Journal of Information Systems and Informatics
Vol. 7, No. 3, September 2025 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v7i3.1234 Published By DRPM-UBD

Public Opinion Sentiment Analysis Towards Government Budget Efficiency Policy on Twitter (X) Using the Naïve Bayes Classifier Algorithm

Rizki Ramadani Ritonga[1], Sriani[2]

[1,2]Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: [1]rizkiramadanirtg@gmail.com, [2]sriani@uinsu.ac.id

Abstract

The government's budget efficiency program, mandated through Presidential Instruction No. 1 of 2025, represents a strategic initiative to maximize the effectiveness of national (APBN) and regional (APBD) spending while minimizing waste. This policy has triggered diverse public responses, particularly on Twitter (X), which serves as one of the most widely used platforms in Indonesia for expressing opinions openly. This study investigates public sentiment toward the policy by applying the Multinomial Naïve Bayes Classifier algorithm. A total of 1,000 tweets were collected through crawling between January and March 2025 using the keywords "government budget efficiency" and "APBN savings." The analytical process involved several steps, including text preprocessing, automatic labeling with the Indonesian InSet lexicon-based dictionary, and TF-IDF weighting. The dataset was divided into 80% training data and 20% testing data. Labeling results identified 703 positive tweets and 297 negative tweets. Model performance evaluation using a confusion matrix achieved an accuracy of 77%, precision of 57.14%, recall of 82.76%, and an F1-score of 67.6%. Although this study focuses only on binary sentiment classification (positive and negative), the findings demonstrate that the proposed method is sufficiently effective in classifying public sentiment related to the government's budget efficiency policy. The results also provide significant insights into public opinion and can serve as a reference for policymakers as well as for future research on social media-based sentiment analysis.

Keywords: analysis, budget efficiency, inset lexicon, naïve bayes, social media, twitter.

1. INTRODUCTION

The government's budget efficiency policy represents a critical initiative designed to optimize public spending, ensuring that state resources are allocated effectively, fairly, and without waste. As outlined in Presidential Instruction Number 1 of 2025, the policy mandates a substantial reduction in both the state budget (APBN) and regional budgets (APBD) by a total of Rp306.7 trillion. This includes cuts of Rp256.1 trillion in ministry and institutional spending, as well as Rp50.6 trillion in regional transfers [2]. The goal of these budget cuts is to improve public financial governance, but implementing such significant reductions presents practical

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

challenges and raises concerns regarding their impact on public welfare [3], [4]. While these efforts are designed to strengthen fiscal health, there is a gap in understanding the public's reaction to these changes, particularly in how people perceive and respond to such policy shifts.

Social media platforms, particularly Twitter (X), have become crucial spaces for public discourse, where citizens freely express their opinions, concerns, and support for government policies. The vast amount of data generated on Twitter makes it an ideal platform for analyzing public sentiment, providing valuable insights into how people feel about various government initiatives [5]–[7]. Sentiment analysis, a process that identifies and classifies emotions in text, plays a key role in extracting meaningful insights from this data. It enables researchers to categorize public attitudes—whether positive, negative, or neutral—toward specific issues. While sentiment analysis has been successfully applied in various domains, there remains a significant gap in research focused on analyzing public sentiment regarding fiscal policies, especially those related to government budget efficiency.

The Naïve Bayes algorithm has demonstrated high performance in text classification tasks, offering competitive accuracy compared to other methods [9], [10]. Previous studies have successfully utilized Naïve Bayes for sentiment analysis on social media, such as in analyzing public reactions to the 2023 Hajj cost increase [9] and public opinion on the 2024 presidential election [10]. However, there is limited research on applying this algorithm specifically to analyze public sentiment regarding fiscal policies, such as the government's budget efficiency measures. Given this gap, there is a clear need for a study that explores how sentiment analysis can provide insights into public opinion on fiscal policy, thereby guiding government decisions.

This study aims to fill this gap by applying the Multinomial Naïve Bayes algorithm, combined with the TF-IDF feature representation, to analyze public sentiment regarding the government's budget efficiency policy as expressed on Twitter. The primary goal is to develop a classification model capable of accurately capturing public sentiment, which could then be used to provide data-driven feedback to the government on its budget efficiency measures. This research will help bridge the existing gap by offering an objective and quantifiable picture of public sentiment, thereby enabling more informed policy decisions.

To achieve this, the study is guided by the following research questions:

1. How can the Multinomial Naïve Bayes Classifier algorithm be applied to analyze public sentiment (positive and negative) toward the government's budget efficiency policy using Twitter data?

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

2. To what extent do the TF-IDF weighting method and the InSet Lexicon-based labeling improve the accuracy of the Multinomial Naïve Bayes Classifier in classifying public sentiment toward the policy?
3. What are the accuracy, precision, recall, and F1-score values produced by the Multinomial Naïve Bayes Classifier algorithm in classifying public sentiment toward the government's budget efficiency policy?

In line with these research questions, the study seeks to achieve the following objectives:

1. To evaluate the effectiveness of the Multinomial Naïve Bayes Classifier algorithm in analyzing public sentiment (positive and negative) toward the government's budget efficiency policy using Twitter data.
2. To assess the impact of the TF-IDF weighting method and the InSet Lexicon-based labeling on enhancing the accuracy of the Multinomial Naïve Bayes Classifier in classifying public sentiment on the policy.
3. To measure the accuracy, precision, recall, and F1-score achieved by the Multinomial Naïve Bayes Classifier in classifying public sentiment regarding the government's budget efficiency policy.

By addressing these questions and objectives, the study aims to contribute valuable insights into the public's views on fiscal policy, providing the government with data-driven input to improve its budget efficiency strategies and enhance public financial governance.

