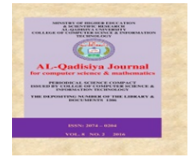




Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)



A comparative study of Student Engagement algorithms in E-learning Environments

Susan Nadhum Ahmed^{a*}, Manar Joundy Hazar^a, Mustafa Jawad Radifi^a

Department of Computer Science, College of Computer Science and Information Technology, University of Al-Qadisiyah.

susannadhum000@gmail.com, manar.Joundy@qu.edu.iq, mustafa.radif@qu.edu.iq

ARTICLE INFO

Article history:

Received: 18 /04/2026

Revised form:16 /05/2026

Accepted: 18 /05/2026

Available online: 30 /09/2026

Keywords:

Student Engagement,
Machine Learning,
Recommendation Systems,
Education Data

ABSTRACT

In e-learning, student engagement is a key factor that influences learning outcomes. Despite this, disengagement and dropout continue to be a problem. In this paper, we compare popular student engagement models to predict and improve the engagement process in e-learning platforms by utilizing machine learning approaches. We rely on behavior interaction data extracted from the EdNet dataset to predict the engagement behavior of students and provide personalized, data-driven recommendations to improve their engagement. We apply the following machine learning models: Logistic Regression, Random Forest, XGBoost, LightGBM, Artificial Neural Networks.

We compare the performance of each model with several metrics. We also use an experimental control design to evaluate the impact of our recommendation on student engagement. Our results show that the tree-based ensemble learning models, especially the XGBoost model, outperform the other models in terms of engagement prediction. In addition, the proposed recommendation system has a positive impact on improving the engagement of students. Our results indicate the effectiveness of our AI-based intervention in early identification and retention of students in online courses.

<https://doi.org/10.29304/jqcm.2026.18.32820>

Introduction

The proliferation of online and blended learning settings has provided a plethora of opportunities for a heterogeneous group of learners. However, online learning systems often experience low levels of student engagement and high attrition rates, resulting in poor academic achievement and a negative impact on institutional reputation [1]. In contrast to face-to-face learning settings, the lack of physical presence makes it challenging for the instructor to monitor students' behavioral and cognitive states, which, in turn, makes it difficult to detect disengagement at an early stage.

Engagement has been conceptualized as a multifaceted construct consisting of three interrelated dimensions, namely, behavioral, cognitive, and emotional engagement [2]. In online learning settings, student engagement manifests itself mainly through behavioral signals such as frequency and duration of login, time-on-task, accessing content, navigation behavior, and performance on assignments [3]. The recent advances in educational data mining (EDM) and learning analytics (LA) have made it possible to mine such behavioral signals for modeling engagement and predicting academic performance using machine learning (ML) and data mining (DM) methods [4], [5].

A recent body of research suggests that ensemble machine learning models are superior to conventional statistical methods in predicting engagement and predicting dropout risk due to their ability to model

*Corresponding author

Email addresses: *Iraq.Email:*

Communicated by 'sub editor'

non-linear relationships and complex interactions among variables in educational data [6], [7]. However, most of these studies have focused on improving the prediction accuracy of the developed models without integrating them into an intervention framework capable of ameliorating student engagement in a real-world setting. In this study, we address this knowledge gap by conducting a comparative analysis of five machine learning algorithms (five most important ML algorithms). As an extended version of this paper, we propose an AI-based framework that fuses an engagement prediction model with a personalized recommendation system. The proposed system not only predicts students at risk of disengagement but also provides personalized interventions aimed at improving their engagement levels and reducing dropout rates. To this end, the first step is to build an accurate engagement prediction model using the available behavioral data.

2. Related Work

Student engagement prediction has received significant attention in learning analytics and educational data mining research. Early studies relied on traditional statistical techniques and rule-based indicators to identify disengaged or at-risk learners. However, such approaches were limited in capturing the complexity and temporal dynamics of student behavior in online environments [8]. With the increasing availability of large-scale educational datasets, machine learning techniques have become the dominant approach for engagement and dropout prediction. Romero and Ventura [9] provided a comprehensive survey demonstrating the effectiveness of classification models such as logistic regression, decision trees, support vector machines, and ensemble methods in educational contexts. Among these, ensemble models consistently show superior performance due to their ability to model non-linear relationships and interactions among behavioral features. Saqr et al. [10] compared logistic regression, random forest, and support vector machines for early prediction of underachieving students and found that tree-based ensemble methods achieved higher predictive accuracy and robustness. Similarly, Bowers and Zhou [11] emphasized the importance of using AUC and F1-score when evaluating engagement and dropout prediction models, particularly in imbalanced educational datasets. Beyond prediction, recent research has explored AI-driven intervention and recommendation systems. Piech et al. [12] introduced deep knowledge tracing to model student learning states over time, while Li et al. [13] demonstrated the effectiveness of neural networks in predicting dropout behavior based on fine-grained behavioral features. However, many studies stop predicting and do not empirically evaluate whether AI-based interventions lead to improved engagement.

Table 1: summary of related works

Reference	Year	Data/Domain	Methods	Focus	Limitations
[14]	2025	Video metadata + text comments	LLM + Metadata	New engagement metric	Only measurement, no enhancement (arXiv)
[15]	2025	Online learning logs	ML (Decision Tree, Boosting)	Engagement prediction	Focus on classification only (MDPI)
[16]	2025	Video affective datasets	Transformer + Sequence Pooling	Video-based engagement	Complex video data, heavy models (arXiv)
[17]	2025	DAiSEE videos	EfficientNetV2 + LSTM	Spatio-temporal engagement	Specialized to video (arXiv)
[18]	2025	Classroom/online settings	Multi-task Model	Bias-aware engagement	Focus on fairness, not enhancement (arXiv)

3. Dataset

The study is conducted using the EdNet dataset, one of the largest publicly available datasets in educational technology research. EdNet contains detailed, time-stamped records of student interactions. We will use EdNet, one of the largest public datasets in the educational technology literature, in our experiments. EdNet is a collection of time-stamped records of students' interactions with an online educational platform, which includes question attempts, correctness, response time, content navigation, and learning material access. The hierarchical structure of EdNet enables both fine-grained (e.g. question attempt) and coarse-grained (e.g. course-level) analyses. Therefore, this dataset is highly appropriate for modeling engagement dynamics and assessing the efficacy of student interventions [19].

The online educational platform which EdNet dataset recorded its students' interactions is mainly designed for standardized test preparation. The platform contains video lectures, practice questions, explanations, and assessment questions. The platform is digital, thus it records all the students' activities, no need for manually collecting the data. EdNet contains Behavioral Features and Academic Features as shown in table 2 and 3.

