Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

Comparative Analysis of Baseline IndoBERT, Class-Weighted IndoBERT, and SMOTE with Support Vector Machine for Handling Imbalanced Sentiment Classification in Indonesian

Riya Widayanti*[1], Fitriana Cendra Kasih[2]

[1,2]Department of Informatics, Faculty of Computer Science, Esa Unggul University, Jakarta, Indonesia

Email: [1]riya.widayanti@esaunggul.ac.id

Received : Feb 2, 2026; Revised : Mar 17, 2026; Accepted : Mar 25, 2026; Published : Jun 15, 2026

Abstract

Imbalanced data distribution is a common issue in Indonesian sentiment classification and significantly affects the performance of classification models. This study investigates three approaches, namely SMOTE combined with Support Vector Machine (SMOTE + SVM), Baseline IndoBERT, and Class-Weighted IndoBERT. The dataset consists of Google Maps reviews, which are categorized into positive, neutral, and negative sentiments. Prior to model training, the data undergo preprocessing steps including cleaning, normalization, and tokenization. Model performance is evaluated using confusion matrix analysis and macro-averaged F1-score. The results show that Baseline IndoBERT achieves a macro F1-score of 0.598, followed by Class-Weighted IndoBERT with 0.582, while SMOTE + SVM obtains the lowest performance at 0.545. Despite having slightly lower overall performance, Class-Weighted IndoBERT demonstrates a more balanced capability in recognizing minority classes. These findings indicate that incorporating class-weighting mechanisms into transformer-based models can help mitigate bias toward majority classes and improve minority class recognition. From a scientific perspective, this study provides empirical evidence on how imbalance-aware learning strategies influence the behavior of transformer-based models in imbalanced text classification tasks. Furthermore, this study highlights the importance of using macro-averaged evaluation metrics to ensure a more comprehensive and fair assessment of model performance, particularly in low-resource and imbalanced language settings.

Keywords: *Class Weighting, Imbalanced Dataset, IndoBERT, SMOTE, Sentiment Classification, SVM.*

1. INTRODUCTION

Finding the opinion, feeling, and attitude expressed in a piece of writing is the focus of sentiment analysis [1]. With the rapid growth of user-generated content on social media platforms, online review sites and discussion forums, sentiment analysis has become an important tool to understanding user's perception and experiences [2]. In Indonesia, sentiment analysis has been widely applied in various domains, including product reviews, public services, education, tourism, and government-related research, where respondents offer text input in a range of informal and varied language forms [3], [4], [5], [6], [7], [8]. Despite its wide range of applications, sentiment analysis in real-world scenarios still faces several challenges. One of the most significant issues is the problem of class imbalance dataset [9]. In many real-world sentiment analysis datasets, the distribution of sentiment classes is often imbalanced, where one class may dominate the dataset while the remaining classes are underrepresented [10]. As a result, machine learning models tend to become biased toward the majority class, which can reduce the effectiveness of classification models in identifying minority classes [11]. In such situations, high overall accuracy may be misleading because the model may perform well only in predicting the majority class while failing to detect the minority class thatare often more critical for analysis [12].

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

In recent years, transformer-based language models have demonstrated outstanding performance in a variety of natural language processing applications. One of the most widely used architectures is Bidirectional Encoder Representations from Transformers (BERT), which is capable of capturing contextual relationships within text effectively [13]. To better represent linguistic characteristic specific to the Indonesian language, IndoBERT was introduced as a pre-trained language model trained on large-scale Indonesian corpora [14], [15]. Previous studies have shown that IndoBERT achieves superior performance in Indonesian sentiment classification tasks compared to traditional machine learning approaches and earlier deep learning models [16], [17].

However, despite its strong representation capability, fine-tuning IndoBERT on imbalanced sentiment datasets does not necessarily solve the class imbalance problem. During the training process, the standard optimization objective based on cross-entropy loss does not explicitly consider class distribution [18]. Consequently, the model may focus more on the majority classes, which can lead to lower recall and F1 score values for minority sentiment classes, even when the overall accuracy appears satisfactory [19].

Several techniques have been proposed to address the class imbalance problem in sentiment classification tasks. These approaches can generally be categorized into data-level and algorithm-level methods. Data level approaches aim to modify the class distribution within the dataset through techniques such as oversampling or undersampling [20]. One of the most widely used oversampling techniques is the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for minority classes based on similarities within the feature space [21]. SMOTE has been shown to perform well when combined with traditional machine learning classifiers such as Support Vector Machines (SVM), which are known for their effectiveness in handling high-dimensional text representations [22].

In contrast, algorithm-level approaches focus on modifying the learning process without changing the dataset itself. One common technique in this category is class-weighted learning, where higher penalties are assigned to misclassification of minority classes through weighted loss functions [23]. In deep learning models, class weighting can encourage model to learn more balanced decision boundaries and improve its ability to recognize minority classes [24]. This approach is particularly suitable for transformer-based models such as IndoBERT because it allows the model to handle class imbalance while preserving the original data distribution and linguistic characteristic.

Although class-weighted learning and SMOTE-based resampling have been extensively studied separately, there is still a lack of comprehensive comparative studies that evaluate the effectiveness of these approaches within a unified experimental framework for Indonesian sentiment analysis. In particular, only a limited number of studies have compared the performance of baseline transformer-based without imbalanced handling, class-weighted transformer-based models and traditional machine learning models combined with data level resampling techniques such as SMOTE.

Furthermore, many previous studies have relied heavily on overall accuracy as the main evaluation metric, which may not adequately reflect model performance on minority sentiment classes. For imbalanced classification problems, evaluation metrics such as macro-averaged precision, recall and F1-score, along with confusion matrix analysis, proved a more comprehensive assessment of model performance across all sentiment classes [25].

Therefore, this study aims to conduct a comparative analysis of three different approaches for handling imbalanced sentiment classification in Indonesian text: baseline IndoBERT, class-weighted IndoBERT and SMOTE-SVM. The models are evaluated using multiple performance metrics, with particular emphasis on macro-averaged evaluation metrics to assess their effectiveness in identifying minority sentiment classes. By comparing data-level and algorithm-level approaches within a unified experimental framework, this study contributes to a deeper understanding of how transformer-based

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

models and traditional machine learning methods perform when dealing with imbalanced sentiment datasets in the Indonesian language.

2. METHOD

To evaluate the effectiveness of different approaches in handling imbalanced sentiment classification in Indonesian text, this study conducts a comparative experimental study. Three classification models, namely Baseline IndoBERT, Class-Weighted IndoBERT and Support Vector Machine with SMOTE, are implemented and compared to examine their respective performances. To ensure a fair and consistent evaluation, all models are trained and tested using the same dataset and evaluation protocol.

