International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com
Volume 3 Issue 9, September 2026
E-ISSN : 3049-2424
DOI: https://doi.org/10.71366/ijwos

PREDICTIVE MODEL FOR STUDENT DROPOUT RISK IN ONLINE LEARNING ENVIRONMENT

K.Santhiya[1], Dr. L. Baby Victoria[2]
[1] II-MCA, Department of Computer Applications, Vellalar College for Women (Autonomous), Erode, Tamil Nadu, India
[2] Department of Computer Applications, Vellalar College for Women (Autonomous), Erode, Tamil Nadu, India.

Article Info

Article History:
Published: 22 Sept 2026

Publication Issue:
Volume 3, Issue 9
September-2026

Page Number:
376-380

Corresponding Author:
K.Santhiya

Abstract:

Online learning environments have become an important part of modern education by providing flexible access to courses, learning materials, assessments, and communication facilities. However, student dropout remains a major challenge because learners may discontinue their courses due to poor academic performance, low engagement, irregular participation, lack of interaction, personal difficulties, or technical problems. This paper proposes a predictive model for identifying student dropout risk in online learning environments using machine learning techniques. The proposed approach analyzes student-related attributes such as attendance, assignment submission, assessment performance, course activity, login frequency, and interaction with learning resources. The system performs data preprocessing, feature selection, model training, and classification to identify students who may be at risk of dropping out. Machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine can be evaluated using standard classification metrics. The proposed model can assist educators and administrators in identifying students who may require early academic support. It provides a data-driven decision-support approach for improving student monitoring and retention in online education.

Keywords: Classification, Dropout Prediction, E-Learning, Machine Learning, Student Engagement, Student Retention

1. INTRODUCTION

The rapid development of internet technologies has significantly increased the use of online learning platforms in education. Online learning provides students with flexibility to access educational content from different locations and at different times. Learning management systems can record information related to student activities, including course access, assignment submission, assessment performance, attendance, discussion participation, and learning-resource usage.

Although online education provides several advantages, student dropout continues to be a significant challenge. A student may stop participating in an online course because of poor academic performance, lack of motivation, insufficient interaction, irregular attendance, personal difficulties, or technical issues. When a student gradually reduces participation, it may be difficult for instructors to identify the problem at an early stage.

Machine learning provides an opportunity to analyze historical student data and identify patterns associated with dropout. Predictive models can process multiple student attributes and estimate dropout risk. Early identification can allow instructors and institutions to provide appropriate academic guidance, reminders, counselling, learning

resources, or other support.

The objective of this study is to develop a predictive model that analyzes student learning behaviour and identifies students who may be at risk of dropping out from an online learning environment.

2. EXISTING SYSTEM

Many online learning environments monitor students using conventional methods such as examination results, attendance records, assignment marks, and instructor observation. These methods provide useful information but may not identify dropout risk sufficiently early.

Manual monitoring becomes difficult when the number of students increases. Different students may also demonstrate different patterns of disengagement. For example, one student may show reduced login frequency, while another may continue logging in but stop submitting assignments.

Traditional monitoring systems generally focus on current academic performance rather than combining multiple behavioural and academic indicators. Therefore, an automated approach is needed to analyze several student characteristics simultaneously.

3. PROPOSED SYSTEM

The proposed system uses machine learning to predict student dropout risk in an online learning environment. The system collects relevant academic and behavioural attributes, cleans and transforms the data, selects important features, and trains classification models. The trained model predicts whether a student may be at risk of dropping out.

The prediction is intended as decision-support information for educators rather than a final judgment about a student's future behaviour.

The proposed system can be represented as:

Student Data
↓
Data Collection
↓
Data Preprocessing
↓
Feature Selection
↓
Machine Learning Model
↓
Dropout Risk Prediction
↓
Risk Classification
↓
Early Warning / Intervention