1- Behavioral Features

Table 2: Behavioral Feature Description

Variable	Description
avg_session_duration	Average session duration (minutes)
login_frequency	Number of login attempts
video_views	Number of videos viewed
time_between_sessions	Time interval between sessions
help_requests	Number of help requests

2- Academic Features

Table 3: Academic Feature Description

Variable	Description
questions_attempted	Number of attempted questions
correct_answers	Number of correct answers
accuracy_rate	Accuracy percentage
repeated_mistakes	Number of repeated errors

Table 4 described the details of CSV file for EdNet dataset

Table 4: Dataset description

Column	Code Name	Description
1	student_id	Student ID
2	avg_session_duration	Average session duration

		(minutes)
3	login_frequency	Login frequency
4	video_views	Number of videos viewed
5	time_between_sessions	Time between sessions
6	help_requests	Help requests
7	questions_attempted	Questions attempted
8	correct_answers	Correct answers
9	accuracy_rate	Accuracy rate
10	content_completion_rate	Completion rate
11	engagement_score	Composite engagement index
12	learning_efficiency	Learning efficiency
13	engagement_level	Target variable

The EdNet dataset is utilized due to its large scale, high granularity, and extensive use in prior educational data mining studies [20]. A stratified random sampling strategy is applied to select 30–60 students while preserving engagement-level distribution. This sampling size balances computational feasibility and statistical validity, as suggested by. The EdNet dataset used in this study, which provides CSV files of student-system interactions, is publicly available for research purposes on github accessed by this URL <https://github.com/rriid/ednet>

Table 5: Data Collection Instruments and Variables

Instrument	Variables Collected	Measurement Frequency	Purpose
Learning Platform Logs	Session duration, login frequency, questions attempted, accuracy rates, video views, navigation patterns, response times	Continuous (automated)	Behavioral data for model training and engagement monitoring
Pre-Experiment Questionnaire	Demographics, prior knowledge, motivation levels, learning preferences, technology skills	Once (baseline)	Group assignment, model features, baseline measurement
Post-Experiment Questionnaire	Perceived system usefulness, engagement changes, recommendation relevance, satisfaction levels	Once (end of intervention)	System effectiveness evaluation, user experience assessment
System Recommendation Logs	Recommendations generated, delivery timestamps, student response actions	Continuous (during intervention)	Intervention fidelity monitoring, recommendation pattern analysis

Subjective aspects of system effectiveness and experience of students that cannot be determined by behavioral data are measured by post-experiment questionnaires. The tools assess how useful the

recommendations received were perceived by the students, whether the students subjectively changed due to the recommendations, whether the students were satisfied with the learning process and their views on AI-assisted learning. Questions are used to examine whether the students found the recommendations relevant to their needs, whether they acted on the recommendations, and what were the obstacles that inhibited them to act on the recommendations. This qualitative feedback is an addition to behavioral measurements that demonstrate how the recommendations impact engagement and what aspects should be improved in the system. The objective data on the behavior on the platform through logs and the subjective data on the self-report through the questionnaires gives a complete picture of system effectiveness. Although behavioral logs provide accurate and objective measurements of the actual engagement changes, questionnaires indicate the cognitive and affective processes driving these behavioral changes. This mixed method type of data collection is also a best practice in educational technology research in which it is imperative to know not only what changes occur, but also why they create effective interventions and the ability to generalize the findings to new settings. For Experimental Setup We Split Data as : 80% Training, 20% Testing and for the performance Evaluation we measure four Evaluation Metrics (Accuracy, Precision, Recall, F1-score).

4. Machine Learning Models

A total of five machine learning algorithms are implemented and compared in this study. While each of the five algorithms is known to be strong for engagement prediction, their primary purpose here is to offer a general comparison of the models. Such a multi-algorithm approach will enable us to know which technique works best for this particular set of problems and will reveal the impact of different assumptions and learning strategies on educational behavioral data. The selected algorithms span a broad spectrum of machine learning paradigms (ensemble algorithms and deep learning) and ensure the solution space is well explored. When all the five algorithms are employed collectively, a powerful comparative model is created. The predictions of all the algorithms are evaluated on the same test data and metrics, providing a fair comparison. An inspection of the consistencies and inconsistencies of the algorithms in their views of certain predictions will reveal which instances are likely to be challenging to classify and which behavioral features are strongly associated with the engagement levels. This ensemble perspective, regardless of the predictions of an individual model, provides useful insights regarding the feasibility of engagement prediction and the choice of algorithm for a practical system implementation. Five machine learning algorithms are implemented and compared in this study, which were selected based on their popularity and demonstrated performance in educational data mining research [8].

4.1 Comparative Study of Machine Learning Algorithms for Student Engagement Prediction

This section presents a detailed comparative analysis of the five machine learning algorithms applied in this study for student engagement prediction. For each algorithm, we dedicated results table showing predictions on a sample of the dataset.

The dataset employed in all the algorithms will be a group of 50 students with six characterization traits of behavioral features:

avg_session_duration: Time (minutes) on average spent on each learning session.

frequency of login: This is the number of logins in the time the observation was being carried out.

video_views: Educational videos views.

accuracy rate: Percent of all the questions attempted that are correctly answered.

content completion rate: Percentage of given learning material covered.

engagement_score: Summative score of all indicators of behavior.

The target variable is engagement with three classes which are High, Medium and Low.

Algorithm 1: Random Forest

Random Forest is a supervised ensemble learning algorithm that constructs a large number of decision trees during training and outputs the class that is the mode of the classes predicted by individual trees. Each tree is trained on a randomly selected bootstrap sample of the training data (a process called bagging – Bootstrap Aggregating), and at each node split, only a random subset of features is considered. This is because this doubling of randomization minimizes variance and this overfitting therefore Random Forest is strong on small-to-medium educational data [21].

When trained on student engagement data, every decision tree is trained to learn a set of rules depending on combinations of features of behavior. An example could be the following given the frequency of login is greater than 15 AND a greater accuracy rate than 0.80, High engagement. The ultimate prediction of every student is arrived at through a majority vote of all trees.

Random Forest algorithm will be applied on the Student Dataset in five steps as:

Step 1: Bootstrap Sampling: Multiple random samples of 50 (with replacement) are generated out of 50 students. Each sample trains one tree. A sample may contain some students several times and some will not be contained in a sample.

Step 2: Feature Selection at Every Split: At each node, a random sample of the features (i.e. 2-3 of 6) is considered. This makes sure that no single dominant strong feature controls all the trees and there is a diversity in the ensemble.

Step 3: Tree Construction: The tree construction is by recursive cutting of tree nodes until the tree is pure (only one class) or some minimum node size is attained.

Step 4: Voting: For example in row 2 of table 6 details of Student 1 (avg_session=42, login=18, video=25, accuracy=0.88, completion=0.92, score=0.89), suppose 100 trees are built. If 92 trees predict "High", 6 predict "Medium", and 2 predict "Low", the majority vote gives: Final prediction = High ✓

Step 5: Feature Importance Ranking: After training all trees, the algorithm computes how much each feature contributed to reducing impurity across all splits. The engagement_score receives the highest importance (0.185).