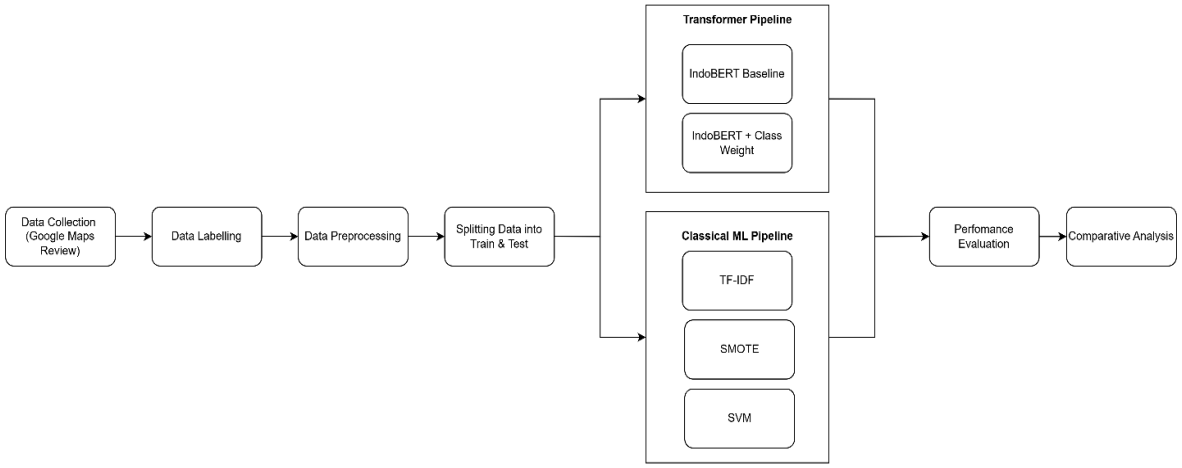

Figure 1. Research Method

2.1. Data Collection

The data used in this study was gathered from Google Maps assessments of an Indonesian university. The Apify platform, which provides automated mining of publicly accessible user-generated content, was used in the web scraping technique used to obtain the data [26].

2.2. Data Labelling

The rating-based approach used in the sentiment categorization process of this study was able to categorize the sentiment classes depending on the star ratings provided by the users in Google Maps reviews [27]. Reviews with ratings between four and five stars were categorized as positive sentiment, reviews with ratings between three and two stars as neutral sentiment, and reviews with ratings between one and two stars as negative sentiment. Based on this labeling approach, mid-range ratings indicate neutral opinions, lower ratings indicate dissatisfaction, and higher ratings indicate positive opinions from the users. To minimize subjectivity in the labeling process, the rating-based labeling approach was employed.

2.3. Data Preprocessing

Data preprocessing is a crucial stage in sentiment analysis to improve data quality and ensure that textual data can be effectively processed by machine learning and deep learning models. In this study, preprocessing was applied to Google Maps review texts to reduce noise, standardize text formats, and enhance semantic representation. The preprocessing steps consisted of text normalization, stopwords removal, text cleaning, and tokenization.

2.3.1. Text Cleaning

Text cleaning was used to eliminate irrelevant features that could have a negative impact on the learning process. This text cleaning process involved the elimination of punctuation signs, special characters, numbers, emojis, URLs, and unnecessary whitespace. In addition to this, the elimination of non-alphabetic characters and repeated whitespaces was done to ensure that the text data contained only relevant linguistic information. This text cleaning process is used to enhance the robustness of IndoBERT and traditional machine learning models employed in this study [28].

2.3.2. Text Normalization

Text normalization was performed to ensure that the review texts were in a standardized form. This process included converting all characters to lowercase to avoid duplication based on case sensitivity. In addition, text normalization entailed the removal of informal words, duplicated characters, and spelling variations that are normally prevalent in user-generated content. The purpose of this step was to mitigate vocabulary sparsity and ensure that words with similar meanings were regarded as identical [28].

2.3.3. Stopwords Removal

The removal of stopwords was conducted to remove common Indonesian words that do not have significant semantic meaning in the sentiment classification process. In this research, the removal of stopwords was conducted using the Sastrawi library, which has a standardized Indonesian stopwords list that is widely used. Words such as conjunctions, prepositions, and common functional words were removed to remove noise from the text data. The removal of stopwords is conducted to improve the emphasis of sentiment words and to enhance the effectiveness of both deep learning and classical machine learning models used in this research [29].

2.3.4. Tokenization

Tokenization is the step where the text is divided into smaller units called tokens. This is where the model is able to understand the relationships between words in context, and is the last step before training the model.

A. BertTokenizer

For the deep learning method, the tokenization step was carried out using the BertTokenizer from the pre-trained IndoBERT model. BertTokenizer is a subword-level tokenization technique used for the pre-processing of text data into a form of tokens for BERT-based models. It relies on the WordPiece algorithm, which breaks down words into smaller units called subwords to deal with out-of-vocabulary words and morphological variations [30].

Given an input sentence $\boldsymbol{S}$, the tokenization process can be defined as a mapping function:

$$\boldsymbol{S} \rightarrow \{\boldsymbol{t_1}, \boldsymbol{t_2}, \dots, \boldsymbol{t_n}\} \quad (1)$$

where $\boldsymbol{t_i}$represents the resulting subword tokens.

Each token $\boldsymbol{t_i}$ is then mapped to a unique integer index using the vocabulary $\boldsymbol{V}$ of the pre-trained IndoBERT model:

$$\boldsymbol{t_i} \rightarrow \boldsymbol{id_i} \in \boldsymbol{V} \quad (2)$$

The input sequence is further transformed into numerical representations consisting of input IDs and an attention mask, which can be expressed as:

$$\text{Input} = \{(\boldsymbol{id_1}, \boldsymbol{id_2}, \dots, \boldsymbol{id_n}), (\boldsymbol{m_1}, \boldsymbol{m_2}, \dots, \boldsymbol{m_n})\} \quad (3)$$

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

where $\boldsymbol{m_i} \in \{\mathbf{0}, \mathbf{1}\}$ denotes the attention mask value indicating whether the token corresponds to actual text (1) or padding (0). This tokenization mechanism allows IndoBERT to capture contextual information at the subword level, enabling effective semantic representation for sentiment classification tasks.

B. Term Frequency – Inverse Document Frequency (TF-IDF)

For the classical machine learning approach, feature extraction was performed using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical technique employed to convert text documents into numerical feature vectors based on the significance of a term in a document compared to the entire corpus. TF-IDF is composed of two primary components, which are Term Frequency (TF) and Inverse Document Frequency (IDF) [31]. The Term Frequency (TF) is the measure of the term's frequency in a document and is given by the formula:

$$\boldsymbol{TF(t,d)} = \frac{\boldsymbol{f(t,d)}}{\sum_{\boldsymbol{k}} \boldsymbol{f(k,d)}} \quad (4)$$

where $\boldsymbol{f(t,d)}$ denotes the frequency of term $\boldsymbol{t}$ in document $\boldsymbol{d}$, and the denominator represents the total number of terms in document $\boldsymbol{d}$.

The Inverse Document Frequency (IDF) reflects the importance of a term across the entire corpus and is calculated as:

$$\boldsymbol{IDF(t)} = \mathbf{log}\left(\frac{\boldsymbol{N}}{\boldsymbol{df(t)}}\right) \quad (5)$$

where $\boldsymbol{N}$ is the total number of documents in the corpus and $\boldsymbol{df(t)}$ denotes the number of documents containing the term $\boldsymbol{t}$.

The TF-IDF value for a term is obtained by multiplying TF and IDF as follows:

$$\boldsymbol{TF\text{-}IDF(t,d)} = \boldsymbol{TF(t,d)} \times \boldsymbol{IDF(t)} \quad (6)$$

TF-IDF gives more importance to words that are more frequent in a specific document but less frequent in the entire corpus. This is useful for extracting discriminative features in text classification problems. In this research, TF-IDF feature vectors were used as input attributes for the Support Vector Machine classifier, and the Synthetic Minority Over-sampling Technique was used in the feature space to handle the problem of class imbalance.

