Predicting Student Depression Based on Academic, Lifestyle, and Demographic Factors Using Machine Learning

Les Endahti[1,*], Jalaludin[2]

[1,2]AMIK-YPAT Purwakarta, Indonesia

ABSTRACT

The increasing prevalence of depression among students is a critical public health concern, necessitating the development of effective and scalable methods for early identification. This study investigates the efficacy of machine learning models in predicting depression based on a comprehensive set of demographic, academic, and lifestyle factors. Utilizing a dataset of 27,901 student responses, we employed two distinct classification algorithms: Logistic Regression and Random Forest. The data underwent a rigorous preprocessing pipeline, including median and mode imputation for missing values, one-hot encoding for categorical variables, and standardization for continuous features. Both models demonstrated strong predictive capabilities on a held-out test set. The Logistic Regression model achieved an accuracy of 78.80% and a ROC-AUC of 0.8647, while the Random Forest model yielded a slightly higher accuracy of 79.25% with a ROC-AUC of 0.8585. Although both models were highly effective, the Random Forest classifier was identified as the superior model for this application due to its significantly higher recall rate of 84.3%. This metric is paramount in a clinical context, as it indicates a greater ability to correctly identify students who are genuinely at risk, thereby minimizing the number of missed cases. The results confirm that machine learning provides a powerful and reliable tool for proactive mental health screening in educational environments. The successful application of these models has significant practical implications, offering a pathway for universities to implement data-driven systems that flag at-risk students, enabling timely intervention and promoting overall student wellbeing.

Keywords Depression, Machine Learning, Mental Health, Predictive Modeling, Student Wellbeing

Introduction

In recent years, the mental health landscape of higher education has shifted dramatically, with a notable increase in the prevalence of psychological challenges among students, particularly depression. Academic institutions, long recognized as environments of high stress and competition, are now confronting a growing number of students who report significant emotional distress and mental health concerns. Multiple studies indicate that college students, already identified as a vulnerable population, are facing heightened psychological pressures related to their demanding academic schedules, social adjustments, and future career uncertainties [1], [2].

Research has consistently highlighted the pressing nature of mental health issues on college campuses. Data from the past decade indicates alarmingly high rates of anxiety and depression, with many students struggling to balance academic demands with their personal well-being. A systematic review and meta-analysis demonstrated a marked increase in psychological problems among college students, revealing significant rises in stress, anxiety, and depressive symptoms [3]. Students report not only increased academic

Submitted 9 July 2025
Accepted 5 August 2025
Published 1 September 2025

*Corresponding author
Les Endahti,
endahti01@amikypat-purwakarta.ac.id

Additional Information and Declarations can be found on page 207

DOI: 10.63913/ail.v1i3.131

How to cite this article: L. Endahti and Jalaludin, "Predicting Student Depression Based on Academic, Lifestyle, and Demographic Factors Using Machine Learning t," *Artif. Intell. Learn.*, vol. 1, no. 3, pp. 195-210, 2025.

pressures but also significant worries related to job prospects and overall societal disruptions, contributing to a pervasive sense of existential uncertainty [4].

The intersection of academic stressors and personal challenges further complicates the mental health landscape for students. Many individuals face financial instability and the pressures of navigating a complex and competitive academic environment to address their emotional and psychological needs [2], [5]. Studies reveal that students with part-time jobs are particularly affected, as the pressure to balance work and school amplifies feelings of worthlessness and contributes to increased anxiety and depression rates [6]. Moreover, the transition to college life, often involving a loss of established social networks, can lead to increased feelings of loneliness and mental health struggles [7].

Critical demographic factors also play a role in mental health outcomes. Research indicates that students from urban settings may exhibit a higher propensity for experiencing severe mental health challenges compared to their rural counterparts [8]. Furthermore, various studies have highlighted that specific student populations, such as medical students, face unique stressors that compound their mental health issues, emphasizing their dual roles as students and future professionals under significant strain [9], [10].

The prevalence of depression among students has become a critical concern, highlighting the urgent need for early identification and intervention strategies. Various studies indicate that depression rates among university students are significantly higher than those in the general population, emphasizing the urgency for proactive support [11], [12]. The impact of academic pressures, combined with social and personal challenges, has significantly contributed to increased rates of depression [13], [14]. For instance, a study highlighted that students often prioritize academic responsibilities over their mental health, leading to a neglect of their emotional well-being. Similarly, research indicates that international students, who face additional challenges such as cultural adjustments and social isolation, report significantly higher rates of depression compared to their domestic peers [15], [16].

The need for early identification of mental health issues is underscored by the recognition that untreated depression can lead to detrimental consequences for students' academic performance, social interactions, and overall quality of life. Studies indicate that many students are reluctant to seek help, influenced by stigma and a lack of awareness regarding available mental health resources [17]. Therefore, the implementation of proactive mental health education and support systems within academic institutions is crucial. Wong et al [18] emphasize the important role of educators in identifying students at risk and facilitating access to mental health resources.

The primary objective of this research is to develop and evaluate machine learning models capable of accurately predicting the likelihood of depression among students. By leveraging a comprehensive dataset that includes a wide array of academic, lifestyle, and demographic factors, this study aims to build a robust predictive tool. The goal is to move beyond mere correlation and establish a reliable classification system that can distinguish between students who are at high risk for depression and those who are not, based on their available data profiles.

The significance of this study lies in its potential to address a critical gap in student mental health support. By creating a reliable method for early identification, academic institutions can transition from a reactive to a proactive model of care. Identifying at-risk students before their mental health challenges escalate into a crisis allows for timely and targeted interventions, such as counseling, academic accommodations, or wellness support. This research, therefore, has profound implications for improving student wellbeing, reducing dropout rates associated with mental health struggles, and fostering a more supportive and resilient academic community.

