OPEN ACCESS

Volume 10 (3), July 2026, 1225-1238

E-ISSN:2580-1643

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: https://doi.org/10.35870/jtik.v10i3.6418

IndoBERT-Based Natural Language Processing for Early Detection of Mental Disorders among Indonesian Gen-Z Students: A Mobile Application Approach with Logistic Regression Baseline

Athif Basyar Mussafa [1*], Widi Hastomo [2]

[1*,2] *Department of Information Technology, Institut Teknologi dan Bisnis Ahmad Dahlan, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia.*

a r t i c l e   i n f o

Article history:
Received 25 December 2025
Received in revised form 20 January 2026
Accepted 1 February 2026
Available online July 2026.

Keywords:
IndoBERT Natural Language Processing; Logistic Regression; Mental Healthl; Detection
Generation Z.

Kata Kunci:
IndoBERT Pemrosesan Bahasa Alami; Regresi Logistik; Kesehatan Mental; Deteksi
Generasi Z.

a b s t r a c t

Mental health issues have become a growing concern among young adults, while access to professional psychological services remains limited. Most existing digital mental health applications rely mainly on self-report questionnaires and lack the ability to interpret contextual emotional expressions found in user-written text, which reduces their effectiveness for early screening. This study proposes the design and implementation of a mobile-based mental health detection system that integrates contextual natural language processing with interactive assessment features. The system analyzes Indonesian-language textual reflections using an IndoBERT-based classification model and complements the results with a rule-based psychological scoring mechanism derived from questionnaire responses. Logistic Regression with TF–IDF features is employed as a baseline model for comparative evaluation. System performance is assessed using accuracy, precision, recall, and F1-score metrics. Experimental results show that the IndoBERT model outperforms the baseline, achieving an accuracy of 97.79%, compared to 94.17% for Logistic Regression. The proposed system is implemented as a Flutter-based mobile application to improve accessibility to early mental health screening among Indonesian university students. This study integrates two complementary approaches: NLP-based text classification using IndoBERT and rule-based psychological scoring derived from self-report questionnaires.

a b s t r a k

Masalah kesehatan mental telah menjadi perhatian yang meningkat di kalangan dewasa muda, sementara akses ke layanan psikologis profesional masih terbatas. Sebagian besar aplikasi kesehatan mental digital yang ada sebagian besar bergantung pada kuesioner laporan diri dan kurang mampu menafsirkan ekspresi emosional kontekstual yang ditemukan dalam teks yang ditulis pengguna, yang mengurangi efektivitasnya untuk skrining dini. Studi ini mengusulkan desain dan implementasi sistem deteksi kesehatan mental berbasis mobile yang mengintegrasikan pemrosesan bahasa alami kontekstual dengan fitur penilaian interaktif. Sistem ini menganalisis refleksi tekstual berbahasa Indonesia menggunakan model klasifikasi berbasis IndoBERT dan melengkapi hasilnya dengan mekanisme penilaian psikologis berbasis aturan yang berasal dari respons kuesioner. Regresi Logistik dengan fitur TF–IDF digunakan sebagai model dasar untuk evaluasi komparatif. Kinerja sistem dinilai menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa model IndoBERT mengungguli model dasar, mencapai akurasi 97,79%, dibandingkan dengan 94,17% untuk Regresi Logistik. Sistem yang diusulkan diimplementasikan sebagai aplikasi mobile berbasis Flutter untuk meningkatkan aksesibilitas skrining kesehatan mental dini di kalangan mahasiswa universitas Indonesia. Studi ini mengintegrasikan dua pendekatan yang saling melengkapi: klasifikasi teks berbasis NLP menggunakan IndoBERT dan penilaian psikologis berbasis aturan yang diperoleh dari kuesioner laporan diri.

Corresponding Author. Email: athifbaz@gmail.com [1].

1. Introduction

Mental health problems among Generation Z have become a growing global concern, particularly among students who experience academic pressure, social demands, and rapid lifestyle changes. The World Health Organization reports that approximately one in seven adolescents worldwide experiences a mental health condition, with anxiety and depression being the most common disorders (World Health Organization, 2026). Furthermore, nearly half of all mental health disorders begin before the age of 18, yet many remain undetected and untreated, leading to long-term psychological and social consequences (World Health Organization, 2026). These findings highlight the importance of early detection and timely intervention to prevent the progression of mental health disorders. Generation Z university students represent a particularly vulnerable population due to the intersection of academic pressure, identity development, and intensive digital technology usage. Previous studies indicate that Gen-Z students experience higher levels of anxiety, depression, and emotional distress compared to previous generations, while simultaneously exhibiting lower rates of professional help-seeking behavior.

