probability ($argmaxP(c|d)$). This process is performed on the Test Set (20%), which is separated separately during the SMOTE process, to ensure that the evaluation reflects the model's performance on previously unseen real-world data.

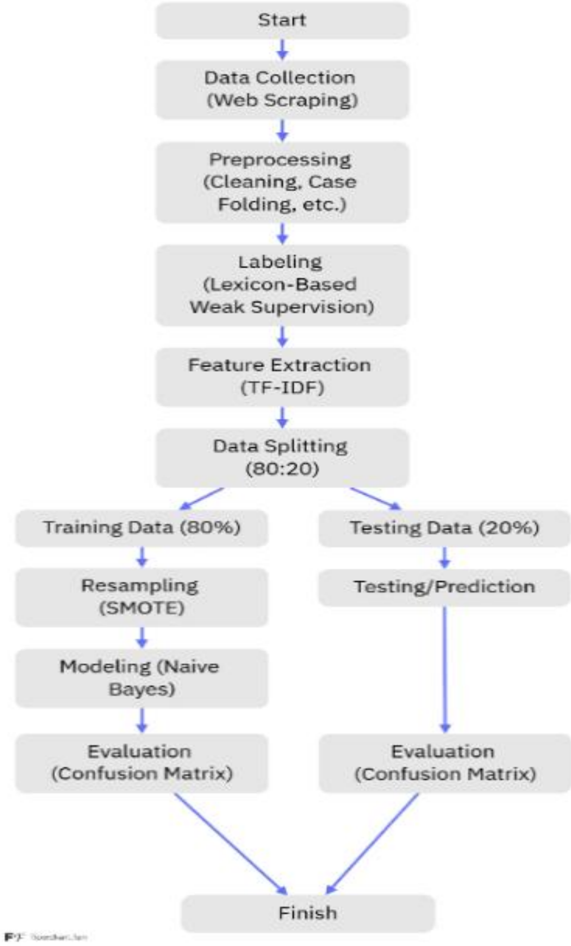

Figure 1. Research methodology flowchart

3.7 *Model Evaluation*

To verify the effectiveness of the proposed sentiment analysis model, an evaluation was conducted using a Test Set consisting of 172 original documents. This data was not processed by the SMOTE process to ensure that the evaluation results reflected the model's ability to handle imbalanced data distribution in the real world (Kurnia et al., 2023). The assessment was based on a Confusion Matrix, which categorizes predictions into four types: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) (Rohmatun & Baita, 2025). The following metrics were used to measure the model's performance:

1. Accuracy

   This metric represents the overall percentage of correct predictions out of the total dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

2. Precision:

   Precision measures the model's exactness, indicating how many of the items identified as a certain class actually belong to that class.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

3. Recall
   Recall measures the model's completeness, signifying the capability to identify all relevant instances within a specific class.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

4. F1-Score
   The F1-Score is the harmonic mean of precision and recall, providing a balanced assessment for imbalanced datasets (Ulhaq & Suprayogi, 2025).

$$f1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

Description :
TP (True Positive) : The number of documents correctly predicted as the target class.
TN (True Negative) : The number of documents correctly predicted as not belonging to the target class.
FP (False Positive) : The number of documents incorrectly predicted as the target class.
FN (False Negative) : The number of documents belonging to the target class but incorrectly predicted as another class.

4. RESULTS AND DISCUSSION

This section presents the results of sentiment analysis experiments on public opinion regarding road damage in North Sumatra in online media. The discussion covers the results of each stage of the methodology that has been carried out, starting from the results of automatic labeling using a lexicon, data visualization before and after balancing with SMOTE, to the evaluation of the performance of the Naive Bayes algorithm on the test data. In addition, this chapter also presents an in-depth analysis of the effectiveness of the model in detecting various sentiment categories.

