Indonesian Journal of Data and Science

Volume 6 Issue 3
ISSN 2715-9930
https://doi.org/10.56705/ijodas.v6i3.349

Research Article

Public Response on X to the Revocation of Indonesia's 3-Kg LPG Retail Ban: A Support Vector Machine Study

Ni Nyoman Asti Sri Wahyuni[1]; I Gede Iwan Sudipa[2*]; Ni Nyoman Ayu J. Sastaparamitha [3]; Ayu Gede Willdahlia[4]; I Gusti Ayu Agung Mas Aristamy[5]

[1] Institut Bisnis dan Teknologi Informatika, Denpasar, Bali 80225, Indonesia, astisriwahyuni56@gmail.com
[2]Institut Bisnis dan Teknologi Informatika, Denpasar, Bali 80225, Indonesia, iwansudipa@instiki.ac.id
[3]Institut Bisnis dan Teknologi Informatika, Denpasar, Bali 80225, Indonesia, ajsasta@gmail.com
[4] Institut Bisnis dan Teknologi Informatika, Denpasar, Bali 80225, Indonesia, willdahlia@gmail.com
[5]Institut Bisnis dan Teknologi Informatika, Denpasar, Bali 80225, Indonesia, agungmas.aristamy@instiki.ac.id

Correspondence should be addressed to I Gede Iwan Sudipa: iwansudipa@instiki.ac.id

Received 02 August 2025; Accepted 12 November 2025; Published 31 December 2025

© Authors 2025. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes).
License: https://creativecommons.org/licenses/by-nc/4.0/ — Published by Indonesian Journal of Data and Science.

Abstract:
This study examines public responses on X to the 3-Kg LPG retail ban implemented on February 1, 2025, and revoked on February 4, 2025, which caused widespread shortages, long queues, and limited access, particularly for citizens living far from official distribution points. A total of 2,524 Indonesian-language tweets were collected via crawling and systematically processed through text cleaning, tokenization, normalization, stopwords removal, and stemming, followed by automatic labeling using the Indonesian Sentiment (InSet) Lexicon. After removing 229 neutral tweets, 1,405 tweets (61.2%) were classified as negative and 890 tweets (38.8%) as positive, with the study focusing on these two sentiment classes. Text features were extracted using TF-IDF, and classification was conducted using a linear-kernel Support Vector Machine (C = 0.1) with an 80:20 train-test split. The model achieved an overall accuracy of 84%, with precision, recall, and F1-score of 0.82, 0.94, and 0.88 for the negative class, and 0.87, 0.68, and 0.76 for the positive class. Results indicate that negative sentiment was dominated by criticism related to LPG shortages and insufficient policy communication, while positive sentiment reflected user relief over restored supply and hopes for fairer distribution in the future. These findings suggest that revoking the ban did not fully restore public perception, highlighting the necessity for more effective policy dissemination and stricter monitoring of 3-Kg LPG distribution. The study also emphasizes the importance of leveraging social media, particularly X, as a real-time source for monitoring public opinion and evaluating the effectiveness of energy distribution policies in Indonesia.

Keywords: Revocation; Sentiment Analysis; Public Response; 3-Kg LPG; Support Vector Machine

Dataset link: https://bit.ly/dataset-sentiment-lpg-3kg

1. Introduction

Liquefied Petroleum Gas (LPG) is one of the main energy sources for the majority of households in Indonesia, especially for cooking purposes. Data from the Central Statistics Agency shows that more than 88% of households use 3-Kg LPG subsidized by the government as their primary fuel. The high dependency on subsidized LPG makes any changes in its distribution policy highly sensitive and potentially prone to public unrest.

This sensitivity was evident when the Ministry of Energy and Mineral Resources implemented a policy prohibiting retailers from selling 3-Kg LPG starting February 1, 2025. The policy aimed to regulate the distribution channels so that subsidies could be better targeted. However, its implementation caused various problems, such as shortages, long queues, and limited access for people living far from official distribution points. Numerous public complaints on social media prompted President Prabowo to revoke the policy on February 4, 2025, and replace it with a "subpangkalan" system to restore smoother distribution.

After the policy revocation, public opinion on X showed mixed responses. Some people felt relieved as LPG supply became accessible again, while others considered the distribution still unstable and the policy dissemination insufficient. X was chosen as the data source for this study because it is open, provides real-time information flow, and allows users to express their opinions spontaneously [1]. Although tweets are brief and informal, they still reflect user opinions and emotions, making them relevant for sentiment analysis.

This study uses Support Vector Machine to classify public opinions on X into two sentiment classes positive and negative. Support Vector Machine was selected because it is effective for binary classification and can find the optimal separator between two sentiment classes. Its advantages have also been demonstrated in previous studies, such as Zakaria et al., study on fuel subsidy policy sentiment analysis with 85% accuracy; Subarkah et al., study on renewable energy discourse, where the use of the Optimized Selection (OS) method improved Support Vector Machine accuracy from 93% to 96%; and Zahra et al., study on the 2024 Indonesian Presidential Dispute Trial reported Support Vector Machine as the best-performing algorithm, achieving 91.1% accuracy with Sastrawi stemming. [2], [3], [4]. However, most sentiment analysis studies in Indonesia typically focus on periods before or during policy implementation and do not evaluate perception changes after the policy is revoked. Therefore, this study analyzes public responses following the policy revocation using the Support Vector Machine method on Indonesian-language data.

