ENTHUSIASTIC
INTERNATIONAL JOURNAL OF APPLIED STATISTICS AND DATA SCIENCE
Volume 5, Issue 2, October 2025, pp. 153-165

Analyzing Sentiments on IISMA Discontinuation Rumors with SVM, Random Forest Classifier, and XGBoost Classifier

Michelle Intan Handa [a,1], Maria Zefanya Sampe [a,2,*], Syafrudi [b,3]

[a] Business Mathematics, School of Applied STEM Universitas Prasetiya Mulya, Edu Town Kavling Edu I No. 1, Banten 15339, Indonesia
[b] Artificial Intelligence and Robotics, School of Applied STEM Universitas Prasetiya Mulya, Edu Town Kavling Edu I No. 1, Banten 15339, Indonesia
[1] 23102210036@student.prasetiyamulya.ac.id; [2] maria.zefanya@prasetiyamulya.ac.id*; [3] syafrudi@prasetiyamulya.ac.id
* Corresponding author

ARTICLE INFO

Article history
Received: February 15, 2025
Revised: April 22, 2025
Accepted: April 26, 2025

Keywords
IISMA
Sentiment Analysis
SVM
RFC
XGBoost Classifier

ABSTRACT

Indonesian International Student Mobility Award (IISMA) is a government-run student exchange program. Recently, rumors regarding its discontinuation have sparked various public opinions. This study aims to analyze these public sentiments and evaluate which machine learning model is most suitable for classifying sentiment labels in the dataset. The models tested included support vector machine (SVM), random forest classifier (RFC), and extreme gradient boosting (XGBoost) classifier. The dataset consisted of 630 tweets scraped from Twitter and was split into an 80:20 ratio, with 80% allocated for training and 20% for testing. The results indicated that both SVM and RFC were the most effective models, achieving the highest accuracy of 85.44%. Sentiment analysis reveals that the majority of public opinion is positive, suggesting that most people agree with the discontinuation of the IISMA program because the program is perceived as nonurgent and not a current national priority. These findings provide insights into public sentiment and highlight the utility of machine learning models in classifying such sentiment data effectively.

1. Introduction

The Indonesian International Student Mobility Award (IISMA) is a part of the Merdeka Belajar Kampus Merdeka (MBKM) program launched by Indonesia's Ministry of Education, Culture, Research, and Technology in 2021. This program offers Indonesian diploma and undergraduate students the opportunity to study abroad for one semester at partner universities to broaden their perspectives, enhance soft skills, and foster international relations [1]. In 2024, more than 9,000 students applied for this program, with 2,029 students being selected. While IISMA has received significant appreciation for its potential to enrich student's academic and personal development, it has also faced criticism related to issues such as accessibility, its relevance to domestic educational challenges, and disparities in educational facilities across Indonesia's various regions [2].

Public opinions about the program are frequently shared on platform X (Twitter), one of Indonesia's largest social media platforms with over 19 million active users. These opinions, ranging

from positive to negative, provide valuable insights into public sentiment regarding IISMA. Sentiment analysis offers a powerful method to evaluate the opinions, emotions, and attitudes expressed by the public [3]. In this study, sentiment analysis was applied to examine the public's views and reactions to the IISMA program, specifically focusing on the recent controversy surrounding the program's dissolution.

Several previous studies have examined sentiment analysis related to IISMA using different methodologies. One study applied support vector machine (SVM) with various kernel functions, such as linear, polynomial, and radial basis function (RBF) to classify sentiment on Twitter. The study found that while SVM could effectively categorize positive and neutral sentiments, it struggled with negative sentiment classification due to dataset imbalance. The RBF kernel performed best with a precision of 70.52% and an accuracy of 62.92%, while the linear kernel produced the highest F1-score of 53%, indicating a more balanced performance between precision and recall. Furthermore, sentiment analysis on English-language tweets yielded higher accuracy (71.11%) than on Indonesian-language tweets (62.44%), suggesting a need for better sentiment analysis tools tailored to the Indonesian language [4].

Another study utilized both Twitter and TikTok comments to analyze sentiment toward IISMA using SVM and FastText. The study highlighted that FastText's ability to generate word vectors complemented SVM's classification capability. Results indicated that sentiment analysis performed better on TikTok data, with an accuracy of 85.0%, compared to 74.4% on Twitter. When combining both datasets, the accuracy reached 84.24%, demonstrating the advantage of integrating multiple social media sources. The study also revealed that IISMA sentiment was predominantly neutral and recommended program improvements, such as providing English language test assistance for underprivileged students [5].

The issue arises from the large number of discussions happening on Twitter, which can be difficult to analyze manually due to the variety of opinions and the volume of posts [6]. The complexity of mixed sentiments makes it challenging for individuals to process and understand all perspectives comprehensively. This urges the need for an automated approach to efficiently process these opinions, especially considering the limitations observed in previous research regarding dataset imbalance and language-specific performance issues.

