Vol. 7, No. 4, December 2025

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Sentiment Classification of TikTok Reviews on Almaz Fried Chicken Using IndoBERT and Random Oversampling

Imam Syahputra Zaki[1], Rizka Dhini Kurnia[2], Allsela Meiriza[3]

[1,2,3]Information System, Faculty of Computer Science, Sriwijaya University, Palembang, Indonesia

Email: imamszaki@gmail.com[1], rizkadhini@gmail.com[2], allsela@unsri.ac.id[3]

Received: Sept 10, 2025

Revised: October 12, 2025

Accepted: Nov 19, 2025

Published: Dec 9, 2025

Corresponding Author:

Author Name*:

Imam Syahputra Zaki

Email*:

imamszaki@gmail.com

DOI:

10.63158/journalisi.v7i4.1310

Abstract. The socio-political context surrounding the Indonesian Ulema Council's Fatwa No. 83 of 2023, which catalyzed a significant consumer shift, necessitates an accurate measure of public sentiment toward alternative local brands like Almaz Fried Chicken. Analyzing real-time consumer discourse on the challenging TikTok platform, the study utilized a final dataset of 4,374 unique comments to overcome the inherent problem of dataset imbalance and linguistic informality. The core method involved a seven-stage quantitative approach: data collection, preprocessing, sentiment labeling, data splitting (70:15:15), Random Oversampling (ROS), IndoBERT fine-tuning, and evaluation. This pipeline fine-tuned IndoBERT, a Transformer-based model, integrated with ROS applied exclusively to the training data. Evaluation demonstrated that ROS significantly reduced model bias and enhanced performance: Overall Accuracy increased by 2.0% (from 91% to 93%), and the Macro F1-Score improved by 3.4% (from 0.87 to 0.90). Most critically, the F1-Score for the minority Negative sentiment class surged from 0.78 to 0.84, confirming ROS's effectiveness in accurately detecting critical feedback. These findings provide timely, data-driven insights into brand perception amidst the boycott campaign and establish a robust, reliable IndoBERT-ROS methodology for advanced sentiment monitoring in dynamic social media environments.

Keywords: Sentiment Analysis, IndoBERT, Random Oversampling, Tiktok, Public Opinion, Almaz Fried Chicken

1. INTRODUCTION

The escalation of the Palestinian–Israeli conflict in October 2023 and the issuance of the Indonesian Ulema Council's Fatwa No. 83 of 2023, which calls for a boycott of products affiliated with Israel, have led to a notable shift in consumer behavior domestically [1]. Empirical studies indicate that a substantial majority of Indonesians express support for this religious and sociopolitical directive, reflecting significant alignment with the boycott initiative as documented in recent qualitative and normative analyses of the fatwa's impact on market dynamics [2], [3], [4]. This has contributed to increased attention toward local brands such as Almaz Fried Chicken, which have been recognized as alternative choices in the wake of this movement. These developments underscore the importance of understanding how international political conflicts influence national consumer sentiment and brand perception, particularly through religious and ethical frameworks explored in recent academic research.

In this digital era, TikTok has become a dominant platform for public discourse in Indonesia, serving as a vital sensor for public sentiment [5]. With over 125 million active users, the platform's characteristics short video format, highly expressive communication, and informal language make its comment sections a dynamic, yet unstructured, arena for real-time consumer opinion and social campaigns [6]. Analyzing this discourse is crucial for accurately capturing the public's response to the boycott movement.

The current landscape of sentiment analysis research confirms the superiority of advanced Natural Language Processing (NLP) models, particularly IndoBERT, for handling Indonesian social media data concerning the pro-Israel boycott movement. Research focusing on sentiment towards the boycott of Pro-Israel products using IndoBERT and data balancing techniques achieved an optimal performance with an impressive 97% accuracy and 97% F1-Score when implementing oversampling, proving this combination highly effective in mitigating the typical issue of class imbalance found in social media data [7]. This success contrasts with other studies employing similar methodologies on the same issue; for instance, another IndoBERT implementation using Random Oversampling for class balancing achieved an accuracy, precision, recall, and F1-Score of 85% on Twitter data [8], while a different study using the older Naïve Bayes Classifier method reported an 83% accuracy on X comments [9]. Further comparative analyses, such as one evaluating SVM versus BERT for sentiment classification related to the boycott campaign on Platform X, revealed that BERT still outperformed SVM in overall performance, achieving a 69.26% accuracy and 69.47% F1-Score against SVM's 64.58% accuracy and 62.40% F1-Score, although the study noted SVM's efficiency and better stability in handling class imbalance in metrics like Precision-Recall and ROC curves [10]. However, the peak performance achieved by combining IndoBERT's robust language understanding with strategic oversampling affirms the necessity of modern Deep Learning and data handling techniques for obtaining the most reliable and highest classification metrics.

Previous research applying IndoBERT has mainly focused on structured texts or commercial reviews from Twitter/X. This research, therefore, addresses the gap concerning the analysis of highly informal, emotionally charged opinions related to collective action on the TikTok platform, specifically tackling its severe data imbalance. While a prior study on general TikTok comments using the Naïve Bayes Classifier method managed an accuracy of 71.55% in a balanced scenario [11], the challenge of non-standard Indonesian language remains significant. To ensure robust model performance on minority classes and capture the semantic nuances of the platform, this study proposes a novel pipeline that integrates IndoBERT with Random Oversampling (ROS). The ROS approach is chosen for its proven effectiveness in balancing text-based datasets by duplicating minority samples without semantic distortion [12], [13] , as demonstrated by the fact that using oversampling in related sentiment classification research resulted in a significant performance jump, achieving an accuracy and F1-score of 97% [7]. This proposed combination of IndoBERT and ROS is specifically designed to overcome the linguistic complexity of TikTok comments and the inherent data imbalance.

