Leveraging Machine Learning to Forecast Candidate Selection Outcomes

Chaimae Ouhaddou*, Asmaâ Retbi, and Samir Bennani

RIME TEAM, Mohammadia School of Engineers (EMI) Mohammed V University in Rabat, Rabat, Morocco
Email: chaimae.ouhaddou@research.emi.ac.ma (C.O.); retbi@emi.ac.ma (A.R.); sbennani@emi.ac.ma (S.B.)
*Corresponding author
Manuscript received December 27, 2024; revised February 11, 2025; accepted April 28, 2025; published August 22, 2025

***Abstract*—Machine learning has emerged as a transformative tool in education, driving personalized learning experiences. This study focuses on its application in the educational sector, particularly through the lens of peer learning systems. Our research presents a systematic approach to predicting candidate success during the selection phase, also known by the immersive evaluation phase. In this context, five distinct machine-learning algorithms (Decision Trees, Support Vector Machines, Logistic Regression, Random Forest, and Stacking) were employed to assess their effectiveness in classifying candidates as retained or rejected. Additionally, we explored attribute importance to provide insights into the key factors influencing candidate selection. The findings reveal that Stacking (Random Forest (RF), K-Nearest Neighbors (KNN), Naive Bayes (NB) + XGboost) model proved to be the most effective after evaluating performance metrics using bootstrapping methods.**

***Keywords*—machine learning, predictive modelling, academic performance, student success prediction**

I. Introduction

In recent years, Machine Learning (ML) has made progress across various sectors with the goal of transforming industries like finance, healthcare and logistics. One of its most effective applications is in the education field, where it is used to study and predict student performance throughout the academic year or during the selection phase. Utilizing data mining algorithms on datasets can be advantageous in identifying students' performance at an early stage and enable educational institutions to implement strategies [1]. In this study, we aim to explore the concept of peer-learning schools, particularly those that utilize systems based on the immersive evaluation phase in their admission process.

Traditionally, the candidate selection process may rely on subjective evaluations generally based on written exam tests, and in some cases face-to-face interviews. However, peer-learning schools focus on organizing a selection phase that typically lasts between 4 to 6 weeks, where candidates are evaluated through multiple assessments. This approach, known as immersive evaluation, involves placing candidates in real or simulated learning environments where they take part in problem-solving tasks, group projects, and peer assessments. While delving deeper into this subject, we found that many studies have focused on introducing ML into the traditional selection processes, however, our paper intend to explore the selection process, which represents an interesting niche for the application of ML. Specifically, we investigate the relationship between various candidate features that are gathered through the 4 weeks and their impact on predicting success during this phase.

The decision to apply machine learning to this context stems from several challenges with the current system. One major issue is the variation in how candidates are assessed. Despite the existence of clear criteria and rating scale, evaluators may interpret them in a different way which can lead to varied judgments for the same candidate. In addition, the significant volume of data related to applicants makes it difficult to process and evaluate candidate performance and can result in potential errors. Evaluators may also overlook crucial factors like problem solving skills or creativity and focus more on attributes such as previous academic performance or level. By applying ML to this context, our paper aims to achieve a fairer and data-driven selection process while highlighting the innovative perspective and system of these schools.

The rest of this paper is structured as follows: We begin by describing the purpose of the study through four different questions related to the application of ML in this context. In the second section, we provide an overview of the existing review and survey articles related to students, the concept of immersive evaluation and peer learning system. Next, the methodology section follows detailing the data collection and discretization step, the feature selection step, and the machine learning techniques applied. In section IV, we present the findings by assessing the performance metrics. Section V provides a discussion addressing the research questions posed in the study. Finally, the paper concludes with a summary of the work and directions for future work.

Purpose of the Study: The purpose of this study is to explore the application of ML in predicting student success by the end of the immersive evaluation phase in a peer learning school. Our focus is on analyzing different candidate features and studying their impact on students' success. To this end, we have formulated four questions to outline the objective of this paper and segmented these inquiries into two main sections. The first set focuses only on the algorithms used for the application of ML, and the analysis of their effectiveness and accuracy. The second set aims to identify the most influential student attributes that contribute directly to success, with particular attention to the role of soft skills.

- What machine learning techniques are most effective in predicting candidate success in a peer learning school?
- How do feature selection techniques affect the performance of machine learning models?
- Which candidate attributes are the most significant predictors of success during the evaluation process?
- What role do soft skills, such as collaboration and communication, play in predicting candidate success?

Fig. 1 illustrates how strategic questioning plays a vital role in enhancing model precision and improving predictive outcomes.

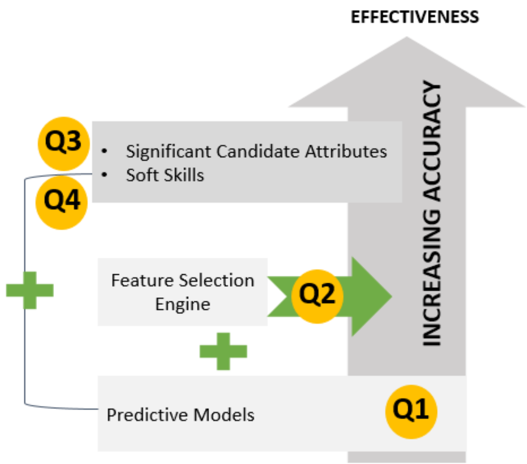

Fig. 1. The role of key questions in increasing predictive model accuracy.

