

Article

A Computational Analysis of Information Reduction in Academic Risk Prediction

Orlando Arroyo ^{1,*}, José M. Benjumea ¹  and Jesús Villalba ² ¹ Escuela de Ingeniería Civil, Universidad Industrial de Santander, Bucaramanga 680006, Colombia; josbenro@uis.edu.co² Departamento de Ingeniería Civil, Pontificia Universidad Javeriana, Bogotá 110231, Colombia; jesus.villalba@javeriana.edu.co

* Correspondence: orlando.arroyo@uis.edu.co

Abstract

Academic risk prediction often assumes that richer educational data yields better performance, yet many institutions operate with strictly limited information. This study quantifies how predictive behavior degrades when information is progressively constrained within the feature space. Using prerequisite-based academic records from three Civil Engineering courses at Universidad Industrial de Santander, Colombia, we evaluate nested information regimes ranging from a no-information baseline to a full ten-variable prerequisite representation. We assessed logistic regression models through repeated cross-validation, measuring continuous discrimination, probabilistic calibration, and threshold-dependent classification. Additional paired and chronology-controlled robustness analyses evaluated whether the reduced and full regimes differed materially across folds, nonlinear learners, first-attempt cohorts, and target-year blocks. The results demonstrate predictive compressibility within prerequisite-only records: transitioning from the baseline to a one-dimensional model (the average prerequisite grade) drives the largest performance gain, while subsequent feature additions yield rapidly diminishing returns. Despite a 90% reduction in dimensionality, this sparse model retains between 98.0% and 100.1% of the full-model ROC-AUC and maintains stable probabilistic calibration. These findings expose substantial structural redundancy within broader prerequisite representations, suggesting that sparse, interpretable academic summaries capture most of the measurable prerequisite-based predictive signal. Ultimately, academic risk estimation remains feasible under severe data constraints, although the moderate absolute ROC-AUC values indicate that these models are best interpreted as transparent triage tools rather than complete explanations of academic performance.

Keywords: academic risk prediction; educational data mining; logistic regression; predictive compressibility; early-warning systems



Academic Editor: Sandra Ferreira

Received: 9 May 2026

Revised: 23 June 2026

Accepted: 30 June 2026

Published: 2 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Academic risk prediction is central to educational data mining and learning analytics. Early identification of at-risk students enables timely intervention, improves retention, and supports academic management [1–4]. Predictive models utilize heterogeneous data—such as academic, demographic, behavioral, and administrative records—to forecast outcomes like final grades, course failure, and timely graduation [4–8]. Predictive performance often improves when behavioral variables complement academic records, and when feature selection isolates dominant predictors [5,6,9,10].

Despite these advances, the literature predominantly focuses on performance optimization—emphasizing model comparison, hyperparameter tuning, and metrics like accuracy and AUC [11–14]. However, accuracy alone can be misleading in imbalanced educational data, as minority-class recall often remains poor [8,15]. This performance-centric approach overlooks a critical question: how does predictive behavior change when information is progressively removed from the feature space? This question is highly relevant given the disparity in educational data availability [4,6–8,16,17]. While some institutions collect extensive behavioral and online-activity logs, many operate with strictly limited or incomplete Student Information System (SIS) records [7,8,17]. Moreover, some institutions only have information about academic student performance. Thus, data-intensive predictive techniques are simply not viable for institutions lacking the necessary volume and diversity of information [8,16,17].

Computationally, feature reduction is typically deployed to mitigate high dimensionality, remove redundant variables, and maximize model performance [9,10,18,19]. This study deploys feature reduction as an analytical instrument to examine informational redundancy and predictive degradation. The objective is to quantify exactly how discrimination, calibration, and classification behavior evolve under strictly constrained informational regimes.

We use the term predictive compressibility to describe the empirical situation in which a reduced representation preserves most of the measurable predictive behavior of a richer representation. This concept is closely related to redundancy analysis and interpretable feature reduction, but the goal here is not to introduce a new classification algorithm. Rather, the contribution is to measure how much prerequisite-based academic signal remains after progressively removing information, and to evaluate whether the resulting low-dimensional model remains calibrated and stable enough for transparent risk triage. This distinction matters because academic performance is influenced by many omitted behavioral, demographic, instructional, and temporal factors; compressibility within prerequisite records does not imply that these other factors are unimportant.

This study presents a computational analysis of information reduction in academic risk prediction. We investigate how predictive behavior shifts as feature subsets progressively shrink, complementing traditional evaluations of peak performance. The analysis relies on course-level academic datasets from the Civil Engineering program at Universidad de Santander, Colombia, and evaluates three key dynamics: predictive degradation (the performance loss from removing information), marginal information contribution (the incremental gain from adding each variable), and calibration stability (the reliability of probabilistic predictions at lower dimensionalities).

This work contributes a mathematically structured, empirically grounded characterization of predictive evolution under feature reduction. By analyzing degradation curves, marginal gains, calibration behavior, and informational redundancy, the study provides a computational perspective that enriches existing performance-oriented analyses in educational prediction. In addition, this study highlights the practical value of simple models for institutions with constrained data environments, which may not have access to extensive behavioral or online-activity logs.

The manuscript is organized as follows: Section 2 reviews related work on academic risk prediction, sparse modeling, and information-constrained inference. Section 3 details the mathematical formulation of the framework, describes the datasets, models, and experimental design. Section 4 presents the results. Section 5 discusses the implications and limitations of the findings, and Section 6 concludes the study.

2. Literature Review

Information reduction is a computationally consequential component of academic risk prediction. Broad reviews of educational prediction highlight feature selection as a core challenge alongside student attributes and modeling strategies [1,20]. Researchers consistently use feature selection to reduce high dimensionality, manage uncertainty, remove redundant variables, and isolate dominant predictors that improve model performance [9,10,18]. Early-warning systems similarly screen and reduce features before applying ensemble learning [4]. Thus, feature reduction simplifies models and explicitly determines which information functions as the predictive signal.

Crucially, predictive gains do not scale indefinitely with larger feature sets or greater model complexity. For instance, one feedforward network achieved 91.95% accuracy using only behavioral attributes, closely matching the 93.17% achieved without feature selection [21]. Another study found that a minimalistic feature set outperformed larger alternatives, reaching 91.2% classification accuracy [22]. Similarly, in online programming courses, simple logistic regression achieved the highest test AUC-ROC, outperforming complex models like SVMs and neural networks [23]. Recent work on student engagement reinforces this pattern, arguing that near-optimal accuracy becomes preferable when paired with high computational efficiency and robustness [24]. While no universal saturation threshold exists, these findings demonstrate that predictive performance often plateaus, meaning additional complexity does not guarantee better educational models.

A related line of research distinguishes discrimination from calibration. Many educational prediction studies heavily emphasize discriminative measures like precision, recall, F1-score, and ROC-AUC [11,14,25]. However, probabilistic models increasingly prioritize predicted risk probabilities and their associated uncertainty over hard classification accuracy. For example, a probabilistic logistic regression study for at-risk identification found that model uncertainty decreases significantly later in the academic calendar, improving prediction reliability [26]. Recent work on multimodal engagement assessment also emphasizes that stability and balanced evaluation are critical even when accuracy is high [24]. Broader uncertainty-aware classification research evaluates expected calibration error, proving that strong classification performance and reliable probabilities are distinct achievements [27]. This distinction remains vital for academic risk prediction because institutions typically trigger interventions based on predicted probabilities.

Furthermore, sparse and interpretable educational models prove both feasible and practically valuable. Many approaches intentionally rely on a restricted set of early or readily available indicators. Early warning models based on course grades demonstrate that failing a single marking period strongly predicts graduation status [28]. Engineering models similarly identify at-risk students using compact combinations of grade history, LMS interactions, and entry profiles [29,30]. Interpretability has become a primary requirement: recent studies apply explainable machine learning tools like SHAP to identify key predictive factors [12,25], and design belief-rule-based approaches specifically to ensure transparency [31]. Institutional data availability also varies widely. While some datasets integrate academic, demographic, behavioral, and administrative records [5,8], others rely strictly on standard SIS data without external surveys or LMS inputs [7]. Because missing educational data remains a persistent challenge, researchers often require dedicated imputation prior to prediction [17]. Consequently, sparse educational models represent practical, deployable solutions tailored for varied data environments.

Table A1 (Appendix A) summarizes the methodologies reported in the literature on academic performance prediction, highlighting the datasets employed, the predictive approaches adopted, the principal findings, and the main strengths of each study. Despite the significant progress achieved in this field, most studies have focused on reducing

the feature space as a means of improving predictive performance. Consequently, there remains a need to investigate how model behavior evolves, degrades, or stabilizes when the amount of available information is progressively restricted.

3. Materials and Methods

3.1. Academic Dataset

This study uses academic records from the undergraduate Civil Engineering program at Universidad Industrial de Santander (UIS), Colombia. We analyze three foundational courses: Statics (*Estática*), Solid Mechanics (*Mecánica de Sólidos*), and Fluid Mechanics (*Mecánica de Fluidos*). These core courses share strong prerequisite relationships and require progressive mechanical reasoning.

The dataset spans from the first academic term of 2007 (2007-1) to the first term of 2025 (2025-1). The Civil Engineering program has operated under essentially the same structural curricular grid since October 2007, when a major curricular reform was implemented. Although a subsequent adjustment was introduced in 2013, it did not alter the core courses analyzed in this study; rather, it was limited to the reorganization of elective technical courses in the final semesters. Consequently, despite the 2007-1 grades belonging to the pre-2007 curricular grid, the longitudinal dataset remains highly homogeneous. The effect of the pre-2007 data is considered negligible. Finally, it is worth noticing that a comprehensive structural reform is under development and is expected to be deployed in 2027.

We analyze each course independently to evaluate whether information-reduction behavior remains consistent across different academic contexts and approval-rate distributions. Table 1 reports the primary modeling cohorts used in the repeated-validation experiments. These cohorts include students with an observed final grade in the target course and complete records for the four prerequisite courses used in the information regimes.

