Jurnal Masyarakat Informatika, 17(1), 2026
DOI:10.14710/jmasif.17.1.77146
ISSN: 2777-0648

Aspect-Based Sentiment Analysis on Nickel Mining Activities in Raja Ampat to Support Sustainable Development Goals

Dzakiah Aulia Karima, Dwi Rosa Indah*, Mgs Afriyan Firdaus, M Rudi Sanjaya, Gabriel Ekoputra Hartono Cahyadi

Department of Information Systems, Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia

* *Corresponding author: indah812@unsri.ac.id*

Abstract

Nickel mining in Raja Ampat has triggered significant public reaction, particularly on social media, due to its environmental and social impacts. However, public opinion on this issue has not been systematically analyzed. This study aims to examine public sentiment toward this issue using the Aspect-Based Sentiment Analysis (ABSA) approach with four classification algorithms: Support Vector Machine, K-Nearest Neighbor, Naïve Bayes, and Random Forest, all optimized through the Particle Swarm Optimization (PSO) method. Data was collected from X between June 1 and June 30, 2025, and analyzed based on three main aspects, namely environmental, social, and economic, with a total of 4,025 datasets. The analysis shows that negative sentiment dominates over positive sentiment, with the environmental aspect being the main focus, especially regarding coral reef damage and marine pollution. Among the four models used, the optimized Support Vector Machine algorithm achieved the highest performance with an accuracy of 87.5%. These findings are expected to serve as an evaluation for the government regarding mining permits to formulate policies that support the achievement of SDG 14 (Life Below Water) and SDG 15 (Life on Land).

Keywords: *Aspect-Based Sentiment Analysis, Nickel Mining, Sustainable Development Goals (SDGs), Support Vector Machine, Particle Swarm Optimization.*

1 Introduction

In the midst of growing global demand for low-carbon energy, nickel demand has risen rapidly in line with the transition to clean energy [1]. This situation has driven the growth of mining activities in various mineral-producing countries. However, nickel mining often has negative impacts on the environment, such as deforestation, ecosystem degradation, and complex social impacts [2]

This phenomenon is also evident in Raja Ampat, West Papua, which is an important global conservation area. Despite being designated a UNESCO Global Geopark in 2023, Raja Ampat continues to face the threat of nickel mining expansion. According to a report by Auriga Nusantara cited by Antara News, this mining has resulted in the loss of 494 hectares of forest, accompanied by a threefold increase in mining activity since 2020 [3]. The public reacted strongly after Greenpeace released a documentary video and investigative report on the mining's impacts, sparking the #SaveRajaAmpat campaign, which ultimately led the government to revoke four mining permits on June 10, 2025. These permits had been issued to several companies operating on Gag Island, Kawe Island, Manuran Island, Batang Pele Island, and Manyaifun Island [4].

Rehabilitation of former nickel mining sites requires a minimum of four years of intensive effort to restore soil fertility [5]. This prolonged recovery period has contributed to emotionally

Received: 20 August 2025; Revised: 9 January 2026; Accepted: 12 January 2026; Published: 11 February 2026

charged public discourse on social media platforms, although such expressions of public concern have not yet been analyzed in a systematic and data-driven approaches. The absence of structured sentiment analysis makes it difficult for stakeholders to identify the main issues of concern to the public. Sentiment analysis has also been used by governments to gauge public reactions to specific issues and to design and target public information or propaganda campaigns more effectively [6]. Sentiment analysis is used to determine positive or negative views on a topic [7]. Moreover, routine sentiment monitoring must be conducted daily for at least three weeks to accurately capture the emotional dynamics of society [8]. Most previous studies have focused on analyzing messages on X, as the platform is widely utilized by users from diverse backgrounds to express their views on a wide range of issues [9] Based on this, this study collected data from the X platform over a period of one month, from June 1 to June 30, 2025.

However, most previous studies only focused on general sentiment classification, namely positive, negative, or neutral, without exploring specific aspects of the issues discussed. In complex issues such as mining exploitation, the Aspect-Based Sentiment Analysis (ABSA) approach enables a more focused and comprehensive mapping of public opinion. This method generates more actionable insights that can support the development of effective and targeted policy responses [10]. According to [11], the use of ABSA can increase accuracy to 97%, which is higher than the general approach. Therefore, this study will apply ABSA to fill this gap.

Four classification models are used in this study: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forest (RF), each offering distinct advantages. These algorithms were selected based on the findings in [12], as they are among the most commonly applied methods in sentiment analysis and demonstrate strong capability in handling data variation. As explained in [13], SVM performs well on both linear and non-linear data, KNN relies on proximity between data points, NB is a fast and efficient probabilistic classifier, and RF is robust against overfitting due to its ensemble approach. To optimize model performance, Particle Swarm Optimization (PSO), a swarm intelligence-based optimization method, was employed and has been proven highly effective [14]. The effectiveness of PSO in optimizing parameters for aspect-based sentiment analysis has also been demonstrated in [15], [16].

This study aims to analyze public sentiment toward nickel mining exploitation in Raja Ampat using ABSA and four classification algorithms, namely SVM, KNN, NB, and RF, optimized with PSO. The novelty of this research lies in the implementation of ABSA and the comparison of the performance of the four algorithms, each optimized by PSO, which is still rarely done. The results of this analysis are expected to provide a more focused understanding of public opinion based on aspects, as a basis for formulating policies that favor the community and the environment, and support the achievement of SDG 14 (Life Below Water) and SDG 15 (Life on Land).

2 Research Methods

The research phases were arranged structurally to ensure that the analysis process could achieve the predetermined research objectives. This study applied the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as the main framework in the data analysis process. This methodology consists of six sequential, iterative stages. These are Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

The study in [17] highlights that CRISP-DM has become a standard methodology widely adopted across industries for structuring data mining projects. Figure 1 illustrates the CRISP-DM phases implemented in this study.

