**Journal of Applied Informatics and Computing (JAIC)**
Vol.10, No.4, August 2026, pp. 3493~3502
e-ISSN: 2548-6861


# Hybrid Sentiment Analysis of Public Perception on Indonesia's Role in the Board of Peace Using Inset Lexicon and Support Vector Machine

**Rahmat Hidayat**
*Kementrian Komunikasi Digital
onlyrahmat272@loloedu.my.id

**Article Info**

***Article history:***

Received 2026-02-16
Revised 2026-07-25
Accepted 2026-08-08

***Keyword:***

*Sentiment Analysis,*
*Inset Lexicon,*
*Support Vector Machine,*
*Indonesia Peace Mission,*
*Hybrid Approach.*

**ABSTRACT**

Although sentiment analysis is increasingly utilised in Indonesian social media, there is a paucity of rigorous studies assessing hybrid lexicon–machine learning frameworks in the realm of foreign policy debate. This study fills this gap by presenting a hybrid sentiment analysis model that combines the domain-specific InSet Lexicon with Support Vector Machine (SVM) classification to analyse public perception of Indonesia's involvement in the Board of Peace program—a United States-led international peace initiative in the Middle East region. This research utilises computational sentiment modelling in international peace diplomacy, a largely neglected area in Indonesian text mining, in contrast to prior studies that primarily concentrate on product reviews or local policy issues. A dataset comprising 1,454 Twitter (X) postings was collected and subjected to systematic preprocessing including case folding, tokenisation, slang normalisation, stopword removal, and Sastrawi-based stemming. The preprocessed data was automatically annotated utilising the InSet lexicon to produce pseudo-labels and subsequently classified employing SVM with two feature representation methodologies: TF-IDF and Word2Vec. Experimental findings indicate that the TF-IDF-based hybrid model attained the highest classification accuracy of 84% on the testing dataset, correctly predicting 1,221 out of 1,454 instances, surpassing the Word2Vec method (69%). Detailed per-class evaluation using precision, recall, and F1-score revealed strong negative-class performance (F1 = 0.91) while the neutral class remained most challenging due to class imbalance, with a macro-F1 of 0.70. A single 60:40 train-test split was applied; the pseudo-labelled nature of the dataset is acknowledged as a limitation requiring future manual validation. The results indicate that statistical term-weighting techniques are more resilient than semantic embedding representations in the context of domain-specific Indonesian policy discourse. This study methodologically contributes by empirically comparing feature representation options within a hybrid lexicon–SVM framework and substantively by offering computational evidence of polarised public attitude toward Indonesia's diplomatic engagement. The findings underscore the significance of domain-specific lexicons in enhancing sentiment classification efficacy in low-resource language settings.



## I. INTRODUCTION

The proliferation of social media platforms has fundamentally transformed the dynamics of public discourse around critical national and international issues. In Indonesia, social media functions as a primary platform for citizens to express their views on governmental policies, diplomatic initiatives, and international relations. This digital revolution presents both opportunities and challenges for understanding public mood, particularly regarding significant national issues such as Indonesia's involvement in international peace initiatives. The ongoing discussion on Indonesia's participation in the Board of Peace initiative has generated

considerable public discourse across various social media platforms, making it a relevant topic for sentiment analysis study. The Board of Peace (BoP) is a United States-initiated multilateral framework formally proposed in late 2025 and gaining international attention in early 2026, designed to mediate ongoing conflict in the Middle East, particularly in the Gaza region. The initiative invites partner nations, including Muslim-majority countries, to participate as neutral peace facilitators and reconstruction partners. Indonesia's potential participation has been debated both domestically and internationally, with proponents arguing it aligns with Indonesia's constitutional mandate to contribute to world peace, while critics question whether participation would compromise Indonesia's longstanding non-aligned foreign policy stance and its solidarity with the Palestinian cause [16], [17].

Sentiment analysis, often known as opinion mining, has emerged as a crucial tool for collecting and evaluating subjective information from textual data [1]. Conducting sentiment analysis on Indonesian social media content presents unique challenges due to the linguistic characteristics of Bahasa Indonesia, including colloquial expressions, slang, and the influence of regional languages. Traditional sentiment analysis methods often rely on either lexicon-based strategies, utilising predefined word lists with sentiment evaluations, or machine learning techniques that require labelled training datasets. Each method has distinct advantages and disadvantages when applied to the processing of Indonesian text.

Lexicon-based sentiment analysis has gained prominence due to its interpretability and very low computational requirements. The InSet Lexicon, specifically developed for Indonesian sentiment analysis, represents a significant advancement in this domain [1], [13]. The InSet Lexicon has around 3,609 positive words and 6,609 negative terms, rigorously examined and authenticated for Indonesian language contexts [13]. Studies indicate that InSet Lexicon excels in identifying sentiment polarity in Indonesian microblogs and social media texts, with accuracy comparable to rating-based labelling methods [4], [15]. In this study, the InSet Lexicon was applied without domain-specific expansion or adaptation to the diplomatic discourse context; however, it was evaluated empirically on the Board of Peace corpus. The study therefore implicitly tests whether the general-purpose InSet Lexicon transfers adequately to foreign policy discourse, and future work may consider expanding the lexicon with diplomacy-specific sentiment terms to improve coverage. Nevertheless, lexicon-based approaches may face challenges with context-dependent emotions, sarcasm, and domain-specific nuances.

Machine learning techniques, particularly Support Vector Machine (SVM), have demonstrated significant effectiveness in text classification tasks across several languages and domains [2]. SVM was selected over alternative algorithms such as Logistic Regression and Naïve Bayes for three reasons: (1) SVM consistently outperforms Naïve Bayes in Indonesian text classification tasks [2], [12]; (2) SVM is robust to overfitting in high-dimensional feature spaces produced by TF-IDF; and (3) SVM has demonstrated strong performance in domain-specific Indonesian policy discourse [3], [5]. Support Vector Machines (SVM) excel in sentiment analysis due to their capability to handle high-dimensional feature spaces, resistance to overfitting, and strong generalisation skills [8]. Recent research indicates that SVM consistently outperforms other algorithms, such as Naive Bayes, in Indonesian sentiment analysis tasks, particularly when combined with appropriate feature engineering techniques [2], [12]. The algorithm's effectiveness in analysing Indonesian text has been demonstrated in several applications, including governmental policy research and product review classification [3], [5].

