



# **Modelos de analítica predictiva para la gestión del rendimiento académico en estudiantes de posgrado en modalidad virtual**

Predictive Analytics Models for Managing Academic Performance in  
Graduate Students in Virtual Learning Environments

**Esther Rosa Saenz Arenas**

ID 0000-0003-0340-2198

Universidad Científica Del Sur

[esaenza@cientifica.edu.pe](mailto:esaenza@cientifica.edu.pe)

## Resumen

**Objetivo:** Evaluar la eficacia de los modelos de analítica predictiva en la identificar factores vinculados al rendimiento académico en alumnos de maestría virtual de universidades privadas de Lima Metropolitana. **Metodología:** El estudio efectuó un enfoque de tipo cuantitativo, de diseño correlacional-predictivo. Teniendo por muestra de 120 alumnos de programas de posgrado de manera virtual en universidades privadas en el período académico 2024-2025. Asimismo, para valorar la relación de las variables académicas, sociodemográficas y modelos de cooperación en plataformas digitales se efectuaron técnicas de regresión logística y árboles de decisión. El instrumento usado se consiguió a través de cuestionarios organizados y validados por análisis sistemático de registros institucionales. **Resultados:** El modelo de analítica predictiva efectuado alcanzó una precisión 79.2% de precisión en la clasificación interna, lo que sugiere viabilidad técnica en este contexto. Las variables que mostraron mayor nivel de significancia estadística fueron: tiempo dedicado al estudio semanal ( $\beta=0.49$ ,  $p<0.01$ ), participación activa en actividades simultáneas ( $\beta=0.56$ ,  $p<0.01$ ), y experiencia previa en modelos de educación virtual ( $\beta=0.45$ ,  $p<0.05$ ). **Conclusión:** La implementación táctica de modelos predictivos consiente a las instituciones diseñar e implementar intervenciones formativas focalizadas y oportunas.

**Palabras clave:** aprendizaje automático, educación superior digital, deserción académica, educación en línea, tecnología educativa, intervención pedagógica, minería de datos.

## Abstract

**Objective:** To evaluate the effectiveness of predictive analytics models in identifying factors related to academic performance among master's students in virtual programs at private universities in Lima Metropolitan Area. **Methodology:** The study employed a quantitative approach with a correlational-predictive design. The sample consisted of 120 postgraduate students enrolled in virtual programs at private universities during the 2024–2025 academic period. To assess the relationship between academic, sociodemographic variables and patterns of collaboration in digital platforms, logistic regression and decision tree techniques were applied. The instrument was developed through structured questionnaires validated by a systematic analysis of institutional records. **Results:** The predictive analytics model achieved an accuracy rate of 79.2% in classifying academic performance. The variables that demonstrated the highest statistical significance were weekly study time ( $\beta = 0.49$ ,  $p < 0.01$ ), active participation in synchronous activities ( $\beta = 0.56$ ,  $p < 0.01$ ), and previous experience with virtual education models ( $\beta = 0.45$ ,  $p < 0.05$ ). **Conclusion:** The strategic implementation of predictive models enables institutions to design and execute targeted and timely educational interventions.

**Keywords:** machine learning, digital higher education, academic dropout, online education, educational technology, pedagogical intervention, data mining.



## Introducción

En la última década, la educación superior ha experimentado una metamorfosis rápida que ha posicionado a la modalidad virtual opción consolidada para la enseñanza en posgrado. La pandemia de 2020 impulsó este cambio al demandar que las universidades modifiquen sus modelos presenciales a plataformas digitales (García-Peñalvo et al., 2021). En América Latina, y particularmente en Perú, las instituciones privadas han liderado la implementación de maestrías totalmente virtuales, ante la necesidad de profesionales que requieren equilibrar su formación académica con sus responsabilidades laborales. (Cabero-Almenara & Valencia-Ortiz, 2021).

El crecimiento de la educación virtual en posgrado ha expuesto problemas evidentes en la continuidad estudiantil y en el rendimiento académico. Estudios coinciden que entre el 40 % y el 60 % de los matriculados en modelos virtuales interrumpen sus estudios, cifras que superan ampliamente los índices de la educación presencial (Henderikx et al., 2019; Muljana & Luo, 2019). Esta realidad compromete la gestión institucional, desajusta las proyecciones profesionales de los estudiantes y debilita la percepción de calidad asociada a la modalidad virtual de posgrado (Bañeres et al., 2020).

Frente al aumento sostenido de la deserción en posgrado virtual, la analítica predictiva surge como una opción de gestión orientada a la prevención de problemas académicos. A través de algoritmos de aprendizaje automático se estudian grandes volúmenes de datos producidos por los estudiantes, reconociendo señales tempranas que inciden en el rendimiento académico (Hellas et al., 2018; Waheed et al., 2020). La modalidad virtual registra de forma sistemática datos de interacción, navegación y desempeño, por lo que se cuenta con datos extensos que permiten una alta resolución para construir modelos de predicción sin alterar la dinámica del aprendizaje. (Viberg et al., 2018).

Investigaciones actuales, muestran que los modelos de predicción llegan a niveles de precisión del 80 %, al identificar alumnos de posgrado con riesgo académico en la modalidad virtual (Aljohani et al., 2019; Nuankaew et al., 2019). Investigaciones de Europa y norteamericana concuerdan en destacar como determinantes clave la frecuencia de ingreso a la plataforma, la participación en foros, el tiempo en actividades simultáneas y el rendimiento en pruebas iniciales (Ifenthaler & Yau, 2020; Tsai et al., 2020). No obstante, la mayoría de estos trabajos se efectuaron en situaciones con infraestructura tecnológica afianzada y culturas académicas específicos, lo que reduce la posibilidad de trasladar sin ajustes sus modelos predictivos al entorno latinoamericano.

La investigación regional ha comenzado a examinar si los modelos predictivos desarrollados en el Norte Global mantienen su eficacia en posgrado virtual cuando se incorporan realidades latinoamericanas. Zambrano et al. (2021) evidenciaron en instituciones colombianas que la brecha digital, el ancho de banda reducido y las responsabilidades laborales explican una proporción significativa de la varianza en el rendimiento académico. En México, Pérez-López et al. (2021) confirmaron que el apoyo institucional y la competencia digital inicial



moderan la relación entre participación y éxito en la modalidad virtual. Estos resultados exigen modelos predictivos que incluyan variables de contexto propias de América Latina.

Desde el plano teórico, la investigación adopta la perspectiva del aprendizaje autorregulado de Zimmerman (2002) y su reformulación digital de Broadbent y Poon (2015), que vinculan el rendimiento académico en posgrado virtual con la capacidad del estudiante para gestionar cognitivamente su propio aprendizaje. Dicha visión se complementa con la ecuación de distancia transaccional de Moore (1993) revisada por Huang et al. (2020), que establece que la combinación de estructura del curso, diálogo educativo y autonomía configura el resultado académico en modelos virtuales. Estas referencias teóricas otorgan el sustento conceptual para entender por qué ciertas variables emergen como predictoras en la analítica predictiva.

