

# Ensemble Learning for Early Academic Performance Prediction: A Case Study in Engineering Leveling Courses

Iván Carrera<sup>1</sup>; Rossy Armendáriz<sup>2</sup>; Valentina Ramos<sup>3</sup>

<sup>1,2,3</sup>Escuela Politécnica Nacional, Ecuador,

[ivan.carrera@epn.edu.ec](mailto:ivan.carrera@epn.edu.ec), [rossy.armendariz@epn.edu.ec](mailto:rossy.armendariz@epn.edu.ec), [valentina.ramos@epn.edu.ec](mailto:valentina.ramos@epn.edu.ec)

**Abstract**— Student attrition in engineering leveling courses is a persistent challenge in Latin American higher education, often exacerbated by the lack of timely identification mechanisms. Traditional Early Warning Systems (EWS) frequently rely on mid-term grades, limiting their capacity for preventive intervention. This paper presents an Ensemble Learning model designed to predict final academic performance using exclusively pre-enrollment data, combining sociodemographic, previous academic records, and socioeconomic and psychometric variables. This study analyzed historical records of 1,533 students (7,210 instances) from a polytechnic university in Ecuador. Five machine learning algorithm families were evaluated through rigorous hyperparameter optimization involving 9,972 configurations. The proposed Stacking architecture, which integrates the top-performing XGBoost models via a Ridge Regression meta-learner, achieved a Coefficient of Determination ( $R^2$ ) of 0.749 and a Mean Absolute Error (MAE) of 3.783 on a 0–40 scale, demonstrating superior generalization compared to single models. Furthermore, SHAP (SHapley Additive exPlanations) analysis revealed that non-cognitive factors, such as study habits and academic self-confidence, are critical predictors alongside high school GPA and mathematics admission scores. To validate its practical utility, the model was deployed for the 2025-B cohort ( $N=900$ ), identifying 9.2% of incoming students as "High Risk" before the start of classes. These results demonstrate the technical feasibility and ethical viability of using advanced ensemble methods for proactive, evidence-based academic management.

**Keywords**— Educational Data Mining, Stacking Ensemble, Early Warning System, Student Performance Prediction, Explainable AI (XAI).

## I. INTRODUCTION

Student retention and academic success in engineering programs remain critical challenges for higher education institutions globally, particularly in Latin America. High failure rates in leveling and first-year courses not only represent a financial loss for institutions and governments but also a significant setback for students' professional aspirations [1]. In this context, Early Warning Systems (EWS) have emerged as essential tools to identify at-risk students and deploy timely interventions. In the context of polytechnic universities, "Leveling" or preparatory courses serve as an important filter for academic rigor; however, failure rates often remain alarmingly high. Traditional approaches to academic management tend to be reactive, identifying at-risk students only after the first partial evaluations or mid-term grades are available [2].

However, a significant limitation in current Educational Data Mining (EDM) literature is the reliance on intra-semester data. Most predictive models achieve high accuracy by incorporating mid-term grades or Learning Management System (LMS) logs collected weeks or months into the academic period. While effective, these "late stage" predictions often arrive too late for preventive strategies, transforming the intervention from proactive support into reactive remedial action [3]. There is a scarcity of robust models capable of predicting academic performance with high precision before the first day of classes, relying exclusively on pre-enrollment data.

Although socioeconomic variables are commonly used, recent reviews indicate they offer only moderate predictive power when used in isolation [4] [5]. To bridge this gap, integrating psychological constructs such as academic self-concept and motivation has proven effective in enhancing prediction accuracy [6] [7]. Furthermore, while traditional Machine Learning algorithms like Decision Trees or Logistic Regression are widely used in education, recent studies suggest they may struggle to capture the complex, non-linear relationships inherent in student demographic and behavioral data [8]. Emerging research indicates that Ensemble Learning techniques, specifically Stacking, consistently outperform single classifiers in EDM tasks by combining the strengths of multiple algorithms [9] [10] [11]. Despite this, the application of Stacking architecture in the specific context of university leveling courses remains under-explored.