The focus of this study is divided into two main aspects: first, to measure public sentiment tendencies after the policy revocation; second, to evaluate the performance of the Support Vector Machine model in classifying tweets that were automatically labeled using the InSet Lexicon. The discussion includes dataset preparation, the design of preprocessing and automatic labeling workflow, and the implementation of the Support Vector Machine model accompanied by interpretable visualizations to provide an objective overview of public perception. Thus, this study emphasizes the importance of monitoring public opinion via social media as a basis for evaluating energy distribution policies in Indonesia in the future.

2. Method

Figure 1 illustrates the research workflow, which consists of problem identification, literature review, data crawling, text preprocessing, data labeling, word weighting, model classification, data visualization, and model evaluation. The data were collected from X through a crawling process using an authentication token in Google Colab, and then processed through several stages, including cleansing, case folding, tokenizing, normalization, stopwords removal, and stemming. Each tweet was labelled as either positive or negative using the Lexicon-based method with the InSet Lexicon dictionary, followed by word weighting using the TF-IDF method. The dataset was divided into training data (80%) and testing data (20%) for classification using the Support Vector Machine. The analysis results were evaluated using a Cross Validation and Confusion Matrix to measure model performance in terms of accuracy, precision, recall, and F1-score, while the findings were visualized using pie charts and Word Cloud.

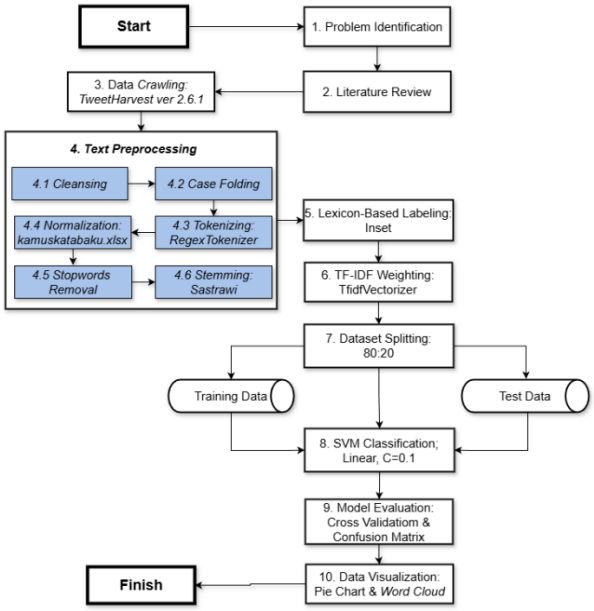


Figure 1: Research Workflow

Data Collection

Data were collected by crawling from X using TweetHarvest version 2.6.1, developed by Helmi Satria, through a personal authentication token in Google Colab with Python. Crawling is an automated process for gathering data from web pages [5]. The crawling process was conducted from February 4 to March 31, 2025, using the code $lang = id$ to filter Indonesian-language tweets, with five main keywords: "larangan pengecer LPG dicabut," "*antrean* LPG 3 kg," "*subpangkalan*," "LPG 3 kg *langka*," and "*distribusi* LPG 3 kg." During crawling, TweetHarvest automatically paused for 10–15 minutes when the data retrieval limit (rate-limit) from X was reached, then resumed, and stopped automatically once the targeted number of relevant tweets was obtained. After crawling, irrelevant tweets were manually filtered, and duplicates were removed using Microsoft Excel. All tweets obtained from the keyword search were used, including original tweets, replies, and quote tweets, while tweets unrelated to the targeted topics were removed. Bot and spam accounts were handled manually through individual inspection in Microsoft Excel. The final dataset comprised 2,524 tweets, free from duplicates, including original tweets and replies, and was saved in CSV format for further preprocessing and analysis. Table 1 below presents several sample tweets obtained from the data crawling results.

Table 1. Sample of Crawled Tweets

No.	Full Text
1	*Ini peringatan darurat untuk Pak Prabowo agar segera ambil tindakan tegas buat mentrinya yang bikin gas elpiji 3 kg langka. Rakyat memantau jangan tunggu rakyat murka*
2	*Kalau nggak berisik mungkin lpg 3 kg kemarin masih langka rakyat masih susah cari gas.. Evaluasi lah https://t.co/XZ5IXZGQ8k*
3	*@InfoJktCom Mantap sekarang gak perlu bingung cari pangkalan resmi LPG 3 kg lagi!*
4	*Gimana gas udah bisa beli di pengecer? #gas #lpg #gaslpg*
5	*Kebijakan Subpangkalan Gas Elpiji 3 Kg Sepatutnya Disosialisasikan Terlebih Dulu https://t.co/087QNGPTGp*
6	*@BebySoSweet Nah yang harus dikontrol itu di agennya bukan di pengecer*
7	*Kementerian ESDM Bantah Kurang Sosialisasi soal Larangan Pengecer Jual Elpiji https://t.co/fnY6B4xJN1*
8	*Pemerintah melarang pengecer menjual gas melon tanpa sosialisasi yang jelas membuat warga kesulitan mendapat gas untuk memasak dan berjualan. Antrean panjang di pangkalan jadi pemandangan sehari-hari bahkan menimbulkan korban jiwa. #GasLangka #Elpiji3kg #GasMelon #OkeGas https://t.co/y6en3pPsCP*

No.	Full Text
9	*Nah harusnya emang ini yg tepat jangan malah pengecer gas LPG 3 kg yg jadi korban. Perang terus berlanjut.*
10	*@OposisiCerdas Gas Melon 3 Kg Masih Tetap Langka Meski Prabowo Cabut Larangan Pengecer Jual Elpiji. Jangan gara - gara LPG 3 kg terjadi Revolusi!*

Preprocessing

Text preprocessing is a crucial initial stage in text data processing that aims to clean and prepare raw data to be more structured, consistent, and ready for analysis [6], [7]. This stage aims to remove irrelevant elements and standardize the text so that the main meaning can be analyzed more accurately. The preprocessing steps performed include cleansing, case folding, tokenization, normalization, stopwords removal, and stemming.

a. Cleansing

Cleansing is performed to remove irrelevant elements such as special symbols, URLs (Uniform Resource Locators), numbers, and non-alphabetic characters that are not needed in the analysis process [8].