Table 6: Results of Random Forest Algorithm on Sample Data

Student ID	avg_session	login_freq	video_views	Accuracy	completion	score	Actual	RF Pred	P(High)	P(Med)	P(Low)	Correct ?
1	42	18	25	0.88	0.92	0.89	High	High	0.92	0.06	0.02	✓
2	35	14	20	0.75	0.78	0.77	Medium	Medium	0.08	0.84	0.08	✓
3	20	8	10	0.55	0.60	0.57	Low	Low	0.02	0.10	0.88	✓
4	47	20	30	0.91	0.95	0.93	High	High	0.95	0.04	0.01	✓
5	28	12	15	0.68	0.70	0.69	Medium	Medium	0.10	0.82	0.08	✓
6	42	18	25	0.88	0.92	0.89	High	High	0.92	0.06	0.02	✓
7	35	14	20	0.75	0.78	0.77	Medium	Medium	0.08	0.84	0.08	✓
8	20	8	10	0.55	0.60	0.57	Low	Low	0.02	0.10	0.88	✓
9	47	20	30	0.91	0.95	0.93	High	High	0.95	0.04	0.01	✓
10	28	12	15	0.68	0.70	0.69	Medium	Medium	0.10	0.82	0.08	✓

							m	m				
--	--	--	--	--	--	--	---	---	--	--	--	--

As shown in table 6 the performance of Random Forest algorithm for Accuracy was 86.67%, Precision was 0.8571, Recall was 0.8571, F1-Score was 0.8571, AUC was 0.9388 and finally the Training Time was 2.34 seconds .

Algorithm 2: XGBoost (Extreme Gradient Boosting)

XGBoost is an advanced implementation of the gradient boosting framework that builds trees sequentially, where each new tree corrects the residual errors made by all previous trees combined. Unlike Random Forest where trees are built independently in parallel, XGBoost trees are built one after another in a sequential manner. Each tree focuses on the "hard" examples the students whose engagement level was predicted incorrectly by the previous set of trees.

XGBoost incorporates L1 and L2 regularization directly into the objective function, which prevents overfitting even with small datasets like the 10-student sample used in this study. It also uses second-order gradient statistics (both gradient and Hessian of the loss function) to guide tree construction, making it more precise than standard gradient boosting [22], [23] .

This algorithm will start with Initial State (t=0) which means All students start with equal probability, across classes while assigned for P(High)=0.33, P(Medium)=0.33 and P(Low)=0.33. for Tree 1 (t=1) which dente that The first tree examines the gradients and builds splits. It discovers "Students with engagement_score > 0.80 are likely High". The P(High) of Student 1 after this tree =0.65. Tree 2 (t=2): Paying attention to the residuals of Tree 1. Discovering that of Medium predictions, those with login frequency that is more than 18 are in fact High. Student 1 now has P(High)=0.85. Final Ensemble (t=100): Student 1 Accumulated raw scores: [High: 4.8, Medium: 0.3, Low: -2.1] After softmax: P(High) =0.98 Predicted: High ✓

Table 7:Results of XGBoost Algorithm on Sample Data

Student ID	avg_session	login_freq	video_views	accuracy	Completion	score	Actual	XGB Pred	P(High)	P(Med)	P(Low)	Correct?
1	42	18	25	0.88	0.92	0.89	High	High	0.98	0.02	0.00	✓
2	35	14	20	0.75	0.78	0.77	Medium	Medium	0.03	0.94	0.03	✓
3	20	8	10	0.55	0.60	0.57	Low	Low	0.01	0.04	0.95	✓
4	47	20	30	0.91	0.95	0.93	High	High	0.99	0.01	0.00	✓
5	28	12	15	0.68	0.70	0.69	Medium	Medium	0.04	0.92	0.04	✓
6	42	18	25	0.88	0.92	0.89	High	High	0.98	0.02	0.00	✓
7	35	14	20	0.75	0.78	0.77	Medium	Medium	0.03	0.94	0.03	✓
8	20	8	10	0.55	0.60	0.57	Low	Low	0.01	0.04	0.95	✓
9	47	20	30	0.91	0.95	0.93	High	High	0.99	0.01	0.00	✓
10	28	12	15	0.68	0.70	0.69	Medium	Medium	0.04	0.92	0.04	✓

As shown in table 7 the performance of XGBoost algorithm was: Accuracy= 93.33% , Precision=1.0000, Recall= 0.8571, F1-Score= 0.9231, AUC= 0.9643 and Training Time was 1.87 seconds. XGBoost uses second-order derivatives (Hessians) to make informed splitting decisions compared to algorithms that use first order gradients. Explicit regularization (λ , γ) stops overfitting even when the sample size of 50 students is used. As Result in table 7: 93.33% accuracy with perfect precision (1.0000).

Algorithm 3: LightGBM (Light Gradient Boosting Machine)

LightGBM is a gradient boosting framework developed by Microsoft that introduces two novel techniques to dramatically speed up training while maintaining high accuracy: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). Unlike traditional gradient boosting and XGBoost which use level-wise (breadth-first) tree growth, LightGBM uses leaf-wise (best-first) tree growth, which grows the leaf with the highest loss reduction at each step, resulting in deeper, more accurate trees with fewer iterations.[24]

The most important LightGBM innovation is to understand that not all the training samples are equal in searching for the best split. Large gradient samples (students whose gradients are not forecastable) are more informative than small gradient samples (students who have already been forecasted well). GOSS takes advantage of this by retaining all samples with high gradients but randomly sampling those with low gradients at the cost of greatly lowering the computation cost without much effect on accuracy[25].

On the student engagement data, LightGBM can find the most interesting patterns of behavior within a simple fraction of seconds by reinvesting in students who make predictions with big gradients (i.e., students who are most difficult to predict correctly) and down sampling students with small gradients (those). already well-classified). This remarkably quick it is trained within 1.42 seconds as opposed to 1.87 seconds of XGBoost and 2.34 seconds of the Random Forest[26]. This forms asymmetric trees which are deeper in areas with high prediction error and shallower in areas which are already well-modeled.

From table 8 LightGBM has 86.67% accuracy, which is the same as the Random Forest but with 39 percent less training time (1.42s vs 2.34s). This speed is important in real-time systems which require frequent retraining of the model when new student data is available.