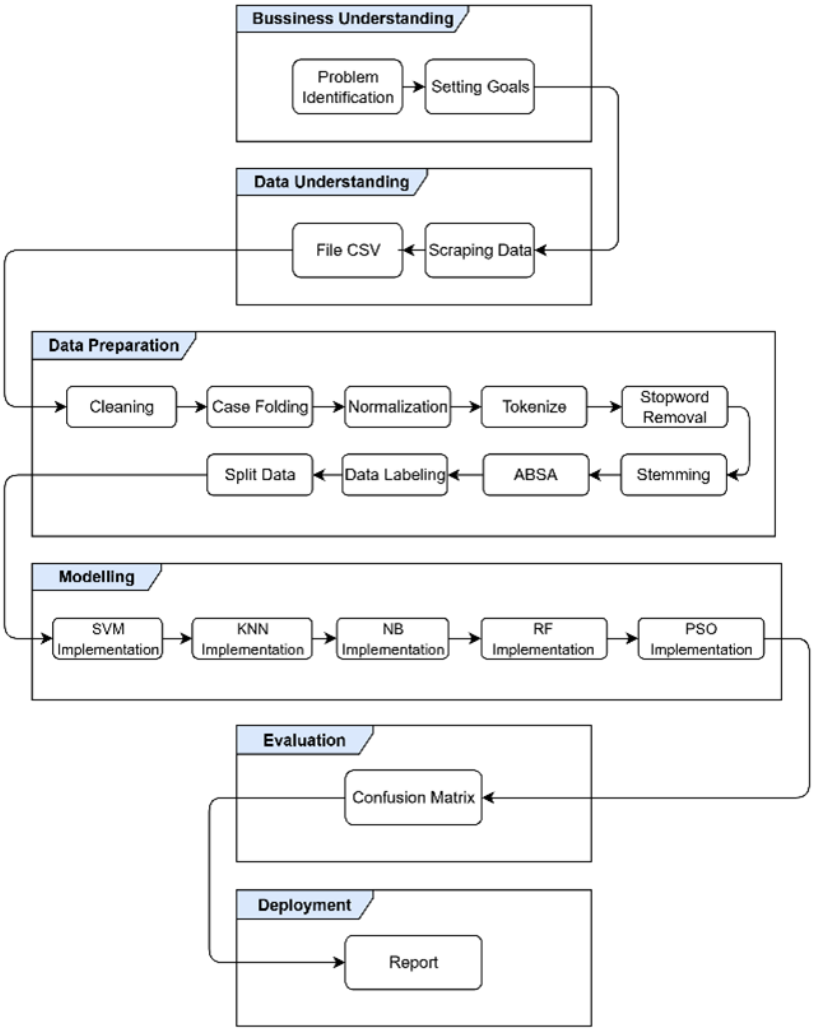


Figure 1 Research Phase

2.1. Business Understanding Phase

The business understanding phase is the initial stage in the CRISP-DM methodology, which aims to understand problems from a business perspective or issues faced. At this stage, there are two main steps, namely problem identification and goal setting.

2.2. Data Understanding Phase

The data understanding phase aims to provide an initial overview of the data's characteristics and quality. In this study, data were collected from X using the keyword #SaveRajaAmpat from June 1 to 30, 2025, resulting in 4,025 Indonesian-language posts in text format. This stage also includes data verification to ensure the validity of the analysis by removing empty data, duplicates, and irrelevant content, so that only clean, relevant data is used.

2.3. Data Preparation Phase

The data preparation phase aims to prepare data for the modeling process, starting with text cleaning using several stages, which include cleaning to remove punctuation marks, numbers, symbols, URLs, and emojis. Case folding to convert all text to lowercase. Word normalization to convert non-standard words into standard forms. Tokenization to separate words in a sentence. Stopword removal to remove common words without significant meaning and stemming to return words to their base form.

Subsequently, aspect extraction is carried out based on predefined categories such as environmental, social, and economic. Sentiment labeling is then performed using a lexicon-based approach, where the data is classified into two classes, positive and negative. The dataset is later split into training and testing sets using an 80:20 ratio, as recommended in [18], since this proportion is considered to provide more representative results.

2.4. Modelling Phase

In this study, the modeling phase aims to build an analysis model by applying four classification algorithms, such as SVM, KNN, Naïve Bayes, and Random Forest, optimized using PSO. The entire modeling process is conducted on Google Colab as a cloud-based data processing platform.

2.5. Evaluation Phase

The evaluation phase aims to assess the extent to which the model achieves the analysis objectives. In this study, the evaluation was conducted using a confusion matrix, which is a common method for assessing the performance of classification models with four main components, such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Based on these components, several key evaluation metrics were used, as shown in Table 1 [19].

Table 1 Confusion Matrix

	Actual Negative	**Actual Positive**
Predicted Negative	TN	FP (Type I Error)
Predicted Positive	FN (Type II Error)	TP

a. Accuracy, to measure the proportion of correct predictions:

$$Accuracy = \frac{TP+TN}{(TP+FP+FN+TN)} \tag{1}$$

b. Precision, to measure the accuracy of positive predictions:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

c. Recall, to measure how much positive data is successfully recognized:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

d. F1-Score, which is the average of precision and recall, which is very useful when there is class imbalance:

$$F - 1\, Score = \frac{2\, x\, precision\, x\, recall}{precision + recall} \tag{4}$$

2.6. Deployment Phase

The implementation phase represents the final stage of the research, aiming to communicate the analytical results to relevant stakeholders, particularly governmental institutions. In this study, the outcomes of the aspect-based sentiment analysis are compiled into a comprehensive report encompassing sentiment classification, interpretation of public perception, and word cloud visualizations. These outputs are intended to serve as data-driven input for policymakers in the development of strategies and regulatory frameworks that support the attainment of SDG 14 (Life Below Water) and SDG 15 (Life on Land).

3 Results and Discussion

3.1. Business Understanding Phase

As an initial stage, this study began by identifying the core issues underlying the research focus. The exploitation of nickel in Raja Ampat has generated significant controversy and elicited diverse public responses on social media, particularly concerning its potential impacts on the environment and local communities. This situation indicates that public opinions on sensitive environmental issues tend to be complex and heterogeneous, making general sentiment analysis insufficient to capture the depth and variation of public perceptions.

Studies by [20][21] demonstrate that Aspect-Based Sentiment Analysis (ABSA) offers more targeted and actionable insights that can better support policy formulation. Accordingly, because the controversy produces public opinions that span multiple issue dimensions, ABSA is necessary to identify sentiment more precisely at the aspect level. The purpose of this study is to analyze public sentiment toward nickel mining activities in Raja Ampat based on relevant aspects, as part of supporting the achievement of SDG 14 and SDG 15, and to provide stakeholders with a clearer and more comprehensive understanding for responding to environmental issues sustainably.

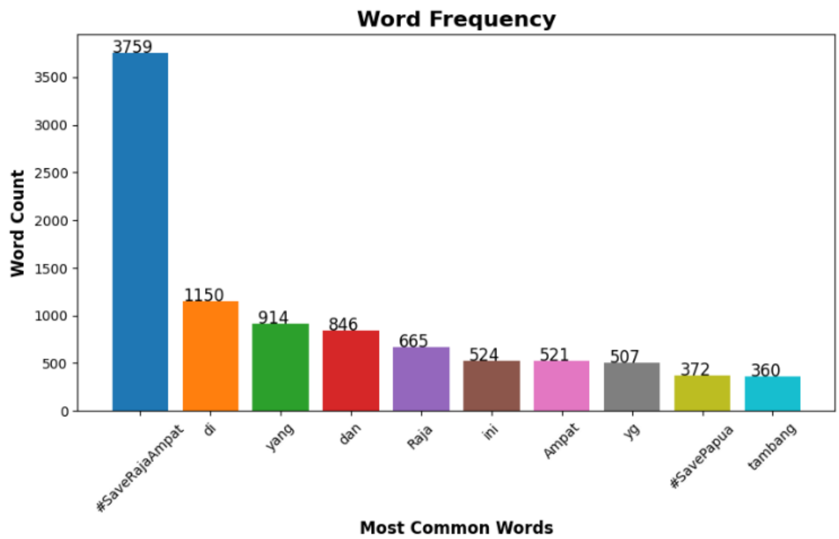


Figure 2 Word Frequency Distribution of Pre-processed Posts

3.2. Data Understanding Phase

In this study, data were collected from X through a data scraping process using the keyword #SaveRajaAmpat during the period of June 1 to 30, 2025, resulting in 4.025 posts in Indonesian text form and saved in CSV format files. Subsequently, data verification was conducted by removing empty data, duplicates, and irrelevant content to ensure the feasibility of the analysis, resulting in a final dataset of 3.942 posts. Furthermore, a word frequency analysis was performed to identify the most commonly occurring terms in the dataset, as illustrated in Figure 2.