Table 2. Cleansing

Before	After
@boii_mmm Tp skrg kan udah bisa beli di pengecer	*Tp skrg kan udah bisa beli di pengecer*

b. Case Folding

Case folding converts all letters to lowercase to maintain data consistency and prevent differences in meaning caused by the use of uppercase and lowercase letters [9].

Table 3. Case Folding

Before	After
Tp skrg kan udah bisa beli di pengecer	*tp skrg kan udah bisa beli di pengecer*

c. Tokenizing

Tokenization splits sentences into smaller units or tokens so that the algorithm can read and process the text more efficiently [10]. In this study, tokenization was performed using the NLTK library with RegexpTokenizer.

Table 4. Tokenizing

Before	After
Tp skrg kan udah bisa beli di pengecer	*['tp', 'skrg', 'kan', 'udah', 'bisa', 'beli', 'di', 'pengecer']*

d. Normalization

Normalization aims to convert non-standard words into their standard forms according [11], to the *Kamus Besar Bahasa Indonesia* (KBBI) or additional dictionaries from the kamuskatabaku.xlsx dataset available on Kaggle website.

Table 5. Normalization

Before	After
['tp', 'skrg', 'kan', 'udah', 'bisa', 'beli', 'di', 'pengecer']	*['tapi', 'sekarang', 'kan', 'sudah', 'bisa', 'beli', 'di', 'pengecer']*

e. Stopwords Removal

This step removes common words that do not provide significant meaning, such as "*di*", "*ke*", or "*kan*", so that the analysis focuses more on meaningful words [12]. The stopwords list used was manually created and adapted to the research context.

Table 6. Stopwords Removal

Before	After
['tp', 'skrg', 'kan', 'udah', 'bisa', 'beli', 'di', 'pengecer']	['tapi', 'sekarang', 'kan', 'sudah', 'bisa', 'beli', 'di', 'pengecer']

f. Stemming

Stemming is the final stage of text preprocessing, which converts affixed words into their root forms to ensure consistent representation, such as "*menjual*," "*pengecer*," and "*larangan*" becoming "*jual*," "*ecer*," and "*larang*" [13]. In this study, stemming was performed using the Sastrawi library StemmerFactory. *S*temmerFactory.

Table 7. Stemming

Before	After
['tapi', 'sekarang', 'sudah', 'bisa', 'beli', 'pengecer']	tapi sekarang sudah bisa beli ecer

Labeling

Sentiment in this research was labeled automatically using the InSet Lexicon developed by Koto and Rahmaningtyas [14]. This lexicon consists of 3,609 positive words and 6,609 negative words that have been stemmed and assigned polarity weights ranging from -5 (very negative) to +5 (very positive). The polarity score for each tweet is calculated by summing the weights of all detected words, then classified as positive if the score is greater than 0, negative if the score is less than 0, and neutral if the score is exactly 0 [15]. The initial labeling was subsequently reviewed to correct any misclassified tweets before being used for model training. Tweets with a total polarity score of 0, considered neutral, were removed from the dataset to focus the analysis on binary sentiment classification (positive vs negative) and to improve model consistency and accuracy. Slang words or terms not found in the InSet Lexicon have been handled through normalization and stopwords removal stages, so they do not affect the polarity score calculation.

Features

The Term Frequency–Inverse Document Frequency (TF-IDF) method was used to assign weights to each word to reflect its level of importance within a document [16]. Features were extracted using $TfidfVectorizer$ with $ngram_range = (1,2)$ to consider unigrams (single words) and bigrams (word pairs), allowing better capture of word context. Words that appear in fewer than 4 documents were ignored ($min_df = 4$) to reduce noise from rarely occurring terms, while words appearing in more than 90% of documents ($max_df = 0.9$) were removed as too common to be informative. After vectorization, the vocabulary size reached 2,939 tokens, which were then used as features for sentiment analysis.

Modelling

The dataset was split using $train_test_split$ with an 80:20 ratio, employing stratified splitting to ensure balanced class distribution and $random_state = 42$ for reproducibility. This split also retains unseen data for evaluation, helping to prevent overfitting and ensuring that the Support Vector Machine model can optimally classify texts into positive and negative categories [17], [18]. Stratification was applied in the train/test split and 10-fold cross-validation (K=10), and the Support Vector Machine kernel and C parameter were selected via $GridsearchCV$ to optimize classification performance.

Support Vector Machine, introduced by Vapnik and colleagues in 1992, works by finding an optimal hyperplane that separates two classes with the maximum margin, where the closest data points (support vectors) determine the

hyperplane's position [19], [20], [21]. The model performed hyperparameter search for kernel (Linear, Polynomial, Radial Basis Function, Sigmoid) and C (0.1, 1, 10) using $GridsearchCV$, and was configured with $class_weight = 'balanced'$ to handle class imbalance. All analyses were conducted using Python 3.12.12 and scikit-learn 1.6.1.