II. Literature Review

A. Student Selection Using ML

Many studies have explored the use of ML for predicting student performance across diverse educational contexts. These studies vary in terms of datasets, techniques, and evaluation metrics. Appendix Table A1 provides an overview of selected works, highlighting their sources, objectives, datasets, ML methods used, and results. Among the reviewed works, study [2] compares ensemble-learning methods by using a dataset of 245 students. As a result, authors demonstrate that XGBoost outperformed in terms of predictive accuracy. In study [3], researchers applied five methods: K-Nearest Neighbor (KNN), Decision Trees (DT), Support Vector Classifier (SVC), Random Forest, Gradient Boosting, and Linear Discriminant, to a dataset of 1179 students, the study identified KNN and DT as the best methods, with accuracies of 89,74% and 94,44% respectively. In a different way, paper [4] analyzes 70 studies and provides a review of the application of ML techniques in this context. Authors observed that Artificial Neural Network (ANN) and Support Vector Machine (SVM) were the most frequently used methods followed by Collaborative Filtering (CF), DT, and Naïve Bayes.

On the other hand, study [5] focused on predicting students' GPA based on both personal and academic characteristics while using a dataset of 525 students. Authors used Naïve Bayes, SMO, Multilayer Perceptron (MLP), J48, and Random Forest and concluded that Naïve Bayes had the lowest accuracy, whereas ANN consistently outperformed J48 in accuracy predictions. For [6], the dataset includes students' information from five classes (a total of 309 students) to test the predictive power of ML techniques in forecasting student performance. It was concluded that DT had the highest accuracy. Study [7] analyzed Scholastic Assessment Test (SAT) Math scores using linear regression, DT, and Naïve Bayes applied to a dataset of 403 students. Interestingly, Naïve Bayes achieved the best accuracy.

In study [8], authors focus on developing a machine learning-based system to evaluate high school student performance using a dataset of 459 students. As for study [9] where authors use a larger dataset (1000 students), the focus was detecting the most significant factor influencing student success by finalizing those factors like absence rate and risk score had minimal impact. Papers [10] and [11] used several machine learning models, including Random Forest, Decision Tree, Neural Network, and others, to determine the most effective model for predicting student performance. Both studies proved that Decision Tree were the top-performing models with accuracy superior to 96%. To go further, study [12] proposes a novel method for predicting students' future performance based on current and past academic achievements, using a dataset of 1196 students. The ensemble-based progressive prediction method used in the study demonstrated superior performance compared to traditional models. In [13], authors explored forecasting final grades in first-semester courses with a dataset of 1282 students and the use of ML models such as Decision Tree (J48), SVM, and Naïve Bayes. The findings showed that SVM ensemble models produced greater accuracy. Where in paper [14], the authors proposed early segmentation of students based on performance levels, by analyzing a large dataset of 2459 students. As a result, they found that Random Forest has to be the best classification techniques. Similarly, study [15] developed an adaptive recommendation system to guide 725 students in choosing the best academic program. The authors applied SVM, KNN, and RF techniques, with the Quadratic Discriminant Analysis (QDA) algorithm achieving the highest F-measure of 0.91.

Since analyses on a fairly large dataset provide better results, the authors of paper [16] applied tree-based methods and ML models to analyze the PISA 2015 test scores of students across nine countries, with a total sample size of 97,000. The study demonstrated that tree-based models enhanced the predictive power of linear regression models. In the same level, study [17] used random forest techniques on a large dataset of 165,715 students to predict dropout risk. The predictive model achieved an excellent accuracy of 0.95, making it a strong tool for dropout prediction. Both Paper [18] and [19] present a survey on ML techniques used to predict student dropout rates in online courses and classify students. The results of the studies showed that BP achieved the highest performance with an accuracy of superior to 87%.

Study [20] conducted another systematic literature review to explore the use of machine learning algorithms in predicting student performance, and provide a detailed analysis of trends, methodologies, and challenges in this area. The review examines how factors such as dataset and features selection influence the accuracy of predictions. Similarly, in scope to study [20], paper [21] focuses on summarizing key developments in the field of student performance prediction using ML. Finally, papers [22–24] aim to predict student performance by the application of numerous machine learning techniques such as Decision Trees, Naïve Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbor and Neural Networks. Among these, the logistic regression and ANN classifiers were found to be the most accurate in predicting final grades.

Despite the extensive research on ML applications in student performance prediction, it has been noted that most studies focus on the use of structured academic data specifically grades and test scores, rather than behavioral and skill-based data. Our research extends this work by applying ML techniques to an immersive evaluation context, analyzing real-time engagement, and problem-solving ability to enhance candidate selection accuracy.

B. Immersive Evaluation Concept

Since we have explored the concept of immersive evaluation, it was considered important to define its steps and main rules. Hence, this selection process, which is generally implemented in peer-learning Schools, is designed to evaluate and focus on candidates' skills and learning ability. It begins with an initial screening phase, where candidates get through basic assessments and conduct interviews with the committee to present their personal projects. Once successfully completed, candidates enter an intensive four-week bootcamp, where they are immersed in competitive challenges. The regulations during this phase require students to work individually but also in-group projects to assess their collaboration, teamwork and communication. In addition, this phase includes specific evaluation to quantify the ability of students in solving problems, their adaptability and perseverance. At the end of the four weeks, student performance is measured using metrics such as task completion rate, peer reviews, engagement, etc. Those who excel across these areas are offered admission to the school.

The application of machine learning in this immersive evaluation process offers great potential for enhancing the accuracy of the evaluation system. Predicting students' performance across various phases of immersive evaluation enables institutions and mentors to identify patterns in individual and group tasks as well as assessing skill-based attributes like perseverance and problem-solving ability. Furthermore, advanced ML techniques such as Natural Language Processing (NLP) could also be used to analyze interview responses or project presentations.

