Dynamic Sentiment Analysis on the Emergence of Pre-Trained Generative Model-Based Applications in Indonesia

Frans Mikael Sinaga[1*], Jefri Junifer Pangaribuan[2], Kelvin[3], Ferawaty[4], Andree Emmanuel Widjaja[5]
Faculty of AI and Data Sciences-Informatics Department, Universitas Pelita Harapan, Indonesia[1, 4]
Faculty of AI and Data Sciences-Information Systems Department, Universitas Pelita Harapan, Indonesia[5]
Information Systems Department-School of Information Systems, Bina Nusantara University, Indonesia[2]
Informatics Faculty-Informatics Department, Universitas Mikroskil, Indonesia[3]

***Abstract*—The emergence of pre-trained generative model–based applications has intensified sentiment dynamics within Indonesia's multi-platform digital ecosystem, where sentiment intensity and temporal fluctuations occur simultaneously. To overcome these challenges, this study extends IndoBERT by incorporating a time-aware tokenization mechanism within a fine-grained dynamic sentiment analysis framework. This mechanism is designed to explicitly capture the evolution of sentiment over time. Instead of relying on external embeddings or implicit timestamps, temporal information is injected directly into the IndoBERT tokenizer through explicit temporal tokens, enabling end-to-end temporal adaptation during fine-tuning. We utilized a large-scale dataset harvested from various platforms—including TikTok, Twitter (X), YouTube, and forums—alongside AI-generated content from Gemini, ChatGPT, and Copilot. The dataset was annotated into five fine-grained sentiment classes: very positive, positive, neutral, negative, and very negative. The experimental evaluation demonstrates that the proposed time-aware IndoBERT model attains an average accuracy of 96.38%, exceeding the performance of the baseline BERT and RoBERTa models. Furthermore, ablation studies validate that the inclusion of time-aware tokenization yields quantifiable performance gains, proving that explicit temporal encoding refines sentiment sensitivity and offers sharper insights into the shifting public opinion in Indonesia.**

***Keywords*—Dynamic sentiment; fine-grained; IndoBERT; multi-platform big data; sentiment analysis**

I. Introduction

Analyzing sentiment from large-scale, multi-platform data in Indonesia becomes increasingly complex when higher sentiment granularity is required [1]. Large volumes of unstructured text produced by social media interactions and AI-generated responses demand advanced modeling techniques to capture contextual sentiment variations accurately [2], [3]. Conventional sentiment analysis frameworks frequently lack classification precision. This deficit is often attributed to three primary factors: suboptimal preprocessing, rigid vectorization techniques, and a shallow grasp of textual context [4], [5], [6]. Sentiment analysis across multiple platforms is crucial because opinions expressed by users differ across digital spaces due to varying audience engagement and platform-specific language use. Studies have shown that collecting data from diverse sources, such as Twitter, TikTok, and forums, improves the robustness of sentiment classification models [7]. Sentiment data are collected from multiple platforms to capture broader and more dynamic expressions of sentiment, enabling better adaptability in different digital platforms [8]. This research focuses on sentiment analysis of AI-generated responses from multiple pre-trained generative models, including ChatGPT, Gemini, Copilot, and others, rather than limiting the study to ChatGPT alone [5], [6]. Recent studies emphasize the growing impact of AI-generated content on public sentiment and the necessity of analyzing various generative models to gain a comprehensive understanding [9].

Fine-grained sentiment analysis advances beyond standard classification methods by transcending simple ternary labels (positive, negative, neutral). Instead, it adopts a more nuanced spectrum that includes five distinct intensity levels: very positive, positive, neutral, negative, and very negative [10]. This approach enables a deeper understanding of sentiment intensity, emotional variations, and nuanced expressions in text [11]. A study on sentiment classification highlights that fine-grained sentiment analysis improves model interpretability by capturing more subtle differences in user opinions, which is particularly useful for complex and multi-faceted sentiment expressions [12]. In the context of Indonesian multi-platform environments, employing a fine-grained analytical approach is indispensable. This is largely due to the complexity of user-generated content, which is frequently characterized by heterogeneous linguistic styles, conflicting sentiments, and subtle, implicit emotional cues [10], [12], [13].

Dynamic sentiment analysis focuses on modeling sentiment changes over time, recognizing that user opinions and emotions fluctuate due to external events, trends, and contextual influences [14]. Traditional sentiment classification often assumes static sentiment representation, which does not account for real-world changes in user emotions [15]. The need for dynamic sentiment analysis arises from various challenges, including the rapid evolution of online discussions, real-time sentiment shifts due to breaking news, and context-dependent emotional expressions [16]. By integrating time-token embeddings [17], sentiment models can account for sentiment shifts across different periods, making classification more adaptive and reflective of real-world changes, where sentiment dynamics play a critical role in decision-making [18], [19].

*Corresponding author.

Recent advancements in deep learning, particularly transformer-based architectures such as BERT (Bidirectional Encoder Representations from Transformers), have significantly improved sentiment classification performance [20]. BERT eliminates the need for extensive preprocessing [21] and enhances classification accuracy through contextual embedding [22], [23]. IndoBERT represents a specialized adaptation of the BERT architecture, pre-trained extensively on Indonesian corpora. This specific training regimen allows it to effectively address the unique linguistic nuances and structural characteristics inherent to the language [24], [25]. Compared with the standard BERT model, IndoBERT provides more effective contextual representations for Indonesian-language text. By modeling the unique morphological, syntactic, and semantic characteristics of the language, IndoBERT attains notably improved accuracy in sentiment classification tasks [26], [27], [28]. Optimizing IndoBERT for sentiment analysis is essential for effectively addressing the complexities of the Indonesian language, reducing misclassifications that arise due to linguistic ambiguity and diverse language structures [29]. IndoBERT enhances sentiment classification through its generation of context-aware embeddings. These representations are particularly effective at modeling long-range dependencies and detecting dynamic semantic shifts within the narrative [30], [31].

The principal aim of this study is to develop and optimize a specialized IndoBERT framework for Fine-Grained and Dynamic Sentiment Analysis, explicitly adapted for Indonesian multi-platform big data. A central component of this research is the integration of time-aware token embeddings, which are utilized to capture sentiment fluctuations and enhance the precision of dynamic analysis [17], leveraging IndoBERT's contextual capabilities to improve Indonesian sentiment classification, utilizing multi-platform data for better generalization across diverse digital environments [32], applying fine-grained sentiment labeling to capture nuanced emotional expressions, and analyzing sentiment trends in AI-generated content across multiple pre-trained generative models (ChatGPT, Gemini, Copilot, etc.). By addressing these challenges, this study offers valuable insights into various societal and economic trends in Indonesia [25] and advances the methodology for sentiment analysis in dynamic online environments. Although prior studies have explored fine-grained and dynamic sentiment analysis, existing approaches largely treat temporal information as auxiliary metadata, with limited investigation into its explicit integration within IndoBERT for multi-platform Indonesian data, particularly in the context of AI-generated content. To address this gap, this study introduces a time-aware tokenization mechanism embedded directly into IndoBERT's tokenizer and embedding space, proposes a fine-grained dynamic sentiment framework with five sentiment intensity levels, validates the contribution of temporal tokens through ablation-based analysis, and performs large-scale evaluations on multi-platform Indonesian big data, including content generated by multiple generative AI models. This research explicitly serves as a significant extension of our preliminary study, which focused primarily on single-platform analysis. The current work expands that scope by integrating temporal dynamics and enlarging the dataset to encompass heterogeneous sources.

