Joinforcs: Journal of informatics and information system
Volume 1, (5) 2026, p. 196-201
E-ISSN : XXXX-XXXX
©2026. The Author(s). This is an open acces article under cc-by-sa

Classification and detection of mental health disorders in university students using a machine learning approach based on text datasets

Rosales Kevin Hilarius[1], Heny Pratiwi[2]
Department of Information Systems, Universitas STMIK Widya Cipta Dharma, Indonesia[1]
Master of Information Technology Department, Universitas STMIK Widya Cipta Dharma, Indonesia[2]
2341036wicida.ac.id[1], henypratiwi@wicida.ac.id[2]
* corresponding author

Article Info

Article History:
Received: August 26, 2026
Revised: September 09, 2026
Accepted: September 17, 2026

Keywords:
Early Detection, Machine Learning, Mental Health, Text Classification, Students.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Abstract

Student mental health is an increasingly critical issue in the academic environment, yet early detection remains very limited. This study aims to automatically classify student mental health conditions using machine learning techniques based on text analysis and survey data. The datasets used consist of five sources: (1) Student Mental Health Survey (101 respondents), (2) Combined Mental Health Dataset (53,043 statements with 7 mental condition classes), (3) Mental Health Text Dataset (27,977 binary classification text samples), (4) Mental Health FAQ Dataset (98 question-answer pairs), and (5) Chatbot Training Dataset (3,512 context-response pairs). The results of the survey analysis indicate that 34.6% of students experience depression and 33.7% experience anxiety, but only 5.9% seek professional help—indicating a worrying treatment gap. Classification experiments using the Gradient Boosting algorithm achieved the best performance with 89.2% accuracy and an F1-score of 89.1% on the binary classification task, and a macro F1-score of 81.4% on the multiclass classification task. The application of the SMOTE technique consistently improved performance on the minority class by an average of 8.3 percentage points. This research contributes to the development of an artificial intelligence-based early detection framework to support mental health interventions in Indonesian campus environments.

1. Introduction

Mental health is a fundamental component of overall human wellbeing, yet it is often overlooked in academic contexts [1]. University students face a range of unique pressures: heavy academic workloads, life transitions, financial pressures, and social adjustment, all of which can trigger mental health problems. The World Health Organisation (WHO) estimates that one in four people worldwide will experience a mental disorder at some point in their lives, with young people being the most vulnerable group [2].

In Indonesia, awareness of mental health issues remains relatively low, and access to mental health services is very limited, particularly in university settings. Early detection of conditions such as depression, anxiety, and stress is essential for timely intervention. However, conventional methods based on clinical interviews are costly, time-consuming, and cannot be implemented on a large scale [3].

Advances in artificial intelligence, particularly natural language processing (NLP) and machine learning, have opened up new opportunities for the automated detection of mental health disorders through text analysis. A person's verbal and written expressions often reflect their psychological state, making text-based analysis a promising approach [4]. This study integrates five datasets from various sources to develop a comprehensive and accurate mental health classification system for university students, with the aim of supporting the development of a scalable screening system in Indonesian university settings.

Mental health is a fundamental component of overall human well-being, yet it is often overlooked in academic contexts. University students face various unique pressures: heavy academic workloads, life transitions, financial pressures, and social adjustment, all of which can trigger mental health problems. The World Health Organisation (WHO) estimates that 1 in 4 people worldwide will experience a mental disorder at some point in their lives, with young people being the most vulnerable group [2].

Joinforcs: Journal of informatics and information system
Volume 1, (5) 2026, p. 196-201
E-ISSN : XXXX-XXXX
©2026. The Author(s). This is an open acces article under cc-by-sa

In Indonesia, awareness of mental health issues remains relatively low, and access to mental health services is very limited, particularly in university settings. Early detection of conditions such as depression, anxiety, and stress is crucial for timely intervention. However, conventional methods based on clinical interviews are costly, time-consuming, and cannot be implemented on a large scale [3]. Research into machine-learning-based mental health detection has developed rapidly over the past decade. Its early development focused on rule-based approaches and basic sentiment analysis to detect negative emotions in text. These approaches have limitations in accuracy because they cannot capture the complex contextual nuances of human expression [5]. [6] demonstrated the use of a BERT model to classify depression from social media posts, achieving an accuracy of 87.3%. Meanwhile, Coppersmith et al. (2019) explored the use of Twitter data to detect signs of mental disorders by combining linguistic features with deep-learning models. These studies show that NLP-based text analysis has significant potential for mental health screening. In the specific context of university students, several studies have used structured surveys. [7] investigated the relationship between students' GPAs and levels of anxiety and depression using statistical correlation analysis. However, that study was limited by a small dataset and did not employ a machine-learning approach. Our research addresses this gap by integrating structured survey data with free-text analysis for more comprehensive multiclass classification, while also addressing the issue of class imbalance, which is often a challenge in this domain.