A pesar de los progresos reportados, persiste una brecha significativa de investigación en universidades privadas del Perú. Las universidades privadas de Lima Metropolitana han incrementado rápidamente sus programas de posgrado en modalidad virtual, sin contar con evidencia local que fundamente el uso de modelos de analítica predictiva para la gestión académica (Rodríguez-Abitia & Bribiesca-Correa, 2021). Esta carencia adquiere especial relevancia dado que el perfil del estudiante de posgrado virtual en estas universidades privadas se caracteriza por ser profesional en ejercicio, con edades comprendidas entre 28 y 45 años, que organiza estudios con obligaciones laborales y familiares (García-Chitiva & Suárez-Guerrero, 2019).

La formulación del problema de estudio sería: ¿En qué medida los modelos de analítica predictiva permiten identificar componentes relacionados al rendimiento académico de alumnos de maestría virtual en universidades privadas de Lima Metropolitana? Esta pregunta principal se desglosa en problemas específicos: ¿Qué variables académicas, demográficas y de participación digital pronostican de forma significativa el rendimiento académico? ¿Qué grado de precisión logran los modelos de predicción en este escenario determinado? ¿Cuáles son las implicaciones prácticas de estos hallazgos para el diseño de intervenciones pedagógicas?

La hipótesis principal sostiene que la implementación de modelos de analítica predictiva permite identificar con exactitud superior al 75% los componentes críticos vinculados al rendimiento académico, teniendo a las variables de participación digital como las de mayor peso de predicción. Concretamente, hipotetizar que la participación en acciones sincrónicas, el tiempo de estudio a la semana y la experiencia previa en educación virtual componen los predictores más reveladores del rendimiento académico en este escenario.

El objetivo general de esta investigación es Analizar el uso de modelos predictivos para reconocer componentes relacionados al rendimiento académico en estudiantes de maestría virtual de universidades privadas de Lima Metropolitana. Los objetivos específicos serían: 1. Determinar el grado de precisión de modelos predictivos establecidos en regresión logística y árboles de decisión, (2) identificar las variables con mayor poder de predicción del rendimiento académico, y (3) establecer implicaciones prácticas para el diseño de sistemas de alerta temprana institucionales. La investigación enriquece el conocimiento existente al presentar evidencia empírica colocada sobre la aplicabilidad y los resultados de la analítica predictiva en



los posgrados virtuales de universidades privadas del Perú. (4) Contrastar la hipótesis planteada respecto al nivel de precisión y la significancia estadística de las variables predictoras.

## Materiales y métodos

La investigación se apoyó en un Enfoque cuantitativo, diseño no experimental, transversal y de alcance predictivo, dirigido a identificar relaciones estadísticamente relevantes de variables independientes y el rendimiento académico en la modalidad virtual de posgrado. Se empleó un diseño no experimental de corte transversal, recogiendo información una sola vez durante el segundo semestre del período 2024-2025 en universidades privadas de Lima. Asimismo, se utilizó un muestreo probabilístico estratificado, considerando universidad y programa académico como estratos. Esta estrategia permitió registrar el comportamiento académico sin alterar las variables de estudio, asegurando la validez ecológica de los resultados en contextos educativos reales.

La unidad de análisis la integraron estudiantes de maestría en modalidad virtual de universidades privadas de Lima Metropolitana. La población accesible abarcó 287 estudiantes de posgrado que reunieron los criterios de selección. El tamaño de la muestra se calculó a través de la fórmula para poblaciones finitas, con un 95 % de confianza, 5 % de error y varianza máxima, dando por resultado 120 casos. La elección se efectuó por medio del muestreo probabilístico estratificado conveniente, organizando la muestra de acuerdo a la universidad de origen y programa académico. Esta estratificación garantizó la representatividad de diferentes áreas disciplinares y perfiles institucionales.

Los criterios de inclusión exigieron estar activos en programas de maestría completamente virtuales de universidades privadas, haber aprobado al menos dos ciclos académicos, tener edad entre 25 y 50 años, y disponer de acceso estable a internet y dispositivos tecnológicos adecuados. Se excluyó a estudiantes con licencias académicas vigentes, a quienes cursaban solo cursos de nivelación o propedéuticos, a quienes registraban menos del 70 % de asistencia virtual y a quienes no firmaron el consentimiento informado. Estos filtros garantizaron una muestra con trayectorias académicas suficientemente consolidadas para alimentar modelos predictivos válidos en modalidad virtual.

La técnica de recolección integró encuestas estructuradas con el raspado de datos generados en las plataformas de gestión del aprendizaje de las universidades privadas. El instrumento cuantitativo, 45 ítems, cubrió cuatro dimensiones: características sociodemográficas, experiencia en modalidad virtual, estrategias de autorregulación del aprendizaje y percepción de apoyo institucional. La validación de contenido, mediante juicio de cinco especialistas en educación virtual, dio por resultado un V de Aiken  $> 0.85$  en todos los reactivos. Un estudio piloto con 30 estudiantes de posgrado virtual confirmó la confiabilidad ( $\alpha$  de Cronbach = 0.89), asegurando alta consistencia interna antes de la aplicación definitiva.



Simultáneamente, se recuperaron datos objetivos almacenados en las plataformas educativas: accesos semanales, duración de conexión, participaciones en foros asincrónicos, asistencia a clases sincrónicas, cumplimiento de fechas de entrega, notas obtenidas en evaluaciones parciales y finales, y frecuencia de descarga de recursos. Esta combinación de fuentes, registros sistemáticos y respuestas al cuestionario, aumentó la validez del estudio y controló la deseabilidad social. Ambos conjuntos de datos (cuestionario y registros digitales) fueron integrados en una base unificada mediante identificadores anónimos, lo que permitió correlacionar respuestas subjetivas con métricas objetivas de desempeño virtual.

La variable dependiente, rendimiento académico, se categorizó en alto (promedio ponderado  $\geq 15$  sobre 20) y bajo ( $< 15$ ), conforme al reglamento de las universidades privadas que fija 15 como límite satisfactorio en posgrado. Las independientes se agruparon en: (1) sociodemográficas, edad, género, estado civil, situación laboral y nivel educativo anterior; (2) indicadores de participación digital en entornos virtuales, accesos por semana, tiempo total en línea, intervenciones en sincrónicas y asincrónicas, puntualidad en entregas; y (3) antecedentes académicos, experiencia previa en educación virtual, promedio de pregrado y años desde el último estudio formal.

El estudio contó con autorización previa de los comités de ética de las universidades privadas participantes. Los estudiantes fueron informados sobre los fines de la investigación, la naturaleza libre de su participación, el tratamiento confidencial de sus datos personales y académicos, y la posibilidad de abandonar en cualquier momento. Aceptaron mediante consentimiento informado digital con firma electrónica verificable. Para garantizar privacidad, se asignaron códigos alfanuméricos a cada caso, eliminando identificadores personales antes del análisis del rendimiento académico en modalidad virtual.

El procesamiento se realizó en tres etapas consecutivas empleando SPSS 27 y Python 3.9 con librerías especializadas en analítica predictiva. En la etapa descriptiva se calcularon medidas de tendencia central y dispersión de todas las variables. La inferencia incluyó prueba Kolmogorov-Smirnov para normalidad, chi-cuadrado para asociación entre categorías y correlación Spearman para datos ordinales. La fase predictiva contrastó dos modelos: regresión logística binaria y árboles de decisión CART. Se empleó validación cruzada  $k=10$  para estimar precisión, sensibilidad, especificidad y área bajo la curva ROC. Los algoritmos se entrenaron con el 70 % de los casos y se probaron con el 30 % restante, siguiendo el protocolo habitual en analítica predictiva de rendimiento académico. Se fijó  $p < 0.05$  como umbral de significancia para garantizar robustez inferencial.



