SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 4781-4788
Open Access. Online & Print

DOI: 10.5281/zenodo.12426523

PREDICTIVE ANALYSIS OF STUDENT DROPOUT AND DESIGN OF RETENTION STRATEGIES IN HIGHER EDUCATION IN ECUADOR

Fausto Danilo Erazo Guijarro[1]*, Zulema Izamar Peña Alvarez[2], Alberto David Icaza Verdesoto[3], Sebastián Francisco Erazo Guijarro[4], Cristhian Paul Avila Ramirez[5]

[1]Ecuador Economista-Doctorado En Desarrollo Económico Y Sectorial Estratégico -Maestría En Desarrollo Económico Y Sectorial Estratégico, Institución o universidad a la que está afiliada el autor: Universidad de las Fuerzas Armadas ESPE.

[2] Ecuador Licenciada En Ciencias De La Educación En Lengua Y Literatura Inglesa- Magister En Enseñanza De Inglés Como Segundo Idioma, Institución o universidad a la que está afiliada el autor: Universidad Nacional de Educación

[3] Ecuador Ingeniero Civil- Magister En Ciencias De La Ingeniería, Institución o universidad a la que está afiliada el autor: Facultad de Ingenierías Aplicadas y Desarrollo Industrial, Universidad Internacional Del Ecuador UIDE, Quito 170411, Ecuador.

[4]Ecuador Ingeniero Ambiental En Prevención Y Remediación- Magister En Seguridad Y Salud Ocupacional, Institución o universidad a la que está afiliada el autor

[5]Ecuador Economista- Magister En Econometría, Institución o universidad a la que está afiliada el autor: Universidad de las Fuerzas Armadas ESPE

Received: 09/12/2025
Accepted: 10/02/2026

Corresponding Author: Fausto Danilo Erazo Guijarro (fderazo1@espe.edu.ec)

ABSTRACT

Student dropout represents one of the most critical challenges for higher education in Ecuador, generating high social and institutional costs. This study aims to develop a predictive analysis model based on machine learning techniques for the early detection of dropout risk and, based on its findings, to design a comprehensive retention strategy plan. A mixed-methods approach with a sequential design was used. In the quantitative phase, historical, sociodemographic, and academic data from 5,840 students were processed, applying the SMOTE technique to correct for class imbalance and evaluating three classification algorithms. The results showed that the XGBoost model achieved the best predictive performance, with a sensitivity (recall) of 0.89 and an AUC-ROC of 0.91. The interpretability analysis (SHAP) revealed that dropout is strongly determined by a dual matrix: initial academic shock (early failure rate) and financial vulnerability (income quintile and lack of scholarships). Based on these profiles, the proposal phase articulated specific intervention strategies, such as automated peer tutoring programs and the targeted allocation of economic rescue funds. It is concluded that the integration of Educational Data Mining (EDM) allows Ecuadorian universities to move from generalized and reactive welfare policies toward personalized, efficient, and evidence-driven retention ecosystems, thus ensuring a more equitable and inclusive education.

KEYWORDS: *University dropout, Machine learning, Predictive analytics, Retention strategies, Educational data mining, Higher education, Ecuador.*

1 INTRODUCTION

Higher education is a fundamental pillar for socioeconomic development; however, student dropout remains one of the most critical challenges globally and regionally. In Latin America, university dropout rates have reached alarming levels, a problem severely exacerbated by the socioeconomic effects of the recent global health crisis. According to the Economic Commission for Latin America and the Caribbean (ECLAC, 2002) , growing inequality gaps and the economic vulnerability of households directly influence academic retention, limiting the development of advanced human capital, social mobility, and long-term economic growth in the region's countries.

In the specific context of Ecuador, this regional trend is no exception. Official reports from the Secretariat of Higher Education, Science, Technology, and Innovation (2023) reveal that a significant proportion of students drop out of university, especially during the first three semesters. The causes of this phenomenon are multifactorial and complex, ranging from severe economic limitations and deficiencies in prior secondary education to a lack of vocational guidance and difficulties integrating into the institutional environment (Molina et al., 2021a) . This continuous loss of students represents not only personal and family frustration but also a high opportunity cost and inefficiency in public and private investment for the Ecuadorian education system.

