International Journal of Artificial Intelligence in Medical Issues

Volume 4 Issue 1
ISSN 3025-4167
https://doi.org/10.56705/ijaimi.v4i1.432

Research Article

Explainable Machine Learning for Predicting the Mental Health Impact of AI and Digital Platform Usage among Students

Agus Halid[1,*]; Dwi Amalia Purnamasari[2]; Ade Chandra Saputra[3]; Nicodemus Mardanus Setiohardji[4]

[1] Universitas Almarisah Madani, Kota Makassar, Sulawesi Selatan 90245, Indonesia, agushalid@gmail.com
[2] Politeknik Negeri Batam, Kota Batam, Kepulauan Riau 29461, Indonesia, dwiamaliaps@gmail.com
[3] Universitas Palangka Raya, Kota Palangka Raya, Kalimantan Tengah 74874, Indonesia, adechandra@itupr.ac.id
[4] Politeknik Negeri Kupang, Kota Kupang, Nusa Tenggara Timur. 85258, Indonesia, nicoluck81@gmail.com

Correspondence should be addressed to Agus Halid; agushalid@gmail.com

Received 02 December 2025; Revised 06 January 2026; Accepted 25 April 2026; Published 30 November 2026

Abstract:
The increasing use of artificial intelligence and digital platforms among students has created new opportunities for learning support, academic assistance, and digital interaction. However, intensive platform usage may also be associated with mental health concerns, sleep disruption, and negative effects on students' daily life. This study aims to develop and evaluate machine learning models for predicting the overall impact of AI and digital platform usage among students by integrating demographic, behavioral, sleep-related, and mental health-related variables. The dataset consisted of 1,705 student records with features including age, gender, academic level, country, average daily usage hours, most-used platform, sleep hours per night, and mental health score. The target variable was Overall_Impact, categorized into Negative, Neutral, and Positive classes. Six supervised machine learning algorithms were evaluated: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Gradient Boosting. Model performance was assessed using accuracy, precision, recall, F1-score, Cohen's Kappa, MAE, RMSE, ROC-AUC, and confusion matrix. The results showed that Random Forest achieved the best performance, with an accuracy of 99.71%, F1-macro of 99.52%, Cohen's Kappa of 0.9950, and ROC-AUC of 0.9994 on the testing set. Feature importance analysis revealed that Mental_Health_Score, Sleep_Hours_Per_Night, and Avg_Daily_Usage_Hours were the most influential predictors. The findings indicate that machine learning can effectively predict the impact of digital platform usage and provide useful insights for AI-driven health informatics and student well-being monitoring. However, further validation using longitudinal and clinically grounded datasets is recommended.

Keywords: Artificial Intelligence, Digital Platform Usage, Machine Learning, Mental Health, Student Well-Being, Sleep Behavior, Health Informatic.

Dataset link: -

1. Introduction

The rapid adoption of artificial intelligence and digital platforms has transformed the way students access information, complete academic tasks, communicate, and manage daily activities. Platforms such as ChatGPT, Gemini, and other digital learning tools provide immediate access to knowledge and personalized assistance, making them increasingly embedded in students' academic and personal routines. While these technologies offer substantial benefits for learning efficiency and digital literacy, their intensive use also raises concerns regarding psychological well-being, sleep behavior, academic balance, and social interaction.

Mental health among young people has become a major global concern [1], [2]. The World Health Organization reports that approximately one in seven adolescents experiences a mental health condition, with depression, anxiety, and behavioral disorders among the leading causes of illness and disability in this age group. Mental health problems during adolescence and young adulthood may also affect educational participation, social functioning, physical health, and long-term quality of life. Therefore, identifying behavioral and digital factors associated with students' mental well-being is increasingly important in the context of AI-supported education and digital lifestyles.

Digital technology use has been widely discussed as a potential factor associated with mental health and sleep outcomes among students and adolescents. Recent evidence from the WHO Regional Office for Europe reported an increase in problematic social media behavior among adolescents, from 7% in 2018 to 11% in 2022, indicating growing concern regarding unhealthy online habits. However, the relationship between digital platform use and mental health is complex. Some studies suggest that digital platforms can support learning, communication, and access to information, while excessive or problematic use may be associated with poorer well-being, anxiety, sleep disturbance, and reduced daily functioning.

Sleep is another important factor in student well-being. Adequate sleep supports cognitive performance, emotional regulation, concentration, and daily functioning. A systematic review on digital media use and sleep among late adolescents and young adults reported that digital media use was generally associated with shorter sleep duration and poorer sleep quality, particularly when use occurred at bedtime or during the night. This suggests that sleep behavior should be considered when analyzing the broader impact of digital platform usage on students' mental and personal life.

The emergence of generative AI tools has introduced a new dimension to the discussion of students' digital behavior. Unlike conventional social media or search engines, generative AI systems provide interactive, conversational, and task-oriented support that may influence learning behavior, dependency, motivation, and psychological responses. Recent research on compulsive ChatGPT usage has examined its relationship with anxiety, burnout, and sleep disturbance, suggesting that AI platform usage should be studied not only from an academic performance perspective but also from a health informatics and mental well-being perspective.