## Resultados

Los hallazgos mostraron relaciones significativas entre variables independientes y rendimiento académico en entornos virtuales. Del total de 120 estudiantes de maestría de universidades privadas, 68 presentaron alto rendimiento (56.7 %) y 52 bajo rendimiento (43.3 %), balance que permitió generar modelos predictivos robustos y adecuadamente representativos del comportamiento académico en la modalidad virtual.

**Tabla 1.** Características sociodemográficas y académicas de la muestra según nivel de rendimiento

| Característica                      | Alto rendimiento (n=68) | Bajo rendimiento (n=52) | Total (n=120) |
|-------------------------------------|-------------------------|-------------------------|---------------|
| <b>Edad (años)</b>                  |                         |                         |               |
| Media (DE)                          | 35.2 (5.4)              | 33.8 (6.3)              | 34.6 (5.8)    |
| <b>Género</b>                       |                         |                         |               |
| Femenino                            | 41 (60.3%)              | 28 (53.8%)              | 69 (57.5%)    |
| Masculino                           | 27 (39.7%)              | 24 (46.2%)              | 51 (42.5%)    |
| <b>Situación laboral</b>            |                         |                         |               |
| Empleado tiempo completo            | 62 (91.2%)              | 45 (86.5%)              | 107 (89.2%)   |
| Empleado medio tiempo               | 6 (8.8%)                | 7 (13.5%)               | 13 (10.8%)    |
| <b>Responsabilidades familiares</b> |                         |                         |               |
| Sí                                  | 46 (67.6%)              | 33 (63.5%)              | 79 (65.8%)    |
| No                                  | 22 (32.4%)              | 19 (36.5%)              | 41 (34.2%)    |
| <b>Experiencia previa virtual</b>   |                         |                         |               |
| Sí                                  | 35 (51.5%)              | 15 (28.8%)              | 50 (41.7%)    |
| No                                  | 33 (48.5%)              | 37 (71.2%)              | 70 (58.3%)    |
| <b>Promedio pregrado</b>            |                         |                         |               |
| Media (DE)                          | 14.8 (1.2)              | 13.6 (1.5)              | 14.3 (1.4)    |

Nota: DE=Desviación estándar. Valores expresados en n (%) para variables categóricas.

La Tabla 1 presenta las características descriptivas de la muestra estratificada según nivel de rendimiento académico. En variables sociodemográficas básicas la muestra presentó homogeneidad entre grupos; en cambio, la experiencia previa en educación virtual mostró disparidades significativas. Mientras que la mitad (51.5%) de quienes alcanzaron alto rendimiento ya había estudiado en formatos virtuales, solo 28.8% de los de bajo rendimiento reportaba dicha experiencia. El dominio previo de la modalidad virtual aparece así como factor diferenciador clave en el rendimiento académico de universidades privadas. Asimismo, el promedio de pregrado mostró una diferencia de 1.2 puntos entre grupos, indicando que el desempeño académico histórico mantiene cierta influencia predictiva incluso en niveles de posgrado.



**Gráfico 1.** Distribución de experiencia previa en educación virtual según rendimiento académico



Nota:  $\chi^2 = 6.42$ ,  $p = 0.011$ . Diferencia estadísticamente significativa en la distribución de experiencia previa.

El contraste muestra una diferencia significativa en experiencia previa entre ambos segmentos ( $\chi^2 = 6.42$ ,  $p = 0.011$ ). En el grupo de alto rendimiento la proporción es prácticamente simétrica: 51.5 % con experiencia previa en modalidad virtual y 48.5 % sin ella. En cambio, el grupo de bajo rendimiento concentra 71.2 % de estudiantes noveles. Este patrón indica que la adaptación a entornos virtuales actúa como barrera inicial; quienes dominan herramientas digitales, estrategias de autorregulación y comunicación asincrónica acceden con mayor facilidad al rendimiento académico esperado en posgrado.

**Tabla 2.** Variables de participación digital según nivel de rendimiento académico

| Variable                         | Alto<br>(n=68)   | rendimiento | Bajo<br>(n=52)   | rendimiento | Diferencia   | Valor p |
|----------------------------------|------------------|-------------|------------------|-------------|--------------|---------|
| Frecuencia acceso sem (días)     | M=5.8 (DE=0.9)   |             | M=3.2 (DE=1.3)   |             | 2.6 días     | <0.001  |
| Tiempo conexión sem (horas)      | M=12.4 (DE=2.6)  |             | M=7.1 (DE=2.8)   |             | 5.3 horas    | <0.001  |
| Participación foros (interv/mes) | M=18.6 (DE=4.2)  |             | M=8.3 (DE=3.7)   |             | 10.3 interv. | <0.001  |
| Asistencia sesiones sincrón (%)  | M=87.3 (DE=8.1)  |             | M=62.4 (DE=12.5) |             | 24.9%        | <0.001  |
| Puntualidad entregas (%)         | M=91.5 (DE=6.8)  |             | M=71.2 (DE=11.3) |             | 20.3%        | <0.001  |
| Descargas materiales (% total)   | M=94.2 (DE=5.6)  |             | M=78.5 (DE=9.8)  |             | 15.7%        | <0.001  |
| Tiempo prom. x sesión (min.)     | M=68.4 (DE=12.3) |             | M=42.7 (DE=15.6) |             | 25.7 min     | <0.001  |

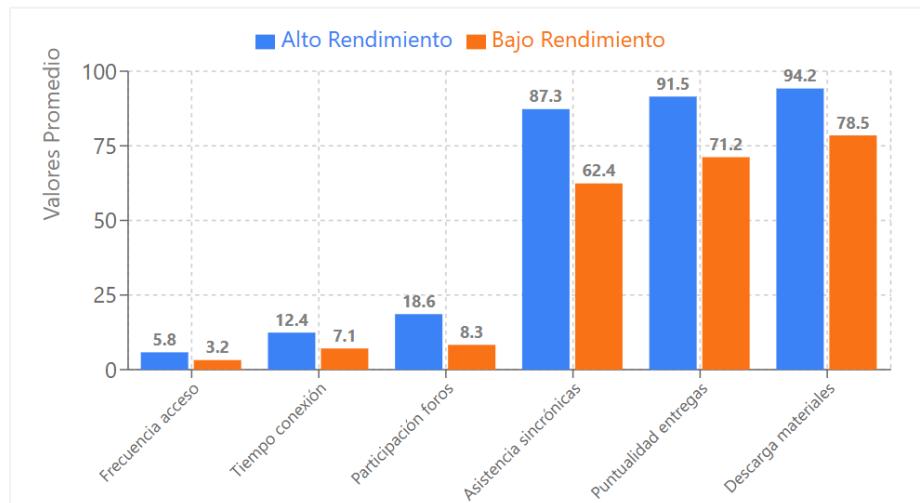
Nota: M=Media, DE=Desviación estándar. Prueba U de Mann-Whitney para variables continuas.