The structure of this study is outlined as follows: Section II surveys existing literature relevant to the study. Section III elaborates on the proposed methodology, while Section IV analyzes the experimental results and discussion. The study concludes with a summary of findings in Section V.

II. Related Work

Prior sentiment analysis research has examined fine-grained classification, temporal sentiment modeling, and transformer-based architectures, including BERT and its variants, across various application domains. Fine-grained sentiment analysis has been shown to provide richer emotional representation compared to binary or ternary classification schemes, particularly in social media contexts.

Many dynamic sentiment analysis approaches integrate temporal signals through post-hoc aggregation or external time-series models rather than embedding temporal context directly within language representations. However, such methods often treat temporal signals as auxiliary metadata rather than as integral linguistic features.

In the field of Indonesian natural language processing, IndoBERT has demonstrated strong performance across various applications, with notable effectiveness in sentiment classification. Nevertheless, existing studies largely rely on static representations and do not explicitly integrate temporal information into the tokenization or embedding process. Moreover, limited attention has been given to sentiment analysis on multi-platform Indonesian big data involving AI-generated content.

This study addresses these gaps by introducing a time-aware tokenization mechanism embedded directly into IndoBERT, enabling explicit modeling of sentiment evolution over time within a fine-grained sentiment framework.

III. Methods

As illustrated in Fig. 1, our proposed framework operates through four sequential phases: data acquisition, preprocessing, the injection of time-aware token embeddings, and fine-grained sentiment annotation. To achieve high-resolution sentiment granularity, we employ a lexicon-based approach that classifies text into a five-tier spectrum, ranging from very negative to very positive. A key innovation in this workflow is the integration of time-aware embeddings, which allows the system to explicitly model temporal fluctuations and track the evolution of public opinion. Finally, IndoBERT is deployed as the core classifier; its robust contextual embedding capabilities are essential for detecting nuanced emotional shifts and time-dependent variations across heterogeneous social media platforms.

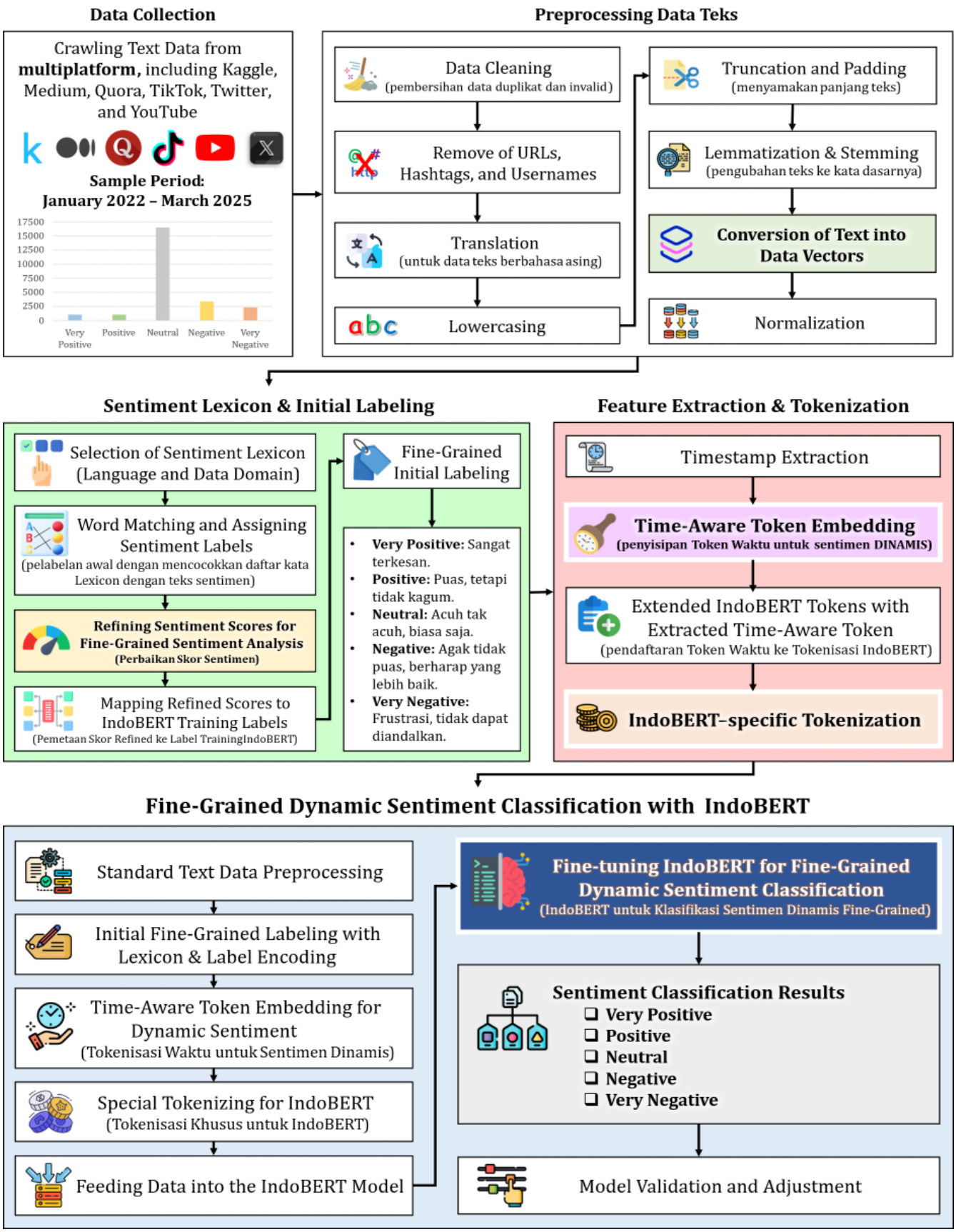

Fig. 1. Procedural framework for the proposed approach.

A. Data Acquisition

To construct a robust and diverse corpus, we executed a comprehensive data harvesting strategy across heterogeneous digital ecosystems. The dataset aggregates textual content from social media, discussion forums, and open repositories, specifically targeting TikTok, Twitter, YouTube, Medium, Quora, and Kaggle. The specific search queries and terms utilized for extraction are enumerated in Table I. Furthermore, to broaden the analytical scope beyond prior studies, we augmented the initial collection with supplementary datasets retrieved using an expanded set of keywords.

TABLE I. KEYWORDS EMPLOYED FOR DATASET COLLECTION

Index	Search Keywords	Dataset Size
1	ChatGPT-based Systems	November 2022 – 2023 25,000 datasets
2	OpenAI Technologies	
3	Education Sector	
4	Academic Research	
5	ChatGPT Copyright Issues	November 2022 – 2024 27,000 datasets
6	Textual Similarity	
7	ChatGPT Usage Disclaimer	
8	Unauthorized AI-Generated News	
9	Politik	May 2024 – 2025 100,000 datasets
10	ChatBot	
11	Gemini (Bard)	
12	Copilot	
13	Generative Model	

Note: Example sentences are presented in the original Indonesian language to preserve semantic and linguistic characteristics relevant to sentiment analysis.