2. METHODS

This study applies a quantitative approach with text data from Twitter (X) regarding public opinion on the government's budget efficiency policy. Data were collected through a crawling process using Python and the Tweepy library to access the official Twitter API. The keywords used were "*efisiensi anggaran pemerintah*" and "*penghematan APBN*" during January–March 2025. A total of 1,000 tweets were retrieved for analysis. The research uses quantitative methods with text data from Twitter(X), includes stages preprocessing, word weighting with TF-IDF, as well as evaluation of accuracy, precision, recall, and F1-score through Confusion Matrix. Figure 1 the flow of is research.

2.1. Identification of problems

The researchers have identified the main problem that forms the basis of this research. The main focus of this research is the implementation of the algorithm Naïve Bayes Classifier in the process and processing and classifying public opinion based on the sentiments conveyed through Twitter(X) This problem was identified

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

based on the urgency to understand how public perception can be used as material for evaluating the effectiveness of policies implemented by the government.

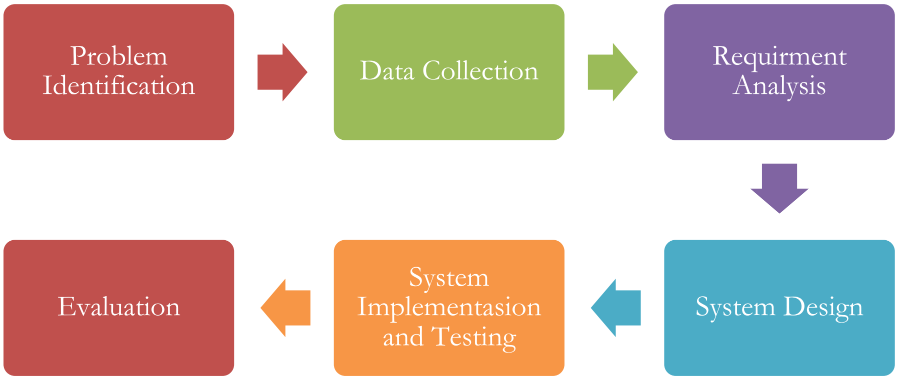

Figure 1. Research Framework

2.2. Data Collection

Research data was obtained throughcrawlingagainst 1,000 tweetrelated to the government's budget efficiency policy. The data is then processed through stagespreprocessing (cleansing, case folding, tokenizing, stopword removal, normalization, andstemming), weighted using TF-IDF, and labeled with a combination of manual andlexicon-based. Furthermore, algoritma Naïve Bayes Classifierimplemented to classify public opinion into positive and negative categories, with the process flow visualized throughflowchart system.

2.3. System Planning

The system design for implementing the Naïve Bayes Classifier (NBC) algorithm in sentiment analysis is an intricate and structured process, encompassing several key stages from data preprocessing to model evaluation. This design ensures that the sentiment classification of public opinions, specifically on government budget efficiency policies, using Twitter data is accurate, efficient, and reliable. The process is clearly visualized in Figure 2 and Figure 3, which guide the entire sentiment analysis workflow, helping to break down the steps involved in transforming raw, unstructured text into meaningful insights.

1) Text Preprocessing in Sentiment Analysis

The first critical step in the system design is text preprocessing, which plays a vital role in preparing the raw data for effective analysis. In the context of sentiment analysis, text preprocessing ensures that the data is cleaned, organized, and adjusted

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

to eliminate errors, noise, and inconsistencies, which could otherwise compromise the accuracy of the analysis [10]. This step involves several sub-processes, starting with data cleaning, where unnecessary characters such as numbers, symbols, hashtags, and emoticons are removed. This ensures that only meaningful content is considered for further analysis [18]. Next, case folding is applied to convert all text to lowercase, which reduces the variability caused by uppercase and lowercase distinctions in the text [21].

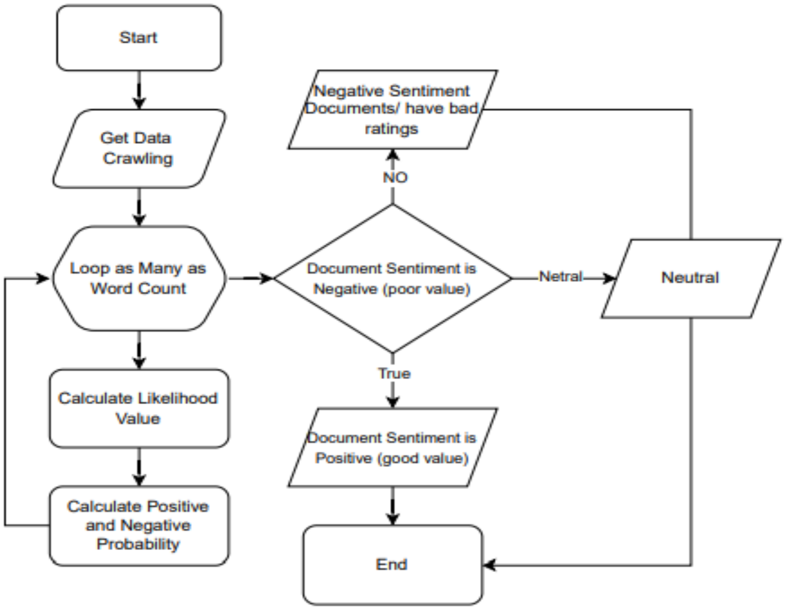

Figure 2. Naïve Bayes Classifier (NBC) Flowchart

Following case folding, the system applies tokenization, where the text is split into smaller units, known as tokens. These tokens are typically words or phrases, allowing the algorithm to process each word independently for sentiment analysis [10]. The next step is stopword removal, which eliminates common words (like "the," "is," and "and") that do not carry significant meaning in sentiment analysis [22]. Normalization follows, adjusting non-standard or inconsistent words to conform to recognized language standards, further ensuring the accuracy of the analysis [22]. Finally, stemming is performed to reduce words to their root form (e.g., "running" becomes "run"), making the text more uniform and ensuring that the algorithm processes variations of the same word consistently [22].

