

Inteligencia artificial y analítica del aprendizaje:

Del dato a la decisión en la educación superior



**Paucar Córdova, Rosman José
Delgado Ramírez, Jorge Cristopher
Zambrano Pachay, Jorge Francisco
Sánchez Rodríguez, Alma Delia**

INTELIGENCIA ARTIFICIAL Y ANALÍTICA DEL APRENDIZAJE:

DEL DATO A LA DECISIÓN EN LA EDUCACIÓN SUPERIOR



Inteligencia artificial y analítica del aprendizaje:

Del dato a la decisión en la educación superior

Paucar Córdova, Rosman José

Delgado Ramirez, Jorge Cristopher

Zambrano Pachay Jorge Francisco

Sánchez Rodríguez, Alma Delia

Editorial GIPTEDU SAS

Website: <https://institutogiptedu.com/omp/>

Email: centrogiptedu@gmail.com

Telf. (+593)958829631

Machala, Ecuador

Primera edición, 2026

ISBN: [978-9907-809-00-8](#)

DOI: [10.66059/9789907809008](#)

Tipo de obra: *Libro académico*

Formato: *Digital*

Distribución: *Online*

© 2026 Editorial GIPTEDU SAS. Todos los derechos reservados.



Esta obra ha sido sometida a un riguroso proceso de evaluación académica bajo la modalidad de doble par ciego, con el propósito de garantizar la calidad científica y editorial de su contenido. El texto y las ideas aquí desarrolladas están protegidos por la normativa vigente en materia de propiedad intelectual. Queda estrictamente prohibida su reproducción total o parcial, distribución, comunicación pública o cualquier otra forma de utilización no autorizada, en cualquier medio o soporte electrónico, mecánico, óptico, de grabación, fotocopia u otros. Toda infracción constituirá una vulneración de los derechos exclusivos de su(s) autor(es) y de la editorial, dando lugar a las acciones legales correspondientes.

Todos los derechos reservados © 2026

Rosman José Paucar Córdova

<https://orcid.org/0000-0001-5254-4921>

Docente universitario con más de nueve años de trayectoria al servicio de la formación profesional de futuros docentes y profesionales ecuatorianos, con una sólida instrucción formal como Ingeniero en Sistemas, Máster en Tecnología Educativa y Competencias Digitales, Magíster en Interconectividad de Redes de Computadoras y Doctor en Educación. Posee amplias competencias en pedagogía, didáctica e investigación, así como en tecnología educativa, redes de computadoras, programación, inteligencia artificial y análisis de datos. Ha desempeñado cargos en instituciones públicas y privadas, ejerciendo la docencia en los niveles de educación secundaria, pregrado y posgrado, además de participar activamente en la asesoría y publicación de artículos científicos, tesis, libros y ponencias en eventos académicos nacionales e internacionales

Delgado Ramirez Jorge Cristopher

<https://orcid.org/0000-0002-0123-4031>

Docente universitario en el nivel pregrado y postgrado es un profesional del ámbito educativo con orientación académica e investigativa, enfocado en el análisis y fortalecimiento de los procesos de enseñanza–aprendizaje. Su trabajo se centra en la investigación educativa como medio para la mejora pedagógica y la toma de decisiones basadas en evidencia. Cuenta con experiencia en el diseño y desarrollo de proyectos de investigación, aplicando metodologías cualitativas y mixtas, así como en la elaboración de artículos científicos y propuestas pedagógicas innovadoras. Se caracteriza por su rigor metodológico, pensamiento crítico y compromiso con la ética académica, manteniendo una actualización constante y una visión orientada a la calidad y transformación educativa.

