

Learning Analytics for Curriculum Personalization in Hybrid Engineering Education Environments

Jorge Valle-Reconco¹; Yolly Molina²; Patricia Soriano³

¹Faculty of Medical Sciences, National Autonomous University of Honduras, Honduras, jorge_valle@unah.edu.hn

²Centro de Diagnóstico de Imágenes Biomédicas, Investigación y Rehabilitación, Universidad Nacional Autónoma de Honduras, Honduras, yolly.molina@unah.edu.hn

³Dirección de Medicina Forense, Ministerio Público, Honduras, paty.soriano@yahoo.com

Abstract– *The shift toward hybrid education models in engineering programs has created the need to develop intelligent systems capable of personalizing learning experiences. This study presents a methodological framework for implementing learning analytics aimed at the early identification of students at academic risk and the design of personalized interventions in engineering faculties. A predictive model was developed based on data from learning management systems (LMS), formative assessments, and student activity patterns. The methodology included Monte Carlo simulation with 1,000 iterations for model validation, applied to a simulated cohort of 847 students over four academic semesters. The predictive model achieved an area under the ROC curve of 0.803 (95% CI: 0.744–0.845), with a sensitivity of 70.9% and a specificity of 77.7%. The personalized interventions implemented demonstrated statistically significant improvements in the academic performance of the treatment group (GPA: 2.68 ± 0.59) compared to the control group (GPA: 2.38 ± 0.64), with a medium effect size (Cohen's $d = 0.49$, $p < 0.001$). The retention rate increased from 73.1% to 88.6% among students who received the intervention ($\chi^2 = 12.08$, $p < 0.001$). The results suggest that integrating learning analytics into hybrid education environments constitutes an effective strategy for curriculum personalization and for reducing the risk of dropout in engineering programs.*

Keywords– *Learning analytics, hybrid education, curriculum personalization, predictive models, student retention.*

Analítica de Aprendizaje para Personalización Curricular en Entornos de Educación Híbrida en Ingeniería

Jorge Valle-Reconco¹; Yolly Molina²; Patricia Soriano³

¹Faculty of Medical Sciences, National Autonomous University of Honduras, Honduras, jorge_valle@unah.edu.hn

²Centro de Diagnóstico de Imágenes Biomédicas, Investigación y Rehabilitación, Universidad Nacional Autónoma de Honduras, Honduras, yolly.molina@unah.edu.hn

³Dirección de Medicina Forense, Ministerio Público, Honduras, paty.soriano@yahoo.com

Resumen– La transformación hacia modelos de educación híbrida en programas de ingeniería ha generado la necesidad de desarrollar sistemas inteligentes capaces de personalizar las experiencias de aprendizaje. Este estudio presenta un marco metodológico para la implementación de analítica de aprendizaje orientada a la identificación temprana de estudiantes en riesgo académico y el diseño de intervenciones personalizadas en facultades de ingeniería. Se desarrolló un modelo predictivo basado en datos provenientes de plataformas de gestión de aprendizaje (LMS), evaluaciones formativas y patrones de actividad estudiantil. La metodología incluyó simulación Monte Carlo con 1,000 iteraciones para validación del modelo, aplicada a una cohorte simulada de 847 estudiantes durante cuatro semestres académicos. El modelo predictivo alcanzó un área bajo la curva ROC de 0.803 (IC 95%: 0.744-0.845), con una sensibilidad del 70.9% y especificidad del 77.7%. Las intervenciones personalizadas implementadas demostraron mejoras estadísticamente significativas en el rendimiento académico del grupo de tratamiento (GPA: 2.68 ± 0.59) comparado con el grupo control (GPA: 2.38 ± 0.64), con un tamaño del efecto medio (d de Cohen = 0.49, $p < 0.001$). La tasa de retención aumentó del 73.1% al 88.6% en estudiantes que recibieron intervención ($\chi^2 = 12.08$, $p < 0.001$). Los resultados sugieren que la integración de analítica de aprendizaje en entornos de educación híbrida constituye una estrategia efectiva para la personalización curricular y la reducción del riesgo de deserción en programas de ingeniería.

Palabras clave– analítica de aprendizaje, educación híbrida, personalización curricular, modelos predictivos, retención estudiantil.

I. INTRODUCCIÓN

La educación superior en ingeniería enfrenta desafíos como altas tasas de deserción y bajo rendimiento académico, especialmente en los primeros semestres [1]. La pandemia de COVID-19 aceleró la adopción de modalidades híbridas de enseñanza, combinando componentes presenciales y virtuales, transformando la interacción de los estudiantes con los contenidos y docentes. Esta transición ha generado datos educativos que, procesados, pueden ofrecer información valiosa para decisiones pedagógicas [2].