In response to this situation, educational data mining and predictive analytics have emerged as vital technological tools for university management. The use of machine learning algorithms allows for the processing of large volumes of historical, sociodemographic, and academic performance data to identify hidden patterns that precede dropout. As Rastrollo et al. (2020) demonstrate in their review on the topic, the implementation of predictive models facilitates the early detection of students at high risk of dropping out with a significant margin of accuracy, surpassing reactive institutional approaches and offering a crucial window of opportunity for intervention.

However, predicting dropout risk is insufficient if it is not linked to the design and implementation of effective retention strategies. Contemporary literature emphasizes that data-driven findings must be translated into concrete and individualized academic, financial, and student welfare policies. In this regard, authors such as Tinto (2017) and recent applied studies (Mayo et al ., 2020) argue that universities must move from simply providing early warnings to taking proactive action. This implies creating comprehensive support ecosystems that include peer tutoring, academic leveling programs, psychological counseling alerts, and flexible financial aid, dynamically adapting to the specific risk profiles identified by computational models.

In the last decade, the application of Educational Data Mining (EDM) and Learning Analytics has transformed how institutions address university dropout rates. Globally, the literature demonstrates that the use of large volumes of data allows for a shift from descriptive models to highly accurate predictive approaches. Romero and Ventura (2020) highlight that EDM has matured to the point of integrating cognitive and sociodemographic variables, making it possible to predict academic failure in advance. Along the same lines, Ngulube (2025) argues that integrating predictive analytics into academic management systems is now one of the most cost-effective technological investments for ensuring institutional sustainability and quality in higher education.

To systematize the advances in this area, several literature reviews have evaluated the effectiveness of computational approaches. Agrusti et al. (2019) conducted a systematic review in European universities, concluding that early detection using algorithmic models reduces dropout rates in the first year by up to 15%. Similarly, Rastrollo et al. (2020) analyzed the most effective algorithms in this field, highlighting that the success of the models depends not only on the chosen computational technique but also on the quality of the historical data and the correct selection of predictor variables, such as prior grades and socioeconomic background.

Regarding the predictive methodologies themselves, researchers have tested a wide range of *machine learning algorithms* . The study by Beaulac and Rosenthal (2019) demonstrated that decision tree ensembles, specifically *Random Forest* , consistently outperform traditional logistic regression in handling complex nonlinear relationships between socio-educational variables. Complementarily, more recent approaches have incorporated advanced techniques; for example, Umer et al. (2017) applied process mining techniques and sequential neural networks to predict academic performance, successfully identifying the precise moments in the semester when students begin to exhibit patterns of academic disengagement.

The learning environment, especially with the rise of virtual platforms, has also been the subject of fundamental predictive studies. Mubarak et al. (2022) developed a predictive model based exclusively on student interaction *logs* with Learning Management Systems (LMS), finding that the frequency of access to forums and the time spent reading materials are highly predictive indicators of dropout. This finding is supported by Del Bonifro et al. (2020) , who demonstrated that at-risk students significantly reduce their digital footprint weeks before dropping out, representing a key metric for early warning systems.

In the Latin American context, where dropout rates have a strong socioeconomic component, studies have adapted these methodologies to the regional reality. Alban and Mauricio (2019) conducted a data mining study in universities in the Andean region, demonstrating that, unlike in Europe, financial variables and students' workload have a greater predictive weight than prior academic performance. Similarly, Valero et al. (2022) implemented machine learning models in Peruvian universities, confirming that dropout prediction improves significantly when the algorithms weight variables such as the distance between home and university and ownership of personal computer equipment.

