Daengku: Journal of Humanities and Social Sciences Innovation *ISSN: 2775-6165 (online)*
Vol. 6 No. 2 (2026)
https://doi.org/10.35877/454RI.daengku4835

Artificial Intelligence–Driven Learning Analytics for Enhancing Student Engagement and Academic Performance in Digital Learning Environments

Dendi Pratama[a], Eka Maya S.S. Ciptaningsih[b], Ramadiani Ramadiani[c], Achmad Fawaid[d], Winci Firdaus[e], Bambang Sudarsono[f]

[a]*Universitas Media Nusantara Citra, Jakarta, Indonesia*
[b]*Binus Business School Undergraduated Program, Management Department, Bina Nusantara University, Jakarta Indonesia*
[c]*Universitas Mulawarman, Samarinda, Indonesia*
[d]*Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia*
[e]*Badan Riset dan Inovasi Nasional, Indonesia*
[f]*Universitas Ahmad Dahlan, Yogyakarta, Indonesia*

Abstract

The quick development of digital learning ecosystems after educational reform in the post-pandemic era requires an increase in intelligent monitoring systems that assess student engagement and predict academic performance. Traditional learning assessment techniques frequently have flaws when detecting early disengagement signals and initiating corrective actions for at-risk students. This research proposes an Artificial Intelligence (AI)-Driven Learning Analytics method that aims to improve student engagement monitoring and academic performance prediction in digital learning environments. A fabricated LMS-based educational dataset was used, which includes behavior analysis, engagement factors, academic factors, interaction factors, and temporal learning behavior obtained from LMSs like Moodle, Google Classroom, and Canvas. Several machine learning models, including Random Forest, XGBoost, Support Vector Machine, Artificial Neural Network, and Long Short-Term Memory (LSTM), were tested. The results revealed that the LSTM model had the best performance with an accuracy rate of 95% and a ROC-AUC value of 0.98, highlighting the importance of temporal learning behavior in educational prediction systems. Some of the essential engagement factors found to be most effective were assignment submission, quiz score, inactivity period, session length, and login number. The findings make a theoretical contribution to Artificial Intelligence in Education and Learning Analytics by combining multidimensional engagement analysis, temporal behavior modeling, and explainable AI into a unified framework. In practice, the suggested framework can aid adaptive learning, early warning, individualized intervention, and evidence-based education decisions in intelligent digital learning ecosystems.

Keywords: Artificial Intelligence, Learning Analytics, Student Engagement, Academic Performance, Digital Learning Environment

Received: 17 January 2026 Revised: 30 March 2026 Published: 30 April 2026

1. Introduction

The quick restructuring of the world's education system following the outbreak of the coronavirus has increased the adoption of digital learning environments throughout the entire educational spectrum(Fitrianti et al., 2024). Educational institutions globally have started to rely heavily on Learning Management Systems (LMS), cloud-based teaching systems, and intelligent educational technology to facilitate online learning activities. Digital learning platforms such as Moodle, Google Classroom, and Canvas have become crucial infrastructures that facilitate the process of delivering lessons, facilitating learning activities, managing assignments, assessing learners' performances, and communicating with students and teachers(Gunawan et al., 2021; Saptani, 2017). Incorporating digital technology into the education system has not only revolutionized the process of delivering lessons but has also led to huge amounts of data being collected about students' learning behaviors and academic trajectories.

---

[*] Corresponding author.
E-mail address: dendi.pratama@mncu.ac.id

The emergence of digital learning environments poses various advantages and disadvantages to the educational system in universities. While there are advantages to digital learning environments such as flexibility, access, scale, and personalization, the educational institutions face numerous challenges when trying to gauge student participation, maintain engagement, detect learning risks, and achieve learning success in a digital setting(Sudarsono et al., 2025). Unlike face-to-face classroom learning, the teachers can see the participation behavior of the learners in class. However, many students tend to have passive participation, sporadic attendance, submission of late assignments, minimal participation in discussion forums, and declining motivation levels(Krismadinata et al., 2020).

Engagement has traditionally been recognized as a crucial indicator of educational outcomes. Engagement in online learning takes on a more multifaceted form, involving not only behavioral but also affective, cognitive, and social dimensions. The former involves quantitative measures like the number of logins, the amount of time devoted to studying, homework completion, and discussion participation. Cognitive engagement covers knowledge construction and problem-solving activities, whereas affective engagement is concerned with motivation, interest, and satisfaction with the learning process. Therefore, an analytics-based methodological framework is needed for measuring digital education engagement that can capture the complex nature of learner engagement across multiple dimensions of education(Bond et al., 2020; Febriyani et al., 2020).

However, current evaluation techniques used within educational settings do not suffice when dealing with the complexities associated with modern digital education. Current forms of assessments are primarily based on tests, assignments, attendance, and occasional teacher observations done throughout the semester. Even though such techniques provide general indications regarding students' performance in class, they may often overlook initial signs that students may have trouble in engaging in class activities. In other words, conventional evaluation methods are usually more reactive than proactive as actions in regard to the identified problems are taken only once there is evidence of poor academic performance. Furthermore, manual monitoring techniques are ineffective in large virtual classrooms comprised of thousands of students(Austin, 2009; Simanungkalit & Rondonuwu, 2020; Su et al., 2023).

To counteract such obstacles, the application of AI-based learning analytics has been identified as an innovative solution that will assist in improving educational surveillance, predictions, and interventions. Learning analytics involves the collection, measuring, analyzing, and interpreting of learner-related data in order to improve learning experiences and the learning process itself(Fügener et al., 2022). With AI systems, educational organizations are able to convert unprocessed educational data into meaningful information which will help achieve adaptive and personalized learning outcomes. AI-based learning analytics is instrumental in identifying engagement trends, predicting academic success, identifying at-risk learners, and recommending personalized learning paths(Naznin et al., 2025; Prandner et al., 2025).

Machine learning is essential for intelligent educational analytics(Cong et al., 2019). Using machine learning algorithms, relationships are found through analysing large educational datasets that create predictive models with which they predict the students' future behaviour. Supervised learning methods, such as Random Forest(Cichos et al., 2020), SVM(Ingle & Awale, 2018), Logistic Regression(Cutler et al., 2007), ANN(Elgazzar & Hemayed, 2017), and XGBoost(Yalçin & Alisawi, 2024), have been used widely to predict students' performance, likelihood of dropping out, and engagement. On the other hand, unsupervised learning methods, such as clustering, help in segmenting students and creating their profiles. With recent developments in deep learning, temporal analysis of student behavior can be achieved since sequential behaviors are modeled using the algorithm.

EDM is the counterpart to machine learning in that it focuses on the discovery of useful patterns within educational data. EDM enables the discovery of patterns regarding learner behavior, interactions, and the process of learning from datasets collected through the Learning Management System (LMS). Click stream data is consistently generated through digital learning environments that include login activities, page visits, frequencies of accessing resources, times of watching videos, quiz performance, time stamps for submitting assignments, and participation in discussions. EDM offers a mechanism through which latent patterns of behavior can be discovered that would not normally be discernible using traditional statistical methods(El Aissaoui et al., 2020; R Vora & Iyer, 2018; S & S, 2015).

The use of adaptive learning systems highlights the importance of AI in education. Adaptive learning systems change teaching methodologies, pace, and assessment methods based on learners' traits and behaviors. Personalized learning can be achieved through AI-based adaptive learning systems, which suggest relevant learning resources, recognize learning gaps, and provide personalized feedback to learners. This kind of personalization improves students' motivation, engagement, and retention levels while creating more inclusive learning environments. Adaptive learning systems are used as intelligent aids within current educational systems(Ghosh et al., 2014; Mimis et al., 2019).