3.3. Data Preparation Phase

This phase involves several key steps in the data pre-processing workflow. The cleaning process includes removing URLs, mentions, hashtags, punctuation, numbers, and extraneous characters to reduce noise. Case folding is applied by converting all text to lowercase, followed by normalization to standardize informal Indonesian words commonly found in social media content. Stopword removal is performed to eliminate high-frequency terms that do not contribute to aspect or sentiment identification. Stemming is then conducted to reduce words to their root form, ensuring consistent representation of word variations. This study uses the Sastrawi stemming algorithm, which is widely adopted for Indonesian text processing due to its accuracy and rule-based approach. The results of data preprocessing can be viewed in Table 2.

Table 2 Pre-processing Data

Before	After
@RadioElshinta Lebih elok @Haniffaisol_N @KLH_BPLH merestorasi Pulau Gag dan Pulau Waigeo dan mengembalikan biodiversitas yang rusak akibat kegiatan eksploitasi sumberdaya alam nikel @GreenpeaceIndo @jatamnas @walhinasional #SaveRajaAmpat #savepapua	restorasi pulau gag pulau waigeo kembalik biodiversitas rusak akibat eksploitasi sumberdaya alam nikel saverajaampat savepapua
(@RadioElshinta It would be better if @Haniffaisol_N @KLH_BPLH restored Gag Island and Waigeo Island and recovered the damaged biodiversity caused by nickel exploitation activities @GreenpeaceIndo @jatamnas @walhinasional #SaveRajaAmpat #savepapua)	(restoration gag island waigeo island restore damaged biodiversity due to nickel natural resource exploitation)
@unmagnetism hastag #SaveRajaAmpat jangan berhenti sampe mereka berhenti juga!! #SaveRajaAmpat	berhenti sampe berhenti juga saverajaampat
(@umnagnetism hashtag #SaveRajaAmpat should not stop until they stop too!! #SaveRajaAmpat)	(stop until they stop too saverarjaampat)

After the data pre-processing step is complete, the next stage is aspect extraction as part of the application of ABSA. This stage aims to group public opinion based on aspects that have been manually determined, such as environmental, social, and economy aspects. The selection of these three aspects aligns with one of the aspect extraction approaches described by [22], namely the Rule-Based Approach, which identifies aspects using predefined syntactic or semantic patterns such as keyword lists, and corresponds with the *Triple Bottom Line* theoretical framework, which has been extensively applied in social media analysis [23]. The complete list of analyzed aspects is presented in Table 3.

Table 3 Aspect for Analysis

Aspect	Word in Text
Environmental	'sedimentasi', 'siltasi', 'abrasi', 'erosi', 'pencemaran', 'logam berat', 'kontaminasi', 'limbah', 'residu', 'beracun', 'air', 'keruh', 'ekosistem', 'keanekaragaman hayati', 'terumbu', 'karang', 'habitat', 'mangrove', 'lahan basah', 'gambut', 'hutan', 'deforestasi', 'pembukaan lahan', 'degradasi lahan', 'longsor', 'tanah', 'kehilangan', 'satwa liar', 'ekologi', 'lingkungan', 'alam', 'kualitas udara', 'kawasan lindung', 'pesisir', 'pantai', 'laut', 'kelautan', 'degradasi', 'terdampak', 'rusak' ('sedimentation', 'siltation', 'abrasion', 'erosion', 'pollution', 'heavy metal', 'contamination', 'waste', 'residue', 'toxic', 'water', 'turbid', 'ecosystem', 'biodiversity', 'reef', 'coral', 'habitat', 'mangrove', 'wetland', 'peat', 'forest', 'deforestation', 'clearing', 'land degradation', 'landslide', 'soil', 'loss', 'wildlife', 'ecology', 'environment', 'nature', 'air quality', 'protected area', 'coastal', 'beach', 'sea', 'marine', 'degradation', 'impacted', 'damaged')
Social	'konflik', 'masyarakat', 'mata pencaharian', 'penurunan pendapatan', 'ketergantungan', 'kesehatan', 'penyakit pernapasan', 'kriminalisasi', 'lsm', 'tuntutan', 'kompensasi', 'sosialisasi', 'klaim', 'hak', 'pelanggaran', 'hak asasi manusia', 'budaya', 'kesenjangan', 'masyarakat', 'keadilan', 'kebijakan', 'pemerintah', 'demo', 'khalayak', 'organisasi', 'etika', 'toleransi', 'kepercayaan', 'ancam', 'menolak', 'lembaga', 'korban', 'menjaga', 'merawat', 'petisi', 'lokal', 'dampak', 'hidup' ('conflict', 'community', 'livelihood', 'income decline', 'dependency', 'health', 'respiratory disease', 'criminalization', 'ngo', 'demands', 'compensation', 'socialization', 'claim', 'rights', 'violation', 'human rights', 'culture', 'inequality', 'people', 'justice', 'policy', 'government', 'protest', 'public', 'organization', 'ethics', 'tolerance', 'trust', 'threaten', 'reject', 'institution', 'victim', 'protect', 'care', 'petition', 'local', 'impact', 'life')
Economy	'nilai', 'ekspor', 'investasi', 'royalti', 'pertumbuhan ekonomi', 'harga', 'biaya operasional', 'modal', 'modal asing', 'subsidi', 'infrastruktur', 'logistik', 'produksi', 'kapasitas produksi', 'volume', 'keuntungan', 'uang', 'jual', 'beli', 'konsumsi', 'pendapatan', 'pengeluaran', 'kewajiban keuangan', 'pinjaman', 'pengangguran', 'bangkrut', 'rupiah', 'lowongan kerja', 'utang', 'utang negara', 'kerugian', 'keuntungan', 'bisnis', 'kredit', 'usaha', 'pajak', 'juta', 'ribu', 'miliar', 'triliun' ('value', 'export', 'investment', 'royalty', 'economic growth', 'price', 'operational cost', 'capital', 'foreign capital', 'subsidy', 'infrastructure', 'logistics', 'production', 'production capacity', 'volume', 'profit', 'money', 'sell', 'buy', 'consumption', 'income', 'expenditure', 'financial obligation', 'loan', 'unemployment', 'bankrupt', 'rupiah', 'job vacancy', 'debt', 'national debt', 'loss', 'gain', 'business', 'credit', 'enterprise', 'tax', 'million', 'thousand', 'billion', 'trillion')

This stage, which focuses on aspect extraction, serves to map sentiment in a more targeted manner, so that public perceptions of each aspect of nickel mining exploitation in Raja Ampat can be identified. The extraction results are visualized in the form of data distribution by aspect, as shown in Figure 3. The aspect extraction show that environmental issues dominate public discourse, accounting for 66.01% of the posts, followed by social aspects (27.30%), economic aspects (6.60%), and a small “other” category (0.10%). The “others” category consists of posts that could not be reliably assigned to any of the three main aspects due to limited or unclear contextual information. The strong dominance of the environmental aspect reflects the high level of public concern regarding ecological conditions in Raja Ampat and provides a richer contextual foundation for interpreting the sentiment results.