B. Standard Text Data Preprocessing

1) Data cleaning and noise reduction: To ensure the purity of the dataset, we implemented a filtration process to strip away non-textual artifacts. This included the excision of punctuation, emojis, and special characters that offer no semantic value to the classification task, thereby standardizing the input for the model.

2) Removal of metadata artifacts: Metadata markers, specifically hyperlinks (URLs), platform-specific user tags (@username), and topic hash markers (#), were systematically purged. These elements are typically regarded as noise that can obscure the linguistic patterns required for effective feature extraction.

3) Language Unification (Translation): To guarantee linguistic homogeneity across the corpus, we translated all English textual segments into Indonesian. This unification step is critical for mitigating ambiguity caused by code-mixing and ensures the model analyzes a consistent grammatical structure.

4) Case normalization: We applied a uniform case normalization strategy, converting the entire corpus to lowercase. This approach mitigates sparsity issues caused by case sensitivity, ensuring that capitalized and non-capitalized iterations of the same word are recognized as identical tokens, which enhances matching efficiency.

5) Sequence standardization (truncation and padding): To accommodate the fixed input dimension requirements of the deep learning architecture, sequence lengths were standardized. Texts exceeding the maximum length were truncated, while shorter sequences were padded using specific tokens (e.g., [PAD] or zero vectors) to achieve uniformity.

6) Lemmatization and Stemming: We employed both lemmatization and stemming to reduce inflectional variability. Lemmatization was used to revert words to their dictionary base (e.g., restoring "running" to "run"), while stemming stripped suffixes to isolate the root form. This dual approach preserves semantic integrity while minimizing vocabulary size.

C. Initial Labeling with Sentiment Lexicon

In this step, sentiment values are combined using heuristic rules [33], as described in Eq. (1), to generate sentiment labels. This process ensures that sentiment aggregation aligns with contextual meaning, refining the dataset for improved classification performance [see Eq. (1) to Eq. (5)].

$$S_{very\ positive} \sum_{i \in t}^{n} \ very\ positive\ score_i \quad (1)$$

$$S_{positive} \sum_{i \in t}^{n} \ positive\ score_i \quad (2)$$

$$S_{neutral} \sum_{i \in t}^{n} \ neutral\ score_i \quad (3)$$

$$S_{negative} \sum_{i \in t}^{n} \ negative\ score_i \quad (4)$$

$$S_{very\ negative} \sum_{i \in t}^{n} \ very\ negative\ score_i \quad (5)$$

In this framework, initial sentiment labels are generated using a lexicon-based scoring mechanism that aggregates word-level polarity values and contextual modifiers [34]. By utilizing a lexicon-based scoring mechanism to measure intensity, the system assigns one of the five distinct fine-grained labels (very positive, positive, neutral, negative, or very negative) to every sentence, as outlined in Table II. This fine-grained labeling provides a more detailed sentiment distribution, enabling deeper insights into sentiment intensity and emotional variations across multi-platform data [12]. These labels also serve as training data, enabling IndoBERT to capture nuanced sentiment expressions with higher accuracy.

TABLE II. LEXICON-DERIVED DATASET LABELING RESULTS

Lexicon Scores	Lexicon Result	Tweet	Date	Username
-0.6	very negative	Tukang coding 10 orang bisa jd cukup d handle 1 orang aja, mending chatgpt dan ai d musnahkan, karena dapat menghilangkan banyak pekerjaan manusia, yg menyebabkan pengangguran meningkat	Sun Jul 16 03:54:19 +0000 2023	JusDooIt
-0.2	negative	Caution is required when disclosing personal information to conversational AI platforms. Although data sharing may enhance model learning and performance, concerns persist regarding possible unauthorized access to or exploitation of user data.	Thu Jun 29 22:52:50 +0000 2023	leniptri__
0.3	positive	dampak positif AI pada bidang pendidikan terutama di kalangan mahasiswa sangat lah membantu sekali, karena dengan AI akan sangat efektif dan efisien jika digunakan oleh mahasiswa dalam mengerjakan tugas dll.	Thu May 11 03:17:43 +0000 2023	bagussptyno
0.1	neutral	Sebagai sebuah model bahasa buatan, saya tidak belajar seperti manusia pada umumnya. Saya telah dilatih dengan deep learning yang melibatkan pengolahan besar jumlah data dan membangun jaringan saraf tiruan kompleks untuk memahami pola dan struktur bahasa.	Sat Jan 07 10:00:08 +0000 2023	morasaki12
0.4	positive	Syukur lah ada chat gpt semua jadi mudah	Tue Jun 06 05:53:39 +0000 2023	xxbrightvcxx
0.8	very positive	As artificial intelligence continues to advance, conversational agents have emerged as a prominent application. ChatGPT is a recent and sophisticated chatbot designed to interact naturally with users and support various everyday activities. https://t.co/88LLi3nITV	Fri May 24 12:25:58 +0000 2024	user18587

Note: Example sentences are presented in the original Indonesian language to preserve semantic and linguistic characteristics relevant to sentiment analysis

The sentiment labeling process begins by analyzing each word within the text and assigning a score based on its presence in predefined sentiment lexicons. A token identified within the positive lexicon contributes a positive unit increment to the overall score, while a token appearing in the negative lexicon contributes a negative unit decrement. Words that do not belong to either category do not contribute to the overall sentiment score. This word-level scoring serves as the foundation for determining the sentiment at the sentence level. The pseudo-algorithm for this process is presented in Algorithm 1.

Algorithm 1. Pseudo-Algorithm to Label Each Word [35]

```
For each word in the text do
    if the word is in the positive list, then
        sum = sum + 1
    else
        if the word is in the negative list, then
            sum = sum - 1
        End if
    End if
End for
```

Following the word-level sentiment scoring, the overall sentiment of a sentence is determined using a predefined scoring equation, as referenced in Eq. (7). The sentiment initial label is determined based on predefined score thresholds. If the score is 0.6 or higher, the sentence is labeled as very positive (4). Scores between 0.4 and 0.6 are classified as positive (3), while scores ranging from -0.4 to 0.4 are labeled as neutral (2). Sentences with scores between -0.6 and -0.4 are categorized as negative (1), and those with scores of -0.6 or lower are assigned a very negative (0) label.

D. Conversion of Text into Vector Data

We utilized Word2Vec to map words into a continuous vector space, a technique that effectively captures semantic relationships. This process enhances the model's comprehension of word similarities and contextual meanings, thereby improving the overall efficacy of sentiment classification [36]. Furthermore, the generated numerical embeddings streamline subsequent analysis, enabling machine learning algorithms to process textual data with greater efficiency. The results of this transformation are detailed in Table III.

