Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Public Opinion Sentiment Analysis of Government Fuel Purchasing Policy by the Private Sector Using Support Vector Machine (SVM) Methods

Muhammad Rossi Satria Fitrah[1], Afif Al Qifary[2*], Ahmad Maulana Wahyudi[3], Dea Deswina Sumarna[4], Muhammad Nabiel Alfarizi[5], Fuad Nur Hasan[6]

[1,2,3,4,5,6] *Prodi Studi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika*
ocicuyot@gmail.com[1], apiippwoe@gmail.com[2*], ahmadmaulanachen@gmail.com [3], deadeswinasumarna@gmal.com[4], nabilfarixi@gmail.com[5], *fuad.fnu@bsi.ac.id*[6]

Abstract

Government policies that provide opportunities for the private sector to participate in the purchasing and distribution of fuel oil (BBM) have triggered various reactions within society. The diversity of opinions expressed on social media reflects public perceptions of the effectiveness and potential impacts of these policies. This study aims to examine public sentiment toward the government policy by applying the Support Vector Machine (SVM) method. Data were collected from various social media platforms containing public responses to the issue of private sector involvement in fuel purchasing. The analysis process consisted of several stages, including data collection, data preprocessing (comprising cleansing, tokenizing, stopword removal, and stemming), feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) approach, and sentiment classification using the SVM algorithm. The results show that the SVM algorithm performs well in classifying public opinions into two sentiment categories, positive and negative, with a relatively high level of accuracy. The analysis indicates that the majority of public opinions tend to be negative, driven by concerns over potential price disparities, weakened government oversight, and possible socio-economic impacts. The findings of this study are expected to provide constructive input for the government in evaluating and developing energy policies that are more transparent and oriented toward public interest.

***Keywords**: Sentiment Analysis, Public Opinion, Energy Policy, Fuel Oil, Support Vector Machine (SVM)*

1. Introduction

Government policies that provide opportunities for the private sector to participate in the purchasing and distribution of fuel oil (BBM) have generated diverse responses from the public. Some members of society perceive this policy as a means to strengthen national energy supply and improve the efficiency of fuel management, while others believe it may lead to price increases and reduce government control over the energy sector.

Social media has become one of the most active public spaces for accommodating public opinions regarding government policies. One frequently used platform is YouTube, where users can directly post comments on news, statements, or videos discussing the policy. These comments contain spontaneous public opinions that can be analyzed to determine the direction of public sentiment.

According to Fauzan et al. (2023), sentiment analysis is a computational process used to identify an individual's emotions or opinions toward a particular topic based on textual data. This analysis classifies opinions into two categories, positive and negative, thereby providing a general overview of public perceptions of an issue[1]. Previous research conducted by Ramlan et al. (2023) indicates that the Support Vector Machine (SVM) algorithm achieves high accuracy in classifying public opinions on fuel oil policy issues. In addition, Muhayat (2023) demonstrates that YouTube comment data can serve as an effective data source for understanding public perceptions of government policies.[2], [3]

Based on these considerations, this study aims to analyze public opinion regarding government policies on private sector involvement in fuel oil purchasing using the Support Vector Machine (SVM) algorithm. The results of this study are expected to provide input for the government in evaluating energy policies to make them more adaptive to public needs and aspirations.

2. Literature Review

2.1 Related Research

A study by Silbaqolbina and Najicha (2022) found that the government can address the global scarcity of oil resources and maintain state budget stability by implementing a policy of increasing fuel prices, particularly Pertamax. To maintain the ability of the community, especially the lower-middle economic group, the government decided to increase the price of non-subsidized fuel while maintaining subsidies for certain fuels. The study shows that the fuel price adjustment policy is also in line with the principles of natural resource management stipulated in Article 33 of the 1945 Constitution and aims to ensure national sustainable development. Furthermore, the study emphasizes future socio-economic impacts, such as the possibility of increased inflation and a shift in public consumption from Pertamax to Pertalite, so that subsidies must be more targeted to prevent excessive fiscal burdens.[4]

