Research in Education, Technology, and Multiculture

ISSN: 3025-6763 | Journal's DOI: 10.61436/rietm

Homepage: https://imrecsjournal.com/journals/index.php/rietm

Analysis of Student Sentiment towards the Gratispol Program Using Support Vector Machine (SVM) and TF-IDF Algorithms

Kristian Vandi Hermawan[1], Heny Pratiwi[2], Ahmad Fahrijal Pukeng[1]

[1]Computer Science Program, STMIK Widya Cipta Dharma, Samarinda

[2]Information Systems Program, STMIK Widya Cipta Dharma, Samarinda

ARTICLE INFO

Article History:

Received: 17 July 2026

Accepted: 08 August 2026

Published: 18 August 2026

Keywords:

Sentiment analysis;
Gratispol program;
Support vector machine;
TF-IDF.

Corresponding Author:

2343016@wicida.ac.id

ABSTRACT

Public policies in the education sector require ongoing evaluation to ensure their effective implementation, including the Gratispol Program organized by the East Kalimantan Provincial Government. This study aims to analyze student sentiment based on 539 comments collected from 21 posts on the official Instagram accounts of 14 higher education institutions in East Kalimantan. The comments were initially labeled using IndoBERT, manually validated, converted into TF-IDF features, and class imbalance was then addressed using Random Oversampling on the training data before classification. They were then classified using a linear-kernel Support Vector Machine (SVM) optimized via GridSearchCV. The results show that neutral sentiment dominates (50.5%), followed by negative sentiment (28.9%) and positive sentiment (20.6%). The best model (C = 100) achieved an accuracy of 66.67 %, precision of 66.57%, recall of 66.67%, and an F1-score of 66.45%. Negative comments primarily concerned delays in the disbursement of funds, while positive comments reflected appreciation for the program. These findings provide empirical evidence of student perceptions that can support the East Kalimantan Provincial Government in evaluating and improving the implementation of the Gratispol Program.

To cite this article:

Hermawan, K. V., Pratiwi, H., & Pukeng, A. F. (2026). Analysis of student sentiment towards the gratispol program using support vector machine (svm) and tf-idf algorithms. *Research in Education, Technology, and Multiculture*, *5*(3), 264-279. doi: https://doi.org/10.61436/rietm/v5i3.pp264-279

■ INTRODUCTION

Public policy refers to government decisions that guide the implementation of programs to address societal issues, including those in the education sector. The success of a public policy is largely determined by a continuous evaluation process that involves direct feedback from target groups; without adequate evaluation mechanisms, potential gaps between policy objectives and beneficiaries' experiences are difficult to detect early on. One of the education policies currently being implemented is the Gratispol Program, organized by the East Kalimantan Provincial Government as a tuition assistance program for college students. This program aims to improve access to and equity in higher education, alleviate students' financial burden of tuition, and improve the quality of human resources in East Kalimantan (East Kalimantan Provincial Government, 2025). Higher education institutions play a strategic role in producing competitive, high-quality human resources; thus, the Gratispol Program is expected to expand public access to higher education while supporting improvements in HR quality in East Kalimantan (Abdillah, 2024). As of the third disbursement phase in May 2026, the program has reportedly reached 63,603 students from 52 higher education institutions in East Kalimantan with a budget of Rp288.5 billion (East Kalimantan Provincial Department of Communication and Information Technology, 2026); given the scale of its coverage, evaluating beneficiaries' perceptions has become an urgent and policy-relevant need.

Alongside the implementation of the Gratispol Program, various student responses and opinions have been widely expressed on social media, particularly on Instagram. As one of the social media platforms actively used by students, Instagram serves not only as a channel for disseminating official information from higher education institutions but also as a medium for students to share their experiences, criticisms, suggestions, and support regarding the implementation of the Gratispol Program. These various opinions form positive, neutral, or negative sentiments that reflect students' perceptions of the program's implementation; thus, this information has the potential to serve as valuable input for the government in evaluating and refining policies, provided that these opinions can be systematically and

objectively summarized. Social media sentiment analysis has become one approach for identifying public perceptions of government policies. The results of such analysis can help the government understand public responses and serve as input for evaluations and decision-making regarding policy implementation (Irawan et al., 2023). This approach has also been applied to various government programs in Indonesia to understand public perceptions of policy implementation (Pratiwi et al., 2023).

However, analyzing social media comments is no easy task due to the data's unstructured nature. Instagram comments typically use informal language, abbreviations, slang, emojis, and non-standard spellings and grammatical structures, making sentiment identification more complex (Cortis & Davis, 2021; Alita et al., 2019; Khan et al., 2023). Therefore, a sentiment analysis method is needed to systematically process text data and identify trends in public opinion on a particular issue.

One widely used approach for processing opinion data is sentiment analysis, a technique in the field of Natural Language Processing (NLP) that identifies and classifies opinions in text into specific sentiment categories, such as positive, neutral, and negative (Shaik et al., 2023). In this study, sentiment analysis was conducted through several stages, including data collection, text preprocessing, sentiment labeling, separation of training and test data, feature extraction using TF-IDF, and classification using the Support Vector Machine (SVM) algorithm. A detailed explanation of each stage is presented in the Research Methods section.

During the sentiment labeling stage, this study utilized a pre-trained IndoBERT model to generate initial sentiment labels in the categories of positive, neutral, and negative. Given that Instagram comments often contain informal language, abbreviations, spelling variations, and sarcastic expressions, the initial labeling results were manually reviewed to ensure the labels aligned with each comment's context. The labels from this manual review were subsequently used as ground truth for training and testing the Support Vector Machine (SVM) classification model. During the feature weighting stage, TF-IDF was used to quantify a word's importance based on its frequency in a document and its rarity across the entire corpus, thereby producing a numerical representation suitable for classification (Mee et al., 2021). The numerical representation produced by TF-IDF weighting is then used as input to the SVM algorithm, since it cannot process text data directly. The combination of TF-IDF and SVM is widely used in sentiment analysis research because it produces effective feature representations, thereby improving classification performance over raw text data (Nafis & Awang, 2021).

Support Vector Machine (SVM) is a widely used machine learning algorithm in sentiment analysis because it can handle high-dimensional data and achieve competitive classification performance across various text domains (Philip et al., 2023). These capabilities keep SVM relevant in various current sentiment analysis studies, including those involving social media and education data (Br Sembiring et al., 2026). Various studies have shown that SVM is capable of delivering competitive classification performance across various sentiment analysis domains, including social media, app reviews, and public opinion (Das et al., 2023; Nafis & Awang, 2021; Oladele & Ayetiran, 2023; Paul et al., 2023; Rustam et al., 2019; Br Sembiring et al., 2026). SVM testing in the mobile app review domain using a linear kernel has been shown to achieve accuracy rates exceeding 90% (Ranjan & Mishra, 2022; Srivastava et al., 2022), while in the analysis of public opinion on the social media platform X (Twitter), the use of TF-IDF-based SVMs consistently delivers high classification efficiency (Cam et al., 2024; Shyrokykh et al., 2023). These research findings indicate that SVMs consistently deliver high classification performance across various types of text data and social media platforms in Indonesia. Beyond the Indonesian context, sentiment analysis of student opinions has also been studied internationally: A systematic review on sentiment analysis and opinion mining in educational data highlights its increasingly important role in supporting pedagogical and institutional decision-making (Shaik et al., 2023). Louati et al. (2023) reported that the Support Vector Machine algorithm classified the sentiment of student reviews in higher education settings with good performance, thereby reinforcing the relevance of using SVM in the education domain. The combination of TF-IDF and SVM applied to university students' Google app reviews achieved an accuracy of 93.41% (Ranjan & Mishra, 2022), and a comparative study of TF-IDF-based SVM models and Transformer-based models regarding Indonesian higher education students' opinions on AI adoption reported an SVM accuracy of 82.14% (Ramadhan et al., 2026). In addition to being used independently, the combination of TF-IDF and SVM has also proven effective in various studies. One piece of empirical evidence comes

from a study on public sentiment analysis of Indonesian horror films, in which this combination achieved 82.51% accuracy (Br Sinulingga & Sitorus, 2024). These findings are further supported by Ginanjar et al. (2026), who demonstrated that Support Vector Machines continue to deliver strong classification performance in analyzing social media users' sentiment regarding local policies. These research results indicate that SVM remains a relevant and consistent method across various sentiment analysis contexts. Overall, these studies indicate that the combination of TF-IDF and SVM is a consistent and effective approach for various sentiment analysis tasks across diverse domains. Therefore, the combination of TF-IDF and Support Vector Machine was selected as the methodological basis for classifying students' sentiments toward the Gratispol Program.