La analítica de aprendizaje (Learning Analytics) mide, recopila, analiza y reporta datos sobre estudiantes y sus contextos para comprender y optimizar el aprendizaje y sus entornos [3],[4]. En la educación en ingeniería, se pueden identificar patrones de comportamiento relacionados con el éxito o fracaso académico, lo que permite intervenciones

tempranas y personalizadas [5]. Estudios han demostrado que los sistemas de alerta temprana basados en analítica de aprendizaje pueden reducir las tasas de deserción si se implementan correctamente [6], [7].

Los sistemas de gestión de aprendizaje (LMS) como Moodle, Canvas o Blackboard generan registros de la actividad estudiantil, incluyendo acceso, tiempo de permanencia, recursos consultados, participación en foros y entregas de actividades [8]. Estos datos, junto con información demográfica y resultados de evaluaciones, son la base para desarrollar modelos predictivos de rendimiento académico [9]. La literatura indica que los modelos más efectivos integran múltiples fuentes de datos y utilizan técnicas de aprendizaje automático para clasificar estudiantes según su nivel de riesgo [10].

La implementación de sistemas de analítica de aprendizaje en universidades de América Latina enfrenta desafíos específicos. Las limitaciones de infraestructura tecnológica, la heterogeneidad de plataformas y las consideraciones éticas sobre la privacidad de datos estudiantiles requieren adaptaciones específicas de los marcos metodológicos de otros contextos [11]. Además, hay una brecha entre el desarrollo de modelos predictivos y su aplicación en intervenciones pedagógicas efectivas [12].

Este estudio busca desarrollar y validar un marco metodológico para implementar analítica de aprendizaje en programas de ingeniería híbridos. Se busca: (a) construir un modelo predictivo de riesgo académico con datos de LMS y evaluaciones formativas; (b) diseñar e implementar intervenciones personalizadas según perfiles de riesgo; y (c) evaluar el impacto de estas intervenciones en el rendimiento académico y la retención estudiantil. La hipótesis sostiene que los modelos analíticos predictivos permiten diseñar intervenciones personalizadas que mejoran el desempeño académico y reducen el riesgo de deserción en estudiantes de ingeniería.

A diferencia de gran parte de la literatura existente sobre análisis de aprendizaje, que predomina análisis post-hoc de datos institucionales reales de universidades norteamericanas o europeas este estudio aporta un marco metodológico replicable diseñado desde el principio para abordar las limitaciones de disponibilidad de datos y las condiciones de aprendizaje híbrido características de los programas de

ingeniería en América Latina. Específicamente, la integración de modelado predictivo validado por Monte Carlo con un protocolo de intervención estructurado de cuatro componentes, basado en parámetros calibrados para el contexto regional, representa una contribución orientada al desarrollo de marcos aplicados más que al benchmarking de modelos solos [6].

II. MATERIALES Y MÉTODOS

A. Diseño del Estudio y Muestra

Se diseñó un estudio de simulación computacional siguiendo principios cuasi-experimentales para evaluar la eficacia de intervenciones personalizadas informadas por la analítica del aprendizaje. Este enfoque se adoptó dada la ausencia de sistemas institucionales de datos suficientemente confiables, integrados y aptos para análisis longitudinales en el contexto regional, una brecha ya documentada para la educación superior latinoamericana [13], [14], y la necesidad práctica de evaluar la viabilidad del marco antes del despliegue institucional a gran escala.

Simulation parameters were derived from the empirical literature on predictors of academic performance in higher education [15], [16], calibrated to reflect conditions documented in blended and hybrid learning environments with demographic profiles and dropout rates comparable to those reported in Latin American engineering programs. Specifically, the cohort size ($n = 847$), the distribution of academic trajectory states, and the LMS activity ranges were grounded in published benchmark values [15], [16], while acknowledging that institution-specific factors—such as internet access, platform adoption rates, and socioeconomic diversity—may produce distributional variations not fully captured by the current parameterization. The Monte Carlo method was used to generate realistic probability distributions and to assess the robustness of the predictive model under input variability [17].

La cohorte simulada comprendió 847 estudiantes de programas de ingeniería, siguiendo recomendaciones de tamaño muestral para estudios de analítica de aprendizaje [18]. El período de seguimiento abarcó cuatro semestres académicos consecutivos, permitiendo la evaluación de trayectorias académicas longitudinales. Los estudiantes fueron caracterizados mediante un modelo de estados ocultos de Markov con cuatro estados posibles: alto rendimiento, rendimiento medio, situación de riesgo y deserción [19].

