Evaluating Policies to Improve Relational-Level Learning in Engineering Mathematics: An Approach based on Machine Learning and the Monte Carlo Method

Isaac Caicedo-Castro*†‡ Rubby Castro-Púche†§ Oswaldo Vélez-Langs*‡
*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 | oswaldovelez}@correo.unicordoba.edu.co

***Abstract*—In this research, we combine machine learning and Monte Carlo simulation to estimate the probability that Systems Engineering students at the University of Córdoba fail to attain the relational level of the Structure of the Observed Learning Outcomes (SOLO) taxonomy in mathematics courses. Attainment of this level indicates the ability to integrate mathematical concepts and techniques in order to solve engineering problems. Logistic Regression is employed to model this probability using measures of student self-efficacy. Because the available dataset provides only limited coverage of the input space, Monte Carlo simulation is used to generate additional student profiles and evaluate counterfactual intervention scenarios aimed at improving self-efficacy. The results indicate that the baseline probability of failing to attain the relational level is 56.94%. Among the individual interventions considered, reducing mathematics anxiety yields the largest reduction in this probability, followed by strengthening mastery-learning beliefs and improving perceived teaching support. The simultaneous implementation of all interventions produces the greatest effect, reducing the estimated probability to 47.69%. These findings suggest that the combination of machine learning and Monte Carlo simulation provides a computational framework for evaluating educational interventions prior to their implementation and supports evidence-based decision-making in engineering education.**

Keywords-machine learning; logistic regression; support vector machines; Gaussian processes for classification; educational data mining; Monte Carlo method; learning outcomes.

I. INTRODUCTION

Learning applied mathematics is challenging across educational levels, from elementary education to postgraduate study. This difficulty stems from both the formal nature of mathematics and the complexity involved in translating real-world situations into mathematical representations. Consequently, engineering students often experience difficulties in integrating concepts, techniques, and methods when addressing problems within their respective fields, such as industrial, electrical, and mechanical engineering.

This research was conducted within the Bachelor's programme in Systems Engineering at the University of Córdoba, Colombia, where an Outcome-Based Education (OBE) approach is integrated with the Structure of the Observed Learning Outcomes (SOLO) taxonomy. Within this framework, learning achievement is classified into five levels according to the grade scale [0, 5], as follows:

i) Prestructural (0.0–2.0): The student has not yet demonstrated an understanding of the key concepts.
ii) Unistructural (2.1–2.9): The student demonstrates an understanding of a single aspect of the task.
iii) Multistructural (3.0–3.7): The student demonstrates an understanding of several aspects of the task, although these are not yet integrated.
iv) Relational (3.8–4.5): The student is able to integrate multiple aspects of the task into a coherent whole.
v) Extended Abstract (4.6–5.0): The student demonstrates a deep understanding of the underlying concepts and is able to apply them in novel contexts.

A student is deemed to have passed an assessment upon achieving at least the multistructural level, corresponding to a grade of 3.0 or higher. Similarly, progression within a course requires a final grade of at least 3.0. However, the development of effective engineering problem-solving skills is generally associated with the relational and extended abstract levels, particularly in mathematics-related courses.

Accordingly, the first objective of this research is to estimate the probability that a student fails to attain at least the relational level in mathematics as a function of self-efficacy, defined as an individual's belief in their capability to perform actions required to achieve a particular outcome [1]. Self-efficacy is shaped by several sources, including mastery experiences, social modelling through the observation of peers, social persuasion such as instructor feedback, and emotional and motivational states.

Regarding the mathematical notation, let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{M}$ be a dataset comprising observations of students' self-efficacy perceptions and their performance in engineering mathematics. The ith student is represented by a real-valued D-dimensional feature vector $x_i \in \mathcal{X} \subseteq \mathbb{R}^D$, whose jth component corresponds to a particular self-efficacy attribute. For example, x_{ij} may represent the extent to which the ith student believes they can successfully solve most mathematics problems during an assessment.

The corresponding target variable is defined as $y_i = 1$ if the ith student failed to attain the relational level, as determined by the mean grade obtained in previous mathematics courses, and

 ISBN: 978-1-68558-402-3

$y_i = 0$ otherwise. Here, M denotes the number of observations in the dataset.

Given a new student represented by the feature vector x_k, the objective is to estimate the probability $P(\hat{y}_k = 1 \mid x_k)$, where $\hat{y}_k = 1$ indicates that the model predicts the student will fail to attain the relational level. Here, x_k denotes the input vector, while y_k and $\hat{y}_k$ represent the observed and predicted outcomes, respectively.

A further objective of this research is to estimate the probability of failing to attain the relational level under a range of hypothetical interventions designed to reduce this risk. To this end, we pondered the following research questions:

i) How would this probability change if students perceived greater teaching support?
ii) How would this probability change if mathematics anxiety were reduced?
iii) How would this probability change if mastery confidence were increased?
iv) How would this probability change if greater teaching support, reduced mathematics anxiety, and increased mastery confidence were considered simultaneously?

To address the two objectives outlined above, a dataset comprising 93 observations and 26 variables was collected (i.e., $M = 93$ and $D = 26$).

Several machine learning methods were evaluated for estimating the probability that a student x_k fails to attain the relational level, that is, $P(\hat{y}_k = 1 \mid x_k)$. Across the 10-fold cross-validation experiments, Logistic Regression achieved the highest accuracy among the methods considered. A similar pattern was observed for precision.

Finally, the Monte Carlo method is employed to simulate the proposed intervention scenarios and thereby address the research questions outlined above. According to the simulation outcomes, the baseline probability of failing to attain the relational level in mathematics is estimated at 57.94%. The simulation results suggest that this estimate decreases to 54.54% under improved teaching support, 53.64% under reduced mathematics anxiety, and 53.99% under increased mastery confidence. The largest reduction is observed when all three interventions are considered simultaneously, yielding an estimated probability of 47.69%.

The remainder of the paper is organised as follows. Section II reviews the relevant literature, Section III describes the methodology, and Sections IV and V present the results and discussion. Section VI concludes the paper.

II. Prior Research

Educational data mining methods have been widely employed to predict student dropout, delayed graduation [2][3][4], and the risk of course failure or withdrawal [5][6][7][8][9][10][11][12][13][14]. In contrast, we examine the relationship between self-efficacy and engineering students' attainment of the relational level in mathematics, which reflects their ability to integrate mathematical concepts and apply them to problem-solving tasks within the SOLO framework.

Self-efficacy, introduced by Albert Bandura in 1986, refers to an individual's belief in their capability to successfully achieve a particular level of performance [1].

The relationship between self-efficacy and mathematics has been extensively investigated in the literature [15][16][17][18][19][20]. Similar associations have also been reported for reading skills [21].

Several of these studies have focused on mathematics at the undergraduate level [15][16][20], whereas others have examined self-efficacy in middle and secondary school settings [17][18][19].