2.4. Splitting Data into Train & Test

The dataset was separated into training and testing sets in order to provide an unbiased evaluation of the suggested models' performance. An 80:20 ratio was used to partition the dataset for this research project. This indicates that 20% of the dataset was used for testing, and the remaining 80% was used for training. This ratio was used to make sure the training dataset was adequate for learning and training the models, including training the SMOTE-SVM classifier and fine-tuning the IndoBERT models. Simultaneously, the testing dataset was used only for testing in order to assess the models' performance.

2.5. Model Training

To determine how to handle sentiment classification on imbalanced Indonesian text data, this study used three distinct classification techniques: baseline IndoBERT, class-weighted IndoBERT, and SMOTE when combined with Support Vector Machine (SVM). To ensure a fair performance comparison, the same training dataset was used to train each model.

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

2.5.1. IndoBERT Baseline

Without applying any specific technique for handling the class imbalance problem, the baseline model employed in this research was an IndoBERT model fine-tuned for sentiment classification. IndoBERT, with 12 transformer layers, 768 hidden dimensions, and 12 attention heads, is a pre-trained language model designed for Indonesian using the BERT-base (uncased) architecture. A large Indonesian corpus consisting of online news articles, Wikipedia articles, and the Indonesian Web Corpus was used to pre-train the IndoBERT model in the baseline model. The standard cross-entropy loss function was used to train the IndoBERT model in the baseline model, assigning equal importance to each sentiment class. The IndoBERT model was trained using the tokenized input sequences with backpropagation for fine-tuning [32]. This baseline configuration serves as a reference point for evaluating the effectiveness of class imbalance handling techniques applied in subsequent experiments.

In this study, the IndoBERT-base-p1 model was employed for sentiment classification with specific training parameters. The model was trained using a learning rate of 2e-5, batch size of 8 for both training and evaluation, and 3 training epochs. A weight decay value of 0.01 was applied to reduce overfitting. Model evaluation and checkpoint saving were performed at each epoch. Training logs were recorded every 100 steps and stored for monitoring the training process. This configuration was selected to achieve stable and effective model performance.

2.5.2. Class-Weighted IndoBERT

The class-weighted training method is used in this study on the IndoBERT model to counter the issue of class imbalance in the sentiment classification problem. In this method, a weight is given to each class during training, unlike in the standard method where all classes are given equal weight. This ensures that the minority classes have more importance in the loss function. The bias of the model towards the majority class is overcome using this method [24]. The weights of the classes are determined based on the number of samples in each class. The formula used for determining the class weights is as follows:

$$\boldsymbol{w_i} = \frac{\boldsymbol{N}}{\boldsymbol{C} \times \boldsymbol{n_i}} \tag{7}$$

where:

- $\boldsymbol{w_i}$ denotes the weight of class $\boldsymbol{i}$,
- $\boldsymbol{N}$ represents the total number of samples,
- $\boldsymbol{C}$ is the total number of classes,
- $\boldsymbol{n_i}$ is the number of samples in class $\boldsymbol{i}$.

These class weights are then incorporated into the cross-entropy loss function, resulting in a weighted cross-entropy loss defined as:

$$\boldsymbol{\mathcal{L}} = -\sum_{\boldsymbol{i=1}}^{\boldsymbol{C}} \boldsymbol{w_i}\ \boldsymbol{y_i} \boldsymbol{\log}\ (\boldsymbol{\hat{y}_i}) \tag{8}$$

where:

- $\boldsymbol{y_i}$ is the ground truth label for class $\boldsymbol{i}$,
- $\boldsymbol{\hat{y}_i}$ is the predicted probability for class $\boldsymbol{i}$,
- $\boldsymbol{w_i}$ is the weight assigned to class $\boldsymbol{i}$.

Cxc Class weights were used in the model's training process to address the problem of class imbalance. The class weights' computed values were [1.90, 7.39, 0.43]. The minority classes were given

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

more weight in the loss function by using these values. The addition of class weights in the proposed IndoBERT model only affects the loss function; the model's architecture remains unchanged.

2.5.3. Synthetic Minority Over-sampling Technique (SMOTE) + Support Vector Machine (SVM)

For the classical machine learning approach, sentiment classification was performed using a Support Vector Machine (SVM) classifier combined with the Synthetic Minority Over-sampling Technique (SMOTE).

2.5.3.1. Synthetic Minority Over-sampling Technique (SMOTE)

This study employs the Synthetic Minority Over-sampling Technique (SMOTE) to address the problem of class imbalance in the traditional machine learning approach. Instead of replicating the existing samples in the feature space, SMOTE interpolates among them to generate new examples for the minority classes [33].

Given a minority class sample $\mathbf{x}_i$, a new synthetic sample $\mathbf{x}_{new}$is generated using one of its k-nearest neighbors $\mathbf{x}_{nn}$as follows:

$$\mathbf{x}_{new} = \mathbf{x}_i + \lambda \times (\mathbf{x}_{nn} - \mathbf{x}_i) \quad (9)$$

where:

- $\mathbf{x}_i$ is an original minority class sample,
- $\mathbf{x}_{nn}$ is one of the k-nearest neighbors of $\mathbf{x}_i$,
- λ is a random value sampled from a uniform distribution in the range $[0, 1]$.

This process is repeated until the desired level of class balance is achieved. In this study, SMOTE is applied only to the training data after feature extraction using TF-IDF, in order to prevent data leakage.

2.5.3.2. Support Vector Machine (SVM)

After balancing the training data using SMOTE, sentiment classification is performed using a Support Vector Machine (SVM) classifier. SVM aims to find the optimal hyperplane that maximally separates data points from different classes in a high-dimensional feature space [34].
For a binary classification problem, the decision function of SVM is defined as:

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b \quad (10)$$

where:

- $\mathbf{w}$ is the weight vector,
- $\mathbf{x}$ is the input feature vector,
- b is the bias term.

The optimal hyperplane is obtained by solving the following optimization problem:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2} \| \mathbf{w} \|^2 + C \sum_{i=1}^{n} \xi_i \quad (11)$$

subject to:

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (12)$$

where:

- $y_i \in \{-1, +1\}$ is the class label,
- ξ_i are slack variables allowing misclassification,
- C is the regularization parameter controlling the trade-off between maximizing the margin and minimizing classification error.

For multi-class sentiment classification, SVM is extended using the one-vs-rest (OvR) strategy, where a separate classifier is trained for each class.

2.5.3.3. Integration of SMOTE and SVM

In this study, textual data are first transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) representation. SMOTE is then applied to the TF-IDF feature space to balance the class distribution in the training set. The balanced dataset is subsequently used to train the SVM classifier. This approach serves as a traditional machine learning baseline for comparison with deep learning-based models, namely baseline IndoBERT and class-weighted IndoBERT.

2.6. Performance Evaluation

The performance of the proposed models was tested using a number of standard classification measures to obtain a complete evaluation, especially when there is class imbalance. The evaluation was conducted on the test data that was not used during training. The measures are calculated based on the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [35].

A. Accuracy

Accuracy measures the proportion of correctly classified instances over the total number of instances and is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

B. Precision

Precision evaluates the correctness of positive predictions and is formulated as:

$$\boldsymbol{Precision = \frac{TP}{TP+FP}} \quad (14)$$

C. Recall

Recall measures the model's ability to correctly identify all relevant samples of a given class and is defined as:

$$\boldsymbol{Recall = \frac{TP}{TP+FN}} \quad (15)$$

D. F1-score

The F1-score represents the harmonic mean of precision and recall, providing a balanced evaluation metric:

$$\boldsymbol{F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall}} \quad (16)$$

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

E. Confusion Matrix

A confusion matrix was also used to analyze the classification results in more detail. It provides a class-wise breakdown of true and false predictions, allowing further inspection of misclassification patterns among sentiment classes [36].

