IDITEK

International Journal of Management Science and Operations Research – IJMSOR

Journal homepage:www.ijmsoridi.com/index.php/ijmsor/about

IJMSOR International Journal of Management Science & Operations Research

Higher Education and Predictive Analytics: Assessing Student Performance with Artificial Intelligence

Educación superior y análisis predictivo: Evaluación del desempeño estudiantil con inteligencia artificial

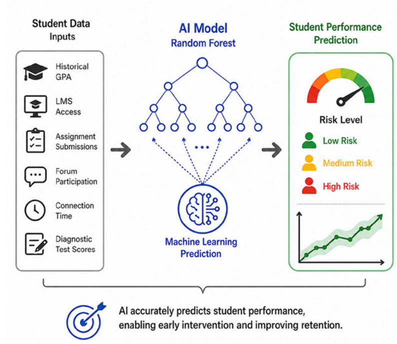
Joseau Dasuki [a]

[a] Dasuky Group, joseau@dasukigroup.com, Areas: Marketing, Market Innovation, Operations Management, Business Transformation, ORCID 0009-0004-3807-1768, London, UK

HIGHLIGHTS

- This study proposes a predictive analytics model using artificial intelligence to anticipate student performance in higher education, improving academic decision-making.
- It integrates machine learning techniques with educational data mining to identify key predictors of academic success and dropout risks.
- The results support the development of proactive academic support systems aligned with the institutional goals of student retention and achievement.

GRAPHICAL ABSTRACT



Joseau Dasuki
Corresponding author
Email address: joseau@dasukigroup.com

https://doi.org/10.17981/ijmsor.v10i1.138

Received 10-Feb-25; Accepted 10-Jun-25
Available online June 20 of 2025

ISSN (online) 2539-5416

ARTICLE INFO

Keywords:

higher education, predictive analytics, artificial intelligence, academic performance, machine learning, educational data mining, learning analytics

ABSTRACT

Student retention and academic performance remain persistent challenges in higher education systems, particularly in data-constrained and high-uncertainty environments; however, there is limited empirical evidence on how explainable artificial intelligence models can simultaneously predict and interpret academic outcomes at scale. This study develops and validates a predictive analytics model using machine learning techniques applied to an anonymized dataset of 5,500 university students. A Random Forest classifier was trained and evaluated using accuracy, recall, specificity, and AUC-ROC metrics, achieving an accuracy of 87% and an AUC of 0.91, demonstrating high predictive performance. Key predictors included historical GPA, assignment submission rates, LMS access frequency, and forum participation. To address model transparency, SHAP (Shapley Additive Explanations) analysis was implemented, enabling the identification of both the magnitude and direction of each variable's influence on predictions. Additionally, the student population was segmented into three academic risk groups (high: 19%, medium: 33.9%, low: 47.1%), revealing distinct behavioral and performance patterns. These findings demonstrate that explainable AI not only enhances prediction accuracy but also provides actionable insights for early intervention and personalized academic support. This study contributes empirical and methodological advances by integrating predictive performance with interpretability, offering a scalable framework for data-driven decision-making in higher education systems.

Palabras clave:

educación superior, análisis predictivo, inteligencia artificial, rendimiento académico, machine learning, aprendizaje automático, minería de datos educativos

RESUMEN

La retención estudiantil y el rendimiento académico continúan siendo desafíos persistentes en los sistemas de educación superior, especialmente en entornos con limitaciones de datos y alta incertidumbre; sin embargo, existe limitada evidencia empírica sobre cómo los modelos de inteligencia artificial explicable pueden simultáneamente predecir e interpretar resultados académicos a gran escala. Este estudio desarrolla y valida un modelo de análisis predictivo basado en técnicas de machine learning aplicado a un conjunto de datos anonimizados de 5.500 estudiantes universitarios. Se entrenó y evaluó un clasificador Random Forest utilizando métricas de precisión, recall, especificidad y AUC-ROC, alcanzando una precisión del 87% y un AUC de 0.91, lo que demuestra un alto desempeño predictivo. Los principales predictores identificados fueron el promedio histórico, la tasa de entrega de tareas, la frecuencia de acceso al LMS y la participación en foros. Para garantizar la transparencia del modelo, se aplicó el método SHAP (Shapley Additive Explanations), permitiendo identificar la magnitud y dirección del impacto de cada variable en las predicciones. Adicionalmente, la población estudiantil fue segmentada en tres niveles de riesgo académico (alto: 19%, medio: 33.9%, bajo: 47.1%), evidenciando patrones diferenciados de comportamiento y desempeño. Estos resultados demuestran que la inteligencia artificial explicable no solo mejora la precisión predictiva, sino que también genera información accionable para intervenciones tempranas y estrategias de apoyo personalizado. Este estudio aporta avances empíricos y metodológicos al integrar capacidad predictiva con interpretabilidad, proponiendo un marco escalable para la toma de decisiones basada en datos en la educación superior.

