OPEN ACCESS
Indo-JC Indonesia Journal of Computing
ISSN 2460-9056
journals.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 10, Issue. 1, August 2025. pp. 88-109
doi:10.21108/indojc.v10i1.9710

Emotion Recognition from Text and Gesture Generation for an Early Marriage Counseling Chatbot in Lombok Using BERT

Adam Zahran Ramadhan[1], Rifki Wijaya[2], Shaufiah[3]

School of Computing, Telkom University
Telecommunication Street No. 1, Bandung 40257, Indonesia

[1]adamzahranr@student.telkomuniversity.ac.id
[2] rifkiwijaya@telkomuniversity.ac.id
[3] shaufiah@telkomuniversity.ac.id

Abstract

Early marriage remains a pressing issue among adolescents in Lombok, Indonesia, influenced by cultural norms, educational barriers, and economic challenges. This study develops an emotion classification and reason identification framework for a virtual counseling chatbot to support prevention efforts. Five functional emotion categories 'Enthusiastic', 'Gentle', 'Analytical', 'Inspirational', and 'Cautionary' were defined to capture counseling tones. The system leverages IndoBERT with a two-phase fine-tuning strategy. Phase 1 used a balanced dataset of 2,000 samples and achieved a macro F1-score of 0.95, while Phase 2 refined the model using 10,000 imbalanced pseudo-labeled samples, yielding a macro F1-score of 0.88 and improved sensitivity to minority classes. In addition, a semantic similarity-based reason identification module was implemented to classify user inputs into Education, Economy, Religion, or Culture categories, enhancing context awareness beyond simple keyword matching. Performance evaluation employed accuracy, precision, recall, and F1-score, supported by confusion matrices and training plots for generalization analysis. A descriptive emotion-to-gesture mapping was also designed to link each emotion category with static body pose visualizations, providing a conceptual basis for future multimodal applications.

Keywords: Early Marriage, Emotion Classification, Gesture Mapping, IndoBERT, NLP, Virtual Chatbot

Abstrak

Pernikahan dini masih menjadi masalah mendesak di kalangan remaja di Lombok, Indonesia, yang dipengaruhi oleh norma budaya, hambatan pendidikan, dan tekanan ekonomi. Penelitian ini mengembangkan kerangka klasifikasi emosi dan identifikasi alasan untuk chatbot konseling virtual guna mendukung upaya pencegahan. Lima kategori emosi fungsional 'Antusias', 'Lembut', 'Analitis', 'Inspiratif', dan 'Peringatan' digunakan untuk mencerminkan nada komunikasi konseling. Sistem ini memanfaatkan IndoBERT dengan strategi fine-tuning dua tahap. Tahap pertama menggunakan 2.000 sampel seimbang dan mencapai macro F1-score 0,95, sedangkan tahap kedua menggunakan 10.000 sampel pseudo-labeled yang tidak seimbang dengan macro F1-score 0,88, serta meningkatkan sensitivitas pada kelas minoritas. Selain itu, modul identifikasi alasan berbasis kesamaan semantik dikembangkan untuk mengklasifikasikan masukan pengguna ke dalam kategori Pendidikan, Ekonomi, Agama, atau Budaya, sehingga lebih kontekstual dibanding pencocokan kata kunci sederhana. Evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score yang diperkuat dengan confusion matrix dan grafik pelatihan. Pemetaan emosi-ke-gestur secara deskriptif juga dirancang untuk menghubungkan tiap kategori emosi dengan visualisasi pose tubuh statis sebagai dasar pengembangan aplikasi multimodal di masa depan.

Keywords: Chatbot Virtual, IndoBERT, Klasifikasi Emosi, NLP, Pemetaan Gestur, Pernikahan Dini

Received on January, 2025. Accepted on August, 2025

I. Introduction

EARLY marriage continues to pose a significant societal concern in Indonesia, particularly in Lombok, West Nusa Tenggara (NTB), where socio-cultural norms exert strong influence. The consequences include intergenerational poverty, limited educational attainment, and heightened exposure to physical and psychological health risks. Despite ongoing efforts by governmental and non-governmental bodies, early marriage rates remain high [1]. Recent statistics from the Central Bureau of Statistics (2022) identify NTB as having the highest national percentage 16.23% of women aged 20–24 who entered marriage or cohabitation before age 18 [2].

Advancements in digital technology, particularly in Natural Language Processing (NLP) have paved the way for new forms of social assistance. One such innovation is the development of virtual chatbot systems [3], [4], which simulate human dialogue and have found applications in emotional support, education, and health services. The constant availability and anonymous nature of chatbots make them highly effective for addressing sensitive issues like early marriage.

This study presents the development of an NLP-driven counseling chatbot that integrates emotion recognition and reason identification into its dialogue system. IndoBERT, a pre-trained transformer model for Bahasa Indonesia, functions as the core classification engine and as the embedding model for semantic similarity-based reason detection. Initial training was conducted using 2,000 manually labeled synthetic dialogues generated by a Large Language Model (LLM), designed to reflect typical exchanges between counselors and youth. To enhance contextual flexibility and improve emotion inference, an additional 10,000 unlabeled samples were incorporated during a second training phase.

A central contribution is the introduction of function-oriented emotional categories such as ‘Enthusiastic’, ‘Gentle’, ‘Analytical’, ‘Inspirational’, and ‘Cautionary’ which better capture the communicative intent of counseling conversations than conventional emotion labels, together with a semantic similarity-based reason identification module that classifies user inputs into the categories of Education, Economy, Religion, or Culture. Unlike conventional keyword-matching approaches, the module leverages IndoBERT-based sentence embeddings to improve generalization and context awareness. These two components work in tandem to provide a richer understanding of user context, enabling the chatbot to deliver responses that are both emotionally sensitive and contextually informed. Additionally, the system introduces a gesture-mapping framework that associates each emotional category with representative nonverbal expressions, reinforcing the connection between verbal and physical communication. Leveraging IndoBERT’s contextual modeling capabilities, the framework is able to detect subtle emotional nuances while simultaneously identifying semantic relationships between user inputs and predefined reason categories, laying the foundation for a more empathetic and context-aware virtual counseling agent for early marriage prevention [5], [6], [7].

Following the increasing effectiveness of Transformer-based architectures, recent research has validated the application of BERT and its derivatives for emotion and text classification tasks. In multi-class classification settings, the integration of BERT with active learning techniques has resulted in an F1-score of 0.83, while concurrently reducing annotation costs by up to 85%, underscoring its efficiency in resource-constrained environments [8]. In multilingual emotion recognition, architectures such as RoBERTa-MA and XLNet-MA have demonstrated accuracy rates of 62.4% and 60.5%, respectively, reflecting the contributions of multi-attention mechanisms to emotional inference from text [9].

For Bahasa Indonesia, IndoBERT pretrained on domain-relevant corpora has emerged as a robust baseline. When combined with deep learning structures such as BiLSTM, BiGRU, and attention layers, IndoBERT-based models have achieved up to 91% accuracy and 78% on benchmarks such as IndoNLU, confirming their capacity to detect nuanced emotional content in Indonesian text [10]. Parallel approaches employing CNN and BiGRU hybrids have reported F1-scores exceeding 80% on general emotion datasets like ISEAR and WASSA,

demonstrating the continued relevance of deep learning under tailored configurations [11]. Additionally, emerging methods that incorporate large language models with pseudo-labeling, such as ChatGPT-supported few-shot annotation using CamemBERT have achieved promising F1-scores up to 0.6662 in multilingual emotion classification tasks [12].

While prior studies have made considerable progress in emotion classification and facial expression synthesis, the translation of functional text-based emotions into embodied gestures remains largely underexplored particularly in culturally sensitive and low-resource environments. Addressing this gap, the present study proposes a novel framework that utilizes IndoBERT-derived emotional outputs to drive synchronized body and hand gestures, grounded in principles of nonverbal communication and attuned to local socio-cultural dynamics.

This study aims to investigate whether a multi-phase fine-tuning strategy combined with class balancing techniques can improve emotion recognition performance, particularly for minority emotion categories critical to early marriage prevention. By aligning textual emotion recognition with expressive gestural responses, this approach introduces a multimodal layer of interaction that enhances the emotional authenticity of virtual counseling agents. Rather than treating gesture generation as a cosmetic addition, the framework positions it as a core communicative feature. In doing so, it advances the design of emotionally intelligent systems that not only understand but also embody emotion crucial for fostering empathy, trust, and connection in sensitive counseling dialogues.