TABLE III. VECTOR DATA RESULT

Before Vectorized	Vector Data Result
chat gpt adalah gamechanger tidak sabar untuk melihat apa yang akan dapat dilakukan di masa depan	[0.23, -0.12, 0.34, 0.07, -0.49, 0.15, 0.23, 0.04, 0.56]

Note: Example sentences are presented in the original Indonesian language to preserve semantic and linguistic characteristics relevant to sentiment analysis.

E. Embedding Time-Aware Token

Integrating time-aware tokens into the sentiment analysis pipeline enables the model to capture temporal variations in sentiment expression more explicitly [19]. This approach allows IndoBERT to dynamically adjust its contextual understanding by incorporating time-specific markers, such as `<hari-senin>`, `<waktu-sore>`, or `<waktu-pagi>`, into the tokenized input [17]. These tokens provide additional contextual information that helps the model recognize sentiment shifts influenced by time-dependent events, social trends, or daily discourse patterns. The process is illustrated in Fig. 2.

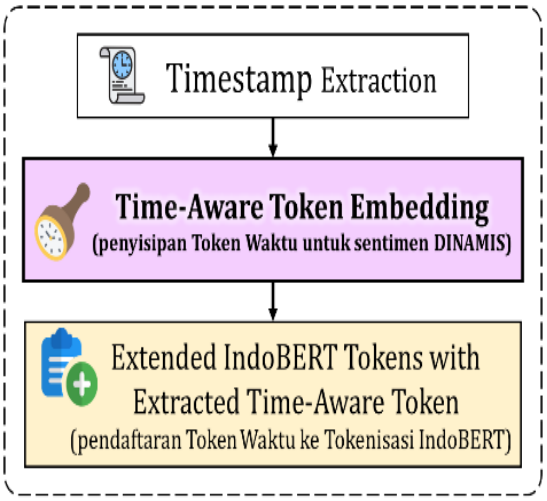

Fig. 2. Time-aware tokenization process.

The process of integrating time-aware tokens into IndoBERT involves the following steps:

1) Extracting temporal information: The initial stage of the procedure focuses on deriving temporal attributes from the timestamp linked to each text instance. In most cases, textual data such as social media posts or online reviews are accompanied by metadata containing date and time information. Based on this metadata, essential temporal features—including the day of the week (e.g., Monday, Tuesday) and time-of-day segments (e.g., morning, afternoon, evening, night). These features provide crucial context, as sentiment expression often varies depending on the time of posting. For example, users may express frustration more frequently on Monday mornings due to work-related stress, while positive sentiments may be more prevalent on Friday evenings due to weekend anticipation.

2) Mapping temporal features to time-aware tokens: Once the relevant time-based features are extracted, they are mapped to predefined time-aware tokens. Each time unit is assigned a corresponding token representation that IndoBERT can recognize. For example, the day of the week is converted into tokens such as `<hari-senin>` or `<hari-sabtu>`, while the time of day is mapped to tokens like `<waktu-pagi>` or `<waktu-sore>`. These tokens serve as additional input features that help the model understand the temporal context in which a given text was written. For example, a tweet posted on "Wednesday at 08:30 AM" would be mapped to `<hari-senin> <waktu-pagi>`. By incorporating these tokens, the model can better capture variations in sentiment that correlate with specific time periods.

3) Appending time-aware tokens to text data: After the tokens are generated, they are appended to the original text before further processing. The tokens are typically inserted at the beginning of the sentence to ensure that IndoBERT processes the temporal context before analyzing the main content of the text. Table IV provides an example of tokenized input with time-aware tokens applied to different sentences.

TABLE IV. PREPROCESSED SENTENCES WITH EMBEDDED TIME-AWARE TOKENS

Timestamp	Sentence	Tokenized Input with Time-Aware Token
Wed Mar 25 14:21:55 +0000 2022	ChatGPT sangat membantu dalam pekerjaan saya!	<hari-rabu> <waktu-siang> chatgpt sangat membantu dalam pekerjaan saya !
Tue Aug 8 17:59:38 +0000 2023	Diskusi dengan ChatGPT tentang sejarah dunia sangat menarik, tetapi terkadang jawabannya terlalu umum dan kurang mendalam.	<hari-sabtu> <waktu-sore> diskusi dengan chatgpt tentang sejarah dunia sangat menarik , tetapi terkadang jawabannya terlalu umum dan kurang mendalam .
Sun Feb 24 23:47:16 +0000 2024	Semakin sering saya pakai, semakin terasa kurang akurat. Dan juga, kenapa hari ini ChatGPT jadi sering error?	<hari-minggu> <waktu-malam> semakin sering saya pakai , semakin terasa kurang akurat . dan juga , kenapa hari ini chatgpt sering error ?

Note: Example sentences are presented in the original Indonesian language to preserve semantic and linguistic characteristics relevant to sentiment analysis.

This ensures that the model is aware of when the statement was made before it begins extracting sentiment-related features. By explicitly including this information, IndoBERT can learn to associate specific sentiment trends with different time periods.

4) Expanding IndoBERT's Tokenizer Vocabulary: Since IndoBERT's default tokenizer does not contain these new time-aware tokens, it must be extended to recognize and process them correctly. This involves modifying the tokenizer's vocabulary by adding the new tokens to the model's predefined token set. The process includes:

a) Updating the tokenizer vocabulary: The new tokens, such as `<hari-jumat>`, `<waktu-pagi>` and `<waktu-malam>`, are explicitly added to IndoBERT's tokenizer to prevent them from being broken into subword units. This ensures that they are treated as single, meaningful tokens rather than being split into unrelated subwords.

b) Adjusting token embeddings: Since IndoBERT's existing embeddings do not include representations for the new tokens, additional embeddings are initialized for them. These embeddings can be either randomly initialized and learned during fine-tuning or mapped using similar contextual embeddings from pre-existing words that represent temporal concepts.

c) Ensuring consistent token handling: The modified tokenizer ensures that every input text containing time-aware tokens is processed consistently. Without this step, the new tokens might be treated as unknown (`[UNK]`) and ignored by the model, reducing the effectiveness of time-based sentiment learning.

Experimental results indicate that the inclusion of time-aware tokens improves the classification performance across various sentiment categories. The model demonstrates an increased ability to distinguish between sentiment fluctuations occurring at different times, particularly in social media data where user opinions may change based on external influences such as current events, holidays, or market trends. Comparative evaluations between IndoBERT with and without time-aware

embedding reveal a notable improvement in sentiment classification accuracy, especially for categories with high temporal sensitivity, such as political discussions or real-time product reviews.

F. Special Tokenization for IndoBERT

Text is tokenized according to IndoBERT's linguistic structure, ensuring proper segmentation for effective processing. The updated IndoBERT tokenizer recognizes time-aware tokens as distinct entities, preventing them from being broken into subwords. This modification preserves the temporal context embedded in the input.

During encoding, each token, including the new time-aware markers, is mapped to an embedding vector. Since these tokens were not in the original IndoBERT vocabulary, their embeddings are either initialized randomly or derived from semantically related words. This enables IndoBERT to incorporate temporal context dynamically, refining sentiment predictions based on time-sensitive factors. The outcomes of this tokenization process are presented in Table V.