4.1 *Automatic Labeling and Pre-Processing Results*

The initial stage of the research involved collecting online news data related to road infrastructure, which then underwent pre-processing, including case folding, data cleaning, and tokenization. At this stage, word integrity was maintained to ensure that the lexicon-based tagging process could accurately identify sentiment markers without losing semantic context.

From a total of 859 documents collected, automatic tagging was performed using a custom dictionary tailored to the infrastructure context. This process produced a sentiment distribution that reflected real-world conditions, where news reports about road damage were significantly dominated by complaints and criticism. Table 2 shows examples of words and weights used in the custom lexicon dictionary.

Table 2. Example of custom lexicon dictionary weights

Category	Example Word (Indonesian)	English Translation	Weight
Extremely negative	*hancur, mati, longsor, roboh, biadab*	destroyed, dead, landslide, collapsed, barbaric	-5

Category	Example Word (Indonesian)	English Translation	Weight
Severe negative	*terbengkalai, kecelakaan, darurat, genangan*	abandoned, accident, emergency, puddle	-4
Moderately negative	*berlumpur, lubang, lumpur, macet, retak*	muddy, potholes, mud, stuck, cracked	-3
Moderately positive	*perbaikan, dukungan, bangun, benah, bantu*	repair, support, build, fix, help	+3
Strong positive	*mulus, mantap, apresiasi, nyaman, selesai*	smooth, steady, appreciation, comfortable, completed	+5

Based on the tagging results, the initial sentiment distribution was obtained as shown in Figure 2.

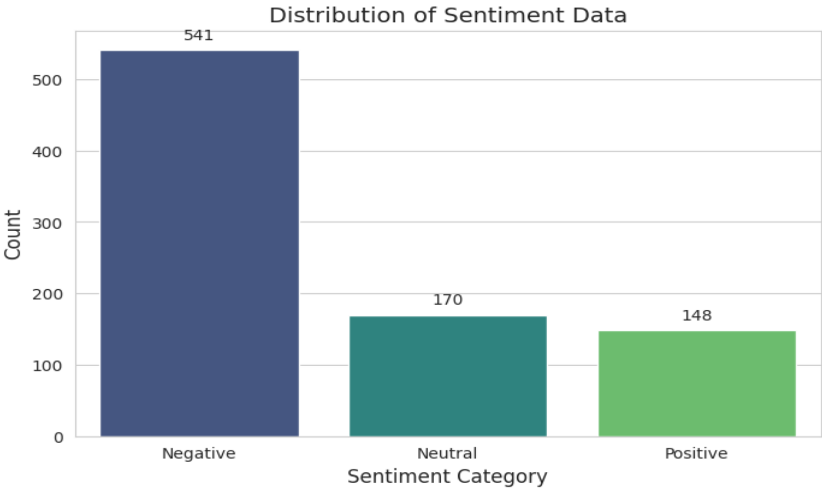

Figure 2. Distribution of sentiment data from lexicon labeling ($N = 859$)

Figure 2 illustrates the unbalanced nature of the dataset, where negative sentiment (63.0%) is far more dominant than neutral sentiment (19.8%) and positive sentiment (17.2%). This distribution shows that public discourse on road infrastructure in North Sumatra is largely driven by negative experiences and complaints.

4.2 Word Cloud Visualization

To identify the most frequently occurring topics and keywords in the dataset, a Word Cloud visualization was generated. In response to the need for clear thematic mapping, a special filtering layer was applied at this stage. Although the main pre-processing pipeline (as described in Section 3.3) retains all words to maintain the lexical integrity of the Naive Bayes model, this visualization uses a custom stopword list to filter out non-thematic functional words in Indonesian, such as 'yang' (that), 'di' (in), 'dan' (and), and 'untuk' (for). This refinement ensures that the visualization highlights substantial issues such as infrastructure conditions rather than grammatical noise.

