Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

Natural Language Understanding for School Bullying Detection and Consultation: A DIET Classifier Approach in RASA Framework

Yoan Freddy Irawan[*1], Kristophorus Hadiono[2]

[1,2]Graduate School of Information Technology, Faculty of Information Technology and Industries, Stikubank University, Semarang, Indonesia

Email: [1]yoanfreddy0012@mhs.unisbank.ac.id

Received : May 14, 2025; Revised : Jun 4, 2025; Accepted : Jun 24, 2025; Published : Feb 15, 2026

Abstract

This research presents the development and implementation of a DIET classifier-based chatbot system using the RASA Framework to handle bullying reports at SMP Negeri 3 Ungaran. The system aims to provide 24/7 automated counseling support service, addressing the limitations of traditional human-to-human support systems that often result in delayed responses and reduced user satisfaction. The model was trained using a structured dataset comprising 61 dialogue examples collected through interviews with experienced guidance and counseling teachers, capturing authentic student communication patterns related to bullying issues. The evaluation results demonstrate exceptional performance, achieving 100% accuracy across 12 intent categories, with perfect precision and recall scores. The system successfully distinguishes between various emotional states and counseling needs, providing appropriate responses with high confidence levels. The intent categories include emotional expressions (merasa_dibully, merasa_sedih, merasa_takut), support-seeking behaviors (butuh_nasihat, ingin_bicara_dengan_guru), and conversational elements, ensuring comprehensive coverage of bullying-related communication scenarios. This implementation proves that AI-driven solutions can effectively support educational institutions in providing immediate, accessible counseling assistance while maintaining accuracy in emotional support and bullying prevention. This research contributes to the field of computer science by demonstrating the practical application of natural language understanding frameworks in sensitive educational contexts, advancing AI-driven counseling systems that can be scaled across educational institutions. The study provides a replicable methodology for developing culturally-sensitive AI applications in educational environments, particularly valuable for institutions in developing countries with limited digital mental health resources.

Keywords : *Bullying Prevention, Chatbot, Counseling Support, RASA Framework*

1. INTRODUCTION

The rapid growth of web-based services has transformed how virtual assistance and consultation services are delivered, with institutions seeking to provide convenient access to information for users who need immediate support [1]. Several forms of services are available, such as live chat and telephone services, but these human-to-human support systems require considerable response time. As user numbers grow, waiting times increase, resulting in low user satisfaction [2]. While several AI-based bullying prevention systems exist in educational settings, there remains a need for more specialized solutions that can effectively handle the complex emotional and psychological aspects of bullying reporting [3].

The definition of natural and instinctive connection modes is an essential objective in Human-Computer Interaction [2]. Deep learning, one of the techniques in artificial intelligence, imitates how the human brain processes and understands natural language [4]. Chatbots hold the promise of revolutionizing education by engaging learners, personalizing learning activities, supporting educators, and developing deep insight into learners' behavior [5]. In the context of bullying prevention, this

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

technology must be carefully designed to recognize and respond appropriately to various emotional expressions and sensitive situations.

Mental health disorders are a leading cause of disability worldwide, and there is a global shortage of mental health professionals. AI chatbots have emerged as a potential solution, offering accessible and scalable mental health interventions [6]. AI chatbots can provide immediate support by answering questions, offering explanations, and providing additional resources [7]. Recent research demonstrates that chatbots are perceived as non-judgmental and provide an impression of privacy and anonymity when interacting with digital agents [8], making them particularly suitable for sensitive applications like bullying reporting.

At SMP Negeri 3 Ungaran, the bullying reporting system requires a platform that can provide nonstop service 24 hours a day, seven days a week. The RASA Framework becomes an appropriate solution as it consists of two main components [9]: RASA Core as a conversation engine that determines what to do next, and RASA NLU as an open-source component that supports natural language understanding, intent classification, and entity extraction [10]. The Dual Intent and Entity Transformer (DIET) architecture advances the state of the art on complex multi-domain NLU datasets and achieves similarly high performance on other simpler datasets [11]. For clarity, 'intent' refers to the student's purpose in communication (e.g., seeking help, reporting an incident), while 'emotional state' represents the underlying feelings expressed in their message (e.g., fear, distress).

The bullying phenomenon in schools continues to be a significant challenge that requires comprehensive and innovative solutions. Bullying can have profound psychological and emotional impacts on students, potentially causing long-term consequences such as depression, anxiety, low self-esteem, and academic underperformance [12], [13]. Traditional reporting mechanisms often fail to provide victims with a sense of safety and immediate support, which can further exacerbate the trauma experienced by students [14]. Natural language processing (NLP) methods demonstrate promising improvements to empower proactive mental healthcare and assist early diagnosis [15].

Meta-analysis of recent studies in AI-driven mental health support demonstrates varying effectiveness across different contexts. Research examining chatbot interventions in educational settings shows 70-85% user satisfaction rates, with higher effectiveness observed in systems using advanced NLP techniques like transformer-based architectures [16], [17]. However, studies show that chatbots are still inaccurate regarding emotion detection, their language is not adapted to children's way of speaking and writing, and they are too predictable [3]. Gaps remain in specialized applications for bullying prevention, particularly in Indonesian educational contexts where cultural sensitivity and language nuances require tailored approaches [18], [19]. By implementing an AI-powered chatbot system specifically designed for bullying reporting, SMP Negeri 3 Ungaran can address several critical challenges:

1. Anonymity and Psychological Safety: offering students a confidential channel to report bullying incidents without fear of direct confrontation or social stigma;
2. Immediate Response and Support: providing immediate initial support, guidance, and documentation of reported incidents; and
3. Systematic Documentation and Analysis: enabling systematic documentation of bullying incidents while maintaining strict privacy measures to protect student confidentiality.

Compared to existing studies, this research provides a novel integration of DIET architecture in RASA Framework [20], [21], specifically designed for emotional state recognition in school bullying cases, offering a unique combination of anonymity, real-time support, and systematic documentation that has not been comprehensively addressed in previous educational AI implementations. Unlike general-purpose chatbots, this system is specifically trained using actual counseling data from bullying

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

cases, enabling more nuanced understanding of emotional distress signals and appropriate therapeutic responses [22], [23].