Evaluation

The model performance was evaluated using cross-validation to assess the stability and reliability of the built model. This implementation employed K-fold cross validation with $K = 10$, where the dataset was divided into ten subsets, and each subset was used alternately as the test data while the remaining subsets served as training data [22]. Subsequently, evaluation was conducted using a Confusion Matrix, which compares the model's predictions with the actual data to calculate accuracy, precision, recall, and F1-score [23]. Accuracy indicates the overall correctness of the model, precision measures the accuracy of positive class predictions, recall shows the model's ability to identify all positive instances, and F1-score balances precision and recall [24]. Since this study only involves two sentiment classes, a 2×2 Confusion Matrix was used.

Visualization

Word clouds were created to display the most frequently occurring words in the dataset, both overall and for each sentiment class [25]. The words were taken from text that had been preprocessed and stemmed, so each word was cleaned of special characters and variations in word forms. For each word cloud, the text from a single class (or the entire dataset) was combined into a single string. The Python $WordCloud$ library was used with an image size of 900×450 pixels, displaying a maximum of 150 words and applying different colormaps for each class. Words with very low frequency were automatically filtered by the max_words parameter, making the visualization clearer and highlighting the most relevant words [26].

3. Result and Discussion

Results

Table 8 below presents the results of text preprocessing from 2,524 tweets that have undergone several stages, including cleansing to remove non-alphabetic characters such as links, symbols, and hashtags, case folding to convert all letters to lowercase, tokenizing to split sentences into individual words, normalization to replace non-standard words such as "*nggak*" with standard forms like "*tidak*", stopwords removal to eliminate common words such as "*ke*" or "di", and stemming to reduce inflected words to their root forms, for example, "*pengecer*" to "*ecer*."

This process is essential to ensure that the data are more consistent and ready for analysis, as social media text is generally unstructured and contains a significant amount of noise [27]. Therefore, the results of text preprocessing play a crucial role in improving data quality before proceeding to the labeling and classification stages [28].

Table 8. Training and Validation Performance per Epoch

No.	Full Text	Stemming
1	*Ini peringatan darurat untuk Pak Prabowo agar segera ambil tindakan tegas buat mentrinya yang bikin gas elpiji 3 kg langka. Rakyat memantau jangan tunggu rakyat murka*	*ini ingat darurat untuk bapak prabowo agar segera ambil tindak tegas buat menteri yang buat gas elpiji kilogram langka rakyat pantau jangan tunggu rakyat murka*
2	*Kalau nggak berisik mungkin lpg 3 kg kemarin masih langka rakyat masih susah cari gas.. Evaluasi lah https://t.co/XZ5IXZGQ8k*	*kalau tidak berisik mungkin lpg kilogram kemarin masih langka rakyat masih susah cari gas evaluasi*
3	*@InfoJktCom Mantap sekarang gak perlu bingung cari pangkalan resmi LPG 3 kg lagi!*	*mantap sekarang tidak perlu bingung cari pangkal resmi lpg kilogram lagi*
4	*Gimana gas udah bisa beli di pengecer? #gas #lpg #gaslpg*	*bagaimana gas sudah bisa beli ecer*
5	*Kebijakan Subpangkalan Gas Elpiji 3 Kg Sepatutnya Disosialisasikan Terlebih Dulu https://t.co/087QNGPTGp*	*bijak subpangkalan gas elpiji kilogram patut sosialisasi lebih dulu*
6	*@BebySoSweet Nah yang harus dikontrol itu di agennya bukan di pengecer*	*yang harus kontrol itu agen bukan ecer*

7	*Kementerian ESDM Bantah Kurang Sosialisasi soal Larangan Pengecer Jual Elpiji https://t.co/fnY6B4xJN1*	*menteri energi sumber daya mineral bantah kurang sosialisasi soal larang ecer jual elpiji*
8	*Pemerintah melarang pengecer menjual gas melon tanpa sosialisasi yang jelas membuat warga kesulitan mendapat gas untuk memasak dan berjualan. Antrean panjang di pangkalan jadi pemandangan sehari-hari bahkan menimbulkan korban jiwa. #GasLangka #Elpiji3kg #GasMelon #OkeGas https://t.co/y6en3pPsCP*	*perintah larang ecer jual gas melon tanpa sosialisasi yang jelas buat warga sulit dapat gas untuk masak dan jual antre panjang pangkal jadi pandang hari hari bahkan timbul korban jiwa*
9	*Nah harusnya emang ini yg tepat jangan malah pengecer gas LPG 3 kg yg jadi korban. Perang terus berlanjut.*	*harus memang ini yang tepat jangan jadi ecer gas lpg kilogram yang jadi korban perang terus lanjut*
10	*@OposisiCerdas Gas Melon 3 Kg Masih Tetap Langka Meski Prabowo Cabut Larangan Pengecer Jual Elpiji. Jangan gara - gara LPG 3 kg terjadi Revolusi!*	*gas melon kilogram masih tetap langka meski prabowo cabut larang ecer jual elpiji jangan gara gara lpg kilogram jadi revolusi*

Figure 2 shows the distribution of sentiment labeling in a pie chart, while **Figure 3** presents the same distribution in a bar chart, both generated using the InSet Lexicon after the initial labeling and consistency review. Out of the 2,524 collected tweets, 1,405 tweets (61.2%) were classified as negative and 890 tweets (38.8%) as positive, while 229 neutral tweets were excluded due to their ambiguous nature, which could reduce model accuracy. To compare classification results using two-class and three-class setups, refer to **Table 9**.