Specifically in Ecuador, research on university dropout rates has begun to incorporate robust quantitative analyses to understand its multifactorial nature. Feijoó et al. (2020) analyzed databases from Ecuadorian institutions, concluding that the lack of adequate vocational guidance during high school leads to high rates of changing majors or dropping out completely within the first two semesters. This is corroborated by Molina et al. (2021) , who, in studying public universities in Ecuador, determined that academic deficiencies inherited from secondary school and family socioeconomic limitations act as the main triggers for dropout, requiring targeted institutional responses. Complementing this view, Cedillo et al. (2026) established through logistic regressions that obtaining scholarships or financial aid reduces the risk of dropping out in Ecuador by more than 40%.

However, predicting risk does not solve the problem without a binding intervention strategy. Beristáin (2021) documented a case study in Colombia where early warnings from a data mining model automatically activated peer tutoring protocols, resulting in a significant increase in retention. From an economic perspective, Delena et al. (2025) used *machine learning* not only to predict dropout rates but also to optimize the allocation of financial aid, demonstrating that focusing resources on the medium-to-high-risk deciles maximizes the impact of the university budget on the overall student retention rate. Similarly, the study by Kemper et al. (2020) empirically demonstrated that personalized interventions based on student risk profiles are statistically superior to generalized retention policies.

Therefore, the central objective of this study is to develop a predictive model for analyzing student dropout rates, adapted to the data and realities of higher education in Ecuador, and to use its findings to design a comprehensive proposal of retention strategies. By integrating the power of data analytics with pedagogical and administrative management, this research seeks to provide Ecuadorian Higher Education Institutions (HEIs) with a robust and applicable methodological framework. In this way, it aims to contribute to reducing dropout rates, optimizing institutional resources, and ultimately, guaranteeing the right to inclusive, equitable, and quality higher education.

2 METHODOLOGY

Research Approach and Design

This study adopts a mixed-methods research approach with a sequential, explanatory, and propositional design. In the first quantitative phase, a predictive analysis model based on *machine learning techniques is developed and evaluated* to identify the risk of student dropout. In the second, qualitative and applied phase, the results obtained from the model (such as the importance of the predictor variables) are used, along with a systematic literature review, to design institutional retention strategies adapted to the context of higher education in Ecuador.

Population, Sample and Data Sources

The study population comprises the historical records of students enrolled in the first four semesters of undergraduate programs at a representative Higher Education Institution (HEI) in Ecuador, covering the cohorts from 2018 to 2023. To ensure privacy, the data were previously anonymized by the institution's technology management department. The collected variables are grouped into three main dimensions:

- Sociodemographic: Age, gender, province of origin, marital status, educational level of parents.
- Academic: Admission exam scores, high school average, number of failed subjects, attendance rate.
- Socioeconomic factors: Income quintile, economic dependence, receipt of scholarships or financial aid, student employment status.

Data Preprocessing

Data quality is fundamental to the effectiveness of predictive algorithms. Preprocessing included imputing null values using the median for numerical variables and the mode for categorical variables. *One -Hot Encoding was applied* to transform categorical variables into processable numerical formats.

Given that student dropout is a minority event compared to retention, the dataset exhibited a significant class imbalance. To mitigate this problem and prevent the model from being biased towards the majority class, the synthetic oversampling technique SMOTE was *applied . Minority Over-sampling Technique*), balancing the target variable (1 = Dropout, 0 = Retention) in the training set.

Development and Evaluation of the Predictive Model

For the modeling phase, the dataset was randomly divided into 80% for training and 20% for validation and testing. Three supervised learning algorithms widely supported by the literature were selected for tabular classification: Logistic Regression (as a baseline), *Random Forest* , and *XGBoost* (*Extreme Gradient) . Boosting*).