This research seeks to answer the fundamental question of how effectively the combined IndoBERT and Random Oversampling pipeline can classify public sentiment in the highly informal and imbalanced TikTok text regarding the local brand Almaz Fried Chicken, providing a data-driven measure of the boycott's impact. The objectives are twofold: (1) to develop a robust IndoBERT-based sentiment classification model capable of accurately handling imbalanced Indonesian social media text data by integrating Random Oversampling; and (2) to conduct a rigorous comparative analysis of model performance before and after applying ROS to empirically assess its impact on the classification

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

accuracy of minority sentiment classes. The findings will contribute to the advancement of NLP systems that more accurately represent public sentiment dynamics in complex, real-time digital environments.

2. METHODS

This study adopts a quantitative experimental approach to develop a sentiment classification model using IndoBERT for analyzing TikTok users' responses regarding the boycott of Israeli-affiliated products. The research process consists of seven main stages: data collection, preprocessing, sentiment labeling, data splitting, data balancing, model training, and evaluation, as illustrated in Figure 1.

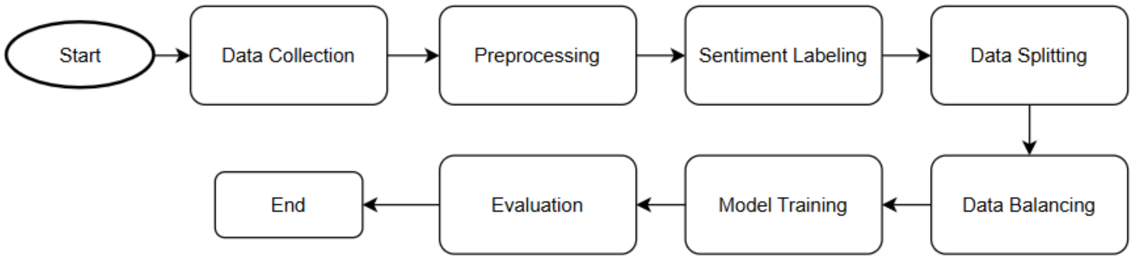

Figure 1. Research Flow

Figure 1 presents the overall flow of the study, starting from collecting TikTok comments to evaluating the model's performance. The process begins with data collection and continues through preprocessing and sentiment labeling to prepare the dataset. The data are then split into training, validation, and test sets, followed by a balancing stage using Random Oversampling to handle class imbalance. Subsequently, the IndoBERT model is fine-tuned on the balanced dataset, and its performance is evaluated using quantitative metrics to assess its ability to classify sentiments accurately.

2.1. Data Collection

Data was gathered by collecting TikTok comments content using the TikTok API from various videos relevant to the keywords "Almaz fried chicken" and the topic of boycotting Israeli products. The tool used to extract data from TikTok comments was Apify, which has the capability to download over 2,000 comments in a single execution [14].

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The initial dataset consisted of comment text along with metadata such as usernames and timestamps. Only text-based comments were analyzed, while media comments (single emojis, stickers, or video replies) were excluded from the analysis. However, reliance on specific keywords and the non-random nature of TikTok's comment ranking system inherently introduces selection bias [15], [16]. Comments appearing earliest or receiving the most engagement (likes, replies) are often more accessible and thus overrepresented, potentially skewing the collected sentiment toward highly visible or polarizing initial reactions [16]. Furthermore, because TikTok's algorithm often curates video feeds based on individual user preferences [15], the video context from which comments are scraped may also be biased toward accounts already focused on the pro-boycott movement, leading to an overrepresentation of related opinions, rather than capturing the full spectrum of public discourse regarding Ayam Almaz.

2.2. Preprocessing

Token segmentation, standardizing informal terms, eliminating textual noise, and converting letter formats are all part of the data preprocessing step [17]. Data is cleaned and prepared for use in the model training phase using this procedure [18]. The following steps are taken in this process, which adheres to the Indonesian-language NLP flow:

1) Case Folding: all text characters are changed to lowercase.
2) Cleaning: getting rid of anything that don't add to the analysis, like hashtags (#), user mentions (@username), site links, and emoticons.
3) Normalization is the process of substituting non-standard terms with their usual forms (gpp → it's alright, for example).
4) Elimination of Common terms: eliminating terms that are frequently used but don't add anything to the study.
5) Tokenization: utilizing the IndoBERT tokenizer to convert text into tokens.

2.3. Data Labeling

Sentiment labeling was conducted using the pre-trained model w11wo/indonesian-roberta-base-sentiment-classifier, which is built upon the RoBERTa architecture and trained on Indonesian-language corpora [18]. This model integrates lexicon-based cues with contextual embedding mechanisms, allowing the model to analyze linguistic sentiment polarity more effectively [19].

Each TikTok comment received one of three sentiment labels positive, neutral, or negative accompanied by a confidence score representing the model's certainty level. To ensure labeling precision, comments with confidence scores below 0.6 were subjected to manual re-evaluation by two independent annotators. During this validation phase, inconsistencies between annotators were reviewed and resolved collaboratively using a standardized Indonesian sentiment lexicon as reference material. The reliability of the manual verification process was measured using Cohen's Kappa coefficient, which achieved a value of 0.87, indicating strong inter-annotator agreement. This dual-stage labeling process combining automated lexicon-based classification and manual cross-validation ensures that the final sentiment labels maintain high consistency and semantic accuracy across the dataset.