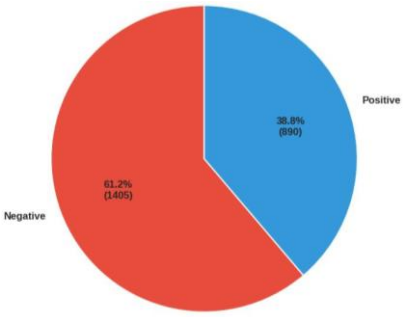


Figure 2. Pie Chart of Sentiment Distribution

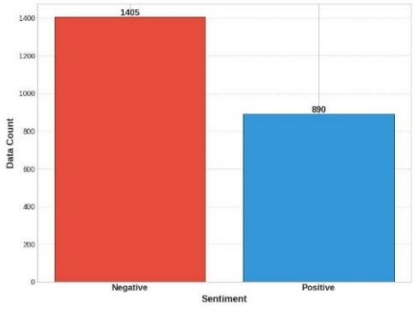


Figure 3. Pie Chart of Sentiment Distribution

The classification results using Support Vector Machine with a linear kernel and C=0.1 demonstrated stable performance. Hyperparameter selection was performed using $GridsearchCV$, using $K = 10$, to ensure model consistency and mitigate the impact of data imbalance between negative and positive classes. **Figure 4** presents the classification report of the model tested on 20% of the data, achieving an accuracy of 84%, with precision, recall, and F1-score for the negative class of 0.82, 0.94, and 0.88, respectively, and for the positive class of 0.87, 0.68, and 0.76, indicating the model's ability to accurately recognize sentiment.

```
Classification Report:
              precision    recall  f1-score   support

    negative       0.82      0.94      0.88       281
    positive       0.87      0.68      0.76       178

    accuracy                           0.84       459
   macro avg       0.85      0.81      0.82       459
weighted avg       0.84      0.84      0.83       459
```

Figure 4. Classification Report

The Support Vector Machine model demonstrates the best performance on the two-class scheme, achieving an accuracy of 0.84 and a macro F1-score of 0.82, which is significantly more stable compared to the three-class scheme, where accuracy drops to 0.69 and the macro F1-score to 0.57. The performance decline in the three-class scheme is mainly due to the difficulty in distinguishing the neutral class and data imbalance. Based on these results, the two-class scheme is chosen as the optimal configuration as it provides more accurate and consistent performance [29]. A comparison of Support Vector Machine performance between the two-class and three-class schemes is presented in **Table 9**.

Table 9. Performance Comparison of Two-Class and Three-Class

Metric	Two-Class	Three-Class
Accuracy	0.84	0.69
Negative F1-score	0.88	0.81
Positive F1-score	0.76	0.29
Neutral F1-score	-	0.61
Macro F1-score	0.82	0.0.57
Weighted F1-score	0.83	0.69

The Confusion Matrix shows that the model classifies most instances correctly, with 263 true negatives and 121 true positives. However, some errors remain, including 18 negative instances predicted as positive and 57 positive instances predicted as negative. These results indicate that the model is stronger at detecting negative sentiment than positive sentiment.

Table 10. Confusion Matrix

Actual	Prediction	
	Positive	Negative
Positive	121	57
Negative	18	263

The error analysis shows that the model tends to focus on words with clear surface meaning without fully understanding the overall context. In the first tweet, the negative indicator "*bodoh*" is overshadowed by words like "*publik*," which often appear in neutral or positive patterns, causing the negative meaning to be misinterpreted. In the second tweet, the presence of words such as "*berisik*," "*susah*," and "*langka*" leads the model to classify it as a negative complaint, even though the overall context is actually positive.

Table 11. Error Analysis

Tweet	Actual Label	Prediction Label
perintah gas lpg kilogram adalah bodoh publik	Negative	Positive

Tweet	Actual Label	Prediction Label
kalau tidak berisik mungkin lpg kilogram kemarin masih langka rakyat masih susah cari gas evaluasi	Positive	Negative

Confidence Interval (CI) were calculated using bootstrap to evaluate the robustness of the model's performance metrics against data variability. **Table 12** presents the CI results, including the mean, standard deviation, and 95% CI for accuracy, as well as the macro and micro averages of precision, recall, and F1-score. The results indicate good model robustness, with a macro F1-score of 0.819 and a micro F1-score of 0.837, reflecting consistent performance across both classes.

Table 12. Confidence Interval Result

	Mean	Std	CI 95% Lower	CI 95% Upper
Accuracy	0.83692	0.01783	0.80174	0.86928
Macro Precision	0.84630	0.01816	0.81031	0.87963
Micro Precision	0.83692	0.01783	0.80174	0.86928
Macro Recall	0.80830	0.01957	0.76871	0.84550
Micro Recall	0.83692	0.01783	0.80174	0.86928
Macro F1-score	0.81931	0.01957	0.77976	0.85610
Micro F1-score	0.83692	0.01783	0.80174	0.86928

To evaluate the superiority of the main model, a performance comparison was conducted between Support Vector Machine, Naïve Bayes, and Logistic Regression. The results show that Support Vector Machine, achieved the highest accuracy (0.84), slightly outperforming Naïve Bayes (0.83) and Logistic Regression (0.81). However, the McNemar test indicates that the performance differences are not statistically significant (Support Vector Machine vs Logistic Regression: $p = 0.0987$; Support Vector Machine vs Naïve Bayes: $p = 0.6291$), suggesting that the three models perform within a similar range [30]. Despite this, Support Vector Machine remains the numerically best-performing model across all evaluations.