Our research differs from prior one in three principal respects.

First, the questionnaire employed in our research captures broader beliefs regarding mathematical self-efficacy, rather than focusing on task-specific skills such as histogram construction [20].

Second, the relationship between self-efficacy and the outcome variable is formulated in our research as a classification problem, whereas previous studies have typically adopted a regression framework. In addition, our analysis is based on an educational process extending over at least two and a half years, reflecting the period during which students might complete the core mathematics curriculum. Our research further combines empirical observations with simulation to investigate intervention scenarios that would be difficult to evaluate experimentally because of the time, resources, and ethical considerations involved.

Third, our analysis focuses specifically on engineering mathematics, thereby providing a more clearly defined application context than that considered in related recent work [20].

A similar combination of machine learning and Monte Carlo methods has previously been employed to analyse the risk of failing physics courses [22][23] and student demotivation in scientific computing [24][25].

Logistic Regression and the Monte Carlo method have been applied to estimate the risk of failing physics courses based on students' academic progress during a semester [23]. This work extends an earlier study in which Monte Carlo simulation alone was used for the same purpose on a smaller dataset [22]. In that context, student progress is measured using at most two grades, making visualisation feasible. By contrast, our research considers a multivariate input space comprising more than three input variables. Consequently, both studies combine predictive modelling and Monte Carlo simulation, although we additionally evaluate Gaussian Processes and Support Vector Machines alongside Logistic Regression.

Prior research has also combined regression models with Monte Carlo simulation to analyse the risk of student demotivation within a multivariate setting [24][25]. While both lines of research employ supervised learning methods, our research addresses a classification problem, whereas the aforementioned studies focus on regression.

The Monte Carlo method has also been applied to predict student performance in engineering, medical, and mixed student cohorts, where it was found to outperform linear regression, particularly in capturing non-linear grade transitions and stochastic variability [26].

To our knowledge, the combination of methods adopted in our research has not previously been applied to investigate the relationship between self-efficacy and the probability of attaining the relational level in engineering mathematics.

III. Research Methodology

In this research, supervised machine learning methods were used to estimate the probability that the kth student fails to attain the relational level in mathematics, which is denoted as $P(\hat{y}_k = 1 \mid x_k)$. The analysis considered probabilistic classifiers, including Logistic Regression and Gaussian Processes [27], together with Support Vector Machines. Although Support Vector Machines do not provide probabilistic outputs inherently, their predictions might be calibrated to yield probability estimates [28]. Further details of these methods may be found in [29][30].

A dataset was collected at the beginning of 2026 to train the models described above. The corresponding questionnaire was designed such that its items represented the input variables included in the dataset. Responses were recorded on a Likert scale ranging from 0 (*strongly disagree*) to 5 (*strongly agree*).

The questionnaire was administered to 93 students enrolled between the fifth and eighth semesters of the programme, which is designed to be completed over 10 semesters. Participant identities were anonymised prior to analysis. Table I presents the questionnaire and the correspondence between its items and the 26 input variables of the models, represented by the components of the vector x_i, for $i = 1, \ldots, 93$.

The target variable is derived from the mean grade obtained across the core mathematics courses of the programme, namely Calculus I (Differential Calculus), Calculus II (Integral Calculus), Calculus III (Vector Calculus), Differential Equations, and Linear Algebra. The distribution of the average grades used to define the target variable is presented in Figure 1, while Figure 2 shows the resulting class distribution.

As the dataset is not perfectly balanced, model performance is evaluated using stratified 10-fold cross-validation. This procedure partitions the data into 10 folds, enabling model training and evaluation across 10 iterations. The distribution of records within each test fold is reported in Table II. Prior to model fitting, all input variables are standardised by centring them to zero mean and scaling them to unit variance.

Within the collected dataset, the proportion of students who failed to attain the relational level is 56.99%, corresponding to $P(y = 1) = 56.99\%$. However, Figure 3 shows that the input variables do not uniformly cover the input space, with some values unobserved in the available data. For example, the variables $x_{i,1}$, $x_{i,2}$, $x_{i,3}$, and $x_{i,15}$ never take the value 2. To explore regions of the input space that are sparsely represented or absent from the observed dataset, the Monte Carlo method is employed [31], assuming a multivariate normal distribution for the input variables.

The second objective of the research is addressed by estimating the probability that students fail to attain the

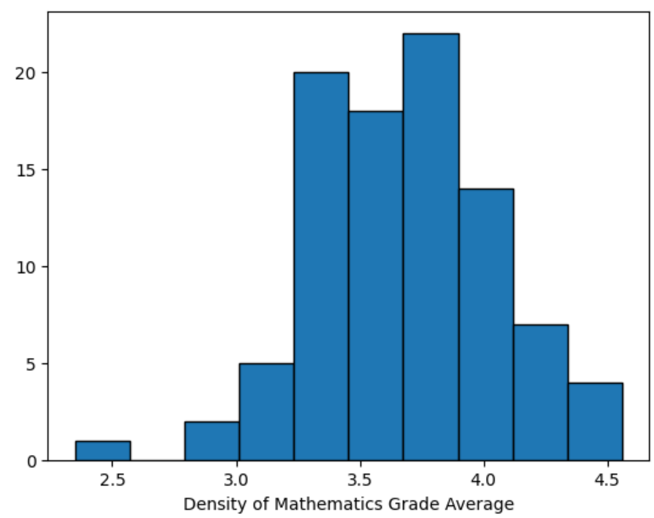

Figure 1. Histogram of the students' Average Grade obtained in mathematics courses

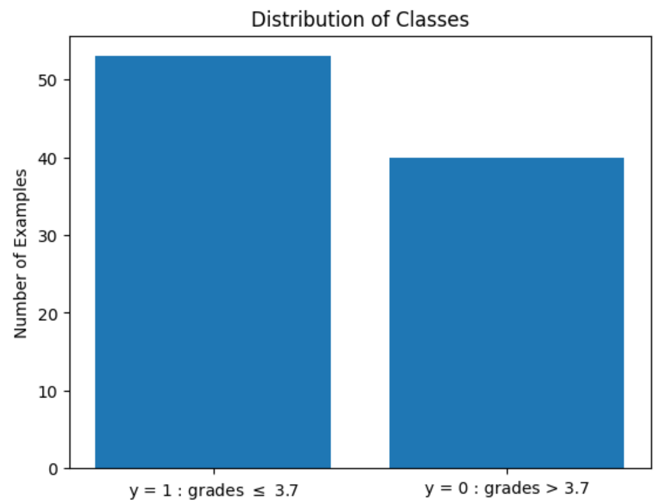

Figure 2. 53 out of 93 students (57%) did not reach the relational level in mathematics

relational level under a range of intervention scenarios. To this end, the Monte Carlo method is employed to simulate the implementation of these interventions.

