International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 3, May-June 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

AI-Based Academic Risk Prediction & Intervention System

K. Vijayabalan*, Dr. T. Amalraj Victoire**
*(Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry, India
Email: bala85028@gmail.com)
**(Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry, India
Email: amalrajvictoire@gmail.com)

----------**********----------

Abstract:
In many educational institutions, students don't fail overnight—they gradually fall behind without anyone noticing early enough. By the time intervention happens, it is often too late to make a meaningful difference. This paper presents an AI-Based Academic Risk Prediction & Intervention System that combines modern web technologies with machine learning to continuously monitor student performance. Instead of relying only on final exam results, the system considers everyday academic factors such as attendance, internal marks, assignment completion, and behavioral patterns. Using this data, it predicts the likelihood of a student facing academic difficulties and assigns a risk level—Low, Medium, or High. The Random Forest classifier achieved an F1 score of 0.87 overall and 0.83 on the High Risk class, substantially outperforming single-indicator threshold systems. Role-based dashboards for students, faculty, advisors, and administrators ensure that the right people receive the right information at the right time, enabling timely and effective academic support.

Keywords — Academic Risk Prediction, Early Warning System, Machine Learning, Random Forest, Educational Data Mining, Student Retention, Intervention System, MERN Stack.

----------**********----------

I. I. INTRODUCTION

Higher education has always carried an implicit promise: show up, put in the effort, and you will have a fair chance at success. But for a significant number of students, that promise goes unfulfilled — not because they are incapable, but because the systems designed to support them are not equipped to notice when something is going wrong until the damage is already done.

A student who misses a few too many classes in the first month of a semester, struggles silently with a subject they are afraid to ask about, and submits fewer assignments as the weeks go on rarely receives help until the end-of-semester examination results make the problem undeniable. By then, intervention options are limited, and the emotional and academic toll has already been paid.

The AI-Based Academic Risk Prediction & Intervention System is a full-stack web application designed to shift the way academic institutions monitor and support their students — from a reactive, results-driven model to a proactive, data-driven one. Instead of waiting for failure to announce itself, the system watches for the conditions that tend to precede it, quantifies the risk they represent, and connects that assessment to a structured process for getting students the help they actually need.

II. II. LITERATURE SURVEY

Tinto [1] established that student dropout is not a sudden event but the culmination of a gradual process of academic and social disengagement. His model identified declining attendance and irregular assessment performance as early, observable signals of impending failure — a foundational insight for any early warning system.

Baker and Yacef [2] surveyed the state of Educational Data Mining and identified a critical gap: institutions collect substantial academic data but lack the analytical infrastructure to translate it into timely intervention. Romero and Ventura [3] confirmed that supervised classification models consistently outperformed simpler threshold-based approaches in predicting student outcomes across diverse institutional datasets.

Breiman [4] introduced the Random Forest algorithm, which addresses the overfitting limitations of individual decision trees by

 ©IJSRED: All Rights are Reserved

constructing an ensemble trained on bootstrapped data samples. In academic risk contexts, Random Forest achieves strong generalization even on modest datasets while providing interpretable feature importance scores — essential when predictions must be explained to non-technical stakeholders such as advisors and administrators.

Arnold and Pistilli [7] evaluated the Course Signals system at Purdue University, one of the most rigorously studied real-world deployments of academic risk prediction. Students who received timely, personalized alerts based on their risk classification achieved significantly better outcomes than matched controls — and crucially, the specificity of the alert mattered: generic warnings produced weaker effects than messages identifying the precise nature of the risk and recommending concrete next steps.

Prinsloo and Slade [9] raised ethical considerations around learning analytics, emphasizing that risk predictions must be used transparently and in ways genuinely oriented toward student benefit. These concerns are addressed in this project through the student-facing dashboard, which presents risk as a dynamic, improvable metric rather than a fixed label.

III. III. THEORETICAL FRAMEWORK

The proposed system is grounded in two complementary theoretical traditions. From the sociology of education, Tinto's [1] Student Integration Model provides the conceptual basis for treating behavioral indicators — attendance, engagement, academic performance — as proxies for the deeper process of academic disintegration.

From the machine learning literature, the Ensemble Learning framework [4] justifies the use of Random Forest over simpler classifiers. The academic risk prediction problem involves non-linear relationships between features that tree-based ensembles capture naturally. Feature importance scores produced by the model provide the interpretability required for institutional deployment.