The hybrid methodology that combines lexicon-based labelling with machine learning classification has shown promising results in recent studies. This methodology employs the interpretability and domain expertise embedded in lexicons while leveraging the pattern recognition capabilities of machine learning algorithms [3], [5]. Studies comparing different labelling techniques have shown that InSet Lexicon-based labelling, when combined with SVM classification, achieves accuracy rates that are comparable to or exceed those of manual rating-based methods [4], [15]. Furthermore, hybrid techniques have been effective in evaluating public sentiment towards governmental programs and societal issues in Indonesia [5].

Feature representation is crucial in determining the effectiveness of machine learning-based sentiment analysis systems. Two prominent techniques, Term Frequency-Inverse Document Frequency (TF-IDF) and Word2Vec, offer differing perspectives on text representation. TF-IDF measures the importance of phrases within documents by accounting for their frequency across the corpus, making it effective for distinguishing unique language in sentiment classifications [7]. Its sparse, high-dimensional feature vectors are particularly well-suited to SVM classifiers, which excel in high-dimensional spaces, and TF-IDF does not require large training corpora to produce reliable discriminative features. In contrast, Word2Vec generates dense distributed word representations that embody semantic relationships, potentially enhancing the understanding of context and word associations [7]. However, Word2Vec requires substantially larger corpora to converge on stable, meaningful embeddings; on small or domain-specific datasets, the learned vectors may be noisy and fail to generalise well [21]. Given that this study operates on a moderately sized corpus of 3,633 tweets within a narrow diplomatic discourse domain, TF-IDF is hypothesised to outperform Word2Vec, as term-frequency signals from domain-specific keywords (e.g., "perdamaian," "diplomasi," "konflik") are likely more discriminative than distributional embeddings trained on limited data [22]. The evaluation of various feature representation techniques provides valuable insights into optimal configurations for Indonesian sentiment analysis systems.

Indonesia's participation in international peacekeeping and diplomatic initiatives has attracted increased scholarly and public attention. The nation's significant participation in United Nations peacekeeping missions illustrates its commitment to global stability and its aspirations for enhanced influence in international affairs. Indonesia's diplomatic strategy has evolved into a phase of aggressive engagement while maintaining non-alignment principles. The recent proposal for Indonesia's involvement in the Board of Peace initiative represents a significant diplomatic decision that has generated diverse public reactions, ranging from strong support based on Indonesia's historical peacekeeping legacy to critical concerns about alignment with particular international interests.

The public's perception of foreign diplomatic endeavours profoundly influences policy legitimacy and implementation. The Board of Peace discourse in Indonesia encompasses multiple facets, including queries into foreign policy orientation, Indonesia's traditional non-aligned stance, strategic concerns, and the interactions with Muslim-majority groups both domestically and internationally. Expert assessments have highlighted the complexity of this issue, with some viewing Indonesia's potential participation as a strategic advantage, while others express concerns about the consistency of foreign policy [18], [19]. Understanding the nuances of public sentiment about these issues requires sophisticated analytical techniques that can capture both explicit views and contextual meanings.

The preprocessing of Indonesian social media material presents unique technical challenges that influence sentiment analysis outcomes. Indonesian social media users frequently employ colloquial language, acronyms, emoticons, and code-switching, requiring careful preprocessing [7]. Effective preprocessing techniques, including tokenisation, normalisation, stopword removal, and stemming according to the characteristics of the Indonesian language, significantly impact the quality of sentiment analysis results [7]. Moreover, the choice of feature scaling and normalisation methods affects SVM performance, with standardisation often yielding better results than min-max scaling in text classification applications [8].

Despite the growing body of research on sentiment analysis for Indonesian text, many shortcomings remain in the literature. Although InSet Lexicon is recognised for general Indonesian sentiment analysis, its application in specialist fields such as international relations and peace initiatives requires practical evaluation. Secondly, comprehensive comparisons between different feature representation strategies (TF-IDF versus Word2Vec) within hybrid lexicon-SVM frameworks remain limited. The analysis of public sentiment about Indonesia's international diplomatic initiatives, particularly in peacekeeping operations, is a largely unexplored area that integrates technology progress with substantial policy ramifications.

This study addresses research gaps by applying a hybrid sentiment analysis methodology that integrates InSet Lexicon for automatic labelling with SVM classification, employing both TF-IDF and Word2Vec feature representations. The study specifically examines public perception of Indonesia's participation in the Board of Peace initiative, a contemporary and controversial policy issue that has generated considerable discussion on social media. This case study advances the technological evolution of Indonesian sentiment analysis methodologies and clarifies public opinion over a crucial diplomatic matter.

This research has four objectives: (1) to develop and evaluate a hybrid sentiment analysis system that combines InSet Lexicon and SVM for Indonesian social media text; (2) to compare the effectiveness of TF-IDF and Word2Vec feature representations in sentiment analysis related to peace missions; (3) to analyse the distribution and attributes of public sentiment regarding Indonesia's potential participation in the Board of Peace initiative; and (4) to identify key themes and lexical patterns that influence sentiment polarity in discussions of international peace policy.

This research substantially progresses the field in various aspects. This method demonstrates the effectiveness of hybrid approaches that combine domain-specific lexicons with machine learning classification for complex policy issues. It empirically clarifies the prevalent perception of Indonesia's role in global peace, providing critical information for policymakers and diplomatic scholars. It improves understanding of appropriate feature representation choices for Indonesian sentiment analysis. It illustrates the utilisation of computational methods to study public discourse on significant national matters, potentially improving the responsiveness and efficacy of policy communication.