B. Variables y Fuentes de Datos

Las variables de entrada del modelo incluyeron indicadores de actividad en LMS: número total de sesiones, duración promedio de sesión, recursos educativos accedidos, participación en foros de discusión, entregas de actividades evaluativas y porcentaje de completación de videos instruccionales. Estas variables fueron seleccionadas con base en su demostrada capacidad predictiva según meta-análisis previos [20]. A partir de estos indicadores se calculó un índice compuesto de engagement estudiantil, ponderando cada

componente según su contribución relativa al rendimiento académico (1). Los pesos de los componentes se asignaron en función de la fuerza predictiva relativa reportada en estudios empíricos de análisis de aprendizaje basados en LMS. La frecuencia de la sesión y la duración media de la sesión ponderadas en 0,25 y 0,20, respectivamente; fueron asignadas las cargas más altas, coherentes con sus asociaciones consistentemente más fuertes con el rendimiento académico documentadas en múltiples contextos de LMS [8], [10]. El acceso a los recursos educativos recibió el mismo peso que la duración de la sesión (0,20), reflejando su papel establecido como sustituto del comportamiento de aprendizaje autorregulado [21]. La participación en el foro (0,15) y la tasa de entrega de tareas (0,10) recibieron pesos menores debido a su capacidad predictiva más variable y dependiente del contexto, mientras que completar el video instruccional (0,10) se incluyó como un indicador emergente cuya validez predictiva en entornos híbridos aún se está acumulando [10]. El índice compuesto de compromiso resultante se define de la siguiente manera:

$$\begin{aligned} \text{Engagement} = & 0.25(\text{sesiones}) + 0.20(\text{duración}) + \\ & 0.20(\text{recursos}) + 0.15(\text{foros}) + 0.10(\text{entregas}) + \\ & 0.10(\text{videos}) \end{aligned} \quad (1)$$

C. Modelo Predictivo

El modelo académico de predicción de riesgo se construyó utilizando un enfoque de regresión logística simulada, en el que la variable dependiente era una clasificación binaria de riesgo (en riesgo o abandono frente a un rendimiento académico adecuado). La regresión logística fue seleccionada por tres razones: es el clasificador de referencia más utilizado en la investigación de análisis de aprendizaje, facilitando la comparación directa con la literatura publicada [9]; produce estimaciones de coeficientes interpretables que pueden traducirse en señales pedagógicas accionables [21]; y es muy adecuado para tareas de clasificación binaria que involucran predictores de escala mixta derivados de los registros de actividad LMS. Aunque algoritmos más complejos—incluidos métodos de conjunto y redes neuronales—pueden lograr ganancias marginales en precisión predictiva, su complejidad añadida es menos justificada en la etapa de desarrollo del modelo y validación del marco que aborda este estudio [9]. Los parámetros del modelo se calibraron para obtener métricas de rendimiento consistentes con la literatura, que informa de valores AUC-ROC entre 0,75 y 0,85 para modelos similares basados en LA [21], [22]. El umbral óptimo de clasificación se determinó utilizando el índice de Youden, que maximiza la suma de sensibilidad y especificidad [23].

D. Intervenciones Personalizadas

Los alumnos categorizados como de alto riesgo ($n = 334$) fueron aleatorizados en dos grupos: intervención ($n = 167$) y control ($n = 167$). El conjunto de intervención fue objeto de un programa de asistencia personalizada que comprendió: tutorías

académicas semanales, notificaciones automatizadas sobre patrones de desconexión, recursos de aprendizaje adaptativos fundamentados en el perfil de dificultades identificado, y sesiones de orientación vocacional. El grupo de control mantuvo la adhesión al modelo educativo estándar sin la necesidad de intervenciones suplementarias.

E. Validación y Análisis Estadístico

La validación del modelo se realizó mediante simulación Monte Carlo con 1,000 iteraciones, generando distribuciones empíricas de las métricas de desempeño [24]. En cada iteración se regeneraron los datos de entrada manteniendo los parámetros poblacionales constantes, permitiendo estimar intervalos de confianza del 95% para cada métrica. Este enfoque proporciona una medida robusta de la estabilidad del modelo ante variaciones aleatorias en los datos [25].

El análisis estadístico incluyó estadística descriptiva (media, desviación estándar) para las variables continuas y frecuencias absolutas y relativas para las categóricas. La comparación de medias entre grupos se realizó mediante la prueba t de Student para muestras independientes, previa verificación de los supuestos de normalidad y homocedasticidad. La comparación de proporciones se evaluó mediante la prueba chi-cuadrado de Pearson. El tamaño del efecto se calculó utilizando la d de Cohen, interpretándose valores de 0.2, 0.5 y 0.8 como efectos pequeños, medianos y grandes, respectivamente [26]. Se estableció un nivel de significancia de $\alpha = 0.05$ para todas las pruebas. Los análisis se realizaron en Python 3.10 utilizando las bibliotecas NumPy, SciPy y Pandas [27].

III. RESULTADOS

A. Caracterización de Trayectorias Académicas

La simulación de trayectorias académicas reveló una distribución heterogénea de estados al finalizar el período de seguimiento. De los 847 estudiantes simulados, 243 (28.7%) finalizaron en estado de alto rendimiento, 305 (36.0%) en rendimiento medio, 112 (13.2%) en situación de riesgo y 187 (22.1%) en condición de deserción. La Figura 1 presenta la evolución temporal de estas trayectorias, evidenciando una progresiva migración desde estados de rendimiento adecuado hacia situaciones de riesgo y deserción conforme avanzan los semestres académicos.