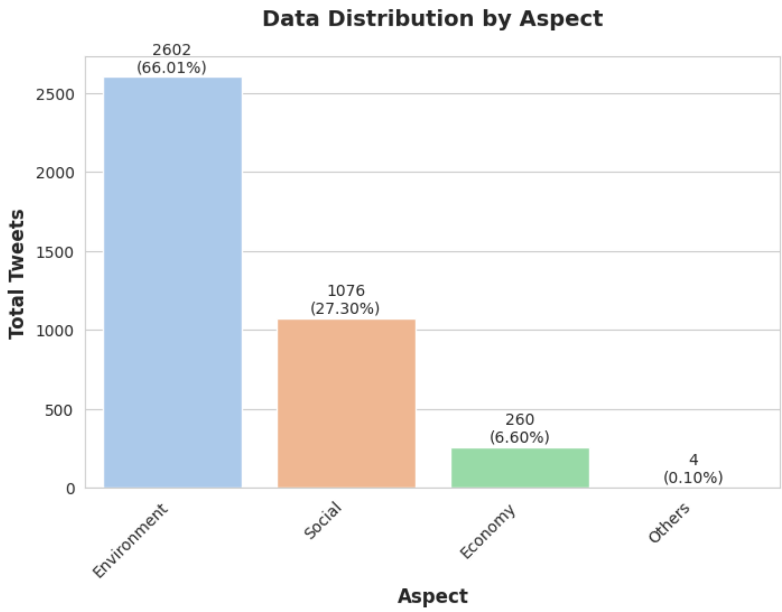


Figure 3 Aspect Distribution in the Sentiment Dataset

In this study, ABSA is employed to identify the tendency of public sentiment toward each aspect, aligning with recent approaches that regard aspect extraction and sentiment polarity classification as two independent subtasks within ABSA [24], [25].This separation allows the sentiment classification process to operate effectively without requiring aspect labels during model training, ensuring that the imbalance in aspect distribution does not influence the performance of the sentiment classifier.

After the aspect extraction stage, each tweet was labeled with sentiment using a lexicon-based approach. Lexicon-Based is a sentiment analysis approach that relies on a predefined dictionary of terms, each assigned a polarity weight, positive, neutral, or negative, to identify emotional orientation within a text, which is then quantified and categorized [26]

This labeling process classified posts into two main classifications, positive and negative, based on the opinion tendencies found in the text content. The distribution of the labeling results, both overall and by aspect, is shown in Figures 4 and 5.

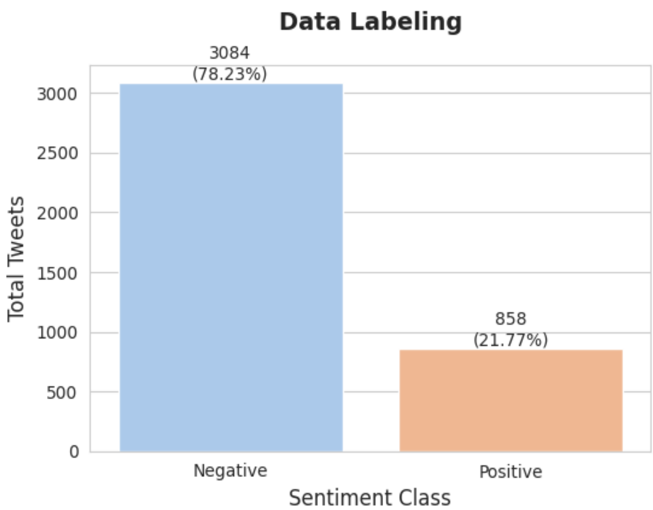


Figure 4 Data Labeling

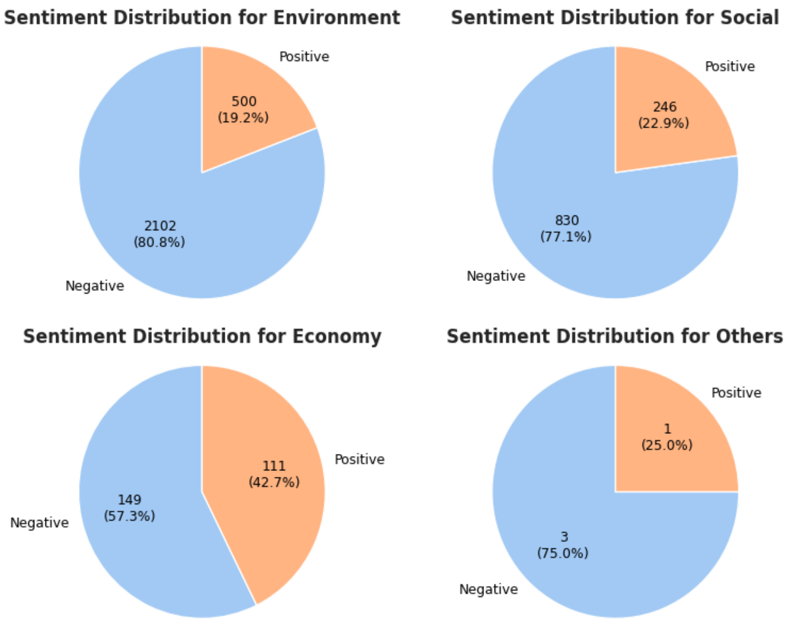


Figure 5 Data Labeling by Aspect

Figures 4 and 5 show that negative sentiment is dominant, both in the overall distribution and within each aspect. The environmental aspect recorded the highest number of posts, with negative sentiment being more prominent than in the other aspects. After the labeling stage, the dataset was divided into training and testing sets using an 80:20 ratio to ensure optimal model performance during training and evaluation. As a result, 3,153 entries were used for training and 789 for testing. The data split is visualized in Figure 6.

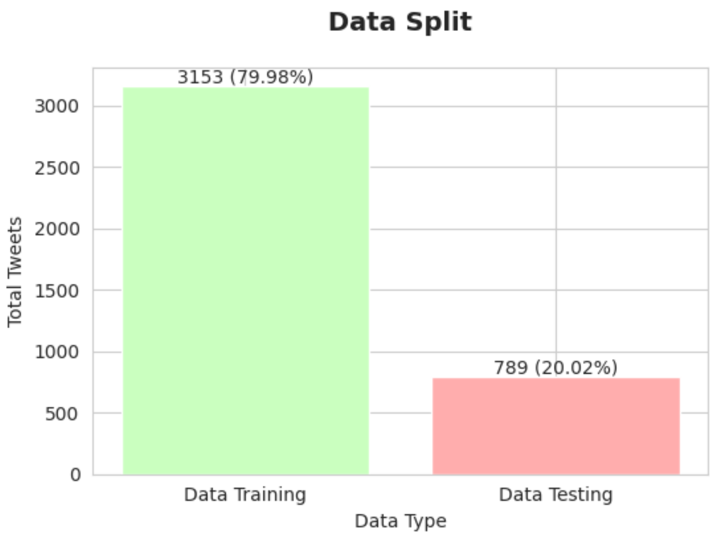


Figure 6 Data Split

3.4. Modelling Phase

In the modeling phase, this study used four classification algorithms, such as Support Vector Machine, K-Nearest Neighbor, Naïve Bayes, and Random Forest. Each algorithm was initially applied with standard configurations, such as SVM with a linear kernel effective for linear data, KNN with five nearest neighbors ($k = 5$) as the basis for classification, Naïve Bayes using MultinomialNB suitable for text data, and Random Forest with 100 decision trees and random settings to maintain consistency of results. The model was trained using training data and tested on test data to obtain prediction results before the optimization process was conducted.