Multiclass Confusion Matrix		Predicted		
		S_1	S_2	S_3
Actual	S_1	TP	FN	FN
	S_2	FP	TN	TN
	S_3	FP	TN	TN

Figure 2. Confusion Matrix

2.7. Comparative Analysis

Comparative analysis was conducted to evaluate the effectiveness of different imbalance handling strategies applied to Indonesian sentiment classification. In this study, three models were compared, namely the baseline IndoBERT model trained without any explicit class imbalance handling strategy, the class-weighted IndoBERT model that integrates class weights into the loss function to reduce bias toward majority classes, and the SMOTE+SVM model, which represents a traditional machine learning approach combining data-level imbalance handling with a Support Vector Machine classifier. The comparison focuses on differences in performance across multiple evaluation metrics, particularly macro-average precision, recall, and F1-score, which are more appropriate for imbalanced datasets than accuracy alone. Special attention is given to the performance of minority classes to assess how effectively each method mitigates the impact of class imbalance. Through this comparative evaluation, the study aims to provide empirical insights into the strengths and limitations of deep learning-based approaches versus traditional machine learning techniques for handling imbalanced sentiment classification tasks in Indonesian text.

3. RESULT

This section presents the results of the experiment and discuss the comparison of the performance of Baseline IndoBERT, Class-Weighted IndoBERT, and SMOTE-SVM in handling the imbalanced sentiment classification problem in Indonesian text. The experimental results will be analyzed using confusion matrix analysis and classification metrics to see the performance of the methods in dealing with the problem, especially in identifying the minority classes.

3.1. Dataset

A total of 1,831 review records were successfully gathered using Apify scrapper tools. The review text, rating score (stars), review date, and geographical details are among the elements that are present in every record. The gathered reviews serve as the unprocessed textual data for sentiment analysis and reflect users' opinions and emotions about the organization.

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

After the data collection process, the collected reviews were prepared for the sentiment analysis task through a data labelling process. Data labelling is an essential step in sentiment analysis because each review must be assigned an appropriate sentiment category before being used for model training. In this study, the labelling process was conducted using a rating-based approach, utilizing the star ratings provided by users in Google Maps reviews. This approach was chosen because star ratings generally reflect the level of user satisfaction toward the reviewed entity.

Table 1. Example of Labelled Dataset Structured

Review	Rating	Label
Kampusnya bersih fasilitasnya lengkap	5	Positive
Biasa saja, tidak terlalu bagus atau buruk	3	Negative
Pelayanan sangat lambat dan membingungkan	0.780	Neutral

3.2. Preprocessing Results

Before the dataset is used for model training, a preprocessing stage is conducted to clean and standardize the textual data. This process aims to reduce noise and improve the overall quality of the dataset so that it can be effectively processed by sentiment classification models. In this study, several preprocessing techniques are applied, including text cleaning, text normalization, stopword removal, and tokenization.

As shown in Table 2, the preprocessing steps transform the raw review text into a cleaner and more structured form by removing unnecessary characters, converting informal words into their standard forms, eliminating stopwords, and splitting the text into tokens that can be further processed by the classification models.

Table 2. Example of Data Preprocessing Steps

Step	Label
Original Text	kampusnya bagus bgt!!! pelayanannya cepet dan sangat membantu 😊
Cleaning	kampusnya bagus bgt pelayanannya cepet dan sangat membantu
Normalization	kampusnya bagus banget pelayanannya cepat dan sangat membantu
Stopwords Removal	kampusnya bagus banget pelayanannya cepat membantu
BERTokenizer	[CLS], kampus, ##nya, bagus, bgt, pelayanan, ##nya, cepet, dan, sangat, membantu, [SEP], [PAD] [2, 4731, 57, 1305, 1123, 4592, 57, 8762, 106, 1198, 3210, 3, 0, ...] [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, ...]
TF-IDF Vectorizer	[0.00, 0.21, 0.34, 0.15, 0.00, 0.27, ...]

3.3. Class Distributions

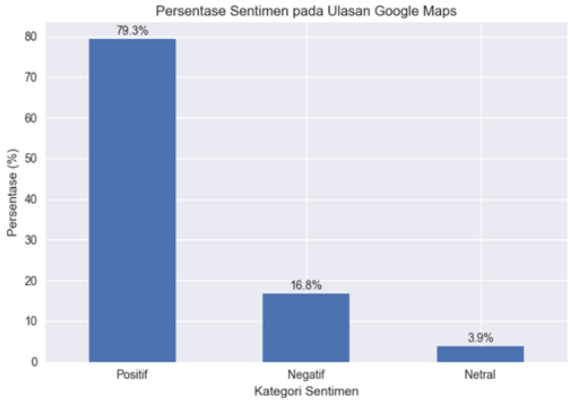

Figure 3. Class Distribution

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

After the preprocessing step, the original dataset of 1,831 Google Maps reviews was reduced to 1,000 clean and valid instances. The dataset was then divided into three classes of sentiment: positive, negative, and neutral. Figure 2 illustrates the distribution of the classes in the dataset. The results indicate that the dataset is highly imbalanced, where the positive sentiment class has the largest number of instances with 79.3% of the total data, followed by the negative sentiment class with 16.8%, and the neutral sentiment class with only 3.9%. The imbalance in the dataset shows that the dataset is biased towards the positive class, which may result in biased classification models that are inefficient in identifying the minority classes of sentiment.

3.4. Model Training

Three approaches—Baseline IndoBERT, Class-Weighted IndoBERT, and SMOTE-SVM—are investigated in this study to address the imbalance issue in sentiment categorization on Indonesian text. Metrics including accuracy, precision, recall, and F1-score are used to evaluate the models' performance.

3.5. Model Evaluation

This section provides the performance comparison of Baseline IndoBERT, Class-Weighted IndoBERT, and SMOTE-SVM in terms of accuracy, precision, recall, F1-score, and macro F1-score. Table 1 displays the overall classification result of the three models.

Table 3. Performance Comparison of Classification Models

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
Baseline IndoBERT	0.905	0.5963	0.6010	0.5980
Class-Weighted IndoBERT	0.825	0.5656	0.6091	0.5819
SMOTE + SVM	0.780	0.5420	0.5898	0.5445

The performance comparison of the three sentiment classification methods, namely Baseline IndoBERT, Class-Weighted IndoBERT, and SMOTE + SVM, is shown in Table 3. The performance metrics used for the comparison are accuracy, macro precision, macro recall, and macro F1-score. The performance comparison shows that the Baseline IndoBERT method has the highest accuracy of 0.905. However, the macro values of precision (0.5963), recall (0.6010), and F1-score (0.5980) are still low. This indicates that, although the Baseline IndoBERT method is efficient for the majority class, it is not efficient for the minority classes. This is expected because the Baseline IndoBERT method is designed for imbalanced datasets, and accuracy is not a good performance metric in such datasets, due to the dominance of the majority class.

The Class-Weighted IndoBERT model indicates a drop in accuracy to 0.825, but with a slightly better macro recall value of 0.6091 than the baseline model. This is an indication that the use of class weights in the model has improved its ability to focus on the minority classes, resulting in a balanced detection of instances among the classes of sentiment. However, the macro precision of 0.5656 and macro F1-score of 0.5819 do not indicate a significant improvement.

