Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

Comparing TabNet and CatBoost Models for Explainable Student Depression Prediction

Triana Dewi Salma[1,*], Rizqi Darmawan[1], Fauzan Natsir[2], Esa Kurniawan[3]

[1] Faculty of Science and Business, Informatics, Universitas LIA, Jakarta, Indonesia
[2] Faculty of Engineering and Computer Science, Informatics Engineering, Universitas Indraprasta PGRI, Jakarta, Indonesia
[3] Faculty of Science and Business, Information Systems, Universitas LIA, Jakarta, Indonesia
Email: [1,*]triana.salma@universitaslia.ac.id, [2]rizqi.darmawan@universitaslia.ac.id, [3]fauzan.natsir@gmail.com, [4]esakurniawan@universitaslia.com
Correspondence Author Email: triana.salma@universitaslia.ac.id
Submitted: **15/05/2026**; Accepted: **22/05/2026**; Published**: 23/05/2026**

Abstract−Early identification of depression risk among students is increasingly important for educational institutions because mental health problems may affect academic engagement, well-being, and learning continuity. This study aims to compare TabNet and CatBoost in predicting student depression risk and to examine their explainability in identifying influential predictors from structured tabular data. The dataset used in this study consists of student-related variables covering personal characteristics, academic conditions, psychological indicators, and lifestyle factors. The experimental procedure included data cleaning, missing value treatment, categorical feature transformation, feature scaling, model training, testing, and interpretation. Model performance was assessed using accuracy, precision, recall, F1-score, and ROC-AUC. Meanwhile, model explainability was examined through attention-based feature importance for TabNet and SHAP-based interpretation for CatBoost. The experimental results indicate that CatBoost produced better overall classification performance, achieving 84.54% accuracy compared with 83.44% for TabNet. CatBoost also obtained higher precision, F1-score, and ROC-AUC values. In contrast, TabNet showed slightly better recall, suggesting stronger sensitivity in detecting students classified as at risk. The interpretation results show that suicidal thoughts, financial stress, academic pressure, sleep duration, study satisfaction, dietary habits, and workload-related variables were consistently relevant to the prediction process. These findings indicate that model selection for student depression prediction should consider not only accuracy, but also sensitivity and interpretability.

Keywords: Student Depression Prediction; Explainable Machine Learning; TabNet; CatBoost; SHAP Analysis

1. INTRODUCTION

Student depression has become a critical issue in educational environments because depressive symptoms may impair psychological well-being, reduce academic performance, weaken learning motivation, disrupt social interaction, and affect students' long-term educational development [1], [2]. University students are frequently exposed to multiple sources of pressure, including academic workload, financial constraints, family expectations, lifestyle changes, and uncertainty regarding future careers, which may increase their vulnerability to depression [1], [3]. Previous studies have also shown that students' mental health conditions became more complex during and after the COVID-19 pandemic, as changes in learning environments, limited social interaction, and prolonged psychological distress disrupted students' academic and social adaptation processes [4], [5], [6]. Therefore, student depression should not be viewed merely as an individual psychological problem, but also as an institutional issue that requires early identification, systematic monitoring, and data-informed support mechanisms.

Traditional depression assessment commonly relies on clinical interviews, psychological screening instruments, and self-reported questionnaires. These approaches remain important in counseling and clinical practice because they allow direct evaluation of psychological symptoms and individual experiences. However, their implementation as institutional monitoring mechanisms may face several limitations. Self-reported instruments can be influenced by response bias, social stigma, delayed disclosure, and students' reluctance to report sensitive psychological conditions [7], [8]. In addition, conventional screening procedures are often conducted periodically rather than continuously, making it difficult for educational institutions to identify risk patterns at an early stage. From an information technology perspective, these limitations indicate the need for data-driven decision support systems that can assist early screening by identifying depression risk patterns from structured student-related data. Such systems are not intended to replace professional diagnosis, but they may support counselors, academic advisors, and institutional decision-makers in prioritizing preventive interventions. Machine learning has gained increasing attention in mental health analytics because it can model complex and non-linear relationships among demographic, academic, psychological, and lifestyle-related variables.

In student depression prediction, machine learning models can process multiple indicators simultaneously and produce risk estimations that may support early warning mechanisms. Prior research has demonstrated that machine learning approaches can be applied to depression prediction and mental health-related classification tasks using structured data, survey-based variables, and heterogeneous behavioral indicators [9], [10], [11]. Recent studies have also emphasized the relevance of explainable and machine learning-based approaches in depression-risk prediction, particularly in health-related contexts where predictive reliability and model transparency are both required [12], [13]. In addition, previous works using XLNet and transformer-based sentiment analysis have shown that advanced learning architectures are effective for classification tasks involving complex digital data [14], [15].

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

However, unlike those studies that focused on textual classification and sentiment analysis, the present study emphasizes explainable tabular prediction by comparing TabNet and CatBoost in student depression prediction. These studies indicate that predictive modeling has the potential to complement conventional assessment by providing scalable and data-informed insights. Nevertheless, the use of machine learning in mental health contexts also introduces important methodological and ethical challenges. High predictive performance alone is not sufficient because depression prediction involves sensitive human-centered decisions that require transparency, accountability, and interpretability.