Figure 3. Word cloud of the most frequently occurring terms

The visualization in Figure 3 shows that terms such as 'warga' (residents), 'jalan' (roads), 'rusak' (damaged), 'perbaikan' (repairs) are very dominant. In addition, location-specific terms such as "Medan" and 'Sumatera Utara' (North Sumatra) appear prominently, confirming the high relevance of this dataset to the research area. The dominance of these substantive keywords proves the effectiveness of selective filtering, which successfully reveals important thematic issues that are of public concern regarding road infrastructure.

4.3 Data Division and SMOTE Application

Prior to model training, the dataset consisting of 859 documents was divided into training and testing sets using stratified random sampling with a ratio of 80:20. This process yielded 687 samples for the training set and 172 samples for the testing set. This separation was strictly enforced prior to SMOTE application to prevent data leakage, ensuring that model evaluation was performed on pure, authentic, and realistic data. Preliminary analysis of the training data revealed significant class imbalance. As seen in the initial distribution, negative sentiment dominated the dataset, while positive and neutral sentiment were minority classes. This imbalance risked causing model bias, where classification might favour the majority class (negative) and perform poorly on the minority classes. To address this, the Synthetic Minority Over-sampling (SMOTE) technique was applied exclusively to the training data. SMOTE synthesizes new samples for the minority classes based on the k-nearest neighbors algorithm, balancing the Positive and Neutral classes to match the number of Negative classes. The shift in distribution is described in Table 3 and visualized in Figure 4.

Table 3. Training Data Distribution Before and After SMOTE

Sentiment	Count (Before SMOTE)	Percentage (%)	Count (After SMOTE)	Percentage (%)
Negative	433	63.03	433	33.33
Neutral	136	19.80	433	33.33
Positive	118	17.17	433	33.33
Total	687	100.00	1,299	100.00

As shown in Table 3, SMOTE successfully increased the number of minority samples, resulting in a perfectly balanced training dataset with 1,299 documents. This balanced distribution, further illustrated in Figure 4, ensures that the Naive Bayes algorithm can learn feature representations from all classes with equal weight, thereby improving the model's ability to distinguish between different sentiment nuances.

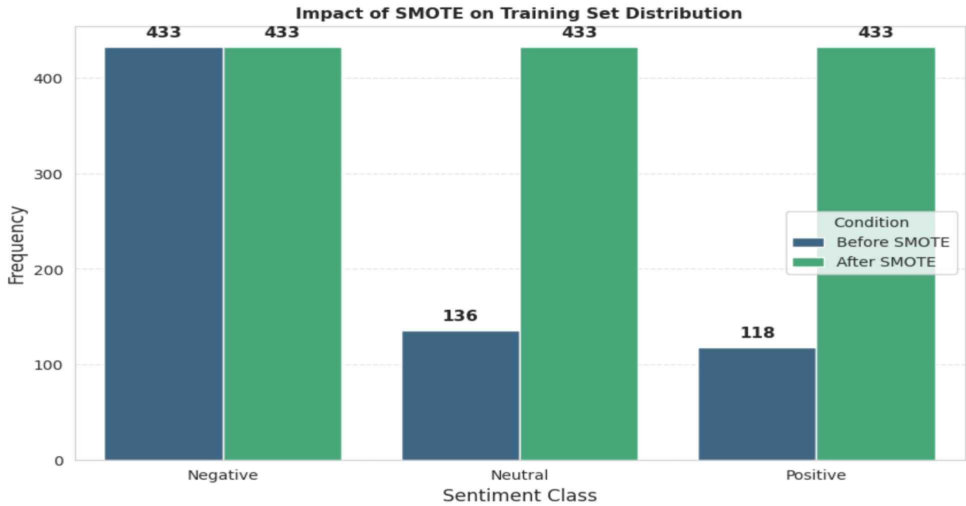

Figure 4. Comparison of training data class distribution before and after SMOTE

Figure 4 shows the successful transition from an unbalanced distribution to an even distribution through the oversampling process. By achieving this balance, the Multinomial Naive Bayes algorithm is given a fair basis for learning and distinguishing feature patterns across all three sentiment categories. This step is crucial to prevent the model from developing a bias towards the majority class, ensuring that the classification process remains sensitive to the nuances of positive and neutral sentiments that were previously underrepresented.