2.2 Sentiment Analysis

Sentiment analysis is a technique used to identify and measure opinions or emotions contained in textual data. This technique is commonly applied to determine public opinion tendencies toward a particular policy or event. According to Adilah and Alit (2023), sentiment analysis can assist researchers in objectively understanding public attitudes toward government policies through social media data. [5] Research conducted by Fauzan et al. (2023) indicates that the majority of public opinions regarding fuel price increases tend to be negative. This finding demonstrates that sentiment analysis can be used as a tool to assess the level of public satisfaction and trust in policies being implemented. [1]

2.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a learning system that operates by constructing a hypothesis space composed of linear functions in a high-dimensional feature space and is trained using optimization theory–based learning algorithms. SVM was first introduced in 1992 by Vapnik as a combination of several prominent concepts in the field of pattern recognition. SVM is a classification algorithm that utilizes a hypothesis space consisting of linear functions within a high-dimensional feature set. The primary objective of SVM is to identify the optimal separating function between classes. The optimal separating function is defined as the one that produces the maximum margin between vectors belonging to different classes and lies equidistant between them. In this study, the separating function sought is a linear function.[6] According to Ramlan et al. (2023), the main advantage of Support Vector Machine (SVM) lies in its ability to classify high-dimensional data such as text. SVM is capable of grouping data into positive and negative sentiment categories with a high level of accuracy without requiring a complex training process.[3] Supian (2023) also shows that SVM is more stable than other algorithms such as Naïve Bayes or Decision Tree in the context of Indonesian-language text analysis. In addition, Tjut Adek (2025) emphasizes that the use of SVM produces more consistent classification results because this method focuses the learning process on the most influential data points, namely the class boundary points known as support vectors.[7]

2.4 YouTube as a Source of Public Opinion Data

YouTube is now not only a video-sharing platform, but also a space for the public to discuss and respond to policy issues. Comments appearing in video feeds reflect the public's direct, unfiltered views.[2] Research by Andhini et al. (2023) states that data from YouTube has advantages because it is natural, up-to-date, and covers various levels of society, making it highly relevant for use in public opinion research.[8]

3. Methodology

3.1 Research Design

This research uses a descriptive quantitative approach with a sentiment analysis method based on the Support Vector Machine (SVM) algorithm. The goal is to classify public opinion regarding government policy on private fuel purchases into two sentiment categories: positive and negative.

3.2 Data Collection

Data were obtained from YouTube user comments on videos discussing government policies regarding private sector involvement in fuel purchases. Data were collected using the YouTube Data API using keywords such as "private sector fuel purchases" and "government fuel policy." Comments containing advertising, spam, or irrelevant content were removed. This collection process followed the method used by Muhayat (2023), who also used YouTube comment data as the primary source for public opinion research.[2]

3.3 Data Pre-Processing

Before being analyzed, the comment data underwent several pre-processing stages to improve the quality of the classification results, namely:

1. Cleansing: removing unnecessary symbols, URLs, punctuation, and emoticons.
2. Case Folding: converting all letters to lowercase.
3. Tokenizing: separating sentences into words.
4. Stopword Removal: removing common words that do not affect the sentiment.

5. Stemming: returning words to their base form using the Nazief–Adriani algorithm.

This step refers to research by Adilah & Alit (2023) and Tjut Adek (2025) who successfully applied similar steps to Indonesian language texts.[5], [9]

3.4 Classification with Support Vector Machine

Support Vector Machine (SVM) is a statistical method that can be used for classification. SVM is a technique for understanding hyperplanes that can be used to analyze two data sets from two different classes (Vapnik, 1999). One of the advantages of SVM is its ability to quickly determine distances using support vectors, which speeds up the computational process (Vapnik, 1995). Rustam et al. (2003) conducted a study on SVM, comparing the K-Nearest Neighbor (KNN) classification method with SVM. The study concluded that SVM had superior performance because it could correctly classify odor data 100% of the time based on its class. Furthermore, Rachmandan Purnami (2012) conducted a study on cancer grade classification using logistic regression and support vector machines (SVM). The results showed that SVM had a higher accuracy, at around 98.11%.[10]

Each comment will be classified into two categories:

1. Sentiment Positive : comments that support or agree with government policies.
2. Sentiment Negative : comments that reject, criticize, or express disagreement with the policy.