Another critical barrier to the adoption of AI in education is the "black box" problem. Stakeholders—academic advisors and students—are often reluctant to trust algorithmic predictions without understanding the underlying reasons [3] [10]. Consequently, the integration of Explainable AI (XAI) techniques, such as Shapley Additive exPlanations (SHAP), has become a requisite for valid EWS, enabling the interpretation of global risk factors and local individual predictions [3] [9].

This paper presents a Stacking Ensemble Learning model designed to predict academic performance in engineering leveling courses using solely pre-enrollment data. The proposed system integrates five families of algorithms—Random Forest, XGBoost, ElasticNet, SVR, and KNN—optimized through a rigorous stacking architecture. To ensure practical utility, the model is augmented with SHAP analysis to provide interpretable insights into the socioeconomic and academic factors driving student risk.

The main contributions of this work are: (i) the development of a high-precision predictive model ( $R^2$  approx 0.75) that requires no academic history within the university; (ii) a comparative analysis demonstrating the superiority of Stacking over single models in this domain; and (iii) the deployment of an explainable framework that identifies key risk drivers, such as high school background and study habits, to support evidence-based academic management.

## II. RELATED WORK

The development of Early Warning Systems (EWS) in higher education has evolved from simple statistical analyses to complex machine learning pipelines. This section reviews recent advancements in three key areas: data scope, the integration of psychometric variables, and the shift towards ensemble learning architectures.

### A. Prediction Scope: Pre-enrollment vs. Intra-semester Data

A significant portion of current Educational Data Mining (EDM) research relies on data collected *during* the academic term. Studies demonstrate that including dynamic metrics such as mid-term grades, Learning Management System (LMS) interaction logs, and attendance records significantly boosts predictive performance, with recent ensemble models achieving accuracies exceeding 90% in these scenarios [9] [12]. However, models dependent on such longitudinal data suffer from a critical "late-action" limitation: they often identify at-risk students only when the semester is well underway and failure is imminent, transforming interventions from proactive support into reactive remedial action [3]. Conversely, research focused exclusively on *pre-enrollment* data faces the challenge of "temporal decay," where the strong predictive power of prior academic achievement (e.g., high school GPA) diminishes as students' progress through their university trajectory [1]. Furthermore, while socioeconomic status (SES) is a foundational predictor, recent reviews indicate that SES variables alone offer only moderate predictive power and are best utilized in combination with other factors to explain variance in student performance [5], [4]. This performance gap highlights the urgent need for robust modeling strategies—such as Stacking Ensembles and the integration of psychometric constructs—to achieve high-precision predictions before the first day of classes [6].

### B. The Role of Psychometric Variables

To compensate for the absence of intra-semester academic history in pre-enrollment scenarios, recent literature advocates for the integration of non-cognitive or psychometric variables. Studies have consistently demonstrated that constructs such as academic self-concept, intrinsic motivation, and self-efficacy are not merely secondary traits but strong, independent predictors of performance [6] [7]. For instance, Séllei *et al.* [6] provided empirical evidence that positive psychological factors possess incremental predictive power, capable of explaining

significant variance in student outcomes that traditional demographic and pre-enrollment achievement data fail to capture. Furthermore, these psychometric variables play a critical role in educational equity; research by Marini *et al.* [5] and Ma'ruf & Juhaidi [2] indicates that self-efficacy and resilience often mediate the impact of socioeconomic status, effectively acting as buffers that allow students from disadvantaged backgrounds to outperform statistical expectations based solely on their financial metrics. Consequently, integrating these indicators into Early Warning Systems transforms them from simple regressive models into holistic tools capable of identifying "invisible" at-risk students who might otherwise be overlooked by demographic profiling alone [3].