Nevertheless, evaluative studies of the Gratispol Program still have fundamental research gaps. First, there is a gap in the representation of the primary target group. Previous research related to this program used the Naïve Bayes algorithm with data sourced from the TikTok platform (Widayani & Arriyanti, 2026), thereby focusing only on general public sentiment and not specifically capturing the perceptions of students, the primary beneficiaries. Second, there is a gap in the selection of data platforms. The use of Instagram as a research data source has not been widely explored, even though this platform serves as the official medium for disseminating information about the Gratispol Program by various universities in East Kalimantan and is actively used by students to express their empirical experiences. Third, there are limitations in methodological optimization, as previous studies have not applied robust combined methods, such as TF-IDF feature weighting and the Support Vector Machine (SVM) classification algorithm, to address the complexity of informal language in evaluating this policy.

To address these gaps, this study offers solid innovations in three fundamental aspects. First, novelty in data sources: this study pioneers the extraction of data from 539 comments on 21 posts by the official Instagram accounts of 14 public and private universities in East Kalimantan. This ensures that the data is drawn from a highly targeted information ecosystem, unlike previous research that relied on general TikTok comments (Widayani & Arriyanti, 2026). Second, novelty in the research subject: this analysis focuses exclusively on student perceptions, ensuring that the resulting sentiment output accurately represents the voices of direct beneficiaries without the bias of general public opinion. Third, methodological novelty: this study applies a hybrid method never before used on the Gratispol topic, specifically, a combination of TF-IDF and linear kernel SVM, which is innovatively enhanced by preliminary sentiment labeling using IndoBERT (manually validated), TF-IDF feature extraction, linear kernel SVM classification, and class imbalance handling via Random Oversampling. Ultimately, this innovation is expected to contribute to research on public policy sentiment analysis, particularly in evaluating students' perceptions of the implementation of the Gratispol Program.

Although sentiment analysis is increasingly applied in evaluating public policy, studies that specifically examine students' perceptions of the Gratispol Program using Instagram data and a combination of TF-IDF and Support Vector Machine (SVM) remain limited. To address this research gap, empirical evidence is needed to identify the distribution of student sentiment, evaluate the classification performance of the proposed approach, and explore its contribution to policy evaluation. Therefore, this study aims to answer the following research questions:

RQ1. What is the distribution of positive, neutral, and negative sentiments expressed by students toward the Gratispol Program based on comments collected from the official Instagram accounts of public and private higher education institutions in East Kalimantan?

RQ2. How does the TF-IDF feature weighting method, combined with the Support Vector Machine (SVM) algorithm, perform in classifying student sentiments toward the Gratispol Program?

RQ3. What insights can be gained from the sentiment classification results to support the evaluation and improvement of the Gratispol Program's implementation?

Based on the identified research gaps and the research questions formulated above, this study aims to analyze student sentiment toward the Gratispol Program by examining comments collected from the official Instagram accounts of public and private higher education institutions in East Kalimantan. This study uses TF-IDF for feature weighting and a Support Vector Machine (SVM) for sentiment classification. The findings are expected to provide a comprehensive understanding of students' perceptions of the Gratispol Program and offer empirical evidence to help the East Kalimantan Provincial Government evaluate and improve the program's implementation.

METHOD

Research Design

This study employed a quantitative approach using sentiment analysis to identify students' sentiments toward the Gratispol Program based on comments collected from Instagram. The proposed framework combines initial labeling with IndoBERT, which was manually validated; feature extraction with TF-IDF; class imbalance handling with RandomOverSampler on the TF-IDF-extracted feature matrix; and classification with a Support Vector Machine (SVM) optimized via GridSearchCV. The entire workflow, including data collection, preprocessing, sentiment labeling, training-test data splitting, class balancing, feature extraction, model training, hyperparameter optimization, and performance evaluation, was implemented in Python on Google Colab. The overall research workflow is illustrated in Figure 1.

Data Collection

The data collection process began by opening the Instagram platform and searching for the keyword "Gratispol" to find posts related to the Gratispol Program. Based on the search results, 21 posts were identified, consisting of 13 posts and 8 Reels from the official accounts of various universities in East Kalimantan, namely Mulawarman University (UNMUL), Muhammadiyah University of East Kalimantan (UMKT), Samarinda Health Sciences College (STIKSAM), Samarinda State Polytechnic (POLNES), Sultan Aji Muhammad Idris State Islamic University of Samarinda (UINSI), Dirgahayu Samarinda Health Sciences College, Mulia University, Borneo International School of Information Technology (STMIK Borneo Internasional), Samarinda Islamic College (STAI Samarinda), Mutiara Mahakam Health Sciences College, Balikpapan Oil and Gas Technical School (STT Migas Balikpapan), Kalimantan Institute of Technology (ITK), Balikpapan State Polytechnic (Poltekba), and Tarakan State Polytechnic (Poltara). These 14 higher education institutions include both public and private institutions across several cities in East Kalimantan; thus, the collected data reflect the geographical and institutional diversity of the beneficiaries rather than originating from a single campus, which could introduce sample bias. Furthermore, comments on each post were collected via web scraping using the IGCommentsExport Tool extension in Google Chrome. The scraping process yielded 754 raw comments, as summarized in Table 1.

Of these 754 raw comments, a filtering process was conducted to ensure only Indonesian-language comments directly related to the Gratispol Program were included, resulting in a final dataset of 539 comments, as summarized in Table 1. This filtering process is crucial for maintaining the internal validity of the study, as irrelevant comments (e.g., promotions for other accounts, spam, or

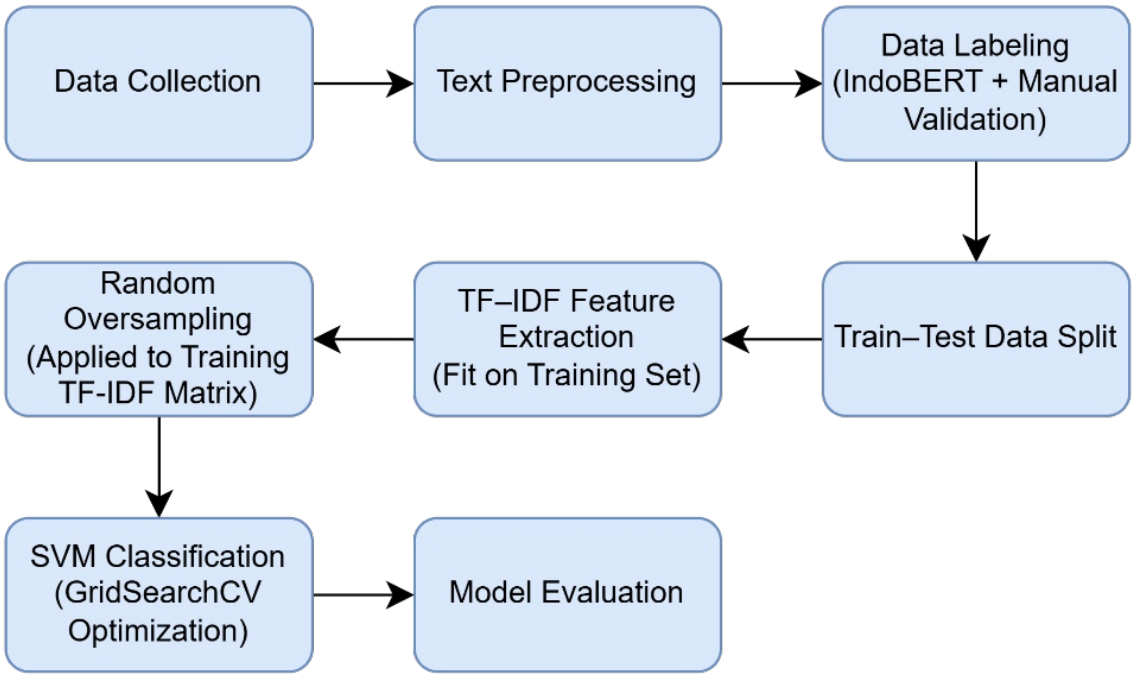

Figure 1. Research workflow

comments in regional or foreign languages that cannot be processed by the IndoBERT model and Indonesian-language NLP libraries) could compromise the quality of labeling and classification if left unchecked.