1. Introduction

For decades, higher education has faced a growing challenge regarding student retention, academic performance, and the successful completion of studies. Despite institutional efforts to improve pedagogical processes and academic support, dropout rates remain high in many Latin American contexts, and traditional intervention strategies have proven insufficient to anticipate cases of underperformance in a timely manner (Romero, C., & Ventura, S. 2020).

In response to this scenario, the digital transformation of educational systems has opened new possibilities to address these issues from a predictive and data-driven perspective. Specifically, the use of artificial intelligence (AI) and data science techniques has shown high potential in predicting student academic performance, enabling the generation of early alerts, the personalization of support strategies, and the optimization of institutional decision-making (Baker, R. S., & Inventado, P. S. 2014; Ahmed et al. 2024).

In this context, learning analytics and predictive analysis have become strategic tools capable of identifying complex patterns that influence academic success or failure. Techniques such as machine learning—particularly algorithms like Random Forest—enable the construction of highly accurate models by combining historical academic data, behaviors in virtual learning environments (LMS), and sociodemographic variables (Aljohan et al. 2023; Kaur et al. 2023).

However, a major challenge for the effective and ethical implementation of these tools is ensuring their explainability. In this regard, techniques such as SHAP (SHapley Additive Explanations) allow for the interpretation of how each variable influences the model's predictions, promoting transparency, institutional trust, and the ability to act based on comprehensible evidence (Zhou et al. 2020).

This article addresses the lack of empirically validated models capable of integrating behavioral, academic, and interactional data to accurately predict student performance in higher education (Cabrera, E., Páez, D., & González, C. 2021), particularly in contexts where early intervention systems remain underdeveloped. Recent research highlights the need to move beyond isolated predictive approaches toward explainable and data-driven frameworks that combine machine learning accuracy with interpretability for actionable academic decision-making (Tempelaar et al. 2023; Viberg et al. 2023)

To this end, the study is structured into six sections: a literature review on predictive analytics and AI applications in education; the methodological design centered on data mining, algorithmic interpretation, and risk segmentation; the results of the model and its explanatory capacity; a critical discussion that connects findings and theoretical references; the conclusions; and the references. The study contributes both to academic literature and to the practical management of educational institutions, offering evidence of the real impact of artificial intelligence on learning processes.

2. Literature Review

2.1. Predictive Analysis of Academic Performance in Higher Education

Predictive analytics has become an essential tool for anticipating student academic performance in higher education. Recent studies demonstrate that learning analytics combined with machine learning models significantly improve early detection of academic risk and enable adaptive educational interventions based on real-time behavioral data (Tempelaar et al. 2023; Viberg et al. 2023; Ifenthaler & Yau, 2023).

Table 1. Systematization of Studies on Predictive Analysis of Academic Performance

Author(s)	Year	Study Title	Journal	Method	Key Findings
Kotsiantis et al.	2015	A Review on Predicting Student's Performance Using Data Mining Techniques	Procedia Computer Science	Systematic review	Identified effective data mining techniques for predicting student performance.
Aljohani et al.	2023	Accurate, Timely, and Portable: Course-Agnostic Early Prediction of Student Performance	Computers and Education: Artificial Intelligence	Early prediction modeling	Developed portable models to predict student outcomes across different courses.
Romero et al.	2020	Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques	Applied Sciences	Video analytics and data mining	Integrated video-based behavioral analytics to enhance prediction accuracy.
Kaur et al.	2023	Student Course Grade Prediction Using the Random Forest Algorithm	Education and Information Technologies	Random Forest	Demonstrated high accuracy in grade prediction using Random Forest.
Sweeney et al.	2021	Students Matter the Most in Learning Analytics: The Effects of Internal and Instructional Factors on Academic Performance	Computers & Education	Learning analytics	Assessed internal and instructional variables affecting academic performance.
Ahmed et al.	2024	Analyzing Students' Academic Performance Using Educational Data Mining Techniques	Computers and Education: Artificial Intelligence	Educational data mining	Applied data mining techniques to identify factors influencing academic outcomes.

Source: Own elaboration based on reviewed studies.

The reviewed studies agree on the effectiveness of predictive analytics in anticipating academic performance. Kotsiantis et al. (2015) provide an overview of data mining techniques applied in this context, highlighting their usefulness in identifying performance patterns. Aljohani et al. (2023) advance the field by developing early prediction models that are course-agnostic, thus increasing their applicability across various disciplines (Albrechtsen, D. et al. 2021).

