IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

A Feasibility-Aware Counterfactual Framework for Explainable Early Academic Risk Assessment: A Synthetic-Data Prototype

Prof. Payal Shivaji Ulhare

MCA, Adsul Technical Campus, Ahilyanagar, Maharashtra, India

Abstract: Early academic risk prediction can support timely educational intervention, but a risk score alone does not explain why a student is flagged or what realistic changes may reduce that risk. This paper presents a feasibility-aware counterfactual framework that combines supervised risk prediction, local explainability, constrained counterfactual generation, and action-oriented reporting. A fully synthetic dataset of 2,500 student records was generated for controlled methodology validation. Four classifiers were evaluated using a stratified 75:25 hold-out split. Logistic Regression achieved the highest F1-score (0.567) and ROC-AUC (0.805) in this prototype. SHAP was used to explain model behaviour, while the counterfactual module searched bounded changes in potentially actionable academic-behavioural variables. Valid counterfactuals were found for 70.8% of the evaluated high-risk cases, with a mean of 1.86 changed features. The study is explicitly a synthetic-data prototype; its results demonstrate methodological feasibility and reproducibility rather than effectiveness on real students.

Keywords: Academic Risk Prediction, Explainable Artificial Intelligence (XAI), Counterfactual Explanations, Educational Data Mining, Early Warning Systems, SHAP, Actionable Recourse, Machine Learning.

I. INTRODUCTION

Machine learning is increasingly used in educational data mining to identify students who may require academic support. Recent research has combined early-warning prediction with explainability, fairness analysis, and counterfactual recommendations [1]–[4]. These developments show that predictive accuracy alone is insufficient when a model is intended to support human intervention.

A key distinction is between explanation and recourse. An explanation describes factors associated with a model prediction, whereas a counterfactual proposes a hypothetical change that would alter that prediction. A counterfactual can nevertheless be unrealistic if it changes a historical or uncontrollable variable. This work therefore focuses on feasibility constraints.

The paper does not claim that explainable academic-risk prediction or counterfactual explanations are new research areas. Instead, it presents a reproducible prototype that explicitly connects prediction, local explanation, and constrained recourse, and reports quantitative properties of the counterfactual search.

II. RESEARCH PROBLEM AND OBJECTIVES

The research problem is how an early academic-risk model can provide explanations and counterfactual recommendations while respecting predefined intervention constraints.
The objectives are: (1) construct a reproducible academic-risk prediction prototype; (2) compare multiple classifiers; (3) generate local explanations; (4) generate feasibility-aware counterfactuals; and (5) quantify counterfactual validity, sparsity, and feasibility.

III. RELATED WORK

Recent educational early-warning research has demonstrated machine-learning prediction and dynamic academic alert systems [1], [2]. Explainable AI has also been applied to student performance prediction using SHAP and domain knowledge [3]. More recent work has explicitly incorporated counterfactual explanations and prescriptive recommendations into educational-risk systems [4], [5].
For example, El Jihaoui et al. combine explainability, fairness auditing, and counterfactual explanations in an early-warning setting [1]. Muñoz-Muñoz et al. present a prescriptive educational dashboard with individual and group

counterfactual explanations [4], while Lara and Romero investigate group counterfactual explanations for students at risk of dropout [5]. These studies demonstrate that the general problem addressed here is already active. The present work therefore positions its contribution as a feasibility-constrained integration and reproducible prototype rather than a claim of universal novelty.

IV. PROPOSED FRAMEWORK

The proposed workflow contains five stages: (1) early-semester data preparation; (2) supervised risk prediction; (3) local explanation; (4) feasibility-aware counterfactual generation; and (5) action-oriented reporting.
The prediction model estimates P(Risk=1|X). SHAP values are used to identify influential features. For a high-risk instance, the counterfactual module searches for X' such that predicted risk falls below 0.50. Only predefined actionable variables may be changed, each within a specified maximum change. Candidate solutions are ranked first by the number of changed variables and then by weighted normalized change cost.
The permitted variables include attendance, assignment completion, learning activity, submission consistency, quiz performance, practical performance, and internal assessment. Historical performance and performance trend are not directly modified.

V. SYNTHETIC DATASET DESIGN

The experiment uses 2,500 synthetic records and contains no real student information. The features represent early academic and behavioural indicators: attendance percentage, internal assessment, assignment completion, practical score, quiz average, previous performance, learning activity, submission consistency, performance trend, and missed assignments.

The synthetic target contains 628 high-risk cases (25.1%) and 1,872 low-risk cases. A fixed random seed of 42 was used. The dataset, experiment outputs, figures, and supporting code are included with this manuscript package.

VI. EXPERIMENTAL METHODOLOGY

The evaluated classifiers were Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. A stratified 75:25 train-test split was used. Performance was measured using accuracy, precision, recall, F1-score, ROC-AUC, and precision-recall AUC.

Logistic Regression was selected for the explainability and counterfactual stage because it achieved the highest F1-score in this experiment (F1=0.567). SHAP was used for feature attribution. Counterfactual generation permitted at most two changed variables and used feature-specific maximum-change constraints.

