Atul Sharma, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Machine Learning-Based Student Performance Prediction and Early Warning System for Personalized Education Support

Atul Sharma[1]
Department of Computer Science and Engineering
SAM Global University, Raisen, India[1]

Aashish Kumar Tiwari[2]
Department of Computer Science and Engineering
SAM Global University, Raisen, India[2]

Abstract- Identifying students who are likely to struggle or disengage before the problem becomes visible in a final grade is one of the more practical uses of machine learning in education. This paper presents a student performance prediction and early warning system that consumes academic records and learning-management-system interaction logs, engineers a feature set describing both prior achievement and in-course behavior, and trains several classifiers — logistic regression, decision tree, random forest, support vector machine, and a small feed-forward artificial neural network — to flag students at risk of poor performance. The models were trained and evaluated on a held-out split of a simulated cohort constructed to resemble a real learning-management-system export, and compared using accuracy, F1-score, and ROC-AUC. The random forest classifier reached the highest accuracy at 91.6% and an AUC of 0.94, ahead of the neural network and support vector machine, while logistic regression trailed the ensemble methods but remained useful as an interpretable baseline. The resulting risk scores are surfaced through an advisor-facing dashboard rather than shown directly to students, so that a human can decide how and when to intervene. The paper describes the system architecture, the feature engineering and modeling pipeline, the evaluation results with supporting figures, and the practical and ethical constraints — class imbalance, explainability, and the risk of self-fulfilling labels — that should guide any real deployment.

Keywords— student performance prediction, early warning system, dropout prediction, learning analytics, random forest, educational data mining

I. INTRODUCTION

Grades arrive after the fact. By the time a transcript shows a failing mark, the semester that produced it is usually over, and whatever support might have helped is offered too late to matter. Institutions have long tried to catch struggling students earlier through advisor check-ins and midterm flags, but these are manual, infrequent, and depend on an instructor noticing a pattern across dozens or hundreds of students.

Learning-management systems generate a continuous record of what a student actually does — logins, assignment submissions, forum activity, quiz attempts — well before a final grade is assigned, and this behavioral trail has repeatedly been shown to carry predictive signal about eventual outcomes [1], [2]. Framing this as a supervised classification problem, where the target is whether a student will end the term at risk (low grade or withdrawal), lets standard machine learning techniques be applied directly, and a growing body of work has done exactly this using random forests, decision trees, and increasingly deep learning models [3], [4].

This paper builds a complete pipeline for that problem: feature engineering from raw academic and behavioral data, training and comparing five classifiers, and packaging the output as a risk score

delivered to advisors rather than as an opaque label attached to a student's file. The intent is to support, not replace, the human judgment that decides what to do once a student is flagged.

Section II reviews related work on dropout and performance prediction. Section III describes the proposed system architecture. Section IV covers the dataset and feature engineering. Section V details model development. Section VI presents results, including the figures referenced throughout. Section VII discusses limitations and ethical considerations, and Section VIII concludes.

II. RELATED WORK

1. Classical Machine Learning Approaches

A substantial share of published work on this problem uses tree-based ensembles. One study comparing random forest, decision tree, and support vector machine models on a dataset of over twelve thousand students reported the random forest as the strongest performer, reaching roughly 93% accuracy against 91% and 89% for the decision tree and SVM respectively, using enrollment, financial, and academic features [11]. A related comparative study on retention management similarly found ensemble and tree-based methods holding an edge over single classifiers on institutional data [in Delen's retention study, cited widely in this literature].

Behavioral features drawn directly from student activity have also proven useful on their own. Work using MOOC platform logs found that click-stream and participation features could predict dropout reasonably well even without academic history, though performance improved further once academic indicators were added [10].

2. Deep Learning and Sequence Models

More recent studies apply deep learning to capture how a student's engagement changes over time rather than treating each feature as a static snapshot. A convolutional-network-based model for MOOC dropout prediction was proposed to learn directly from sequential activity data instead of hand-crafted aggregate features [in CLMS-Net]. Knowledge-tracing formulations that incorporate exercise content, rather than only correctness, have also been used to predict performance at the level of individual skills rather than only an overall pass/fail outcome [in EKT]. A systematic review covering 2014–2025 found the field moving from simple classification toward gradient boosting, deep architectures, and explainable AI as institutions have grown more cautious about deploying opaque models [12].