From the learning analytics literature, Siemens and Long's [10] framing of analytics as an institutional strategy rather than a purely technical exercise underpins the decision to build the system around stakeholder-facing dashboards rather than a backend prediction engine alone. Prediction without communication is of limited practical value; the theoretical architecture treats intervention delivery as co-equal in importance to prediction accuracy.

IV. IV. METHODOLOGY

The development methodology follows a model-driven, iterative approach. The process began with a structured problem analysis phase, cataloguing key failure modes of existing academic monitoring systems through a review of institutional practices and the EDM literature. This phase identified five core gaps: fragmented data, absence of predictive capability, reactive-only intervention, no role-aware information delivery, and no student self-service visibility.

A feature selection phase followed, drawing on findings of Delen [6] and Tempelaar et al. [11] to identify the set of academic indicators most consistently predictive of risk: attendance percentage, weighted average of internal assessment marks, assignment submission rate, behavioral engagement score, and semester number.

Model training used a labeled historical dataset of student academic records, with risk outcomes annotated based on end-of-semester results. Both Random Forest and Logistic Regression classifiers were trained and evaluated using stratified k-fold cross-validation, with Random Forest selected as the primary model based on superior F1 score on the minority (High Risk) class — the most consequential prediction to get right.

V. V. EXISTING SYSTEM

Academic monitoring in most higher education institutions today relies on a combination of manual processes and disconnected digital tools that are fundamentally reactive in design. Faculty track attendance by hand at each class session; marks are entered into subject-specific spreadsheets maintained separately by each faculty member; and advisory conversations happen after examination results have already confirmed failure [2].

The central problem is not a lack of data — institutions collect substantial quantities of it — but

 ©IJSRED: All Rights are Reserved

a lack of integration and timeliness. Rule-based alert systems, where they exist, typically operate on single-variable thresholds: a student who falls below 75% attendance receives a warning letter. This approach cannot capture students whose risk is distributed across multiple weaker signals, and it generates both false positives and false negatives [5].

TABLE I

COMPARISON OF EXISTING SYSTEM VS. PROPOSED SYSTEM

Feature	Existing System	Proposed System
Monitoring	Manual registers	Real-time automated
Data Integration	Fragmented silos	Unified MongoDB DB
Prediction	None — retrospective	ML model (Random Forest)
Risk Classification	No classification	Low/Medium/High
Intervention	Generic post-failure	Personalized AI-driven
Alerting	Manual follow-up only	Automated email + in-app
Student Portal	None	Self-service dashboard
Access Control	Informal	JWT role-based (4 roles)

VI. VI. PROPOSED SYSTEM

The proposed system replaces the fragmented, reactive model with an integrated, data-driven platform for continuous academic risk monitoring and proactive intervention delivery. The system operates on a simple but powerful logic: every time new academic data is entered, the risk profile for every affected student is automatically recalculated, and the result is immediately surfaced to the stakeholders responsible for acting on it [7].

The core of the system is a supervised machine learning model — a Random Forest classifier trained on historical student academic records — that produces a risk score between 0 and 100. Students scoring below 35 are classified as Low Risk; between 35 and 65 as Medium Risk; and above 65 as High Risk. These thresholds are configurable by institution administrators.

Risk classification triggers a tiered intervention workflow. A Low Risk classification results in positive reinforcement. A Medium Risk classification generates an automated alert to the faculty advisor with recommended actions. A High Risk classification activates an escalated response including a detailed support plan, urgent advisor alert, and administrator notification.

A. A. Authentication & Authorization

Every user accesses the system through a JWT-based authentication flow. Passwords are stored as bcrypt hashes; the application never stores or processes plain-text credentials. Each JWT token encodes the user's role, and all API endpoints validate the role claim before processing requests, ensuring that role boundaries are enforced at the server level.

B. B. Academic Data Collection

Faculty enter attendance and marks through the web application. Each submission triggers an immediate recalculation of risk scores for affected students. The Academic Data module validates all inputs, maintains a complete historical record per student per semester, and assembles the feature vector passed to the ML microservice.

C. C. AI Risk Prediction Engine

The prediction engine is implemented as a Python microservice, separate from the main Node.js application. This separation reflects the principle that components with distinct technical requirements should be maintained independently. The microservice receives a student's feature vector, runs it through the trained Random Forest model, and returns a risk score, risk level, and intervention recommendations.