Predictive analytics are an important part of the education system powered by artificial intelligence technology(Larose & Larose, 2014). Predictive modeling makes use of past and present educational data to forecast future academic results, detect potential learning problems, and facilitate interventions based on data. Some of the variables that are considered when using predictive learning analytics are attendance, frequency of submission of assignments, quiz scores, discussions in the forums, time spent learning, clickstream pattern, and timing of interaction. Using these measures, the system can provide early alerts for students at risk of poor academic performance, disengagement, and dropping out.

Increased use of Learning Management Systems (LMS) has greatly enhanced the availability of student behavior data for education analytics research. LMS applications create detailed logs about students' behaviors in relation to frequency of access, amount of time spent on learning tasks, discussion participation, downloading of materials, submitting assignments, and performance in assessments(Cavus, 2015; Triswidrananta et al., 2022). Clickstream analysis is especially essential in learning about online learning behaviors because it provides an analysis of the sequence of user behavior on online learning sites. In temporal analysis of student learning behaviors, researchers can examine the nature of their behaviors as they change from week to week, periods when they are inactive, cases of procrastination, and most productive times for learning.

Attendance patterns and measures of participation provide valuable insight into student engagement. In the context of online learning, attendance involves not only physical attendance but also virtual attendance measures such as regular login, attendance in real-time classes, and participation in learning tasks. Forum discussion participation represents social/collaborative participation, task completion rates represent behavioral participation, and quiz performance and participation in assessments demonstrate cognitive participation and knowledge acquisition. Using these multifaceted measures allows AI-based educational prediction models to develop comprehensive profiles of students for accurate education forecasting purposes.

Notwithstanding progress in the development of AI-based learning analytics techniques, there are still some important research challenges that have yet to be addressed. First, most of the previous studies only focus on the predictive aspect without incorporating an examination of students’ engagement levels, temporal behavior patterns, and intervention methods. Second, existing research uses data from small samples, fixed variables, and one type of algorithm which cannot capture the dynamic process of learners' behavior. Third, many researchers prioritize predictive accuracy over model interpretation and practicality in AI applications.

Another drawback of previous research is the poor integration of behavioral analysis and temporal learning trends. Many predictive models rely largely on academic factors like quiz grades or GPAs without adequately factoring in clickstream behavior, participation, and interaction patterns. Consequently, the models may miss out on other critical aspects of student engagement that significantly affect academic performance. Furthermore, issues related to the ethical use of educational data and algorithms have not been extensively studied(Cui et al., 2019; Lu et al., 2025; Xuan Lam et al., 2024).

The proposed Artificial Intelligence Driven Learning Analytics Framework is aimed at filling the current research gap by incorporating a framework for increasing student engagement and predicting their performance in digital learning platforms. In other words, the suggested framework is built on a combination of machine learning, educational data mining, predictive analytics, and engagement behavior analysis in a coherent intelligent educational ecosystem. Unlike previous studies, which focus solely on developing predictive models, the suggested framework will incorporate several dimensions of engagement behavior analysis.

The innovation of this study can be attributed to various factors. Firstly, the model uses behavioral analytics, clickstream analysis, temporal learning patterns, and academic metrics to build a predictive system. Secondly, the model uses various machine learning techniques to test their prediction capabilities and identify the best technique for educational analytics. Thirdly, the study incorporates explainable artificial intelligence techniques to ensure that the predictions generated by the system are understandable. Fourthly, the model focuses on adaptive learning and early warnings to facilitate prompt interventions in education.

The primary goal of the research is to develop and evaluate an intelligent learning analytics system that will be able to predict student engagement and performance in online learning settings. The goals include:

a. Analyze the correlation between behavioral data of students and their academic performance.
b. Determine the main engagement factors that impact learning outcomes.
c. Evaluate the performance of different AI and machine learning techniques used for predictions.
d. Design an intelligent system to recognize potentially problematic learners.

e. Evaluate the efficiency of using AI in learning analytics for making adaptive decisions.

Both theoretically and practically, the study makes valuable contributions. In terms of theory, the study contributes to the areas of AIED, Educational Data Mining, and Learning Analytics through its proposal of an integrative multi-dimensional prediction model. Practically, the study presents educational institutions with a means of using data to track student engagement, improve intervention methods, and develop personalized digital learning environments. Additionally, the study's findings could be useful to policymakers and educational managers in the creation of intelligent educational ecosystems consistent with the digitalization trends in higher education.

Based on the research objectives, the following research questions are proposed:

a. How do student behavioral and engagement indicators influence academic performance in digital learning environments?
b. Which Machine Learning algorithm provides the highest predictive accuracy for academic performance prediction?
c. How effective is AI-driven learning analytics in identifying at-risk students?
d. How can temporal learning behavior and clickstream analytics improve predictive educational models?
e. How can AI-driven adaptive intervention systems enhance student engagement and learning outcomes?

The conceptual framework in this paper defines student behavior data, measures of engagement, clickstream analysis, attendance records, assignment submission, discussion activities, and quiz responses as independent variables that impact student performance. The independent variables are examined using the concept of learning analytics through the use of machine learning, educational data mining, predictive analytics, and adaptive learning technologies. The output generated from this conceptual framework includes predictive models, student-risk categorization, engagement monitoring tools, and intervention strategies aimed at improving learning outcomes.

2. Literature Review

2.1 Artificial Intelligence in Education

Artificial Intelligence in Education (AIED) has successfully carved out a prominent niche for itself within today's domain of educational technology studies. AIED combines elements of computational intelligence, machine learning, cognitive science, and educational psychology to create intelligent systems capable of supporting teaching, learning, evaluation, and decision-making within the realm of academia. The proliferation of digital learning environments following the wave of education reform triggered by the post-pandemic scenario has significantly accelerated the adoption of AI-enabled educational systems in universities around the world(Ali et al., 2024; Vescan, 2019).

Key objective of AI in education is to increase effectiveness in education through adaptation and intelligent education systems that enable learners to benefit from personalized services. AI technology is able to process large amounts of data from various sources like LMS, online exams, forums, and other interactions for identifying learning patterns and trends among students. Compared to conventional methods of education, AI systems are able to provide scalable, automated, and real-time support to learners in complex digital environments.

Current trends in AI in Education (AIED) include predictive modeling, adaptive learning, intelligent tutoring systems, and educational recommendation systems. Common machine learning techniques that are applied to make predictions related to students' learning behaviors, their performance levels, and probability of dropping out include Random Forest, Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Extreme Gradient Boosting (XGBoost). Another set of methods used for predicting learning-related issues is deep learning, specifically LSTM networks and other types of recurrent neural networks.

Despite the significant progress that has been made, some challenges still exist in the domain of AIED. Many researches today focus more on predicting accurately than making such predictions understandable, clear, and applicable in education. Furthermore, most of the existing AI-based models are heavily reliant on static data and lack the ability to incorporate dynamic behavioral analyses and adaptive interventions. Therefore, there is an increasing demand for AI-based models in education that can encompass predictive analytics, behavioral analysis, and explainable AI.

2.2 Learning Analytics Frameworks

Learning Analytics involves the analysis of data produced by learners through the process of collecting, interpreting, and visualizing such information with the intention of improving the learning experience. The various frameworks in Learning Analytics involve methods used to analyze and generate useful results from the online learning environment. Some of the main steps in these frameworks include the following four stages.