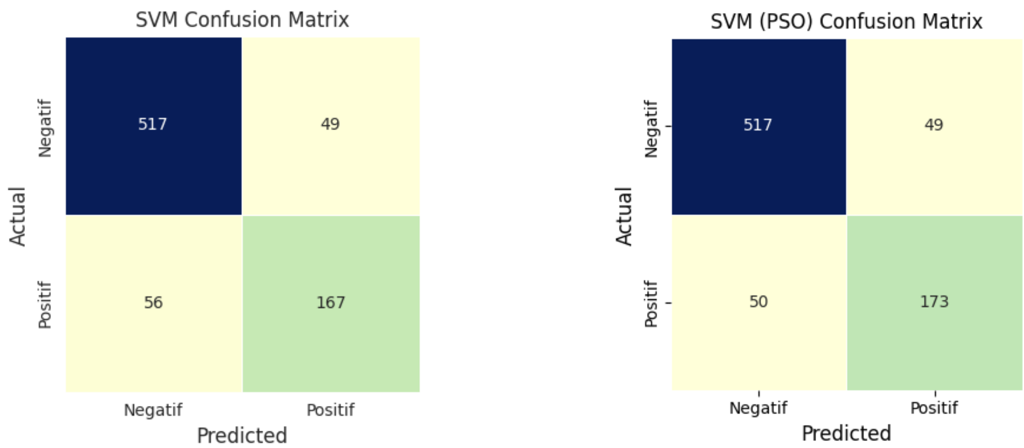
Then, hyperparameter optimization was conducted using Particle Swarm Optimization to find the best parameter combination with validation using 5-Fold Cross-Validation. The optimized parameters included C in SVM, the number of neighbors (k) in KNN, the alpha value in Naïve Bayes, and the combination of n_estimators, max_depth, and min_samples_split in Random Forest. After obtaining the optimization results, each model was retrained using the best parameters found and tested against test data to evaluate performance improvements. This approach was conducted so that each algorithm could produce optimal classification in the analysis of public sentiment related to nickel mining exploitation in Raja Ampat.

3.5. Evaluation Phase

In the evaluation phase, model performance is measured using a confusion matrix as a reference for analysis. The confusion matrix serves to compare the model's prediction results with the actual labels. From this matrix, several evaluation metrics are obtained, such as accuracy, precision, recall, and F1-Score. All models are evaluated both before and after optimization to determine how effectively each model classifies public sentiment toward the issue of nickel mining exploitation in Raja Ampat. The following sections detail the discussion of the evaluation results for each model.

3.5.1. Support Vector Machine (SVM)

The evaluation of the SVM model resulted in a classification report and a confusion matrix, presented in Tables 4 and 5. Based on the classification report, the accuracy of the SVM model increased from 86.7% to 87.5% after optimization with PSO. The precision, recall, and F1-score values also showed improvement, especially in the positive class. This performance improvement is supported by the results of the confusion matrix in Figure 7, where the number of correct predictions for the positive class increased from 167 to 173, and the number of errors decreased from 56 to 50, while predictions for the negative class remained stable at 517 correct and 49 incorrect.



(a) (b)

Figure 7 Confusion matrices of the SVM classifier: (a) baseline SVM without PSO optimization; (b) SVM optimized using PSO

Table 4 Comparison Classification Report of SVM Before and After PSO Optimization

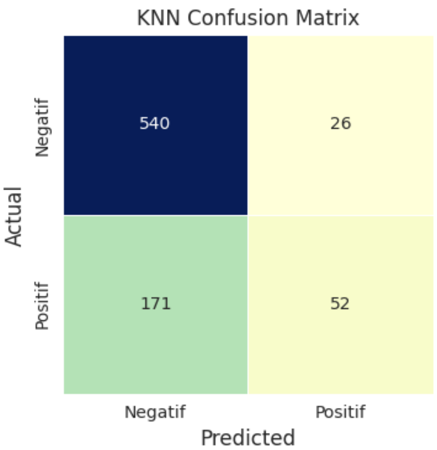
Class	Precision (Baseline)	Precision (PSO)	Recall (Baseline)	Recall (PSO)	F1-Score (Baseline)	F1-Score (PSO)
Negative	0.902	0.912	0.913	0.913	0.908	0.913
Positive	0.773	0.779	0.749	0.776	0.761	0.778
Macro Avg	0.838	0.846	0.831	0.845	0.834	0.845
Weighted Avg	0.866	0.874	0.867	0.875	0.866	0.874

Table 5 Comparison of the Accuracy of SVM Before and After PSO Optimization

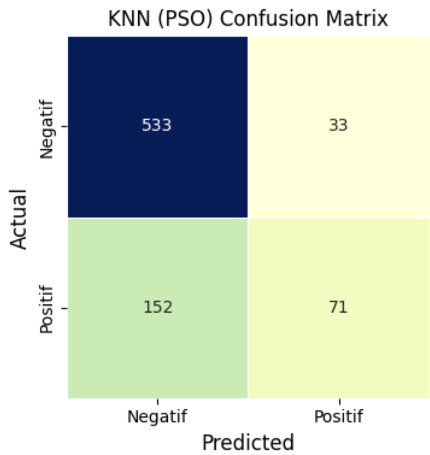
Accuracy (Baseline)	Accuracy (PSO)
0.867	0.875

3.5.2. K-Nearest Neighbor (KNN)

The evaluation of the KNN model produced a confusion matrix and a classification report, which are shown in Tables 6 and 7. The performance change of the KNN model after optimization with PSO can be seen in the classification report, where the accuracy increased from 75% to 76.6%. This improvement was also accompanied by improvements in precision, recall, and F1-score values in the positive class, although the changes were not very significant. These results are reinforced by the confusion matrix in Figure 8, which shows that correct predictions in the positive class increased from 52 to 71, while positive data errors decreased from 171 to 152. However, in the negative class, there was a slight decrease, from 540 to 531 data points correctly classified.



(a)



(b)

Figure 8 Confusion matrix of the KNN model: (a) without PSO optimization; (b) with PSO optimization

Table 6 Comparison Classification Report of KNN Before and After PSO Optimization

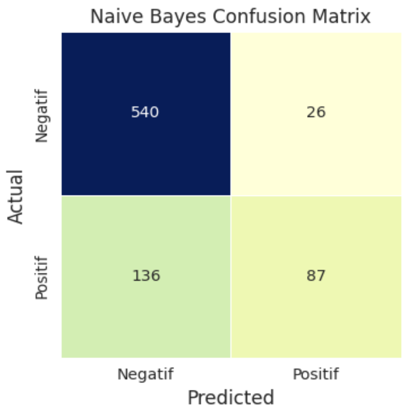
Classes	Precision (Baseline)	Precision (PSO)	Recall (Baseline)	Recall (PSO)	F1-Score (Baseline)	F1-Score (PSO)
Negative	0.759	0.778	0.954	0.942	0.846	0.852
Positive	0.667	0.683	0.233	0.318	0.346	0.434
Macro Avg	0.713	0.730	0.594	0.630	0.630	0.643
Weighted Avg	0.733	0.751	0.750	0.766	0.704	0.734