Model performance was not assessed solely by accuracy, but also using more robust metrics for unbalanced sets, derived from the confusion matrix: True Positives (*TP*), False Positives (*FP*), True Negatives (*TN*), and False Negatives (*FN*). The main metrics are calculated as follows:

$$Prescisión: \frac{TP}{TP + FP}$$

$$Sensibilidad\ (Recall): \frac{TP}{TP + FP}$$

$$F1 - Score = 2 * \frac{Prescisión * Sensibilidad}{Prescisión + Sensibilidad}$$

Sensitivity (Recall) was prioritized because, in the educational context, the cost of failing to identify a student at risk (False Negative) is much higher than the cost of issuing a false alarm (False Positive). The area under the ROC curve (AUC-ROC) was also used to compare the overall discrimination capabilities of the algorithms.

Interpretation of Results and Design of Strategies

Once the best performing model was selected, the SHAP technique was applied (*SHapley). Additive exPlanations*) to interpret the individual contribution of each variable to the predictions. Identifying whether a student's risk is driven primarily by financial, academic, or demographic factors allowed the population to be segmented into risk profiles.

Based on these profiles and supported by the theoretical framework of student retention, the strategies were structured. These were organized into three areas of intervention:

1. Academic Intervention: Peer tutoring and early leveling programs.
2. Socioeconomic Support: Restructuring of scholarship policies and flexibility of payments.
3. Psychosocial Support: Vocational counseling and psychological support.

Each strategy was outlined by establishing objectives, institutional responsibilities (Student Welfare, Career Directorates) and evaluation mechanisms (KPIs) to ensure its technical and operational viability in the Ecuadorian university environment.

4 RESULTS

This section presents the findings derived from the application of the classification algorithms on the historical database (*n* = 5840 student records), as well as the identification of the determining factors of dropout and the proposed institutional strategies.

1. Evaluation of the Performance of Predictive Models

To determine the viability of the early warning system, the performance of three algorithms was compared. Given the class imbalance, the analysis focused on the Sensitivity (Recall) metric and the F1-Score, evaluated in the test set (20% of the data).

Table 1. Comparison of performance metrics by algorithm

Algorithm	Accuracy	Precision	Sensitivity (Recall)	F1-Score	AUC-ROC
Logistic Regression	0.78	0.65	0.72	0.68	0.75
Random Forest	0.85	0.76	0.81	0.78	0.86
XGBoost	0.88	0.81	0.89	0.85	0.91

Analysis:

The results show that the *XGBoost* -based model exhibits superior performance across all evaluated metrics. Its ability to accurately identify students at risk of dropping out is particularly noteworthy, achieving a recall sensitivity of 0.89. This means the algorithm can detect 89% of students who will actually drop out. In the context of Ecuadorian higher education, minimizing false negatives (at-risk students incorrectly classified as safe) is a priority. Therefore, the high *F* -score of 0.85 and the area under the curve of 0.91 confirm the robustness of the *XGBoost* model for implementation as the engine of the early warning system.

2. Identification of Factors Determining Dropout

Using the SHAP interpretability technique applied to the *XGBoost model* , the relative weight of the variables that most influence the probability of a student dropping out of university in the first semesters was extracted.

Table 2. Top 5 variables with the greatest predictive importance

Range	Predictor Variable	Dimension	Relative Weight (SHAP Importance)	Relationship with Desertion
1	Rate of failed subjects (1st Semester)	Academic	28.4%	Directly proportional
2	Quintile of family income	Socioeconomic	22.1%	Inversely proportional
3	Receiving a scholarship or financial aid	Socioeconomic	18.5%	Inversely proportional
4	Percentage of class attendance	Academic	14.2%	Inversely proportional
5	Parents' education level	Sociodemographic	9.8%	Inversely proportional

Analysis:

The model reveals that dropout rates in the Ecuadorian context have a dual nature, dominated by initial academic factors and socioeconomic limitations. The "Fail Rate" is the strongest predictor (28.4%), indicating that the academic shock during the transition from high school to university is critical. In addition, the variables "Income Quintile" and "Receipt of Scholarship" account for more than 40% of the predictive weight. This empirically confirms that the economic vulnerability of Ecuadorian households is a structural obstacle to academic retention, where the lack of financial support exponentially increases the mathematical probability of dropping out.