### C. Ensemble Learning and Explainability (XAI)

While single classifiers like Decision Trees and Logistic Regression remain common in educational research, they often struggle to model the high dimensionality and non-linearity typical of student data, particularly when integrating heterogeneous sources (e.g., demographics mixed with psychometrics). Recent comparative studies indicate that Ensemble Learning methods consistently outperform individual models. Specifically, stacking architectures have demonstrated superior efficiency compared to homogeneous ensembles like Bagging or Boosting. For instance, Saleem *et al.* [10] and Abiodun & Wreford [9] reported that stacking diverse base learners (e.g., combining KNN, Random Forest, and XGBoost) achieved higher accuracy and F1-scores than any constituent model alone, effectively mitigating individual algorithm biases. Furthermore, Öz *et al.* [11] validated this approach on large-scale educational datasets (PISA 2022), finding that Stacking minimized error rates (MAE/RMSE) more effectively than Gradient Boosting across diverse student populations. However, as model complexity increases, so does the "Black Box" opacity problem. To resolve the tension between high accuracy and low interpretability, best practices in EDM now mandate the integration of Explainable AI (XAI). Techniques such as SHapley Additive exPlanations (SHAP) have become the gold standard, allowing institutions to move beyond binary predictions to actionable insights by visualizing how specific factors—like low self-efficacy—drive risk at both the global and individual levels [3] [8].

## III. METHODOLOGY

To ensure a structured, reproducible, and industry-standard workflow, we adopted the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology [13]. This framework guided the study through an iterative lifecycle comprising six distinct phases, from Business Understanding to Model Deployment. While the process covered the entire pipeline, particular emphasis was placed on two critical stages. First, the Data Preparation phase was rigorously executed to harmonize the heterogeneous nature of the dataset, which integrates sociodemographic records, academic history, and

psychometric survey data into a coherent feature space. Second, the Modeling phase focused on the architectural design and implementation of a Stacking Ensemble, prioritizing the optimization of hyper-parameters to maximize predictive accuracy beyond the capabilities of traditional single-classifier approaches.

#### A. Data Description and Ethical Considerations

The dataset comprises historical academic records from the leveling courses of a polytechnic university in Ecuador, covering three academic periods (2024-A, 2024-B, and 2025-A). The final dataset contains 7,210 records corresponding to 1,533 unique students. The variables are categorized into four dimensions:

1. Sociodemographic: Age, gender, marital status, and residence.
2. Previous Academic Background: High school grade point average (GPA) and university admission scores (mathematical, verbal, and spatial components).
3. Socioeconomic: Employment status, income level, and economic dependence.
4. Self-Perception: Psychometric variables measured via Likert scales (1-5), including academic self-concept, study habits, and motivation.

**Ethical Statement:** This study relied on the secondary analysis of pre-existing institutional data collected for academic management purposes. To ensure strict compliance with the university's data protection policies, a rigorous anonymization protocol was implemented prior to the modeling phase. All Personally Identifiable Information (PII), such as names and national ID numbers, was removed and replaced with random alphanumeric codes (e.g., est0001). As the research involved no direct interaction with subjects and used only de-identified historical records, it presented no risk to participants.

#### B. Data Preprocessing and Differentiated Pipelines

Data preparation was critical due to the heterogeneous nature of the variables. The feature engineering process expanded the original dataset to a final vector of 173 features, primarily due to the One-Hot Encoding of high-cardinality categorical variables such as "Degree Program" (26 categories) and "Course" (7 categories). Ordinal variables (e.g., parental education level) were mapped to numerical sequences.

A distinct contribution of this work is the implementation of differentiated preprocessing pipelines tailored to the mathematical requirements of each algorithm family:

- **Tree-based models (Random Forest, XGBoost):** Required minimal preprocessing. For XGBoost, the native handling of missing values (sparsity-aware split finding) was utilized, and no scaling was applied to preserve the original feature distribution.

- **Distance and Gradient-based models (KNN, SVR, ElasticNet):** Required rigorous standardization. Missing values were imputed using the median, and all numerical features were transformed using StandardScaler to prevent features with larger magnitudes from dominating the objective function.

#### C. Modeling Strategy

To ensure the statistical robustness and generalizability of the predictive models, the dataset was partitioned into a training set (80%) and a hold-out testing set (20%). This approach preserves the underlying proportion of academic risk categories across both subsets, ensuring that the test set remains truly representative of the heterogeneous student population and preventing biases related to class imbalance.

