

# Explainable AI-Based Student Result Prediction: A Review of Machine Learning Approaches Using CatBoost and Logistic Regression

*Review Paper on Intelligent Education Systems Using Explainable AI and Hybrid Machine Learning*

**Heena M. Pathan<sup>1</sup>, Pournima E. Gawade<sup>2</sup>**

<sup>1</sup>(Computer Science & Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajnagar — heenas0812@gmail.com)

<sup>2</sup>(Computer Science & Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajnagar — pournimagawade@dietms.org)

**Abstract**—Student academic performance prediction has become a critical priority for educational institutions seeking to improve learning outcomes and reduce dropout rates. Traditional evaluation methods are inadequate for handling the large volumes of student data generated by modern digital academic environments. While numerous machine learning models have been proposed for result prediction, a persistent gap exists between predictive accuracy and model interpretability. This paper presents a comprehensive review of machine learning approaches for student result prediction, culminating in the proposal of a hybrid framework combining CatBoost and Logistic

Regression. The reviewed models are assessed for accuracy, interpretability, computational cost, and suitability for real educational deployment. Key limitations — including the absence of explainability in high-accuracy models, poor handling of categorical student data, and class imbalance — are analyzed in depth. The paper concludes by identifying open research problems and proposing directions toward scalable, transparent, and computationally efficient student performance prediction systems.

**Index Terms**—Student Result Prediction, Explainable AI, CatBoost, Logistic Regression, Educational Data Mining, Machine Learning, Hybrid Model, Class Imbalance, SHAP.

## I. INTRODUCTION

Educational institutions accumulate diverse categories of student data including internal assessment marks, attendance records, assignment scores, and examination results. Traditionally, educators analyzed these records manually to evaluate student progress; however, such approaches are time-consuming, error-prone, and unable to identify at-risk students early enough for effective academic intervention. The rapid growth of digital learning platforms has intensified this challenge, making manual analysis practically infeasible at scale.

Artificial intelligence and machine learning offer powerful solutions for automating student performance analysis. Supervised learning models trained on historical academic records can predict outcomes such as pass/fail status or final grade with considerable accuracy. However, a fundamental tension persists: high-accuracy models tend to operate as black boxes, offering no interpretable insight into their decisions. In educational settings this opacity is especially problematic, because teachers and administrators must understand the reasoning behind predictions in order to design appropriate and targeted academic interventions.

Explainable Artificial Intelligence (XAI) addresses this gap by providing methods that make model decisions understandable to human stakeholders. The challenge is to build systems that are simultaneously accurate, interpretable, and deployable within the resource constraints of real educational institutions. This paper reviews machine learning approaches for student result prediction and proposes a hybrid framework that reconciles these competing demands through the combination of CatBoost and Logistic Regression.

## II. BACKGROUND

### A. Educational Data Mining

Educational Data Mining (EDM) is a discipline concerned with developing methods for exploring data originating in

educational environments. The primary goal is to discover patterns that illuminate student learning processes, institutional effectiveness, and factors predictive of academic outcomes. EDM draws on statistics, machine learning, and data visualization to transform raw academic records into actionable institutional knowledge.

### ***B. Machine Learning in Education***

In the education domain supervised learning techniques are most commonly applied: a model is trained on records of students whose outcomes are known, then used to predict outcomes for incoming students. Classical algorithms such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines have been widely adopted, alongside more recent methods including gradient boosting and deep learning architectures.

### ***C. Explainable AI (XAI)***

Explainable AI refers to techniques that render machine learning predictions interpretable to human users. Methods such as SHAP (SHapley Additive exPlanations) quantify the contribution of each feature to a model output. In educational contexts explainability is not merely desirable but essential: institutions must justify data-driven decisions that directly affect student futures, making transparency a fundamental system requirement.

### ***D. CatBoost and Logistic Regression***

CatBoost is a gradient boosting algorithm designed to handle categorical features natively without requiring extensive preprocessing. Its ordered boosting mechanism reduces overfitting and delivers strong classification performance on moderate-sized datasets. Logistic Regression is a classical statistical model that provides transparent, probability-based predictions; each coefficient directly represents the influence of a specific feature on the predicted outcome. The complementary strengths of these two approaches motivate their combination in a hybrid prediction framework.

## **III. LITERATURE REVIEW**

Research on student result prediction spans a wide range of algorithmic approaches. Early work relied primarily on Logistic Regression, which delivers clear probabilistic outputs and transparent feature coefficients. Kumar and Sharma demonstrated accuracy in the range 72–78% for binary pass/fail classification, establishing it as a reliable baseline but insufficient for complex academic environments [1].