Interpretability constitutes an essential requirement in educational decision support systems because predictive outputs must be understandable to the stakeholders involved in follow-up actions. In student depression prediction, the model should not merely assign a risk category, but should also indicate the variables that shape the prediction. When this explanatory information is absent, machine learning models may be perceived as opaque systems, which can reduce trust among educators, counselors, and institutional decision-makers. Explainable machine learning responds to this concern by enabling the analysis of model behavior and feature contribution. For structured student data, interpretability is particularly important because depression-related predictions may reflect the combined influence of academic pressure, study satisfaction, financial stress, suicidal thoughts, sleep duration, dietary habits, and family history of mental illness. Accordingly, model selection should consider not only classification performance, but also the ability of the model to generate interpretable and actionable explanations.

TabNet is one of the machine learning models designed specifically for structured tabular data. It uses a sequential attention mechanism that enables adaptive feature selection during the decision-making process [16]. Unlike conventional deep neural networks that often require additional post-hoc interpretation techniques, TabNet provides built-in interpretability through attention-based feature masks. These feature masks allow the model to indicate which input variables receive greater attention across decision steps. Previous studies have shown that TabNet can achieve competitive performance in structured prediction tasks while maintaining interpretable feature selection behavior [16], [17]. This characteristic makes TabNet relevant for student depression prediction because the model can identify influential variables while preserving a certain level of transparency in its internal decision process.

CatBoost is another strong model for structured tabular prediction. It is a gradient boosting algorithm designed to handle categorical variables effectively and reduce prediction bias through ordered boosting [18]. CatBoost has been widely used in heterogeneous tabular data problems because of its robustness, strong predictive performance, and ability to capture non-linear feature interactions. In health-related predictive modeling, CatBoost has demonstrated reliable performance in classification tasks involving clinical, behavioral, and psychological variables [10], [13]. Although CatBoost does not provide an attention mechanism similar to TabNet, it can be interpreted using post-hoc explanation techniques such as Shapley Additive Explanations, commonly known as SHAP [19], [20]. SHAP provides a theoretically grounded approach to estimate the contribution of each feature to model predictions, making it useful for interpreting complex ensemble models in decision support applications.

Despite the growing number of studies applying machine learning to depression prediction, several research gaps remain. First, many previous studies primarily emphasize classification performance and treat interpretability as an additional or secondary component. This creates a limitation because predictive accuracy does not necessarily guarantee practical usability in sensitive domains such as student mental health. Second, existing studies often evaluate individual machine learning models without sufficiently comparing how different explainability mechanisms influence model selection. Third, comparative studies involving TabNet and CatBoost in explainable student depression prediction remain limited, although both models are highly relevant for structured tabular data. TabNet offers built-in attention-based interpretability, while CatBoost offers strong predictive performance supported by SHAP-based explanation. Comparing these two models can provide useful insights into the trade-off between predictive accuracy and interpretability in educational decision support systems.

Based on these gaps, this study compares TabNet and CatBoost for explainable student depression prediction using a structured tabular dataset containing demographic, academic, psychological, and lifestyle-related attributes. The comparison is conducted under the same experimental setting to ensure a fair evaluation of both models. Predictive performance is assessed using accuracy, precision, recall, F1-score, and ROC-AUC, while interpretability is analyzed using TabNet attention-based feature importance and CatBoost SHAP values. This study aims to examine how both models perform in predicting student depression risk, identify influential predictors produced by different explainability mechanisms, and provide system-oriented recommendations for selecting machine learning models in educational mental health decision support systems.

Based on this objective, the main contributions of this study are threefold. First, this study provides a comparative evaluation of TabNet and CatBoost for student depression prediction using structured tabular data. Second, this study analyzes the interpretability behavior of both models by comparing attention-based feature importance and SHAP-based explanation. Third, this study offers practical insights for the development of explainable machine learning-based decision support systems in educational environments, where predictive reliability, transparency, and responsible use must be considered together. Considering the urgency of early depression risk identification among students, this study positions comparative model evaluation as a way to examine how predictive accuracy, sensitivity to at-risk students, and interpretable outputs can be balanced in data-informed screening. Therefore, this research not only contributes to the technical comparison of two machine learning models, but also demonstrates how explainable prediction can support responsible decision-making in educational mental health

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

contexts. This approach supports early screening while keeping model predictions interpretable for educational decision-making.

2. RESEARCH METHODOLOGY

2.1 Research Stages

This study employed a comparative experimental research design to evaluate the predictive performance and explainability of TabNet and CatBoost in student depression prediction. The research was conducted using a structured machine learning workflow consisting of five main stages: dataset preparation, data preprocessing, model development, performance evaluation, and explainability analysis. These stages were designed to ensure that both models were evaluated under comparable experimental conditions and that the resulting findings could provide reliable insights into the relationship between predictive accuracy and interpretability.

The first stage was dataset preparation, in which a publicly available student depression dataset was selected as the empirical basis of the study. The dataset contains student-related attributes representing demographic, academic, psychological, and lifestyle-related conditions. The second stage was data preprocessing, which included missing value treatment, categorical feature transformation, numerical feature scaling, and train-test data splitting. The third stage was model development, where TabNet and CatBoost were trained using the same preprocessed dataset. The fourth stage was model evaluation, in which both models were assessed using accuracy, precision, recall, F1-score, and ROC-AUC. The fifth stage was explainability analysis, where TabNet attention-based feature importance and CatBoost SHAP values were analyzed to identify influential predictors of student depression risk. To provide a clearer overview of the experimental procedure, the research stages are visualized in Figure 1. The figure illustrates the sequential process from dataset preparation to explainability analysis, showing how both TabNet and CatBoost were evaluated under the same experimental workflow.