4.4 *Evaluation of Naïve Bayes Model Performance*

The testing phase was conducted using a test data set consisting of 172 original documents. These documents represent pure and realistic data, as they were separated from the training data prior to the application of SMOTE. Based on the evaluation, the Multinomial Naive Bayes model achieved a global accuracy of 70.93%. Detailed performance metrics are presented in Table 4.

Table 4. Classification Report

Sentiment Class	Precision	Recall	F1-Score	Support
Negative	0.86	0.74	0.80	108
Neutral	0.58	0.62	0.60	34
Positive	0.49	0.70	0.58	30
accuracy	-	-	0.71	172
macro avg	0.64	0.69	0.66	172
weighted avg	0.74	0.71	0.72	172

To analyze specific prediction errors between the original labels and the model predictions, a Confusion Matrix is used, as shown in Figure 5.

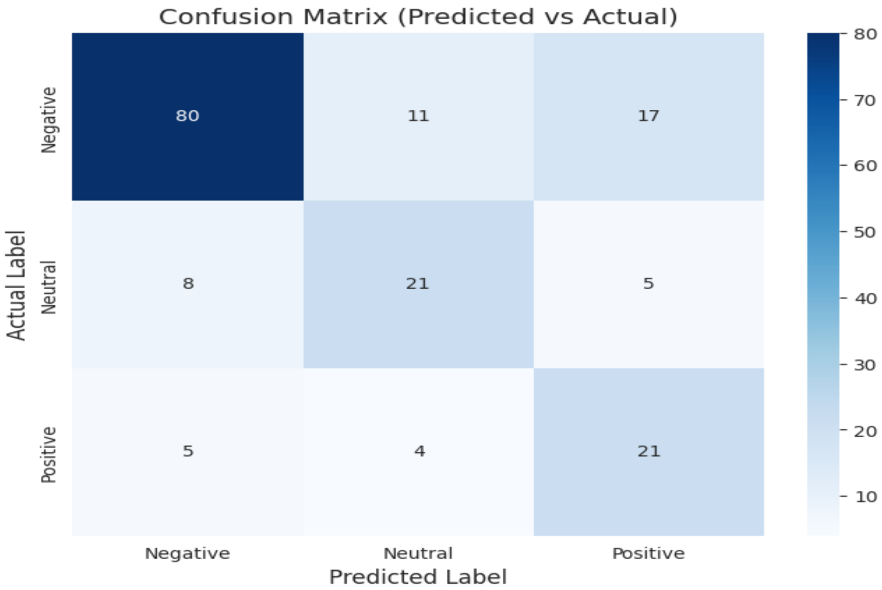

Figure 5. Confusion matrix model Multinomial Naïve Bayes on test data

Figure 5 shows that the model has a high success rate in predicting the Negative class, as most correct predictions are concentrated along the main diagonal. However, some classification errors occur, especially in the Neutral class, which is sometimes confused with other categories. In addition, the overall distribution of sentiment predictions on the test data is shown in Figure 6.