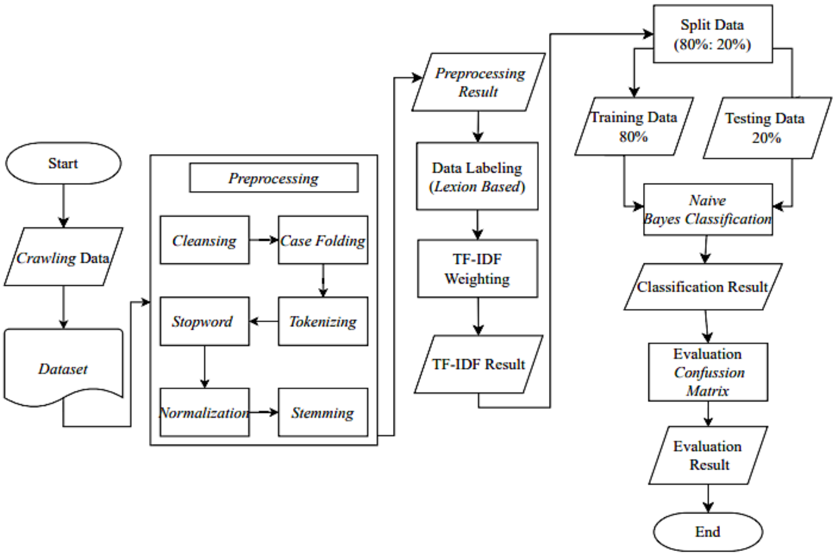

Figure 3. System Flowchart

Together, these two flowcharts demonstrate the systematic approach taken to apply the Naïve Bayes algorithm to sentiment analysis. Figure 2 focuses on the specific steps of the Naïve Bayes Classifier in classifying sentiment based on word probabilities, while Figure 3 outlines the overall system workflow, from data collection and preprocessing to training, classification, and evaluation. This structured design ensures that the sentiment analysis process is comprehensive, accurate, and data-driven, offering valuable insights into public sentiment regarding the government's budget efficiency policies.

2) Labeling Lexicon-Based in Sentiment Analysis

Once the data has been preprocessed, the next step is data labeling, where each document is assigned a sentiment category—positive, negative, or neutral. In sentiment analysis, the InSet Lexicon is often employed to assign sentiment scores to words within the text, ranging from -5 (strong negative sentiment) to +5 (strong positive sentiment) [24]. This lexicon-based labeling helps quantify the sentiment strength expressed in the document, making it easier for the Naïve Bayes algorithm to classify sentiment accurately.

3) Word Weighting TF-IDF

Another critical step in the sentiment analysis system is word weighting, which determines the significance of each word in the context of the entire corpus. The Term Frequency-Inverse Document Frequency (TF-IDF) method is employed

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

here, where Term Frequency (TF) measures the frequency of a word within a document, and Inverse Document Frequency (IDF) evaluates the importance of that word across all documents in the dataset. Words that appear frequently in a document but infrequently across the corpus are given more weight, ensuring that the sentiment analysis algorithm focuses on the most meaningful words in each document [16]. This combined approach ensures that the Naïve Bayes algorithm effectively handles the importance of different words, improving the classification accuracy.

4) Naïve Bayes Classifier Algorithm

At the heart of this system is the Naïve Bayes Classifier (NBC), a statistical algorithm based on probability theory. It is known for its simplicity, speed, and efficiency in handling text classification tasks, especially when the data has been represented in the form of word frequencies or weights, such as those obtained using TF-IDF. In this context, the Multinomial Naïve Bayes variant is most suitable because it handles text data where the features are represented by word frequencies [3]. The algorithm works by calculating the probability of a document belonging to a specific sentiment class (positive or negative) based on the occurrence of words within the document.

The Naïve Bayes algorithm can be represented mathematically using Bayes' Theorem, which calculates the probability of each sentiment class (positive or negative) based on the likelihood of individual words appearing in each class. The class with the highest probability is selected as the sentiment classification for the document. The Naïve Bayes algorithm is mathematically described by Equation 1, 2, and 3.

$$P(V_j) = \frac{|docs_j|}{|example|} \tag{1}$$

$$P(V_j) = \frac{n_k}{n + |vocabulary|} = +1 \tag{2}$$

$$V_{map} = \frac{armax}{v_{phenomenon}} \prod_{i=1}^{n} P(V_j)P(V_j) \tag{3}$$

5) Model Evaluation

After classifying the sentiment, the performance of the Naïve Bayes model must be evaluated using several metrics, such as accuracy, precision, recall, and F1-score. A confusion matrix is typically used for this evaluation, which compares the predicted sentiment classifications with the actual sentiment labels in the test dataset. The confusion matrix provides insight into the true positives (TP), true

Journal of Information Systems and Informatics
Vol. 7, No. 3, September 2025
p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

negatives (TN), false positives (FP), and false negatives (FN) of the classification. From the confusion matrix, the following Equation 4 to 7 for calculating.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}\times 100\% \tag{4}$$

$$\text{Precision}=\frac{TP}{TP+FP}\times 100\% \tag{5}$$

$$\text{Recall}=\frac{TP}{TP+FN}\times 100\% \tag{6}$$

$$\text{F1 Score}=2\times\frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}} \tag{7}$$

3. RESULTS AND DISCUSSION

3.1. Data Collection

Data was collected from social media X (formerly Twitter) through a processcrawling using a script Python. This process is carried out through Google Colabratory platformby utilizing the librarytweepy, which is used to access the official API ofplatformX via token (Application Programming Interface). This API token allows the retrieval of public data in the form oftweetdirectly to connect to X. Keywords used in the search includekeyword"government budget efficiency", "budget cuts", and "state budget savings", which are in line with the research focus and have successfully collected as many as 1,000tweetThe data is then saved in .csv format using the DataFrame function from the pandas library. Pandas is used to store the resulting data.crawlinginto .csv format. Before using the library, it must be installed first using the PIP command and the system command becausetweet-harvestbuilt using the Node.js environment. The processcrawlingdata on application X usinggoogle collab.