Table 1. Summary of the research data set

Description	Total
Number of raw comments	754
Number of comments after filtering	539
Platform	Instagram
Topic	Gratispol Program

Text Preprocessing

The text preprocessing stage aims to clean and standardize unstructured comment data for subsequent processing. This stage is performed using the Python programming language in the Google Colab environment, in the following order: (1) data cleaning, (2) case conversion, (3) normalization of non-standard words, (4) tokenization, (5) stopword removal, and (6) stemming.

Data Cleaning

Data cleaning is performed to remove elements from the comment text that are irrelevant to sentiment analysis, including URLs, mentions (@username), hashtags, HTML tags, emojis, numbers, punctuation, and excess spaces. Emoji removal is performed using the Python emoji library to ensure that all Unicode emoji variants, including those from the latest Unicode range, are completely removed.

Case Folding

Case folding is the process of converting all letters in the comment text to lowercase so that words with different capitalization (e.g., "Gratispol" and "GRATISPOL") are recognized as identical.

Normalization of Non-Standard Words

Normalization aims to convert non-standard words, abbreviations, and slang into their standard forms. This process is performed by matching each word or phrase in the comment text against a general Indonesian slang normalization dictionary (slang_indo.xlsx), thereby expanding the range of words that can be normalized.

Tokenization

Tokenization is the process of splitting the comment text into word units (tokens) using the `word_tokenize` function from the Natural Language Toolkit (NLTK) library.

Stopword Removal

Stopword removal is performed to eliminate common words that do not significantly contribute to the sentiment meaning of a sentence, such as conjunctions and pronouns. The stopword list is sourced from the Sastrawi library and supplemented with a custom list tailored to the characteristics of social media language (such as "yg," "nya," "sih," "deh," "wkwk").

Stemming

Stemming is the process of converting a word into its root form by removing affixes, prefixes, and suffixes using the Sastrawi library, which is specifically designed for morphological processing of the Indonesian language. Preprocessing steps are performed to remove irrelevant textual elements, such as usernames, links, emojis, and non-standard words, resulting in a cleaner, standardized text representation suitable for subsequent sentiment analysis. This standardized text is then used in the sentiment labeling and TF-IDF feature extraction stages.

To illustrate the implementation of the data preprocessing workflow, an original comment reading "Kapan pencairan gratispol luar kaltim tahap 2 Pak " begins with a cleaning stage to remove numbers, special characters, and emojis, resulting in the text "Kapan pencairan gratispol luar kaltim tahap Pak." The text is then processed through a case-folding stage to convert all characters to lowercase, followed by a non-standard word-normalization stage that changes the abbreviation "kaltim" to "East Kalimantan." Next, the stopword removal stage is applied to filter out common words, followed by the stemming process to convert inflected words like "pencairan" to their base form, "cair." This series of steps ultimately produces a standardized text representation (token) in the form of: "kapan cair gratispol luar kalimantan timur tahap pak."

Data Labeling

Sentiment labeling in this study was conducted in two stages. In the first stage, all comments were automatically assigned an initial sentiment label using the IndoBERT model, which was specifically tailored for three-class sentiment classification (positive, neutral, and negative) in Indonesian via the

Hugging Face Transformers library. This model generated sentiment labels and confidence scores for each comment. The use of transformer-based contextual embeddings, such as IndoBERT, reflects the growing adoption of transformer architectures for sentiment analysis, owing to their ability to capture contextual and semantic information more effectively than traditional feature-based approaches (Naseem et al., 2020).

The initial labeling of 539 comments was performed automatically using the pre-trained IndoBERT language model. Specifically, the model was accessed via the Hugging Face Transformers library through the repository mdhugol/indonesia-bert-sentiment-classification. During the inference stage, the model was configured with default parameters, enabling text truncation (truncation=True) and limiting the maximum input length to 512 tokens to ensure uniform processing of all text. Although the dataset consisted of only 539 comments, IndoBERT was used as an annotation-support tool rather than as a substitute for human assessment. The use of these automatically generated initial labels allows annotators to verify, confirm, or correct the predicted sentiment, rather than assigning each label entirely from scratch. This semi-automated annotation procedure reduces repetitive manual decision-making, improves annotation efficiency, and helps maintain consistent labeling criteria, while ensuring the reliability of the final sentiment labels through comprehensive manual validation.

In the second stage, all automatically generated labels are manually validated by re-reading the context of each comment. Manual validation is performed because automatic predictions can still make errors in interpreting sarcastic expressions, implicit meanings, neutral but informative questions, or highly context-dependent language (Khan et al., 2023; Alita et al., 2019). As a result, each comment is manually reviewed to ensure that the final sentiment label accurately reflects the intended meaning of the text.

This two-stage annotation procedure was intentionally adopted to combine IndoBERT's computational efficiency with the contextual understanding of human annotators. Thus, IndoBERT is used to support a structured and reproducible annotation process, while manual validation remains the final authority in determining sentiment labels. Although the dataset consists of only 539 comments, manually labeling sentiment from scratch still requires repeated interpretation of ambiguous expressions, slang, and context-dependent opinions. IndoBERT reduces the cognitive workload by providing initial sentiment suggestions, allowing annotators to focus on verification and correction rather than independently determining the sentiment of each comment. This approach improves annotation consistency, reduces labeling bias caused by decision-making fatigue, and establishes a reproducible annotation workflow while maintaining the reliability of the final labels through thorough manual validation. The final distribution of sentiment labels obtained after manual validation is presented in Table 2.

Table 2. Distribution of sentiment labels from manual validation

Sentiment	Total
Positive	111
Neutral	272
Negative	156
Total	539

TF-IDF Feature Extraction

Text data that has undergone preprocessing and labeling is then converted into a numerical representation using the TF-IDF method via the `TfidfVectorizer` function from the scikit-learn library, which automatically calculates Term Frequency (TF), Inverse Document Frequency (IDF), and the final TF-IDF weight for each word across all comment documents. This weighting scheme has also proven effective for large-scale analysis of social media text related to politics and policy, such as in the sentiment analysis of British Members of Parliament's tweets regarding Brexit (Mee et al., 2021), as well as in Indonesian-language text with an imbalanced class distribution (Fikri & Sarno, 2019). To avoid data leakage, the TF-IDF adjustment process (vocabulary construction and IDF value calculation) is performed only on the training data. In contrast, the test data undergoes only the transformation process using the vocabulary and IDF values previously learned from the training data, without affecting the resulting weight construction.

$$
\begin{aligned}
W_{tf}(t,d) &= \begin{cases} 1+\log_{10}(tf(t,d)), & \text{if } tf(t,d)>0 \\ 0, & \text{if } tf(t,d)=0 \end{cases} \\
idf(t) &= \log_{10}\left(\frac{N}{df(t)}\right) \\
W(\mathrm{t},\mathrm{d}) &= W_{\mathrm{tf}}(\mathrm{t},\mathrm{d}) \times idf(\mathrm{t})
\end{aligned}
\tag{1}
$$

Description: tf(t,d) is the frequency of occurrence of term t in document d. Wtf(t,d) is the term frequency weight. df(t) is the number of documents containing term t, and N is the total number of documents in the corpus. W(t,d) is the final TF-IDF weight of term t in document d (Mee et al., 2021). The result of the TF-IDF weighting process is a numerical matrix that represents each document as a word weight vector, which is then used as input features for classification algorithms such as Support Vector Machines. To prevent information leakage from the test data to the training data (data leakage), the vocabulary and IDF values in this study were specifically adjusted from the training data. In contrast, the test data underwent only a transformation process using the previously learned vocabulary (Mee et al., 2021), as described in the Data Separation section.

Separation of Training and Test Data

After the preprocessing and sentiment labeling stages, the dataset is split into training data (training set) and test data (test set) using a 70:30 ratio with the `train_test_split` function from the scikit-learn library. Data splitting was performed in a stratified manner based on sentiment labels to ensure that the proportion of each sentiment class remained representative in both the training and test sets, thereby reducing potential bias from random data splitting. The resulting training data was then used to extract TF-IDF features. Next, the RandomOverSampler was applied to the TF-IDF-derived feature matrix to balance the class distribution before SVM training.