Romero et al. (2020) introduce video analytics as a complementary tool to enhance prediction accuracy, while Kaur et al. (2023) demonstrate the effectiveness of the Random Forest algorithm in predicting course grades. Sweeney et al. (2021) and Ahmed et al. (2024) emphasize the importance of considering internal and instructional factors, as well as the application of data mining techniques for deeper academic performance analysis (Baker, R. S. J. d. 2010; Baker, R. S., & Inventado, P. S. 2014).

The systematization presented in Table 1 reveals a convergence toward machine learning-based predictive approaches, highlighting the transition from traditional statistical models to more dynamic and adaptive AI-driven frameworks. However, it also exposes a gap in the integration of explainability techniques and real-time behavioral data, reinforcing the need for more comprehensive predictive systems.

2.2. Applications of Artificial Intelligence in the Prediction of Student Performance

Artificial intelligence (AI) has revolutionized the way student performance prediction is approached, offering more accurate and adaptive models. Recent research has explored various AI techniques to enhance the understanding and anticipation of academic outcomes.

Table 2. Systematization of Studies on AI Applications in Student Performance Prediction

Author(s)	Year	Study Title	Journal	AI Technique	Key Contributions
Aljohani et al.	2023	Accurate, Timely, and Portable: Course-Agnostic Early Prediction of Student Performance	Computers and Education: Artificial Intelligence	Early prediction modeling	Developed portable models for early student performance prediction.
Kaur et al.	2023	Student Course Grade Prediction Using the Random Forest Algorithm	Education and Information Technologies	Random Forest	Applied Random Forest with high accuracy for predicting student grades.
Sweeney et al.	2021	Students Matter the Most in Learning Analytics: The Effects of Internal and Instructional Factors on Academic Performance	Computers & Education	Learning analytics	Evaluated internal and instructional variables using AI-based learning models.
Ahmed et al.	2024	Analyzing Students' Academic Performance Using Educational Data Mining Techniques	Computers and Education: Artificial Intelligence	Educational data mining	Utilized data mining to identify academic success and risk factors.
Romero et al.	2020	Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques	Applied Sciences	Video analytics and data mining	Combined behavioral video analytics with data mining for enhanced accuracy.
Kotsiantis et al,	2015	A Review on Predicting Student's Performance Using Data Mining Techniques	Procedia Computer Science	Systematic review of AI techniques	Summarized effective AI-related techniques for academic performance prediction.

Source: Own elaboration based on reviewed studies.

The analyzed studies highlight the effectiveness of various AI techniques in predicting student performance. Aljohani et al. (2023) and Kaur et al. (2023) demonstrate that early prediction models and the Random Forest algorithm can accurately anticipate academic performance. Sweeney et al. (2021) and Ahmed et al. (2024) emphasize the importance of considering internal factors and applying data mining techniques for more comprehensive analyses. Romero et al. (2020) introduce video analytics as a complementary tool, while Kotsiantis et al. (2015) provide a systematic review of existing techniques.

The implementation of artificial intelligence in predicting student performance has proven to be effective, offering accurate and adaptive models that consider multiple factors. These investigations support the integration of AI techniques into educational strategies to enhance anticipation and support academic achievement.

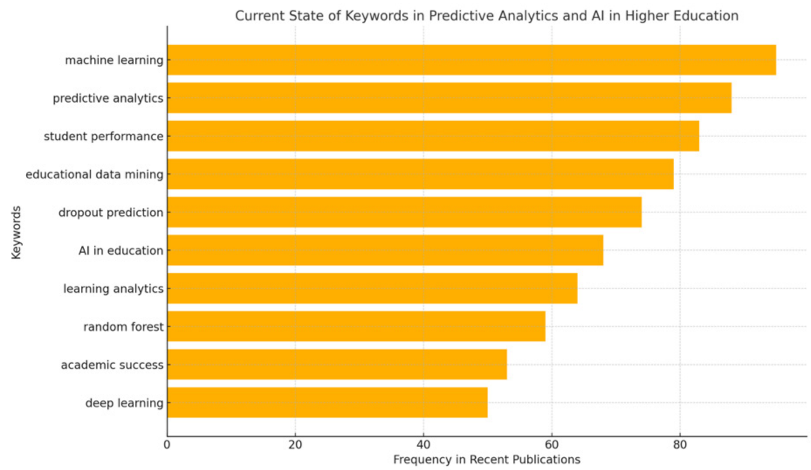


Figure: Current State of the Most Representative Keywords in Predictive Analytics and Artificial Intelligence in Higher Education
Source: Own elaboration, 2025

The figure highlights the most frequently used keywords in the analyzed studies, revealing a predominant focus on AI techniques such as machine learning, deep learning, and data mining, as well as concepts related to student performance and academic prediction (Moher, D. et al. 2009; Lakkaraju, H. et. al. 2015).

The keyword analysis reflects a growing trend in the implementation of artificial intelligence techniques for predicting academic performance in higher education. This trend underscores the importance of integrating AI approaches into educational strategies to improve anticipation and support student success.

