Estimating the Risk of Failing Physics Courses Through Machine Learning and Monte Carlo Simulation

Isaac Caicedo-Castro*†‡ Rubby Castro-Púche†§ Samir Castaño-Rivera*‡
*Socrates Research Team
†Research Team: Development, Education, and Healthcare
‡Department of Systems and Telecommunications Engineering
§Department of Social Sciences
University of Córdoba
Carrera 6 No. 76-305, 230002, Montería, Colombia
e-mail: {isacaic | rubycastro | sacastano}@correo.unicordoba.edu.co

***Abstract*—This research is conducted within the Systems Engineering undergraduate program at the University of Córdoba (Colombia) with the aim of estimating the risk of failing physics courses, which are widely regarded as particularly challenging for students. At this university, each semester is divided into three sessions, each contributing equally to the final course grade. Our objective is to estimate the probability of course failure based on student performance in the early sessions. To this end, we compiled a dataset containing session grades and final outcomes for students enrolled in Physics I, II, and III during the 2024 and 2025 academic years. A logistic regression model was employed to estimate the probability that a student fails a course given their grades in the first two sessions. Using the predictive function obtained from the fitted model, we subsequently conducted a Monte Carlo simulation to estimate both absolute and relative risks of course failure across a broader range of academic scenarios. The results suggest that poor performance in early sessions is strongly associated with an increased probability and absolute risk of course failure, particularly in Physics I and Physics III. In several scenarios, the estimated risk of failure approaches critical levels when students fail multiple early assessments. These findings highlight the importance of early academic performance as a predictor of final outcomes and provide evidence that might support instructors and academic administrators in designing early interventions aimed at reducing course failure and improving student success.**

Keywords-machine learning; logistic regression; education; Monte Carlo method; absolute and relative risk.

I. Introduction

Nowadays, academic success is a major concern for universities worldwide. Consequently, identifying effective strategies and conducting research to predict the risk of academic failure has become an active area of study within social computing, particularly through educational data mining approaches. This research builds upon a previous study [1], extending it through the use of machine learning methods and a larger dataset.

Educational data mining methods are commonly utilised to predict student dropout, delayed graduation [2][3][4], and the likelihood of course failure or withdrawal [5][6][7][8][9][10][11][12][13][14].

Our research is centred on the academic context of the University of Córdoba in Colombia, where each academic semester is split into three sessions, each lasting six weeks and contributing equally to the final grade. The final course grade is calculated as the mean of the student's grades across the three sessions. Within each session, no single assessment exceeds 40% of the session grade, meaning that each student undergoes at least nine evaluations during a semester.

This structure is designed to achieve several pedagogical goals: reducing the pressure of final exams, diversifying assessment strategies, encouraging consistent study habits, enabling continuous monitoring of learning progress, facilitating early interventions, and providing timely support to students. This approach is supported by several educational theories and instructional strategies, including constructivism, formative assessment, multiple intelligences theory, cognitive load theory, active learning, and outcome-based education.

According to constructivist theory, students build their understanding through interaction with their environment. Frequent evaluations help instructors monitor this evolving understanding and adjust teaching strategies accordingly.

Formative assessment emphasises the use of ongoing evaluations throughout the instructional period to monitor learning, identify challenges, and guide teaching. This method offers continuous feedback to both students and instructors, aligning well with the university's evaluation strategy.

The theory of multiple intelligences posits that students possess diverse talents and learning preferences. A variety of assessments throughout the semester provides a more inclusive way to evaluate these varied strengths.

Cognitive load theory suggests that students learn more effectively when information is presented in manageable segments. Multiple evaluations distributed over time align with this principle by reducing cognitive overload.

Active learning promotes student engagement through problem-solving, discussions, and hands-on activities. Multiple evaluations throughout the semester may reinforce this approach by encouraging students to actively engage with the thorough material.

At the University of Córdoba, Outcome-Based Education (OBE) is the foundational approach. It emphasises clearly defined learning outcomes and assessments aligned with those outcomes. Dividing the semester into multiple sessions allows for a more granular alignment of evaluations with specific goals.

The university's OBE model is integrated with the Structure of the Observed Learning Outcomes (SOLO) taxonomy, which categorises learning into five levels:

1) Prestructural (0.0–2.0): The student has not yet grasped the key concepts.
2) Unistructural (2.1–2.9): The student understands a single aspect of the task.
3) Multistructural (3.0–3.7): The student understands several aspects, but without integration.
4) Relational (3.8–4.5): The student can integrate multiple aspects meaningfully.
5) Extended Abstract (4.6–5.0): The student demonstrates deep understanding and applies concepts to new contexts.

To pass an evaluation, a student must achieve at least the multistructural level, corresponding to a grade above 3.0.

By structuring learning outcomes around SOLO taxonomy principles, the curriculum offers a coherent and progressively challenging learning experience. This structure emphasises the development of deeper understanding as students progress. However, despite this pedagogical framework, physics courses remain particularly challenging for systems engineering students, who often struggle to reach the relational or extended abstract levels. For example, during 2025 and 2024, the average final grades for Physics I, II, and III were 3.42, 3.55, and 3.73, respectively, suggesting limited integration of concepts or application to real-world contexts.

In an effort to mitigate course failure and student dropout, the university considers students to be at risk of failing a course if they fail either of the first two sessions. This institutional criterion motivates the following research questions:

- What is the risk of failing a physics course if a student fails the first session?
- What is the risk if a student fails the second session?
- What is the risk if a student fails the first and the second sessions?

These questions were previously addressed in [1] by simulating a wide range of grade scenarios across the three sessions using the Monte Carlo method. In contrast, in our research, a predictive model is fitted using a larger dataset to estimate the probability of failing a course given a student's grades in the first two sessions. We then employ the Monte Carlo method to simulate a wide range of grade scenarios across these sessions in order to analyse the research questions posed above.

The Monte Carlo method has been applied in related educational contexts, for instance, to evaluate curriculum effectiveness [15] and to estimate students' motivation when learning scientific computing [16][17]. As far as we know, no prior research has addressed these questions using the methodological approach adopted in this study.

The results of our simulations suggest the following findings:

- For Physics I, approximately 62 out of 100 students are likely to fail the course. Among students who failed the first session, about 81 out of 100 are at risk of failing the course; this risk increases to 83 out of 100 for those who failed the second session, and to 99 out of 100 for those who failed both sessions.
- For Physics II, approximately 59 out of 100 students are likely to fail the course. Among students who failed the first session, about 72 out of 100 are at risk of failing; this increases to 85 out of 100 for those who failed the second session, and to 99 out of 100 for those who failed both sessions.
- For Physics III, approximately 52 out of 100 students are likely to fail the course. Among students who failed the first session, about 71 out of 100 are at risk of failing; this increases to 73 out of 100 for those who failed the second session, and to 95 out of 100 for those who failed both sessions.

These insights contribute to implement early intervention strategies and improve academic support in physics courses.

Finally, the rest of this paper is outlined as follows: in Section II, we present the research and simulation methodology adopted in this research, while we present and discuss the results in Section III. The article concludes in Section IV.

II. Research Methodology

We adopted a quantitative approach, collecting the session and final grades of 611 students enrolled in physics courses at the University of Córdoba between 2024 and 2025. Specifically, 280 students were enrolled in Physics I, 193 in Physics II, and 138 in Physics III. The relatively small dataset size reflects the recent implementation of the previously described Outcome-Based Education (OBE) framework at the institution.

Let $\mathcal{D}$ denote the dataset used in this research, defined as follows: $\mathcal{D} = \{(x_i, y_i) \mid x_i \in \mathcal{X} \subset \mathbb{R}^D \wedge y_i \in \mathcal{Y} = \{0,1\}; \forall i = 1, \ldots, M\}$, where the grades of the ith student in the first two sessions are represented by the vector x_i, which is a two-dimensional real-valued vector. Thus, $D = 2$ because there are only two independent variables corresponding to the grades obtained in these sessions. The target or dependent variable y_i, might take one of two possible values: $y_i = 1$ if the ith student failed the course, and $y_i = 0$ otherwise. Furthermore, M is the size of the dataset.

We preprocessed the independent variables by standardising each component for later practical convenience. Thus, for a given ith student's grade in the jth session, the component x_{ij} is yielded as follows:

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{1}$$

where $\bar{x}_j$ and s_j denote the mean and standard deviation of the grades obtained in the jth session, respectively.

Given the dataset $\mathcal{D}$, our goal is to estimate the probability that the ith student fails the course given the grades obtained in the first two sessions, that is, $P(y_i = 1 \mid x_i)$.

