IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

EduPulse AI: AI-Based Early Detection of Student Academic Stress Using LMS and Attendance Data

Gurav Aparna Sunil[1], Mr. Manoj V. Nikum[2], Prof. K. C. Bhave[3]

Research Scholar, MCA, Shri Jaykumar Rawal Institute of Technology, Dondaicha, India[1]

HOD MCA & Assistant Professor, Shri Jaykumar Rawal Institute of Technology, Dondaicha, India[2]

Assistant Professor, Shri Jaykumar Rawal Institute of Technology, Dondaicha, India[3]

Abstract: EduPulse AI is a role-based web application prototype that combines attendance, learning-managementsystem (LMS) engagement, assignment, quiz and academic-performance indicators to support early review of students who may need academic assistance. The system presents three indicator categories—Low, Moderate and High—and provides separate interfaces for administrators, faculty members and students. A synthetic dataset of 2,000 records was used to compare Random Forest, Extra Trees and Gradient Boosting classifiers through stratified five-fold crossvalidation. Extra Trees achieved the highest mean cross-validation accuracy of 0.859 (standard deviation 0.027) and 0.880 accuracy on a held-out set of 400 records. The implementation uses Python, Flask, SQLite, SQLAlchemy and Scikit-learn. Since the target labels were generated using illustrative rules on synthetic data, these results demonstrate the technical workflow only; they do not establish real-world or clinical prediction performance.

Keywords: Learning Analytics, Academic Early Warning, Student Support, Machine Learning, Extra Trees, Attendance, LMS Engagement.

I. INTRODUCTION

Educational institutions generate student information through attendance systems, learning-management platforms, assignments, quizzes and academic assessments. These records can provide useful indications of participation and academic progress. When the information is maintained in separate systems or reviewed manually, however, it can be difficult for educators to identify changes early and coordinate appropriate support.

EduPulse AI was developed as a web-based academic early-warning prototype to consolidate selected academic indicators and present them through a role-based interface. The system classifies records into Low, Moderate and High indicator categories and allows authorised users to review student information, inspect progress and record supportive actions with follow-up dates. The project uses the term academic stress in an early-warning context; it is not a psychological or clinical diagnostic system.

The main objective is to demonstrate an integrated workflow that connects academic data preparation, machinelearning classification and human-centred student-support functions. The current experiment is based on synthetic data and should be interpreted as a technical prototype rather than evidence of validated performance on real students.

II. LITERATURE REVIEW

Learning analytics concerns the measurement, collection, analysis and reporting of data about learners and their contexts to understand and improve learning. Long and Siemens [1] discuss the potential of analytics to support evidence-informed decisions in education. This provides a conceptual basis for presenting attendance, LMS engagement and assessment indicators together rather than considering a single measure in isolation.

Educational data mining applies data-mining and machine-learning techniques to educational data. Romero and Ventura [2] survey the field and describe data-driven approaches for understanding educational processes. In an earlywarning setting, such approaches can help identify patterns that merit further review. However, model outputs depend on the quality and representativeness of the input data and labels.

Prior work on student dropout prediction and educational early-warning systems illustrates the use of predictive modelling to identify students who may require support [3]. Such outputs should be interpreted with care and reviewed by educators.

 This work is licensed under a Creative Commons Attribution 4.0 International License

EduPulse AI adopts a support-first approach: its categories are intended to prompt human review, not to make definitive judgements about students.

III. METHODOLOGY

A. System Architecture and Application Workflow

The application uses Python 3 and Flask for request handling and backend workflows. HTML5, CSS3, JavaScript and Bootstrap 5 provide the user interface, while Chart.js is used for visualisations. SQLite with SQLAlchemy ORM stores application records. Pandas and NumPy support tabular data preparation, and Scikit-learn is used for model comparison, training and prediction. Joblib is used to persist the selected trained model.

Three roles are supported: Administrator, Faculty and Student. The administrator manages student records and faculty accounts; faculty members review permitted academic information and record support actions; students access their own permitted information and progress views. The prototype includes an overview dashboard, student management, student support centre, faculty accounts, student creation and CSV-based academic-data import.