C. Peer-Learning and Student Performance: The Role of Self-efficacy

Through experience, peer learning has positively affected student performance. This concept fosters environments that enhance motivation, engagement, and strengthen understanding of academic content. Research consistently highlights the positive effects of peer learning on student performance across various academic domains. Many studies demonstrate that peer learning groups foster a collaborative space where students support each other in understanding difficult concepts and solving problems while developing their thinking skills. In this context, many papers provide tangible results of how peer learning can improve grades and lead to better academic outcomes.

Study [25] demonstrates that students in the peer-learning group achieved higher final exam scores and semester grades compared to those in the lecture-based group.

Furthermore, authors in [26] focused on nursing education and concluded that institutions should utilize peer-led learning as a supplemental strategy to enhance academic performance, critical thinking, and confidence in nursing coursework. Several studies have explored the interaction between peer learning and self-efficacy. As a result, it was found that students who participated in peer learning environments showed higher self-efficacy compared to those who worked individually during their academic journey.

While several studies emphasize the role of peer learning in academic success, there is limited research on how ML can assess and enhance peer-learning experiences specifically in immersive evaluation environments. This study addresses that gap by integrating ML models to analyze student behavioral and skill-based data.

III. Materials and Methods

In this paper, we adopted a structured methodology for analyzing and predicting candidate success in the immersive evaluation process. Fig. 2 outlines three main steps: Data Collection & Preparation, Feature Engineering and Model Training, and Performance Analysis.

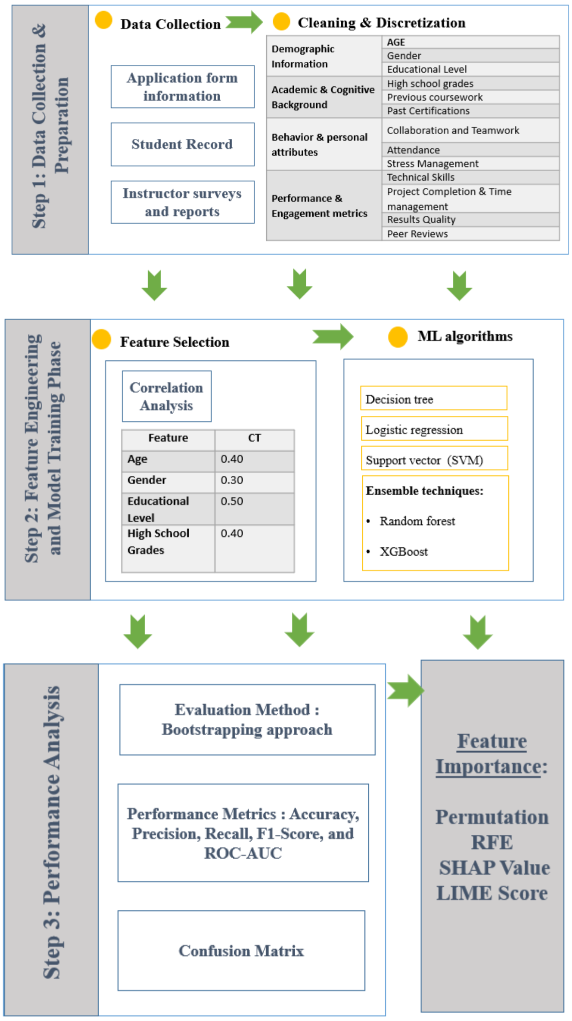

Fig. 2. Step-by-step process for student performance analysis using machine learning techniques.

1) Data Collection & Preparation: this phase includes collecting information from application forms, student records, and instructor surveys. The data is cleaned and structured into four categories. Then we worked on data discretization to simplify the analysis. Next, these metrics are prepared for threshold-based analysis.
2) Feature Engineering and Model Training: this second phase highlights the feature selection method that is Correlation Analysis. Also, we selected five machine learning algorithms including decision tree, logistic regression, support vector machines SVM, and random forest. Additionally, we incorporated Stacking as an ensemble hybrid method to further enhance predictive performance
3) Performance Analysis: The last phase evaluates the model's performance through bootstrapping methods. We

used various performance metrics such as accuracy, precision, recall, F1-Score, and Receiver Operating Characteristic – Area Under the Curve (ROC-AUC). To reinforce our measurements, we also applied confusion in this phase.

The study also focuses on measuring Feature Importance generated by permutation and Recursive Feature Elimination RFE techniques, to rank the significance of each feature in contributing to prediction.

Fig. 2 illustrates the step-by-step process of analyzing student performance using machine learning. The proposed workflow involves the step of collecting and preprocessing relevant data, followed by feature selection step and finally the model training step.

A. Data Collection

The data for this research were collected from a Moroccan peer-learning school that operates under the immersive evaluation concept. Our data collection process covered three distinct sessions (in 2024 and 2025). In the first session, we gathered information from 87 candidates. The second session, conducted in August, included data from 132 candidates and 102 candidates of the third session. In total, we collected a dataset including information about 321 students. While gathering information, we ensured that we went beyond the basic demographic details, such as age, gender or educational level, and focused on additional data points that could prove valuable for our analysis. This approach allowed us to identify and categorize four distinct groups of attributes: demographic information (age, gender, educational level), academic and cognitive background, behavior and personal attributes and Performance & engagement metrics.

Fig. 3 illustrates the 13 key attributes collected at various stages, beginning with the pre-selection screening, followed by immersive evaluations, and concluding with the assessment phase. Each attribute is crucial in contributing to model predictions.