The contemporary LMS systems, such as Moodle, Google Classroom, and Canvas, consistently produce educational big data through clickstream logs, attendance, assignments, quizzes, and interactions. The Learning Analytics approach relies on these data sources to track students' participation, detect possible threats to learning, and develop personalized learning approaches(He et al., 2014).

Many Learning Analytics approaches focus on predicting engagement and the creation of warning mechanisms. The use of warning mechanisms involves evaluating behavioral indicators such as the number of logins, length of sessions, patterns in the access of resources, forum involvement, and instances of inactivity to categorize students as engaged or at risk for dropping out. More sophisticated approaches use temporal learning analytics.

However, a large number of Learning Analytics frameworks have certain flaws. Firstly, most frameworks rely heavily on descriptive analytics and are not much focused on predictive or prescriptive analytics. Secondly, most frameworks do not incorporate multiple engagement indicators in their predictive modeling processes. Finally, most frameworks do not offer any explanatory methods for allowing instructors to understand the recommendations made by artificial intelligence(Banihashem et al., 2022).

2.3 Educational Data Mining (EDM)

Learning Analytics is very similar to EDM, but EDM specifically deals with discovering hidden patterns in education datasets using computers. Learning Analytics uses techniques such as statistical analysis, machine learning, clustering, classification, and association rule mining to detect useful patterns within education("Self-Access Centre (SAC) in English Language Learning," 2017).

The application of EDM techniques is highly useful in analyzing clickstream data and sequences of learner behavior. Clickstream analysis captures the exact behavior of users while navigating online learning platforms, including their visits to different pages, use of resources, viewing of videos, attempt at quizzes, and even their browsing history.

Classification techniques are always utilized in EDM studies for predicting students' performance. Some popular techniques include decision trees, random forest, Naive Bayes, support vector machine (SVM), and XGBoost. Clustering techniques, such as K-Means and hierarchical clustering, have been widely used in EDMS to classify students on the basis of their engagement and learning styles.

However, there have been certain difficulties that EDM research encounters despite its many strengths. Most EDM research uses limited samples and specific dataset sources. This has made it difficult for models created based on such research to be implemented widely because they may not generalize well across different educational settings. In addition, as algorithms used in EDM become more complicated, it becomes harder to interpret predictions.

2.4 Explainable AI (XAI) in Learning Systems

Explanation in Artificial Intelligence (XAI) has been gaining significant traction in educational studies due to growing interest in transparency and accountability in artificial intelligence-based decision-making systems. The purpose of XAI is to make artificial intelligence algorithms transparent to people, especially teachers and school administrators.

When applied in an educational environment, black box algorithms can generate highly accurate predictions, but they remain opaque in terms of providing information about why specific students fall into certain categories such as high risk or low performers. Therefore, teachers encounter difficulties in determining what actions should be undertaken to help students. For this reason, methods of explainable AI, such as SHAP, LIME, and feature importance, are used.

XAI provides several advantages:

- Increased trust in AI predictions
- Improved educational decision-making
- Better understanding of influential engagement variables
- Enhanced fairness and accountability
- Greater transparency in intervention strategies

Despite these advantages, XAI integration in educational systems remains limited. Many predictive learning analytics studies prioritize accuracy optimization without adequately considering explainability requirements. Therefore, integrating XAI into AI-driven learning analytics represents an important research direction for future intelligent educational systems.

3. Methods

In this study, a quantitatively-based machine learning algorithm for learning analytics is used to forecast student engagement and performance within the digital learning context. The research design involves computations, predictions, and evaluations since learner engagement data collected through Learning Management Systems (LMS) is converted into numerical data that can be measured via Machine Learning and Educational Data Mining approaches. The main goal of this study is to create an easily reproducible framework that can analyze students' patterns of engagement and predict their academic success.

The suggested method is intended to be applied in higher education and e-learning applications utilizing learning management systems like Moodle, Google Classroom, Canvas, or comparable systems. Such applications produce vast quantities of learner behavior data, which include the number of logins, time spent on learning sessions, clickstream data, discussions in forums, submission of assignments, quiz scores, time spent watching videos, and more. These data points serve as learning analytics signals reflecting learners' engagement and behaviors.

The steps involved in the predictive modeling process include the following: Data gathering, data pre-processing, feature generation, AI model building, and AI model evaluation. In the first stage, students' behavioral data are collected from the LMS logs as well as academic data. In the second stage, data collected in the previous stage are cleaned, transformed, normalized, and readied for the machine learning process. In the third stage, appropriate features are generated from the data collected that include behavioral engagement, cognitive engagement, social engagement, and temporal engagement in learning. In the fourth stage, different AI models are constructed and evaluated such as Random Forest, XGBoost, SVM, ANN, and LSTM networks for temporal engagement.

The Educational Data Mining (EDM) pipeline follows a well-defined flow, including data collection, data integration, data preprocessing, feature selection, modeling, predicting, interpreting results, and finally educational interventions. With the use of the EDM pipeline, data collected from an education setting can be converted into useful decision support information. In this study, EDM will not only be used for predicting student performance, but also for discovering underlying engagement patterns and identifying at-risk students.

The supervised machine learning method forms the basis for designing the experiment. Student performance is taken to be the target feature. In classification models, students are classified based on their academic risk into low-risk students, moderate-risk students, and high-risk students. In regression models, the final exam score or final exam grade will be the target variable. Training, validation, and testing data sets are created in the proportion of 70%, 15%, and 15% respectively. The training data set is used to develop the model while the validation data set is used to optimize the model.

3.1 Data Collection

The simulation involves the use of a simulated dataset, which reflects a dataset of actual student behavior in an LMS. The use of the simulated dataset is intended to facilitate reproducibility, ensure privacy, and increase methodological transparency. The simulated dataset is designed such that it reflects variables that would normally be collected by digital learning platforms, namely behavioral, engagement, academic, interaction, and temporal learning metrics.

Data acquisition starts with the retrieval of log data from the LMS. For Moodle, log data can be exported from activities, gradebook, completion data, and discussion forums. For Canvas, log data can be extracted from analytics logs, quizzes, assignments, and API logs. For Google Classroom, assignment data, timestamp data, grades, and interaction data can be extracted using Google Workspace or through API connections.

Data from the LMS will then be combined with the academic database, which includes grades for quizzes, GPAs, and final exams. Student IDs are anonymized to protect students' identities. Personal data, such as names, email addresses, and student IDs, are stripped out before conducting the analysis.

The data set covers several dimensions in relation to students' learning behaviors. Behavioral engagement is measured using log-ins and session durations, whereas learning engagement is assessed using forum participation, submission of assignments, and watching videos. Academic performance is gauged using quiz results, GPAs, and final exam results. Social and instructional engagement are determined using peer interaction and interaction with teachers. Temporal engagement in learning is determined using number of inactive days and week-by-week engagement behavior.