Artificial intelligence and machine learning provide a promising approach for analyzing multidimensional behavioral health data [3], [4]. In mental health research, machine learning has been increasingly used to detect risk patterns, classify psychological conditions, and support early screening based on digital traces, self-reported information, and behavioral indicators. Compared with traditional statistical analysis, machine learning can model nonlinear relationships among multiple variables, such as age, gender, academic level, daily usage duration, sleep hours, platform preference, and mental health score. Furthermore, explainable AI techniques can help identify which features contribute most strongly to model predictions, making the results more interpretable for researchers, educators, and health professionals.

Despite increasing attention to digital platform use and student well-being, several gaps remain. First, many studies focus separately on academic performance, screen time, or social media behavior, while fewer studies integrate

AI/digital platform usage, sleep patterns, and mental health indicators into a single predictive framework. Second, existing studies often emphasize association rather than predictive modeling, leaving room for machine learning-based approaches that can classify the overall impact of platform usage. Third, model interpretability remains important because prediction accuracy alone is insufficient for health-related decision support. Researchers need to understand which behavioral factors are most influential in predicting negative, neutral, or positive outcomes.

To address these gaps, this study proposes a machine learning-based framework to predict the overall impact of AI and digital platform usage among students. The dataset includes demographic variables, academic level, country, average daily usage hours, most-used platform, sleep hours per night, mental health score, academic performance impact, and overall impact category. Several machine learning algorithms are compared, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Gradient Boosting. Model performance is evaluated using accuracy, precision, recall, F1-score, Cohen's Kappa, MAE, RMSE, ROC-AUC, and confusion matrix. In addition, feature importance analysis is applied to identify the most influential factors associated with the predicted impact [5], [6].

The main contributions of this study are threefold. First, this study develops a predictive model for assessing the impact of AI and digital platform usage on students' mental health-related outcomes. Second, it integrates behavioral, demographic, academic, sleep, and mental health variables within a single machine learning framework. Third, it provides interpretable insights into the factors that contribute to positive, neutral, or negative digital platform impact, supporting future research in AI-driven health informatics and student well-being analytics.

2. Method

Research Design:

This study employed a quantitative experimental research design using a supervised machine learning approach. The main objective was to develop and evaluate predictive models for classifying the overall impact of AI and digital platform usage among students [7], [8]. The classification task focused on predicting whether the use of digital platforms had a positive, neutral, or negative impact on students' academic and personal life, with particular attention to mental health-related indicators and sleep behavior.

The general workflow of this study consisted of several stages: dataset preparation, exploratory data analysis, preprocessing, data splitting, model development, performance evaluation, and feature importance analysis. Several machine learning algorithms were compared to identify the most effective model for predicting the overall impact of digital platform usage.

Data Description:

The dataset used in this study contains information on students' demographic characteristics, AI and digital platform usage behavior, academic impact, sleep duration, mental health score, and overall impact category. The dataset consists of 1,705 records and 11 variables. No missing values were found in the dataset.

The attributes in the dataset include student identity, age, gender, academic level, country, average daily usage hours, most-used digital platform, academic performance impact, sleep hours per night, mental health score, and overall impact. The target variable in this study was Overall_Impact, which consists of three classes: Positive, Neutral, and Negative.

The distribution of the target variable showed that the Negative class was the most frequent class, followed by Positive and Neutral. Specifically, the dataset contained 939 Negative, 499 Positive, and 267 Neutral records. This indicates that the classification task involved an imbalanced multi-class prediction problem.

A summary of the dataset variables is presented in **Table 1**.

Table 1. Distribution of Depression Severity Labels

Variable	Description
Student_ID	Unique identifier for each student
Age	Age of the student in years
Gender	Gender of the student
Academic_Level	Education level of the student
Country	Country or location of the student
Avg_Daily_Usage_Hours	Average number of hours spent daily on AI or digital platforms
Most_Used_Platform	Platform most frequently used by the student
Affects_Academic_Performance	Whether platform usage affects academic performance
Sleep_Hours_Per_Night	Average sleep duration per night
Mental_Health_Score	Score representing students' mental well-being
Overall_Impact	Overall effect of platform usage (Positive, Neutral, or Negative)

Target Variable and Feature Selection:

The target variable used in this study was Overall_Impact. This variable represents the overall effect of AI and digital platform usage on students' lives and consists of three categories: Positive, Neutral, and Negative. The classification of this variable was considered suitable for evaluating how behavioral and health-related factors contribute to the overall impact of digital platform usage.

The feature variables used for model development included:

- Age
- Gender
- Academic_Level
- Country
- Avg_Daily_Usage_Hours

- Most_Used_Platform
- Sleep_Hours_Per_Night
- Mental_Health_Score

The variable Student_ID was removed because it only served as an identifier and did not provide meaningful predictive information. The variable Affects_Academic_Performance was also excluded from the main feature set because it represents another outcome-related variable. Removing this variable helped reduce the possibility of target leakage and ensured that the model focused on demographic, behavioral, sleep, and mental health-related predictors.