To accomplish this goal, we employ logistic regression, a supervised machine learning method. In this model, the logistic function $\sigma : \mathbb{R} \to [0, 1]$ applied to a linear combination of the model weights and the independent variables to compute $P(y_i = 1 \mid x_i)$ as follows:

$$P(y_i = 1 \mid x_i) = \sigma(w^T \hat{x}_i + w_0) = \sigma(w_2 \hat{x}_{i2} + w_1 \hat{x}_{i1} + w_0) \tag{2}$$

where the components of the D-dimensional vector $w \in \mathbb{R}^D$ are the weights associated with the independent variables (D = 2), and $w_0 \in \mathbb{R}$ is the intercept of the prediction function. The logistic function is defined as follows:

$$\sigma(u) = \frac{1}{1 + \exp(-u)} \tag{3}$$

The values of the weights and intercept are estimated by finding w and w_0 that minimise the objective function, namely the regularised cross-entropy error $\mathcal{L}$ with L_2 regularisation:

$$\begin{aligned} \min_{w_0, w} \mathcal{L}(w_0, w) = -\sum_{i=1}^{M} \{ y_i \ln[\sigma(w^T \hat{x}_i + w_0)] + \dots \\ \dots + (1 - y_i) \ln[1 - \sigma(w^T \hat{x}_i + w_0)] \} + \frac{\lambda}{2} ||w||^2 \end{aligned} \tag{4}$$

where $\lambda \in \mathbb{R}$ is the regularisation parameter used to shrink the vector w and prevent overfitting.

The model weights may be estimated using gradient-descent–based optimisation algorithms, which rely on the partial derivatives of the objective function $\mathcal{L}$. We standardised the independent variables to speed up the convergence of the gradient-descent-based optimisation process. In particular, the partial derivatives are as follows:

$$\frac{\partial \mathcal{L}(w_0, w)}{\partial w_j} = \sum_{i=1}^{M} [\sigma(w^T \hat{x}_i + w_0) - y_i] \hat{x}_{ij} + \lambda w_j \tag{5}$$

and

$$\frac{\partial \mathcal{L}(w_0, w)}{\partial w_0} = \sum_{i=1}^{M} [\sigma(w^T \hat{x}_i + w_0) - y_i] \tag{6}$$

A broader description of logistic regression is provided in [18]. We used the implementation available in the machine learning library known Scikit-learn [19] and employed its default optimisation algorithm, namely L-BFGS [20], to minimise the above-mentioned objective function.

The probability threshold used for classification in [19] is 0.5. Accordingly, for a new input vector x_i corresponding to the ith student, whose data were not used to fit the model, the predicted label is $\hat{y}_i = 1$ (failure) if $P(\hat{y}_i = 1 \mid x_i) \geq 0.5$; otherwise, $\hat{y}_i = 0$.

On the other had, to select the regularisation parameter and evaluate the performance of the logistic regression model, we adopted K-fold Cross-Validation. In this procedure, the dataset $\mathcal{D}$ is divided into K subsets, i.e., $\mathcal{D}_1$, $\mathcal{D}_2$, ..., $\mathcal{D}_K$, and the model is evaluated K times. At each iteration, $K - 1$ subsets are used to fit the model, whereas the remaining subset is employed for testing. More formally, in the kth evaluation the model is fitted using $\cup_{l=1; l \neq k}^{K} \mathcal{D}_l$ and tested on the subset $\mathcal{D}_k$.

We chose $K = 10$ for the models fitted to the datasets corresponding to Physics I and Physics II, while $K = 3$ was selected for Physics III due to the small number of positive examples. To address the class imbalance in the dataset, we used the stratified K-fold cross-validation procedure implemented in Scikit-learn. The dataset partitions used in each experimental setting are presented in Tables I to III.

TABLE I. 10-FOLD CROSS-VALIDATION SETTING FOR THE DATASET CORRESPONDING TO THE STUDENTS WHO ENROLLED PHYSICS I COURSE

Fold	**Number of Examples for Fitting the Model**			**Number of Examples for Testing**		
	Positive	**Negative**	**Total**	**Positive**	**Negative**	**Total**
1	68	184	252	8	20	28
2	68	184	252	8	20	28
3	68	184	252	8	20	28
4	68	184	252	8	20	28
5	68	184	252	8	20	28
6	68	184	252	8	20	28
7	69	183	252	7	21	28
8	69	183	252	7	21	28
9	69	183	252	7	21	28
10	69	183	252	7	21	28

TABLE II. 10-FOLD CROSS-VALIDATION SETTING FOR THE DATASET CORRESPONDING TO THE STUDENTS WHO ENROLLED PHYSICS II COURSE

Fold	**Number of Examples for Fitting the Model**			**Number of Examples for Testing**		
	Positive	**Negative**	**Total**	**Positive**	**Negative**	**Total**
1	29	144	173	3	17	20
2	28	145	173	4	16	20
3	28	145	173	4	16	20
4	29	145	174	3	16	19
5	29	145	174	3	16	19
6	29	145	174	3	16	19
7	29	145	174	3	16	19
8	29	145	174	3	16	19
9	29	145	174	3	16	19
10	29	145	174	3	16	19

TABLE III. 3-FOLD CROSS-VALIDATION SETTING FOR THE DATASET CORRESPONDING TO THE STUDENTS WHO ENROLLED PHYSICS III COURSE

Fold	**Number of Examples for Fitting the Model**			**Number of Examples for Testing**		
	Positive	**Negative**	**Total**	**Positive**	**Negative**	**Total**
1	2	90	92	1	45	46
2	2	90	92	1	45	46
3	2	90	92	1	45	46

The evaluation aims to measure the accuracy, precision, recall, and harmonic mean or F_1-score of the classification model. Moreover, the regularisation parameter is tuned by selecting the value that minimises the error (defined as $1 - accuracy$) using the elbow rule. Since the dataset is imbalanced in terms of classes, we analysed the Receiver Operating Characteristic (ROC) curve and calculated the area under this curve.

Once w and w_0 are estimated, we employed the Monte Carlo simulation method [21] to explore the space of possible student performance outcomes during the first two sessions. Instead of relying solely on the limited number of observed cases, we reconstructed the grade distribution using random numbers drawn from a uniform distribution over the interval [0, 5]. This allowed us to simulate a large number of plausible grade combinations and estimate the associated risks under uncertainty.

In essence, the Monte Carlo simulation serves as a data-informed method for approximating risk in underrepresented or unobserved configurations, enabling generalisation beyond the empirical observations while remaining grounded in the observed statistical characteristics of the data.

This method may be used to estimate the probability that a student fails a physics course as follows:

$$P(\hat{y}_i = 1) = \int_{\mathcal{X}} P(\hat{y}_i = 1 \mid x_i) P(x_i) dx_i \tag{7}$$

where $\hat{y}_i = 1$ if the model predicts that a new ith student fails the course. Since this integral cannot be solved analytically, the numerical method known as Monte Carlo simulation can approximate its value by generating an $N \times 2$-dimensional matrix $X \in [0,5]^{N\times 2}$, whose components $X_{ij} \sim \mathcal{U}(0,5)$ are independently drawn from a uniform distribution over the interval $[0,5]$.

Then, by the law of large numbers, the integral above is approximated by computing the expected value of $P(\hat{y}_i = 1 \mid X_{i,:})$ (here $X_{i,:}$ is the ith row of X) as follows:

$$P(\hat{y}_i = 1) \approx \mathbb{E}_{X_{i,:} \sim \mathcal{U}[0,5]^2}[P(\hat{y}_i = 1 \mid X_{i,:})] \tag{8}$$

or

$$P(\hat{y}_i = 1) \approx \frac{1}{N} \sum_{i=1}^{N} P(\hat{y}_i = 1 \mid X_{i,:}) \tag{9}$$

The grades are sampled from a uniform distribution because our goal is not to reproduce the empirical distribution of grades observed in the dataset. Instead, the purpose of the simulation is to systematically explore the full domain of admissible grade values when evaluating the risk predicted by the logistic regression model. Under this assumption, all combinations of grades in the first two sessions are treated as equally plausible a priori, preventing the simulation from being biased toward patterns that might be overrepresented in the observed data.

Assuming a uniform distribution is consistent with the goal of risk analysis under uncertainty, which focuses on understanding how the predictive model behaves across the full range of possible inputs rather than only within the empirical distribution of the data used to fit the logistic regression model. Consequently, the Monte Carlo method enables the approximation of expected failure probabilities while allowing the analysis of scenarios that may be rare or even absent in the available data.

Thus, the probability that the ith student fails a physics course given that they failed the jth session is denoted as $P(\hat{y}_i = 1 \mid x_{ij} < 3)$, where $x_{ij} < 3$ indicates that the ith student failed the jth session because the minimum grade required to pass the course is 3.