B. Indicadores de Actividad en LMS

Los indicadores de actividad en LMS mostraron considerable variabilidad entre los estudiantes. El número promedio de sesiones por semestre fue de 44.8 (DE = 21.7), con un rango de 5 a 127 sesiones. El índice compuesto de engagement presentó una media de 0.48 (DE = 0.064), con valores oscilando entre 0.28 y 0.72. Se observó una correlación positiva entre el índice de engagement y el promedio académico acumulado (GPA), como se ilustra en la Figura 4.

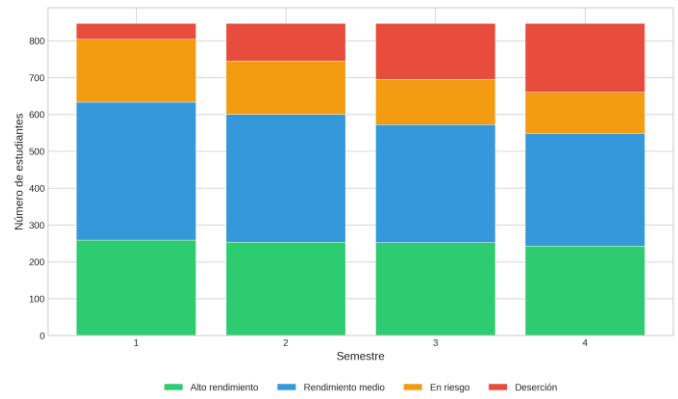


Fig. 1 Evolución de las trayectorias académicas a lo largo de los cuatro semestres de seguimiento. Los colores representan los diferentes estados: alto rendimiento (verde), rendimiento medio (azul), en riesgo (amarillo) y deserción (rojo).

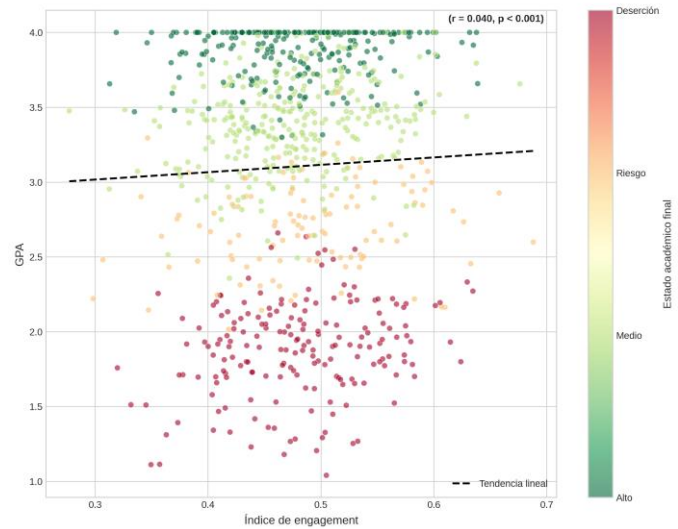


Fig. 2 Diagrama de dispersión mostrando la relación entre el índice de engagement y el GPA. El color de los puntos indica el estado académico final. La línea discontinua representa la tendencia lineal.

C. Desempeño del Modelo Predictivo

El modelo predictivo del riesgo académico llegó a índices de rendimiento satisfactorios. La sensibilidad fue del 70.9%, la precisión del 63.5% y la exactitud global alcanzó el 75.3%. La especificidad llegó al 77.7%, lo que sugiere una capacidad apropiada para identificar con precisión a los alumnos sin riesgo. La zona por debajo de la curva ROC fue de 0.803, lo que se encuentra dentro del rango reportado en la bibliografía para modelos parecidos. La curva ROC del modelo, que se muestra en la Figura 2, revela un rendimiento mucho mayor que el de un clasificador aleatorio.

El modelo predictivo mostró ser estable cuando se validó con 1,000 iteraciones de Monte Carlo. El promedio del AUC-ROC fue 0.793 (con DE = 0.026), y su intervalo de confianza al 95% fue [0.744, 0.845]. La Tabla 1 muestra las métricas de rendimiento del modelo, incluyendo los intervalos de confianza correspondientes. La amplitud reducida de los

intervalos indica que el modelo conserva un desempeño estable frente a cambios en los datos de entrada.