Formally, the estimation of $P(\hat{y} = 1)$ requires the evaluation of the following integral:

$$P(\hat{y} = 1) = \int_{\mathcal{X}} P(\hat{y} = 1 \mid x) P(x) dx. \quad (1)$$

As the integral cannot be evaluated analytically, the Monte Carlo method is employed to approximate it. To this end, a matrix $X \in \mathbb{R}^{N \times D}$ is generated, whose rows correspond to samples drawn from a multivariate normal distribution with zero mean and covariance matrix S. The zero-mean assumption reflects the prior standardisation of the input variables. Here, N denotes the number of generated samples and D the number of input variables.

TABLE I. DESCRIPTION OF THE DATASET COLLECTED IN THIS RESEARCH

Input Variable	Question	Mean	Standard Deviation	Minimum Value	Maximum Value	Mode	Median
$x_{i,1}$	I can successfully solve most of the mathematics problems assigned to me.	3.65	0.71	2.00	5.00	4.00	4.00
$x_{i,2}$	If I try hard enough, I can complete even difficult mathematics tasks.	4.23	0.69	2.00	5.00	4.00	4.00
$x_{i,3}$	I can learn new mathematics topics without needing constant help.	3.62	0.83	2.00	5.00	4.00	4.00
$x_{i,4}$	When I make mistakes in mathematics, I can identify them and improve.	3.96	0.70	3.00	5.00	4.00	4.00
$x_{i,5}$	I am confident in my ability to understand complex mathematical concepts.	3.63	0.88	1.00	5.00	4.00	4.00
$x_{i,6}$	I often have difficulty successfully completing mathematics problems.	3.09	0.95	1.00	5.00	3.00	3.00
$x_{i,7}$	Compared with most of my classmates, I am able to achieve good results in mathematics.	3.43	0.74	2.00	5.00	3.00	3.00
$x_{i,8}$	I learn mathematics as quickly as other students in my class.	3.60	0.88	2.00	5.00	4.00	4.00
$x_{i,9}$	Seeing other students succeed in mathematics increases my belief that I can succeed as well.	4.19	0.95	1.00	5.00	5.00	4.00
$x_{i,10}$	When other students understand mathematics faster than I do, I begin to doubt my own ability.	2.54	1.17	1.00	5.00	3.00	3.00
$x_{i,11}$	I feel that my mathematical ability is inferior to that of most of my classmates.	2.58	1.08	1.00	5.00	3.00	3.00
$x_{i,12}$	My teachers believe that I can succeed in mathematics.	3.60	0.87	1.00	5.00	3.00	4.00
$x_{i,13}$	When my teacher encourages me, my confidence in mathematics increases.	4.11	0.90	1.00	5.00	5.00	4.00
$x_{i,14}$	Positive feedback about my work in mathematics strengthens my belief in my ability.	4.22	0.81	1.00	5.00	4.00	4.00
$x_{i,15}$	When I receive negative feedback in mathematics, I begin to doubt my competence.	2.70	1.20	1.00	5.00	3.00	3.00
$x_{i,16}$	I feel supported by my teacher in developing my mathematical skills.	3.65	1.01	1.00	5.00	3.00	4.00
$x_{i,17}$	I feel calm when solving mathematics problems.	3.78	0.96	1.00	5.00	3.00	4.00
$x_{i,18}$	I become anxious when mathematics problems are very challenging.	3.44	1.09	1.00	5.00	3.00	3.00
$x_{i,19}$	I feel overwhelmed when I face difficult mathematics topics.	3.29	1.19	1.00	5.00	3.00	3.00
$x_{i,20}$	I feel confident even when working with unfamiliar mathematical material.	2.94	0.97	1.00	5.00	3.00	3.00
$x_{i,21}$	Even relatively simple mathematics tasks can make me feel stressed.	2.38	1.06	1.00	5.00	2.00	2.00
$x_{i,22}$	I enjoy solving challenging mathematics problems.	3.14	1.10	1.00	5.00	3.00	3.00
$x_{i,23}$	I believe mathematics is important for becoming a successful systems engineer.	4.22	0.81	2.00	5.00	5.00	4.00
$x_{i,24}$	I study mathematics only to obtain good grades.	2.98	1.11	1.00	5.00	3.00	3.00
$x_{i,25}$	I study mathematics only to fulfil the requirements for obtaining the Systems Engineering degree.	2.94	1.06	1.00	5.00	3.00	3.00
$x_{i,26}$	I feel curious about learning mathematics.	3.63	1.03	1.00	5.00	4.00	4.00

The multivariate normal distribution is assumed in order to preserve the dependence structure among the input variables. For example, variables $x_{i,18}$ and $x_{i,21}$ are expected to be statistically dependent because both relate to the anxiety experienced by the ith student.

This assumption is motivated by the mathematical tractability of the multivariate Gaussian distribution and its ability to represent relationships among variables through the covariance matrix S. Consequently, the dependence structure of the input variables is approximated using a latent multivariate Gaussian model parametrised by the empirical covariance matrix. Although this approximation might not possibly capture every aspect of the observed dependence, it provides a computationally efficient and statistically reasonable basis for Monte Carlo simulation.

The integral introduced above is approximated using Monte Carlo integration:

$$P(\hat{y}=1) \approx \frac{1}{N}\sum_{i=1}^{N} P(\hat{y}_i = 1 \mid X_{i,:}), \tag{2}$$

where $X_{i,:}$ denotes the ith row of the matrix X.

The evaluation of a policy π leads to the conditional probability

$$P(\hat{y}=1 \mid \pi) = \int_{\mathcal{X}} P(\hat{y}=1 \mid x,\pi)P(x \mid \pi)dx. \tag{3}$$

This quantity is estimated by Monte Carlo simulation after fixing selected variables to values consistent with the policy under consideration. The resulting approximation is given by

$$P(\hat{y}=1 \mid \pi) \approx \frac{1}{N}\sum_{i=1}^{N} P(\hat{y}_i = 1 \mid X_{i,:},\pi). \tag{4}$$

To simulate improved teaching support, variables $X_{:,12}$, $X_{:,13}$, $X_{:,14}$, $X_{:,15}$, and $X_{:,16}$ are fixed at their respective 75th percentiles, namely 0.44, 0.97, 0.96, 0.33, and 0.33.

To simulate reduced mathematics anxiety, variables $X_{:,17}$ and $X_{:,20}$ are fixed at their 75th percentiles (1.31 and 0.01, respectively), whereas variables $X_{:,18}$, $X_{:,19}$, and $X_{:,21}$ are fixed at their 25th percentiles (-0.37, -0.24, and -0.38, respectively).

To simulate increased mastery confidence, variables $X_{:,1}$, $X_{:,2}$, $X_{:,3}$, $X_{:,4}$, and $X_{:,5}$ are fixed at their 75th percentiles (0.55, 1.22, 0.39, 0.8, and 0.44, respectively), while $X_{:,6}$ is fixed at its 25th percentile (-1.19).