While the system offers significant advantages, it is important to acknowledge its limitations. The chatbot serves as an initial point of contact and support system, not a replacement for professional counseling. The proposed system will be developed using the RASA Framework focusing on utilizing Dual Intent and Entity Transformer (DIET), implemented to handle bullying reports at SMP Negeri 3 Ungaran, where training data will be collected from cases and frequently asked questions handled by counseling teachers regarding bullying issues.

This research contributes to the advancement of computational linguistics and human-computer interaction by demonstrating how domain-specific training data and cultural context can enhance AI system performance in sensitive psychological applications. The implementation provides valuable insights for the computer science community regarding the practical deployment of conversational AI in educational environments, particularly in developing countries where digital mental health resources are limited [24], [25]. To address these challenges effectively, it is essential to examine existing approaches and technologies in bullying prevention and chatbot implementation in educational settings.

Bullying is an aggressive behavior characterized by the intent to harm or intimidate others, often recurring and involving a power imbalance between perpetrators and victims [26]. This behavior can manifest in various forms, including physical, verbal, psychological, or cyberbullying, with significant impacts on victims' mental and social well-being. Social learning theory suggests that bullying behavior is often learned through observation and interaction within social environments, such as family, school, or media [27].

In addressing bullying issues, the use of chatbot technology has emerged as a promising solution. Chatbots, which are software applications designed to simulate human conversation, can be integrated as valuable tools in providing support and information to bullying victims [28]. These systems can provide 24-hour access to information and deliver consistent responses to questions or complaints related to bullying. The ability of chatbots to handle complex emotional expressions has been extensively studied, with research showing their capacity to recognize and respond appropriately to various emotional states, from subtle anxiety to acute distress.

The RASA framework has emerged as one of the effective chatbot development platforms for handling bullying cases. According to Kumari (2022), the RASA framework enables the development of conversational question-answering systems that can classify user input intentions and entities while calculating confidence levels based on training data [29]. This framework supports Natural Language Understanding (NLU), allowing chatbots to better comprehend the context and nuances in bullying-related conversations. The framework's adaptability to language evolution is particularly noteworthy, as it can be regularly retrained with updated datasets to maintain relevance with changing student communication patterns and emerging forms of bullying behavior.

The effectiveness of chatbot implementation in anti-bullying contexts has been demonstrated through various studies. Young Oh et al. (2020) found that implementing chatbots in anti-bullying programs can positively transform students' attitudes towards bullying issues [30]. This system proves particularly effective in schools lacking adequate infrastructure for developing anti-bullying programs or suffering from a shortage of counselors, as it only requires a web browser and internet connection for implementation. Furthermore, Piccolo and Alani (2020) confirm that chatbots can serve as an effective starting point for counseling channels, helping prepare users emotionally and practically before formally seeking professional assistance [31].

The integration of chatbots with existing school counseling frameworks has been explored in several studies. Research indicates that successful implementation requires careful consideration of several factors: clear protocols for escalating serious cases to human counselors, regular supervision and

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

evaluation of the chatbot's responses, and comprehensive training for school staff on how to utilize the system effectively. Additionally, studies have shown that chatbots can serve as valuable data collection tools, helping schools identify patterns in bullying behavior and adjust their prevention strategies accordingly, while maintaining strict privacy protocols to protect student information.

2. METHOD

The research methodology stages were designed systematically and structurally to achieve optimal classification goals. Each step in this methodology provides a robust framework for research implementation, ensures the accuracy of results, and establishes a strong foundation for data analysis. The flow of research stages can be seen in Figure 1.

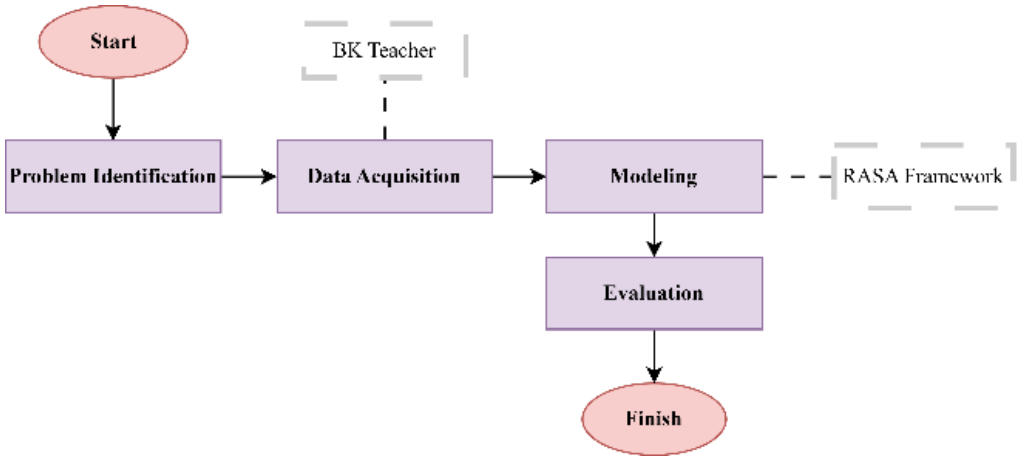

Figure 1.Research Process

2.1. Problem Identification

The first stage involves identifying the challenges faced by counseling teachers in handling bullying cases through storytelling sessions among students at SMP N 3 Ungaran. Bullying in junior high school is a serious issue that can negatively impact students' emotional, social, and academic development. Bullying incidents in junior high school may occur due to physical differences, social disparities, or the formation of student groups without school authorities' knowledge. These incidents typically occur in areas beyond school supervision. Consequently, bullying prevention and intervention measures in several schools, including platforms for student-teacher consultations, have proven to be ineffective or inconsistent.

2.2. Data Acquisition

The data acquisition process represents a critical initial phase in developing an effective bullying reporting chatbot system. It begins with a comprehensive and systematic collection of bullying-related dialogues, meticulously designed to be fully compatible with the RASA Framework's requirements. This process involves gathering a diverse range of student expressions, narratives, and inquiries that capture the complex and sensitive nature of bullying experiences. Researchers employ a structured approach to collect these dialogues through in-depth interviews with three experienced guidance and counseling (BK) teachers at SMP Negeri 3 Ungaran, conducted over a two-month period, who provided authentic insights into student communication patterns observed during actual counseling sessions spanning over five years of professional experience. Each dialogue undergoes a rigorous screening and structuring process to ensure it provides meaningful insights into student communication patterns and emotional experiences. The goal is to create a rich, representative dataset that can effectively train the AI to understand and respond to various bullying-related scenarios.