II. Literature Review

Developing a chatbot for early marriage counseling requires an interdisciplinary approach that integrates emotion recognition and nonverbal communication, particularly in linguistically and culturally sensitive environments. While prior studies have leveraged Transformer-based architectures such as BERT for text-based emotion classification across languages including Indonesian, most have prioritized generic emotional taxonomies and facial expressions, often neglecting the communicative intent underlying emotional states in counseling contexts. This study addresses those limitations by introducing a function-oriented emotion classification framework tailored to the dialogic functions of empathy, encouragement, and caution, with direct mappings to culturally informed body gestures. By extending emotion understanding beyond static text or facial animation, the system enables contextually grounded, multimodal virtual counseling in Bahasa Indonesia, contributing to more emotionally responsive and socially attuned human–AI interaction.

A. Emotion Classification in Text-Based Counseling

Emotion classification in textual data has gained relevance across multiple domains, particularly in applications such as counseling, where recognizing affective states is central to effective interaction. In the Indonesian language context, prior studies have employed various machine learning strategies and feature extraction techniques to classify emotion in general-purpose text.

One study focused on Indonesian tweets related to the 2024 general election, employing a six-category emotion framework: joy, love, surprise, anger, fear, and sadness. Feature extraction was conducted using Term Frequency–Inverse Document Frequency (TF-IDF) and Bag-of-Words (BoW), with classification via the K-Nearest Neighbors (KNN) algorithm. Results demonstrated that TF-IDF outperformed BoW, yielding 58% accuracy in multi-class classification and 79% accuracy in binary classification using an 80:20 data split [13].

Another study introduced a deep learning framework based on a Convolutional Neural Network (CNN), optimized through Hyperband Tuning. The model evaluated several input encoding methods such as CountVectorizer, TF-IDF, and Keras Tokenizer prior to training. In comparison with conventional classifiers such as Decision Tree, Naive Bayes, and Boosting SVM, the CNN-based approach achieved the highest performance, recording an F1-score of 71.00% [14].

While these approaches indicate substantial progress in emotion classification for Indonesian text, their focus has largely been confined to general content domains such as social media. Counseling dialogues present additional challenges, including implicit affective cues, evolving tone, and speaker-specific language styles. As such, future research must emphasize domain adaptation and model refinement, particularly through transformer-based architectures capable of handling emotionally layered and context-sensitive language in therapeutic interactions.

B. Emotionally Aware Language Processing with IndoBERT

Emotionally aware language processing requires a deep understanding of how affective meaning is conveyed through linguistic context. BERT (Bidirectional Encoder Representations from Transformers) marked a pivotal advancement in this domain by introducing bidirectional context modeling via transformer-based pretraining [15]. Despite its success across a broad range of natural language processing tasks, BERT's reliance on English and multilingual training corpora limits its adaptability to languages with distinct grammatical and cultural structures, such as Bahasa Indonesia.

To address this limitation, IndoBERT was introduced as a language-specific adaptation of the original BERT architecture, trained on extensive Indonesian text corpora [16], [17]. This adaptation enhances the model's ability to process Indonesian-specific syntax, morphology, and contextual nuance, thereby improving its performance on tasks requiring emotional interpretation. Empirical studies have demonstrated IndoBERT's superiority over multilingual alternatives in recognizing emotions in Indonesian texts, including informal and conversational data such as counseling interactions [18].

In the present study, IndoBERT is employed as the foundational model not only for classifying emotional tones in early marriage counseling dialogues but also as the embedding model for semantic similarity-based reason identification. Its context-sensitive embeddings facilitate the accurate identification of affective categories such as Lembut (Gentle), Peringatan (Cautionary), and Inspiratif (Inspirational), all of which are critical for designing virtual agents capable of delivering culturally attuned and empathetic responses in sensitive counseling environments [6].

C. Synthetic Data Generation with Large Language Models (LLMs)

In domains such as counseling, access to labeled conversational datasets is restricted due to ethical, legal, and privacy concerns, particularly when dealing with culturally sensitive issues like early marriage. To address this, recent research has examined synthetic data generation using large language models (LLMs) as a practical and ethical alternative [19]. This method enables the creation of high-quality, diverse, and emotionally rich datasets without compromising user confidentiality.

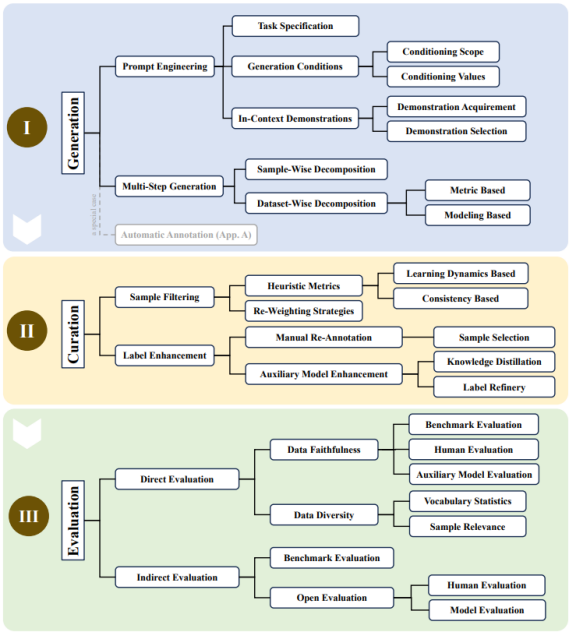

Fig. 1. A classification of synthetic data generated by LLMs [19].

As illustrated in Fig 1, the synthetic data generation framework adopted in this study is structured into three interdependent phases: Generation, Curation, and Evaluation. This taxonomy supports the systematic development of high-quality, emotionally resonant dialogue datasets for NLP applications. During the Generation phase, prompt engineering and in-context demonstrations are used to guide large language models (LLMs) in producing outputs aligned with targeted emotional tones and conversational intents. To enhance contextual depth and lexical variety, multi-step generation strategies both at the sample and dataset level sare employed, enabling richer and more nuanced outputs.

In the Curation phase, the generated samples are refined through a combination of automated and manual techniques. Heuristic and consistency-based filtering methods are applied to exclude low-quality or ambiguous entries. Additionally, sample re-weighting ensures balanced label distributions, while label quality is improved through manual annotation or verification via auxiliary models. This phase is critical to aligning the synthetic outputs with the emotional and contextual fidelity required for classification tasks. The Evaluation phase focuses on validating the quality and utility of the curated dataset. Direct evaluations assess fidelity, diversity, and label accuracy using a combination of benchmark metrics and expert human judgment. Indirect evaluations are also conducted by examining downstream model performance when trained on the synthetic data. Together, these validation strategies establish the dataset's reliability and its applicability to emotion-aware language processing systems.

Empirical studies have shown that when guided by structured prompts and informed by context-specific emotional taxonomies, LLM-generated data can function as effective training material in low-resource environments [19], [20]. Typically, synthetic data creation follows a structured multi-stage pipeline, consisting of generation, curation, and evaluation phases. Techniques such as emotion-guided generation and heuristic filtering are used to ensure emotional authenticity, linguistic coherence, and thematic relevance. Moreover, this framework allows precise control over class balance and contextual variability key factors in training robust emotion classification models.

Recent findings further validate the efficacy of synthetic data. Benchmarks show that LLM-generated datasets can match or surpass real-world data in classification accuracy, particularly when followed by domain-specific

fine-tuning [21]. These outcomes support the growing consensus that synthetic generation is not a provisional substitute, but a scalable methodology capable of improving model generalizability and performance in emotionally complex NLP tasks such as those found in counseling systems.

D. Gesture Mapping Based on Emotional Label

Emotion recognition systems in natural language processing have traditionally relied on textual input as the primary source for detecting affective states. However, emerging research emphasizes the critical role of non-verbal modalities such as facial expressions and body gestures in conveying emotional meaning. These cues often precede or reinforce verbal communication and are central to achieving empathy, clarity, and emotional resonance in human interaction [22].

Gesture mapping, which refers to the systematic alignment between functional emotional labels and physical expressions, is a significant advancement in the development of embodied conversational agents. This concept is based on multimodal communication theory [23], which asserts that emotions are conveyed not only through words but also through a combination of verbal and non-verbal behaviors. As a design strategy, gesture mapping enables virtual agents to display behaviors that are socially coherent and culturally appropriate, thereby enhancing the perceived authenticity and emotional intelligence of these systems.

