

Mental Health Chatbot for Detecting Depression Symptoms Using Natural Language Processing and DASS-21

Fadilah Nuraini^{1*}, Muhammad Najamuddin Dwi Miharja², Abdul Halim Anshor³

^{1,2,3}*Department of Informatics Engineering, Faculty of Engineering, Pelita Bangsa University, Bekasi, Indonesia*

*Corresponding author email: fadilah.nuraini13@mhs.pelitabangsa.ac.id

Jurnal Teknologi is used only:

Received 30 May 2025; Revised 24 June 2025; Accepted 21 July 2025

ABSTRACT

The prevalence of depressive symptoms among university students continues to rise, driven by academic pressure, social isolation, and limited access to psychological support. Early detection and intervention remain critical challenges in mental health services. This study presents the design and implementation of an intelligent chatbot that integrates the Depression Anxiety Stress Scale-21 (DASS-21) with Natural Language Processing (NLP) techniques to enable non-clinical mental health screening. The chatbot processes user input through intent classification and text preprocessing pipelines to dynamically assess indicators of depression, anxiety, and stress. Utilizing a hybrid rule-based and machine learning architecture, the system provides a self-assessment interface that delivers personalized feedback based on the DASS-21 scoring rubric. Two models were evaluated: a TF-IDF-based Neural Network and a fine-tuned BERT model. The TF-IDF model achieved an accuracy of ninety-one percent with a weighted F1-score of 0.91, while the BERT model outperformed it with an accuracy of ninety-four percent and a weighted F1-score of 0.94. Notably, the BERT model demonstrated a recall of ninety-eight percent in identifying moderate depression cases. However, both models showed limitations in detecting mild depression due to data imbalance. The approach prioritizes usability, anonymity, and accessibility, key factors in promoting help-seeking behavior among young adults. The results demonstrate the potential of NLP-powered conversational agents as scalable, low-cost tools for early detection of mental health risks in academic environments.

Keywords: DASS-21, Natural Language Processing, Chatbot, Psychological Screening, Intent Classification.

Introduction

Depression is a common mental health disorder characterized by persistent sadness, loss of interest in previously enjoyable activities, and impaired daily functioning [1]. It affects individuals across all age groups, particularly adolescents and university students, and can be triggered by traumatic experiences, emotional stress, or substance abuse. When left untreated, depression may lead to significant negative outcomes, including decreased productivity, social dysfunction, and even suicidal ideation.

According to the World Health Organization (WHO), over 280 million people suffer from depression globally, and approximately 800,000 suicide cases are reported annually, with a large proportion involving young adults. In Indonesia, the National Police's Criminal Investigation Agency (Bareskrim) reported 431 suicide cases between January and May 2024, 51 of which involved students. These figures reflect the growing concern over academic pressure, economic insecurity, and the negative impact of social media use on students' mental well-being [2].

Furthermore, recent developments in digital mental health solutions have led to the emergence of self-assessment tools aimed at increasing awareness and early detection of psychological disorders. One of the most widely used instruments is the Depression Anxiety Stress Scale-21 (DASS-21), which provides a simple and practical self-report method for evaluating mental well-being. This tool has been validated for various populations, including university students, and is effective in identifying early signs of depression, anxiety, and stress [3].

Given the increasing demand for accessible mental health services, technology such as chatbots powered by Natural Language Processing (NLP) has shown promising potential. NLP allows machines to process, understand, and respond to human language naturally and empathetically. When integrated with DASS-21, these intelligent systems can offer interactive mental health screenings while maintaining anonymity and accessibility. This reduces the stigma surrounding mental health and encourages users to engage in self-reflection and seek professional help when needed [4][5].

In Indonesia, the urgency for such tools is amplified by concerning national data. A notable case in Malang, East Java, involved a ninth-semester university student (MAS, 24 years old) who tragically took his own life due to untreated depression. This tragedy highlights the need for more responsive, stigma-free, and scalable tools to detect and manage mental health issues among students.

To address these challenges, this study proposes the implementation of a chatbot that utilizes the DASS-21 scale through NLP methods. The goal is to provide an intelligent mental health support tool capable of assessing emotional states, delivering personalized responses, and guiding students toward professional mental health services. Through this system, it is expected that early detection and intervention can be achieved, ultimately reducing the impact of depression on student populations.