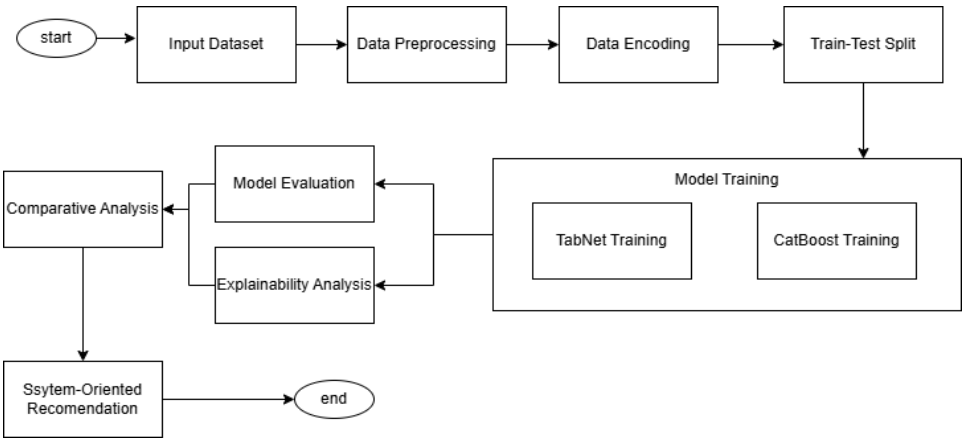

Figure 1. Research Stages for Explainable Student Depression Prediction

The workflow begins with the input dataset and proceeds through preprocessing, model training, performance evaluation, and interpretability analysis. This sequential design allows the study to compare not only classification results but also the explanatory behavior of each model. Therefore, the methodology supports a balanced evaluation of predictive performance and explainability, which is essential in student mental health decision support systems.

2.2 Dataset and Data Preprocessing

The dataset employed in this study was sourced from a publicly available benchmark dataset on student depression [21]. It is presented in a structured tabular format and comprises several categories of student-related variables, namely demographic characteristics, academic indicators, psychological conditions, and lifestyle factors. The demographic variables cover age, gender, and geographic information, while the academic indicators include academic pressure, study satisfaction, cumulative academic performance, and study or work duration. The psychological and lifestyle-related variables consist of suicidal thoughts, family history of mental illness, sleep duration, dietary habits, and financial stress. The outcome variable is depression status, which is formulated as a binary classification target. In this study, label 1 represents students who show depressive symptoms, whereas label 0 represents students without such symptoms.

Prior to model development, preprocessing was conducted to enhance data quality and maintain consistency throughout the experimental workflow. Missing values in numerical attributes were treated using median imputation, as the median is more robust against extreme values than the mean. Meanwhile, missing values in categorical attributes were filled using the mode, which corresponds to the most frequently occurring category in each variable. This strategy was adopted to retain the overall distributional characteristics of the dataset while minimizing potential information loss from incomplete records.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

Furthermore, categorical attributes were transformed into numerical representations using one-hot encoding to ensure compatibility with the modeling pipeline. The first category was omitted during encoding to reduce redundancy and minimize the possibility of multicollinearity among generated features. Numerical attributes were then standardized using z-score normalization to ensure that all numerical features were represented on a comparable scale. This standardization was particularly relevant for TabNet because neural network-based models can be affected by differences in feature magnitude during training. After the preprocessing stage, the dataset was split into training and testing subsets using an 80:20 proportion. Stratified sampling was applied to preserve the class distribution in both subsets, thereby supporting a fairer and less biased evaluation process.

2.3 Model Development

TabNet and CatBoost were developed and compared as the main algorithms for explainable student depression prediction. Both models were selected because they represent different approaches to tabular data classification, where TabNet uses an attention-based deep learning architecture and CatBoost applies a gradient boosting approach for structured and categorical data. TabNet was developed using an attentive interpretable tabular learning architecture [16]. The TabNet architecture was configured with 16-dimensional decision and attention layers and five decision steps to support sequential feature selection during prediction. Sparse regularization of 10^{-4} was applied to encourage focused feature utilization, while the Adam optimizer with a learning rate of 0.02 was used during model training. A batch size of 1024 and a virtual batch size of 128 were applied to support training stability and computational efficiency.

CatBoost was developed as a gradient boosting model that is effective for structured tabular data and categorical feature representation [18]. For CatBoost, the main configuration consisted of 200 boosting iterations, a tree depth of six, a learning rate of 0.1, and Logloss as the objective function for binary classification. The number of iterations and tree depth were selected to allow the model to capture non-linear feature interactions while controlling excessive model complexity. Overall, these configurations were used to ensure that both models were evaluated under comparable experimental conditions in terms of predictive performance, generalization capability, and interpretability analysis. Table 1 summarizes the main model configurations used in this study.