3.5 Model Evaluation

Model testing is necessary to evaluate the performance of the Support Vector Machine (SVM) method. In this study, the evaluation was conducted using accuracy values. Accuracy is an evaluation metric used to measure the extent to which a classification or prediction model can produce correct results or match existing data. Accuracy indicates the percentage of success of the model in predicting the correct class or label. This measures the extent to which the model can correctly classify the data as a whole. In other words, accuracy indicates the percentage of success of the model in predicting the correct class.[11]

4. Results and Discussion

4.1 Data Collection and Pre-processing Results

A total of 9,198 initial comments from a predetermined set of YouTube videos were successfully collected during the initial data collection process. In a rigorous preprocessing stage, this raw data was cleaned of noise and its formatting adjusted. The amount of data relevant for analysis was reduced to 8,036 comments after the cleaning and language filtering process, leaving only Indonesian comments. Next, an automatic labeling process was performed using a pretrained transformer model (mdhugol/indonesian-roberta-base-sentiment-classifier). In this process, neutral data were ignored, prioritizing binary classification analysis (positive and negative). The final result of this stage was a labeled dataset of 6,698 comments, ready for model training. Table 2 shows the class distribution of this final dataset.

Tabel 1. Public Comments on Videos Regarding the Policy of Collaboration Between Private Fuel Stations and Pertamina

No	video_id	parent_id	Text	Author	Published_at	Like Count	Video Title
0	wIFheB4Pk8E	None	Bgini klo menteri gak nguasai fakta lapangan, perbaiki dulu kualitas BBM Pertamina, orang lari ke SPBU swasta gara2 BBM Pertamina busuk smua.	@indohunter2075	2025-10-13 23:56:26+00:00	0	Bahlil: SPBU Swasta Harus Setuju untuk Kolaborasi dengan Pertamina
1	wIFheB4Pk8E	None	Dan memang harus setuju.. Itu namanya terpaksa cuk	@gunawanwibisono5049	2025-10-12 11:38:23+00:00	0	Bahlil: SPBU Swasta Harus Setuju untuk Kolaborasi dengan Pertamina
2	wIFheB4Pk8E	None	Menteri 😅	@robbyaydy8364	2025-10-12 16:58:02+00:00	0	Bahlil: SPBU Swasta Harus Setuju untuk Kolaborasi dengan Pertamina
3	wIFheB4Pk8E	None	Segera lempar aja orang ini, bagaimana kompetitor dipaksa harus setuju. Makin lama, makin bobrok saja negara ini.	@Zainal-v7q	2025-10-12 16:23:15+00:00	0	Bahlil: SPBU Swasta Harus Setuju untuk Kolaborasi dengan Pertamina
4	wIFheB4Pk8E	None	Setuju nggak beli dan kabur dr ind😂😂😂😂 inilah kalau tdk kompeten dg bidangnya	@Sebastian-b9irh	2025-10-09 15:16:13+00:00	0	Bahlil: SPBU Swasta Harus Setuju untuk Kolaborasi dengan Pertamina

Tabel 2. Sentiment Class Distribution in the Final Dataset

Label	Jumlah Komentar	Persentase
Negatif	5.837	87,1%
Positif	853	12,9%
Total	6.698	100%

From Table 2, it is clear that there is a significant class imbalance, where negative sentiment dominates the dataset.

4.2. Model Training and Evaluation Results

A pipeline consisting of the LinearSVC algorithm and feature extraction using TfidfVectorizer, which combines word- and character-level n-grams, was employed to build the sentiment classification model. Hyperparameter optimization was conducted using HalvingGridSearchCV with the f1_macro evaluation metric, which is robust to class imbalance, to optimize model performance. This process yielded the best cross-validation Macro F1 score of 0.7839.