Table 1. Primary modeling cohorts for the prerequisite-based prediction experiments.

Target Course	Students	Approved	Failed	Pass Rate
Statics	3839	2565	1274	0.668
Mechanics of Solids	3508	2716	792	0.774
Fluid Mechanics	3344	3004	340	0.898

UIS employs a grading system from 0.0 to 5.0, with 3.0 or higher constituting a passing grade. Consequently, we formulate the prediction target as a binary classification problem where:

$$y_i = \begin{cases} 1 & \text{if student } i \text{ passed the course} \\ 0 & \text{if student } i \text{ failed the course} \end{cases}$$

We anonymized the dataset prior to analysis and excluded all personally identifiable information from model training and evaluation. The primary feature-reduction experiments used a student-level table in which each row represents one anonymized student and each course column stores the final observed grade available for that course. The institutional records contain official final course grades rather than disaggregated assessment components. The university grade spreadsheet stores grades only for students who completed the courses and does not include withdrawal records or a withdrawal-status field. Therefore, withdrawn students were not modeled as failures; the failure class consists of completed course attempts with official final grades below 3.0. For Fluid Mechanics, any group-project or peer-assessment component that may have contributed to the official course grade is embedded in the final individual grade recorded by the university; group identifiers

and project-component scores were not available and were not modeled separately. To address chronology and repeated-enrollment concerns, we additionally used the original long-format academic-record workbook, which contains student identifier, enrollment year, enrollment period, course name, and final grade for each course registration.

Figure 1 presents the grade distributions for the three courses. Although all courses concentrate around the passing threshold, they exhibit important differences in approval rates and lower-tail behavior. Statics and Solid Mechanics show broader distributions with higher densities near the threshold and pronounced lower-grade tails. Conversely, Fluid Mechanics concentrates strongly toward higher grades, resulting in higher approval prevalence. These differences directly affect class imbalance, threshold sensitivity, and calibration behavior in our predictive experiments. Particularly, Statics and Mechanics of Solids have a significant number of students with grades below 3.0, indicating a higher degree of academic risk.

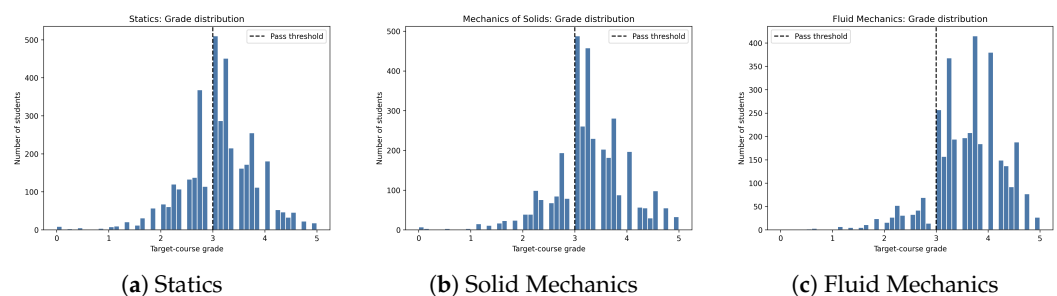


Figure 1. Grade distributions for the analyzed engineering courses. The dashed vertical line indicates the institutional passing threshold of 3.0. The distributions illustrate differences in approval prevalence and lower-tail behavior across courses, which directly influence class imbalance and threshold-dependent predictive metrics.

3.2. Prerequisite-Based Feature Representation

This study focuses on progressively enriched representations of prerequisite-course performance. We restrict the feature space strictly to prerequisite-derived academic indicators. This formulation systematically evaluates how predictive behavior evolves as models incorporate richer prerequisite data.

We construct multiple statistical descriptors from the foundational prerequisite courses, including average performance, minimum grade, variability indicators, and counts of failed prerequisites. The objective is to quantify predictive degradation, marginal information contribution, and calibration behavior under these progressively enriched prerequisite representations.

3.3. Information Regimes

We construct a hierarchy of nested information regimes to analyze predictive behavior, as summarized in Table 2. Each regime captures prerequisite performance at a different level of detail, ranging from a no-information baseline (R_0) to a complete prerequisite-derived representation (R_5). By structuring the feature space this way, we systematically evaluate predictive degradation, marginal predictive contributions, calibration stability, and informational redundancy.

The ten variables in the full regime R_5 are therefore the four observed prerequisite grades (Calculus I, Calculus II, Physics I, and Linear Algebra I) and six deterministic summaries of those grades: mean, minimum, maximum, standard deviation, number of prerequisite grades below the passing threshold of 3.0, and number of prerequisite grades below 3.5. The term “information reduction” is used in this bounded sense: the study does not claim that ten variables constitute high-dimensional educational data in general. Instead, it asks whether the full prerequisite-derived representation available in this setting

can be reduced to a much smaller, more transparent representation without materially changing predictive behavior.

Table 2. Hierarchy of information regimes defining the feature spaces.

Regime	Type	Variables Included
R_0	Baseline	None (predicts global training-fold passing probability)
R_1	Minimal	Average prerequisite grade
R_2	Low	Average prerequisite grade, Count of prerequisite grades below 3.0
R_3	Medium	Average prerequisite grade, Count of prerequisite grades below 3.0, Minimum prerequisite grade, Standard deviation of prerequisite grades, Count of prerequisite grades below 3.5
R_4	Raw Grades	Raw prerequisite-course grades
R_5	Full	Four raw prerequisite-course grades plus six aggregate descriptors: average, minimum, maximum, standard deviation, count below 3.0, and count below 3.5

3.4. Mathematical Formulation of Information Reduction in Academic Risk Prediction

Consistent with the preceding sections, we formalize information reduction as an ordered sequence of inference problems. We seek to quantify how discrimination, calibration, classification behavior, and computational complexity evolve as the predictive model is exposed to progressively richer representations of the academic environment.

Let $\mathcal{C} = \{1, \dots, C\}$ index the target courses. For each course $c \in \mathcal{C}$, we observe a dataset $\mathcal{D}^{(c)} = \{(x_i^{(c)}, y_i^{(c)})\}_{i=1}^{n_c}$, where $x_i^{(c)} \in \mathbb{R}^d$ is the complete feature vector for student i , and $y_i^{(c)} \in \{0, 1\}$ denotes the binary academic outcome, with $y_i^{(c)} = 1$ indicating course approval.

We consider a nested family of information regimes $\mathcal{R} = \{R_0, R_1, R_2, R_3, R_4, R_5\}$. A mapping $k : \mathcal{R} \rightarrow \mathbb{N}$ defines the dimensionality of each regime, with $k(R_0) = 0$, $k(R_1) = 1$, $k(R_2) = 2$, $k(R_3) = 5$, $k(R_4) = 4$, and $k(R_5) = 10$. For each regime $r \in \mathcal{R}$, a representation map $\phi_r : \mathbb{R}^d \rightarrow \mathbb{R}^{k(r)}$ yields a regime-specific feature vector $z_{i,r}^{(c)} = \phi_r(x_i^{(c)})$. This nesting ensures that lower-information representations are derived by restricting or aggregating the full feature space. Specifically, R_0 serves as a no-information baseline, R_1 relies solely on the prerequisite grade average, R_2 incorporates the count of failed prerequisites, R_3 employs five aggregate prerequisite descriptors, R_4 uses the four raw prerequisite grades, and R_5 utilizes the full ten-variable representation.

Under each regime r , the objective is to estimate a mapping $f_{c,r} : \mathbb{R}^{k(r)} \rightarrow [0, 1]$ such that $p_{i,r}^{(c)} = f_{c,r}(z_{i,r}^{(c)})$ approximates the probability of course approval. For $r \neq R_0$, we employ a logistic regression model,

$$p_{i,r}^{(c)} = \sigma(\beta_{0,c,r} + \beta_{c,r}^\top z_{i,r}^{(c)}), \quad \sigma(u) = \frac{1}{1 + e^{-u}},$$

which has $p(r) = k(r) + 1$ fitted parameters. The baseline regime R_0 ignores student-specific covariates and simply predicts the empirical pass rate of the training sample, $p_{i,R_0}^{(c)} = \bar{y}_{\text{train}}^{(c)}$, establishing a reference to measure the signal recovered by introducing minimal information. Although formulated here for logistic regression, the framework relies solely on out-of-fold predicted probabilities, allowing direct extensions to tree-based and ensemble learners.

To ensure robust estimation, we evaluate models using repeated cross-validation over splits $s = 1, \dots, S$ into disjoint training sets $I_{\text{tr}}^{(s)}$ and validation sets $I_{\text{va}}^{(s)}$. The same split

indices are reused across information regimes and model classes within each course, so regime comparisons are paired at the fold level. A model $f_{c,r}^{(s)}$ is trained on $I_{tr}^{(s)}$, generating validation predictions $\tilde{p}_{i,r}^{(c,s)}$ for $i \in I_{va}^{(s)}$. For any evaluation metric M , its split-level value is denoted by $M_{c,r}^{(s)}$. Performance is summarized via the repeated-validation mean $\hat{M}_{c,r} = \frac{1}{S} \sum_{s=1}^S M_{c,r}^{(s)}$ and its standard deviation $SD_{c,r}(M)$. Empirical 95% intervals are obtained from the 2.5% and 97.5% quantiles of the split-level metric distribution.

Model performance is evaluated across discrimination, calibration, and threshold-dependent classification metrics. Discrimination is assessed by the ROC-AUC. Calibration is quantified via the Brier score,

$$B_{c,r}^{(s)} = \frac{1}{|I_{va}^{(s)}|} \sum_{i \in I_{va}^{(s)}} \left(\tilde{p}_{i,r}^{(c,s)} - y_i^{(c)} \right)^2,$$

and through the calibration intercept $\alpha_{c,r}^{(s)}$ and slope $\beta_{c,r}^{(s)}$ derived from the logistic recalibration model, $\text{logit}(\Pr(Y^{(c)} = 1 \mid \tilde{p}_r^{(c,s)})) = \alpha_{c,r}^{(s)} + \beta_{c,r}^{(s)} \text{logit}(\tilde{p}_r^{(c,s)})$. Ideal calibration corresponds to $\alpha_{c,r}^{(s)} = 0$ and $\beta_{c,r}^{(s)} = 1$. Threshold-dependent metrics, such as the F1 score and balanced accuracy, are evaluated at a fixed decision boundary of $t = 0.5$. Furthermore, model stability is reflected in $SD_{c,r}(M)$, where lower variance indicates greater robustness across data splits.