To address these gaps, this study applied sentiment analysis techniques using SVM, random forest classifier (RFC), and extreme gradient boosting (XGBoost) classifier. These machine learning models have been proven effective in classifying text data and can be leveraged to analyze and interpret the sentiments expressed on Twitter regarding IISMA [7]. The selection of these three models was driven by the need to determine the most appropriate classification approach for sentiment analysis on this topic, considering their distinct characteristics. SVM, a hyperplane-based model, is effective in high-dimensional spaces and excels at finding optimal decision boundaries for separating sentiment classes. Meanwhile, RFC is a tree-based ensemble method, thereby providing robustness against noisy data and offers feature importance insights, making it valuable for handling complex sentiment patterns. Meanwhile, XGBoost classifier, a gradient boosting-based model, enhances classification performance through iterative learning, efficiently capturing non-linear relationships within the data. These models represent the three fundamental approaches to classification; hyperplane-based, tree-based, and gradient boosting-based, which allow this study to identify which method is best suited for analyzing public sentiment regarding IISMA. By incorporating a more balanced dataset and evaluating multiple models, this research aims to enhance sentiment classification accuracy and provide a more comprehensive understanding of public opinion on the program's discontinuation.

2. Method

The research method is illustrated in Fig. 1.

p-ISSN 2798-253X
e-ISSN 2798-3153

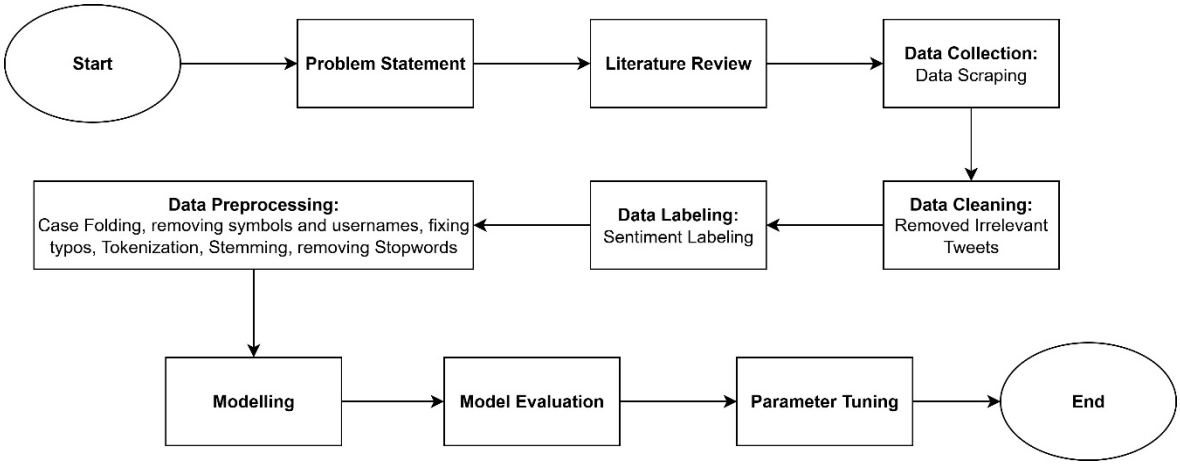

Fig. 1 Research method flowchart.

2.1. Support Vector Machine (SVM)

SVM is a widely used supervised machine learning method that is known for its ability to handle complex and large datasets, including high-dimensional data such as text, images, or signals, making it a versatile tool for classification and prediction tasks [8]. This method employs a kernel-based approach to find an optimal hyperplane or separating plane for classifying data. The main objective of SVM is to maximize the margin between the hyperplane and the closest data points from each class, ensuring a well-defined separation between the classes. Through optimization, SVM trains the algorithm to identify the ideal hyperplane.

At first, the principle of SVM works as a linear classifier, where classification is achieved by separating classes with a straight line [9]. SVM was later developed to handle nonlinear problems by mapping data x into a function $\phi(x)$, which is a transformation function that maps the data x into a higher-dimensional space, where a linear separation becomes possible. Typically, the new values resulting from the transformation $\phi(x)$ are unknown and difficult to comprehend. However, according to Mercer's theorem, the dot product calculation between functions ϕ can be replaced with the kernel trick, which directly defines the transformation ϕ.

The kernel trick helps to compute this transformation efficiently. The trained SVM model can then be used to classify new, unseen test data. Table 1 shows the kernels that are often used in SVM.

Table 1. Kernel

Kernel Name	Kernel Function
Linear	$K(x_i, x_j) = x_i \cdot x_j$
Polynomial	$K(x_i, x_j) = (x_i \cdot x_j)^d$
Gaussian RBF	$K(x_i, x_j) = exp(\frac{-\lvert x_i \cdot x_j \rvert^2}{2\sigma^2})$
Sigmoid	$K(x_i, x_j) = tanh(\sigma(x_i, x_j) + c)$

2.2. Random Forest Classifier (RFC)

RFC is a powerful model used to classify data into specific classes by using decision trees. A decision tree is constructed by determining the root node and ending with several leaf nodes to arrive at the final result. RFC belongs to the ensemble family of methods, combining the simplicity and interpretability of individual decision trees with improved accuracy [10]. By using bootstrap aggregating (bagging), random forest reduces variance in noisy datasets by creating a bootstrapped dataset for each tree and training the decision tree models in parallel. The final classification is determined by a majority voting mechanism, where the most frequent result among all decision trees becomes the final prediction. This ensemble approach significantly enhances the model's accuracy and robustness compared to a single decision tree model [11].

p-ISSN 2798-253X
e-ISSN 2798-3153

The process of forming a decision tree in the RFC method is similar to the process in classification and regression tree (CART), except that in RFC, pruning (trimming) is not performed. The Gini index is used to select features at each internal node of the decision tree. The equation to calculate the Gini index is shown in (1).

$$Gini\ (D) = 1 - \sum_{i=1}^{m} {P_i}^2 \quad (1)$$

where P_i is the probability of a tuple D from a class and m is the amount of class labels.