As digital natives, this generation increasingly expresses emotional states through short-form reflective texts in digital environments, making them a relevant target population for NLP-based mental health screening systems. Therefore, focusing on Indonesian Gen-Z university students allows this study to address both a high-risk demographic group and a population that naturally interacts with text-based digital platforms. The increasing availability of mobile health technologies has encouraged the development of digital mental health applications aimed at improving accessibility to psychological support. Recent reviews show that mobile mental health applications, including AI-based apps for children and adolescents, can support self-monitoring, enhance mental health awareness, and facilitate engagement with mental wellness tools, although rigorous clinical validation remains limited (Yang *et al.*, 2025; Endriyani & Susanti, 2024). However, most existing applications primarily rely on manual self-assessment questionnaires and subjective user input. Although questionnaire-based self-assessment is widely used, recent studies have shown that NLP-based approaches are effective in identifying mental health indicators from user-generated text (Chancellor & De Choudhury, 2020). Recent advancements in natural language processing (NLP) have enabled automatic analysis of textual data for mental health detection. Traditional machine learning approaches, such as Support Vector Machines and Logistic Regression combined with bag-of-words or TF–IDF features, have been widely applied for text classification tasks (Zwerenz *et al.*, 2019). Nevertheless, these methods struggle to capture semantic and contextual meaning, particularly in low-resource languages such as Indonesian, which lack large-scale annotated datasets and standardized linguistic resources. As a result, traditional machine learning approaches often exhibit limited performance and reduced effectiveness in real-world mental health text classification (Koto *et al.*, 2020). The emergence of deep learning-based language models, particularly Bidirectional Encoder Representations from Transformers (BERT), has significantly improved contextual text understanding.

BERT-based models have demonstrated superior performance in various NLP tasks, including sentiment analysis and mental health detection, by learning bidirectional contextual representations (Devlin *et al.*, 2019). In the Indonesian language context, IndoBERT has been introduced as a pre-trained language model that achieves strong performance across multiple downstream NLP tasks (Cahyawijaya *et al.*, 2021). Several recent studies have explored the application of BERT-based and other deep learning models for mental health-related text classification, highlighting their potential effectiveness in analyzing psychological conditions from textual data (Le Glaz *et al.*, 2021). Recent studies have demonstrated that NLP-based deep learning models are effective in identifying and categorizing mental stress patterns from textual data through multi-label classification frameworks, enabling more accurate detection of complex mental health conditions (John *et al.*, 2025). In parallel, cross-platform mobile development frameworks such as Flutter have been increasingly adopted for health application development due to their efficiency, performance, and unified codebase across platforms (Huang &

Savkin, 2020). While prior research has explored Flutter-based health applications focusing on usability and system integration, the incorporation of contextual Indonesian NLP models for automatic mental health detection within mobile applications remains limited. Moreover, comparative evaluations between deep learning-based language models and classical baseline methods in this domain are still rarely reported. Recent studies have shown that deep learning-based architectures, such as graph attention networks, are effective for mental health text classification by capturing complex contextual relationships within textual data (Ahmed *et al.*, 2023). Transformer-based models, including BERT variants, have demonstrated superior performance in mental illness text classification tasks through contextualized representation learning and transfer learning approaches (Sao & Lim, 2024). Text classification techniques have been widely applied to social media data for mental health surveillance, demonstrating the feasibility of detecting mental health indicators from large-scale user-generated text (Couto *et al.*, 2025).

Despite the recent success of deep learning approaches, classical machine learning models such as Logistic Regression combined with TF–IDF features remain strong baseline methods for text classification due to their simplicity, interpretability, and competitive performance on limited datasets (Minaee *et al.*, 2022). For the Indonesian language, which is considered a low-resource language, pre-trained and transfer learning-based deep learning models have been shown to effectively capture linguistic characteristics and significantly improve performance across various downstream NLP tasks (Ekakristi *et al.*, 2025). Despite recent advances in deep learning-based NLP models, mental health text classification in low-resource languages remains challenging due to limited annotated datasets, linguistic diversity, and domain-specific variations (Pakray *et al.*, 2025). In this context, the selection of an appropriate language-specific model becomes critical, particularly for Indonesian. While multilingual BERT (mBERT) provides cross-lingual representations, it often struggles to capture language-specific nuances, informal expressions, and culturally contextualized terms commonly found in Indonesian user-generated text. IndoBERT, which is pre-trained on large-scale Indonesian corpora, is therefore better suited to model contextual representations for mental health-related text classification in this study. To address these limitations, this study proposes a mobile mental health detection system that integrates contextual Indonesian natural language processing into a Flutter-based mobile application named Alomind. The proposed system automatically classifies user-generated reflection texts into six predefined mental health categories using an IndoBERT-based contextual text classification model, while Logistic Regression combined with TF–IDF features is employed as a classical baseline for empirical performance comparison. In addition, rule-based psychological scoring is applied to questionnaire responses to generate interpretable mental health indicators and personalized recommendations.

2. Research Methodology

This study employs an applied quantitative experimental approach to design and evaluate a mobile-based mental health detection system. The primary objective is to implement a Natural Language Processing (NLP) method to analyze textual data for the early identification of mental health conditions. The system classifies user-written reflections into predefined categories, with the results presented through a mobile application. The core of the classification process utilizes IndoBERT, a pre-trained contextual language model specifically designed for the Indonesian language. Logistic Regression is used as a baseline for performance comparison. Additionally, a rule-based psychological scoring mechanism is applied to questionnaire responses, which helps generate personalized recommendations for the users. The entire system is implemented as a mobile application called Alomind, developed using the Flutter framework.