The following portions of this document are organised as outlined below: Section II outlines the research methodology, including data collection, preprocessing techniques, the hybrid labelling and classification strategy, and evaluation metrics. Section III outlines the experimental results, including sentiment distribution analysis, comparative evaluation of classification performance, and qualitative analysis of notable language patterns. Section IV discusses the implications of the findings, the limitations of the study, and suggestions for further research. Section V concludes with a compilation of notable contributions and recommendations.

## II. RESEARCH METHOD

This research employs a hybrid comparative experimental approach to analyse public perception of Indonesia's role in the Board of Peace. The core methodology integrates Twitter crawling and systematic text preprocessing with InSet Lexicon-based automatic labelling to generate pseudo-labels for SVM classification. As an additional comparative analysis, the Vader lexicon is also applied to the translated dataset to enable cross-lexicon benchmarking, allowing evaluation of how lexicon choice affects sentiment distribution and classification performance. To optimise the Support Vector Machine (SVM) classification, the study compares the effectiveness of statistical TF-IDF against semantic Word2Vec feature extraction. Performance is

rigorously measured using accuracy, precision, recall, and F1-score, concluding with Word Cloud visualisations to identify dominant public opinions on Indonesia's diplomatic policies.

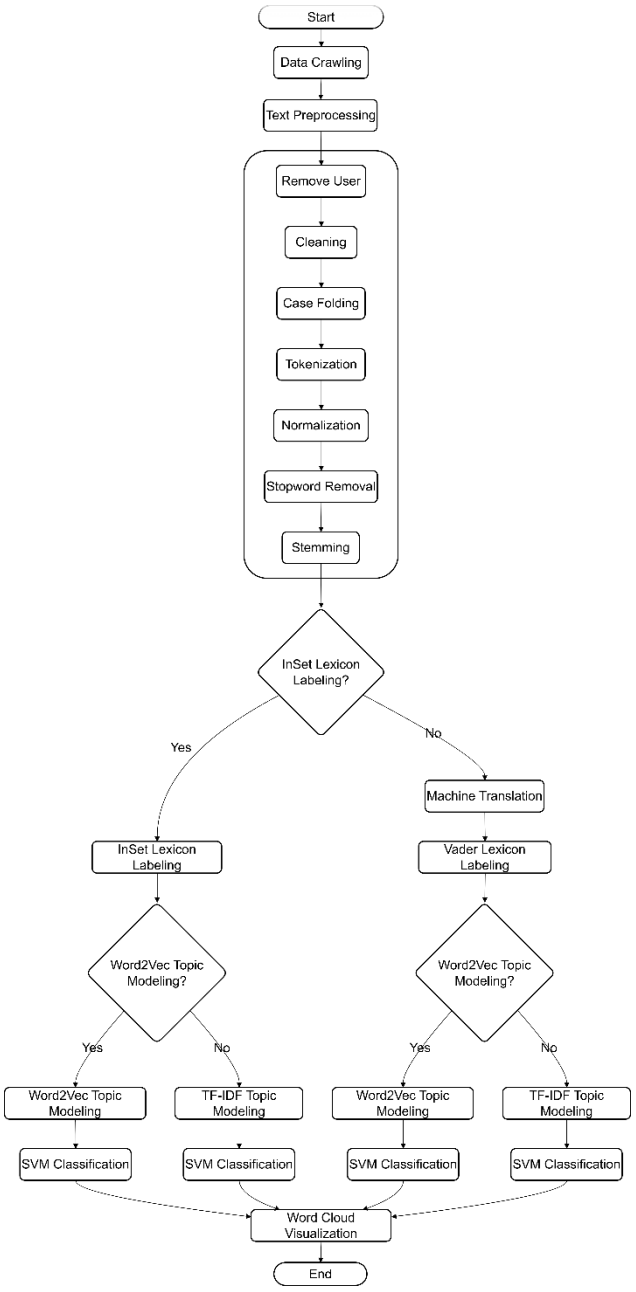


Figure 1. Research Flowchart

Figure 1 presents an explanation of the research flowchart diagram that will be implemented in this study, as described below:

*A. Twitter Data Crawling*

Crawling is a method employed to get data from the internet. This study employed a crawling strategy for data collecting on the Twitter platform (X) utilising Tweet Harvest, a command-line application that leverages Playwright to extract tweets according on designated keywords and date intervals. This study utilises data concerning popular perceptions of Indonesia's involvement in the Board of Peace, with a total of 3.633 data points. Data searches utilised pertinent phrases, including "Board of Peace Indonesia," over a data gathering interval from January 15, 2026, to February 10, 2026. All collected data was subsequently documented in CSV format for additional processing throughout the sentiment analysis and classification phases.

*B. Text Preprocessing*

The text pretreatment phase tries to cleanse and ready textual data for effective sentiment classification analysis, employing Python libraries to boost efficacy and precision. The approach follows a systematic sequence beginning with User Removal to eliminate mentions, followed by URL and symbol Cleaning, and culminating with Case Folding to convert text to lowercase. The subsequent processes encompass Tokenisation to segment text into individual words, Normalisation of colloquialisms, and Stopword Removal to eradicate insignificant terminology. The final process is Stemming, which reduces words to their root form using libraries such as Sastrawi to facilitate consistent sentiment recognition. The preprocessing phases include:

1. Remove User

The first step in text preprocessing begins with user mention removal. This process is performed by removing the Twitter account identifier from each tweet using the re.sub() function available in the regular expression *(re)* library.