To rigorously compare predictive degradation across regimes, we introduce a sign indicator $\eta_M \in \{+1, -1\}$, where $+1$ corresponds to metrics to be maximized (e.g., ROC-AUC, F1) and -1 to those to be minimized (e.g., Brier score). The degradation of regime r relative to the full representation R_5 is defined as

$$\Delta_{c,r}^{(M)} = \eta_M \left(\hat{M}_{c,R_5} - \hat{M}_{c,r} \right).$$

A degradation near zero implies that the reduced regime preserves the predictive fidelity of the full model. Similarly, the marginal gain of advancing from a preceding regime r^- to r is $G_{c,r}^{(M)} = \eta_M (\hat{M}_{c,r} - \hat{M}_{c,r^-})$. To formalize the capacity of a low-dimensional regime to capture the majority of the available signal, we define the predictive compressibility ratio for higher-is-better metrics as

$$\Pi_{c,r}^{(M)} = \frac{\hat{M}_{c,r} - \hat{M}_{c,R_0}}{\hat{M}_{c,R_5} - \hat{M}_{c,R_0}}, \quad \text{for} \quad \hat{M}_{c,R_5} > \hat{M}_{c,R_0}.$$

Values of $\Pi_{c,r}^{(M)}$ close to one indicate that regime r retains nearly all the performance improvement offered by the full ten-variable model.

Finally, the computational impact of information reduction is quantified by the relative dimensionality reduction, $\delta(r) = 1 - k(r)/k(R_5)$, alongside the mean fitting and prediction times per fold. By structuring feature reduction as a sequence of measurable comparisons, this framework enables a comprehensive analysis of the trade-offs between predictive capacity, probabilistic reliability, and model complexity. Within this study, these formulations are applied as empirical descriptors of specific educational settings, demonstrating that high-performing risk prediction may often be recovered from sparse, interpretable academic summaries.

3.5. Computational Workflow

Figure 2 summarizes the architecture of the experimental pipeline. The workflow begins with anonymized academic records, constructs target-specific prerequisite represen-

tations, evaluates nested information regimes under paired repeated validation, and then summarizes discrimination, calibration, stability, and computational-cost metrics.

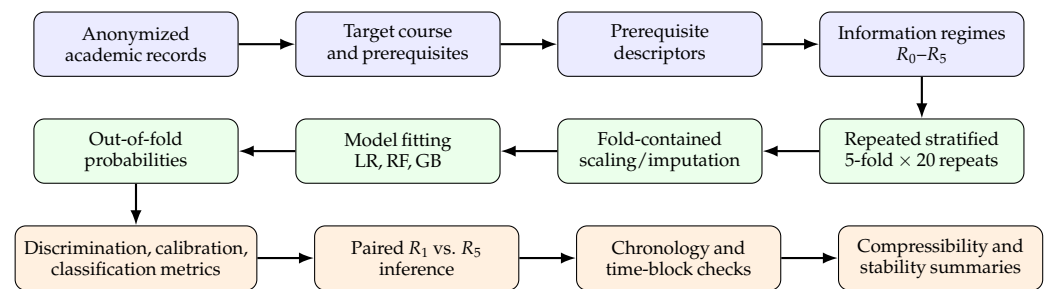


Figure 2. Computational workflow for the prerequisite-based information-reduction experiments. Each target course is modeled independently. Regime comparisons are paired because the same validation folds are reused across information regimes and model classes.

3.6. Data Preprocessing

Prior to model training, the academic records underwent systematic filtering and transformation. The main preprocessing sequence was:

1. Select the target course and the four prerequisite courses: Calculus I, Calculus II, Physics I, and Linear Algebra I. This anchors every experiment to the same curricular dependency structure.
2. Remove observations without an observed target-course grade, because the binary approval outcome cannot be defined for those students. Since the source spreadsheet stores grades only for completed course attempts, withdrawn students are outside the modeled cohorts.
3. For the primary feature-reduction experiments, retain students with complete grades for the four prerequisite courses so that all regimes are compared on the same cohort rather than on samples that differ because of missingness.
4. Define course approval as $y_i = 1$ when the target-course grade is at least 3.0 and $y_i = 0$ otherwise, matching the institutional grading policy.
5. Compute prerequisite-derived descriptors: average, minimum, maximum, standard deviation, number of prerequisites below 3.0, and number of prerequisites below 3.5. These row-level summaries quantify central tendency, lower-tail risk, and grade dispersion without using target-course information.
6. Generate the regime-specific feature matrices R_0 – R_5 from these descriptors and raw prerequisite grades, allowing each information regime to be trained and evaluated under identical validation folds.

These steps are necessary because the study compares information regimes rather than optimizing a single predictive model. Keeping the same students across regimes prevents apparent performance differences from being caused by changing sample composition. Aggregate descriptors are computed directly from observed prerequisite grades before cross-validation because they are deterministic row-level summaries and do not use target-course information. Fold-dependent preprocessing, including median imputation for raw-feature models and feature scaling for logistic regression, is fitted only on the training partition and then applied to the corresponding validation partition.

3.7. Predictive Modeling and Experimental Design

The predictive task functions as a probabilistic estimation of course approval. The computational workflow independently trains and evaluates predictive models for each specific course and information regime.

Our experiments explicitly quantify predictive discrimination, threshold-dependent classification behavior, calibration stability, computational efficiency, and robustness under paired repeated validation.

The main repeated-validation design used repeated stratified cross-validation with five folds and twenty repeats, yielding 100 validation folds per course. The same fold assignments were reused across information regimes and model classes within each course. The primary split was not time-based; it was designed to compare information regimes under identical paired validation folds over the available institutional distribution. To evaluate whether the resulting pattern depended on temporal mixing, we conducted the chronology-controlled and target-year block analyses described below. Logistic regression used standardized predictors and a maximum of 5000 iterations. Random forest models used 200 trees, a minimum leaf size of 10, and the same random seed. Gradient boosting used 120 estimators, learning rate 0.05, maximum depth 2, and minimum leaf size 10. These modest fixed configurations were selected to test whether the information-reduction pattern persisted beyond the linear logistic model without making the results dependent on extensive hyperparameter tuning.

Performance evaluation relies on ROC-AUC, PR-AUC, balanced accuracy, F1-score, Brier score, log loss, calibration slope, and calibration intercept. Because the validation folds are paired across regimes, we additionally computed paired fold-level comparisons between the minimal regime R_1 and the full regime R_5 using Wilcoxon signed-rank tests for ROC-AUC and Brier score, rank-biserial effect sizes, empirical 95% intervals for paired differences, and Holm-adjusted p -values within metric families. These tests are used to quantify whether small observed differences are practically negligible or systematic across validation folds.

3.8. Experimental Environment and Reproducibility

All analyses were implemented in Python 3.12.12 on macOS 26.5 with an arm64 platform. The computational environment used NumPy 1.26.4, pandas 3.0.2, scikit-learn 1.8.0, SciPy 1.17.1, and Matplotlib 3.10.8. Random seeds were fixed for repeated cross-validation and for stochastic model classes. The fixed validation design and model configurations described above were used for all course-regime comparisons to make the experiments reproducible and to ensure that observed differences between information regimes were not caused by changing folds, changing preprocessing rules, or model-specific tuning choices.

3.9. Chronology-Controlled Robustness Analysis

For each target course, we reconstructed a first-attempt cohort by selecting each student's first observed enrollment in the target course. For every prerequisite, we used the latest prerequisite attempt completed strictly before that first target-course attempt. Students without all four prerequisites before the target attempt were excluded from this robustness cohort. This procedure prevents target-course retakes or post-target prerequisite records from contributing information to the predictors. The institutional records do not include a separate indicator for equivalent online-course completion. Terms affected by pandemic-era remote instruction are therefore retained within the 2020–2025 target-year block rather than modeled as a separate course modality.

The chronology-controlled cohorts contained 3704 students for Statics, 3484 for Mechanics of Solids, and 3314 for Fluid Mechanics. We then repeated the R_1 versus R_5 analysis for logistic regression, random forest, and gradient boosting. To evaluate time-related stability, we also repeated the logistic comparison within three target-year blocks: 2007–2013, 2014–2019, and 2020–2025.

4. Results

4.1. Predictive Performance Across Information Regimes

Table 3 summarizes the repeated cross-validation results for the logistic models across the primary information regimes. We define “full” as the ten-variable prerequisite representation (R5), “medium” as the five aggregate prerequisite variables (R3), “low” as the two-variable representation using prerequisite average and the count of prerequisites below 3.0 (R2), and “minimal” as the single-variable prerequisite-average representation (R1). The baseline regime (R0) employs no predictors, assigning the training-fold pass rate to each validation observation.

Table 3. Predictive performance of logistic models across information regimes. Values are repeated-cross-validation means.