data

	full_text
0	BGN Pastikan Tak Ada Kebijakan Menu MBG Berupa...
1	Dilema Kebijakan Truk Zero ODOL https://t.co/A...
2	...Akibat pendangkalan di Pelabuhan Pulau Baai...
3	biarin aja dia ngomongin apapun itukan dia lag...
4	Dalam rangka memperingati Hari Jadi ke-241 Pek...
...	...
995	Bener sekali nih negara hebat itu kalo kita se...
996	Bro yuk kita buktikan kalau kerjasama itu kunc...
997	Bener banget bro bangsa tangguh itu kalo kita ...
998	@CherryCheolie95 Yup kebijakan terkait scam me...
999	Benar banget bro bangsa hebat itu kalo masyara...

1000 rows × 1 columns

Figure 4. Raw Dataset (Indonesia)

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

3.2. Preprocessing

Preprocessing Data processing is the stage for eliminating several problems during data processing where the format is inconsistent or not clean enough from all elements that can reduce the level of accuracy of the analysis results. The pre-processing stage, namely text pre-processing, is carried out in several stages, namely cleaning (remove unnecessary elements), case folding (make lower case), tokenizing (break down sentences), stopword removal (omitting common words), normalization (standardize words), and stemming (returning the word to its base form). The following are the stages in the process preprocessing

1) Text Cleaning

The Cleaning and Case Folding stage is carried out to remove irrelevant elements in the data, such as URLs, emoticons, mentions (@username), hashtags (#), punctuation marks, and numbers. In addition, all letters in the data are removed.tweet also converted to lowercase (lowercase) so that the analysis process does not differentiate between capital letters and lowercase letters which actually have the same meaning.

	full_text	text_clean
0	BGN Pastikan Tak Ada Kebijakan Menu MBG Berupa...	bgn pastikan tak ada kebijakan menu mbg berupa...
1	Dilema Kebijakan Truk Zero ODOL https://t.co/A...	dilema kebijakan truk zero odol
2	...Akibat pendangkalan di Pelabuhan Pulau Baai...	akibat pendangkalan di pelabuhan pulau baai se...
3	biarin aja dia ngomongin apapun itukan dia lag...	biarin aja dia ngomongin apapun itukan dia lag...
4	Dalam rangka memperingati Hari Jadi ke-241 Pek...	dalam rangka memperingati hari jadi ke pekanba...
5	@msaid_didu Tidak Ada Yg salah. kebijakan Meng...	didu tidak ada yg salah kebijakan mengelola sd...
6	Sebelumnya Gubernur Jakarta Pramono Anung mem...	sebelumnya gubernur jakarta pramono anung mem...
7	@Bambangmulyonoo Anak nya banteng..saat sesama...	anak nya bantengsaat sesama banteng si edy mi...
8	@Intel_Imut @barengwarga apa yg bisa di downgr...	imut apa yg bisa di downgrade dr pemerintah t...
9	Indonesia Maju milik kita! Mari kawal bersama ...	indonesia maju milik kita mari kawal bersama k...

Figure 5. Clean Data (Indonesia)

2) Tokenizing

At the level tokenizing, clean text splitting is performed (text_clean) into word units using the word_tokenize function from the NLTK library. The aim is to convert a sentence or phrase into a list of tokens, which are smaller pieces of words that can be analyzed separately.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

	full_text	text_clean	text_tokens
0	BGN Pastikan Tak Ada Kebijakan Menu MBG Berupa...	bgn pastikan tak ada kebijakan menu mbg berupa...	[bgn, pastikan, tak, ada, kebijakan, menu, mbg...
1	Dilema Kebijakan Truk Zero ODOL https://t.co/A...	dilema kebijakan truk zero odol	[dilema, kebijakan, truk, zero, odol]
2	...Akibat pendangkalan di Pelabuhan Pulau Baai...	akibat pendangkalan di pelabuhan pulau baai se...	[akibat, pendangkalan, di, pelabuhan, pulau, b...
3	biarin aja dia ngomongin apapun itukan dia lag...	biarin aja dia ngomongin apapun itukan dia lag...	[biarin, aja, dia, ngomongin, apapun, itukan, ...
4	Dalam rangka memperingati Hari Jadi ke-241 Pek...	dalam rangka memperingati hari jadi ke pekanba...	[dalam, rangka, memperingati, hari, jadi, ke, ...
...	...	...	...
995	Bener sekali nih negara hebat itu kalo kita se...	bener sekali nih negara hebat itu kalo kita se...	[bener, sekali, nih, negara, hebat, itu, kalo,...
996	Bro yuk kita buktikan kalau kerjasama itu kunc...	bro yuk kita buktikan kalau kerjasama itu kunc...	[bro, yuk, kita, buktikan, kalau, kerjasama, i...
997	Bener banget bro bangsa tangguh itu kalo kita ...	bener banget bro bangsa tangguh itu kalo kita ...	[bener, banget, bro, bangsa, tangguh, itu, kal...
998	@CherryCheolie95 Yup kebijakan terkait scam me...	yup kebijakan terkait scam memang rata diliha...	[yup, kebijakan, terkait, scam, memang, rata, ...
999	Benar banget bro bangsa hebat itu kalo masyara...	benar banget bro bangsa hebat itu kalo masyara...	[benar, banget, bro, bangsa, hebat, itu, kalo,...