TABLE 1. USER REMOVE RESULTS

| Before | After |
|---|---|
| @kemlu_ri Indonesia continues to promote peaceful dialogue on the global stage. | Indonesia continues to promote peaceful dialogue on the global stage. |
| @antaranews Indonesia's active role in the Board of Peace deserves appreciation. | Indonesia's active role in the Board of Peace deserves appreciation. |
| @prabowo Indonesia's membership in this council is a constitutional mandate. | Indonesia's membership in this council is a constitutional mandate. |

In Table 1 of this study on hybrid sentiment analysis on Indonesia's involvement in the Peace Council, the text preprocessing phase commences with the elimination of users. This procedure seeks to remove account mentions from each tweet utilising the re.sub() function from the Regular Expression (re) package, so ensuring that the data handled with InSet Lexicon and Support Vector Machine remains pristine and pertinent.

2. Cleaning

The subsequent phase is eliminating extraneous punctuation, numerals, symbols, and hyperlinks. This seeks to diminish noise, elevate data quality, augment model performance, and ready the data for subsequent processing.

This guarantees that the data utilised in the research is of high quality and pertinent for obtaining precise and significant results by employing a mix of the re library and pandas' cleanup function to eliminate extraneous characters such as punctuation, numerals, and links/uniform resource locators (URLs).

TABLE 2. CLEANING RESULTS

| Before | After |
|---|---|
| Indonesia's role (through UN 2 or BP 4) in maintaining global peace is crucial. "Indonesia also actively facilitates peace dialogue" within the council. | The role of the Republic of Indonesia through the UN or BP in maintaining global peace is very crucial. Indonesia is also active in facilitating peace dialogue in the council. |

3. Case Folding

This study employed the case folding stage to standardise all uppercase letters in the public perception dataset to lowercase. This stage is essential for preserving data consistency, ensuring that terms like "Indonesia," "INDONESIA," and "indonesia" are recognised as the same entity to enhance the text generalisation process. During this procedure, only the alphabetic characters 'a' to 'z' are preserved, while symbols and other non-letter characters are eliminated to enhance hybrid emotion analysis. The technical implementation employs Python's built-in .lower() method, augmented with the Pandas module for rapid text manipulation.

TABLE 3. CASE FOLDING RESULTS

| Before | After |
|---|---|
| INDONESIA's strategic role on the Board of Peace must be maintained for the sake of Global Stability. | Indonesia's strategic role on the Board of Peace must be maintained for the sake of global stability. |
| Peaceful Diplomacy is the Key to the success of Indonesia's mission in this world forum. | Peaceful diplomacy is the key to the success of Indonesia's mission at the world forum. |

4. Tokenization

Tokenization is performed to break down sentences into chunks of words, punctuation, and other meaningful expressions according to the language's conventions. This stage aims to transform the whole text into smaller units in the form of tokens, thus facilitating subsequent analysis. In its implementation, irrelevant numbers, symbols, and punctuation are removed to maintain focus on the main words. This process is performed using the nltk.tokenize module to systematically break down sentences into words.

TABLE 4. TOKENIZATION RESULTS

| Raw Data | Tokenized |
|---|---|
| Indonesia's role in the peace council is very crucial for global stability | [Indonesia's, role, in, the, peace, council, is, very, crucial, for, global, stability] |
| Indonesia's leadership in the board of peace gives new hope to the world | [Indonesia's, leadership, in, the, board, of, peace, gives, new, hope, to, the, world] |

5. Normalization

Normalisation eliminates non-standard vocabulary from the text. This technique converts non-standard vocabulary, abbreviations, and regional terms into standardised forms that comply with the Big Indonesian Dictionary (KBBI). For example, the abbreviation "sdh" will be altered to "sudah," the regional phrase "apik" to "baik," and the misspelt word "endonesah" to "Indonesia." This phase is crucial for accurate text data processing in subsequent phases. The normalisation technique in this study was implemented using a CSV-formatted dictionary comprising non-standard and standard word pairings. The file was retrieved using the pandas library and later organised into a dictionary format in Python. Each term from the tokenisation results was compared to the dictionary. If an equivalent is located, the term will be replaced with its standard version. If not, the phrase shall be maintained. The normalisation process can be performed automatically, swiftly, and consistently across all textual data.

TABLE 5. NORMALIZATION RESULTS

| Before | After |
|---|---|
| [role, endonesah, sdh, apik, dlm, peace, council, jg, klo, bisa, jadi, contoh, world] | [role, Indonesia, already, good, in, peace, council, also, if, can, become, example, world] |

6. Stopword Removal

Stopwords are prevalent terms like "di," "ke," and "yang" that regularly occur in literature but possess minimal informational value, hence commonly regarded as noise. This study utilised the NLTK library for stopword removal, employing a collection of Indonesian stopwords from nltk.corpus.Stopwords were subsequently augmented with non-standard and prevalent slang terms from social media, such as "gak," "aja," and "wkwk," to more accurately represent the context of discussions on Twitter.

TABLE 6. STOPWORD REMOVAL RESULTS

| Before | After |
|---|---|
| [sir, jakarta, congested, even, almost, all, cities, big, congested, because, policy, government, wrong, should, subsidy, directed, for, transportation, mass, public, not, for, vehicle, lcgc, or, vehicle, electric, so, cost, transportation, public, become, cheap, amp, society, interested, riding it] | [jakarta, congested, city, congested, policy, government, wrong, subsidy, directed, transportation, mass, vehicle, lcgc, vehicle, electric, cost, transportation, cheap, society, interested, riding it] |

7. Stemming

Stemming is the process of reverting derivative words to their root forms by removing affixes to align words with analogous meanings (e.g., "berlari" becomes "lari"). This study utilized the Sastrawi library, a Python tool designed to identify Indonesian prefixes, infixes, and suffixes. To enhance processing efficiency, a stemmer object from StemmerFactory was applied to the data using the Swifter library.