Reddy and Patel investigated K-Nearest Neighbours and Support Vector Machine classifiers, reporting accuracy improvements of 75–82% through non-linear decision boundaries [2]. Singh and Verma explored Decision Tree models, which offer high interpretability but are prone to overfitting, achieving 70–76% accuracy [3]. Gupta and Jain demonstrated that Random Forest substantially reduces overfitting through ensemble aggregation, reaching 80–85% accuracy while providing feature importance rankings [4].

Kord et al. examined course-planning recommendation combined with performance prediction through educational data mining and reported promising results for multi-objective academic support systems [9]. Khan and Ali applied SMOTE-based oversampling to address class imbalance, demonstrating improved recall for failing students when combined with standard classifiers [14].

More recent work turned to deep learning. Bidirectional LSTM networks applied to temporal academic data achieve 87–92% accuracy but at the cost of interpretability and substantial computational requirements [5]. AutoML frameworks that search over algorithm and hyperparameter spaces report 84–88% accuracy but share the opacity of other complex models [6].

The LASA framework integrates deep learning with SHAP-based explanations for long-term performance prediction, achieving 88–91% accuracy while partially addressing the transparency gap [7]. Graph Neural Network approaches, which

model relational patterns among students and courses, report accuracy up to 89–93% but require specialized infrastructure impractical for most institutions [8]. Despite these advances, a recurring limitation is the failure to simultaneously achieve high accuracy, interpretability, and ease of deployment in resource-constrained settings.

## **IV. COMPARATIVE ANALYSIS**

A systematic review reveals a consistent trade-off between predictive accuracy and model transparency. Simple models such as Logistic Regression and Decision Trees are easily understood by educators but underperform on complex academic datasets. Advanced models including Bi-LSTM and GNN achieve superior accuracy but function as black boxes, limiting practical utility where predictions must be explained and justified to stakeholders.

Class imbalance presents a persistent challenge. Educational datasets are skewed toward passing students, biasing models toward the majority class. SMOTE-based oversampling addresses this at the data level but does not simultaneously resolve interpretability or deployment challenges. The comparative summary in Table I captures the key characteristics of reviewed approaches and highlights the gap the proposed hybrid framework is designed to fill.

TABLE I  
COMPARATIVE SUMMARY OF ML APPROACHES FOR STUDENT PREDICTION

Approach	Accuracy	Interpre- tability	XAI	Deplo y
Logistic Reg. [1]	72–78%	High	Moderate	High
KNN / SVM [2]	75–82%	Low	None	Moderate
Decision Tree [3]	70–76%	High	None	High
Random Forest [4]	80–85%	Moderate	Low	Moderate
Bi-LSTM [5]	87–92%	None	None	Low
AutoML [6]	84–88%	None	None	Low
LASA [7]	88–91%	Low	High	Low
GNN [8]	89–93%	None	None	Very Low
<b>Proposed</b>	<b>High</b>	<b>High</b>	<b>High</b>	<b>High</b>

### V. RESEARCH GAPS

A careful analysis of existing literature exposes several structural limitations that must be resolved before ML-based student result prediction achieves widespread practical adoption. The most significant gap is the persistent separation of accuracy and

interpretability as design objectives: systems optimize for one at the expense of the other, rather than seeking a principled balance.

Class imbalance remains a major challenge. Many high-performing models fail to reliably identify the minority class of at-risk students — the primary group prediction systems are intended to serve. Categorical feature handling is also frequently overlooked; academic records contain numerous categorical attributes such as department, course type, and admission category that standard models handle poorly without complex preprocessing.

Most proposed systems are designed as research prototypes with little consideration for integration into existing institutional infrastructure or usability by non-technical faculty. Complex deep learning models demand computational resources unavailable in typical educational institutions, creating a gap between reported benchmark performance and real-world applicability.

categorical features, robustness to class imbalance through built-in class weighting, and consistently strong classification performance across diverse datasets without demanding extensive preprocessing.

In the proposed framework CatBoost functions as the primary accuracy-focused component, learning complex non-linear patterns from student academic data. Its output probabilities are subsequently processed through Logistic Regression, which provides clear, coefficient-based explanations of the key factors influencing each prediction. This two-stage design ensures that the system is both predictively powerful and fully interpretable.

The framework is specifically engineered for practical deployment in educational institutions. It requires no high-end computational infrastructure, integrates into existing academic management systems, and produces outputs that faculty can act upon without data science expertise. Probability-based risk scores enable graded interventions — targeted tutoring, counselling — rather than purely binary classifications, supporting more nuanced academic support strategies.