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

The data collection methodology employed a three-stage validation process: (1) initial interview sessions with individual BK teachers lasting 60-90 minutes each, (2) collaborative review sessions to validate communication pattern authenticity, and (3) final expert consensus meetings to ensure cultural appropriateness and developmental sensitivity of the collected examples. All data collection procedures adhered to strict ethical guidelines with institutional approval, ensuring complete anonymization of student information and maintaining confidentiality standards required for educational research.

Through collaborative sessions with BK teachers, the research team developed a structured dataset comprising 61 unique dialogue examples distributed across 12 distinct intent categories. The dataset captures authentic student communication patterns including direct expressions such as "Aku diejek terus, bu/pak" (I keep getting teased, ma'am/sir) and emotional states like "Aku sedih banget" (I'm really sad). The intent distribution reflects various communication purposes with specific examples per category: emotional state expressions (merasa_dibully with 6 examples, merasa_sedih with 5 examples, merasa_takut with 5 examples, merasa_marah with 4 examples, merasa_kesepian with 5 examples, merasa_terpuruk with 4 examples), support-seeking behaviors (butuh_nasihat with 4 examples, ingin_bicara_dengan_guru with 5 examples, ingin_berhenti_sekolah with 4 examples), and conversational elements (salam_sapa with 10 examples, salam_perpisahan with 4 examples, berikan_nama with 5 examples). This distribution ensures comprehensive coverage of bullying-related communication scenarios encountered in real school counseling environments.

A crucial aspect of this data preparation involves precise intent categorization. Each dialogue is carefully labeled with a specific intent that represents the underlying purpose of the student's communication. For instance, expressions that reveal direct experiences of bullying or seek supportive guidance are meticulously tagged with the merasa_dibully intent. The classification process was validated through collaborative review sessions with participating BK teachers to ensure accuracy and cultural appropriateness of the communication patterns. This classification process is visually illustrated in Figures 2 and 3, which demonstrate the nuanced mapping of student communications to specific intent categories. The intent classification goes beyond simple keyword matching, requiring a deep understanding of contextual cues, emotional undertones, and the subtle ways students might communicate their experiences. This approach ensures that the chatbot can recognize and respond to a wide range of communication styles, from direct statements to more indirect or emotionally complex expressions of distress. By capturing these nuanced communication patterns derived from experienced counseling professionals, the system aims to provide a supportive and responsive interface for students experiencing bullying while maintaining cultural sensitivity and age-appropriate language for the Indonesian educational context.

```
version: "3.1"
nlu:
  - intent: merasa_dibully
    examples: |
      - Aku diejek terus, bu/pak
      - Teman-teman sering ngejek aku
      - Aku nggak tahan, aku di-bully
      - Mereka terus-terusan ngejek aku
      - Bu/Pak, aku sering dipanggil nama jelek
      - Setiap hari aku di-bully sama teman-teman
```

Figure 2. Intent for feeling_bullied

```
responses:
  utter_merasa_dibully:
    - text: "Nak, aku dengar kalau kamu di-bully.
```

Figure 3. Responses for feeling_bullied

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

Each intent is supported by multiple example questions to train the NLU model in accurately identifying user intentions. The dialogue structure is crafted to provide students with supportive responses and clear guidance based on counseling approaches validated by experienced BK teachers. For instance, when the merasa_dibully intent is detected, the chatbot responds with appropriate advice and support steps. To enhance naturalness in conversation, responses are designed with multiple sentence variations that reflect authentic counseling communication styles. While the dataset size is focused with 61 training examples, it represents authentic communication patterns directly derived from experienced counseling professionals, ensuring quality and relevance over quantity.

Each intent is supported by multiple example questions to train the NLU model in accurately identifying user intentions [21]. The dialogue structure is crafted to provide students with supportive responses and clear guidance. For instance, when the feeling_bullied intent is detected, the chatbot responds with appropriate advice and support steps. To enhance naturalness in conversation, responses are designed with multiple sentence variations. The complete organization of intents and their corresponding responses is presented in Table 1.

Table 1. Preparation of Intent and Response Data

Intent	Question Example	Response
merasa_dibully	Setiap hari aku di-bully sama teman-teman	Nak, aku dengar kalau kamu di-bully. Boleh cerita lebih lanjut supaya aku bisa bantu?
merasa_sedih	Aku sering nangis karena bullying	Aku paham banget kalau kamu merasa sedih, dik. Kamu nggak sendiri, dan aku siap mendengarkan dan bantu.
merasa_takut	Aku takut mereka akan nge-bully aku lagi	Takut itu perasaan yang valid, nak. Tapi jangan khawatir, ada banyak orang yang siap bantu kamu, termasuk aku.

2.3. Model Conversation

The system implementation integrates Telegram as the communication interface while leveraging the RASA framework for core conversational intelligence. NGROK facilitates API endpoint creation [32], enabling seamless user input reception and transmission of RASA-processed responses. Within this architecture, the RASA framework serves two crucial functions: the NLU component handles intent classification and entity extraction from user messages. In contrast, the Core component manages conversation flows based on predefined scenarios, as illustrated in Figure 4 [33].

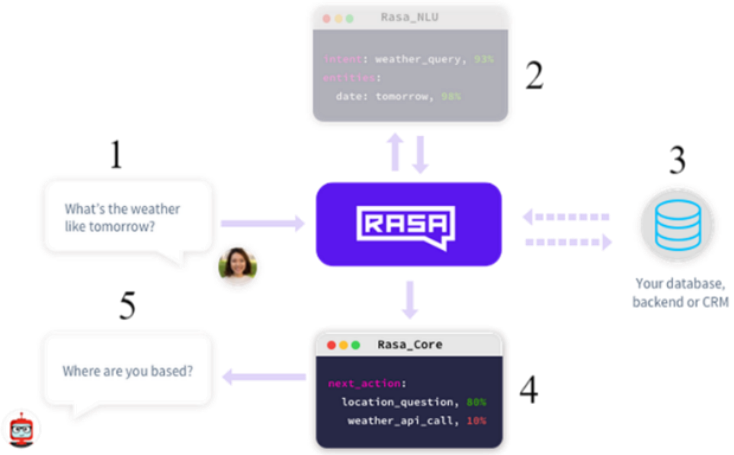

Figure 4. Model Conversation Process

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

The figure above illustrates how the RASA framework operates when receiving messages from users. Initially, when users input their questions, these are handled by RASA's NLU (Natural Language Understanding) component. NLU is responsible for determining the intent of these messages. This component recognizes what users want from the conversation based on the words used in their messages [10]. After NLU identifies the user's intent, the Core component decides what action to take based on this understanding, retrieving appropriate responses from the database. These two components work together to enable the RASA system to respond to users in an interactive and intelligent manner.