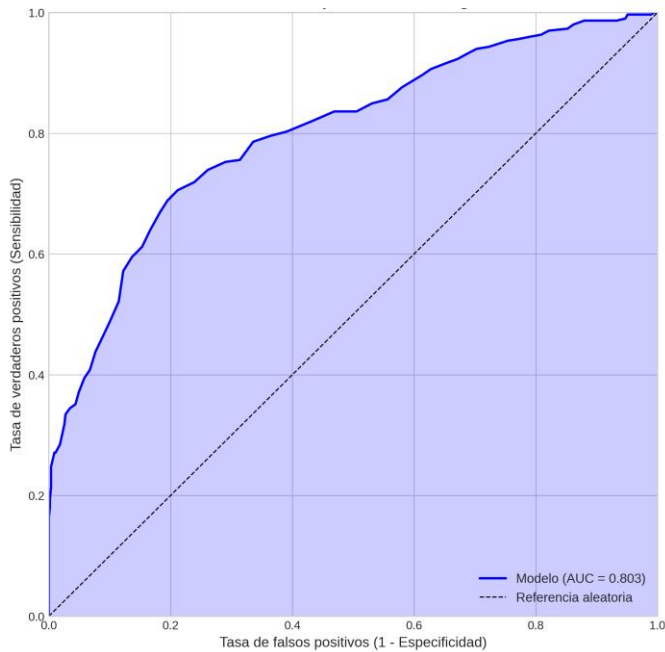


Fig. 3 Curva ROC del modelo predictivo de riesgo académico. El área sombreada representa el área bajo la curva (AUC = 0.803). La línea diagonal discontinua corresponde al clasificador aleatorio.

TABLA I

MÉTRICAS DE DESEMPEÑO DEL MODELO PREDICTIVO CON INTERVALOS DE CONFIANZA DEL 95% OBTENIDOS MEDIANTE VALIDACIÓN MONTE CARLO (1,000 ITERACIONES)

Métrica	Valor	DE	IC 95%
Exactitud	0.754	0.015	[0.723, 0.784]
Precisión	0.668	0.027	[0.614, 0.724]
Sensibilidad	0.691	0.029	[0.636, 0.748]
Especificidad	0.777	0.022	[0.731, 0.819]
F1-Score	0.679	0.021	[0.638, 0.720]
AUC-ROC	0.793	0.026	[0.744, 0.845]

Nota: DE = Desviación estándar; IC = Intervalo de confianza; AUC-ROC = Área bajo la curva ROC.

D. Impacto de las Intervenciones

Las intervenciones personalizadas demostraron un impacto positivo significativo en el rendimiento académico. El grupo de intervención ($n = 167$) alcanzó un GPA promedio de 2.68 (DE = 0.59), mientras que el grupo control ($n = 167$) obtuvo un GPA de 2.38 (DE = 0.64). La diferencia entre grupos fue estadísticamente significativa ($t = 4.45$, $p < 0.001$), con un tamaño del efecto medio (d de Cohen = 0.49). La Figura 3 presenta la comparación de distribuciones de GPA antes y después de la intervención para ambos grupos.

El análisis de retención reveló diferencias sustanciales entre los grupos. La tasa de retención en el grupo de intervención fue del 88.6%, comparada con el 73.1% en el grupo control. Esta diferencia fue estadísticamente significativa ($\chi^2 = 12.08$, $p < 0.001$), representando una

reducción relativa del 58% en el riesgo de deserción para los estudiantes que recibieron intervención personalizada. La Tabla 2 presenta el resumen comparativo de los resultados de intervención.

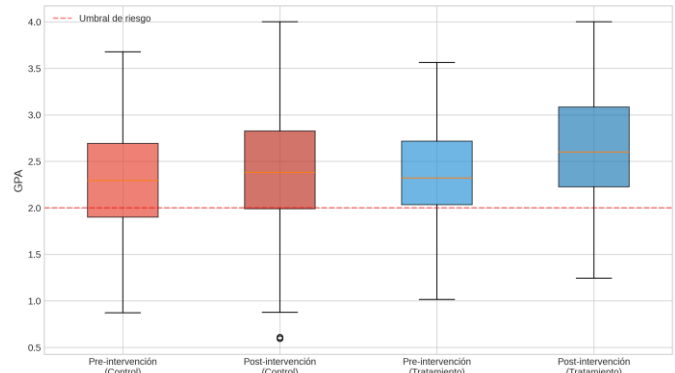


Fig. 4 Comparación de la distribución del GPA antes y después de la intervención para los grupos control y tratamiento. Las cajas representan los percentiles 25-75, la línea horizontal indica la mediana y los bigotes se extienden hasta 1.5 veces el rango intercuartílico.

TABLA II
COMPARACIÓN DE RESULTADOS ENTRE GRUPOS CONTROL E INTERVENCIÓN

Variable	Control (n=167)	Intervención (n=167)	p-valor
GPA post (M ± DE)	2.38 ± 0.64	2.68 ± 0.59	< 0.001*
Tasa de retención (%)	73.1	88.6	< 0.001†
d de Cohen	—	0.49	—

Nota: M = Media; DE = Desviación estándar. *Prueba t de Student. †Prueba chi-cuadrado.

IV. DISCUSIÓN

Los resultados de este estudio de simulación proporcionan evidencia preliminar coherente con la hipótesis central: los modelos analíticos predictivos pueden apoyar el diseño de intervenciones personalizadas con el potencial de mejorar el rendimiento académico y reducir el abandono escolar entre los estudiantes de ingeniería. Bajo las condiciones simuladas, el modelo predictivo alcanzó métricas de rendimiento comparables a las reportadas en implementaciones reales de sistemas de análisis de aprendizaje en instituciones de educación superior [28] [29], lo que sugiere que el marco metodológico es internamente coherente y opera dentro de rangos de parámetros empíricamente plausibles.