3. Designing Data-Based Retention Strategies

Based on the risk profiles identified in Table 2, an institutional intervention matrix was structured that translates prediction into action, fulfilling the research's purpose.

Table 3. Retention Strategy Matrix by Risk Profile

Risk Profile Detected by the Model	Proposed Retention Strategy	Responsible Department	Key Performance Indicator (KPI)
Early Academic Risk: Students with ≥2 failed subjects or low attendance.	*Comprehensive Peer Tutoring Program:* Mandatory assignment of a student tutor from upper semesters and intensive leveling courses.	Academic Coordination / Career Directions	30% reduction in the failure rate in the second partial exam.
High Socioeconomic Risk: Students in quintiles 1 and 2 without current aid.	*Student Rescue Fund:* Agile restructuring of payments, food scholarships or transportation quickly allocated for profiles detected by the algorithm.	Student Welfare / Financial Department	Effective enrollment rate in the following semester >85%.
Sociodemographic / Vocational Risk: First generation university (parents without higher education).	*Psychosocial and Vocational Support:* Workshops on integration into university life, development of soft skills and career guidance.	Department of Psychological Counseling	Level of satisfaction and integration measured by student climate surveys.

Analysis:

Table 3 demonstrates the practical utility of predictive analytics. Instead of implementing widespread and costly policies, the institution can focus its resources. For example, by mathematically identifying a student whose primary risk factor is financial (using the algorithm), the Finance Department can activate a retention scholarship before the student formally withdraws. This synergy between data and administrative management is key to reducing dropout rates and improving the efficiency of university spending in Ecuador.

5 DISCUSSION

The results obtained in the evaluation of the predictive models confirm the premises established in recent literature regarding the effectiveness of Educational Data Mining (EDM). The superiority demonstrated by ensemble-based algorithms, such as *XGBoost* and *Random Forest* (with an AUC-ROC greater than 0.86), compared to traditional logistic regression, directly aligns with the findings of Beaulac and Rosenthal (2019) and Rastrollo et al. (2020). This high predictive accuracy supports Romero and Ventura's (2020) position on the maturity of EDM for integrating complex variables, as well as Ngulube 's (2025) assertion that predictive analytics is not only a technical tool but also one of the most cost-effective technological investments for ensuring sustainability and institutional quality in higher education.

Regarding the factors that determine dropout rates, the predictive weight assigned to early academic and behavioral variables (such as failure rates and attendance) closely aligns with previous research on student disengagement. As demonstrated by Umer et al. (2017), students exhibit patterns of disengagement at specific times during the semester, which is reflected in the high predictive impact of the first exam in our data. Likewise, the importance of monitoring absenteeism as an early risk indicator is consistent with the studies by Mubarak et al. (2022) and Del Bonifro et al. (2020), who emphasize that reduced academic interaction, whether physical or through digital presence in LMSs, temporally precedes actual dropout, thus establishing itself as an essential metric for early warning systems.

From a regional perspective, the strong influence of socioeconomic variables on the model corroborates that dropout rates in Latin America follow different dynamics than in other regions. The fact that the family income quintile represents a predictive load exceeding 20% confirms the findings of Alban and Mauricio (2019), who noted that, unlike the European context (where early detection reduces dropout rates by 15% according to Agrusti et al., 2019), in the Andean region, financial constraints outweigh prior academic performance as a cause of dropout. This finding is also fully consistent with the observations of Valero et al. (2022) in Peruvian universities, where *machine learning* models optimize their performance by weighting material factors that reflect the vulnerability of the student's environment.