2.3. Extreme Gradient Boosting (XGBoost) Classifier Model

XGBoost is a decision- tree based machine learning algorithm designed to improve efficiency, error, and accuracy over other gradient boosting methods [12]. This algorithm uses ensemble learning techniques by combining several decision trees to produce more accurate predictions. XGBoost optimizes the model with loss function and regularization methods, which helps in avoiding overfitting and improving model performance.

Mathematically, the prediction of XGBoost can be calculated as the sum of the outputs of K decision trees.

$$\hat{yi} = \sum_{k}^{K} f_k(x_i), f_k \epsilon F \quad (2)$$

where F is the set of all possible decision trees. To optimize the model, XGBoost minimizes the objective function consisting of the loss function of $l(yi, \hat{yi})$ and the regulation model of $\Omega(f_k)$. XGBoost also implements improvement techniques such as shrinkage (learning rate), automatic feature selection, and parallelization in training, which makes it a very effective algorithm for classification and regression tasks [13].

2.4. Data Collection

In this research, the collection of data was done by web scraping. Web scraping is an automated technique used to gather information and data from websites. The collected data can then be saved in various formats such as text, CSV, or JSON [14]. Through web scraping, it becomes easy to collect pertinent tweets from platform Twitter related to the IISMA program. In this step, relevant data were gathered from platform Twitter using scraping techniques, allowing access to public posts and user-generated content related to the IISMA program. The keywords used to gather the data were "IISMA," "MBKM," and "Kampus Merdeka." The dataset was limited to 2,000 data points and collected from January 1, 2021, to February 1, 2025. The goal was to collect a large dataset of tweets that express opinions about IISMA in general. However, the selection of keywords may introduce bias in the sentiment distribution, as users expressing opinions about the discontinuation of the IISMA program may not always use these specific keywords. This limitation may affect the representativeness of the dataset and the diversity of opinions captured.

2.5. Data Cleaning

The data gathered through web scraping is often messy, unorganized, and unstructured. Web scraping tools typically gather a large volume of data, but this data may contain numerous inconsistencies, duplicates, and irrelevant tweets that do not align with the research objectives. For instance, some tweets may be off-topic, not related to the IISMA program, or contain noise such as spam or irrelevant hashtags. Additionally, there may be repeated tweets, which can lead to redundancy in the dataset, affecting the quality of the analysis.

To ensure the accuracy and reliability of the data, it is essential to perform a thorough data cleaning process. The first step in cleaning the dataset involved removing duplicates, which are often present due to the nature of web scraping. Duplicate entries can distort the sentiment analysis results by artificially inflating the frequency of certain sentiments.

Next, it was important to filter out irrelevant tweets. These might include tweets that did not mention the keywords or were not related to the public sentiment regarding its discontinuation.

Irrelevant data also included tweets in languages other than the one used for the study or tweets that were not properly formatted for analysis.

2.6. Data Labelling

After collecting the data, the next step was dataset labeling. In this phase, each tweet was manually annotated or categorized based on the sentiment it expresses positive, negative, or questioning [15]. Positive label was for tweet that agreed with the discontinuation of IISMA, negative label was for tweet that did not agree with the discontinuation of IISMA, and questioning label was for tweet that questioned whether IISMA was really being discontinued.

The annotation process was conducted by the author, who served as a single annotator, using domain knowledge and predefined criteria, such as expressions of support, complaints, or inquiries. This approach helps to provide a reliable ground truth for training the machine learning models. Labeling the data ensures that the model learns to distinguish between different sentiments and can classify future tweets accurately [16].

2.7. Data Preprocessing

Data preprocessing is a crucial step in preparing the dataset for machine learning model training. Several tasks are performed to clean and standardize the data [17].

2.7.1. Case Folding

This process involves converting all texts to lowercase. This aims to ensure uniformity and avoid the model treating the same word in different cases as distinct.

2.7.2. Remove Symbols and Usernames

Tweets on Twitter often contain hashtags or symbols in the sentence that represent usernames. These symbols must be removed since the analysis focused solely on the text of the tweet itself.

2.7.3. Fixing Typos

Many social media posts contain misspellings or informal language. This step corrects common typos by using a predefined dictionary to map each word's typo. In this research, SymSpell was utilized to enhance the quality of the text data, ensuring that the model could accurately interpret the words.

2.7.4. Tokenization

Tokenization involves breaking down text into smaller units, such as words or phrases (tokens), then assigning a token into each word. This helps the model understand the structure of the language and the relationships between different words.

2.7.5. Stemming

Stemming reduces words to their base or root form. This helps to minimize the complexity of the data and ensures that variations of the same word are treated as equivalent. Sastrawi is the library used for stemming Indonesian words [18].

2.7.6. Removing Stopwords

Stopwords are common words that are removed from the dataset as they do not carry significant meaning for sentiment analysis. Common Indonesian words like "yang," "di," "ke," and "dan" were removed. In this research, natural language toolkit (NLTK) was used to remove the stopwords [19].