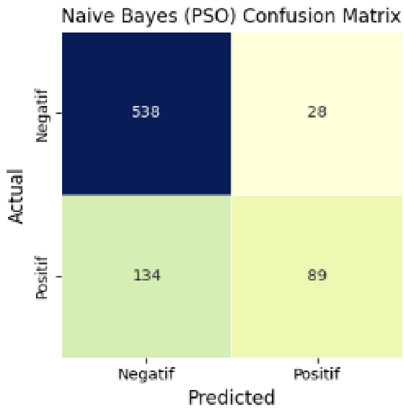
Table 7 Comparison Accuracy of KNN Before and After PSO Optimization

Accuracy (Baseline)	Accuracy (PSO)
0.750	0.766

3.5.3. Naïve Bayes (NB)

The evaluation of the NB model produced a confusion matrix and a classification report, as illustrated in Tables 8 and 9. Based on the classification report, the accuracy of the NB model remained unchanged at 79.5% both before and after PSO optimization. Nevertheless, a slight improvement was observed in precision and F1-score values, particularly for the positive class. This trend is reflected in the confusion matrix presented in Figure 9. The baseline NB model correctly classified 87 positive instances, whereas the PSO-optimized model increased the number of correctly classified positive instances to 89. Conversely, correct predictions for the negative class decreased marginally from 540 to 538. Overall, PSO optimization resulted in minor performance gains for the positive class, although it did not lead to an improvement in overall classification accuracy.





(a) (b)

Figure 9 Confusion matrix of the Naïve Bayes (NB) model: (a) without PSO; (b) with PSO

Table 8 Comparison Classification Report of NB Before and After PSO Optimization

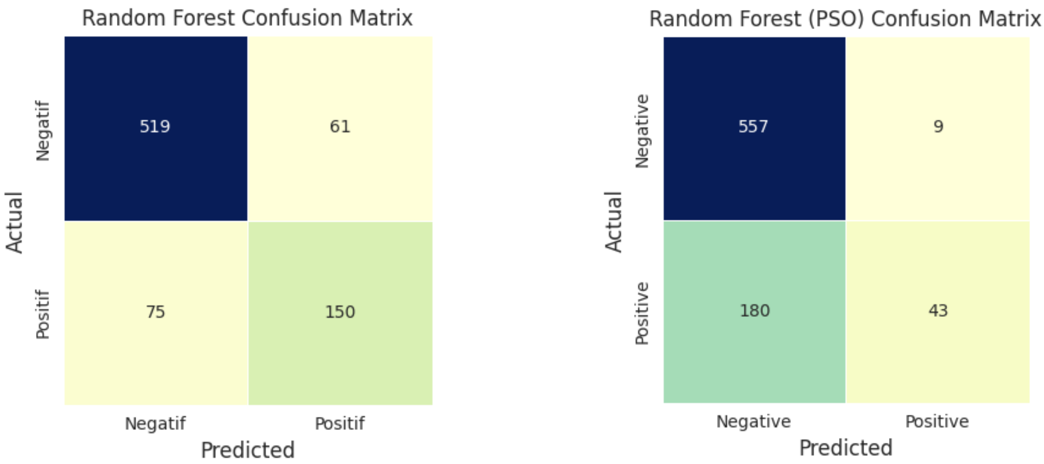
Class	Precision (Baseline)	Precision (PSO)	Recall (Baseline)	Recall (PSO)	F1-Score (Baseline)	F1-Score (PSO)
Negative	0.799	0.801	0.954	0.951	0.870	0.869
Positive	0.770	0.761	0.390	0.399	0.518	0.524
Macro Avg	0.784	0.781	0.672	0.675	0.694	0.696
Weighted Avg	0.791	0.789	0.795	0.795	0.770	0.771

Table 9 Comparison Accuracy of NB Before and After PSO Optimization

Accuracy (Baseline)	Accuracy (PSO)
0.795	0.795

3.5.4. Random Forest (RF)

The evaluation of the RF model resulted in a confusion matrix and a classification report, presented in Tables 10 and 11. Based on the classification report, the RF model exhibited degraded performance after PSO optimization, with accuracy decreasing from 80.9% to 77.8%. This behavior is clearly illustrated in Figure 10. The baseline RF model correctly classified 519 negative and 150 positive instances. In contrast, the PSO-optimized model substantially reduced the number of correctly classified positive instances to only 43, while the number of misclassified samples increased from 75 to 180. Although the classification performance for the negative class remained relatively high, a slight decline was observed. Overall, these results indicate that PSO optimization did not enhance the performance of the RF model in this scenario.



(a) (b)

Figure 10 Confusion matrix of the Random Forest (RF) model: (a) without PSO; (b) with PSO

Table 10 Comparison Classification Report of SVM Before and After PSO Optimization

Class	Precision (Baseline)	Precision (PSO)	Recall (Baseline)	Recall (PSO)	F1-Score (Baseline)	F1-Score (PSO)
Negative	0.857	0.773	0.880	0.977	0.868	0.863
Positive	0.673	0.824	0.628	0.274	0.650	0.411
Macro Avg	0.765	0.799	0.754	0.625	0.759	0.637
Weighted Avg	0.805	0.809	0.807	0.788	0.778	0.735

Table 11 Comparison of the Accuracy of SVM Before and After PSO Optimization

Accuracy (Baseline)	Accuracy (PSO)
0.809	0.778

3.5.5. Model Comparison Evaluation Results

The model performance comparison uses the F1-score because this metric provides a more balanced representation of precision and recall, especially when dealing with imbalanced data. The comparison of F1-score values for each model, both before and after PSO optimization, is presented in Figure 11.

The comparison results show that PSO optimization improves the performance of several algorithms. For SVM, the F1-score increases from 0.834 to 0.845 after PSO tuning. A similar improvement occurs in KNN, rising from 0.596 to 0.643, and in Naive Bayes, where the score slightly increases from 0.694 to 0.696. These results indicate that PSO contributes positively to enhancing the predictive capability of these models.

In contrast, Random Forest exhibits a different trend. The baseline F1-score of 0.759 drops significantly to 0.637 after PSO optimization. This decline suggests that PSO does not consistently yield performance gains across all algorithms and may even disrupt models that already operate effectively with their default parameters. Similar findings were reported in [27], [28], where Random Forest accuracy declined after being optimized using PSO. This outcome suggests that optimization does not always guarantee improvement particularly when a model such as Random Forest is already performing near optimally with its default parameters. In such cases, PSO may inadvertently lead to suboptimal parameter combinations or increase the tendency toward overfitting, ultimately reducing generalization performance on unseen data. Therefore, a strict and well-designed validation procedure is crucial to ensure the stability and robustness of model performance after optimization.