Zambrano-Pachay Jorge Francisco

<https://orcid.org/0000-0001-9456-2765>

Master Zambrano es un educador en el campo de la enseñanza del idioma Inglés, tecnología educativa y necesidades educativas especiales con una amplia experiencia en procesos educativos. Ha enseñado durante 15 años en diferentes instituciones y contextos educativos, incluyendo escuelas, secundarias públicas y privadas y universidades. Fue director y consultor académico en algunas instituciones públicas. En el ámbito de la investigación relacionada con la tecnología, Master Zambrano ha contribuido al diseño y ejecución de cuatro proyectos de servicios comunitarios con la sociedad, desarrollo de cursos asincrónicos para el perfeccionamiento de metodología y del idioma inglés y promotor de estrategias de aprendizaje y conductuales para la población con necesidades educativa especiales. También es un colaborador activo de la comunidad TESOL y miembro de la ETIA - Educación con Tecnología e IA (México). Ha desarrollado y prestado capacitaciones profesionales a profesores en Ecuador, México, Suiza, Colombia y los EE.UU., en áreas relacionadas con la Inteligencia Artificial, el Desarrollo de Software, la Realidad Aumentada y virtual y el currículo de aprendizaje en línea.

Alma Delia Sánchez Rodríguez

<https://orcid.org/0000-0003-3371-4600>

Docente universitario con una trayectoria de más de 12 años al servicio de la formación profesional de los futuros docentes y profesionales ecuatorianos.

Instrucción formal como Ing. en Computación e Informática, Master en Informática empresarial, Especialista En Redes De Comunicación De Datos, y Diploma Superior En Sistemas De Información Empresarial, Competencias profesionales en áreas de Tecnología educativa, Redes de computadoras, Inteligencia Artificial y Análisis de Datos.

Cargos desempeñados que incluyen el ejercicio profesional en instituciones públicas y privadas, así como en el ejercicio de la docencia en instituciones educativas de nivel pregrado y posgrado.

Participación en asesoría y publicación de artículos científicos, tesis de pre y posgrado, libros y ponencias a nivel nacional e internacional.

SINOPSIS

En la educación actualmente, ya no basta con registrar calificaciones, debido a que cada clic en el aula virtual, cada entrega, foro, actividad y evaluación dejan huellas que, interpretadas, pueden transformar la enseñanza. Este libro guía al lector “del dato a la decisión”, integrando analíticas del aprendizaje e inteligencia artificial para convertir información dispersa en acciones pedagógicas responsables.

A lo largo de los capítulos, se presentan los fundamentos de la analítica del aprendizaje, su evolución y sus tipos (descriptiva, diagnóstica, predictiva y prescriptiva) en distintos niveles (estudiante, curso e institución). Posteriormente, se explica con lenguaje claro y accesible, cómo la IA (en especial el aprendizaje automático), amplía el análisis con tutorías inteligentes, retroalimentación automatizada, plataformas adaptativas, sistemas de alerta temprana y chatbots, además de la llegada de la IA generativa y sus implicaciones para la integridad académica, la evaluación y la autonomía del estudiante.

El manuscrito pone especial énfasis en la “materia prima”, que son los datos, mismos que actualmente son conocidos como el nuevo oro por su gran valor e importancia. Se detalla, qué clases de datos se generan en educación, de dónde provienen (LMS, sistemas académicos, evaluaciones en línea, encuestas y repositorios), y por qué la calidad con respecto a la completitud, consistencia y representatividad, determinan si una visualización (dashboard) destaca o engaña.

El libro ofrece orientación práctica, pues recorre preparación y limpieza de datos, métodos de análisis e interpretación pedagógica, visualización y criterios para diseñar dashboards comprensibles y elegir herramientas, desde hojas de cálculo y BI hasta entornos de programación, cuidando la transparencia y reproducibilidad de los mismos.

El cierre del libro, se completa con modelos de clasificación y regresión, sistemas de recomendación y el diseño de indicadores y métricas alineados con competencias y evaluación formativa, y, por último, aborda desafíos éticos y organizacionales, como privacidad, transparencia, consentimiento, sesgos y gobernanza. El resultado resulta una ruta clara y aplicable para el estudio donde se deseen implementar proyectos de analítica e IA con impacto real.