TABLE V. TOKENIZING RESULT

Before tokenizing text	After tokenizing text
ah temen gue jg ada yg essay nya dibuat pure pake ChatGPT lolos and he being proud of it soalnya dia tinggal kirim prompt juga ujungnya lolos jd awardee	['ah' , 'temen' , 'gue' , 'jg' , 'ada' , 'yg' , 'essay' , 'nya' , 'dibuat' , 'murni' , 'pake', 'ChatGPT' , 'lolos' , 'dan' , 'dia', 'merasa' , 'bangga' , 'dengan' , 'itu' , 'soalnya' , 'dia' , 'tinggal' , 'kirim' , 'prompt' , 'juga' , 'ujungnya' , 'lolos' , 'jd' , 'pemenang']

G. Classification Using IndoBERT

For the core classification engine, we deployed IndoBERT [25], a transformer architecture pre-trained specifically on Indonesian corpora via masked language modeling. Unlike the standard BERT model, which is optimized for English, IndoBERT offers superior contextual embedding for Indonesian linguistics. By fine-tuning this model on our fine-grained annotated dataset, we enabled it to distinguish between the five distinct sentiment tiers with high precision. The detailed architecture and processing stages are described below:

1) Text Processing Pipeline for IndoBERT: Each token must be transformed into a vector-based representation through an embedding mechanism [see Eq. (6)].

$$V_t = Trans\ (W_t X_a + b_t) \tag{6}$$

2) Initial fine-grained labeling with sentiment lexicon: The use of sentiment lexicons for labeling is a prevalent method in sentiment classification. This technique involves mapping words and phrases to specific sentiment scores according to established polarity values, as in Eq. (7) [34]:

$$\sum_{i=1}^{k} \frac{(p-n)_i}{k} \tag{7}$$

where, k represents the total number of words in the text that are listed in the lexicon, and $p - n$ denotes the sentiment polarity score associated with each word in the lexicon, which is determined based on its occurrence in positive and negative categories [37].

3) Time-aware token embedding: Temporal tokens (e.g., `<hari-senin>`, `<waktu-siang>`) are incorporated into the text to enhance dynamic sentiment analysis by capturing variations influenced by time-based factors.

4) IndoBERT-specific tokenization: The tokenized text, including standard tokens and newly introduced time-aware tokens, is processed using the IndoBERT tokenizer.

5) Feeding data into the IndoBERT Model: IndoBERT generates vector embeddings for individual tokens derived from the input text, as in Eq. (8):

$$V_b = Bert(W_b X_b + b_b) \tag{8}$$

6) Self-Attention: The self-attention mechanism enhances token representations by incorporating contextual information from the surrounding text, as in Eq. (9):

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{9}$$

In this equation, Q, K, V denote the query, key, and value matrices corresponding to the input tokens.

7) Position-wise feed-forward network: Applied after the Self-Attention stage, FFNN layers refine tokens into more intricate representations. This transformation is critical for capturing nonlinear patterns within the text, which in turn leads to improved model understanding and robustness in generalization [see Eq. (10)].

$$FFN(H) = ReLU(HW_1 + b_1)W_2 + b_2 \tag{10}$$

H denotes the contextual embedding generated by self-attention, whereas W1, W2, b1, and b2 define the feed-forward network parameters.

8) Fine-tuning IndoBERT: The fine-tuning phase entails retraining IndoBERT using domain-specific sentiment datasets. This process is designed to optimize classification metrics while preserving the model's original contextual integrity [24], [25].

9) Sentiment classification: In the classification stage, IndoBERT analyzes the input text and assigns it to one of five fine-grained sentiment categories, utilizing the vector representations generated during processing.

H. Data Testing

The evaluation process employs a confusion matrix to summarize prediction results into TP, TN, FP, and FN. In this framework, TP and TN indicate correct predictions, whereas FP and FN reflect classification errors. The effectiveness of the proposed model is then quantified through the accuracy measure, as presented in Eq. (10).

IV. PERFORMANCE EVALUATION AND DISCUSSION

The proposed methodology integrates several stages, including data collection, text preprocessing, sentiment annotation using a lexicon-based approach, time-aware token embedding, and fine-grained sentiment classification based on the IndoBERT model. In particular, the preprocessing step involves text standardization, selective translation of English expressions into Indonesian, token segmentation, sequence

truncation, and padding, aiming to enhance the quality of input representations for IndoBERT.

A. Fine-Tuned IndoBERT Model

The fine-tuning process adjusts IndoBERT's parameters using task-specific data, thereby augmenting its capacity to recognize complex linguistic structures. As illustrated in Fig. 3, the proposed architecture supports this optimization process by adapting the model to task-specific characteristics, thereby leading to improvements in accuracy, precision, and recall.

A total of roughly 24,000 data samples are included in the dataset and are distributed across fine-grained sentiment categories, as presented in Table VI. Analyzing this label distribution, which is visualized in Fig. 4, is essential for assessing IndoBERT's capacity to manage class imbalances and understanding their potential influence on the training process.

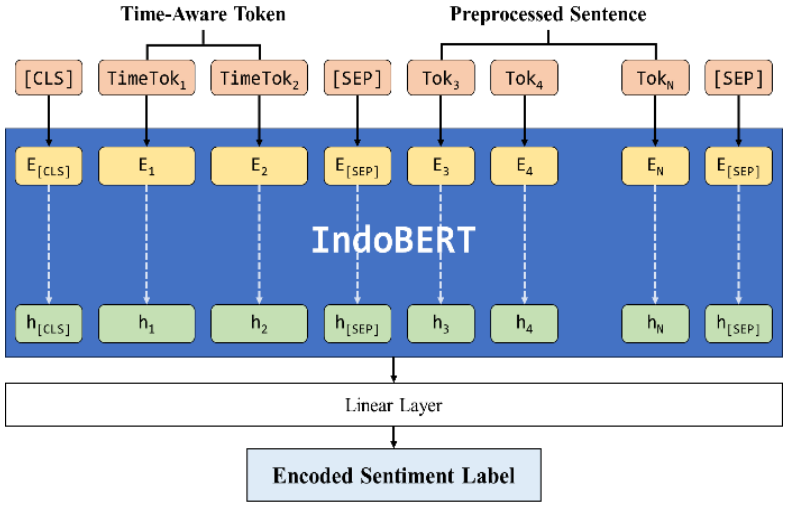

Fig. 3. Time-aware token–enhanced fine-tuned IndoBERT.

TABLE VI. FINE-GRAINED SENTIMENT LABEL STATISTICS

Label	Training Set	Validation Set	Testing Set
Very Positive	1,153	346	149
Positive	920	276	118
Neutral	5,684	1,705	731
Negative	2,213	663	285
Very Negative	1,419	426	182
Total Data	11,389	3,416	1,465

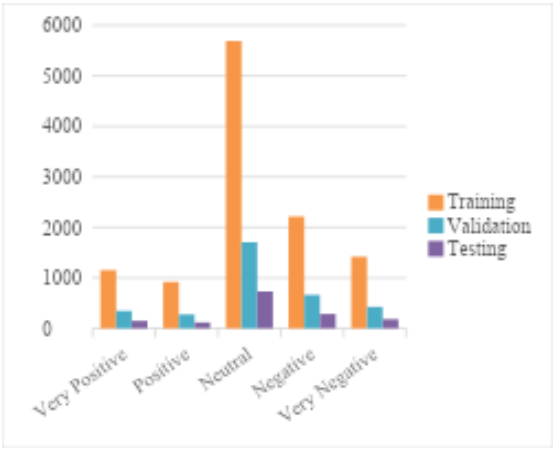

Fig. 4. Distribution of fine-grained sentiment labels.