To capture the complex, multi-dimensional relationships within the educational data, five distinct families of regression algorithms were selected, representing a diverse range of inductive biases:

1. Ensemble Methods: Random Forest (RF) was employed for its capacity to reduce variance through bagging, while Extreme Gradient Boosting (XGBoost) was selected for its state-of-the-art performance in minimizing bias through sequential learning.
2. Regularized Linear Models: ElasticNet was utilized to handle potential multicollinearity among features by combining L1 (Lasso) and L2 (Ridge) penalties.
3. Kernel-based Methods: Support Vector Regression (SVR) was included to map non-linear boundaries in high-dimensional spaces.
4. Instance-based Learning: K-Nearest Neighbors (KNN) was evaluated to capture local data structures based on similarity metrics.

The hyperparameter optimization phase was conducted using an exhaustive GridSearchCV approach coupled with 5-Fold Cross-Validation to prevent overfitting to specific data partitions. The experimental design covered a search space of 9,972 unique configurations, resulting in a total computational load of approximately 49,860 model fits. This extensive search ensured that the base learners selected for the subsequent Stacking architecture were fully optimized to minimize the Mean Absolute Error (MAE).

#### D. Stacking Ensemble Architecture

The operational flow of the proposed Stacking strategy is formalized in Fig. 1. To minimize generalization error and enhance robustness against noise, a Stacking Ensemble strategy was implemented. Unlike simple voting mechanisms, Stacking involves a hierarchical learning process where a high-level model learns how to best combine the predictions of multiple base models. The proposed architecture is organized into two distinct learning stages:

1. **Level-0 (Base Learners):** This layer is responsible for the initial feature extraction and prediction. From the extensive optimization phase, the Top 25 best-performing models were selected based on minimizing the Mean Absolute Error (MAE) on the validation folds. A notable finding from this selection process was the overwhelming dominance of gradient boosting methods; specifically, XGBoost configurations occupied all top 25 slots, surpassing Random Forest and SVR. This homogeneity suggests that the underlying function mapping pre-enrollment variables to academic performance is best approximated by the sequential tree-building capability of gradient boosting, which effectively handles the non-linear interactions between sociodemographic and psychometric factors.
2. **Level-1 (Meta-Learner):** The independent predictions generated by the 25 base models were consolidated into a new dataset of "meta-features." To synthesize these outputs into a final grade prediction, a Ridge Regression (Linear Regression with L2 Regularization) was selected as the meta-learner. The choice of Ridge was strategic: since the base learners are all variations of XGBoost, their predictions are highly correlated. The L2 regularization penalty effectively handles this multicollinearity, assigning stable weights to the most reliable models without overfitting. Crucially, the meta-model was trained using out-of-fold predictions generated during cross-validation. This technique ensures that the inputs for the meta-learner are unbiased predictions on data instances that the base models did not see during their training, thereby preventing data leakage and ensuring rigorous generalization to the test set.

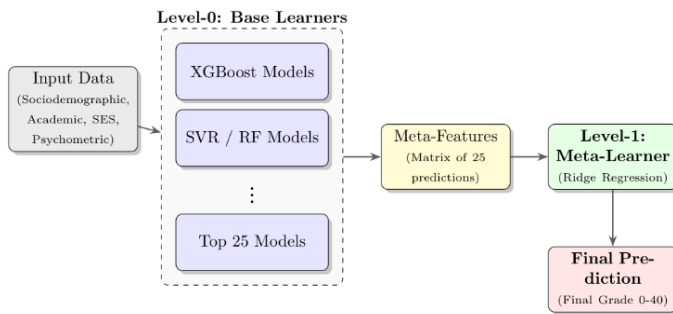


Fig. 1 Architecture of the Stacking Ensemble Model.

## IV. RESULTS

### A. Model Performance

The exhaustive hyperparameter optimization process, involving 9,972 configurations, revealed a clear hierarchy in

algorithm performance. As summarized in Table I, XGBoost demonstrated superior predictive capability, achieving the lowest Mean Absolute Error (MAE) of 4.065 in cross-validation, significantly outperforming traditional algorithms like Random Forest (MAE 4.987). This dominance is reflected in the composition of the Stacking ensemble, where all top 25 models selected for the Level-0 layer were XGBoost variations.