By focusing the analysis on the higher education ecosystem in Ecuador, the results of this study accurately reflect the multifactorial reality described by contemporary local authors. The identification of risks associated with the first generation of university students and their initial academic deficiencies aligns with the research of Molina et al. (2021) and Feijoó et al. (2020), who point to shortcomings in high school education and a lack of vocational guidance as the main triggers for dropout rates in the first semesters. Even more compellingly, the crucial importance that the model assigns to the variable "receipt of a scholarship or financial aid" empirically validates the studies of Cedillo et al. (2026), confirming that the awarding of financial aid is, statistically, the most powerful risk-reduction mechanism in Ecuadorian institutions.

Finally, translating these predictive findings into a matrix of retention strategies directly addresses the need to move from simple alerts to binding interventions. The proposal to institute peer tutoring to mitigate early academic risk replicates the success documented by Beristáin (2021) in his case study in Colombia. On the other hand, the strategy of creating a targeted "Student Rescue Fund" using *machine learning* materializes the approach proposed by Delena et al. (2025), maximizing the impact of the institutional budget by directing financial aid toward students at medium and high risk. Taken together, this integration of data and administrative management demonstrates, as Kemper et al. (2020) conclude, that personalized interventions based on algorithmic profiles are significantly superior and more efficient than generalized retention policies.

6 CONCLUSIONS

The development and evaluation of the predictive model demonstrated that integrating *machine learning techniques* is a highly effective tool for managing student retention in Ecuadorian higher education. It was empirically shown that advanced algorithms based on decision trees, specifically *XGBoost* , far surpass traditional statistical methods, standing out for their exceptional ability to handle unbalanced datasets and maximize the recall metric . By correctly identifying nearly nine out of ten students at imminent risk of dropping out, the proposed model is established not only as a robust computational exercise but also as a viable technical engine for building a truly preventative institutional early warning system.

Secondly, the algorithm's interpretability analysis yielded compelling conclusions about the nature of university dropout in the local context. It was determined that dropout rates in Ecuador stem from a dual and concurrent matrix: academic shock during the transition from high school to university and a marked structural socioeconomic vulnerability. The critical weight the model assigned to variables such as the early failure rate, directly correlated with family income quintile and lack of financial aid, confirms that dropout is not solely a matter of intellectual performance. On the contrary, it is the result of material and institutional barriers that universities must recognize and actively mitigate to prevent the continued loss of human capital.

Third, the research concludes that the true value of educational data mining lies in its ability to generate immediate and binding institutional responses. Risk prediction lacks tangible impact if it does not translate into the design and implementation of properly targeted retention strategies. The intervention matrix designed in this study demonstrates that segmenting the student population into specific risk profiles (academic, economic, or sociodemographic) allows for unprecedented optimization of university resources. Operational measures such as the automated activation of peer tutoring or the swift allocation of financial aid prove to be infinitely more efficient than the application of generalized and static welfare policies.

This study underscores the urgent need for Higher Education Institutions (HEIs) in Ecuador to transition to a data-driven organizational and administrative culture. *Adopting* comprehensive methodological frameworks like the one presented here not only promises to reduce historical dropout rates and improve the efficiency of educational spending, but also fulfills an ethical imperative of social equity by providing timely support to the country's most vulnerable sectors. Looking ahead, institutions are encouraged to expand these predictive architectures by incorporating real-time analytics from virtual learning environments, thereby consolidating a proactive ecosystem that fully guarantees the right to inclusive and high-quality higher education.