On the other hand, the SMOTE + SVM method produces the lowest accuracy of 0.780 with macro precision of 0.5420, macro recall of 0.5898, and macro F1-score of 0.5445. Despite the fact that SMOTE is intended to handle class imbalance by oversampling, the SMOTE + SVM method does not perform better than IndoBERT-based models. This can be explained by the fact that the semantic representation capacity of traditional machine learning features is limited compared to the contextual representation offered by IndoBERT, which is more adept at handling the nuances of Indonesian text.

Although the evaluation metrics provide a broad comparison of the performance of the models, they do not give a complete representation of the behavior of each model with respect to the individual

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

classes of sentiment. Hence, a more detailed evaluation is carried out using the confusion matrices to analyze the performance of classification with respect to the individual classes.

Table 4. Per-Class Performance Based on Confusion Matrix

Model	Class	Precision	Recall	F1-Score
Baseline IndoBERT	Negative	0.8788	0.8286	0.8530
	Neutral	0.0000	0.0000	0.0000
	Positive	0.9102	0.9744	0.9412
Class-Weighted IndoBERT	Negative	0.6818	0.8571	0.7590
	Neutral	0.0714	0.1111	0.0870
	Positive	0.9437	0.8590	0.8993
SMOTE + SVM	Negative	0.5000	0.8571	0.6316
	Neutral	0.2000	0.1111	0.1429
	Positive	0.9259	0.8013	0.8590

For the Baseline IndoBERT model, the class with the highest performance metric, with an F1-score of 0.9412 is Class Positive, which is an indication of its high capability for the majority class. However, the precision and recall for Class Neutral are zero, which implies that the model is not capable of correctly classifying instances belonging to this class. The Class-Weighted IndoBERT model shows a slight improvement in the detection of the minority class. Class Neutral reaches a non-zero F1-score of 0.0870, which indicates that with the use of class weighting, the model is able to detect the minority class, although it is not very effective. Class Positive continues to perform well with an F1-score of 0.8993. The SMOTE+SVM method also enhances the performance of minority class detection over the baseline model, with Class Neutral achieving an F1-score of 0.1429. Nevertheless, the method performs less well than IndoBERT models overall, especially in Class Positive, which achieves an F1-score of 0.8590. This indicates that while oversampling can be used to handle the class imbalance problem, machine learning models may not be able to leverage the contextual semantic information as well as transformer models.

3.5.1. Confusion Matrix

This subsection presents a detailed evaluation of each model using confusion matrix analysis to examine classification performance at the sentiment class level.

3.5.1.1. Evaluation of IndoBERT Baseline

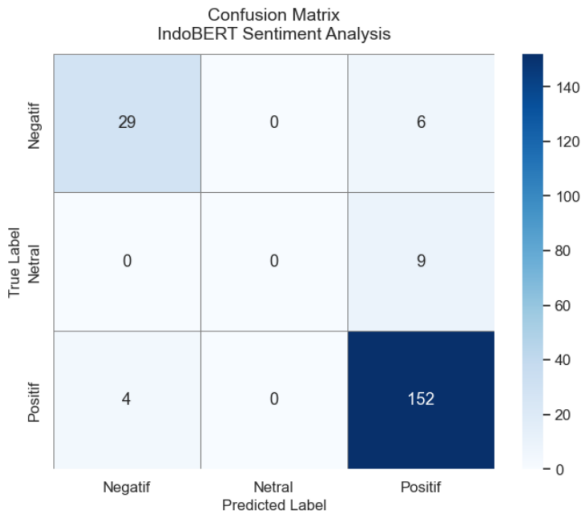

Figure 4. Confusion Matrix of IndoBERT Baseline Model Evaluation

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

3.5.1.2. Evaluation of Class-Weighted IndoBERT

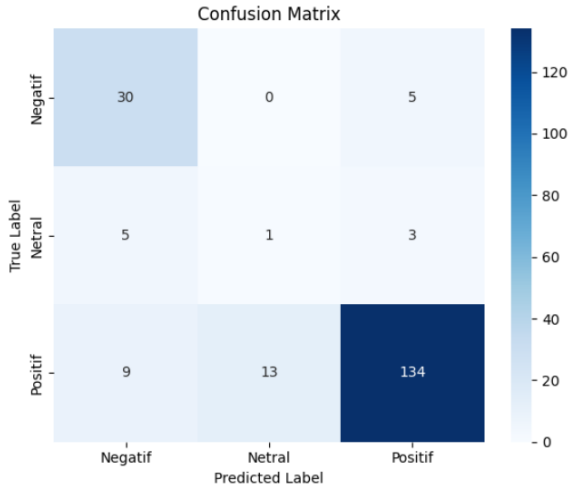

Figure 5. Confusion Matrix of Class Weighted IndoBERT Model Evaluation

The confusion matrix for the Class-Weighted IndoBERT model is shown in Figure 5. The Class-Weighted IndoBERT model performs better in identifying the minority classes than the baseline model. The confusion matrix reveals that the model is able to classify at least one instance of the Neutral class correctly, which means that the effect of class weighting is positive in improving the sensitivity of the model towards the minority classes. The misclassifications are more uniformly distributed over the sentiment classes, which indicates that the effect of the majority Positive class has been diminished. However, the number of correctly classified instances of the Neutral class is still very low, which means that although class weighting has improved the classification accuracy of the minority class, the overall classification accuracy of the Neutral class remains poor.

3.5.1.3. Evaluation of SMOTE + SVM

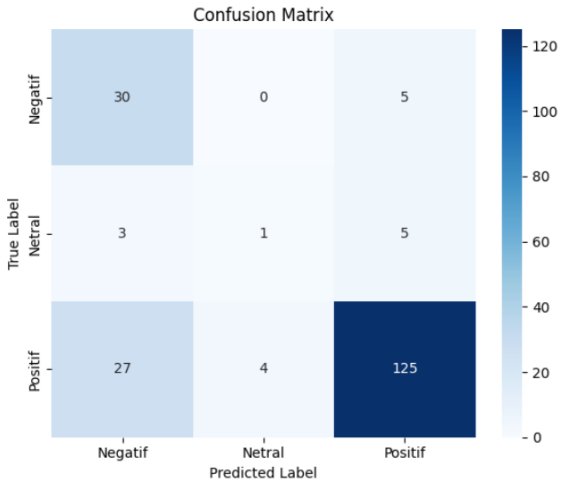

Figure 6. Confusion Matrix of SMOTE + SVM

The confusion matrix of the SMOTE + SVM model is shown in Figure 6. The confusion matrix shows that there is a moderate improvement in the Neutral class over the Baseline IndoBERT model. However, there is a large number of Positive instances that are predicted as Negative, which affects the performance of the model. This observation shows that although SMOTE is able to handle the class imbalance problem in the training data by generating new samples from the minority class, the SVM

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

classifier is not able to capture the semantic meaning of the context as effectively as transformer models. Because of that, the SMOTE + SVM approach is not effective compared to both IndoBERT models, particularly for Positive and Negative classes. These observations together indicate that the oversampling approach along with traditional machine learning classifiers may not be sufficient for sentiment classification tasks on Indonesian text.

4. DISCUSSIONS

This research focuses on the performance of three techniques: Baseline IndoBERT, Class-Weighted IndoBERT, and SMOTE with SVM to handle the problem of imbalanced sentiment classification in Indonesian text. The experimental results indicate that class imbalance significantly affects model learning behavior, particularly in recognizing minority classes such as the Neutral class. This is consistent with other research on imbalanced text classification, which indicates that accuracy as the optimization criterion can lead to biased results towards the majority classes [37], [38].