Recent empirical findings indicate that specific gestures strongly influence how emotions are interpreted. For example, expressions of happiness are often conveyed through a genuine, broad smile, an upright and open posture, and energetic, animated movements, all of which signal joy and contentment. In contrast, sadness is typically reflected in a slouched or drooping posture, slow and subdued movements, a lowered head, and avoidance of eye contact, which together communicate a sense of sorrow or melancholy [24]. These non-verbal cues not only express affective states but also play a crucial role in regulating and modulating emotions during social interaction.

Gestures play a key role in modulating emotional intensity, shaping how feelings are conveyed and perceived during interactions. Subtle changes in hand movement, posture, gaze, or facial tension can either amplify or soften verbal messages, making emotional communication more intuitive and context-aware [24].This is especially crucial in domains like mental health or early marriage counseling, where trust and empathy are essential. By integrating gesture mapping, virtual agents can express emotions more authentically, enhancing user engagement and emotional resonance [25].

E. Reason Identification in Counseling Context

Identifying underlying reasons for early marriage is essential in counseling systems because it provides counselors and automated agents with deeper contextual understanding beyond surface-level emotional tone. A systematic review and empirical studies in Indonesia have consistently shown that education, economy, religion, and culture are major structural drivers of early marriage decisions, and these categories are widely used in public health and sociological interventions [26], [27], [28].

In natural language processing (NLP), early approaches to reason identification often relied on rule-based keyword matching, which is interpretable but fails to generalize across linguistic variations and implicit expressions. For instance, keyword-based methods might fail to detect economic hardship if a user says “I had to stop school to support my family” because the word “economy” is not explicitly mentioned. Recent research therefore advocates for semantic similarity frameworks using contextual embeddings [7], [29], which encode entire sentences into high-dimensional vectors and allow cosine similarity-based matching with predefined categories. These approaches are particularly effective in low-resource and culturally diverse settings because they capture meaning rather than just exact word matches.

In this study, we leverage IndoBERT embeddings to perform reason identification via cosine similarity, offering context-aware categorization without additional supervised annotation. This module complements the emotion classification pipeline by providing structured insights into the underlying motivations of user inputs, enabling more adaptive, empathetic, and targeted counseling responses.

III. Research Method

A. Emotion and Reason Classification in Text-Based Counseling

The proposed research framework aims to develop an emotion classification model and a semantic similarity-based reason identification module to support emotionally responsive and context-aware virtual counseling in early marriage scenarios. The system development comprises a series of sequential stages: data generation, annotation, model training, reason identification, and emotion-to-gesture mapping, as illustrated in Fig 2.

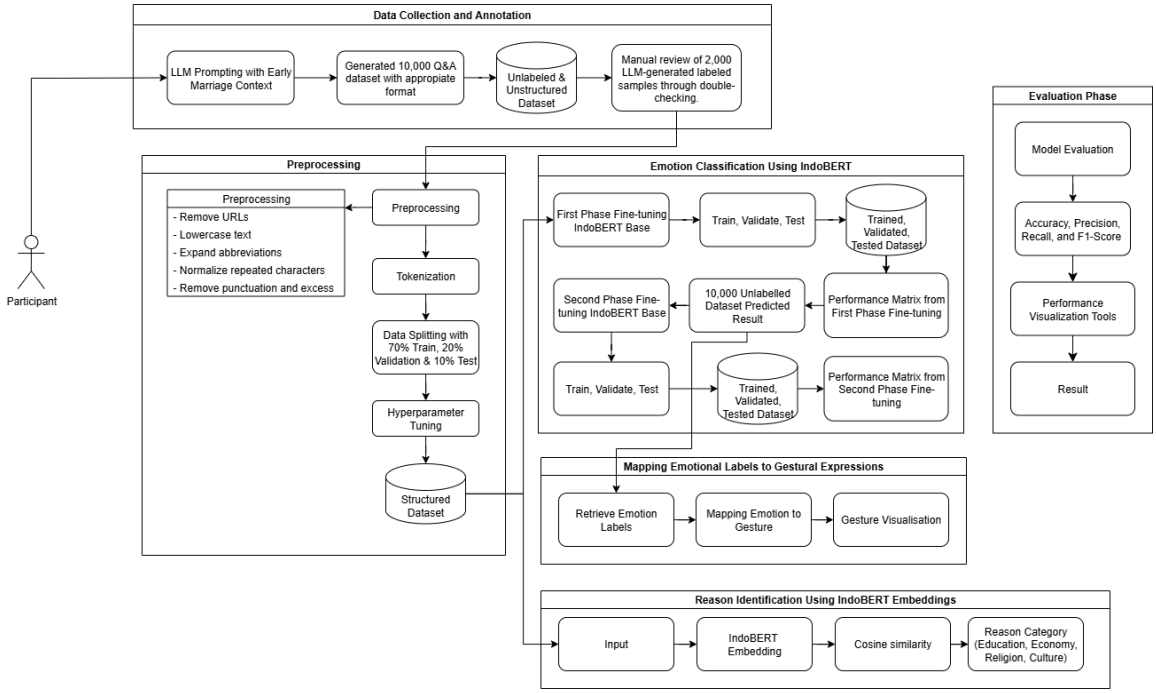

Fig. 2. Architecture of the Proposed Model

To construct the dataset, a Large Language Model (LLM) generates 10,000 synthetic question-answer pairs designed to simulate realistic counseling exchanges. From this dataset, 2,000 samples undergo a double-checking manual review process to ensure alignment with a predefined emotional taxonomy tailored to the counseling domain. These refined samples are then used for initial model fine-tuning. This same dataset also serves as input for the reason identification module, which later maps user inputs to one of four structural factors associated with early marriage: education, economy, religion, and culture.

Before training, the data is subjected to text normalization and tokenization, and subsequently split into training, validation, and test sets. A pretrained IndoBERT Base model is fine-tuned using the verified 2,000 samples in a multi-class emotion classification task. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score, along with visual diagnostics like confusion matrices and loss curves to identify potential misclassifications.

In the second phase, the emotion labels from the 2,000 samples are removed, and the trained model is used to re-predict emotion labels for the entire 10,000-sample dataset. At this stage, the reason identification module is also applied. It uses contextual embeddings from IndoBERT and cosine similarity to compare each input with pre-defined semantic representations of the four reason categories. This process enables context-aware assignment of reason categories even in the absence of explicit keywords. The updated dataset, now pseudo-labeled with both emotion and reason annotations, is then used to conduct a second fine-tuning phase. This iterative refinement improves generalization and contextual depth across classification tasks.

In the final stage, each emotion label predicted by the model is associated with a set of expressive body movements through a gesture mapping module. These gestures are designed to reflect the emotional nuance of each label, described in rich detail to ensure natural alignment with the chatbot's verbal output. This integration enhances the multimodal capabilities of the system, allowing it to respond in a way that feels more humanlike and emotionally resonant through synchronized nonverbal expressions.

B. Data Collection and Annotation

To construct a relevant and context-sensitive dataset for emotion classification and reason identification in early marriage counseling scenarios, a total of 10,000 synthetic question–answer pairs were generated using a Large Language Model (LLM). The prompts were designed to emulate realistic conversations between adolescents and counselors, emphasizing emotionally complex themes such as hesitation, encouragement, fear, reflection, and social pressure. These prompts were written in Bahasa Indonesia and adapted to reflect the local expressions and sociocultural characteristics of Lombok and similar Indonesian regions, ensuring linguistic authenticity.

Given the absence of publicly available datasets related to early marriage counseling in Indonesian contexts, and considering ethical restrictions around accessing real counseling transcripts, this study relied entirely on synthetic data generation. Collaboration with psychologists or institutions was not conducted due to concerns regarding data privacy, access limitations, and time constraints. The use of LLM-generated dialogues provided a scalable and ethical alternative to create a varied dataset that mirrors typical counseling discourse.

From the 10,000 synthetic samples, a stratified subset of 2,000 was selected for manual annotation. This subset was labeled based on a function-oriented emotion taxonomy comprising five categories: 'Enthusiastic', 'Gentle', 'Analytical', 'Inspirational', and 'Cautionary'. These labels were chosen not only for their affective tones but also for their communicative roles in a counseling setting, for instance 'Gentle' responses are often soothing, while Cautionary messages deliver empathetic warnings. In contrast, the reason categories (Education, Economy, Religion, Culture) were not manually labeled, but rather inferred automatically using semantic similarity and IndoBERT embeddings in a separate module, as described in Section II-E.