The *Absolute Risk (AR)* of failing the course given failure in session jth is defined as:

$$AR(\hat{y}_i = 1 \mid x_{ij} < 3) = \int_{\mathcal{X}} \frac{P(\hat{y}_i = 1, x_{ij} < 3)}{P(x_{ij} < 3)} dx_i \tag{10}$$

Similarly, the absolute risk of failing the course given that the student did *not* fail session jth is:

$$AR(\hat{y}_i = 1 \mid x_{ij} \geq 3) = \int_{\mathcal{X}} \frac{P(\hat{y}_i = 1, x_{ij} \geq 3)}{P(x_{ij} \geq 3)} dx_i \tag{11}$$

whereas, the *Relative Risk (RR)* is defined as the ratio of these two quantities:

$$RR(\hat{y}_i = 1 \mid x_{ij} < 3) = \frac{AR(\hat{y}_i = 1 \mid x_{ij} < 3)}{AR(\hat{y}_i = 1 \mid x_{ij} \geq 3)} \tag{12}$$

Using the Monte Carlo method, the absolute risk $AR(\hat{y}_i = 1 \mid x_{ij} < 3)$ is approximated as:

$$AR(\hat{y}_i = 1 \mid x_{ij} < 3) \approx \frac{\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i = 1 \wedge X_{ij} < 3)}{\sum_{i=1}^{N} \mathbf{1}(X_{ij} < 3)} \tag{13}$$

Herein we assume that $\hat{y}_i = 1$ if $P(\hat{y}_i = 1 \mid X_{i,:}) \geq 0.5$, and $\mathbf{1}(u) = 1$ if the condition u is true, and 0 otherwise. Similarly, the absolute risk for students who did *not* fail session j is:

$$AR(\hat{y}_i = 1 \mid x_{ij} \geq 3) \approx \frac{\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i = 1 \wedge X_{ij} \geq 3)}{\sum_{i=1}^{N} \mathbf{1}(X_{ij} \geq 3)} \tag{14}$$

Finally, the relative risk is calculated as:

$$RR(\hat{y}_i = 1 \mid x_{ij} < 3) \approx \frac{\frac{\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i=1 \wedge X_{ij}<3)}{\sum_{i=1}^{N} \mathbf{1}(X_{ij}<3)}}{\frac{\sum_{i=1}^{N} \mathbf{1}(\hat{y}_i=1 \wedge X_{ij}\geq 3)}{\sum_{i=1}^{N} \mathbf{1}(X_{ij}\geq 3)}} \tag{15}$$

This simulation-based approach enables us to estimate the conditional risks associated with failing individual sessions and provides a probabilistic understanding of academic outcomes based on partial performance.

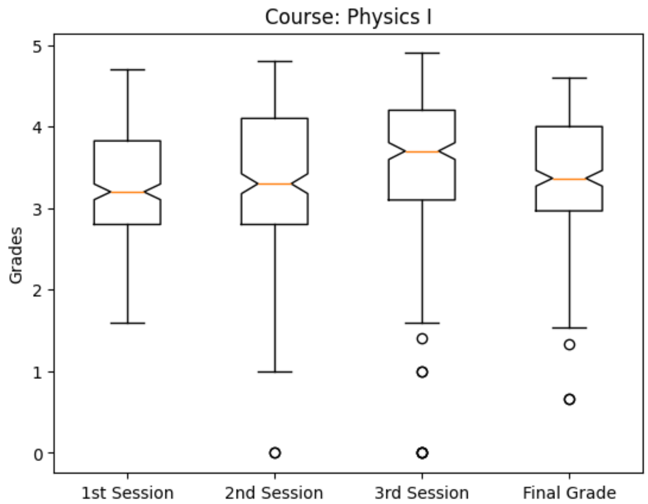

Figure 1. Grades of the students enrolled in Physics I during 2024 and 2025

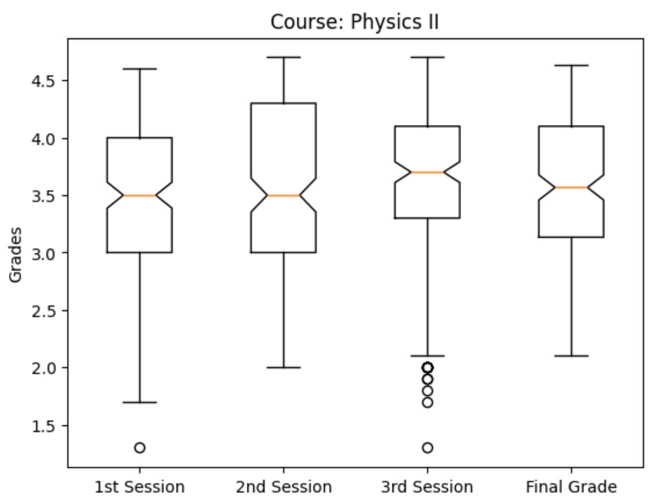

Figure 2. Grades of the students enrolled in Physics II during 2024 and 2025

The simulation was implemented in Python using the NumPy and Matplotlib libraries. The anonymised dataset and corresponding source code are available upon request from the first author.

III. The Research Results and Discussion

Based on the collected dataset, the mean grades for the first, second, and third sessions of the Physics I course were 3.29, 3.36, and 3.59, respectively, with corresponding standard deviations of 0.73, 0.81, and 0.85. As shown in Figure 1, the box plot of the final grade indicates that most students passed the course, while exceptionally high grades were relatively uncommon. Furthermore, a general trend of improved performance can be observed in the final session.

Similarly, the mean grades for Physics II were 3.51, 3.56, and 3.56 for the first, second, and third sessions, respectively, with standard deviations of 0.69, 0.72, and 0.70. Figure 2 shows a performance pattern comparable to that observed in Physics I, where exceptionally high performances were relatively infrequent.

Additionally, the dataset shows that the mean grades for the first, second, and third sessions in the Physics III course were 3.68, 3.66, and 3.86, respectively, with standard deviations of 0.66, 0.60, and 0.49. Figure 3 illustrates a pattern similar to that of the previous two physics courses, where students tend to gradually improve their grades across sessions.

TABLE IV. 10-Fold Cross-Validation Results Corresponding to the Students who Enrolled Physics I Course

Fold	**Accuracy** (%)	**Precision** (%)	**Recall** (%)	$\mathbf{F_1}$ **score** (%)
1	100.00	100.00	100.00	100.00
2	92.86	100.00	100.00	100.00
3	96.43	88.89	100.00	94.12
4	89.29	72.73	100.00	84.21
5	67.86	40.00	25.00	30.77
6	92.86	100.00	75.00	85.71
7	92.86	85.71	85.71	85.71
8	89.29	100.00	57.14	72.73
9	92.86	100.00	71.43	83.33
10	89.29	75.00	85.71	80.00
Mean	90.36	86.23	77.50	80.23

On the other hand, concerning the machine learning method employed in this research, the results of the 10-fold cross-validation for the logistic regression models of Physics I and Physics II are presented in Tables IV and V, respectively. Table VI reports the results of the 3-fold cross-validation for the Physics III model.

The results reveal that all models achieve high accuracy despite the class imbalance in the datasets. Moreover, recall exceeds 70% for the models corresponding to the Physics I and Physics II courses, and reaches at least 66% for the model corresponding to the Physics III course. In addition, the area under the Receiver Operating Characteristic (ROC) curve is above 0.95 for all models (see Figures 4 to 6). These results indicate that the models are effective at predicting whether a student is likely to fail a given physics course based on their grades in the first two sessions.

These results are consistent with the confusion matrices presented in Tables VII to IX. In particular, the model correctly identified 69 out of 76 positive examples for the Physics I course (see Table VII), 26 out of 32 positive examples for the Physics II course (see Table VIII), and 2 out of 3 positive examples for the Physics III course (see Table IX). The results for the latter case are aligned with those reported for the second fold in Table VI, where the precision, recall, and F_1 score were equal to zero.

Figures 10 to 12 visualise the classification of all examples in the dataset. In particular, Figure 12 shows that only one positive example lies above the decision boundary (a false negative), which might correspond to an exceptional case.