Palabras clave: Analítica del aprendizaje; Inteligencia artificial educativa; Calidad de datos educativos; Dashboards y visualización; Indicadores y métricas de aprendizaje.

PSINOPSIS

In today's education, it is no longer enough to simply record grades, because every click in the virtual classroom, every submission, forum, activity, and assessment leaves traces that, when interpreted, can transform teaching. This book guides readers “from data to decision,” integrating learning analytics and artificial intelligence to turn scattered information into responsible pedagogical action.

Throughout the chapters, the foundations of learning analytics are presented, along with its evolution and its main types, descriptive, diagnostic, predictive, and prescriptive, across different levels, student, course, and institution. Next, in clear and accessible language, the book explains how artificial intelligence, especially machine learning, expands analysis through intelligent tutoring, automated feedback, adaptive platforms, early warning systems, and chatbots, as well as the emergence of generative AI and its implications for academic integrity, assessment, and student autonomy.

The manuscript places special emphasis on the “raw material,” data, now widely regarded as the new gold because of its value and importance. It details what kinds of data are produced in education, where they come from, such as learning management systems, academic information systems, online assessments, surveys, and repositories, and why data quality in terms of completeness, consistency, and representativeness determines whether a visualization, a dashboard, reveals insights or misleads.

The book offers practical guidance by covering data preparation and cleaning, analysis methods and pedagogical interpretation, visualization, and criteria for designing understandable dashboards and selecting tools, from spreadsheets and business intelligence platforms to programming environments, while safeguarding transparency and reproducibility.

The closing section is completed with classification and regression models, recommendation systems, and the design of indicators and metrics aligned with competencies and formative assessment. Finally, it addresses ethical and organizational challenges such as privacy, transparency, consent, bias, and governance. The result is a clear and applicable path for contexts where learning analytics and AI projects with real impact are intended to be implemented.

Keywords: Learning analytics; Educational artificial intelligence; Educational data quality; Dashboards and data visualization; Learning indicators and metrics.

ÍNDICE GENERAL

CAPÍTULO I: FUNDAMENTOS DE LA ANALÍTICA DEL APRENDIZAJE Y LA IA EN EDUCACIÓN	15
Introducción	15
1.1. Evolución de la analítica del aprendizaje	16
1.1.1. De la minería de datos educativos a la analítica del aprendizaje	17
1.1.2. Hitos históricos en la analítica del aprendizaje	17
1.1.3. Relación entre analítica del aprendizaje y evaluación educativa	18
1.2. Conceptos clave de IA aplicada al proceso de aprendizaje	20
1.2.1. Definiciones básicas de inteligencia artificial	20
1.2.2. Enfoques de IA relevantes para la educación	21
1.2.3. Rol de la IA en la personalización del aprendizaje	22
1.3. Tipos y niveles de analítica del aprendizaje	23
1.3.1. Analítica descriptiva, diagnóstica, predictiva y prescriptiva	23
1.3.2. Analítica a nivel de estudiante, curso e institución	24
1.3.3. Analítica del aprendizaje vs. analítica académica e institucional	25
1.4. Tendencias y desafíos actuales	26
1.4.1. Analítica del aprendizaje en entornos híbridos y virtuales	27
1.4.2. IA generativa y nuevas formas de interacción educativa	27
1.4.3. Retos para la adopción en instituciones de educación superior	28
Cierre del capítulo	29
CAPÍTULO II: INTELIGENCIA ARTIFICIAL APLICADA AL ANÁLISIS DEL PROCESO DE APRENDIZAJE	31
Introducción del capítulo	31
2.1. Fundamentos de la inteligencia artificial	32
2.1.1. Definición operativa de inteligencia artificial	32
2.1.2. Principales enfoques y técnicas de IA	32
2.1.3. Aprendizaje automático como eje de la IA actual	34
2.2. Relevancia de la IA en educación	35