2.8. Modelling

Once the data are preprocessed and clean, the next step was to apply machine learning models for sentiment classification. In this research, three popular classifiers models were used to evaluate sentiment from the tweets: SVM, RFC, and XGBoost Classifier.

p-ISSN 2798-253X
e-ISSN 2798-3153

2.9. Model Evaluation

In this research, the evaluation of the machine learning models was conducted using accuracy, precision, recall, and F1-score, which are essential metrics for assessing classification performance. Precision measures the accuracy of positive predictions by calculating the proportion of correctly predicted positive instances out of all predicted positives. A high precision value indicates that the model makes fewer false positive errors. Recall, on the other hand, evaluates the model's ability to correctly identify all relevant instances by measuring the proportion of correctly predicted positives out of all actual positive instances. A high recall value suggests that the model effectively captures most of the positive samples but may produce more false positives. F1-score is the harmonic mean of precision and recall, providing a balanced measure when there is an uneven class distribution. It is particularly useful in sentiment analysis, where some sentiment classes may be more frequent than others [20]. By analyzing these three metrics, this study ensures a comprehensive evaluation of the classifiers' effectiveness in accurately predicting sentiment in tweets about IISMA.

2.10. Parameter Tuning

Parameter tuning is a crucial part in machine learning, as the performance of a model heavily depends on the parameters chosen. These parameters include number of estimators, leaf, impurity and model configuration parameters. Optimizing them is often time-consuming, requiring multiple training runs over large datasets, leading to high computational costs and CO_2 emissions. Selecting the right hyper-parameters is critical because they influence the model accuracy, training efficiency, and ensure the model performs well on unseen data [21].

In this research, the parameter tuning used was Optuna. Optuna is an open-source hyperparameter optimization framework designed to automate the search for optimal parameters in machine learning models [22]. It uses a technique called Bayesian optimization to efficiently explore the hyperparameter space, making it a more computationally efficient alternative to methods like GridSearch or random search. Instead of trying every possible combination of parameters, Optuna learns from previous trials to propose more promising parameter sets, often leading to better results in fewer iterations. This approach can save a lot of time and reduce computational costs, especially in complex models with many hyperparameters. Additionally, Optuna's integration with other machine learning libraries, like scikit-learn, XGBoost, or LightGBM, makes it highly adaptable to various use cases. By leveraging Optuna, model's performance can be fine-tune more effectively while minimizing resource consumption [23].

3. Results and Discussion

3.1. Scrapping and Labeling Result

The data collection process which was done through web scraping resulted in a dataset containing 1,879 entries. This dataset holds three distinct sentiment classes: positive, negative, and questioning. A positive label indicates agreement with the discontinuation of the IISMA program, while negative signifies disagreement, and questioning reflects uncertainty about it. Table 2 shows the result of the scrapping and labeling process.

Table 2. Dataset Labeling Result

Tweet	Label
@ferrykoto Program nadiem kampus merdeka MSIB dan IISMA dihapus menteri yg ini nih. Padahal itu bagus2 semua.	Negative
@kenapsbangkesel mau dihapus AH pler nih ajg duitnya habis ke iisma & lpdp #hapuskaniisma ga guna relevansi ke disiplin ilmu gue hampir gada anying	Positive
Damn.... iisma n msib cuma postpone apa beneran dihapus ya.....	Questioning

Some entries did not specifically focus on IISMA and were then removed. After filtering the data to concentrate only on IISMA, the resulting dataset contained 630 entries, as shown in Fig. 2.

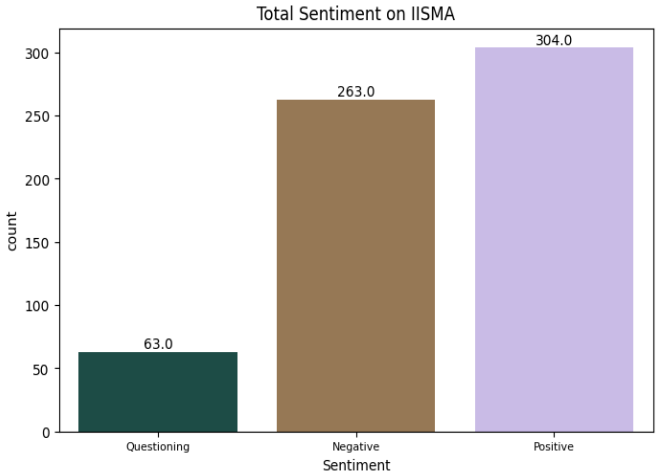

Fig. 2 Visualization of the total sentiment of IISMA.

The data included 63 entries categorized as questioning, 263 entries classified as negative, and 304 entries marked as positive. This distribution shows an imbalance in the data, with the questioning class having significantly fewer entries and the negative class having slightly fewer entries than the positive class. These imbalances could potentially affect the model's accuracy in predicting these classes. To address this issue, the questioning class label was removed, and resampling was performed on the negative and positive classes to balance the dataset. Specifically, 263 samples were randomly resampled between both the negative and positive classes, ensuring equal data size.

3.2. Preprocessing Result

3.2.1. Case Folding

This step converted all text into lowercase. Table 3 shows the result of this step.

