



Public Sentiment Analysis on the Issuance of Panda Bonds as an Effort for Rupiah Stability Using SVM Algorithm on Youtube Social Media

Junjung Rahmat Santosa^{1*}, Rangga Apriwijaya², Ilham Ardiasyah³, Rangga Apriansyah⁴, Destiarini⁵

^{1,2,3,4} Department of Informatics, Faculty of Engineering and Computer Science, Universitas Baturaja, Indonesia
junjungrahmat8@gmail.com^{1*}, ranggabta15@gmail.com²

Abstract

The stability of the Rupiah exchange rate is a crucial indicator of Indonesia's economic health, one of which is pursued through the issuance of Panda Bonds. However, this policy has triggered dynamic discourse on social media, particularly YouTube. This study aims to map public perception and test the performance of the Support Vector Machine (SVM) algorithm in classifying sentiments related to this issue. The research methodology includes scraping YouTube comment data, text preprocessing, automated labeling using the Lexicon-based method, and classification using SVM with a Linear kernel. From a total of 659 collected data, the results show that public sentiment is dominated by positive responses at 51.9%, followed by neutral sentiment at 29.0%, and negative sentiment at 19.1%. While public concerns focus on the debt burden and foreign currency dependence, there is overall support for economic stability efforts. The model evaluation demonstrates excellent performance, achieving an accuracy rate of 87.86%, precision of 88.79%, and an F1-score of 87.96%. This proves that a hybrid approach between Lexicon-based and SVM is effective in analyzing complex public opinions within the economic domain on social media.

Keywords: *Lexicon-based, Panda Bonds, Rupiah Stability, Sentiment Analysis, Support Vector Machine, YouTube*

1. Introduction

Exchange rate stability is one of the fundamental indicators that determine the economic health of a country. For Indonesia, fluctuations in the Rupiah exchange rate against foreign currencies, particularly the United States Dollar (USD), are often a major challenge in maintaining macroeconomic stability. Dynamic global pressures demand the government to continuously innovate in finding alternative financing sources that are not only efficient but also capable of diversifying currency risks. One of the strategic steps taken by the Indonesian Government through the Ministry of Finance is the plan to issue Panda Bonds, which are bond instruments denominated in Renminbi (Yuan) issued in China's domestic market.

The strategy of diversifying debt instruments into non-traditional currencies such as the Yuan is a crucial preventive measure to reduce dependence on USD dominance (de-dollarization) and mitigate extreme exchange rate volatility. Theoretically, the issuance of Panda Bonds provides Indonesia with access to reach a wider investor base in the world's second-largest financial market. However, this policy is not free from public discourse. Any policy related to foreign debt instruments, especially those involving partnerships with China, tends to trigger highly dynamic and polarized sentiments among the Indonesian public [1].

In the current digital era, social media platforms such as YouTube have transformed into the primary digital public sphere where the public expresses opinions, criticisms, or support for government policies in real-time. Comments on economic news videos regarding Panda Bonds reflect diverse collective perceptions, ranging from optimism about strengthening foreign exchange reserves to deep concerns about long-term debt burdens and geopolitical issues. This phenomenon generates massive textual data. Processing this data manually to obtain objective conclusions is impossible to do efficiently, necessitating a computationally based technology approach. The use of Natural Language Processing (NLP) through sentiment analysis serves as a scientific solution to automatically map public opinion. In this study, the Support Vector Machine (SVM) algorithm was chosen as the primary classification method. As proven by research [2], SVM has robustness in handling unstructured Indonesian text data with high feature dimensions, thereby providing more stable classification results compared to other probabilistic algorithms. However, the main obstacle in large-scale sentiment analysis often lies in the data labeling (annotation) process, which is time-consuming and costly.

To overcome labeling issues, this study applies the Lexicon-based method as an initial stage for automated labeling. This approach utilizes a sentiment word dictionary to determine text polarity without requiring substantial manual intervention. According to [3], integrating the Lexicon-based method as an automated labeler, which is then validated by a machine learning model such as SVM, represents a highly

effective hybrid approach to increase data processing speed without significantly sacrificing classification accuracy. By integrating YouTube data scraping techniques, rigorous text preprocessing, lexicon-based labeling, and SVM classification, this study aims to present an in-depth analysis of how the public responds to the Panda Bonds policy. The results of this study are expected not only to contribute to the advancement of informatics science, particularly in data science, but also to serve as a reference for policymakers in understanding the pulse of public opinion to improve the quality of economic policy communication in the future.