2.2.1. Personalización de rutas y ritmos de aprendizaje.....	35
2.2.2. Automatización de procesos de evaluación y seguimiento.....	36
2.2.3. Apoyo a la toma de decisiones pedagógicas basada en datos	37
2.3. Aplicaciones de IA al análisis del proceso de aprendizaje.....	38
2.3.1. Sistemas de tutoría inteligente.....	39
2.3.2. Plataformas adaptativas y sistemas de alerta temprana	39
2.3.3. Chatbots educativos y asistentes conversacionales	40
2.4. Relación entre IA y analítica del aprendizaje	41
2.4.1. Diferencias conceptuales y operativas entre IA y analítica del aprendizaje	42
2.4.2. Puntos de convergencia en el uso de datos educativos.....	42
2.4.3. Escenarios de integración IA y analítica del aprendizaje.....	43
Cierre del capítulo.....	44
CAPÍTULO III: RECOPIACIÓN, TIPOS Y CALIDAD DE LOS DATOS	
EDUCATIVOS.....	46
Introducción del capítulo.....	46
3.1. Naturaleza de los datos en educación.....	47
3.1.1. Concepto de dato en el contexto educativo	48
3.1.2. Tipos de datos	49
3.1.3. Datos cuantitativos, cualitativos y mixtos	49
3.2. Estructura y dinámica de los datos educativos	50
3.2.1. Datos estructurados y no estructurados	50
3.2.2. Datos estáticos y dinámicos en el proceso educativo	51
3.2.3. Implicaciones de la estructura y dinámica para el análisis	52
3.3. Fuentes de datos en entornos educativos digitales	52
3.3.1. LMS y plataformas institucionales.....	53
3.3.2. Formularios, encuestas y registros docentes.....	53
3.3.3. Otras fuentes digitales de datos de aprendizaje	54
3.4. Calidad de los datos y problemas de recolección	55

3.4.1. Dimensiones de calidad: completitud, precisión, coherencia, actualidad y relevancia	55
3.4.2. Problemas frecuentes en la captura y almacenamiento de datos educativos	56
3.4.3. Construcción de datasets educativos confiables	56
Cierre del capítulo	58
CAPÍTULO IV: PREPARACIÓN, ANÁLISIS Y VISUALIZACIÓN DE DATOS EDUCATIVOS	
60	
Introducción del capítulo	60
4.1. Preparación y limpieza de datos.....	61
4.1.1. Tratamiento de datos perdidos (missing data)	61
4.1.2. Manejo de valores atípicos e inconsistencias	62
4.1.3. Detección y eliminación de instancias duplicadas	63
4.2. Métodos de análisis de datos educativos	65
4.2.1. Análisis descriptivo del rendimiento y la participación.....	65
4.2.2. Análisis predictivo para anticipar resultados de aprendizaje	66
4.2.3. Interpretación pedagógica de los resultados	67
4.3. Visualización de datos educativos	68
4.3.1. Selección de gráficos según tipo de variable y objetivo	68
4.3.2. Uso de gráficos de barras, histogramas, diagramas de caja y otros	69
4.3.3. Principios de diseño de visualizaciones comprensibles para docentes y gestores ..	70
4.4. Herramientas para el análisis de datos educativos	71
4.4.1. Hojas de cálculo para limpieza y análisis básico	72
4.4.2. Plataformas de BI y dashboards educativos.....	73
4.4.3. Entornos de programación con IA.....	74
Cierre del capítulo	75
CAPÍTULO V: MODELOS DE IA, INDICADORES Y MÉTRICAS PARA EL SEGUIMIENTO DEL APRENDIZAJE.....	
78	
Introducción del capítulo	78
5.1. Modelos básicos de IA aplicados a datos educativos.....	79
5.1.1. Modelos de clasificación para identificación de perfiles y riesgo	79