To establish a baseline representing traditional linear approaches, the ElasticNet model was evaluated, yielding an MAE of 5.850 and an  $R^2$  of 0.448. When compared to this linear baseline, the final Stacking Ensemble (MAE 3.783,  $R^2$  0.749) demonstrated a substantial improvement—explaining approximately 30% more variance and reducing the absolute error by over 2 points. This significant performance gap validates the necessity of utilizing complex, non-linear ensemble modeling for this educational dataset, as simpler linear assumptions fail to capture the intricate dynamics between psychometric and academic variables.

The final Stacking Meta-Model (using Ridge Regression) further improved predictive accuracy. On the unseen test set (20% of data), the system achieved an MAE of 3.783 points (on a 0-40 scale) and an  $R^2$  of 0.749, indicating that the model explains approximately 75% of the variance in final grades using solely pre-enrollment data.

TABLE I  
PERFORMANCE COMPARISON OF INDIVIDUAL BASE ALGORITHMS (5-FOLD CROSS-VALIDATION)

| Algorithm Family                | Best MAE | Best $R^2$ | Rank |
|---------------------------------|----------|------------|------|
| XGBoost                         | 4.065    | 0.710      | 1    |
| Support Vector Regression (SVR) | 4.212    | 0.683      | 2    |
| K-Nearest Neighbors (KNN)       | 4.647    | 0.624      | 3    |
| Random Forest                   | 4.987    | 0.607      | 4    |
| ElasticNet                      | 5.850    | 0.448      | 5    |

A critical finding is the model's generalization capability. Comparing performance metrics between training and testing sets reveals no signs of overfitting; in fact, the model exhibited slightly better performance on the test set (MAE 3.783) compared to the training set (MAE 4.011). This unusual but positive phenomenon can be attributed to the rigorous regularization applied (Ridge meta-learner and XGBoost shrinkage) and the effective stratified randomization of the data splits.

### B. Explainability and Feature Importance (SHAP)

To overcome the "black box" limitation, SHapley Additive exPlanations (SHAP) were employed to quantify the marginal contribution of each feature. The analysis of global feature importance, illustrated in Fig. 2, reveals a hybrid driver of academic success:

1. **Academic Antecedents as Primary Predictors:** As expected, historical academic performance remains the strongest predictor. High School GPA and the Admission Test: Math exhibit the highest mean absolute SHAP values, occupying the top ranks. This

confirms that prior knowledge foundations are critical for navigating the engineering curriculum.

2. **Psychometric Variables as Critical Modulators:** Crucially, the model identified non-cognitive variables as top-tier predictors, validating the inclusion of self-perception surveys. Good Study Habits and Problem Solving Efficacy emerged as highly influential features, surpassing demographic variables like Age. While the figure highlights the magnitude of their impact, further analysis indicates that students reporting high discipline and self-efficacy tend to have positive Shapley values, pushing their predicted grades upward and effectively compensating for moderate gaps in prior academic records.

This interpretability confirms that the risk of failure is not solely determined by cognitive ability but is significantly modulated by behavioral and motivational factors, which serve as actionable targets for university intervention programs.

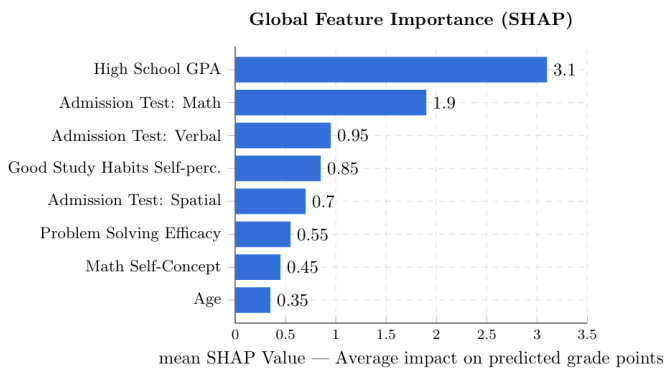


Fig. 2 Global feature importance based on the mean absolute SHAP value.