System Development Method

A prototyping-based development approach is adopted for this study, facilitating iterative design and implementation of the proposed mental health detection system. This method allows for the rapid development of system components, which are then tested, evaluated, and refined based on the results obtained. Although the stages of development are outlined in a linear sequence, the prototyping method enables continuous improvements throughout the development process, ensuring that the system evolves to meet the desired objectives.

Figure 1. Overview Proposed Research Method

Figure 1 illustrates the overall research methodology employed in this study. The figure outlines a sequential workflow, beginning with problem analysis and data collection, followed by text preprocessing, model training and classification, system implementation, and system evaluation. This framework demonstrates how the proposed mental health detection system is developed using a prototyping approach, where each stage contributes to the iterative refinement of both the NLP model and the mobile application. The stages are as follows:

1) Start
   This stage establishes the research objectives and defines the scope of the proposed mental health detection system.
2) Problem Analysis
   This phase identifies key limitations in existing mental health applications, particularly their reliance on manual self-assessments and the absence of contextual analysis in user-written emotional expressions.
3) Data Collection
   Data is gathered through Indonesian-language text reflecting mental health expressions. The dataset consists of user-generated reflections representing various emotional conditions. Labeling is based on predefined psychological indicators from established mental health literature. The labels correspond to self-reported tendencies, rather than formal clinical diagnoses, and the dataset is anonymized to protect participant privacy. The texts are categorized into six predefined labels: depression, anxiety, anger-related issues, obsessive-compulsive disorder, narcissistic personality disorder, and normal condition.
4) Text Preprocessing
   The collected data is processed to ensure consistent and reliable input for the model. Text normalization is followed by tokenization using the IndoBERT tokenizer, which converts the text into contextual token representations required by the transformer-based model.
5) Model Training & Classification
   The IndoBERT model is trained on the dataset to perform multi-class mental health text classification. The model is fine-tuned to learn contextual and semantic patterns from the reflections, and its performance is compared to Logistic Regression using metrics such as accuracy, precision, recall, and F1-score.
6) System Implementation
   After the training phase, the IndoBERT model is integrated into a mobile application, Alomind, developed with the Flutter framework. Users submit written reflections through an interactive interface, and the system processes the text in real time to display mental health analysis results along with confidence levels.
7) System Evaluation
   The final stage assesses the system's overall performance and functionality, including the classification accuracy of the integrated model and functional testing of the application. The evaluation ensures that the system meets the defined requirements and provides reliable mental health analysis, with feedback used to refine the system in subsequent prototype iterations.

Data Analysis Method

This study employs a hybrid data analysis approach that combines natural language processing (NLP) with rule-based analysis for mental health detection. IndoBERT, a pre-trained contextual language model, is used as the primary classification model due to its ability to capture contextual and semantic meanings in Indonesian-language text. Logistic Regression with TF-IDF features serves as the baseline model for performance comparison.

In addition to text-based analysis, a rule-based psychological scoring method is applied to questionnaire responses. This method generates complementary mental health assessments and personalized recommendations. By separating the analysis of structured questionnaire data from unstructured text data, the system allows both types of data to be processed independently before integration at the system level, ensuring a comprehensive approach to mental health evaluation.

System Interface and Application Features

The design of a user-centered interface is essential for the success of mobile mental health applications, as intuitive and interpretable interfaces improve user engagement, trust, and long-term system usage. This section outlines the implementation of the proposed system within a mobile application and its interactive interface. The application is designed to provide a straightforward and user-friendly experience for mental health self-reflection and analysis. The main interface offers users an overview of available features, such as text-based reflection submission and questionnaire input. Users can navigate to the reflection page to submit written emotional expressions, which are then processed by the IndoBERT-based text classification model.

The system analyzes the input and presents the classification results along with confidence information and brief recommendations. The interface is designed for clarity and accessibility, ensuring that users can easily understand the analysis outcomes without needing technical knowledge of the underlying NLP processes. Beyond text-based analysis, the application integrates questionnaire-based mental health assessments using a rule-based psychological scoring approach. Responses to structured questionnaires are processed through predefined scoring rules to generate mental health indicators such as calmness, anxiety, and stress levels, which are displayed on the main dashboard. This combination of structured questionnaire analysis and unstructured text processing provides users with a more comprehensive mental health overview while maintaining interpretability and usability within the mobile application.

Main Page Interface

Figure 2 illustrates the main page interface of the Alomind mobile application, which functions as the primary dashboard for displaying the results of questionnaire-based mental health assessments. This dashboard provides users with a snapshot of their current mental health status based on the responses submitted through structured questionnaires. The system applies a rule-based psychological scoring method to process these responses, assigning weights to each answer according to predefined psychological indicators. The aggregated results are then used to generate mental health indicators, such as calmness, anxiety, and stress levels. The dashboard presents these indicators in a summarized and easy-to-interpret format, enabling users to quickly grasp their mental health condition without needing technical expertise. By integrating rule-based scoring on the main page, the system ensures transparency and interpretability, as the results are derived from explicit scoring rules rather than opaque algorithms. This approach fosters self-awareness and allows users to track changes in their mental health over time. Moreover, the main page serves as an entry point, providing access to additional analytical features of the application, including text-based reflection and condition-specific mental health assessments. By centralizing these features in one interface, the dashboard promotes a cohesive user flow and maintains a clear distinction between questionnaire-based assessments and NLP-based analysis. This design enables users to interact with various forms of mental health evaluation within a unified interface, reinforcing the main page's role as an integrative component of the proposed system.