5.1.2. Modelos de regresión para predicción de rendimiento académico	80
5.1.3. Riesgos de sesgo y consideraciones éticas en modelos educativos	82
5.2. Sistemas de recomendación y apoyo a la decisión.....	83
5.2.1. Enfoques de recomendación en contextos educativos	84
5.2.2. Recomendación de recursos y actividades personalizadas	85
5.2.3. Interpretación de recomendaciones por parte de docentes y estudiantes	85
5.3. Diseño de indicadores de aprendizaje.....	86
5.3.1. Concepto y propiedades de calidad de los indicadores educativos	87
5.3.2. Indicadores de proceso, resultado e impacto	89
5.3.3. Alineación de indicadores con resultados de aprendizaje y competencias.....	90
5.4. Métricas, dashboards y evaluación continua.....	91
5.4.1. Métricas cuantitativas y cualitativas para evaluar el aprendizaje.....	91
5.4.2. Diseño y uso de dashboards para el seguimiento del aprendizaje.....	92
5.4.3. Automatización de reportes y análisis de tendencias para la mejora continua.....	93
Cierre del capítulo.....	93
REFERENCIAS	95

CAPÍTULO I: FUNDAMENTOS DE LA ANALÍTICA DEL APRENDIZAJE Y LA IA EN EDUCACIÓN

“Los datos constituyen siempre una representación limitada de la realidad, y los resultados de un sistema de IA dependen de los datos que utiliza.” UNICEF, 2021.

Introducción

En las dos últimas décadas, los sistemas educativos han pasado de registrar únicamente calificaciones finales a generar enormes volúmenes de datos sobre la actividad de los estudiantes en aulas físicas, entornos virtuales de aprendizaje, plataformas de evaluación en línea y redes sociales académicas. En este contexto surge la analítica del aprendizaje como campo que busca medir, recolectar, analizar e interpretar datos para comprender y optimizar el aprendizaje y los entornos donde este ocurre (Long & Siemens, 2011).

Paralelamente, la inteligencia artificial (IA) aplicada a la educación (AIED) ha desarrollado herramientas capaces de ofrecer tutoría inteligente, retroalimentación automatizada y experiencias adaptativas de aprendizaje (Holmes et al., 2019).

Comprender los fundamentos conceptuales de la analítica del aprendizaje y la IA en educación es crucial por al menos tres razones. Primero, porque las decisiones pedagógicas y de política institucional basadas en datos pueden mejorar el acompañamiento a estudiantes, pero también amplificar sesgos y desigualdades si se usan sin criterio crítico (Miao & Holmes, 2021).

Segundo, porque la rápida expansión de estas tecnologías obliga a docentes y directivos a desarrollar nuevas competencias de alfabetización de datos y de IA para interpretar dashboards, modelos predictivos y recomendaciones generadas algorítmicamente. Tercero, porque las discusiones éticas en torno a privacidad, transparencia y agencia estudiantil se han vuelto centrales para la investigación y la práctica educativa contemporáneas (Alfredo et al., 2024).

Este capítulo se organiza en cuatro secciones. La sección 1.1 recorre la evolución histórica de la analítica del aprendizaje, desde la minería de datos educativos hasta los desarrollos recientes, y

explora su relación con la evaluación educativa. La sección 1.2 introduce conceptos clave de inteligencia artificial relevantes para el análisis del proceso de aprendizaje, destacando los enfoques que han tenido mayor impacto en educación. La sección 1.3 presenta los tipos y niveles de analítica del aprendizaje, diferenciando entre analítica descriptiva, diagnóstica, predictiva y prescriptiva, y entre niveles de estudiante, curso e institución. Finalmente, la sección 1.4 aborda tendencias actuales (entornos híbridos, la IA generativa y la adopción institucional) y los principales desafíos que enfrentarán las instituciones de educación superior.

1.1. Evolución de la analítica del aprendizaje

La analítica del aprendizaje no surge en el vacío, sino como convergencia de tradiciones previas de minería de datos educativos, evaluación, psicometría y aprendizaje en línea. En sus inicios, el foco estuvo puesto en aprovechar los registros de interacción de los estudiantes con entornos virtuales para identificar patrones de éxito y abandono. Con el tiempo, el campo se ha ido desplazando desde una mirada predominantemente técnica hacia una visión más pedagógica y centrada en los actores, donde la pregunta clave ya no es solo qué se puede predecir, sino cómo se mejora el aprendizaje a partir de esa información (Gašević et al., 2015; Ferguson, 2019).