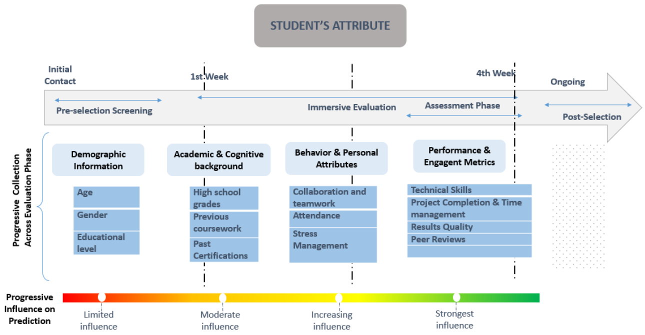

Fig. 3. Timeline of attribute collection and their increasing influence on final prediction: From pre-selection to post-selection.

B. Discretization and Coding Schemes

After this stage, we proceeded to the discretization of student metrics into a rating scale of 4 values (from 0 to 3) to standardize the representation of data (especially categorical data). The main goal is to reduce complexity of the second step of the framework and enable the algorithms to make accurate and precise predictions about student success in the immersive evaluation step.

Table 1 presents a categorization of student attributes and their corresponding levels. Each attribute is grouped into specific categories, with varying levels that represent different degrees of proficiency or involvement.

Table 1. Categorization of attributes and levels

Group of Attribute	Attributes	Values/levels	Description
Demographic Information	AGE	18 – 20 (0)	The schools accept candidates between the ages of 18 and 30.
		20 – 25 (1)	
		>25 (2)	
	Gender	Male (0)	-
		Female (1)	
	Educational Level	Non-Bac (0)	Admission is based purely on merit, thus, it is not necessary to have a baccalaureate (high school diploma) to be admitted.
		Bac+ 2 (1)	
		Bac +3 (2)	
		Bac +5 (3)	
Academic & Cognitive Background	High school grades	High (3)	High grades and strong academic performance in high school.
		Medium (2)	Mid-range grades and average academic performance
		Low (1)	Poor grades and low academic performance
		Fail (0)	Failed to meet academic standards in high school.
	Previous coursework	Yes (1)	The Candidate has relevant prior coursework.
		No (0)	Candidate does not have relevant prior coursework.
	Past Certifications	High (3)	Candidate holds high-value certifications (more than 5)
		Medium (2)	Candidate holds some certifications, but they are mid-level
		Low (1)	Candidate holds basic or low-level certifications
		Fail (0)	Candidate has not any certifications (or failed to obtain them)
Behavior & Personal Attributes	Collaboration and Teamwork	High (3)	Strong capacity in working on a team and takes initiative I tasks
		Medium (2)	Adequate collaborator but may not lead in some situations
		Low (1)	Struggles to work with others or show minimal collaboration
		Fail (0)	Uncooperative candidate
	Attendance	Poor (0)	Candidate often misses sessions (above 2 absences during the immersive phase)
		Good (1)	Candidate attends sessions regularly (below 2 absences during the immersive phase)
	Stress Management	High (3)	Handles stress extremely well
		Medium (2)	Can manage stress, but may have occasional difficulties
		Low (1)	Struggles to manage stress
		Fail (0)	Fails to manage stress
Performance &	Technical Skills	High (3)	Demonstrates high-level technical skills, with strong understanding of topics

Engagement Metrics		Medium (2)	Has solid fundamentals but struggles with advanced topics
		Low (1)	Basic understanding of concepts
		Fail (0)	Performs poorly in technical tasks.
	Project Completion & Time management	High (3)	Completes all or nearly all projects and respect deadlines
		Medium (2)	Completes a moderate number of projects, and respect deadlines
		Low (1)	Completes a few basic projects
		Fail (0)	Does not complete projects or misses deadlines
	Results Quality	High (3)	Results are efficient and follow best practices.
		Medium (2)	Results are good but lacks efficiency or clarity;
		Low (1)	Results are functional but lacks organization,
		Fail (0)	Results are non-functional or poorly structured
	Peer Reviews	High (3)	Receives positive feedback, with helpful content
		Medium (2)	Mixed feedback, positive, but needs improvement
		Low (1)	Frequently receives negative feedback
		Fail (0)	Consistently poor peer reviews

C. Integration Of Threshold-Based Analysis

After discretizing student attributes, we established thresholds for candidate performance metrics, in order to decide whether it is a retention or rejection respecting the original core system. This approach, known as threshold-based analysis, involves setting predefined cutoff values for key performance indicators to systematically classify candidates based on their results. We set up a **success threshold** by determining the number of metrics for which a candidate's performance falls below acceptable level.

- Rejected = (4 or more "fail" or "low")
- Retained = (Less than 4 "fail" or "low" metrics)

Table 2 shows an example with 4 performance metrics:

Table 2. Applying threshold to 4 performance metrics

Stress Management	**LOW**	**Peer Reviews**	**LOW**
Collaboration and Teamwork	FAIL	Results Quality	Medium

In this scenario, since the candidate has failed or scored low in **3 out of 4 metrics**, they would still be considered for retention based on the model's prediction.

While this rule-based method follows the logic of traditional evaluation systems, it cannot handle complex or non-linear relationships between multiple factors. Therefore, it is used as a baseline to show how much better machine learning methods perform, as demonstrated in the evaluation section.

D. Feature Selection Methods

Table 3. Correlation of features with the target variable

Feature	Correlation with Target	Selected (Yes/No)
Age	0.40	Yes
Gender	0.30	Yes
Educational Level	0.50	Yes
High School Grades	0.40	Yes
Previous Coursework	0.75	Yes
Past Certifications	0.60	Yes
Collaboration and Teamwork	0.80	Yes
Attendance	0.55	Yes
Stress Management	0.60	Yes
Technical Skills	0.85	Yes
Project Completion	0.75	Yes
Results Quality	0.90	Yes
Peer Reviews	0.70	Yes

In this section, we employ correlation analysis as a straightforward and effective method for feature selection, this method allows us to evaluate the linear relationship between individual features and the target variable, facilitating the identification of the most relevant features for our selected models. We used Pearson's correlation coefficient to measure the strength and direction of the relationship between each feature and the target variable.