## V. CASE STUDY APPLICATION

To demonstrate the model's translational value from theoretical validation to institutional strategy, the proposed Stacking ensemble was applied to a new, unseen cohort of students (Period 2025-B). This application illustrates how predictive analytics can be operationalized to support academic management, enabling a shift from reactive remedial measures to proactive resource allocation before the academic term begins.

### A. Strategic Risk Segmentation

The trained model was utilized to process the pre-enrollment profiles of 6,123 registers of 900 incoming students. By automating the risk assessment process, the system serves as a high-precision filter for academic advisors. The model generated predictions with a cohort mean of 30.32/40 and a standard deviation of 4.35. To actionable management, the population was segmented into three distinct risk categories based on the institution's approval threshold (24/40):

1. High Risk (Predicted Grade < 24): A total of 561 registers (9.2%) were identified as the student is likely

to fail a specific course. This group represents the primary target for immediate preventive interventions. Identifying these students prior to the first day of classes allows the institution to enroll them in mandatory tutoring programs or psycho-pedagogical support tracks from "Day 0."

2. Medium Risk (24 Predicted Grade < 30): 1,854 registers (30.3%) were classified in an "uncertainty zone." While predicted to pass, their safety margin is narrow, making them vulnerable to academic setbacks. This segment would require periodic monitoring and access to elective support workshops.
3. Low Risk (Predicted Grade 30): The remaining 3,708 registers (60.6%) demonstrated a high probability of success. Accurately identifying this self-sufficient group allows the university to optimize limited resources, diverting efforts away from those who do not need them and concentrating retention strategies on the 39.5% of the population that is truly vulnerable.

This successful segmentation validates the potential of the proposed framework not just as a predictive exercise, but as a practical instrument for reducing dropout rates and enhancing educational quality.

## VI. DISCUSSION

This study aims to develop a preventive predictive model capable of identifying at-risk students before the start of the academic term. The obtained results offer several critical insights into the dynamics of student performance prediction.

### A. Overcoming the "Cold Start" vs. Accuracy Trade-off

A recurring theme in EDM literature is the high accuracy (>90%) achieved by models utilizing intra-semester data, such as partial grades or LMS logs. However, these models suffer from high latency, providing alerts when remedial actions may already be ineffective. In contrast, our approach accepts a moderate trade-off in accuracy to achieve "Day 0" latency. The obtained R2 of 0.75 indicates that three-quarters of the variance in final academic performance is determined by pre-existing conditions (academic background and psychometric profile). This finding challenges the assumption that university performance is solely an emergent property of the current semester, validating the feasibility of pre-emptive interventions rather than reactive ones.

### B. The Modulating Role of Psychometrics

The SHAP analysis provided crucial interpretability, revealing that while historical academic performance (High School GPA, Admission Tests) remains the dominant predictor, it is not deterministic. The high ranking of psychometric variables—specifically Study Habits and Problem-Solving Efficacy—confirms that non-cognitive skills play a compensatory role. Students with lower entry scores but high self-efficacy demonstrated positive Shapley values, suggesting

that resilience and discipline can mitigate gaps in prior knowledge. This aligns with recent findings by Séllei *et al.* [6] and supports the integration of soft-skills workshops as a core component of retention strategies, moving beyond purely academic tutoring.

### C. Robustness of Stacking in Noisy Environments

Educational data is inherently noisy and heterogeneous. The experimental results demonstrated that single highly optimized models, such as XGBoost (MAE 4.065), struggle to capture the full complexity of the data compared to the Stacking Ensemble (MAE 3.783). By combining the diverse inductive biases of tree-based, kernel-based, and linear models, the Stacking architecture effectively reduced the generalization error. This suggests that for sociodemographic and behavioral datasets, where no single mathematical assumption holds perfectly true for all students, heterogeneous ensembles offer a superior architectural standard for Early Warning Systems.