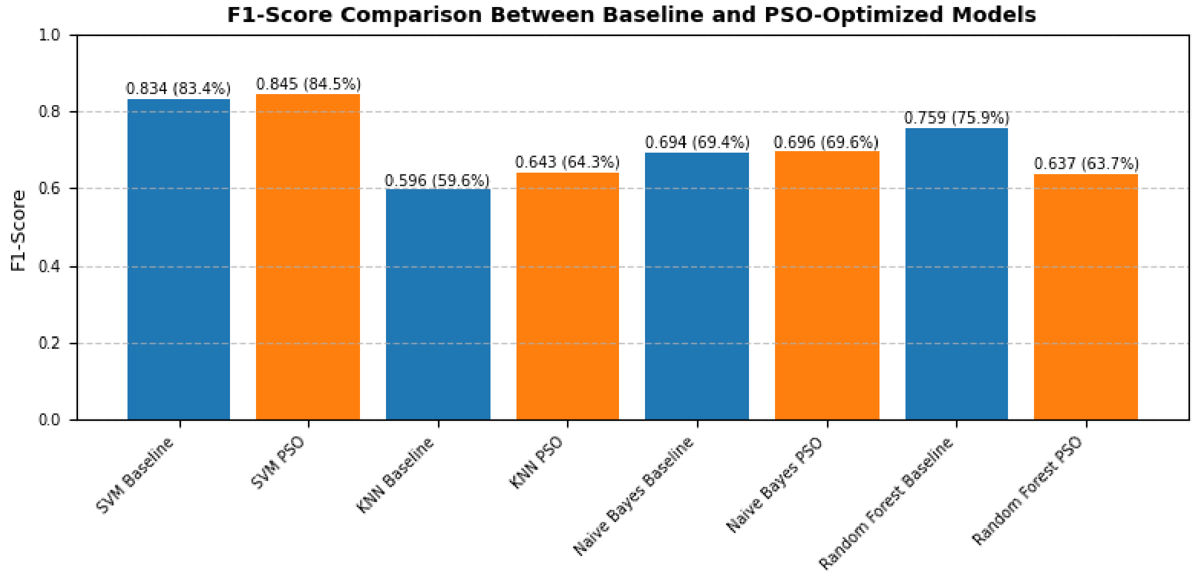


Figure 11 Comparison of Model F1-Scores

3.6. Deployment Phase

In the Deployment phase, the focus shifts from model evaluation to translating technical findings into actionable insights for stakeholders, particularly local government and environmental agencies. This study delivers a structured executive report that utilizes visualization tools, including Word Cloud and LDA topic modeling, to interpret the underlying themes of public sentiment.

Through this process, LDA generates probabilistic distributions of words within each topic, allowing the identification of dominant themes emerging from the data. These themes are then interpreted as representations of public perception within two main sentiment categories, namely positive sentiment and negative sentiment. The results of the LDA model are presented in Tables 12 and 13.

Table 12 LDA Model for Positive Sentiment

Topic	Dominant Words (LDA Gensim Format)	Theme / Interpretation
1	*0.085*"beautiful" + 0.061*"scenic" + 0.044*"lush" + 0.033*"blue" + 0.027*"nature" + 0.024*"paradise" + 0.018*"clear" + 0.014*"beach" + 0.011*"serene"*	Appreciation of Raja Ampat's Natural Beauty
2	*0.072*"rehabilitation" + 0.048*"restoration" + 0.035*"recovery" + 0.028*"biodiversity" + 0.022*"conservation" + 0.018*"preservation" + 0.014*"protect" + 0.012*"action"*	Support for Ecosystem Recovery and Conservation
3	*0.067*"tourism" + 0.052*"potential" + 0.036*"heritage" + 0.028*"green" + 0.022*"sustainable" + 0.019*"hope" + 0.013*"revive" + 0.011*"local"*	Optimism Toward Ecotourism and Sustainable Management

Table 13 LDA Model for Negative Sentiment

Topic	Dominant Words (LDA Gensim Format)	Theme / Interpretation
1	*0.094*"damage" + 0.073*"coral" + 0.058*"reef" + 0.041*"pollution" + 0.032*"waste" + 0.026*"ecosystem" + 0.019*"deforestation" + 0.014*"destruction" + 0.010*"water"*	Ecosystem Damage & Marine Pollution
2	*0.089*"conflict" + 0.064*"land" + 0.049*"customary" + 0.036*"community" + 0.028*"rights" + 0.022*"protest" + 0.017*"government" + 0.014*"inequality"*	Customary Rights, Land Issues, and Community Welfare Impacts
3	*0.083*"investment" + 0.061*"profit" + 0.047*"royalty" + 0.031*"loss" + 0.026*"unemployment" + 0.020*"cost" + 0.015*"income" + 0.012*"local_economy"*	Economic Disparities and Losses Experienced by Local Communities

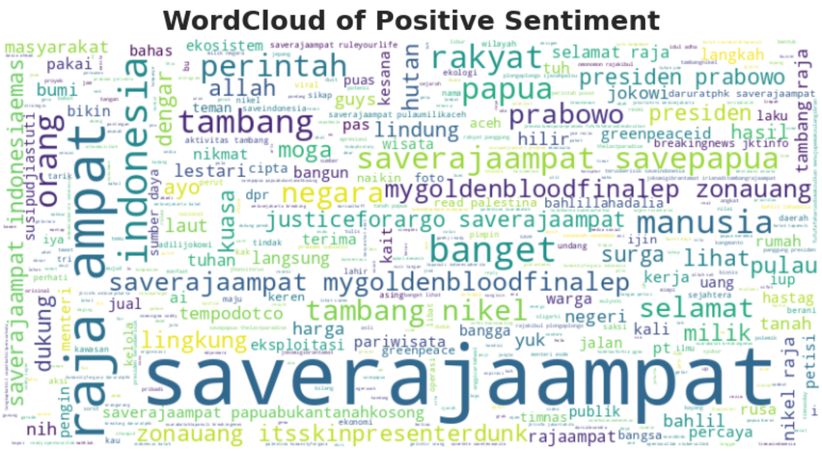


Figure 12 WordCloud of Positive Sentiment

Analytically, the visualizations reveal that positive sentiment is dominated by terms such as beautiful and conservation. In contrast, negative sentiment is predominantly clustered around keywords such as damage, pollution, and greed. This dominance indicates that public distress is specifically driven by visible environmental degradation rather than general economic opposition, as shown in Figure 12.

For policymakers, these aspect-specific insights function as an evidence-based monitoring tool. Instead of broad policy shifts, stakeholders can utilize these findings to prioritize environmental audits on mining sites that trigger high volumes of negative sentiment. This targeted approach ensures that mining regulations are enforced adaptively, directly supporting the conservation targets of SDG 14 and SDG by mitigating specific ecological risks identified by the community.

4 Conclusion

This study indicates that nickel mining activities in Raja Ampat have generated notable public attention, particularly regarding environmental, social, and economic aspects. By ABSA approach that integrates four classification algorithms (SVM, KNN, NB, and RF) and optimization using Particle Swarm Optimization (PSO), the research provides a structured overview of public sentiment across these aspects.

Evaluation results show that the PSO-optimized SVM model achieved the highest accuracy at 87.5%. The KNN model also experienced a slight improvement after tuning, while Naive Bayes remained stable. In contrast, the Random Forest model demonstrated a decrease in accuracy, suggesting that optimization does not always yield better performance for more complex algorithms.

Sentiment analysis reveals that negative opinions are predominantly associated with environmental issues, such as coral reef degradation and marine pollution. These findings align with the study's objective to assess public sentiment toward nickel mining in relation to SDG 14 (Life Below Water) and SDG 15 (Life on Land). The results may serve as a consideration for policymakers when reviewing mining practices in the region.

