

## PREDICTIVE ANALYTICS FOR STUDENT PERFORMANCE IN DIGITAL MEDIA COURSES

Dr. Swati Desai <sup>1</sup>, Dr. Satyawan Changadev Hembade <sup>2</sup>, Dr. Shweta Joglekar <sup>3</sup>, Dr. Ramchandra Vasant Mahadik <sup>4</sup>,  
Deepti Deshmukh <sup>5</sup>, Dr Anita Desai <sup>6</sup>

<sup>1</sup> Assistant Professor, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development, Pune-411038, India

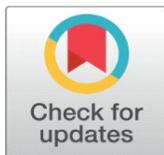
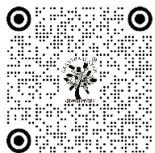
<sup>2</sup> Associate professor, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development, Pune-411038, India

<sup>3</sup> Assistant professor, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development, Pune-411038, India

<sup>4</sup> Associate professor, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development, Pune-411038, India

<sup>5</sup> Assistant Professor, Bharati Vidyapeeth (Deemed to be) University, IMED, Pune, India

<sup>6</sup> Sr. Assistant Professor, School of Computer Studies, Sri Balaji University, Tathawade, Pune, India



**Received** 07 September 2025

**Accepted** 02 December 2025

**Published** 17 February 2026

### Corresponding Author

Dr. Swati Desai,  
[swati.desai@bharatividyaapeeth.edu](mailto:swati.desai@bharatividyaapeeth.edu)

### DOI

[10.29121/shodhkosh.v7.i1s.2026.6804](https://doi.org/10.29121/shodhkosh.v7.i1s.2026.6804)

**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Copyright:** © 2026 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.

## ABSTRACT

The fast adoption of the digital media technologies has altered the practices of learning, creative production, and assessment in higher education. Traditional evaluation methods are mostly retrospective and do not offer much assistance in the early detection of academic vulnerability. This paper seeks to solve this problem of poor predictive visibility by suggesting a predictive analytics model of student outcomes in digital media learning. The methodology uses feature engineering and normalization, time-series trend and supervised machine learning classification, feature engineering and normalization, model explainability. Models to be used are Random Forest, Gradient Boosting, and Logistic Regression to analyze multisource data, such as assignment submissions, logs on digital tool use, attendance, assessment scores, peer collaboration measures, and creative project assessment, and SHAP-based interpretability to analyze feature impacts. According to experimental findings, the given framework is able to reach an average prediction accuracy of 92.56 and the values of precision and recall of 91.48% and 92.02, respectively. Early-warning predictions of mid-semester are 86.94% accurate, which can be used to implement academic interventions in time. The analysis of contribution to a feature indicates that the contribution of project iteration frequency, degree of engagement, and punctual submission is more than 64% predictive. The results prove that predictive analytics would be a useful tool to facilitate early intervention, personalized comments, and data-driven instruction methods. The results of the research are that the AI-based performance prediction improves the performance of learners and the planning of instruction.

**Keywords:** Predictive Analytics, Student Performance Prediction, Digital Media Education, Learning Analytics, Machine Learning, Educational Data Mining



## 1. INTRODUCTION

The high rate of technological advancement in the digital media has redefined modern day education, especially in areas like graphic design, animation, video creation, interactive media and digital storytelling. Such courses focus on creativity, repetitive practice and a process of work mediated by technology, so the process of student learning is highly dynamic and non-linear [Syed Mustapha \(2023\)](#). As learning spaces become more and more dependent on learning management systems, digital ateliers, and online collaborative models, high amounts of educational information are constantly produced. Predictive analytics has become the potent tool to convert this data into actionable information through the detection of patterns, forecasting, and evidence-based academic decision-making [Sunawar et al. \(2025\)](#). Predictive analytics is of a particular importance in digital media education since it can allow instructors and institutions to abandon the intuition-based evaluation methods in favor of well-planned, data-driven approaches, which have the potential to increase both student achievement and the efficiency of an instructional process [Poh and Khor \(2024\)](#). In spite of such a possibility, there are unique issues concerning student performance in digital media courses. Creative and practice-based courses have a wide range of influencing factors on performance, as opposed to traditional theory-based subjects where technical skill enhancement, creative experimentation, level of engagement, time, and working with peers are the main factors [Lee et al. \(2025\)](#). Students tend to develop at a slow pace, have a different creativity style and show irregular performance at the project milestones. This variability in nature complicates the ability of the instructors to detect failing students early through the normal form of grading [Maraza-Quispe et al. \(2022\)](#). In addition, the creative work is not necessarily directly reflected in numeric scores, which makes it even more difficult to assess performance. Consequently, most students get feedback once they have taken major tests, so that it is not possible to intervene and help them at an early stage through personalized intervention [Bravo-Agapito et al. \(2020\)](#).