2. Research Method

2.1. Dataset Sources and Characteristics

This study draws on five complementary datasets to develop a comprehensive understanding of university students' mental health. All datasets were collected from publicly available open sources and had been previously validated. Table 1 summarises the characteristics of each dataset used.

Table 1. Summary of the Research Datasets

Dataset	Data Volume	Key Features	Description
Student Mental Health Survey	101 respondents	Gender, CGPA, Depression, Anxiety, Panic	Structured survey data from Malaysian university students
Combined Mental Health Dataset	53.043 statements	Statement, Status (7 classes)	Text labelled with 7 mental helath conditions
Mental Health Text Dataset	27.977 texts	Text, Label (0/1)	Binary calssification of mental helath texts
Mental Health FAQ	98 Q&A pairs	Question, Answer	Knowledge base for a mental health chatbot
Chatbot Training Dataset	3.512 dialogue pairs	Context, Response	Training data for a conversational model

2.2. Data Preprocessing Stages

Data preprocessing was carried out in several systematic stages to ensure quality and consistency before modelling. The stages applied to the text data included: (1) tokenisation and lowercasing, (2) removal of English stopwords, (3) stemming using the Porter Stemmer, (4) removal of special characters and irrelevant numbers, and (5) space and encoding normalisation [8]. For the survey dataset, missing values were handled using mode imputation, and categorical variables were encoded using one-hot encoding [9]. The significant class imbalance in the Combined Mental Health Dataset (Normal: 30.8%, Depression: 29.0%, Suicidal: 20.1%, Anxiety: 7.3%, Bipolar: 5.4%, Stress: 5.0%, Personality Disorder: 2.3%) was addressed using the SMOTE (Synthetic Minority Over-sampling Technique) technique to improve the representation of minority classes and reduce prediction bias [10].

Joinforcs: Journal of informatics and information system
Volume 1, (5) 2026, p. 196-201
E-ISSN : XXXX-XXXX
©2026. The Author(s). This is an open acces article under cc-by-sa

2.3. Machine Learning Models Implemented

This study implements and compares five machine learning algorithms: Naive Bayes, Logistic Regression, Support Vector Machine (SVM) with a linear kernel, Random Forest, and Gradient Boosting. Text representation is performed using TF-IDF (Term Frequency–Inverse Document Frequency) with unigram and bigram n-grams. The models are evaluated using accuracy, precision, recall, and F1-score, with 10-fold cross-validation to ensure the generalisability of the results [11].

3. Results and Discussion

3.1. Descriptive Analysis of the Student Survey Data

Descriptive analysis of the Student Mental Health Survey (n=101) revealed significant patterns in the mental health conditions of students. The majority of respondents were female (74.3%), compared with male respondents (25.7%). The distribution by year of study was dominated by first-year students (42.6%), followed by third-year students (23.8%) and second-year students (25.7%). Academically, 46.5% of students had a CGPA of 3.50–4.00, while 42.6% fell within the range of 3.00–3.49, indicating a relatively high-achieving academic population.

Table 2. Prevalence of Mental Health Disorders Among University Students

Condition	Number (n=101)	Percentage (%)
Depression (Yes)	35	34,6%
Anxiety (Yes)	34	33,7%
Panic Attacks (Yes)	33	32,7%
Sought Professional Help (Yes)	6	5,9%
No Disorder Experienced	44	43,6%

The most concerning finding is the low rate of seeking professional help: only 5.9% of students experiencing mental health problems sought specialist assistance. This treatment gap is likely due to social stigma, limited access to services, or a lack of awareness of the importance of mental health. These findings reinforce the urgency of developing an early detection system that can operate anonymously and non-invasively [11].

3.2. Class Distribution in the Combined Mental Health Dataset

Analysis of the Combined Mental Health Dataset (n=53,043) shows an imbalanced class distribution. Table 3 presents the complete distribution of the seven mental health condition classes in this dataset.

Table 3. Class Distribution of the Combined Mental Health Dataset (N=53.043)

Class	Sample Count	Proportion (%)
Normal	16.351	30,8%
Depression	15.404	29,0%
Suicidal	10.653	20,1%
Anxiety	3.888	7,3%
Bipolar	2.877	5,4%
Stress	2.669	5,0%
Personality Disorder	1.201	2,3%
TOTAL	53.043	100%

3.3. Comparison of Machine Learning Model Performance

Classification experiments were conducted on the Mental Health Text Dataset (n=27,977) for a binary classification task. Table 4 presents a comparison of the performance of all the algorithms tested. All models were evaluated using 10-fold cross-validation with TF-IDF bigram feature representations.