Subsequently, the model with the optimal hyperparameters was evaluated on a held-out test dataset comprising twenty percent of the total data. The model's performance on the test set is presented through a classification report (Table 3) and a confusion matrix (Figure 1).

Tabel 3. Model Classification Report on Test Data

Kelasifikasi	Precision	Recall	F1-Score
Negatif	0.948	0.937	0.943
Positif	0.603	0.649	0.625

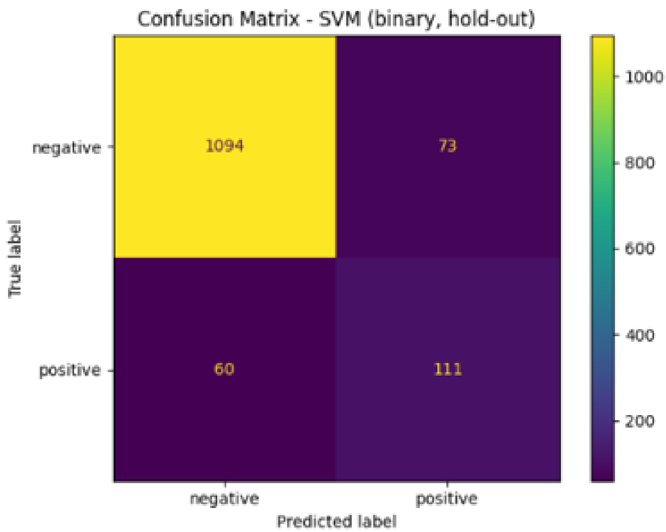

Fig. 1. Confusion Matrix of Model Performance on the Test Data

The results indicate that the model performs very well in identifying negative sentiment (F1-score = 0.943), which represents the majority class. However, its performance is lower for the positive class (F1-score = 0.625), which is expected given the smaller number of samples. Nevertheless, the Macro Average F1-score of 0.784 demonstrates that the model maintains a balanced generalization capability across both classes. Overall, the model achieves an accuracy of 90.1%.

4.3. Links and bookmarks

To identify general sentiment patterns, the trained and calibrated model was subsequently applied to 6,698 comments. The analysis results indicate that negative sentiment dominates the overall sentiment landscape regarding this subject.

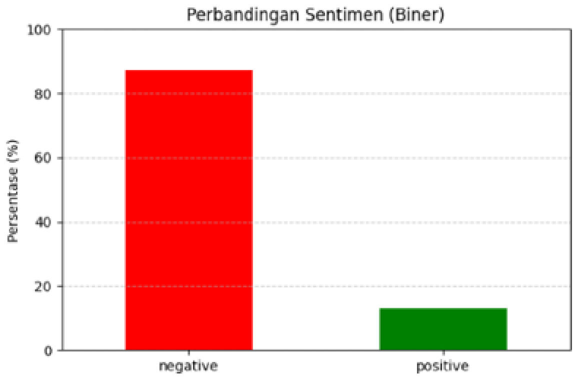

Fig. 2. Comparison of Overall Sentiment Percentages

The overall proportion of public sentiment is divided into 87.01% negative and 12.99% positive, as illustrated in Figure 2. The high number of negatively toned comments indicates widespread dissatisfaction and criticism among the audience regarding the issues discussed in the related videos.

To obtain a deeper qualitative understanding, comments with the highest sentiment probability were further examined. Figures 3 and 4 present the results of this analysis.