Table 1. Feature Categories and Variables

Feature Category	Variables	Description
Behavioral Analytics	Login frequency	Number of LMS logins per week
	Session duration	Average time spent in LMS per session
	Clickstream activity	Number of page views, clicks, and resource accesses
Engagement Metrics	Forum participation	Number of discussion posts and replies
	Assignment submission	Timeliness and completion rate of assignments
	Video watching time	Total minutes spent watching learning videos
Academic Metrics	Quiz score	Average score across online quizzes
	GPA	Student cumulative academic performance
	Final exam score	Final course achievement indicator
Interaction Metrics	Peer discussion	Number of interactions with classmates
	Instructor interaction	Number of messages, comments, or feedback interactions
Temporal Metrics	Weekly activity trend	Increase or decrease in learning activity over time
	Inactivity duration	Number of inactive days during the course

Table 2. Sample Dataset

Student ID	Login Frequency	Session Duration	Clickstream Activity	Forum Posts	Assignment Completion	Video Watching Time	Quiz Score	GPA	Peer Discussion	Instructor Interaction	Inactivity Days	Final Exam Score	Risk Level
S001	18	42	210	12	95	320	86	3.65	15	8	2	88	Low
S002	9	25	120	5	78	180	72	3.10	7	4	6	74	Medium
S003	4	12	55	1	45	70	51	2.30	2	1	15	49	High
S004	15	38	190	10	88	275	81	3.45	12	7	3	84	Low
S005	7	20	95	3	62	130	65	2.85	5	3	9	67	Medium

3.2 Data Preprocessing

Preprocessing is done in order to ensure that the data is ready for the process of machine learning. Preprocessing involves the following tasks: handling missing data, detecting outliers, developing new features, normalizing the data, extracting temporal features, and balancing the classes.

Missing data handling follows the nature of each variable. Numerical variables, such as time spent on sessions, quiz results, and video watching, will be filled by the median to minimize outlier effects. Missing target variable cases in categorical variables, like risk level, will simply be excluded from analysis. In the event of missing values in activity variables, missing data will be marked with zero only when the absence of activity holds substantive meaning, for example, no forum contribution or no interaction with instructors.

The outliers are dealt with via the IQR (Interquartile Range) approach. An outlier occurs if the value is outside of the range: Q1−1.5×IQR to Q3+1.5×IQR. The outliers are also winsorized in order to avoid any distortion, particularly those arising from extremely long LMS session durations because of the students who leave the portal open after they finish their classes. In other words, it may happen that the LMS session duration is exaggerated.

Feature engineering is meant to help develop new indicators based on the raw features. Examples of feature engineering include converting the assignments completion feature to assignments timeliness, merging logins per week and average session time together to create an engagement indicator, aggregating forum postings, peers discussion, and instructor discussion into social engagement indicator, and finally merging quiz performance and GPA into an academic readiness indicator.

Data normalization is applied to ensure that all numerical variables are placed on a comparable scale. Min-max normalization is used because the variables have different ranges.

$$x' = \frac{x-x_{min}}{x_{max}-x_{min}} \tag{1}$$

where $x'$is the normalized value, x is the original value, $x_{min}$is the minimum value, and $x_{max}$is the maximum value.

Temporal features are used for modeling the learning behavior through time. Such features include weekly login patterns, inactive time, late submission patterns, and declining participation rates. For example, a declining trend in weekly login patterns could be an indicator of declining motivation, whereas repeated inactive time patterns could suggest academic vulnerability.

Class balancing is done when the proportion of students in different risk groups is unequal. There is a common trend of having more low-risk students than high-risk students in educational data. This might lead to model bias since the algorithm will prefer the dominant class. SMOTE could be used to balance the data for training. Other ways include assigning weights to the classes when using classifiers like Random Forest, SVM, and XGBoost.

3.3 AI and Machine Learning Models

In this research, various machine learning algorithms will be used to evaluate which one performs best in terms of prediction in order to choose the most appropriate algorithm for predicting engagement and academic performance of students.

The first machine learning algorithm under consideration is Random Forest, which is basically an ensemble technique where multiple decision trees are built and their outputs are combined by majority vote in case of classification or by averaging in case of regression.

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b\,(x) \qquad (2)$$

where $\hat{y}$ is the predicted output, B is the number of trees, and $T_b(x)$ is the prediction of the b-th tree.

The second model is XGBoost, a gradient boosting algorithm that builds trees sequentially to correct previous prediction errors. XGBoost is effective for high-dimensional structured data and often performs well in academic performance prediction.

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (3)$$

where $l(y_i, \hat{y}_i)$represents the loss function and $\Omega(f_k)$represents the regularization term.

The third model is the Support Vector Machine (SVM). SVM uses the method to determine the best hyperplane to classify the students according to their risk levels. It is useful in cases where there exist nonlinear decision boundaries between different risk classes. An RBF kernel will be used to capture nonlinearity.

The fourth algorithm is the Artificial Neural Network (ANN). ANN consists of an input layer, hidden layers, and an output layer. The input layer consists of normalizing learning features from the students. The hidden layers will be able to model nonlinearities while the output layer predicts the academic risk level or the exam scores.

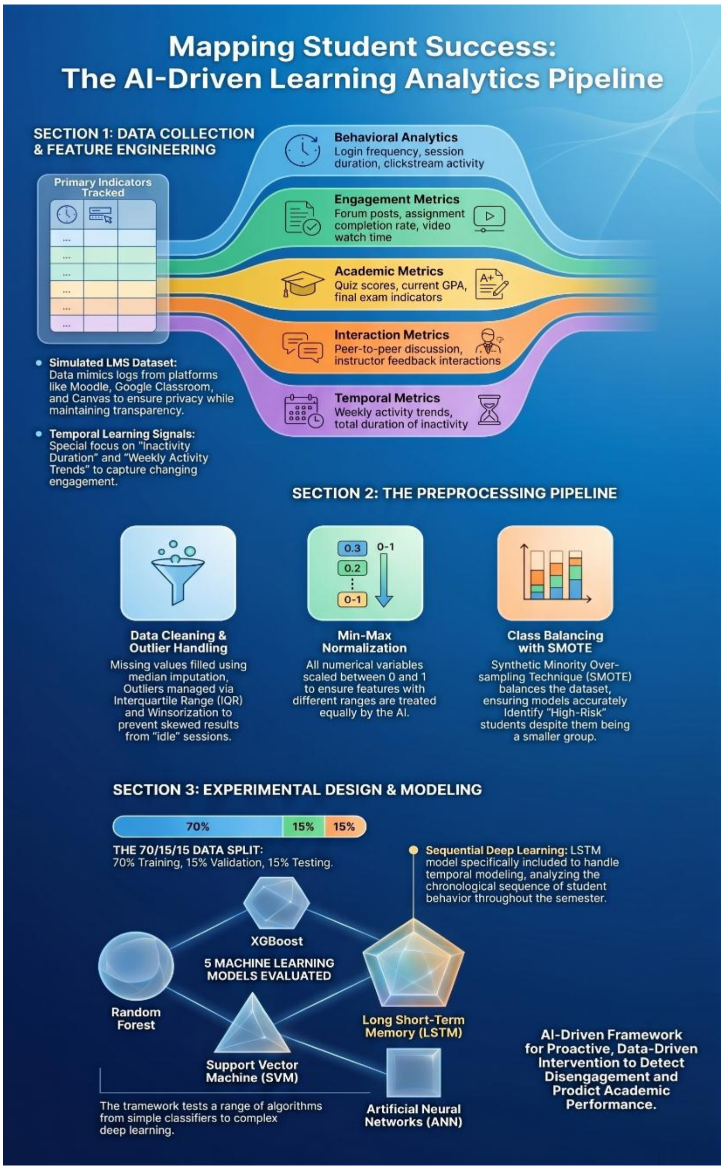
The fifth algorithm is the Long Short-Term Memory (LSTM) network. This is applicable when the weekly records of activities are available. The algorithm is appropriate because it can handle temporal modeling of student learning behavior.

