Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

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

Sentiment Analysis of Public Opinion on Rupiah Redenomination Policy Using Support Vector Machine and SMOTE

Haidar Aslam[1], Haikal Nurul Barki[2], Adhi Prasetyo Wibowo[3], Faqih Al Araf [4], Abdul Hamid Musawir[5], Fuad Nur Hasan[6]*

[1,2,3,4,5,6]Prodi Studi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

aslamhaidar98@gmail.com[1], haikalbarki0@gmail.com[2], adhiprasetyow07@gmail.com[3], faqihalaraf51@gmail.com[4], Hamid03abd@gmail.com[5], fuad.fnu@bsi.ac.id[6*]

Abstract

The government's planned rupiah redenomination has generated a substantial wave of public opinion across social media platforms. This study aims to analyze public sentiment by examining comments on YouTube and classifying them into two categories: positive and negative. The data are collected through web scraping conducted on December 21, 2025, using the keyword "rupiah redenomination." Given the pronounced imbalance between negative and positive opinions, this study applies the Synthetic Minority Over-sampling Technique (SMOTE) to balance the class distribution within the training data. The research pipeline consists of text preprocessing, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), and classification using a linear-kernel Support Vector Machine (SVM). Experimental results indicate that the SVM model achieves an accuracy of 88.28%. The application of SMOTE is shown to effectively enhance the model's ability to identify the minority class, with the recall for positive sentiment reaching 0.71. Furthermore, the analysis reveals that public opinion is predominantly negative (83.93%), reflecting widespread concern regarding the potential economic implications of the policy.

Keywords: *Sentiment Analysis; Support Vector Machine; Rupiah Redenomination; SMOTE; YouTube.*

1. Introduction

The discourse surrounding the rupiah redenomination policy defined as the simplification of the currency's nominal value without diminishing public purchasing power has become a prominent topic of discussion in Indonesia over the past several years. Although monetary authorities emphasize that redenomination is intended to enhance transaction efficiency and strengthen the prestige of the national currency, the proposed transition frequently generates public concern, particularly regarding the risk of inflation or confusion with past currency reform policies such as sanering. This situation has led to a substantial surge of public opinion across various social media platforms, with YouTube being one of the primary sources. YouTube is selected as a data source due to users' freedom to provide extended comments that reflect genuine emotions and public perceptions in response to news videos or official government communications [1], [2].

Sentiment analysis of public comments plays a crucial role for governments and relevant institutions in monitoring opinion trends and developing more effective risk communication strategies to mitigate public anxiety [3], [4]. Nevertheless, textual data derived from social media presents inherent challenges, particularly due to its unstructured nature and the frequent occurrence of class imbalance. In the context of the redenomination policy, negative comments tend to be significantly more prevalent than positive ones. This imbalance can introduce bias into conventional classification models, leading them to favor the majority class (negative) while inadequately capturing patterns associated with the minority class (positive) [5], [6].

To address these challenges, this study employs a Support Vector Machine (SVM) algorithm in conjunction with the Synthetic Minority Over-sampling Technique (SMOTE). SVM is selected due to its strong capability in handling high-dimensional data, such as Indonesian-language text [6]. Meanwhile, SMOTE is applied to balance the training data distribution by generating synthetic samples for the minority class, a method that has been shown in prior studies to significantly improve Recall and F1-Score performance metrics [7], [8], [9]. This research is expected to contribute to the academic literature by demonstrating the application of a hybrid SVM–SMOTE approach to a national monetary policy issue, as well as providing empirical insights for policymakers regarding public perceptions of rupiah redenomination in Indonesia.

2. Research Methods

This study is conducted systematically through several main stages, ranging from data collection to the evaluation of the classification model. The overall research workflow is illustrated in Figure 1.