Out of the three approaches, the Baseline IndoBERT model has the best accuracy. However, this performance is largely driven by its ability to correctly predict the dominant class, namely Positive, while failing to adequately recognize the Neutral class, as evidenced by the confusion matrix analysis. This conclusion aligns with the findings of [39], which reported that transformer-based models trained on imbalanced datasets tend to favor majority class patterns when no imbalance handling strategy is applied.

The Class-Weighted IndoBERT model addresses this limitation by assigning higher penalties for misclassifying minority classes during training. As a result, the model demonstrates improved recall for the Neutral class and a more balanced distribution of predictions across sentiment categories. Similar findings have been reported in previous research on loss-sensitive learning, which show that class weighting enhances minority class recognition in neural network-based classifiers [40], [41]. These findings are also supported by theoretical analyses of loss functions in deep learning, which emphasize that modifying the loss function can partially mitigate the impact of class imbalance [42]. However, the improvement in this study remains limited, as the Neutral class still exhibits relatively low precision and F1-score. This indicates algorithm-level approaches alone are insufficient when the minority class data is severly underrepresented, which aligns with limitations reported in prior studies.

In contrast, the SMOTE combined with SVM approached applies a data-level strategy by generating synthetic samples to balance the dataset. The confusion matrix shows a significant increase in misclassification between Positive and Negative classes, even if this strategy increases minority class representation in the training set. This finding differs from previous studies [43], which reported that SMOTE outperformed conventional classifiers on structured datasets. This discrepancy may be explained by the limitations of SMOTE in handling textual data. Unlike structured data, textual data contains complex semantic and contextual relationships that cannot be effectively preserved through simple interpolation in feature space. As a result, the generated synthetic samples may not accurately reflect the underlying linguistic patterns. Previous studies have also reported that SMOTE may introduce noise and does not consistently improve performance in text classification tasks, particularly when applied to high-dimensional linguistic data [44], [45] .

Furthermore, the results demonstrate that transformer-based models, such as IndoBERT, consistently outperform traditional machine learning approaches. This result is consistent with the experimental investigation conducted in [46], which examined how well the BERT and LSTM models performed on sentiment analysis tasks. When compared to the conventional TF-IDF feature representation, the authors discovered that contextual embeddings significantly enhance the classification model's performance. This suggests that contextual understanding plays a critical role in sentiment classification task involving natural language processing.

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

In addition, this study emphasizes the importance of using macro averaged metrics and confusion matrix analysis in evaluating imbalanced sentiment classification models. It has been found in recent empirical studies in the area of imbalanced learning that accuracy alone can be a misleading measure of performance, as it may hide poor performance on the minority class while overstating the performance of the classifier [47]. The macro F1-score obtained in this study is consistent with prior IndoBERT-based sentiment analysis research, which commonly adopts macro-averaged F1 as a primary evaluation metric. Previous studies have reported that IndoBERT achieves strong F1 performance across various Indonesian datasets and generally outperforms traditional machine learning models, although challenges in recognizing minority classes still remain [48], [49].

From an informatics perspective, this study provides important insights into the comparative effectiveness of datal-level and algorithm-level approaches for handling class imbalance in natural language processing. These findings contribute to the field of informatics by demonstrating that the effectiveness of imbalance handling techniques is highly dependent on the underlying model architecture. Specifically, algorithm-level approaches are more suitable for transformer-based models, while data-level approaches may be less effective for text data due to the loss of semantic coherence. The findings indicate that transformer-based models benefit more from algorithm-level strategies, such as class weighting, as they preserve the original data distribution while improving minority class sensitivity. In contrast, data-level approaches such as SMOTE may introduce noise in textual feature space, leading to reduced semantic consistency and classification performance. This study extends existing knowledge by highlighting the interaction between imbalance handling strategies and contextual language models, providing empirical evidence that supports the theoretical limitations of data-level resampling in high-dimensional linguistic feature spaces.

Nevertheless, this study has several limitations. The performance improvement for the Neutral class remains relatively low, indicating that the applied methods are not sufficient to fully address serve class imbalance. In addition, the dataset size and class distribution may limit the generalizability of the findings. Future research may explore more advance imbalance handling techniques, such as focal loss, hybrid sampling methods and ensemble approaches integrated with transformer-based models. Increasing dataset size and incorporating domain-adaptive pretraining may also enhance model robustness and improve minority class recognition.

Overall, the results confirm that handling imbalanced sentiment data remains a fundamental challenge in Indonesian natural language processing. The results indicate that a more theoretically and empirically validated approach to attaining balanced sentiment classification performance across all sentiment categories is to use transformer-based models in conjunction with imbalance-aware training techniques.

5. CONCLUSION

This study presents a comparative analysis of three different approaches, namely Baseline IndoBERT, Class-Weighted IndoBERT, and SMOTE combined with SVM, for addressing the problem of imbalanced sentiment classification in Indonesian text. The main objective of this research was to evaluate how transformer-based models and traditional machine learning methods perform under imbalanced data conditions and to identify effective strategies for improving minority class recognition.

The experimental results indicate that the Baseline IndoBERT achieved the highest overall accuracy; however, confusion matrix analysis revealed that this performance was heavily influenced by its strong prediction of the majority Positive class, while completely failing to correctly classify Neutral sentiment samples. This finding confirms that high accuracy does not necessarily represent robust classification performance when the dataset is highly imbalanced. Therefore, the use of macro-averaged

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

evaluation metrics and per-class performance analysis is essential for obtaining a more comprehensive and fair assessment of sentiment classification models.

The Class-Weighted IndoBERT model showed greater sensitivity to the minority classes by giving them greater weights during training. This led to a better macro recall value and a more evenly distributed prediction for the sentiment classes than the baseline model. Although the improvement was not very significant, it can be concluded that class weighting is an effective technique for reducing majority class bias in transformer-based sentiment classification models. This result also indicates that modifying the loss function plays an important role in shaping the decision boundaries of transformer models under imbalanced data conditions.

In contrast, the SMOTE + SVM approach showed limited effectiveness in handling imbalanced sentiment data. While synthetic oversampling improved minority class representation, the model experienced a significant increase in misclassification between Positive and Negative sentiments. This outcome highlights the limitations of traditional machine learning classifiers and feature-based representations in capturing the semantic complexity of Indonesian text. The findings further emphasize the superiority of contextual language models, such as IndoBERT, for sentiment analysis tasks involving nuanced linguistic patterns.

From scientific perspective, this study contributes to the field of Informatics by providing empirical evidence on how imbalance-aware learning strategies influence the behaviour of transformer-based models in linguistic classification tasks. The results demonstrate that transformer architecture, despite their strong contextual representation capabilities, remain sensitive to class distribution and require explicit imbalance handling to achieve balanced and unbiased predictions. Furthermore, this study establishes a comparative benchmark between algorithm-level and data-level approaches for handling class imbalance in Indonesia sentiment analysis, offering insights into the trade-offs between these strategies.

Despite the contributions, this study has several limitations. From the results of the Neutral class, which still has a very low precision and F1-score, the improvement in this study is still very small. The size of the dataset and the class distribution may limit the generality of the results. Moreover, only two methods of imbalance handling were considered in this study. Future studies may investigate more sophisticated methods, for instance, focal loss, adaptive resampling, ensemble learning, or hybrid deep learning methods, to further improve the minority class classification. Increasing the size of the dataset and domain-specific pretraining may also help to improve the robustness of the models.