Handling Imbalanced Data (Random Oversampling)

The distribution of sentiment classes in the training data shows class imbalance, where the Neutral class contains more instances than the Positive and Negative classes. Class imbalance can cause the model to be more inclined to predict the majority class, thereby reducing its ability to recognize minority classes. This problem is common in sentiment analysis datasets, necessitating the addressing of class imbalance before classification (Srinivasan & Subalalitha, 2023; Taskiran et al., 2025). Therefore, the class distribution imbalance in the training data is addressed using the RandomOverSampler method from the imbalanced-learn library. The oversampling process is performed after TF-IDF feature extraction, so samples are added to the numerical representation produced by vectorization rather than to the raw text data. This approach preserves the Inverse Document Frequency (IDF) calculation, ensuring that feature weights remain unchanged due to document duplication at the text stage. The oversampling process is applied only to the training set to avoid data leakage during model training. This method balances the class distribution by duplicating samples in the numerical feature matrix, derived from TF-IDF extraction, for the minority class, thereby making the number of samples per class more balanced. To ensure an unbiased evaluation, Random Oversampling is applied only to the training data, while the test data retains its original class distribution. Consequently, model performance is evaluated using previously unseen data that reflects the distribution of student comments in the real world.

Data Analysis

Support Vector Machine (SVM)

Sentiment classification in this study uses the Support Vector Machine (SVM) algorithm, which finds the optimal hyperplane—or separating plane—with the largest margin (maximum margin) that separates the data into different classes, thereby improving the model's ability to classify previously unseen data. The SVM classification model can be expressed as follows.

$$f(x) = w^T x + b \quad (2)$$

Description: f(x) is a decision function used to determine the class of a data point. w is a weight vector that determines the direction of the hyperplane. x is a data feature vector represented using TF-IDF; $w^T x$ is the result of multiplying the weight vector by the feature vector. b is a bias value that shifts the position of the hyperplane. If the value of f(x) lies on the positive side of the hyperplane, the data is classified into the positive class, and vice versa. In this study, classification was performed on three sentiment classes (positive, neutral, negative) using a one-versus-rest approach. The model was trained using a linear kernel with parameter C=1 as the base model, and then further optimized as described in the Model Optimization stage. Grid search remains a widely used and reliable strategy for tuning SVM hyperparameters compared with other search algorithms (Wainer & Fonseca, 2021), particularly for short-text sentiment analysis on social media (Liu et al., 2024).

Model Evaluation

The performance of the built SVM classification model is evaluated on test data unseen during training. The evaluation is conducted using a Confusion Matrix to

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1\text{-}Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (6)$$

calculate accuracy, precision, recall, and F1-score, using the following formulas.

In this equation, TP (True Positive) is the number of data points correctly predicted to belong to a specific class, TN (True Negative) is the number of data points correctly predicted not to belong to that class, and FP (False Positive) is the number of data points that do not actually belong to that class but are predicted to belong to it. At the same time, FN (False Negative) is the number of data points that actually belong to that class but are predicted to belong to another class. Precision indicates the model's accuracy in predicting a class; recall indicates the model's ability to identify all data points that truly belong to that class. At the same time, the F1-score is the harmonic mean of precision and recall, providing a more balanced measure of the model's performance. The evaluation metrics are calculated for three-class sentiment classification (positive, neutral, and negative) and reported using macro-averages and weighted averages so that performance on the minority class is not overshadowed by the dominance of the majority class.

■ RESULTS AND DISCUSSION

Research Dataset, Preprocessing Results, and Sentiment Labeling Results

The research dataset was obtained from Instagram user comments on posts discussing the Gratispol Higher Education Program, a tuition-waiver program organized by the East Kalimantan Provincial Government. Data were collected via web scraping from the comments sections of official program posts, then filtered to obtain 539 Indonesian-language comments relevant to the research topic, which were used in all subsequent stages of analysis, as previously described in Table 1.

The results of applying the preprocessing steps to one of the comments in the dataset are shown in the following workflow example.