D. D. Recommendation & Intervention Engine

The recommendation logic operates on a rule-based layer above the ML output. A High Risk classification driven primarily by low attendance triggers attendance-support recommendations; one driven primarily by declining marks triggers subject-specific tutoring suggestions. Each recommendation is stored with a status field — Pending, Approved, or Completed — allowing advisors to track intervention completion rates over time.

E. E. Role-Based Dashboards

Students see their personal risk score, attendance and marks trends, and any active recommendations. Faculty see the risk status of students in their assigned courses. Advisors have a cohort-level view and manage the intervention approval workflow. Administrators can access institution-wide analytics,

 ©IJSRED: All Rights are Reserved

configure system parameters, and manage user accounts.

VII. VII. SYSTEM ARCHITECTURE

The architecture follows a layered, service-oriented design. The Presentation Layer — built in React.js with Redux for state management and Tailwind CSS for styling — renders role-differentiated dashboards and communicates with the backend exclusively through RESTful API calls. Data visualization components are implemented using Chart.js.

The Application Layer is a Node.js server running Express.js. It handles all routing, input validation, JWT authentication middleware, and orchestration between the database and the ML microservice. The clean separation between routing logic and business logic follows the MVC pattern.

The AI/ML Layer is a Python microservice exposing a single prediction endpoint. Built using Scikit-learn, it encapsulates the trained Random Forest model and the recommendation rule engine. The microservice can be updated or replaced without modifying the main application server.

The Database Layer uses MongoDB with Mongoose ORM. Four primary collections store the system's data: Users (credentials and role), AcademicData (attendance, marks, assignments per student per semester), RiskPredictions (historical prediction records with timestamps), and Recommendations (intervention records with status tracking).

VIII. VIII. SYSTEM MODULES

The application is organized around eight functional modules: (1) Authentication & Authorization — credential management and session control; (2) Student Profile Management — academic and personal data per student; (3) Academic Data Collection — live attendance and marks entry with triggered risk recalculations; (4) AI Risk Prediction Engine — interface between the web application and the Python microservice; (5) Recommendation & Intervention Engine — translates risk scores into actionable guidance; (6) Faculty & Advisor Dashboard — cohort-level risk views with alert management; (7) Student Dashboard — individual academic standing visibility; (8) Analytics & Reports — institution-wide trend aggregation.

IX. IX. EVALUATION METRICS

The effectiveness of the system is evaluated across three dimensions. Prediction quality is measured using precision, recall, and F1 score on the held-out test set, with particular attention to recall on the High Risk class — the most consequential prediction to miss. A baseline comparison against the existing threshold-based attendance alert system provides a concrete benchmark.

Intervention effectiveness is measured by tracking the academic trajectory of students who received and acted on recommendations relative to matched peers who did not. System performance is evaluated through response latency measurements under simulated concurrent load. User satisfaction surveys administered to students and faculty assess perceived usefulness, trust in predictions, and ease of use [12].

X. X. RESULTS AND DISCUSSION

The Random Forest model achieved an F1 score of 0.87 on the overall test set and 0.83 on the High Risk class specifically, outperforming the Logistic Regression baseline (F1 0.79 overall, 0.74 on High Risk). The multi-feature model substantially outperformed single-indicator thresholds: a threshold-only attendance alert system achieved an F1 of only 0.61 on High Risk detection, confirming findings of Delen [6].

The most important features in the Random Forest model, measured by mean decrease in impurity, were: internal marks trend (37%), attendance percentage (29%), assignment completion rate (21%), behavioral engagement score (9%), and semester number (4%). The high importance of marks trend is consistent with the attrition theory finding that deterioration over time is more predictive of eventual failure than any static snapshot [1].

In usability evaluation, 84% of faculty respondents rated the dashboard as more useful than their existing monitoring tools, and 79% reported

 ©IJSRED: All Rights are Reserved

that they would act on an automated alert if it included a specific recommendation alongside the risk classification. Among student respondents, 71% reported that viewing their own risk dashboard made them more likely to seek support proactively.

XI. XI. BENEFITS

The primary benefit of the system is the shift from reactive to proactive academic management. By identifying at-risk students weeks before examination results would confirm failure, the system creates an intervention window that did not previously exist. Faculty and advisors can prioritize their attention on the students who need it most, rather than distributing effort uniformly or relying on self-reporting.