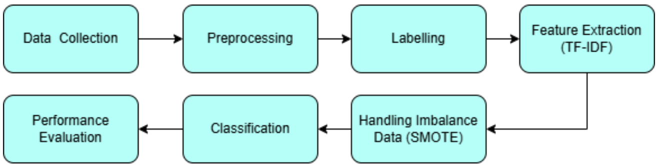

Fig. 1. Research methods

2.1. Data Collection

The data utilized in this study consist of public comments collected from the YouTube social media platform. Data acquisition was conducted through web scraping using the YouTube Data API v3 on videos related to the "redenominasi rupiah" policy. YouTube data are considered relevant as they effectively capture open and real-time public opinions [1], [2]. In total, 3,752 raw comment entries were successfully obtained for analysis.

2.2. Preprocessing

Raw textual data obtained from social media are generally unstructured and contain a substantial amount of noise. Therefore, a preprocessing stage is applied to clean and standardize the data prior to classification modeling [9]. This process begins with case folding, where all characters are converted to lowercase to ensure consistency. Subsequently, a cleaning step is performed to remove numbers, punctuation marks, symbols, excessive whitespace, and URLs. The text is then tokenized by segmenting sentences into individual word units. Normalization is applied to correct slang or non-standard expressions into formal Indonesian, with particular attention to economic terminology. Next, stopword removal is conducted to eliminate common words that do not contribute meaningful information to sentiment interpretation, such as conjunctions and prepositions. Finally, stemming is carried out to reduce inflected words to their root forms using the Sastrawi library.

2.3. Labelling (lexicon-based)

Data labeling is performed automatically using a lexicon-based approach. Each comment is categorized into two sentiment classes, namely Positive and Negative. This labeling process is conducted prior to the application of machine learning models in order to provide an initial basis for sentiment classification [5] .

2.4. Feature Extraction (TF-IDF)

The textual data are transformed into numerical representations using the Term Frequency–Inverse Document Frequency (TF-IDF) method. TF-IDF assigns weights to each term based on its frequency within a document relative to its occurrence across the entire dataset. This approach has been widely recognized as effective in capturing term importance for text classification tasks [4], [3].

2.5. Handling Imbalanced Data (SMOTE)

Following the feature extraction stage, a substantial class imbalance was identified, with negative sentiments accounting for 83.93% of the data and positive sentiments comprising only 16.07%. To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. SMOTE operates by generating new synthetic samples for the minority class (positive) based on its k-nearest neighbors, thereby producing a balanced class distribution in the training dataset at a 50:50 ratio [7], [8], [10].

2.6. Classification (Support Vector Machine)

The Support Vector Machine (SVM) algorithm is employed to construct the classification model. SVM operates by identifying an optimal hyperplane that separates the two sentiment classes while maximizing the margin between them. In this study, a linear kernel is adopted due to its effectiveness in handling high-dimensional text data [6], [4]. The dataset is partitioned into 80% for training and 20% for testing purposes.

2.7. Performance Evaluation

Model performance is evaluated using a confusion matrix to compute Accuracy, Precision, Recall, and F1-Score. The primary emphasis of the evaluation is to assess the extent to which SMOTE enhances the model's ability to identify positive sentiment as the minority class, while maintaining overall classification accuracy [5], [10].

3. Result and Discussion

This section presents the results of the experimental procedures conducted in this study, including an analysis of data distribution, keyword visualization, and a comparative evaluation of SVM model performance before and after the application of SMOTE.

3.1. Lexicon based labeling results

The lexicon-based labeling process yields a sentiment distribution that reflects the public's initial perception of the rupiah redenomination discourse. As illustrated in Figure 2, the labeling results reveal a pronounced class imbalance. Of the total comments analyzed, 3,149 instances (83.93%) are classified as negative sentiment, while only 603 instances (16.07%) fall into the positive category. The predominance of negative sentiment suggests a collective public concern regarding economic stability and the potential for price increases (inflation), which are often misinterpreted as being analogous to past currency policies such as sanering.