La Tabla 2 evidencia diferencias estadísticamente significativas en todas las variables de participación digital entre ambos grupos de rendimiento. El análisis muestra que quienes alcanzaron alto rendimiento acceden 2,6 días más por semana a la plataforma educativa. La diferencia más marcada aparece en la asistencia a sesiones sincrónicas: 25 puntos porcentuales de distancia (87,3 % frente a 62,4 %;  $p < 0,001$ ). Esta brecha subraya el valor de la interacción directa en tiempo real con docentes y compañeros para el aprendizaje colaborativo en entornos virtuales. Paralelamente, ese mismo grupo duplicó sus aportes en foros asincrónicos,



evidenciando que la participación activa en espacios discursivos correlaciona fuertemente con el éxito académico en programas de posgrado virtual.

**Gráfico 2.** Comparación de variables de participación digital entre grupos de rendimiento académico



Nota: Todas las diferencias estadísticamente significativas ( $p < 0.001$ ) según prueba U de Mann-Whitney.

El gráfico evidencia brechas consistentes y sustanciales de los dos grupos en todas las dimensiones de participación digital valoradas. La asistencia a sesiones sincrónicas muestra la diferencia más pronunciada en términos absolutos (24.9 puntos porcentuales), seguida de la puntualidad en entregas (20.3 puntos). Este patrón confirma que el compromiso activo y sostenido con actividades académicas virtuales constituye un diferenciador crucial del rendimiento. Resulta significativo que la descarga de materiales, conducta aparentemente pasiva, muestre una brecha de 15,7 puntos, confirmando que los estudiantes exitosos gestionan de forma más exhaustiva los recursos educativos. La disparidad se acentúa en el tiempo de estudio: más de una hora por sesión para alto rendimiento contra poco más de 40 minutos para bajo rendimiento, lo que refleja diferencias en profundidad de procesamiento y en la aplicación de estrategias autorregulatorias decisivas para el desempeño académico en posgrado virtual.

**Tabla 3.** Matriz de correlación entre variables predictoras y rendimiento académico

| Variable independiente                   | Rho Spearman | Valor p | Nivel significancia |
|--|--------------|---------|---------------------|
| Participación en actividades sincrónicas | 0.682        | <0.001  | ***                 |
| Tiempo de estudio semanal                | 0.624        | <0.001  | ***                 |
| Frecuencia de acceso a plataforma        | 0.591        | <0.001  | ***                 |
| Puntualidad en entregas                  | 0.563        | <0.001  | ***                 |
| Participación en foros asincrónicos      | 0.542        | <0.001  | ***                 |
| Experiencia previa en educación virtual  | 0.506        | <0.001  | ***                 |
| Promedio de pregrado                     | 0.428        | <0.001  | ***                 |
| Descarga de materiales educativos        | 0.387        | <0.001  | ***                 |

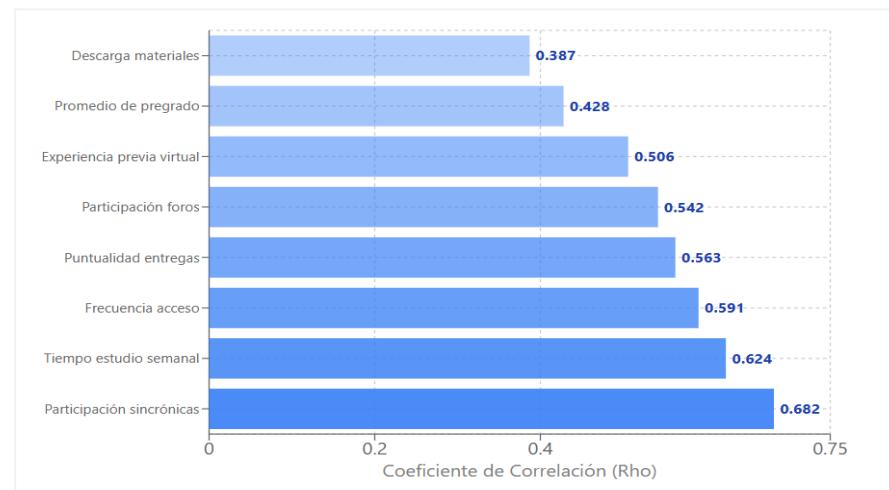


|                              |        |       |    |
|------------------------------|--------|-------|----|
| Responsabilidades familiares | -0.156 | 0.089 | ns |
| Edad del estudiante          | 0.178  | 0.052 | ns |
| Género                       | 0.094  | 0.308 | ns |
| Situación laboral            | 0.121  | 0.189 | ns |

Nota: p<0.001, ns=no significativo. Rendimiento académico codificado como variable ordinal (1=bajo, 2=alto).

La Tabla 3 presenta el análisis correlacional que identificó las variables con mayor asociación al rendimiento académico. Los resultados confirman que la asistencia a actividades sincrónicas es el predictor más potente ( $\rho = 0.682$ ,  $p < 0.001$ ), mostrando una asociación positiva fuerte con el rendimiento académico en modalidad virtual. Le siguen el tiempo semanal de estudio ( $\rho = 0.624$ ) y la frecuencia de acceso a la plataforma ( $\rho = 0.591$ ), ambos altamente significativos. Llama la atención que variables sociodemográficas, edad, género y situación laboral, no alcanzaron correlaciones significativas, indicando que los patrones de participación digital explican mejor el desempeño que las características personales. Incluso las responsabilidades familiares, habitualmente consideradas barrera en la literatura, no mostraron relación estadísticamente relevante en este contexto de universidades privadas.

**Gráfico 3. Importancia relativa de variables predictoras según coeficiente de correlación**



Nota: Todas las correlaciones mostradas son estadísticamente significativas ( $p < 0.001$ ).

La visualización ordenada de correlaciones confirma que indicadores de actividad en línea son los principales predictores, mientras que datos demográficos aportan escasa capacidad explicativa. Ocho variables de comportamiento superan  $\rho = 0,40$  y resultan significativas; ninguna sociodemográfica alcanza ese nivel. La asistencia a sesiones sincrónicas ( $\rho = 0,682$ ) domina el espectro, y el bloque de las tres variables de conexión activa, todas  $\rho > 0,59$ , revela que la implicación continuada con la plataforma define el rendimiento académico en la modalidad virtual de las universidades privadas estudiadas. Un hallazgo contraintuitivo es la



correlación relativamente modesta del promedio de pregrado ( $\rho=0.428$ ), indicando que el rendimiento histórico tiene menor capacidad predictiva que los comportamientos actuales en el entorno virtual.