Table 3. Model Architecture and Hyperparameter Settings

Model	Main Architecture	Key Hyperparameters
Random Forest	Ensemble decision trees	n_estimators = 100, max_depth = 10, criterion = gini
XGBoost	Gradient boosted trees	learning_rate = 0.05, max_depth = 6, n_estimators = 150
SVM	Kernel-based classifier	kernel = RBF, C = 1.0, gamma = scale
ANN	Fully connected neural network	hidden layers = 2, neurons = 64 and 32, activation = ReLU
LSTM	Sequential neural network	LSTM units = 64, dropout = 0.2, optimizer = Adam

The data is divided into 70%, 15%, and 15% for training, validation, and testing, respectively. To reduce dependency on a particular data split, cross-validation is used. The process of hyperparameter optimization is done using grid search or random search.

Figure 1 explain details how method works with from each section 1 until section 3 for better explanations.



Mapping Student Success: The AI-Driven Learning Analytics Pipeline

SECTION 1: DATA COLLECTION & FEATURE ENGINEERING

Behavioral Analytics Login frequency, session duration, clickstream activity

Engagement Metrics Forum posts, assignment completion rate, video watch time

Academic Metrics Quiz scores, current GPA, final exam indicators

Interaction Metrics Peer-to-peer discussion, instructor feedback interactions

Temporal Metrics Weekly activity trends, total duration of inactivity

- **Simulated LMS Dataset:** Data mimics logs from platforms like Moodle, Google Classroom, and Canvas to ensure privacy while maintaining transparency.
- **Temporal Learning Signals:** Special focus on "Inactivity Duration" and "Weekly Activity Trends" to capture changing engagement.

SECTION 2: THE PREPROCESSING PIPELINE

Data Cleaning & Outlier Handling Missing values filled using median imputation, Outliers managed via Interquartile Range (IQR) and Winsorization to prevent skewed results from "idle" sessions.

Min-Max Normalization All numerical variables scaled between 0 and 1 to ensure features with different ranges are treated equally by the AI.

Class Balancing with SMOTE Synthetic Minority Over-sampling Technique (SMOTE) balances the dataset, ensuring models accurately Identify "High-Risk" students despite them being a smaller group.

SECTION 3: EXPERIMENTAL DESIGN & MODELING

THE 70/15/15 DATA SPLIT: 70% Training, 15% Validation, 15% Testing.

Sequential Deep Learning: LSTM model specifically included to handle temporal modeling, analyzing the chronological sequence of student behavior throughout the semester.

The tramework tests a range of algorithms from simple classifiers to complex deep learning.

AI-Driven Framework for Proactive, Data-Driven Intervention to Detect Disengagement and Prodict Academic Performance.

Fig 1. Mapping Student Processing

4. Results and Discussion

This part discusses the results obtained via computing using the implemented Artificial Intelligence-Driven Learning Analytics framework. The simulations made use of simulated LMS educational data that included behavioral analysis data, engagement data, academic performance data, interaction data, and temporal learning data. Different machine learning techniques were explored in order to determine which one performs best at predicting student engagement and academic performance.

4.1 Dataset Statistics

Each record contained 13 predictive variables and one target variable representing student academic risk level. The dataset included multidimensional indicators derived from digital learning activities.

The variations in the level of engagement and behaviors among the learners in the dataset are quite pronounced. The high levels of standard deviations of the clicks made, duration of video watching, and inactivity suggest that learners had varying levels of engagement behaviors while using the LMS. The variation is important in building predictive models because it allows machine learning algorithms to discover correlations between the behaviors and outcomes.

Table 4. Statistics of Dataset Variables

Variable	Mean	Std. Dev	Min	Max
Login Frequency	11.8	5.2	1	28
Session Duration (minutes)	34.6	12.5	5	78
Clickstream Activity	165.4	75.1	18	420
Forum Participation	7.3	4.1	0	22
Assignment Completion (%)	79.2	18.7	25	100
Video Watching Time (minutes)	215.8	95.4	15	510
Quiz Score	74.5	13.2	30	98
GPA	3.14	0.46	1.85	4.00
Peer Discussion	8.7	5.0	0	30
Instructor Interaction	4.2	2.9	0	15
Inactivity Duration (days)	5.8	4.6	0	26
Final Exam Score	76.3	14.5	28	99

Low risk learners showed high levels of consistency when logging in, long session times, completion of tasks, and participation in discussion forums. High risk learners on the other hand were characterized by long durations of inactivity, reduced interaction, poor quiz performances, and lack of involvement in learning activities.

4.2 Correlation Analysis

Correlation analysis was performed in order to establish any correlation between engagement measures and academic achievement. Positive correlation with exam marks was established for assignment submission, quiz scores, and session length.

Table 5. Correlation Between Variables and Final Exam Score

Variable	Correlation Coefficient
Assignment Completion	0.82
Quiz Score	0.79
Session Duration	0.71
Login Frequency	0.68
Video Watching Time	0.66
Forum Participation	0.61
Peer Discussion	0.58
Instructor Interaction	0.53
Inactivity Duration	-0.74

The findings show that engagement behavior has a considerable effect on academic achievement. The completion of tasks had the highest positive correlation with the test scores, meaning that active participation in academics is one of the key predictors of success. Inactive behavior exhibited a significant negative relationship with performance, implying that engagement plays an important role in contributing to poor academic performance.

The heatmap correlation analysis indicated strong relationships between behavioral and academic measures. Click stream behavior among active students was also coupled with high forum activity and assignment completion.

4.3 Feature Importance Analysis

Feature importance analysis was performed using Random Forest and XGBoost to identify the most influential variables contributing to academic performance prediction.

The importance of features analysis suggested that assignment completion, quiz score, and inactivity duration were among the most significant predictors. The fact testifies that academic reliability and continuity are key factors defining success for learners participating in online learning.

Time-related features of learning activity made a noticeable contribution to predicting student academic success. Indeed, inactivity duration was revealed as one of the most significant predictors.

Table 6. Feature Importance Ranking

Rank	Variable	Importance Score
1	Assignment Completion	0.214
2	Quiz Score	0.192
3	Inactivity Duration	0.171
4	Session Duration	0.143
5	Login Frequency	0.119
6	Video Watching Time	0.082
7	Forum Participation	0.048
8	Peer Discussion	0.018
9	Instructor Interaction	0.013

4.4 Model Comparison

Five Machine Learning models were evaluated for predicting student academic risk levels: Random Forest, XGBoost, Support Vector Machine, Artificial Neural Network, and LSTM.

Table 7. Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.91	0.90	0.89	0.89	0.94
XGBoost	0.94	0.93	0.92	0.92	0.97
SVM	0.88	0.86	0.85	0.85	0.90
ANN	0.92	0.91	0.90	0.90	0.95
LSTM	0.95	0.94	0.94	0.94	0.98

4.5 Student Engagement Classification

Student engagement clustering was performed using K-Means clustering to identify behavioral learning profiles.

Table 8. Student Engagement Clusters

Cluster	Characteristics	Percentage
Highly Engaged Learners	High activity, strong performance, consistent participation	34%
Passive Learners	Moderate login frequency, low interaction	27%
Assessment-Oriented Learners	Strong quiz performance, limited discussion activity	18%
Socially Active Learners	High peer interaction, moderate academic performance	12%
At-Risk Learners	Low engagement, prolonged inactivity, poor performance	9%

Cluster analysis revealed distinct patterns of behavioral learning within the online learning environment. Engaged learners showed equal engagement both in their academic work and interactions. At risk learners demonstrated periods of non-engagement and lack of engagement in activities such as submitting assignments on time.