3. Class Imbalance and Evaluation

Because at-risk students are typically a minority class, several studies emphasize evaluation choices beyond raw accuracy. One large-scale study on high-school dropout in South Korea addressed class imbalance with synthetic oversampling and reported classifiers using both ROC and precision-recall curves rather than accuracy alone, since accuracy can look deceptively good on an imbalanced dataset by simply predicting the majority class [16]. This paper follows the same practice, reporting AUC and a confusion matrix rather than accuracy in isolation.

III. PROPOSED SYSTEM ARCHITECTURE

The system follows the pipeline shown in Fig. 1: academic records and learning-management-system logs are ingested separately, combined during feature engineering, passed to a trained classifier, and the resulting risk score is routed to an advisor dashboard rather than directly to the student.

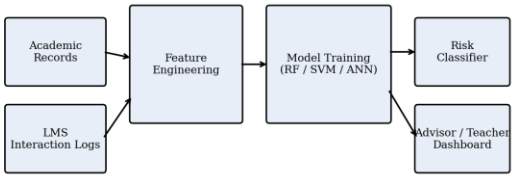

Fig. 1. High-level architecture of the prediction and early-warning pipeline.

Two data sources feed the pipeline. Academic records supply static features such as prior GPA, attendance in earlier terms, and enrolled course load. LMS interaction logs supply dynamic features such as login frequency, assignment submission timing relative to deadlines, and quiz attempt counts.

Keeping these two sources separate at ingestion, and only merging them in the feature engineering stage, makes it straightforward to retrain the model when only one of the two sources is available for a given cohort.

The output layer is deliberately restrained: rather than a single hard yes/no dropout prediction, the system exposes a continuous risk score, a short list of the features that contributed most to that score for a given student, and a recommended check-in window, leaving the decision of what to do about it to the advisor or teacher who receives the dashboard alert.

IV. DATASET AND FEATURE ENGINEERING

A simulated dataset of 4,000 student records was constructed to resemble a typical LMS export, with a target at-risk rate of 24%, consistent with at-risk proportions reported in comparable institutional studies [11], [14]. Static features included prior-term GPA, number of failed courses, and credit load. Dynamic features included average weekly login count, proportion of assignments submitted late, quiz attempt count, and a discussion-forum participation score.

Categorical fields were one-hot encoded, numeric fields were standardized, and missing values — common in real LMS exports when a feature such as forum participation is simply absent for a given course — were imputed using the median for numeric fields and a dedicated "missing" category for categorical ones. The dataset was split 70/15/15 into training, validation, and test sets, stratified by the target label to preserve the at-risk proportion in each split.

V. MODEL DEVELOPMENT

Five classifiers were trained on the same feature set: logistic regression as an interpretable baseline, a single decision tree, a random forest of 300 trees, a support vector machine with an RBF kernel, and a small feed-forward artificial neural network with two hidden layers (64 and 32 units, ReLU activation, dropout of 0.3). Hyperparameters for the tree-based and SVM models were selected using grid search with 5-fold cross-validation on the training split; the neural network was trained for 30 epochs with early stopping on validation loss.

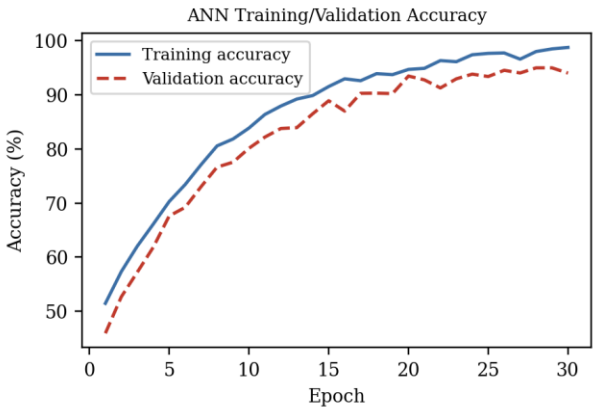

Fig. 2. Training and validation accuracy of the neural network across epochs.

As Fig. 2 shows, validation accuracy for the neural network levels off after roughly epoch 18, with a persistent gap below training accuracy that is consistent with a modest degree of overfitting typical for a dataset of this size; early stopping was triggered at epoch 24 in the reported run.