Table 1. Model Configuration for TabNet and CatBoost

Model	Main Configuration
TabNet	Decision dimension = 16; attention dimension = 16; decision steps = 5; sparse regularization = 10^{-4}; learning rate = 0.02; batch size = 1024; virtual batch size = 128
CatBoost	Iterations = 200; depth = 6; learning rate = 0.1; loss function = Logloss

2.4 Model Evaluation and Explainability Analysis

Each model was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. Accuracy represents the overall correctness of predictions on the test set. Precision shows the reliability of positive depression-risk predictions, while recall indicates the ability of the model to identify actual positive cases. The F1-score was used to summarize the balance between precision and recall, especially when false positive and false negative errors must be considered together. ROC-AUC was included to assess the discriminative capability of the model across different classification thresholds. The evaluation metrics used in this study are shown in Equations (1)–(4).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$F1score = 2x\frac{(Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

In these equations, TP represents true positive, TN represents true negative, FP represents false positive, and FN represents false negative. In the context of student depression prediction, recall is particularly important because false negatives may cause students at risk to remain unidentified. However, precision and F1-score are also relevant because institutional decision support systems should avoid excessive false alarms that may reduce trust in the system.

Explainability analysis was conducted using two model-specific approaches. For TabNet, feature importance was obtained from its attention-based feature masks [16]. These masks indicate which features receive greater attention during the prediction process, thereby providing insight into the internal feature selection behavior of the model. For CatBoost, interpretability was analyzed using SHAP values [13], [19], [20]. SHAP quantifies the contribution of each feature to the model's prediction and enables the identification of variables that increase or decrease depression risk classification. By comparing TabNet feature importance and CatBoost SHAP values, this study examines whether both models consistently identify similar influential predictors and how their interpretability mechanisms differ.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

This methodological design allows the study to evaluate TabNet and CatBoost from two complementary perspectives: predictive performance and explainability. The results of this evaluation are presented and discussed in the following section.

3. RESULT AND DISCUSSION

This section presents the experimental results and discussion of the two algorithms compared in this study, TabNet and CatBoost, for explainable student depression prediction. Both algorithms were evaluated using the same testing dataset and the same performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. In addition, their explainability mechanisms were analyzed through TabNet attention-based feature importance and CatBoost SHAP interpretation. The discussion covers model performance, feature importance analysis, SHAP-based interpretation, and comparative implications for educational decision support systems.

3.1 Model Performance Evaluation

The performance evaluation was conducted using the held-out test dataset after the preprocessing and training stages. Both TabNet and CatBoost were evaluated under the same experimental setting to ensure that the comparison reflected model capability rather than differences in data preparation. The performance results are presented in Table 2.

Table 2. Performance Comparison between TabNet and CatBoost

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
TabNet	0.8344	0.8405	0.8853	0.8623	0.8239
CatBoost	0.8454	0.8576	0.8825	0.8699	0.8377

As shown in Table 2, CatBoost achieved an accuracy of 0.8454, while TabNet achieved an accuracy of 0.8344. This result indicates that CatBoost produced a slightly higher proportion of correct predictions on the test data. The difference in accuracy is 0.0110, or approximately 1.10 percentage points. Although this difference is not large, it suggests that CatBoost was more stable in classifying both depressed and non-depressed student cases under the given experimental condition. This finding is consistent with previous studies showing that CatBoost performs strongly on structured tabular data due to its gradient boosting mechanism, ordered boosting strategy, and robustness in handling heterogeneous feature interactions [18].

CatBoost also achieved higher precision than TabNet, with a score of 0.8576 compared to 0.8405. Precision is important because it indicates how reliable the model is when predicting a student as being at risk of depression. A higher precision value means that among the students predicted as positive cases, a larger proportion were actually positive. In institutional decision support systems, precision is relevant because excessive false positive predictions may create unnecessary follow-up burdens for counselors, academic advisors, or student support units. Therefore, the higher precision achieved by CatBoost indicates its advantage in producing more reliable positive predictions.

In terms of recall, TabNet obtained a slightly higher value of 0.8853 compared to CatBoost with 0.8825. Although the difference is small, this result is meaningful in the context of early depression screening. Recall measures the model's ability to correctly identify students who are actually at risk. In mental health-related prediction, false negatives are particularly important because students who are truly at risk but classified as not at risk may not receive timely support. Therefore, TabNet's slightly higher recall suggests that its attention-based architecture may be more sensitive in detecting positive depression-risk cases. However, this advantage should be interpreted carefully because the difference between both models is marginal.

The F1-score results show that CatBoost achieved 0.8699, while TabNet achieved 0.8623. Since F1-score combines precision and recall, this result indicates that CatBoost provided a slightly better balance between correctly identifying positive cases and reducing false positive predictions. This finding is also supported by the ROC-AUC result, where CatBoost achieved 0.8377 compared to TabNet's 0.8239. ROC-AUC measures the model's discriminative ability across classification thresholds. The higher ROC-AUC score of CatBoost indicates that it was slightly better at distinguishing between students with and without depressive symptoms.

Overall, both models demonstrated strong classification capability. CatBoost showed better performance in accuracy, precision, F1-score, and ROC-AUC, while TabNet achieved slightly better recall. This pattern suggests that CatBoost may be more suitable when the system prioritizes balanced predictive performance, whereas TabNet may be considered when sensitivity to at-risk students is the primary concern. However, the small performance gap also indicates that model selection should not be based solely on numerical metrics. In explainable student depression prediction, interpretability and usability must also be considered, particularly because predictions may influence institutional intervention decisions.