Finally, the simultaneous implementation of all interventions is simulated by fixing the corresponding variables according to the specifications described above.

IV. The Research Results

The evaluation results suggest that Logistic Regression achieves higher accuracy and precision than both Support Vector Machines (SVMs) and Gaussian Processes (GPs) (Table III). A plausible explanation is the relatively small sample size in relation to the dimensionality of the input space. With only 93 records and 26 input variables, more flexible non-linear models are likely to be more susceptible to overfitting, whereas Logistic Regression provides a simpler decision boundary that might generalise more effectively under limited-data conditions.

TABLE II. 10-Fold Cross-Validation Setting for the Dataset

Fold	Number of Examples for Fitting the Model			Number of Examples for Testing		
	Positive	Negative	Total	Positive	Negative	Total
1	47	36	83	6	4	10
2	47	36	83	6	4	10
3	47	36	83	6	4	10
4	48	36	84	5	4	9
5	48	36	84	5	4	9
6	48	36	84	5	4	9
7	48	36	84	5	4	9
8	48	36	84	5	4	9
9	48	36	84	5	4	9
10	48	36	84	5	4	9

Furthermore, Logistic Regression is particularly suitable when the relationship between the input variables and the log-odds of the target variable can be approximated linearly. Under such conditions, the model typically exhibits lower variance and greater interpretability than more flexible non-linear alternatives.

The results reported in Table III correspond to the hyperparameter configurations selected through stratified 10-fold cross-validation. All models were implemented using the scikit-learn library [32].

For SVMs, the best-performing polynomial kernel had degree one and was therefore equivalent to the linear kernel, which explains the identical results obtained with both kernels. The regularisation parameter C was 7.8×10^{-3} for the linear kernel and 0.25 for the polynomial kernel. For the sigmoid kernel, the optimal values of C and γ were 0.5 and 1.6×10^{-2}, respectively. When the radial basis function (RBF) kernel was employed, the corresponding values were $C = 2$ and $\gamma = 1.9 \times 10^{-3}$.

For Gaussian Processes, the best-performing configuration consisted of the sum of two kernels: an RBF kernel scaled by 3.1×10^{-3} with length scale equal to 1, and a Matérn kernel scaled by 1.9×10^{-3} using the same length scale. For the Rational Quadratic kernel, the length scale and shape parameter α were 1.5×10^{-2} and 1.9×10^{-3}, respectively.

The fold-wise results obtained for Logistic Regression during the 10-fold cross-validation procedure are reported in Table IV. The model achieved an average accuracy of 72.22% and an area under the receiver operating characteristic (ROC) curve of 0.72 (Figure 4), indicating performance substantially better than random classification. In addition, the recall of 91% corresponds to the correct identification of 48 out of 53 positive instances, as shown in the confusion matrix reported in Table V.

The regularisation parameter of the Logistic Regression model was selected using the elbow criterion illustrated in Figure 5. The comparison reported in Table III was conducted using $\lambda = 2 \times 10^{-2}$.

After fitting the Logistic Regression model to the complete dataset using the selected regularisation parameter, Monte Carlo

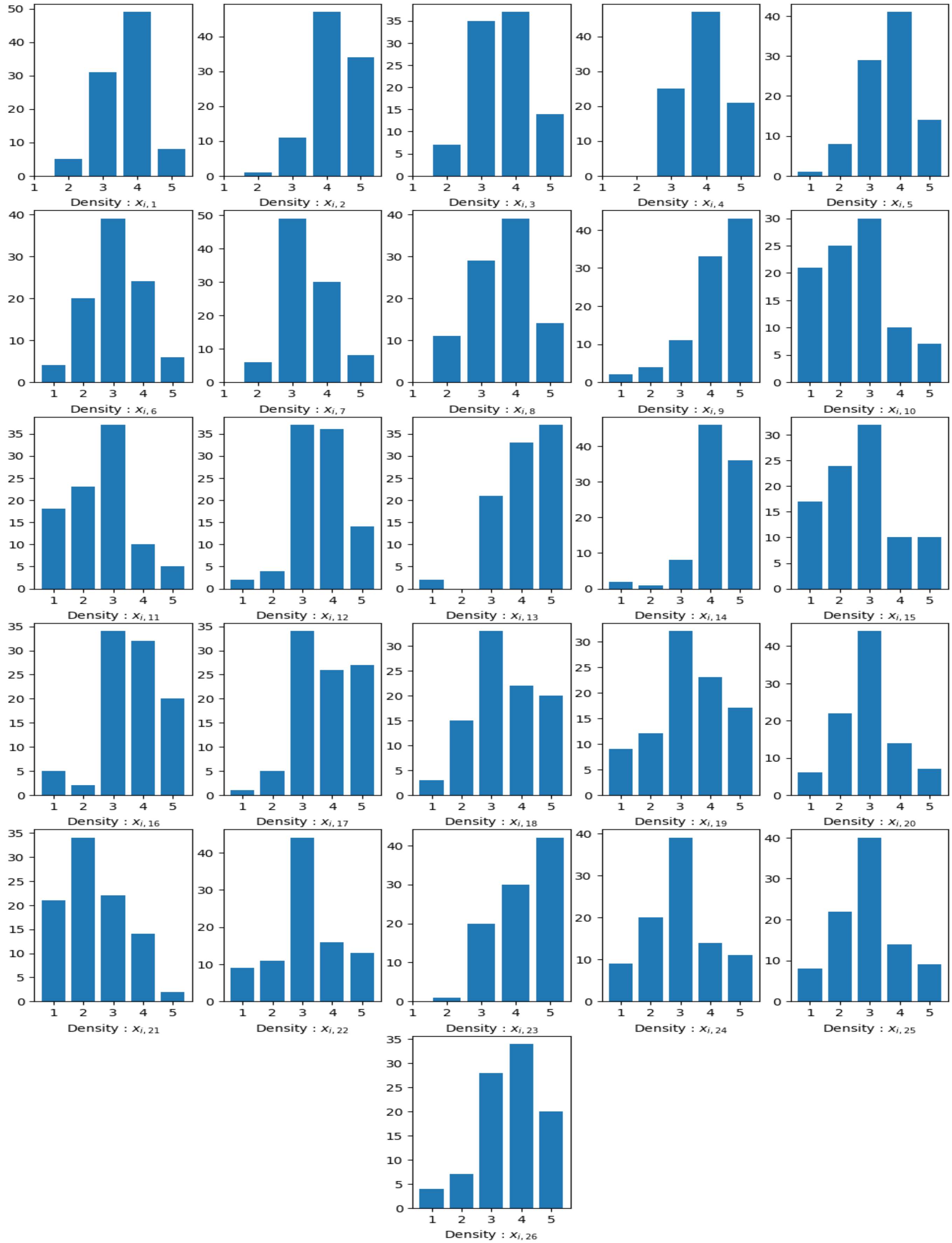

Figure 3. Histogram for each input variable

TABLE III. 10-FOLD CROSS-VALIDATION RESULTS

Machine Learning Model	Mean Precision (%)	Mean Recall (%)	Mean F_1 Score (%)	Mean Accuracy (%)
LR	**70.32**	91.00	78.94	**72.22**
SVM+Lin	67.33	**96.33**	79.00	70.11
SVM+Sig	68.00	**96.33**	**79.50**	71.11
SVM+Pol	67.33	**96.33**	79.00	70.11
SVM+RBF	67.33	**96.33**	79.00	70.11
GP+RQ	68.83	92.67	78.59	71.00
GP+DP	69.86	65.67	66.82	63.33
GP+M+RBF	69.04	83.00	74.80	68.78
GP+M	68.14	83.00	74.16	67.67
GP+RBF	69.39	84.67	75.68	69.78