Each sample in the annotated subset was labeled using a detailed rubric outlining the linguistic features, emotional functions, and conversational contexts associated with each category. Although the initial labeling leveraged automated emotion prediction using a LLM, every label underwent manual verification. This double-checking process ensured consistency with predefined emotional definitions and minimized classification noise. The finalized annotated dataset was then employed to fine-tune the IndoBERT base model, chosen for its suitability in capturing contextual nuances in the Indonesian language. Below are examples of the Q&A pairs used in the dataset, written in Bahasa Indonesia, and accompanied by their English translations, as shown in Table I and Table II.

TABLE I
SAMPLE Q&A

Input	Response	Label
Nikah muda bikin aku lebih gampang punya temen, nggak? "Does getting married young make it easier to make friends?"	*Punya teman datang dari interaksi sosial, nggak cuma dari nikah. Menikah cepat bisa bikin fokus ke keluarga. Coba gabung komunitas sekarang, kamu pasti punya banyak koneksi!* "Making friends comes from social interaction, not just marriage. Getting married early might shift your focus to family. Try joining a community now, you'll build lots of connections!"	Enthusiastic
Orang tuaku menikah muda dan baik-baik saja, apakah aku juga bisa? "My parents got married young and they're doing fine, can I do the same?"	*Kondisi setiap orang berbeda. Pastikan kamu memiliki kesiapan mental dan finansial sesuai kebutuhan zaman sekarang.* Everyone's situation is different. Make sure you're mentally and financially prepared for today's challenges.	Gentle
Apakah menikah muda memudahkan aku memiliki rumah sendiri? "Will getting married young make it easier for me to own a house?"	*Memiliki rumah membutuhkan stabilitas keuangan yang solid. Menikah tanpa perencanaan finansial bisa mempersulit tujuan ini.* "Owning a house requires strong financial stability. Getting married without a financial plan could make this goal harder to achieve."	Analytical
Apakah menikah muda bisa membantu meningkatkan karierku? "Can early marriage help boost my career?"	*Karier berkembang melalui dedikasi dan peluang, bukan hanya karena menikah. Pastikan kamu bisa menyeimbangkan keduanya.* "A career grows through dedication and opportunity, not just by getting married. Make sure you can balance both well."	Inspirational
Aku sulit mengatur keuangan, apakah menikah bisa membuatku lebih disiplin? "I struggle with money management, will marriage help me be more disciplined?"	*Keterampilan mengelola keuangan perlu dilatih sebelum menikah. Menikah tanpa kemampuan ini bisa menambah tantangan.* "Financial management is a skill to build before marriage. Getting married without this ability might add more challenges."	Cautionary

Table II displays the distribution of the initial 2,000 labeled samples, which were intentionally balanced across the five emotional categories. In comparison, table III highlights the resulting imbalance after automated labeling, reflecting the model's inherent biases influenced by the structure and content of the question–response pairs.

TABLE III
LABEL DISTRIBUTION OF SAMPLE DATASET

Label	Count	Percentage
Enthusiastic	400	20%
Gentle	400	20%
Analytical	400	20%
Inspirational	400	20%
Cautionary	400	20%

TABLE IIIII
LABEL DISTRIBUTION OF FIRST PHASE FINE-TUNING DATASET

Label	Count	Percentage
Enthusiastic	4348	20%
Gentle	2250	20%
Analytical	1985	20%
Inspirational	783	20%
Cautionary	643	20%

To facilitate model training and evaluation, the complete dataset was divided into training, validation, and test sets through stratified sampling to maintain proportional representation across emotion classes. This split 70% for training, 20% for validation, and 10% for testing allowed the model to generalize effectively by learning from a representative sample while being assessed on previously unseen data.

C. Preprocessing

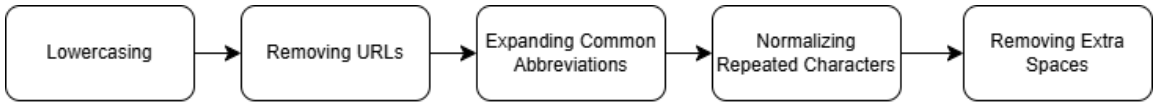

Fig. 3. Preprocessing Stage.

Based on Fig 3, the raw textual data is subjected to a systematic preprocessing pipeline to enhance its quality and ensure it aligns with the input expectations of the IndoBERT model. This pipeline begins with lowercasing, where all characters are converted to lowercase to reduce redundancy caused by case sensitivity for instance, "Marriage" and "marriage" are treated uniformly. Next, URLs are removed to eliminate extraneous content that does not carry emotional or semantic weight in the context of counseling dialogues, such as "click this https://indo.com" to “click this url”. The third step involves expanding common abbreviations, converting informal expressions like “krn” to “karena” which means “because, which improves the model's comprehension of colloquial input.

These preprocessing steps are applied not only to the emotion classification task, but also to the reason identification module, which uses IndoBERT embeddings to perform semantic similarity matching. Clean and normalized input ensures that the resulting sentence embeddings are contextually consistent, reducing variability caused by informal language patterns or noisy formatting.

Subsequently, repeated characters are normalized to handle exaggerated or expressive writing styles commonly found in social or informal texts. For example, expressions like "seru bangetttt" which means “so funnnn” to "seru banget" which means “so fun”. This step helps maintain semantic integrity without losing emotional nuances. Finally, extra spaces are removed to ensure formatting consistency and prevent parsing issues. Collectively, these preprocessing steps aim to reduce noise, preserve meaningful linguistic structures, and optimize the dataset for emotion classification using IndoBERT.

D. IndoBERT Embedding for Emotion Classification and Semantic Similarity

Word embedding is the process of converting words into dense numerical vectors that can be interpreted by machine learning models. In natural language processing, traditional methods such as bag-of-words or TF-IDF are limited by their inability to capture contextual relationships. More advanced models like Word2Vec and GloVe offer static embeddings but lack context sensitivity. Contextual embedding methods such as BERT,

ELMo, and IndoBERT [30] overcome this by representing words in context, producing different vector representations depending on the surrounding words.

IndoBERT is a contextual language model adapted from the original BERT architecture and optimized for processing Bahasa Indonesia [17]. In this study, IndoBERT serves as the core component for both emotion classification and reason identification modules. For classification, the model is fine-tuned using labeled conversational data to detect function-oriented emotional categories. For semantic similarity, IndoBERT is employed in inference-only mode to generate sentence-level embeddings that can be matched to predefined reason categories using cosine similarity.

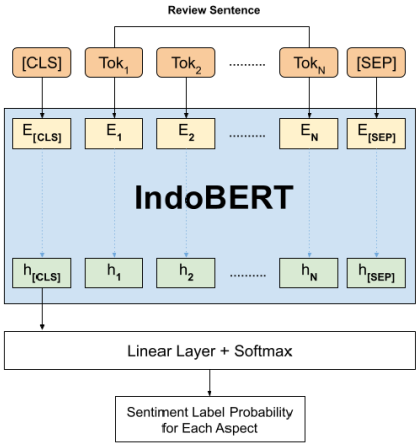

Fig. 4. IndoBERT Architecture [31].

Fig 4 illustrates the overall architecture of IndoBERT when applied to the emotion classification task. The process begins by passing the input sentence already embedded using token, segment, and positional encodings into IndoBERT's Transformer encoder layers. Each token in the sequence generates a contextualized hidden state (h_1, h_2, ..., h_n), with special attention given to the hidden state corresponding to the [CLS] token. This [CLS] representation functions as a condensed summary of the entire sentence and is commonly used in classification tasks [31].

In this study, the [CLS] token's final hidden state is passed into a linear layer followed by a softmax function, which outputs a probability distribution over five predefined emotion categories 'Enthusiastic, Gentle, Analytical, Inspirational, and Cautionary. The classification process is optimized through two fine-tuning stages: one using 2,000 manually labeled samples and another using pseudo-labeled outputs from a 10,000-sample dataset. This architecture enables IndoBERT to serve as a robust emotion classifier while maintaining its language understanding capabilities in Indonesian contexts.

To train the model for the emotion classification task, the dataset was split into three parts: 70% training, 20% validation, and 10% testing. This split ensures sufficient data for learning while reserving data for hyperparameter tuning and unbiased evaluation.