The main assessment methods of the traditional digital media education are largely retrospective as they are based on end-term grades, cumulative project reviews, or subjective evaluation by an instructor [Barlaskar et al. \(2025\)](#). Although these approaches offer summative information, they give little predictive transparency of future performance and do not account variations in the course of learning behaviors. These drawbacks deny teachers the autonomy to actively respond to learning shortcomings, academic vulnerability, and instructional adjustments in response to immediate demands [Fahd and Miah \(2025\)](#). To address these issues, the research is informed by the fact that there is a need of an intelligent, interpretable, and early-stage predictive framework, specific to digital media education. The suggested framework will use supervised machine learning models to process academic, behavioral, and engagement-related data and combine explainable AI techniques to provide transparency and pedagogical significance [Tjandra et al. \(2024\)](#). The most important contribution of this paper is that it shows how predictive analytics can efficiently predict student performance, identify at-risk learners, and give interpretable information that underlies specific interventions. The alignment of data-driven prediction and educational decision-making in the framework will push the scalable learning analytics solutions to creative and practice-oriented academic setting [Tiwari and Jain \(2024\)](#).

## 2. RELATED WORK

Student performance prediction is a dynamic field of research in higher education analytics because of the increased access to institutional and learning management system data. Linear regression, correlation analysis are the main statistical methods that were used in the early research involving the modeling of academic results through attendance, former grades, and demographic factors [Skalka and Drlik \(2020\)](#). As the computational techniques improved, machine learning models such as decision trees, support vector machines and the ensemble techniques have shown to predict better due to their ability to harness non-linear relationships and multi-facet interactions of features [Tjandra et al. \(2024\)](#). These methods have been highly used to anticipate final grades, course completion, and dropout possibility with promising enhancement in the accuracy and prompt identification of academically vulnerable learners. The research of learning analytics has over time grown beyond the conventional STEM and other theory-based courses to incorporate creative and digital media studies. Nevertheless, the use of predictive analytics in these areas is relatively small [Maraza-Quispe et al. \(2021\)](#). The digital media education on project based learning, iterative design process and creative exploration focus on generation of rich yet unstructured information like version histories, design artifacts and interaction logs. The current research in the field of creative education analytics has been dedicated to the descriptive analysis of the engagement patterns and qualitative analysis of the creative output, but not predictive modeling [Santos and Henriques \(2023\)](#). Consequently, an absence exists in the use of behavioural and process-oriented data to predict

student achievement in creative processes despite creativity having a high potential in informing instructional design and feedback processes.

Educational data mining methods that utilize machine learning processes have also advanced with the most powerful and famous methods of ensemble like the Random Forest and Gradient Boosting because of their strength and the predictive accuracy [Alalawi et al. \(2024\)](#). The latter models are especially useful in processing heterogeneous educational data that are a combination of academic, behavioral, and temporal characteristics. Logistic Regression remains a popular choice of the baseline model due to its interpretability and simplified implementation in educational institutions. Recent research has also researched into hybrid models and temporal learning methods to enhance the level of early-warning by adding time-series aspects based on student activity logs [Sagala et al. \(2022\)](#). Though these approaches have been significantly successful in their predictive accuracy, most studies put more importance on the performance measures in lieu of model transparency and educational interpretability [Adegbite \(2025\)](#).

The limitation which is critical in the current literature on student performance prediction is that it does not have a prediction which can be explained to the educator and taken action on. Numerous machine learning models that perform well are black box models, and they provide a poor insight into why some students are expected to be at risk [Karmagatri et al. \(2023\)](#). It makes predictive systems less trustworthy to the instructors and makes it harder to implement the predictive systems practically in academic settings. Moreover, early warning systems tend to be binary with no background information that would give instructions on specific intervention measures. These gaps also have the implication of predictive frameworks that are capable of not only producing high accuracy but also of combining the explainable AI methods with the intention of presenting feature-wise influence and learning behavioral patterns. The response to these shortcomings is especially significant in the education of digital media when subtle awareness of engagement, creativity, and iteration are crucial in providing scholarly support.

**Table 1**

Table 1 Summary of Related Work on Student Performance Prediction and Learning Analytics					
Ref. No.	Educational Context	Data Sources Used	ML / Analytical Methods	Key Focus Area	Identified Limitations
<a href="#">Skalka and Drlik (2020)</a>	General Higher Education	Attendance, grades, demographics	Linear Regression, Statistical Models	Grade prediction	Limited handling of non-linearity
<a href="#">Tjandra et al. (2024)</a>	Undergraduate Courses	LMS activity, assessments	Decision Trees, SVM	Performance forecasting	Poor early-stage prediction
<a href="#">Maraza-Quispe et al. (2022)</a>	Creative & Digital Media Education	Project artifacts, engagement logs	Descriptive Analytics	Learning behavior analysis	Lack of predictive modeling
<a href="#">Santos and Henriques (2023)</a>	Design & Media Studies	Interaction logs, submissions	Rule-based Analytics	Engagement measurement	No risk prediction capability
<a href="#">Alalawi et al. (2024)</a>	Higher Education (Mixed Domains)	Academic + behavioral data	Random Forest, Gradient Boosting	Accuracy optimization	Black-box model behavior
<a href="#">Sagala et al. (2022)</a>	Online Learning Platforms	Temporal activity logs	ML + Time-Series Analysis	Early-warning detection	Limited interpretability
<a href="#">Karmagatri et al. (2023)</a>	Academic Risk Prediction	LMS, performance metrics	Ensemble Models	Dropout risk identification	No explainable insights
<a href="#">Tjandra et al. (2024)</a>	Engineering Education	Quiz scores, attendance	Logistic Regression	Baseline prediction	Lower accuracy than ensembles
<a href="#">Alalawi et al. (2024)</a>	Multidisciplinary Programs	Mixed structured data	Ensemble Learning	Robust prediction	High computational complexity