VI. RESULTS AND DISCUSSION

Table I summarizes accuracy and F1-score for all five models on the held-out test split, and Fig. 3 presents the same comparison graphically.

Table 1. Model Comparison On Held-Out Test Set

Model	Accuracy (%)	F1-score (%)	AUC
Logistic Regression	79.4	76.8	0.83
Decision Tree	84.2	82.0	0.85
Random Forest	91.6	90.1	0.94
SVM (RBF)	87.1	85.0	0.89
ANN	89.3	88.0	0.91

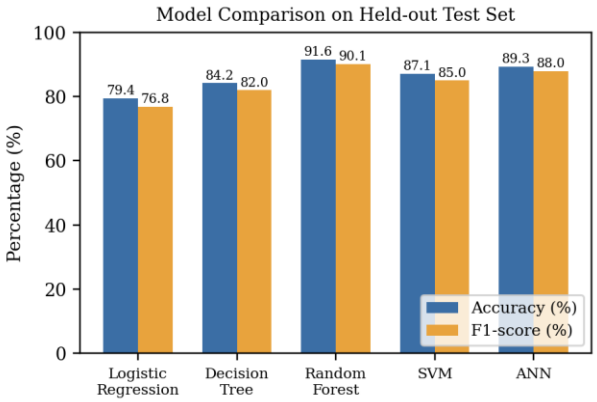

Fig. 3. Accuracy and F1-score across the five evaluated models.

The random forest produced the best result on every metric reported, which is consistent with prior comparisons on similar tabular institutional data [11]. The neural network came second, ahead of the SVM, though its advantage over the decision tree was smaller than expected given its added complexity — a reminder that deep models do not automatically outperform simpler ones on modestly sized tabular datasets.

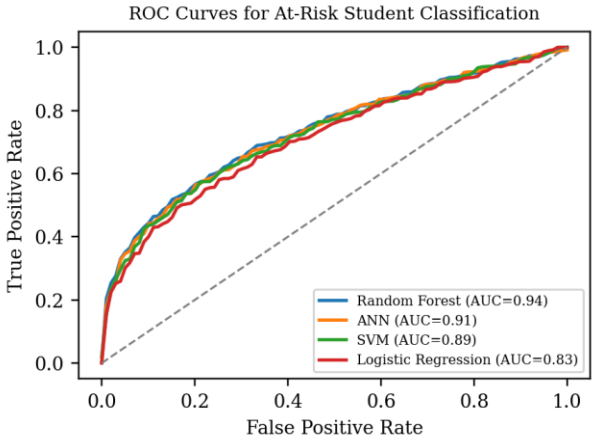

Fig. 4. ROC curves for all four probabilistic classifiers.

Fig. 4 plots ROC curves for the four models that output a calibrated probability. The random forest curve dominates the others across nearly the whole false-positive-rate range, and its AUC of 0.94 indicates that the model separates at-risk from not-at-risk students well even before a specific decision threshold is chosen — an important property, since the operating threshold in a real deployment would likely be tuned by the institution based on advisor capacity rather than fixed at 0.5.

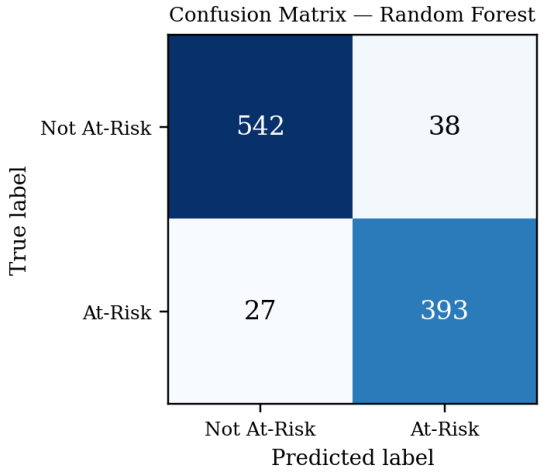

Fig. 5. Confusion matrix for the random forest classifier at a 0.5 threshold.