### D. Limitations

Despite the model's robustness, several limitations exist. First, approximately 25% of the variance remains unexplained, likely attributable to unobserved dynamic factors during the semester, such as financial shocks or health issues. Second, while the optimization process maximized empirical accuracy, the dominance of XGBoost variations among the base learners presents a theoretical limitation. This homogeneity reduces the algorithmic diversity typically desired in stacking ensembles, potentially causing the meta-learner to over-rely on gradient boosting biases and miss patterns that a more heterogeneous mix might capture.

Third, the reliance on self-reported psychometric surveys inherently introduces social desirability bias. Because self-perception can vary significantly across socioeconomic and cultural backgrounds, future deployments must incorporate fairness audits to ensure these metrics do not inadvertently penalize specific demographic groups. Finally, the dataset is confined to a single academic institution in Ecuador. Consequently, while the 'Day 0' predictive framework is highly transferable, specific feature weights—such as the dominance of the Mathematics Admission Score—exhibit limited generalizability and would require recalibration before application in non-STEM environments.

## VII. CONCLUSIONS AND FUTURE WORK

This study empirically demonstrates the technical viability of developing a high-precision Early Warning System (EWS) that solves the "cold start" problem inherent in higher education. Unlike traditional models that wait for mid-term failure signals, the proposed framework relies exclusively on pre-enrollment data to achieve an  $F_1$  of 0.749 and a Mean Absolute Error (MAE) of 3.783. This proves that student academic performance is not an emergent property solely of the university experience but can be anticipated with high confidence before the first day of classes. Furthermore, the successful

implementation of a rigorous anonymization protocol—replacing identifiers with random codes—confirms that granular predictive analytics can be conducted ethically, ensuring strict data privacy while empowering institutions with actionable intelligence to democratize educational support.

The comparative analysis highlights the limitations of single-family algorithms in capturing the heterogeneous nature of educational datasets. While XGBoost emerged as the most capable individual learner (MAE 4.065), it was still outperformed by the Stacking Ensemble approach (MAE 3.783). This performance gap suggests that academic success is a multi-dimensional phenomenon involving linear and non-linear interactions that no single inductive bias can fully capture. The success of the Ridge Meta-Learner in consolidating the top 25 base models underscores the potential of ensemble methods to reduce generalization error and mitigate overfitting, establishing Stacking as a robust standard for future Educational Data Mining tasks where data noise is prevalent.

Beyond accuracy, the integration of SHAP successfully addressed the "black box" dilemma, a critical barrier to AI adoption in education. The analysis revealed that while prior academic history is the dominant predictor, psychometric variables—specifically Study Habits and Self-Efficacy—act as significant modulators of performance. This finding has profound managerial implications: it validates the institutional investment in self-perception surveys and provides academic advisors with interpretable risk profiles. Interventions can now be tailored not just to *who* is at risk, but *why* (e.g., targeting motivational gaps vs. knowledge gaps), shifting the paradigm from generic remedial actions to personalized support strategies.

While the current model offers a robust preventive tool, approximately 25% of the variance in final grades remains unexplained, a limitation stemming from the exclusive use of static, pre-semester data. Future iterations of this work will aim to bridge this gap through two key avenues:

1. Hybrid Modeling: Integrating dynamic variables accumulated during the academic period—such as real-time attendance records, Virtual Learning Environment (VLE) logs, and early partial evaluation scores—to create a continuous learning pipeline that recalibrates risk probabilities week-by-week.
2. Prescriptive Analytics: Moving beyond prediction to prescription. Future research will focus on developing recommendation systems that suggest specific pedagogical interventions based on the student's SHAP profile, thereby closing the loop between data science and student success.

## ACKNOWLEDGMENT

This research was conducted as part of the Educational Innovation Project of the Vice-Rectorate of Teaching with the

strategic support of the ADA Data Science Laboratory. The authors gratefully acknowledge the laboratory for providing the computing infrastructure. Furthermore, we thank the Vice-Rectorate for facilitating access to the institutional data and academic records used in this research.