1000 rows × 3 columns

Figure 6. Tokenizing (Indonesia)

3) Normalization

In the normalization stage, informal or slang words in public opinion tokens are converted into formal words using the kamusGaul.txt dictionary. This process equates words with the same meaning but written informally so they can be recognized by the classification algorithm. The kataGaul() function replaces slang words with their formal equivalents, then stores the results in the columntext_normalization.

text_tokens	text_normalisasi
[bgn, pastikan, tak, ada, kebijakan, menu, mbg...	[bgn, pastikan, tidak, ada, kebijakan, menu, m...
[dilema, kebijakan, truk, zero, odol]	[dilema, kebijakan, truk, zero, odol]
[akibat, pendangkalan, di, pelabuhan, pulau, b...	[akibat, pendangkalan, di, pelabuhan, pulau, b...
[biarin, aja, dia, ngomongin, apapun, itukan, ...	[biarkan, saja, dia, mengomongkan, apapun, itu...
[dalam, rangka, memperingati, hari, jadi, ke, ...	[dalam, rangka, memperingati, hari, jadi, ke, ...
...	...
[bener, sekali, nih, negara, hebat, itu, kalo,...	[benar, sekali, nih, negara, hebat, itu, kalo,...
[bro, yuk, kita, buktikan, kalau, kerjasama, i...	[saudara laki-laki, yuk, kita, buktikan, kalau...
[bener, banget, bro, bangsa, tangguh, itu, kal...	[benar, banget, saudara laki-laki, bangsa, tan...
[yup, kebijakan, terkait, scam, memang, rata, ...	[ya, kebijakan, terkait, scam, memang, rata, d...
[benar, banget, bro, bangsa, hebat, itu, kalo,...	[benar, banget, saudara laki-laki, bangsa, heb...

Figure 7. Normalization (Indonesia)

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

4) Stopwords

At the level stopword removal, common words that do not contribute to sentiment analysis, such as"And", "Which", or "That", removed. The stopword list was taken from the Indonesian NLTK library, then expanded with external files and slang words from social media. Each token in the text_normalization column was matched against the list, and any included tokens were removed. The results were stored in the text_stopwords column.

text_normalisasi	text_stopwords
[bgn, pastikan, tidak, ada, kebijakan, menu, m...	[bgn, pastikan, kebijakan, menu, mbak, bahan, ...
[dilema, kebijakan, truk, zero, odol]	[dilema, kebijakan, truk, zero, odol]
[akibat, pendangkalan, di, pelabuhan, pulau, b...	[akibat, pendangkalan, pelabuhan, pulau, baai,...
[biarkan, saja, dia, mengomongkan, apapun, itu...	[biarkan, mengomongkan, apapun, itukan, mengom...
[dalam, rangka, memperingati, hari, jadi, ke, ...	[rangka, memperingati, pekanbaru, walikota, ag...
...	...

Figure 8. Stopword (Indonesia)

5) Stemming

At the level stemming, words in columnstext_stopwordsreduced to its base form using an algorithm from the Sastrawi library. This process removes affixes so that words like"manage", "management", And “managed” become “manage”This step is important so that words with similar meanings are processed as one entity, thereby increasing classification accuracy.

text_stopwords	text_steamming
[bgn, pastikan, kebijakan, menu, mbak, bahan, ...	[bgn, pasti, bijak, menu, mbak, bahan, mentah,...
[dilema, kebijakan, truk, zero, odol]	[dilema, bijak, truk, zero, odol]
[akibat, pendangkalan, pelabuhan, pulau, baai,...	[akibat, dangkal, labuh, pulau, baai, warga, p...
[biarkan, mengomongkan, apapun, itukan, mengom...	[biar, omong, apa, itu, omong, uneguneg, resah...
[rangka, memperingati, pekanbaru, walikota, ag...	[rangka, ingat, pekanbaru, walikota, agung, nu...
...	...

Figure 9. Stemming Text (Indonesia)

Journal of Information Systems and Informatics
Vol. 7, No. 3, September 2025

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

6) Sentiment Labeling

At the stage of determining sentiment labels, a dictionary-based approach is used (lexicon-based) by comparing the words in the columndetokenizedagainst two lexicon lists, positive.tsv and negative.tsv. The determine_label() function counts the number of positive and negative words in each text, then assigns a positive label if the number of positive words is greater, negative if the number of negative words is greater, and None if the number is equal or not detected. The labeling results are stored in the Label column, which will be used as a classification target in the machine learning stage.

text_Steamming	detokenized	Label
[bgn, pasti, bijak, menu, mbak, bahan, mentah,...	bgn pasti bijak menu mbak bahan mentah tangsel	positif
[dilema, bijak, truk, zero, odol]	dilema bijak truk zero odol	negatif
[akibat, dangkal, labuh, pulau, baai, warga, p...	akibat dangkal labuh pulau baai warga pulau en...	positif
[biar, omong, apa, itu, mong, uneguneg, resah...	biar omong apa itu omong uneguneg resah kau de...	negatif
rangka, ingat, pekanbaru, walikota, agung, nu...	rangka ingat pekanbaru walikota agung nugroho ...	positif
...	...	...

Figure 10. Labeling Using Inset Lexicon

The number of positive and negative comments can be seen in Figure 4.13 below, which is a visualization of the number of positive and negative sentiment analyses, where the process is carried out using programming.python Of google colaboratory.

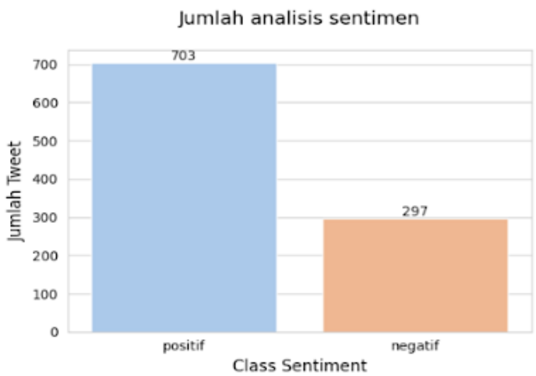

Figure 11. Visualization of the Number of Sentiment Analysis

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

3.3. Modelling

The results of the sentiment analysis on public opinions regarding the government's budget efficiency policy are derived from a comprehensive modeling process that involved several stages, including data preprocessing, feature extraction using TF-IDF, model training with Naïve Bayes Classifier, and performance evaluation. As seen in Figure 11, the sentiment distribution reveals that 703 tweets fall into the positive sentiment category, while 297 tweets are classified as negative sentiment. This distribution provides valuable insights into public opinion, showing that the majority of Twitter users (70.3%) express support for the government's budget efficiency policy, indicating general approval of the policy's objectives. However, the 29.7% of negative sentiment suggests that some segments of the public remain skeptical or dissatisfied, which could be due to concerns over its implementation or perceived negative impacts.