The results of the correlation analysis are summarized in Table 3:

All features listed exhibit a positive correlation with the target, suggesting their influence on the outcome.

IV. Result and Discussion

A. Machine Learning Application

In our context, five algorithms were chosen based on their strengths and suitability for classification tasks, allowing for a comprehensive comparison of different approaches. The algorithms are: **decision tree**, **logistic regression**, **Support Vector Machine (SVM) and random forest**. To further enhance performance, we implemented a stacking approach, which combines multiple machine learning models to improve overall prediction performance. In stacking, predictions from several base models are used as inputs to a meta-model, which then makes the final prediction.

The base models selected for stacking were Random Forest, K-Nearest Neighbors (KNN), and Naive Bayes. The predictions from these base models were then combined using XGBoost as the meta-learner. XGBoost was chosen for its capacity to model complex interactions between the base models' predictions and its overall superior performance in classification tasks.

B. Evaluation Methodologies

In this study, we employed the Bootstrapping technique in order to evaluate the performance of the machine learning models used. This concept aims to train each model on different bootstrap samples and evaluate it using the data not included in those samples. We repeated the bootstrapping approach 50 times across all models before starting calculating models' performance. We then proceeded with the derivation of the confusion matrix which allows us to analyze the true positives, true negatives, false positives, and false negatives. The results of the confusion matrix using the bootstrapping approach for each machine learning algorithm are presented in Table 4.

To gain a deeper understanding of the classification outcomes, we also used the following performance metrics: Accuracy, Precision, Recall, F1 Score, and ROC-AUC. These metrics were calculated for each iteration of bootstrapping (repeated 50times), and the results were obtained by calculating the average of the values across all bootstrap iterations. The results are presented in Table 5, where we provide a comprehensive summary of the model's effectiveness, showing how the performance metrics vary across the different iterations and their mean values.

Table 4. Confusion matrix

Algorithm	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)	Total
Decision Tree	121	131	19	50	321
SVM	137	138	25	21	321
Logistic Regression	128	126	18	49	321
Random Forest	141	134	18	28	321
Stacking (RF, KNN, NB+ XGboost)	148	142	15	16	321
Threshold-Based System	110	120	30	61	321

The analysis clearly shows that Stacking (RF, KNN, NB + XGBoost) shows the highest overall performance across all metrics.

Based on the results presented in the table, we highlight the importance of ROC-AUC as a key metric for evaluating model performance, particularly in the context of imbalanced datasets. The ROC-AUC scores indicate different levels of model effectiveness, with Stacking achieving the highest score. Furthermore, we examine the trade-offs between precision and recall for each model, which reveal the balance between reducing false positives and false negatives.

Table 5. Performance metrics for different models using bootstrapping

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Decision Tree	0.78	0.86	0.70	0.79	0.88
Support Vector Machine (SVM)	0.86	0.85	0.87	0.86	0.87
Logistic Regression	0.79	0.88	0.72	0.80	0.89
Random Forest	0.86	0.89	0.83	0.86	0.89
Stacking (RF, KNN, NB+ XGboost)	0.90	0.90	0.90	0.90	0.90
Threshold-Based	0.72	0.79	0.64	0.71	N/B

C. Feature Importance

For the feature importance step, we utilized four key techniques: permutation importance, RFE (Recursive Feature Elimination), SHAP (Shapley Additive Explanations) analysis and LIME (Local Interpretable Model-agnostic Explanations)

- Permutation importance consists in assessing the effect of randomly shuffling feature values, in other words, we repeated the process of rearranging values of feature X while keeping the values of all other features the same.
- RFE leads us to eliminate less important features based on their importance or contribution to the model's predictions.
- SHAP Values measure the impact of each feature on a model's predictions.
- LIME provides local interpretability by approximating the model around a specific prediction and showing which features contributed most to that individual outcome.

Table 6 below presents the results for all techniques used to assess feature importance: Permutation Importance, Recursive Feature Elimination (RFE), SHAP (Shapley Additive Explanations) Values and LIME.

Table 6. Feature importance

Feature	Permutation Importance (%)	Rank (RFE)	SHAP VALUE	LIME Score
Technical Skills	16	1	027	0.26
Peer Reviews	15	2	0.24	0.23
Results Quality	11	3	0.21	0.21
Collaboration and Teamwork	8	4	0.18	0.18
Project completion & time management	7	5	0.16	0.17
Attendance	5	6	0.15	0.14
High School Grades	5	7	0.12	0.13
Past Certifications	4	8	0.10	0.11
Previous Coursework	3	9	0.09	0.09
Stress Management	2	10	0.07	0.08
Age	1	11	0.05	0.06
Gender	0	12	0.03	0.03
Educational Level	0	13	0.02	0.02

Table Explanation*:*

- Feature: The candidate attributes being evaluated.
- Permutation Importance (%): The percentage drop in model accuracy when the feature values were permuted, indicating its importance.
- Rank (RFE): The rank assigned to each feature based on the Recursive Feature Elimination process.

The SHAP analysis reinforced the importance rankings obtained from permutation importance and RFE, it represents the average impact of each feature on the model's output. A higher SHAP value indicates that the feature has a stronger influence on the prediction of student success. For example, the "Technical Skills" feature, with a SHAP value of 0.27, significantly contributes to the model's prediction, meaning students with stronger technical skills are more likely to be predicted as successful. Similarly, "Peer Reviews" (0.24) and "Results Quality" (0.21) also play important roles, reinforcing that both interpersonal perception and output quality are strong indicators of potential success.