## REFERENCES

- [1] Ó. Espinoza, L. González, N. McGinn, L. Sandoval, B. Corradi y Y. Larrondo, «Academic Performance of Persistent University Students in Chile,» *Higher Education Quarterly*, 2025.
- [2] H. Ma'ruf y A. Juhaidi, «Multigroup analysis of Islamic Higher Education and General Higher Education in Indonesia: the influence of family socioeconomic background, possibility of success, and self-efficacy on student academic performance with resilience as a mediator,» *Cogent Education*, 2025.
- [3] E. Alyahyan y D. Düşteğör, «Predicting academic success in higher education: literature review and best practices,» *International Journal of Educational Technology in Higher Education*, vol. 17, pp. 1-21, 2020.
- [4] A. Vázquez, B. Alba, D. Lois, C. García-Ael y A. Bustillos, «The Role of Subjective Socioeconomic Status in Predicting Academic Performance: Exploring the Mediating Influence of Sense of Belonging among Students in a Distance Learning University,» *The Spanish Journal of Psychology*, vol. 27, 2024.
- [5] M. Marini, C. Parisse, S. Pagliaro, I. Giovannelli, D. Pietroni y S. Livi, «The role of social class in academic university contexts: exploring students' academic self-concept, educational expectations, and achievement goals,» *Frontiers in Psychology*, vol. 16, 2025.
- [6] B. Séllei, N. Stumphauer y R. Molontay, «Traits versus Grades—The Incremental Predictive Power of Positive Psychological Factors over Pre-Enrollment Achievement Measures on Academic Performance,» *Applied Sciences*, 2021.
- [7] J. Guo, L. Yang, J. Zhang y Y. Gan, «Academic self-concept, perceptions of the learning environment, engagement, and learning outcomes of university students: relationships and causal ordering,» *Higher Education*, vol. 83, 2021.
- [8] N. Butt, Z. Mahmood, K. Shakeel, S. Alfarhood, M. Safran y I. Ashraf, «Performance Prediction of Students in Higher Education Using Multi-Model Ensemble Approach,» *IEEE Access*, vol. 11, pp. 136091-136108, 2023.
- [9] O. Abiodun y A. Wreford, «Student's Performance Evaluation Using Ensemble Machine Learning Algorithms,» *Engineering and Technology Journal*, 2024.
- [10] F. Saleem, Z. Ullah, B. Fakieh y F. Kateb, «Intelligent Decision Support System for Predicting Student's E-Learning Performance Using Ensemble Machine Learning,» *Mathematics*, vol. 9, 2021.
- [11] E. Öz, O. Bulut, Z. Cellat y H. Yürekli, «Stacking: An ensemble learning approach to predict student performance in PISA 2022,» *Education and Information Technologies*, vol. 30, pp. 7753-7779, 2024.
- [12] B. Saini, «Optimizing Student Academic Performance Prediction Using Heterogeneous Ensemble Learning,» *European Journal of Artificial Intelligence and Machine Learning*, vol. 4, n° 4, 2025.
- [13] R. Wirth y J. Hipp, «CRISP-DM: Towards a standard process model for data mining,» Manchester, 2000.
- [14] H. Mengash, «Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems,» *IEEE Access*, vol. 8, pp. 55462-55470, 2020.
- [15] M. Jury, A. Smeding, N. Stephens, J. Nelson, C. Aelenei y C. Darnon, «The Experience of Low-SES Students in Higher Education: Psychological Barriers to Success and Interventions to Reduce Social-Class Inequality,» *Journal of Social Issues*, vol. 73, pp. 23-41, 2017.
- [16] M. Jihaoui, O. Abra y K. Mansouri, «Factors Affecting Student Academic Performance: A Combined Factor Analysis of Mixed Data and Multiple Linear Regression Analysis,» *IEEE Access*, vol. 13, pp. 15946-15964, 2025.
- [17] A. Browman y M. Destin, «The Effects of a Warm or Chilly Climate Toward Socioeconomic Diversity on Academic Motivation and Self-Concept,» *Personality and Social Psychology Bulletin*, vol. 42, pp. 172-187, 2016.