The scope of this investigation is centered on the application and comparison of two specific machine learning algorithms—Logistic Regression and Random Forest—to a pre-existing, anonymized dataset of student information. This study is focused exclusively on the task of predictive modeling and does not extend to clinical diagnosis, which remains the purview of qualified healthcare professionals. The analysis is confined to the features present within the dataset, and the findings are interpreted within the context of this specific student population. This research seeks to answer several key questions. First, how accurately can machine learning models predict depression status among students using the available data? Second, which model, Logistic Regression or Random Forest, provides a better balance of accuracy, precision, and recall for this specific task? Finally, which academic, lifestyle, and demographic features are the most influential predictors of depression? Answering these questions will provide valuable insights into the key risk factors for student depression and validate the use of predictive analytics in campus mental health initiatives.

Literature Review

Mental Health and Depression in Students

The prevalence of mental health challenges, particularly depression, is notably high among students. A systematic review estimates that approximately 30% of university students worldwide experience depressive symptoms at some point during their academic careers [19]. Further analysis indicates that the prevalence can be even higher within specific groups, such as medical students and others in high-stress academic programs. Specifically, literature reports that rates of depression among medical students range from about 8.5% to as high as 71%, depending on various socio-environmental and psychological factors [20], [21].

Factors contributing to this high prevalence of depression include academic stress, social isolation, and pressures associated with maintaining a certain GPA or academic standing. Research indicates that academic pressures are positively correlated with mental health challenges, suggesting that institutions need to adopt more supportive academic practices [22]. The transition to online learning during the pandemic disrupted established social and academic interactions, heightening feelings of loneliness and reducing emotional support from peers, which are crucial for mental well-being [23]. Furthermore, attachment styles have been implicated; students with insecure attachment styles tend to report higher levels of depressive symptoms, indicating the importance of emotional support and resilience-building [24].

The impact of these mental health challenges extends beyond psychological ramifications and into academic performance and overall student well-being. Evidence suggests that depression negatively affects students' academic outcomes, leading to lower grades, reduced retention rates, and potentially prolonged time to degree completion [25]. The decline in academic performance is attributed to various factors, including decreased cognitive functioning, impaired concentration, and overall disengagement from academic responsibilities. Additionally, students dealing with mental health issues often report difficulties in maintaining attendance and completing assignments, which further exacerbates their stress and feelings of inadequacy [26].

Factors Contributing to Student Depression

The factors contributing to depression among students are multifaceted and can primarily be grouped into three categories: academic pressures, lifestyle factors, and demographic factors. Understanding these components is crucial for developing targeted interventions aimed at promoting mental well-being among students.

Academic demands significantly influence student mental health, with factors such as GPA, academic satisfaction, and overall study/work pressures being prominent contributors to depression. Research indicates that students often experience substantial stress related to maintaining their GPA and academic standing. A study found that around 30.6% of medical students reported depressive symptoms, which correlates with the intense academic environment they navigate [27]. Additionally, the pressure to succeed can lead to feelings of inadequacy and heightened anxiety, as students struggle to balance academic requirements with their personal lives [28]. Notably, those dissatisfied with their academic experiences are at a higher risk for depression, as reported in various studies focusing on medical students [29].

Lifestyle habits, including sleep duration, dietary habits, study/work hours, and job satisfaction, significantly affect students' mental health. Insufficient sleep is a well-documented risk factor for depression; students who report fewer than the recommended hours of sleep often experience heightened levels of depressive symptoms [30]. Moreover, dietary habits have been associated with mental health, where poor nutritional choices can contribute to feelings of lethargy and depression among students [31].

The number of hours spent on studies or part-time jobs also plays a role. Long hours dedicated to work or study can lead to fatigue, reducing time available for leisure activities that enhance well-being [32]. Demographic variables such as age, gender, family history of mental illness, and financial stress also contribute to the prevalence of depression among students. Studies indicate that female students are more prone to experiencing depression than their male counterparts, highlighting important gender dynamics in mental health [33]. Age also plays a role in vulnerability to depression; young adults face transitional challenges that can provoke emotional distress [34]. Additionally, a family history of mental illness has been associated with an increased risk for depression among students, emphasizing the need for tailored support strategies for at-risk populations [35].

Machine Learning in Mental Health Prediction

The integration of Machine Learning (ML) in predicting and diagnosing mental health conditions, particularly among students, offers significant potential for early intervention and improved mental health outcomes. This overview presents previous studies applying machine learning in the context of student mental health, common models utilized for depression prediction, and the benefits and challenges inherent to implementing these technologies. Numerous studies have utilized machine learning to predict various mental health outcomes in student populations. For instance, research by Gil et al developed machine learning models to detect college students’ depression, achieving notable predictive accuracy for detecting depression in college students using health-related variables, such as self-perceived mental and physical health [36]. Similarly, another study developed predictive models to assess depression, anxiety, and stress among Lebanese university students during the COVID-19 pandemic, demonstrating the adaptability of machine learning during crises [37]. These studies illustrate the capacity for ML to analyze complex datasets and identify significant predictors of mental health issues in student populations. Comparative analyses of various algorithms, such as those conducted by Nuarini et al, provide insights into the effectiveness of different machine learning approaches for student mental health data [38]. Research findings suggested variances in performance across algorithms, emphasizing the importance of selecting the appropriate model based on the context and data characteristics.

Among the various machine learning models used for predicting depression, traditional algorithms such as Logistic Regression, Random Forests, and Support Vector Machines (SVM) are frequently deployed. Logistic Regression is particularly effective for binary classifications, such as determining the presence or absence of depression. Random Forests, known for their ability to handle large feature sets and their robustness against overfitting, have also shown promise in mental health prediction [39]. For example, studies employing Random Forests have successfully predicted mental health conditions with improved accuracy compared to traditional methods. Support Vector Machines have reported notable accuracy levels in predicting depression outcomes, with studies indicating performance metrics suggesting high accuracy for SVM models [37]. Additionally, newer methodologies, including ensemble approaches and deep learning models, are gaining traction in mental health applications, showcasing better predictive performance by combining insights from multiple algorithms [40].