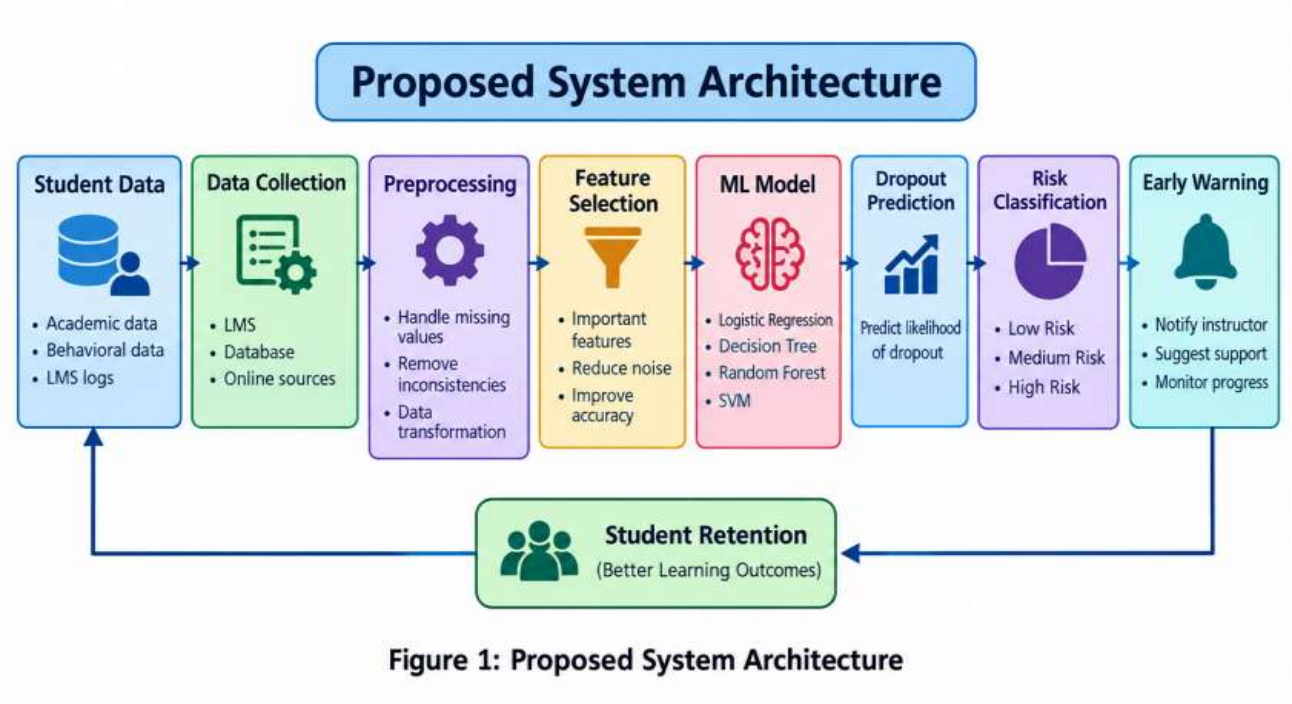

Fig 1

4. METHODOLOGY

4.1. Data Collection

The dataset may contain academic and behavioural information related to students enrolled in an online learning environment. Possible attributes include student ID, attendance, assignment submission rate, assessment marks, login frequency, course activity, time spent on learning resources, discussion participation, previous academic performance, and final dropout status. The target variable represents the student's dropout or completion status.

4.2. Data Preprocessing

Raw datasets may contain missing values, duplicate records, inconsistent values, or categorical attributes. The preprocessing stage includes removing duplicate records, handling missing values, converting categorical data into numerical form, removing irrelevant attributes, scaling numerical features when required, and separating input features from target labels.

4.3. Feature Selection

Feature selection identifies attributes that are useful for predicting dropout. Important features may include attendance, assignment submission, assessment performance, login frequency, course activity, learning-resource usage, and discussion participation. Feature selection can reduce unnecessary information and focus the model on relevant student characteristics.

4.4. Machine Learning Models

Logistic Regression can estimate the probability of a dropout outcome. Decision Tree represents classification decisions through a tree structure. Random Forest combines multiple decision trees and can handle several student features. Support Vector Machine identifies a decision boundary between different classes. These algorithms can be trained and compared using the same dataset and evaluation procedure.

5. MODEL EVALUATION

The processed dataset is divided into training and testing datasets. The training data is used to build the models, while the testing data is used to evaluate their performance. Accuracy, precision, recall, and F1-score can be used as evaluation metrics.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1-score = 2 × (Precision × Recall) / (Precision + Recall)

For dropout-risk prediction, recall is particularly useful because failing to identify an at-risk student may reduce the opportunity for early support.

6. RESULTS AND DISCUSSION

The experimental evaluation should compare the selected machine learning algorithms using the same dataset and evaluation procedure. The actual values obtained from the experiment should be reported for accuracy, precision, recall, and F1-score. A confusion matrix can also be generated to analyze correct and incorrect classifications. Feature-importance analysis can help identify the student attributes that contribute to the prediction.