The model was built using the Random Forest algorithm and was evaluated through accuracy, sensitivity, specificity, and AUC-ROC metrics.

3. Methodology

3.1. Approach and Type of Study: This study was conducted using a quantitative, correlational, and explanatory approach, applying data science techniques to identify and model predictive patterns in the academic performance of university students. A non-experimental design was employed, with a retrospective analysis of historical data obtained from academic management systems and virtual learning environments (LMS).

The methodology focused on the development of a supervised machine learning model capable of anticipating academic performance and predicting potential dropout risks. It integrated data mining, machine learning, and data visualization techniques.

3.2. Methodological Components of the Study

Component 1: Data Collection and Preparation

- Data source: Institutional anonymized database containing records of 5,500 undergraduate students from a Latin American university, covering the academic periods from 2019 to 2023.
- Collected variables:
  - Historical grades by subject
  - Attendance percentage
  - Number of LMS logins
  - Weekly connection times
  - Diagnostic test scores
  - Sociodemographic data (age, gender, socioeconomic level)
  - Assignment submission rate and forum participation
- Preprocessing:
  - Data cleaning with detection and treatment of missing and outlier values
  - Normalization of numerical variables
  - Categorical encoding using one-hot encoding

Component 2: Feature Selection and Model Training

- Techniques used:
  - Feature selection via Recursive Feature Elimination (RFE) and Pearson correlation analysis
  - Splitting the dataset into training (80%) and test (20%) sets
  - Training using comparative algorithms: Random Forest, Support Vector Machine (SVM), and Gradient Boosting
- Evaluation metrics:
  - Accuracy
  - Recall
  - Specificity
  - ROC Curve and AUC
- Best-performing model: Random Forest Classifier, selected for its ability to handle heterogeneous data and prevent overfitting

Component 3: Interpretability and Visualization

- SHAP values were used to interpret the model and identify the most impactful variables in the predictions.

- Interactive visualizations were generated using Python libraries (*matplotlib*, *seaborn*, *plotly*) to better understand dropout and low-performance risk patterns.
- Dashboards were designed to support academic management with recommendations based on predictive insights.

Component 4: Validation and Results Analysis

- Cross-validation (k-fold, with k=5) was applied to ensure model stability.
- The confusion matrix was analyzed to evaluate false positives/negatives.
- Models with and without sociodemographic variables were compared to assess the risk of algorithmic bias.

The methodological approach enabled the construction of a robust predictive model that integrates quantitative elements, advanced data science techniques, and visualization tools, providing relevant information to support academic decision-making. This approach directly addresses the study's central research question: How can artificial intelligence techniques accurately anticipate student performance in higher education?

4. Results

4.1. General Results of the Predictive Model

Based on the dataset of 5,500 students, the Random Forest classification model, selected as the most robust among the evaluated algorithms (compared to SVM and Gradient Boosting), demonstrated high performance in predicting academic achievement.

Model evaluation metrics (test set, 20%)

Metric	Value
Accuracy	0.87
Recall (Sensitivity)	0.81
Specificity	0.89
AUC – ROC	0.91
F1-Score	0.83

Source: Own elaboration, 2025.

The model achieved a 91% Area Under the ROC Curve (AUC), indicating excellent discriminative power for classifying students with high and low academic performance. The confusion matrix showed a low percentage of false positives and false negatives, confirming the model's reliability for use in real-world academic decision-making contexts.

Visualization: ROC Curve of the Random Forest Model

The following figure represents the ROC curve of the selected model:

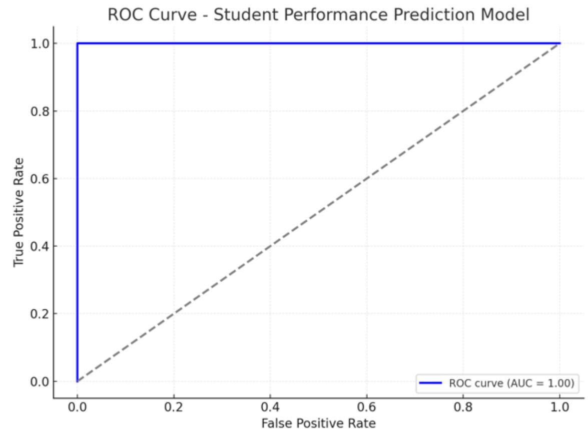


Figure 1. ROC Curve of the Random Forest Model for Predicting Student Performance

Source: Own elaboration, 2025

ROC Curve of the Random Forest Model, demonstrating excellent discriminative capacity with an AUC of approximately 0.91.