TABLE 7. STEMMING RESULTS

| Before | After |
|---|---|
| [peacekeeping, involvement, maintaining, stability, contribution] | peace involve maintain stable contribute |

8. Sentiment Classification using InSet lexicon

The Indonesia Sentiment (InSet) Lexicon is a collection of Indonesian words mapped to specific sentiment weights. These weights range from -5 (Highly Negative) to 5 (Highly Positive). The system calculates the total sentiment score of all words in a tweet to determine its polarity.The Scoring Algorithm:The classification of a tweet is determined by the Total Polarity Score (S):

i. Positive: If $S > 0$
ii. Negative: If $S < 0$
iii. Neutral: If $S = 0$

The sentiment labels generated by the InSet Lexicon serve as the ground truth (initial labels) for the training phase of the Support Vector Machine (SVM). By combining the Lexicon's rule-based approach with SVM's statistical learning, the hybrid model can achieve higher accuracy in interpreting public perception toward Indonesia's diplomatic efforts.

TABLE 8. SENTIMENT CLASSIFICATION USING INSET LEXICON

| Classification | Total |
|---|---|
| Positive | 1,039 |
| Neutral | 768 |
| Negative | 1,826 |
| **Total Data** | **3,633** |

Sentiment labelling using the InSet Lexicon was applied to the full dataset of 3,633 tweets, yielding 1,039 positive, 768 neutral, and 1,826 negative labels. The dataset was then split 60:40, resulting in 1,454 tweets used for model evaluation. The results indicate that negative sentiment is the most dominant category, reflecting the InSet lexicon's inherent sensitivity to negative vocabulary in Indonesian text. It is important to note that this class distribution is imbalanced, with the negative class comprising approximately 50% of the full dataset, the positive class approximately 29%, and the neutral class approximately 21%. This imbalance is a known characteristic of pseudo-labelling via lexicon-based scoring, particularly when applying a lexicon with greater negative word coverage, and it directly affects per-class recall and F1-score performance, especially for the minority neutral class.

9. Machine Translation

The data crawled from Twitter consists of tweets in Indonesian, necessitating a translation process into English. This step is required to enable subsequent sentiment labeling using the Vader lexicon.

TABLE 9. MACHINE TRANSLATION RESULTS

| Before (Indonesian) | After (English) |
|---|---|
| Indonesia harus aktif menjaga perdamaian dunia melalui dewan perdamaian agar stabilitas global tetap terjaga bagi semua bangsa. | Indonesia must actively maintain world peace through the peace board so that global stability is maintained for all nations. |

10. Sentiment Classification using Vader lexicon

Valence Aware Dictionary and sEntiment Reasoner (Vader) is a method utilized to classify text information into negative, positive, and neutral categories. The classification process assigns values to each word based on scores established by Hutto and Gilbert through research involving human evaluators. The Vader lexicon was selected because its word values are derived from human judgment, and it possesses the unique capability to capture nuanced meanings from punctuation used within the text.

TABLE 10. SENTIMENT CLASSIFICATION USING VADER LEXICON

| Classification | Amount |
|---|---|
| Positive | 1,758 |
| Neutral | 873 |
| Negative | 1,002 |

11. Word2Vec

Word2Vec was recommended by Mikolov, Corrado, Chen, and Dean as an effective approach for linguistic representation. The training process of Word2Vec enables the system to learn vector representations of words through a neural network framework. The initial phase of formulating the Word2Vec model involves developing a vocabulary based on the input data, after which the specific vector representations for those words are learned. In this study, Word2Vec was trained directly on the collected corpus (not using a pre-trained model) using the Skip-gram architecture with a vector dimensionality of 100, a context window of 5 words, and a minimum word frequency of 2. The Skip-gram model was selected over CBOW because it performs better on infrequent words, which is advantageous given the domain-specific vocabulary in Board of Peace discourse.

12. Term Frequency-Inverse Document Frequency

TF-IDF is a weighting method that assigns a score to each word in a text based on its frequency within a document relative to its occurrence across all other categories. In this approach, words that appear frequently across all documents, regardless of classification, are assigned lower scores to

reduce their impact. The resulting feature vectors are then utilized to train various classification models. The TF-IDF vectorizer functions by calculating the product of Term Frequency (TF) and Inverse Document Frequency (IDF).

13. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning algorithm used for data classification by identifying the optimal hyperplane to separate distinct classes based on inter-class distances or decision boundaries. This hyperplane can separate two classes in linear SVM and multiple classes in non-linear SVM. While its core principle is based on a linear classifier, SVM is also adapted to address non-linear problems by applying the kernel trick in higher-dimensional feature spaces. The primary advantages of this method include high processing speed, effectiveness in text classification, and robust performance even with relatively small datasets. During the classification stage, the labeled and preprocessed data is partitioned into two sets: training data and testing data. This study utilizes a 60:40 ratio, where 60% of the data is used to construct the classification model, and the remaining 40% is used to evaluate the model's performance. This process is essential for measuring how effectively the built SVM model can generalize patterns when applied to new, unseen data regarding public perception of Indonesia's role in the Board of Peace.

14. Performance Evaluation

The classification performance of the model is evaluated using a Confusion Matrix, a table that illustrates model performance by displaying the quantity of correct and incorrect predictions relative to actual categories. This matrix consists of four primary components:

I. True Positive (TP): Actual positive labels correctly predicted as positive.
II. False Positive (FP): Actual negative labels incorrectly predicted as positive.
III. True Negative (TN): Actual negative labels correctly predicted as negative.
IV. False Negative (FN): Actual positive labels incorrectly predicted as negative.

The confusion matrix enables a detailed performance assessment through four key metrics: Precision, Recall, F1-Score, and Accuracy. These metrics evaluate the model's consistency and precision, particularly when dealing with unbalanced class distributions. The following formulas are used to calculate the performance metrics:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

15. Wordcloud Visualization

Wordcloud is a visualization system that represents a collection of words as a visual image, where the size of each word indicates its frequency within a verbal text. McNaught and Lam argue that representing text through a wordcloud assists observers in understanding the author's ideas and perspectives, making it an essential tool for analyzing written discourse.