Data Preprocessing:

Data preprocessing was conducted to prepare the dataset for machine learning analysis [9], [10]. First, the dataset was inspected for missing values, duplicated information, data types, and class distribution. Since no missing values were found, no imputation method was required.

The numerical variables, including Age, Avg_Daily_Usage_Hours, Sleep_Hours_Per_Night, and Mental_Health_Score, were standardized using StandardScaler. Standardization was applied to ensure that numerical variables had comparable scales, especially for algorithms that are sensitive to feature magnitude, such as Support Vector Machine and K-Nearest Neighbors.

Categorical variables, including Gender, Academic_Level, Country, and Most_Used_Platform, were transformed using one-hot encoding. This encoding technique converted categorical values into binary numerical representations so that they could be processed by machine learning algorithms. Unknown categories in the testing phase were ignored to maintain compatibility between training and testing data.

The preprocessing process was implemented using a pipeline approach to ensure that all transformations were applied consistently during cross-validation, training, and testing [11], [12].

Data Splitting:

The dataset was divided into training and testing sets using an 80:20 split ratio. A stratified sampling strategy was applied to maintain the same class distribution of the target variable in both training and testing sets [13], [14]. As a result, 1,364 records were used for model training, while 341 records were used for independent model testing.

Stratified splitting was important because the target variable was imbalanced. This ensured that each class, including the minority class, remained represented in both subsets.

Machine Learning Models:

Six supervised machine learning algorithms were developed and compared in this study. These models were selected to represent linear, tree-based, distance-based, kernel-based, and ensemble learning approaches.

The models used in this study were:

- Logistic Regression was used as a baseline classifier because of its simplicity, interpretability, and effectiveness in multi-class classification tasks [15].
- Decision Tree was applied because it can model nonlinear relationships and provide interpretable decision rules [16].
- Random Forest was used as an ensemble learning model that combines multiple decision trees to improve classification performance and reduce overfitting [17].
- Support Vector Machine was applied using a radial basis function kernel to capture nonlinear relationships among the features [18].
- K-Nearest Neighbors was used as a distance-based classifier that predicts class membership based on the closest observations in the feature space [19].
- Gradient Boosting was included as a boosting-based ensemble model that sequentially improves weak learners to enhance predictive performance [20].

For models that support class weighting, balanced class weights were applied to reduce the effect of class imbalance.

Model Validation Strategy:

To obtain a more reliable estimate of model performance, 5-fold stratified cross-validation was conducted on the training data [21]. In this process, the training data were divided into five folds while preserving the proportion of each target class. Each model was trained and validated five times, with each fold serving once as the validation subset.

After cross-validation, each model was retrained using the full training set and then evaluated on the independent testing set. This strategy allowed the study to assess both the stability of the models during training and their generalization ability on unseen data.

Evaluation Metrics:

The models were evaluated using several performance metrics to provide a comprehensive assessment of classification performance. The main metrics used in this study were accuracy, precision, recall, F1-score, Cohen's Kappa, MAE, RMSE, ROC-AUC, and confusion matrix [22], [23].

Accuracy was used to measure the proportion of correctly classified instances. Precision measured the proportion of correctly predicted instances among all instances predicted as a specific class. Recall measured the ability of the model to correctly identify instances from each actual class. F1-score was used as the harmonic mean of precision and recall, making it particularly useful for imbalanced classification problems.

Cohen's Kappa was used to measure the agreement between predicted and actual labels while considering the possibility of agreement occurring by chance [24]. In addition, MAE and RMSE were calculated by mapping the

target classes into an ordinal scale, where Negative = 0, Neutral = 1, and Positive = 2. This allowed the study to measure the magnitude of classification errors based on the ordered nature of impact categories.

ROC-AUC was calculated using the one-vs-rest strategy for multi-class classification. Confusion matrices were also generated to provide a detailed view of correct and incorrect predictions for each class.

Feature Importance Analysis:

Feature importance analysis was conducted using the Random Forest model. This analysis was performed to identify the most influential variables contributing to the prediction of overall impact. The importance score of each feature was calculated based on its contribution to reducing impurity across the trees in the Random Forest model.

This step was important because model interpretability is essential in health-related predictive analytics. By identifying influential factors such as Mental_Health_Score, Sleep_Hours_Per_Night, and Avg_Daily_Usage_Hours, the study provides insights into the behavioral and mental health-related variables associated with positive, neutral, or negative digital platform impact.

Experimental Workflow:

The overall experimental workflow of this study can be summarized as follows:

1. The dataset was loaded and inspected.
2. Missing values, data types, and target class distribution were examined.
3. Exploratory data analysis was conducted to understand the distribution of digital platform usage, sleep duration, mental health score, and overall impact.
4. Irrelevant and outcome-related variables were removed from the feature set.
5. Numerical variables were standardized and categorical variables were encoded using one-hot encoding.
6. The dataset was split into training and testing sets using stratified sampling.
7. Six machine learning models were trained and validated using 5-fold stratified cross-validation.
8. The trained models were evaluated on the independent testing set.
9. Model performance was compared using multiple evaluation metrics.
10. Feature importance analysis was conducted to identify the most influential predictors.