In this research, we have adopted Logistic Regression (LR), Support Vector Machines (SVM) with Linear (SVM+Lin), Sigmoid (SVM+Sig), Polynomial (SVM+Pol), and Radial Basis Function (SVM+RBF) kernels. Additionally, we employed Gaussian Processes (GP) for classification with Rational Quadratic (GP+RQ), Dot Product (GP+DP), Matern (GP+M), and Radial Basis Function (GP+RBF) Kernels.

TABLE IV. 10-FOLD CROSS-VALIDATION RESULTS CORRESPONDING TO THE LOGISTIC REGRESSION MODEL

Fold	Precision (%)	Recall (%)	F_1 score (%)	Accuracy (%)
1	71.43	83.33	76.92	70.00
2	71.43	83.33	76.92	70.00
3	62.50	83.33	71.43	60.00
4	62.50	100.00	76.92	66.67
5	83.33	100.00	90.91	88.89
6	71.43	100.00	83.33	77.78
7	80.00	85.71	80.00	77.78
8	71.43	100.00	83.33	77.78
9	62.50	100.00	76.92	66.67
10	66.67	80.00	72.73	66.67
Mean	70.32	91.00	78.94	72.22

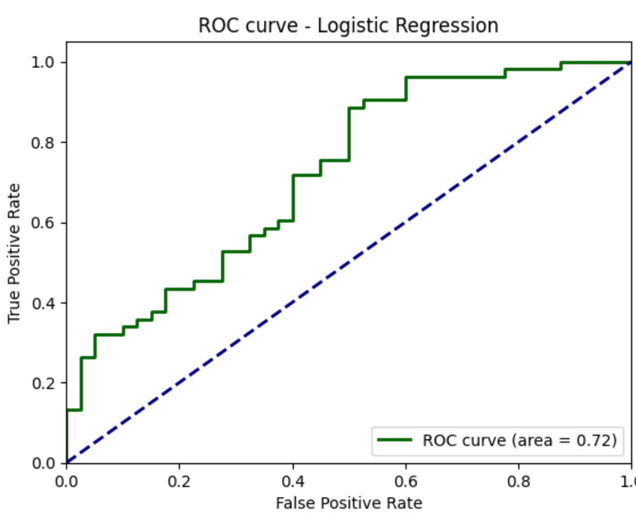

Figure 4. Receiver Operating Characteristic (ROC) curve for the Logistic Regression model

TABLE V. CONFUSION MATRIX ILLUSTRATING THE PERFORMANCE OF LOGISTIC REGRESSION MODEL

True Class	Forecasted Class		Total
	$\hat{y}_i = 0$	$\hat{y}_i = 1$	
$y_i = 0$	19	21	40
$y_i = 1$	5	48	53
Total	24	69	93

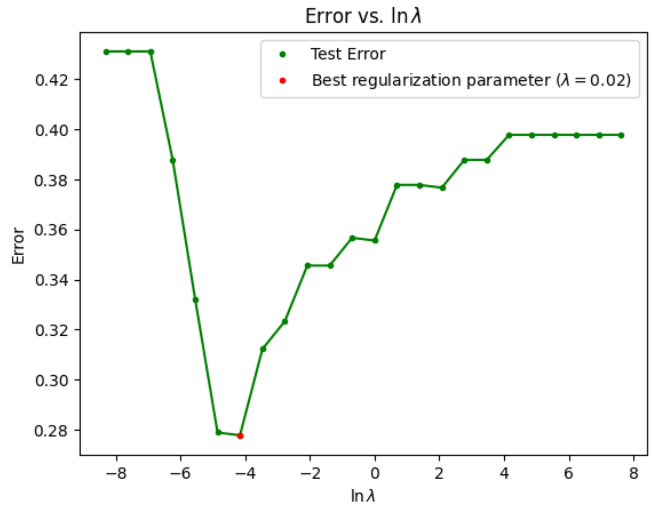

Figure 5. Elbow rule used to select the regularisation parameter for the logistic regression model

TABLE VI. MONTE CARLO SIMULATION OUTCOMES

Policy π	Probability of Failing to Reach the Relational Level (%)	95 % CI	Standard Error
Without Policy	$P(\hat{y} = 1) = 56.94$	[56.80, 57.06]	6.5×10^{-4}
Teaching Support	$P(\hat{y} = 1 \mid \pi) = 54.54$	[54.23, 54.56]	6×10^{-4}
Anxiety Reduction	$P(\hat{y} = 1 \mid \pi) = 53.64$	[53.52, 53.75]	8.1×10^{-4}
Mastery Learning	$P(\hat{y} = 1 \mid \pi) = 53.99$	[53.89, 54.10]	5.2×10^{-3}
All Previous Policies	$P(\hat{y} = 1 \mid \pi) = 47.69$	[47.55, 47.82]	7×10^{-3}

CI stands for Confidence Interval

simulation was employed to explore a broader region of the input space. The resulting estimates are reported in Table VI. All intervention scenarios reduce the probability of failing to attain the relational level. Among the individual interventions, anxiety reduction yields the largest decrease in this probability, whereas the simultaneous implementation of all interventions produces the greatest overall reduction.

The distributions shown in Figures 6–10 are consistent with the estimates reported in Table VI. In particular, Figure 6, corresponding to the baseline scenario, exhibits a distribution centred at higher predicted probabilities than the intervention scenarios. By contrast, Figure 10, which corresponds to the simultaneous implementation of all interventions, is centred at lower predicted probabilities, indicating the lowest estimated risk of failing to attain the relational level. Furthermore, the Monte Carlo estimates exhibit convergence in all scenarios, as illustrated in Figures 11–15.

V. ANALYSIS OF THE RESULTS

In this section, the performance of the classification models is analysed and the implications of the Monte Carlo simulation results are discussed.