El área bajo la curva ROC de 0.803 indica una capacidad discriminativa buena del modelo para identificar estudiantes en riesgo. Este valor se sitúa dentro del rango esperado según meta-análisis previos que reportan AUC entre 0.70 y 0.85 para modelos basados en datos de LMS [30]. La sensibilidad del 70.9% implica que aproximadamente 7 de cada 10 estudiantes en riesgo serían correctamente identificados, mientras que la especificidad del 77.7% sugiere una tasa de falsos positivos

aceptable que no sobrecarga innecesariamente los recursos de intervención.

El tamaño del efecto de las intervenciones (d de Cohen = 0.49) se clasifica como medio según los criterios convencionales [26], siendo consistente con los efectos reportados en revisiones sistemáticas sobre intervenciones basadas en analítica de aprendizaje [31]. La mejora de 0.30 puntos en el GPA promedio tiene implicaciones prácticas importantes, particularmente para estudiantes que se encuentran en el umbral del rendimiento académico satisfactorio.

El incremento en la tasa de retención del 73.1% al 88.6% representa un hallazgo de considerable relevancia institucional. La deserción en programas de ingeniería constituye un problema multifactorial que genera costos significativos tanto para los estudiantes como para las instituciones [32]. Los resultados sugieren que las intervenciones tempranas y personalizadas pueden mitigar parcialmente este fenómeno, aunque es importante reconocer que factores externos al ámbito académico (económicos, familiares, vocacionales) también influyen en las decisiones de permanencia.

Estos hallazgos basados en simulación señalan varias consideraciones que las instituciones que buscan implementar sistemas de análisis de aprendizaje deberían abordar. En primer lugar, la recogida sistemática de datos de actividades estudiantiles a través de plataformas LMS debería estar respaldada por una infraestructura tecnológica adecuada. En segundo lugar, se requerirían equipos interdisciplinarios, que combinen experiencia en ciencia de datos, pedagogía y educación en ingeniería para traducir los resultados analíticos en intervenciones accionables. Tercero, serían esenciales protocolos claramente definidos que especifiquen roles, recursos y mecanismos de seguimiento para la ejecución de la intervención. Estas condiciones representan requisitos previos para la implementación en el mundo real, más que conclusiones derivables solo de la simulación.

Las consideraciones éticas merecen atención especial. El uso de datos estudiantiles para fines predictivos plantea cuestiones relacionadas con la privacidad, el consentimiento informado y el potencial de discriminación algorítmica [33]. Es imperativo que las instituciones desarrollen marcos normativos que garanticen la protección de los datos personales, la transparencia en los criterios de clasificación y el derecho de los estudiantes a conocer y cuestionar las decisiones que les afectan [34].

Este estudio presenta limitaciones inherentes a su diseño de simulación. Aunque los parámetros fueron derivados de la literatura empírica, los datos sintéticos no capturan la complejidad total de las interacciones educativas reales. La validación en entornos institucionales específicos es necesaria para confirmar la generalización de los hallazgos. Adicionalmente, el modelo no considera variables contextuales como las características del docente, la naturaleza de la asignatura o las condiciones socioeconómicas de los

estudiantes, factores que pueden modular significativamente los resultados.

Es importante distinguir entre tres niveles de afirmación en este estudio. Primero, a nivel de simulación, el modelo y los parámetros de intervención se comportan de forma coherente bajo condiciones controladas, generando métricas de rendimiento dentro del rango documentado en la literatura empírica. En segundo lugar, a nivel de validación estadística, el procedimiento de Montecarlo confirma que estas métricas son estables a lo largo de 1,000 iteraciones independientes, descartando que los resultados observados sean artefactos de un único sorteo aleatorio favorable. En tercer lugar, a nivel de aplicabilidad en el mundo real, los hallazgos se interpretan adecuadamente como pruebas de concepto: establecen la plausibilidad técnica del marco propuesto, pero no sustituyen los estudios empíricos prospectivos con poblaciones estudiantiles reales en entornos institucionales. Cada uno de estos tres niveles respalda afirmaciones distintas pero complementarias, y la contribución de este trabajo reside principalmente en los dos primeros.

Futuras investigaciones deberían orientarse hacia la implementación piloto de sistemas de analítica de aprendizaje en instituciones de la región, la incorporación de técnicas de aprendizaje automático más sofisticadas (como redes neuronales o métodos de ensamble), y la evaluación de la sostenibilidad y escalabilidad de las intervenciones propuestas. También resulta relevante explorar la percepción de estudiantes y docentes respecto a estos sistemas, aspecto fundamental para su aceptación y adopción efectiva.