At the data division stage, 80% of the data was used for model training (X_train, y_train) and 20% for testing (X_test, y_test). This 80:20 split is a standard practice in machine learning, balancing the need for a robust model training set while providing enough testing data to evaluate the model's generalization capabilities. To ensure reproducibility, a random state parameter of 0 was applied. The training and testing data subsets were saved as agg_latih.csv and agg_uji.csv, respectively, as shown in Figure 12.

```
# Menampilkan jumlah data latih & uji
print(f"Jumlah Data Latih : {len(X_train)}")
print(f"Jumlah Data Uji   : {len(X_test)}\n")

Jumlah Data Latih : 800
Jumlah Data Uji   : 200
```

Figure 12. Split Data

Word weighting was performed using the TF-IDF method, which measures the importance of words in a document relative to the entire corpus. The process involved transforming the detokenized text into a vector format using CountVectorizer, followed by conversion into TF-IDF values through the Tf-idfTransformer. The weighted results were then arranged into a data frame containing documents, words, and their corresponding weights, as shown in Figure 13.

After preprocessing, labeling, and transforming the data, the dataset was split for training and testing. The model was trained on 80% of the data and tested on 20%, a common practice to ensure sufficient data for both learning and evaluation. The model was implemented using the Multinomial Naïve Bayes Classifier, a suitable choice for text data where word frequencies or weights, like those from TF-IDF, are used. The training results were applied to predict labels on the test data, and

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

the model's performance was evaluated using key metrics such as accuracy, precision, recall, and F1-score, as shown in Figure 14.

```
[29] from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer,TfidfTransformer
     cv = CountVectorizer()
     term_frequency = cv.fit_transform(dataset['detokenized'])
     tfidf_transformer = TfidfTransformer().fit(term_frequency)
     tfidf = tfidf_transformer.transform(term_frequency)
     vocabulary = cv.vocabulary_

[30] datatfidf = []
     count = 0
     for i, j in zip(*tfidf.nonzero()):
         datatfidf.append([i, list(vocabulary.keys())[list(vocabulary.values()).index(j)], tfidf[i, j]])
     colum_labels = ['Dokumen', 'Kata', 'TF-IDF']
     data_clean = pd.DataFrame(datatfidf, columns=colum_labels)

[31] dataset.to_csv('TF-IDFagg.csv')
```

Figure 13. Calculating TF-IDF

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report, accuracy_score

# Memisahkan data fitur dan label
x = df['detokenized']
y = df['label']

# Mengubah teks ke bentuk vektor
vectorizer = countvectorizer()
x_vectorized = vectorizer.fit_transform(x)

# Membagi data ke dalam training dan testing set
x_train, x_test, y_train, y_test = train_test_split(x_vectorized, y, test_size=0.2, random_state=42)

# Melatih model Naive Bayes
model = MultinomialNB()
model.fit(X_train, y_train)

# Memprediksi data test
y_pred = model.predict(x_test)

# Evaluasi model
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

Figure 14. Naive Bayes Model

The evaluation revealed that the model achieved an accuracy of 77%, indicating that the classifier is quite effective in distinguishing between positive and negative sentiment regarding the government's budget efficiency policy. Table 1 shows the confusion matrix used to assess the model's performance by comparing predicted labels with actual labels on the 200 test tweets.

Table 1. Confussion Matrix

Actual / Prediction	Positive (1)	Negative (0)	Total
Positive	TP = 48	FN = 10	58
Negative	FP = 36	TN = 106	142
Total	84	116	200

From the confusion matrix, we calculated the following performance metrics:

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = x100\%$$

$$= \frac{48+106}{48+10+36+106} = \frac{154}{200} = 0.77\%$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} = \text{x}100\%$$

$$= \frac{48}{48+36} = \frac{48}{84} = 0.57.14\ \%$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} = \text{x}100\%$$

$$= \frac{48}{48+10} = \frac{48}{58} = 0.8276\% = \frac{48}{48+10} = \frac{48}{58} = 0.8276\%$$

$$\text{F1 Score} = 2 \text{ x } \frac{\text{precision x recall}}{\text{precision+recall}}$$

$$= 2 \text{ x } \frac{0.5714 \text{ x } 0.8276}{0.5714+0.8276} = 0.676$$

The model's performance showed a relatively high recall of 82.76%, indicating that the Naïve Bayes Classifier was effective at identifying the relevant positive and negative sentiment tweets. However, the precision was lower at 57.14%, which suggests that the model tends to classify neutral or ambiguous tweets as either positive or negative. This is common in social media sentiment analysis, where the data is informal, short, and can be ambiguous.

The results also provide practical insights into public opinion dynamics. The dominance of positive sentiment (70.3%) suggests that the government's budget efficiency policy is generally well received by Twitter users. However, the significant proportion of negative sentiment (29.7%) highlights the presence of concerns or dissatisfaction among some segments of the public, which could be attributed to issues such as transparency, impact on public services, or the implementation process.