Once the logistic regression models were fitted for each course, Monte Carlo simulations were performed as described in the previous section. Thus, the results of the numerical simulations indicate that the probabilities of failing Physics I, II, and III are 61.62%, 58.61%, and 51.82%, respectively (see

TABLE V. 10-Fold Cross-Validation Results Corresponding to the Students who Enrolled Physics II Course

Fold	Accuracy (%)	Precision (%)	Recall (%)	F_1 score (%)
1	100.00	100.00	100.00	100.00
2	95.00	100.00	75.00	85.71
3	95.00	80.00	100.00	88.89
4	94.74	75.00	100.00	85.71
5	94.74	100.00	66.67	80.00
6	89.47	100.00	33.33	50.00
7	84.21	50.00	33.33	40.00
8	100.00	100.00	100.00	100.00
9	94.74	100.00	66.67	80.00
10	89.47	100.00	33.33	50.00
Mean	93.74	90.50	70.83	76.03

TABLE VI. 10-Fold Cross-Validation Results Corresponding to the Students who Enrolled Physics II Course

Fold	Accuracy (%)	Precision (%)	Recall (%)	F_1 score (%)
1	100.00	100.00	100.00	100.00
2	97.83	0.00	0.00	0.00
3	100.00	100.00	100.00	100.00
Mean	99.28	66.67	66.67	66.67

TABLE VII. Confusion Matrix Illustrating the Performance of Logistic Regression Model Corresponding to the Course of Physics I.

True Class	Forecasted Class		Total
	$\hat{y}_i = 0$	$\hat{y}_i = 1$	
$y_i = 0$	194	10	204
$y_i = 1$	17	59	76
Total	211	69	280

TABLE VIII. Confusion Matrix Illustrating the Performance of Logistic Regression Model Corresponding to the Course of Physics II.

True Class	Forecasted Class		Total
	$\hat{y}_i = 0$	$\hat{y}_i = 1$	
$y_i = 0$	158	3	161
$y_i = 1$	9	23	32
Total	167	26	193

TABLE IX. Confusion Matrix Illustrating the Performance of Logistic Regression Model Corresponding to the Course of Physics III.

True Class	Forecasted Class		Total
	$\hat{y}_i = 0$	$\hat{y}_i = 1$	
$y_i = 0$	135	0	135
$y_i = 1$	1	2	3
Total	136	2	138

TABLE X. Probability of Failing a Physics Course Obtained from the Monte Carlo Simulation Results.

Course	Probability of Failing $P(\hat{y}_i = 1)$	Standard Error	95% CI
Physics 1	61.73%	9.3×10^{-4}	[61.60%, 61.86%]
Physics 2	58.44%	6.3×10^{-4}	[58.31%, 58.56]
Physics 3	51.82%	5.0×10^{-4}	[51.72%, 51.92%]

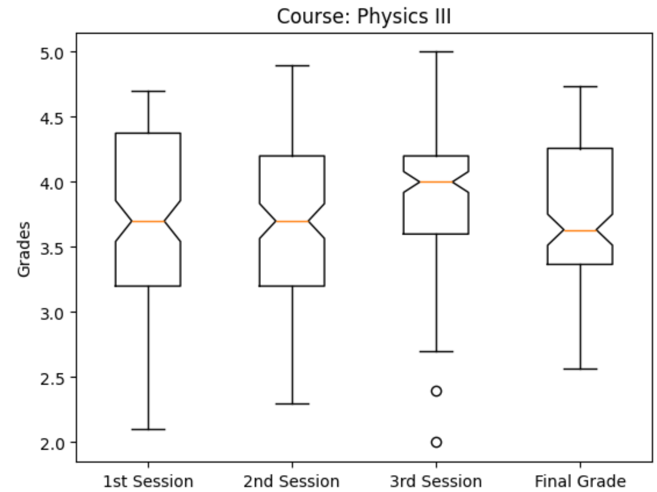

Figure 3. Grades of the students enrolled in Physics III during 2024 and 2025

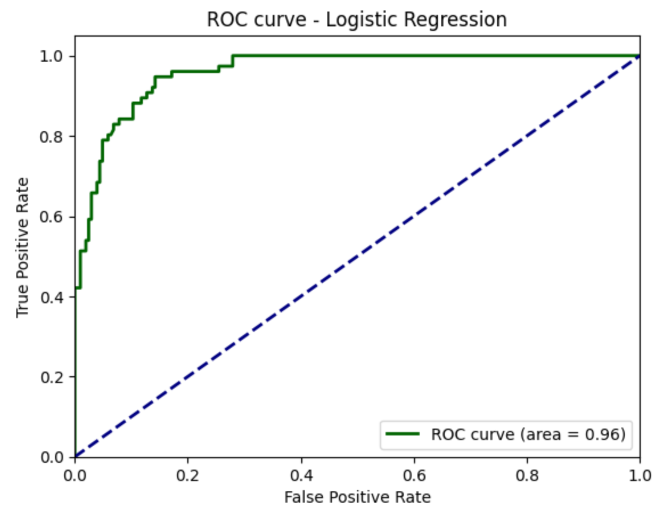

Figure 4. Receiver Operating Characteristic (ROC) curve for the logistic regression model corresponding to the course of Physics I

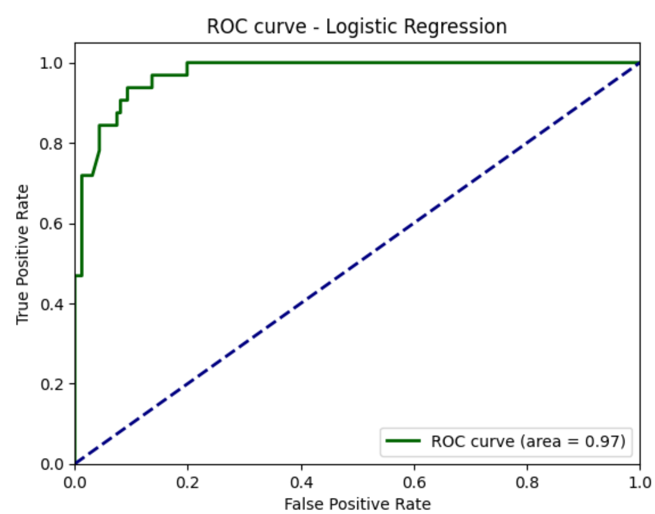

Figure 5. Receiver Operating Characteristic (ROC) curve for the logistic regression model corresponding to the course of Physics II

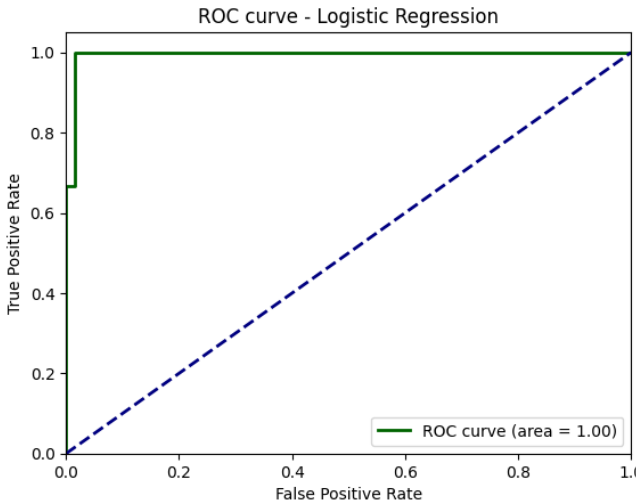

Figure 6. Receiver Operating Characteristic (ROC) curve for the logistic regression model corresponding to the course of Physics III

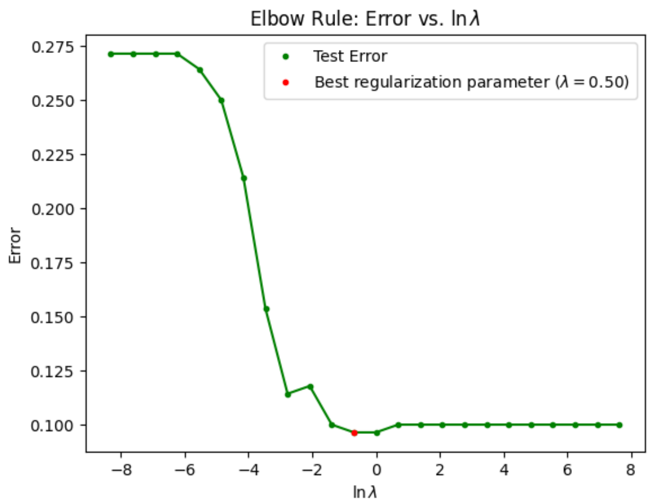

Figure 7. Elbow rule used to select the regularisation parameter for the logistic regression model corresponding to the course of Physics I

Table X). Figures 13–15 illustrate the convergence behaviour of the Monte Carlo simulations towards these estimates.

Figures 16–18 show the histograms of the failure probability obtained from simulated inputs to the logistic regression model. Instead of a bell-shaped pattern, the histograms exhibit a U-shaped distribution, reflecting the bimodal tendency typically observed in well-separated classification problems, where most randomly generated inputs yield probabilities close to either 0 or 1. Under these conditions, the simulated outcomes follow a Bernoulli distribution, and the sample mean of the simulations provides a Monte Carlo estimate of $P(y_i = 1)$, interpreted here as the probability that a student fails the physics course.