2. Theoretical Review

Studies regarding sentiment analysis of public opinion toward national economic policies have been widely conducted using various computational approaches. Prior research by Guild [1] demonstrates that policies involving foreign debt instruments, particularly partnerships with foreign capital, invariably trigger highly dynamic sentiment polarization among the Indonesian public. These sentiments reflect structural anxieties regarding fiscal sovereignty and long-term financial burdens. In the context of classification algorithms, the Support Vector Machine (SVM) algorithm has been proven to exhibit highly robust and stable performance in handling the characteristics of non-standard Indonesian text on social media. Bangkalang [2] demonstrates that through the construction of an optimal linear separating boundary (*hyperplane*), SVM is capable of maintaining stable accuracy rates even within high-dimensional feature spaces compared to other conventional probabilistic algorithms. The superiority of SVM in minimizing structural risk errors makes it the premier choice for modeling public opinion text analysis. However, the greatest constraint in traditional supervised machine learning approaches lies in the absolute dependence on the availability of manually labeled training data (*ground truth*). This manual annotation process is highly time-consuming and prone to human subjectivity bias. To overcome these limitations, hybrid approaches that integrate Lexicon-based methods with SVM have begun to be widely developed. Nugraha et al. [3] prove that utilization of an Indonesian sentiment dictionary as an initial automated labeler, which is subsequently re-classified by SVM, is capable of slashing annotation time by up to 80% while maintaining optimal model precision and F1-score levels.

The utilization of social media as a digital public sphere is also a major focus in sentiment research. Arsi and Waluyo [4] assert that text-and-video-based platforms such as YouTube archive massive volumes of opinion data that reflect immediate, real-time public reactions toward crucial socio-economic issues. Furthermore, the de-dollarization trend underpinning the Panda Bonds issuance policy aligns with the study by Abdurokhim [5], which states that diversifying non-USD currency instruments is critical to safeguarding monetary sovereignty and cushioning exchange rate stability from global macroeconomic shocks. The theoretical foundation of the Lexicon-based methodology itself relies on the study by Taboada et al. [6], explaining that dictionary-based calculations of word polarity weights are highly effective for extracting contextual text opinion trends. The evaluation of kernel functions within the SVM algorithm is further supported by research from Rabbani et al. [7], confirming that the Linear kernel consistently yields the best performance and computational efficiency on sparse text data. Other comparative studies by Fitriyana et al. [8], Mahendrajaya et al. [9], and Husada and Paramita [10] heavily reinforce the argument that integrating comprehensive text preprocessing, dictionary-based automated labeling, and SVM classification successfully generates highly precise visualizations and mappings of public sentiment across various digital services and public domains in Indonesia.

3. Research Method

The methodological framework of this study is structured systematically to ensure complete data integrity, procedural transparency, and optimal classification accuracy. The sequential workflow spans from initial raw data acquisition on social media to the final rigorous evaluation of the Support Vector Machine (SVM) machine learning model.

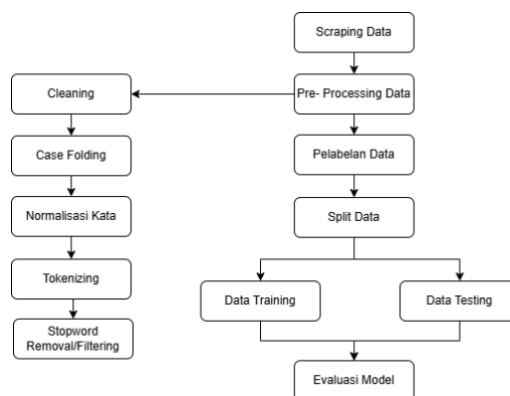


Fig. 1: Research Flowchart Diagram

The research procedure is explicitly divided into two major tracks as illustrated in the flowchart. The primary track involves data acquisition, automated labeling, dataset partitioning, and model training/testing. The secondary parallel track details the rigorous text preprocessing pipeline, which is fundamental to transforming unstructured public noise into highly informative computational features. The empirical dataset utilized in this research comprises public comments extracted directly from the YouTube social media platform. YouTube was purposefully selected as the digital public sphere due to its massive user base in Indonesia and its high density of policy-related debates. The data extraction process was executed programmatically using the Python programming language, leveraging the official google-api-python-client library to interact with the YouTube Data API v3. The data collection targeted comment threads from

popular, verified digital news channels broadcasting discussions on macroeconomic policies. The search criteria were bound by strict thematic and temporal constraints: Thematic Keywords: "Panda Bonds" and "Stabilitas Rupiah" (Rupiah Stability); Temporal Window: May 7, 2026, to May 14, 2026. Textual data derived from social media platforms is notoriously noisy, highly unstructured, and heavily saturated with non-standard linguistic forms. Therefore, text preprocessing represents a critical phase to enhance data quality, eliminate computational noise, and standardize text inputs before feature extraction. The strict preprocessing pipeline implemented in this study consists of five consecutive stages: Cleaning, Case Folding, Word Normalization, Tokenizing, and Stopword Removal/Filtering. Following the completion of the preprocessing pipeline, the cleaned dataset must be annotated with ground truth sentiment labels. To circumvent the extensive time, financial costs, and subjectivity associated with manual human annotation, this study implements an automated, algorithmic labeling approach using a Lexicon-based method. This approach leverages a comprehensive, pre-validated Indonesian sentiment lexicon dictionary containing a vast index of words pre-assigned with specific numerical polarity weights. The net sentiment score determines the final categorical label (Positive, Negative, or Neutral).