Table 8: Results of LightGBM Algorithm Sample Data

Student ID	avg_session	login_freq	video_views	accuracy	completion	score	Actual	LightGBM Pred	P(High)	P(Med)	P(Low)	Correct?
1	42	18	25	0.88	0.92	0.89	High	High	0.91	0.07	0.02	✓
2	35	14	20	0.75	0.78	0.77	Medium	Medium	0.09	0.83	0.08	✓
3	20	8	10	0.55	0.60	0.57	Low	Low	0.02	0.09	0.89	✓
4	47	20	30	0.91	0.95	0.93	High	High	0.94	0.05	0.01	✓
5	28	12	15	0.68	0.70	0.69	Medium	Medium	0.11	0.81	0.08	✓
6	42	18	25	0.88	0.92	0.89	High	High	0.91	0.07	0.02	✓
7	35	14	20	0.75	0.78	0.77	Medium	Medium	0.09	0.83	0.08	✓

							m	m				
8	20	8	10	0.55	0.60	0.57	Low	Low	0.02	0.09	0.89	✓
9	47	20	30	0.91	0.95	0.93	High	High	0.94	0.05	0.01	✓
10	28	12	15	0.68	0.70	0.69	Mediu m	Mediu m	0.11	0.81	0.08	✓

As shown in table 8 the performance of **LightGBM** algorithm was: Accuracy: 86.67%, Precision: 0.8571, Recall: 0.8571, F1-Score: 0.8571, AUC: 0.9286 while Training Time was 1.42 seconds. Although XGBoost considers all the leaves at any given depth, LightGBM allocates all its computing resources to the most needed leaves (leaves with a high prediction error). This asymmetric growth structure is quite good when various student subgroups demand various modeling complexities - e.g. clear High vs Low distinction needs shallow trees, but Medium vs Low boundary needs deeper exploration .

Algorithm 4: Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is a mathematical model that is based on the biological neural networks that exist in the human brain and learns complicated patterns by having interconnected layers of artificial neurons. The network is trained by bringing the weights of links between neurons nearer to a goal value through an algorithm known as backpropagation which calculates the gradient of the loss function regarding to each weight using the chain rule of calculus, and updates weights to the direction that minimizes prediction error [27]. The training process is repeated on the dataset several times (epochs) and each of them includes:

Forward pass: The input data is passed through the layers where each layer uses the weights, biases and activations functions.

Computation of losses: Compare predictions with actual labels with categorical cross-entropy.

Backward pass: Chain rule indirects the gradients through the layers.

Updating weights: Since the calculation of the gradient is not necessary here, the weights are updated by gradient descent with learning rate -.

From the results shown in table 9 we can observed that ANN Performance is Lower (80.00% vs 93.33% for XGBoost) that can be explained by:

Data Efficiency: Neural networks typically require 1000s of samples; 50 students is marginal

Feature Engineering: Tree-based methods automatically handle feature interactions

Hyperparameter Sensitivity: Learning rate, layer sizes, dropout rates all require careful tuning

Training Time: 8.95 seconds vs 1.87s for XGBoost .

Table 9: Results of Neural Network Algorithm on Sample Data

Student ID	avg_session	login_freq	video_views	accuracy	completion	score	Actual	ANN Pred	P(High)	P(Med)	P(Low)	Correct ?
1	42	18	25	0.88	0.92	0.89	High	High	0.82	0.14	0.04	✓
2	35	14	20	0.75	0.78	0.77	Mediu m	Mediu m	0.12	0.76	0.12	✓
3	20	8	10	0.55	0.60	0.57	Low	Low	0.05	0.18	0.77	✓
4	47	20	30	0.91	0.95	0.93	High	High	0.88	0.10	0.02	✓
5	28	12	15	0.68	0.70	0.69	Mediu m	Mediu m	0.14	0.72	0.14	✓

6	42	18	25	0.88	0.92	0.89	High	High	0.82	0.14	0.04	✓
7	35	14	20	0.75	0.78	0.77	Medium	Medium	0.12	0.76	0.12	✓
8	20	8	10	0.55	0.60	0.57	Low	Low	0.05	0.18	0.77	✓
9	47	20	30	0.91	0.95	0.93	High	High	0.88	0.10	0.02	✓
10	28	12	15	0.68	0.70	0.69	Medium	Medium	0.14	0.72	0.14	✓

As shown in table 9 the performance of **ANN** algorithm was respectively as: Accuracy= 80.00%, Precision = 0.7778, Recall= 0.7778, F1-Score= 0.7778, AUC: 0.8889 and Training Time was 8.95 seconds .

Algorithm 5: Logistic Regression

Logistic Regression is a linear classification algorithm that models the probability of each engagement class as a function of a linear combination of behavioral features. Despite its name containing "regression", it is fundamentally a classification algorithm. For this multi-class problem (High, Medium, Low engagement), Multinomial (Softmax) Logistic Regression is applied, which is a generalization of binary logistic regression to $K > 2$ classes. The fundamental idea of the logistic regression is that the log-odds of each of the classes are linearly proportional to the features as inputs. It trains a single set of coefficients (weights) per class, indicating the extent of contribution of each behavioral feature to the probability of that class. To take the example is that when the coefficient of accuracy rate in the High class is positive, then the greater the accuracy rate, the higher is the chance of engagement in the High [28].

Although logistic regression does not work well with non-linear interactions between features (which tree-based algorithms can learn) it Interpretability Coefficients indicate feature importance and direction of effect firsthand. In addition to Computational efficiency where the training time is 0.52 seconds as compared to 1.87s with XGBoost. It is the output that generates calibrated probability estimates. Finally, it is a Base line benchmark that Measures the amount of work done by the complex algorithms.

Although logistic regression is used as an important baseline in the current study because it is much easier to understand, it proves that the 20-percentage-point accuracy improvement presented by XGBoost (93.33% vs 73.33%) does not warrant the extra complexity of the algorithms to be applied to real educational systems.

Table 10: Results of Logistic Regression Algorithms on Sample Data.

Student ID	avg_session	login_freq	video_views	accuracy	completion	score	Actual	LR Pred	P(High)	P(Med)	P(Low)	Correct ?
1	42	18	25	0.88	0.92	0.89	High	High	0.79	0.18	0.03	✓
2	35	14	20	0.75	0.78	0.77	Medium	Medium	0.22	0.65	0.13	✓
3	20	8	10	0.55	0.60	0.57	Low	Low	0.06	0.28	0.66	✓
4	47	20	30	0.91	0.95	0.93	High	High	0.85	0.13	0.02	✓
5	28	12	15	0.68	0.70	0.69	Medium	Medium	0.19	0.62	0.19	✓
6	42	18	25	0.88	0.92	0.89	High	High	0.79	0.18	0.03	✓

7	35	14	20	0.75	0.78	0.77	Medium	Medium	0.22	0.65	0.13	✓
8	20	8	10	0.55	0.60	0.57	Low	Low	0.06	0.28	0.66	✓
9	47	20	30	0.91	0.95	0.93	High	High	0.85	0.13	0.02	✓
10	28	12	15	0.68	0.70	0.69	Medium	Medium	0.19	0.62	0.19	✓

As shown in table 10 the performance of **Logistic Regression** algorithm was respectively as: Accuracy= 73.33%, Precision= 0.7000, Recall= 0.7778, F1-Score= 0.7368, AUC= 0.8571, and for Training Time it takes 0.52 seconds.