Additionally, Table XI summaries the relative and absolute risk estimates derived from the Monte Carlo simulation. The corresponding relative risks for each course are illustrated using forest plots in Figures 19–21. A similar pattern is observed

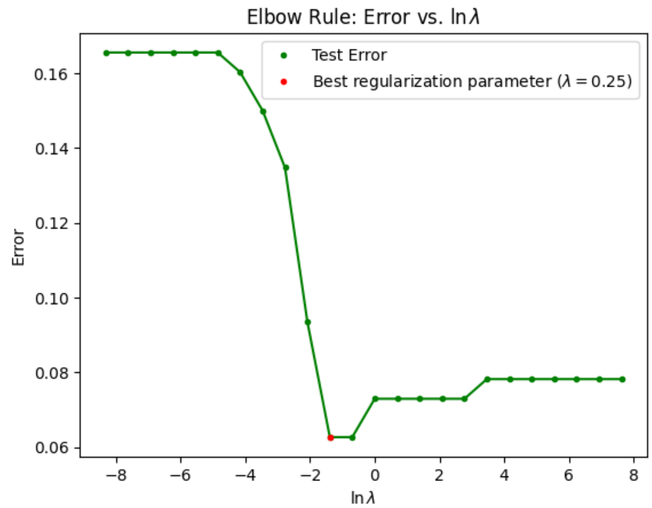

Figure 8. Elbow rule used to select the regularisation parameter for the logistic regression model corresponding to the course of Physics II

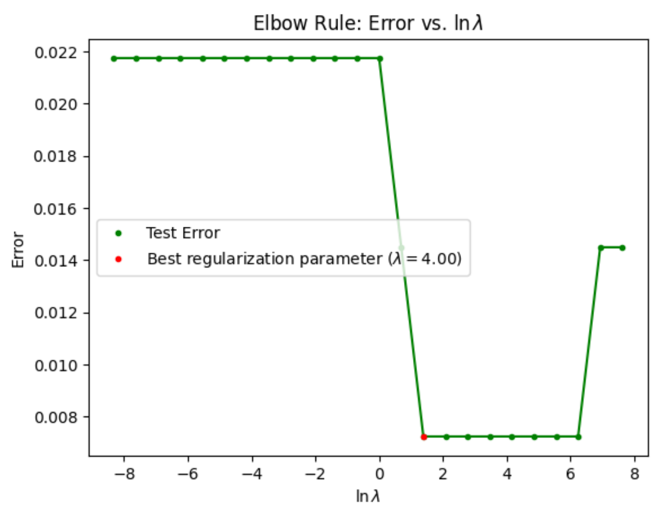

Figure 9. Elbow rule used to select the regularisation parameter for the logistic regression model corresponding to the course of Physics III

TABLE XI. ABSOLUTE AND RELATIVE RISK BY SESSION AND COURSE.

Course	Session(s) Failed	Absolute Risk (%)	Relative Risk (%)	95% CI (RR)
Physics I	S1	80.89	2.32	[2.31, 2.34]
	S2	82.82	2.61	[2.59, 2.63]
	S1 and S2	99.37	2.86	[2.84, 2.88]
Physics II	S1	71.94	1.85	[1.83, 1.85]
	S2	85.17	4.47	[4.42, 4.51]
	S1 and S2	98.75	2.54	[2.52, 2.55]
Physics III	S1	70.63	2.97	[2.95, 2.98]
	S2	73.23	3.67	[3.64, 3.69]
	S1 and S2	95.54	4.01	[3.98, 4.04]

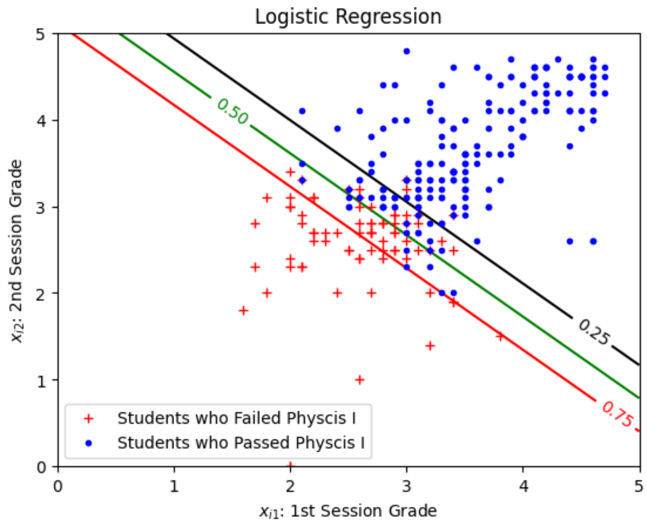

Figure 10. Classification of students by the logistic regression model for the Physics I course. Regions above the black line correspond to failure probabilities below 25%, while regions below the red line correspond to probabilities above 75%. The green line indicates the decision boundary, i.e., $P(\hat{y}_i \mid x_i) = 50\%$

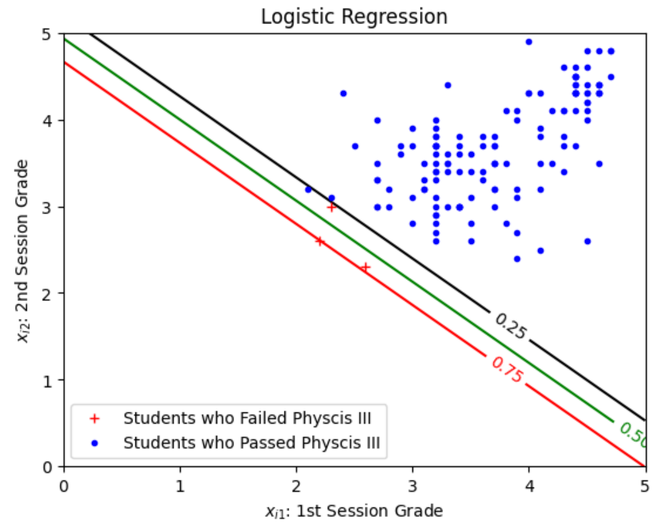

Figure 12. Classification of students by the logistic regression model for the Physics III course. Regions above the black line correspond to failure probabilities below 25%, while regions below the red line correspond to probabilities above 75%. The green line indicates the decision boundary, i.e., $P(\hat{y}_i \mid x_i) = 50\%$

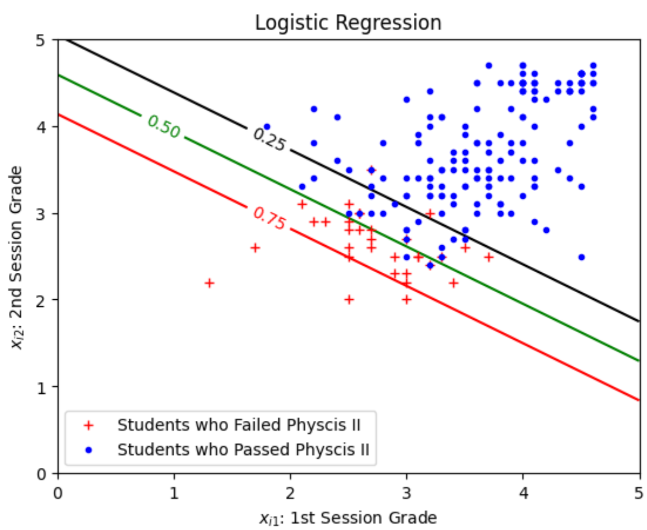

Figure 11. Classification of students by the logistic regression model for the Physics II course. Regions above the black line correspond to failure probabilities below 25%, while regions below the red line correspond to probabilities above 75%. The green line indicates the decision boundary, i.e., $P(\hat{y}_i \mid x_i) = 50\%$

for Physics I and Physics III. When comparing the risk of failing the course given that a student fails either of the first two sessions ($RR(\hat{y}_i \mid x_{i1} < 3) = 2.32$ and $RR(\hat{y}_i \mid x_{i2} < 3) = 2.61$ for Physics I, whereas $RR(\hat{y}i \mid x_{i1} < 3) = 2.97$ and $RR(\hat{y}_i \mid x_{i2} < 3) = 3.67$ for Physics III), failing both sessions is associated with a substantially higher risk of failing Physics I ($RR(\hat{y}_i \mid x_{i1} < 3, x_{i2} < 3) = 2.86$) and Physics III ($RR(\hat{y}_i \mid x_{i1} < 3, x_{i2} < 3) = 4.01$).