3.2 TabNet Feature Importance Analysis

TabNet provides interpretability through its attention-based feature selection mechanism. Unlike conventional deep neural networks that often require post-hoc explanation methods, TabNet generates feature masks that indicate which input variables receive greater attention during the prediction process [16]. This mechanism enables the model to

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

select relevant features sequentially across decision steps. In this study, TabNet feature importance was used to identify variables that contributed most strongly to student depression prediction.

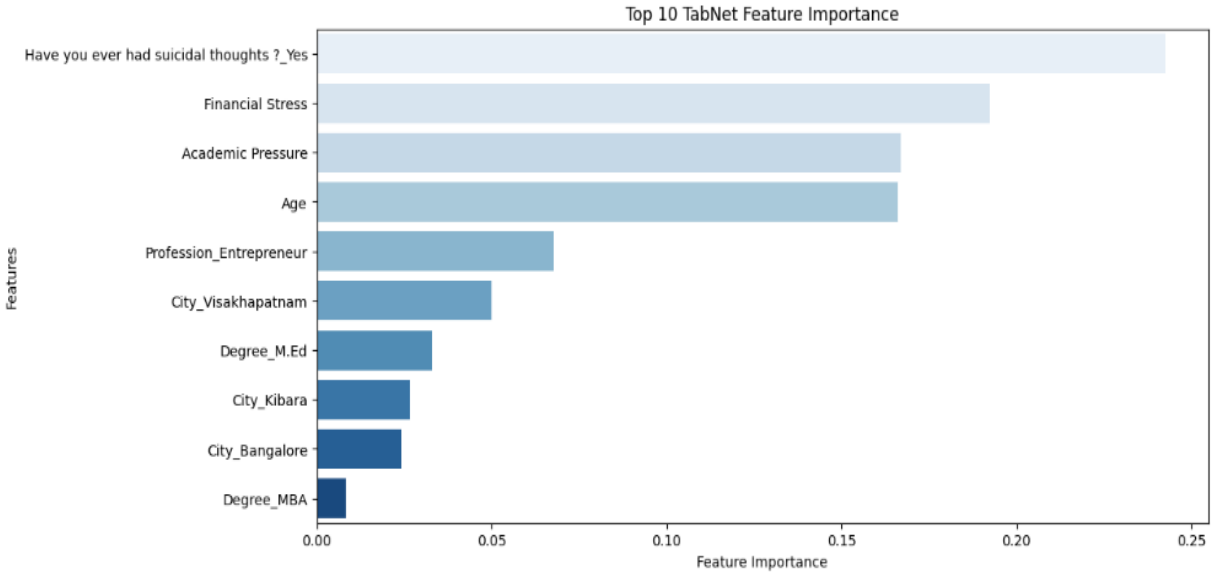

Figure 2. Top 10 TabNet Feature Importance

Based on the TabNet feature importance analysis shown in Figure 2, several variables emerged as important predictors of student depression risk. The most influential features included suicidal thoughts, financial stress, academic pressure, and sleep duration. These variables reflect psychological, socioeconomic, academic, and lifestyle-related dimensions of student mental health. The presence of suicidal thoughts as a dominant feature is clinically and conceptually reasonable because suicidal ideation is closely associated with severe psychological distress and depressive symptoms. In a predictive modeling context, this variable provides a strong signal for identifying students who may require urgent attention.

Financial stress also appeared as an important predictor. This finding supports the argument that socioeconomic pressure can affect students' psychological well-being. Students who experience financial difficulty may face additional burdens related to tuition fees, living costs, family expectations, or the need to work while studying. Such conditions can increase emotional stress and reduce learning focus. Academic pressure was also identified as an influential variable, indicating that educational demands, workload, performance expectations, and study-related anxiety may contribute to depression risk. This aligns with previous findings showing that academic and family stress can significantly affect students' depression levels and academic performance [1].

Sleep duration was another important feature identified by TabNet. This finding is relevant because sleep patterns are closely related to psychological well-being, emotional regulation, concentration, and academic functioning. Inadequate or irregular sleep may increase vulnerability to depressive symptoms, while depression itself may also disrupt sleep quality. The identification of sleep duration as an influential predictor indicates that lifestyle-related factors should not be ignored in student mental health analytics. Student depression prediction should therefore be understood as a multidimensional problem involving psychological, academic, socioeconomic, and behavioral factors.

The attention-based explanation produced by TabNet is valuable because it provides a direct interpretation of feature relevance from the model's internal decision process. This characteristic makes TabNet attractive for applications requiring transparent feature selection. In an educational decision support system, attention-based feature importance can help stakeholders understand which factors are considered important by the model. For example, if the model emphasizes academic pressure and sleep duration, institutions may consider strengthening academic counseling, workload monitoring, time management programs, or student well-being campaigns. However, attention-based importance should not be interpreted as causal evidence. It indicates model-based relevance, not direct causal influence. Therefore, the findings should be used as decision support information rather than definitive clinical conclusions.