### 3. DATASET DESCRIPTION AND FEATURE ENGINEERING

#### 3.1. DIGITAL MEDIA COURSE STRUCTURE AND DATASET DESCRIPTION

In this research, the OULAD (Open University Learning Analytics Dataset), which is a publicly available and widely tested dataset, is used to model the performance in digitally mediated and practice-oriented courses. OULAD also has extensive data on interactions between students in virtual learning spaces and thus can be used in an educational setting involving digital media and creative technologies. The data used is of several courses which are provided on-line and have more than 32,000 students, 22 course presentations and additional than 10 million interaction records. Types of

data can be demographic data, evaluation data and detailed logs of activity like clicks, frequency of accessing content, and times of submissions. The performance of students is put in outcome classes like pass, fail, withdraw, and distinction to facilitate supervised learning and early-risk prediction activities in digital learning processes.

### **3.2. ACADEMIC, BEHAVIORAL, ENGAGEMENT AND CREATIVE PERFORMANCE CHARACTERISTICS**

Features are categorized into four complementary categories in order to provide multidimensional learning behaviour. The academic aspects are assignment marks, assessment percentage, cumulative marks, and timely submission. Behavioral characteristics are the learning habits including frequency of using the platform, time spent on the platform, and regularity of weekly use. The engagement features are based on the intensity of interaction with digital resources, discussion forums, and other learning material, which represent active participation in course work. Aspects of exploratory learning and repetitive enhancement associated with the use of digital media in education are reflected in creative performance features, which can be approximated by the number of project submissions and by assessment rubrics. Combined, these groups of features will give a comprehensive description of student learning processes, with outcome and process-related cues that are vital to the modeling of performance within courses that are creative and practice focused.

### **3.3. PREPROCESSING AND CONSTRUCTION OF DATA**

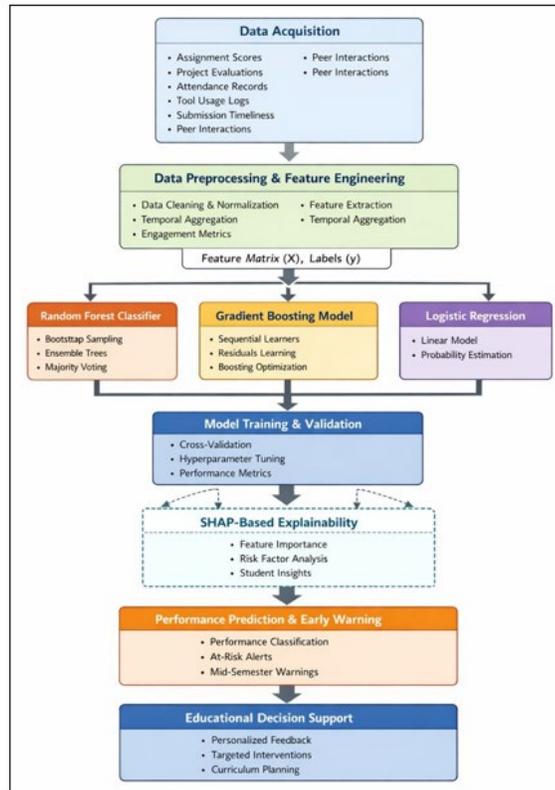
The preprocessing of data consists of eliminating duplicate records, processing missing data with the use of statistical imputation, and separation of incomplete student records. Min max scaling is used to normalize numerical features to provide equal contribution across models and the one-hot and ordinal encoding are used to encode categorical variables. Stratified sampling and synthetic minority oversampling are used to handle class imbalance in order to enhance predictive fairness. The feature construction is based on the temporal and engagement-based features, including rolling averages of weekly activity, trend slope of engagement intensity, submission delay ratio, early-to-mid semester change rate of participation and so on. These artificial capabilities allow prediction of early warnings since they are not only able to record the performance level that is static but also dynamic learning behavior.

## **4. PREDICTIVE MODELLING METHODOLOGY**

### **4.1. OVERVIEW OF THE PREDICTIVE ANALYTICS PIPELINE**

The predictive analytics pipeline will be a workflow that is structured into an end-to-end process to convert raw educational data into performance predictions that should be used. It starts with processed and engineering-based student data that include academic performance, engagement, and time related factors. These characteristics are divided into training and validation parts to assure the objective learning of models. Several predictive models are then trained in parallel to model various learning pattern evident in courses in digital media. Both ensemble and linear models work in the same feature space, allowing them to be predictive with good accuracy and at the same time be comparatively interpretable. In the training, the quality of the model outputs is constantly measured in terms of standard classification metrics to determine early-warning possibilities and final performance forecasting. The pipeline also incorporates explainability by means of a post-hoc analysis that enables model expectations to be explained with regards to the influence of features. The architecture allows predicting at mid-semester based on the temporal features and refines predictions with the help of validation feedback. In general, the pipeline is scalable, modular and geared towards the needs of creative, practice oriented, learning settings in which learning behaviours change dynamically over time.