Interestingly, for Physics II (see Figure 20), failing the second session alone ($RR(\hat{y}_i \mid x_{i2} < 3) = 4.47$) is associated with a higher relative risk than failing either the first session alone ($RR(\hat{y}_i \mid x_{i1} < 3) = 1.85$) or both sessions combined ($RR(\hat{y}_i \mid x_{i1} < 3, x_{i2} < 3) = 2.54$). This pattern suggests that interventions targeting students who fail the first two sessions might have mitigated their risk of ultimately failing the course.

The absolute risk of course failure among students exposed to session failures versus those unexposed is compared in Table XII. The simulation results suggest that the probability of failing Physics I is 80.89% for students who fail the first session. This corresponds to a risk difference of 46.09 percentage points (95% CI: [45.81%, 46.36%] – CI stands for Confidence Interval –) relative to students who pass the first session, whose absolute risk of failure is 34.80%. This difference is statistically significant, indicating a meaningful association between failing the first session and ultimately failing Physics I. Furthermore, the relative risk is 2.32, suggesting that students who fail the first session are at least twice as likely to fail the course as those who do not (see Figure 19).

TABLE XII. COMPARISON OF ABSOLUTE RISK (AR) OF COURSE FAILURE BETWEEN STUDENTS EXPOSED AND UNEXPOSED TO FAILING PREVIOUS SESSIONS, WITH CORRESPONDING RISK DIFFERENCES (RD)

Course	Session(s) Failed	AR (%) exposed	AR (%) unexposed	RD (%)	95% CI (RD)
Physics I	S1	80.89	34.80	46.09†	[45.81, 46.36]
	S2	82.82	31.75	51.07†	[50.80, 51.34]
	S1 and S2	99.37	34.80	64.57†	[64.34, 64.81]
Physics II	S1	71.94	38.94	32.99†	[32.70, 33.29] †
	S2	85.17	19.07	66.10†	[65.86, 66.34]
	S1 and S2	98.75	38.94	59.81†	[59.57, 60.05]
Physics III	S1	70.63	23.82	46.81†	[46.62, 47.01]
	S2	73.23	19.98	53.26†	[53.07, 53.44]
	S1 and S2	95.54	23.82	71.72†	[71.56, 71.89]

(p-value < 0.05)

The absolute risk of failing the course increases to 82.82%

among students who fail the second session. This corresponds to a risk difference of 51.07 percentage points (95% CI: [50.80%, 51.34%]) compared with students who pass the second session, whose absolute risk of failure is 31.75%. The relative risk is 2.61, indicating that students who fail the second session are 2.61 times as likely to fail the course as those who pass it (see Figure 19).

When students fail both the first and second sessions of Physics I, the absolute risk of failing the course increases to 99.37%. This corresponds to a risk difference of 64.57 percentage points (95% CI: [64.34%, 64.81%]) compared with students who do not fail the first two sessions, whose absolute risk of failure is 34.80%. The relative risk is 2.86, further highlighting the substantially increased likelihood of course failure under these conditions (see Figure 19).

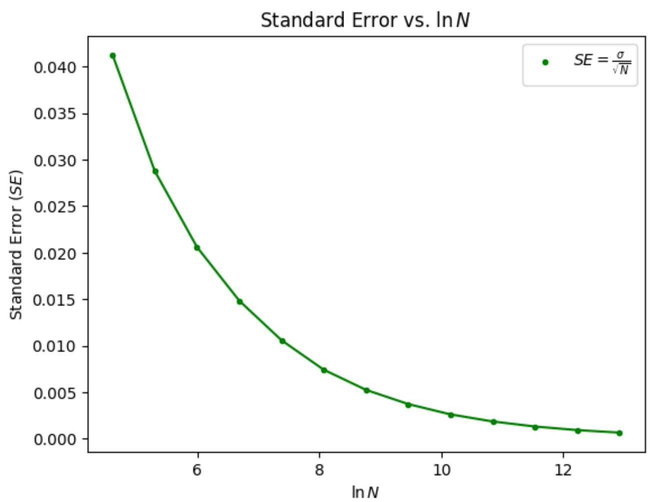

Figure 13. The simulation converges to the expected probability $P(\hat{y}_i = 1) = 61.73\%$ in the Physics I course as $N = 409,600$, with a standard error of 6.6×10^{-4}. The result lies within the 95% confidence interval of [61.61%, 61.86%].

For Physics II, the absolute risk of failing the course among students who fail the first session is 71.94%. This corresponds to a risk difference of 32.99 percentage points (95% CI: [32.70%, 33.29%]) compared with students who pass the first session, whose absolute risk of failure is 38.94%. The relative risk is 1.85, indicating that students who fail the first session are almost twice as likely to fail Physics II (see Figure 20).

The absolute risk of failing the course among students who fail the second session increases up to 85.17%. This corresponds to a risk difference of 66.10 percentage points (95% CI: [65.86%, 66.34%]) compared with students who pass the second session, whose absolute risk of failure is 19.07%. The relative risk is 4.47, indicating a substantially higher likelihood of failing the course under this condition than the previous one (see Figure 20).

Failing both sessions in Physics II rises the absolute risk of failing the course until 98.75%. This corresponds to a risk difference of 59.81 percentage points (95% CI: [59.57%, 60.05%]) compared with students who pass both sessions,

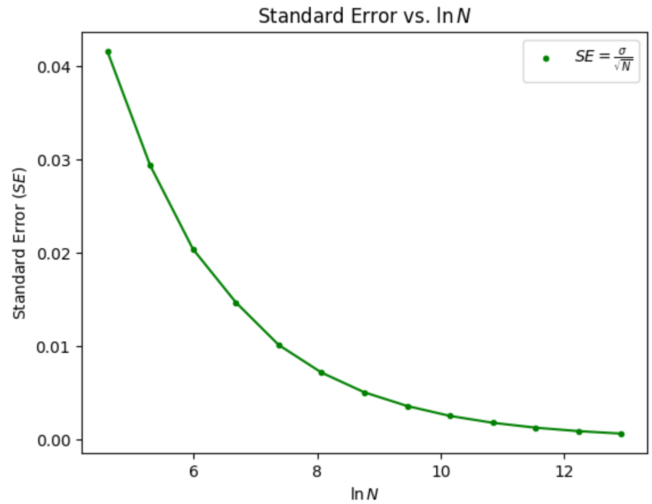

Figure 14. The simulation converges to the expected probability $P(\hat{y}_i = 1) = 58.44\%$ in the Physics II course as $N = 409,600$, with a standard error of 6.3×10^{-4}. The result lies within the 95% confidence interval of [58.34%, 58.56%].

whose absolute risk of failure is 38.94%. However, the relative risk decreases to 2.54, suggesting that students who fail the first two sessions are more than twice as likely to fail Physics II (see Figure 20).

For Physics III, the absolute risk of failing the course among students who fail the first session is 70.63%. This corresponds to a risk difference of 46.81 percentage points (95% CI: [46.62%, 47.01%]) compared with students who pass the first session, whose absolute risk of failure is 23.82%. The relative risk is 2.97, suggesting that students who fail the first session are nearly three times as likely to fail the course (see Figure 21).

For students who fail the second session, the absolute risk of course failure increases to 73.23%, compared with 19.98% among those who pass that session. This corresponds to a risk difference of 53.26 percentage points (95% CI: [53.07%, 53.44%]). The corresponding relative risk of 3.67 suggests that students who fail the second session are far more than three times as likely to fail the course (see Figure 21).

Furthermore, for students who fail the first two sessions, the absolute risk of failing Physics III increases to 95.54%, compared with 23.82% among students who pass both sessions. This corresponds to a risk difference of 71.72 percentage points (95% CI: [71.56%, 71.89%]). The relative risk of 4.01 suggests that students who fail both sessions are about four times as likely to fail the course (see Figure 21).

The pattern of absolute risk in all courses increases across sessions, whereas the relative risk does not follow this pattern in the case of Physics II. Despite the pattern observed in absolute risk, the risk differences follow the same pattern as the relative risk.

Besides, although the dataset contains few instances of students failing physics courses, especially Physics III, the Monte Carlo simulations suggest a noteworthy risk of failure.

In particular, Physics II requires further analysis to understand the pattern of lower failure risk among students who failed in the previous two sessions.

Several threats to validity should be considered when interpreting the results. First, the simulation model does not account for potential confounding variables such as prior preparation, study habits, or instructor differences, which may influence course performance. Second, the evaluation structure analysed is specific to the University of Córdoba, limiting the generalisability of the findings to other institutions or disciplines. Finally, because observed cases of course failure, particularly in Physics III, are relatively scarce in the dataset, the risk estimates rely heavily on the assumptions of the Monte Carlo simulation model.