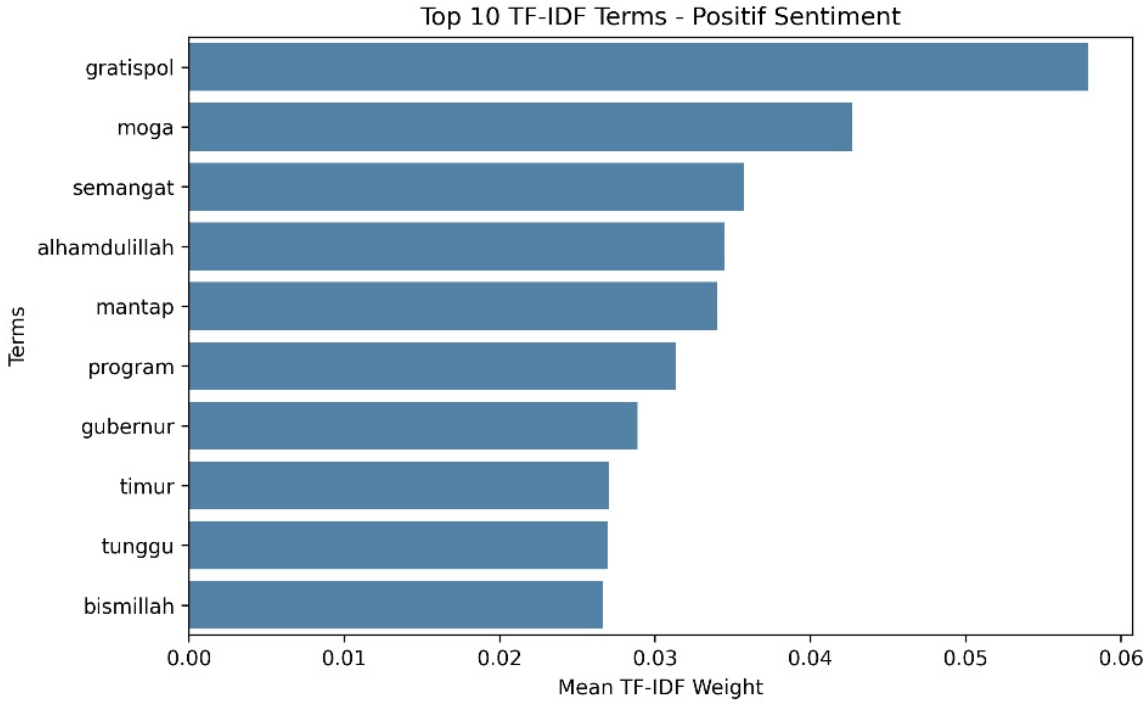

Figure 2. Top 10 IF-IDF terms for positive sentiment

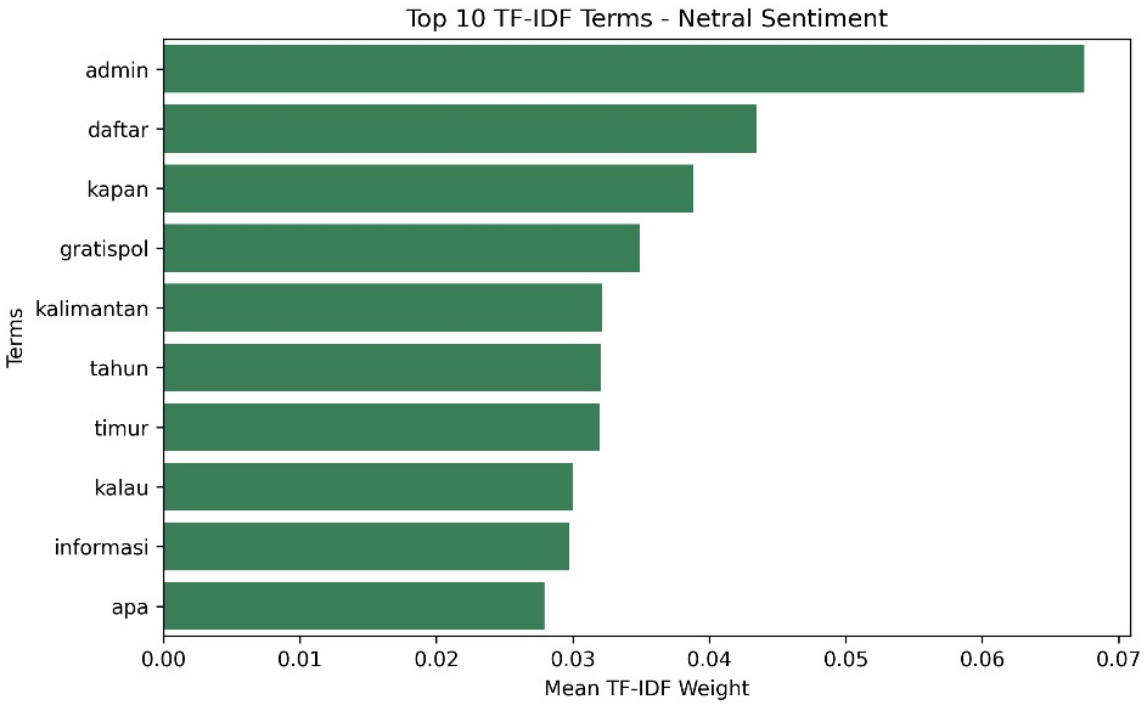

Figure 3. Top 10 TF-IDF terms for neutral sentiment

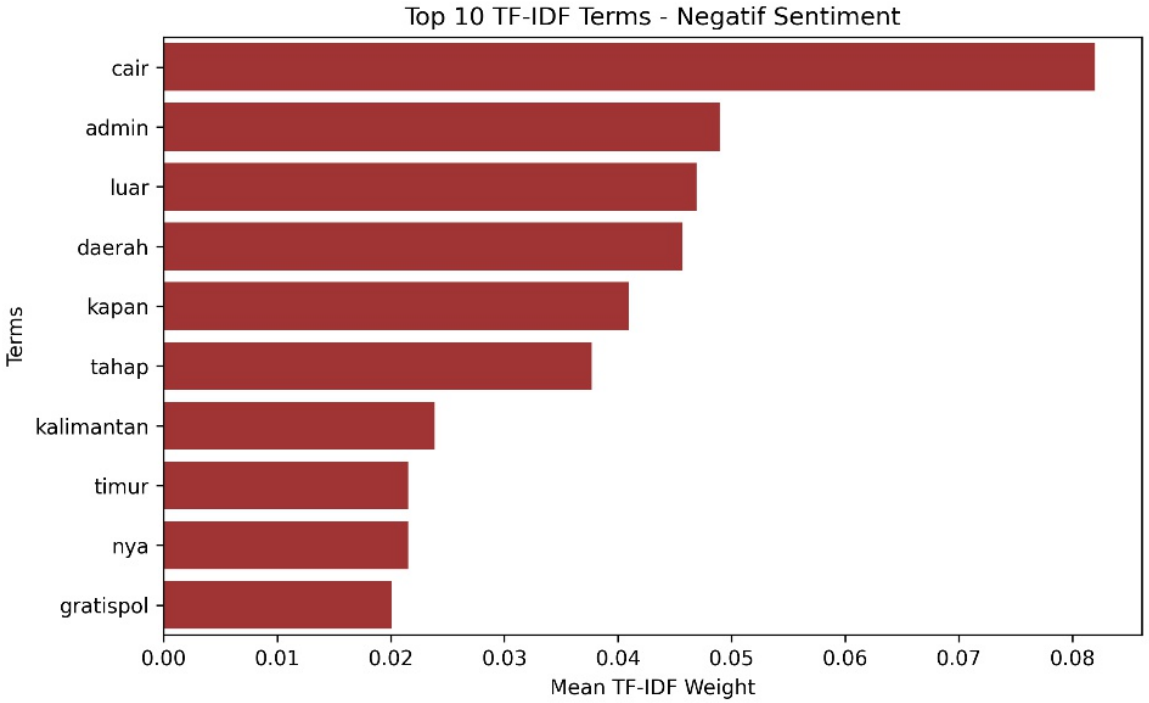

Figure 4. Top 10 TF-IDF terms for negative sentiment

Consistent with these steps, all 539 comments were processed through the same six stages, yielding a uniform text representation (text_final) as input to the labeling and feature extraction processes in subsequent stages. The distribution of sentiment labels generated from manual validation of the 539 comments is dominated by the Neutral class (272 comments, 50.5%), followed by the Negative class (156 comments, 28.9%), and the Positive class (111 comments, 20.6%), as shown in Table 2.

Top 10 TF-IDF Terms per Sentiment Class

A horizontal bar chart was used to display the 10 words with the highest average Term Frequency-Inverse Document Frequency (TF-IDF) scores in each sentiment class. This approach allows readers to more clearly compare the exact mathematical weights between words. In the positive sentiment class, the words with the highest weights that appeared included "gratispol," "moga" (hopefully), "semangat" (spirit), "alhamdulillah" (praise be to God), "mantap" (great), and "gubernur" (governor). This indicates that positive comments generally contain appreciation, prayers, enthusiasm, gratitude, and public support for the continuation of the Gratispol Program, as shown in Figure 2. In the neutral sentiment class, the words with dominant weights are "admin," "register," "when," "year," "if," and "information." This indicates that most comments in this category are questions directed to the organizers (admin) about registration information, program mechanisms, and schedules, rather than subjective evaluations of the program, as shown in Figure 3. In the negative sentiment class, the words with the highest weights that appeared were "disbursement," "admin," "outside," "region," "when," and "phase." This indicates that most user complaints centered on questions about when funds would be disbursed, delays at each stage, and confusion regarding requirements, particularly for students residing outside East Kalimantan, as shown in Figure 4.

TF-IDF Weighting

In this study, TF-IDF weighting was applied to the training and testing text data using the TfidfVectorizer library. After the training data was converted to numerical representations using TF-IDF, class distribution was balanced using RandomOverSampler. The feature matrix resulting from oversampling was then used as input for the SVM model training process. This process produced feature matrices of sizes (570, 855) for the training data and (162, 855) for the test data. The resulting feature space consists of 855 unique vocabulary terms learned from the training dataset. No additional feature selection or document frequency thresholds were applied because the resulting feature space remained manageable for linear Support Vector Machine (SVM) classification. Linear SVM is widely recognized for its effectiveness in handling sparse, high-dimensional text data without requiring explicit feature reduction. At the same time, TF-IDF inherently reduces the influence of less informative terms by assigning lower weights to words that appear frequently across the corpus. Therefore, all extracted TF-IDF features are retained to preserve complete lexical information for sentiment classification. This decision also allows the proposed TF-IDF and SVM framework to be evaluated without introducing additional feature selection bias.

Data Split

Following the preprocessing stage and manual sentiment labeling, the dataset was split into training and testing sets at a 70:30 ratio using a stratified split to preserve the proportion of each sentiment class in both subsets. The split resulted in 377 training samples and 162 testing samples, while maintaining the original class distribution. The training data was then converted into a numerical representation using TF-IDF. Next, a RandomOverSampler was applied to the TF-IDF-derived feature matrix to balance the class distribution before training the SVM model.

Application of Random Oversampling

To prevent data leakage and preserve the mathematical integrity of word weights, the feature extraction and class imbalance handling steps were performed sequentially. First, the data was divided into training and test sets. Next, Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction was performed exclusively on the training data using the scikit-learn library. This step is crucial to ensure that the Document Frequency (DF) values are calculated from the natural text distribution. After the TF-IDF numerical vector matrix is formed, handling of the imbalanced data is performed by applying the Random Oversampling algorithm (using the RandomOverSampler module from the imbalanced-learn library) exclusively to the resulting numerical matrix from the extraction. This approach avoids direct text duplication before extraction, which could artificially increase the Document Frequency (DF) for the minority class and compromise the validity of the IDF weights. The change in the distribution of sample counts in the training data before and after applying the Random Oversampling algorithm is shown in Figure 5.