By leveraging technology that is both accessible and adaptable, this chatbot aims to

bridge the gap between mental health needs and available support, especially in underserved academic communities. It not only functions as a preliminary screening tool but also serves as an empathetic virtual companion that can encourage self-awareness, reduce stigma associated with seeking help, and facilitate early intervention through timely recommendations thereby aligning with broader public health goals to integrate mental wellness into digital platforms for increased reach and effectiveness.

The purpose of this study is to design and implement a chatbot system that integrates the DASS-21 scale and Natural Language Processing (NLP) techniques to provide accessible, non-clinical mental health screening specifically for university students. This system aims to support early detection of psychological distress, deliver empathetic and personalized feedback, and promote proactive mental well-being strategies within academic environments, ultimately reducing the stigma and barriers to seeking professional help [6].

The development of a depression detection chatbot based on DASS-21 and Natural Language Processing (NLP) requires integration of several foundational components and prior studies. This section outlines the key materials and previous works that support the research.

- a. DASS-21 as Diagnostic Material: The Depression Anxiety Stress Scale-21 (DASS-21) is a validated psychometric tool used globally to assess mental health conditions such as depression, anxiety, and stress. It consists of 21 questions evenly distributed across the three dimensions. Each response is scored on a Likert scale from 0 to 3 and provides insights into the user's emotional condition. This study adopts DASS-21 to standardize early screening through the chatbot interface [7].
- b. Natural Language Processing (NLP): NLP allows machines to interpret, analyze, and generate human language. In this research, NLP is used to process user inputs, identify emotional cues, and deliver relevant and empathetic

- responses. Techniques such as tokenization, TF-IDF vectorization, and intent classification are applied to ensure contextual understanding and appropriate reaction [8].
- c. Chatbot Infrastructure: The chatbot is designed to simulate human-like conversations and is trained using a dataset from Kaggle that includes mental health-related interactions. The dataset includes intents such as anxiety questions, depression affirmations, and therapy suggestions. This conversational base allows the chatbot to mirror real-world mental health consultations [9].

Table 1. State of the art of interactive map

Published			Development				
No	Publisher	Author	Year	Journal Title	Data	Tools	Results
1.	Journal of Advanced Technology and Application	Teddy Surya Gunawan, et. al.	2023	Development of Intelligent Telegram Chatbot Using Natural Language Processing	User text inputs used to train the chatbot for emotion recognition.	Telegram API and Natural Language Processing (NLP) algorithms.	The chatbot successfully interacted naturally, recognized user emotions, and responded empathetically.
2.	IEEE Access	Girija Attigeri, et. al.	2024	Advanced NLP Models for Technical University Chatbot Development and Conversational Assistant	Academic-related queries such as class schedules, exam information, and technical questions from university students and staff.	Advanced NLP models using deep learning algorithms and machine learning techniques.	The chatbot was able to understand complex academic conversations, respond quickly and accurately, and provide a more natural, context-aware, and human-like interaction.
3.	Digital Transformation Technology	Dody Indra Sumantiawan, et.al.	2024	Factors Affecting Depression in Students and the Role of Machine Learning in Mental Health Analysis	Student data related to social status, academic pressure, and campus environment.	Statistical analysis and Machine Learning algorithms, specifically Naive Bayes.	The study identified key depression predictors in students and highlighted the potential of automated systems for early mental health intervention in education.

Research related to the development of chatbots for mental health and academic assistance has gained increasing attention in recent years. Several studies have explored the integration of Natural Language Processing

(NLP) and Machine Learning (ML) in chatbot systems to improve user interaction, emotional understanding, and mental health support. A summary of selected relevant studies is presented in Table 1.

A study conducted by Teddy Surya Gunawan et al [4]. focuses on the development of a Telegram-based intelligent chatbot using NLP. The chatbot is designed to understand user input and respond empathetically by recognizing emotional cues. Utilizing Telegram API and NLP algorithms, this chatbot demonstrated effective natural interaction and emotional response generation.

Another study by Girija Atigeri et al., published in IEEE Access [10], introduced advanced NLP models to develop a chatbot capable of assisting university students with academic tasks. The chatbot leverages deep learning to comprehend complex queries related to class schedules, examinations, and technical issues, providing a more human-like and context-aware conversational experience.