Top POSITIVE:

	video_title	clean	positive	negative
307	Alasan SPBU Swasta Kompak Ogah Beli BBM dari Pertamina Meski Kehabisan Stok	bagus	0.998936	0.001064
3420	Warga Gugat Menteri ESDM Bahlil Lahadalia Karena Stok BBM di SPBU Swasta Kosong | Liputan 6	bagus	0.998936	0.001064
4808	Panas! Purbaya Vs Bahlil Perkara Pertamina Belum Tambah Kilang	pak purbaya mantap	0.998440	0.001560
5798	Curhat ke DPR, Menkeu Purbaya Kesal Pertamina Malas Bangun Kilang, Sering Impor Minyak	mantap pak purbaya	0.996293	0.003707
326	Alasan SPBU Swasta Kompak Ogah Beli BBM dari Pertamina Meski Kehabisan Stok	mantap	0.993525	0.006475

Fig 3. Examples of Comments with the Strongest Positive Sentiment

Top NEGATIVE:

	video_title	clean	positive	negative
4494	Panas! Purbaya Vs Bahlil Perkara Pertamina Belum Tambah Kilang	si bahlul tau apa tau ee datanya aja ada ko akal an cuma biyas korupsi dri impor emo_pos emo_pos pecat aja pecat lah si bahlul	0.000110	0.999890
4514	Panas! Purbaya Vs Bahlil Perkara Pertamina Belum Tambah Kilang	aneh apa istemewanya si bahlul ini jadi smpai sekarang belum di ganti juga sm prabowo byk banget pernytaan dia yang tidak NEG_masuk NEG_akal bahkan pertamina itu sarang mafia parah ancur negara si bahlul klw masih menjabat	0.000265	0.999735
3649	Warga Gugat Menteri ESDM Bahlil Lahadalia Karena Stok BBM di SPBU Swasta Kosong | Liputan 6	pecat aja si bahlul ini	0.000293	0.999707
6157	Viral PHK Pekerja SPBU Swasta, Ini Keterangan Menteri ESDM Bahlil Lahadalia | KOMPAS PAGI	banyak ngobrol tapi ngarti kagak bloon nya itu lho jangan NEG_diulang NEG_lah bos tau diri kalau tidak NEG_becus NEG_mending mundur aja	0.000323	0.999677
1149	Bahlil Buka Suara Soal SPBU Swasta Tak Jadi Beli BBM dari Pertamina | SAPA MALAM	sama sama import tapi maunya menang sendiri import juga swasta pakai duit sendiri tidak NEG_rugikan NEG_pemerintah atau pertamina sudah monopoli pasar tapi masih rugi dan sekarang ruginya ngajak ngajak yang lain pula	0.000328	0.999672

Fig. 4. Examples of Comments with the Strongest Negative Sentiment

From the table above, several important points can be discussed:

1. Characteristics of Positive Comments: Comments identified as strongly positive tend to be short, direct, and use general expressions of praise such as "good" and "excellent."

2. Complexity of Negative Comments: In contrast, negative comments tend to be longer, more argumentative, and often contain specific critical terms. The model's ability to correctly handle comments containing negation demonstrates the effectiveness of the preprocessing stage, particularly the negation handling feature.

3. Implications of the Findings: The overwhelming dominance of negative sentiment, supported by qualitative comment examples, provides a strong signal of underlying issues or fundamental concerns attracting public attention regarding the video topic. This analysis not only measures sentiment but also successfully identifies concrete examples of complaints and praise expressed by the audience.

5. Conclusion and Suggestions

5.1 Conclusion

For future research development, it is recommended that efforts focus on methodological refinement and expansion of the analytical scope. From a methodological perspective, the application of advanced oversampling techniques such as SMOTE should be explored to more effectively address class imbalance, along with evaluating more complex deep learning model architectures, such as LSTM or IndoBERT, to capture deeper semantic context. In addition, the data scope should be expanded by integrating sources from other social media platforms, such as Twitter and Facebook, to obtain a more holistic view of public opinion. Finally, the analysis could be extended from binary classification to multi-class classification by including a neutral category, or further developed into Aspect-Based Sentiment Analysis (ABSA) to identify sentiment toward more specific and detailed aspects of the topic.