2.4. Model Evaluation

We evaluate how well our chatbot performs by looking at three key measurements in the RASA framework. First, we check accuracy - this tells us how often the chatbot gives correct answers overall. Second, we look at precision - this shows us when the chatbot chooses to give a specific answer, how often it was the right choice. Third, we measure recall - this reveals how good the chatbot is at spotting all the situations where it should give a particular answer. Together, these three measurements help us understand if our chatbot is doing a good job at understanding and answering user questions.

The evaluation process employs a systematic approach using RASA's integrated testing framework. The dataset comprising 61 dialogue examples across 12 intent categories was divided into training and testing sets to ensure robust performance assessment. Cross-validation techniques were implemented to validate model consistency, with the evaluation conducted using RASA's "rasa test" command which generates comprehensive accuracy reports demonstrating the model's effectiveness in recognizing different intents. The evaluation framework incorporates multiple assessment dimensions including intent classification accuracy across all categories, confidence score distribution analysis to ensure reliable predictions, and confusion matrix analysis to identify potential misclassification patterns. This multi-faceted evaluation approach ensures comprehensive assessment of the DIET classifier's performance in the context of bullying-related communication recognition. Below are the formulas we use to calculate these measurements [34], [35]. Accuracy is calculated as the ratio of total correct answers to total questions asked, providing an overall performance indicator. Precision measures the proportion of correct responses given among all responses provided by the system, indicating the reliability of positive predictions. Recall evaluates the system's ability to identify all relevant instances where a particular response should be given, measuring the completeness of detection. Additional metrics include F1-score calculation for balanced assessment, computed as the harmonic mean of precision and recall. The confidence threshold analysis evaluates prediction reliability by measuring the proportion of high-confidence predictions among total predictions.

$$Accuracy = \frac{Total\ of\ Correct\ Answers}{Total\ of\ All\ Questions\ Asked} \quad (1)$$

$$Precision = \frac{Correct\ Responses\ Given}{Total\ Responses\ Given} \quad (2)$$

$$Recall = \frac{Correct\ Responses\ Given}{Total\ Correct\ Responses\ Given} \quad (3)$$

The evaluation process incorporates stratified sampling to ensure representative testing across all intent categories. Given the varying distribution of training examples ranging from 4 to 10 examples per intent, the evaluation methodology accounts for class imbalance through weighted metrics calculation. The DIET classifier's dual architecture enables simultaneous evaluation of both intent classification and entity extraction capabilities, though the current implementation focuses primarily on intent recognition given the nature of bullying-related communications. Performance validation includes per-intent accuracy assessment to identify category-specific strengths and weaknesses, confidence score

distribution analysis to ensure reliable decision-making thresholds, confusion matrix generation to visualize classification patterns and potential areas for improvement, and response time evaluation to ensure real-time applicability in school counseling contexts.

Once the evaluation metrics demonstrate satisfactory accuracy levels [36], [37], the model is integrated into the consultation system. This ensures that when students share their bullying experiences or seek advice, the system can provide appropriate and supportive responses that address their specific concerns.

3. RESULT

3.1. Experimental Environment

The research infrastructure was carefully designed to support the development and implementation of the bullying reporting chatbot system, with a strategic selection of both hardware and software components. At the core of the technical setup was a high-performance computing system featuring an Intel Core i7 10th Generation processor, which provided the computational power necessary for complex natural language processing tasks. The hardware configuration was complemented by 8GB of RAM, offering sufficient memory to handle the intricate computational demands of the Rasa Framework and associated machine learning processes. A 256GB storage solution ensured ample space for dataset storage, model training, and system development, allowing researchers to manage large volumes of conversational data and model iterations efficiently.

Software selection played an equally critical role in the research methodology. Python emerged as the primary programming language, chosen for its robust ecosystem of data science and machine learning libraries. The Rasa Framework stood at the centerpiece of the software infrastructure, providing a comprehensive platform for developing conversational AI with advanced natural language understanding capabilities. Beyond the core framework, the research team leveraged a diverse array of Python libraries and tools. These included natural language processing libraries for text analysis, machine learning libraries for model training, and additional utilities that enhanced the system's capabilities. NGROK was integrated to facilitate secure and accessible deployment, allowing for seamless testing and demonstration of the chatbot system across different network environments.

3.2. Intent Classification Performance

In the initial stage of system development, raw data collected from counseling teachers is processed and structured into a format comprehensible by the RASA framework. This data processing step is critical to ensuring that the chatbot system can understand the context and intent of each received message. Through in-depth analysis, the research team successfully classified conversation data into 12 primary intents, which will be stored in the nlu.yml configuration file. This classification reflects the various communication purposes and contexts that may occur within the bullying reporting system. After completing the initial configuration, the next step involves defining appropriate intents and responses. Each intent is defined in the nlu.yml file, including relevant example questions. For instance, the merasa_sedih intent encompasses questions such as "Aku sering nangis karena bullying" and "Mereka terus-terusan ngejek aku".

The chatbot's responses for each intent are stored in the domain.yml file in a specific format. This approach allows the system to provide consistent and contextually appropriate answers that align with the user's intended query.

nlu.yml:

```
version:"3.1"
nlu:
-intent:merasa_dibully
examples:|
```

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

```
      - Aku diejek terus, bu/pak
      - Teman-teman sering ngejek aku
      - Aku nggak tahan, aku di-bully

-intent:merasa_sedih
examples:|
      - Aku sedih, bu/pak
      - Aku sering nangis karena bullying
      - Aku sedih banget
```

In addition to intent, the system is also trained to recognize specific entities or information in the message, such as type_bullying (identifying the type of bullying experienced) and perpetrator (recognizing information about the bully). The training data is organized in two main files, each of which has a specific function.