To rigorously train and evaluate the machine learning classifier, the completely labeled dataset is partitioned into two distinct subsets using an 80:20 distribution ratio: Training Data (80%) and Testing Data (20%). The final phase of the methodology involves evaluating the predictive performance of the Support Vector Machine model against the ground truth labels of the testing dataset using the Confusion Matrix framework to calculate Accuracy, Precision, Recall, and F1-Score.

4. Results and Discussion

Following the automated scraping and the execution of the rigorous text preprocessing pipeline, a final dataset consisting of exactly 659 public comments was secured for analysis. This clean dataset explicitly captures the dynamic public discourse, immediate anxieties, and shifting perceptions of Indonesian citizens regarding the Ministry of Finance's strategic plan to issue Renminbi-denominated bonds in China's domestic market. To gain an initial qualitative understanding of the central themes dominating the public digital sphere, a WordCloud visualization was generated from the preprocessed tokens. This visualization highlights word frequency through font size, allowing for immediate identification of core public focus points.



Fig. 2: WordCloud of Panda Bonds Sentiment Discourse

An analytical review of the WordCloud demonstrates that the textual landscape is heavily dominated by a specific cluster of keywords: "utang" (debt), "negara" (country), "yuan", "bunga" (interest), and "indonesia". The prominent, massive visibility of "utang" and "bunga" explicitly underscores that the primary underlying anxiety of the general public centers on the potential long-term financial liabilities, debt-servicing burdens, and sovereignty risks associated with entering partnerships involving foreign loans. To provide precise empirical data supporting this visualization, the absolute frequencies of the top 15 words were calculated as shown in Table 1.

Table 1. Frequency of the 15 Most Frequently Occurring Words

No	Word	Frequency	No	Word	Frequency	No	Word	Frequency
1	INDONESIA	145	6	YUAN	98	11	BOND	60
2	UTANG	137	7	CHINA	85	12	PAKAI	60
3	HUTANG	112	8	DOLLAR	83	13	BELI	57
4	RUPIAH	111	9	UANG	75	14	PERINTAH	57
5	NEGARA	105	10	DOLAR	71	15	EKONOMI	57

By deploying the automated Lexicon-based annotation algorithm across the 659 preprocessed comments, the macro-level distribution of public opinion regarding the Panda Bonds policy was successfully mapped into three distinct categories as detailed in Figure 3.

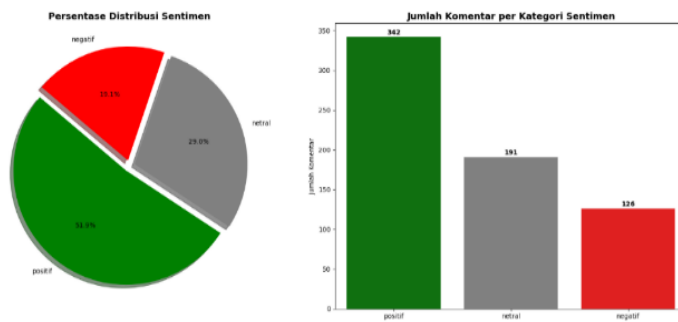


Fig. 3.: Percentage of Public Sentiment Distribution

The quantitative distribution of public sentiment shows Positive Sentiment at 51.9% (342 comments), Neutral Sentiment at 29.0% (191 comments), and Negative Sentiment at 19.1% (126 comments). To validate the capability of the hybrid framework, the SVM model with a Linear kernel was evaluated using the testing dataset (206 unseen comments) as shown in Table 2.

Table 2. Overall SVM (Linear) Model Evaluation Results

Metric	Value
ACCURACY	0.8786 (87.86%)
PRECISION	0.8879
RECALL	0.8786
F1-SCORE	0.8796

The overall classification accuracy achieved is 87.86%, proving that the combination of structured text preprocessing and a linear SVM hyperplane is highly robust. To dissect the performance with granular detail, a class-specific classification report was generated in Table 3.