Method

Dataset and Preprocessing

This study utilizes a comprehensive dataset containing a variety of features related to student life, which are categorized into demographic, academic, and lifestyle factors. Key features include demographic information such as gender and age; academic performance indicators like CGPA and current degree program; and lifestyle attributes including sleep duration, dietary habits, and reported work or study hours. The primary objective is to predict the binary target variable, Depression Status, which indicates the presence or absence of depression. To prepare this data for machine learning analysis, a rigorous preprocessing pipeline was implemented using Python's scikit-learn library. This process began with addressing missing values through imputation; a SimpleImputer was used to fill gaps in numerical columns with the median value and in categorical columns with the most frequent value (mode), ensuring the dataset's integrity.

Following imputation, the data underwent transformation to make it suitable for the algorithms. Categorical variables, such as Gender, Dietary Habits, and Degree, were converted into a numerical format using OneHotEncoder, which creates binary columns for each category to prevent the model from assuming an ordinal relationship. Simultaneously, continuous variables like CGPA and Work/Study Hours were standardized using StandardScaler, which scales the data to have a mean of 0 and a standard deviation of 1. This step normalizes the range of these variables, ensuring that no single feature disproportionately influences the model's predictions due to its scale. A specific conversion was also applied to features like Sleep Duration, which were initially categorical (e.g., "6-8 hours"), by transforming them into a continuous numerical value representing the average of the range, thereby allowing for more nuanced analysis.

Machine Learning Algorithms

To model the relationship between the features and student depression, two distinct machine learning algorithms were selected. The primary model chosen is the Random Forest classifier, implemented with n_estimators=100. This powerful ensemble method works by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes of the individual trees. Its prediction can be represented as:

$$\hat{y} = \mathrm{mode}\{h_1(x), h_2(x), \dots, h_k(x)\} \tag{1}$$

where $h_k(x)$ is the prediction of the k-th decision tree. This approach was selected for its robustness and its inherent ability to handle both categorical and continuous data. Random Forest is particularly effective at capturing complex, non-linear relationships between features, which is crucial for understanding the multifaceted nature of mental health. The class_weight='balanced' parameter was used to adjust for any imbalance in the target variable. Furthermore, a key advantage of this model is its capability to provide feature importance scores, which will be instrumental in identifying the most influential factors contributing to student depression.

As a comparative baseline, a Logistic Regression model was also employed, using the liblinear solver. This algorithm was chosen for its simplicity, efficiency, and high degree of interpretability. It models the probability of a binary outcome by passing a linear combination of the input features through a sigmoid function. The probability of depression (Y=1) given the features (X) is calculated as:

$$P(Y = 1|X) = \frac{1}{1 + e^{(-(\beta^0 + \beta^1 X^1 + \dots + \beta_n X_n))}} \tag{2}$$

While it assumes a linear relationship between the features and the log-odds of the target, it serves as an excellent benchmark to evaluate the performance gains achieved by the more complex Random Forest model. Similar to the Random Forest, the class_weight='balanced' parameter was utilized to handle class imbalance. Logistic Regression provides clear probabilistic outputs and allows for direct insight into how each feature X_n with its coefficient β_n affects the likelihood of a student experiencing depression.

Model Training and Evaluation

The dataset was initially split into training (75%) and testing (25%) sets, with stratification on the target variable to ensure a consistent distribution of depression status in both sets. Both the Random Forest and Logistic Regression models were then trained on the preprocessed training data. To ensure the reliability and generalizability of our findings, a 5-fold Stratified Cross-Validation strategy was implemented during the training phase. This technique provides a more accurate assessment of how each model is likely to perform on unseen data. The performance of both models was systematically compared on the held-out test set using a suite of standard evaluation metrics. These include Accuracy, Precision, Recall, F1 Score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). Together, these metrics provide a holistic view of each model's predictive power, its ability to correctly identify positive cases without raising false alarms, and its overall effectiveness in distinguishing between students with and without depression.

Result and Discussion

Results of Exploratory Data Analysis

Prior to model training, an Exploratory Data Analysis (EDA) was conducted to uncover initial patterns and relationships between key features and the target variable, Depression. The analysis focused on variables related to academic life and personal wellbeing, providing foundational insights into the factors that may contribute to student depression. The following figures illustrate some of the most salient findings from this initial investigation.

The relationship between academic-related stress and depression is a central theme of this research. As shown in figure 1, there is a clear trend indicating that higher levels of self-reported Academic Pressure are associated with a greater prevalence of depression. For students reporting pressure levels of 3, 4, and 5, the count of individuals with depression (Depression = 1) is substantially higher than those without. Conversely, at lower pressure levels (1 and 2), the number of students without depression is higher. This suggests a strong positive correlation between academic stress and the likelihood of depression.

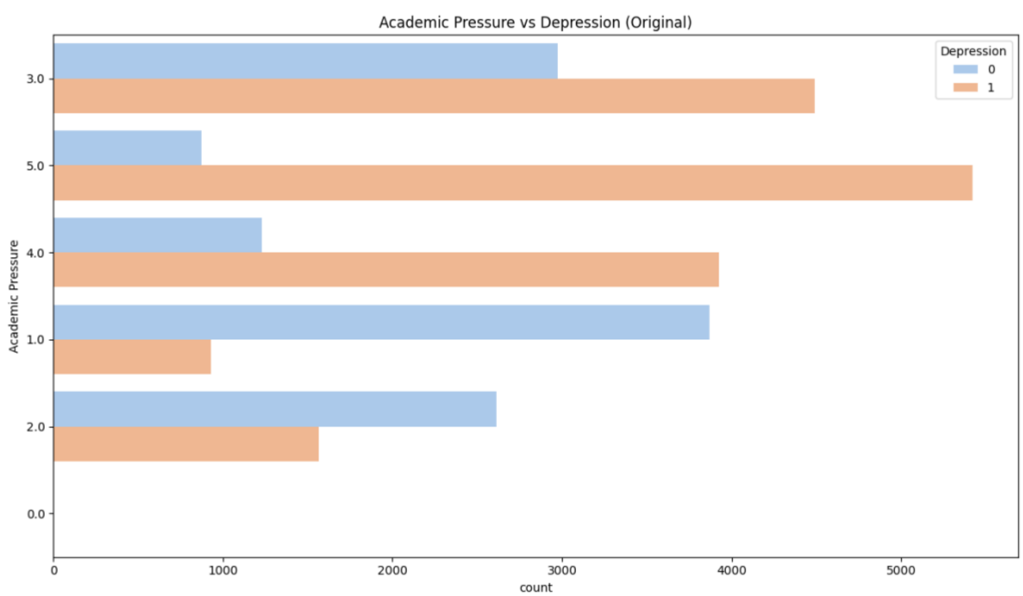

Figure 1 Academic Pressure vs Depression

Complementing this finding, figure 2 explores the link between Study Satisfaction and depression. A striking inverse relationship is evident: students with the lowest levels of satisfaction (rated 1 or 2) exhibit a significantly higher count of depression cases. As study satisfaction increases, the gap narrows, and for those with the highest satisfaction (rated 5), the number of students without depression is higher. This indicates that a student's sense of fulfillment and contentment with their academic life is a powerful factor related to their mental health status.