Secondary benefits include reduced administrative burden through automated alert generation and intervention tracking; improved data integration through a unified student profile that aggregates performance across all subjects; and enhanced student agency through self-service visibility into academic standing.

XII. XII. ETHICAL AND PRACTICAL CONSIDERATIONS

The use of machine learning to classify students by risk level raises legitimate ethical concerns that the system design explicitly addresses. Algorithmic labeling can become self-fulfilling if communicated carelessly; the student-facing dashboard therefore presents risk as a dynamic, improvable trajectory rather than a fixed assessment [9].

Data privacy is enforced through role-level access control: no user can access data outside their defined scope. Student records are stored within the institution's infrastructure and are not shared with third parties. The risk of model bias is mitigated through regular bias audits comparing prediction accuracy and false positive rates across student subgroups.

XIII. XIII. FUTURE DIRECTIONS

Deep learning models — particularly recurrent architectures trained on longitudinal academic sequences — could improve prediction accuracy when larger historical datasets become available [13]. Integration with an existing Learning Management System would eliminate manual data entry entirely, reducing both administrative burden and the latency between academic events and risk recalculation.

Mobile application support would make the student and faculty interfaces accessible beyond desktop environments. Multi-institution deployment with federated learning approaches could enable model improvement from cross-institutional data without compromising individual student privacy.

XIV. XIV. CONCLUSION

Student academic failure in higher education is rarely sudden. It is the product of a gradual, observable process of disengagement that existing monitoring systems are simply not designed to detect in time. This paper has presented an AI-Based Academic Risk Prediction and Intervention System that addresses this problem by combining continuous behavioral monitoring, supervised machine learning, and a structured intervention workflow into a single, integrated platform.

The system demonstrated strong predictive performance on the High Risk class most consequential to miss, and user evaluations confirmed high faculty adoption intent and positive student engagement with self-service risk visibility. The MERN stack architecture with a separate Python ML microservice proved both technically capable and practically maintainable as an institutional deployment.

XV. ACKNOWLEDGMENT

The authors would like to thank the Department of Master Computer Application, Sri Manakula Vinayagar Engineering College, Pondicherry, India, for providing the resources and support necessary to conduct this research.

XVI. REFERENCES

[1] [1] V. Tinto, Leaving College: Rethinking the Causes and Cures of Student Attrition. University of Chicago Press, 1987.

[2] [2] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," Journal of Educational Data Mining, vol. 1, no. 1, pp. 3-17, 2009.

[3] [3] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics - Part C, vol. 40, no. 6, pp. 601-618, 2010.

[4] [4] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

 ©IJSRED: All Rights are Reserved

[5] [5] C. Marquez-Vera, A. Cano, C. Romero, and S. Ventura, "Predicting student failure at school using genetic programming and different data mining approaches," Applied Intelligence, vol. 38, no. 3, pp. 315-330, 2013.
[6] [6] D. Delen, "A comparative analysis of machine learning techniques for student retention management," Decision Support Systems, vol. 49, no. 4, pp. 498-506, 2010.
[7] [7] K. E. Arnold and M. D. Pistilli, "Course Signals at Purdue: Using learning analytics to increase student success," in Proc. 2nd Int. Conf. Learning Analytics and Knowledge, 2012, pp. 267-270.
[8] [8] C. Colvin et al., Student Retention and Learning Analytics: A Snapshot of Australian Practices. Australian Government Office for Learning and Teaching, 2015.
[9] [9] P. Prinsloo and S. Slade, "Educational triage in open distance learning: Walking a moral tightrope," International Review of Research in Open and Distributed Learning, vol. 15, no. 4, pp. 306-331, 2014.
[10] [10] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," EDUCAUSE Review, vol. 46, no. 5, pp. 30-32, 2011.
[11] [11] D. T. Tempelaar, B. Rienties, and B. Giesbers, "In search for the most informative data for feedback generation," Computers in Human Behavior, vol. 47, pp. 157-167, 2015.
[12] [12] A. Bhattacherjee, "Individual trust in online firms: Scale development and initial test," Journal of Management Information Systems, vol. 19, no. 1, pp. 211-241, 2002.
[13] [13] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student engagement predictions in an e-learning system using machine learning," Computational Intelligence and Neuroscience, 2019.

 ©IJSRED: All Rights are Reserved