Fig. 5 further demonstrate the convergence characteristics and training progression of the model across successive epochs.

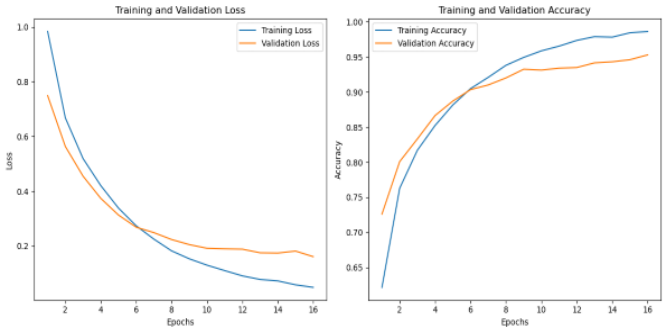

Fig. 5. Validation accuracy and loss of the IndoBERT model.

Table VII presents the evaluation results on the test set in terms of accuracy, precision, recall, and F1-score, highlighting the performance of the fine-tuned IndoBERT model.

TABLE VII. PERFORMANCE EVALUATION OF THE FINE-TUNED INDOBERT MODEL ON THE TEST DATASET

Sentiment Class	Precision	Recall	F1-Score	Sample Count
Very Positive Class	99%	94%	96%	149
Positive Class	91%	97%	94%	118
Neutral Class	98%	98%	98%	731
Negative Class	94%	95%	95%	285
Very Negative Class	94%	95%	95%	182
Overall Testing Accuracy				96.38%

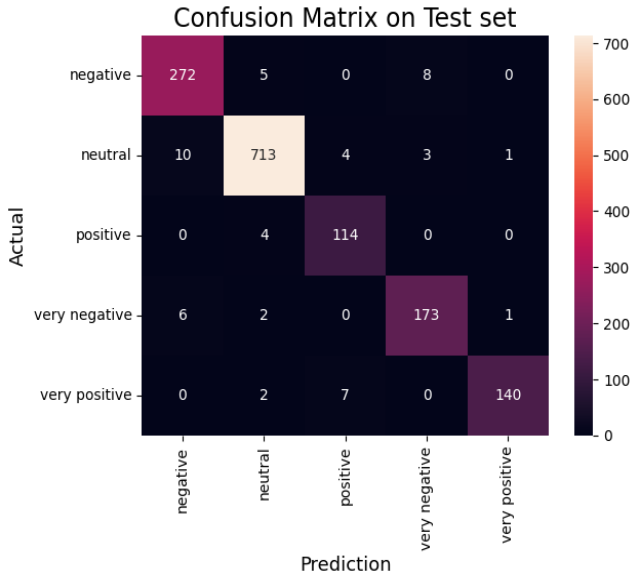

Fig. 6. Confusion matrix of the model on the test set.

B. Influence of Dataset Size on Model Performance

To examine the effect of training data volume on model performance, a series of experiments was performed using multiple dataset subsets with varying sizes and label distributions across the training, validation, and testing stages. As summarized in Table VIII, the results reveal a consistent pattern in which increasing the size of the training dataset generally leads to improved classification accuracy.

TABLE VIII. MODEL PERFORMANCE ACROSS DIFFERENT DATASET SCALES

Raw Data	Data Size			
	850*	***2,000***	***25,000***	***36,000***
Training set	577	980	17,091	21,867
Validation set	159	294	5,128	6,747
Testing set	88	126	2,198	2,624
Classification Accuracy	100%	75%	96%	89%

*Data collected from a single platform.

The attainment of perfect accuracy on the smallest dataset consisting of 850 samples indicates a strong tendency toward overfitting, suggesting that the model memorizes individual instances rather than capturing patterns that generalize well. While smaller datasets may reduce variability and temporarily simplify the classification process, they ultimately compromise model robustness. Furthermore, increasing the dataset size beyond a certain point yields diminishing performance gains. These observations underscore the importance of maintaining a balanced trade-off among dataset size, class distribution, and computational efficiency during the fine-tuning of the IndoBERT model.

C. Performance Evaluation with and without Time Tokens

To evaluate the impact of embedding time-aware tokens in the sentiment analysis model, we conducted experiments with and without time token embeddings. The model was trained, validated, and tested on the same dataset under both conditions.

When trained without time token embeddings, the model yielded a training accuracy of 99.01%, followed by 95.57% for validation and 95.91% for testing. These metrics demonstrate the model's robust generalization capabilities, preserving high accuracy levels on unseen data.

When time token embeddings were incorporated, the model demonstrated slight improvements in accuracy. The training accuracy increased to 99.53%, while validation and test accuracy improved to 95.73% and 96.38%, respectively. The classification performance is visualized in the confusion matrix shown in Fig. 6 and the comparison shown in Fig. 7.

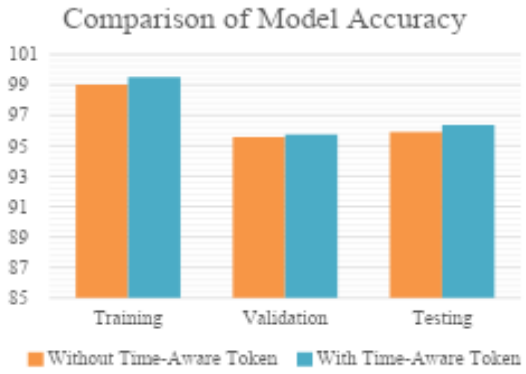

Fig. 7. Comparison of model accuracy.

The inclusion of time tokens led to marginal improvements in both validation and test accuracy. This suggests that incorporating temporal context allows the model to better capture sentiment variations over different time frames. While the gain in accuracy is relatively small, it highlights the potential of leveraging temporal information in sentiment classification tasks. Overall, the results indicate that embedding time-tokens enhances model performance, albeit modestly.

D. Comparison of Different Base Model Performance

We conducted a comparative evaluation of BERT, IndoBERT, and RoBERTa to identify the optimal architecture for practical fine-grained sentiment classification. To guarantee a rigorous assessment, each model underwent fine-tuning with hyperparameters optimized for its specific design. The data in Table IX confirms IndoBERT as the top performer, achieving superior accuracy and outperforming its counterparts in all sentiment classes.

TABLE IX. BENCHMARKING PERFORMANCE AMONG DISTINCT ARCHITECTURES

Base Model	BERT Model	RoBERTa Model	IndoBERT Model
Training Data Size	12,000	24,000	24,000
Number of Epoch	16	16	16
Learning Rate	2e-6	9e-6	2e-6
Training Set Accuracy	95.53%	97.90%	99.53%
Validation Set Accuracy	92.94%	95.68%	95.73%
Testing Accuracy	93.19%	95.90%	**96.38%**

IndoBERT's superior performance stems from its pretraining on Indonesian corpora, which facilitates a deeper grasp of context and more accurate word representations. Conversely, the English-pretrained BERT (bert-base-uncased) encounters difficulties with Indonesian tokenization and semantic interpretation. Similarly, RoBERTa (roberta-base), despite its optimized architecture for longer sequences, fails to deliver substantial improvements due to a lack of linguistic adaptation. These results demonstrate the critical importance of language-specific pretraining, validating IndoBERT as the most suitable architecture for addressing the complexities of fine-grained and dynamic sentiment analysis in Indonesian.