3. Result and Discussion

Dataset Characteristics:

The dataset used in this study consisted of 1,705 student records with demographic, behavioral, academic, sleep-related, and mental health-related attributes. The target variable was Overall_Impact, which classified the impact of AI and digital platform usage into three categories: Negative, Neutral, and Positive.

The distribution of the target variable showed that the Negative class was the most dominant category, with 939 records, followed by the Positive class with 499 records, and the Neutral class with 267 records. This distribution

indicates that the dataset was imbalanced, with more than half of the samples categorized as having a negative overall impact.

Table 2. Distribution of the Target Variable

Overall Impact	Number of Records	Percentage
Negative	939	55.07%
Positive	499	29.27%
Neutral	267	15.66%
Total	1.705	100%

The descriptive statistics showed that the average age of students was 20.85 years, with a range between 18 and 24 years. The average daily usage of AI and digital platforms was 5.10 hours, while the average sleep duration was 6.60 hours per night. The mean mental health score was 6.22, indicating a moderate level of mental well-being among the students.

Table 3. Descriptive Statistics of Numerical Variables

Variable	Mean	Std. Dev.	Min	Max
Age	20.85	1.76	18.00	24.00
Avg_Daily_Usage_Hours	5.10	1.68	1.50	8.50
Sleep_Hours_Per_Night	6.60	1.21	3.80	9.60
Mental_Health_Score	6.22	1.28	4.00	9.00

These characteristics suggest that the dataset is suitable for predictive modeling because it contains a combination of demographic, behavioral, and health-related variables that may contribute to the overall impact of digital platform usage.

Correlation Analysis of Numerical Features:

The correlation analysis was conducted to examine the relationship among numerical variables, including age, average daily usage hours, sleep hours per night, and mental health score. The results are presented in **Figure 1**.

The strongest negative correlation was found between Avg_Daily_Usage_Hours and Mental_Health_Score, with a correlation coefficient of -0.83. This indicates that students with higher daily usage of AI and digital platforms tended to have lower mental health scores. A strong negative correlation was also observed between Avg_Daily_Usage_Hours and Sleep_Hours_Per_Night, with a correlation coefficient of -0.82, suggesting that longer digital platform usage was associated with shorter sleep duration.

In contrast, Sleep_Hours_Per_Night and Mental_Health_Score showed a strong positive correlation of 0.79. This finding indicates that students with longer sleep duration tended to report better mental health scores. Meanwhile, age

showed only weak correlations with the other numerical variables, suggesting that age was not a major factor in explaining variations in digital platform usage, sleep duration, or mental health score in this dataset.

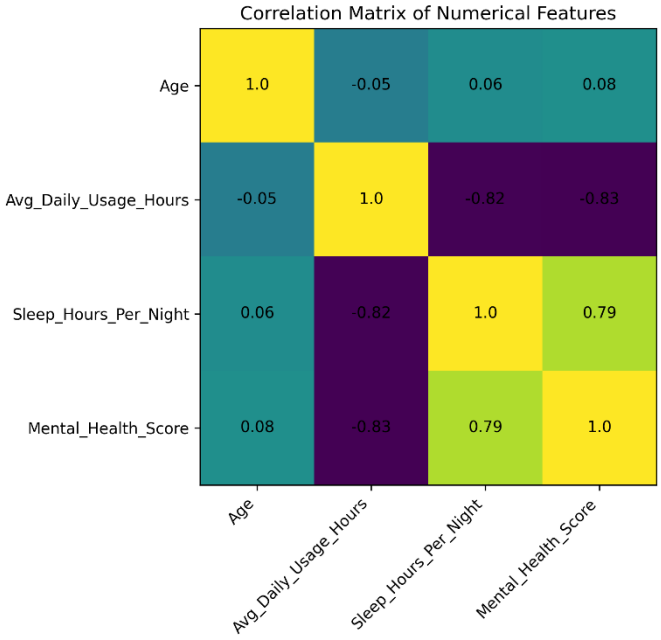

Figure 1. Correlation matrix of numerical features

These findings support the assumption that digital platform usage, sleep behavior, and mental health are closely related. In the context of health informatics, this relationship is important because excessive digital platform usage may be associated with reduced sleep duration and lower mental well-being.

Cross-Validation Performance:

A 5-fold stratified cross-validation was conducted on the training data to evaluate the stability of each machine learning model. The results are shown in **Table 4**.

Table 4. Cross-Validation Results of Machine Learning Models

Model	Accuracy	Precision Macro	Recall Macro	F1-Macro	Cohen's Kappa
Random Forest	9,875	9,881	9,809	9,842	9,787
Gradient Boosting	9,809	9,808	9,721	9,758	9,675
Decision Tree	9,736	9,704	9,627	9,659	9,549
Support Vector Machine	9,700	9,581	9,684	9,629	9,493
K-Nearest Neighbors	9,560	9,468	9,350	9,405	9,245
Logistic Regression	9,531	9,306	9,504	9,398	9,210

Based on the cross-validation results, Random Forest achieved the best overall performance, with an accuracy of 0.9875, F1-macro of 0.9842, and Cohen's Kappa of 0.9787. These results indicate that Random Forest was highly stable across different validation folds.