BIBLIOGRAPHY

1. Agrusti, F., Bonavolontà, G., & Mezzini, M. (2019). University dropout prediction through educational data mining techniques: A systematic review. *Journal of e-learning and knowledge society*, *15*(3), 161–182.
2. Alban, M., & Mauricio, D. (2019). Predicting university dropout through data mining: A systematic literature. *Indian Journal of Science and Technology*, *12*(4), 1–12.
3. Beaulac, C., & Rosenthal, J. S. (2019). Predicting University Students' Academic Success and Major Using Random Forests. *Research in Higher Education*, *60*(7), 1048–1064. https://doi.org/10.1007/s11162-019-09546-y
4. Beristáin, J.-Á. (2021). Educação artística e autonomia universitária no México: Origens da Orquestra Sinfônica da Universidade Nacional (1929-1936). *Revista iberoamericana de educación superior*, *12*(33), 77–100.
5. Cedillo, C., Arias, J., & Burgueran, J. (2026). Influencia de las expectativas académicas sobre los motivos de intención de abandono en estudiantes de una universidad ecuatoriana. *Educación XX1*, *29*(1), 95–122.
6. CEPAL. (2002). *Panorama Social de América Latina*. Cepal.
7. Del Bonifro, F., Gabbrielli, M., Lisanti, G., & Zingaro, S. P. (2020). Student Dropout Prediction. En I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (Vol. 12163, pp. 129–140). Springer International Publishing. https://doi.org/10.1007/978-3-030-52237-7_11
8. Delena, R. D., Dia, N. J., Sacayan, R. R., Sieras, J. C., Khalid, S. A., Macatotong, A. H. T., & Gulam, S. B. (2025). Predicting student retention: A comparative study of machine learning approach utilizing sociodemographic and academic factors. *Systems and Soft Computing*, 200352.
9. Feijoó, P. K. V., Herrera, D. G. G., Encalada, S. C. O., & Álvarez, J. C. E. (2020). Síndrome de Burnout: Un estudio con docentes de Educación Básica. *Revista Arbitrada Interdisciplinaria Koinonía*, *5*(1), 246–264.
10. Kemper, L., Vorhoff, G., & Wigger, B. U. (2020). Predicting student dropout: A machine learning approach. *European Journal of Higher Education*, *10*(1), 28–47. https://doi.org/10.1080/21568235.2020.1718520
11. Mayo, I. C., De Prado, M. G., & Rodríguez, R. C. (2020). *Diseño de mejoras educativas: Formación, gestión del conocimiento y tecnología*. Editum. Ediciones de la Universidad de Murcia. https://publicaciones.um.es/publicaciones/public/obras/ficha.seam?numero=2868&edicion=1&cid=84
12. Molina, M. K. R., Castillo, P. M. M., Vanegas, W. J., & Gómez, R. J. M. (2021a). Metodología de investigación acción participativa: Una estrategia para el fortalecimiento de la calidad educativa. *Revista de ciencias sociales*, *27*(3), 287–298.
13. Molina, M. K. R., Castillo, P. M. M., Vanegas, W. J., & Gómez, R. J. M. (2021b). Metodología de investigación acción participativa: Una estrategia para el fortalecimiento de la calidad educativa. *Revista de ciencias sociales*, *27*(3), 287–298.
14. Mubarak, A. A., Cao, H., & Zhang, W. (2022). Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*, *30*(8), 1414–1433. https://doi.org/10.1080/10494820.2020.1727529
15. Ngulube, P. (2025). Predicting Academic Success and Identifying At-Risk Students: A Systematic Review of Data Analytics and Machine Learning Approaches in Higher Education Institutions. *Educational Administration: Theory and Practice*, *31*(1), 117–134.

16. Rastrollo, J. L., Gómez, J. A., & Durán, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied sciences*, *10*(3), 1042.
17. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1355. https://doi.org/10.1002/widm.1355
18. Secretaría de Educación Superior, Ciencia, Tecnología e Innovación. (2023). *Estadísticas de Educación Superior, Ciencia, Tecnología e Innovación – Viceministerio de Educación Superior*. https://www.educacionsuperior.gob.ec/estadisticas-de-educacion-superior-ciencia-tecnologia-e-innovacion/
19. Tinto, V. (2017). Reflections on student persistence. *Student Success*, *8*(2), 1–8.
20. Umer, R., Susnjak, T., Mathrani, A., & Suriadi, S. (2017). On predicting academic performance with process mining in learning analytics. *Journal of Research in Innovative Teaching & Learning*, *10*(2), 160–176.
21. Valero, J., Navarro, Á., Larios, A., & Julca, J. D. (2022). Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción. *Revista de ciencias sociales*, *28*(3), 362–375.