VII. RESULTS

Table I summarizes the hold-out model comparison.

Model	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
Logistic Regression	0.731	0.476	0.701	0.567	0.805	0.586
Random Forest	0.774	0.562	0.465	0.509	0.762	0.519
Decision Tree	0.600	0.341	0.637	0.444	0.638	0.369
Gradient Boosting	0.771	0.597	0.274	0.376	0.773	0.531

The selected Logistic Regression model achieved accuracy=0.731, precision=0.476, recall=0.701, F1=0.567, ROC-AUC=0.805, and PR-AUC=0.586.

The five largest mean absolute SHAP contributors in the prototype were internal_score, assignment_completion_pct, attendance_pct, previous_performance, submission_consistency_pct.
Among 120 evaluated high-risk test cases, 85 received a valid counterfactual, corresponding to a validity rate of 70.8%. The mean number of changed features was 1.86, and the mean feasibility score was 0.182. These statistics describe the prototype's constrained search and are not evidence that the recommended changes improve actual student outcomes.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

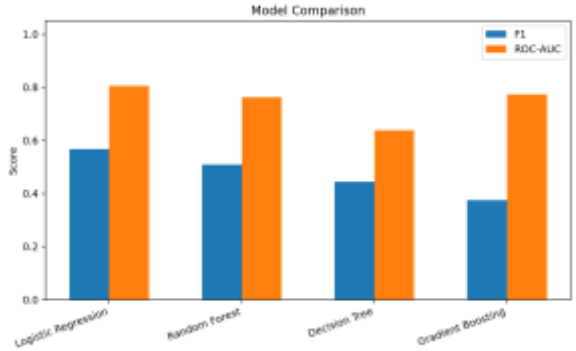


Fig. 1. Model comparison using F1-score and ROC-AUC.

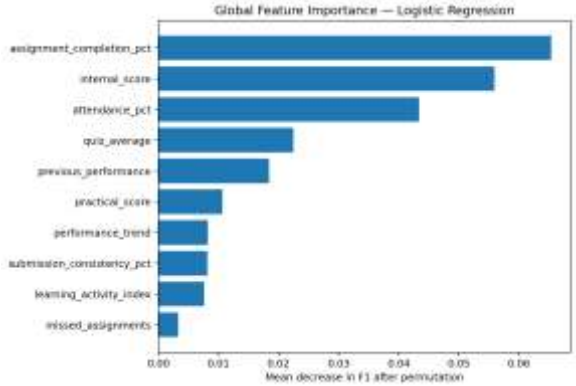


Fig. 2. Permutation-based global feature importance.

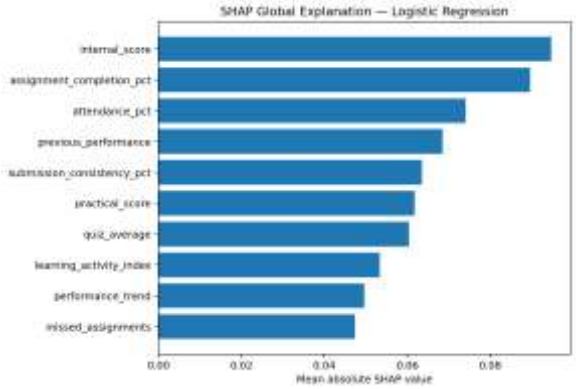


Fig. 3. Mean absolute SHAP importance.

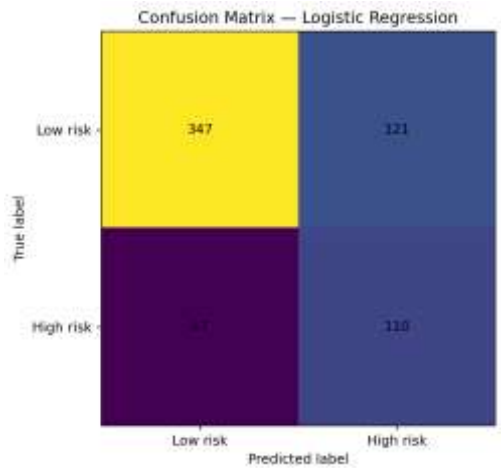


Fig. 4. Confusion matrix of the selected model.

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

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

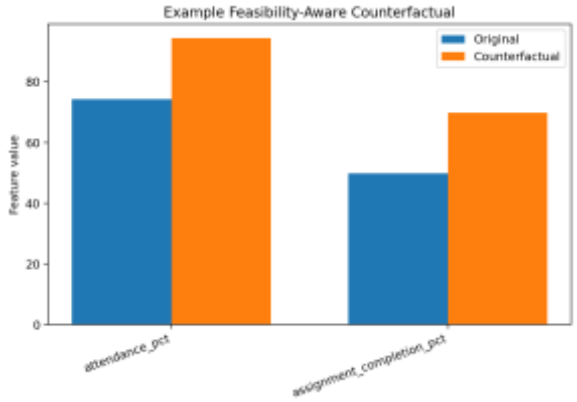


Fig. 5. Example of a bounded feasibility-aware counterfactual.