3.3 CatBoost SHAP Analysis

CatBoost interpretability was analyzed using SHAP values. SHAP provides a post-hoc explanation approach based on Shapley values, which estimate the contribution of each feature to the model's prediction [19], [20]. Unlike simple feature importance measures, SHAP can explain both the magnitude and direction of feature influence. This makes it useful for interpreting complex ensemble models such as CatBoost, especially in decision support contexts where stakeholders need to understand why a model produces a certain prediction.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

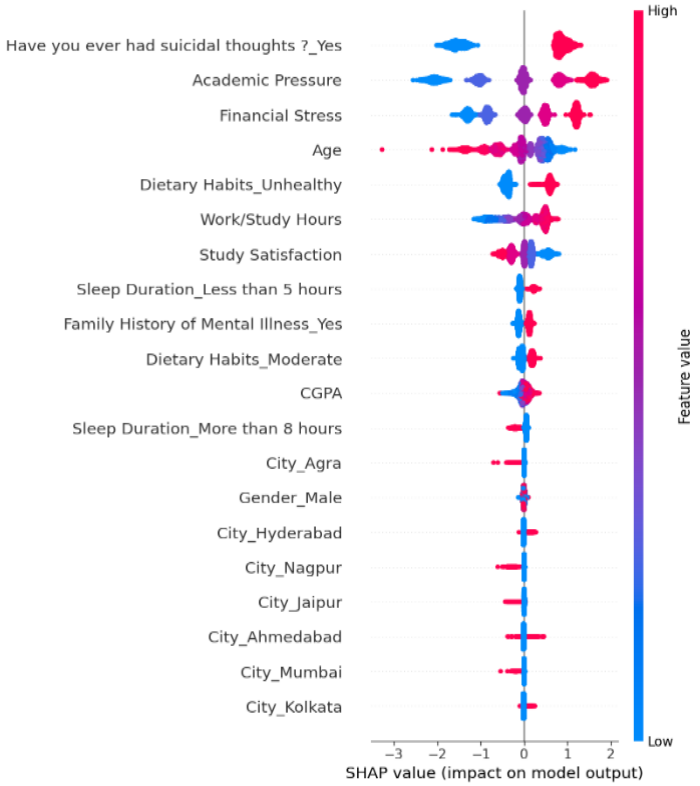

Figure 3. SHAP Summary Plot of CatBoost Model

As shown in Figure 3, the SHAP analysis identified several influential predictors, including suicidal thoughts, financial stress, study satisfaction, dietary habits, sleep duration, and workload-related variables. Similar to TabNet, CatBoost highlighted suicidal thoughts and financial stress as major contributors to depression prediction. This consistency strengthens the reliability of the findings because both models, despite using different learning mechanisms, identified similar psychological and socioeconomic factors as important. Such convergence indicates that these variables contain strong predictive information in the dataset. The SHAP analysis also provided more granular insights into lifestyle and academic-related predictors. Study satisfaction appeared as an influential variable, suggesting that students' subjective evaluation of their academic experience may be related to depression risk. Lower study satisfaction may reflect academic disengagement, dissatisfaction with learning environments, perceived lack of achievement, or mismatch between expectations and actual academic experience. Dietary habits were also highlighted, indicating that lifestyle patterns may contribute to the model's prediction. Although dietary habits should not be interpreted as a standalone determinant of depression, their appearance in the SHAP analysis suggests that behavioral patterns may interact with psychological and academic variables in predicting depression risk.

Workload-related variables also contributed to CatBoost predictions. Students with excessive work or study hours may experience fatigue, reduced rest time, and increased stress. This finding is relevant in contexts where students may combine academic responsibilities with part-time work or family obligations. CatBoost's ability to capture such variables demonstrates its strength in modeling non-linear relationships among heterogeneous features. Since gradient boosting models build sequential decision trees, they can capture feature interactions that may not be easily represented by simpler linear models.

Compared to TabNet's attention-based explanation, SHAP analysis offers a more detailed view of feature contribution. While TabNet identifies which features receive attention during model decision-making, SHAP explains how each feature contributes to increasing or decreasing predicted risk. This distinction is important. TabNet provides built-in interpretability through its architecture, whereas CatBoost relies on post-hoc explanation after the model is trained. In practical terms, TabNet may be easier to justify as an inherently interpretable model, while CatBoost may provide richer explanation outputs through SHAP visualization. Therefore, the choice between both models depends not only on prediction performance but also on the type of explanation required by the decision support system.

3.4 Comparative Discussion

The comparative results indicate that CatBoost and TabNet offer different strengths in explainable student depression prediction. CatBoost achieved slightly higher accuracy, precision, F1-score, and ROC-AUC, suggesting better overall predictive performance. This result is consistent with the general strength of gradient boosting methods in structured tabular data analysis. CatBoost can model complex feature interactions effectively and has been shown to perform well in heterogeneous datasets [13], [18]. Its performance advantage in this study may be explained by the nature of

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

the dataset, which consists of structured survey-based variables, categorical attributes, and mixed psychological, academic, and lifestyle indicators.

TabNet, although slightly lower in overall performance, achieved higher recall. In student depression prediction, this is an important result because recall reflects sensitivity in identifying at-risk students. If the objective of a system is early screening, a model with higher recall may be preferable because it reduces the likelihood of missing students who may need support. However, higher recall should be balanced with precision to avoid generating excessive false positives. In educational institutions, false positives may lead to unnecessary follow-up, increased workload for counselors, and possible anxiety if risk communication is not handled carefully. Therefore, a model should be evaluated not only based on one metric but also based on the institutional purpose of deployment.