The comparison clearly demonstrates that pre-training on the target language is a prerequisite for high accuracy. Consequently, IndoBERT is validated as the superior architecture for addressing the complex nuances of dynamic and fine-grained sentiment analysis in Indonesian.

E. Analysis and Discussion

Current literature underscores the efficacy of IndoBERT in sentiment analysis, noting its ability to perform well even with minimal preprocessing. This performance is further augmented by the incorporation of knowledge-based strategies, such as Sentiment Lexicons. The integration of such lexicons is particularly advantageous for fine-grained and dynamic sentiment analysis, as it significantly refines the granularity required to classify text into multiple sentiment levels.

Conventional classifiers like Naïve Bayes, LSTM, and SVM often face challenges when attempting to capture sentiment nuances in complex multilingual datasets. In contrast, transformer architectures—specifically BERT, RoBERTa, and IndoBERT—demonstrate significantly higher efficacy. Their advantage lies in the use of self-attention mechanisms and deep contextual embeddings, which allow for a more sophisticated understanding of text.

BERT (Bidirectional Encoder Representations from Transformers) introduced the concept of deep bidirectional learning, enabling better contextual understanding compared to previous sequential models like LSTMs [38]. However, BERT was trained primarily on English corpora, making it less effective for sentiment tasks in Indonesian text.

RoBERTa (Robustly Optimized BERT Pretraining) improves upon BERT by employing a longer training duration, dynamic masking, and removal of next-sentence prediction (NSP). While it has demonstrated improvements in general NLP tasks, its effectiveness in sentiment analysis for Indonesian remains an area of exploration. Some studies suggest that RoBERTa excels in capturing subtle sentiment shifts due to its improved pretraining strategies, particularly in domain-specific applications [39].

IndoBERT, a variant of BERT specifically pre-trained on Indonesian corpora, addresses the limitations of standard BERT by providing better tokenization and contextual understanding for the Indonesian language. Its self-attention mechanisms enable it to retain long-range dependencies, making it highly effective for processing unstructured text [30], [31]. Moreover, Sentiment Lexicon-based labeling enhances IndoBERT's ability to capture nuanced emotions [2], [3], complementing its contextual representations [27], [28]. While LSTM and CNN-based models struggle with handling lengthy and context-dependent texts [5], [6], IndoBERT's deep contextual embeddings dynamically adjust word representations based on surrounding text, ensuring more accurate sentiment classification [26].

By incorporating time-aware tokenization, the proposed method effectively adapts to temporal shifts, rendering it highly suitable for monitoring evolving discourse in multi-platform environments. Furthermore, the strategic synergy of IndoBERT, Sentiment Lexicons, and Dynamic Sentiment modeling establishes a robust architecture designed for accurate, fine-grained sentiment analysis.

V. Conclusion

The findings indicate that combining sentiment lexicons with a fine-grained, dynamic labeling strategy significantly improves sentiment classification accuracy for Indonesian multi-platform data. By utilizing a five-point intensity scale—ranging from very negative to very positive—this framework offers a more detailed and expressive representation than traditional binary or ternary models. Furthermore, the synergy between lexicon-based labeling and IndoBERT's contextual understanding allows for the precise identification of subtle sentiment nuances across various social media contexts.

By embedding temporal information directly into the tokenization process, IndoBERT is able to model sentiment evolution more effectively than static representation-based approaches. The findings indicate that temporal context should be treated as a linguistically meaningful feature rather than merely auxiliary metadata, particularly in rapidly evolving online discussions, including content influenced by generative AI technologies. Ablation-based analysis confirms that the incorporation of temporal tokens contributes directly to the observed performance improvements.

Despite these contributions, this study has several limitations. The reliance on lexicon-based initial labeling may introduce bias in ambiguous or context-dependent sentiment expressions, and the temporal tokens are currently limited to coarse-grained time units. Future work will explore adaptive temporal granularity, enhanced sentiment scoring mechanisms, and the integration of multimodal information to further improve dynamic sentiment understanding and long-term sentiment trend analysis.

Acknowledgment

The authors extend their gratitude to all institutions and individuals who supported this study. Specifically, Frans Mikael Sinaga (Universitas Pelita Harapan) spearheaded the research design and constructed the dynamic sentiment analysis framework; Jefri Junifer Pangaribuan (Bina Nusantara University) contributed to data collection, preprocessing, and sentiment lexicon development; Kelvin (Universitas Mikroskil) supported the implementation of fine-grained sentiment analysis and labeling strategies; Ferawaty (Universitas Pelita Harapan) contributed to linguistic validation and contextual sentiment interpretation; and Andree Emmanuel Widjaja (Universitas Pelita Harapan) was responsible for model optimization, validation, and performance evaluation using

IndoBERT. The authors also acknowledge the institutional support provided by Universitas Pelita Harapan, Bina Nusantara University, and Universitas Mikroskil.

REFERENCES

[1] F. M. Sinaga, R. Purba, S. J. Pipin, W. S. Lestari, and S. Winardi, "Optimization of sentiment analysis classification of chatgpt on big data twitter in Indonesia using BERT," *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, p. 1665, July 2024.

[2] R. Nakka, T. S. Lakshmi, D. Priyanka, N. R. Sai, S. P. Praveen, and U. Sirisha, "LAMBDA: Lexicon and aspect-based multimodal data analysis of tweet," *Ingénierie des systèmes d information*, vol. 29, no. 3, pp. 1097–1106, June 2024.

[3] D. Bao and W. Su, "Optimizing deep learning-based natural language processing for sentiment analysis," *International Journal of High Speed Electronics and Systems*, January 2025.

[4] S. Efendi and P. Sihombing, "Sentiment analysis of food order tweets to find out demographic customer profile using SVM," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 583–594, July 2022.

[5] F. M. Sinaga, S. J. Pipin, S. Winardi, K. M. Tarigan, and A. P. Brahmana, "Analyzing sentiment with self-organizing map and long short-term memory algorithms," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 1, pp. 131–142, November 2023.

[6] S. J. Pipin, F. M. Sinaga, S. Winardi, and M. N. Hakim, "Sentiment analysis classification of ChatGPT on twitter big data in Indonesia using fast R-CNN," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 4, p. 2137, October 2023.

[7] A. C. Nanayakkara, B. T. G. S. Kumara, and R. M. Kapila Tharanga Rathnayaka, "Cross-platform topic analysis: identifying trends and divergences in social media discourse," in *2024 8th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI)*, IEEE, pp. 1–6, December 2024.

[8] M. M. Agüero-Torales, J. I. Abreu Salas, and A. G. López-Herrera, "Deep learning and multilingual sentiment analysis on social media data: An overview," *Applied Soft Computing*, vol. 107, p. 107373, August 2021.