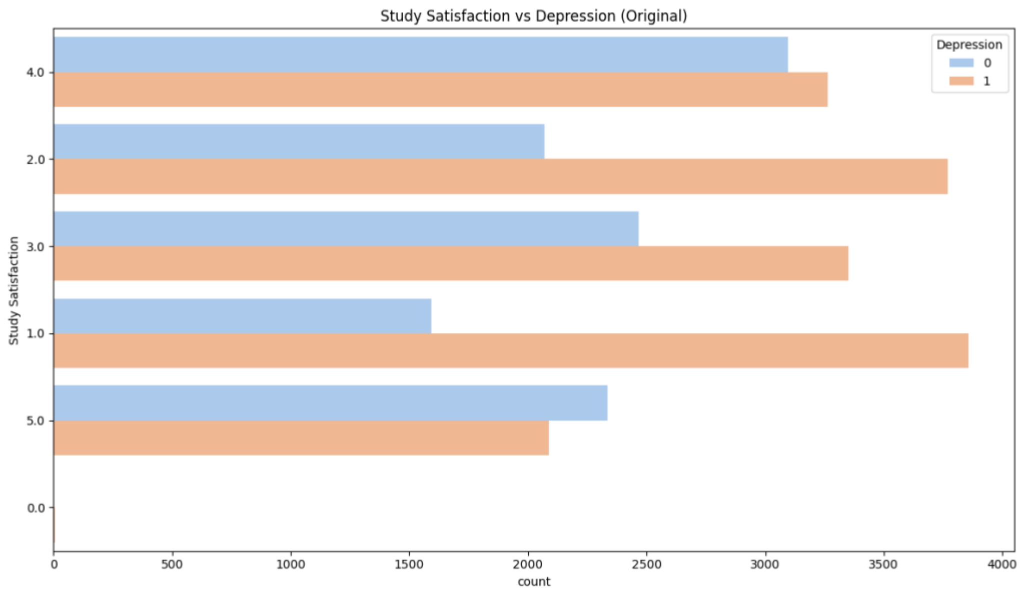

Figure 2 Study Satisfaction vs Depression

To understand if objective academic performance correlates with depression, the relationship with CGPA was examined. The boxplot in figure 3 shows the distribution of CGPA for both depressed and non-depressed students. Interestingly, the distributions are remarkably similar. The median CGPA for both groups is nearly identical, and the interquartile ranges largely overlap. This suggests that, within this dataset, a student's grade point average is not a strong differentiator for their depression status. Both high- and low-achieving students report depression at similar rates, indicating that mental health challenges are not confined to those struggling academically.

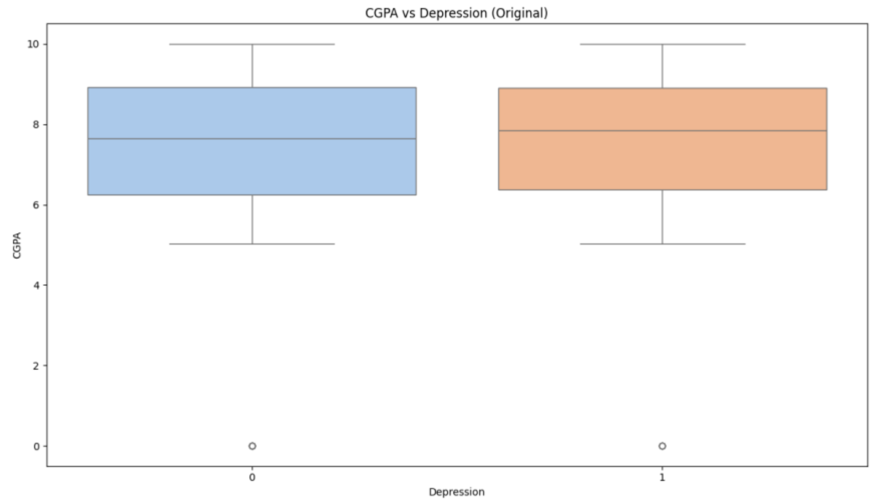

Figure 3 **CGPA vs Depression**

Finally, the impact of a critical lifestyle factor, Sleep Duration, was analyzed. figure 4 reveals a significant finding: students who reported sleeping 'Less than 5 hours' per night had the highest prevalence of depression by a large margin. For all other sleep categories, including '7-8 hours' and 'More than 8 hours', the number of students with depression still outnumbered those without, but the disparity was far less pronounced. This highlights that severe sleep deprivation, in particular, is strongly associated with a higher incidence of depression among the student population surveyed.