The precision and accuracy of the classifiers are below 80%, suggesting that the predictive information contained in the dataset is limited, albeit still informative, as both metrics exceed

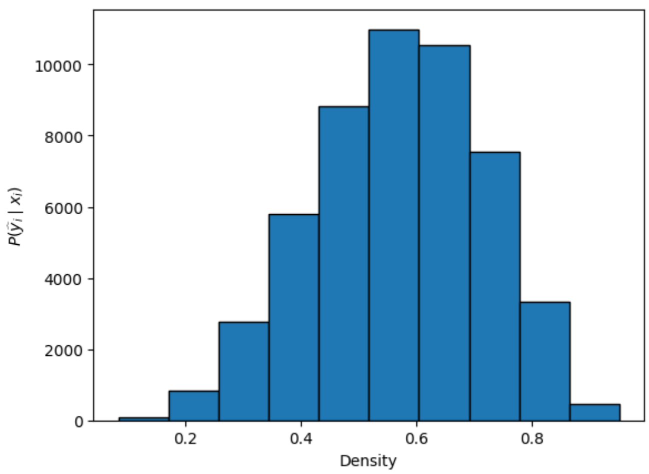

Figure 6. Distribution of probabilities $P(\hat{y}_i = 1 \mid x_i)$ obtained from the simulation without a policy.

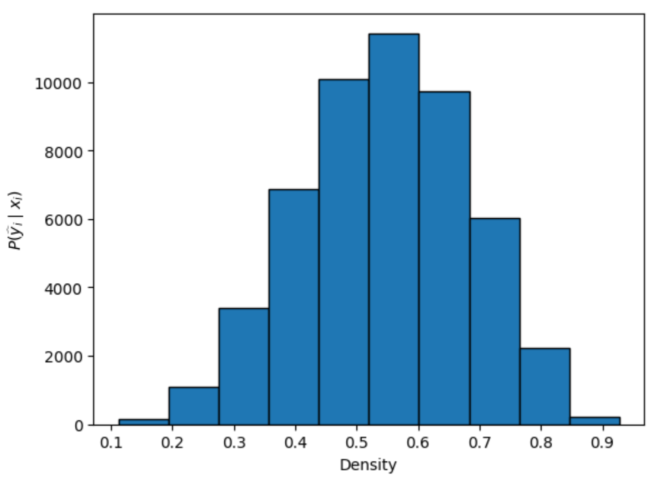

Figure 7. Distribution of probabilities $P(\hat{y}_i = 1 \mid x_i)$ obtained from the simulation for the application of the teaching support policy.

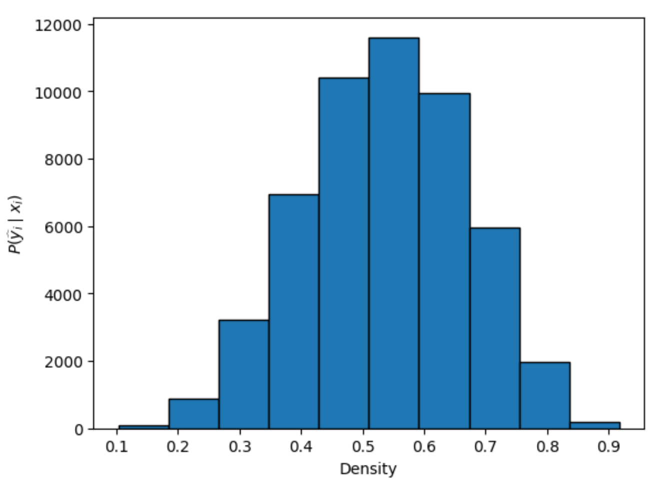

Figure 8. Distribution of probabilities $P(\hat{y}_i = 1 \mid x_i)$ obtained from the simulation for the application of anxiety reduction policy.

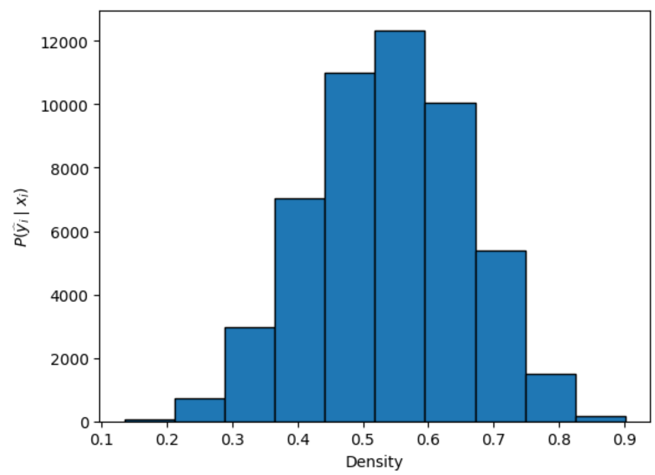

Figure 9. Distribution of probabilities $P(\hat{y}_i = 1 \mid x_i)$ obtained from the simulation for the application of mastery learning policy.

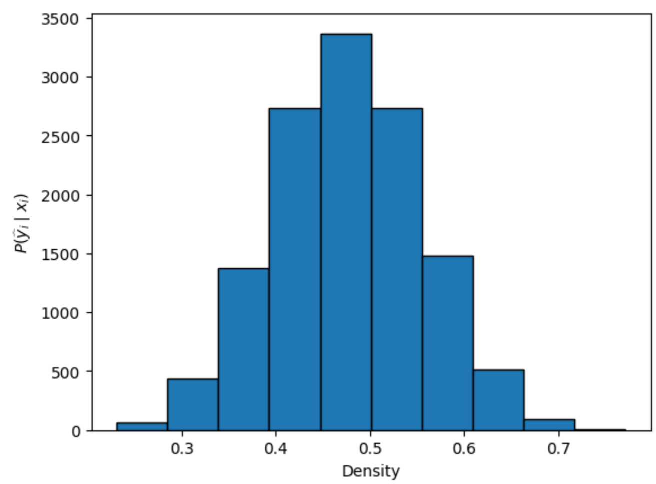

Figure 10. Distribution of probabilities $P(\hat{y}_i = 1 \mid x_i)$ obtained from the simulation for the application of all policies.

60%. Furthermore, the recall of Logistic Regression and the other models exceeds 90%, indicating that most students who are unlikely to attain the relational level are correctly identified. In this context, failing to identify a struggling student may have more serious consequences than providing additional support to a student who may not strictly require it.

The moderate predictive performance of the models is reflected in the distributions shown in Figures 6–10. Models with greater discriminative power might produce more polarised distributions, concentrating probability mass closer to the extremes. Nevertheless, such levels of uncertainty are not uncommon in educational psychology research, where human perceptions and behaviours are often characterised by substantial variability.