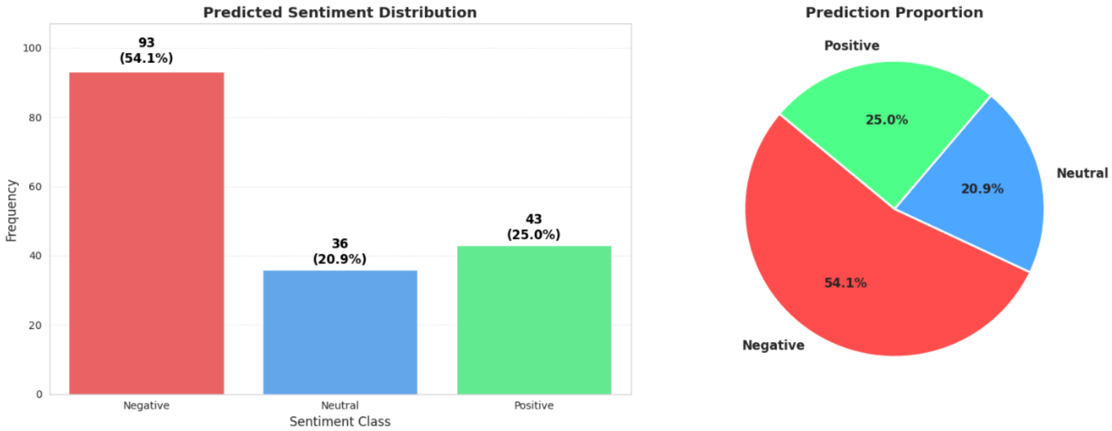

Figure 6. Distribution of sentiment in model predictions ($N = 172$)

Based on Figure 6, negative sentiment dominated the prediction results with 93 data points (54.1%), followed by positive sentiment with 43 data points (25.0%), and neutral sentiment with 36 data points (20.9%). These results are in line with the initial hypothesis that road infrastructure problems tend to trigger mostly negative public responses.

4.5 Discussion

Based on the experimental results, this study successfully developed a sentiment classification model using the Weak Supervision approach with the Multinomial Naïve Bayes algorithm, which was optimized through SMOTE. The following is an in-depth analysis of the model's performance and its practical implications:

1. Effectiveness of Complaint Detection as a Monitoring Priority

The main objective of this infrastructure monitoring system is to accurately capture public complaints. The evaluation results show that this model performs best in the Negative Sentiment class, with a Precision of 0.86, Recall of 0.74, and F1 Score of 0.80. High precision indicates that when the model identifies an article as a complaint, there is an 86% probability that the classification is correct.

This is very important for the Inspectorate to minimize false alarms and ensure that resources are directed to valid reports. In addition, a Recall of 0.74 indicates that the system successfully captures nearly three-quarters of all public complaints in the media, thereby effectively minimizing the risk of losing critical information about severe infrastructure damage.

2. Significant Impact of SMOTE on Positive Opinion Detection
   The main challenge in this study was significant data imbalance, where positive sentiment (appreciation for road improvements) was in the minority. The application of SMOTE to the training data proved transformative. This is evidenced by a Recall value of 0.70 for the Positive class, indicating that despite the scarcity of original data, the model successfully identified 70% of positive news correctly.

   Although the Precision value (0.49) for this class is relatively lower, indicating that the model sometimes misclassifies neutral news as positive, this is considered an acceptable compromise in the context of government. Strategically, it is more beneficial for the system to slightly “overzealous” the detection of performance appreciation than to completely fail to recognize the success of infrastructure projects.

3. Linguistic Challenges in Neutral Classification
   The analysis shows that the Neutral class produces the lowest performance, with an F1-score of 0.60. This is due to the specific linguistic characteristics of news reports. Neutral sentences often consist of technical and descriptive data, such as “panjang jalan 5 km” (5 km road length), or “anggaran 2 miliar” (2 billion budget) that overlap with positive and negative contexts.

   In addition, the inherent ambiguity of news language, such as the use of future tense verbs like “direncanakan” (planned) or “akan” (will), poses a challenge for the Naive Bayes algorithm, which relies on word frequency and probability. However, since the main objective of this study is to distinguish between extreme polarities (complaints vs. appreciation), the moderate performance of the Neutral class does not diminish the overall usefulness of the model.

4. Validity of Weak Supervision Approach
   These results prove the effectiveness of the Hybrid/Weak Supervision framework. Even though it was trained using labels automatically generated by a lexicon dictionary rather than manually tagged by humans, the Naive Bayes model achieved an overall accuracy of 70.93% on previously unseen test data.