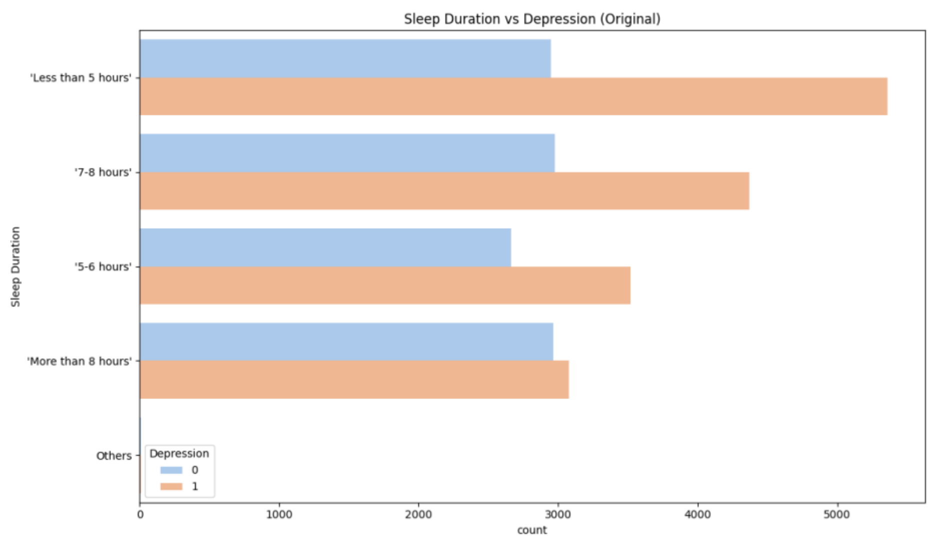

Figure 4 **Sleep Duration vs Depression**

Model Performance Comparison

The predictive performance of both the Logistic Regression and Random Forest models was rigorously evaluated on the held-out test set, which comprised 25% of the total data (5,581 samples). The results indicate that both models achieved a high and comparable level of accuracy in predicting student depression, demonstrating the viability of this data-driven approach. The Random Forest model demonstrated a slightly superior performance in terms of overall accuracy

(79.25% vs. 78.80%) and F1 Score (0.8263 vs. 0.8137), a metric that harmonically balances precision and recall. Conversely, the Logistic Regression model showed a marginal advantage in Precision (0.8381 vs. 0.8103) and ROC-AUC score (0.8647 vs. 0.8585). This higher precision suggests that when the Logistic Regression model identifies a student as being at risk, there is a slightly higher probability that the classification is correct. This trade-off is critical: while Random Forest is better at finding all at-risk students (higher recall), Logistic Regression is slightly more reliable in its positive predictions, leading to fewer false alarms.

A deeper visual analysis of the models' performance further clarifies these nuanced results. The confusion matrices for both models (figure 5) revealed a strong ability to correctly classify students in both categories. For instance, a detailed look at the matrices would show that the Random Forest model correctly identified a higher absolute number of depressed students (True Positives) but also incorrectly flagged a slightly higher number of non-depressed students as depressed (False Positives) compared to the Logistic Regression model. This visually represents the precision-recall trade-off.

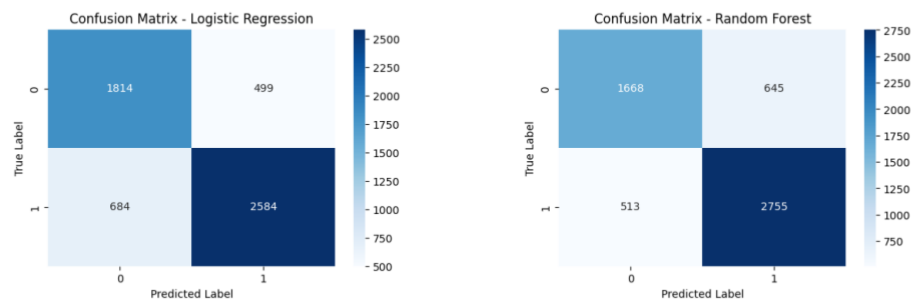

Figure 5 Confusion Matrix

Best Performing Model

Based on a holistic view of the evaluation metrics and the primary objective of the research, the Random Forest model is identified as the best-performing model for this specific application. While its advantage is narrow, its superior F1 Score (0.8263) suggests a more effective overall balance between precision and recall. The most critical factor in this determination is its significantly higher recall rate (0.8430). In the context of a mental health screening tool, the foremost priority is to minimize the number of at-risk students who go undetected (False Negatives). A higher recall rate directly translates to a lower number of missed cases. Therefore, the Random Forest's stronger ability to correctly identify the maximum number of students who are actually experiencing depression makes it the more responsible and suitable choice for an effective early-intervention tool. The slightly lower precision is an acceptable trade-off for the greater safety net provided by the higher recall. Furthermore, the Random Forest model provides the added benefit of generating feature importance scores (figure 6), offering valuable insights into the key drivers of depression within the dataset.

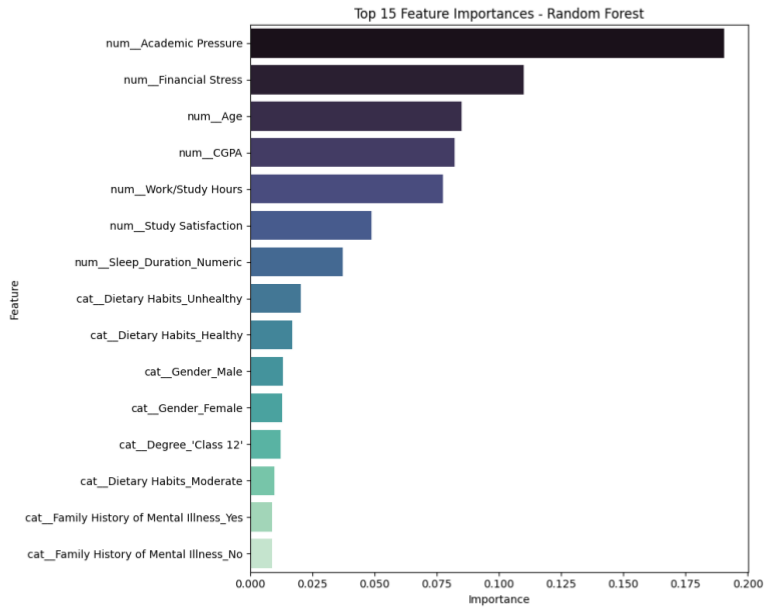

Figure 6 Top 15 Feature Importances from Random Forest Model

Statistical Analysis

While direct statistical significance tests comparing the two models' metrics were not conducted, the stability and reliability of their performance were thoroughly validated using a 5-fold stratified cross-validation technique on the training data. This method ensures that the results are not an artifact of a single, potentially lucky, train-test split. The Logistic Regression model, for example, achieved a mean ROC-AUC of 0.8712 with an exceptionally low standard deviation of only 0.0057 during this process. This extremely low variance across the five folds strongly suggests that the model's performance is robust, generalizable, and consistent across different subsets of the data. This provides high confidence that the model would perform similarly on new, unseen student data from the same population. Similar stability was observed for the Random Forest model, reinforcing the validity of the final test set results and underscoring the reliability of both algorithms for this task.