The training employed the AdamW optimizer and a learning rate scheduler to progressively reduce the learning rate throughout training, which helps stabilize updates and preserve IndoBERT's pre-trained knowledge. The initial learning rate was set at 2e-5, a configuration commonly used for fine-tuning Transformer-based models on small datasets. The model was fine-tuned for 5 epochs during Phase 1 using a batch size of 8, and early stopping with a patience of 1 epoch was applied to prevent overfitting. In Phase 2, the model was fine-tuned again for 3 epochs using the pseudo-labeled dataset, also with the same early stopping configuration. To address the significant class imbalance present in this dataset, weighted loss functions were

incorporated during fine-tuning. The complete hyperparameter configurations used in both training phases are summarized in table IV and table V for better clarity.

TABLE IV
HYPERPARAMETER TUNING ON FIRST PHASE FINE-TUNING

Hyperparameter	Value
Optimizer	AdamW
Learning Rate	0 - 2e-5(with scheduler)
Epochs	5
Batch Size	8

TABLE V
HYPERPARAMETER TUNING ON SECOND PHASE FINE-TUNING.

Hyperparameter	Value
Optimizer	AdamW
Learning Rate	0 - 2e-5(with scheduler)
Epochs	3
Batch Size	8

In this study, the training process was carried out in three main stages. In the first phase, the IndoBERT model was fine-tuned using a labeled dataset consisting of 2,000 samples, allowing the model to adapt its pre-trained knowledge to the specific context of early marriage counseling conversations. Once the first fine-tuning phase was completed, the model was used to perform inference on a larger unlabeled dataset containing 10,000 samples, generating emotion label predictions for each sentence. In the final phase, the IndoBERT model underwent a second fine-tuning using the 10,000 samples from the inference stage, which now contained pseudo-labels. This two-step fine-tuning approach, combined with pseudo-labeling, allowed the model to gradually expand the amount of labeled data and ultimately improved its generalization capability on the target task.

E. Evaluation Phase

To evaluate the performance of the fine-tuned IndoBERT model in recognizing function-based emotional categories, four standard classification metrics were used: accuracy, precision, recall, and F1-score [32]. Together, these metrics offer a well-rounded view of how effectively the model identifies the five emotion classes ‘Enthusiastic’, ‘Gentle’, ‘Analytical’, ‘Inspirational’, and ‘Cautionary’, each capturing a distinct communicative purpose within counseling conversations.

a) Accuracy

Accuracy represents the proportion of correct predictions relative to the total number of predictions made. It offers a simple yet valuable indicator of overall performance, particularly when class distributions are relatively balanced [33]. The formula to calculate Accuracy is as follows:

$$\boldsymbol{Accuracy} = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Total\ Samples\ (TP + TN + FP + FN)} \quad (1)$$

b) Precision

Precision measures the proportion of correct positive predictions out of all positive predictions made by the model [34]. This metric is particularly important when false positives have significant consequences, such as incorrectly labeling a cautionary message as inspirational. Precision is computed as follows:

$$\boldsymbol{Precision} = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \tag{2}$$

A high precision score ensures that the emotional labels assigned by the model, especially those with sensitive counseling intent, are accurate and contextually appropriate [34].

c) Recall

Recall, on the other hand, evaluates the model's ability to correctly identify all relevant instances of a particular emotional class [32]. It is calculated using formula as follows:

$$\boldsymbol{Recall} = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negative\ (FN)} \tag{3}$$

This metric is crucial for applications such as early marriage counseling, where overlooking emotional cues (false negatives) may result in responses that are inadequate or lack empathy.

d) F1-Score

The F1-score provides a balanced measure of precision and recall by calculating their harmonic mean. This metric is particularly beneficial in scenarios with imbalanced class distributions or when reducing both false positives and false negatives is equally important [32]. It is given by the following equation:

$$\boldsymbol{F1 - Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

This balanced metric is particularly appropriate for multi-class emotion classification, where each emotional label may have differing degrees of representation and predictive difficulty.

e) Confusion Matrix and Plot

In addition to the standard metrics, visual evaluation tools such as confusion matrices and performance curves are essential for understanding how effectively a Transformer-based model like IndoBERT generalizes. A confusion matrix identifies where predictions match or deviate from true labels, highlighting misclassification trends. This is especially valuable for closely related emotion categories like Gentle and Inspirational, which often overlap in counseling contexts.

Plotting training and validation accuracy and loss curves across epochs also provide important insights. These plots help detect underfitting, where accuracy remains low and loss remains high, or overfitting, where training accuracy continues to improve but validation accuracy stagnates or declines. Ideally, both accuracy and loss metrics should demonstrate consistent improvement and

remain closely aligned between training and validation sets, indicating that the model is learning robust patterns. Reducing both false positives and false negatives is equally important [28].

In addition to evaluating the emotion classification module, this study also assesses the performance of the reason identification component. As this module operates using an unsupervised semantic similarity approach based on IndoBERT embeddings, it does not require additional supervised training or labeled data for fine-tuning, as it operates in inference mode using IndoBERT embeddings and semantic similarity. Therefore, evaluation is conducted qualitatively by examining the plausibility and coherence of the assigned categories (Education, Economy, Religion, Culture) across a representative sample of user inputs. The results suggest that the semantic similarity method effectively maps inputs to their appropriate reason categories, even when explicit keywords are absent. While no ground-truth labels are currently available to support a quantitative evaluation, this module demonstrates strong potential for identifying underlying drivers of early marriage decisions based on linguistic context alone. Future work may include human-annotated labels or crowd-sourced validation to quantify performance more rigorously.

F. Mapping Emotional Labels to Gestural Expressions

The integration of body gestures as a complement to textual emotional expression plays a crucial role in improving the interpretability and naturalness of counseling interactions. Each emotion category is mapped to specific non-verbal gestures that convey its underlying communicative intent. Based on table VI, Enthusiastic is associated with a broad smile, open eyes, and raised eyebrows, projecting excitement and engagement. Meanwhile, Gentle uses a soft smile and relaxed eyes to establish empathy and a calming presence, which is especially valuable in sensitive counseling contexts.

TABLE VI
REFERENCE FOR MAPPING EMOTIONS TO BODY GESTURES

Emotion	Body Gestures	Description
Enthusiastic	Wide smile, eyes fully open, eyebrows lifted	Shows excitement and involvement through expressive facial expressions and an open demeanor.
Gentle	Subtle smile, relaxed eye expression, neutral eyebrows	Demonstrates tranquility and reassurance, conveying empathy and a comforting presence.
Analytical	Eyebrows drawn together, lips closed, concentrated gaze	Indicates deep thinking and attentiveness, representing analysis or problem-solving focus.
Inspirational	Genuine smile, brightened eyes, slightly elevated eyebrows	Projects positivity and encouragement, fostering a forward-looking and optimistic atmosphere.
Cautionary	Mild frown, intense gaze, steady eye contact	Signals concern and cautious advice, reflecting attentiveness and measured warning.

On the other hand, Analytical is represented by furrowed eyebrows, closed lips, and a focused gaze, illustrating critical thinking and attention to detail. Similarly, Inspirational combines a sincere smile, bright eyes, and slightly raised eyebrows to express motivation and optimism, encouraging the user toward positive actions. Finally, Cautionary is characterized by a slight frown, a serious gaze, and focused eye contact, signaling careful warning and measured concern.

This mapping not only enhances emotional clarity but also provides virtual agents with multimodal capabilities, allowing them to communicate intent and empathy beyond text. Such integration of gestures

strengthens user trust and engagement, as it mirrors natural human communication patterns and supports the emotional depth required for counseling applications.

IV. Results And Discussion

A. Fine-tuning Phase 1

The first fine-tuning phase was conducted using the balanced labeled dataset previously described in Section III-C, which consisted of 2,000 samples distributed evenly across the five emotion classes. This initial stage aimed to establish a strong baseline model that could be used for inference on a much larger unlabeled dataset.

The fine-tuning was performed for 5 epochs using the AdamW optimizer with a learning rate of 2e-5, together with a learning rate scheduler to gradually reduce the learning rate during training. The batch size was set to 8, and an early stopping mechanism (patience = 1 epoch) was applied to avoid overfitting. Following training, the model achieved a validation accuracy of 96.5% and a test accuracy of 95% across the five emotion categories. Table VII presents the performance metrics obtained from the test set.