VIII. DISCUSSION

The experiment demonstrates that prediction, explanation, and constrained recourse can be implemented as separate but connected components. SHAP explains model behaviour; the counterfactual module proposes model-relative changes that may reduce predicted risk. The feasibility layer limits unrestricted numerical manipulation by restricting variables and bounding changes.

The framework should be treated as decision support rather than an automated decision maker. A high-risk flag should trigger human review and supportive academic intervention. Before institutional use, calibration, subgroup fairness, privacy, data governance, and longitudinal effectiveness should be evaluated.

IX. ETHICAL AND PRIVACY CONSIDERATIONS

The prototype uses synthetic data and therefore does not involve human-subject data collection. If real institutional data are introduced, the study should obtain the approvals required by the institution and applicable regulations, minimize collected data, document the purpose of processing, apply access controls, and define retention and deletion procedures. Predictions should not be used as the sole basis for punitive or exclusionary decisions.

X. LIMITATIONS

The principal limitation is the synthetic dataset. Its data-generating process cannot reproduce all institutional, behavioural, socioeconomic, curricular, and temporal patterns present in real education. The decision threshold and feasibility constraints are also manually specified. A counterfactual that lowers model probability does not establish that the proposed action will improve actual academic outcomes. External and temporal validation are therefore necessary.

XI. FUTURE WORK

Future research should validate the framework using ethically approved institutional data, perform temporal and cross-institutional validation, evaluate calibration and subgroup fairness, and compare multiple counterfactual-generation algorithms. Longitudinal studies should test whether recommendations are understandable and feasible and whether they are associated with improved student outcomes. Privacy-preserving learning may also be investigated where centralized data sharing is inappropriate.

XII. CONCLUSION

This paper presented a feasibility-aware counterfactual framework for explainable early academic risk assessment. On synthetic data, Logistic Regression achieved the strongest F1-score (0.567) and ROC-AUC (0.805) among the four evaluated models. The constrained counterfactual module found valid recourse for 70.8% of the evaluated high-risk cases. The work provides a reproducible methodology for connecting prediction, explanation, and feasible recourse, while explicitly avoiding claims of real-world effectiveness or universal novelty.

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

ACKNOWLEDGMENT AND AI-USE DISCLOSURE

AI-assisted development disclosure: OpenAI ChatGPT was used to assist with manuscript drafting, language editing, experimental-code generation, data-analysis workflow development, and figure/document preparation. The authors are responsible for reviewing and verifying the manuscript, references, experimental methodology, results, interpretations, and final submitted version. The exact disclosure should be adjusted to match the authors' actual use of AI and the requirements of the target IEEE publication.

DATA AND CODE AVAILABILITY

The dataset used for the reported experiment is synthetic and is supplied with the manuscript package. The experiment outputs and reproducibility materials are also supplied. Before submission, the authors should decide whether and how to publish the code and data according to the target venue's data-sharing requirements.

REFERENCES

[1] M. El Jihaoui, O. E. K. Abra, and K. Mansouri, “Towards a fair and transparent early warning system for identifying students at risk of academic failure,” Frontiers in Education, vol. 11, Art. no. 1878686, 2026, doi: 10.3389/feduc.2026.1878686.

[2] Q. Li, Y. Liu, R. Ma, and Q. Wu, “Academic early warning model based on machine learning and model application,” Scientific Reports, vol. 16, Art. no. 25621, 2026, doi: 10.1038/s41598-026-51767-6.

[3] M. Qiang, Z. Liu, and R. Zhang, “Explainable AI in education: integrating educational domain knowledge into the deep learning model for improved student performance prediction,” Scientific Reports, vol. 16, Art. no. 9515, 2026, doi: 10.1038/s41598-026-40538-y.

[4] M. Muñoz-Muñoz, C. Luna, J. A. Lara, and C. Romero, “Dropout insight: Educational risk dashboard with counterfactual explanations,” SoftwareX, vol. 33, Art. no. 102551, 2026, doi: 10.1016/j.softx.2026.102551.

[5] J. Lara and C. Romero, “Group counterfactual explanations: A use case to support students at risk of dropping out in online education,” Electronics, vol. 15, no. 1, Art. no. 51, 2026, doi: 10.3390/electronics15010051.

[6] V. Hegde, A. Palaniswamy, D. Bhar, M. Revanna, N. Prakash, A. Prabhu, and P. M. Shanthappa, “An explainable machine learning framework for early student dropout risk prediction and stratification,” Discover Artificial Intelligence, vol. 6, Art. no. 1224, 2026, doi: 10.1007/s44163-026-02288-8.

[7] IEEE Author Center, “Structure Your Paper,” IEEE, 2026.

[8] IEEE Author Center, “Submission and Peer Review Policies,” IEEE, 2026.