SVM Classification Results and Model Evaluation

The prediction results from the best SVM model, obtained via parameter optimization (GridSearchCV with C=100), are summarized in the confusion matrix shown in Figure 6. This model was evaluated on 162 test comments that retained their original class distribution,

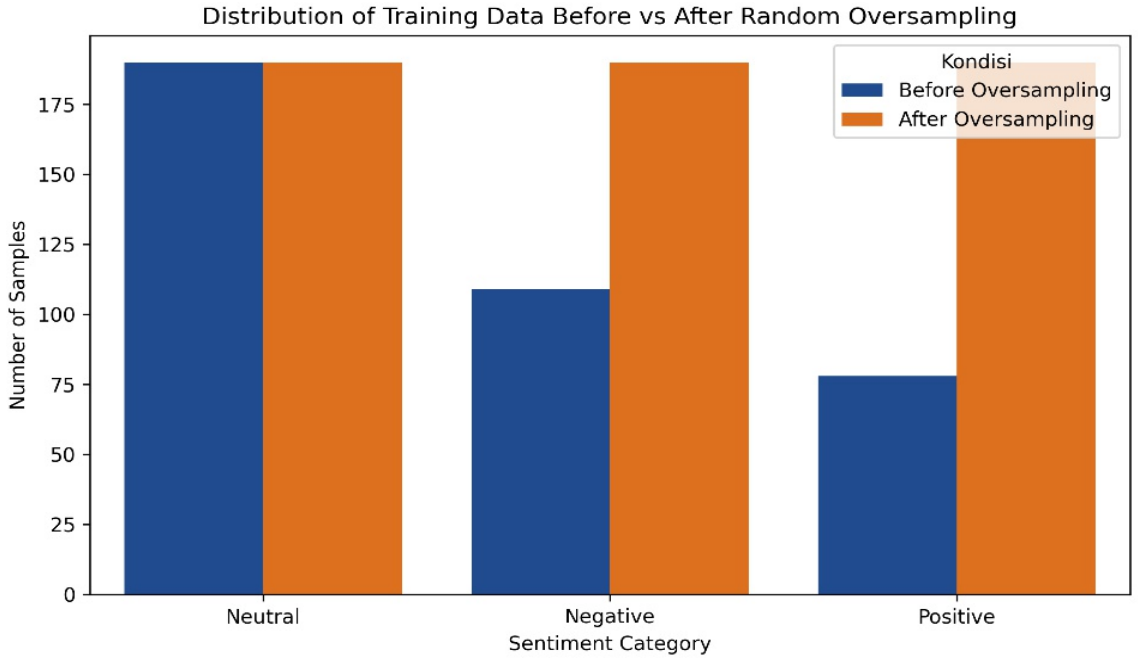

Figure 5. Bar chart showing the distribution of training data before and after random oversampling

providing a realistic assessment of performance. Based on a detailed evaluation using the `classification_report` module from scikit-learn, the model performed best at recognizing the Neutral class, as evidenced by the highest recall value of 0.74 and a precision of 0.69 (successfully classifying 61 out of 82 Neutral comments correctly). For the Positive class, the model achieved a precision of 0.67 but a lower recall of 0.55, indicating that although its positive predictions were fairly accurate, it still missed some comments that should have been labeled positive. Meanwhile, the Negative class showed balanced performance with precision and recall values of 0.62 each. This distribution of classification errors indicates that applying the Random Oversampling algorithm to the TF-IDF matrix effectively helps the SVM model recognize patterns in the minority class, even though the dominant class still achieves the optimal recognition rate.

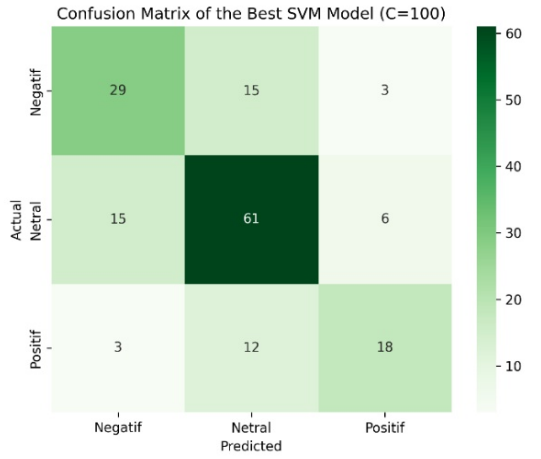

Figure 6. Confusion matrix for the optimized SVM model (linear kernel, c=100)

To obtain the optimal model parameters, a search for the best parameters was conducted using GridSearchCV with 5-fold cross-validation on the linear kernel, with candidate C values of 0.01, 0.1, 1, 10, and 100. The average cross-validation score (weighted F1) for each C value is presented in Table 3.

Table 3. Results of parameter search using GridSearchCV

Value	Average F1 (Cross-Validation)
0.01	0.6445
0.10	0.6428
1.00	0.8216
10.00	0.8332
100.00	0.8334

The search results indicate that the best parameters were obtained at C = 100, yielding the highest average weighted F1 score of 0.8334 in 5-fold cross-validation. The optimized model was then evaluated on an

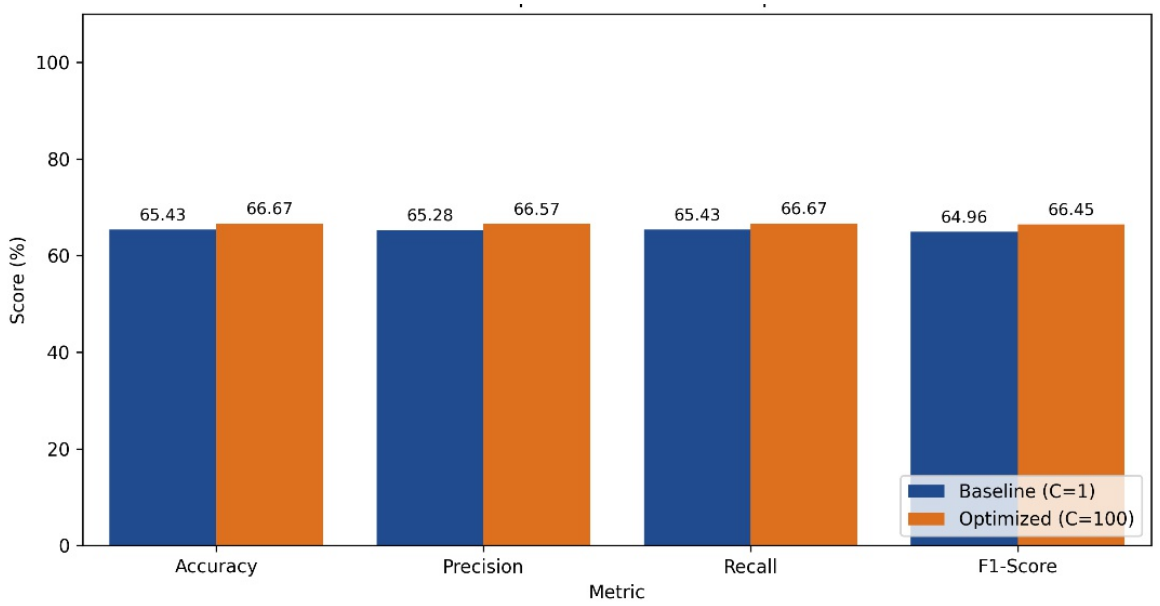

Figure 7. Comparison of the performance of the baseline svm model and the results of parameter optimization

independent test set and compared with the baseline model's performance, as illustrated in Figure 7.

Why Is the Neutral Class the Largest Class?

As illustrated in Figure 8, neutral comments primarily relate to the fund disbursement schedule (22.0%) and eligibility requirements (21.5%), followed by general information inquiries (14.5%), registration procedures (6.0%), and user interactions (18.0%). This percentage distribution clearly shows that the majority of neutral comments are more focused on seeking administrative information than on expressing opinions, which ultimately explains the consistent dominance of the Neutral class in the dataset.

Why Is the Positive Class the Smallest Class?

The Positive Class is the smallest, comprising 111 comments (20.6% of the total data). This is because comments with positive sentiment on social media are generally expressed briefly, such as "Good" or "Hope it goes well," resulting in fewer comments that explicitly express appreciation compared to comments containing questions (neutral) or complaints (negative). This observation is consistent with previous research showing that users post far more negative content than positive content on social media platforms (Cortis & Davis, 2021).

Why Is the Negative Class Difficult for the Model to Recognize?