**Tabla 4.** Rendimiento comparativo de modelos predictivos y coeficientes de regresión logística

| <b>A. Métricas de rendimiento de modelos</b> |           |              |               |         |  |
|--|-----------|--------------|---------------|---------|--|
| Modelo                                       | Precisión | Sensibilidad | Especificidad | AUC-ROC |  |
| Regresión logística                          | 79.2%     | 82.4%        | 75.0%         | 0.842   |  |
| Árbol de decisión (CART)                     | 76.7%     | 78.8%        | 73.8%         | 0.818   |  |

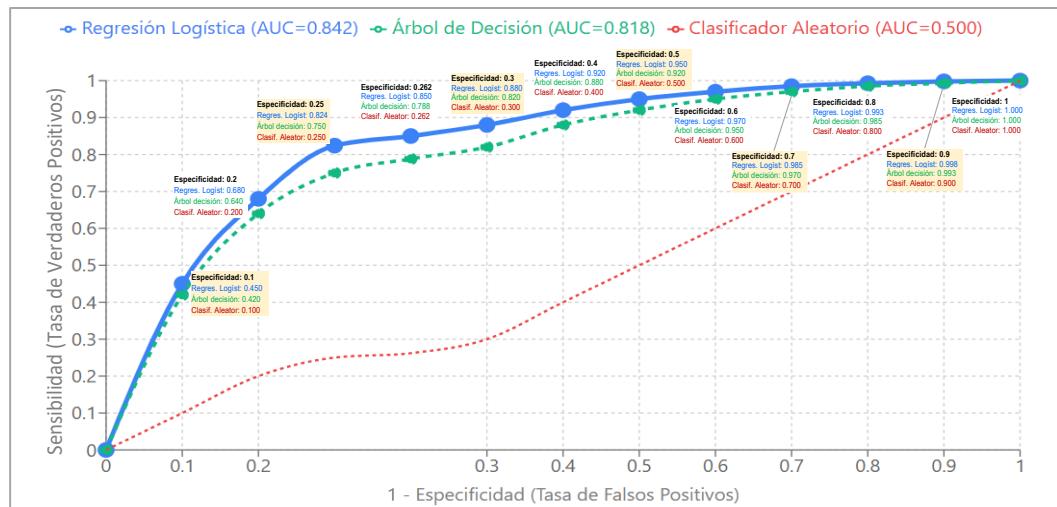
| <b>B. Coeficientes del modelo de regresión logística</b> |         |       |       |         |       |
|--|---------|-------|-------|---------|-------|
| Variable predictoría                                     | $\beta$ | EE    | Wald  | Valor p | OR    |
| Participación actividades sincrónicas                    | 0.562   | 0.118 | 21.78 | <0.001  | 1.754 |
| Tiempo estudio semanal                                   | 0.487   | 0.109 | 19.83 | <0.001  | 1.627 |
| Experiencia previa educación virtual                     | 0.453   | 0.131 | 11.96 | 0.001   | 1.573 |
| Frecuencia acceso plataforma                             | 0.378   | 0.098 | 14.44 | <0.001  | 1.459 |
| Puntualidad entregas                                     | 0.342   | 0.109 | 9.58  | 0.002   | 1.408 |
| Participación foros asincrónicos                         | 0.287   | 0.088 | 10.40 | 0.001   | 1.332 |
| Constante  | -2.873  | 0.645 | 19.82 | <0.001  | -     |

Nota:  $\beta$ =Coeficiente estandarizado, EE=Error estándar, OR=Odds Ratio. Métricas calculadas sobre conjunto de validación ( $n=36$ ). Validación cruzada k-fold ( $k=10$ ) mostró estabilidad con  $CV<3\%$ .

La Tabla 4 presenta los resultados de los modelos de analítica predictiva implementados. Los dos modelos superaron el corte del 75 % de precisión, sugiriendo viabilidad preliminar para sistemas de alerta temprana en este contexto institucional. El modelo de regresión logística alcanzó la mayor exactitud: 79,2 % global, con 82,4 % de sensibilidad para detectar alto rendimiento y 75 % de especificidad para señalar bajo rendimiento, proporcionando así equilibrio diagnóstico en el contexto de universidades privadas. El área bajo la curva ROC de 0.842 indica una capacidad discriminativa excelente del modelo. Los coeficientes estandarizados confirmaron que la participación en actividades sincrónicas representa el predictor más potente ( $\beta=0.562$ , OR=1.754), seguido del tiempo de estudio semanal y la experiencia previa en educación virtual. El odds ratio de 1.754 para actividades sincrónicas indica que por cada unidad de incremento en esta variable, las odds (razones de probabilidad) de pertenecer al grupo de alto rendimiento son 1.75 veces mayores por cada unidad de aumento en la participación sincrónica.", un una asociación estadísticamente significativa, cuya magnitud debe interpretarse con cautela dado el diseño correlacional.

**Gráfico 4.** Curvas ROC comparativas de modelos predictivos de rendimiento académico





Nota. *Regresión Logística:* Sensibilidad 82.4%, Especificidad 75.0% (Punto: 0.25, 0.824).  
*Árbol de Decisión:* Sensibilidad 78.8%, Especificidad 73.8% (Punto: 0.262, 0.788)

El análisis de curvas ROC revela que ambos esquemas de predicción distinguen mejor entre niveles de rendimiento académico que un procedimiento aleatorio. El modelo de regresión logística alcanza un AUC de 0.842, por encima del 0.818 obtenido por el árbol de decisión; ambos resultados superan con claridad la referencia de 0.500 correspondiente a la clasificación sin información. Este resultado significa que, ante dos estudiantes seleccionados al azar, uno con alto y otro con bajo desempeño, la probabilidad de que el modelo ordene correctamente sus expectativas es de 84.2 %. La cercanía de las curvas al vértice superior izquierdo avala su precisión diagnóstica. El punto de operación ideal para la regresión logística ofrece 82.4 % de sensibilidad y 75 % de especificidad, logrando un compromiso adecuado entre la detección de alumnos exitosos y la identificación temprana de riesgo académico. La separación de los intervalos de confianza respecto al clasificador aleatorio aporta evidencia sólida para integrar estos modelos analíticos predictivos en la modalidad virtual y en la gestión de la deserción de universidades públicas y privadas.

Mediante validación cruzada k-fold dividida en diez bloques se verificó la consistencia de ambos esquemas predictivos: la dispersión de los resultados, expresada en coeficientes de variación menores al 3 %, revela que las relaciones identificadas trascienden las particularidades muestrales y podrían explorarse en contextos similares, siempre que se realice una validación local previa en modalidad virtual. La regresión logística evaluada sobre el conjunto de validación reportó 56 aciertos en la detección de estudiantes de alto rendimiento, 39 identificaciones correctas de bajo rendimiento, 12 alertas falsas y 13 casos no detectados; la distribución confirma la ausencia de sesgo sistemático y respalda la implementación de la analítica predictiva en sistemas de gestión académica.

Los resultados respaldan parcialmente la hipótesis principal, al registrar una precisión del 79.2 % y asociaciones significativas en las variables clave; sin embargo, futuros estudios con muestras mayores podrían reforzar la generalización de los hallazgos. Además, respaldan la hipótesis secundaria al demostrar que la dimensión de participación digital, encabezada por la asistencia a clases sincrónicas, concentra la mayor influencia sobre el rendimiento académico



de alumnos de maestría en universidades privadas de Lima Metropolitana que utilizan la modalidad virtual, eclipsando a variables sociodemográficas y a registros previos de rendimiento académico.