Figure 2. Main Page Interface

Questionnaire Page Interface

Figure 3 illustrates the interface of the questionnaire page, designed to collect structured data regarding users' mental and emotional conditions through a series of predefined questions. This page serves as the primary input component for the rule-based psychological scoring mechanism integrated into the system. Users are prompted to respond to statements that reflect emotional states, stress levels, and behavioral tendencies, using scaled response options. The questionnaire instrument is adapted from the DASS-21 indicators and clinical psychology literature, intended solely as a non-diagnostic early screening tool. Each response is processed based on well-defined psychological rules, with assigned weighting values. This enables the system to generate mental health indicators in a clear and transparent manner. The resulting scores are then summarized and displayed on the main dashboard as part of an overall mental health assessment. The use of rule-based scoring ensures that the process remains interpretable, as the generated indicators are directly derived from explicit scoring rules, rather than complex, opaque computational models. In addition to the core affective dimensions derived from DASS-21, the questionnaire incorporates supplementary indicators, such as obsessive-compulsive tendencies (OCD), narcissistic personality traits (NPD), and anger-related issues. These dimensions were included to complement the affective categories that are not fully covered by DASS-21. The inclusion of these indicators is based on established clinical psychology literature, with the aim of providing additional context for early mental health screening. However, these indicators were not intended as diagnostic criteria, and no formal clinical validation or DSM-5-based diagnosis was performed in this study.

Figure 3. Questionnaire Page Interface

Questionnaire Result Page Interface

Figure 4 illustrates the interface displaying the results of the questionnaire, which are generated through a rule-based psychological scoring mechanism applied to user responses. Each questionnaire item is processed by assigning weighted values based on predefined psychological criteria. The aggregated scores are then used to determine the dominant mental health indicator. In addition to identifying the primary condition, the interface also presents a percentage value reflecting the intensity of the detected indicator in comparison to other assessed conditions. To enhance clarity and user comprehension, the condition with the highest score is visually emphasized. It is accompanied by concise, actionable recommendations, helping users to understand their current mental state. This design approach ensures that users can interpret the results intuitively, without needing specialized knowledge in psychology or technical details. Furthermore, the rule-based nature of the scoring process promotes transparency, as the assessment outcomes are derived from clear, predefined rules rather than opaque computational models. When integrated with the IndoBERT-based text analysis module, this interface bolsters the system's role as an accessible tool for early mental health screening. Importantly, it maintains its function as a supportive resource rather than a substitute for professional clinical diagnosis.

Figure 4. Questionnaire Result Page Interface

Check Page Interface

Figure 5 illustrates the Check Page Interface, which provides access to targeted mental health screening modules focusing on specific conditions, including obsessive-compulsive disorder, narcissistic personality disorder, and anger-related issues. Each screening module is based on structured symptom-related questions and applies a rule-based scoring mechanism to evaluate the likelihood of each condition. The scoring rules are defined according to symptom severity and frequency, allowing the system to produce structured and interpretable assessment results. This interface complements the main dashboard and text-based analysis by offering condition-specific evaluations. While the Catatan page focuses on contextual emotional analysis through natural language processing, the Check page emphasizes structured psychological assessment. The integration of these features enables the system to provide a more comprehensive mental health evaluation by combining unstructured text analysis and structured rule-based screening within a single mobile application.

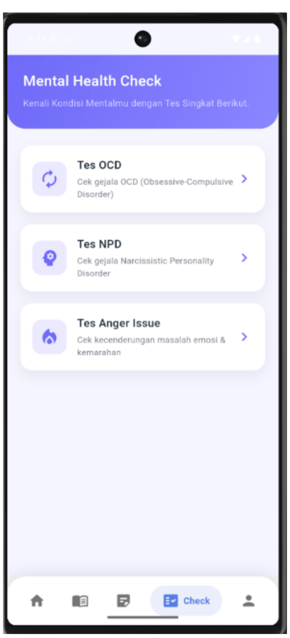

Figure 5. Check Page Interface

Catatan Page Interface

Figure 6 illustrates the Catatan (Smart Notes) page interface, designed to facilitate text-based mental health analysis using natural language processing (NLP). This page allows users to freely write daily reflections, describing their thoughts, emotions, and experiences in natural language. Unlike questionnaire-based assessments, this feature focuses on analyzing unstructured text data to capture contextual emotional expressions that may not be explicitly reflected in structured responses. User-submitted text is processed through an IndoBERT-based classification model, customized for analyzing mental health-related text in Indonesian. This model identifies semantic and contextual patterns within the text to classify it into predefined mental health categories. This approach enables the system to detect subtle emotional cues and contextual meanings often overlooked by traditional feature-based methods. The results of the analysis are then presented to users as preliminary indications of mental health, accompanied by recommendations based on the content of their reflections. This feature supports early detection through reflective writing, providing users with insights into their mental state. The text analysis utilizes the IndoBERT-based model, which converts user input into contextual representations, enabling the model to estimate the probability of various predefined mental health categories. Each probability value indicates the degree to which the text corresponds to a specific mental health condition. The category with the highest probability score is then selected and displayed as the primary mental health indication for the user.