In addition, LIME complemented these results by offering local insights into how each feature influenced individual predictions.

D. Visualization of Importance

To enhance understanding, we also created a bar chart to visualize the importance of each feature based on the drop in performance. Fig. 4 illustrates this, showing the relative importance of each feature using permutation analysis.

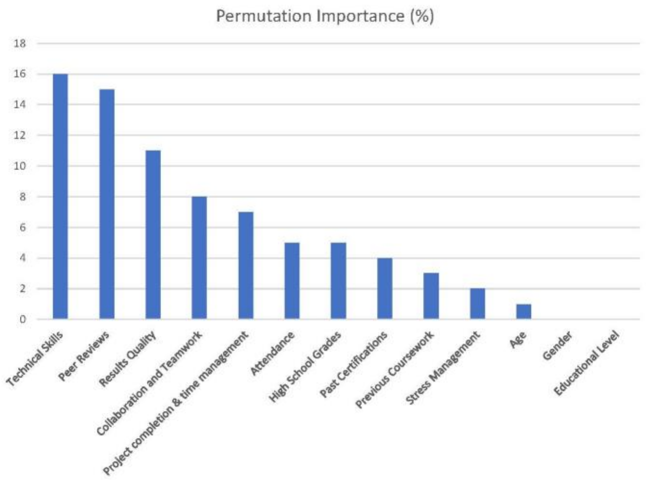

Fig. 4. Visualization of variable importance using permutation analysis.

E. Discussion

Beyond evaluating the predictive performance of various machine learning techniques, it is crucial to consider the practical implications of integrating ML into candidate selection processes. Traditional selection methods often rely on subjective assessments, manual evaluations which can introduce biases, inconsistencies, and inefficiencies. In contrast, our study seeks to explore predictive methods used to predict candidate success in the evaluation phase and aims to enhance fairness by relying on objective data rather than human assessment. We have carefully chosen attributes, ML techniques and various performance metrics for analysis, to obtain more relevant answers to the questions raised previously. In the following section, we delve into the key research questions, illustrating how our methodology addresses these challenges and contributes to a more reliable and scalable selection process:

Q1: What machine learning techniques are most effective in predicting candidate success in a self-directed learning school?

The results of our analysis showed that Stacking ensemble method outperformed all other models and considered to be one of the most effective machine learning techniques for predicting candidate success in peer learning schools. These results highlight the strengths of multiple base learners (Random Forest, KNN, and Naïve Bayes) and combining them with XGBoost as a meta-learner. It was also concluded that Random Forest is effective especially when it was applied to data from different sources such as student's records, behavioral traits and instructor feedback. However, decision tree algorithm showed weaker accuracy and did not appear well-suited for this context.

Q2: How do feature selection techniques affect the performance of machine learning models in predicting student outcomes?

As mentioned previously, we have applied four feature selection techniques (Permutation Importance RFE, SHAP Analysis and LIME score) which have a crucial role in enhancing model accuracy. Thus, it was observed that the utilization of the most relevant features leads to higher performance and relevant results. Equivalent terms, focusing on significant attributes such as Technical Skills and Peer Reviews, and reducing the dimensionality of the input data, allowed the models to generate better predictions.

Q3: Which candidate attributes are the most significant predictors of success during the evaluation process?

Our results demonstrate that Technical Skills Results Quality and Peer Reviews data are the most influential predictors of success in this context. Candidates who excelled in these tasks were more likely to pass through the selection phase. In addition, attributes like collaboration and teamwork were also classified as strong predictors, which highlights the importance of both interpersonal and technical capabilities in predicting success. However, some attributes such as age and gender, have little to no significant impact.

Q4: What role do soft skills, such as collaboration and communication, play in predicting candidate success?

Our study highlights that soft skills such as teamwork, communication, and collaboration play a key role in candidate success in addition to technical abilities. This demonstrates that candidates with strong interpersonal skills perform better, especially in group tasks, and confirms that self-directed learning requires more than just technical knowledge. Working well with others helped candidates share knowledge, adapt, and solve problems more effectively. These findings suggest that assessing soft skills in the selection process could improve predictions and provide a more complete evaluation of candidates.

V. CONCLUSION

Predicting candidate success is an important and challenging mission. The goal of this paper is to apply ML classification models, on a peer system school, to predict students' retention during the evaluation phase. The research highlights the importance of considering various factors in the prediction process and describes the sources and types of data but also the modelling engine stage with the application of predictive algorithms. We include the application of various models such as Stacking and Random Forest and demonstrate that machine learning techniques can effectively identify patterns with strong correlation with student retention. Our study also demonstrates that applying machine learning in the context of selection phase effectively addresses the challenges associated with candidate assessment and promotes fairness to the process.