Table 3. Case Folding Result

Tweet	Case Folding Result
@ferrykoto Program nadiem kampus merdeka MSIB dan IISMA dihapus menteri yg ini nih. Padahal itu bagus2 semua.	@ferrykoto program nadiem kampus merdeka msib dan iisma dihapus menteri yg ini nih. padahal itu bagus2 semua.
@kenapsbangkesel mau dihapus AH pler nih ajg duitnya habis ke iisma & lpdp #hapuskaniisma ga guna relevansi ke disiplin ilmu gue hampir gada anying	@kenapsbangkesel mau dihapus ah pler nih ajg duitnya habis ke iisma & lpdp #hapuskaniisma ga guna relevansi ke disiplin ilmu gue hampir gada anying
Damn.... iisma n msib cuma postpone apa beneran dihapus ya.....	damn.... iisma n msib cuma postpone apa beneran dihapus ya.....

3.2.2. Remove Symbols and Usernames

Every symbol, hashtag, usernames, and URL were removed in this step. Table 4 shows the result of this step.

Table 4. Symbol and Username Removal Result

Tweet	Symbol and Username Removal Result
@ferrykoto program nadiem kampus merdeka msib dan iisma dihapus menteri yg ini nih. padahal itu bagus2 semua.	program nadiem kampus merdeka msib dan iisma dihapus menteri yg ini nih padahal itu bagus2 semua
@kenapsbangkesel mau dihapus ah pler nih ajg duitnya habis ke iisma & lpdp #hapuskaniisma ga guna relevansi ke disiplin ilmu gue hampir gada anying	mau dihapus ah pler nih ajg duitnya habis ke iisma lpdp ga guna relevansi ke disiplin ilmu gue hampir gada anying
damn.... iisma n msib cuma postpone apa beneran dihapus ya.....	damn iisma n msib cuma postpone apa beneran dihapus ya

3.2.3. Fixing Typos

Most tweets contained informal language and typos, so they needed to be fixed. Table 5 shows the result of this step.

Table 5. Fixing Typos Result

Tweet	Symbol and Username Removal Result
program nadiem kampus merdeka msib dan iisma dihapus menteri yg ini nih padahal itu bagus2 semua	program nadiem kampus merdeka msib dan iisma dihapus menteri yang ini nih padahal itu bagus2 semua
mau dihapus ah pler nih ajg duitnya habis ke iisma lpdp ga guna relevansi ke disiplin ilmu gue hampir gada anying	mau dihapus ah pler nih anjing duitnya habis ke iisma lpdp tidak berguna relevansi ke disiplin ilmu gue hampir tidak ada anying
damn iisma n msib cuma postpone apa beneran dihapus ya	damn iisma dan msib hanya postpone apa beneran dihapus ya

3.2.4. Tokenization

In this step, each word in the sentence was broken down into tokens. Table 6 shows the result of this step.

Table 6. Tokenization

Tweet	Tokenization Result
program nadiem kampus merdeka msib dan iisma dihapus menteri yang ini nih padahal itu bagus2 semua	[program, nadiem, kampus, merdeka, msib, dan, iisma, dihapus, menteri, yang, ini, nih, padahal, itu, bagus2, semua]
mau dihapus ah pler nih anjing duitnya habis ke iisma lpdp tidak berguna relevansi ke disiplin ilmu gue hampir tidak ada anying	[mau, dihapus, ah, pler, nih, anjing, duitnya, habis, ke, iisma, lpdp, tidak, berguna, relevansi, ke, disiplin, ilmu, gue, hampir, tidak, ada, anying]
damn iisma dan msib hanya postpone apa beneran dihapus ya	[damn, iisma, dan, msib, hanya, postpone, apa, beneran, dihapus, ya]

3.2.5. Stemming

In this step, each word that had been tokenized was reverted to its original form. Table 7 shows the result of this step.

Table 7. Stemming

Tweet	Tokenization Result
[program, nadiem, kampus, merdeka, msib, dan, iisma, dihapus, menteri, yang, ini, nih, padahal, itu, bagus2, semua]	[program, nadiem, kampus, merdeka, msib, dan, iisma, hapus, menteri, yang, ini, nih, padahal, itu, bagus2, semua]
[mau, dihapus, ah, pler, nih, anjing, duitnya, habis, ke, iisma, lpdp, tidak, berguna, relevansi, ke, disiplin, ilmu, gue, hampir, tidak, ada, anying]	[mau, hapus, ah, pler, nih, anjing, duit, habis, ke, iisma, lpdp, tidak, berguna, relevansi, ke, disiplin, ilmu, gue, hampir, tidak, ada, anying]
[damn, iisma, dan, msib, hanya, postpone, apa, beneran, dihapus, ya]	[damn, iisma, dan, msib, hanya, postpone, apa, beneran, hapus, ya]

3.2.6. Removing Stopwords

Indonesian stopwords, such as “dan,” “atau,” “akan,” and “ya,” were removed from the tokenization result. Table 8 shows the result of this step.