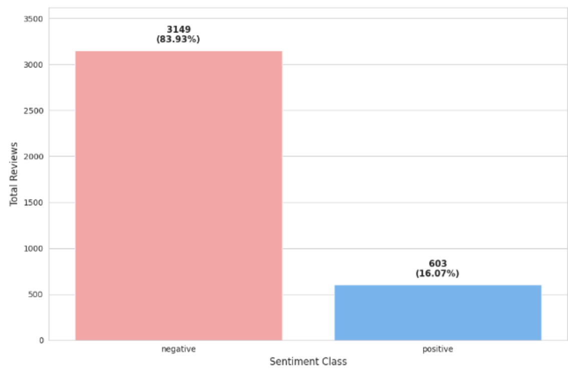

Fig. 2. Result labelling lexicon based.

3.2. Text Preprocessing and wordcloud Analysis

The preprocessing stage is shown to be effective in reducing noise within the YouTube text data. A comparison of the WordCloud visualizations before and after preprocessing demonstrates the removal of non-informative terms (stopwords) and the emergence of more substantive keywords. Within the negative sentiment group, frequently occurring terms such as "harga", "naik", "susah", and "takut" dominate, reflecting public anxiety related to economic conditions.

Fig. 3. Wordcloud before preprocessing.

In contrast, the positive sentiment category features keywords including "simpel", "maju", "dukung", and "transaksi" which indicate public expectations regarding improved efficiency in national currency administration.

Fig. 4. Wordcloud after preprocessing.

3.3. Performance Evaluation of SVM with SMOTE

The evaluation is conducted using an 80% training and 20% testing data split. Prior to the application of SMOTE, the SVM model exhibits a bias toward the majority (negative) class, resulting in a low recall for positive sentiment. After data balancing is applied to the training set using SMOTE, the model demonstrates a more equitable performance across classes.

The SVM–SMOTE model achieves an overall accuracy of 88.28%. Although this value is slightly lower than the pseudo-accuracy obtained without SMOTE, a closer examination of the evaluation metrics reveals an overall improvement in model quality. As illustrated in Figure 5, the recall for the positive class reaches 0.71, representing a substantial increase compared to the pre-balancing condition. Moreover, the macro-average F1-score of 0.80 indicates that the model maintains stable and proportional performance in recognizing both sentiment classes.

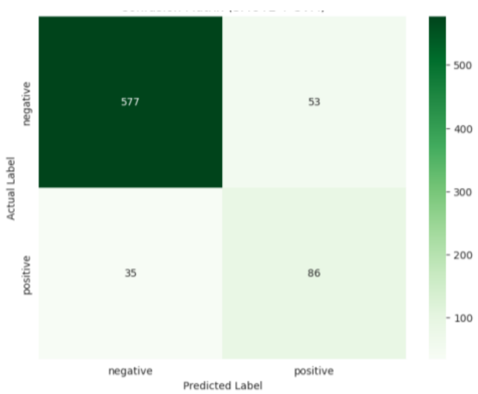

Fig. 5. Confusion matrix.

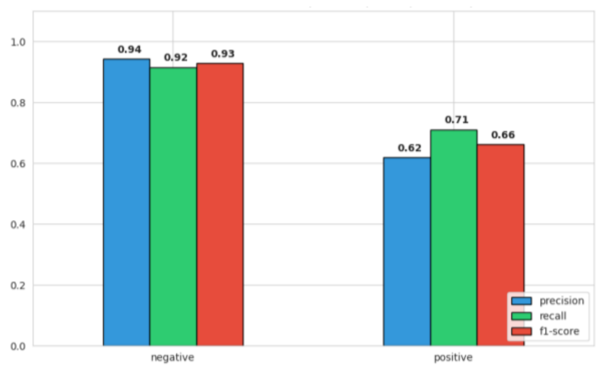

Fig. 6. Evaluation matrix

3.4. Comparison with Previous Studies

The findings of this study demonstrate competitive performance when compared with prior research. The achieved accuracy of 88.28% exceeds the results reported by Lalupanda, who obtained an accuracy of 80.00% in sentiment analysis of AI application reviews. These results indicate that the combination of TF-IDF features with unigram and bigram representations, a linear-kernel SVM classifier, and the SMOTE technique is highly effective when applied to heavily imbalanced public opinion datasets related to national monetary policy.