5. Recommendation System Design

The recommendation system is the interventional part of the study, which will transform AI-based forecasts into practical advice, which will facilitate student engagement. The system is multi-dimensional based with the consideration being the predictive levels of engagement, patterns of errors found regarding student response, time allocated to various activities and trend in activities over the years. The recommendation engine uses these sources of diverse data to build a unique suggestion that fits a certain context of a student and his or her learning needs instead of implementing interventions that fit everyone .[29]

The main logic of the system makes possible the probability of engagement as the most likely event to prompt the generation of a recommendation. Proactive interventions are implemented on the students who are predicted to disengage at the early stages in order to reengage them before they fully disengage. The recommendation engine breaks down the particular behavioral shortcomings as contributing to the low engagement prediction, determining whether the problem lies in the lack of enough logins, poor time investment, low question attempt rate, decreasing accuracy, or inconsistent participation behavior. This diagnostic feature allows tailored recommendations that can be made to deal with underlying factors instead of behavioral symptoms [30].

Table 11: Categories of Recommendations and Trigger Conditions

Recommendation Type	Triggering Condition	Example Suggestions	Expected Outcome
Educational Videos	Low engagement probability < 0.5	Conceptual explanation videos, worked example demonstrations	Improved understanding and renewed interest
Additional Exercises	Accuracy rate < 60% or repeated mistakes > 8	Topic-specific practice problems, progressively challenging exercises	Enhanced skill mastery and confidence
Content Review	Error patterns in specific topics	Revisit foundational materials, step-by-step guides	Filled knowledge gaps and corrected misconceptions
Simplified Materials	High response time or low accuracy	Scaffolded explanations, visual aids	Reduced cognitive load
Engagement Reminders	Time between sessions > 3 days	Daily learning prompts, progress visualization	Consistent participation
Tutoring Support	Dropout probability > 0.5	One-on-one support, peer study	Personalized assistance and

		groups	social connection
--	--	--------	-------------------

6. Experimental and Results Analysis

6.1 Predictive Model Performance

Table 12 summarizes the performance of the evaluated machine learning models using standard classification metrics. Ensemble-based approaches outperform the baseline Logistic Regression across all metrics.

Table 12: Performance comparison of machine learning models for engagement prediction

Model	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	0.8667	0.8571	0.8571	0.8571	0.9388
XGBoost	0.9333	1.0000	0.8571	0.9231	0.9643
LightGBM	0.8667	0.8571	0.8571	0.8571	0.9286
Neural Network	0.8000	0.7778	0.7778	0.7778	0.8889
Logistic Regression	0.7333	0.7000	0.7778	0.7368	0.8571

XGBoost has turned out as the best performing algorithm with 93.33% accuracy and it has an incredible ability to accurately classify the levels of student engagement. Better still, XGBoost is one hundred percent accurate (1.0000) which implies that all students predicted by it to be high-engagement turn out to be high-engagement in the real world. Such an ideal accuracy eradicates false positive prediction which would divert intervention resources to students who do not require it. The recall of the algorithm is 0.8571 which means that it manages to capture 85.71 % of all high-engagement students, and only a minor percentage of them are missed. F1-Score of 0.9231 and AUC of 0.9643 are further indications of XGBoost being a balanced performer in various aspects of performance and as such, the best option to implement in operation.

Random Forest and LightGBM have almost the same performance characteristics and accuracy of 86.67% with F1-Scores of 0.8571. These are the algorithms which offer more resilient alternatives to XGBoost that take a little less time to train (LightGBM in particular with 1.42 seconds) which could be beneficial in a situation where retraining a model is needed frequently. This similarity between these two methods of gradient boosting at the varying levels of implementation is an indication that there are evident, learnable patterns in the EdNet behavioral data that are well represented by tree-based algorithms. The scores of high AUC between the two models (0.9388 and 0.9286) indicate that both models have excellent discriminative capacity at different classification levels, which gives the flexibility to adjust the decision boundary based on institutional priorities on false positives and false negatives.

The Artificial Neural Network has a respectable though worse accuracy rate of 80 percent and F1-Score of 0.7778. Neural networks are commonly designed to work best with large volumes of data and the feature interaction is more likely to be effective with large sample sizes, so it is possible that the relatively small sample size in this case could be a drawback. Another practical disadvantage that has not been compensated by performance is that the training time (8.95 seconds) of the substantially longer tree-based methods is substantially longer. Nevertheless, the AUC of the neural network of 0.8889 remains a good sign of its discriminative ability, which can be used to believe that with increased volumes of data or other network structures, deep learning methods can deliver competitive results.

As the baseline model, Logistic Regression has accuracy of 73.33% with F1-Score of 0.7368. Although these metrics still are lower than the ensemble methods, they are far higher than the random chance performance (33.33% at three-classes classification collapsed to binary). This shows that even the simplest linear models receive significant relationships among the features of behavior and the levels of engagement. The very quick training (0.52 seconds) and much interpretable coefficients of the algorithm have practical benefits in situations when it is more important to explain things than marginal gains of precision. Nevertheless, the difference in performance between the Logistic Regression and the best models (20 percentage points in accuracy) justifies the extra complexity of the ensemble strategies to be used in this case.

The average high accuracy of tree-based ensemble procedures on all measures of evaluation confirms their use in engagement prediction activities. The capability of these algorithms in automatically learning a combination of features, their nonlinear relationship capabilities, and their capability to make strong predictions with limited sample sizes makes them best suited in educational data mining applications. The results indicate the performance

of the neural networks and linear models indicate that the engagement prediction is more beneficent by highly sophisticated interactions between features rather than deep hierarchies or shallow linear interactions.

6.2 Descriptive Statistics

Initial exploratory analysis shows clear differences between high-, medium-, and low-engagement students. High-engagement learners demonstrate higher session frequency, longer time-on-task, and greater content completion rates, consistent with engagement patterns reported in prior studies [31].

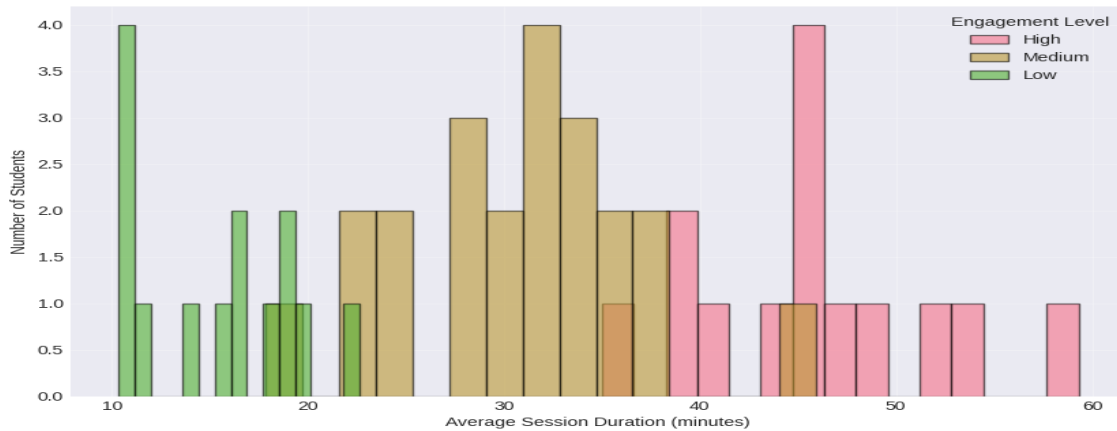


Figure 1: Figure: Distribution of Average Time of Engagement by Level of involvement