Furthermore, Dody Indra Sumantrawan et al [11]. examined the factors influencing depression in students using a Machine Learning approach. By applying statistical analysis and Naive Bayes classification, the research identified key predictors of student depression, emphasizing the importance of predictive systems in educational institutions for early intervention and mental health monitoring.

These studies collectively highlight the evolving role of intelligent chatbots in supporting mental health and educational communication, showing how the combination of NLP, ML, and user data can lead to responsive, adaptive, and context-sensitive systems.

The integration of Neural Networks (NN) and Term Frequency-Inverse Document Frequency (TF-IDF) plays a critical role in developing intelligent mental health chatbots [12]. Neural Networks, inspired by the structure of the human brain, are composed of interconnected artificial neurons arranged in layer input, hidden, and output [13]. These architectures learn patterns from data through backpropagation and gradient descent, enabling them to handle complex classification tasks. In the context of this study, a Dense Feedforward Neural Network is used to classify text inputs that have been converted into numerical vectors via the TF-IDF method.

The model uses ReLU activation functions, Dropout to reduce overfitting, and Softmax on the output layer for multi-class classification tasks [14].

TF-IDF, on the other hand, is a statistical method used to evaluate the importance of a word within a document relative to a collection of documents (corpus). It combines two metrics: Term Frequency (TF), which measures how frequently a word appears in a document, and Inverse Document Frequency (IDF), which penalizes common words across the corpus. This technique is particularly effective in feature extraction for text classification, sentiment analysis, and information retrieval. In this study, TF-IDF helps the model to focus on contextually significant words within user input, particularly in detecting patterns of depression, anxiety, and stress symptoms [15].

The combined use of TF-IDF and Neural Networks improves the chatbot's ability to understand and respond to nuanced user inputs in natural language, making the system more adaptive and accurate in mental health assessments. Despite TF-IDF's limitations in capturing word semantics, it remains a robust method when paired with deep learning models in domains with limited or structured datasets such as psychological self-assessment tools.

The TF-IDF value is calculated using the following formula (1):

$$TF - IDF(w, d) = TF(w, d) \times IDF(w) \quad (1)$$

- Term Frequency measures how often a word appears in a document relative to the total number of words in that document.
- Inverse Document Frequency (IDF) is used to reduce the weight of words that frequently occur across many documents, as such words tend to carry less meaning in text classification tasks.
- The final score represents the importance of a word within the context of the document and the entire corpus.

Methods

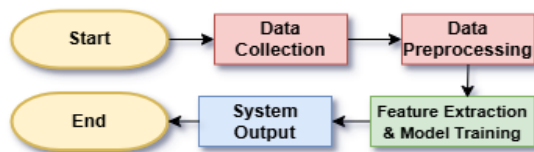


Figure 1. Research Framework for Chatbot-Based Mental Health Detection System

The research framework illustrated in Figure 1 guides the development of a mental health chatbot system based on the DASS-21 scale and Natural Language Processing (NLP). This study adopts an applied research approach aimed at building a chatbot capable of classifying user intent and assessing levels of depression, anxiety, and stress. The system was developed using Python and trained on a publicly available dataset from Kaggle, containing a variety of intent-based and mental health-related conversations.

1. Dataset and Preprocessing

The primary dataset was obtained from Kaggle (Mental Health Conversational Data), containing various conversation intents such as anxiety inquiries, depressive expressions, and therapeutic suggestions. The data preprocessing process involved:

- Removing irrelevant characters and symbols
- Tokenization and case normalization
- Stopword removal
- Vectorization using TF-IDF for feature extraction

This stage ensured the dataset was clean, structured, and ready for model training.

2. Chatbot Development Process

The system was built using Python and integrated with a custom interface that presents users with the DASS-21 questionnaire. The process flow consists of:

- User Input (chat + DASS-21 responses)
- Response Classification using NLP
- Scoring DASS-21 based on Likert scale (0–3 for each item)
- Interpretation of emotional state (normal to extremely severe)

- Chatbot Response tailored to detected intent and emotional score

3. Model Training and Evaluation

After preprocessing, the dataset was divided into training and testing subsets. A neural network model was trained to classify user intent and assess emotional states based on DASS-21 responses. The model's performance was evaluated using a confusion matrix, which provided insight into the classification accuracy across multiple intent categories. Additional metrics such as precision, recall, and F1-score were also calculated to measure the model's effectiveness in handling imbalanced classes and to ensure the robustness of the system when applied to unseen data.