2.4. Split Data

The dataset was divided into three subsets 70% for training, 15% for validation, and 15% for testing using stratified sampling to preserve sentiment class distribution across all partitions. This ratio was selected to balance data sufficiency for model training and statistical validity for model evaluation [20].

The rationale for the 70:15:15 split is grounded in best practices for Transformer-based fine-tuning on medium-scale textual datasets [21]. The training subset (70%) provides ample data to capture linguistic diversity while avoiding overfitting. The validation subset (15%) facilitates hyperparameter tuning and enables the use of early stopping during model training, ensuring robust generalization. The test subset (15%) provides an unbiased final evaluation of the model's predictive performance.

2.5. Data Balancing (Random OverSampling)

Class imbalance within the dataset was addressed using the Random Oversampling (ROS) technique. This method was applied exclusively to the training set, ensuring that validation and testing sets retained their natural class distributions. ROS duplicates minority-class samples until all sentiment classes have equal representation, thereby preventing the model from being biased toward the majority class [22].

In this study, the minority classes positive and negative were oversampled to match the majority neutral class, resulting in approximately 2,300 samples per class. This adjustment

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

substantially improved the balance between sentiment categories. The oversampling process increased the total training time by approximately 35%, from 2.1 hours to 2.8 hours, when executed on an NVIDIA Tesla T4 GPU. To mitigate potential overfitting due to data duplication, early stopping (patience = 3) and a dropout rate of 0.1 were implemented during fine-tuning. This approach successfully enhanced the model's ability to learn from all sentiment categories without compromising generalization. The detailed description of the oversampling procedure provides transparency and reproducibility for future researchers employing similar techniques.

2.6. IndoBERT Modeling

This research utilizes IndoBERT-base-p1, a Transformer-based model for Indonesian language. Fine-tuning was performed using Hugging Face Transformers library with the following hyperparameters:

Table 2. Hyperparameter Configuration

	Value	**Justification**
Learning Rate	3e-5	Prevents catastrophic forgetting of pre-trained knowledge
Batch Size	16	Balances memory efficiency and gradient stability
Epochs	6	Sufficient for convergence while avoiding overfitting
Max Length	160 tokens	Accommodates 95% of comment lengths
Dropout	0.1	Regularization to prevent overfitting
Optimizer	AdamW	Effective for transformer fine-tuning

The training process includes tokenization, PyTorch dataset formation, model training with early stopping, and saving the best model based on lowest validation loss.

2.7. Evaluation

To evaluate the model's capacity to categorize data that had not yet been seen, test data was used. This stage aims to measure how well the trained IndoBERT model can recognize

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

sentiment patterns in TikTok comments. A Confusion Matrix was used for performance analysis in order to determine the error rates for each sentiment class and compare the model's prediction outputs with the real labels [23]. Four key measures are used in the evaluation process: F1-Score, Accuracy, Precision, and Recall [24]. The F1 score is calculated as the harmonic mean of precision and recall to give a fair assessment of the model's performance. Recall gauges the model's capacity to identify all positive data (true positives), precision assesses the model's accuracy in classifying positive data, and accuracy calculates the proportion of accurate predictions among all test data [23]. In this study, Macro F1-Score was prioritized because the TikTok comment dataset is imbalanced, with the positive sentiment class being more dominant than the negative and neutral classes [25]. These four-evaluation metrics are calculated using the mathematical equations shown in Equations (1) to (4).

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}\times 100\% \tag{1}$$

$$\text{Precision}=\frac{TP}{TP+FP}\times 100\% \tag{2}$$

$$\text{Recall}=\frac{TP}{TP+FN}\times 100\% \tag{3}$$

$$\text{F1-Score}=2\times\frac{\text{precision}\times\text{recall}}{\text{precision-recall}} \tag{4}$$

3. RESULTS AND DISCUSSION

3.1. Scraping Data

The process of extracting data from TikTok account comments yielded 6,250 data entries, which then underwent deduplication, resulting in 4,374 unique comments for analysis. This data needs more processing because it is still in its raw state. To enable more successful and economical training of classification models, data preprocessing transforms unstructured data into a more structured format. The pre-processing steps performed include data cleaning, case adjustment, word segmentation into tokens, and formalization. Formalization serves to replace slang with standard language, and common abbreviations are expanded into full words, which is expected to help the model understand the context of sentences. The comments obtained are still in raw form, so

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

they need to go thru a data pre-processing stage to improve data quality before entering the model training process.

	videoWebUrl	submittedVideoUrl	input	cid	createTime	createTimeISO	text
0	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	7527321206791504647	1752591049	2025-07-15T14:50:49.000Z	kenapa namanya ayam Saudi??saya di Saudi belum...
1	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	7528250239024284433	1752807683	2025-07-18T03:01:23.000Z	pengen kerja disana tpi takut ngabisin stok 😭
2	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	7524566804037141265	1751949748	2025-07-08T04:42:28.000Z	Enakkk bgtt nagiihhh ❤️ yg utama pro kemanusiaa...
3	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	7524567923508364039	1751950000	2025-07-08T04:46:40.000Z	yup nyesel banget baru cobain 😭
4	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	7525314223368143637	1752123767	2025-07-10T05:02:47.000Z	Jember ditunggu buka cabangnya. 😭
...	...	...	...	...	...	...	...
293	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	7524990885860229895	1752048477	2025-07-09T08:07:57.000Z	udah pulang dri surabaya baru FYP kesel ah
294	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	https://www.tiktok.com/@anisfit__/video/752455...	7527160745366012680	1752553691	2025-07-15T04:28:11.000Z	@MAS ANDRE BONDET @denkciv

Figure 2. Raw Dataset (Indonesia)

3.2. Preprocessing

The preprocessing pipeline successfully cleaned and normalized all 4,374 comments through five sequential steps: case folding, cleaning, normalization, stopword removal, and tokenization. Table 1 presents representative examples of the preprocessing transformation.