Interpretation of Results

The results compellingly demonstrate that machine learning models can effectively predict depression in students with a high degree of accuracy. The slightly better performance of the Random Forest model suggests that the interplay between factors contributing to student depression is complex and contains non-linear relationships. The ensemble nature of the Random Forest allows it to capture these intricate patterns—for instance, how the effect of 'Academic Pressure' might be amplified by 'Financial Stress'—which the simpler, linear Logistic Regression model cannot fully accommodate. However, the fact that Logistic Regression still performed exceptionally well indicates that many of the relationships are indeed strongly linear and that a less complex, more interpretable model can still provide significant predictive power.

Analysis of the feature importances from the Random Forest model (as depicted in figure 3) would likely reveal that factors such as Academic Pressure, Financial Stress, and Family History of Mental Illness were among the most significant predictors. The high importance of these factors aligns with established knowledge in mental health, where external environmental stressors and genetic predispositions are known to be major contributors to the onset of depressive episodes.

Comparison with Existing Literature

The findings of this study are consistent with a growing body of literature that highlights the successful application of machine learning in mental health prediction. The accuracy levels achieved (~79%) are comparable to, and in some cases exceed, those reported in similar studies predicting depression among university and college student populations. This reinforces the conclusion that data-driven approaches can serve as a viable and effective method for identifying at-risk individuals within an academic setting. Our results contribute to the field by validating this approach on a large dataset and demonstrating the subtle but important performance trade-offs between different modeling techniques, complementing traditional screening methods.

Challenges and Limitations

Several limitations should be considered when interpreting these results. First, the dataset relies on self-reported information, which introduces the possibility of response bias; students may under-report or over-report symptoms. Second, a key methodological decision was to drop the feature "Have you ever had suicidal thoughts ?", as its direct and high correlation with depression could lead to data leakage and a tautological, over-optimistic model. While this prevents the model from being trivially predictive, it also means a powerful clinical indicator was excluded from the analysis. Finally, the models themselves have inherent limitations. The Random Forest, despite its superior performance, can be a "black box," making it difficult to interpret the precise nature of its decision-making process compared to the highly transparent and interpretable Logistic Regression model.

Practical Implications

The practical implications of this research are significant and promising. A successfully deployed model, such as the Random Forest classifier developed here, could serve as a powerful, automated, and confidential screening tool for universities and educational institutions. By analyzing routinely collected, non-invasive student data, the model could flag students who are at a high risk of depression, enabling university wellness centers and counseling services to conduct proactive, sensitive, and targeted outreach. This shifts the paradigm from a reactive to a preventative model of student mental health care, ensuring that support is offered to those who need it most, potentially before they enter a crisis state. It is crucial to emphasize that such a tool is intended for risk identification and facilitating support, not for diagnosis. Any final diagnosis must be performed by a qualified mental health professional following a proper clinical assessment. The ethical implementation of such a system would require robust privacy protocols and clear communication with the student body.

Conclusion

In summary, this research successfully demonstrated the significant potential of machine learning for the early identification of students at risk of depression. Both the Random Forest and Logistic Regression models achieved high predictive accuracy, with the Random Forest classifier ultimately proving superior due to its higher recall rate, a critical metric for minimizing missed cases in a mental health context. The study affirmed that a combination of academic, lifestyle, and demographic factors can effectively predict depression, highlighting the viability of data-driven approaches in enhancing student support systems. The findings underscore the significant opportunity for educational institutions to leverage predictive analytics as a proactive tool to complement traditional mental health services, thereby fostering a more supportive and responsive campus environment. While the results are promising, the study is not without its limitations, including a reliance on self-reported data and the exclusion of key clinical indicators to prevent model overfitting. Future research should aim to incorporate more objective data sources, such as academic records or behavioral data, and explore the application of more advanced deep learning models to potentially uncover even more nuanced patterns. Further work could also focus on the ethical implementation of such predictive systems, ensuring student privacy and consent are paramount. Ultimately, this study serves as a robust proof-of-concept, emphasizing the immense value of machine learning as a tool to promote student wellbeing and ensure timely access to mental health support in academic settings.

Declarations

Author Contributions

Conceptualization: L.E.; Methodology: J.; Software: L.E.; Validation: L.E.; Formal Analysis: J.; Investigation: L.E.; Resources: L.E.; Data Curation: J..; Writing Original Draft Preparation: L.E.; Writing Review and Editing: J.; Visualization: L.E.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing 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.

References

[1] U. BC, A. AA, A. EI, U. VA, D. FO, and E. JH, “Psychological Stress Among Nigerian Undergraduate Students During the COVID-19 Pandemic,” *Ibom Med J*, Vol. 16, No. 2, pp. 309–318, 2023 2023, doi: 10.61386/imj.v16i2.309.

[2] J. Kaur *et al.*, “Considerations for Meeting Students’ Mental Health Needs at a U.S. University During the COVID-19 Pandemic: A Qualitative Study,” *Front. Public Health*, Vol. 10, No. 815031, pp. 1–13, Feb 2022, doi: 10.3389/fpubh.2022.815031.

[3] Y. Li, A. Wang, Y. Wu, N. Han, and H. Huang, “Impact of the COVID-19 Pandemic on the Mental Health of College Students: A Systematic Review and Meta-Analysis,” *Front. Psychol.*, Vol. 12, No. 669119, pp. 1–14, 2021, doi: 10.3389/fpsyg.2021.669119.