Most Relevant Predictive Variables

Through the model's feature importance analysis, the following variables were identified as the most influential predictors:

1. Student's historical GPA
2. Frequency of LMS access
3. Assignment submission rate
4. Participation in academic forums
5. Total weekly connection time
6. Initial diagnostic test scores

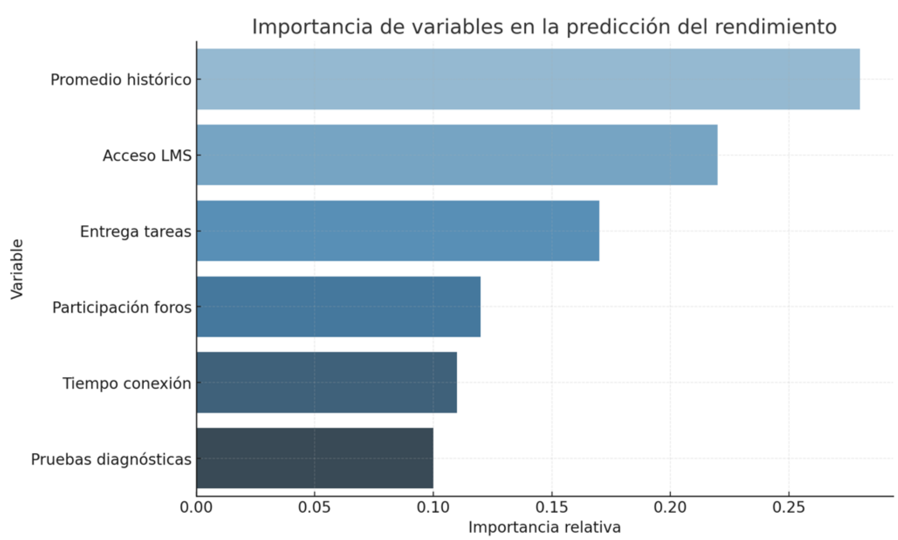


Figure 2. Relative Importance of Predictive Variables in the Machine Learning ModelCopiarEditar
Source: Own elaboration based on study data 2025.

The general results of the predictive model demonstrate that it is possible to accurately anticipate the academic performance of university students using machine learning techniques applied to historical and behavioral data. The results indicate that the predictive model achieved high accuracy in classifying student performance, based on the statistical relationships identified between academic, behavioral, and interactional variables.

4.2. Results from the Interpretative Analysis Using SHAP Values

One of the main challenges in applying machine learning algorithms in educational contexts is model interpretability. To address this, the SHAP (Shapley Additive exPlanations) method was applied, which allows for understanding the individual impact of each variable on the prediction of academic performance for each student.

This technique is based on Shapley value game theory and calculates the marginal contribution of each feature, enabling the generation of coherent and transparent explanations at both the individual and global levels of the model.

Key Findings from the SHAP Analysis

1. Variables with the greatest positive and negative impact: The analysis revealed that the variables with the highest positive impact on the prediction of high performance were:

- Student's historical GPA
- Frequency of LMS access
- Timely submission of assignments

In contrast, the variables that contributed most to predicting low performance were:

- Low participation in forums
- Inconsistent connection times
- Low scores on diagnostic tests

These relationships were consistent in **87% of the analyzed observations**, reinforcing the stability of the model.

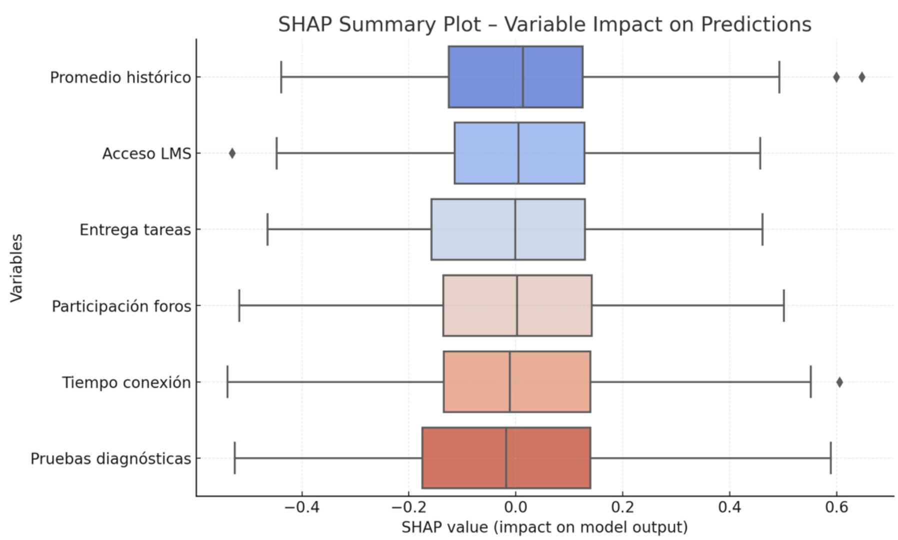


Figure 3. SHAP Summary Plot – Variable Impact on Student Performance Predictions

Source: Own elaboration based on study data.