From the interpretability perspective, TabNet and CatBoost also differ. TabNet provides built-in interpretability through attention-based feature importance. This is useful when a system designer wants a model whose explanation is embedded in the learning architecture. CatBoost, in contrast, provides strong predictive performance and can be explained using SHAP. SHAP-based explanation offers more detailed information regarding feature contribution but requires an additional post-hoc analysis layer. This difference reflects an important design consideration in explainable machine learning. Built-in interpretability may be simpler to communicate, while post-hoc explainability may provide more flexible and granular interpretation.

The feature importance results from both models show substantial convergence. Suicidal thoughts and financial stress consistently appeared as influential predictors. Academic pressure, sleep duration, study satisfaction, dietary habits, and workload also contributed to depression prediction. These findings indicate that student depression risk is not shaped by a single factor, but by the interaction of psychological, academic, socioeconomic, and lifestyle-related conditions. This multidimensional nature supports the need for holistic student support systems. Educational institutions should not interpret machine learning output only as a classification result, but as an opportunity to identify patterns that may inform preventive programs, counseling priorities, and academic policy interventions.

Compared with previous studies, the findings of this study support the relevance of machine learning for depression prediction and mental health-related decision support. Prior studies have demonstrated that machine learning models can be used to identify depression risk based on structured data and behavioral indicators [9], [10], [11]. This study extends previous work by focusing not only on predictive performance but also on explainability comparison between two different tabular learning paradigms. The comparison between TabNet and CatBoost provides additional insight into how different explanation mechanisms may influence model selection. Rather than treating accuracy as the only indicator of model quality, this study emphasizes the balance between performance, sensitivity, interpretability, and system usability.

The results also have practical implications for the design of student mental health decision support systems. If the primary objective is to maximize balanced classification performance, CatBoost may be a stronger option because it achieved higher accuracy, precision, F1-score, and ROC-AUC. If the system prioritizes sensitivity and built-in interpretability, TabNet may be considered because it achieved slightly higher recall and provides attention-based feature importance. In some cases, a hybrid strategy may also be useful. For example, CatBoost can be used as the main predictive model because of its stronger overall performance, while TabNet can be used as a complementary model to validate whether similar features are selected through attention mechanisms. Such a strategy may improve confidence in model interpretation.

Nevertheless, several limitations should be considered when interpreting the results. First, the dataset used in this study is a structured public dataset, which may not fully represent the complexity of student mental health conditions across different institutions, countries, or cultural contexts. Second, the prediction task is based on static tabular data, whereas depression risk may change over time due to academic cycles, personal events, social conditions, or institutional support. Third, although feature importance and SHAP analysis provide useful interpretability, they do not establish causal relationships. A feature identified as important by the model does not necessarily cause depression; it only indicates predictive relevance within the dataset. Fourth, the use of machine learning for mental health-related prediction requires careful ethical consideration, particularly regarding privacy, data security, fairness, and responsible communication of risk scores.

Based on these considerations, the results of this study should be viewed as a contribution to explainable machine learning evaluation rather than a clinical diagnostic framework. The proposed comparison provides evidence that both TabNet and CatBoost can support student depression prediction, but their deployment in real-world educational settings should involve counselors, psychologists, data governance teams, and institutional policymakers. Machine learning predictions should be used as supporting information for early screening and prevention, not as a substitute for professional assessment. Future studies should validate the findings using longitudinal datasets, institution-specific student records, and multimodal data sources such as academic performance trajectories, learning management system activity, and behavioral indicators. Further research may also explore ensemble or hybrid explainable learning frameworks that combine the predictive strength of CatBoost with the inherent interpretability of TabNet.

Overall, the findings demonstrate that CatBoost provides slightly stronger predictive performance, while TabNet offers competitive performance with attention-based interpretability and slightly higher recall. The consistent identification of suicidal thoughts, financial stress, academic pressure, sleep duration, study satisfaction, dietary habits, and workload as influential predictors indicates that explainable machine learning can provide meaningful insights

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

into student depression risk. Therefore, the main value of this study lies not only in comparing two algorithms, but also in showing how predictive models can be interpreted and positioned responsibly within educational decision support systems.

4. CONCLUSION

This study concludes that TabNet and CatBoost are both applicable for explainable student depression prediction using structured tabular data, but they offer different strengths in predictive performance and interpretability. CatBoost showed stronger overall classification performance by achieving higher accuracy, precision, F1-score, and ROC-AUC, with an accuracy of 84.54% compared to 83.44% for TabNet, indicating its advantage when balanced predictive reliability is prioritized. However, TabNet achieved slightly higher recall, suggesting better sensitivity in identifying students at risk of depression, which is important for early screening contexts where false negatives should be minimized. The explainability analysis showed that suicidal thoughts, financial stress, academic pressure, sleep duration, study satisfaction, dietary habits, and workload-related variables were consistently relevant predictors, confirming that student depression risk is influenced by psychological, academic, socioeconomic, and lifestyle-related factors. These findings indicate that CatBoost may be more suitable for decision support systems requiring stronger overall performance, while TabNet may be considered when built-in interpretability and higher sensitivity are needed. Nevertheless, the findings should be interpreted as decision-support evidence rather than clinical diagnosis because the study used a static public dataset and did not capture longitudinal or institution-specific mental health dynamics. Future studies should validate the models using longitudinal, institutional, and multimodal student data while addressing privacy, fairness, and responsible deployment in educational mental health screening systems.