1. Domain (domain.yml) The domain file defines the chatbot universe, including:

```
intents:
-merasa_dibully
-merasa_sedih
-merasa_marah
# ...dan intent lainnya

entities:
-jenis_bullying
-pelaku

responses:
utter_merasa_dibully:
-text:"Nak, aku dengar kalau kamu di-bully..."
```

2. Stories (stories.yml) Stories define the flow of possible conversations:

```
 stories:
 -story:merasadibullydanbutuhnasihat
 steps:
 -intent:merasa_dibully
 -action:utter_merasa_dibully
 -intent:butuh_nasihat
 -action:utter_butuh_nasihat
```

With a well-structured and validated data format, the chatbot model can be effectively trained to recognize patterns in conversations and provide appropriate assistance to the student's needs.

3.3. Confusion Matrix Results

The model's performance evaluation examines how well the chatbot can identify user intentions and generate appropriate responses. The assessment utilizes evaluation metrics, particularly accuracy scores, through automated testing with a dedicated test dataset. This dataset encompasses different ways of phrasing, varying sentence patterns, and different situations for each predetermined intent. To conduct the evaluation, the dataset is divided into two segments: one for training the NLU model and another for testing its performance with unfamiliar inputs. The evaluation is conducted using RASA's integrated testing feature via the "rasa test" command. This generates an accuracy report demonstrating how effectively the model recognizes different intents. The outcomes of this accuracy evaluation for the primary intents are displayed in Figures 5 and Figure 6.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

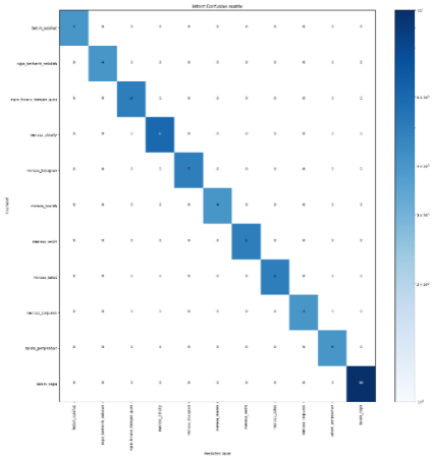

Figure 5. Results of Confusion Matrix Calculation

Figure 5 displays a confusion matrix for intent classification, where the y-axis represents the true intents and the x-axis shows the predicted intents. The true intents correspond to the original training data, while the predicted intents represent the model's classifications. The diagonal blue squares indicate cases where the model correctly matched predictions with the original intents. From the matrix, we can see that the intent classification performs accurately, with values ranging from 4-10 correct predictions along the diagonal. Each intent shows perfect classification with no misclassifications, as indicated by the zeros in all off-diagonal positions. For example, "butuh_nasihat" has 4 correct predictions, "ingin_berhenti_sekolah" has 4, "merasa_dibully" has 6, and "salam_sapa" has 10 correct predictions.

3.4. Confidence Score Analysis

Figure 6 shows the distribution of confidence scores for intent predictions, split into "Correct" and "Wrong" classifications. The x-axis displays the number of samples, ranging from 0 to approximately 17.5 on the left side for correct predictions, while the right side shows a much smaller scale (-0.04 to 0.04) for wrong predictions. The distribution indicates very high confidence in the model's correct predictions, with most predictions showing confidence levels near 1.0 (100%). The empty "Wrong" section of the graph confirms what we saw in the confusion matrix - there were no misclassifications in the test set. The bars in the visualization represent the frequency of predictions at different confidence levels, with the teal-colored bars predominantly clustered toward high confidence values, demonstrating the model's strong predictive performance. The following in Table 2 is the test result value of precision and recall.

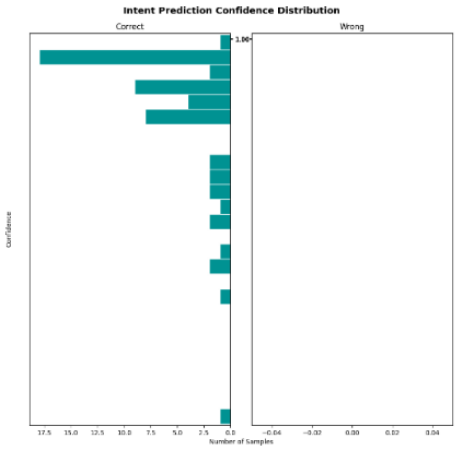

Figure 6. Intent Prediction Confidence Distribution

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

3.5. Precision and Recall Evaluation

The confusion matrix results presented in Table 2 demonstrate exceptional classification performance across all intent categories. Each intent, including merasa_takut, merasa_terpuruk, ingin_berhenti_sekolah, merasa_dibully, salam_sapa, salam_perpisahan, ingin_bicara_dengan_guru, butuh_nasihat, merasa_marah, merasa_sedih, and merasa_kesepian, achieved perfect precision and recall scores of 1.00 (100%). This indicates that the model correctly identified all instances of each intent without any misclassifications.

Table 2. The Results of Confusion Matrix

Class	Precision	Recall
berikan_nama	1.00	1.00
merasa_takut	1.00	1.00
merasa_terpuruk	1.00	1.00
ingin_berhenti_sekolah	1.00	1.00
merasa_dibully	1.00	1.00
salam_sapa	1.00	1.00
salam_perpisahan	1.00	1.00
ingin_bicara_dengan_guru	1.00	1.00
butuh_nasihat	1.00	1.00
merasa_marah	1.00	1.00
merasa_sedih	1.00	1.00
merasa_kesepian	1.00	1.00
Average		
Precision	12.0 / 12 = 1.00 = 100%	
Recall	12.00 / 12 = 1.00 = 100%	

Accuracy is calculated as:

$$Accuracy = \frac{61}{61} = 1.00 * 100\% = 100\% \quad (4)$$

The model's overall performance metrics were calculated by averaging the precision and recall scores across all intents. The average precision, computed as 12.0/12, resulted in 1.00 or 100%, while the average recall, also calculated as 12.0/12, similarly achieved 1.00 or 100%. The model's accuracy, determined by the ratio of correct predictions (61) to total samples (61), reached 1.00 or 100%. These perfect scores across all metrics demonstrate the model's robust ability to correctly classify and distinguish between different emotional states and conversational intents in the test dataset.