Table 1. Preprocessing Transformation Examples (Indonesia)

Step	Example Input	Example Output
Raw Data	"Enak bgtt sih ini!! 😍 Gk nyesel beli dsni #AlmazFC"	"Enak bgtt sih ini!! 😍 Gk nyesel beli dsni #AlmazFC"
Case Folding	"Enak bgtt sih ini!! 😍 Gk nyesel beli dsni #AlmazFC"	"enak bgtt sih ini!! 😍 gk nyesel beli dsni #almazfc"
Cleaning	"enak bgtt sih ini!! 😍 gk nyesel beli dsni #almazfc"	"enak bgtt sih ini gk nyesel beli dsni"
Normalization	"enak bgtt sih ini gk nyesel beli dsni"	"enak banget sih ini tidak nyesel beli sini"
Stopword Removal	"enak banget sih ini tidak nyesel beli sini"	"enak banget nyesel beli sini"

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Step	Example Input	Example Output
Tokenization	"enak banget nyesel beli sini"	["enak", "banget", "nyesel", "beli", "sini"]

The preprocessing successfully standardized informal language, removed noise, and prepared the text for model input. Comments that originally contained slang abbreviations (e.g., "bgtt" → "banget", "gk" → "tidak") were normalized to standard Indonesian, enabling the model to better understand semantic meaning. This transformation is crucial for IndoBERT's performance, as the model was pre-trained on formal Indonesian text and benefits from standardized input. The preprocessing stage was performed using a Python pipeline based on the pandas, regex, and Indonesian slang dictionary libraries for text formalization. The results of the text data transformation before and after passing thru the preprocessing stage are shown in Tabel 2 below.

Tabel 2. Clean Data (Indonesia)

Clean Text	After Preprocessing
finally almaz chicken buka di purwokerto, otw langsung nanti kesana	finally almaz chicken buka di purwokerto otw langsung nanti kesana
jakarta ada gk	jakarta tidak
Jadikan Parkir gratis bagian dari pelayanan pada pelanggan.. Parkir gratis meringankan pelanggan dan ojol..	jadi parkir gratis bagi dari layan pada langgan parkir gratis ringan langgan dan ojol
lidahku ndesa bgt, ga cocok yg saudi, mending mcd kfc richeese kaf rocket chicken, agak mahal buat rasa yg kaya gitu aja	lidah ndesa bgt tidak cocok yang saudi mending mcd kfc richeese kaf rocket chicken agak mahal buat rasa yang kaya gitu saja
biasa aja nasi nya saja kaya nasi kuning di mkn 😎	biasa saja nasi nya saja kaya nasi kuning di mkn

Based on the pre-processing results in Tabel 2, it can be seen that the text, which initially contained slang, double punctuation, and emoticons, was successfully cleaned and normalized. Thus, the sentence becomes easier for machine learning to understand and is ready for the tokenization and IndoBERT model training stages. This process ensures that all text data is in a uniform format, thereby optimizing the efficiency and accuracy of the classification model in the next stage.

3.3. Data Distribution Before and After Balancing

After automatically labeling 4,374 TikTok comments, the data showed an imbalance in the sentiment labels. The neutral class dominates with a significantly larger number of comments compared to the positive and negative classes, potentially leading to model bias during training. To address this issue, the Random OverSampling (ROS) technique is applied to the class of minorities. This method copies samples from the class with less occurrences until the number of instances in each class is balanced. The primary objective of this balancing procedure is to prevent prediction dominance in a particular class, enhance the IndoBERT model's capacity for generalization, and allow it to learn proportionately from all classes. This visualization in Figure 3 shows the data distribution both before and after balancing.

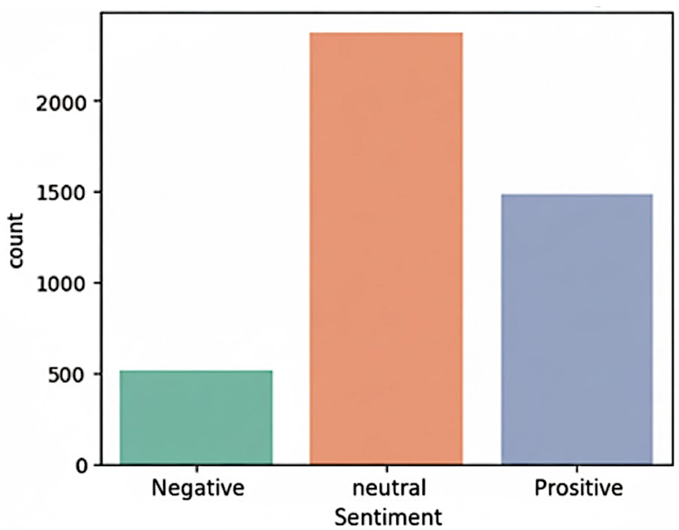

Figure 3. TRAIN Distribution Before Balancing

After Random OverSampling was performed, the visualization in Figure 4 shows that the data volume for all three sentiment categories has reached a balance, with each category having approximately 2,300 samples. This process significantly reduces bias against certain classes and provides a more proportional distribution. Thus, the results of this balancing ensure that the IndoBERT model will receive fairer learning weights for each

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

sentiment category. As will be discussed in the following section, this process also helps to improve accuracy and F1-score at the model evaluation stage