These findings are valuable for policymakers, as sentiment analysis offers a real-time, data-driven tool to assess public response to policy initiatives. The government can use these insights to refine communication strategies, address public concerns more effectively, and improve the transparency and accountability of the budget efficiency program. The classification results also underscore the importance of continued monitoring and adjustments based on public sentiment, ensuring that policies align with public expectations and needs. As shown in Figure

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

16, the classification report confirms the overall effectiveness of the Naïve Bayes Classifier, while also providing detailed performance metrics to guide future improvements in the analysis.

```
Confusion Matrix:
[[ 48  10]
 [ 36 106]]

Test Accuracy  : 77.00%
Test Recall    : 82.76%
Test precision : 57.14%
Test F1-Score  : 67.61%

Classification Report:
              precision    recall  f1-score   support

     negatif       0.57      0.83      0.68        58
     positif       0.91      0.75      0.82       142

    accuracy                           0.77       200
   macro avg       0.74      0.79      0.75       200
weighted avg       0.81      0.77      0.78       200
```

Figure 16. Classification Report

3.4. Discussion

The sentiment analysis conducted on public opinions regarding the government's budget efficiency policy provides important insights into how the policy is perceived by Twitter users. The majority of tweets (70.3%) expressed positive sentiment, indicating that most people support the government's efforts to optimize state spending. This suggests that the policy is largely well-received by the public, with many perceiving it as a positive step toward improving financial governance. However, a significant portion of the tweets (29.7%) expressed negative sentiment, indicating that some people are skeptical or dissatisfied with the policy's implementation. This divide highlights the complexity of public opinion and the need for government officials to consider both positive and negative reactions when refining the policy.

The preprocessing steps applied to the data were crucial in ensuring that the sentiment analysis model could effectively process the raw text data from social media. By cleaning the data, removing irrelevant elements like hashtags, mentions, and URLs, and converting the text to lowercase, we ensured that the analysis was not skewed by inconsistencies in the formatting of the tweets. Additionally, tokenization, stopword removal, and stemming helped break down the text into smaller, meaningful units, reducing the impact of common words and ensuring that the model focused on the core elements of sentiment. These steps are essential in transforming unstructured social media data into a format suitable for machine learning algorithms.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

The lexicon-based sentiment labeling approach used in this analysis was effective in classifying the sentiments of the tweets into positive or negative categories. By comparing the words in each tweet with predefined positive and negative word lists, the model was able to assign sentiment labels. This method provided a clear structure for classification, although it is worth noting that more complex sentiment expressions—such as sarcasm or mixed emotions—might not have been fully captured. The use of TF-IDF for word weighting further enhanced the model's ability to focus on important terms, helping improve the accuracy of sentiment classification.

The Multinomial Naïve Bayes Classifier was an appropriate choice for this task, as it works well with word frequency data and is efficient for text classification tasks. The model's performance, with an accuracy of 77%, indicates that it was able to correctly classify most tweets. However, the model's relatively low precision of 57.14% suggests that it often misclassified neutral or ambiguous tweets as either positive or negative. This is a common challenge in social media sentiment analysis, where tweets can be short, informal, and sometimes unclear in their tone. Despite this, the model's high recall of 82.76% shows that it was successful at identifying most of the relevant positive and negative sentiment tweets, which is useful for understanding overall public sentiment.

The findings from this analysis have significant implications for policymakers. While the majority of public sentiment is positive, the significant portion of negative sentiment calls attention to areas where the government may need to improve communication or address concerns. Sentiment analysis, as demonstrated in this study, can serve as a real-time, data-driven tool for monitoring public opinion and informing policy decisions. Policymakers can use these insights to refine communication strategies, increase transparency, and address any areas of dissatisfaction that may exist within the public. Future work could focus on improving the model's precision by incorporating more advanced techniques, such as n-grams or deep learning models like BERT, to better capture nuanced sentiment in social media data.

4. CONCLUSION

This study successfully analyzed public sentiment toward the government's budget efficiency policy using the Multinomial Naïve Bayes algorithm, coupled with TF-IDF weighting and InSet Lexicon labeling. The model achieved an accuracy of 77%, with a recall of 82.76% and a precision of 57.14%, demonstrating the algorithm's effectiveness in classifying public opinion on Twitter regarding the budget efficiency policy. These results highlight the ability of sentiment analysis to capture a broad range of public opinions, providing a solid foundation for

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

understanding public perception of government policies. From a practical perspective, the findings offer actionable insights for the government. With a majority of positive sentiment, there is clear support for the budget efficiency policy. However, the significant portion of negative sentiment signals areas of concern that need attention, such as improving transparency and addressing public dissatisfaction. These insights can help refine communication strategies, enhance public engagement, and ensure that the policy resonates with a broader segment of the population.

For future research, it is recommended to expand the dataset to include a more diverse range of opinions and a larger sample size, which could improve the robustness of the analysis. Additionally, conducting comparative evaluations with other classification algorithms, such as Support Vector Machines (SVM) or ensemble methods, could provide deeper insights into the strengths and weaknesses of different models in the context of sentiment analysis. Implementing multi-class sentiment analysis, which includes a neutral sentiment category, would offer a more nuanced understanding of public sentiment, accounting for mixed or indifferent opinions. Such improvements could enhance the accuracy and reliability of sentiment analysis in monitoring public opinion on social media in future studies.

REFERENCES

[1] S. M. Harahap and R. Kurniawan, "Analisis sentimen komentar Youtube terhadap food vlogger dengan menggunakan metode Naïve Bayes," *MEANS (Media Informasi Analisa dan Sistem)*, vol. 9, no. 1, pp. 87–96, 2024, doi: 10.54367/means.v9i1.3912.

[2] A. N. Hasanah and B. N. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Jasa Ojek Online Maxim Pada Google Play Dengan Metode Naïve Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 1, pp. 90–96, 2024, doi: 10.23960/jitet.v12i1.3628.