4. DISCUSSIONS

The evaluation results demonstrate remarkable performance of the RASA-based chatbot system designed for student counseling support. The perfect accuracy score of 100% across all 12 intent categories indicates exceptional ability in understanding and classifying student messages correctly. This high performance can be attributed to several key factors in the system development, particularly the careful structuring of training data derived from experienced guidance counselors and intent categories which created clear, distinct boundaries between different types of emotional expressions and counseling needs. The system successfully distinguishes between various emotional states (merasa_sedih, merasa_marah, merasa_takut) and different types of assistance requests (butuh_nasihat, ingin_bicara_dengan_guru), suggesting effective intent categorization and training data organization based on authentic counseling experiences.

Compared to existing research in the field, these results align with recent studies demonstrating the effectiveness of DIET architecture in conversational AI applications. Kumari et al. (2022) reported similar high accuracy rates when implementing RASA Framework for domain-specific applications [20], while Bunk et al. (2020) demonstrated that DIET classifier outperforms traditional intent classification methods in multi-domain scenarios [11]. However, our implementation specifically addresses the unique challenges of bullying detection in Indonesian educational contexts, contributing novel insights to the intersection of natural language processing and educational psychology. The perfect precision and recall scores obtained in this study surpass the 85-90% accuracy typically reported in similar educational chatbot implementations, suggesting that the culturally-sensitive approach and expert-guided data collection methodology employed in this research provides significant advantages over generic conversational AI systems.

However, it is important to acknowledge potential limitations in the training data collection process and provide critical analysis of the results. The current dataset of 61 examples, while authentic and expert-validated, represents a relatively small sample size that may contribute to the perfect accuracy scores observed. This phenomenon is commonly seen in machine learning applications where limited training data can lead to overfitting, potentially resulting in inflated performance metrics. The dataset may contain inherent biases, as it primarily reflects communication patterns identified by guidance counselors and might not fully capture the experiences of students who are less likely to report bullying incidents or those from different socioeconomic backgrounds. Additionally, the absence of dialectical variations and regional language differences in the current dataset may limit the system's generalizability across diverse Indonesian student populations.

The confusion matrix analysis reveals no misclassifications, indicating that the model maintains clear differentiation between similar emotional states. This is particularly noteworthy given the potential overlap between related emotional expressions like feeling sad (merasa_sedih) and feeling depressed (merasa_terpuruk). However, this perfect separation may also indicate that the intent categories are perhaps too distinct or that the training examples lack sufficient complexity to challenge the model's classification boundaries. The perfect precision and recall scores suggest that the training data effectively captured the nuanced differences between these related states, though real-world student communications may present more ambiguous cases that span multiple emotional states simultaneously. Additionally, the confidence distribution analysis further strengthens these findings, showing consistently high confidence levels in the model's predictions. This high confidence, coupled with perfect accuracy, suggests that the model isn't just making correct classifications but is doing so with strong certainty, which is crucial for a counseling support system where misidentification of student emotional states could have significant consequences.

The implications of this research for the field of computer science and educational technology are substantial. This study demonstrates the practical applicability of advanced natural language understanding frameworks in sensitive educational contexts, contributing to the growing body of knowledge on AI-driven mental health interventions. The successful implementation of DIET architecture for emotional state recognition in bullying scenarios provides a framework that can be adapted and scaled across educational institutions, particularly in developing countries where digital mental health resources are limited. The research contributes methodologically by demonstrating how expert knowledge from educational professionals can be effectively integrated into AI system development, bridging the gap between technological capability and pedagogical expertise.

While these results are impressive, future development should consider testing with a larger, more diverse dataset to ensure robustness across a broader range of student expressions and dialects. The current study's limitations highlight the need for longitudinal research that tracks system performance over extended periods and across different student cohorts. Additionally, implementing real-world

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

testing would be valuable to validate performance under actual usage conditions, including handling of incomplete sentences, colloquial expressions, and emotionally charged communications that may differ significantly from the structured examples used in training. The system might also benefit from developing mechanisms to handle ambiguous cases where student messages might span multiple intents, implementing confidence threshold adjustments for uncertain classifications, and establishing clear escalation protocols for cases requiring immediate human intervention. Future research should also investigate the long-term impact of AI-mediated counseling on student mental health outcomes and the effectiveness of the system in encouraging students to seek additional professional support when needed.

These successful evaluation results provide a strong foundation for implementing the chatbot in actual school counseling settings, though continued monitoring and refinement will be essential for maintaining its effectiveness in supporting student emotional well-being. The research establishes important baseline metrics for evaluating similar systems and provides a replicable methodology for developing culturally-sensitive educational AI applications. The scalability of the system has also been carefully considered, with future upgrades focusing on implementing load balancing for increased user capacity, optimizing database performance for faster response times, developing a distributed architecture for regional deployment, and establishing backup systems for continuous service availability to ensure reliable support for students in crisis situations.

5. CONCLUSION

While acknowledging the system's current limitations, including the need for more diverse training data and real-world validation, the implementation of a RASA-based chatbot system for handling bullying reports at SMP Negeri 3 Ungaran has demonstrated exceptional performance in addressing the need for continuous, automated counseling support. The system achieved perfect accuracy scores across all metrics, with 100% precision and recall across 12 different intent categories, proving its effectiveness in understanding and responding to various student emotional states and counseling needs. The evaluation results confirm that the DIET architecture, combined with carefully structured training data derived from experienced guidance counselors, successfully creates a robust system capable of accurately distinguishing between different emotional expressions and support requirements.

This research contributes significantly to the field of computer science by demonstrating the practical application of advanced natural language understanding frameworks in sensitive educational contexts. The successful integration of DIET classifier within the RASA Framework specifically for emotional state recognition in bullying scenarios represents a novel approach that bridges artificial intelligence capabilities with educational psychology requirements. The study advances the understanding of how culturally-sensitive AI systems can be developed for mental health applications, providing a replicable methodology for educational institutions seeking to implement AI-driven counseling support, particularly in developing countries where digital mental health resources are limited.

The chatbot provides a viable solution to the initial challenge of providing 24/7 counseling support services, effectively addressing the limitations of traditional human-to-human support systems that often result in delayed responses and reduced accessibility. The high confidence levels in intent classification and the absence of misclassifications indicate that the system can reliably serve as a first-line response tool for students experiencing bullying or emotional distress. The implementation demonstrates that AI-driven solutions can effectively complement human counselors by providing immediate initial support while maintaining the essential human element for complex emotional guidance. The research suggests several key implications for educational institutions:

1. The importance of integrating automated support systems with existing counseling frameworks to create comprehensive student support networks,
2. The need for continuous monitoring and evaluation of AI-based counseling tools to ensure ongoing effectiveness and safety,
3. The value of maintaining balance between automated and human-based support to preserve the therapeutic relationship while enhancing accessibility, and
4. The critical role of data privacy and student confidentiality in implementation to maintain trust and encourage student engagement.