The analysis results demonstrate that the system can differentiate between multiple mental health conditions by examining the semantic meaning and contextual relationships within the text. By adopting a probabilistic multi-class classification approach, the system can capture overlapping emotional expressions that may correspond to multiple conditions simultaneously. This approach provides more adaptive and informative outputs compared to traditional single-label classification methods, especially in cases where emotional symptoms are implicitly conveyed or overlap. On the Catatan Page interface, the analysis results are presented in a clear, user-friendly layout. The dominant mental health category is emphasized, with probability scores for other relevant categories displayed alongside. This design enhances the transparency of the analytical process, allowing users to easily observe the confidence levels associated with each potential condition, thereby increasing user trust in the system's output. The effectiveness of this feature is largely attributed to the BERT-based architecture, which models bidirectional contextual information within textual data. This capability is crucial for Indonesian-language text, where emotional meanings are often conveyed indirectly or through contextual nuances. By utilizing contextual embeddings, the model can detect implicit emotional cues and linguistic patterns that are often missed by conventional machine learning methods relying on surface-level features. Beyond providing analytical results, the Catatan Page fosters self-reflection and mental health awareness. By integrating qualitative mental health indications with quantitative probability scores, the system encourages users to reflect on their emotional state and consider potential psychological concerns. Rather than acting as a definitive diagnostic tool, this feature is designed to support early mental health screening and responsible self-assessment. In this way, the Catatan Page bridges automated text analysis with user comprehension, aiding informed decision-making within the mobile mental health detection system.

Figure 6. Catatan Page Interface

Dataset and Data Acquisition

The dataset used in this study is a labeled Indonesian text corpus, stored in *dataset_curhat.csv*. This dataset was compiled from non-clinical sources to support early mental health screening research. It consists of 3,172 text samples in Indonesian, representing user reflections related to mental and emotional conditions. Data were collected via voluntarily distributed questionnaires among university students, ensuring anonymity and the absence of personally identifiable information. The labels for this dataset were obtained through self-report questionnaires and a rule-based psychological scoring system. These labels are intended as early screening indicators rather than formal clinical diagnoses. The mental health categories used in this study were defined based on established psychological literature and commonly recognized emotional indicators. Each text sample was manually labeled into one of six categories: Depression, Anxiety, Anger Issues, Obsessive Compulsive Disorder (OCD), Narcissistic Personality Disorder (NPD), and Normal Condition. It is important to note that this dataset does not contain clinical diagnostic data and was exclusively used for research purposes. For experimental evaluation, the dataset was split into training and testing sets, with 2,537 samples allocated for training and 635 samples used for testing, following an 80:20 ratio.

Data Processing

The data processing stage ensures that the collected Indonesian text data is prepared for analysis and model training. This stage involves text cleaning and normalization to remove noise and ensure consistency, followed by tokenization using the IndoBERT tokenizer. Tokenization converts the cleaned text into contextual token representations, enabling effective learning during classification. For the baseline model, TF–IDF (Term Frequency-Inverse Document Frequency) feature extraction is applied to represent the text numerically. This method helps quantify the text data for comparison with the IndoBERT-based model. Once processed, the dataset is divided into training and testing sets to facilitate model evaluation.

NLP-Based Text Analysis Method

The proposed system employs a Natural Language Processing (NLP) approach to analyze user-generated reflections related to mental health conditions. NLP allows unstructured Indonesian text to be transformed into meaningful data that can be processed computationally for classification. This method is particularly valuable for detecting emotional expressions and contextual clues embedded within user-written narratives. The primary NLP model used in this study is IndoBERT, a deep learning-based contextual language model pre-trained on large-scale Indonesian text datasets. IndoBERT utilizes a bidirectional transformer architecture, enabling the model to understand word meanings based on their surrounding context, rather than treating words as isolated tokens. This capability is especially crucial in mental health text analysis, where emotional cues are often subtle and conveyed through complex sentence structures. Recent studies have shown that transformer-based deep learning models, like IndoBERT, are highly effective in capturing semantic and emotional patterns in text, especially when compared to traditional machine learning methods (Devlin *et al.*, 2019; Cahyawijaya *et al.*, 2021). For baseline comparison, a traditional machine learning approach using Logistic Regression with TF–IDF features is also employed. Unlike IndoBERT, this method relies on surface-level lexical features and does not account for deep contextual meaning. The comparison between IndoBERT and Logistic Regression underscores the superiority of contextual NLP models in handling nuanced emotional expressions in Indonesian mental health texts, highlighting the advantages of using deep learning for this type of analysis.