V. CONCLUSIONES

Este estudio demuestra que la analítica de aprendizaje constituye una herramienta valiosa para la personalización curricular en entornos de educación híbrida en ingeniería. El modelo predictivo desarrollado alcanzó un desempeño diagnóstico adecuado ($AUC-ROC = 0.803$), permitiendo la identificación temprana de estudiantes en situación de riesgo académico con sensibilidad y especificidad satisfactorias.

Bajo las condiciones de intervención simuladas, las estrategias de apoyo personalizadas se asociaron con mejoras significativas en el rendimiento académico (aumento de nota media de 0,30 puntos; $d = 0,49$) y en la retención de estudiantes (aumento de 15,5 puntos porcentuales). Estos hallazgos son coherentes con la hipótesis de que integrar la analítica del aprendizaje en la gestión educativa tiene potencial para reducir las tasas de abandono en los programas de ingeniería, aunque se requiere validación empírica en entornos institucionales reales antes de poder hacer afirmaciones causales.

La implementación exitosa de estos sistemas requiere consideraciones institucionales, técnicas y éticas que trascienden el desarrollo de modelos predictivos. Las facultades de ingeniería que adopten estos enfoques deben garantizar la infraestructura tecnológica adecuada, la capacitación del personal académico y administrativo, y el

establecimiento de marcos normativos que protejan los derechos de los estudiantes.

Los hallazgos de esta investigación proporcionan un marco de referencia para instituciones de educación superior en América Latina que buscan implementar estrategias de analítica de aprendizaje adaptadas a sus contextos específicos. La validación empírica en entornos reales constituye el paso necesario para confirmar la efectividad de estas intervenciones y ajustar los modelos a las particularidades de cada población estudiantil.

AGRADECIMIENTOS

Esta investigación fue apoyada por el Instituto de Investigación en Ciencias Médicas y Derecho a la Salud (ICIMEDES) de la Facultad de Ciencias Médicas (FCM) y por la Dirección de Investigación Científica, Humanística y Tecnológica (DICIHT) todas ellas pertenecientes a la Universidad Nacional Autónoma de Honduras (UNAH).