©2026. The Author(s). This is an open acces article under cc-by-sa

Table 4. Comparison of Machine Learning Model Performance (Binary Classification)

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	78,4%	79,1%	77,2%	78,1%
Logistic Regression	82,7%	83,2%	81,9%	82,5%
SVM (Linear)	85,3%	85,8%	84,7%	85,2%
Random Forest	87,6%	88,1%	87,0%	87,5%
Gradient Boosting	89,2%	89,7%	88,6%	89,1%

3.4. Discussion

Gradient Boosting achieved the best performance, with an accuracy of 89.2% and an F1 score of 89.1%, followed by Random Forest (87.6%). The superiority of Gradient Boosting can be explained by its ability to handle complex non-linear feature interactions and its iterative boosting mechanism, which gradually minimises residual error. This result is consistent with the findings of [11], which demonstrated the superiority of boosting-based ensemble methods on high-dimensional data. SVM with a linear kernel also showed competitive performance (85.3%), consistent with literature demonstrating the advantages of SVM in high-dimensional feature spaces such as TF-IDF [12]. On the other hand, although Naive Bayes is the simplest model, it still achieved an accuracy of 78.4%, indicating that the assumption of conditional independence retains predictive value in this domain [13].

In multiclass classification (7 classes), Gradient Boosting also outperformed the other models, with a macro F1 score of 81.4%. However, the Personality Disorder class remained the most difficult to classify (F1 score: 71.3%) due to the limited number of samples and semantic overlap with other classes [14]. The consistent application of SMOTE improved performance on minority classes by an average of 8.3 percentage points, confirming the effectiveness of synthetic oversampling techniques in addressing class imbalance [9]. The finding of a low rate of seeking professional help (5.9%) indicates that AI-based systems could serve as a crucial complementary solution. An anonymous, accessible automated screening system that does not require direct interaction with healthcare professionals could reduce the psychological barriers faced by students when undertaking self-screening [15].

4. Conclussion

This study successfully demonstrated the potential of text-analysis-based machine learning for detecting and classifying mental health disorders among university students. By utilising five complementary datasets (comprising more than 84,000 samples in total), the study developed a comprehensive understanding of linguistic patterns correlated with various mental health conditions.

Key findings include: (1) a significant prevalence of mental health disorders among university students, with 34.6% experiencing depression and 33.7% experiencing anxiety; (2) a concerning treatment gap, with only 5.9% seeking professional help; (3) Gradient Boosting achieving the best performance, with an accuracy of 89.2% for binary classification and a macro F1 score of 81.4% for multiclass classification; and (4) the application of SMOTE effectively improving performance on minority classes by an average of 8.3 percentage points. For future research, it is recommended to: (1) explore transformer-based deep learning models such as IndoBERT for Indonesian-language text; (2) integrate multimodal data (audio and digital behaviour); (3) conduct clinical validation in collaboration with trained psychologists; and (4) develop a real-time, mobile application-based system that can be widely implemented in Indonesian university environments.

Declaration

AI Statement

The authors declare that generative AI tools were used solely to assist with language editing and grammatical improvement. All research ideas, analyses, interpretations, and conclusions are the authors' own. The authors take full responsibility for the content of this manuscript.

Joinforcs: Journal of informatics and information system
Volume 1, (5) 2026, p. 196-201
E-ISSN : XXXX-XXXX
©2026. The Author(s). This is an open acces article under cc-by-sa

Author contribution
Rosales Kevin Hilarius: Original Draft, Writing, Editing
Heny Pratiwi: Supervision and Reviewing

Funding statement
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Additional information
No additional information is available.