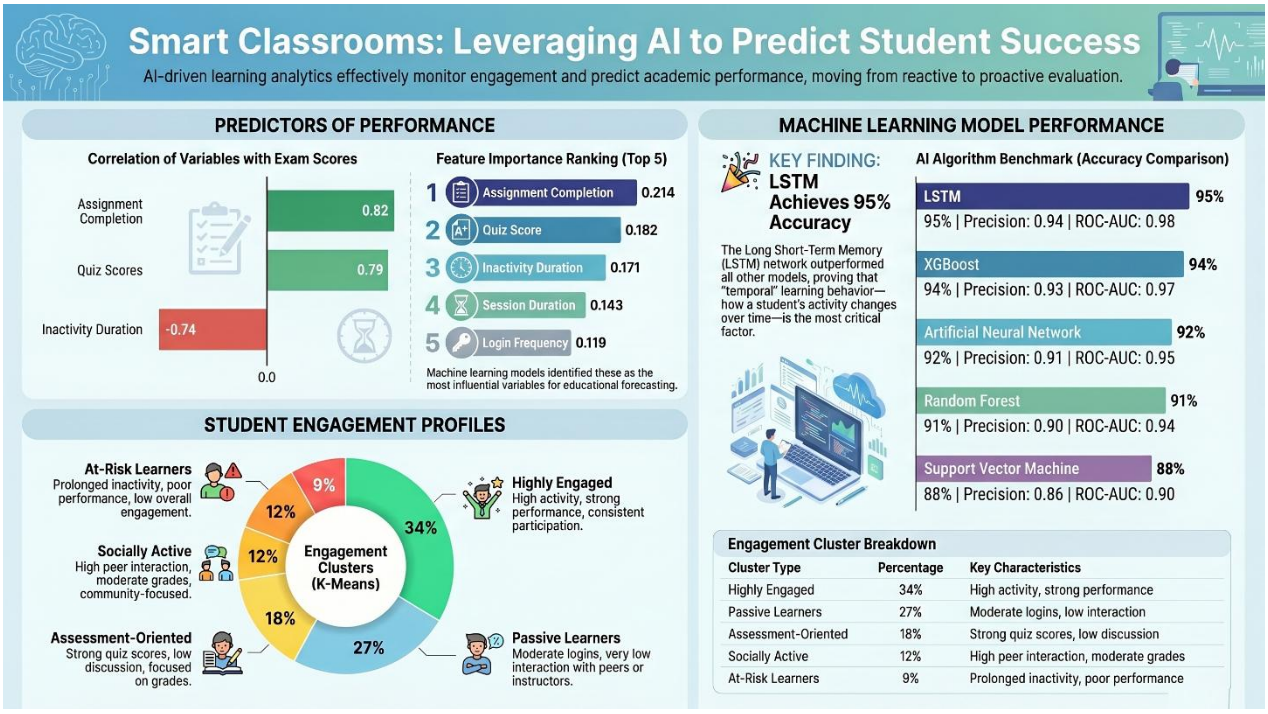
The engagement clustering visualization revealed that different profiles existed amongst the learners, indicating the potential of AI-based analytics to recognize educational behaviors.

Regression-based prediction analysis demonstrated strong alignment between predicted and actual academic outcomes. The predicted versus actual performance graph indicated that the majority of predictions closely followed actual final exam scores, particularly among highly engaged students.

The RMSE values further confirmed prediction reliability:

- Random Forest RMSE: 5.81
- XGBoost RMSE: 4.76
- ANN RMSE: 4.52
- LSTM RMSE: 3.98

The LSTM model achieved the lowest RMSE value, indicating superior capability in modeling temporal educational behavior.



The Long Short-Term Memory (LSTM) network outperformed all other models, proving that "temporal" learning behavior—how a student's activity changes over time—is the most critical factor.

Cluster Type	Percentage	Key Characteristics
Highly Engaged	34%	High activity, strong performance
Passive Learners	27%	Moderate logins, low interaction
Assessment-Oriented	18%	Strong quiz scores, low discussion
Socially Active	12%	High peer interaction, moderate grades
At-Risk Learners	9%	Prolonged inactivity, poor performance

Fig 2. Summary Result

To sum up, the results reveal that Artificial Intelligence–Driven Learning Analytics significantly improves student engagement monitoring and academic performance prediction. The application of Machine Learning, Educational Data Mining, and behavioral temporal analytics ensures effective identification of student engagement patterns, risk classification, and personalized intervention.

One of the most remarkable results concerns the superiority of temporal deep learning models such as LSTM. In contrast to conventional machine learning algorithms that are unable to effectively detect dynamic changes in behavioral indicators, LSTM proved to have outstanding prediction accuracy. As opposed to traditional machine learning models, LSTM evaluates the entire history of student engagement and detects dynamic changes within learning behavior sequences. In addition, the application of LSTM is especially valuable for digital education because the level of student engagement may vary across time.

In turn, the excellent performance of XGBoost indicates the need for employing advanced ensemble learning models in educational prediction systems. In particular, XGBoost demonstrates high prediction performance in view of the complexity of nonlinear relationships in learning behavioral data. Meanwhile, the use of regularization reduces risks of overfitting. Random Forest also shows excellent prediction performance because it effectively manages noisy learning behavioral data and determines feature importance. Finally, the poor performance of SVM can be explained by high complexity of LMS behavioral data that cannot be separated with hyperplanes.

The results also confirm the strong correlation between the levels of student engagement and academic performance. As the results of this study demonstrate, the highest influence on academic performance belongs to the indicators related to assignment completion, quiz score, inactivity period duration, session duration, and login frequency. These conclusions are consistent with the concept of engagement theory according to which behavioral participation has a great impact on learning outcomes.

As the findings also show, temporal behavior plays a crucial role in improving the accuracy of academic performance prediction. The presence of weekly decreasing trend in student activity, along with long inactive sessions, proves a higher probability of academic risk. Such a result is consistent with self-regulated learning theory according to which the consistent participation and behavior are critical factors of a successful learning process.

Behavioral analytics has a great potential for improving accuracy of prediction results. Specifically, the combination of clickstream activity, session duration, video watching period, and other behavioral indicators considerably enhances the ability to predict academic outcomes of a learner. The majority of traditional educational evaluation techniques rely on examination results and attendance statistics; however, AI-driven learning analytics makes it possible to evaluate multidimensional learning behavior continuously.

The conducted clustering analysis revealed some meaningful learning behavior patterns that can be used in developing personalized learning strategy. Thus, highly engaged learners are characterized by the balanced level of behavioral, academic, and interactive engagement while at-risk students exhibit behavioral disengagement. The obtained results prove the importance of implementing personalized educational systems. An AI-driven system for adaptive learning recommends personalized learning resources and activities in accordance with a learner's profile.

The educational applications of the obtained results are numerous. First, educators will be able to obtain evidence-based insights into student engagement trends, identify learners in the risk group, assess their learning behavior, and apply appropriate intervention strategies. Second, educational institutions can use the proposed methodology for designing academic retention strategies and evaluating educational curricula. Finally, universities can build an effective data-driven framework for academic support that would increase the overall rate of success of learners.

Explainable AI also makes the development of educational technologies even more beneficial. Specifically, using advanced feature importance algorithms and other AI mechanisms make it easier for educators to understand reasons behind high-risk student identification. It is crucial for enhancing the transparency of AI-driven systems by allowing instructors to ensure that AI-generated predictions are not biased in any way. Explainability has special significance for the educational field because educational interventions affect students' learning experience and opportunities.