Gradient Boosting obtained the second-best performance, with an accuracy of 0.9809 and F1-macro of 0.9758. Decision Tree and Support Vector Machine also produced strong results, while K-Nearest Neighbors and Logistic Regression showed slightly lower performance compared with the ensemble-based models.

The strong performance of Random Forest and Gradient Boosting suggests that ensemble learning methods are more effective in capturing the nonlinear relationships among digital platform usage, sleep duration, mental health score, and overall impact.

Testing Performance:

After cross-validation, each model was trained using the full training set and evaluated on the independent testing set. The testing set consisted of **341 records**. The results are shown in **Table 5**.

Table 5. Testing Results of Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Macro	Kappa	MAE	RMSE	ROC-AUC
Random Forest	9,971	9,967	9,937	9,952	9,950	29	542	9,994
Gradient Boosting	9,912	9,917	9,856	9,885	9,850	117	1,326	9,985
Decision Tree	9,853	9,767	9,835	9,800	9,750	176	1,532	9,894
SVM	9,736	9,662	9,660	9,657	9,551	352	2,298	9,972
KNN	9,560	9,369	9,309	9,337	9,245	469	2,298	9,951
Logistic Reg.	9,443	9,204	9,451	9,318	9,064	733	3,294	9,912

The testing results confirm that Random Forest was the best-performing model, achieving an accuracy of 99.71%, F1-macro of 99.52%, Cohen's Kappa of 0.9950, and ROC-AUC of 0.9994. The very low MAE value of 0.0029 and RMSE value of 0.0542 further indicate that the model produced very small classification errors.

Gradient Boosting also demonstrated excellent predictive performance, with an accuracy of 99.12% and F1-macro of 98.85%. Decision Tree achieved an accuracy of 98.53%, showing that even a single tree-based model was able to classify the target variable effectively. However, compared with Random Forest and Gradient Boosting, the Decision Tree model may be more prone to overfitting because it relies on a single decision structure.

Support Vector Machine, K-Nearest Neighbors, and Logistic Regression also produced acceptable results, with accuracy values above 94%. However, their performance was lower than the ensemble-based models. This suggests that the relationship between the features and the target variable may be better represented by nonlinear and ensemble learning approaches.

Confusion Matrix Analysis:

The confusion matrices provide a detailed view of how each model classified the three target classes: Negative, Neutral, and Positive. Overall, the models showed strong classification ability, particularly for the Negative and Positive classes.

The Random Forest model produced the best confusion matrix. It correctly classified all 188 Negative samples and all 100 Positive samples. For the Neutral class, it correctly classified 52 out of 53 samples, with only one Neutral sample misclassified as Positive. This means that Random Forest made only one incorrect prediction out of 341 testing samples.

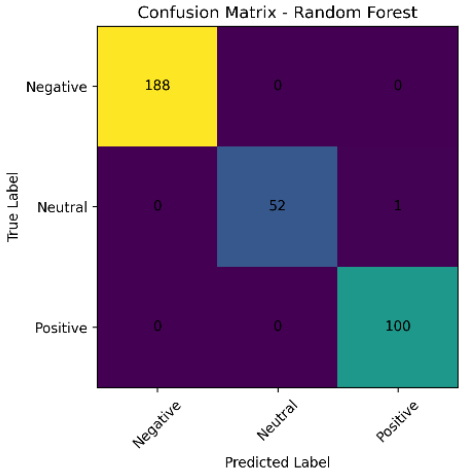

Figure 2. Confusion matrix of Random Forest

The Gradient Boosting model also showed excellent performance. It correctly classified 187 Negative, 51 Neutral, and 100 Positive samples. Only three samples were misclassified. This confirms that Gradient Boosting was also highly effective for this classification task.

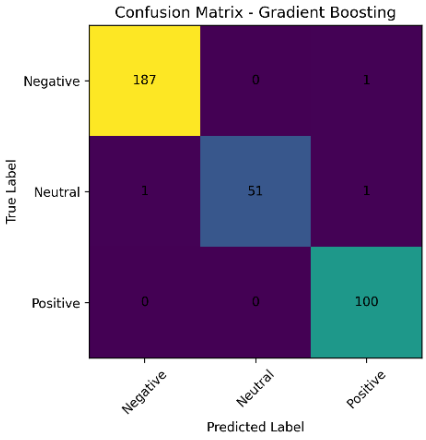

Figure 3. Confusion matrix of Gradient Boosting

The Decision Tree model correctly classified 336 out of 341 testing samples. Although its performance was slightly lower than Random Forest and Gradient Boosting, it still showed strong classification ability. The model made small errors across the Negative, Neutral, and Positive classes.