Table 13. Comparison of Models and McNemar Test Results

Model Comparison	Accuracy Support Vector Machine	Accuracy Naïve Bayes	Accuracy Logistic Regression	McNemar p-value
Support Vector Machine vs Logistic Regression	0.84	-	0.81	0.0987
Support Vector Machine vs Naïve Bayes	0.84	0.83	-	0.6291
Naïve Bayes vs Logistic Regression	-	0.83	0.81	0.2962

Figure 5 presents a Word Cloud of all words from the collected tweets, with terms such as "*lpg*," "*kilogram*," "*langka*," and "*ecer*" appearing dominantly. These words represent the main topics discussed by users, particularly related to the revocation of the policy banning retailers from selling 3-Kg LPG, which previously caused shortages and limited public access to subsidized LPG. The frequency of each word can be seen in **Figure 6**, which complements the Word Cloud visualization.

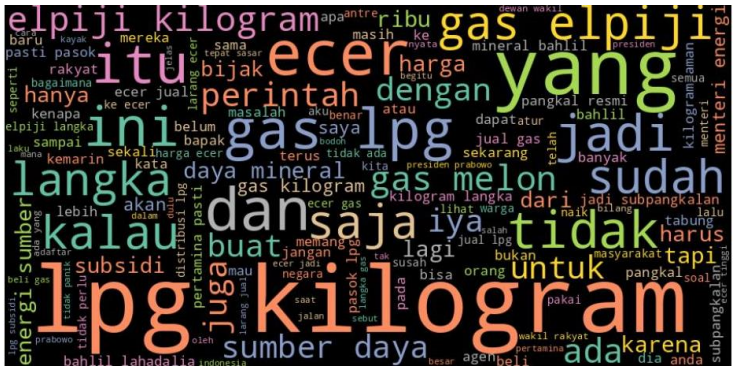


Figure 5. Overall Sentiment Words

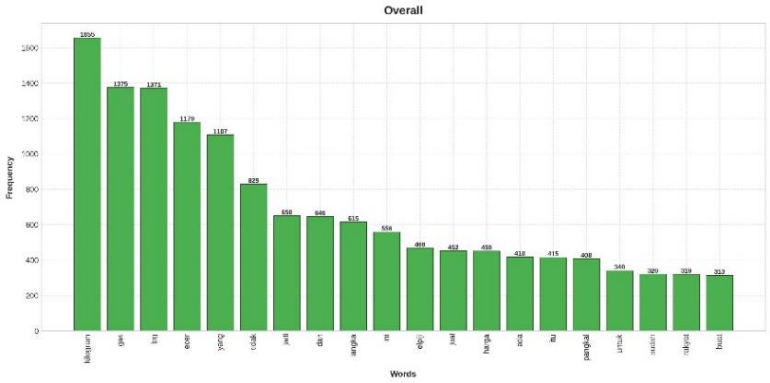


Figure 6. Frequency of All Words

Figure 7 presents a Word Cloud of positive sentiments, with words such as "*distribusi*," "*pasok*," and "*aman*" appearing dominantly. These words reflect the reasons behind the positive perceptions, as many users provided suggestions to the government to strengthen oversight of the 3-Kg LPG distribution so that the supply remains secure and does not experience shortages again. The frequency of each word can be seen in **Figure 8**, which complements the Word Cloud visualization.

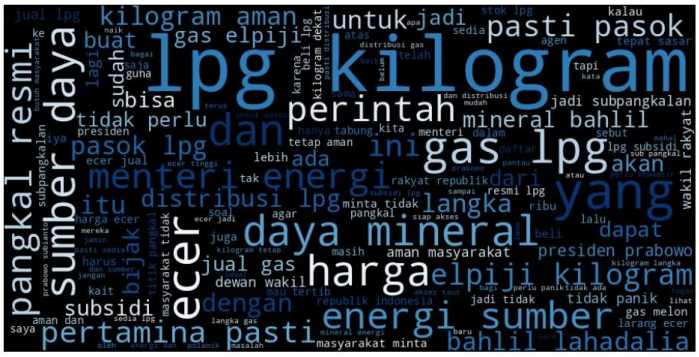


Figure 7. Positive Sentiment Words

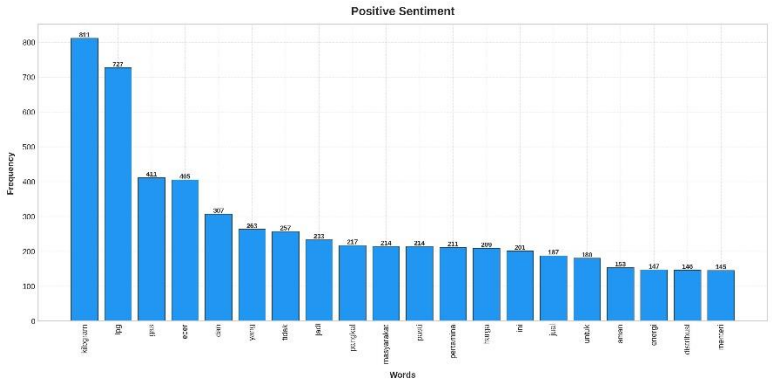


Figure 8. Frequency of Positive Words

Figure 9 presents the Word Cloud of negative sentiments, showing dominant words such as "*susah*," "*langka*," and "*masalah*." These words indicate that negative sentiment persists because some users continued to experience LPG shortages even after the policy was revoked, faced various issues that were perceived as unresolved, and encountered problems related to policy socialization, which was considered ineffective. The frequency of each word's occurrence can be seen in **Figure 10** as a complementary visualization to the Word Cloud.