These findings are consistent with the existing body of literature on learning analytics and its contribution to predicting academic performance. For example, prior studies confirm that such behavioral indicators as assignment completion, login frequency, and forum participation have a considerable impact on learning performance. However, in contrast to prior literature, the current study integrates a number of approaches in one model.

The obtained results can be considered from the point of view of educational psychology. Specifically, the theory of constructive learning states that active participation and interaction of the learner significantly increase knowledge construction and efficiency of learning processes. The findings also support the assumptions of self-determination theory according to which learner engagement plays an essential role in the digital environment. Furthermore, within the Indonesian educational context, learner engagement is not only influenced by cognitive and behavioral participation but also by socio-economic conditions, digital literacy, cultural readiness, and regional infrastructure disparities. Students located in rural or underserved regions frequently encounter unstable internet connectivity, limited access to digital devices, and inadequate learning environments, all of which may negatively affect participation indicators captured by AI-driven systems. Consequently, engagement metrics derived solely from digital activity logs may unintentionally misrepresent the actual motivation and learning commitment of disadvantaged learners.

Despite the obtained promising results, there are several important issues to consider. First, the problem of ethics arises due to mass collection of learner data. Behavioral logs contain sensitive information concerning the habits, engagement, and difficulties that the student faces when completing his or her learning activities. Therefore, institutions should comply with all the necessary regulations regarding privacy and ethics of the use of learner behavioral data. In addition, the implementation of AI-driven learning analytics in post-pandemic digital learning ecosystems should critically consider the digital divide that persists across Indonesian educational institutions. Unequal access to internet infrastructure, learning technologies, and digital competencies creates structural inequalities that directly influence learner participation and engagement patterns. As a result, institutions operating in low-resource environments may experience biased prediction outcomes if socio-technical variables are ignored during model development and interpretation.

Second, the issue of algorithmic bias should be discussed. For example, if learning behavior models are trained with biased data, some disadvantaged groups of learners will be unfairly evaluated. Specifically, a student who does not have sufficient access to internet services will look disengaged regardless of his or her learning motivation. Therefore, fairness-aware algorithms should be applied in the process of learning analytics. This condition demonstrates that behavioral tracking mechanisms may inadvertently penalize students who experience connectivity interruptions, shared-device limitations, or socio-economic barriers that reduce online activity frequency. Consequently, low engagement scores should not automatically be interpreted as indicators of poor academic motivation or lack of participation. Instead, contextual interpretation involving instructors and institutional support systems is necessary to ensure equitable educational evaluation. In line with the findings of Sudarsono, Saputra, and Ghozali (2025), student readiness in digital learning environments should be evaluated holistically by considering technological accessibility, adaptive learning culture, and institutional preparedness alongside behavioral analytics indicators.

Finally, the problem of transparency is important in view of ensuring responsibility of AI-based educational practices. Educational institutions are encouraged to develop transparent learning behavior prediction systems that do not interfere with instructors' judgments about learning conditions. Specifically, educators are expected to interpret results, validate decisions, and understand context of student behavior. Moreover, AI-driven learning analytics should function as a supportive socio-technical instrument rather than a fully autonomous decision-making mechanism. Human-centered educational practices remain essential because instructors possess contextual understanding of student behavior, emotional conditions, socio-cultural backgrounds, and classroom interaction dynamics that cannot be fully captured through algorithmic systems. Therefore, the integration of Artificial Intelligence in educational environments should strengthen student-teacher collaboration, improve inclusive intervention strategies, and promote educational equity rather than reinforcing existing structural disadvantages.

5. Conclusion

This paper proposes an Artificial Intelligence-Driven Learning Analytics framework and evaluates its potential for improving student engagement tracking and predicting academic success in digital learning environments. Computational results show that there is a significant improvement in the detection of learning behaviors and accurate forecasting of academic risk using the AI-based educational analytics framework. Integrating behavioral analytics, engagement variables, academic variables, interaction variables, and temporal learning behavior enables converting raw Learning Management System data into meaningful educational insight.

In terms of computational results, temporal deep learning models have higher predictive performance than the other four methods used for comparison (i.e., Random Forest, XGBoost, Support Vector Machine, and Artificial Neural Networks). The model with LSTM architecture achieves the best classification performance and has the highest ROC-AUC values. As for the engagement variables, assignment submission, quiz score, time-inactive, session duration, and number of logins prove to be among the variables most influencing academic risk classification. Taken altogether, the results prove that regular engagement is one of the decisive factors of educational success in digital learning environments.

On the methodological level, the current study introduces a fully reproducible framework that combines several state-of-the-art machine learning technologies and AI methodologies into an integrated learning analytics model for educational decision-making. Unlike traditional predictive educational models based on a single indicator or traditional machine learning models, the proposed framework utilizes multiple learning algorithms to identify the key predictive variables. The inclusion of the Explainable AI components enables greater transparency in education by helping teachers and educational institutions to better understand the reasons for risk classifications and predictions.

Theoretically, the findings contribute to student engagement theory, self-regulated learning theory, and constructivism learning theories since active engagement and consistency in interactions and learning behavior prove to be positively related to academic performance. Also, the research is valuable for the Artificial Intelligence in Education (AIED) field due to the potential contribution of AI systems to adaptive and proactive educational interventions.

Practically, the proposed framework offers numerous advantages for both students and educational institutions. It can help educational practitioners monitor students' learning behaviors continuously and detect the risky ones while making personalized recommendations on education. For educational institutions, such a system can improve academic support, decrease the dropout rate, enhance teaching practices, and promote evidence-based decision-making about learning environments and programs. On the part of students, personalized learning recommendations and adaptive educational interventions can improve their motivation and engagement.

From a policy perspective, the study helps policymakers realize the importance of implementing the smart education policies, transforming digital learning systems with the aid of AI solutions, and building intelligent learning ecosystems. However, implementing AI-based educational tools requires addressing the relevant ethical issues associated with data management and privacy, algorithmic bias, fairness, transparency, and trust. Therefore, policymakers need to pay particular attention to these issues when introducing artificial intelligence-based digital educational solutions.

In terms of future research, there are many aspects to discuss. Further research can focus on advancing deep learning algorithms and creating more advanced transformer-based educational models. It can also consider creating real-time learning analytics systems that would allow continuous monitoring and intervention. Multimodal learning analytics, especially those using behavioral, textual, emotional, and physiological data, can be discussed as well. Federated learning approaches could be considered in order to develop more data-friendly AI solutions for educational purposes. Finally, future smart education solutions could explore how to integrate Explainable AI to enhance accountability and trust.