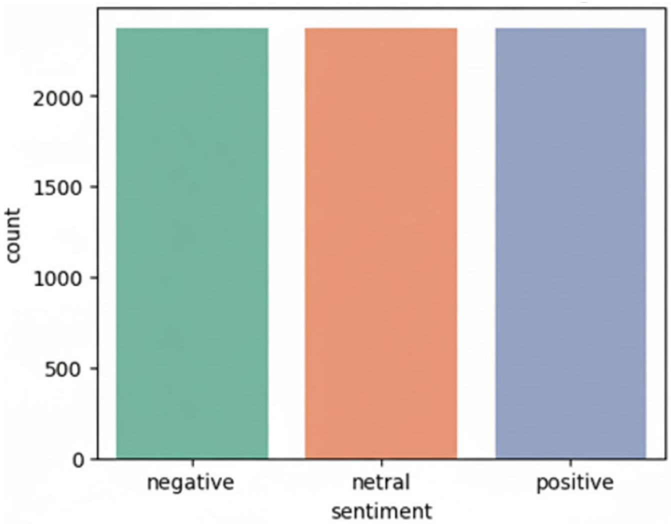

Figure 4. TRAIN Distribution After Balancing

3.4. Performance Evaluation

During the evaluation phase, the IndoBERT model's efficacy in sentiment Measures including accuracy, precision, recall, and F1-score are used to assess classification. Testing is conducted under two different data conditions, namely before and after data balancing is applied (imbalanced and balanced), to observe the impact of the balancing technique on classification results.

Tabel 3. Classification Report (TEST) (IMBALANCED)

Sentiment Class	Precision	Recall	F1-Score	Support
Negative	0.86	0.75	0.78	110
Neutral	0.92	0.95	0.93	509
Positive	0.93	0.89	0.91	319
Accuracy			**0.91**	**938**
Macro Average	**0.90**	**0.86**	**0.87**	**938**
Weighted Average	**0.91**	**0.91**	**0.91**	**938**

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

In Tabel 3, under imbalanced conditions, the model shows a tendency to perform with a bias toward the majority class, which is the neutral class. Based on the classification report, this study was successfully implemented with a success rate of 91%, with an average macro F1 score of 87%. With an F1 value of 0.93, the neutral class had the greatest score, while the negative class received the lowest, with an F1 score of 0.78. This finding indicates that the model is less effective at recognizing negative sentiment, which is most likely influenced by an imbalanced data distribution, leading to classification bias.

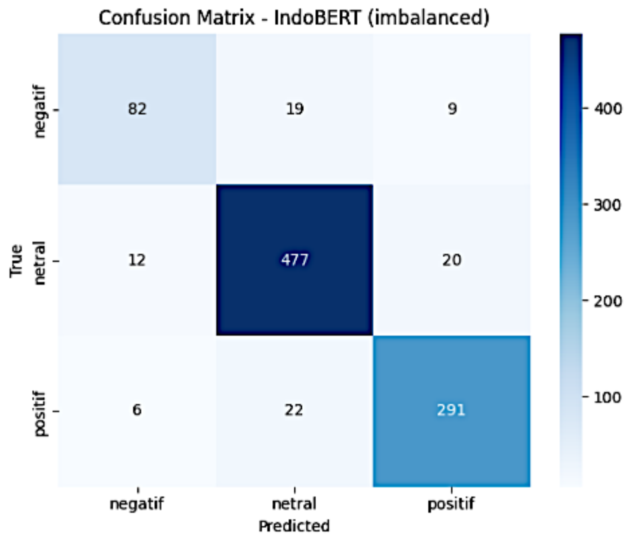

Figure 5. Evaluation Results of the Model before Balancing

From the confusion matrix results, it can be seen that the model more frequently predicts comments as neutral, even though they are actually negative or positive. For example, out of 110 comments labeled negative, only 82 were successfully predicted correctly, while the rest were often misclassified as neutral.

Table 4. Classification Report (TEST) (BALANCED)

Sentiment Class	Precision	Recall	F1-Score	Support
Negative	0.89	0.80	0.84	110
Neutral	0.92	0.96	0.94	509
Positive	0.94	0.92	0.93	319
Accuracy			**0.93**	**938**
Macro Average	**0.92**	**0.89**	**0.90**	**938**
Weighted Average	**0.93**	**0.93**	**0.93**	**938**

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The model performed much better once the data was balanced using the Random OverSampling technique. The accuracy value increased to 93%, with a macro F1-score of 90%. Additionally, improvement was also seen in the negative class, which previously had a low F1-score, but now reached 0.84 with a precision of 0.89 and a recall of 0.80.

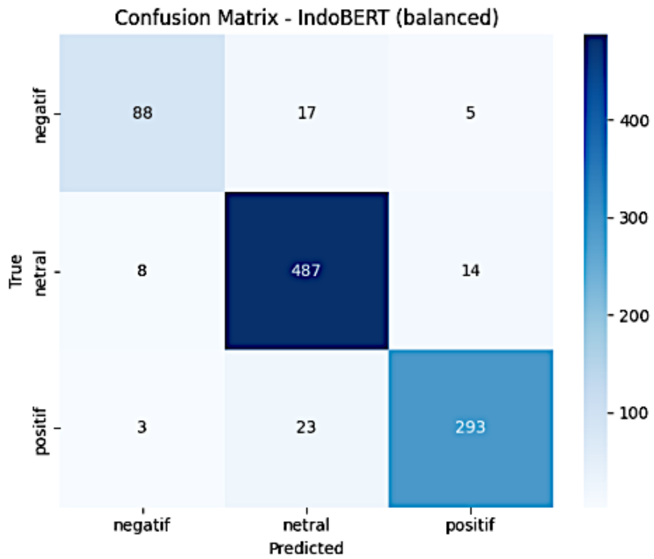

Figure 6. Evaluation Results of the Model after Balancing

Based on the confusion matrix for the data that has undergone the random oversampling or balancing technique, the number of correct predictions in all three classes has increased. The model is now able to recognize negative comments in 88 out of 110 cases, neutral comments in 487 out of 509 cases, and positive comments in 293 out of 319 cases. This increase indicates that the balancing process helps the model understand the sentiment distribution more proportionally, resulting in more accurate and fair classification results for each class.