4. System Output Generation

Based on the classification results and DASS-21 scoring, the system generates personalized responses through the chatbot interface. Each user input is processed in real time, and the chatbot provides output in two forms:

- Emotion Classification Result: The emotional state (normal, mild, moderate, severe) based on the total DASS-21 score.
- Response Recommendation: Automated messages or coping strategies tailored to the user's intent and emotional condition.

To evaluate the performance and impact of the developed chatbot system, this study applied both quantitative and qualitative data analysis methods:

1. Quantitative Analysis

The primary analysis was conducted using evaluation metrics derived from the classification model, including [16]:

- Accuracy: To determine the overall correctness of intent and emotional classification.

- Precision, Recall, and F1-Score: To assess the model's effectiveness, particularly in handling imbalanced categories of emotional states.
- Confusion Matrix: Used to identify misclassifications and validate the consistency of predicted intents and sentiment categories.
- DASS-21 Scoring Analysis: The responses to DASS-21 items were statistically analyzed to determine the distribution of users' emotional states across the sample population.

2. Qualitative Analysis

A small group of users was selected to interact with the chatbot, and feedback was collected to assess [17]:

- Relevance of chatbot responses based on detected emotional state.
- User perception of empathy and helpfulness in generated responses.
- Usability and engagement based on interaction satisfaction.

This mixed-method approach ensures that the system is evaluated not only based on algorithmic accuracy but also on its practical usability and psychological relevance for mental health support.

Results and Discussions

This study implements two approaches to build a classification model for detecting depressive symptoms based on user conversation data and DASS-21 scores: a Neural Network with TF-IDF and a fine-tuned BERT model. The training process was conducted over three epochs, and the results were evaluated using common classification metrics such as precision, recall, and F1-score. Evaluation was focused on three main classes: class 0 (normal), class 2 (mild depression), and class 4 (moderate depression), since there were no respondent inputs for class 1 and 3.

Table 2. Training and Validation Loss of the Neural Network Model (TF-IDF)

Epoch	Training Loss	Validation Loss
1	1.466300	1.172545
2	0.585900	0.365195
3	0.327600	0.247934

As shown in Table 2, both the training and validation loss decreased significantly across the epochs. This indicates that the model was able to learn effectively from the data and showed improved performance during training.

Table 3. Classification Evaluation of Neural Network Model (TF-IDF)

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
0	0.91	0.95	0.93	73
2	1.00	0.50	0.67	4
4	0.92	0.90	0.91	49
Accuracy	-	-	0.91	126
Macro Avg	0.94	0.78	0.83	126
Weighted Avg	0.91	0.91	0.91	126

As presented in Table 3, the TF-IDF-based neural network model achieved an accuracy of 91% and a weighted average F1-score of 0.91. However, the model struggled to classify class 2 (mild depression), as evidenced by its recall score of only 0.50. This suggests a limitation in recognizing mild cases, likely due to the small and imbalanced dataset for that class.

Table 4 Training and Validation Loss of the Fine-Tuned BERT Model

Epoch	Training Loss	Validation Loss
1	0.260200	0.392201
2	0.242500	0.184030
3	0.216800	0.204222

As seen in Table 4, the validation loss decreased most significantly during the second epoch, reaching its lowest value before increasing slightly in the third epoch. This suggests that the optimal learning point was achieved around epoch 2, beyond which the model may have started overfitting.

Table 5. Classification Evaluation of Fine-Tuned BERT Model

Class	Precisio n (%)	Recal l (%)	F1-Score (%)	Support
0	0.96	0.95	0.95	73
2	1.00	0.50	0.67	4
4	0.92	0.98	0.95	49
Accur acy	-	-	0.94	126
Macro Avg	0.96	0.81	0.86	126
Weigh ted Avg	0.95	0.94	0.94	126

As shown in Table 5, the fine-tuned BERT model performed better overall, with an accuracy of 94% and a macro average F1-score of 0.86. The model achieved excellent recall for class 4 (moderate depression) at 0.98, indicating high sensitivity in detecting these cases. Despite class 2 still showing low recall, the overall stability and adaptability of BERT significantly outperformed the TF-IDF model.