Student engagement indicators such as attendance, assignment submission, login frequency, and course activity can provide useful information about learning behaviour. Combining these attributes with academic performance may provide a broader representation of student participation. The prediction should be used together with instructor knowledge and individual student circumstances when considering support.

7. ADVANTAGES AND LIMITATIONS

The proposed system supports early identification of possible dropout risk, reduces dependence on manual monitoring, processes multiple student attributes simultaneously, provides measurable machine learning predictions, supports early intervention, and can be integrated with online learning platforms.

However, prediction quality depends on dataset quality and representativeness. Missing or incorrect information can affect performance, and behavioural data may not capture all reasons for dropout. A prediction does not guarantee that a student will actually drop out. Student data must also be handled carefully to protect privacy.

8. FUTURE ENHANCEMENT

Future work can incorporate larger and more diverse datasets from different online learning platforms. Additional machine learning and deep learning techniques can be evaluated. A real-time monitoring dashboard could display engagement indicators and updated risk predictions. Personalized recommendations, explainable machine learning, and privacy-preserving techniques can also be incorporated.

9. CONCLUSION

This paper presents a machine learning-based approach for predicting student dropout risk in online learning environments. The proposed system analyzes academic and behavioural characteristics such as attendance, assessment performance, assignment submission, login frequency, and course activity. Data preprocessing and feature selection are performed before applying classification algorithms.

The proposed approach can assist educational institutions in identifying students who may require additional attention at an earlier stage. The actual effectiveness of the models should be established through experiments using a suitable

dataset and standard evaluation metrics. The system can be further enhanced through real-time monitoring, explainable predictions, larger datasets, and integration with online learning platforms.

10. ACKNOWLEDGMENT

The author would like to express sincere gratitude to the Department of Computer Applications, [Vellalar College For Women], for providing the facilities and support required to carry out this work.

11. AUTHOR CONTRIBUTION

The author contributed to the conceptualization, literature study, methodology design, data preprocessing, model development, analysis, and preparation of the manuscript.

12. CONFLICT OF INTEREST

The author declares that there is no conflict of interest associated with this research work.

References

[1] Prenkaj, B., Velardi, P., Stilo, G., Distante, D., & Faralli, S. (2020). A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses. *ACM Computing Surveys*, 53(3). https://doi.org/10.1145/3388792.

[2] Vaarma, M., & Li, H. (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. *Technology in Society, 76*, 102474. https://doi.org/10.1016/j.techsoc.2024.102474.

[3] Alharahsheh, H. F., et al. (2023). Predictive Modeling of Student Dropout in MOOCs and Self-Regulated Learning. *Computers, 12*(10), 194. https://doi.org/10.3390/computers12100194.

[4] Predicting Student Dropout in Self-Paced MOOC Course Using Random Forest Model. (2021). *Information, 12*(11), 476.

[5] MOOC dropout prediction using machine learning techniques: Review and research challenges. (2018). *IEEE Global Engineering Education Conference (EDUCON)*. https://doi.org/10.1109/EDUCON.2018.8363340.

[6] Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. (2020). *Decision Support Systems, 135*, 113325.

[7] Machine Learning and Deep Learning for Dropout Prediction in Higher Education: A Review. (2026). *Computers, 15*(3), 164.

[8] Nti, I. K., & Ramanayake, S. (2026). Explainable machine learning for student dropout prediction and tailored interventions in online personalized education. *Discover Artificial Intelligence, 6*, 288.

[9] Romdhoni, R. D., Majid, N. W. A., & Permatasari, A. F. (2026). Machine Learning-Based Early Warning for Student Dropout: Evidence from LMS Behavioral Engagement Patterns in Online Higher Education. *Journal of Informatics and Vocational Education, 9*(2), 55–66. https://doi.org/10.20961/joive.v9i2.3508.

[10] Ovtsarenko, O. (2026). Using AI to forecast student dropout risk in technical education using a learning analytics approach. *Scientific Reports, 16*, 14616.

[11] A Real-Time Early Warning Intervention System for At-Risk Students Using Predictive Modeling and Academic Engagement Metrics. (2026). *Procedia Computer Science, 283*, 2486–2495. https://doi.org/10.1016/j.procs.2026.06.318.