This proves that the algorithm does not simply "memorize" the lexicon, but is capable of generalizing patterns by learning contextual word associations. This finding reinforces previous research by Syah et al. (2023), which suggests that combining lexicon-based tagging with machine learning is a viable solution to overcome resource limitations in manual data tagging.

5. CONCLUSION

Based on the results of research and testing conducted on the analysis of public opinion sentiment regarding road damage in North Sumatra, several conclusions can be drawn as follows:

1. This study successfully proved that the Weak Supervision approach (a combination of Lexicon-based automatic labeling and Naive Bayes classification) is an effective solution to overcome the obstacle of a lack of labeled data. The model built was able to produce an accuracy of 70.93% on previously unseen test data.
2. The model showed superior performance in detecting negative sentiment (public complaints) with a Precision value of 0.86 or 86%. This indicates that the system has a high level of confidence; when the system detects a news item as a "complaint," it is highly likely that the prediction is correct.
3. The application of the SMOTE technique to the training data proved crucial in increasing the model's sensitivity to positive sentiment (appreciation/improvement). This is indicated by a positive class Recall value of 0.70 or 70%, which means that the model is able to capture the majority of information about road improvements even though the amount of original data is very small compared to complaints.
4. The results of the analysis show that the dominance of public sentiment in online media related to road infrastructure in North Sumatra is Negative (54.1%), followed by Positive (25.0%) and Neutral (20.9%). This confirms the high level of public attention and criticism regarding the condition of infrastructure.

Based on the research that has been conducted, the following recommendations can be formulated for further development:

1. It is necessary to add to and curate the Lexicon dictionary in greater depth, particularly to address slang words or the regional language of North Sumatra, so that neutral sentiment detection is more accurate.
2. It is recommended to compare the performance of Naive Bayes with other algorithms such as Support Vector Machine (SVM) or Long Short-Term Memory (LSTM) on the same dataset to see the potential for accuracy improvement.

REFERENCES

Afuan, L., Khanza, M., & Zahira Hasyati, A. (2025). Enhancing Sentiment Analysis of the 2024 Indonesian Presidential Inauguration on X Using Smote-Optimized Naive Bayes Classifier. *Jurnal Teknik Informatika (JUTIF)*, *6*(1), 325–333. https://doi.org/10.52436/1.jutif.2025.6.1.4290

Albab, M. U., P., Y. K., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, *20*(2), 1–12. https://doi.org/10.26623/transformatika.v20i2.5374

Amal, I., & Jayanta. (2023). Perbandingan Pelabelan Otomatis Dan Manual Untuk Analisis Sentimen Terhadap Kenaikan Harga BBM Pertamina Pada Twitter Menggunakan Algoritma Support

Vector Machine. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, *4*(2), 473–487. https://conference.upnvj.ac.id/index.php/senamika/article/view/2562

Badriyah, Chamidy, T., & Suhartono. (2025). Application of SMOTE in Sentiment Analysis of MyXL User Reviews on Google Play Store. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *10*(1), 74–86. https://doi.org/10.14421/jiska.2025.10.1.74-86

Budaya, I. G. B. A., & Suniantara, I. K. P. (2024). Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(3), 1077–1086. https://doi.org/10.57152/malcom.v4i3.1459

Fatharani, D., & Syahrul, E. (2025). HYBRID SENTIMENT ANALYSIS OF MAXIM APP USERS USING SUPPORT VECTOR MACHINE AND LEXICON- BASED APPROACH. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, *13*(3), 1930–1937. https://doi.org/http://dx.doi.org/10.23960/jitet.v13i3S1.8148 HYBRID

Fikri, M. R., Handayanto, R. T., & Irwan, D. (2022). Web Scraping Situs Berita Menggunakan Bahasa Pemograman Python. *Journal of Students 'Research in Computer Science*, *3*(1), 123–136.