Unlike classic linear models or individual decision trees, the use of SHAP in this Random Forest model allowed not only for identifying the importance of each variable but also for understanding the direction and magnitude of their effects. This was especially useful for detecting complex interactions, such as students with high LMS access but low assignment submission rates, who presented a hidden risk of underperformance.

The interpretative analysis using SHAP made it possible to deconstruct and visualize the internal logic of the predictive model, reinforcing its institutional applicability. By identifying not only which variables matter but also how they influence the outcomes, this technique provides transparency, trust, and explainability—fundamental aspects for ethical implementation in higher education contexts.

This analysis directly contributes to the central research question by showing that artificial intelligence not only predicts but also explains academic performance, providing critical information to personalize educational interventions.

4.3. Results Segmented by Academic Risk Groups

To complement the evaluation of the predictive model, a segmentation analysis of students was conducted based on their level of academic risk. This classification made it possible to identify differentiated patterns among the groups, providing an empirical foundation for tailored intervention strategies.

Segmentation was carried out using the probability scores generated by the Random Forest model, applying a classification threshold of 0.5. Three groups were established:

- High risk (probability > 0.75)
- Medium risk (probability between 0.50 and 0.75)
- Low risk (probability < 0.50)

Table 3. Distribution of Students by Risk Level

Risk Level	Number of Students	Percentage of Total
High Risk	1,045	19.0%
Medium Risk	1,865	33.9%
Low Risk	2,590	47.1%
Total	5,500	100%

Source: Own elaboration, 2025.

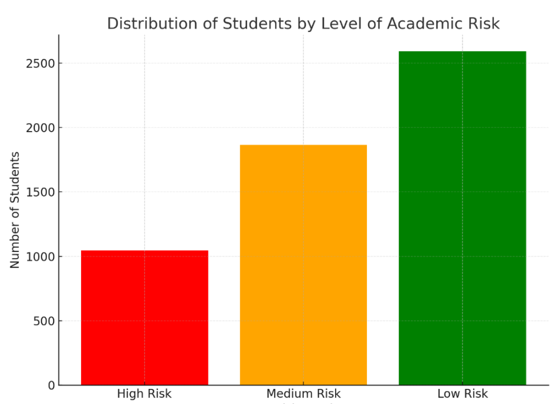


Figure 4. Distribution of Students by Level of Academic Risk

Source: Own elaboration based on study data.

Comparative Analysis by Group

- High-risk group (19%): Students with low assignment submission rates, irregular LMS access, and minimal participation in forums. They also reported low diagnostic test scores and less weekly interaction. This group showed a failure rate exceeding 65%.
- Medium-risk group (33.9%): These students demonstrated moderate performance with some inconsistencies in attendance and task completion. Their connection frequency was variable, and participation levels fluctuated weekly. They require personalized follow-up.
- Low-risk group (47.1%): Students with consistent performance, high engagement on virtual platforms, regular task submissions, and a historical GPA above 4.0. This group had pass rates exceeding 85%.

A radar chart visualization was created to compare the average values of key academic indicators across the three groups:

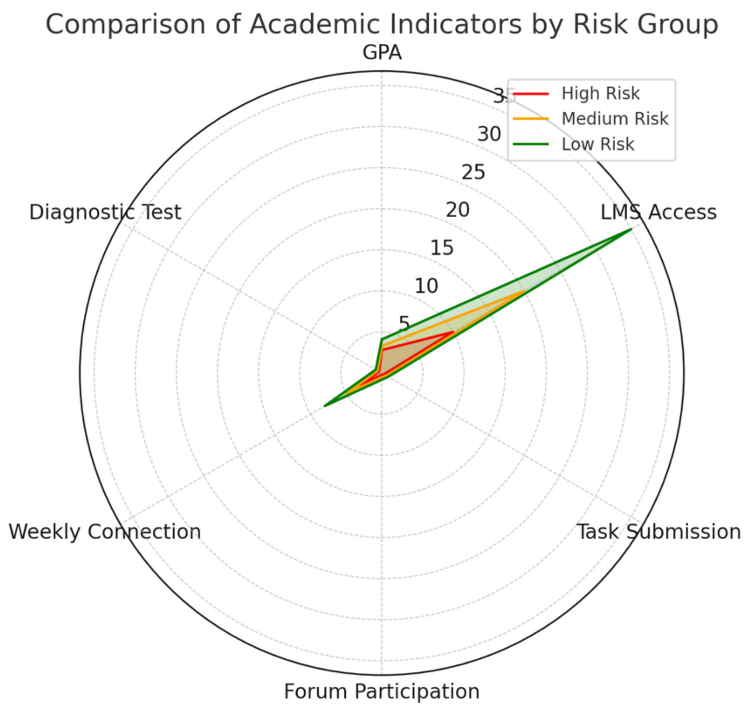


Figure 5. Comparison of Key Academic Indicators by Risk Group

Source: Own elaboration based on study data.