[9] W. Feng, Y. Li, C. Ma, and L. Yu, "From ChatGPT to sora: Analyzing public opinions and attitudes on generative artificial intelligence in social media," *IEEE Access*, vol. 13, pp. 14485–14498, 2025.

[10] L. Zhu, Y. Xu, Z. Zhu, Y. Bao, and X. Kong, "Fine-grained sentiment-controlled text generation approach based on pre-trained language model," *Applied Sciences*, vol. 13, no. 1, p. 264, December 2022.

[11] W. Sofiya and E. B. Setiawan, "Fine-grained sentiment analysis in social media using gated recurrent unit with support vector machine," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 3, pp. 511–519, June 2023.

[12] R. Chauhan, A. Gusain, P. Kumar, C. Bhatt, and I. Uniyal, "Fine grained sentiment analysis using machine learning and deep learning," in *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)*, IEEE, pp. 423–427, Sep. 2023.

[13] Y. Xiao, C. Li, M. Thürer, Y. Liu, and T. Qu, "User preference mining based on fine-grained sentiment analysis," *Journal of Retailing and Consumer Services*, vol. 68, p. 103013, September 2022.

[14] H. Liang, U. Ganeshbabu, and T. Thorne, "A dynamic bayesian network approach for analysing topic-sentiment evolution," *IEEE Access*, vol. 8, pp. 54164–54174, 2020.

[15] Jency Jose and Simritha R, "Sentiment analysis and topic classification with lstm networks and textrazor," *International Journal of Data Informatics and Intelligent Computing*, vol. 3, no. 2, pp. 42–51, June 2024.

[16] K. MOHAMAD and K. M. KARAOĞLAN, "Enhancing deep learning-based sentiment analysis using static and contextual language models," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 12, no. 3, pp. 712–724, September 2023.

[17] C. Volpetti, K. Vani, and A. Antonucci, "Temporal word embeddings for narrative understanding," in *Proceedings of the 2020 12th International Conference on Machine Learning and Computing*, New York, NY, USA: ACM, February 2020.

[18] X. ShiXiao, M. Muwafak Alobaedy, S. B. Goyal, S. Singla, S. Kang, and R. Chadha, "Real-time sentiment analysis on social networks using meta-model and machine learning techniques," *Scalable Computing: Practice and Experience*, vol. 24, no. 4, pp. 819–834, November 2023.

[19] M. Rodriguez-Ibanez, F.-J. Gimeno-Blanes, P. M. Cuenca-Jimenez, S. Munoz-Romero, C. Soguero, and J. L. Rojo-Alvarez, "On the statistical and temporal dynamics of sentiment analysis," *IEEE Access*, vol. 8, pp. 87994–88013, 2020.

[20] O. Ndama, I. Bensassi, and E. Mokhtar En-Naimi, "The impact of BERT-infused deep learning models on sentiment analysis accuracy in financial news," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 2, pp. 1231–1240, April 2025.

[21] M. Pota, M. Ventura, R. Catelli, and M. Esposito, "An effective BERT-based pipeline for twitter sentiment analysis: A case study in Italian," *Sensors (Switzerland)*, vol. 21, no. 1, pp. 1–21, January 2021.

[22] A. Zhao and Y. Yu, "Knowledge-enabled BERT for aspect-based sentiment analysis," *Knowledge Based System*, vol. 227, September 2022.

[23] M. P. Geetha and D. Karthika Renuka, "Improving the performance of aspect based sentiment analysis using fine-tuned BERT base uncased model," *International Journal of Intelligent Networks*, vol. 2, pp. 64–69, 2021.

[24] F. Baharuddin and M. F. Naufal, "Fine-tuning IndoBERT for Indonesian exam question classification based on bloom's taxonomy," *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, no. 2, pp. 253–263, November 2023.

[25] E. Yulianti and N. K. Nissa, "ABSA of Indonesian customer reviews using IndoBERT: single-sentence and sentence-pair classification approaches," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3579–3589, October 2024.

[26] H. Imaduddin, F. Y. A'la, and Y. S. Nugroho, "Sentiment analysis in Indonesian healthcare applications using IndoBERT approach," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, 2023.

[27] Taufiq Dwi Purnomo and Joko Sutopo, "Comparison of pre-trained BERT-based transformer models for regional language text sentiment analysis in Indonesia," *International Journal Science and Technology*, vol. 3, no. 3, pp. 11–21, November 2024.

[28] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, "Hybrid models for emotion classification and sentiment analysis in Indonesian language," *Applied Computational Intelligence and Soft Computing*, vol. 2024, no. 1, January 2024.

[29] A. F. Hidayatullah, "Parameter efficient fine-tuning using low-rank adaptation for emotion classification in Indonesian texts," in *2024 9th International Conference on Information Technology and Digital Applications (ICITDA)*, IEEE, pp. 1–6, November 2024.

[30] P. M. Gavali and S. K. Shiragave, "Text representation for sentiment analysis: From static to dynamic," in *2023 3rd International Conference on Smart Data Intelligence (ICSMDI)*, IEEE, pp. 99–105, March 2023.

[31] N. Alamsyah, E. Noersasongko, G. F. Shidik, and N. Rijati, "Fine-grained sentiment classification of public opinion on electric cars in Indonesia using IndoBERT," in *2024 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, pp. 502–508, Sep. 2024.

[32] C. Zhao, S. Wang, and D. Li, "Multi-source domain adaptation with joint learning for cross-domain sentiment classification," *Knowledge Based System*, vol. 191, p. 105254, March 2020.

[33] M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. AlSobeh, K. Mansour, and H. Migdady, "Emo-SL framework: Emoji sentiment lexicon using text-based features and machine learning for sentiment analysis," *IEEE Access*, vol. 12, pp. 81793–81812, 2024.

[34] S. Limboi and L. Dioşan, "A lexicon-based feature for twitter sentiment analysis," in *2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP)*, IEEE, pp. 95–102, September 2022.

[35] B. Y. Ziwei and H. N. Chua, "A Depression Diagnostic System using Lexicon-based Text Sentiment Analysis," 2022.

[36] Q. Wang, "Research on the Application Algorithm of Natural Language Processing Technology in Text Analysis," in *2024 IEEE 6th International*

Conference on Civil Aviation Safety and Information Technology (ICCASIT), IEEE, pp. 1429–1433, October 2024.

[37] S. Consoli, L. Barbaglia, and S. Manzan, "Fine-grained, aspect-based sentiment analysis on economic and financial lexicon," *Knowledge Based System*, vol. 247, p. 108781, July 2022.

[38] O. Ndama, I. Bensassi, and E. Mokhtar En-Naimi, "The impact of BERT-infused deep learning models on sentiment analysis accuracy in financial news," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 2, pp. 1231–1240, April 2025.

[39] J. Sirusstara, N. Alexander, A. Alfarisy, S. Achmad, and R. Sutoyo, "Clickbait headline detection in indonesian news sites using robustly optimized BERT pre-training approach (RoBERTa)," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, IEEE, pp. 1–6, Sep. 2022.