The combination of machine learning and Monte Carlo simulation provides a virtual environment for evaluating educational interventions without the time and resource requirements associated with large-scale experimental studies. Notably, the

 ISBN: 978-1-68558-402-3

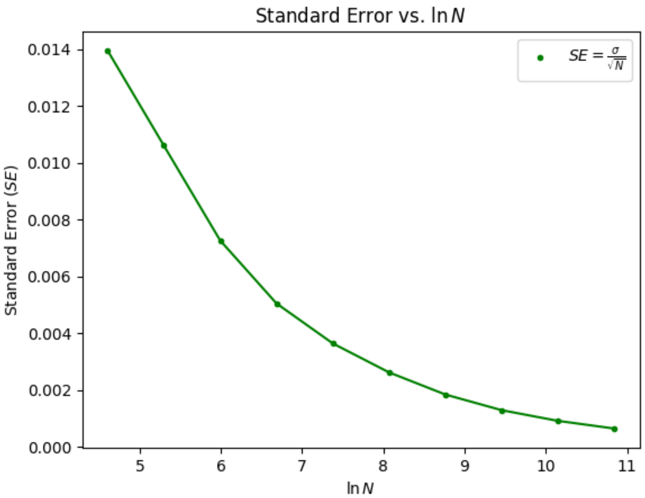

Figure 11. The simulation converges to the expected probability $P(\hat{y} = 1) = 56.94\%$ when no policy is applied as $N = 51,200$, with a standard error of 6.5×10^{-4}. The result lies within the 95% confidence interval of [56.80, 57.06].

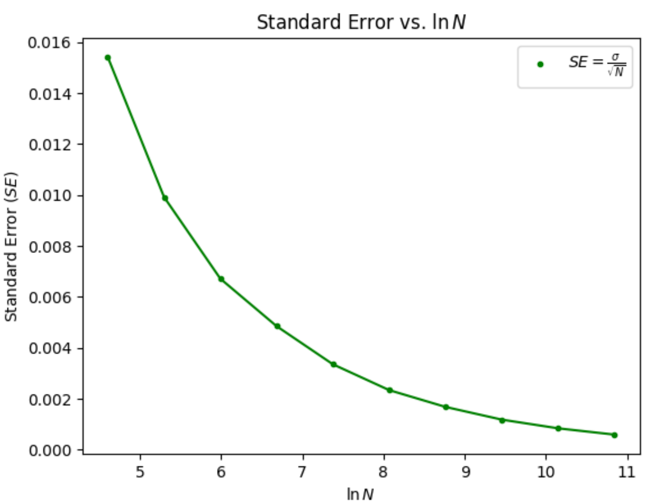

Figure 12. The simulation converges to the expected probability $P(\hat{y} = 1 \mid \pi) = 54.54\%$ when teaching support policy is applied as $N = 51,200$, with a standard error of 6×10^{-4}. The result lies within the 95% confidence interval of [54.23, 54.56].

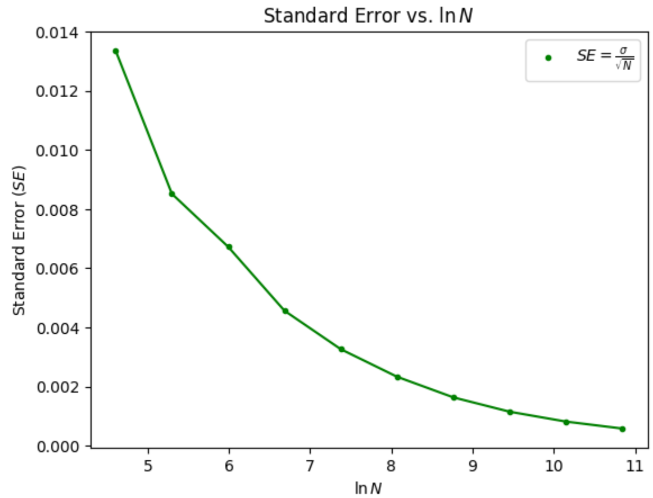

Figure 13. The simulation converges to the expected probability $P(\hat{y} = 1 \mid \pi) = 53.64\%$ when anxiety reduction policy is applied as $N = 51,200$, with a standard error of 8.1×10^{-4}. The result lies within the 95% confidence interval of [53.52, 53.75].

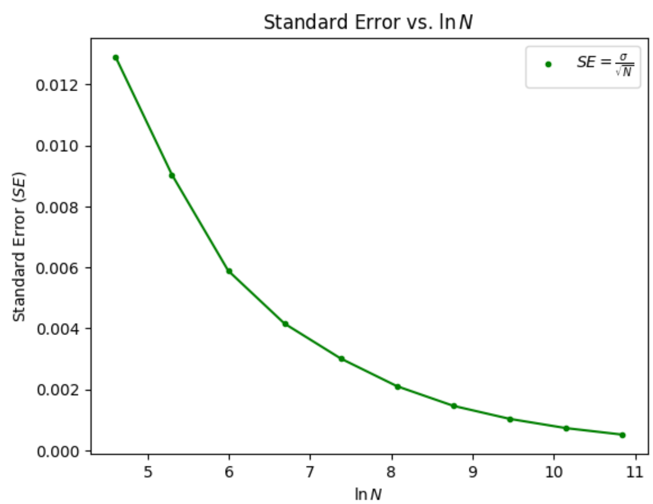

Figure 14. The simulation converges to the expected probability $P(\hat{y} = 1 \mid \pi) = 53.99\%$ when mastery learning policy is applied as $N = 51,200$, with a standard error of 5.2×10^{-3}. The result lies within the 95% confidence interval of [53.89, 54.10].

probability estimated under the baseline simulation is very close to the empirical proportion of students who failed to attain the relational level in the observed dataset. Specifically, $|P(\hat{y} = 1) - P(y = 1)| = |56.94 - 56.99| = 0.05$.

The small discrepancy between these values suggests that the multivariate Gaussian assumption provides a reasonable approximation for the purposes of simulation.

Despite these encouraging results, further research is required to improve predictive performance and strengthen the validity of the proposed framework. In particular, collecting a larger dataset and investigating lower-dimensional representations of the input space may improve model robustness and predictive accuracy.

The simulation outcomes provide evidence that might assist decision-makers in prioritising interventions aimed at helping students attain at least the relational level in engineering mathematics. Within the context of the present study, reducing mathematics anxiety appears to yield the largest individual reduction in risk, followed by increasing mastery confidence and improving teaching support. The greatest reduction is achieved when all interventions are implemented simultaneously.