Tabke 8. Stopwords Removal

Tweet	Stopwords Removal Result
[program, nadiem, kampus, merdeka, msib, dan, iisma, hapus, menteri, yang, ini, nih, padahal, itu, bagus2, semua]	[program, nadiem, kampus, merdeka, msib, iisma, hapus, menteri, padahal, bagus2, semua]
[mau, hapus, ah, pler, nih, anjing, duit, habis, ke, iisma, lpdp, tidak, berguna, relevansi, ke, disiplin, ilmu, gue, hampir, tidak, ada, anying]	[hapus, pler, anjing, duit, habis, iisma, lpdp, tidak, berguna, relevansi, disiplin, ilmu, gue, hampir, tidak, ada, anying]
[damn, iisma, dan, msib, hanya, postpone, apa, beneran, hapus, ya]	[damn, iisma, msib, hanya, postpone, beneran, hapus]

p-ISSN 2798-253X
e-ISSN 2798-3153

3.3. Model Performance Result

The dataset was into train and test data with the ratio of 80:20 to ensure a sufficient amount of training data. Cross-validation was not used in this study due to the limited dataset size, which consisted of only 368 data points. Splitting the data into multiple folds could result in excessively small training and validation sets, potentially leading to unstable model performance and reduced generalizability. The models performed in this research are SVM, RFC, and XGBoost Classifier. The performance of each model was evaluated based on the accuracy, precision, recall, and F1-score.

3.3.1. Support Vector Machine (SVM)

The SVM model has the accuracy of 83%. This score is considered good and indicates that the model performs well in predicting the classification. In addition to accuracy, the model's performance can be evaluated as shown in Table 9.

Table 9. SVM Performance Result

	Precision	Recall	F1-Score
Positive	0.79	0.86	0.82
Negative	0.87	0.81	0.84

These results highlight the trade-offs between precision, recall, and F1-score for different classes, emphasizing the model's effectiveness in handling class imbalance. It has higher precision for the negative class (fewer false positives) but also has slightly lower recall for this class (misses some true positives). To improve the model's performance, parameter tuning was conducted, resulting in an increase in accuracy to 85.44%. The updated SVM model scores after tuning are shown below in Table 10.

Table 10. Updated SVM Performance Result

	Precision	Recall	F1-Score
Positive	0.83	0.86	0.85
Negative	0.88	0.85	0.86

After parameter tuning, there was a noticeable improvement in precision, recall, and F1-score for both classes with the most significant improvement observed in the negative class recall and the F1-scores of both classes. So, the model became more effective in both identifying and classifying the positive and negative class, with the negative class showing the most improvement after tuning.

3.3.2. Random Forest Classifier (RFC)

The RFC model achieved an accuracy of 83%, which is the same as the SVM model. This result indicates that the model performed well in predicting the classification. Besides accuracy, the model's performance was also assessed, as shown in Table 11.

Table 11. RFC Performance Result

	Precision	Recall	F1-Score
Positive	0.82	0.82	0.82
Negative	0.84	0.84	0.84

The model performed similarly for both classes with slight improvements in Negative class. The F1-scores are balanced, indicating a decent performance of the Random Forest Classifier in distinguishing between the two classes. However, there is room for improvement, especially if the application is sensitive to misclassifications of either class. Therefore, parameter tuning was performed to improve the model.

After parameter tuning, the accuracy of the model increased to 85.44%. The updated RFC model scores after tuning are shown below in Table 12.

p-ISSN 2798-253X
e-ISSN 2798-3153

Table 12. Updated RFC Performance Result

	Precision	Recall	F1-Score
Positive	0.81	0.89	0.85
Negative	0.90	0.82	0.86

Overall, the parameter tuning has led to a model that is better at identifying both classes, with a slight trade-off in recall for negative class but a notable increase in precision. The F1-scores for both classes have improved, indicating that the tuning has enhanced the balance of the model's performance.

3.3.3. XGBoost Classifier

The XGBoost classifier model achieves an accuracy of 78.3%. Besides accuracy, the model's performance was also assessed, as shown in Table 13.

Table 13. XGBoost Classifier Performance Result

	Precision	Recall	F1-Score
Positive	0.76	0.78	0.77
Negative	0.80	0.79	0.80

While both classes exhibited decent performance, the negative class emerged as the stronger performer, with slightly better precision and recall values. The model was more adept at distinguishing negative class instances, resulting in higher accuracy when predicting this class. To further improve performance, additional tuning was applied, leading to a significant boost in the model's effectiveness. After the tuning, the model's performance increased significantly to 84.18%. The updated XGBoost model scores after tuning are shown below in Table 14.

Table 14. XGBoost Classifier Performance Result

	Precision	Recall	F1-Score
Positive	0.79	0.90	0.84
Negative	0.91	0.79	0.84

The model showed a significant improvement after tuning, particularly in precision and recall for the positive class, which had a higher recall value of 0.90. While precision for negative class was higher, recall for positive class was significantly better. This result suggests better performance in identifying positive class instances. Therefore, the model performed well for both classes by maintaining a strong balance as evidenced by similar F1-scores for both classes. This result indicates that the model has effectively learned to balance both classes after tuning.

This improvement highlights the importance of hyperparameter tuning, especially for a parameter-sensitive model like XGBoost. By optimizing key parameters such as learning rate, max depth, and number of estimators, the model was able to refine its classification performance, increasing its overall accuracy from 78.3% to 84.18%. However, despite these improvements, XGBoost still did not surpass SVM and RFC, suggesting that tree-based boosting methods may not be the most effective approach for this particular text classification task. This could be attributed to the high-dimensional nature of text data, where models like SVM, which excel in handling sparse representations, demonstrated superior performance.