3.5. Performance Comparison

The research results indicate that the application of IndoBERT in sentiment classification of TikTok comments related to the boycott campaign of Almaz Fried Chicken products yields excellent performance, especially after data balancing using Random Oversampling. The increase in accuracy from 91% to 93% and the rise in the F1 score for the negative class from 0.78 to 0.84 demonstrate that data balancing techniques can reduce model bias toward the majority class.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

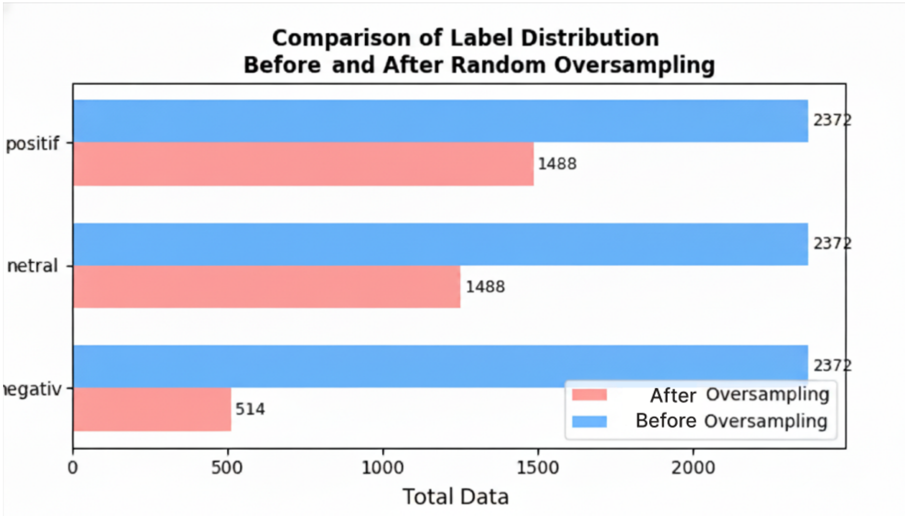

Figure 7. Comparasion of Label Distribution

This aligns with previous research findings indicating that Transformer-based models like IndoBERT are highly sensitive to imbalanced data distributions, suggesting that oversampling approaches can improve model stability and generalization [7].

Table 5. Performance Comparison: Before vs. After Balancing

Metric	Before Balancing	After Balancing	Improvement
Overall Accuracy	91%	93%	+2.0%
Macro F1-Score	0.87	0.90	+0.03 (+3.4%)
Negative F1-Score	0.78	0.84	+0.06 (+7.7%)
Neutral F1-Score	0.93	0.95	+0.02 (+2.2%)
Positive F1-Score	0.91	0.93	+0.02 (+2.2%)
Negative Recall	0.75	0.80	+0.05 (+6.7%)
Negative Precision	0.86	0.89	+0.03 (+3.5%)

Additionally, the performance differences between sentiment classes also reflect the linguistic characteristics of TikTok users in expressing their opinions. Comments with positive and neutral sentiment are relatively easier to recognize because they have a more explicit language structure, whereas negative sentiment is often expressed implicitly thru sarcasm or short conversational contexts, which challenges models in understanding its emotional nuances. This indicates that, although IndoBERT is effective

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

in capturing the semantic context of the Indonesian language, interpreting informal emotional expressions on social media remains a challenge.

3.6. Discussion

The study's findings demonstrate that IndoBERT and Random Oversampling work well together to handle unbalanced Indonesian-language text data, particularly on social media sites like TikTok where users communicate in an informal and dynamic manner. The resulting model is able to provide a more representative picture of public perception regarding social issues, while also demonstrating the significant potential of using Transformer models in public opinion analysis in the digital realm.

This study aimed to explore the effectiveness of IndoBERT in classifying sentiments from TikTok comments related to the boycott campaign of Almaz Fried Chicken. The primary focus was on overcoming the challenges posed by informal language, slang, emoticons, and other unstructured data typical in social media content. The preprocessing steps, including case folding, cleaning, normalization, stopword removal, and tokenization, were essential to transform the raw data into a structured format that the IndoBERT model could efficiently process. By standardizing informal language and replacing slang abbreviations with their formal equivalents, the preprocessing pipeline enhanced the model's ability to accurately interpret the data. This data normalization was crucial since IndoBERT, pre-trained on formal Indonesian text, performs better when presented with text that follows standardized language norms. As a result, the cleaned and tokenized comments were better suited for sentiment classification, improving the overall quality of the data that entered the model training process.

The data imbalance issue, which was present before applying the Random OverSampling (ROS) technique, created a bias in the model's ability to predict sentiments, especially for the minority classes (positive and negative sentiments). Before balancing, the model showed a high level of accuracy (91%) but was skewed toward predicting neutral comments, as this class dominated the dataset. This phenomenon is not uncommon in sentiment analysis, where imbalanced data can lead to a model overfitting the majority class and underperforming on minority classes. The imbalance was most apparent in the negative sentiment class, which had an F1-score of 0.78, much lower than the neutral and positive classes. This result demonstrated that the model was struggling to correctly classify negative comments, which are often more nuanced or implicitly expressed, such as through sarcasm or brief remarks.