[3] H. Ashari, D. Arifianto, H. Azizah, and A. Faruq, "Perbandingan kinerja algoritma Multinomial Naïve Bayes, Multivariate Bernoulli dan Rocchio dalam klasifikasi konten berita hoax berbahasa Indonesia pada media sosial," *Proceedings/Repository Unmuh Jember*, pp. 1–12, 2020.

[4] W. M. P. Dhuhita and F. Zone, "Perbandingan Kinerja Algoritma Multinomial dan Bernoulli Naïve Bayes dalam Mengklasifikasikan Komentar Cyberbullying," *Komputika J. Sist. Komput.*, vol. 12, no. 2, pp. 109–117, 2023, doi: 10.34010/komputika.v12i2.9767.

[5] N. K. K. Ardana et al., "Perbandingan Metode KNN, Naive Bayes, dan Regresi Logistik Binomial dalam Pengklasifikasian Status Ekonomi

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Negara," *Jambura J. Math.*, vol. 5, no. 2, pp. 404–418, 2023, doi: 10.34312/jjom.v5i2.2110.

[6] I. Juventius, T. Gurning, P. P. Adikara, and R. S. Perdana, "Analisis Sentimen Dokumen Twitter menggunakan Metode Naïve Bayes dengan Seleksi Fitur GU Metric," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2169–2177, 2023.

[7] W. H. Prasetyo and A. P. Nugraheni, "Analisis Realisasi Anggaran Belanja Dalam Rangka Mengukur Efektivitas Dan Efisiensi Penyerapan Anggaran Belanja Dinas Kependudukan Dan Pencatatan Sipil (Disdukcapil) Kota Magelang Periode 2015 – 2019," *JIMAT (Jurnal Ilm. Mhs. Akuntansi) Univ. Pendidik. Ganesha*, vol. 11, no. 1, pp. 1–10, 2020.

[8] M. Z. Arifin and M. SH, *Tindak pidana korupsi kerugian ekonomi dan keuangan negara (Perspektif hukum dan praktik)*, PT Publica Indonesia Utama, 2024.

[9] V. Agustina and A. Herliana, "Analisis Sentimen Publik atas Kebijakan Efisiensi Anggaran 2025 dengan Text Mining dan Natural Language Processing," *Jurnal Media Informatika*, vol. 6, no. 3, pp. 2182–2194, 2025.

[10] I. Gunawan and M. Furqan, "Analisis Sentimen Terhadap Tindakan Pemerintah Indonesia Untuk Menampung Sementara Pengungsi Etnis Rohingya Menggunakan Naïve Bayes Classifier Sentiment Analysis of the Indonesian Government's Actions to Temporarily Accommodate Rohingya Ethnic Refugees," vol. 9, no. 2, pp. 737–749, 2024.

[11] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4902.

[12] R. Harahap and E. D. Kurniawan, "Analisis sentimen komentar terhadap kebijakan pemerintah mengenai Tabungan Perumahan Rakyat (TAPERA) pada aplikasi X menggunakan metode Naïve Bayes," *Jurnal Teknik Informatika Unika ST. Thomas*, vol. 9, no. 1, pp. 2657–1501, 2024.

[13] A. Salsabila Juwita, A. Rizky Kurniawan, A. Aryaputra Ashari, D. Tyan Putro, V. Nurcahyawati, and H. Artikel, "Implementasi Data Mining untuk Memprediksi Kesehatan Mental Mahasiswa menggunakan Algoritma Naïve Bayes," *KOMPUTEK J. Tek. Univ. Muhammadiyah Ponorogo*, vol. 8, no. 1, pp. 61–70, 2024.

[14] D. Puspita Sari, S. Shofia Hilabi, and A. Hananto, "Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama," *SMARTICS J.*, vol. 9, no. 1, pp. 14–19, 2023, doi: 10.21067/smartics.v9i1.8088.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[15] M. A. R. Sriani and Armansyah, "Analisis sentimen pada pengguna aplikasi Dana menggunakan algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1470–1476, 2024, doi: 10.36040/jati.v8i2.9041.
[16] C. A. Misrun, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen komentar Youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode Naïve Bayes Classifier," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 207–215, 2023, doi: 10.37859/coscitech.v4i1.479.
[17] A. Laia, "Analisis pemahaman mahasiswa terhadap prinsip transparansi dan akuntabilitas dalam manajemen keuangan publik: Studi kasus pada mahasiswa," *Jurnal Ekonomi Revolusioner*, vol. 7, no. 11, pp. 217–224, 2024.
[18] U. Wulandari and R. Paramitalaksmi, "Analisis efisiensi anggaran pasca penggabungan satuan kerja pemerintah: Investigasi pada UPT Balai Pelestarian Kebudayaan Wilayah X," *Jurnal Mutiara Akuntansi*, vol. 9, no. 2, pp. 50–61, 2024.
[19] M. Yasir and R. Suraji, "Perbandingan metode klasifikasi Naïve Bayes, Decision Tree, dan Random Forest terhadap analisis sentimen kenaikan biaya haji 2023 pada media sosial Youtube," *Jurnal Cahaya Mandalika (JCM)*, vol. 3, no. 2, pp. 180–192, 2023.
[20] F. Zamzami, R. Hidayat, and R. Fathonah, "Penerapan algoritma Naïve Bayes Classifier untuk analisis sentimen komentar Twitter proyek pembangunan IKN," *Faktor Exacta*, vol. 17, no. 1, pp. 47–57, 2024, doi: 10.30998/faktorexacta.v17i1.22265.
[21] S. A. Albahli, "Sentiment analysis of Twitter data using Naïve Bayes and logistic regression," *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 11, no. 7, pp. 573–578, 2020, doi: 10.14569/IJACSA.2020.0110770.
[22] Y. Zhang, M. Wang, and L. Jin, "Public opinion mining on social media: A case study of government policy sentiment analysis," *Gov. Inf. Q.*, vol. 40, no. 1, pp. 101–110, 2023, doi: 10.1016/j.giq.2022.101705.