3.4. Sentiment Analysis Word Cloud

In this research, in addition to comparing three models for classifying labels for the IISMA program, sentiment analysis was conducted using a word cloud visualization. This approach aims to provide a deeper understanding of the sentiment surrounding the IISMA program by analyzing the frequency and context of key terms, offering valuable insights into public perception. The combination of model comparison and sentiment analysis will contribute to a more comprehensive evaluation of the program's reception and overall sentiment.

p-ISSN 2798-253X
e-ISSN 2798-3153

3.4.1. Negative Label

The word cloud in Fig. 3 visualizes the negative sentiment expressed by individuals who disagree with the discontinuation of the IISMA program. The word cloud highlights the most frequently mentioned terms from each tweet.

Fig. 3 Visualization of most frequent words in negative label.

The word cloud above reflects the key concerns and emotional responses from those opposing the decision. From the word cloud above, frequent appearance of words such as “iisma,” “magang,” “msib,” “merdeka,” “program,” and “kampus” indicates strong emotions related to the loss of the program and its value in terms of independence and opportunities for university students. Words like “kalo” and “gak” also hint at casual and frustrated tones in expressing disbelief or disagreement with the decision. Other words such as “studi independent,” “exchange,” and “manfaat” show that the IISMA program gave positive impact and new experiences to the students. This analysis helps illustrate the emotional weight behind the discourse, reflecting concerns over the impact of this change on students and their future opportunities.

3.4.2. Positive Label

The word cloud in Fig. 4 visualizes the positive sentiment expressed by individuals who agree with the discontinuation of the IISMA program. The word cloud highlights the most frequently mentioned terms from each tweet.

Fig. 4 Visualization of most frequent words in positive label.

Fig. 4 exhibits the sentiment of individuals who agree with the discontinuation of the IISMA program. Keywords like “iisma” and “hapus” appear as the most frequent words which indicates the strong focus on the idea of removing or discontinuing the program. Additionally, words like “msib,” “mending,” and “magang” suggest that people might feel the need to shift towards other educational

opportunities. There is also a significant use of terms like "output" and "anggar," which may imply concerns related to the allocation of resources and national priorities in education. This word cloud represents a more pragmatic and supportive tone toward the changes, perhaps focusing on efficiency, program restructuring, and new opportunities in line with the government's vision.

4. Conclusion

The sentiment analysis and visualizations revealed that positive sentiment toward the discontinuation of the IISMA program outweighed negative sentiment, suggesting general public agreement. Contributing factors included concerns about the program's urgency and measurable outcomes, and a desire to prioritize other national education issues. These findings can inform policymakers in reassessing IISMA's impact and exploring more targeted educational investments, such as improving access to quality education, domestic university infrastructure, or vocational training. SVM and RFC models performed best for sentiment classification, both achieving 85.44% accuracy.

However, limitations exist, including the reliance on Twitter data, a small and demographically skewed sample, and a short observation period. Future research should expand to multiple platforms, include larger and more diverse datasets, and analyze long-term sentiment trends. Additionally, exploring multilingual sentiment models, temporal shifts in opinion, and applying topic modeling to policy discourse would further enrich educational policy analysis.

Acknowledgment

The authors sincerely thank their academic lecturer, Maria Zefanya Sampe, M.Si., M.M., for her invaluable guidance and support throughout the research process. Her insightful feedback played a crucial role in the successful completion of this paper. The authors are deeply grateful for her mentorship and contributions.

References

[1] S.A. Dawya and A. Okvitawanli, "More than studying abroad: The impacts of Indonesian International Student Mobility Awards (IISMA) program to its alumni and society," in *Proc. 2023 Brawijaya Int. Conf.*, Jan. 2024, pp. 617–626, doi: 10.2991/978-94-6463-525-6_69.

[2] N. Risa, D.D. Prasetya, W.N. Hidayat, P.I. Maula, I.M. Wirawan, and S.Y. Setiawan, "Sentiment analysis of 'Kampus Merdeka' on Twitter using support vector machine (SVM) algorithm," in *2024 IEEE 2nd Int. Conf. Elect. Eng. Comput. Inf. Technol. (ICEECIT)*, Nov. 2024, pp. 163–168, doi: 10.1109/ICEECIT63698.2024.10859978.

[3] E. Nurmiati, M.Q. Huda, and S. Mitsalina, "Sentiment analysis and topics modeling on mobile banking reviews of Sharia Bank in Indonesia using naive Bayes and latent dirichlet allocation," in *2024 12th Int. Conf. Cyber IT Serv. Manag. (CITSM)*, Oct. 2024, pp. 1–6, doi: 10.1109/CITSM64103.2024.10775521.

[4] D.A. Fidian, T.N. Fatyanosa, and A. Ridok, "Analisis sentimen terhadap program Indonesian International Student Mobility Award (IISMA) di platform X/Twitter menggunakan TF-IDF dan SVM," *J. Pengemb. Teknol. Inf. Ilmu Komput.,* vol. 1, no. 1, pp. 1–7, Jan. 2025.

[5] P.A.F. Sara, "Analisis sentimen program Kampus Merdeka IISMA berbasis komentar Titkok dan Tweets Twitter menggunakan metode support vector machine dan FASTTEXT," B.S. thesis, Dept. Inform. Eng. Univ. Pendidik. Ganesha, Bali, Indonesia, 2024.