TABLE VII
Evaluation Matrix on First Phase Fine-tuning

Label	Precision	Recall	F1-Score	Support
Enthusiastic	1.00	1.00	1.00	40
Gentle	0.95	0.90	0.92	40
Analytical	0.90	0.90	0.90	40
Inspirational	0.93	0.97	0.95	40
Cautionary	0.97	0.97	0.97	40

Based on the confusion matrix shown in Fig 5, the model demonstrates strong performance, achieving a 95% test accuracy on the 200-sample test set. The Enthusiastic class was classified perfectly with no mispredictions, while the Gentle and Analytical classes still exhibited some classification errors. Specifically, four samples from the Gentle class were misclassified as either Analytical or Inspirational, whereas the Analytical class had four misclassified samples distributed across the Gentle, Inspirational, and Cautionary classes. The Inspirational and Cautionary classes achieved high performance with 39 correct predictions each, although each had one sample misclassified into the Analytical class.

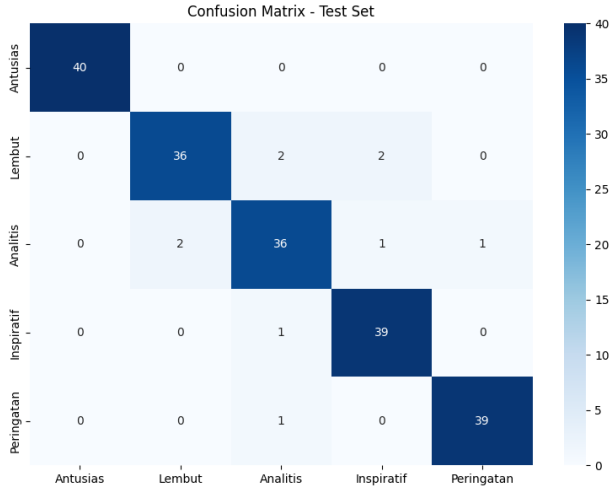

Fig. 5. Confusion Matrix from First Phase Fine-tuning.

Most of the classification errors can be attributed to overlapping linguistic characteristics between certain classes, particularly between Gentle and Analytical. These two categories often employ neutral and formal language styles, making them difficult for the model to distinguish. In addition, the class imbalance observed in earlier phases, where 'Gentle' and 'Cautionary' contained fewer samples compared to other categories likely impacted the model's ability to recognize minority classes. Some test samples also contained ambiguous contexts that could plausibly belong to more than one emotional category, further increasing the likelihood of misclassification.

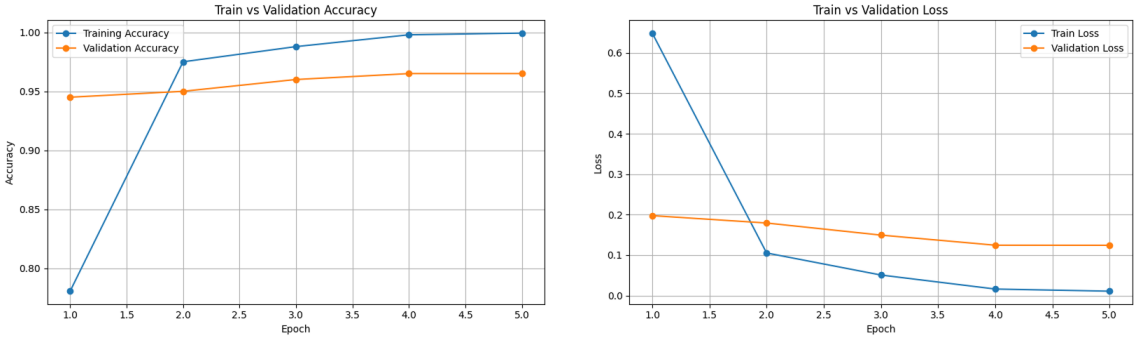

Fig. 6. Plots of Accuracy and Loss in First Phase.

The training and validation accuracy and loss curves presented in Fig 6 provide additional insight into the model's performance. Both accuracy and loss trends show consistent improvement across epochs, with the validation metrics closely tracking the training metrics, suggesting that the model was learning effectively without overfitting. The early stabilization of the validation curves also confirms the effectiveness of the applied learning rate scheduler and early stopping mechanism in maintaining generalization capability. These results indicate that the fine-tuned model has achieved stable and reliable performance, making it suitable for the next phase of experimentation.

B. Inference on Unlabeled Dataset

In this phase, the fine-tuned IndoBERT model from Phase 1 was used to predict emotion labels for 10,000 previously unlabeled counseling text samples. In addition, each input was also passed through the Reason

Identification module, which utilized IndoBERT embeddings to compute semantic similarity between the input sentence and predefined category: Education, Economy, Religion, and Culture. This dual prediction process allowed the system to enrich the dataset with both emotional and contextual reasoning labels, providing a richer context for further analysis.

The resulting pseudo-labeled dataset, now annotated with both emotion and reason categories, was then used to support the second fine-tuning phase for emotion classification and enable downstream analysis of contextual patterns in user responses. This approach not only expanded the dataset but also helped ensure that the system's predictions were more contextually aware of the reasons behind the emotions expressed, thereby improving the model's overall understanding of user motivations.

Table IX presents the distribution of predicted emotion labels within the unlabeled dataset. As can be seen, the dataset shows an imbalance in the representation of different emotion categories. Enthusiastic is the most frequent label, comprising 43.4% of the dataset, while Cautionary represents the smallest proportion at 6.4%. This distribution reflects the inherent bias in the original dataset, which was not specifically balanced to ensure equal representation across emotion categories.

TABLE VIII
Label Distribution of First Phase Fine-tuning

Label	Count	Percentage
Enthusiastic	4348	43.4%
Gentle	2250	22.5%
Analytical	1985	19.8%
Inspirational	783	7.8%
Cautionary	643	6.4%

Moreover, table IX shows the distribution of the reason categories predicted using the Reason Identification module. Education is the most prevalent reason, comprising 59.94% of the dataset, followed by Culture (16.78%) and Religion (16.09%). The Economy category is the least represented, accounting for only 7.19% of the dataset. This imbalance in reason categories may have an impact on the model's performance, especially in the next fine-tuning phase where the model will be trained on a larger, potentially more diverse dataset.

TABLE IX
Reason Categories Distribution

Label	Count	Percentage
Education	5994	59.94%
Economy	719	7.19%
Culture	1678	16.78%
Religion	1609	16.09%

The imbalance observed in both emotion labels and reason categories is noteworthy. The Education category, for instance, represents a disproportionately high percentage of the dataset compared to the other categories. This is a result of the initial dataset construction, where labels were not evenly distributed. To address this issue, class-weighted loss was applied during the training of the model to ensure that the loss function did not overly penalize the model for misclassifying the majority classes. This approach allowed the model to focus more on the minority classes, which are crucial for understanding the broader range of emotional responses in sensitive contexts such as early marriage counseling.

C. *Fine-tuning Phase 2*

The second fine-tuning phase was conducted using the 10,000 pseudo-labeled samples obtained from the inference stage. Unlike the balanced dataset used in Phase 1, this dataset exhibited significant label imbalance, table IX. To mitigate this issue, class weight balancing was applied during training, ensuring that minority classes received proportionally higher loss contributions to counteract the dominance of majority classes. This step was crucial for improving sensitivity to underrepresented labels.

The fine-tuning was performed for 3 epochs using the AdamW optimizer with an initial learning rate of 2e-5 and a learning rate scheduler to progressively reduce the learning rate during training. The batch size was maintained at 8, and an early stopping mechanism (patience = 1 epoch) was again applied to prevent overfitting. Following training, the model achieved a validation accuracy of 88.8% and a test accuracy of 89.6% across the five emotion categories. Table X presents the detailed performance metrics from the test set.

TABLE X
EVALUATION MATRIX ON SECOND PHASE FINE-TUNING

Label	Precision	Recall	F1-Score	Support
Enthusiastic	0.89	0.89	0.89	199
Gentle	0.81	0.85	0.83	78
Analytical	0.91	0.93	0.92	435
Inspirational	0.90	0.85	0.87	225
Cautionary	0.92	0.90	0.90	63

Based on the confusion matrix shown in Fig 7, the model achieved strong performance during the second fine-tuning phase, correctly classifying the majority of samples across all emotion categories. The ‘Enthusiastic’ class achieved the highest accuracy, with 177 of 199 samples correctly predicted. The ‘Analytical’ class also performed strongly, with 406 correct predictions. However, the ‘Gentle’ and ‘Inspirational’ classes showed slightly more misclassifications, with several samples predicted as ‘Analytical’ or ‘Cautionary’.