B. Dataset and Preprocessing

A synthetic dataset containing 2,000 records was generated for model development. The records were assigned to three target categories: Moderate (1,151; 57.55%), Low (649; 32.45%) and High (200; 10.00%). The dataset was divided using a stratified train-test split, with 1,600 records for model development and 400 records held out for final evaluation. The target labels were produced using illustrative rules; they are not independently assessed studentreported stress labels.

C. Model Training and Evaluation

Random Forest, Extra Trees and Gradient Boosting classifiers were compared using stratified five-fold crossvalidation on the training portion. The classifier with the highest mean cross-validation accuracy was selected. The selected model was then evaluated on the held-out test set, and a classification report and confusion matrix were generated. The model and evaluation metrics were saved for use by the web application.

IV. RESULT AND DISCUSSION

The cross-validation results are summarised in Table I. Extra Trees achieved the highest mean accuracy (0.859), followed by Gradient Boosting (0.853) and Random Forest (0.832). The selected Extra Trees model achieved 0.880 accuracy on the 400-record held-out test set.

TABLE I. CROSS-VALIDATION MODEL COMPARISON

Model	Mean 5-fold CV Accuracy	Standard Deviation
Random Forest	0.832	0.010
Extra Trees	0.859	0.027
Gradient Boosting	0.853	0.011

The class-wise results for the selected model are shown in Table II. The model obtained an F1-score of 0.90 for Moderate and 0.88 for Low. For High, precision was 0.90, recall was 0.68 and F1-score was 0.77. The lower recall for the High category indicates that some High records were predicted as Moderate, so overall accuracy alone is not sufficient to assess the system.

TABLE II. CLASSIFICATION REPORT FOR EXTRA TREES

Class	Precision	Recall	F1-score	Support
Low	0.88	0.88	0.88	130
Moderate	0.88	0.92	0.90	230
High	0.90	0.68	0.77	40
Accuracy	—	—	0.88	400
Macro avg	0.89	0.82	0.85	400
Weighted avg	0.88	0.88	0.88	400

The confusion matrix, with rows representing actual classes and columns representing predicted classes in the order Low, Moderate and High, is presented in Table III. It shows 114 correctly predicted Low records, 211 correctly predicted Moderate records and 27 correctly predicted High records. Thirteen actual High records were classified as Moderate. This pattern reinforces the need to examine class-specific recall and to improve the model before any operational use.

TABLE III. CONFUSION MATRIX

Actual / Predicted	Low	Moderate	High
Low	114	16	0
Moderate	16	211	3
High	0	13	27

These results are limited by the use of synthetic records and illustrative target labels. They do not establish realworld academic-stress prediction performance, fairness across student groups or clinical validity. Before deployment, the model would require evaluation on representative, consented institutional data with independently defined outcomes, privacy safeguards, human oversight and monitoring of errors across categories.

V. CONCLUSION

EduPulse AI demonstrates a prototype workflow for consolidating attendance, LMS engagement and academic indicators in a role-based web application. The project combines synthetic data generation, preprocessing, comparison of three machine-learning classifiers, evaluation of the selected Extra Trees model and student-support workflows for administrators, faculty and students. The selected model achieved 85.9% mean five-fold cross-validation accuracy and 88.0% held-out accuracy on synthetic test data.

The results show that the implemented pipeline can reproduce patterns represented by the generated labels; they do not demonstrate reliable detection of academic stress in a real student population. Future work should use validated and consented institutional data, improve recall for the High category, assess fairness and model drift, add explainability and strengthen audit, privacy and data-retention controls. Predictions should be used only to guide timely supportive review and never to stigmatise or penalise students.

ACKNOWLEDGMENT

The author expresses sincere gratitude to the project guide, Prof. K. C. Bhave, the Head of the Department, Prof. Manoj V. Nikum, and the faculty members of Shri Jaykumar Rawal Institute of Technology, Dondaicha, for their guidance and support during the project.

REFERENCES

[1] P. Long and G. Siemens, “Penetrating the Fog: Analytics in Learning and Education,” EDUCAUSE Review, vol. 46, no. 5, 2011.

[2] C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” WIREs Data Mining and Knowledge Discovery, vol. 10, e1355, 2020.

[3] “Predicting student dropouts with machine learning: An empirical study in Finnish higher education,” Technology in Society, vol. 76, 102474, 2024.

[4] Scikit-learn Documentation. Available: https://scikit-learn.org/stable/.[5] Flask Documentation. Available: https://flask.palletsprojects.com/.