This study has certain limitations, particularly the dataset, which was collected over a one-month period and sourced from a single social media platform, thereby limiting the breadth of public opinions captured. Therefore, future research is encouraged to extend the data collection timeframe and incorporate multiple social media platforms to obtain a more comprehensive and representative understanding of public discourse.

Bibliography

[1] M. G. Y. Lo *et al.*, "Nickel mining reduced forest cover in Indonesia but had mixed outcomes for well-being," *One Earth*, vol. 7, no. 11, pp. 2019–2033, Nov. 2024, doi: 10.1016/j.oneear.2024.10.010.

[2] M. S. Agussalim, A. Ariana, and R. Saleh, "Kerusakan Lingkungan Akibat Pertambangan Nikel di Kabupaten Kolaka Melalui Pendekatan Politik Lingkungan," *Palita: Journal of Social Religion Research*, vol. 8, no. 1, pp. 37–48, Apr. 2023, doi: 10.24256/pal.v8i1.3610.

[3] Antara News, "Tambang dan ekosistem di jantung biodiversitas Raja Ampat," Antara News.

[4] Kementerian ESDM, "Minerba One Map Indonesia." Accessed: Jun. 24, 2025. [Online]. Available: https://geoportal.esdm.go.id/minerba

[5] T. Priono, R. Rosariastuti, and W. Sih Dewi, "The Influence of Length of Rehabilitation Process for Ex-Nickel Mining Land on Soil pH, Soil Organic Matter, Population and Distribution of Soil Microbes," *Jurnal Teknik Pertanian Lampung*, vol. 14, no. 1, pp. 99–106, 2024, doi: 10.23960/jtep-l.v14i1.99-106.

[6] J. Y. M. Nip and B. Berthelier, "Social Media Sentiment Analysis," *Encyclopedia*, vol. 4, no. 4, pp. 1590–1598, Oct. 2024, doi: 10.3390/encyclopedia4040104.

[7] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," Dec. 01, 2021, *Springer*. doi: 10.1007/s13278-021-00776-6.

[8] M. Pellert, H. Metzler, M. Matzenberger, and D. Garcia, "Validating daily social media macroscopes of emotions," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-14579-y.

[9] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," Aug. 01, 2023, *Elsevier Ltd*. doi: 10.1016/j.eswa.2023.119862.

[10] N. Kumar, R. Talwar, S. Tiwari, and P. Agarwal, "Aspect-based sentiment analysis of Twitter mobile phone reviews using LSTM and Convolutional Neural Network," *International Journal of Experimental Research and Review*, vol. 43, pp. 146–159, 2024, doi: 10.52756/ijerr.2024.v43spl.011.

[11] Salsabila, S. M. P. Tyas, Y. Romadhona, and D. Purwitasari, "Aspect-based Sentiment and Correlation-based Emotion Detection on Tweets for Understanding Public Opinion of Covid-19.," *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, no. 1, pp. 84–94, 2023, doi: 10.20473/jisebi.9.1.84-94.

[12] J. Wang, J. Wei, F. Tian, and Y. Wei, "A comparative study of machine learning models for sentiment analysis of transboundary rivers news media articles," *Soft comput*, vol. 28, no. 23, pp. 13331–13347, Dec. 2024, doi: 10.1007/s00500-024-10357-2.

[13] Z. Ahmad, W. Haider Bangyal, K. Nisar, M. Reazul Haque, and M. Adil Khan, "Comparative Analysis Using Machine Learning Techniques for Fine Grain Sentiments," *Journal on Artificial Intelligence*, vol. 4, no. 1, pp. 49–60, 2022, doi: 10.32604/jai.2022.017992.

[14] D. Freitas, L. G. Lopes, and F. Morgado-Dias, "Particle Swarm Optimisation: A historical review up to the current developments," Mar. 01, 2020, *MDPI AG*. doi: 10.3390/E22030362.

[15] J. W. Iskandar and Y. Nataliani, "Comparison of Naïve Bayes, SVM, and k-NN for Aspect-Based Gadget Sentiment Analysis," *Jurnal RESTI*, vol. 5, no. 6, pp. 1120–1126, Dec. 2021, doi: 10.29207/resti.v5i6.3588.

[16] A. Fauzi and A. Y. Heri, "Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K-Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 3, Dec. 2022, doi: 10.26418/jp.v8i3.56656.

[17] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.

[18] S. N. Selamat, N. Abd Majid, and A. Mohd Taib, "A Comparative Assessment of Sampling Ratios Using Artificial Neural Network (ANN) for Landslide Predictive Model in Langat River Basin, Selangor, Malaysia," *Sustainability (Switzerland)*, vol. 15, no. 1, Jan. 2023, doi: 10.3390/su15010861.

[19] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/ajbr.v27i4s.4345.

[20] A. Wafda, D. H. Fudholi, and J. Nugraha, "ASPECT-BASED SENTIMENT ANALYSIS ON TWITTER TWEETS ABOUT THE MERDEKA CURRICULUM USING INDOBERT," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 3, Feb. 2025, doi: 10.33480/jitk.v10i3.5692.

[21] A. Rajani and A. H. Hasugian, "Sentiment Analysis on the Planned Nickel Mining Development in Raja Ampat Using the Sentiment Analysis on the Planned Nickel Mining Development in Raja Ampat Using the Random Forest Algorithm," *Jatilima : Jurnal Multimedi dan Teknologi Informasi*, vol. 07, no. 03, 2025, doi: 10.54209/jatilima.v7i03.1565.

[22] W. Ahmad, H. U. Khan, F. K. Alarfaj and M. Alreshoodi, "Aspect-Based Sentiment Analysis: A Comprehensive Review and Open Research Challenges," in *IEEE Access*, vol. 13, pp. 65138-65182, 2025, doi: 10.1109/ACCESS.2025.3555744.

[23] K. Pilgrim and S. Bohnet-Joschko, "Corporate Social Responsibility on Twitter: A Review of Topics and Digital Communication Strategies' Success Factors," Dec. 01, 2022, *MDPI*. doi: 10.3390/su142416769

[24] Y. C. Hua, P. Denny, J. Wicker, and K. Taskova, "A systematic review of aspect-based sentiment analysis: domains, methods, and trends," *Artif Intell Rev*, vol. 57, no. 11, Nov. 2024, doi: 10.1007/s10462-024-10906-z.

[25] I. Perikos and A. Diamantopoulos, "Explainable Aspect-Based Sentiment Analysis Using Transformer Models," *Big Data and Cognitive Computing*, vol. 8, no. 11, Nov. 2024, doi: 10.3390/bdcc8110141.

[26] S. A. S. Mola, R. V. K. I. O. Roma, and T. Widiastuti, *Text Mining Analisis Sentimen dengan Lexicon*. Kaizen Media Publishing, 2025.

[27] R. Maulana and E. Eliyani, "Diabetes Classification Algorithm Optimization Using Particle Swarm Optimization on Naïve Bayes, C4.5 and Random Forest," *Jurnal Sisfokom (Sistem*

Informasi dan Komputer), vol. 14, no. 4, pp. 499–509, Oct. 2025, doi: 10.32736/sisfokom.v14i4.2431.

[28] R. Faizal Amir and I. Agus Sobari, "Penerapan PSO Over Sampling Dan Adaboost Random Forest Untuk Memprediksi Cacat Software," *IJSE-Indonesian Journal on Software Engineering*, vol. 6, no. 2, pp. 230–239, 2020.