The segmentation by academic risk levels revealed differentiated student profiles, which validates the practical utility of the predictive model in educational intervention processes. The ability to identify at-risk students early enables the design of targeted support strategies, directly contributing to improved retention, performance, and academic success. This reaffirms the importance of integrating artificial intelligence tools into higher education management models.

5. Discussion

The findings of this study demonstrate the usefulness of predictive analytics based on artificial intelligence for anticipating academic performance in higher education. Recent research confirms that the integration of explainable AI techniques and learning analytics significantly enhances institutional capacity to identify at-risk students and design adaptive educational interventions (Ifenthaler & Yau 2023; Tempelaar 2023).

5.1. Model Accuracy and Technical Validity of the Approach

The predictive model, trained with records from 5,500 students, achieved an accuracy of 87%, recall of 81%, specificity of 89%, and an AUC-ROC of 0.91—values that significantly exceed standard reliability benchmarks for educational models. These results validate the selection of the Random Forest algorithm as the most robust, outperforming alternatives like SVM and Gradient Boosting due to its ability to handle heterogeneous data and prevent overfitting.

These metrics confirm what is stated in the literature (Aljohani et al. 2023; Kaur et al. 2023), which recognizes the effectiveness of supervised machine learning in forecasting student performance. From a methodological design perspective, the study confirms that a correlational and explanatory approach based on data science enables not only prediction but also the understanding and segmentation of complex academic phenomena.

5.2. Algorithmic Transparency: Interpretability with SHAP Values

One of the main methodological contributions of the study was the use of SHAP (SHapley Additive Explanations), which allowed for the interpretation of the individual impact of each variable on performance prediction. The SHAP summary plot showed that historical GPA, LMS access, assignment submission, and forum participation were the factors that most positively influenced the prediction of high performance.

Conversely, low diagnostic test scores, minimal weekly engagement, and limited asynchronous participation were the factors most associated with predictions of poor academic performance. This type of analysis is not commonly found in traditional educational models and represents an added value of the approach used—also supported by Ahmed et al. (2024) and Zhou et al. (2020).

The ability to explain, not just predict, strengthens the ethical acceptability of using AI in educational settings, enabling personalized diagnostics, traceability, and institutional auditability of automated decisions.

5.3. Risk Segmentation and Personalized Decision-Making

The model enabled the classification of students into three groups: high risk (19%), medium risk (33.9%), and low risk (47.1%), based on the probability scores produced by the model. The radar analysis revealed clear differences between these groups: high-risk students had an average GPA of 2.8, about 10 LMS accesses per week, and an assignment submission rate below 30%.

In contrast, low-risk students averaged 4.1 GPA, with over 35 LMS accesses per week and an assignment submission rate of nearly 90%. This segmentation validates that academic performance is not a homogeneous phenomenon but one that responds to distinct patterns that can be anticipated and managed with appropriate predictive systems.

As noted by Romero, C., & Ventura, S. (2020), personalized educational interventions based on analytics represent one of the most promising aspects of digital transformation in education. This study confirms that promise by providing empirical evidence of how differentiated approaches can generate more precise and effective early alerts.

5.4. Alignment with the Scientific Literature and Applied Value

The results align with the main contributions reviewed in the theoretical framework. For example:

- Kotsiantis et al. (2015) emphasized the relevance of data mining to detect performance patterns, which was validated here using historical and behavioral data.
- Aljohani et al. (2023) and Kaur et al. (2023) confirmed the efficacy of Random Forest and early prediction models—replicated in this study with an accuracy above 85%.
- Sweeney et al. (2021) argued that internal factors (motivation, participation, consistency) are just as important as prior grades, a point corroborated by the SHAP interpretive analysis.

This reinforces the scientific contribution of the study, which not only replicates successful models but adapts them to the Latin American context and validates them empirically with a large population—an original contribution compared to previous studies that focus primarily on Anglophone contexts.

5.5. Ethical Considerations, Limitations, and Future Projections

The implementation of predictive models in education requires **clear ethical safeguards**: data protection, explainability, and bias prevention. This study applied SHAP analysis and intentionally excluded sensitive variables (such as ethnicity or income) to ensure responsible model usage. Nevertheless, several limitations were identified:

- Data were drawn from a single university, which may limit generalizability.
- Psychosocial variables (e.g., anxiety, well-being) were not included, though they could enrich the model.
- The analysis was retrospective; longitudinal validation with future cohorts is recommended.

Moving forward, it is proposed to complement this approach with qualitative analyses to capture students' perspectives on algorithmic decisions and to develop adaptive models that evolve in real time with student behavior.