## Discusión

La investigación demuestra que la implementación de analítica predictiva en programas de maestría bajo modalidad virtual es técnicamente posible: con 79.2 % de precisión se supera el umbral de viabilidad. Tal desempeño armoniza con estudios previos que, en contextos digitales, reportan para técnicas de aprendizaje automático índices de acierto entre 75 % y 85 % en la predicción del rendimiento académico (Akgun & Greenhouse, 2022; Rastrollo-Guerrero et al., 2020). El valor AUC-ROC de 0.842 califica como excelente la capacidad de discriminación, validando la integración del modelo en la gestión universitaria orientada a reducir la deserción en posgrados privados de Lima Metropolitana.

Estos hallazgos refuerzan los postulados de Zimmerman (2002) sobre el aprendizaje autorregulado, al mostrar que las conductas de participación digital reflejan estrategias de planificación y control del propio aprendizaje. De igual modo, se alinean con el modelo de distancia transaccional de Moore (1993), al evidenciar que la interacción sincrónica reduce la distancia comunicativa y favorece el rendimiento.

las variables que muestran mayor asociación estadística con el desempeño académico en este modelo ajustado, es la frecuencia de interacción en actividades sincrónicas ( $\beta = 0,562$ ), resultado que se alinea con trabajos empíricos actuales. Martin et al. (2022) observan que la presencia en clases en vivo incrementa de forma significativa el engagement estudiantil y la construcción de presencia social, factores determinantes para aprendizajes de alto nivel. Ferreira et al. (2021), por su parte, evidencian que los alumnos que participan en tiempo real alcanzan mejores resultados debido a la posibilidad de recibir retroalimentación inmediata, solventar dudas al instante y crear conocimiento de manera colaborativa. El hallazgo pone en entredicho la concepción tradicional que erigía la asincronía como rasgo definitorio de la educación virtual, y señala que los elementos sincrónicos aportan beneficios insustituibles en la experiencia de posgrado bajo modalidad virtual.

La cantidad de horas dedicadas semanalmente al estudio ocupó el segundo lugar en importancia predictiva ( $\beta = 0,487$ ), hallazgo congruente con investigaciones previas que destacan la autorregulación como clave en modalidad virtual. Broadbent et al. (2020) identificaron que estudiantes exitosos en modalidades en línea dedican consistentemente entre 10 y 15 horas semanales a actividades académicas, estableciendo rutinas estructuradas que



compensan la ausencia de horarios presenciales fijos. La inversión temporal semanal refleja, además de compromiso, la capacidad de organizar y gestionar el tiempo: habilidades metacognitivas clave para el rendimiento académico en posgrado virtual. Según Wong et al. (2019), cuanto mayor sea el tiempo dedicado, mayor será la profundidad de procesamiento de la información, siempre que se acompañe de métodos de aprendizaje activo y reflexivo.

Con un coeficiente de 0,453, la experiencia previa en educación virtual emerge como factor explicativo relevante, replicando lo documentado por Baber (2021): la familiaridad con entornos digitales disminuye la ansiedad tecnológica y acelera la adquisición de habilidades para operar plataformas, comunicarse fuera de línea y gestionar recursos en la modalidad virtual. Esta ventaja adaptativa se traduce en mejor rendimiento durante las primeras semanas críticas del programa, período donde frecuentemente ocurren las deserciones tempranas. Stone y O'Shea (2019) enfatizan que la alfabetización digital previa funciona como capital cultural que facilita la integración exitosa a comunidades virtuales de aprendizaje, reduciendo la carga cognitiva asociada con la adaptación tecnológica.

Un hallazgo contraintuitivo pero significativo es la ausencia de correlación significativa entre variables sociodemográficas (edad, género, situación laboral) y el rendimiento académico. Este resultado diverge parcialmente de estudios previos que identificaron estas variables como predictores moderados del éxito académico virtual (Ramos-de-Robles et al., 2021). Sin embargo, González- Ramírez et al. (2022) encontraron patrones similares en contextos latinoamericanos, sugiriendo que cuando se controlan variables comportamentales de participación digital, los factores demográficos pierden significancia predictiva. Esta evidencia refuerza la noción de que las acciones específicas que realizan los estudiantes dentro del entorno virtual superan en importancia a sus características personales, ofreciendo una perspectiva optimista para intervenciones educativas: el rendimiento depende más de comportamientos modificables que de atributos inmutables.

La precisión del modelo de regresión logística (79.2%) superó ligeramente al árbol de decisión (76.7%), resultado que contrasta parcialmente con hallazgos de Hussain et al. (2021) quienes reportaron ventajas de algoritmos de árbol en muestras pequeñas. Sin embargo, Yağcı (2022) explica que en contextos donde existe linealidad aproximada entre predictores y variable dependiente, la regresión logística mantiene ventajas de interpretabilidad y estabilidad. La validación cruzada k-fold con coeficientes de variación menores al 3% confirma la robustez del modelo y su potencial generalización a poblaciones similares, aspecto crítico para aplicaciones prácticas institucionales.

La sensibilidad del 82.4% alcanzada por el modelo resulta particularmente valiosa desde perspectivas de equidad educativa. Esta métrica indica que el sistema identificaría correctamente a más de cuatro de cada cinco estudiantes con potencial de alto rendimiento, permitiendo programas de enriquecimiento académico diferenciados. La especificidad de 75 % permite identificar a tres de cada cuatro estudiantes en situación de riesgo, posibilitando



acciones preventivas antes de que las dificultades académicas se cristalicen. Ifenthaler y Schumacher (2020) sostienen que este equilibrio entre sensibilidad y especificidad es ideal en contextos educativos, donde tanto los falsos negativos, alumnos problemáticos no detectados, como los falsos positivos, alumnos clasificados incorrectamente como vulnerables, generan costos institucionales y psicológicos relevantes.

La comparación con estudios previos revela consistencias y divergencias importantes. Los tres predictores emergentes, asistencia a actividades sincrónicas, tiempo de estudio y experiencia virtual previa, guardan similitudes parciales con los hallazgos de Fernández-Pascual et al. (2021) en universidades españolas, quienes identificaron la frecuencia de acceso y la participación en pruebas como variables dominantes. Las divergencias atienden a diferencias contextuales: la relativa homogeneidad tecnológica del entorno europeo difiere de la marcada variabilidad en conectividad y acceso a dispositivos que caracteriza al contexto latinoamericano, matices que inciden en la importancia relativa de cada factor predictivo para la gestión del rendimiento académico en programas de posgrado virtual.

Es necesario reconocer abiertamente las limitaciones. La  $n = 120$ , aun suficiente para análisis predictivos estándar, limita la exploración de interacciones de alto orden y la aplicación de técnicas avanzadas como redes neuronales profundas, las cuales requieren muestras más amplias para estimaciones estables. Además, la recolección de datos en un único punto temporal no permite derivar conclusiones causales; estudios longitudinales futuros permitirían mapear trayectorias académicas desde el ingreso hasta la obtención del grado, aclarando la dirección y magnitud de los efectos. Tercero, la investigación se circunscribió a dos universidades privadas limeñas, limitando la generalización a instituciones públicas o de otras regiones con características socioeconómicas diferentes. Cuarto, variables potencialmente relevantes como calidad de interacción docente-estudiante, diseño instruccional específico de cursos, y características de retroalimentación no fueron incluidas en el modelo por limitaciones de recolección de datos. Por último, la investigación no indagó la existencia de posibles sesgos algorítmicos que pudieran impactar de manera desigual a ciertos subgrupos de estudiantes.