Firda, H., Putra, P., Oktadini, N. R., & Sevtiyuni, P. E. (2025). Perbandingan Pelabelan Rating-based dan Inset Lexicon-based dalam Analisis Sentimen Menggunakan SVM (Studi Kasus: Ulasan Aplikasi GoBiz di Google Play Store). *Sistemasi: Jurnal Sistem Informasi*, *14*(2), 516–528. http://sistemasi.ftik.unisi.ac.id

Gibran, M. K., Rifki, M. I., Hasugian, A. H., Siahaan, A. T. A. A., Sahputra, A., & Ong, R. (2024). Sentiment Analysis of Platform X Users on Starlink Using Naive Bayes. *Instal : Jurnal Komputer*, *16*(3), 210–220. https://doi.org/https://doi.org/10.54209/jurnalinstall.v16i03.240

Hafiz, Y. A., & Sudarmilah, E. (2023). IMPLEMENTASI WEB SCRAPING PADA PORTAL BERITA ONLINE. *Inisiasi*, *12*(1), 55–60. https://doi.org/10.59344/inisiasi.v12i1.120

Haikal, B. F., Hasibuan, M. S., & Rifki, M. I. (2025). Perbandingan Algoritma Naïve Bayes Dan Dempster Shafer Untuk Diagnosis Penyakit ISPA. *JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING*, *04*(3), 147–157. https://doi.org/Doi: https://doi.org/10.55537/cosie.v4i3.1161

Hardiyanti, M., & Fajarlestari, M. K. (2025). Optimasi Analisis Sentimen Komentar Penonton Wayang Digital dengan SMOTE dan Algoritma Naïve Bayes. *SISTEMASI*, *14*(3), 1–11. https://doi.org/https://doi.org/10.32520/stmsi.v14i3.5002

Hermawan, M. A., Faqih, A., & Dwilestari, G. (2025). Implementasi Akurasi Model Naive Bayes Menggunakan Smote Dalam Analisis Sentimen Pengguna Aplikasi Brimo. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(1), 855–862. https://doi.org/10.23960/jitet.v13i1.5748

Jelita, H. P., Saad, M. I., & Wahyuni. (2025). Penerapan Algoritma Naïve Bayes Dalam Analisis Sentimen Masyarakat Terhadap STMIK Widya Cipta Dharma. *Bulletin of Information Technology (BIT)*, *6*(2), 148–160. https://doi.org/10.47065/bit.v5i2.2029

Kurnia, Purnamasari, I., & Saputra, D. D. (2023). Analisis Sentimen Dengan Metode Naïve Bayes, SMOTE Dan Adaboost Pada Twitter Bank BTN. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *7*(2), 235–242. https://doi.org/10.35870/jtik.v7i3.707

Kurniawan, Y. I., Sidiq, R. H., Widadi, A. D., Yubiharto, A. R. A., Gibran, A. K., Aditama, M. R., & Laksono, F. A. T. (2025). Sentiment Analysis of Netflix App Reviews on Google Play Store using the Naive Bayes. *JITAR: Journal of Information Technology and Applications Research*, *1*(2), 01–17. https://doi.org/10.63956/jitar.v1i2.31

Manoppo, M. R., Kolang, I. C., Nur Fiat, D. N., Mawara, R. M. C., Sumarno, A. D. P., Yusupa, A., & Tarigan, V. (2025). ANALISIS SENTIMEN PUBLIK DI MEDIA SOSIAL TERHADAP KENAIKAN PPN 12% DI INDONESIA MENGGUNAKAN INDOBERT. *Jurnal Kecerdasan Buatan Dan Teknologi Informasi*, *4*(2), 152–163. https://doi.org/10.69916/jkbti.v4i2.322

Oktavia, A., Apriani, W., Putri, N., Rina, I., Purnata, H., Martono, S. M., Hanam, E. S., & Qomarotun Nurlaila, S. T. M. T. (2025). *Matematika Teknik*. PT Penerbit Qriset Indonesia.