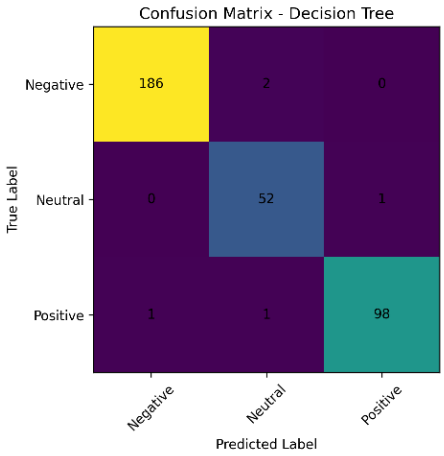

Figure 4. Confusion matrix of Decision Tree

The Support Vector Machine model correctly classified all Positive samples, but it produced several misclassifications in the Negative and Neutral classes. Meanwhile, K-Nearest Neighbors and Logistic Regression showed more classification errors compared with the tree-based ensemble models, particularly in distinguishing the Neutral class from the Negative and Positive classes.

This result indicates that the Neutral class was the most difficult class to classify. This is reasonable because Neutral may represent an intermediate condition between Negative and Positive impacts. In practice, students categorized as Neutral may share characteristics with both groups, making the decision boundary more complex.

Feature Importance Analysis:

Feature importance analysis was conducted using the Random Forest model to identify which variables contributed most strongly to the classification of overall impact. The results showed that the most influential feature was Mental_Health_Score, followed by Sleep_Hours_Per_Night and Avg_Daily_Usage_Hours.

Table 6. Top Feature Importance Based on Random Forest

Rank	Feature	Importance
1	Mental_Health_Score	3,348
2	Sleep_Hours_Per_Night	2,199
3	Avg_Daily_Usage_Hours	1,924
4	Age	293
5	Country_Other	284

Rank	Feature	Importance
6	Most_Used_Platform_LINE	107
7	Academic_Level_High School	98
8	Most_Used_Platform_TikTok	97
9	Most_Used_Platform_Instagram	87
10	Most_Used_Platform_LinkedIn	79

The three most important features, namely Mental_Health_Score, Sleep_Hours_Per_Night, and Avg_Daily_Usage_Hours, contributed approximately 74.71% of the total feature importance. This indicates that the classification of overall impact was mainly determined by mental health condition, sleep duration, and intensity of digital platform usage.

This finding is consistent with the correlation analysis. Higher daily usage was strongly associated with lower mental health scores and shorter sleep duration. Therefore, the model's decision process appears to align with the observed relationships in the data.

In contrast, demographic variables such as age, gender, academic level, and country contributed less to the model. This suggests that behavioral and health-related variables were more influential than demographic characteristics in predicting the overall impact of digital platform usage.

Discussion:

The results of this study demonstrate that machine learning models can effectively predict the overall impact of AI and digital platform usage among students. Among the evaluated models, Random Forest achieved the highest performance in both cross-validation and independent testing. Its superior performance may be attributed to its ability to combine multiple decision trees and capture complex nonlinear relationships among behavioral and health-related variables.

The high predictive performance of Random Forest, Gradient Boosting, and Decision Tree suggests that the dataset has strong separability between the target classes. In particular, the combination of Mental_Health_Score, Sleep_Hours_Per_Night, and Avg_Daily_Usage_Hours appears to provide a clear distinction among Negative, Neutral, and Positive impact categories. Students with higher digital platform usage tended to have lower sleep duration and lower mental health scores, which were associated with negative impact classification. Conversely, students with lower platform usage, longer sleep duration, and higher mental health scores were more likely to be classified as having a positive impact.

The confusion matrix analysis further supports this interpretation. The Negative and Positive classes were classified with very high accuracy across most models, especially Random Forest and Gradient Boosting. However, the Neutral class showed slightly more misclassification. This may occur because Neutral represents a transitional

category between Negative and Positive. Students in this class may have moderate usage patterns, moderate sleep duration, and intermediate mental health scores, making them more difficult to distinguish.

The feature importance analysis provides additional insight into the factors associated with students' digital platform impact. The dominance of Mental_Health_Score indicates that mental well-being was the most important predictor in determining whether digital platform usage had a positive, neutral, or negative impact. Sleep duration was also highly influential, confirming the importance of sleep behavior in student well-being. Furthermore, average daily usage hours played a substantial role, suggesting that usage intensity is an important behavioral indicator in digital health analytics.

From a health informatics perspective, these findings highlight the potential of machine learning as a tool for early identification of students who may be at risk of negative digital platform impact. Predictive models can help educators, health professionals, and institutional decision-makers understand behavioral patterns associated with mental well-being. However, the results should not be interpreted as evidence of causality. The analysis shows predictive association, not direct causal effects.

Although the model performance was very high, this also requires careful interpretation. The strong relationship among daily usage hours, sleep duration, mental health score, and overall impact may indicate that the target classes are highly dependent on these variables. Therefore, future studies should validate the model using external datasets, longitudinal data, and clinically validated mental health instruments. External validation is necessary to ensure that the model can generalize beyond the dataset used in this study.