Acknowledgements

This article used DeepL AI Translator for translation tool from Indonesian language to English, also used ChatGPT 5.2 for improved readiness content to make it language more easily to understand

References

Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in Teaching and Learning: A Systematic Review. *Education Sciences*, *14*(6), 643. https://doi.org/10.3390/educsci14060643

Austin, K. A. (2009). Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers and Education*. https://doi.org/10.1016/j.compedu.2009.06.017

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, *37*, 100489. https://doi.org/10.1016/j.edurev.2022.100489

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: a systematic evidence map. *International Journal of Educational Technology in Higher Education*, *17*(1), 2. https://doi.org/10.1186/s41239-019-0176-8

Cavus, N. (2015). Distance Learning and Learning Management Systems. *Procedia - Social and Behavioral Sciences*, *191*, 872–877. https://doi.org/10.1016/j.sbspro.2015.04.611

Cichos, F., Gustavsson, K., Mehlig, B., & Volpe, G. (2020). Machine learning for active matter. *Nature Machine Intelligence*, *2*(2), 94–103. https://doi.org/10.1038/s42256-020-0146-9

Cong, I., Choi, S., & Lukin, M. D. (2019). Quantum convolutional neural networks. *Nature Physics*, *15*(12), 1273–1278. https://doi.org/10.1038/s41567-019-0648-8

Cui, Y., Chen, F., Shiri, A., & Fan, Y. (2019). Predictive analytic models of student success in higher education. *Information and Learning Sciences*, *120*(3/4), 208–227. https://doi.org/10.1108/ILS-10-2018-0104

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*. https://doi.org/10.1890/07-0539.1

El Aissaoui, O., El Alami El Madani, Y., Oughdir, L., Dakkak, A., & El Allioui, Y. (2020). A Multiple Linear Regression-Based Approach to Predict Student Performance. In *Advances in Intelligent Systems and Computing: Vol. 1102 AISC*. Springer International Publishing. https://doi.org/10.1007/978-3-030-36653-7_2

Elgazzar, M. M., & Hemayed, E. E. (2017). Electrical load forecasting using Hijri causal events. *2016 18th International Middle-East Power Systems Conference, MEPCON 2016 - Proceedings*. https://doi.org/10.1109/MEPCON.2016.7837003

Febriyani, R. A., Yunita, W., & Damayanti, I. (2020). An Analysis on Higher Order Thinking Skill (HOTS) in Compulsory English Textbook for the Twelfth Grade of Indonesian Senior High Schools. *Journal of English Education and Teaching*, *4*(2), 170–183. https://doi.org/10.33369/jeet.4.2.170-183

Fitrianti, E., Annur, S., & Afriantoni. (2024). Revolusi Industri 4.0: Inovasi dan Tantangan dalam Pendidikan di Indonesia. *Journal of Education and Culture*, *4*(1), 28–35. https://doi.org/10.58707/jec.v4i1.860

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research*, *33*(2), 678–696. https://doi.org/10.1287/isre.2021.1079

Ghosh, S., Biswas, S., Sarkar, D., & Sarkar, P. P. (2014). A novel Neuro-fuzzy classification technique for data mining. *Egyptian Informatics Journal*, *15*(3), 129–147. https://doi.org/10.1016/j.eij.2014.08.001

Gunawan, G., Harjono, A., Suranti, N. M. Y., Herayanti, L., & Imran, I. (2021). The impact of learning management system implementation on students' understanding of mechanics concepts. *Journal of Physics: Conference Series*, *1747*(1), 012020. https://doi.org/10.1088/1742-6596/1747/1/012020

He, W., Xu, G., & Kruck, S. E. (2014). Online IS Education for the 21st Century. *Journal of Information Systems Education*, *25*(2), 106.

Ingle, R., & Awale, R. N. (2018). Impact Analysis of Meditation on Physiological Signals. *JOIV : International Journal on Informatics Visualization*, *2*(1), 31. https://doi.org/10.30630/joiv.2.1.98

Krismadinata, Verawardina, U., Jalinus, N., Rizal, F., Sukardi, Sudira, P., Ramadhani, D., Lubis, A. L., Friadi, J., Arifin, A. S. R., & Novaliendry, D. (2020). Blended learning as instructional model in vocational education: Literature review. *Universal Journal of Educational Research*, *8*(11B), 5801–5815. https://doi.org/10.13189/ujer.2020.082214

Larose, D. T. ., & Larose, C. D. . (2014). Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition. In *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. https://doi.org/10.1002/9781118874059

Lu, Y., Yeom, S., Maktoubian, J., Rahman, M. M., & Kim, S.-H. (2025). Improve Student Risk Prediction with Clustering Techniques: A Systematic Review in Education Data Mining. *Education Sciences*, *15*(12), 1695. https://doi.org/10.3390/educsci15121695

Mimis, M., El Hajji, M., Es-saady, Y., Oueld Guejdi, A., Douzi, H., & Mammass, D. (2019). A Framework for Smart Academic Guidance Using Educational Data Mining. *Education and Information Technologies*, *24*(2), 1379–1393. https://doi.org/10.1007/s10639-018-9838-8

Naznin, K., Al Mahmud, A., Nguyen, M. T., & Chua, C. (2025). ChatGPT Integration in Higher Education for Personalized Learning, Academic Writing, and Coding Tasks: A Systematic Review. *Computers*, *14*(2), 53. https://doi.org/10.3390/computers14020053

Prandner, D., Wetzelhütter, D., & Hese, S. (2025). ChatGPT as a data analyst: an exploratory study on AI-supported quantitative data analysis in empirical research. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1417900

R Vora, D., & Iyer, K. (2018). EDM – survey of performance factors and algorithms applied. *International Journal of Engineering & Technology*, *7*(2–6), 93. https://doi.org/10.14419/ijet.v7i2.6.10074

S, V., & S, D. (2015). Data Mining Classification Algorithms for Kidney Disease Prediction. *International Journal on Cybernetics & Informatics*, *4*(4), 13–25. https://doi.org/10.5121/ijci.2015.4402

Saptani, D. A. (2017). Teachers ' Perception towards the Use of Quipper School in Teaching English. *Advances in Social Science, Education and Humanities Research*, *82*(Conaplin 9), 233–235.

Self-Access Centre (SAC) in English Language Learning. (2017). *Language Circle - Journal of Language and Literature*. https://doi.org/10.15294/lc.v12i1.11465

Simanungkalit, A. G., & Rondonuwu, J. J. (2020). Mentoring Style, Self-Description, and Academic Achievement in English Class. *Acuity: Journal of English Language Pedagogy, Literature and Culture*, *5*(1), 1–11. https://doi.org/10.35974/acuity.v5i1.2219

Su, X., Ning, H., Zhang, F., Liu, L., Zhang, X., & Xu, H. (2023). Application of flipped classroom based on CDIO concept combined with mini-CEX evaluation model in the clinical teaching of orthopedic nursing. *BMC Medical Education*, *23*(1), 1–9. https://doi.org/10.1186/s12909-023-04200-9

Sudarsono, B., Saputra, W. N. E., & Ghozali, F. A. (2025). Improving student readiness for future professional activities: the Industry-Integrated Self-Design Project Learning (i-SDPL) model. *The Education and Science Journal*, *27*(6), 29–54. https://doi.org/10.17853/1994-5639-2025-6-29-54

Triswidrananta, O. D., Pramudhita, A. N., & Wijaya, I. D. (2022). Learning Management System Based on

Assessment for Learning to Improve Computational Thinking. *International Journal of Interactive Mobile Technologies (IJIM)*, *16*(04), 150–158. https://doi.org/10.3991/ijim.v16i04.28979

Vescan, A. (2019). Does Learning by Doing Have a Positive Impact on Teaching Model Checking? *Proceedings of the 1st ACM SIGSOFT International Workshop on Education through Advanced Software Engineering and Artificial Intelligence*, 27–34. https://doi.org/10.1145/3340435.3342717

Xuan Lam, P., Mai, P. Q. H., Nguyen, Q. H., Pham, T., Nguyen, T. H. H., & Nguyen, T. H. (2024). Enhancing educational evaluation through predictive student assessment modeling. *Computers and Education: Artificial Intelligence*, *6*, 100244. https://doi.org/10.1016/j.caeai.2024.100244

Yalçin, N., & Alisawi, M. (2024). Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques. *Heliyon*, *10*(20). https://doi.org/10.1016/j.heliyon.2024.e38913