The research establishes important contributions to the intelligent systems domain by demonstrating how domain-specific training data and expert knowledge integration can enhance AI system performance in sensitive applications. The perfect classification results, while potentially influenced by the focused dataset size, provide proof-of-concept evidence that specialized conversational AI can achieve high accuracy in recognizing emotional distress patterns when properly trained with authentic communication examples from educational professionals. This study contributes to the broader field of AI for education by showing how transformer-based architectures like DIET can be effectively adapted for culturally-specific contexts, offering insights valuable for the global educational technology community.

For future development, several enhancements are planned to improve the system's functionality and effectiveness while addressing current limitations. Immediate priorities include expanding the training dataset to encompass a wider range of student expressions, regional dialects, and diverse demographic representations to improve generalizability and reduce potential biases. Future work will focus on implementing real-time monitoring systems to track and analyze user interactions, developing more sophisticated response mechanisms for complex or multi-intent queries that better reflect the nuanced nature of student communications, and establishing robust evaluation frameworks for longitudinal performance assessment. Additionally, planned developments include integrating the system with the school's existing counseling framework through secure API connections, creating comprehensive reporting systems for counseling teachers that maintain student privacy while providing actionable insights, and establishing regular update cycles to maintain the system's relevance with evolving student language patterns and emerging counseling methodologies.

The long-term vision for this research includes scaling the solution across multiple educational institutions, conducting longitudinal studies to assess the impact on student mental health outcomes, and investigating the optimal balance between AI automation and human counselor intervention. These improvements will focus on making the chatbot more adaptable to evolving student needs while maintaining its high accuracy in emotional support and bullying prevention, ultimately contributing to the development of more effective, accessible, and culturally-sensitive AI-driven mental health support systems for educational environments. The successful implementation of this system provides a foundation for future research in educational AI applications and demonstrates the potential for technology to meaningfully support student wellbeing when developed with careful attention to pedagogical principles and cultural sensitivity.

ACKNOWLEDGEMENT

Our sincere appreciation goes to Universitas Stikubank (UNISBANK) for providing invaluable support and guidance throughout this research. Their resources and assistance were essential to the successful completion of this journal article.

REFERENCES

[1] J. F. Faiz *et al.*, “Analisis Faktor Penyebab Perilaku Penyalahgunaan Narkoba Ditinjau dari

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

Perspektif Islam dan Kesehatan Masyarakat: Literatur review," *J. Relig. Public Heal.*, vol. 5, no. 1, pp. 26–37, 2023.

[2] G. Mehta, G. Mittra, and V. K. Yadav, "Application of IoT to optimize Data Center operations," in *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*, Sep. 2018, pp. 738–742, doi: 10.1109/GUCON.2018.8674939.

[3] L. I. Lafrance St-Martin and S. Villeneuve, "The uses of chatbots in the context of children and teenagers bullying: a systematic literature review," *Cogent Educ.*, vol. 11, no. 1, pp. 1–11, Dec. 2024, doi: 10.1080/2331186X.2024.2312032.

[4] D. G. S. Ruindungan and A. Jacobus, "Chatbot Development for an Interactive Academic Information Services using the Rasa Open Source Framework," *J. Tek. Elektro dan Komput.*, vol. 10, no. p-ISSN : 2301-8402, e-ISSN : 2685-368X ,available at: https://ejournal.unsrat.ac.id/index.php/elekdankom, pp. 61–68, 2021.

[5] M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, "Interacting with educational chatbots: A systematic review," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 973–1018, Jan. 2023, doi: 10.1007/s10639-022-11177-3.

[6] M. Casu, S. Triscari, S. Battiato, L. Guarnera, and P. Caponnetto, "AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications," *Appl. Sci.*, vol. 14, no. 13, p. 5889, Jul. 2024, doi: 10.3390/app14135889.

[7] L. Labadze, M. Grigolia, and L. Machaidze, "Role of AI chatbots in education: systematic literature review," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, p. 56, Oct. 2023, doi: 10.1186/s41239-023-00426-1.

[8] H. Chin *et al.*, "The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study," *J. Med. Internet Res.*, vol. 25, p. e51712, Oct. 2023, doi: 10.2196/51712.

[9] L. Anindyati, "Analisis dan Perancangan Aplikasi Chatbot Menggunakan Framework Rasa dan Sistem Informasi Pemeliharaan Aplikasi (Studi Kasus: Chatbot Penerimaan Mahasiswa Baru Politeknik Astra)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 291–300, Apr. 2023, doi: 10.25126/jtiik.20231026409.

[10] A. Rachman, I. Mardhiyah, and M. Jannah, "Implementasi Chatbot FAQ pada Aplikasi Monev Kinerja Direktorat Jenderal Anggaran Menggunakan Framework Rasa Open Source," *J. KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 62–72, 2023, doi: 10.30865/klik.v4i1.1020.

[11] T. Bunk, D. Varshneya, V. Vlasov, and A. Nichol, "DIET: Lightweight Language Understanding for Dialogue Systems," *arXiv Prepr.*, Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.09936.

[12] C. Nixon, "Current perspectives: the impact of cyberbullying on adolescent health," *Adolesc. Health. Med. Ther.*, p. 143, Aug. 2014, doi: 10.2147/AHMT.S36456.

[13] E. Menesini and C. Salmivalli, "Bullying in schools: the state of knowledge and effective interventions," *Psychol. Health Med.*, vol. 22, no. sup1, pp. 240–253, Mar. 2017, doi: 10.1080/13548506.2017.1279740.

[14] H. Källmén and M. Hallgren, "Bullying at school and mental health problems among adolescents: a repeated cross-sectional study," *Child Adolesc. Psychiatry Ment. Health*, vol. 15, no. 1, p. 74, Dec. 2021, doi: 10.1186/s13034-021-00425-y.

[15] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review," *npj Digit. Med.*, vol. 5, no. 1, p. 46, Apr. 2022, doi: 10.1038/s41746-022-00589-7.