**Figure 1**



**Figure 1** Explainable Ensemble-Based Predictive Analytics Architecture for Student Performance in Digital Media Courses

The end-to-end representation of the proposed student performance evaluation predictive analytics framework is represented in the [Figure 1](#). It starts with the multi-source data acquisition, which is followed by the preprocessing and feature engineering to create the organized learning indicators. Random Forest, Gradient Boosting, and Logistic Regression are the ensemble learning models that are trained and qualified to provide performance predictions. SHAP-based explainability layer can be used to provide insights on the features level, allowing to identify at-risk students early and conduct informed learning due to personalized feedback and specific academic intervention.

## 4.2. PREDICTIVE MODELS

### 4.2.1. RANDOM FOREST FOR NON-LINEAR PATTERN LEARNING AND ROBUSTNESS

Random Forest is used to learn complex, non-linear relationships which are inherent in digital media learning data. The model has the effect of capturing interactions between academic performance, engagement intensity and creative iteration behavior by building an ensemble of decision trees, which are trained on bootstrapped samples and random feature subsets. The fact that it is naturally robust to noise and overfitting means that it is best applied in heterogeneous educational data. Random Forest in this study also offers good baseline accuracy and can be used in a wide range of feature distributions and non-uniform learning curves that are typical of creative and practice-based courses.

#### ALGORITHM 1: RANDOM FOREST-BASED STUDENT PERFORMANCE PREDICTION

**Input:**

- Student dataset  $D = \{(x_i, y_i)\}$  for  $i = 1$  to  $N$
- Number of trees  $T$
- Maximum tree depth  $d$

**Output:**

- Predicted student performance class  $\hat{y}$

**Steps:**

- 1) Initialize an empty forest  $F = \{\}$
- 2) For  $t = 1$  to  $T$  do:
  - Draw a bootstrap sample  $D_t$  from dataset  $D$ .
  - Train a decision tree  $h_t$  on  $D_t$ .
  - At each split, randomly select a subset of features.
  - Restrict tree growth using depth  $d$  or minimum leaf size.
  - Add trained tree  $h_t$  to forest  $F$ .
- 3) For a test instance  $x$ :
  - Obtain predictions  $y_t = h_t(x)$  from all trees in  $F$ .
- 4) Compute final prediction using majority voting:

$$y_{hat} = mode(y_1, y_2, \dots, y_T).$$

- 5) Return  $y_{hat}$ .

**4.2.2. GRADIENT BOOSTING FOR HIGH-ACCURACY ENSEMBLE LEARNING**

Gradient Boosting is utilized to achieve high predictive accuracy through sequential ensemble learning. The model iteratively builds weak learners that focus on correcting errors made by previous iterations, enabling refined learning of subtle performance patterns. This approach is particularly effective in identifying early risk signals embedded in temporal engagement features and submission behaviors. In the context of digital media courses, Gradient Boosting captures incremental performance shifts across project milestones, leading to improved early-warning prediction and superior classification performance compared to single-model approaches.

**ALGORITHM 2: GRADIENT BOOSTING-BASED STUDENT PERFORMANCE PREDICTION**

**Input:**

- Training dataset  $D = \{(x_i, y_i)\}$  for  $i = 1$  to  $N$
- Number of boosting iterations  $M$
- Learning rate  $\eta$

**Output:**

- Predicted student performance  $y_{hat}$

**Steps:**

- 1) Initialize the model with a constant prediction:

$$F_{0(x)} = argmin_c \sum L(y_i, c)$$

- 2) For  $m = 1$  to  $M$  do:
  - Compute pseudo-residuals:

$$r_{im} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right] \text{evaluated at } F = F_{m-1}$$

- Train a weak learner  $h_m(x)$  to fit  $r_{im}$ .
- Compute optimal step size  $\gamma_m$ :

$$\gamma_m = \operatorname{argmin}_{\gamma} \sum L(y_i, F_{(m-1)}(x_i) + \gamma h_m(x_i))$$

- Update the ensemble model:

$$F_{m(x)} = F_{(m-1)(x)} + \eta \gamma_m h_m(x)$$

- 3) Use  $F_M(x)$  for prediction.
- 4) Return  $y_{\text{hat}}$ .

### 4.2.3. LOGISTIC REGRESSION AS AN INTERPRETABLE BASELINE CLASSIFIER

As a predictive performance benchmark, the Logistic Regression is taken. Although the decision boundary is linear, the model provides clear information on where the feature impacts on students and the extent. Logistic Regression allows estimating performances in a probability-based manner, allows making predictions in the initial stages, and offers a point of reference against which the advantages of ensemble learning models can be evaluated in a systematic way.