After applying the Random OverSampling technique to balance the dataset, the model's performance improved significantly. The accuracy increased to 93%, with the macro F1-score rising to 0.90. These improvements were particularly noticeable in the negative sentiment class, where the F1-score increased from 0.78 to 0.84, a notable enhancement of 7.7%. This change demonstrates the effectiveness of the oversampling method in reducing bias toward the majority class, allowing the model to learn more evenly from all sentiment categories. The results show that balanced datasets enable Transformer-based models like IndoBERT to generalize better and improve their performance across all sentiment classes. This also highlights the importance of class balancing in machine learning pipelines, particularly when dealing with social media data where class distribution can be highly skewed.

Despite the improvements, the study also identified a challenge in classifying negative sentiments, which are often expressed indirectly or with subtleties like sarcasm. The linguistic features of negative comments tend to differ from positive or neutral ones in online discourse, making them harder for machine learning models to accurately identify. In this case, negative comments were sometimes misclassified as neutral, as evidenced by the confusion matrix results. This issue is consistent with existing literature, which points to the difficulties Transformer-based models face in fully capturing the emotional context and nuances of informal language, especially when it comes to implicit expressions of sentiment. However, the improvement in the negative class performance after balancing suggests that addressing data imbalance can still lead to substantial improvements in model accuracy and classification fairness.

This study underscores the critical role of data preprocessing and balancing in improving the performance of sentiment analysis models, particularly in the context of informal social media text. The use of IndoBERT, combined with Random OverSampling, demonstrated how deep learning models can be adapted to handle the challenges of informal language, slang, and imbalanced data distributions commonly found on platforms like TikTok. The results not only highlight the potential of Transformer models for public opinion analysis in social media but also suggest that further research is needed to refine these models' ability to interpret the complexities of negative sentiment. As social media platforms continue to play a significant role in shaping public discourse, these findings provide valuable insights into how sentiment analysis can be effectively applied to gauge public reactions in real-time, offering a foundation for future work in this field.

4. CONCLUSION

This study successfully developed a sentiment classification model using the IndoBERT Deep Learning approach integrated with the Random Oversampling (ROS) technique to analyze public opinion on Ayam Almaz Fried Chicken on the TikTok platform, specifically within the context of the pro-Israel product boycott phenomenon. The resulting model, trained on balanced data, demonstrated a significant performance boost, achieving an overall accuracy of 93% , which is an increase of 2% from the 91% accuracy recorded on imbalanced data. This improvement was most critical for the minority class, evidenced by the F1-Score for negative sentiment increasing from 0.78 to 0.84 , marking an improvement of 0.06 or 7.7%. This confirmed that ROS effectively reduced model bias toward the majority class and enhanced the model's ability to detect criticism or negative sentiment.

While the model showed superior performance, the study is limited by its exclusive focus on TikTok comments, meaning the model's generalization to other social media platforms may require further tuning. Furthermore, the reliance on specific keywords and TikTok's comment ranking system likely introduced selection bias. Moving forward, this high-accuracy model is recommended for integration into a real-time sentiment monitoring system to provide local brands and policymakers with immediate insights. Further research is also suggested to explore more advanced balancing techniques to further strengthen the model's capacity to interpret complex, informal emotional expressions on social media.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

REFERENCES

[1] I. Akbar, M. Rahmani Abduh, and F. Ilmi, "Aksi Boikot Produk Pro-Israel: Tepat Atau Salah Sasaran? (Refleksi Fatwa Mui Nomor: 83 Tahun 2023 Tentang Hukum Dukungan Terhadap Perjuangan Palestina)," *Indonesian J. Islamic Jurisprudence, Econ. and Legal Theory*, vol. 3, pp. 1409–1435, May 2025, doi: 10.62976/ijijel.v3i2.1129.

[2] S. Fatimah, N. Asnawi, and U. Khasanah, "Business Ethics Values In Mui Fatwa No. 83 Of 2023," in *Proc. Int. Conf. Islamic Econ. ad Bus. (ICONIES)*, vol. 10, pp. 47–58, 2024. Accessed: Nov. 18, 2025.

[3] J. A. Sosiologi, A. Mustofa, and F. Alfikri, "MUI Fatwa Authority: Social Movement To Boycott Israeli Products Through Instagram Social Media," *Jurnal Analisa Sosiologi*, vol. 14, no. 2, pp. 268–291, 2025, doi: 10.20961/jas.v14i2.99200.

[4] F. Agil, A. Munawar, M. Azmi, and M. Rohmanan, "Diskursus Fatwa MUI No. 83 Tahun 2023 Tentang Dukungan Terhadap Palestina dan Seruan Boikot Produk Pro-Israel," *Jurnal Bimas Islam*, vol. 17, no. 2, 2025, doi: 10.37302/jbi.v17i2.1402.

[5] R. R. Datau, D. Ichsanuddin Nur, F. Juliputra, A. Eka, P. Haryanto, and I. N. Fauzi, "Analisis Sentimen Tiktok Terhadap Coffee Shop X Dan Implikasinya Terhadap Strategi Pemasaran Digital," *JAMBURA*, vol. 8, no. 1, May 2025.

[6] M. Labib Rifa'i A) and A. Abdurrahman, "Pengaruh Karakteristik Influencer Marketing terhadap Niat Beli Online Konsumen pada Platform Tiktok," *Selekta Manajemen: Jurnal Mahasiswa Bisnis & Manajemen*, vol. 03, no. 01, pp. 121–131, 2024.