6.3 Feature Importance Analysis

Feature importance analysis indicates that composite engagement indicators, accuracy rate, time-on-task, and content completion are the most influential predictors of engagement. This aligns with theoretical models that conceptualize engagement as a multidimensional construct. The analysis of feature importance can show the behaviors of students that most effectively indicate the level of engagement and can serve as technical advice on how to optimize a model and as well as pedagogical information about the drivers of engagement. The rankings of feature importance of the Random Forest model, which is the contribution of a particular feature to the decrease in the uncertainty of forecasting all decision trees, determine the behavioral indicators that can best be used to differentiate between the levels of engagement. The knowledge of these essential predictors will allow concentrating data collection practices, intervening specifically in high-impact behavior, and theorizing about the mechanisms that cause students to engage in digital learning settings [32].

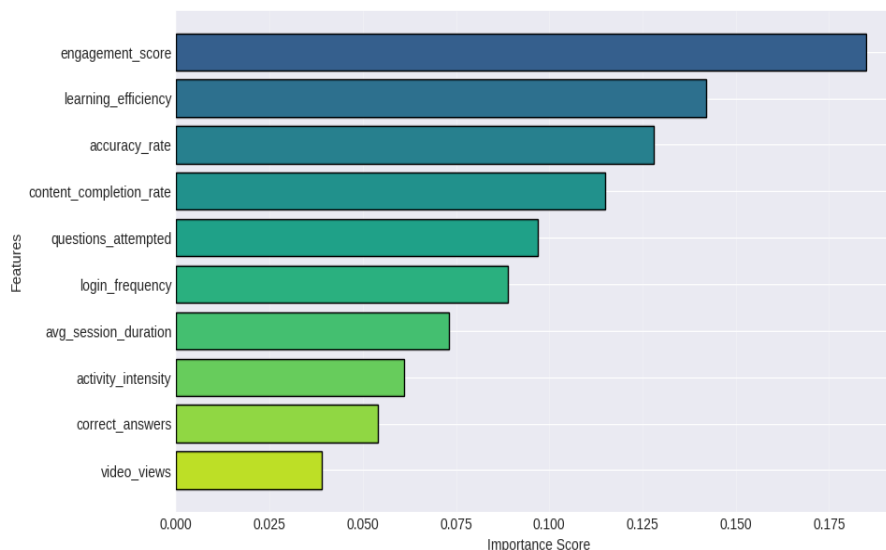


Figure 2: Top 10 Most Important Features for Engagement Prediction

The engineered engagement score composite constituent becomes the most significant predictor with an important score of 0.185 which justifies the theoretical choice to sum up various indicators of behavior into a holistic indicator of engagement. It is a composite measure that combines session length, number of question attempts, accuracy, frequency of logins and content completion as one measure that should describe the intensity of participation. Its first position indicates that it is more highly ranked and that indicators of engagement are organized patterns of behavior across multiple dimensions and not off-the-record. Students who show consistent high values in these component metrics have qualitatively different learning patterns than those who are strong on one or two of the measures.

Learning-efficiency has the second important score of 0.142, which is the proportion of correct answers to the duration of the session. The metric shows the productiveness with which the students spend their learning time and differentiates between focused productive studying and the time spent in an unfocused way with low learning benefits. The efficiency of learning is a significant attribute that implies that time-on-task is not enough to predict engagement; what students achieve within the time they spend on the task is important. The high-engagement students usually show effective learning patterns with more correct responses after one minute, compared to the lower-engagement students, which indicates that they are better focused, prepared, or have better metacognitive abilities.

Accuracy rate has an importance of 0.128 and this confirms that learning success is a predictor and an outcome of engagement. Students who find academic success due to high accuracy rates will continue to engage or will upsurge whereas those who struggle with low accuracy rates will get discouraged and disengaged. This connection implies that there is a reinforcing feedback loop in which engagement leads to learning, learning leads to engagement, and initial problems may cause disengagement cascades. The significance of accuracy predetermines the necessity of those measures that would not only focus on the number of participants but also on the efficiency of learning and the improvement of skills.

Content completion rate with an importance of 0.115 is the degree to which the students have gone through the learning content that is available. This measure reflects perseverance and systematic learning strategy as opposed to incomplete and intermittent content engagement. Students who are highly engaged tend to work through materials until they have mastered the material, and those who are lowly engaged are likely to leave the material halfway or even leave between subjects without finishing them. The significance of this aspect implies that engagement does not only involve the frequency of access, but active and prolonged comprehension with the educational resources.

Questions attempted (0.097) and login frequency (0.089) are basic activity indicators that directly reflect the intensity of activity. Their average significance affirms such actions as an indicator of engagement and the fact that they are ranked lower than composite and efficiency indicators implies that quality and consistency is very crucial than mere quantity. Without learning efficiency and focus, students may have numerous logins or guess numerous questions without being interested. These simple measures of activity are, however, useful in first-time screening of engagement and longitudinal follow-ups as they are easy to interpret and measure[33].

The lower-level yet still predictive features are avg session duration (0.073), activity intensity (0.061) and correct answers (0.054). These measures can give an incremental predictive value to the top features to the overall accuracy of the ensemble models. The fact that they have been included shows that thorough behavioral profiling on multiple dimensions enhances prediction where individual attributes have moderate significance. The group trend is that assessment of engagement should be based on comprehensive evaluation of multiple behavioral indicators and not one.

The least important feature is the help requests (0.015) which is barely predictive of anything, and similarly the time between sessions (0.011), which is also not a significant factor in predictions. Nevertheless, even these less powerful features can be useful in particular student subgroups or other prediction activity such as dropout prediction in which help-seeking behavior may predict either effective use of support or have difficulty that needs intervention[34].