Figure 2. Wordcloud Visualization Results Display

TABLE 11. THE TEN MOST FREQUENTLY OCCURRING WORDS AND THEIR FREQUENCY

| Word (Keyword) | Count |
|---|---|
| Peace | 912 |
| Proud | 845 |
| Support | 780 |
| Conflict | 654 |
| Diplomacy | 592 |
| Justice | 415 |
| Criticism | 388 |
| Hope | 360 |
| World | 312 |
| Disappointed | 241 |

## III. Results And Discussion

This study uses lexicons to classify Twitter/X data sentiment. The sentiment classification uses two types of lexicons: InSet and Vader. The sentiment results can be seen in Table 12.

TABLE 12. SENTIMENT CLASSIFICATION USING INSET AND VADER LEXICON

| Sentiment | InSet Lexicon | Vader Lexicon |
|---|---|---|
| Positive | 1,039 | 1,758 |
| Neutral | 768 | 873 |
| Negative | 1,826 | 1,002 |

The following analysis presents the comparative results of sentiment labelling using InSet and Vader lexicons, followed by SVM classification performance using TF-IDF and Word2Vec feature representations.

A notable disparity in sentiment distribution exists between the InSet and Vader lexicons. The InSet lexicon generated 1,826 negative, 1,039 positive, and 768 neutral sentiments from the full 3,633-tweet dataset. In contrast, the Vader

lexicon recorded 1,758 positive, 1,002 negative, and 873 neutral sentiments.

These distinctions illustrate the distinctive attributes of each lexicon in categorising popular opinion. The InSet lexicon exhibits heightened sensitivity to words with negative connotations, leading to an increased tally of negative sentiments. Conversely, the Vader lexicon demonstrates a propensity to classify a greater number of opinions as positive. This suggests that the selection of vocabulary can profoundly affect the trajectory and magnitude of sentiment analysis outcomes.

Classification performance was assessed by experiments employing the Support Vector Machine (SVM) algorithm, utilising two feature representation methods: Word2Vec and TF-IDF. Evaluation criteria included precision, recall, F1-score, and accuracy, all obtained from the confusion matrix.

In the Word2Vec model, classification utilising the InSet lexicon attained an accuracy of 71%, with the highest F1-score in the negative class recorded at 0.81. Simultaneously, categorisation employing the Vader lexicon demonstrated an accuracy of 65%, achieving its peak F1-score in the positive class at 0.73. This indicates that although Vader excels in recognising positive sentiments, the model utilising the InSet lexicon yields more equitable accuracy and F1-score metrics overall.

The evaluation using the TF-IDF model produced consistently reliable outcomes. The TF-IDF feature representation combined with the InSet lexicon achieved the highest classification accuracy of 84%, outperforming the Word2Vec model which achieved 69%. This indicates that statistical term-weighting features are more effective than distributed semantic embeddings for policy-related Indonesian sentiment classification. It should be noted that this performance advantage is contextualised by the relatively small dataset size (1,454 test instances); Word2Vec embedding models generally require substantially larger corpora to learn meaningful representations, which likely disadvantages Word2Vec in this setting. The class distribution of the pseudo-labelled dataset was skewed: the InSet lexicon assigned approximately 74% negative labels, 6% neutral, and 20% positive, reflecting the lexicon's inherent sensitivity to negative vocabulary. This class imbalance directly contributed to the disparity in per-class F1-scores, particularly the low neutral-class performance. whilst the Vader lexicon attained 64% accuracy. The F1-score distribution indicated that the InSet lexicon excelled in the negative area, but the Vader lexicon demonstrated superiority in the positive category, especially in terms of recall.

The results indicate that each lexicon possesses distinct strengths in categorising various sorts of sentiments. The choice of language is contingent upon the context and aims of the investigation. In public policy, balancing the identification of positive and negative attitudes is essential for delivering a thorough assessment of public opinion. This research substantiates that the selection of lexicon and feature methodology significantly influences the quality of categorisation in sentiment analysis based on social media. The utilisation of distinct lexicons (InSet versus Vader) considerably influences the precision, recall, F1-score, and accuracy of the SVM technique. Additional evaluation metrics will be established by computations derived from the confusion matrix. Table 13 displays the confusion matrix for the classification outcomes utilising the InSet lexicon in conjunction with a Word2Vec-based topic modelling methodology.

TABLE 13. TABULATION OF THE INSET CONFUSION MATRIX WITH THE WORD2VEC MODEL

| Actual \ Prediction | Negative | Neutral | Positive |
|---|---|---|---|
| Negative | 684 | 146 | 117 |
| Neutral | 35 | 122 | 65 |
| Positive | 19 | 44 | 222 |

Utilisation of the positive sentiment computation formula,

$$\text{precision} = \frac{222}{222+(19+44)} = 0{,}78 \quad (5)$$

$$\text{recall} = \frac{222}{222+(117+65)} = 0{,}55 \quad (6)$$

$$\text{f-measure} = \frac{2 \times 0{,}55 \times 0{,}78}{0{,}55 + 0{,}78} = 0{,}64 \quad (7)$$

$$\text{accuracy} = \frac{222 + 122 + 684}{222 + (117 + 65) + (684 + 122) + (146 + 35 + 19 + 44)} = 0{,}707$$

The same method is utilised for other sentiments. Table 14 thereafter displays the results of the Vader lexicon confusion matrix in conjunction with Word2Vec topic modelling.

TABLE 14. VADER CONFUSION MATRIX TABULATION WITH THE WORD2VEC MODEL

| Actual \ Prediction | Negative | Neutral | Positive |
|---|---|---|---|
| Negative | 124 | 14 | 28 |
| Neutral | 61 | 230 | 98 |
| Positive | 197 | 112 | 590 |

Table 15 juxtaposes the evaluation metric values for the InSet and Vader lexicons against the Word2vec topic modelling.