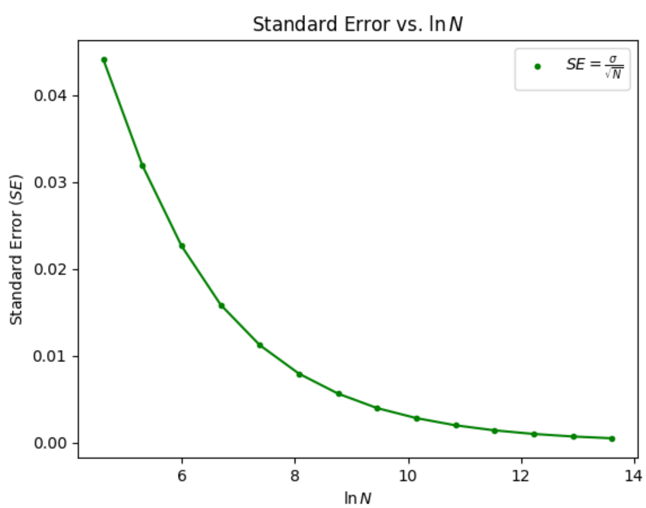

Figure 15. The simulation converges to the expected probability $P(\hat{y}_i = 1) = 51.82\%$ in the Physics II course as $N = 819,200$, with a standard error of 5×10^{-4}. The result lies within the 95% confidence interval of [51.72%, 51.92%].

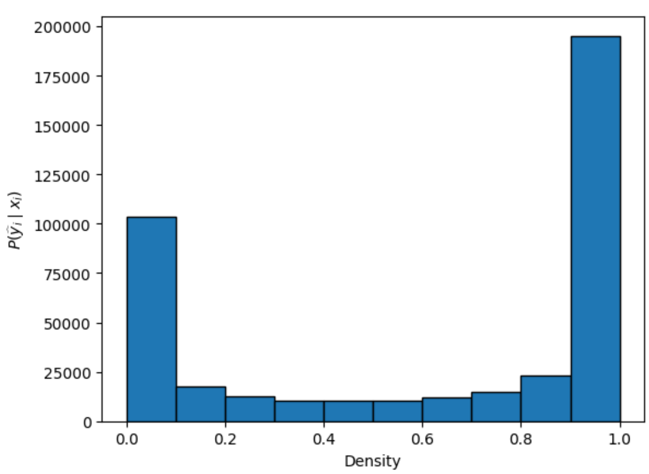

Figure 16. Distribution of probabilities $P(\hat{y}_i = 1 \mid x_i)$ obtained from the simulation for the Physics I course.

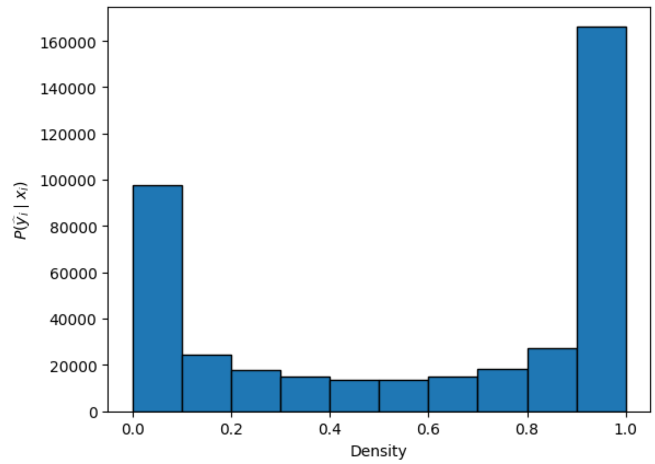

Figure 17. Distribution of probabilities $P(\hat{y}_i = 1 \mid x_i)$ obtained from the simulation for the Physics II course.

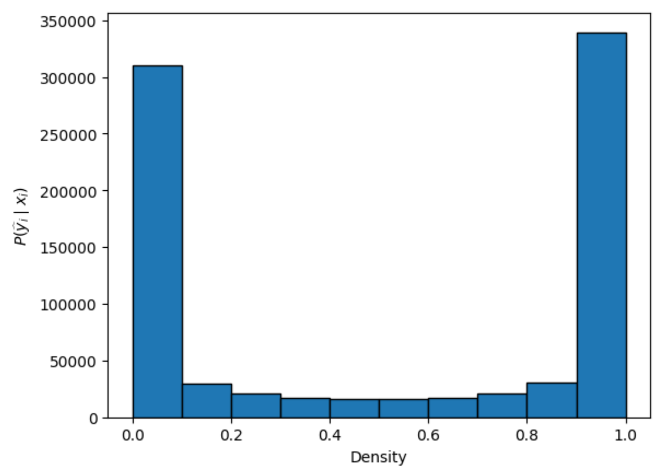

Figure 18. Distribution of probabilities $P(\hat{y}_i = 1 \mid x_i)$ obtained from the simulation for the Physics III course.

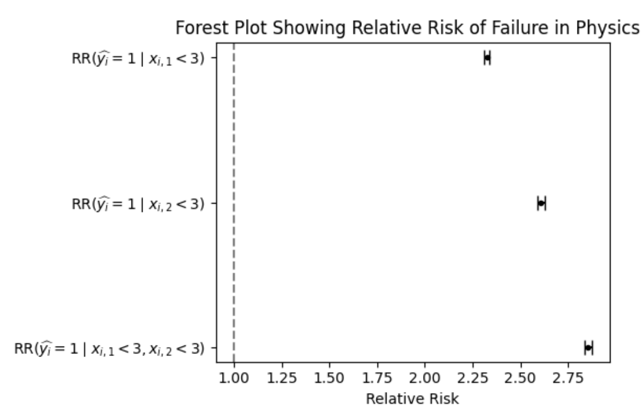

Figure 19. Forest plot showing the relative risk (RR) of failing the Physics I course. $RR(\hat{y}_i \mid x_{i1} < 3) = 2.32$, with a 95% confidence interval of [2.31, 2.34]; $RR(\hat{y}_i \mid x_{i2} < 3) = 2.61$, with a 95% confidence interval of [2.56, 2.63]; and $RR(\hat{y}_i \mid x_{i1} < 3, x_{i2} < 3) = 2.86$, with a 95% confidence interval of [2.84, 2.88]. In all cases, the Wald test p-value is less than 0.05.

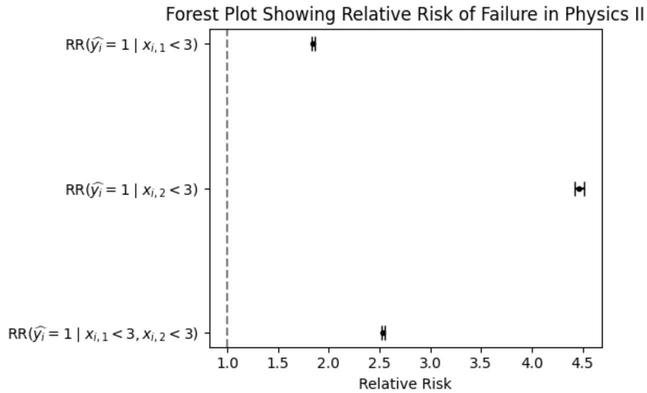

Figure 20. Forest plot showing the relative risk (RR) of failing the Physics II course. $RR(\hat{y}_i \mid x_{i1} < 3) = 1.85$, with a 95% confidence interval of [1.83, 1.85]; $RR(\hat{y}_i \mid x_{i2} < 3) = 4.47$, with a 95% confidence interval of [4.42, 4.51]; and $RR(\hat{y}_i \mid x_{i1} < 3, x_{i2} < 3) = 2.54$, with a 95% confidence interval of [2.52, 2.55]. In all cases, the Wald test p-value is less than 0.05.

IV. Conclusion and Outlook

In this research, we combined machine learning and Monte Carlo simulation to analyse the risk of failing physics courses in the Systems Engineering bachelor's program at the University of Córdoba. This approach was adopted to address the limited number of observed course failures, which makes direct empirical estimation of risk difficult.

The results demonstrate that early-session performance is a strong predictor of course failure, both at the level of individual students and at the population level. The estimated risks might therefore be employed to inform institutional policies aimed at preventing course failure and improving student retention.

The analysis also reveals a progressive increase in absolute risk as students fail additional sessions. In several scenarios, the probability of failing the course reaches alarming levels, approaching approximately 95–99%.

These findings suggest that the grading scheme may strongly amplify the consequences of early session failures. This effect might be associated with factors such as student discouragement, reduced engagement, and diminished resilience when facing subsequent academic challenges [22][23]. More broadly, performance in the first or second session is strongly associated with final course outcomes. Beyond mere statistical correlation, early academic difficulties might signal underlying motivational or behavioural challenges, making them valuable triggers for early intervention and academic support strategies. Further research is needed to design and evaluate targeted measures that could help students recover from early setbacks and improve their overall academic trajectory.