VI. Conclusion and Future Work

In summary, the Monte Carlo simulations estimate that the probability of failing to attain the relational level in mathematics courses within the Bachelor of Systems Engineering programme

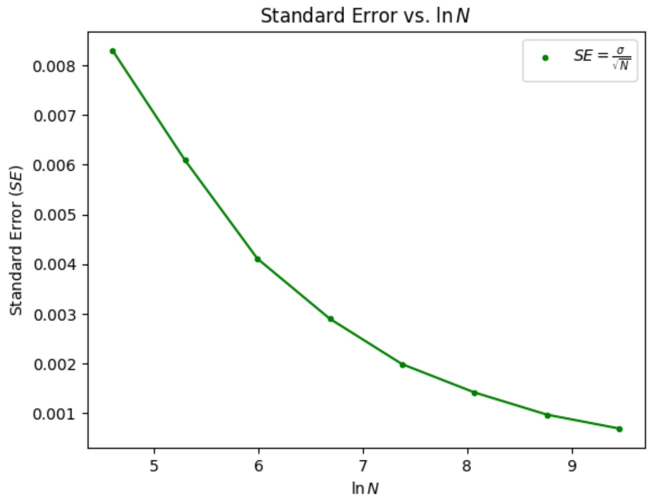

Figure 15. The simulation converges to the expected probability $P(\hat{y} = 1 \mid \pi) = 47.69\%$ when all policies are applied as $N = 12,800$, with a standard error of 7×10^{-3}. The result lies within the 95% confidence interval of [47.55, 47.82].

at the University of Córdoba is 56.94% (95% CI: [56.80, 57.06]).

The simulation of improved teaching support reduces this probability to 54.54% (95% CI: [54.23, 54.56]). This result suggests that students' perceptions of instructional support might play a relevant role in their ability to attain relational-level understanding. In particular, practices that strengthen students' confidence and provide constructive feedback may contribute to reducing the risk of underachievement.

Simulating reduced mathematics anxiety further decreases the estimated probability to 53.64% (95% CI: [53.52, 53.75]). This finding highlights the potential importance of emotional and affective factors in engineering mathematics education. Consequently, instructional strategies that minimise unnecessary stress and support students when confronting challenging material may help mitigate the risk of failing to attain the relational level.

Similarly, improving mastery-learning beliefs reduces the estimated probability to 53.99% (95% CI: [53.89, 54.10]). This result is consistent with the self-efficacy literature, which emphasises the importance of mastery experiences in shaping students' beliefs about their capabilities. Learning environments that promote incremental progress, sustained engagement, and repeated experiences of success might therefore contribute to improved outcomes.

The largest reduction is obtained when all interventions are implemented simultaneously, lowering the estimated probability to 47.69% (95% CI: [47.55, 47.82]). This result suggests that the factors considered in this research should not be viewed in isolation, as their combined influence appears to be greater than that of any individual intervention.

Overall, the results demonstrate that the combination of machine learning and Monte Carlo simulation provides a computational framework for evaluating educational interventions prior to their implementation. Such an approach enables decision-makers to explore the potential consequences of alternative strategies aimed at improving students' attainment of relational-level understanding in engineering mathematics.

Future work should focus on expanding the dataset in order to obtain more accurate and robust estimates. A larger sample would also facilitate the analysis of all levels of the SOLO taxonomy rather than restricting the problem to a binary classification task. In addition, the questionnaire employed in this study should be compared with alternative self-efficacy instruments reported in the literature.

Further research should also investigate dimensionality-reduction techniques, including Principal Component Analysis, Non-negative Matrix Factorisation, t-distributed Stochastic Neighbour Embedding, and Autoencoders, in order to identify latent representations of the input space and potentially improve predictive performance.

Finally, the intervention strategies examined through simulation should be implemented and evaluated empirically in educational settings to assess the extent to which the predicted effects are observed in practice.

Acknowledgement

Caicedo-Castro thanks the Lord Jesus Christ for blessing this project. The authors thank Universidad de Córdoba in Colombia for supporting this study. They also thanks all students who collaborated by answering the survey conducted for collecting the dataset used in this study. Finally, the author thanks the anonymous reviewers for their comments that contributed to improve the quality of this article.

References

[1] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice-Hall series in social learning theory). Englewood Cliffs, NJ: Prentice-Hall, 1986, ISBN: 013815614X.

[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.

[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.

[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.

[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.

[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.

[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.

[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 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*, 2023, pp. 40–51.

[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.

[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.

[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] N. E. Betz and G. Hackett, ‘The Relationship of Mathematics Self-Efficacy Expectations to the Selection of Science-based College Majors’, *Journal of Vocational Behavior*, vol. 23, no. 3, pp. 329–345, 1983.

[16] F. Pajares and D. M. Miller, ‘Mathematics self-efficacy and mathematics performances: The need for specificity of assessment’, *Journal of Counseling Psychology*, vol. 42, no. 2, pp. 190–198, 1995.

[17] F. Pajares and J. Kranzler, ‘Self-Efficacy Beliefs and General Mental Ability in Mathematical Problem-Solving’, *Contemporary Educational Psychology*, vol. 20, no. 4, pp. 426–443, 1995.

[18] F. Pajares and L. R. Graham, ‘Self-Efficacy, Motivation Constructs, and Mathematics Performance of Entering Middle School Students’, *Contemporary Educational Psychology*, vol. 24, no. 2, pp. 124–139, 1999.

[19] I. L. Nielsen and K. A. Moore, ‘Psychometric Data on the Mathematics Self-Efficacy Scale’, *Educational and Psychological Measurement*, vol. 63, no. 1, pp. 128–138, 2003.

[20] M. Coe, R. McFadden, B. Pratt-Sitaula and E. Baer, ‘Geoscience Mathematics Self-Efficacy Scale (GeoMSES) for Majors-Level Undergraduates: Psychometric Data’, *Education Sciences*, vol. 16, no. 2, p. 288, 2026.

[21] E. Rhew, J. Piro, P. Goolkasian and P. Rappold, ‘The effects of a Growth Mindset on Self-Efficacy and Motivation’, *Cogent Education*, vol. 5, no. 1, p. 1 492 337, 2018.

[22] 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*, ser. International Conference on Generative Pretrained Transformer Models and Beyond, Venice, Italy: IARIA: International Academy, Research, and Industry Association, 2025, pp. 15–22, ISBN: 978-1-68558-287-6.

[23] I. Caicedo-Castro, R. Castro-Púche and S. Castaño-Rivera, ‘Estimating the Risk of Failing Physics Courses through Machine Learning and Monte Carlo Simulation’, *International Journal on Advances in Systems and Measurements*, vol. 19, no. 1 and 2, 2026.

[24] 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*, ser. International Conferences on Pervasive Patterns and Applications, Valencia, Spain: IARIA: International Academy, Research, and Industry Association, 2025, pp. 15–22, ISBN: 978-1-68558-263-0.

[25] 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.

[26] M. Hegazi, K. Chebil and S. A. Alqahtanib, ‘A monte carlo-based model for predicting student performance’, 2026.

[27] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (Adaptive Computation and Machine Learning). Cambridge, MA, USA: MIT Press, 2006, ISBN: 026218253X.

[28] J. C. Platt, ‘Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods’, in *Advances in Large Margin Classifiers*, MIT Press, 1999, pp. 61–74.

[29] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2013, ISBN: 9780262018029.

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

[31] 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.

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