While the current study has presented the potential of using predictive models in the educational system to support peer to peer institution in the selection process, there are several areas for future research. Our next direction is to focus on refining the selected models further and to explore other factors that may have a large impact on candidate success or rejection. Hence, future work may explore the influence of non-cognitive factors such as motivation and mindset to offer a better understanding of student performance

APPENDIX

Table A1. Machine learning studies: Key information

N°	Publication source	Purpose	Dataset	ML Techniques	Results
[2]	Journal	enhancing predictive accuracy of student	245 Students	Random Forest, AdaBoost, and	The XGBoost model outperformed

		performance using a new PFA approach that leverages Ensemble Learning methods		XGBoost models.	the original PFA and other algorithms in predictive accuracy
[3]	International Conference	finding out student's current status and predict his/her future results.	1179 Students	K-Nearest Neighbors Decision Tree Classifier SVC Random Forest Gradient Boosting Linear Discriminant	The best result and accuracy with K-Nearest Neighbors, Decision Tree Classifier model with an accuracy of 89.74% & 94.44%.
[4]	Journal	Reviewing different modern techniques widely applied for predicting students' performance, t	70 Papers	ANN SVM CF DR NB	ANN and SVM were more the most applied, followed by CF, DT, and NB.
[5]	Journal	Exploring machine learning techniques to predict students' final GPA based on personal characteristics, entry scores, gap year, and first- and second-year academic performance.	525 Students	Naïve Bayes, SMO, MLP, J48, Random Forest, Random Tree, PART, OneR	Naive Bayes showed lower accuracy, while ANN outperformed J48 in accuracy.
[6]	Journal	apply MLAs for prediction of student results from five different classes	309 Students	Decision Trees (DT), Bayesian Networks (BN), Artificial Neural Networks (ANN), Support Vector Machines (SVM)	Analysis revealed that decision trees, specifically C4.5, are the most suitable algorithm for generating production rules
[7]	International Conference	predicting SAT Math scores for high school students.	403 Students	linear regression, decision tree, and Naive Bayes	The Naive Bayes techniques showed the highest accuracy
[8]	Journal	This research aims to develop a machine learning-based system to evaluate high school students' performance and identify key factors influencing it.	459 Students	Random forest (RF), support vector machines (SVM), logistic regression (LR) and artificial neural network (ANN) techniques	ANN model had the best performance, The LR model was able to provide good performance in some iterations.
[9]	International Conference	Creating a model for predicting student performance using various factors	1000 STUDENTS	linear regression, decision trees, naïve Bayes classification, K nearest neighbors (KNN)	The results highlight GPA as the most significant factor influencing performance, while absence rate, risk score, suspensions, and mobility have minimal impact
[10]	International Conference	Using various machine learning models to detect the more suitable results for predicting student success.	580 Students	Random Forest Decision Tree Neural Network Naive Bayes Logistic Regression Linear Regression AdaBoost K Near Neighbor Quadratic Discriminant Analysis Multi-Layer Perceptron Support Vector Machine	Random Forest 99.50% Decision Tree 99.40% Neural Network 93.98% Naive Bayes 91.93% Logistic Regression 91.87% Linear Regression 91.70% AdaBoost 95.47% K Near Neighbor 98.68% Quadratic Discriminant Analysis 91.40% Multi-Layer Perceptron 93.92% Support Vector Machine
[11]	Journal	This study aims to automate the observation and prediction of students' marks and grades, focusing on achieving higher classification accuracy and lower root mean square error.	9000 students	Decision tree KNN GA	Decision-Tree (DT) 94.39 K-NN 85.74 GA+ Decision-Tree 96.64 GA+K-NN 89.92
[12]	Journal	Proposing a novel method for predicting students' future performance in degree programs given their current and past performance	1196 Students	A developed model-based course clustering method A developed ensemble-based progressive prediction method	the proposed methods achieve superior performance to benchmark approaches.
[13]	International Conference	This paper provides a thorough analysis of machine learning techniques aimed at enhancing predictive accuracy for forecasting final student grades in first-semester courses.	1282 Student	Decision Tree (J48), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbor (kNN), Logistic Regression (LR) and Random Forest (RF)	VM ensemble has produced greater accuracy when predicting students' final grades
[14]	Journal	This study proposes to segment students based on their initial evidence of failure or high performance and the performance levels predicted by the model.	2459 Students	Random Forest decision trees, support vector machines, naive bayes, bagged trees and boosted trees	random forests are superior to the other classification techniques that were considered
[15]	Journal	This paper proposes an adaptive recommendation system to predict suitable educational pathways for students in their college preparatory year.	725 Students	SVM KNN RF RF KNN QDA LR	The QDA algorithm achieved the highest F-measure of 0.91 for the Urban department, while RF scored 0.78 for Mechanical. KNN was selected for Architectural (0.89) and Electrical (0.77), LR for Mining and Petroleum (0.91), SVM for Computer (0.73), and RF for Civil (0.79), yielding an average F-measure of 82.57%
[16]	Journal	This paper develops and applies innovative machine learning and statistical methods to analyze the determinants of students' PISA 2015 test scores across nine countries: Australia, Canada, France, Germany, Italy, Japan, Spain, the UK, and the USA	Australia 14,530 Canada 20,058 France 6108 Germany 6504 Italy 11,583 Japan 6647 Spain 6736 UK 14,157 USA 5712		Tree-based methods enhance linear regression models of educational performance
[17]	Journal	Using the random forests in machine learning to	165 715 students	Random Forest	The predictive model predicts

		predict students at risk of dropping out			students' dropouts with excellent accuracy of 0.95
[18]	Journal	The study proposes a hierarchical classification of existing literature based on design choices in the SDP and introduces formal notation for consistently describing alternative dropout models.		Naive Bayes Linear Regression Rule-based DTrees Survival Analysis Neural Networks Ensemble Adaboost Random Forests	
[19]	International Conference	two datasets were used for predicting and classifying student performance with three machine learning algorithms		Backpropagation (BP), Support Vector Regression (SVR) and Long-Short Term Memory (LSTM)	The results for BP, SVM, and GBC are 87.78%, 83.20%, and 82.44%, respectively
[22]	Conference	Predicting student performance in a bachelor context	499 Students	Decision Tree, Naïve Bayes, Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Sequential Minimal Optimization and Neural Network	logistic regression classifier is the most accurate in predicting the exact final grades of students (68.7% for passed and 88.8% for failed).
[23]	Journal	Looking at the latest ML algorithms and variables used to predict student academic performance		Decision tree (DT), artificial neural networks (ANNs), support vector machine (SVM), K-nearest neighbor (KNN), linear regression (LinR), and Naive Bayes (NB)	ANN outperformed other models and had higher accuracy levels
[24]	Journal	Identifying the students who have poor academic performance in the computer science subject offered by Al-Muthanna University, College of Humanities,	161 Students	Decision Tree; Naïve Bayes; ANN; Logistic Regression,	ANN model achieved the best performance equal to 0.807 and achieved the best classification accuracy equal to 77.04%

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

C.O. collected, analyzed the data and wrote the paper. A.R. is the principal supervisor while S.B. are the co-supervisors. The supervisory team provided guidance throughout the research process.; all authors have approved the final version.