Table 3. Detailed Classification Report per Category

Class	Precision	Recall	F1-Score	Support
NEGATIF	0.97	0.94	0.96	71
NETRAL	0.77	0.91	0.83	65
POSITIF	0.92	0.79	0.85	70
ACCURACY	-	-	0.88	206
MACRO AVG	0.88	0.88	0.88	206
WEIGHTED AVG	0.89	0.88	0.88	206

The Negative Sentiment class achieved the absolute highest precision score of 0.97, demonstrating an incredibly low false-positive rate for hostility detection. To visually inspect the exact distribution of correct predictions and specific misclassifications across the multi-class system, the Confusion Matrix was mapped as a heatmap in Figure 4.

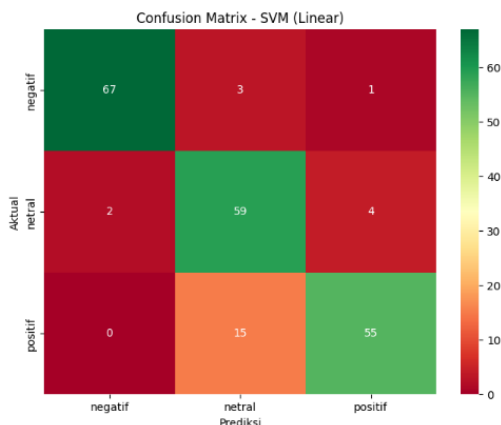


Fig. 4: Confusion Matrix Heatmap - SVM (Linear)

The empirical findings validate that the hybrid integration of automated Lexicon-based annotation with a supervised Support Vector Machine classifier represents an exceptionally potent framework for social media analytics. Achieving an 87.86% accuracy level on short, informal text demonstrates that the high-dimensional linear hyperplane of SVM is perfectly suited for text classification tasks. From a socio-economic perspective, the data proves that public sentiment strongly leans positive toward the de-dollarization strategy.

5. Conclusion

The deployment of the Lexicon-based framework over 659 YouTube comments successfully mapped the public response to the Panda Bonds issuance policy. The digital discourse is explicitly dominated by positive sentiment at 51.9% (342 comments), followed by neutral sentiment at 29.0% (191 comments), and negative sentiment at 19.1% (126 comments). While the absolute majority of the public actively supports national economic stabilization and currency diversification (de-dollarization), the underlying public anxieties absorbed via frequent WordCloud keywords remain heavily fixated on long-term debt liabilities and geopolitical dependencies.

The hybrid classification model integrating automated lexicon labeling with a Linear kernel SVM classifier on 206 independent test instances demonstrated exceptional robustness. The model secured an overall accuracy of 87.86%, a precision of 88.79%, and an F1-score of 87.96%. Crucially, the model proved optimal at isolating public backlash, achieving its maximum precision of 0.97 within the negative sentiment class. Overall, this hybrid methodology is highly effective for processing Indonesian natural language within the financial policy domain on social media.

References

- [1] J. Guild, "The Indonesian state and the strategic use of foreign capital," *Pacific Rev.*, vol. 36, no. 5, pp. 1094–1119, 2023.
- [2] D. H. Bangkalang, "Opinion mining of regional heads in Indonesia using the support vector machine (SVM) method," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, pp. 1622–1627, 2024.
- [3] S. A. Nugraha, A. R. Go, C. Andreas, R. Limantara, I. Maryati, and A. Arjuna, "Public Sentiment on Mining and Energy Development in Indonesia: A Hybrid Machine Learning Perspective," in *2025 8th International Conference on Informatics and Computational Sciences (ICICoS)*, 2025, pp. 92–97.
- [4] P. Arsi and R. Waluyo, "Analisis sentimen wacana pemindahan ibu kota Indonesia menggunakan algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021.
- [5] A. Abdurokhim, "De-Dollarization Trends in Southeast Asia: Assessing Indonesia's Monetary Sovereignty," *Econ. Monet. J.*, vol. 1, no. 2, pp. 64–76, 2025.
- [6] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011.
- [7] 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 Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023.
- [8] V. Fitriyana, L. Hakim, D. C. R. Novitasari, and A. H. Asyhar, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," *J. Buana Inform.*, vol. 14, no. 01, pp. 40–49, 2023.
- [9] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *Komputek*, vol. 3, no. 2, pp. 52–63, 2019.
- [10] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, 2021.
- [11] N. S. Wardani, A. Prahutama, and P. Kartikasari, "Analisis sentimen pemindahan ibu kota negara dengan klasifikasi Naïve Bayes untuk model Bernoulli dan Multinomial," *J. Gaussian*, vol. 9, no. 3, pp. 237–246, 2020.
- [12] V. D. Setiawan, D. U. Iswavigra, and E. Anggiratih, "Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court's Decision Regarding the Minimum Age of Vice Presidential Candidates," *Sci. J. Informatics*, vol. 12, no. 3, pp. 397–406, 2025.
- [13] I. Iwandini, A. Triayudi, and G. Soepriyono, "Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor," *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 543–550, 2023.