This research's refined BERT model outperformed earlier studies with a macro F1-score of 0.86 and an accuracy of 94%, surpassing the average F1-score of approximately 84% seen in comparable studies that used transformer-based architectures [18]. With weighted F1-scores of roughly 84.2%, another study similarly found better performance in depression identification when DistilBERT was paired with auxiliary characteristics [19]. Furthermore, the neural network model in this study that used TF-IDF outperformed conventional machine learning techniques, which often get accuracy rates of about 84%, with a competitive accuracy of 91% [20]. These findings demonstrate the efficacy of deep learning, in particular BERT, for multi-class classification of depressed

symptoms using DASS-21 scores and user chats.

Conclusions

This study presented the development and evaluation of a depression detection chatbot that integrates the DASS-21 psychological scale with Natural Language Processing (NLP) techniques, using two machine learning approaches: a TF-IDF-based Neural Network and a fine-tuned BERT model. The results showed that both models were capable of classifying emotional states with relatively high accuracy. The TF-IDF-based model achieved an accuracy of ninety-one percent and a weighted F1-score of 0.91, while the BERT model outperformed it with an accuracy of ninety-four percent and a weighted F1-score of 0.94. Notably, the BERT model demonstrated superior performance in identifying moderate depression (class 4) with a recall of ninety-eight percent, highlighting its robustness in handling nuanced emotional expressions in user inputs. However, both models struggled to accurately detect mild depression (class 2), primarily due to the limited and imbalanced dataset. This indicates that future research should focus on improving data diversity and representation across all severity levels to enhance model generalization and reliability. In conclusion, the integration of DASS-21 and NLP within a chatbot framework has proven effective for preliminary mental health screening. The BERT model, in particular, offers a promising foundation for building intelligent, empathetic, and accessible tools to support early detection of depressive symptoms especially in academic environments where mental health challenges are increasingly prevalent.

Acknowledgment

I would like to express my gratitude to the Informatics Engineering Department of Pelita Bangsa University who has helped a lot so that this research runs according to expectations and runs well, my deepest gratitude to my friends who have participated in this research, may it be a blessing for all of us.

Funding

This research received no external funding.

Author Contributions

conceptualization, F.N. and M.N.D.M.; methodology, F.N.; software, M.N.D.M.; validation, F.N., M.N.D.M. and A.H.A.; formal analysis, M.N.D.M.; investigation, M.N.D.M.; resources, F.N.; data curation, M.N.D.M.; writing original draft preparation, F.N.; writing review and editing, A.H.A.; visualization, M.N.D.M.; supervision, F.N.; project administration, F.N. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest in this paper.