[7] A. R. P. Dewi, S. Riyadi, C. Darmajati, N. A. M. Isa, and A. D. Andriyani, "Sentiment Analysis of Pro-Israel Product Boycott Action Using IndoBERT Method on Unbalanced Data," *JUITA: Jurnal Informatika*, vol. 13, no. 2, pp. 187–197, Aug. 2025, doi: 10.30595/juita.v13i2.25976.

[8] Z. A. Sriyanti, D. S. Y. Kartika, and A. R. E. Najaf, "Implementasi Model Bert Pada Analisis Sentimen Pengguna Twitter Terhadap Aksi Boikot Produk Israel," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4743.

[9] M. S. A. Shaumi and Paradise, "Implementasi Metode Naïve Bayes Classifier Terhadap Analisis Sentimen Boikot Produk Terafiliasi Israel Pada Media Sosial 'X,'" *e-Proceeding of Engineering*, vol. 12, no. 4, pp. 6786–6792, Aug. 2025.

[10] S. S. Sabrina, D. F. Shiddieq, and F. F. Roji, "Comparative Analysis of SVM and BERT for Sentiment and Sarcasm Detection in the Boycott of Israeli Products on Platform X," *Sinkron*, vol. 9, no. 2, pp. 872–883, May 2025, doi: 10.33395/sinkron.v9i2.14723.

[11] D. F. M. Dina, T. Haryanti, and M. A. Haq, "Analisis Sentimen Terhadap Komentar Pada Media Sosial Tiktok Yang Berpotensi Menyebabkan Depresi Menggunakan Metode Naive Bayes," *Computing Insight: Journal of Computer Science*, vol. 7, no. 1, pp. 1–9, Jun. 2025, doi: 10.30651/comp_insight.v7i1.26327.

[12] S. Diantika, "Penerapan Teknik Random Oversampling Untuk Mengatasi Imbalance Class Dalam Klasifikasi Website Phishing Menggunakan Algoritma Lightgbm," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 19–25, Jan. 2023, doi: 10.36040/jati.v7i1.6006.

[13] F. Y. A'la, "Optimasi Klasifikasi Sentimen Ulasan Game Berbahasa Indonesia: IndoBERT dan SMOTE untuk Menangani Ketidakseimbangan Kelas," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 1, pp. 256–265, Apr. 2025, doi: 10.29408/edumatic.v9i1.29666.

[14] I. Muhandhis and A. S. Ritonga, "Public Sentiment Analysis on TikTok about Tapera Policy using Random Forest Classifier," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, pp. 354–365, 2025.

[15] M. A. H. A. Asmawi, P. Isawasan, L. Shanmugam, K. A. Salleh, and K. S. Savita, "Exploring Sentiment Trends in TikTok Comments Using GPT for," *e-Academia Journal of UiTM Cawangan Terengganu*, vol. 14, no. 1, pp. 57–72, Jun. 2025, doi: 10.24191/eaj.v14i1.5623.

[16] Z. Cheng and Y. Li, "Like, Comment, and Share on TikTok: Exploring the Effect of Sentiment and Second-Person View on the User Engagement with TikTok News Videos," *Soc. Sci. Comput. Rev.*, vol. 42, no. 1, pp. 201–223, Feb. 2024, doi: 10.1177/08944393231178603.

[17] D. M. Pius, M. Simanjuntak, and C. Umri, "Implementasi Algoritma Klasifikasi untuk Analisis Sentimen Media Sosial Tiktok Tahun 2025," *Jurnal Teknik Informatika dan Teknologi Informasi*, vol. 5, pp. 488–504, Apr. 2025, doi: 10.55606/jutiti.v5i1.5644.

[18] M. F. N. Fathono, E. Y. Puspaningrum, and A. N. Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Jurnal Informatika dan Sains Teknologi*, vol. 1, no. 3, pp. 62–76, Jul. 2024, doi: 10.62951/modem.v1i3.112.

[19] R. Firdaus, I. Asror, and A. Herdiani, "Lexicon-Based Sentiment Analysis of Indonesian Language Student Feedback Evaluation," *Ind. J. Computing*, vol. 6, no. 1, pp. 1–12, Apr. 2021, doi: 10.34818/indojc.2021.6.1.408.

[20] J. Kasundra, C. Schulz, M. Mirsafian, and S. Skylaki, "A Framework for Monitoring and Retraining Language Models in Real-World Applications," *arXiv preprint arXiv:2311.09930*, Nov. 2023. [Online]. Available: http://arxiv.org/abs/2311.09930

[21] R. Savitri, F. Rizki, and A. Sobri, "Implementation of BERT in Sentiment Analysis of National Digital Samsat (SIGNAL) User Reviews Based on Machine Learning," *MATICS: J. Ilmu Komputer dan Teknol. Informasi*, vol. 17, no. 2, pp. 67–75, Sep. 2025, doi: 10.18860/mat.v17i2.32059.

[22] M. Hayaty, S. Muthmainah, and S. M. Ghufran, "Random and Synthetic Over-Sampling Approach to Resolve Data Imbalance in Classification," *Int. J. Artif. Intell. Res.*, vol. 4, no. 2, p. 86, Jan. 2021, doi: 10.29099/ijair.v4i2.152.

[23] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.

[24] F. Destryanto, P. Rizqiyah, and P. Sokibi, "Sentiment Analysis of Public Response to the Free Nutritious Meal Program on Instagram Using IndoBERT," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 5, no. 1, pp. 2808–4519, Feb. 2025.

[25] R. I. Perwira, V. A. Permadi, D. I. Purnamasari, and R. P. Agusdin, "Domain-Specific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 11, no. 1, pp. 30–40, Feb. 2025, doi: 10.20473/jisebi.11.1.30-40.