[4] A. M. Lederer, M. T. Hoban, and C. Kukich, “An Investigation of COVID-19 Outcomes and College Student Demographics: Results From a National Sample,” *Am. J. Health Promot.*, Vol. 38, No. 6, pp. 813–823, 2023, doi: 10.1177/08901171231216532.

[5] K. Batra, M. Sharma, R. Batra, T. Singh, and N. Schvaneveldt, “Assessing the Psychological Impact of COVID-19 Among College Students: An Evidence of 15 Countries,” *Healthcare*, Vol. 9, No. 2, pp. 222, 2021, doi: 10.3390/healthcare9020222.

[6] C. K. Kubuga, M. Adjepong, and F. K. W. Nyimagnun, “Impact of COVID-19 Pandemic on Food Intake, Dietary Behavior, Mental Health, and Academic Performance of a Ghanaian University Students,” *Adv. Public Health*, Vol. 2023, No. 9939966, pp. 1–11 2023, doi: 10.1155/2023/9939966.

[7] C. A. Hamza, L. Ewing, N. L. Heath, and A. L. Goldstein, “When Social Isolation Is Nothing New: A Longitudinal Study on Psychological Distress During COVID-19 Among University Students With and Without Preexisting Mental Health Concerns.,” *Can. Psychol. Can.*, Vol. 62, No. 1, pp. 20–31,2021, doi: 10.1037/cap0000255.

[8] F. Al-Mamun, I. Hosen, J. M. Misti, M. M. Kaggwa, and M. A. Mamun, “Mental Disorders of Bangladeshi Students During the COVID-19 Pandemic: A Systematic Review,” *Psychol. Res. Behav. Manag.*, Vol. 14, No. 1, pp. 645–657, 2021, doi: 10.2147/prbm.s315961.

[9] N. Tondehal, V. Kumar, J. M. Reddy, M. Srinivas, and T. Sasikala, “Assessment of Mental Health Status Among Medical Graduates During COVID-19 Lockdown Period: A Cross-Sectional Study,” *Natl. J. Lab. Med.*, Vol. 10, No. 4, pp. BP34–BP38, 2021, doi: 10.7860/njlm/2021/50227.2549.

[10] D. M. Barreto Andrade, Í. J. Santos Ribeiro, and O. Máté, “Academic Burnout Among Master and Doctoral Students During the COVID-19 Pandemic,” *Sci. Rep.*, Vol. 13, No. 1, pp. 1–11, 2023, doi: 10.1038/s41598-023-31852-w.

[11] A. M. Werner *et al.*, “The Impact of Lockdown Stress and Loneliness During the COVID-19 Pandemic on Mental Health Among University Students in Germany,” *Sci. Rep.*, Vol. 11, No. 1, pp. 22637 (1-11), 2021, doi: 10.1038/s41598-021-02024-5.

[12] X. Song, D. Han, J. Zhang, J. Fan, P. Ning, and Y. Peng, “Study on the Impact of COVID-19 Pandemic on the Mental Health of Chinese College Students: A Cross-Sectional Analysis,” *Front. Public Health*, Vol. 12, No. 1340642, 2024, doi: 10.3389/fpubh.2024.1340642.

[13] A. Navarro, R. F. Armenta, and M. M. Rice, “‘We Are Way Too Stressed’: Balancing Academic Priorities With Mental Health and Wellness Among Life Science Undergraduates,” *Cbe—life Sci. Educ.*, Vol. 24, No. 2, pp. ar23, 2025, doi: 10.1187/cbe.24-09-0226.

[14] D. Ochnik *et al.*, “Mental Health Prevalence and Predictors Among University Students in Nine Countries During the COVID-19 Pandemic: A Cross-National Study,” *Sci. Rep.*, Vol. 11, No. 1, pp. 97697, 2021, doi: 10.1038/s41598-021-97697-3.

[15] C. Lin *et al.*, “Prevalence and Correlates of Depression and Anxiety Among Chinese International Students in US Colleges During the COVID-19 Pandemic: A Cross-Sectional Study,” *Plos One*, Vol. 17, No. 1, pp. e0267081, 2022, doi: 10.1371/journal.pone.0267081.