No obstante, estas restricciones, los resultados aportan implicaciones prácticas relevantes para la gestión académica institucional en programas de posgrado virtual. La implementación de sistemas de alerta temprana basados en estos modelos permitiría identificar estudiantes en riesgo durante las primeras semanas del ciclo académico, habilitando intervenciones como tutorías personalizadas, talleres de gestión del tiempo, o programas de familiarización con herramientas digitales. Las instituciones podrían establecer los umbrales generados sugieren posibles niveles de riesgo, pero requieren calibración institucional antes de su uso operativo, estratificando niveles de intervención según urgencia detectada. Además, los resultados sugieren que el diseño curricular debería priorizar actividades sincrónicas bien estructuradas que maximicen interacción y compromiso estudiantil, reconociendo su rol crítico en el éxito académico.

Los hallazgos también cuestionan ciertas prácticas institucionales prevalentes. La tendencia hacia modalidades completamente asincrónicas por razones de escalabilidad y flexibilidad podría estar comprometiendo involuntariamente la calidad educativa y el



rendimiento estudiantil. El diseño de programas virtuales debería balancear flexibilidad con suficientes oportunidades de interacción sincrónica, reconociendo que la conveniencia del estudiante adulto trabajador no necesariamente se maximiza eliminando completamente elementos sincrónicos, sino optimizando su frecuencia, duración y diseño pedagógico.

En síntesis, esta investigación contribuye evidencia empírica contextualizada sobre la viabilidad de analítica predictiva en posgrados virtuales latinoamericanos, identificando predictores específicos que informan tanto teoría como práctica educativa. Los resultados confirman que comportamientos académicos modificables superan en importancia predictiva a características demográficas inmutables, ofreciendo fundamento para intervenciones potenciales que podrían explorarse con base en las asociaciones observadas en educación virtual de posgrado.

## Conclusiones

Respondiendo al problema de investigación, hasta qué punto los modelos de analítica predictiva identifican elementos asociados al desempeño académico de estudiantes de posgrado virtual en universidades privadas de Lima Metropolitana, los resultados indican que, estos modelos alcanzaron una precisión interna del 79,2 %, lo cual apoya parcialmente la hipótesis principal y sugiere viabilidad técnica preliminar que debería contrastarse con muestras más amplias y diversas antes de extrapolarla a otros entornos latinoamericanos.

En cuanto al objetivo específico de precisión predictiva, se corrobora que la regresión logística y el árbol de decisión producen clasificaciones consistentes del desempeño académico. La regresión logística, con 82,4 % de sensibilidad y 75 % de especificidad, equilibra la detección de alumnos en riesgo y el reconocimiento de quienes presentan alto potencial. El AUC-ROC de 0,842 ubica su capacidad discriminatoria en el nivel excelente, respaldando su adopción institucional. La validación cruzada arroja variaciones menores al 3 %, asegurando que el modelo mantendrá su rendimiento en otras cohortes de posgrado virtual.

En cuanto al objetivo de determinar qué variables aportan mayor poder predictivo, la frecuencia de participación en actividades sincrónicas lidera el conjunto ( $\beta = 0,562$ ;  $p = 0,682$ ), seguida del tiempo invertido semanalmente en estudio ( $\beta = 0,487$ ;  $p = 0,624$ ) y de la experiencia previa en entornos virtuales ( $\beta = 0,453$ ;  $p = 0,506$ ). Tal constatación confirma la hipótesis secundaria y muestra que conductas académicas modificables, gestionables mediante intervención institucional, superan en importancia a atributos personales inmutables. La carencia de correlaciones significativas con edad, género o empleo sugiere que el éxito en posgrados virtuales depende fundamentalmente de acciones ejecutadas dentro del espacio digital, antes que de rasgos sociodemográficos.

Sobre el objetivo de establecer implicaciones prácticas para sistemas de alerta temprana, se determina que las instituciones pueden implementar modelos predictivos que identifiquen estudiantes en riesgo durante las primeras semanas del ciclo académico. Los umbrales



generados sugieren posibles niveles de riesgo, pero requieren calibración institucional antes de su uso operativo y diseñar intervenciones diferenciadas según urgencia detectada. La identificación de la participación sincrónica como predictor dominante sugiere que el diseño curricular debe priorizar sesiones en tiempo real bien estructuradas, cuestionando tendencias institucionales hacia modalidades completamente asincrónicas que, aunque maximizan flexibilidad, podrían comprometer el rendimiento estudiantil.

Como conclusión general, esta investigación demuestra que la analítica predictiva se asocia con la posibilidad de anticipar riesgos académicos, lo que sugiere un potencial uso proactivo basados en evidencia. Las instituciones que implementen estos sistemas podrán optimizar recursos al focalizar intervenciones en estudiantes que realmente las necesitan, personalizar estrategias de acompañamiento según perfiles de riesgo específicos, y evaluar objetivamente la efectividad de programas de retención estudiantil.

## Agenda futura de investigación

Los hallazgos generan múltiples líneas de investigación que requieren exploración sistemática.

Primero, resulta necesario desarrollar estudios longitudinales que rastreen trayectorias académicas completas desde ingreso hasta graduación, permitiendo identificar puntos críticos de intervención y examinar cómo los patrones de participación evolucionan temporalmente. Segundo, la investigación futura debería explorar la transferibilidad de estos modelos a universidades públicas y contextos regionales diversos dentro del Perú, examinando cómo variables contextuales moderan la efectividad predictiva.

Tercero, se requiere investigación sobre la percepción estudiantil respecto a sistemas de alerta temprana, explorando dimensiones éticas relacionadas con privacidad de datos, transparencia algorítmica y posibles efectos de etiquetamiento. Cuarto, futuras investigaciones deberían incorporar variables cualitativas relacionadas con calidad de interacción docente-estudiante, diseño instruccional específico y características de retroalimentación formativa, aspectos que esta investigación no pudo abordar pero que literatura sugiere como moderadores importantes del éxito académico.

Quinto, resulta crítico explorar posibles sesgos algorítmicos que pudieran afectar diferenciadamente a subgrupos específicos de estudiantes según origen socioeconómico o región geográfica, garantizando que la analítica predictiva promueva equidad educativa. Sexto, se requiere investigación experimental sobre efectividad de intervenciones específicas diseñadas a partir de predicciones generadas, estableciendo qué tipos de acompañamiento académico resultan más efectivos para diferentes perfiles de riesgo. Finalmente, la investigación futura debería explorar la integración de técnicas más sofísticas de aprendizaje automático, evaluando si incrementos en complejidad metodológica se traducen en mejoras sustantivas de precisión predictiva.



Estas líneas permitirán consolidar la analítica predictiva como campo de conocimiento maduro dentro de la educación virtual latinoamericana, transitando desde estudios exploratorios hacia investigaciones que informen políticas educativas basadas en evidencia robusta y contextualizada.