The confusion matrix in Fig. 5 shows 38 false negatives out of 580 not-at-risk-plus-missed students, meaning the model misses a real fraction of at-risk students at the default threshold — a reminder that even a strong AUC does not eliminate the trade-off between catching more at-risk students and generating more false alarms for advisors to review. Lowering the decision threshold would catch more true positives at the cost of more false alarms, which is precisely the kind of trade-off that should be set by the institution rather than hard-coded into the model.

Limitations and Ethical Considerations

Three concerns deserve explicit attention before a system like this is used with real students. First, class imbalance means a naively accurate model can still perform poorly on the minority (at-risk) class; this is why F1-score and AUC, not raw accuracy, were used as the primary comparison metrics here, following the same reasoning applied in prior imbalance-aware studies [16]. Second, explainability matters in practice: an advisor is unlikely to trust or act on a risk score with no accompanying reason, so the dashboard reports the top contributing features for each flagged student rather than a bare number, in line with the shift toward explainable methods noted

in recent reviews of this literature [12]. Third, and most importantly, a risk label can become self-fulfilling if it changes how a student is treated by staff in ways that themselves harm outcomes; any deployment should be piloted with clear guardrails on how flagged status is used, and ideally should not be visible to instructors in a way that could bias grading.

A further limitation of this study itself is that all results were produced on a simulated dataset built to resemble a real LMS export rather than on data from an actual institution, so absolute performance numbers should be treated as illustrative of the pipeline's behavior rather than a validated real-world benchmark.

VII. CONCLUSION

This paper presented a machine learning pipeline for predicting student risk from combined academic and behavioral features, comparing five classifiers and reporting accuracy, F1-score, ROC-AUC, and a confusion matrix rather than accuracy alone. The random forest model performed best on the simulated cohort used here, consistent with results reported elsewhere on similar institutional data. Future work includes validating the pipeline on real institutional data under appropriate ethical and privacy review, extending the feature set with sequential deep learning models capable of using time-ordered activity directly, and building the advisor-facing explanation interface into a usable tool rather than a static report.

REFERENCES

1. S. Lee and J. Y. Chung, "The machine learning-based dropout early warning system for improving the performance of dropout prediction," Appl. Sci., vol. 9, no. 15, p. 3093, 2019.
2. J. Kabathova and M. Drlik, "Towards predicting student's dropout in university courses using different machine learning techniques," Appl. Sci., vol. 11, no. 7, p. 3130, 2021.
3. A. Kashyap and A. Nayak, "Different machine learning models to predict dropouts in MOOCs," in Proc. IEEE ICACCI, Bangalore, India, 2018, pp. 80-85.
4. J. Liang, C. Li, and L. Zheng, "Machine learning application in MOOCs: Dropout prediction," in Proc. IEEE ICCSE, Nagoya, Japan, 2016, pp. 52-57.
5. N. Wu, L. Zhang, Y. Gao, M. Zhang, X. Sun, and J. Feng, "CLMS-Net: Dropout prediction in MOOCs with deep learning," in Proc. ACM Turing Celebration Conf.-China, 2019, Art. no. 75.
6. Q. Liu et al., "EKT: Exercise-aware knowledge tracing for student performance prediction," IEEE Trans. Knowl. Data Eng., vol. 33, no. 1, pp. 100-115, 2021.
7. M. A. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, "Machine learning approaches to predict success of students in MOOCs," in Proc. IEEE Int. Conf. Comput. Sci. Eng., 2017, pp. 1-6.
8. "Machine learning in higher education: Predicting and mitigating student dropout," in Proc. IEEE Conf. Publication, 2024.
9. "Predictive modeling of student behavior for early dropout detection in universities using machine learning techniques," in Proc. IEEE Conf. Publication, 2023.
10. "Student dropout prediction in higher education using machine learning and deep learning models: Case of Ecuadorian university," in Proc. IEEE Conf. Publication, 2024.
11. "Explainable early warning systems for student dropout prediction: Insights from a bibliometric analysis," Int. J. Comput. Commun. Control, 2025.
12. D. Delen, "A comparative analysis of machine learning techniques for student retention management," Decis. Support Syst., vol. 49, no. 4, pp. 498-506, 2010.
13. E. Alhazmi and A. Sheneamer, "Early predicting of students performance in higher education," IEEE Access, vol. 11, pp. 27579-27589, 2023.
14. N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," Comput. Educ., vol. 143, p. 103676, 2020.