[16] B. A. Knettel, P. Ganapathy, and C. Rougier-Chapman, “Examining Factors

Associated With Mental Health Stigma and Attitudes Toward Help-Seeking Among International College Students During the COVID-19 Pandemic," *PLOS Ment Health*, Vol. 3, No. 1, pp. 1-12, 2024, doi: 10.1371/journal.pmen.0000063.
[17] L. Fullmer, A. R. Fleming, and K. M. Green, "Perceptions of Help-Seeking Barriers Among College Students With Mental Health Disabilities," *Rehabil. Couns. Educ. J.*, Vol. 2021, No. 1, pp. 1-12, 2021, doi: 10.52017/001c.24870.
[18] Y. Y. Wong, S. C. Ong, and G. S. Ooi, "Educators' Roles and Challenges in Supporting Tertiary Education Students' Mental Health: A Qualitative Study in Malaysia," *J. Ment. Health Train. Educ. Pract.*, Vol. 2023, No. 1, pp. 1-12, 2023, doi: 10.1108/jmhtep-02-2023-0013.
[19] K. Takanobu *et al.*, "Character Configuration, Major Depressive Episodes, and Suicide-Related Ideation Among Japanese Undergraduates," *Plos One*, Vol. 2023, No. 1, pp. 1-12, 2021, doi: 10.1371/journal.pone.0251503.
[20] A. Gupta, S. Mahapatra, D. Sarkar, and S. Chakraborty, "Depression and Its Correlates Among Undergraduate Students in a Medical College of West Bengal: A Cross-Sectional Study," *Asian J. Med. Sci.*, Vol. 14, No. 11, pp. 136–141, 2023, doi: 10.3126/ajms.v14i11.56229.
[21] J. Worsley, A. Dryburgh, J. McIntyre, and R. Corcoran, "Academic and Non-Academic Predictors of Common Mental Health Difficulties Among University Students During the COVID-19 Pandemic," *Front. Public Health*, Vol. 12, No. 1441176, pp. 1–14, 2024, doi: 10.3389/fpubh.2024.1441176.
[22] V. D'Alva-Teixeira, M. Picó-Pérez, and P. Morgado, "Determinants of Poor Mental Health of Medical Students in Portugal—A Nationwide Study," *Healthcare*, Vol. 11, No. 14, art. 1991, 2023, doi: 10.3390/healthcare11141991.
[23] N. Syifa, N. Purborini, H. Rachmawati, F. A. Nugrahening Sumadi, and I. Muhammad, "Depression, Anxiety, and Stress Among Health College Students During Covid-19 Outbreak and Online Education," *J. Public Health Dev.*, Vol.2023, No.1, pp. 1-12, 2023, doi: 10.55131/jphd/2023/2103115.
[24] V. Colonnello, E. Fino, and P. M. Russo, "Attachment Anxiety and Depressive Symptoms in Undergraduate Medical Students: The Mediating Role of Emotion Regulation Strategies," *Perspect. Med. Educ.*, Vol. 11, No. 1, pp. 1-12, 2022, doi: 10.1007/s40037-022-00713-z.
[25] S. H. Hamaideh, H. Al-Modallal, M. Tanash, and A. M. Hamdan-Mansour, "Depression, Anxiety and Stress Among Undergraduate Students During COVID-19 Outbreak and 'Home-quarantine,'" *Nurs. Open*, Vol. 8, No. 5, pp. 2372–2381, 2021, doi: 10.1002/nop2.918.
[26] J. Zhu, "The Application Value of Behavior Analysis Based on Deep Learning in the Evaluation of Depression in Art Students," *Appl. Math. Nonlinear Sci.*, Vol 2024, No.1, pp.1-12, 2024, doi: 10.2478/amns-2024-1834.
[27] M. Idkedek *et al.*, "Depression Among Medical Students at West Bank Universities," *Jordan Med. J.*, Vol. 58, No. 1, 2024, doi: 10.35516/jmj.v58i1.2401.
[28] C. T. Suan, A. Khan, and M. Anwar, "Prevalence of Depressive Feelings and Its Associated Risk Factors Among University Students in Malaysia," *Gomal Univ. J. Res.*, Vol. 37, No. 4, 2021, doi: 10.51380/gujr-37-04-01.
[29] S. Phomprasith *et al.*, "Prevalence and Associated Factors of Depression in Medical Students in a Northern Thailand University: A Cross-Sectional Study," *Healthcare*, Vol. 10, No. 3, art. 0488, 2022, doi: 10.3390/healthcare10030488.
[30] Q. Wang *et al.*, "Anxiety and Depression and Their Interdependent Influencing Factors Among Medical Students in Inner Mongolia: The Cross-Sectional Survey," *BMC Med. Educ.*, Vol. 22, No. 1, pp. 383, 2022, doi: 10.1186/s12909-022-03839-0.
[31] M. A. Al Biswas *et al.*, "The Prevalence and Associated Factors of Depressive Symptoms Among Medical Students in Bangladesh During the COVID-19 Pandemic: A Cross-Sectional Pilot Study," *Front. Public Health*, Vol. 9, No. 811345, pp. 1–10, 2022, doi: 10.3389/fpubh.2021.811345.
[32] A. R. Lopes and O. K. Nihei, "Depression, Anxiety and Stress Symptoms in Brazilian University Students During the COVID-19 Pandemic: Predictors and

Association With Life Satisfaction, Psychological Well-Being and Coping Strategies," *Plos One*, Vol. 16, No. 7, pp. e0258493, 2021, doi: 10.1371/journal.pone.0258493.

[33] S. Kandasamy, R. R. John, A. Jayaraman, S. Chinnakutti, M. Chandrasekar, and D. Y. Mallikarjunan, "Immediate Psychological Impact of Dental Students on COVID-19 Epidemic in India – A Cross Sectional Study," *J. Pharm. Bioallied Sci.*, Vol. 13, No. 2, pp. 184–190, 2021, doi: 10.4103/jpbs.jpbs_848_20.

[34] S. Wu, K. Zhang, E. J. Parks-Stamm, Z. Hu, Y. Ji, and X. Cui, "Increases in Anxiety and Depression During COVID-19: A Large Longitudinal Study From China," *Front. Psychol.*, Vol. 12, No. 706601, pp. 1–12, 2021, doi: 10.3389/fpsyg.2021.706601.

[35] M. F. Fauzi *et al.*, "Stress, Anxiety and Depression Among a Cohort of Health Sciences Undergraduate Students: The Prevalence and Risk Factors," *Int. J. Environ. Res. Public. Health*, Vol. 18, No. 6, pp. 3269, 2021, doi: 10.3390/ijerph18063269.

[36] M. Gil, S. Kim, and E. J. Min, "Machine Learning Models for Predicting Risk of Depression in Korean College Students: Identifying Family and Individual Factors," *Front. Public Health*, Vol. 10, No. 1023010, 2022, doi: 10.3389/fpubh.2022.1023010.

[37] C. E. Morr *et al.*, "Predictive Machine Learning Models for Assessing Lebanese University Students' Depression, Anxiety, and Stress During COVID-19," *J. Prim. Care Community Health*, Vol. 2024, No. 1, pp. 1-122024, doi: 10.1177/21501319241235588.

[38] S. Nuarini, S. Fauziah, N. A. Mayangky, and R. Nurfalah, "Comparison Algorithm on Machine Learning for Student Mental Health Data," *J. Med. Inform. Technol.*, Vol. 1, No. 3, pp. 18–26, 2023, doi: 10.37034/medinftech.v1i3.18.

[39] H. Lyu, "Application of Machine Learning on Depression Prediction and Analysis," *Appl. Comput. Eng.*, Vol. 5, No. 20230681, pp. 1, 2023, doi: 10.54254/2755-2721/5/20230681.

[40] J. Chung and J. Teo, "Single Classifier vs. Ensemble Machine Learning Approaches for Mental Health Prediction," *Brain Inform.*, Vol. 10, No. 1, pp. 1–12, 2023, doi: 10.1186/s40708-022-00180-6.