Overall, the findings indicate that machine learning, particularly Random Forest, can provide accurate and interpretable prediction of the impact of AI and digital platform usage on students. The integration of behavioral usage patterns, sleep behavior, and mental health indicators offers a promising direction for AI-driven health informatics and student well-being analytics.

Summary of Findings:

The main findings of this study can be summarized as follows:

1. Random Forest achieved the best performance, with an accuracy of 99.71%, F1-macro of 99.52%, and ROC-AUC of 99.94%.
2. Gradient Boosting was the second-best model, with an accuracy of 99.12% and F1-macro of 98.85%.
3. The Neutral class was the most difficult class to classify, because it represents an intermediate condition between Negative and Positive impacts.
4. Mental_Health_Score, Sleep_Hours_Per_Night, and Avg_Daily_Usage_Hours were the three most important predictors.
5. The correlation analysis showed that higher digital platform usage was strongly associated with lower sleep duration and lower mental health scores.

6. The findings support the use of machine learning for predictive analytics in student mental health and digital behavior monitoring.

4. Conclusion

This study developed and evaluated machine learning models to predict the overall impact of AI and digital platform usage among students, with particular emphasis on mental health-related indicators, sleep behavior, and daily usage patterns. The dataset consisted of demographic attributes, academic level, country, average daily platform usage, most-used platform, sleep hours per night, mental health score, and overall impact category. The prediction task was formulated as a multi-class classification problem with three categories: Negative, Neutral, and Positive.

The experimental results showed that machine learning models were able to classify the overall impact of digital platform usage with high performance. Among the evaluated models, Random Forest achieved the best performance, with an accuracy of 99.71%, F1-macro of 99.52%, Cohen's Kappa of 0.9950, and ROC-AUC of 0.9994 on the testing set. Gradient Boosting also demonstrated strong performance, followed by Decision Tree, Support Vector Machine, K-Nearest Neighbors, and Logistic Regression. These results indicate that ensemble-based models were more effective in capturing the relationship between digital platform usage behavior and student well-being indicators.

The confusion matrix analysis confirmed that Random Forest produced the most accurate classification results, with only one misclassified sample out of 341 testing records. The model successfully classified all Negative and Positive samples, while the only error occurred in the Neutral class. This suggests that the Neutral category was more difficult to distinguish because it may represent an intermediate condition between positive and negative digital platform impact.

The correlation and feature importance analyses revealed that Mental_Health_Score, Sleep_Hours_Per_Night, and Avg_Daily_Usage_Hours were the most influential variables in predicting the overall impact. Higher average daily usage was strongly associated with lower mental health scores and shorter sleep duration, while longer sleep duration was positively associated with better mental health scores. These findings suggest that excessive use of AI and digital platforms may be linked to poorer sleep behavior and lower mental well-being among students.

Overall, this study demonstrates the potential of machine learning as a predictive approach for analyzing digital behavior and student mental health-related outcomes. The findings contribute to AI-driven health informatics by showing how behavioral, sleep-related, and psychological indicators can be integrated into a predictive framework. Such models may support early identification of students who are potentially at risk of negative digital platform impact.

However, this study has several limitations. First, the dataset used in this study was based on structured records and did not include clinically validated mental health assessments. Second, the analysis was predictive rather than causal, meaning that the results cannot be interpreted as direct evidence that digital platform usage causes changes in mental health or sleep behavior. Third, the very high model performance suggests that the target variable may have a

strong relationship with several input variables, particularly mental health score, sleep duration, and daily usage hours. Therefore, external validation using larger and more diverse datasets is needed.

Future studies should consider using longitudinal data, clinically validated psychological instruments, and additional behavioral indicators such as screen time patterns, night-time usage, social interaction quality, academic workload, and stress levels. Future work may also apply explainable AI methods such as SHAP to provide deeper interpretation of model predictions. In addition, the development of an early warning system based on student digital behavior may be explored to support mental health monitoring and digital well-being interventions.

References:

[1] A. P. Wibowo, M. Taruk, T. E. Tarigan, and ..., "Improving Mental Health Diagnostics through Advanced Algorithmic Models: A Case Study of Bipolar and Depressive Disorders," *Indones. J. ...*, 2024, [Online]. Available: https://jurnal.yoctobrain.org/index.php/ijodas/article/view/122.

[2] Z. Zhu, F. K. S. Chan, G. Li, M. Xu, M. Feng, and Y. G. Zhu, "Implementing urban agriculture as nature-based solutions in China: Challenges and global lessons," *Soil Environ. Heal.*, vol. 2, no. 1, p. 100063, 2024, doi: 10.1016/j.seh.2024.100063.

[3] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. Al Mamun, and M. S. Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction," *Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020*, no. January, pp. 1464–1469, 2020, doi: 10.1109/ICECA49313.2020.9297525.

[4] V. Chaurasia, "Prediction of Heart Disease Using Data Mining Technique," *Int. J. Adv. Eng. Res. Dev.*, vol. 3, no. 01, pp. 208–217, 2016, doi: 10.21090/ijaerd.030137.