References

[1] H. Turabieh, "Hybrid machine learning classifiers to predict student performance," in *Proc. 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS)*, 2019.

[2] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interactive Learning Environments*, vol. 31, no. 6, pp. 3360–3379, 2023.

[3] H. M. R. Hasan, A. K. M. S. A. Rabby, and M. T. Islam *et al.*, "Machine learning algorithm for student's performance prediction," in *Proc. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2019.

[4] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," *Applied Sciences*, vol. 10, no. 3, 2020.

[5] D. T. Ha, P. T. T. Loan, and C. N. Giap et al., "An empirical study for student academic performance prediction using machine learning techniques," *International Journal of Computer Science and Information Security*, vol. 18, no. 3, pp. 75–82, April 2020.

[6] D. S. A. Acharya, "Early prediction of students performance using machine learning techniques," *International Journal of Computer Applications*, vol. 107, no. 1, pp. 37–43, December 2014.

[7] J. L. Harvey and S. A. P. Kumar, "A practical model for educators to predict student performance in K-12 education using machine learning," in *Proc. 2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2019, pp. 3004–3011.

[8] M. Zafari, A. Sadeghi-Niaraki, and S. M. Choi *et al.*, "A practical model for the evaluation of high school student performance based on machine learning," *Applied Sciences*, vol. 11, no. 23, 11534, 2021.

[9] K. Kishor, R. Sharma, and M. Chhabra, "Student performance prediction using technology of machine learning," in *Proc. International Conference on Micro-Electronics and Telecommunication Engineering, Singapore: Springer Nature Singapore*, 2021, pp. 541–551.

[10] M. F. Masood, A. Khan, and F. Hussain *et al.*, "Towards the selection of best machine learning model for student performance analysis and prediction," in *Proc. 2019 6th International Conference on Soft Computing & Machine Intelligence*, 2019.

[11] S. Hussain and M. Q. Khan, "Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning," *Annals of Data Science*, vol. 10, no. 3, pp. 637–655, 2023.

[12] J. Xu, K. H. Moon, M. Van Der Schaar "A machine learning approach for tracking and predicting student performance in degree programs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 742–753, 2017.

[13] S. D. A. Bujang, A. Selamat, and R. Ibrahim *et al.*, "Multiclass prediction model for student grade prediction using machine learning," *IEEE Access*, vol. 9, pp. 95608–95621, 2021.

[14] V. L. Miguéis, A. Freitas, and P. J. V. Garcia et al., "Early segmentation of students according to their academic performance: A predictive modeling approach," *Decision Support Systems*, vol. 115, pp. 36–51, 2018.

[15] M. Ezz, and A. Elshenawy, "Adaptive recommendation system using machine learning algorithms for predicting student's best academic program," *Education and Information Technologies*, vol. 25, no. 4, pp. 2733–2746, 2020.

[16] C. Masci, G. Johnes, and T. Agasisti, "Student and school performance across countries: A machine learning approach," *European Journal of Operational Research*, vol. 269, no. 13, pp. 1072–1085, 2018.

[17] J. Y. Chung and S. Lee, "Dropout early warning systems for high school students using machine learning," *Children and Youth Services Review*, vol. 96, pp. 346–353, 2019.

[18] B. Prenkaj, P. Velardi, and G. Stilo *et al.*, "A survey of machine learning approaches for student dropout prediction in online courses," *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.

[19] B. Sekeroglu, K. Dimililer, and K. Tuncal, "Student performance prediction and classification using machine learning algorithms," in *Proc. the 2019 8th International Conference on Educational and Information Technology*, 2019, pp. 7–11.

[20] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci*., vol. 11, no. 9, p. 552, 2021.

[21] H. Pallathadka, A. Wenda, and E. Ramirez-Asís *et al.*, "Classification and prediction of student performance data using various machine learning algorithms," *Materials Today: Proceedings*, vol. 80, pp. 3782–3785, 2023.

[22] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student performance prediction model based on supervised machine learning algorithms," in *Proc. IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 928, no. 3, 2020, 032019.

[23] Y. A. Alsariera, Y. Baashar, and G. Alkawsi *et al.*, "Assessment and evaluation of different machine learning algorithms for predicting student performance," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, 4151487, 2022.

[24] H. Altabrawee, O. A. J. Ali, and S. Q. Ajmi, "Predicting students' performance using machine learning techniques," *Journal of*

University of Babylon, Pure and Applied Sciences, vol. 27, no. 1, pp. 194–205, 2019.

[25] C. Y. Liu and H. L. Chen, "Effects of peer learning on learning performance, motivation, and attitude," *International Journal of Education Economics and Development,* vol. 11, no. 4, pp. 420–443, 2020.

[26] L. M. Cupelli and G. C. Colalillo, "Implementing peer learning to enhance academic performance in first-year nursing students," *Teaching and Learning in Nursing*, vol. 20, no. 1, pp. e150–e158, 2025.