#### ALGORITHM 3: LOGISTIC REGRESSION–BASED STUDENT PERFORMANCE PREDICTION

##### Input:

- Feature matrix  $X \in \mathbb{R}^{(N \times d)}$
- Label vector  $y \in \{0,1\}$
- Learning rate  $\alpha$
- Number of iterations  $K$

##### Output:

- Predicted class label  $y_{\text{hat}}$

##### Steps:

- 1) Initialize weight vector  $w$  and bias  $b$ .
- 2) For  $k = 1$  to  $K$  do:
  - Compute linear score:

$$z_i = w^T x_i + b.$$

- Apply sigmoid function:

$$p_i = \frac{1}{1 + \exp(-z_i)}.$$

- Compute binary cross-entropy loss.
- Update parameters using gradient descent:

$$\begin{aligned} w &= w - \alpha \nabla_w L \\ b &= b - \alpha \nabla_b L \end{aligned}$$

- 3) For a test instance x:
  - Compute probability p.
  - Assign class label:
 if  $p \geq 0.5$ ,  $y_{\text{hat}} = 1$   
 else  $y_{\text{hat}} = 0$
- 4) Return  $y_{\text{hat}}$ .

#### 4.2.4. PROPOSED ENSEMBLE + SHAP

The Ensemble + SHAP model combines the predictive strength of ensemble learning with the transparency of explainable artificial intelligence to deliver accurate and interpretable student performance prediction. This integration makes it robust and better generalization is achieved in relation to various learning behaviors that are usually prevalent in digital media courses. SHAP is used as an interpretability layer to apply post hoc to explain ensemble predictions as a feature-level contribution, thus providing both global and student level explanations. With the disclosure of how engagement consistency, submission punctuality and project iteration have an effect on predictions, the model is useful in making pedagogically significant disclosures. This combinatory solution is more predictive and transparent and provides support to actionable decision-making to support early academic intervention.

##### ALGORITHM 4: SHAP-BASED FEATURE INFLUENCE ANALYSIS

**Input:**

- Trained prediction model  $f(x)$
- Feature set X with F features

**Output:**

- SHAP values  $\phi_j$  for each feature j

**Steps:**

- 1) For each feature j in F:
  - Compute marginal contribution:

$$\phi_j = \sum \left[ |S|! \frac{(|F| - |S| - 1)!}{|F|!} \right] \times (f(S \cup \{j\}) - f(S)),$$

where  $S \subseteq F \setminus \{j\}$ .

- 2) Compute SHAP values for all samples.
- 3) Aggregate mean absolute SHAP values.
- 4) Rank features based on importance.
- 5) Interpret global and local feature effects.
- 6) Return feature influence explanations.

#### 4.3. MODEL TRAINING, VALIDATION STRATEGY, AND HYPERPARAMETER TUNING

Training of the models is stratified to train and validate the model with a stratified split of the training to avoid bias. Five-fold cross-validation has been used to provide a robust and generalizable result. In the case of Random Forest, the most important hyperparameters are the number of trees, tree depth limit, minimum sample per leaf and size of the feature subset. Gradient Boosting tuning is based on boosting iterations, learning rate, tree depth and subsampling ratio to balance between accuracy and overfitting. The regularization strength and type of penalty are used to control the model complexity by minimizing the Logistic Regression. Systematic identification of the best combinations of parameters is done through grid search and validation performance is used to create a final model.

## 5. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

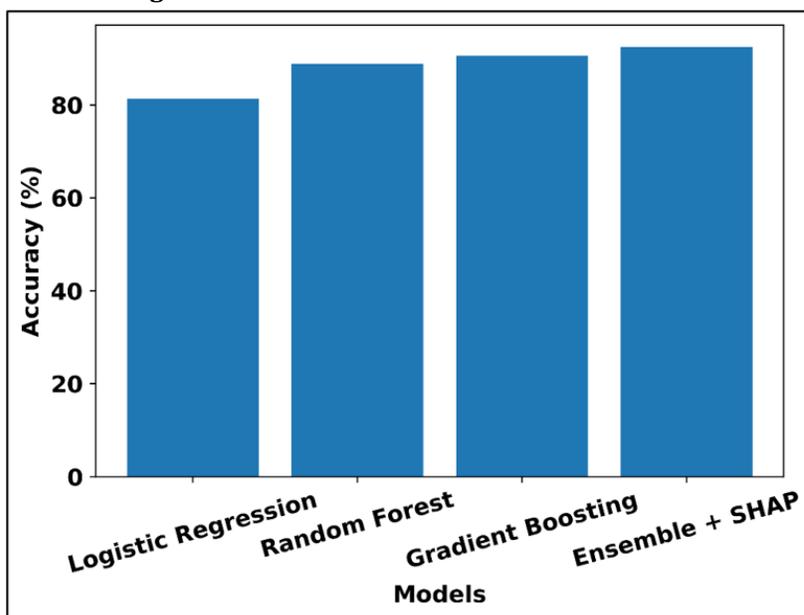
The similar and comparison analysis of predictive models used in digital media courses with predicting student performance was provided in [Table 2](#). The Logistic Regression is a baseline model and has an accuracy of 81.34 indicating its poor modeling ability of non-linear associations between academic, engagement, and time features. Random Forest model has considerably higher predictive accuracy of 88.92 percent and AUC of 0.91 indicating how well ensemble learning can capture the very complex interaction of features.