References

- [1] K. Pleska, J. Wojtania, M. Łepik, Z. Uszok, K. Rosiak, and R. Szygula, "Dyslipidemia among psychiatric patients with depression—common possible reasons and treatment implications—review," *Quality in Sport*, vol. 10, no. 1, pp. 76–87, 2024.
- [2] World Health Organization, "Depression," *WHO Fact Sheets*, Jun. 2023.
- [3] Lovibond, S.H., and Peter F. Lovibond. "Manual for the Depression Anxiety Stress Scales (DASS-21)." *Psychology Foundation of Australia*, 2nd edition (1995).
<https://doi.org/10.1037/t01004-000>
- [4] Gunawan, Teddy Surya, Asaad Babiker, Nanang Ismail, and Mufid Ridlo Effendi. "Development of Intelligent Telegram Chatbot Using Natural Language Processing." *Journal of Advanced Technology and Applications* 7, no. 1 (2023):153–159.
<https://doi.org/10.5281/zenodo.7656768>
- [5] Mahajan, Papiya, Rinku Wankhade, Anup Jawade, Pragati Dange, and Aishwarya Bhoge. "Healthcare Chatbot using Natural Language Processing." *International Research Journal of Engineering and Technology (IRJET)* 7, no. 11 (2020): 250–255.
<https://www.irjet.net/archives/V7/i11/IRJET-V7I1142.pdf>
- [6] X. Li, S. Zhang, X. Zhang, Y. Zhan, and L. Xu, "Design and Evaluation of a CBT-Based AI Chatbot for Mental Health Support among College Students: A Randomized Controlled Trial," *Applied Sciences*, vol. 14, no. 13, p. 5889, 2024. doi: 10.3390/app14135889
- [7] Rakib, A. B., Rumky, E. A., Ashraf, A. J., Hillas, M. M., & Rahman, M. A. "Speech Recognition and Neural Networks based Talking Health Care Bot (THCB): Medibot." In *International Conference on Brain Informatics*. Cham: Springer, 2021, pp. 378–387.
- [8] Najamuddin, Muhammad, and Shohibul Adhkar. "Implementasi Chatbot Deteksi Depresi Dini pada Mahasiswa dengan PHQ-9." *Prosiding SAINTEK: Sains dan Teknologi* 1, no. 1 (2022). Universitas Pelita Bangsa.
- [9] Moh. Abdul Hakim and Nina Vania Aristawati, "Measuring Depression, Anxiety, and Stress in Early Adults in Indonesia: Construct Validity and Reliability Test of DASS-21," *Jurnal Psikologi Ulayat: Indonesian Journal of Indigenous Psychology*, vol. 10, no. 2, pp. 232–250, 2023, doi: [10.24854/jpu553](https://doi.org/10.24854/jpu553).
- [10] Girija Atigeri, Ankit Agrawal, and Sucheta V. Kolekar, "Advanced NLP Models for Technical University Chatbot Development and Conversational Assistant," *IEEE Access*, vol. 12, pp. 29633–29646, 2024, doi: [10.1109/ACCESS.2024.3368382](https://doi.org/10.1109/ACCESS.2024.3368382).
- [11] Dody Indra Sumantrawan, Siska Narulita, Muhammad Khoirul Wais, and Abraham Yano Subandono**, "Factors Affecting Depression in Students and the Role of Machine Learning in Mental Health Analysis," *Digital Transformation Technology*, vol. 4, no. 1, pp. 705–713, Mar. 2024, doi: [10.47709/digitech.v4i1.4577](https://doi.org/10.47709/digitech.v4i1.4577).

- [12] G. Malcher and B. Belatreche, "Weightless Neural Networks for Text Classification Using TF-IDF," in *Proceedings of the 29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, 2021.
- [13] R. Dey, "Classification of Fake News Headlines Based on Neural Networks," *arXiv preprint*, arXiv:2201.09966, 2022.
- [14] C. Feldges, "Text Classification with TF-IDF, LSTM, BERT: A Quantitative Comparison," *Medium*, 2023.
- [15] F. T. John, M. Liu, and A. K. Jaiswal, "Efficient Deep Learning-Based Text Classification with TF-IDF and Neural Networks," *Journal of Intelligent Systems*, vol. 33, no. 1, pp. 45–56, 2023. <https://doi.org/10.1515/jisys-2023-0005>
- [16] M. N. Miftahuddin, A. Fauzi, and A. Rahman, "Mental Health Prediction Using NLP and Machine Learning Techniques: A Comparative Study," *Informatics in Medicine Unlocked*, vol. 46, 2024.
- [17] H. J. Kim and K. Lee, "Design and Evaluation of an NLP-Based Mental Health Chatbot for Early Depression Detection Using the DASS-21," *Scientific Reports*, vol. 14, no. 1, p. 5432, 2025.
- [18] S. Sazan, R. Rahman, M. S. Islam, and M. M. Rahman, "Enhancing Depressive Post Detection in Bangla: A Comparative Study of TF-IDF, BERT, and FastText Embeddings," *IEEE International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2024. <https://doi.org/10.2139/ssrn.4885802>
- [19] T. Kerasiotis, K. Stefanidis, and G. Tsoumakas, "Depression Detection on Social Media Using DistilBERT and Auxiliary Features," *IEEE Access*, vol. 12, pp. 11235–11247, 2024.
- [20] A. Lorenzoni, G. S. Garcia, and L. Rocha, "A Comparative Study of Machine Learning and Deep Learning Models for Mental Health Text Classification," *Scientific Reports*, vol. 15, no. 1, pp. 1–12, 2025.