REFERENCES

[1] Y. Deng et al., “Family and Academic Stress and Their Impact on Students’ Depression Level and Academic Performance,” Frontiers in Psychiatry, vol. 13, p., 2022, doi: 10.3389/fpsyt.2022.869337.
[2] G. Gómez-García, M. Ramos-Navas-Parejo, J. C. D. L. Cruz-Campos, and C. Rodríguez-Jiménez, “Impact of COVID-19 on University Students: An Analysis of Its Influence on Psychological and Academic Factors,” International Journal of Environmental Research and Public Health, vol. 19, p., 2022, doi: 10.3390/ijerph191610433.
[3] G. Ahmed, A. Negash, H. Kerebih, D. Alemu, and Y. Tesfaye, “Prevalence and associated factors of depression among Jimma University students. A cross-sectional study,” International Journal of Mental Health Systems, vol. 14, p., 2020, doi: 10.1186/s13033-020-00384-5.
[4] C. Son, S. Hegde, A. Smith, X. Wang, and F. Sasangohar, “Effects of COVID-19 on College Students’ Mental Health in the United States: Interview Survey Study,” Journal of Medical Internet Research, vol. 22, p., 2020, doi: 10.2196/21279.
[5] X. Wang, S. Hegde, C. Son, B. Keller, A. Smith, and F. Sasangohar, “Investigating Mental Health of US College Students During the COVID-19 Pandemic: Cross-Sectional Survey Study,” Journal of Medical Internet Research, vol. 22, p., 2020, doi: 10.2196/22817.
[6] B. Choi, G. Shim, B. Jeong, and S.-C. Jo, “Data-driven analysis using multiple self-report questionnaires to identify college students at high risk of depressive disorder,” Scientific Reports, vol. 10, p., 2020, doi: 10.1038/s41598-020-64709-7.
[7] C. Hobbs, G. Lewis, C. Dowrick, D. Kounali, T. Peters, and G. Lewis, “Comparison between self-administered depression questionnaires and patients’ own views of changes in their mood: a prospective cohort study in primary care,” Psychological Medicine, vol. 51, pp. 853–860, 2020, doi: 10.1017/S0033291719003878.
[8] C. Sun, S. Li, D. Cao, F.-Y. Wang, and A. Khajepour, “Tabular Learning-Based Traffic Event Prediction for Intelligent Social Transportation System,” IEEE Transactions on Computational Social Systems, vol. 10, pp. 1199–1210, 2023, doi: 10.1109/TCSS.2022.3170934.
[9] H. Nguyen and H. Byeon, “Predicting Depression during the COVID-19 Pandemic Using Interpretable TabNet: A Case Study in South Korea,” Mathematics, p., 2023, doi: 10.3390/math11143145.
[10] C. Zhang, X. Chen, S. Wang, J. Hu, C. Wang, and X. Liu, “Using CatBoost algorithm to identify middle-aged and elderly depression, national health and nutrition examination survey 2011–2018,” Psychiatry Research, vol. 306, p. 114261, 2021, doi: https://doi.org/10.1016/j.psychres.2021.114261.
[11] S. Arik and T. Pfister, “TabNet: Attentive Interpretable Tabular Learning,” ArXiv, vol. abs/1908.07442, p., 2019, doi: 10.1609/aaai.v35i8.16826.
[12] D. Enkhbayar et al., “Explainable Artificial Intelligence Models for Predicting Depression Based on Polysomnographic Phenotypes,” Bioengineering, vol. 12, no. 2, p. 186, Feb. 2025, doi: 10.3390/bioengineering12020186.
[13] L. Yang et al., “Application of machine learning in depression risk prediction for connective tissue diseases,” Sci Rep, vol. 15, no. 1, p. 1706, Jan. 2025, doi: 10.1038/s41598-025-85890-7.
[14] T. D. Salma, G. A. P. Saptawati, and Y. Rusmawati, “Text Classification Using XLNet with Infomap Automatic Labeling Process,” in 2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), 2021, pp. 1–6. doi: 10.1109/ICAICTA53211.2021.9640255.
[15] T. D. Salma, M. F. Kurniawan, R. Darmawan, and A. Basri, “Analisis Sentimen Berbasis Transformer: Persepsi Publik terhadap Nusantara pada Perayaan Kemerdekaan Indonesia yang Pertama,” J. JTIK (Jurnal Teknol. Inf. dan Komunikasi), vol. 9, no. 2, pp. 757–764, 2025.
[16] J. Si, W. Y. Cheng, M. Cooper, and R. Krishnan, “InterpreTabNet: Distilling Predictive Signals from Tabular Data by Salient Feature Interpretation,” ArXiv, vol. abs/2406.00426, p., 2024, doi: 10.48550/arXiv.2406.00426.
[17] J. Hancock and T. Khoshgoftaar, “CatBoost for big data: an interdisciplinary review,” Journal of Big Data, p., 2020, doi: 10.1186/s40537-020-00369-8.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 268−277
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9964

[18] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.09516
[19] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," May 2017, [Online]. Available: http://arxiv.org/abs/1705.07874
[20] S. Somvanshi, S. Das, S. A. Javed, G. Antariksa, and A. Hossain, "A Survey on Deep Tabular Learning," p., 2024.
[21] "Student Depression Dataset." Accessed: May 15, 2026. [Online]. Available: https://www.kaggle.com/datasets/hopesb/student-depression-dataset