Possible interventions include tutoring programs, additional problem-solving sessions, and timely formative feedback implemented shortly after the first assessment to mitigate the risk of course failure.

As directions for future research, we propose the following:

- Collect additional data in order to apply the proposed methodology to other courses and broaden the scope of academic risk analysis.
- Extend the simulation framework to incorporate the specific coursework and evaluation structure of each session, with the aim of estimating risk more accurately.
- Adapt the simulation to consider non-uniform weighting of sessions when calculating final grades, in order to explore grading schemes that may reduce the risk of course failure.
- Incorporate additional data and evaluate alternative machine learning methods to improve the accuracy of the risk estimations.
- Consider additional explanatory variables to further strengthen both the predictive model and the simulation framework.

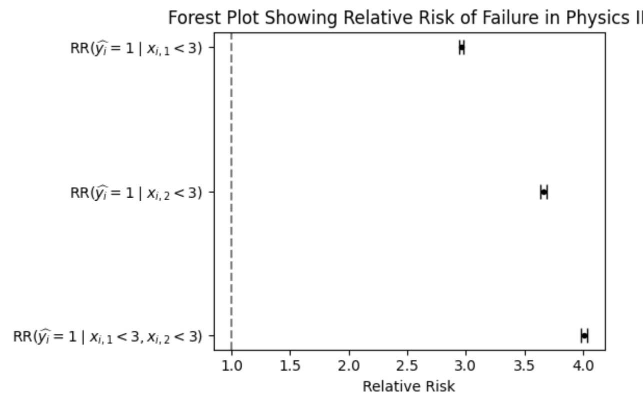

Figure 21. Forest plot showing the relative risk (RR) of failing the Physics III course. $RR(\hat{y}_i \mid x_{i1} < 3) = 2.97$, with a 95% confidence interval of [2.95, 2.98]; $RR(\hat{y}_i \mid x_{i2} < 3) = 3.67$, with a 95% confidence interval of [3.64, 3.69]; and $RR(\hat{y}_i \mid x_{i1} < 3, x_{i2} < 3) = 4.01$, with a 95% confidence interval of [3.98, 4.04]. In all cases, the Wald test p-value is less than 0.05.

Acknowledgement

Caicedo-Castro thanks the Lord Jesus Christ for blessing this project. The authors thank the Universidad de Córdoba, Colombia, for supporting this research. They also thank the Universidad de Córdoba for providing them with the data for conducting this research. Finally, the authors thank the anonymous referees for their comments that contributed to improve the quality of this paper.

References

[1] I. Caicedo-Castro, R. Castro-Púche and S. Castaño-Rivera, ‘Estimating the Risk of Failing Physics Courses through the Monte Carlo Simulation’, in *GPTMB 2025, The Second International Conference on Generative Pre-trained Transformer Models and Beyond*, International Academy, Research, and Industry Association, 2025, pp. 15–22.

[2] D. E. M. da Silva, E. J. S. Pires, A. Reis, P. B. de Moura Oliveira and J. Barroso, ‘Forecasting Students Dropout: A UTAD University Study’, *Future Internet*, vol. 14, no. 3, pp. 1–14, 2022.

[3] I. Caicedo-Castro, O. Velez-Langs, M. Macea-Anaya, S. Castaño-Rivera and R. Catro-Púche, ‘Early Risk Detection of Bachelor's Student Withdrawal or Long-Term Retention’, in *The 2022 IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications*, International Academy, Research, and Industry Association, 2022, pp. 76–84.

[4] S. Zihan, S.-H. Sung, D.-M. Park and B.-K. Park, 'All-Year Dropout Prediction Modeling and Analysis for University Students', *Applied Sciences*, vol. 13, p. 1143, Jan. 2023. DOI: 10.3390/app13021143.

[5] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis and V. Loumos, 'Dropout Prediction in E-Learning Courses through the Combination of Machine Learning Techniques', *Computers and Education*, vol. 53, no. 3, pp. 950–965, 2009, ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2009.05.010.

[6] J. Kabathova and M. Drlik, 'Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques', *Applied Sciences*, vol. 11, p. 3130, Apr. 2021. DOI: 10.3390/app11073130.

[7] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka and P. C. Nshimyumukiza, 'Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization', *Computers and Education: Artificial Intelligence*, vol. 3, p. 100 066, 2022, ISSN: 2666-920X. DOI: https://doi.org/10.1016/j.caeai.2022.100066.

[8] V. Čotić Poturić, I. Dražić and S. Čandrlić, 'Identification of Predictive Factors for Student Failure in STEM Oriented Course', in *ICERI2022 Proceedings*, ser. 15th annual International Conference of Education, Research and Innovation, Seville, Spain: IATED, 2022, pp. 5831–5837, ISBN: 978-84-09-45476-1. DOI: 10.21125/iceri.2022.1441.

[9] V. Čotić Poturić, A. Bašić-Šiško and I. Lulić, 'Artificial Neural Network Model for Forecasting Student Failure in Math Courses', in *ICERI2022 Proceedings*, ser. 15th annual International Conference of Education, Research and Innovation, Seville, Spain: IATED, 2022, pp. 5872–5878, ISBN: 978-84-09-45476-1. DOI: 10.21125/iceri.2022.1448.

[10] I. Caicedo-Castro, M. Macea-Anaya and S. Rivera-Castaño, 'Early Forecasting of At-Risk Students of Failing or Dropping Out of a Bachelor's Course Given Their Academic History - The Case Study of Numerical Methods', in *PATTERNS 2023: The Fifteenth International Conference on Pervasive Patterns and Applications*, ser. International Conferences on Pervasive Patterns and Applications, Nice, France: IARIA: International Academy, Research, and Industry Association, 2023, pp. 40–51, ISBN: 978-1-68558-049-0.

[11] I. Caicedo-Castro, M. Macea-Anaya and S. Castaño-Rivera, 'Early Risk Detection of Bachelor's Student Withdrawal or Long-Term Retention', in *The 2023 IARIA Annual Congress on Frontiers in Science, Technology, Services, and Applications*, International Academy, Research, and Industry Association, 2023, pp. 177–187.

[12] I. Caicedo-Castro, 'Course Prophet: A System for Predicting Course Failures with Machine Learning: A Numerical Methods Case Study', *Sustainability*, vol. 15, no. 18, 2023, 13950. DOI: 10.3390/su151813950.

[13] I. Caicedo-Castro, 'Quantum Course Prophet: Quantum Machine Learning for Predicting Course Failures: A Case Study on Numerical Methods', in *Learning and Collaboration Technologies*, P. Zaphiris and A. Ioannou, Eds., Cham: Springer Nature Switzerland, 2024, pp. 220–240, ISBN: 978-3-031-61691-4. DOI: 10.1007/978-3-031-61691-4_15.

[14] I. Caicedo-Castro, 'An Empirical Study of Machine Learning for Course Failure Prediction: A Case Study in Numerical Methods', *International Journal on Advances in Intelligent Systems*, vol. 17, no. 1 and 2, pp. 25–37, 2024.

[15] D. Torres, J. Crichigno and C. Sanchez, 'Assessing Curriculum Efficiency Through Monte Carlo Simulation', *Journal of College Student Retention*, vol. 22, no. 4, pp. 597–610, 2021. DOI: 10.1177/1521025118776618.

[16] I. Caicedo-Castro, O. Vélez-Langs and R. Castro-Púche, 'Using the Monte Carlo Method to Estimate Student Motivation in Scientific Computing', in *Patterns 2025, The Seventeenth International Conferences on Pervasive Patterns and Applications*, International Academy, Research, and Industry Association, 2025, pp. 15–22.

[17] I. Caicedo-Castro, R. Castro-Púche and O. Vélez-Langs, 'Identifying Factors that Increase the Risk of Demotivation in Scientific Computing Courses Using Monte Carlo Methods', *International Journal on Advances in Systems and Measurements*, vol. 18, no. 3 and 4, pp. 162–171, 2025.

[18] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006, ISBN: 0387310738.

[19] F. Pedregosa et al., 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[20] D. C. Liu and J. Nocedal, 'On the limited memory BFGS method for large scale optimization', *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, 1989.

[21] N. Metropolis and S. Ulam, 'The Monte Carlo Method', *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949, ISSN: 01621459, 1537274X.

[22] V. Tinto, *Leaving College: Rethinking the Causes and Cures of Student Attrition*, 2nd. Chicago, IL: University of Chicago Press, 1994, ISBN: 9780226804521.

[23] M. Yorke, 'Retention, Persistence and Success in On-Campus Higher Education, and their Enhancement in Open and Distance Learning', *Open Learning: The Journal of Open, Distance and e-Learning*, vol. 19, no. 1, pp. 19–32, 2004. DOI: 10.1080/0268051042000177827.