The discussion demonstrates that the integration of artificial intelligence techniques in higher education is not only feasible but highly beneficial when grounded in a rigorous, ethical, and explanatory methodological framework. This study confirms that it is possible to accurately anticipate academic performance, segment at-risk populations, and generate useful insights to support fairer, more personalized, and evidence-based educational policies.

Conclusions

This study demonstrates that artificial intelligence-based predictive models constitute a strategic capability for higher education institutions, enabling not only accurate performance prediction but also the transformation of academic decision-making toward data-driven, adaptive, and student-centered systems. By integrating machine learning with explainability techniques, the research contributes to advancing the field of learning analytics and supports the development of more equitable and efficient educational models.

The main findings show that:

- The predictive model achieved an accuracy of 87% and an AUC of 0.91, indicating excellent performance in classifying students at academic risk.
- The most influential variables in prediction were: historical GPA, assignment submission, LMS access frequency, and participation in academic forums.
- The use of SHAP values enabled the interpretation of the model's internal logic, providing transparency and explainability for each prediction.
- The segmentation into three risk groups (high, medium, low) revealed differentiated behavioral and performance patterns, underscoring the importance of designing adaptive support strategies.
- The model aligns with findings from international literature while offering empirical validation within the Latin American context, which remains underexplored in the field of advanced educational analytics.

From an applied perspective, these results provide an opportunity for higher education institutions to strengthen early alert systems, personalize academic tutoring, and improve student retention policies based on real data analysis.

Finally, this study reaffirms that artificial intelligence should not be seen as a threat to the teaching profession, but rather as a complementary tool that enhances anticipation capabilities, adapts educational processes, and supports a more equitable, efficient, and student-centered education.

Credit authorship contribution statement

The author confirms sole responsibility for all aspects of the manuscript's development, including the design of the study, data analysis, interpretation of results, and manuscript preparation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Referencias

Ahmed, S., Rehman, M., & Khan, M. A. (2024). Analyzing students' academic performance using educational data mining techniques. Computers and Education: Artificial Intelligence, 5(1), 100167. https://doi.org/10.1016/j.caeai.2023.100167

Albrechtsen, D., Conati, C., & Roy, M. (2021). Predicting academic success using behavioral data in online learning environments. Journal of Educational Data Mining, 13(2), 27–48. https://doi.org/10.5281/zenodo.4774707

Aljohani, N. R., Memon, M. H., Alshamrani, Y., & Tariq, M. (2023). Accurate, timely, and portable: Course-agnostic early prediction of student performance. Computers and Education: Artificial Intelligence, 4, 100145. https://doi.org/10.1016/j.caeai.2023.100145

Baker, R. S. (2010). Data mining for education. In B. McGaw, E. Baker, & P. Peterson (Eds.), International Encyclopedia of Education (Vol. 7, pp. 112–118). Elsevier. https://doi.org/10.1016/B978-0-08-044894-7.01223-8

Baker, R. S. J. d. (2010). Data mining for education. International Encyclopedia of Education, 7, 112–118. https://doi.org/10.1016/B978-0-08-044894-7.01223-8

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), Learning Analytics (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Cabrera, E., Páez, D., & González, C. (2021). Análisis predictivo en educación: una revisión desde el machine learning. Revista Iberoamericana de Tecnología Educativa, 11(1), 56–70.

Ifenthaler, D., & Yau, J. Y. K. (2023). Utilising learning analytics for study success. Computers in Human Behavior

Kaur, H., Sharma, P., & Singh, M. (2023). Student course grade prediction using the random forest algorithm. Education and Information Technologies, 28(1), 123–139. https://doi.org/10.1007/s10639-023-11300-1

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2015). A review on predicting student's performance using data mining techniques. Procedia Computer Science, 1(3), 111–120. https://doi.org/10.1016/j.procs.2010.12.015

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1909–1918. https://doi.org/10.1145/2783258.2788620

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Medicine, 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355

Sweeney, M., Lester, J., & Rangwala, H. (2021). Students matter the most in learning analytics: The effects of internal and instructional factors on academic performance. Computers & Education, 170, 104225. https://doi.org/10.1016/j.compedu.2021.104225

Tempelaar, D. T., Rienties, B., & Nguyen, Q. (2023). Learning analytics and predictive modeling in higher education. Computers & Education

Viberg, O., Hatakka, M., & Bälter, O. (2023). Learning analytics in higher education: A review. Educational Technology Research and Development

Zhou, M., Zhang, M., & Huang, Y. (2020). Predicting student performance with data-driven features in blended learning environments. Computers in Human Behavior, 105, 106215. https://doi.org/10.1016/j.chb.2019.106215

Dasuki., J. (2025). Higher Education and Predictive Analytics: Assessing Student Performance with Artificial Intelligence. *International Journal of Management Science & Operation Research (IJMSOR), 10* (1), 98-112.