Course	Regime	Variables	ROC-AUC	PR-AUC	F1	Bal. Acc.	Brier Score
Statics	R0 baseline	0	0.500	0.668	0.801	0.500	0.222
Statics	R1 minimal	1	0.668	0.804	0.792	0.554	0.205
Statics	R3 medium	5	0.669	0.806	0.797	0.556	0.205
Statics	R5 full	10	0.671	0.808	0.798	0.555	0.205
Mechanics of Solids	R0 baseline	0	0.500	0.774	0.873	0.500	0.175
Mechanics of Solids	R1 minimal	1	0.659	0.867	0.871	0.505	0.166
Mechanics of Solids	R3 medium	5	0.663	0.869	0.870	0.502	0.166
Mechanics of Solids	R5 full	10	0.673	0.875	0.872	0.513	0.164
Fluid Mechanics	R0 baseline	0	0.500	0.898	0.946	0.500	0.091
Fluid Mechanics	R1 minimal	1	0.671	0.945	0.946	0.500	0.089
Fluid Mechanics	R3 medium	5	0.668	0.945	0.946	0.500	0.089
Fluid Mechanics	R5 full	10	0.670	0.946	0.946	0.500	0.089

The no-information baseline yielded a 0.500 ROC-AUC across all courses. Transitioning from R0 to the one-dimensional R1 model increased ROC-AUC to 0.668 for Statics, 0.659 for Mechanics of Solids, and 0.671 for Fluid Mechanics. Given the class imbalance observed in the lower-risk courses, we also report the Area Under the Precision–Recall Curve (PR-AUC), which provides a more robust measure of minority-class separation. PR-AUC increased from 0.668 to 0.804 for Statics, 0.774 to 0.867 for Mechanics of Solids, and 0.898 to 0.945 for Fluid Mechanics. Corresponding Brier scores decreased from 0.222 to 0.205, 0.175 to 0.166, and 0.091 to 0.089, respectively. Therefore, a single compressed prerequisite-performance variable captures most of the predictive gain over the baseline.

The full prerequisite-derived representation offered only modest improvements over the minimal model. Compared to R5, R1 lost 0.004 ROC-AUC points for Statics and 0.014 points for Mechanics of Solids, while gaining 0.0004 points for Fluid Mechanics. PR-AUC followed a similar plateau, with R5 showing minimal gains over R1 (0.808 vs. 0.804 for Statics, 0.875 vs. 0.867 for Mechanics of Solids, and 0.946 vs. 0.945 for Fluid Mechanics). Because the courses have high pass rates, a 0.5 probability threshold classified most students as passing, keeping balanced accuracy near 0.50 for Mechanics of Solids and Fluid Mechanics. For R1 out-of-fold predictions, pass-class precision and recall reached 0.695 and 0.920 for Statics, 0.776 and 0.992 for Mechanics of Solids, and 0.898 and 1.000 for Fluid Mechanics. Class imbalance and the arbitrary 0.5 decision threshold strongly skew these threshold-dependent metrics; consequently, high pass-class recall does not necessarily indicate effective separation of failures. We therefore base our primary interpretation on threshold-free ranking metrics (ROC-AUC and PR-AUC), Brier score, and calibration.

4.2. Information–Performance Trade-Off

Figure 3 plots ROC-AUC against the number of prerequisite-derived variables. The curves demonstrate early saturation: the prerequisite average alone drives most of the performance increase over the baseline. Subsequent variables yield diminishing returns, and the full ten-variable representation fails to increase discrimination proportionally. The supplementary marginal-gain analysis in Figure A1 (Appendix B) confirms this pattern. The R0–R1 transition generated ROC-AUC gains of 0.168, 0.159, and 0.171 for Statics, Mechanics of Solids, and Fluid Mechanics, respectively, while later transitions produced minimal or even negative gains.

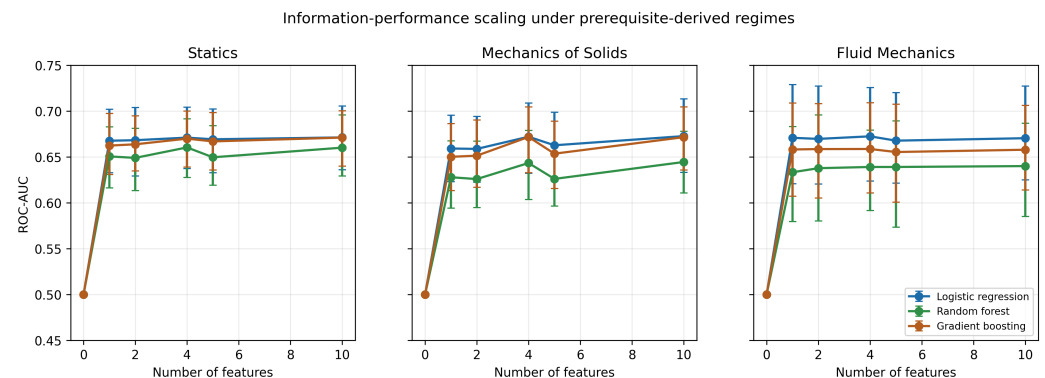


Figure 3. Predictive performance as a function of information dimensionality. ROC-AUC is shown across prerequisite-derived information regimes for each target course, with repeated-validation 95% empirical intervals.

This pattern reveals substantial redundancy within the full feature representation. Predictive performance does not scale linearly with dimensionality; instead, a low-dimensional summary preserves the core ranking signal. Moving from R1 to R5 increased ROC-AUC by just 0.004 points for Statics and 0.014 points for Mechanics of Solids. For Fluid Mechanics, the full representation provided no improvement over the minimal one. These results indicate that prerequisite records function as a compressible representation for information-constrained predictive inference.

We define this empirical pattern as predictive compressibility: severely reducing observed variables preserves most of the measurable prerequisite-based predictive signal. In this dataset, compressibility occurs because the one-dimensional prerequisite average captures nearly all the discrimination provided by the larger prerequisite representations.

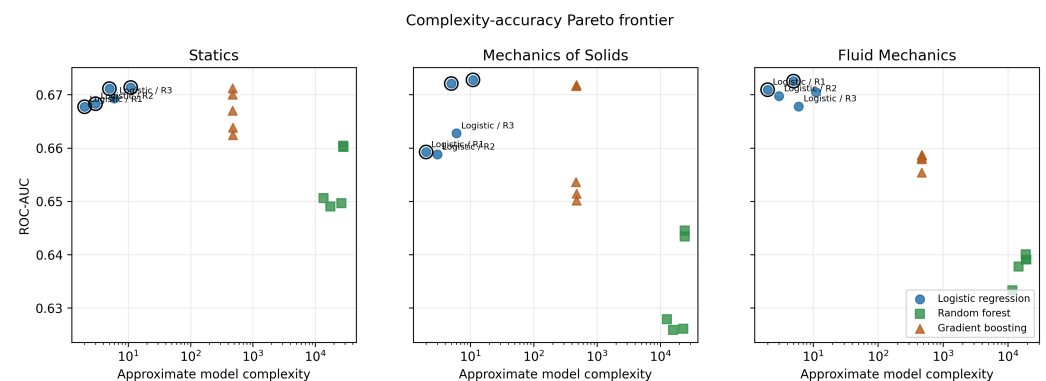
4.3. Minimal Predictive Subsets

The one-variable R1 model, utilizing only the average prerequisite grade, forms the smallest subset with meaningful predictive capacity. This transparent, sparse inference rule relies on a logistic curve with just two fitted parameters: an intercept (β_0) and a single coefficient for the prerequisite average (β_1). The fitted parameters for the R1 models are: $\beta_0 = -4.201$, $\beta_1 = 1.456$ for Statics; $\beta_0 = -3.286$, $\beta_1 = 1.344$ for Mechanics of Solids; and $\beta_0 = -2.547$, $\beta_1 = 1.419$ for Fluid Mechanics. Table 4 compares this minimal representation to the full ten-variable model, highlighting a 90% dimensionality reduction from R5 to R1 across all courses without significant degradation in predictive performance.

Table 4. Minimal subset performance relative to the full prerequisite-derived logistic model.

Course	Minimal Variable	ROC-AUC R1	ROC-AUC R5	Retention of ROC-AUC
Statics	Average prerequisite grade	0.668	0.671	99.5%
Mechanics of Solids	Average prerequisite grade	0.659	0.673	98.0%
Fluid Mechanics	Average prerequisite grade	0.671	0.670	100.1%

Figure 4 visualizes the complexity–accuracy Pareto frontier across model classes. The sparse logistic models occupy highly competitive regions despite relying on only two parameters in R1. Furthermore, complex random forest and gradient boosting models failed to dominate the sparse logistic alternatives in mean ROC-AUC. This demonstrates that the core predictive structure relies on neither high-dimensional features nor heavily nonlinear models.

**Figure 4.** Complexity–accuracy trade-off across information regimes and model classes. Circled points indicate Pareto-efficient configurations, and labeled points identify the sparse logistic regimes.

4.4. Inferential and Chronology-Controlled Robustness

To evaluate whether the small R1–R5 differences were stable across repeated validation folds, we computed paired fold-level comparisons using identical validation splits for both regimes. Table 5 reports the chronology-controlled first-attempt analysis using the original long-format academic records. The full R5 representation produced small systematic gains for Statics and Mechanics of Solids, but the mean paired ROC-AUC differences were only 0.003 and 0.008, respectively. For Fluid Mechanics, the mean paired difference was -0.001 and was not statistically significant after Holm adjustment. Thus, the paired tests show that some differences are statistically detectable because the comparisons are fold-paired, but the practical magnitude remains small relative to the moderate absolute ROC-AUC values.

Table 5. Paired R1–R5 comparisons for chronology-controlled first-attempt logistic models. Positive ROC-AUC differences favor R5; negative Brier differences favor R5.

Course	Metric	Mean R5–R1	Median R5–R1	95% Empirical Interval	Holm-Adjusted p	Effect Size
Statics	ROC-AUC	0.0033	0.0040	[−0.0190, 0.0197]	0.0020	0.411
Statics	Brier	−0.0008	−0.0008	[−0.0032, 0.0027]	<0.001	−0.522
Mechanics of Solids	ROC-AUC	0.0085	0.0087	[−0.0136, 0.0282]	<0.001	0.746
Mechanics of Solids	Brier	−0.0008	−0.0007	[−0.0030, 0.0016]	<0.001	−0.644
Fluid Mechanics	ROC-AUC	−0.0012	0.0019	[−0.0299, 0.0246]	0.734	−0.039
Fluid Mechanics	Brier	0.0002	0.0001	[−0.0008, 0.0013]	0.091	0.250

Figure 5 visualizes the paired ROC-AUC differences across the 100 validation folds. The distributions remain centered close to zero, reinforcing that R5 does not create a qualitatively different discrimination profile from R1.