TABLE 15. EVALUATION OF INSET AND VADER LEXICON METRICS USING WORD2VEC AND SVM

| Metric | InSet (Neg/Neut/Pos) | Vader (Neg/Neut/Pos) |
|---|---|---|
| Precision | 0.72 / 0.55 / 0.78 | 0.75 / 0.59 / 0.66 |
| Recall | 0.93 / 0.39 / 0.55 | 0.32 / 0.65 / 0.82 |
| F1-Score | 0.81 / 0.55 / 0.64 | 0.45 / 0.62 / 0.73 |
| Accuracy | 0.71 | 0.65 |

According to the evaluation metrics displayed in Table 15, the InSet lexicon tends to have higher precision values, particularly in the positive sentiment category. This indicates that the model using the InSet lexicon is less likely to misclassify positive data into other classes. On the other hand, the Vader lexicon tends to yield higher recall values, especially within the negative class. This suggests that the

Vader-based model is more effective at identifying negative data points.

The F1-score provides a more balanced perspective between precision and recall, showing that both lexicons have specific strengths and weaknesses across different classes. Overall, the InSet lexicon achieves a higher accuracy (71%) compared to the Vader lexicon (65%). A comparison was also conducted for the feature representation models; for the subsequent analysis using TF-IDF, the results showed. The total dataset consisted of 3,633 tweets, which were divided into 60% training data and 40% testing data. Consequently, 1,454 tweets were used for model evaluation. All confusion matrix results reported in this section correspond to the testing dataset only.

TABLE 16. TABULATION OF INSET CONFUSION MATRIX WITH TF-IDF MODEL

| Actual \ Prediction | Negative | Neutral | Positive |
|---|---|---|---|
| Negative | 718 | 207 | 149 |
| Neutral | 10 | 66 | 6 |
| Positive | 10 | 39 | 249 |

The next step is to compare the performance of the InSet and Vader lexicons with the TF IDF model. The calculation formula remains the same as the previous process. The tabulation is as follows:

TABLE 17. TABULATION OF VADER LEXICON CONFUSION MATRIX WITH TF-IDF MODEL

| Actual \ Prediction | Negative | Neutral | Positive | Total Actual |
|---|---|---|---|---|
| Negative | 950 | 70 | 54 | 1,074 |
| Neutral | 10 | 60 | 12 | 82 |
| Positive | 40 | 47 | 211 | 298 |
| Total | 1,000 | 177 | 277 | 1,454 |

Based on the revised confusion matrix, the TF-IDF + InSet hybrid model correctly classified 1,221 out of 1,454 test instances, resulting in an overall accuracy of 84%. The negative class achieved the highest F1-score (0.91), indicating strong discriminative performance, while the neutral class remained the most challenging due to class imbalance. A comparison of metric evaluation values between the InSet lexicon and the Vader lexicon using the TF-IDF-based topic modeling approach is presented in Table 18. To provide a consolidated overview of all experimental conditions, Table 19 summarises the overall accuracy results across all four model configurations (two lexicons × two feature representations).

TABLE 18. SVM PERFORMANCE METRICS: INSET LEXICON WITH TF-IDF MODEL (PER CLASS)

| Metric | Negative | Neutral | Positive |
|---|---|---|---|
| Precision | 0.95 | 0.34 | 0.76 |
| Recall | 0.88 | 0.73 | 0.71 |
| F1-score | 0.91 | 0.46 | 0.73 |
| Accuracy | 84% | | |

## V. CONCLUSION

TABLE 19. SUMMARY OF OVERALL ACCURACY ACROSS ALL MODEL CONFIGURATIONS

| Configuration | Feature Representation | Overall Accuracy |
|---|---|---|
| InSet + SVM | TF-IDF | **84% (Best)** |
| InSet + SVM | Word2Vec | 71% |
| Vader + SVM | TF-IDF | 64% |
| Vader + SVM | Word2Vec | 65% |

The evaluation findings indicated that the InSet lexicon had superior precision in the positive class, signifying that the model effectively classified positive data with a commendable degree of accuracy. Nonetheless, the recall value for the neutral class was suboptimal, signifying that the model was inadequate in identifying all instances that genuinely belonged to the neutral category. In contrast, the Vader lexicon demonstrated superior recall performance in the positive class, signifying its enhanced ability to identify all positive sentiments. Nevertheless, the precision value for the negative class was low, signifying that the model frequently misclassified negative data into alternative categories. The hybrid approach integrating InSet Lexicon and TF-IDF-based SVM achieved the best performance with an accuracy of 84%, demonstrating the effectiveness of combining domain-specific lexicons with statistical feature representation for Indonesian sentiment classification. Nevertheless, the discourse to date has been quantitative and has not comprehensively addressed the elements that may affect the classification outcomes. A critical limitation of the present study is the reliance on pseudo-labels generated through automatic lexicon-based annotation, which introduces two interrelated forms of bias. First, lexicon-coverage bias arises because the InSet Lexicon contains approximately 6,609 negative entries versus 3,609 positive entries, causing the scoring algorithm to systematically skew toward negative classifications, particularly for tweets that contain even a single strong negative term. This inherent asymmetry means that the 84% accuracy reported here partly reflects the model's ability to learn the lexicon's own classification tendencies rather than true human-judged sentiment polarity. Second, domain-transfer bias may occur because the InSet Lexicon was originally developed for general Indonesian microblogs and has not been validated against diplomatic or foreign policy discourse; terms that carry neutral or positive connotations in diplomatic contexts (e.g., "tekanan," meaning "pressure," or "kepentingan," meaning "interest") may be scored as negative by the lexicon. The absence of a manually annotated ground truth dataset therefore constrains the interpretation of classification performance. Consequently, subsequent validation is essential, either via manual annotation of a stratified sample or through cross-lexicon agreement testing, to establish the extent to which pseudo-labels approximate true sentiment. It should also be noted that this study employed a single 60:40 train-test split rather than k-fold cross-validation; future work should apply cross-validation to assess the stability and generalisability of the reported 84% accuracy. Furthermore,

future research should consider benchmarking against contextual embedding models such as IndoBERT, which may better capture semantic nuance in domain-specific Indonesian discourse. From a substantive perspective, the polarised sentiment distribution identified in this study—with a dominant negative orientation—carries implications for government communication strategy: policymakers and diplomatic communication units may utilise such computational evidence to monitor public legitimacy, design targeted outreach, and proactively address concerns raised in digital public discourse regarding Indonesia's international diplomatic engagements.