In conclusion, this research confirms that handling imbalanced sentiment data remains a critical challenge in Indonesian text classification. Transformer-based models, particularly IndoBERT combined with class-weighting techniques, provide a promising direction for achieving more equitable and reliable sentiment classification performance across all sentiment categories, while also contributing to the development of more balanced and bias-aware natural language processing systems.

CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the academic supervisors and the Department of Computer Science, Esa Unggul University for their support and guidance during the research process. The authors also thank all contributors who provided assistance in data collection and analysis.

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

REFERENCES

[1] N. A. R. Putri and Ardiansyah, "Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa," *J. Sains dan Inform.*, vol. 9, no. 2, pp. 136–145, 2023, doi: 10.34128/jsi.v9i2.649.

[2] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods , applications , and challenges : A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.

[3] H. Murfi, S. Theresia Gowandi, G. Ardaneswari, and S. Nurrohmah, "BERT-based combination of convolutional and recurrent neural network for indonesian sentiment analysis," *Appl. Soft Comput.*, vol. 151, pp. 1–15, 2024, doi: 10.1016/j.asoc.2023.111112.

[4] R. C. Rivaldi and T. D. Wismarini, "Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP) (Studi Kasus Zalika Store 88 Shopee)," *J. Ilm. Elektron. DAN Komput.*, vol. 17, no. 1, pp. 120–128, 2024, doi: https://doi.org/10.51903/elkom.v17i1.1680 JURNAL.

[5] A. A. Purnama and Y. R. Sipayung, "Sentiment Analysis of Public Service Using Naïve Bayes Classifier," *J. Inf. Syst. Informatics*, vol. 7, no. 3, pp. 2439–2457, 2025, doi: 10.51519/journalisi.v7i3.1207.

[6] M. R. Tanjung, M. Iqbal, and Z. Sitorus, "Analisis Sentimen Google Review terhadap Mutu Kualitas Pendidikan pada Perguruan Tinggi STIE Al-Washliyah Sibolga dengan Metode Lexicon dan Algoritma Naive Bayes," *Jatilima J. Multimed. Dan Teknol. Inf.*, vol. 07, no. 02, pp. 400–412, 2025, doi: https://doi.org/10.54209/jatilima.v7i02.1549.

[7] B. Atmadja, "Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata di Kabupaten Sukabumi Dengan Naïve Bayes," vol. 15, no. 2, pp. 371–382, 2022, doi: https://doi.org/10.51903/elkom.v15i2.872.

[8] K. H. Prastiawan and D. Yuniarto, "Analisis Sentimen Publik terhadap Program Makan Bergizi Gratis dengan Algoritma Naive Bayes," vol. 4, no. 4, pp. 5412–5419, 2025, doi: https://doi.org/10.31004/riggs.v4i4.3652.

[9] G. Shini and S. V, "Performance Evaluation of Sentiment Analysis on Balanced and Imbalanced Dataset Using Ensemble Approach," *INDIAN J. Sci. Technol.*, vol. 15, no. 17, pp. 790–797, 2022, doi: htt ps://doi.org/ 10.17485/IJST/v15i17.2339.

[10] P. A. Perwira and N. I. Widiastuti, "Imbalance Dataset in Aspect-Based Sentiment Analysis on Game Genshin Impact Review," *J. INFOTEL*, vol. 16, no. 1, pp. 71–81, 2024, doi: 10.20895/INFOTEL.V16I1.984.

[11] I. S. Ritonga, Wanayumini, and D. Hartama, "Sentiment Classification in Imbalanced Data : Trade-Offs Between Metrics and Real-World Relevanced," *J. Tek. Inform.*, vol. 18, no. 2, pp. 303–315, 2025, doi: https://doi.org/10.15408/jti.v18i2.46452.

[12] F. Ayu, D. Aryanti, A. Luthfiarta, D. Adiwinata, and I. Soeroso, "Aspect-Based Sentiment Analysis with LDA and IndoBERT Algorithm on Mental Health App : Riliv," *J. Appl. Informatics Comput.*, vol. 9, no. 2, pp. 361–375, 2025, doi: https://doi.org/10.30871/jaic.v9i2.8958.

[13] Sudianto, "PRE-TRAINED BERT ARCHITECTURE ANALYSIS FOR INDONESIAN QUESTION ANSWER MODEL," *J. Appl. Eng. Technol. Sci.*, vol. 6, no. 1, pp. 60–68, 2024, doi: https://doi.org/10.37385/jaets.v6i1.4746.

[14] Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *J. Buana Inform.*, vol. 14, no. 02, pp. 107–116, 2023, doi: 10.24002/jbi.v14i02.7244.

[15] E. Yulianti and N. K. Nissa, "ABSA of Indonesian customer reviews using IndoBERT: single-sentence and sentence-pair classification approaches," *Bull. Electr. Eng. Informatics*, vol. 13, no. 5, pp. 3579–3589, 2024, doi: 10.11591/eei.v13i5.8032.

[16] G. Medantoro and M. Muljono, "Comparative Analysis of IndoBERT and Classic Machine Learning Models for Sentiment Classification of Education Policy on Social Media X," *J. Appl. Informatics Comput.*, vol. 10, no. 1, pp. 548–557, 2026, doi: https://doi.org/10.30871/jaic.v10i1.11723.

[17] Y. Setiawan and L. A. Wulandhari, "Comparative Analysis of IndoBERT and LSTM for Multi-

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

Label Text Classification of Indonesian Motivation Letter," *JOIN (Jurnal Online Inform.*, vol. 10, no. 2, pp. 260–269, 2025, doi: 10.15575/join.v10i2.1499.

[18] M. A. Fathin, Y. Sibaroni, and S. S. Prasetyowati, "Handling Imbalance Dataset on Hoax Indonesian Political News Classification using IndoBERT and Random Sampling," *J. Media Inform. Budidarma*, vol. 8, no. 2021, pp. 352–360, 2024, doi: 10.30865/mib.v8i1.7099.

[19] Y. A. Singgalen, "Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data," *J. Inf. Syst. Res.*, vol. 6, no. 2, pp. 978–988, 2025, doi: 10.47065/josh.v6i2.6505.

[20] F. Pralienka, B. Muhamad, E. Mulyani, M. S. Bunga, and A. Farhan, "Class Balancing Methods Comparison for Software Requirements Classification on Support Vector Machines," *Sink. J. dan Penelit. Tek. Inform.*, vol. 7, no. 2, pp. 1196–1208, 2023, doi: https://doi.org/10.33395/sinkron.v8i2.12415 e-ISSN.

[21] A. Fitri, D. Anggraeni, and I. M. Tirta, "Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ," *J. Nas. Teknol. dan Sist. Inf.*, vol. 02, no. 2022, pp. 163–172, 2023, doi: https://doi.org/10.25077/TEKNOSI.v9i2.2023.163-172.

[22] I. N. Switrayana, D. Ashadi, H. Hairani, and A. Aminuddin, "Sentiment Analysis and Topic Modeling of Kitabisa Applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods," *Int. J. Eng. Comput. Sci. Appl.*, vol. 2, no. 2, pp. 87–98, 2023, doi: 10.30812/IJECSA.v2i2.3406.

[23] A. A. Qolbu, N. Fitriyati, and N. Inayah, "Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang," *J. FOURIER*, vol. 814, no. 1, pp. 29–44, 2025, doi: 10.14421/fourier.2025.141.29-44.

[24] A. B. Siva and L. Hoki, "Comparison of IndoBERT and SVM Performance in Sentiment Analysis of Digital Education Platforms," *Sink. J. dan Penelit. Tek. Inform.*, vol. 10, no. 1, pp. 64–74, 2026, doi: 10.33395/sinkron.v10i1.15472.