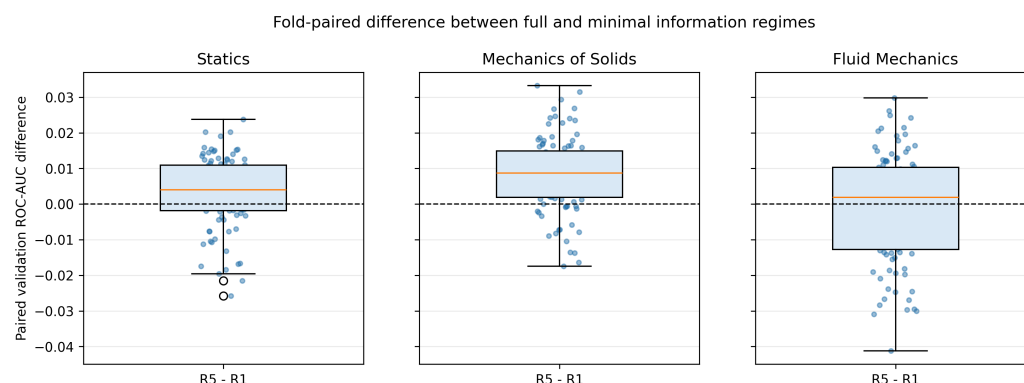


Figure 5. Fold-paired validation ROC-AUC differences between the full R5 and minimal R1 logistic models in the chronology-controlled first-attempt cohorts.

The chronology-controlled cohorts also address repeated-enrollment concerns. The raw academic records identified 4148 first Statics attempts, 3805 first Mechanics of Solids attempts, and 3667 first Fluid Mechanics attempts. After requiring all four prerequisites before the first target attempt, the usable cohorts contained 3704, 3484, and 3314 students, respectively. Repeated target enrollment was common in Statics and Mechanics of Solids but much less common in Fluid Mechanics (Table A2, Appendix B).

Table 6 compares R1 and R5 under logistic regression, random forest, and gradient boosting in the chronology-controlled cohorts. The nonlinear models did not overturn the early-saturation pattern. In Statics and Fluid Mechanics, R5 failed to improve over R1 for at least two model classes. In Mechanics of Solids, R5 improved ROC-AUC by 0.008–0.015 depending on model class, which is consistent with a small but nonzero amount of additional information in the granular prerequisite variables.

Table 6. Chronology-controlled R1 and R5 ROC-AUC across model classes.

Course and Model	R1 ROC-AUC	R5 ROC-AUC	R5–R1
Statics, logistic regression	0.640	0.644	0.004
Statics, random forest	0.622	0.616	−0.006
Statics, gradient boosting	0.636	0.634	−0.002
Mechanics of Solids, logistic regression	0.640	0.648	0.008
Mechanics of Solids, random forest	0.615	0.630	0.015
Mechanics of Solids, gradient boosting	0.631	0.645	0.014
Fluid Mechanics, logistic regression	0.651	0.650	−0.001
Fluid Mechanics, random forest	0.619	0.624	0.005
Fluid Mechanics, gradient boosting	0.637	0.632	−0.004

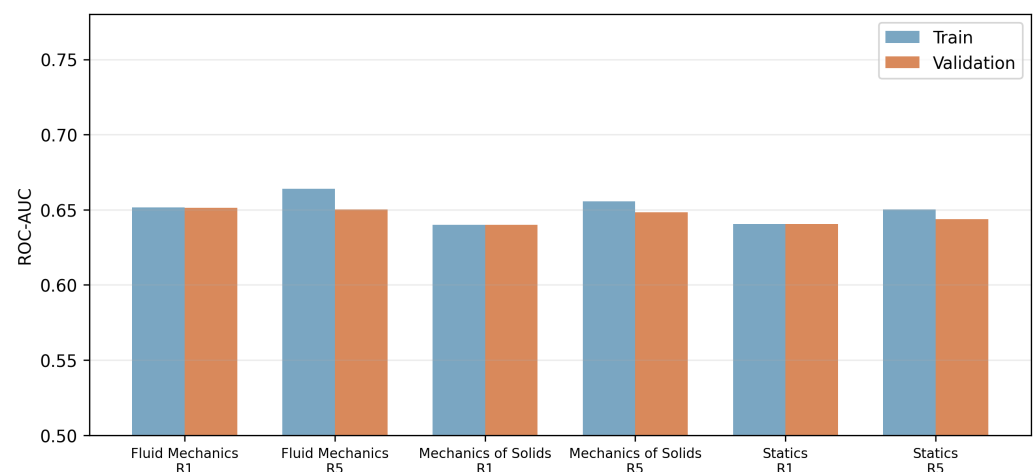
Finally, Table 7 evaluates target-year blocks. The R5–R1 ROC-AUC difference remained small in all nine course-period combinations, ranging from −0.009 to 0.019. These results do not eliminate the possibility of curriculum, grading, or cohort effects, but they show that the early-saturation pattern is not driven solely by one calendar period.

Table 7. R1 and R5 logistic ROC-AUC by target-year block in chronology-controlled first-attempt cohorts.

Course	Period	Students	Failures	R1 ROC-AUC	R5–R1 ROC-AUC
Statics	2007–2013	1153	487	0.623	−0.008
Statics	2014–2019	1361	528	0.582	0.019
Statics	2020–2025	1190	225	0.679	−0.009
Mechanics of Solids	2007–2013	919	282	0.665	0.005
Mechanics of Solids	2014–2019	1351	300	0.618	−0.004
Mechanics of Solids	2020–2025	1214	199	0.609	0.009
Fluid Mechanics	2007–2013	832	47	0.584	0.015
Fluid Mechanics	2014–2019	1312	174	0.700	−0.001
Fluid Mechanics	2020–2025	1170	117	0.633	0.006

Train-validation comparisons further clarify overfitting behavior. For the chronology-controlled logistic models, R1 showed near-zero ROC-AUC optimism across courses, whereas R5 produced small optimism values of 0.006, 0.007, and 0.014 for Statics, Mechanics of Solids, and Fluid Mechanics, respectively. Tree-based models, especially random forests, showed much larger train-validation gaps under R5. Figure 6 summarizes this pattern for logistic regression.

Train-validation consistency in chronology-controlled logistic models

**Figure 6.** Train-validation ROC-AUC consistency for chronology-controlled logistic models. The minimal R1 model shows little optimism, while R5 adds a small train-validation gap.

4.5. Calibration and Stability

Table A3 (Appendix B) and Figure 7 present the calibration results for the R1 logistic model. The Brier scores reached 0.205 for Statics, 0.166 for Mechanics of Solids, and 0.089 for Fluid Mechanics. Calibration slopes hovered near ideal, with repeated-fold means of 1.01, 1.01, and 1.02, respectively. Calibration intercepts closely approached zero, ranging from −0.049 to −0.006.

The empirical 95% confidence intervals for R1 ROC-AUC spanned 0.633–0.702 for Statics, 0.623–0.696 for Mechanics of Solids, and 0.621–0.729 for Fluid Mechanics. These intervals reflect moderate fold-to-fold variation, particularly in Fluid Mechanics, which has the lowest failure rate. Nevertheless, the calibration slopes consistently approximate one, indicating that low-information probabilities maintain coherent probabilistic behavior across repeated validation folds.

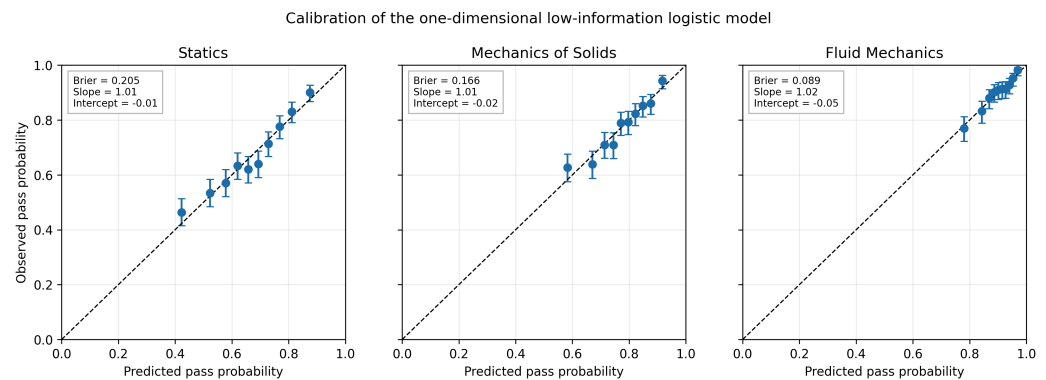


Figure 7. Calibration of the one-dimensional low-information logistic model. Points show quantile-bin empirical pass rates with Wilson 95% intervals; the dashed line denotes ideal calibration.

4.6. Computational Efficiency

Table 8 compares the one-variable R1 logistic model against the ten-variable R5 model. Because mean fold-level fit and prediction times remain extremely small for both models, absolute runtime differences lack practical significance. The primary computational benefit lies in the drastic reduction of observed dimensionality and fitted parameters.

Table 8. Computational efficiency of the minimal logistic model relative to the full prerequisite-derived logistic model. Times are mean seconds per validation fold.

Course	Features R1/R5	Fit R1	Fit R5	Predict R1	Predict R5
Statics	1/10	0.00125	0.12812	0.00016	0.00055
Mechanics of Solids	1/10	0.00171	0.02204	0.00022	0.00048
Fluid Mechanics	1/10	0.00119	0.02700	0.00015	0.00054

Moving from R5 to R1 reduces dimensionality by 90% and cuts logistic model complexity from 11 fitted parameters down to 2. Although R1 executed faster across all three courses, these absolute CPU timing differences are secondary since all logistic configurations fit and predict in milliseconds at this sample size. The true computational and systemic savings of the sparse R1 regime lie at the institutional deployment level. Specifically, predicting risk based solely on a student's rolling prerequisite average drastically reduces data engineering pipeline complexity, minimizes memory allocation, and lowers query execution costs over large institutional databases. It eliminates the need to compile, update, and join sparse, granular historical grade records, simplifying real-time tracking. The critical computational implication is the emergence of a transparent, scalable inference rule that achieves near-optimal predictive precision with a fraction of the data-infrastructure overhead.