Table 1. Hyperparameter Settings for IndoBERT and Logistic Regression

Parameter	IndoBERT	LogisticRegression
Optimizer	AdamW	-
Learning Rate	2e-5	-
Batch Size	16	-
Epochs	3	-
Feature Representations	-	TF-IDF
Solver	-	liblinear

Model Evaluation

The IndoBERT model was fine-tuned for three epochs, as performance improvements plateaued after the third epoch based on validation metrics. This configuration was chosen to balance model performance with computational efficiency, while mitigating the risk of overfitting, particularly given the relatively limited dataset. To evaluate the model's performance, standard classification metrics, including accuracy, precision, recall, and F1-score, were used.

These metrics are commonly applied to assess the effectiveness of machine learning classifiers, with results derived from the confusion matrix (Sujon *et al.*, 2025). The evaluation was conducted using the testing dataset, and the results show that IndoBERT consistently outperformed the Logistic Regression baseline across all metrics. Logistic Regression was employed purely as a comparative baseline to highlight the performance differences between traditional machine learning and deep learning models, rather than as a candidate for deployment in the proposed system. Specifically, IndoBERT achieved an accuracy of 97.79%, while Logistic Regression achieved an accuracy of 94.17%. A summary of the comparative performance is presented in Table 2.

3. Results and Discussion

Results

Experimental Results

This section presents the experimental results derived from the evaluation of the proposed mental health detection system. The performance of the IndoBERT-based classification model was compared against Logistic Regression, which served as the baseline model. Both models were trained and evaluated on the same dataset to ensure consistency and comparability of the results. The evaluation utilized standard multi-class classification metrics, including accuracy, precision, recall, and F1-score. These metrics were selected to offer a comprehensive assessment of model performance, measuring both overall classification accuracy and the quality of class-level predictions. To account for class distribution within the dataset, all evaluation metrics were computed using weighted averaging.

Table 2. Performance Comparison of Classification Models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.9417	0.9427	0.9417	0.9418
IndoBERT	0.9779	0.9783	0.9780	0.9780

Table 2 presents a quantitative comparison of the classification performance between the IndoBERT-based model and the Logistic Regression baseline across all evaluation metrics. The results indicate that the IndoBERT model outperformed the baseline in accuracy, precision, recall, and F1-score. The performance difference was consistently observed across all reported metrics, highlighting the stable and superior classification behavior of the IndoBERT-based model within the experimental setting.

Discussion

The experimental results demonstrate the effectiveness of integrating contextual natural language processing (NLP) techniques into a mobile-based mental health detection system. The superior performance of the IndoBERT-based model indicates that deep learning approaches provide a more robust framework for analyzing unstructured mental health text compared to traditional machine learning methods. This improvement is primarily attributed to IndoBERT's capacity to capture contextual and semantic relationships within Indonesian text, which are crucial for accurately understanding emotional expressions. The findings align with recent studies that highlight the superior performance of BERT-based models over classical machine learning methods in mental health text classification, especially for languages with complex contextual structures like Indonesian (Devlin *et al.*, 2019; Sujon *et al.*, 2025). In contrast to Logistic Regression, which relies on surface-level lexical features, IndoBERT's contextual representations enable a more accurate interpretation of nuanced emotional content. Moreover, the integration of IndoBERT-based text classification with a rule-based psychological scoring mechanism enhances the analytical capability of the proposed system. While the NLP model focuses on extracting implicit emotional patterns from unstructured user reflections, the rule-based approach generates structured, transparent, and interpretable assessments based on questionnaire responses. This hybrid approach allows the system to handle both unstructured and structured data sources, resulting in a more comprehensive and reliable mental health overview for users. Despite the promising results, it is important to note that this study uses non-clinical data aimed at supporting early mental health screening, rather than providing clinical diagnoses.

Model Performance Comparison

The performance comparison between the IndoBERT-based model and the Logistic Regression baseline reveals a clear distinction in classification capability. According to the evaluation results, the IndoBERT model outperformed Logistic Regression in accuracy, precision, recall, and F1-score. This underscores that contextual deep learning models, like IndoBERT, are more effective in capturing the complexity of mental health-related textual expressions than traditional machine learning methods that rely on handcrafted features. Although Logistic Regression combined with TF–IDF features showed strong baseline performance, its effectiveness is constrained by its reliance on surface-level textual representations. IndoBERT, on the other hand, leverages contextual embeddings that allow the model to understand word meanings based on their surrounding context. This is particularly significant in mental health text classification, where emotional expressions are often implicit, ambiguous, and context-dependent. The observed performance improvement highlights that IndoBERT is better suited for multi-class mental health classification tasks, especially for the Indonesian language. Its ability to model bidirectional context enables the model to distinguish between closely related emotional conditions, reducing misclassification caused by overlapping linguistic patterns commonly found in mental health narratives.

Confusion Matrix Analysis

Further analysis of the IndoBERT-based classification model's performance was conducted using a confusion matrix, evaluated on the test dataset with an 80:20 train-test split.