[6] R. Khairunnas, J.A. Pagua, G. Fitriya, and Y. Ruldeviyani, "User sentiment dynamics in social media: A comparative analysis of X and Threads," *IAES Int. J. Artif. Intell. (IJ-AI)*, vol. 14, no. 1, pp. 447–456, Feb. 2025, doi: 10.11591/ijai.v14.i1.pp447-456.

[7] A. Shebl, D. Abriha, A.S. Fahil, H.A. El-Dokouny, A.A. Elrasheed, and Á. Csámer, "PRISMA hyperspectral data for lithological mapping in the Egyptian Eastern Desert: Evaluating the support vector machine, random forest, and XG boost machine learning algorithms," *Ore Geol. Rev.*, vol. 161, Oct. 2023, Art. no 105652, doi: 10.1016/j.oregeorev.2023.105652.

[8] B. AlBadani, R. Shi, and J. Dong, "A novel machine learning approach for sentiment analysis on Twitter incorporating the universal language model fine-tuning and SVM," *Appl. Syst. Innov.*, vol. 5, no. 1, pp. 1–16, Jan. 2022, doi: 10.3390/asi5010013.

[9] R. Obiedat *et al.*, "Sentiment analysis of customers' reviews using a hybrid evolutionary SVM-based approach in an imbalanced data distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.

[10] S. Sönmez, T. Açi, H. Takci, and H. Kekül, "Sentiment analysis for Udemy Reviews with natural language processing and machine learning methods," *Researcher*, vol. 04, no. 02, pp. 184–191, 2024.

[11] N.H. Agjee, O. Mutanga, K. Peerbhay, and R. Ismail, "The impact of simulated spectral noise on random forest and oblique random forest classification performance," *J. Spectroscopy*, vol. 2018, no. 1, pp. 1–8, Jan. 2018, doi: 10.1155/2018/8316918.

[12] A.A. Abdirahman, A.O. Hashi, U.M. Dahir, M.A. Elmi, and O.E.R. Rodriguez, "Comparative analysis of machine learning and deep learning models for sentiment analysis in Somali language," *Int. J. Elect. Electron. Eng.*, vol. 10, no. 7, pp. 41–52, Jul. 2023, doi: 10.14445/23488379/IJEEE-V10I7P104.

[13] S. Liang, "Comparative analysis of SVM, XGBoost and neural network on hate speech classification," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 896–903, Oct. 2021, doi: 10.29207/resti.v5i5.3506.

[14] S. Afrida, "Secondary Data Collection Techniques — Web Scraping and Crawling," Medium. Accessed: April 1, 2025. [Online]. Available: https://yandaafrida.medium.com/secondary-data-collection-techniques-web-scraping-and-crawling-c4071476e62b

[15] M. Wankhade, A.C.S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev*., vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[16] O. Narushynska, V. Teslyuk, A. Doroshenko, and M. Arzubov, "Data sorting influence on short text manual labeling quality for hierarchical classification," Big Data and Cognitive Computing, vol. 8, no. 4, Apr. 2024, Art. no. 8, doi: 10.3390/bdcc8040041.

[17] P.R.A. Savitri, I.M.A. D. Suarjaya, and W.O. Vihikan, "Sentiment analysis of X (Twitter) comments on the influence of South Korean culture in Indonesia," *J. Inf. Syst. Inform*., vol. 6, no. 2, pp. 979–991, Jun. 2024, doi: 10.51519/journalisi.v6i2.749.

[18] Rianto, A.B. Mutiara, E.P. Wibowo, and P.I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J. Big Data*, vol. 8, Dec. 2021, Art. no 26, doi: 10.1186/s40537-021-00413-1.

[19] K. Smelyakov, D. Karachevtsev, D. Kulemza, Y. Samoilenko, O. Patlan, and A. Chupryna, "Effectiveness of preprocessing algorithms for natural language processing applications," in *2020 IEEE Int. Conf. Problems of Infocommun. Sci. Technol. (PIC S&T)*, Oct. 2020, pp. 187–191, doi: 10.1109/PICST51311.2020.9467919.

[20] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep*., vol. 14, Mar. 2024, Art. no 6086 (2024), doi: 10.1038/s41598-024-56706-x.

[21] K. Killamsetty et al., "AUTOMATA: Gradient based data subset selection for compute-efficient hyper-parameter tuning," in *Proc. 36th Int. Conf. Neural Inf. Process. Syst*., Mar. 2022, pp. 28721–287 doi: 10.48550/arxiv.2203.08212.

[22] S. Kumari, Ranganayaki. V. C, S. K, K. Selvi, T. A. Mohanaprakash, and C. Tamilselvi, "Optuna-optimized machine learning technique for accurate diabetes prediction and classification," in *2024 4th Int. Conf. Sustain. Expert Syst. (ICSES)*, Oct. 2024, pp. 1478–1485, doi: 10.1109/ICSES63445.2024.10763036.

[23] R. Guido, M.C. Groccia, and D. Conforti, "A hyper-parameter tuning approach for cost-sensitive support vector machine classifiers," *Soft Comput.*, vol. 27, no. 18, pp. 12863–12881, Sep. 2023, doi: 10.1007/s00500-022-06768-8.

p-ISSN 2798-253X
e-ISSN 2798-3153