TABLE II

RESEARCH GAPS IN ML-BASED STUDENT RESULT PREDICTION

#	Research Gap Identified
1	No unified framework balancing accuracy, interpretability, and efficiency
2	High-accuracy models lack transparency required by educators
3	Class imbalance causes poor detection of at-risk minority students
4	Inadequate handling of categorical academic data in standard models
5	Most systems are research prototypes, not deployment-ready
6	Deep learning requires infrastructure unavailable in most institutions

### VI. PROPOSED RESEARCH DIRECTION

The gaps identified above motivate the design of a hybrid prediction framework integrating CatBoost and Logistic Regression. CatBoost is selected for its native handling of

### VII. FUTURE SCOPE

Future research should explore integration of real-time academic monitoring data to enable continuous prediction updates throughout the semester. Incorporating SHAP and LIME alongside the hybrid framework would provide deeper, instance-level insights into individual student predictions, strengthening trust among educators and administrators.

Extending the framework to support multi-class grade prediction rather than binary pass/fail would broaden its applicability. Adapting the system for school through postgraduate programs and scaling it across multiple institutions with heterogeneous data formats are important future directions. Web-based dashboard deployment and Learning Management System integration would facilitate adoption by staff without programming backgrounds. Ongoing attention to algorithmic fairness, bias reduction, and ethical data governance will be

essential as such systems move toward broader institutional use.

## VIII. CONCLUSION

This paper has presented a structured review of machine learning approaches for student result prediction, tracing the evolution from simple statistical baselines to complex deep learning architectures. The reviewed body of work demonstrates clear technical progress; however, a persistent gap between predictive accuracy and model interpretability continues to limit practical deployment in real educational environments.

The proposed hybrid framework combining CatBoost and Logistic Regression addresses this gap by delivering strong classification performance without sacrificing the transparency that educational stakeholders require. By enabling early identification of at-risk students and supporting data-driven academic planning through interpretable predictions, the proposed approach holds the potential to meaningfully improve student outcomes and institutional decision-making. Addressing the remaining challenges of real-world deployment, fairness, and scalability through continued research will be essential to realising these benefits at scale.

---

## REFERENCES

- [1] A. Kumar and R. Sharma, "Student Result Prediction using Logistic Regression," *IEEE Conference on Educational Data Mining*, 2024.
- [2] V. Reddy and S. Patel, "Student Performance Prediction using KNN and SVM," *Springer – Lecture Notes in Computer Science*, 2025.
- [3] P. Singh and N. Verma, "Decision Tree-Based Academic Performance Prediction," *IEEE Access*, 2024.
- [4] M. Gupta and A. Jain, "Random Forest Model for Student Performance Evaluation," *Elsevier – Computers & Education*, 2024.
- [5] L. Zhou and Y. Wang, "Student Performance Prediction Using Bi-LSTM," *Springer – Applied Artificial Intelligence*, 2024.
- [6] Y. Zhou and Y. Wang, "AutoML-Based Student Academic Result Prediction," *Springer – Applied Artificial Intelligence*, 2025.
- [7] Y. Ren and X. Yu, "Long-Term Student Performance Prediction Using Learning Ability Self-Adaptive Algorithm," *Complex & Intelligent Systems*, Springer, 2024.
- [8] H. Li and Q. Chen, "Graph Neural Networks for Learning Behavior Analysis and Performance Prediction," *Elsevier – Knowledge-Based Systems*, 2024.
- [9] A. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic Course Planning Recommendation and Students' Performance Prediction Based on Educational Data Mining," *Journal of Computing in Higher Education*, Springer, 2025.
- [10] J. Martin and D. Lopez, "Predicting Student Performance in Online Learning Using Multidimensional Time-Series Data," *MDPI – Applied Sciences*, 2024.
- [11] Y. Zhang and M. Liu, "Explainable Student Performance Prediction Using Ensemble Machine Learning," *IEEE Access*, 2024.
- [12] K. Patel and D. Mehta, "Early Prediction of At-Risk Students Using Machine Learning Models," *Elsevier – Computers & Education: Artificial Intelligence*, 2024.
- [13] X. Huang and J. Li, "Explainable AI Models for Academic Performance Prediction in Higher Education," *IEEE Transactions on Learning Technologies*, 2025.
- [14] S. Khan and T. Ali, "Handling Imbalanced Student Data Using SMOTE and Machine Learning," *IEEE International Conference on AI in Education*, 2024.
- [15] N. Omar and R. Hassan, "Student Performance Prediction Using Machine Learning and Explainable AI," *Springer – Artificial Intelligence Review*, 2025.