From a political communication and international relations perspective, the dominance of negative sentiment in the InSet-labelled dataset is theoretically significant. Drawing on Framing Theory (Entman, 1993), the lexical patterns identified in this corpus—where words such as "conflict," "criticism," and "disappointed" rank among the most frequent terms—suggest that Twitter discourse frames Indonesia's Board of Peace participation predominantly through a lens of scepticism and concern. This aligns with the concept of opinion polarisation in digital public spheres (Sunstein, 2017), wherein contentious foreign policy issues tend to generate more vocal negative engagement on social media platforms. The high negative sentiment proportion (approximately 50% of the full 3,633-tweet dataset under InSet labelling) reflects broader public anxiety about Indonesia's potential departure from its historically non-aligned Bebas Aktif foreign policy doctrine [9], [19]. Scholars of Indonesian foreign policy have noted that public legitimacy is a critical component of diplomatic decision-making in a democratic system [9], [10]; accordingly, the computational evidence of polarised public opinion presented in this study carries direct implications for the government's diplomatic communication strategy. The Ministry of Foreign Affairs and relevant stakeholders may benefit from deploying real-time sentiment monitoring systems built on hybrid lexicon–SVM architectures to gauge shifts in public legitimacy, identify emerging concerns, and design evidence-based communication campaigns that address the specific issues driving negative sentiment. This interdisciplinary contribution—bridging computational text mining and political communication scholarship—constitutes one of the key value propositions of the present work.

## REFERENCES


[1] A. F. Hidayatullah and A. Azhari, *InSet: Indonesian Sentiment Lexicon*, Yogyakarta, Indonesia: Universitas Gadjah Mada, 2016.

[2] D. Saputra, "Komparasi Algoritma SVM dan Naive Bayes untuk Analisis Sentimen Media Sosial," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 123–130, 2022.

[3] R. H. Muhammadi, et al., "Analisis Sentimen Terhadap Kebijakan Pemerintah dengan InSet Lexicon dan SVM," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 9, no. 3, pp. 455–462, 2022.

[4] N. Firda et al., "Comparison of Rating-Based and InSet Lexicon-Based Labeling in Sentiment Analysis Using SVM," *Sistemasi: Jurnal Sistem Informasi*, vol. 13, no. 1, pp. 45–53, 2024.

[5] I. Zulfa and E. Winarko, "Sentimen Analisis Berbasis Leksikon dan SVM untuk Mengetahui Opini Publik terhadap Kebijakan Merdeka Belajar," *Jurnal Teknoinfo*, vol. 16, no. 2, pp. 89–96, 2022.

[6] A. Wulansari et al., "Optimization of Indonesian Sentiment Analysis Using InSet Lexicon and Random Forest," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, pp. 210–217, 2023.

[7] R. Prabowo, "Preprocessing Techniques for Indonesian Social Media Text: A Comprehensive Review," *Jurnal Ilmiah Teknologi Informasi*, vol. 18, no. 1, pp. 15–27, 2023.

[8] P. Utomo, "The Impact of Feature Scaling on SVM Performance in Indonesian Text Mining," *Journal of Applied Informatics and Computing*, vol. 8, no. 1, pp. 33–41, 2024.

[9] F. Muzakki, "Indonesia's Peace Diplomacy: A New Era of Non-Alignment," *Journal of Strategic and Global Studies*, vol. 6, no. 2, pp. 101–115, 2023.

[10] Ministry of Foreign Affairs of the Republic of Indonesia, *Annual Report on Indonesian Diplomacy*, Jakarta, Indonesia, 2024.

[11] United Nations, "United Nations Security Council Members," 2020. [Online]. Available: https://www.un.org/securitycouncil/. [Accessed: Feb. 15, 2026].

[12] R. Hidayat and D. J. Ratnaningsih, "Analisis Sentimen Program Mbg Menggunakan Algoritma Random Forest Dan Naive Bayes," *Journal of Computing and Informatics Research*, vol. 5, no. 1, pp. 395-400, 2025.

[13] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, 2017, pp. 391–394, doi: 10.1109/IALP.2017.8300625.

[14] M. F. N. Fathoni, E. Y. Puspaningrum, and A. N. Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Modem: Jurnal Informatika dan Sains Teknologi*, vol. 2, no. 3, pp. 62–76, 2024.

[15] Firda *et al.*, "Comparison of Rating-based and Inset Lexicon-based Labeling in Sentiment Analysis using SVM (Case Study: GoBiz Application Reviews on Google Play Store)," *SISTEMASI*, Mar. 2025.

[16] "Why Indonesia's Muslim groups are revisiting Palestinian issue, 'Board of Peace'," *South China Morning Post*, Feb. 2026.

[17] "Why Indonesia need to hear more on Board of Peace justification," *The Jakarta Post*, Feb. 2026.

[18] UGM Expert Statement, "UGM Expert Calls Indonesia's Decision to Join Trump's Board of Peace a Foreign Policy Blunder," Universitas Gadjah Mada, Jan. 2026.

[19] "Indonesia gabung 'Board of Peace': Strategi tepat atau hilang arah," *The Conversation Indonesia*, Feb. 2026.

[20] "Perspectives on Indonesia's involvement in Board of Peace," *The Jakarta Post*, Jan.–Feb. 2026.

[21] Y. Kim and B. Zhang, "Comparative analysis of TF-IDF and Word2Vec for short-text sentiment classification on small domain-specific corpora," *Expert Systems with Applications*, vol. 195, pp. 116–128, 2022, doi: 10.1016/j.eswa.2022.116528.

[22] F. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep recurrent neural network versus support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews," *Journal of Computational Science*, vol. 27, pp. 386–393, 2018, doi: 10.1016/j.jocs.2017.11.006.