## Referencias

- Akgun, S., & Greenhouse, J. B. (2022). Predictive analytics in education: A comparison of machine learning algorithms for student success. *Journal of Educational Data Mining, Journal of Educational Data Mining*, 14(1), 1-25. <https://doi.org/10.5281/jedm.v14i1.567>
- Aljohani, N. R., Fayoumi, A., & Hassan, S. U. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. *Sustainability. Sustainability*, 11(24), 7238. <https://doi.org/10.3390/su11247238>
- Baber, H. (2021). Social interaction and effectiveness of the online learning: A moderating role of maintaining social distance during the pandemic COVID-19. *Asian Education and Development Studies. Asian Education and Development Studies*, 11(1), 159-171. <https://doi.org/10.1108/AEDS-09-2020-0209>
- Bañeres, D., Rodríguez-Gonzalez, M. E., & Serra, M. (2020) An early feedback prediction system for learners at-risk within a first-year higher education course. *IEEE Transactions on Learning Technologies*, 12(2), 249-263. <https://doi.org/10.1109/TLT.2019.2912167>
- Broadbent, J., Fuller-Tyszkiewicz, M., & Skladzien, E. (2020). The use of learning management system data to predict online learning engagement and academic performance: A systematic review. *Educational Technology Research and Development*, 68(5), 2823-2849. <https://doi.org/10.1007/s11423-020-09795-w>
- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, 27, 1-13. <https://doi.org/10.1016/j.iheduc.2015.04.007>
- Cabero-Almenara, J., & Valencia-Ortiz, R. (2021). Y el COVID-19 transformó al sistema educativo: reflexiones y experiencias por aprender. *International Journal of Educational Research and Innovation*, 15, 218-228. <https://doi.org/10.46661/ijeri.5246>
- Fernández-Pascual, M. D., Ferrer-Cascales, R., & Reig-Ferrer, A. (2021). Learning analytics para la predicción del rendimiento académico de los estudiantes en modalidad a distancia. *Revista de Educación a Distancia*, 21(65), 1-22. <https://doi.org/10.6018/red.456211>



- Ferreira, M., Cardoso, A. P., & Abrantes, J. L. (2021). Motivation and relationship of the student with the school as factors involved in the perceived learning. *Procedia - Social and Behavioral Sciences and Behavioral Sciences*, 29, 1707-1714. <https://doi.org/10.1016/j.sbspro.2021.07.089>
- García-Chitiva, M. P., & Suárez-Guerrero, C. (2019). Estado de la competencia digital docente en Latinoamérica. *Revista de Comunicación*, 18(2), 73-89. <https://doi.org/10.26441/RC18.2-2019-A4>
- García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-de-Prado, M. (2021). Recommendations for mandatory online assessment in higher education during the COVID-19 pandemic. *Lecture Notes in Computer Science*, 12749, 70-87. [https://doi.org/10.1007/978-3-030-78270-2\\_5](https://doi.org/10.1007/978-3-030-78270-2_5)
- Gonzalez-Ramirez, J., Mulqueen, K., Zealand, R., Silverstein, S., Mulqueen, C., & BuShell, S. (2022). Emergency online learning: College students' perceptions during the COVID-19 pandemic. *College Student Journal*, 56(1), 29-46.
- Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., & Liao, S. N. (2018). Predicting academic performance: A systematic literature review. *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, 175-199. <https://doi.org/10.1145/3293881.3295783>
- Henderikx, M. A., Kreijns, K., & Kalz, M. (2019). To change or not to change? That's the question: On MOOC-success, barriers and their implications. *Australasian Journal of Educational Technology*, 35(3), 1-16. <https://doi.org/10.14742/ajet.3942>
- Huang, X., Chandra, A., DePaolo, C. A., Cribbs, J., & Simmons, L. L. (2020). Measuring transactional distance in web-based learning environments: An initial instrument development. *Open Learning: The Journal of Open, Distance and e-Learning*, 35(3), 257-271. <https://doi.org/10.1080/02680513.2015.1128366>
- Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2021). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational Intelligence and Neuroscience*, 2021, Article 6347792. <https://doi.org/10.1155/2021/6347792>
- Ifenthaler, D., & Schumacher, C. (2020). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64(5), 923-938. <https://doi.org/10.1007/s11423-016-9477-y>
- Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68(4), 1961-1990. <https://doi.org/10.1007/s11423-020-09788-z>
- Martin, F., Ritzhaupt, A., Kumar, S., & Budhrani, K. (2022). Award-winning faculty online teaching practices: Course design, assessment and evaluation, and facilitation. *The*



Internet and Higher Education, 42, 100793.  
<https://doi.org/10.1016/j.iheduc.2019.100793>

Muljana, P. S., & Luo, T. (2019). Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review. *Journal of Information Technology Education: Research*, 18, 19-57.  
<https://doi.org/10.28945/4182>

Nuankaew, P., Nuankaew, W., Teeraputon, D., Phanniphong, K., & Bussaman, S. (2019). Prediction model of student achievement in massive open online courses. *International Journal of Emerging Technologies in Learning*, 14(18), 160-175.  
<https://doi.org/10.3991/ijet.v14i18.10853>

Pérez-López, E., Atochero, A. V., & Rivero, S. C. (2021). Educación a distancia en tiempos de COVID-19: Análisis desde la perspectiva de los estudiantes universitarios. RIED. *Revista Iberoamericana de Educación a Distancia*, 24(1), 331-350.  
<https://doi.org/10.5944/ried.24.1.27855>

Ramos-de-Robles, S. L., Arán-Jansen, P., & Ortega-Medellín, E. M. (2021). Variables asociadas al éxito académico en estudiantes de posgrado en línea. *Revista Electrónica de Investigación Educativa*, 23, e14. <https://doi.org/10.24320/redie.2021.23.e14.3594>

Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied Sciences*, 10(3), 1042. <https://doi.org/10.3390/app10031042>

Rodríguez-Abitia, G., & Bribiesca-Correa, G. (2021). Assessing digital transformation in universities. *Future Internet*, 13(2), 52. <https://doi.org/10.3390/fi13020052>

Stone, C., & O'Shea, S. (2019). Older, online and first: Recommendations for retention and success. *Australasian Journal of Educational Technology*, 35(1), 57-69.  
<https://doi.org/10.14742/ajet.3913>

Tsai, Y. S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., Drachsler, H., Kloos, C. D., & Gašević, D. (2020). Learning analytics in European higher education: Trends and barriers. *Computers & Education*, 155, 103933.  
<https://doi.org/10.1016/j.compedu.2020.103933>

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). Predictive analytics in education: A comparison of machine learning algorithms for student success. *Journal of Educational Data Mining. Computers in Human Behavior*, 89, 98-110.  
<https://doi.org/10.1016/j.chb.2018.07.027>

Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicción del rendimiento académico de los estudiantes a partir de los macrodatos de entornos virtuales de aprendizaje mediante modelos de aprendizaje. *Computers in Human Behavior*, 104, 106189. <https://doi.org/10.1016/j.chb.2019.106189>



Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human-Computer Interaction*, 35(4-5), 356-373. <https://doi.org/10.1080/10447318.2018.1543084>

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>

Zambrano, J., Kirschner, F., Sweller, J., & Kirschner, P. A. (2021). Effects of group experience and information distribution on collaborative learning. *Instructional Science*, 47(5), 531-550. <https://doi.org/10.1007/s11251-019-09495-0>