Based on Table 4, the Negative class has the lowest recall (0.55) in the baseline model, indicating that the model still frequently misclassifies negative comments as neutral. This is due to the characteristics of the dataset's negative comments, which contain a lot of sarcasm, abbreviations, and slang. For example, words like "wkwkwk" (laughing), "jir," "lah," and "hadeh" are used to indirectly express sarcasm or disappointment. TF-IDF-based models, which rely on the literal occurrence of words, struggle to capture this kind of implicit meaning because these words do not explicitly convey negative sentiment in isolation. These findings are consistent with the limitations of word-frequency-based approaches documented

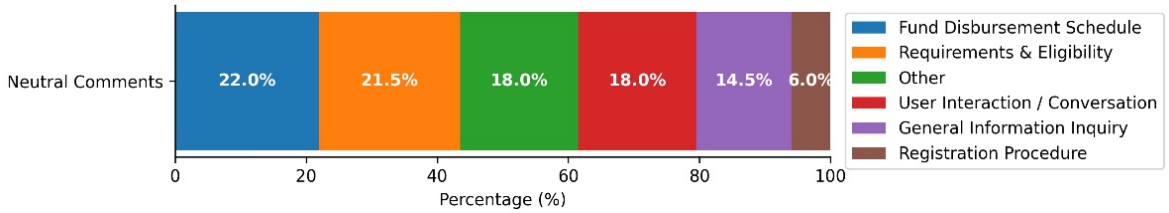

Figure 8. Percentage distribution of topics in the neutral comments category

Table 4. Classification performance of the basic SVM model (linear kernel, c = 1)

Label	Precision	Recall	F1 Score	Support
Negative	0.59	0.55	0.57	47
Neutral	0.67	0.76	0.71	82
Positive	0.65	0.52	0.58	33
Macro Average	0.64	0.61	0.62	162
Weighted Average	0.65	0.65	0.64	162
Overall Accuracy: 65.43%				

in previous literature (Khairani et al., 2024; Khan et al., 2023; Alita et al., 2019; Shyrokykh et al., 2023).

Why Does not Random Oversampling Always Improve Accuracy?

This finding is consistent with previous research showing that addressing class imbalance does not always yield substantial improvements in overall classification accuracy (Taskiran et al., 2025). However, it can enhance the model's ability to recognize the minority class. In this study, Random Oversampling balances the training data by duplicating existing minority-class samples rather than generating synthetic instances. Although this approach improves the representation of the Positive and Negative classes during training, it does not introduce additional linguistic diversity. Consequently, improvements in minority-class recognition are not always accompanied by significant gains in overall model performance. This is reflected in the discrepancy between the high cross-validation scores presented in Table 3 and the lower performance obtained on the independent test set, indicating that the model is less effective at generalizing to unseen data. Nevertheless, Random Oversampling provides a balanced training distribution while preserving the dataset's original textual characteristics, making it suitable for the proposed TF-IDF and SVM classification frameworks. A similar pattern has also been reported when addressing class imbalance in SVM classification of informal social media comments, where resampling techniques for the minority class do not always directly increase overall accuracy (Cervantes et al., 2020; Fernández et al., 2018).

Why Is the Model's Accuracy Moderate?

These findings are consistent with research by Br. Sembiring et al. (2026), who reported that TF-IDF-based SVMs remain a competitive baseline approach. However, their performance generally falls short of Transformer-based models when dealing with complex linguistic contexts and imbalanced class distributions (Br. Sembiring et al., 2026). The moderate classification performance observed in this study can be attributed to the characteristics of Instagram comments, which contain informal language, abbreviations, emojis, sarcasm, spelling variations, and relatively short text. These characteristics limit TF-IDF's ability to capture contextual and semantic information because it relies solely on term frequency and weighting rather than contextual word representations. As a result, the SVM classifier may struggle to distinguish between comments with similar vocabulary but differing underlying meanings. Nevertheless, the results demonstrate that the combination of TF-IDF and SVM remains a practical foundational approach for classifying the sentiment of Indonesian social media text and provides useful insights into public opinion regarding the Gratispol Program. These findings of moderate accuracy are also consistent with other studies showing that TF-IDF-based SVMs perform more moderately on short, informal social media text than on more structured text domains (Rustam et al., 2019; Shyrokykh et al., 2023; Das et al., 2023).

Comparison Between the Base Model and the GridSearchCV Model

The GridSearchCV results presented in Table 4 show that the optimal hyperparameter was C = 100, achieving the highest average weighted F1 score of 0.8334 during 5-fold cross-validation. Compared to the baseline model (C = 1), the optimized model (C = 100) was then evaluated on an independent test set. A higher value of C imposes a stronger penalty on classification errors during training, allowing the classifier to build more discriminative decision boundaries for sentiment classes. Therefore, the SVM model optimized with C = 100 was selected as the final model in this study.

■ CONCLUSION

This study aimed to analyze student sentiment toward the Gratispol Program based on 539 comments from the official Instagram accounts of 14 public and private higher education institutions in East Kalimantan, using a combination of manually validated IndoBERT labeling, TF-IDF feature weighting, class imbalance handling using RandomOverSampler on the TF-IDF feature matrix, and SVM classification. The findings indicate that the research objectives were successfully achieved. The results show that students' perceptions of the Gratispol Program were dominated by neutral sentiment (50.5%), consisting largely of informative questions about the mechanism and schedule for fund disbursement, followed by negative sentiment (28.9%) centered on complaints about delays in fund disbursement, particularly for students outside East Kalimantan, and positive sentiment (20.6%), which consisted of appreciation and hopes that the program would continue. The linear kernel SVM classification model was optimized using GridSearchCV (C = 100), yielding the best performance with an accuracy of 66.67%, precision of 66.57%, recall of 66.67%, and an F1-score of 66.45%. This performance is comparable to the baseline model, with a slight improvement in precision and F1-score after hyperparameter optimization.

In practical terms, these findings provide evaluative insights for the East Kalimantan Provincial Government: the timeliness and transparency of information regarding fund disbursement, particularly for students living outside the campus area, are top priorities that need to be improved in the future implementation of the Gratispol Program. Academically, this study reinforces the empirical evidence that the combination of TF-IDF and SVM remains a competitive approach for Indonesian-language sentiment analysis on social media data, while also highlighting its limitations in handling short, informal, and sarcasm-laden comments.

This study has limitations that warrant attention in future research. First, the relatively small dataset size (539 comments) may affect the model's ability to generalize to a broader population of comments. Second, the TF-IDF-based approach cannot capture semantic context or word order, making it difficult to recognize sarcasm and implicit meanings, as indicated by the low recall for the negative class. Therefore, future research is recommended to (1) expand the dataset scope by extending the data collection period and increasing the number of institutional accounts, (2) explore contextual embedding-based approaches, such as fine-tuning IndoBERT as a classifier, and (3) conduct aspect-based sentiment analysis to identify more specific factors influencing student satisfaction and complaints regarding the Gratispol Program.

■ REFERENCES

Abdillah, F. (2024). *Peran perguruan tinggi dalam meningkatkan kualitas sumber daya manusia di Indonesia* [The role of universities in improving the quality of human resources in Indonesia]. *Educazione: Jurnal Multidisiplin, 1*(1), 13–24. https://doi.org/10.37985/educazione.v1i1.4

Abo, M. E. M., Idris, N., Mahmud, R., & Khan, S. K. (2021). A multi-criteria approach for Arabic dialect sentiment analysis for online reviews: Exploiting optimal machine-learning algorithm selection. *Sustainability, 13*(18), 10018. https://doi.org/10.3390/su131810018

Alita, D., Priyanta, S., & Rokhman, N. (2019). Analysis of emoticon and sarcasm effect on sentiment analysis of Indonesian language on Twitter. *Journal of Information Systems Engineering and Business Intelligence, 5*(2), 100–109. https://doi.org/10.20473/jisebi.5.2.100-109

Br Sembiring, A. B. S., Robet, R., & Hoki, L. (2026). Comparison of IndoBERT and SVM performance in sentiment analysis of digital education platforms. *Sinkron, 10*(1). https://doi.org/10.33395/sinkron.v10i1.15472