Figure 9. Negative Sentiment Words

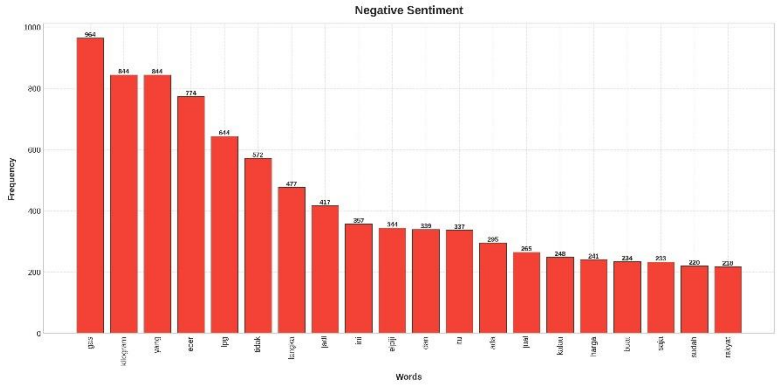


Figure 10. Frequency of Negative Words

Discussion

The experimental results show that linear Support Vector Machine is effective for two-class sentiment classification and more stable than the three-class scheme, which struggled to distinguish neutral tweets due to ambiguity and data imbalance. Among the three models tested, Support Vector Machine achieved the best overall performance, with higher accuracy and F1-score, although the McNemar test indicates that the differences were not

statistically significant. Confusion Matrix evaluation showed that Support Vector Machine detected negative sentiment more reliably than positive, while some misclassifications occurred due to sarcasm, negation, or slang. Error analysis revealed that Support Vector Machine tends to focus on negative words without fully capturing the overall context. Word Cloud were used only as supplementary visualizations to highlight frequently occurring words and do not replace statistical evaluation or model performance metrics.

4. Conclusion

The analysis of 2,524 tweets related to the revocation of the 3-Kg LPG retailer ban showed a predominance of negative sentiment due to gas shortages, distribution problems, and ineffective policy communication. Conversely, positive sentiment emerged from the public who were satisfied with the restored LPG availability and provided feedback on tighter monitoring of 3-Kg LPG distribution. The Support Vector Machine model with a linear kernel and TF-IDF proved effective for two-class sentiment classification in Indonesian, achieving good accuracy and classification performance. These findings can serve as a reference for the government to improve policy communication through effective media and provide quick responses to LPG distribution issues faced by the public.

However, this study has limitations, including the use of Lexicon-based labeling, potential sampling bias from platform X, and a short data collection period. Future research is recommended to use manually labeled test datasets, explore other SVM kernels and additional features, and conduct neutral class drift and time series analyses to assess sentiment stability after March 2025.

Data and Code Availability:

https://bit.ly/lpg-sentiment-colab

https://github.com/helmisatria/tweet-harvest

https://github.com/fajri91/InSet

https://www.kaggle.com/datasets/fornigulo/kamuskatabaku

References:

[1] S. Barreto, R. Moura, J. Carvalho, A. Paes, and A. Plastino, “Sentiment Analysis in Tweets: an Assessment Study from Classical to Modern Word Representation Models,” *Data Min. Knowl. Discov.*, vol. 37, no. 1, pp. 318–380, 2023, doi: https://doi.org/10.1007/s10618-022-00853-0.

[2] Z. Zakaria, K. Kusrini, and D. Ariatmanto, “Sentiment Analysis to Measure Public Trust in the Government Due to the Increase in Fuel Prices Using Naive Bayes and Support Vector Machine,” *Int. J. Artif. Intell. Robot.*, vol. 5, no. 2, pp. 54–62, 2023, doi: https://doi.org/10.25139/ijair.v5i2.7167.

[3] P. Subarkah, B. A. Kusuma, and P. Arsi, “SENTIMENT ANALYSIS ON RENEWABLE ENERGY ELECTRIC USING SUPPORT VECTOR MACHINE (SVM) BASED OPTIMIZATION,” *J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 10, no. 2, pp. 252–260, 2024, doi: https://doi.org/10.33480/jitk.v10i2.5575.

[4] Z. N. Maharani, A. Luthfiarta, and N. Z. Farsya, “Sentiment Analysis of the 2024 Indonesian Presidential Dispute Trial Election using SVM and Naïve Bayes on Platform X,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 440–449, 2024, doi: https://doi.org/10.47065/bits.v6i1.5380.

[5] Y. Qi, “Construction of Online English Corpus Based on Web Crawler Technology,” *Wirel. Commun. Mob. Comput.*, vol. 2022, p. 02, 2022, doi: https://doi.org/10.1155/2022/7589727.

[6] M. A. Palomino and F. Aider, “Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis,” *Appl. Sci.*, vol. 12, p. 8765, 2022, doi: https://doi.org/10.3390/app12178765.

[7] R. R. Raja and H. Darwis, “A Comparative Study of Public Opinion on Indonesian Police: Examining Cases in the Aftermath of the Kanjuruhan Football Disaster,” *Indones. J. Data Sci.*, vol. 6, no. 1, pp. 260–270, 2025,

doi: https://doi.org/10.56705/ijodas.v6i2.235.

[8] A. Tareque, H. H. Siddegowda, D. J. Frank, N. Lee, and R. Moieni, "Overview of Machine Learning Algorithms for Detecting Microaggression in Written Text," *Open J. Soc. Sci.*, vol. 12, no. 07, pp. 347–358, 2024, doi: https://doi.org/10.4236/jss.2024.127025.

[9] A. R. Lubis, Y. Y. Lase, D. A. Rahman, and D. Witarsyah, "Improving Spell Checker Performance for Bahasa Indonesia Using Text Preprocessing Techniques with Deep Learning Models," *Ing. des Syst. d'Information*, vol. 28, no. 5, pp. 1335–1342, 2023, doi: https://doi.org/10.18280/isi.280522.