[5] A. Naswin and A. P. Wibowo, "Performance Analysis of the Decision Tree Classification Algorithm on the Pneumonia Dataset," *... Artif. Intell. Med. ...*, 2023, [Online]. Available: https://jurnal.yoctobrain.org/index.php/ijaimi/article/view/83.

[6] E. Najwaini, T. E. Tarigan, and F. P. Putra, "Application of the K-Nearest Neighbors (KNN) Algorithm on the Brain Tumor Dataset," *... Artif. Intell. ...*, 2023, [Online]. Available: https://www.jurnal.yoctobrain.org/index.php/ijaimi/article/view/85.

[7] R. Setiawan and H. Oumarou, "Classification of Rice Grain Varieties Using Ensemble Learning and Image Analysis Techniques," *Indones. J. Data ...*, 2024, [Online]. Available: https://jurnal.yoctobrain.org/index.php/ijodas/article/view/129.

[8] F. T. Admojo and N. Rismayanti, "Estimating Obesity Levels Using Decision Trees and K-Fold Cross-Validation: A Study on Eating Habits and Physical Conditions," *Indones. J. Data ...*, 2024, [Online]. Available: https://www.jurnal.yoctobrain.org/index.php/ijodas/article/view/126.

[9] B. D. Finley, *Optimizing Data Pre-Processing Transformations with Reinforcement Learning*. search.proquest.com, 2022.

[10] K. N. Myint and Y. Y. Hlaing, "Predictive Analytics System for Stock Data: methodology, data pre-processing and case studies," *2023 IEEE Conf. Comput. ...*, 2023, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10182047/.

[11] N. Rezova, L. Kazakovtsev, G. Shkaberina, and ..., "Data Pre-Processing for Ecosystem Behavior Analysis," *2022 Int. ...*, 2022, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9897105/.

[12] A. Tuppad and S. D. Patil, "Data Pre-processing Issues in Medical Data Classification," *2023 Int. Conf. ...*, 2023, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10275855/.

[13] V. R. Joseph and A. Vakayil, "SPlit: An Optimal Method for Data Splitting," *Technometrics*, vol. 64, no. 2, pp. 166–176, Apr. 2022, doi: 10.1080/00401706.2021.1921037.

[14] E. K. Ampomah, Z. Qin, and G. Nyame, "Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement," *Information*, vol. 11, no. 6, p. 332, Jun. 2020, doi: 10.3390/info11060332.

[15] L. A. Al-Haddad, "An Intelligent Quadcopter Unbalance Classification Method Based on Stochastic Gradient Descent Logistic Regression," *3rd Information Technology to Enhance e-Learning and Other Application, IT-ELA 2022*. pp. 152–156, 2022, doi: 10.1109/IT-ELA57378.2022.10107922.

[16] S. Payabvash, "Machine Learning Decision Tree Models for Differentiation of Posterior Fossa Tumors Using Diffusion Histogram Analysis and Structural MRI Findings," *Front. Oncol.*, vol. 10, 2020, doi: 10.3389/fonc.2020.00071.

[17] X. Yu, "Random forest algorithm-based classification model of pesticide aquatic toxicity to fishes," *Aquat. Toxicol.*, vol. 251, 2022, doi: 10.1016/j.aquatox.2022.106265.

[18] D. Menaka, "A hybrid convolutional neural network-support vector machine architecture for classification of super-resolution enhanced chromosome images," *Expert Syst.*, vol. 40, no. 3, 2023, doi: 10.1111/exsy.13186.

[19] W. Xing and Y. Bei, "Medical Health Big Data Classification Based on KNN Classification Algorithm," *IEEE Access*, vol. 8, pp. 28808–28819, 2020, doi: 10.1109/ACCESS.2019.2955754.

[20] S. Fafalios, P. Charonyktakis, and I. Tsamardinos, "Gradient Boosting Trees," no. April, pp. 1–13, 2020.

[21] Y. Nie, "Deep Melanoma classification with K-Fold Cross-Validation for Process optimization," *IEEE Med. Meas. Appl. MeMeA 2020 - Conf. Proc.*, 2020, doi: 10.1109/MeMeA49120.2020.9137222.

[22] H. Azis, M. Abdullah, S. Ismail, and ..., "A Comparative Study of YOLO Models for Enhanced Vehicle Detection in Complex Aerial Scenarios," *2025 19th Int. ...*, 2025, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10857527/.

[23] A. R. Manga, H. Azis, F. Fattah, Y. Salim, and ..., "ResNet-50 for Flower Image Classification: A Comparative Study of Segmentation and Non-Segmentation Approaches," *2025 19th ...*, 2025, [Online]. Available:

https://ieeexplore.ieee.org/abstract/document/10857520/.

[24] N. Rismayanti, D. D. Prasetya, T. Widiyaningtyas, and T. Hirashima, “Comparative Study of Random Forest and Ordinal Regression in Concept Map Quality Assessment: The Role of TF-IDF, BERT, and SMOTE-based Balancing,” *Ilk. J. Ilm.*, vol. 17, no. 3, pp. 336–345, 2025, doi: http://dx.doi.org/10.33096/ilkom.v17i3.2906.336-345.