REFERENCES

[1] F. Oktavianus, A. Setiawan, A. Khudri, “Design and Development of a Mental Disorder Diagnosis Expert System Using the Dempster-Shafer Method,” Joinforcs: Journal of informatics and information system. Vol. 1, No. 2, pp. 50-54, 2026.
[2] WHO, "World Mental Health Report: Transforming Mental Health for All," World Health Organization, Geneva, 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240049338
[3] M. Wibowo dan M. R. F. Djafar, "Perbandingan Metode Klasifikasi Untuk Deteksi Stress Pada Mahasiswa di Perguruan Tinggi," J. Media Inform. Budidarma, vol. 7, no. 1, pp. 153 159, Jan. 2023, doi: https://doi.org/10.30865/mib.v7i1.5182. [Online]. Available: http://www.stmik budidarma.ac.id/ejurnal/index.php/mib/article/view/5182
[4] K. Rahayu, V. Fitria, D. Septhya, R. Rahmaddeni, dan L. Efrizoni, "Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning," MALCOM: Indones. J. Mach. Learn. Comput. Sci., vol. 3, no. 2, pp. 108-114, 2023, doi: https://doi.org/10.57152/malcom.v3i2.780. [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/view/780
[5] Raynold, "Klasifikasi Mental Mahasiswa Menggunakan Metode Machine Learning," J. QUANCOM: Quantum Comput. J., vol. 1, no. 2, pp. 27-32, Des. 2023. [Online]. Available: https://journal.iteba.ac.id/index.php/jurnal_quancom/article/view/217
[6] H. D. Putra, L. Khairani, dan D. Hastari, "Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Data Kesehatan Mental Mahasiswa," in SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat, vol. 1, no. 1, pp. 120 125, Agt. 2023. [Online]. Available: https://journal.irpi.or.id/index.php/sentimas/article/view/586
[7] A. I. Maulana dan Rismayani, "Mental Health Analysis at the University of Dipa Makassar using Naive Bayes Classifier," IT J. Res. Dev., vol. 8, no. 1, pp. 72-80, Des. 2023, doi: https://doi.org/10.25299/itjrd.2023.11444. [Online]. Available: http://journal.uir.ac.id/index.php/ITJRD
[8] S. Sanjaya, R. Abitio, F. R. Ahmad, H. Pratiwi. “Expert System for Determining Students Interests and Talents Based on Preference Data Using the Certainty Factor Method”, Joinforcs: Journal of informatics and information system, Vol. 1, No. 2, pp. 44-49, 2026.
[9] A. Hapsari, A. S. Nursuwanda, H. Zuhriyah, dan D. J. Vresdian, "Klasifikasi Kesehatan Mental Mahasiswa Model TMAS dengan Algoritma Decision Tree, Logistic Regression, dan Random Forest," INTEK: J. Inform. dan Teknol. Inf., vol. 7, no. 2, pp. 55-64, Nov. 2024. [Online]. Available: https://jurnal.umpwr.ac.id/index.php/intek/article/view/5690
[10] Rismayani, S. Alam, A. H. Endang, H. SY, dan N. Erdianza, "Identifikasi Gangguan Kesehatan Mental Pada Remaja Generasi Z Menggunakan Artificial Neural Network," JUSTIN (J. Sist. dan

©2026. The Author(s). This is an open acces article under cc-by-sa

Teknol. Inf.), vol. 12, no. 4, pp. 776-783, Okt. 2024, doi: https://doi.org/10.26418/justin.v12i4.86650. [Online]. Available: https://jurnal.untan.ac.id/index.php/justin/article/view/86650

[11] W. Kurniawan, A. Musthafa, dan A. Kirani, "Analisis Klasifikasi Tingkat Kesehatan Mental pada Mahasiswi Akhir dalam Menempuh Tugas Akhir Menggunakan Algoritma Support Vector Machine Studi Kasus: Universitas Darussalam Gontor," Prosiding Seminar Nasional Teknoka, vol. 9, no. 1, pp. 17-27, Des. 2024, doi: https://doi.org/10.22236/teknoka.v9i1.17524. [Online]. Available: https://journal.uhamka.ac.id/index.php/teknoka/article/view/17524

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, dan W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," J. Artif. Intell. Res., vol. 16, pp. 321-357, 2002, doi: https://doi.org/10.1613/jair.953.

[13] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., vol. 29, no. 5, pp. 1189-1232, 2001, doi: https://doi.org/10.1214/aos/1013203451.

[14] M. I. T. Azmi, D. A. Musthofa, A. Paristiyo, M. R. Perdana, M. Hanif K, T. D. Putra. “Comparative Analysis of K-Nearest Neighbor (KNN) and Decision Tree Algorithms in Predicting Student Academic Performance”, Joinforcs: Journal of informatics and information system, Vol. 1 No. 4, pp. 139–149, 2026.

[15] D. E. Putra, Nelwati, dan F. Fernandes, "Hubungan depresi, stres akademik dan regulasi emosi dengan ide bunuh diri pada mahasiswa," J. Keperawatan Jiwa, vol. 11, no. 3, pp. 689 706, 2023. [Online]. Available: https://jurnal.unimus.ac.id/index.php/JKJ