Putra, A. H., & Salam, A. (2025). A Comparative Performance of SMOTE, ADASYN and Random Oversampling in Machine Learning Models on Prostate Cancer Dataset. *Journal of Applied Informatics and Computing (JAIC)*, *9*(3), 603–610. https://doi.org/10.30871/jaic.v9i3.9308

Rahmatullah, M. R. F., Andono, P. N., & Soeleman, M. A. (2025). Improving Random Forest Performance for Sentiment Analysis on Unbalanced Data Using SMOTE and BoW Integration: PLN Mobile Application Case Study. *Scientific Journal of Informatics*, *12*(1), 1–10.

Rambe, M. R. A., Zufria, I., & Rifki, M. I. (2025). Analisis Sentimen Masyarakat pada Platform Media Sosial X (Twitter) terhadap Pelantikan Kabinet Merah Putih Menggunakan Bernoulli Naïve Bayes. *Journal Of Information System, Computer Science And Information Technology |*, *6*(1), 83–102. file:///D:/KULIAH SEMSTER 5/ref/ref/Analisis Sentimen Masyarakat pada Platform Media Sosial X (Twitter) terhadap Pelantikan Kabinet Merah Putih Menggunakan Bernoulli Naïve Bayes.pdf

Ratnaswari, S., Wibowo, N. C., & Kartika, Y. S. D. (2025). ANALISIS SENTIMEN MENGGUNAKAN METODE LEXICON-BASED DAN SUPPORT VECTOR MACHINE PADA PRESIDEN DAN WAKIL PRESIDEN INDONESIA PERIODE 2024–2029. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(1), 362–368. https://doi.org/http://dx.doi.org/10.23960/jitet.v13i1.5604 ANALISIS

Rohmatun, L., & Baita, A. (2025). Machine Learning-Based Sentiment Analysis on Twitter (X): A Case Study of the "Kabur Aja Dulu" Issue Using SVM. *Journal of Applied Informatics and Computing (JAIC)*, *9*(4), 1972–1983. https://doi.org/10.30871/jaic.v9i4.9991

Sabrina, D., Sabilla, A. D., & Azizah, N. (2025). Kombinasi Vader Lexicon Dan Support Vector Machine Untuk Klasifikasi Sentimen Komentar Aplikasi Blu Bca. *INSERT: Information System and Emerging Technology Journal*, *6*(1), 22–33. https://ejournal.undiksha.ac.id/index.php/insert/article/view/86240

Safira, A., & Hasan, F. N. (2023). ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATTER MENGGUNAKAN METEODE NAIVE BAYES CLASSIFIER. *Jurnal Sistem Informasi*, *5*(January), 59–70. https://doi.org/10.31849/zn.v5i1.12856

Saputra, C., Irsyad, H., & Rahman, A. (2025). Applied Information Technology and Computer Science Pemrosesan Query dan Pemeringkatan Judul Berita Terkait Gu- bernur Jawa Barat Menggunakan TF-IDF dan Cosine Similarity. *Applied Information Technology and Computer Science*, *4*(1), 25–32. https://doi.org/https://doi.org/10.58466/aicoms.v3i2

Syah, I. R., Hoiriyah, & Walid, M. (2023). ANALISIS SENTIMEN PENGGUNAMEDIA SOSIAL TERHADAP APLIKASI M-HEALTH PEDULI LINDUNGI DENGAn METODE LEXICONBASED DAN NAÏVE BAYES. *Indonesian Journal of Business Intelligence Volume*, *6*(1), 43–54. https://doi.org/http://dx.doi.org/10.21927/ijubi.v6i1.3275

Ulgasesa, R., Negara, A. B. P., & Tursina, T. (2022). Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, *10*(3), 286. https://doi.org/10.26418/justin.v10i3.53880

Ulhaq, A. L. D., & Suprayogi. (2025). Comparing Machine Learning Models for Sentiment Analysis of Textual Reviews. *Journal of Applied Informatics and Computing (JAIC)*, *9*(6), 3642–3647. https://doi.org/https://doi.org/10.30871/jaic.v9i6.11239