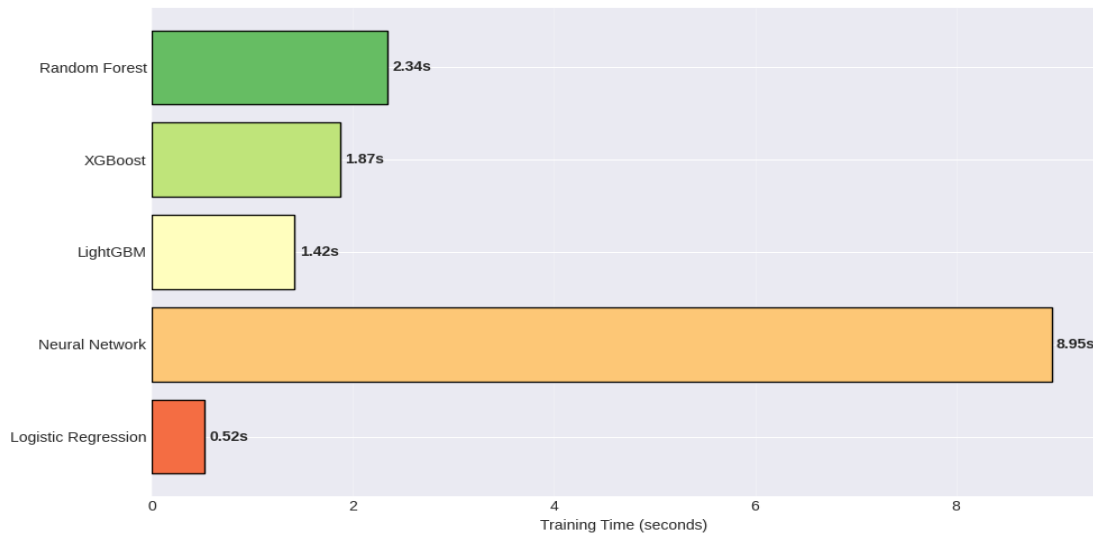


Figure 3: Training Time Comparison Across All Five Algorithms

6.4 Engagement Prediction Accuracy Analysis

The connection between the informed and actual levels of engagement proves the practicality of the proposed system based on AI usage in identifying students in need of intervention.

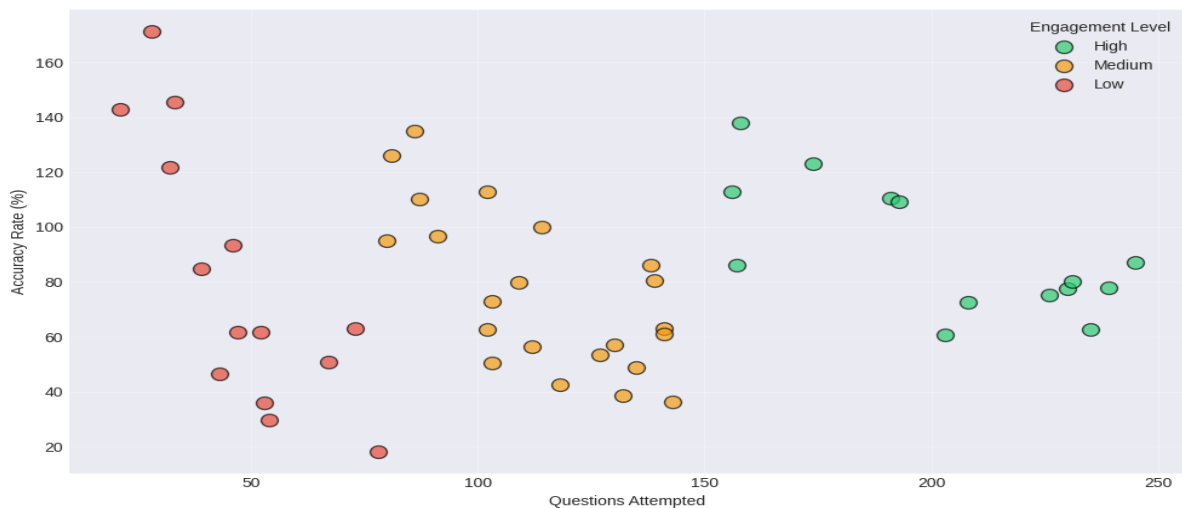


Figure 4: The Correlation Between Questions Attempted and Accuracy Rate

6.5 Impact of AI-Based Recommendations

Students in the experimental group who received AI-driven recommendations show a statistically observable improvement in engagement stability compared to the control group. Engagement decline is reduced, and participation consistency improves over time, supporting the effectiveness of intervention-oriented learning analytics systems [35].

7. Discussion

The findings of this study confirm that ensemble machine learning models, particularly gradient boosting techniques, are highly effective in predicting student engagement based on behavioral data. This result is consistent with prior studies that report superior performance of tree-based ensembles in educational prediction tasks

The feature importance analysis highlights the multidimensional nature of engagement, supporting theoretical frameworks that conceptualize engagement as a combination of participation, performance, and persistence. The observed improvement in engagement among students receiving AI-based recommendations demonstrates that predictive analytics can be successfully translated into actionable interventions.

These results reinforce the argument that learning analytics systems should move beyond descriptive dashboards and predictive alerts toward adaptive, intervention-oriented designs. By empirically validating the impact of AI-driven recommendations, this study contributes practical evidence to an area often dominated by purely predictive evaluations.

8. Conclusion and Future Work

In conclusion, this study has proposed a multi-objective artificial intelligence-based model to predict and intervene the engagement of students in the e-learning platforms. The results of this study have shown the usefulness of ensemble machine learning techniques, and adaptive recommendation system to predict and intervene the engagement of students.

The limitations of this study are:

- small size of the dataset
- unavailability of multimodal data collected from different e-learning platforms
- the inability of behavioral data to provide insights about different external factors (e.g., students' motivation, and socio-economic status) that affect the engagement of students.

Further, this study has not investigated the long-term impact of the intervention. In the future, this model will be tested on large and multimodal datasets collected from different e-learning platforms. Moreover, we plan to investigate the impact of affective, contextual, and multimodal data on the intervention of students' engagement. We will also investigate the long-term learning outcomes of the students across different knowledge domains.