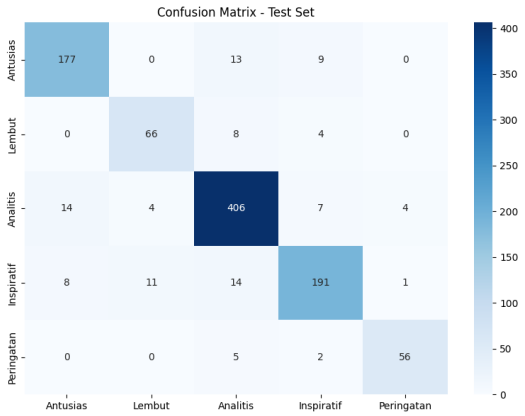

Fig. 7. Confusion Matrix from Second Phase Fine-tuning.

These misclassifications can be attributed to overlapping linguistic cues between the classes, particularly between Gentle-Analytical and Inspirational-Gentle, which often share neutral or supportive language styles. Although class weight balancing helped reduce bias toward the majority classes, the inherent imbalance from the pseudo-labeled dataset still contributed to some performance gaps. Nevertheless, the overall accuracy remained high, and the model demonstrated improved robustness compared to the first fine-tuning phase.

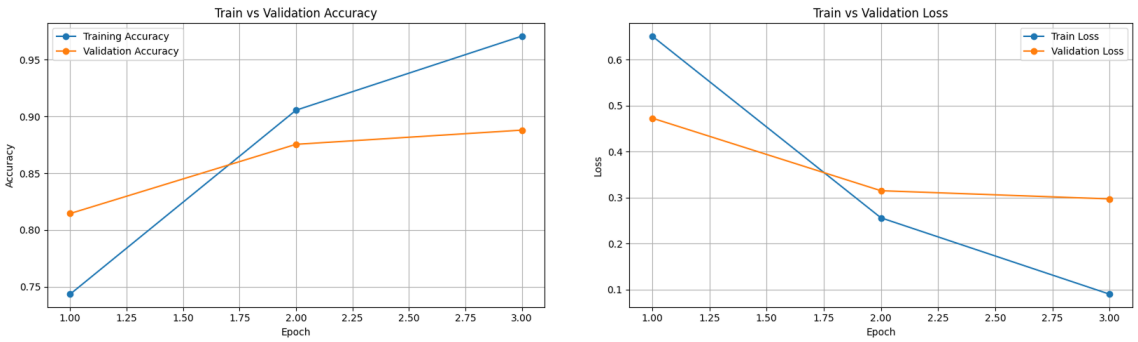

Fig. 8. Plots of Accuracy and Loss in First Phase

The training and validation accuracy and loss curves presented in Fig 8 support these findings. reinforce these findings. Both metrics show steady improvement across epochs, with validation curves closely tracking training curves, indicating that the model learned effectively without overfitting. These results suggest that the Phase 2 fine-tuned model is robust and capable of handling a more complex and imbalanced dataset, making it suitable for deployment in downstream tasks.

D. Gesture Visualization

To enhance the interpretability and expressiveness of the emotion classification model, a gesture mapping framework was developed to link each emotional category with specific nonverbal behaviors. This mapping draws from established counseling theory, which emphasizes the importance of body language in conveying empathy, support, and understanding during interactions. By translating the classified emotions into appropriate gestures, the virtual counseling chatbot can provide a more engaging and human-like user experience.

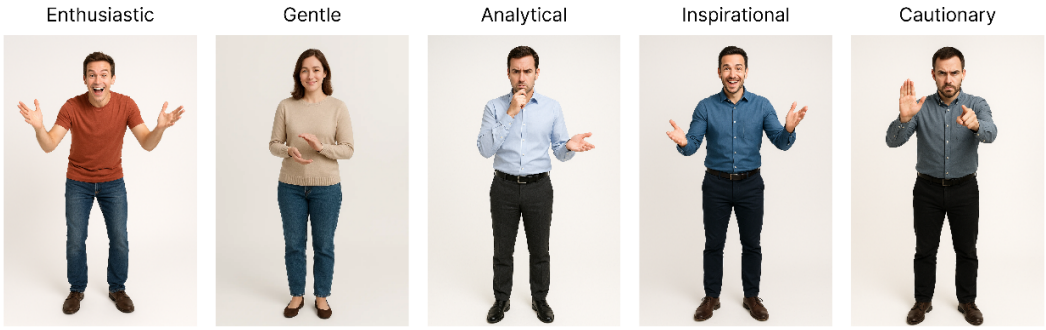

Fig. 9. Emotional representation illustrated via unique body gestures.

As illustrated in Fig 9, each of the five emotion categories was mapped to distinct gesture sets, focusing on facial expressions, eye contact, and posture. For example, the ‘Enthusiastic’ label is depicted with an open body posture, bright smiles, and expressive hand gestures to project energy and enthusiasm. ‘Gentle’ is represented by a calm demeanor, relaxed facial features, and softer hand movements, reflecting comfort and empathy. ‘Analytical’ shows a thoughtful posture, often accompanied by a focused gaze and minimal gestures, signifying contemplation. ‘Inspirational’ combines warm smiles and open-handed gestures to inspire and encourage users, while ‘Cautionary’ category adopts a serious expression, forward hand gestures, and steady eye contact to emphasize the importance of the message being conveyed.

The current emotion-to-gesture mapping is limited to descriptive associations and static pose visualizations. It functions as a conceptual basis for linking the five emotional categories with suitable nonverbal expressions while reinforcing emotional alignment with user inputs. However, it has not yet been expanded to include real-time gesture animation or dynamic rendering. Future work may focus on integrating these mappings into interactive avatar systems to further improve the realism and emotional engagement of the virtual counseling experience.

Overall, the two-stage fine-tuning approach demonstrated the system's ability to adapt from a balanced, smaller dataset to a larger and more varied dataset while maintaining strong generalization. Classes with distinct linguistic patterns were classified accurately, whereas categories with overlapping characteristics proved more challenging, highlighting the importance of data distribution quality. The integration of emotion-to-gesture mapping further reinforces the framework's potential by bridging textual emotion recognition with nonverbal cues. Although this component is currently limited to descriptive and static visualizations, it lays the groundwork for future enhancements toward dynamic, real-time interaction. Collectively, these findings indicate that the framework can serve as a foundation for developing emotionally aware and culturally aligned virtual counseling systems.

V. CONCLUSION

This study developed and evaluated an integrated framework for emotion classification and reason identification in counseling dialogues for early marriage prevention using the IndoBERT architecture. Through a two-phase fine-tuning strategy, the model achieved strong generalization with test accuracies of 95% and 88.8% in Phases 1 and 2, respectively, across five functional emotion categories. These results demonstrate that progressive fine-tuning on increasingly diverse datasets effectively mitigates class imbalance and improves overall performance, particularly for minority classes. The reason identification module, leveraging IndoBERT-based sentence embeddings, further enhanced contextual understanding compared to conventional keyword matching. Beyond text classification, the introduction of emotion-to-gesture mapping provides a conceptual foundation for building more empathetic virtual counseling agents. Overall, this research advances emotion-aware conversational systems in low-resource settings and opens opportunities for future multimodal frameworks capable of delivering culturally adaptive and emotionally intelligent support.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Telkom University for providing the necessary support and the opportunity to carry out this research. We are also deeply thankful to the researchers whose prior work has served as a foundation and inspiration for the concepts and methodology adopted in this study. Furthermore, we wish to extend our appreciation to all individuals and parties who have contributed to this research, both through material support and intellectual input.

REFERENCES

[1] R. Nabila, R. Roswiyani, and H. Satyadi, "A Literature Review of Factors Influencing Early Marriage Decisions in Indonesia," in *Proceedings of the 3rd Tarumanagara International Conference on the Applications of Social Sciences and Humanities (TICASH 2021)*, Atlantic Press, 2022, pp. 1392–1402. doi: 10.2991/assehr.k.220404.223.

[2] D. Fadilah, "Tinjauan Dampak Pernikahan Dini dari Berbagai Aspek," *Pamator Journal*, vol. 14, no. 2, pp. 88–94, Nov. 2021, doi: 10.21107/pamator.v14i2.10590.