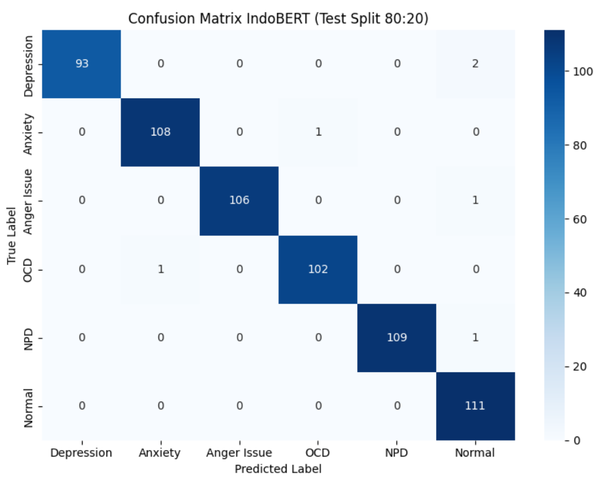

Figure 7. Confusion Matrix of IndoBERT-Based Mental Health Classification (Test Split 80:20)

Confusion Matrix Analysis

The confusion matrix was generated using 635 test samples from the 80:20 train–test split. The matrix reveals strong diagonal dominance across all six mental health categories, indicating robust classification performance. Most predictions fall along the diagonal, reflecting correct classifications for the majority of samples. However, some minor misclassifications were observed between the depression and normal categories, as well as between anxiety and OCD, which reflects the overlapping emotional expressions commonly found in self-reported mental health narratives. These results demonstrate the IndoBERT model's capability to capture contextual emotional cues, while also highlighting the inherent challenge of distinguishing closely related psychological conditions.

Error Analysis and Misclassification

Although the IndoBERT-based model exhibited strong overall performance, several misclassification cases were identified during the evaluation. These errors predominantly occurred in text samples where emotional expressions overlapped, particularly between depression and anxiety categories. Such overlap is frequent in mental health narratives, where individuals may express mixed emotions, such as sadness, worry, and fatigue, within a single reflection. This complexity makes strict category separation challenging, even for contextual models. Additionally, misclassification was influenced by the informal language, slang, and abbreviated expressions often present in user-generated text. While IndoBERT is proficient in capturing contextual semantics, ambiguous sentence structures and short, reflective texts with limited contextual cues can reduce classification accuracy. These findings suggest that while contextual language models significantly enhance performance, challenges persist in handling subtle and implicit emotional expressions. In comparison, the Logistic Regression baseline exhibited a higher tendency to misclassify texts containing complex emotional content. This limitation arises from its reliance on surface-level lexical features, which are less effective in capturing the nuanced semantics found in mental health-related texts. Overall, the error analysis underscores the importance of contextual understanding in mental health text classification and highlights areas for future improvement.

Rule-based Psychological Scoring Discussion

In addition to the text-based analysis, this study integrates a rule-based psychological scoring mechanism to process structured questionnaire responses. This method complements the NLP-based model by providing transparent and interpretable assessments derived from predefined psychological indicators. Each questionnaire response is assigned a weight according to explicit scoring rules, enabling the system to generate easily understandable mental health indicators. A significant advantage of the rule-based approach is its interpretability. Unlike deep learning models, which often function as black-box systems, rule-based scoring allows users to trace the assessment outcomes directly back to their questionnaire responses. This transparency enhances user trust and aligns with established psychological assessment principles. Such clarity is crucial for early mental health screening systems, as it ensures that results support self-awareness and are not mistaken for clinical diagnoses (Guidotti *et al.*, 2019). Rather than replacing the NLP-based classification, the rule-based psychological scoring serves as a complementary analytical component within the proposed system. While the IndoBERT model extracts nuanced emotional expressions from unstructured user-generated text, the questionnaire-based scoring adds structured insights that reinforce the overall mental health evaluation. This hybrid approach allows the system to integrate both unstructured and structured data sources, improving the system's reliability, interpretability, and practical usability in the context of mobile mental health applications.

Ethical Considerations and User Safety

Given the sensitive nature of mental health data, ethical considerations are a central aspect of the proposed system. All textual data used in this study were anonymized, and no personally identifiable information was included in the dataset. The system is designed to process user inputs strictly for analytical purposes, ensuring that sensitive personal information is not stored beyond what is required for system functionality. It is crucial to emphasize that the proposed system is not intended to replace professional mental health diagnosis or treatment. The analysis results generated by the system are meant to support early mental health screening and self-awareness, rather than provide definitive clinical conclusions. Therefore, the application encourages users to seek professional consultation when experiencing severe or persistent mental health concerns. By presenting analysis results transparently and avoiding deterministic diagnostic claims, the system prioritizes user safety and responsible use. This ethical approach ensures that the application serves as a supportive tool, not a substitute for mental health professionals, aligning with best practices in digital mental health research.