Br Sinulingga, J., & Sitorus, Z. (2024). *Analisis sentimen publik terhadap film horor Indonesia menggunakan algoritma Support Vector Machine* [Analysis of public sentiment towards Indonesian horror films using the Support Vector Machine algorithm]. *Jurnal Media Informatika Budidarma, 8*(2), 754–762. https://doi.org/10.30865/mib.v8i2.7541

Cam, H., Cam, A. V., Demirel, U., & Ahmed, S. (2024). Sentiment analysis of financial Twitter posts with machine learning classifiers. *Heliyon, 10*(1), e23784. https://doi.org/10.1016/j.heliyon.2023.e23784

Cervantes, J., García-Lamont, F., Rodríguez-Mazahua, L., & López, A. (2020). Data mining with support vector machines: A review. *Neurocomputing, 408,* 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Cortis, K., & Davis, B. (2021). Over a decade of social opinion mining: A systematic

review. *Artificial Intelligence Review, 54,* 4873–4965. https://doi.org/10.1007/s10462-021-10030-2

Das, R. K., Islam, M., Hasan, M. M., Razia, S., Hassan, M., & Khushbu, S. A. (2023). Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models. *Heliyon, 9*(9), e20281. https://doi.org/10.1016/j.heliyon.2023.e20281

Dinas Komunikasi dan Informatika Provinsi Kalimantan Timur. (2026). *Gratispol cair Rp288 miliar; 63 ribu mahasiswa Kaltim terima manfaat.* https://diskominfo.kaltimprov.go.id/berita/gratispol-cair-rp288-miliar-63-ribu-mahasiswa-kaltim-terima-manfaat

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research, 61,* 863–905. https://doi.org/10.1613/jair.1.11192

Fikri, M., & Sarno, R. (2019). A comparative study of sentiment analysis using SVM and SentiWordNet. *Indonesian Journal of Electrical Engineering and Computer Science, 13*(3), 902–909. https://doi.org/10.11591/ijeecs.v13.i3.pp902-909

Ginanjar, I., Shabir, A. M., Pravitasari, A. A., Pangastuti, S. S., Darmawan, G., & Sukono. (2026). Sentiment analysis of X users regarding Bandung Regency using Support Vector Machine. Applied Sciences, 16(1), 560. https://doi.org/10.3390/app16010560

Irawan, D., Sensuse, D. I., Putro, P. A. W., & Prasetyo, A. (2023). Public response to the legalization of the Criminal Code Bill with Twitter data sentiment analysis. *International Journal of Advanced Computer Science and Applications, 14*(2). https://doi.org/10.14569/IJACSA.2023.0140236

Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). *Pengaruh tahapan preprocessing terhadap model Indobert dan Indobertweet untuk mendeteksi emosi pada komentar akun berita Instagram* [The influence of preprocessing stages on the Indobert and Indobertweet models for detecting emotions in Instagram news account comments]. *Jurnal Teknologi Informasi dan Ilmu Komputer.* https://doi.org/10.25126/jtiik.1148315

Khan, M. Y., Ahmed, T., Siddiqui, M. S., & Wasi, S. (2023). Cognitive relationship-based approach for Urdu sarcasm and sentiment classification. *IEEE Access, 11,* 126661–126690. https://doi.org/10.1109/ACCESS.2023.3325048

Liu, X.-Y., Zhang, K.-Q., Fiumara, G., De Meo, P., & Ficara, A. (2024). Adaptive evolutionary computing ensemble learning model for sentiment analysis. *Applied Sciences, 14*(15), 6802. https://doi.org/10.3390/app14156802

Louati, A., Louati, H., Kariri, E., Alaskar, F., & Alotaibi, A. (2023). Sentiment analysis of Arabic course reviews of a Saudi university using Support Vector Machine. *Applied Sciences, 13*(23), 12539. https://doi.org/10.3390/app132312539

Mee, A., Homapour, E., Chiclana, F., & Engel, O. (2021). Sentiment analysis using TF–IDF weighting of UK MPs' tweets on Brexit. *Knowledge-Based Systems.* https://doi.org/10.1016/j.knosys.2021.107238

Nafis, N. S. M., & Awang, S. (2021). An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *IEEE Access, 9*, 52177–52192. https://doi.org/10.1109/ACCESS.2021.3069001

Naseem, U., Razzak, I., Musial, K., & Imran, M. (2020). Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Generation Computer Systems, 113,* 58–69. https://doi.org/10.1016/j.future.2020.06.050

Oladele, T. M., & Ayetiran, E. F. (2023). Social unrest prediction through sentiment analysis on Twitter using a support vector machine: An experimental study on Nigeria's #EndSARS. *Open Information Science, 7*(1), 20220141. https://doi.org/10.1515/opis-2022-0141

Paul, Y., ul Nisa, S., Hussain, H., & Singh, R. (2023). Twitter sentiment analysis using TF-IDF and machine learning classifiers. *International Journal on Recent and Innovation Trends in Computing and Communication, 11*(3), 693–701. https://doi.org/10.17762/ijritcc.v11i3.11515

Pemerintah Provinsi Kalimantan Timur. (2025). *Peraturan Gubernur Provinsi Kalimantan Timur Nomor 24 Tahun 2025 tentang Bantuan Biaya Pendidikan bagi Mahasiswa pada Perguruan Tinggi.* https://jdih.kaltimprov.go.id/storage/peraturan/PERGUB_24_2025_(1).pdf

Philip, J., VeerasekharReddy, B., Harshini, M., Haritha, I. V. S. L., Patil, S., & Shareef, S. K. (2023). A comparative study of text classification using selective machine

learning algorithms. *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS),* 482–484. https://doi.org/10.1109/ICICCS56967.2023.10142474

Pratiwi, C. D., Damayanti, R. W., & Laksono, P. W. (2023). Public sentiment analysis to support Indonesian government induction stove program. *E3S Web of Conferences, 465,* 02006. https://doi.org/10.1051/e3sconf/202346502006

Ramadhan, H. S., Akbar, A. S., Sinaga, K. Y., Muthoharoh, L., Satria, A., & Manullang, M. C. (2026). *Sentiment analysis of AI adoption in Indonesian higher education using machine learning and transformer-based models. arXiv.* https://doi.org/10.48550/arXiv.2604.27439

Ranjan, S., & Mishra, S. (2022). Perceiving university students' opinions from Google app reviews. *Concurrency and Computation: Practice and Experience, 34*(10), e6800. https://doi.org/10.1002/cpe.6800

Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., & Choi, G. S. (2019). Tweets classification on the base of sentiments for US airline companies. *Entropy, 21*(11), 1078. https://doi.org/10.3390/e21111078

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal, 2,* 100003. https://doi.org/10.1016/j.nlp.2022.100003

Shyrokykh, K., Girnyk, M., & Dellmuth, L. (2023). Short text classification with machine learning in the social sciences: The case of climate change on Twitter. *PLOS ONE, 18*(9), e0290762. https://doi.org/10.1371/journal.pone.0290762

Srinivasan, A., & Subalalitha, C. N. (2023). Sentilyser: Sentiment analysis framework for imbalanced code-mixed social media text. *Journal of King Saud University - Computer and Information Sciences, 35*(1), 120–131. https://doi.org/10.1016/j.jksuci.2022.11.005

Srivastava, A., Jain, A., & Gupta, P. (2022). TF-IDF based feature extraction and machine learning algorithms for sentiment classification. *International Journal of Information Management Data Insights, 2*(2), 100101. https://doi.org/10.1016/j.jjimei.2022.100101

Taskiran, S. F., Turkoglu, B., Kaya, E., & Asuroglu, T. (2025). A comprehensive evaluation of oversampling techniques for enhancing text classification performance. *Scientific Reports, 15,* 21631. https://doi.org/10.1038/s41598-025-05791-7

Wainer, J., & Fonseca, P. (2021). How to tune the RBF SVM hyperparameters? An empirical evaluation of 18 search algorithms. *Artificial Intelligence Review, 54*(6), 4771–4797. https://doi.org/10.1007/s10462-021-10011-5

Widayani, N. H., & Arriyanti, E. (2026). *Analisis sentimen masyarakat terhadap program gratis pol di TikTok menggunakan algoritma Naive Bayes* [Analysis of public sentiment towards the free pol program on TikTok using the Naive Bayes algorithm]. *Bulletin of Information Technology (BIT), 7*(2), 162–171. https://doi.org/10.47065/bit.v7i2.2701