4.7. Summary of Main Findings

Across all three courses, the one-dimensional prerequisite-average model retains meaningful predictive capacity despite severe information reduction. The initial shift from the no-information baseline to the minimal information regime drives the largest performance gain, while subsequent feature additions yield diminishing returns. The full prerequisite representation offers only marginal improvements for Statics and Mechanics of Solids, and provides no improvement for Fluid Mechanics. These results support predictive compressibility and feature redundancy within prerequisite-only records.

The sparse logistic model also achieves stable calibration, generating improved Brier scores and near-ideal calibration slopes. It has a favorable computational profile by minimizing both dimensionality and parameter count while preserving the core ranking signal. The primary limitation of this study is its strictly predictive and observational nature: even

with the chronology-controlled robustness analysis, the models do not identify causal mechanisms or evaluate intervention effectiveness. Ultimately, these findings provide evidence for predictive compressibility within prerequisite-based academic risk modeling.

5. Discussion

Predictive behavior evolved non-linearly as the prerequisite feature space expanded. Across all analyzed courses, transitioning from the no-information baseline (R_0) to the minimal-information model (R_1) produced the largest increase in predictive discrimination. Subsequent feature additions yielded rapidly diminishing marginal returns. Relative to the full representation (R_5), these results demonstrate early predictive saturation. Once models incorporated the average prerequisite grade, adding specific granular variables produced only small changes in student rankings. These experiments indicate that, for the analyzed engineering courses, much of the measurable prerequisite-based predictive signal resides within a compact, low-dimensional subset of academic history.

The marginal-gain analysis supports this structural redundancy. The average prerequisite grade captured the majority of the prerequisite-only predictive signal, making additional prerequisite descriptors only weakly incremental. Computationally, this indicates a high degree of feature overlap within the full prerequisite representation. At the same time, the paired tests show that some R_1 – R_5 differences are statistically detectable because the same validation folds are compared directly. The important point is therefore not that R_1 and R_5 are identical, but that the practical gain from R_5 is small relative to the 90% dimensionality reduction achieved by R_1 .

The contrast between continuous and threshold-dependent metrics highlights a critical methodological distinction. ROC-AUC and Brier scores remained stable across information regimes, confirming that heavily reduced models successfully rank students and maintain coherent probabilistic calibration. Conversely, threshold-dependent metrics like balanced accuracy degraded under information reduction, particularly in courses with high approval rates. This divergence occurs because ROC-AUC evaluates continuous risk separation, whereas balanced accuracy evaluates hard classification at an arbitrary 0.5 probability threshold. In highly imbalanced academic datasets, stable probabilistic ranking provides a far more reliable measure of predictive utility than rigid pass/fail classification. Consequently, sparse models reliably prioritize at-risk students and maintain probabilistic calibration despite threshold-induced classification artifacts.

The practical implications of these findings are substantial but should be interpreted cautiously: academic risk estimation remains feasible under severe data constraints. Many institutions lack advanced learning analytics infrastructure, granular behavioral logs, or continuous data pipelines. For these institutions, sparse predictive models utilizing readily available prerequisite records offer a deployable alternative to data-intensive architectures. By showing that simple prerequisite summaries contain most of the measurable prerequisite-based signal, this study lowers the technical barriers to transparent early-warning triage. However, the absolute ROC-AUC values remain moderate, so these models should support prioritization and monitoring rather than replace comprehensive academic advising or richer diagnostic assessment.

From an educational perspective, the value of the sparse model is not that it explains why a student is at risk, but that it provides a transparent screening signal that can be linked to familiar curricular milestones. A low prerequisite average can prompt instructors, program coordinators, or advisors to review prior conceptual weaknesses, recommend remedial support, monitor early assessments more closely, or prioritize students for tutoring and office-hour outreach. The model should therefore be understood as a triage layer for

allocating attention under limited information, not as a substitute for pedagogical diagnosis, student interviews, or course-specific assessment.

This deployability intrinsically pairs with high interpretability. Sparse models trace risk predictions directly back to recognizable academic milestones, making them transparent and actionable for educators. In our experiments, this transparency came with only small losses in predictive accuracy when compared to the full ten-variable representation. Ultimately, this study shows that prerequisite-based academic risk estimation does not necessarily require maximal prerequisite detail; a carefully selected low-dimensional summary can preserve much of the available signal.

Several limitations frame the scope of these findings:

- **Dataset Specificity:** The observed compressibility is tied directly to the analyzed engineering courses and the specific institutional grading structure. The 20-year span incorporates multiple cohorts, instructors, teaching styles, student backgrounds, and pandemic-era instructional conditions, which increases heterogeneity but also introduces unmodeled temporal variation.
- **External Validity:** The empirical evidence comes from one Civil Engineering program at a single institution and uses only prerequisite-derived academic records. The observed compressibility should therefore be interpreted as a property of this prerequisite-based dataset, not as a general property of academic risk prediction. External validation across institutions, academic programs, grading systems, and richer educational datasets is required before broader generalization.
- **Observational Scope:** The analysis is strictly predictive, not causal. The models estimate risk probabilities but do not identify the underlying causes of academic failure.
- **Curricular Dependence:** Informational redundancy likely depends on institutional curriculum design, grading policies, and prerequisite sequencing, which vary significantly across universities.
- **Metric Sensitivity:** Class imbalance heavily skewed threshold-dependent metrics, isolating the degradation primarily to balanced accuracy rather than probabilistic ranking.
- **Feature Restrictions:** The feature space was deliberately restricted to prerequisite academic records. Integrating behavioral, demographic, or high-frequency LMS data could alter the observed compressibility dynamics and reveal non-redundant signals.
- **Withdrawal Outcomes:** The university grade spreadsheet stores grades only for completed course attempts and does not include withdrawal records or a withdrawal-status field. Consequently, the analysis distinguishes approval from failure among completed attempts, but it does not model course withdrawal as a separate academic outcome.
- **Assessment Aggregation:** The institutional records provide final course grades only. If Fluid Mechanics included group-project or peer-evaluation components in some terms, those components are embedded in the official individual final grade and cannot be separated from individual assessment components in the present data.
- **Temporal and Enrollment Effects:** The chronology-controlled robustness analysis used first target attempts and prerequisite grades completed before those attempts, but it does not fully model instructor changes, curriculum revisions, grading-policy shifts, pandemic-era remote instruction, equivalent online-course completion, or unobserved student-context differences over time.

Consequently, these findings document an empirical phenomenon rather than a universal law. Within this specific academic context, sparse prerequisite summaries capture most of the measurable prerequisite-based predictive power, making broader prerequisite representations only weakly incremental. Verifying whether this precise pattern of predictive compressibility extends to different institutional environments, alternative grading

systems, other academic programs, and broader multimodal datasets remains an important avenue for future research.

6. Conclusions

This study analyzed academic risk prediction under nested information regimes using prerequisite data from three Civil Engineering courses. The results demonstrate a clear pattern of predictive compressibility within prerequisite-only records: severely reducing the feature space preserves most of the measurable predictive signal. Transitioning from a no-information baseline to a minimal, one-dimensional model (the average prerequisite grade) drove the largest predictive gain. Subsequent feature additions provided rapidly diminishing returns, revealing substantial structural redundancy within broader prerequisite representations.

Crucially, academic risk estimation remains feasible under severe data constraints. Despite a 90% reduction in dimensionality, sparse logistic models retained between 98.0% and 100.1% of the full-model ROC-AUC and maintained stable probabilistic calibration. Chronology-controlled robustness checks using first target-course attempts showed that the full representation produced only small practical gains over the one-dimensional model, and nonlinear models did not overturn the early-saturation pattern. Institutions lacking advanced learning analytics infrastructure or high-dimensional behavioral datasets can therefore use low-dimensional academic summaries as transparent triage tools for prioritizing advising, tutoring, and early monitoring, while recognizing that the moderate ROC-AUC values do not support replacing richer diagnostic or advising processes.

Nevertheless, these findings document an empirical phenomenon tied to a specific institutional context and a prerequisite-only feature space. The analysis remains strictly predictive, uses final course grades rather than disaggregated assessment components, excludes withdrawals because the university spreadsheet stores grades only for completed course attempts, and cannot fully isolate instructor, curriculum, group-assessment, repeated-enrollment, online-modality, or pandemic-era effects. Threshold-dependent classification metrics were also highly sensitive to class imbalance in high-approval courses. Future research must externally validate whether this specific pattern of predictive compressibility extends across different universities, alternative grading policies, academic programs, and broader multimodal datasets before it can be treated as a general property of academic risk modeling.

Author Contributions: Conceptualization, O.A., J.M.B. and J.V.; methodology, O.A., J.M.B. and J.V.; software, O.A.; validation, O.A., J.M.B. and J.V.; formal analysis, O.A., J.M.B. and J.V.; investigation, O.A., J.M.B. and J.V.; resources, O.A. and J.M.B.; data curation, O.A.; writing—original draft preparation, O.A.; writing—review and editing, O.A., J.M.B. and J.V.; visualization, O.A., J.M.B. and J.V.; project administration, J.M.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors thanks the support of Vicerrectoría de Investigación y Extensión at Universidad Industrial de Santander through the grant code 4608 of year 2025.