**Table 2**

Table 2 Comparative Performance of Predictive Models					
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Logistic Regression	81.34	79.25	80.18	79.71	0.84
Random Forest	88.92	87.4	88.15	87.77	0.91
Gradient Boosting	90.67	89.85	90.21	90.03	0.93
<b>Proposed Framework (Ensemble + SHAP)</b>	<b>92.56</b>	<b>91.48</b>	<b>92.02</b>	<b>91.75</b>	<b>0.95</b>

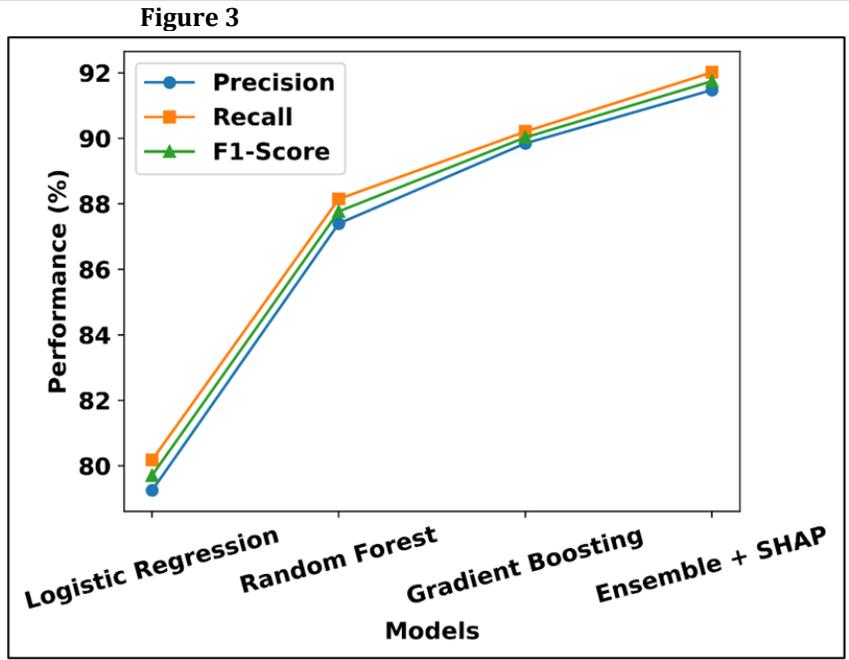
Gradient Boosting also increases the performance more, reaching 90.67% accuracy and a better F1-score of 90.03, which implies a better balance between precision and recall. The suggested Ensemble + SHAP framework is better than all the single models with the highest accuracy of 92.56 percent and AUC of 0.95.

**Figure 2**



**Figure 2** Accuracy Comparison of Predictive Models for Student Performance Prediction

[Figure 2](#) compares the accuracy of various predictive models that are used to predict the performance of students. Logistic Regression has the worst accuracy as it is not very effective at modeling the learning behaviors. Untrained ensemble-based models attain an increasingly better accuracy, and Gradient Boosting and Random Forest win over the baseline. The suggested Ensemble + SHAP framework is significantly the most accurate, which proves that ensemble learning with explainable analytics is effective.



**Figure 3** Precision, Recall, and F1-Score Comparison of Predictive Models

This enhancement serves to emphasize the advantage of multitasking a combination of several ensemble learners with explainability mechanism. The high accuracy and recall scores validate the trustworthy detection of the high performing and at-risk students. On the whole, the findings justify the strength, validity, and usefulness of the suggested framework in predictive analytics in online media education. The [Figure 3](#) shows a comparative study of precision, recall and F1 score of various predictive models. Logistic Regression tends to have the worst performance in all metrics meaning that it has little effectiveness in the consideration of complex learning patterns. There is a steady increase in ensemble-based models, with the highest and more balanced scores assigned to the Random Forest and Gradient Boosting. Its Ensemble + SHAP framework shows the best performance in general, which is an indication of sound and consistent prediction of student performance.

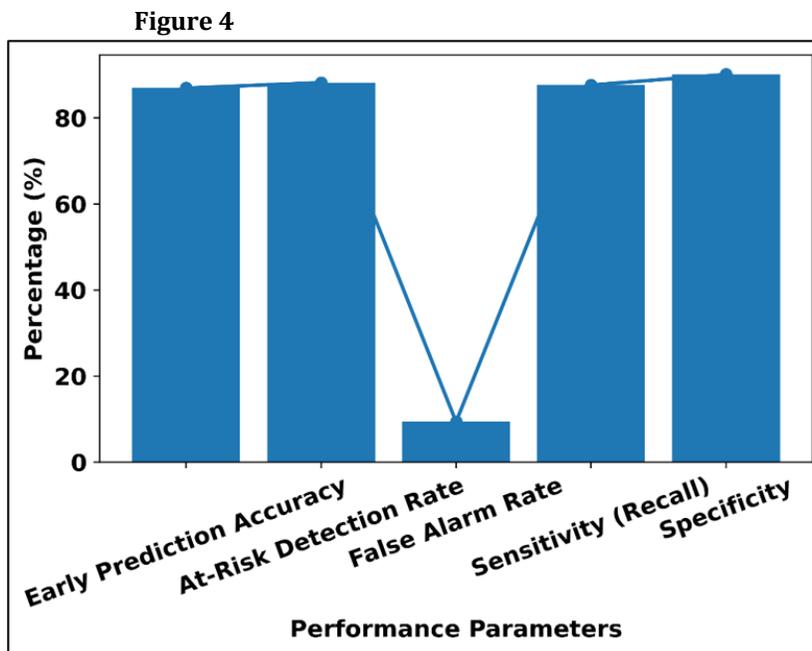
### 5.1. MID-SEMESTER PREDICTION EFFECTIVENESS AND RISK IDENTIFICATION

In order to assess early-warning ability mid-semester forecasts were assessed with various performance indicators, risk-oriented. [Table 3](#) is a summary of the performance of the proposed framework in predicting risks in the mid-semester which shows that the proposed framework is an effective early-warning system. Accuracy in prediction at an early stage of 86.94 per cent proves that the model can provide valid performance predictions even before the course is completed.