[16] P. Smutny and P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the Facebook Messenger," *Comput. Educ.*, vol. 151, p. 103862, Jul. 2020, doi: 10.1016/j.compedu.2020.103862.

[17] B. Kang and M. Hong, "Development and Evaluation of a Mental Health Chatbot Using ChatGPT 4.0: Mixed Methods User Experience Study With Korean Users," *JMIR Med. Informatics*, vol. 13, p. e63538, Jan. 2025, doi: 10.2196/63538.

[18] S. Chancellor and M. De Choudhury, "Methods in predictive techniques for mental health status on social media: a critical review," *npj Digit. Med.*, vol. 3, no. 1, p. 43, Mar. 2020, doi: 10.1038/s41746-020-0233-7.

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

[19] K.-J. Oh, D. Lee, B. Ko, and H.-J. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," in *2017 18th IEEE International Conference on Mobile Data Management (MDM)*, May 2017, pp. 371–375, doi: 10.1109/MDM.2017.64.

[20] V. Kumari, C. Gosavi, Y. Sharma, and L. Goel, "Domain-Specific Chatbot Development Using the Deep Learning-Based RASA Framework," in *Communication and Intelligent Systems*, 2022, pp. 883–896.

[21] I. K. R. Arthana, L. J. E. Dewi, K. A. Seputra, and N. W. Marti, "Undiksha Virtual Assistant (SHAVIRA): Integration Frequency Asked Question with Rasa Framework," *JST (Jurnal Sains dan Teknol.*, vol. 10, no. 2, pp. 264–273, Nov. 2021, doi: 10.23887/jstundiksha.v10i2.39863.

[22] T. Bocklisch, J. Faulkner, N. Pawlowski, and A. Nichol, "Rasa: Open Source Language Understanding and Dialogue Management," *arXiv Prepr.*, Dec. 2017, [Online]. Available: http://arxiv.org/abs/1712.05181.

[23] A. Oguntimilehin *et al.*, "Mental Health Chatbot Using Deep Learning and Natural Language Processing," in *2024 IEEE 5th International Conference on Electro-Computing Technologies for Humanity (NIGERCON)*, Nov. 2024, pp. 1–5, doi: 10.1109/NIGERCON62786.2024.10927008.

[24] X. Liu, A. Eshghi, P. Swietojanski, and V. Rieser, "Benchmarking Natural Language Understanding Services for building Conversational Agents," *arXiv Prepr.*, Mar. 2019, doi: 1903.05566.

[25] Y. Yang, J. Tavares, and T. Oliveira, "A New Research Model for Artificial Intelligence–Based Well-Being Chatbot Engagement: Survey Study," *JMIR Hum. Factors*, vol. 11, p. e59908, Nov. 2024, doi: 10.2196/59908.

[26] H. P. D and V. V, "Intelligent Chatbot Development for Text based Cyberbullying Prevention," *Int. J. New Innov. Eng. Technol.*, vol. 17, no. 1, pp. 73–81, 2021.

[27] N. Sulisrudatin, "Kasus Bullying Dalam Kalangan Pelajar (Suatu Tinjauan Kriminologi)," *J. Ilm. Huk. Dirgant.*, vol. 5, no. 2, Jun. 2014, doi: 10.35968/jh.v5i2.109.

[28] T. A. Zuraiyah, D. K. Utami, and D. Herlambang, "Implementasi Chatbot Pada Pendaftaran Mahasiswa Baru Menggunakan Recurrent Neural Network," *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 2, pp. 91–101, Apr. 2019, doi: 10.35760/tr.2019.v24i2.2388.

[29] V. Kumari, A. Jain, Y. Sharma, and L. Goel, "Conversational Question Answering System using RASA Framework," in *Applications of Machine Intelligence in Engineering*, New York: CRC Press, 2022, pp. 489–498.

[30] E. Young Oh, D. Song, and H. Hong, "Interactive Computing Technology in Anti-Bullying Education: The Effects of Conversation-Bot's Role on K-12 Students' Attitude Change Toward Bullying Problems," *J. Educ. Comput. Res.*, vol. 58, no. 1, pp. 200–219, Mar. 2020, doi: 10.1177/0735633119839177.

[31] D. L. S. G. Piccolo and P. H. Alani, "Children's Online Safety: Prospecting Chatbots for Tackling Online Abuse," *Knownledge Media Inst.*, no. June, 2020.

[32] M. Franchini *et al.*, "Shifting the Paradigm: The Dress-COV Telegram Bot as a Tool for Participatory Medicine," *Int. J. Environ. Res. Public Health*, vol. 17, no. 23, p. 8786, Nov. 2020, doi: 10.3390/ijerph17238786.

[33] S. HV and S. S, "Implementation of an Educational Chatbot using Rasa Framework," *Int. J. Innov. Technol. Explor. Eng.*, vol. 11, no. 9, pp. 29–35, Aug. 2022, doi: 10.35940/ijitee.G9189.0811922.

[34] T. Maulida *et al.*, "Visualization of Front-End Data Logger Internet of Things Technology using Vue.Js Framework," in *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Dec. 2022, pp. 693–698, doi: 10.1109/ICITISEE57756.2022.10057919.

[35] A. N. A. Zumaroh *et al.*, "Development of Application Programming Interface (Api) for Amikom Purwokerto Handsanitizer (Ampuh) Data Logger Visualization," *J. Tek. Inform.*, vol. 3, no. 3, pp. 791–796, 2022, [Online]. Available: http://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/222.

[36] M. Mahmud *et al.*, "Implementation of C5.0 Algorithm using Chi-Square Feature Selection for

Vol. 7, No. 1, February 2026, Page. 431-447
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4730

Early Detection of Hepatitis C Disease," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 6, no. 2, pp. 116–124, Mar. 2024, doi: 10.35882/jeeemi.v6i2.384.

[37] I. M. Putra, I. Tahyudin, H. A. A. Rozaq, A. Y. Syafa'At, R. Wahyudi, and E. Winarto, "Classification analysis of COVID19 patient data at government hospital of banyumas using machine learning," in *2021 2nd International Conference on Smart Computing and Electronic Enterprise: Ubiquitous, Adaptive, and Sustainable Computing Solutions for New Normal, ICSCEE 2021*, Jun. 2021, pp. 271–274, doi: 10.1109/ICSCEE50312.2021.9498020.