Table 1. Classification Report (Precision, Recall, F1-score)

Class	Precision	Recall	F1-Score
Positive	0.62	0.71	0.66
Negative	0.94	0.92	0.93

4. Conclusion and Suggestions

This study concludes that the Support Vector Machine (SVM) algorithm combined with the Synthetic Minority Over-sampling Technique (SMOTE) is effective in classifying sentiment related to the rupiah redenomination, achieving an accuracy of 88.28%. The adoption of this hybrid approach is shown to be critical in addressing the imbalance of public opinions on YouTube, where negative sentiment accounts for 83.93% of the data. The application of SMOTE substantially enhances model performance, particularly by improving the recall of the positive class to 0.71, thereby reducing bias toward the majority class.

The sentiment analysis results reveal public concerns regarding potential inflationary effects and broader economic impacts of the policy, suggesting that relevant authorities should enhance transparency and effectiveness in policy communication. For future research, further development may involve expanding data sources to other platforms such as Twitter or Instagram, as well as exploring deep learning approaches, including IndoBERT, to better capture the complex contextual nuances of the Indonesian language.

References

[1] T. Muhayat, A. Fauzi, and D. J. Indra, “Analisis Sentimen Terhadap Komentar Video Youtube Menggunakan Support Vector Machines”.

[2] A. K. Subagyo and S. Wijayanto, “Analisis Sentimen Komentar Youtube Potensi Gempa Megathrust Di Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” Aug. 2025.

[3] S. Ramadhani Putri *et al.*, “Sistemasi: Jurnal Sistem Informasi Analisis Sentimen Publik terhadap Nadiem Makarim sebagai Mendikbudrisktek menggunakan Support Vector Machine (SVM).” [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[4] 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,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.

[5] F. Destiyanti, A. Id Hadiana, F. Rakhmat Umbara, and U. Jenderal Achmad Yani Jl Terusan Jenderal Jenderal Sudirman, “Penerapan Metode Support Vector Machine dan SMOTE untuk Klasifikasi Sentimen Publik Terhadap Polisi Republik Indonesia,” vol. 8, no. 1, pp. 1–15, Apr. 2024.

[6] D. W. Syahputra, B. Rahayudi, and L. Muflikhah, “Analisis Sentimen Twitter terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat menggunakan Metode Support Vector Machine,” Mar. 2022. [Online]. Available: http://j-ptiik.ub.ac.id`

[7] T. A. Dewi and E. Mailoa, “PERBANDINGAN IMPLEMENTASI METODE SMOTE PADA ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM ANALISIS SENTIMEN OPINI MASYARAKAT TENTANG MIXUE,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 849–855, Sep. 2023, doi: 10.35870/jimik.v4i3.289.

[8] N. Fajriyah, N. T. Lapatta, D. W. Nugraha, and R. Laila, “IMPLEMENTASI SVM DAN SMOTE PADA ANALISIS SENTIMEN MEDIA SOSIAL X TERHADAP PELANTIKAN AGUS HARIMURTI YUDHOYONO,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1359–1370, Mar. 2025, doi: 10.29100/jipi.v10i2.6246.

[9] N. S. Sediatmoko, Y. Nataliani, and I. Suryady, “Sentiment Analysis of Customer Review Using Classification Algorithms and SMOTE for Handling Imbalanced Class,” Aug. 2024.

[10] Y. A. Singgalen, “Performance Evaluation of Sentiment Classification Models: A Comparative Study of NBC, SVM, and DT with SMOTE,” *Media Online*, vol. 4, no. 5, pp. 2539–2548, 2024, doi: 10.30865/klik.v4i5.1827.