Data Availability Statement: Data of this study is available upon reasonable request to the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1 summarizes several key aspects of the studies discussed in the Introduction and the Literature Review section. For each study, information on the dataset employed, the methodology adopted, and the algorithm used is provided. In addition, the main findings and the principal strengths of each approach are highlighted. The following abbreviations

viations are used in the table: ADA/AdaBoost: Adaptive Boosting, ANN: Artificial Neural Network, AODE: Aggregating One-Dependence Estimators, BART: Bayesian Additive Regression Trees, BiLSTM: Bidirectional Long Short-Term Memory, BN: Bayesian Network, BPNN: Backpropagation Neural Network BRB: Belief Rule Base, CART: Classification and Regression Tree CHIMI: A feature selection method combining Chi-square and Mutual Information based on ranked vector scores, CNN: Convolutional Neural Network, Conv1D: One-dimensional Convolutional Neural Network, CTGAN: Conditional Tabular Generative Adversarial Networks, DLBRB-i: Dual-level Progressive Classification Belief Rule Base, DT: Decision Tree, FCN: Fully Convolutional Network, FFNN: Feedforward Neural Network, GAFS: Genetic Algorithm based Feature Selection, GBM: Gradient Boosting Machine, GBR: Gradient Boosting Regressor, GBT: Gradient Boosting Tree, GraphSMOTE: Graph-based Synthetic Minority Over-sampling Technique for imbalanced node classification, IDBN: Improved Deep Belief Network, InceptionTime: A deep learning architecture designed for time-series classification using parallel convolutional pathways, J48: A Java implementation of the C4.5 decision tree algorithm, KNN/k-NN: K-Nearest Neighbors, LightGBM: Light Gradient Boosting Machine, LR: Logistic Regression, LSTM: Long Short-Term Memory, MCNN: Multiscale Convolutional Neural Network that extracts features across different time scales in parallel, MICHI: A feature selection method combining Mutual Information and Chi-square algorithms based on ranked feature scores, MLP: Multilayer Perceptron neural networks, NB: Naive Bayes Opt-MLP: Optimized Multilayer Perceptron (fine-tuned learning parameters such as learning rate and momentum), PCA: Principal Component Analysis, PCR: Principal Component Regression (regression analysis on extracted principal components), PLR: Probabilistic Logistic Regression (quantifies uncertainty in model parameters and predictions), RF: Random Forest, RFBT-RF: Relief-F and Budget Tree Random Forest, Rocket: A time-series classification algorithm used with multivariate centrality measures, SEPN: Sequential Engagement Based Prediction Network (CNN-LSTM architecture), SMO: Sequential Minimal Optimization (used for training Support Vector Machines), SVM: Support Vector Machine, TabDDPM: Tabular Denoising Diffusion Probabilistic Models, TabDDPM-SMOTE: Extension of TabDDPM that includes SMOTE to handle class imbalance, TimeCNN: A Convolutional Neural Network architecture specialized for compact baseline time-series processing, TVAE: Tabular Variational Autoencoder, XGB / XGBoost: Extreme Gradient Boosting.

Table A1. Summary of reviewed studies on academic performance prediction.

Ref.	Year	Dataset(s) Used	Methodology or Model	Key Findings	Strength
[9]	2019	UCI Machine Learning Repository database (Maths/Portuguese) and school records	Hybrid FS (RFBT-RF) with DT, NB, SVM, KNN	RFBT-RF with DT achieved 97.88% accuracy on school records and 88.60% on UCI Maths	Reduces dimensionality and handles uncertainty effectively.
[10]	2020	Survey data from Cambodian high school: ADS and GDS4	Proposed CHIMI FS with KNN, Hybrid RF, IDBN	CHIMI improved accuracy by 1–5%; Hybrid RF reached 99.98% accuracy	Novel algorithm validated on large artificial datasets.
[11]	2020	14 attributes for 9458 students from RMS University in Thailand	Misclassified instance removal and SMOTE	Random Forest achieved 94.7% precision/recall after cleaning training data	Handles class imbalance and noisy instances.

Table A1. Cont.

Ref.	Year	Dataset(s) Used	Methodology or Model	Key Findings	Strength
[18]	2020	Academic performance and 15 attributes of 480 students from Kaggle	Genetic Algorithm FS with RF, NB, DT, KNN	RF with GA FS achieved 82.29% accuracy, outperforming baseline RF	Overcomes local optima using GA global search.
[2]	2020	Survey data from Cambodian high schools	Academic performance prediction system using Hybrid RF and MICHI FS	Hybrid RF with MICHI dominant set achieved 99.98% accuracy	Delivers a real-time web application for institutional intervention.
[15]	2021	Records of 151 students at Al-Buraimi University College, Oman	Undersampling and SMOTE oversampling	SMOTE yielded 91% average F-measure	Shows that high accuracy does not ensure minority-class detection.
[6]	2021	OULAD dataset	SEPN using CNN-LSTM architecture	Engagement detector improved recall by 6% and accuracy by 10%	Captures sequential continuity and periodic features in activities.
[28]	2022	2773 high-school students in an urban U.S. district	Multilevel logistic regression	Failing a single marking period in 8th grade doubles non-graduation odds	Provides early warning indicators after one marking period.
[21]	2023	xAPI dataset	FFNN with variance-based feature selection	Behavioral attributes reached 91.95% accuracy using four features	Uses simple and robust teacher-extractable features.
[26]	2023	Moodle data from 517 students	Five-stage Probabilistic Logistic Regression	Week 6 was optimal for early identification, with 92.81% accuracy	Quantifies prediction uncertainty.
[5]	2023	Kalboard 360 and GRIET datasets	J48, NB, RF, and fine-tuned Opt-MLP	RF achieved 100% accuracy on Kalboard 360 when behavioral features were included	Validates results using local and public datasets.
[1]	2024	N.A.	Systematic literature review	Decision Trees are common due to accuracy and interpretability	Provides an overview of popular models and attributes.
[12]	2024	OULAD, course CCC	XGBoost with SHAP and SMOTE	XGBoost yielded 95.22% accuracy; SHAP identified assessment scores as key drivers	Integrates Explainable AI.
[16]	2024	48,675 University of Jordan records	Conv1D versus LSTM/BiLSTM	Conv1D achieved 94% g-mean in excellence prediction	Uses a large-scale dataset and imbalance analysis.
[23]	2024	591 students from UNEMI, Ecuador	LR, RF, SVM, Neural Networks	Logistic Regression achieved the highest AUC-ROC of 0.9354	Shows that simple models can be practical for educators.

Table A1. Cont.

Ref.	Year	Dataset(s) Used	Methodology or Model	Key Findings	Strength
[29]	2024	Informatics Engineering and Information Technology students	Decision Trees with integrated data sources	Predicted high-risk failure with 100% accuracy on second tests	Identifies students requiring additional support.
[3]	2024	323 students and 10 instructors participating in discussion forums	Rocket and K-neighbours time-series classifiers	Forum network structure did not reliably inform SaR predictions	Captures the socially regulated nature of learning.
[31]	2024	356 students in C programming	Dual-level Belief Rule Base	Accuracy improved to 98.13% through a second classification layer	Maintains explainability and diagnostic accuracy.
[4]	2024	Data from five campus systems	RF, XGBoost, and BP Neural Network	Accurate five-level warnings through holistic behavioral integration	Aggregates information from multiple institutional systems.
[13]	2025	9262 students from Surabaya State University	RF, XGBoost, LightGBM, SVM, KNN, NB, LR	LightGBM consistently outperformed all other models	Integrates academic and socioeconomic factors.
[14]	2025	4093 records from STIEM Bongaya	RF, SVM-RBF, XGBTree, KNN, NB, ANN, AdaBoost	RF with RUS-ROS resampling achieved the best average score of 0.948	Uses rigorous PERMANOVA statistical testing.
[17]	2025	OULAD, course CCC	TVAE, CTGAN, TabDDPM, TabDDPM-SMOTE	TabDDPM preserved distributions best; TabDDPM-SMOTE optimized XGB	Applies advanced generative models to data imputation.
[20]	2025	161 reviewed studies	Systematic literature review	Deep Learning has surged since 2017; transparency remains a challenge	Provides a robustness and interpretability checklist.
[30]	2025	1266 students in Chile	GBR, RF, AdaBoost, KNN, LR	GBR achieved the highest performance, with $R^2 = 0.6109$	Uses characterization surveys to create risk profiles.
[7]	2025	6798 STEM students from the University of Sharjah	RF, GBM, XGBoost, CART, KNN, LR	All models achieved AUC-ROC > 0.91 and F1 > 0.84	First SHAP-based application to SIS data in the MENA region.
[24]	2026	SEA dataset	Stability-centric framework and ensembles	Mean accuracy of 0.901 and macro-F1 of 0.847	Addresses over-optimistic single-run bias through repeated cross-validation.

Appendix B

Table A2 reports the chronology-controlled first-attempt cohorts reconstructed from the raw long-format academic records.

Table A2. Chronology-controlled first-attempt cohorts reconstructed from the raw long-format academic records.

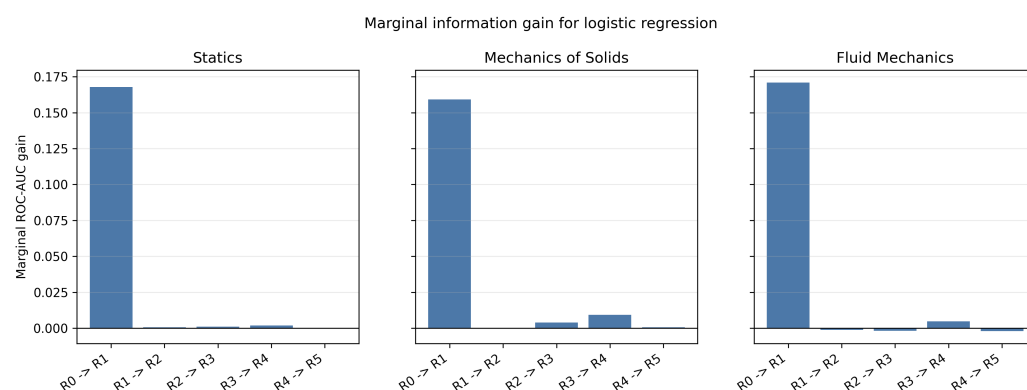
Course	First Target Students	Chronology-Controlled Students	Approved	Failed	Pass Rate	Repeated Target Enrollment
Statics	4148	3704	2464	1240	0.665	1336 (32.2%)
Mechanics of Solids	3805	3484	2703	781	0.776	874 (23.0%)
Fluid Mechanics	3667	3314	2976	338	0.898	321 (8.8%)

Table A3 summarizes calibration and repeated-validation stability for the one-dimensional logistic model.