[10] D. Drašković and S. Milanović, "Aspect-Based Sentiment Analysis of User-Generated Content from a Microblogging Platform," *J. Big Data*, vol. 12, no. 1, p. 02, 2025, doi: https://doi.org/10.1186/s40537-025-01244-0.

[11] J. Khan and S. Lee, "Enhancement of Text Analysis Using Context-Aware Normalization of Social Media Informal Text," *Appl. Sci.*, vol. 11, no. 17, 2021, doi: https://doi.org/10.3390/app11178172.

[12] S. Sarica and J. Luo, "Stopwords in technical language processing," *PLoS One*, vol. 16, no. 8 August, pp. 1–13, 2021, doi: https://doi.org/10.1371/journal.pone.0254937.

[13] N. A. Razmi, M. Z. Zamri, S. S. S. Ghazalli, and N. Seman, "Visualizing Stemming Techniques on Online News Articles Text Analytics," *Bull. Electr. Eng. Informatics*, vol. 10, no. 1, pp. 365–365, 2021, doi: https://doi.org/10.11591/eei.v10i1.2504.

[14] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, no. December, pp. 391–394, 2017, doi: https://doi.org/10.1109/IALP.2017.8300625.

[15] F. Rachmawati, U. Azmi, and R. Azwarini, "Comparison of Lexicon-based Methods and Bidirectional Encoder Representations for Transformers Models in Sentiment Analysis of Government Debt Market Movements," *Int. J. Eng. Comput. Sci. Appl.*, vol. 4, no. 1, pp. 13–28, 2025, doi: https://doi.org/10.30812/ijecsa.v4i1.4832.

[16] H. Liu, X. Chen, and X. Liu, "A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis," *IEEE Access*, vol. 10, pp. 32280–32289, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3160172.

[17] Q. H. Nguyen *et al.*, "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil," *Math. Probl. Eng.*, vol. 2021, p. 02, 2021, doi: https://doi.org/10.1155/2021/4832864.

[18] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, "Trade-off Between Training and Testing Ratio in Machine Learning for Medical Image Processing," *PeerJ Comput. Sci.*, vol. 10, pp. 1–17, 2024, doi: https://doi.org/10.7717/PEERJ-CS.2245.

[19] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Inf.*, vol. 15, no. 4, 2024, doi: https://doi.org/10.3390/info15040235.

[20] N. Amaya-Tejera, M. Gamarra, J. I. Vélez, and E. Zurek, "A distance-based kernel for classification via Support Vector Machines," *Front. Artif. Intell.*, vol. 7, p. 01, 2024, doi: https://doi.org/10.3389/frai.2024.1287875.

[21] S. J. and D. K. U., "Comparison of Sentiment Analysis on Online Product Reviews Using Optimised RNN-LSTM with Support Vector Machine," *Webology*, vol. 19, no. 1, pp. 3883–3898, 2022, doi: https://doi.org/10.14704/web/v19i1/web19256.

[22] Y. Tian, S. Xu, Y. Cao, Z. Wang, and Z. Wei, "An Empirical Comparison of Machine Learning and Deep

Learning Models for Automated Fake News Detection," *Mathematics*, pp. 1–24, 2025, doi: https://doi.org/10.3390/math13132086.

[23] S. Sathyanarayanan and B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, vol. 27, no. 4, pp. 4023–4031, 2024, doi: https://doi.org/10.53555/AJBR.v27i4S.4345.

[24] D. Anggraini, I. Gamayanto, and S. Wibowo, "Comparing Decision Tree and Support Vector Machines in Hospital Satisfaction," *J. Appl. Informatics Comput.*, vol. 9, no. 2, pp. 364–372, 2025, doi: https://doi.org/10.30871/jaic.v9i2.9203.

[25] A. Umair and E. Masciari, "Sentimental and Spatial Analysis of COVID-19 Vaccines Tweets," *J. Intell. Inf. Syst.*, vol. 60, no. 1, pp. 1–21, 2023, doi: https://doi.org/10.1007/s10844-022-00699-4.

[26] T. Malik *et al.*, "Crowd Control, Planning, and Prediction Using Sentiment Analysis: An Alert System for City Authorities," *Appl. Sci.*, vol. 13, no. 3, p. 2, 2023, doi: https://doi.org/10.3390/app13031592.

[27] H. T. Duong and T. A. Nguyen-Thi, "A Review: Preprocessing Techniques and Data Augmentation for Sentiment Analysis," *Comput. Soc. Networks*, vol. 8, no. 1, pp. 1–16, 2021, doi: https://doi.org/10.1186/s40649-020-00080-x.

[28] P. Haryoko, A. Syukur, and N. Rijati, "Development of a Mental Health Classifier Using LSTM and Text Preprocessing Techniques," *Sci. J. Informatics*, vol. 12, no. 1, pp. 77–86, 2025, doi: https://doi.org/10.15294/sji.v12i1.21216.

[29] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment Analysis Methods, Applications, and Challenges: A Systematic Literature Review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024, doi: https://doi.org/10.1016/j.jksuci.2024.102048.

[30] A. R. Aditama and A. F. Wicaksono, "Classification of Customer Complaints on Social Media for E-commerce in Indonesia," *Int. J. Electr. Comput. Eng.*, vol. 15, no. 3, pp. 2977–2985, 2025, doi: https://doi.org/10.11591/ijece.v15i3.pp2977-2985.