**Table 3**

Table 3 Mid-Semester Risk Prediction Performance	
Parameter	Value
Early Prediction Accuracy (%)	86.94
At-Risk Student Detection Rate (%)	88.21
False Alarm Rate (%)	9.37
Sensitivity (Recall) (%)	87.65
Specificity (%)	90.12
Intervention Lead Time (Weeks)	5-6

The rate of at-risk student detection of 88.21% shows that there is a high level of ability to detect students that need academic assistance. The false alarm rate of 9.37 is low, which minimizes unnecessary interventions, which increases its practical use. Use of high sensitivity (87.65) and specificity (90.12) to verify the balanced performance of the correct identification of the at-risk and the non-risk students. Notably, the lead time of 5-6 weeks on an intervention basis offers adequate space to take instructional and mentoring measures in time to ensure proactive academic decision-making in courses in digital media.



**Figure 4** Mid-Semester Risk Prediction Performance Analysis

The [Figure 4](#) shows the risk prediction performance of the proposed framework in mid-semester through a line-bar visualization. The values of the early prediction accuracy, at-risk detection rate, sensitivity, and specificity are high, which indicates that the model has a high ability to detect vulnerable students with high reliability. The relatively small false alarm shows lower false classification and unnecessary interventions. In sum, the number proves the fact that the framework provides balanced and strong early-warning performance, which allows timely academic support and informed decision-making in courses dealing with digital media.

## 5.2. DISCUSSION OF QUANTITATIVE IMPROVEMENTS AND ANALYTICAL FINDINGS

It is evident in the results of the experiments that the performance of the model of ensemble-based learning is better as compared to that of the traditional linear classifier when being used to predict the performance of students in digital media courses. Logistic Regression, which is interpretable, is less accurate than the other because it cannot account for non-linear interactions between engagement, academic and temporal characteristics. Random Forest is better at predicting robustness due to modeling of interaction between complex features whereas Gradient Boosting is even more accurate due to error correction during the iterative process. The suggested structure reaches the optimal performance through the combination of ensemble learning and systematic feature engineering and explainability.

The analysis conducted in the middle of the semester confirms the fact that the framework can offer high-quality early-warning signals and identify at-risk students with a high sensitivity and low false-alarm rates. Potential opportunities to provide specific academic intervention through the possibility to make proper forecasts 5-6 weeks prior to the course completion. Altogether, these results justify the importance of predictive analytics in creative and practice-based education and the importance of the combination of accuracy-driven models with interpretable information to make informed educational decisions.

## 6. CONCLUSION

This paper has developed an elaborate predictive analytics model to model and predict student achievement in digital media classes and has dealt with the special issues of creative and practice-based learning. The proposed approach is based on the baseline of academic, behavioural, engagement, and temporal parameters of learning thus capturing both outcome and process-driven learning outcomes that traditional methods of assessment frequently fail to capture. Enhanced the application of ensemble learning frameworks, namely the Random Forest and Gradient Boosting allowed modeling the complex, non-linear relationships, which emerged in the context of digital media learning, in an efficient manner, and the application of the Logistic Regression allowed the transparent baseline, against which the models could be compared. The experimental findings indicate that the suggested Ensemble + SHAP framework can have the high predictive accuracy of 92.56 and it is incomparably higher than the traditional models. Mid-semester analysis further proves the capability of the framework in detecting the student's in-real-need with high sensitivity and low false-alarm rates and giving actionable early-warning signals a few weeks before the course completion. The current results underscore the importance of predictive analytics in facilitating the active academic interventions, student-specific feedback, and instructional planning. One of the primary works of this study is the incorporation of SHAP-based explainability that converts predictive results into comprehensible output in line with pedagogical decision-making. The framework facilitates the ethical use of AI in teaching since it enables the instructor to trust the tool by revealing the influence at the feature level. All in all, the suggested research approach shows that accuracy-based ensemble models with explainable AI have a scalable and people-oriented solution to enhance the performance of the learners in digital media education.

## CONFLICT OF INTERESTS

None.

## ACKNOWLEDGMENTS

None.