Table A3. Calibration and repeated-validation stability of the one-dimensional logistic model.

Course	Brier	Slope	Intercept	ROC-AUC SD	F1 SD
Statics	0.205	1.01	−0.006	0.018	0.007
Mechanics of Solids	0.166	1.01	−0.016	0.021	0.002
Fluid Mechanics	0.089	1.02	−0.049	0.027	0.0004

Figure A1 reports the marginal contribution of successive information regimes. It is included as supplementary visual evidence because the same early-saturation pattern is already summarized in the main text by Figure 3 and the corresponding ROC-AUC differences.

**Figure A1.** Marginal contribution of successive information regimes for logistic regression. The largest gain occurs when moving from the no-information baseline to the one-dimensional prerequisite-average model.

References

1. Meka, N.K.; Veeranjanyulu, N. A Review on Student Performance Prediction Based on Machine Learning Techniques. In Proceedings of the 2024 International Conference on Social and Sustainable Innovations in Technology and Engineering, SASI-ITE 2024, Tadepalligudem, India, 23–25 February 2024; pp. 307–310. [\[CrossRef\]](#)
2. Sokkhey, P.; Okazaki, T. Developing web-based support systems for predicting poor-performing students using educational data mining techniques. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 23–32. [\[CrossRef\]](#)
3. Flores, N.G.L.; Uc-Cetina, V.; Islind, A.S.; Óskarsdóttir, M. Threads of Complexity: Lessons learnt from Predicting Student Failure through Discussion Forums' Social-Temporal Dynamics. In Proceedings of the International Conference on Educational Data Mining, Atlanta, GA, USA, 14–17 July 2024; pp. 978–981. [\[CrossRef\]](#)
4. Li, X.; Li, C. Research on Construction of Student Academic Early Warning Model Based on Ensemble Learning. *Commun. Comput. Inf. Sci.* **2024**, *1899*, 217–228. [\[CrossRef\]](#)
5. Nayak, P.; Vaheed, S.; Gupta, S.; Mohan, N. Predicting students' academic performance by mining the educational data through machine learning-based classification model. *Educ. Inf. Technol.* **2023**, *28*, 14611–14637. [\[CrossRef\]](#)

6. Song, X.; Li, J.; Sun, S.; Yin, H.; Dawson, P.; Doss, R.R.M. SEPN: A Sequential Engagement Based Academic Performance Prediction Model. *IEEE Intell. Syst.* **2021**, *36*, 46–53. [\[CrossRef\]](#)
7. Al Hashmi, R.A.M.; Zervopoulos, P.D.; Elmehdi, H.M.; Ozturk, I. Predicting Dropout in MENA STEM Higher Education Using Explainable AI: A Machine Learning Approach. *Emerg. Sci. J.* **2025**, *9*, 268–286. [\[CrossRef\]](#)
8. Yáñez, A.; Crawford, B.; Monfroy, E.; Paz, Á.; Soto, R.; Barrera-García, J.; Cisternas-Caneo, F.; López Cortés, B. A New Institutional Dataset Proposal for Timely Graduation in Chilean Higher Education. *Commun. Comput. Inf. Sci.* **2026**, *2702*, 339–351. [\[CrossRef\]](#)
9. Deepika, K.; Sathyanarayana, N. Relief-F and budget tree random forest based feature selection for student academic performance prediction. *Int. J. Intell. Eng. Syst.* **2019**, *12*, 30–39. [\[CrossRef\]](#)
10. Sokkhey, P.; Okazaki, T. Study on dominant factor for academic performance prediction using feature selection methods. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 492–502. [\[CrossRef\]](#)
11. Nuankaew, W.; Thongkam, J. Improving Student Academic Performance Prediction Models using Feature Selection. In Proceedings of the 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2020, Phuket, Thailand, 24–27 June 2020; pp. 392–395. [\[CrossRef\]](#)
12. Choi, W.C.; Lam, C.T.; Mendes, A.J. Enhance Learning Performance Predictions with Explainable Machine Learning. In Proceedings of the Frontiers in Education Conference, FIE, Washington, DC, USA, 13–16 October 2024. [\[CrossRef\]](#)
13. Yustanti, W.; Putra, R.E.; Gusti Putu Asto Buditjahjanto, I.; Nurhidayat, A.I. Exploring Ensemble Classifiers and Filter-Based Feature Selection for Predicting On-Time Graduation Using Multidimensional Student Data. In Proceedings of the 2025 8th International Conference on Vocational Education and Electrical Engineering: Shaping a Sustainable Future with Green Innovation and Industry Collaboration for Education and Intelligent Technology Advancements, ICVEE 2025, Surabaya, East Java, Indonesia, 24–25 September 2025; pp. 246–251. [\[CrossRef\]](#)
14. Bakri, R.; Alam, S.; Astuti, N.P.; Bakhtiar, M.I. Optimizing Machine Learning Models for Graduation on Time Prediction: A Comparative Study with Resampling and Hyperparameter Tuning. *J. Online Inform.* **2025**, *10*, 270–285. [\[CrossRef\]](#)
15. Khan, I.; Ahmad, A.R.; Jabeur, N.; Mahdi, M.N. Minimizing Classification Errors in Imbalanced Dataset Using Means of Sampling. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 13051, pp. 435–446. [\[CrossRef\]](#)
16. Alshamaila, Y.; Alsawalqah, H.; Aljarah, I.; Habib, M.; Faris, H.; Alshraideh, M.; Salih, B.A. An automatic prediction of students' performance to support the university education system: A deep learning approach. *Multimed. Tools Appl.* **2024**, *83*, 46369–46396. [\[CrossRef\]](#)
17. Choi, W.C.; Lam, C.T.; Mendes, A.J. Comparison of Data Imputation Performance in Deep Generative Models for Educational Tabular Missing Data. In Proceedings of the International Conference on Educational Data Mining, Palermo, Italy, 20–23 July 2025; pp. 133–142. [\[CrossRef\]](#)
18. Farissi, A.; Dahlan, H.M.; Samsuryadi. Genetic algorithm based feature selection for predicting student's academic performance. *Adv. Intell. Syst. Comput.* **2020**, *1073*, 110–117. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Rajalaxmi, R.; Valarr, A. Feature selection using hybrid GA-BPSO for engineering students' performance prediction. *Int. J. Appl. Eng. Res.* **2014**, *9*, 19945–19958.
20. Shi, H.; Zhang, N.; Caskurlu, S.; Na, H. Applications of Machine Learning for at-Risk Student Prediction in Online Education: A 10-Year Systematic Review of Literature. *J. Comput. Assist. Learn.* **2025**, *41*, e70058. [\[CrossRef\]](#)
21. Yurtkan, K.; Adalier, A.; Tekgu, U. Student Success Prediction Using Feedforward Neural Networks. *Rom. J. Inf. Sci. Technol.* **2023**, *26*, 121–136. [\[CrossRef\]](#)
22. Chaudhury, P.; Tripathy, H.K. Assessment of Academic Performance of Students Using Classification and Feature Analysis Techniques. *Stud. Comput. Intell.* **2025**, *1212*, 327–349. [\[CrossRef\]](#)
23. Arévalo-Cordovilla, F.E.; Peña, M. Comparative Analysis of Machine Learning Models for Predicting Student Success in Online Programming Courses: A Study Based on LMS Data and External Factors. *Mathematics* **2024**, *12*, 3272. [\[CrossRef\]](#)
24. Almuniri, I.S.; Alhussian, H.; Aziz, N.; Khairy, S.O.F.; AlAbri, A.S.; Jarallah, Z.F.; Yahaya, S.; Adamu, S. Beyond peak accuracy: A stability-centric framework for reliable multimodal student engagement assessment. *Sci. Rep.* **2026**, *16*, 5. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Abdo, W.; Hefny, H.; Darwish, N. Enhancing Student Performance Prediction: Stacking Ensemble with Feature Engineering and SHAP Interpretability. *Int. J. Intell. Eng. Syst.* **2025**, *18*, 419–436. [\[CrossRef\]](#)
26. Nimy, E.; Mosia, M.; Chibaya, C. Identifying At-Risk Students for Early Intervention—A Probabilistic Machine Learning Approach. *Appl. Sci.* **2023**, *13*, 3869. [\[CrossRef\]](#)
27. Effland, A.; Kobler, E.; Pereyra, M.; Peter, J. Uncertainty Estimation for Learning-Based Classification of Corrupted Images. *Siam J. Imaging Sci.* **2025**, *18*, 1828–1856. [\[CrossRef\]](#)
28. Marshall, D.T. Course grades as actionable early warning indicators: Predicting high school graduation in an urban school district. *Prev. Sch. Fail.* **2022**, *66*, 206–213. [\[CrossRef\]](#)

29. De Oliveira, C.F.; Sobral, S.R.; Ferreira, M.J.; Moreira, F. Interpretable Success Prediction in a Computer Networks Curricular Unit Using Machine Learning. In Proceedings of the Procedia Computer Science, Funchal, Portugal, 13–15 November 2024; Volume 239, pp. 598–605. [[CrossRef](#)]
30. Saavedra-Acuna, C.; Quezada-Espinoza, M.; Gomez, D.A. A Predictive Model for Academic Performance in Engineering Students. In Proceedings of the 2025 ASEE Annual Conference and Exposition, Conference Proceedings, Montreal, QC, Canada, 22–25 June 2025. [[CrossRef](#)]
31. Mai, J.; Wei, F.; He, W.; Huang, H.; Zhu, H. An Explainable Student Performance Prediction Method Based on Dual-Level Progressive Classification Belief Rule Base. *Electronics* **2024**, *13*, 4358. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.