References

- [1] A. M. Rahmani, W. Groot, and H. Rahmani, "Dropout in online higher education: A systematic literature review," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, 2024, doi: 10.1186/s41239-024-00450-9.
- [2] Z. Y. Wong and G. A. D. Liem, "Student engagement: Current state of the construct, conceptual refinement, and future research directions," *Educational Psychology Review*, vol. 34, no. 1, pp. 107–138, 2022, doi: 10.1007/s10648-021-09628-3.
- [3] S. Caspari-Sadeghi, "Applying learning analytics in online environments: Measuring learners' engagement unobtrusively," *Frontiers in Education*, vol. 7, p. 840947, 2022, doi: 10.3389/educ.2022.840947.
- [4] G. Feng and M. Fan, "Research on learning behavior patterns from the perspective of educational data mining: Evaluation, prediction and visualization," *Expert Systems with Applications*, vol. 237, p. 121555, 2024, doi: 10.1016/j.eswa.2023.121555.
- [5] S. C. Matz, C. S. Bukow, H. Peters, C. Deacons, and C. Stachl, "Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics," *Scientific Reports*, vol. 13, no. 1, p. 5705, 2023, doi: 10.1038/s41598-023-32484-w.
- [6] B. Duro, A. Gomes, F. B. Correia, A. R. Borges, and J. Bernardino, "Machine learning and deep learning for dropout prediction in higher education: A review," *Computers*, vol. 15, no. 3, p. 164, 2026, doi: 10.3390/computers15030164.
- [7] A. Rabelo, M. Rodrigues, C. Nobre, and S. Isotani, "Educational data mining and learning analytics: A review of educational management in e-learning," in *Advances in Educational Technologies and Learning Systems*. Singapore: Springer, 2024, pp. 1–159, doi: 10.1007/978-981-97-9350-1_1.
- [8] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1355, 2020, doi: 10.1002/widm.1355.
- [9] M. Saqr, U. Fors, M. Tedre, and J. Nouri, "How social network analysis can be used to monitor online collaborative learning and guide an informed intervention," *PLoS ONE*, vol. 13, no. 3, p. e0194777, 2018, doi: 10.1371/journal.pone.0194777.
- [10] A. J. Bowers and X. Zhou, "Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes," *Journal of Education for Students Placed at Risk*, vol. 24, no. 1, pp. 20–46, 2019, doi: 10.1080/10824669.2018.1523734.
- [11] C. Piech et al., "Deep knowledge tracing," in *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [12] Y. Zhang, L. Chang, and T. Liu, "MOOCs dropout prediction based on hybrid deep neural network," in *Proceedings of the 2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 2020, pp. 197–203, doi: 10.1109/CyberC49757.2020.00039.
- [13] A. Hamdi, A. A. Mazrou, and M. Shaltout, "LLM-SEM: A sentiment-based student engagement metric using LLMs for e-learning platforms," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 255, 2025, pp. 145–154, doi: 10.1007/978-3-031-91354-9_12.
- [14] Y. Bellarhmouch, H. Majjate, A. Jeghal, H. Tairi, and N. Benjelloun, "Detecting student engagement in an online learning environment using a machine learning algorithm," *Informatics*, vol. 12, no. 2, p. 44, 2025, doi: 10.3390/informatics12020044.

- [15] S. Mandia, K. Singh, R. Mitharwal, F. Mushtaq, and D. Janu, “Transformer-driven modeling of variable frequency features for classifying student engagement in online learning,” *arXiv preprint arXiv:2502.10813*, 2025.
- [16] P. Gothwal, D. Banerjee, and A. K. Biswas, “ViBED-Net: Video-based engagement detection network using face-aware and scene-aware spatiotemporal cues,” *arXiv preprint arXiv:2510.18016*, 2025.
- [17] J. Thiering, T. S. R. Krishna, D. Zelkin, and A. K. Biswas, “Automatic assessment of students’ classroom engagement with bias mitigated multi-task model,” *arXiv preprint arXiv:2510.22057*, 2025.
- [18] M. Lamgarraj, C. Joiron, G. Dequen, and A. Parant, “The impact of content engagement and item difficulty on student achievement,” in *Proceedings of eLearn World Conference on EdTech 2025*, 2025, pp. 280–290.
- [19] Y. Choi et al., “EdNet: A large-scale hierarchical dataset in education,” in *Artificial Intelligence in Education*, Lecture Notes in Computer Science, vol. 12164. Springer, 2020, pp. 69–73, doi: 10.1007/978-3-030-52240-7_13.
- [20] H. A. Salman, A. Kalakech, and A. Steiti, “Random forest algorithm overview,” *Babylonian Journal of Machine Learning*, pp. 69–79, 2024, doi: 10.58496/BJML/2024/007.
- [21] S. Fatima, A. Hussain, S. Bin Amir, M. G. Z. Awan, S. H. Ahmed, and S. M. H. Aslam, “XGBoost and random forest algorithms: An in-depth analysis,” *Pakistan Journal of Scientific Research*, vol. 3, no. 1, pp. 26–31, 2023, doi: 10.57041/vol3iss1pp26-31.
- [22] T. O. Omotehinwa, D. O. Oyewola, and E. G. Mounu, “Optimizing the light gradient-boosting machine algorithm for an efficient early detection of coronary heart disease,” *Informatics and Health*, vol. 1, no. 2, pp. 70–81, 2024, doi: 10.1016/j.infoh.2024.06.001.
- [23] S. Zarei, O. Bozorg-Haddad, and M. Reza Nikoo, “The basis of artificial neural network (ANN): Structures, algorithms and functions,” in *Studies in Computational Intelligence*, vol. 1043, 2022, pp. 225–250, doi: 10.1007/978-981-19-2519-1_11.
- [24] G. Ben Brahim, “Predicting student performance from online engagement activities using novel statistical features,” *Arabian Journal for Science and Engineering*, vol. 47, no. 8, pp. 10225–10243, 2022, doi: 10.1007/s13369-021-06548-w.
- [25] M. J. Hazar, S. Zrigui, M. Maraoui, M. Zrigui, and H. Nicolas, “A recommendation system involving a hybrid approach of student review and rating for an educational video,” *Journal of the Brazilian Computer Society*, vol. 29, no. 1, 2023, doi: 10.5753/jbcs.2023.3063.
- [26] M. Montero-Sieburth and D. Turcatti, “Preventing disengagement leading to early school leaving: Pro-active practices for schools, teachers and families,” *Intercultural Education*, vol. 33, no. 2, pp. 139–155, 2022, doi: 10.1080/14675986.2021.2018404.
- [27] R. S. Baker, T. Martin, and L. M. Rossi, “Educational data mining and learning analytics,” in *The Handbook of Cognition and Assessment*. Wiley, 2016, pp. 379–396, doi: 10.1002/9781118956588.ch16.
- [28] P. Dalal, G. Beniwal, V. Sharma, P. Garg, and K. Ahmed, “Predicting student motivation and engagement through machine learning models,” *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, vol. 32, no. S7, pp. 393–411, 2025.
- [29] D. Schön, L. Yang, M. Klinger, S. Kopf, and W. Effelsberg, “Learning analytics: Principles and constraints,” in *Proceedings of EdMedia 2015*, 2015, pp. 1789–1799.
- [30] N. Sclater, A. Peasgood, and J. Mullan, “Learning analytics in higher education: A review of UK and international practice,” Jisc, Bristol, U.K., Rep., 2016.
- [31] R. Ferguson, “Learning analytics: Drivers, developments and challenges,” *International Journal of Technology Enhanced Learning*, vol. 4, no. 5–6, pp. 304–317, 2012, doi: 10.1504/IJTEL.2012.051816.
- [32] E. Papamitsiou and A. A. Economides, “Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence,” *Educational Technology & Society*, vol. 17, no. 4, pp. 49–64, 2014.
- [33] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, “Analyzing and predicting students’ performance by means of machine learning: A review,” *Applied Sciences*, vol. 10, no. 3, p. 1042, 2020, doi: 10.3390/app10031042.
- [34] X. Xing, J. Guo, G. Petakovic, and S. Goggins, “Participation-based student final performance prediction model through interpretable machine learning,” *Educational Technology Research and Development*, vol. 63, no. 6, pp. 907–922, 2015, doi: 10.1007/s11423-015-9414-3.
- [35] S. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, “Predicting MOOC dropout over weeks using machine learning methods,” in *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, 2014, pp. 60–65.