5.2 Suggestions

For future research development, it is recommended that attention be directed toward methodological refinement and the expansion of analytical scope. From a methodological standpoint, the application of advanced oversampling techniques such as SMOTE should be explored to more effectively address class imbalance, along with evaluating more sophisticated deep learning model architectures, such as LSTM or IndoBERT, to capture richer semantic context. In addition, the data scope should be expanded by integrating sources from other social media platforms, such as Twitter and Facebook, to obtain a more comprehensive view of public opinion. Finally, the analysis could be extended beyond binary classification to multi-class classification by including a neutral category, or further developed into Aspect-Based Sentiment Analysis (ABSA) to identify sentiment toward more specific and detailed aspects of the topic.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to all parties who provided support in the completion of this research. Special thanks are extended to the academic supervisor for their guidance, valuable input, and continuous support throughout the research process. The authors also thank colleagues and other contributors who assisted with data collection, data processing, and the preparation of this article. In addition, high appreciation is conveyed to the institution and the Journal of Informatics Engineering for providing the opportunity to publish the results of this study. It is hoped that this research will contribute to the advancement of knowledge, particularly in the fields of data analysis and the application of machine learning algorithms in information technology.

References

[1] M. R. Fauzan, H. O. L. Wijaya, and J. Karman, “ANALISIS SENTIMEN MASYARAKAT TERHADAP KENAIKAN HARGA BBM DI MEDIA SOSIAL TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE,” vol. 1, no. 1, 2023.

[2] T. Muhayat, A. Fauzi, and J. Indra, “Analisis Sentimen Terhadap Komentar Video Youtube Menggunakan Support Vector Machines,” *Progresif J. Ilmi. Kom*, vol. 19, no. 1, p. 231, Feb. 2023, doi: 10.35889/progresif.v19i1.1060.

[3] R. Ramlan, N. Satyahadewi, and W. Andani, “Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM,” *Jambura J. Math*, vol. 5, no. 2, pp. 431–445, Aug. 2023, doi: 10.34312/jjom.v5i2.20860.

[4] Y. Zahra Silbaqolbina and F. Ulfatun Najicha, “Kebijakan Pemerintah Dalam Menaikkan Harga Bahan Bakar Minyak Serta Dampaknya Bagi Masyarakat,” *JSF*, vol. 2, no. 06, pp. 604–611, June 2022, doi: 10.54543/fusion.v2i06.198.

[5] H. S. ’Adilah and R. Alit, “Analisis Sentimen Masyarakat Twitter Terhadap Kebijakan Pemerintah Dalam Menaikkan Harga Bahan Bakar Minyak Dengan Menggunakan Metode Support Vector Machine,” *JINACS*, vol. 5, no. 02, pp. 201–215, Sept. 2023, doi: 10.26740/jinacs.v5n02.p201-215.

[6] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, “Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier,” *SISFOKOM*, vol. 10, no. 3, pp. 432–437, Dec. 2021, doi: 10.32736/sisfokom.v10i3.1302.

[7] A. Supian, “Penerapan SVM dan Word2Vec untuk Analisis Sentimen Ulasan Pengguna Aplikasi DANA,” *jikstik*, vol. 23, no. 3, Sept. 2024, doi: 10.32409/jikstik.23.3.3642.

[8] Asmara Andhini, Fadilah Nuria Handayani, Intan Diasih, and Nurmalitasari, “Analisis Sentimen Opini Publik pada Channel Youtube Mata Najwa Menggunakan Metode SVM,” *jutiti*, vol. 5, no. 2, pp. 139–154, July 2025, doi: 10.55606/jutiti.v5i2.5426.

[9] R. Tjut Adek, Z. Fitri, and S. C. Siregar, “Analisis Sentimen Komentar Pada Saluran Youtube Beauty Vlogger Berbahasa Indonesia Menggunakan Metode Support Vector Machine,” *algoritme*, vol. 5, no. 2, pp. 164–175, Apr. 2025, doi: 10.35957/algoritme.v5i2.9692.

[10] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, “PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG”.

[11] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM: Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis,” *MALCOM*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.