[25] M. Cristina, H. Lee, J. Braet, and J. Springael, "Performance Metrics for Multilabel Emotion Classification : Comparing Micro , Macro , and Weighted F1-Scores," *Appl. Sci.*, vol. 14, no. 21, 2024, doi: 10.3390/app14219863.

[26] R. Erama, "Pemanfaatan Platform Cloud Google Colab Untuk Scraping Komentar Tiktok Pada Konten Gorontalo sebagai Dasar Analisis Respons Warganet," *J. Appl. Eng. Sci.*, vol. 1, no. 2, pp. 124–134, 2025, doi: 10.65177/jaes.v1i2.38.

[27] D. S. Utami, A. Erfina, and M. Id, "Analisis Sentimen Ulasan Terkait UNESCO Global Geopark Di Google Maps dengan Algoritma Naive Bayes," *J. Sains Komput. Inform.*, vol. 6, no. 2, pp. 1154–1170, 2022, doi: 10.30645/j-sakti.v6i2.524.

[28] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.

[29] A. Bijaksana and P. Negara, "The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification," *LONTAR Komput. J. Ilm. TEKONOLOGI Inf.*, vol. 14, no. 3, pp. 172–185, 2023, doi: 10.24843/LKJITI.2023.v14.i03.p05.

[30] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Appl. Sci.*, vol. 15, no. 1, pp. 1–28, 2025, doi: 10.3390/app15010172.

[31] O. A. Irmawan, I. Budi, A. B. Santoso, and P. K. Putra, "Improving Sentiment Analysis and Topic Extraction in Indonesian Travel App Reviews Through BERT Fine-Tuning," *J. Nas. Pendidik. Tek. Inform.*, vol. 13, no. 2, pp. 359–370, 2024, doi: 10.23887/janapati.v13i2.77028.

[32] P. Sayarizki and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," *Indones. J. Comput.*, vol. 9, no. August, pp. 61–72, 2024, doi: 10.34818/indojc.2024.9.2.934.

[33] H. Ma'rifah, A. P. Wibawa, and M. I. Akbar, "Klasifikasi Artikel Ilmiah Dengan Berbagai Skenario Preprocessing," *Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, p. 70, 2020, doi: 10.30872/jsakti.v2i2.2681.

[34] I. Daqiqil, H. Saputra, Syamsudhuha, R. Kurniawan, and Y. Andriyani, "Sentiment analysis of

Vol. 7, No. 3, June 2026, Page. 2857-2875
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5692

student evaluation feedback using transformer-based language models," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 2, pp. 1127–1139, 2024, doi: 10.11591/ijeecs.v36.i2.pp1127-1139.

[35] A. Jazuli, Widowati, and R. Kusumaningrum, "Aspect-based sentiment analysis on student reviews using the Indo-Bert base model," *E3S Web Conf.*, vol. 448, pp. 1–10, 2023, doi: 10.1051/e3sconf/202344802004.

[36] N. Sholihah, F. F. Abdulloh, and M. Rahardi, "Sentiment Analysis on KPU Performance Post-2024 Election via YouTube Comments Using BERT," *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. 4, pp. 2222–2232, 2024, doi: 10.33395/sinkron.v8i4.14040.

[37] L. D. Cahya, A. Luthfiarta, J. Imanuel, S. Winarno, and A. Nugraha, "Improving Multi-label Classification Performance on Imbalanced Datasets Through SMOTE Technique and Data Augmentation Using IndoBERT Model," *J. Nas. Teknol. dan Sist. Inf.*, vol. 09, no. 03, pp. 290–298, 2023, doi: 10.25077/TEKNOSI.v9i3.2023.290-298.

[38] A. Kumar, A. Murugappan, T. Esther, A. Murugappan, and T. Esther, "Imbalanced aspect categorization using bidirectional encoder Imbalanced aspect categorization bidirectional encoder representation from using transformers representation from transformers," *Procedia Comput. Sci.*, vol. 218, pp. 757–765, 2023, doi: 10.1016/j.procs.2023.01.056.

[39] K. G. R. Narayan *et al.*, "Attenuating majority attack class bias using hybrid deep learning based IDS framework," *J. Netw. Comput. Appl.*, vol. 230, p. 103954, 2024, doi: https://doi.org/10.1016/j.jnca.2024.103954.

[40] M. N. Razali, N. Arbaiy, and P. Lin, "Optimizing Multiclass Classification Using Convolutional Neural Networks with Class Weights and Early Stopping for Imbalanced Datasets," *MDPI*, vol. 14, no. 4, pp. 1–14, 2025, doi: 10.3390/electronics14040705.

[41] I. Araf, A. Idri, and I. Chairi, "Cost-sensitive learning for imbalanced medical data: a review," *Artif. Intell. Rev.*, vol. 57, no. 4, p. 80, 2024, doi: 10.1007/s10462-023-10652-8.

[42] A. S. Dina, A. B. Siddique, and D. Manivannan, "A deep learning approach for intrusion detection in Internet of Things using focal loss function," *Internet of Things*, vol. 22, p. 100699, 2023, doi: https://doi.org/10.1016/j.iot.2023.100699.

[43] D. A. Lestari, Y. Sibaroni, and S. S. Prasetiyowati, "Sentiment Analysis of Transportation Application Reviews with SVM on Handling Imbalanced Data Using SMOTE," in *2025 International Conference on Data Science and Its Applications (ICoDSA)*, 2025, pp. 287–292. doi: 10.1109/ICoDSA67155.2025.11157024.

[44] S. F. Taskiran, B. Turkoglu, E. Kaya, and T. Asuroglu, "A comprehensive evaluation of oversampling techniques for enhancing text classification performance," *Sci. Rep.*, vol. 15, no. 21631, pp. 1–20, 2025, doi: 10.1038/s41598-025-05791-7 1.

[45] F. Sağlam and M. A. Cengiz, "A novel SMOTE-based resampling technique trough noise detection and the boosting procedure," *Expert Syst. Appl.*, vol. 200, p. 117023, 2022, doi: https://doi.org/10.1016/j.eswa.2022.117023.

[46] G. Wang and M. M. Jaber, "A Deep Learning Approach to Sentiment Analysis of Hotel Reviews : Comparing BERT and LSTM Models," *Int. J. Adv. Artif. Intell. Mach. Learn.*, vol. 2, no. 2, pp. 67–75, 2025, doi: 10.58723/ijaaiml.v2i2.403.

[47] S. Maulana, N. S. Fatonah, G. Firmansyah, A. M. Widodo, and U. E. Unggul, "PREDICTING TECHNICAL INTERN TRAINING PROGRAM TRAINEE SUCCESS : A COMPARATIVE MACHINE LEARNING," *J. INOVTEK POLBENG*, vol. 10, no. 3, pp. 1753–1761, 2025, doi: https://doi.org/10.35314/1r93bf26.

[48] D. Marutho and V. G. Utomo, "Benchmarking IndoBERT and Transformer Models for Sentiment Classification on Indonesian E-Government Service Reviews," *J. Transform.*, vol. 23, no. 1, pp. 85–95, 2025, doi: 10.26623/transformatika.v23i1.12095.

[49] A. Ramadhan and U. Zaky, "Cross-Lingual Sentiment Analysis for Indonesian Monetary Policy," *J. Sci. Res. Educ. Technol.*, vol. 4, no. 4, pp. 2588–2601, 2025, doi: 10.58526/jsret.v4i4.943.