## REFERENCES

- Adegbite, M. A. (2025). Data Privacy and Data Security Challenges in Digital Finance. *Journal of Digital Security and Forensics*, 2(1), 6–19. <https://doi.org/10.29121/digisecforensics.v2.i1.2025.40>
- Alalawi, K., Athauda, R., and Chiong, R. (2024). An Extended Learning Analytics Framework Integrating Machine Learning and Pedagogical Approaches for Student Performance Prediction and Intervention. *International Journal of Artificial Intelligence in Education*, 34. <https://doi.org/10.1007/s40593-024-00429-7>
- Barlaskar, E., Cutting, D., Allen, A., and McDowell, A. (2025). SAPP: Student Academic Performance Predictor. In *Proceedings of the 2025 IEEE Global Engineering Education Conference (EDUCON)* (1–7). IEEE. <https://doi.org/10.1109/EDUCON62633.2025.11016395>
- Bravo-Agapito, J., Romero, S., and Pamplona, S. (2020). Early Prediction of Undergraduate Students' Academic Performance in Completely Online Learning: A Five-Year Study. *Computers in Human Behavior*, 106, Article 106595. <https://doi.org/10.1016/j.chb.2020.106595>
- Fahd, K., and Miah, S. J. (2025). Enhanced Predictive Performance: A Comparative Analysis of ML and DL Models Using Augmented LMS Interaction Data. *Contemporary Educational Technology*, 17(4), Article ep606. <https://doi.org/10.30935/cedtech/17453>
- Karmagatri, M., Kurnianingrum, D., Suciana, M. R., and Utami, S. A. (2023). Predicting Factors Related to Student Performance Using Decision Tree Algorithm. In *Proceedings of the 5th International Conference on Cybernetics and Intelligent System (ICORIS)* ( 1–6). IEEE. <https://doi.org/10.1109/ICORIS60118.2023.10352269>
- Lee, C., Luo, L., Kuhlmann, S. L., Plumley, R. D., Panter, A. T., Bernacki, M. L., Greene, J. A., and Gates, K. M. (2025). Interpretable Predictive Analytics for Online Learning: A Markov-Based Machine Learning Approach. *Journal of Learning Analytics*, 12(2), 259–278. <https://doi.org/10.18608/jla.2025.8375>

- Maraza-Quispe, B., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., and Apaza-Huanca, J. M. (2022). Predictive Model of Student Academic Performance from LMS Data Based on Learning Analytics. In Proceedings of the 13th International Conference on Education Technology and Computers (ICETC '21) (13–19). Association for Computing Machinery. <https://doi.org/10.1145/3498765.3498768>
- Maraza-Quispe, B., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., and Apaza-Huanca, J. M. (2021). Predictive Model of Student Academic Performance from LMS Data Based on Learning Analytics. In Proceedings of the 13th International Conference on Education Technology and Computers (ICETC '21) (13–19). Association for Computing Machinery. <https://doi.org/10.1145/3498765.3498768>
- Poh, Z. X., and Khor, E. T. (2024). Predictive Analytics for Student Online Learning Performance Using Machine Learning and Data Mining Techniques. In G. Marks (Ed.), Proceedings of the International Journal on E-Learning 2024 ( 269–283). Association for the Advancement of Computing in Education (AACE). <https://doi.org/10.70725/686437bpltsf>
- Sagala, T. M. N., Permai, S. D., Gunawan, A. A. S., Barus, R. O., and Meriko, C. (2022). Predicting Computer Science Students' Performance Using Logistic Regression. In Proceedings of the 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (817–821). IEEE. <https://doi.org/10.1109/ISRITI56927.2022.10052968>
- Santos, R. M., and Henriques, R. (2023). Accurate, Timely, and Portable: Course-Agnostic Early Prediction of Student Performance from LMS Logs. Computers and Education: Artificial Intelligence, 5, Article 100175. <https://doi.org/10.1016/j.caeai.2023.100175>
- Skalka, J., and Drlik, M. (2020). Automated Assessment and Microlearning Units as Predictors of at-Risk Students and Students' Outcomes in Introductory Programming Courses. Applied Sciences, 10(13), Article 4566. <https://doi.org/10.3390/app10134566>
- Sunawar Khan, Mazhar, T., Shahzad, T., Amir Khan, M., Waheed, W., Waheed, A., and Hamam, H. (2025). Predictive Analytics in Education: Enhancing Student Achievement Through Machine Learning. Social Sciences and Humanities Open, 12, Article 101824. <https://doi.org/10.1016/j.ssaho.2025.101824>
- Syed Mustapha, S. M. F. D. (2023). Predictive Analysis of Students' Learning Performance Using Data Mining Techniques: A Comparative Study of Feature Selection Methods. Applied System Innovation, 6(5), Article 86. <https://doi.org/10.3390/asi6050086>
- Tiwari, M., and Jain, N. (2024). Student Performance Prediction Using Machine Learning Algorithms. ShodhKosh: Journal of Visual and Performing Arts, 5(6), 1112–1122. <https://doi.org/10.29121/shodhkosh.v5.i6.2024.4552>
- Tjandra, E., Ferdiana, R., and Setiawan, N. A. (2024). Competency Framework in Higher Education: A Bibliometric Analysis from 2000 to 2023. In Proceedings of the 8th International Conference on Education and Multimedia Technology (ICEMT '24) (262–268). Association for Computing Machinery. <https://doi.org/10.1145/3678726.3678736>
- Tjandra, E., Ferdiana, R., and Setiawan, N. A. (2024). OBE-Based Course Outcomes Prediction Using Machine Learning Algorithms. In Proceedings of the 2024 International Conference on Intelligent Cybernetics Technology and Applications (ICICyTA) ( 197–202). IEEE. <https://doi.org/10.1109/ICICYTA64807.2024.10913307>