[3] Z. Ayudiputri, A. Nur, S. Amanda, and F. Hanifa, "Determinants of Child Marriage in Indonesia : A Systematic Review," *Journal of Community Medicine and Public Health Research*, vol. 5, no. 2, pp. 216–227, Nov. 2024, doi: 10.20473/jcmphr.v5i2.45777.

[4] L. Wang *et al.*, "CASS: Towards Building a Social-Support Chatbot for Online Health Community," in *Conference on Computer-Supported Cooperative Work & Social Computing (CSCW)*, Feb. 2021, pp. 1–31. doi: https://doi.org/10.48550/arXiv.2101.01583.

[5] S. Khandelwal, "SOCIAL COMPANION CHATBOT FOR HUMAN COMMUNICATION USING ML AND NLP," *International Journal of Engineering Applied Sciences and Technology*, vol. 8, pp. 321–324, 2023, doi: https://doi.org/10.33564/IJEAST.2023.v08i01.048.

[6] R. E. Guingrich and M. S. A. Graziano, "Chatbots as Social Companions: How People Perceive Consciousness, Human Likeness, and Social Health Benefits in Machines," in *Oxford Intersections: AI in Society*, Oxford University Press, 2025. doi: https://doi.org/10.1093/9780198945215.001.0001.

[7] G. Z. Nabiilah, I. N. Alam, E. S. Purwanto, and M. F. Hidayat, "Indonesian multilabel classification using IndoBERT embedding and MBERT classification," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 1, pp. 1071–1078, Feb. 2024, doi: 10.11591/ijece.v14i1.pp1071-1078.

[8] S. Prabhu, M. Moosa, and H. Misra, *Multi-class Text Classification using BERT-based Active Learning*. 2021. doi: 10.48550/arXiv.2104.14289.

[9] I. Ameer, N. Bölücü, M. H. F. Siddiqui, B. Can, G. Sidorov, and A. Gelbukh, "Multi-label emotion classification in texts using transfer learning," *Expert Syst Appl*, vol. 213, Mar. 2023, doi: 10.1016/j.eswa.2022.118534.

[10] 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, pp. 1–17, 2024, doi: https://doi.org/10.1155/2024/2826773.

[11] S. K. Bharti *et al.*, "Text-Based Emotion Recognition Using Deep Learning Approach," *Comput Intell Neurosci*, vol. 2022, no. 1, p. 2645381, 2022, doi: https://doi.org/10.1155/2022/2645381.

[12] U. Malik, S. Bernard, A. Pauchet, C. Chatelain, R. Picot-Clémente, and J. Cortinovis, "Pseudo-Labeling With Large Language Models for Multi-Label Emotion Classification of French Tweets," *IEEE Access*, vol. 12, pp. 15902–15916, 2024, doi: 10.1109/ACCESS.2024.3354705.

[13] M. Y. Baihaqi, E. Halawa, R. A. S. Syah, A. Nurrahma, and W. Wijaya, "Emotion Classification in Indonesian Language: A CNN Approach with Hyperband Tuning," *Jurnal Buana Informatika*, vol. 14, no. 02, pp. 137–146, Oct. 2023, doi: 10.24002/jbi.v14i02.7558.

[14] A. Zamsuri, S. Defit, and G. W. Nurcahyo, "Classification of Multiple Emotions in Indonesian Text Using The K-Nearest Neighbor Method," *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 4, no. 2, pp. 1012–1021, Jun. 2023, doi: 10.37385/jaets.v4i2.1964.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[16] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," Association for Computational Linguistics, Dec. 2020, pp. 843–857. doi: https://doi.org/10.48550/arXiv.2009.05387.

[17] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, International Committee on Computational Linguistics, Nov. 2020, pp. 757–770. doi: https://doi.org/10.48550/arXiv.2011.00677.

[18] C. Shaw, P. LaCasse, and L. Champagne, "Exploring emotion classification of indonesian tweets using large scale transfer learning via IndoBERT," *Soc Netw Anal Min*, vol. 15, no. 1, p. 22, 2025, doi: 10.1007/s13278-025-01439-6.

[19] M. Nadas, L. Diosan, and A. Tomescu, "Synthetic Data Generation Using Large Language Models: Advances in Text and Code," Mar. 2025, [Online]. Available: http://arxiv.org/abs/2503.14023

[20] L. Long *et al.*, "On LLMs-Driven Synthetic Data Generation, Curation, and Evaluation: A Survey," Jun. 2024, [Online]. Available: http://arxiv.org/abs/2406.15126

[21] Y. Li, R. Bonatti, S. Abdali, J. Wagle, and K. Koishida, "Data Generation Using Large Language Models for Text Classification: An Empirical Case Study," Jul. 2024, [Online]. Available: http://arxiv.org/abs/2407.12813

[22] X. Li, "Recognition Characteristics of Facial and Bodily Expressions: Evidence From ERPs," *Front Psychol*, vol. Volume 12-2021, 2021, doi: 10.3389/fpsyg.2021.680959.

[23] A. Diwan, R. Sunil, P. Mer, R. Mahadeva, and S. P. Patole, "Advancements in Emotion Classification via Facial and Body Gesture Analysis: A Survey," *Expert Syst*, vol. 42, no. 2, p. e13759, 2025, doi: https://doi.org/10.1111/exsy.13759.

[24] C. Forceville, *Visual and Multimodal Communication: Applying the Relevance Principle: Introduction*. 2020. doi: 10.1093/oso/9780190845230.001.0001.

[25] U. Bhattacharya, N. Rewkowski, A. Banerjee, P. Guhan, A. Bera, and D. Manocha, "Text2Gestures: A Transformer-Based Network for Generating Emotive Body Gestures for Virtual Agents," Jul. 2021, pp. 1–10. doi: 10.1109/VR50410.2021.00037.

[26] C. A. Khairunnisa, S. A. M. Sitorus, V. D. Puspita, and I. Maryati, "Exploration of Factors Influencing Early Marriage in Adolescents: A Literature Review," *Care : Jurnal Ilmiah Ilmu Kesehatan*, vol. 12, no. 2, pp. 205–214, Jul. 2024, doi: 10.33366/jc.v12i2.5694.

[27] M. Fitria, A. D. Laksono, I. M. Syahri, R. D. Wulandari, R. Matahari, and Y. Astuti, "Education role in early marriage prevention: evidence from Indonesia's rural areas," *BMC Public Health*, vol. 24, no. 1, p. 3323, 2024, doi: 10.1186/s12889-024-20775-4.

[28] D. A. R. Sojais, J. Suyanto, and H. Rustandi, "Economic, Social, and Cultural Contexts of Early Marriage in Bengkulu Province," *Jurnal Aisyah : Jurnal Ilmu Kesehatan*, vol. 8, no. 2, Jun. 2023, doi: 10.30604/jika.v8i2.2047.

[29] K. D. Rahadika Diana and M. L. Khodra, "IndoSBERT: Enhancing Indonesian Sentence Embeddings with Siamese Networks Fine-tuning," in *2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2023, pp. 1–6. doi: 10.1109/ICAICTA59291.2023.10390469.

[30] E. Luthfi, Z. Yusoh, and B. Aboobaider, "Enhancing the Takhrij Al-Hadith based on Contextual Similarity using BERT Embeddings," *International Journal of Advanced Computer Science and Applications*, vol. 12, Jul. 2021, doi: 10.14569/IJACSA.2021.0121133.

[31] E. Yulianti and N. Nissa, "ABSA of Indonesian customer reviews using IndoBERT: single- sentence and sentence-pair classification approaches," *Bulletin of Electrical Engineering and Informatics*, vol. 13, pp. 3579–3589, Jul. 2024, doi: 10.11591/eei.v13i5.8032.

[32] J. Opitz, "A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice," *Transactions of the Association for Computational Linguistics 2024*, vol. 12, pp. 820–836, Apr. 2024, doi: 10.1162/tacl_a_00675.

[33] T. Schlosser, M. Friedrich, T. Meyer, D. Kowerko, and J. Professorship, *A Consolidated Overview of Evaluation and Performance Metrics for Machine Learning and Computer Vision*. 2024. doi: 10.13140/RG.2.2.14331.69928.

[34] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci Rep*, vol. 14, no. 1, p. 6086, Dec. 2024, doi: 10.1038/s41598-024-56706-x.