Limitations and Future Work

Despite the promising results, this study has several limitations that should be addressed in future research. First, the dataset used in this study, consisting of 3,172 labeled text samples collected from non-clinical participants, may limit the generalizability of the model to broader populations, particularly those diagnosed with clinical mental health conditions. Expanding the dataset to include more diverse and representative samples would help improve the robustness of the classification model and reduce potential biases. Second, the dataset labels were assigned based on predefined mental health categories without clinical validation. While this approach is appropriate for early mental health screening applications, collaboration with mental health professionals could enhance label accuracy and increase the clinical relevance of the dataset. Additionally, the system has not yet undergone real-world user studies to assess usability, long-term engagement, and user perceptions of the feedback provided by the system. Future work could explore the integration of additional mental health categories, the incorporation of multimodal data such as behavioral patterns or interaction logs, and further optimization of model performance. Longitudinal evaluations and clinical validation studies are also necessary to provide deeper insights into the system's effectiveness as a responsible digital mental health support tool. Furthermore, ethical considerations, data bias mitigation, and the responsible use of artificial intelligence remain critical, especially when applying automated analysis to sensitive psychological data (Meadi *et al.*, 2025).

4. Conclusion

This study presents the development and implementation of a mobile-based mental health detection system that integrates contextual natural language processing (NLP) with interactive application features. The system, named Alomind, was implemented as a Flutter-based mobile application, developed using a prototyping approach to facilitate iterative improvement and a user-centered design. By combining text-based analysis and questionnaire-based assessments, the system provides a practical and accessible platform for early mental health screening. For text analysis, IndoBERT was employed as the primary deep learning model due to its capacity to capture contextual and semantic meanings in Indonesian-language text. The results demonstrate that contextual language models, such as IndoBERT, are more effective than classical machine learning approaches in analyzing nuanced emotional expressions within Indonesian mental health-related text. In addition to text classification, a rule-based psychological scoring mechanism was applied to structured questionnaire responses to generate interpretable mental health indicators and personalized recommendations. The integration of unstructured text analysis with structured questionnaire processing enables the system to offer comprehensive and transparent mental health feedback through a unified mobile interface. Overall, the proposed system highlights the feasibility of integrating advanced Indonesian NLP models into a mobile mental health application. While the system is designed for early mental health screening rather than clinical diagnosis, future research may focus on expanding the dataset, incorporating additional mental health categories, and conducting clinical validation to further enhance the system's accuracy and real-world applicability.

5. Acknowledgments

The authors would like to express their sincere gratitude to all those who supported this research, including academic supervisors and contributors who provided valuable feedback throughout the development of this study.

6. References

Cahyawijaya, S., *et al.* (2021). IndoNLG: Benchmark and resources for evaluating Indonesian natural language generation. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing*, 8875–8898. https://doi.org/10.18653/v1/2021.emnlp-main.699.

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. *npj Digital Medicine*, 3(1). https://doi.org/10.1038/s41746-020-0233-7.

Couto, M., Perez, A., Parapar, J., & Losada, D. E. (2025). Temporal word embeddings for early detection of psychological disorders on social media. *Journal of Healthcare Informatics Research*. https://doi.org/10.1007/s41666-025-00186-9

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 Conference on North American Chapter of the Association for Computational Linguistics - Human Language Technologies - Proceedings*, 4171–4186.

Endriyani, S., & Susanti, E. (2024). Android-based application for depression, anxiety, and stress screening at Poltekkes Kemenkes Palembang, South Sumatra Province, Indonesia. *Journal of Health Informatics*, 17(4), 1486–1492.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5). https://doi.org/10.1145/3236009.

Huang, H., & Savkin, A. V. (2020). Autonomous navigation of a solar-powered UAV for secure communication in urban environments with eavesdropping avoidance. *Future Internet*, 12(10), 1–14. https://doi.org/10.3390/fi12100170

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. *COLING 2020 - 28th International Conference on Computational Linguistics - Proceedings*, 757–770. https://doi.org/10.18653/v1/2020.coling-main.66

Le Glaz, A., *et al.* (2021). Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, 23(5). https://doi.org/10.2196/15708.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2022). Deep learning-based text classification. *ACM Computing Surveys*, 54(3). https://doi.org/10.1145/3439726

Pakray, P., Gelbukh, A., & Bandyopadhyay, S. (2025). Natural language processing applications for low-resource languages. *Natural Language Processing Journal*, 31(2), 183–197. https://doi.org/10.1017/nlp.2024.33.

Scherbakov, D. A., Hubig, N. C., Lenert, L. A., Alekseyenko, A. V., & Obeid, J. S. (2025). Natural language processing and social determinants of health in mental health research: AI-assisted scoping review. *JMIR Mental Health*, 12, 1–15. https://doi.org/10.2196/67192.

Shaw, C., LaCasse, P., & Champagne, L. (2025). Exploring emotion classification of Indonesian tweets using large-scale transfer learning via IndoBERT. *Social Network Analysis and Mining*, 15(1), 1–12. https://doi.org/10.1007/s13278-025-01439-6.

Wolf, T., *et al.* (2020). Transformers: State-of-the-art natural language processing. *Transformers for NLP*, 38–45.

Yardley, L., *et al.* (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5), 833–842. https://doi.org/10.1016/j.amepre.2016.06.015