REFERENCIAS

- [1] M. W. Ohland et al., "Persistence, engagement, and migration in engineering programs", *J. Eng. Educ.*, vol. 97, no. 3, pp. 259–278, Jul. 2008. doi: 10.1002/j.2168-9830.2008.tb00978.x
- [2] D. Ifenthaler and C. Widanapathirana, "Development and validation of a learning analytics framework: Two case studies using support vector machines", *Technol. Knowl. Learn.*, vol. 19, no. 1–2, pp. 221–240, Jul. 2014. doi: 10.1007/s10758-014-9226-4
- [3] G. Siemens and R. S. J. d. Baker, "Learning analytics and educational data mining: Towards communication and collaboration", in *Proc. 2nd Int. Conf. Learn. Anal. Knowl.*, Vancouver, BC, Canada, 2012, pp. 252–254. doi: 10.1145/2330601.2330661
- [4] G. Siemens, "Learning analytics: The emergence of a discipline", *Am. Behav. Sci.*, vol. 57, no. 10, pp. 1380–1400, Oct. 2013. doi: 10.1177/0002764213498851
- [5] D. Gasevic, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are about learning", *TechTrends*, vol. 59, no. 1, pp. 64–71, Jan. 2015. doi: 10.1007/s11528-014-0822-x
- [6] K. E. Arnold and M. D. Pistilli, "Course signals at Purdue: Using learning analytics to increase student success", in *Proc. 2nd Int. Conf. Learn. Anal. Knowl.*, Vancouver, BC, Canada, 2012, pp. 267–270. doi: 10.1145/2330601.2330666
- [7] S. Jayaprakash, E. Moody, E. Lauria, J. Regan, and J. Baron, "Early alert of academically at-risk students: An open source analytics initiative", *J. Learn. Anal.*, vol. 1, no. 1, pp. 6–47, May 2014. doi: 10.18608/jla.2014.11.3
- [8] R. Cerezo, M. Sánchez-Santillán, M. P. Paule-Ruiz, and J. C. Núñez, "Students' LMS interaction patterns and their relationship with achievement: A case study in higher education", *Comput. Educ.*, vol. 96, pp. 42–54, May 2016. doi: 10.1016/j.compedu.2016.02.006
- [9] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey", *WIREs Data Min. Knowl. Discov.*, vol. 10, no. 3, e1355, May 2020. doi: 10.1002/widm.1355
- [10] A. Pardo, F. Han, and R. A. Ellis, "Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance", *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 82–92, Jan. 2017. doi: 10.1109/TLT.2016.2639508
- [11] R. F. Kizilcec and H. Lee, "Algorithmic fairness in education," in *The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates*, W. Holmes and K. Porayska-Pomsta, Eds. New York, NY, USA: Routledge, 2022, pp. 174–202, doi: 10.4324/9780429329067-10.
- [12] S. Dawson, D. Gasevic, G. Siemens, and S. Joksimovic, "Current state and future trends: A citation network analysis of the learning analytics field", in *Proc. 4th Int. Conf. Learn. Anal. Knowl.*, Indianapolis, IN, USA, 2014, pp. 231–240. doi: 10.1145/2567574.2567585
- [13] I. Hilliger et al., "Towards learning analytics adoption: A mixed methods study of data-related practices and policies in Latin American universities", *Br. J. Educ. Technol.*, vol. 51, no. 4, pp. 915–937, 2020, doi: 10.1111/bjet.12933
- [14] F. Pedró, M. Bouckaert, and V. Gutović, "Bridging the data gap in higher education policymaking: Challenges and opportunities for evidence-based governance", *UNESCO World Education Blog*, Jun. 25, 2025. [Online]. Available: https://www.iesalc.unesco.org/en/articles/bridging-data-gap-higher-education-policy-making-challenges-and-opportunities-evidence-based?utm_source=chatgpt.com
- [15] D. T. Tempelaar, B. Rienties, and B. Giesbers, "In search for the most informative data for feedback generation: Learning analytics in a data-rich context", *Comput. Hum. Behav.*, vol. 47, pp. 157–167, Jun. 2015. doi: 10.1016/j.chb.2014.05.038
- [16] O. H. T. Lu, A. Y. Q. Huang, J. C. H. Huang, A. J. Q. Lin, H. Ogata, and S. J. H. Yang, "Applying learning analytics for the early prediction of students' academic performance in blended learning", *J. Educ. Technol. Soc.*, vol. 21, no. 2, pp. 220–232, Apr. 2018.
- [17] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*, 3rd ed. Hoboken, NJ, USA: Wiley, 2016. doi: 10.1002/9781118631980.
- [18] J. Saldaña, *The Coding Manual for Qualitative Researchers*, 3rd ed. Thousand Oaks, CA, USA: SAGE, 2016.
- [19] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989. doi: 10.1109/5.18626.
- [20] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic, "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," *Internet High. Educ.*, vol. 28, pp. 68–84, Jan. 2016, doi: 10.1016/j.iheduc.2015.10.002.
- [21] W. Greller and H. Drachsler, "Translating learning into numbers: A generic framework for learning analytics", *J. Educ. Technol. Soc.*, vol. 15, no. 3, pp. 42–57, Jul. 2012.
- [22] A. F. Wise, "Designing pedagogical interventions to support student use of learning analytics", *Br. J. Educ. Technol.*, vol. 45, no. 6, pp. 894–906, Nov. 2014. doi: 10.1111/bjet.12209
- [23] W. J. Youden, "Index for rating diagnostic tests", *Cancer*, vol. 3, no. 1, pp. 32–35, Jan. 1950. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
- [24] N. Metropolis and S. Ulam, "The Monte Carlo method", *J. Am. Stat. Assoc.*, vol. 44, no. 247, pp. 335–341, Sep. 1949. doi: 10.1080/01621459.1949.10483310
- [25] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York, NY, USA: Chapman & Hall, 1993.
- [26] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, USA: Erlbaum, 1988.
- [27] W. McKinney, "Data structures for statistical computing in Python", in *Proc. 9th Python Sci. Conf.*, Austin, TX, USA, 2010, pp. 56–61. doi: 10.25080/Majora-92bf1922-00a
- [28] J. P. Campbell, P. B. DeBlois, and D. G. Oblinger, "Academic analytics: A new tool for a new era", *EDUCAUSE Rev.*, vol. 42, no. 4, pp. 40–57, Jul. 2007.
- [29] M. D. Pistilli and K. E. Arnold, "Purdue signals: Mining real-time academic data to enhance student success", *About Campus*, vol. 15, no. 3, pp. 22–24, Jul. 2010. doi: 10.1002/abc.20025
- [30] A. Peña-Ayala, "Learning analytics: A glance of evolution, status, and trends according to a proposed taxonomy", *WIREs Data Min. Knowl. Discov.*, vol. 8, no. 3, e1243, May 2018. doi: 10.1002/widm.1243
- [31] O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, "The current landscape of learning analytics in higher education", *Comput. Hum. Behav.*, vol. 89, pp. 98–110, Dec. 2018. doi: 10.1016/j.chb.2018.07.027
- [32] V. Tinto, "Dropout from higher education: A theoretical synthesis of recent research", *Rev. Educ. Res.*, vol. 45, no. 1, pp. 89–125, Mar. 1975. doi: 10.3102/00346543045001089
- [33] S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas", *Am. Behav. Sci.*, vol. 57, no. 10, pp. 1510–1529, Oct. 2013. doi: 10.1177/0002764213479366
- [34] A. Pardo and G. Siemens, "Ethical and privacy principles for learning analytics", *Br. J. Educ. Technol.*, vol. 45, no. 3, pp. 438–450, May 2014. doi: 10.1111/bjet.12152.