Las subsecciones que siguen describen esta evolución. Primero, se explica el tránsito desde la minería de datos educativos hacia la analítica del aprendizaje, resaltando similitudes y diferencias. Luego se presentan algunos hitos históricos que han marcado el desarrollo del campo, como la aparición de conferencias especializadas y revisiones sistemáticas recientes. Finalmente, se analiza cómo la analítica del aprendizaje se ha articulado con la evaluación educativa, especialmente con la retroalimentación formativa y la validez de las inferencias basadas en datos. En conjunto, esta sección ofrece un marco histórico que permite situar las discusiones metodológicas y éticas que se profundizarán en capítulos posteriores.

1.1.1. De la minería de datos educativos a la analítica del aprendizaje

La minería de datos educativos (Educational Data Mining, “EDM”) se consolidó a inicios de los años 2000 con el propósito de aplicar técnicas de minería de datos y aprendizaje automático a grandes bases de datos educativos, principalmente para descubrir patrones latentes (por ejemplo, secuencias de clics asociadas al abandono o estilos de resolución de problemas). La analítica del aprendizaje surge posteriormente como campo hermano, más preocupado por la traducción de esos patrones en decisiones pedagógicas y de diseño institucional, así como por la participación de múltiples actores en la interpretación de los datos (Papamitsiou & Economides, 2014).

Si bien EDM y analítica del aprendizaje comparten técnicas como: clasificación, agrupamiento y análisis secuencial, difieren en su énfasis. EDM se centra en la robustez metodológica de los modelos, mientras que la analítica del aprendizaje prioriza el alineamiento con teorías del aprendizaje, la co-creación con docentes y estudiantes y la incorporación de indicadores comprensibles para la práctica educativa (Gašević et al., 2015).

En términos prácticos, un mismo modelo predictivo de riesgo puede considerarse EDM cuando se evalúa su precisión, y analítica del aprendizaje cuando se integra en un sistema de alerta temprana que apoya tutorías, rediseño de actividades y decisiones institucionales.

1.1.2. Hitos históricos en la analítica del aprendizaje

Un hito ampliamente reconocido es la definición de la analítica del aprendizaje formulada por la Society for Learning Analytics Research (SoLAR), quien establece que la analítica del aprendizaje se entiende como el proceso de medir, recopilar, analizar y comunicar datos relativos a los estudiantes y a sus contextos, con el fin de comprender y mejorar el aprendizaje y los entornos en los que este se desarrolla. De manera que, esta definición se difundió en el marco de la primera conferencia LAK en 2011 y del trabajo de Long y Siemens (2011), donde se distinguió explícitamente entre learning analytics y academic analytics, enfatizando el foco de la primera en los procesos de aprendizaje.

Durante la década siguiente, revisiones sistemáticas y estudios de mapeo de la literatura mostraron un crecimiento exponencial de publicaciones, especialmente en educación superior. Viberg et al. (2018) analizaron 252 artículos y evidenciaron que la mayoría de las aplicaciones se centraban en sistemas de gestión del aprendizaje, predicción del rendimiento y visualización de datos para docentes y estudiantes.

Recientemente, Pan et al. (2024) revisaron intervenciones basadas en analítica del aprendizaje integradas en LMS y destacaron un aumento en estudios que vinculan analítica con rediseño instruccional y mejora de la autorregulación.

Estos hitos muestran el tránsito de un campo emergente hacia una fase de consolidación, pero aún con desafíos importantes de impacto y escalabilidad.

1.1.3. Relación entre analítica del aprendizaje y evaluación educativa

La analítica del aprendizaje se ha entrelazado estrechamente con la evaluación educativa, en particular con la evaluación formativa y las prácticas de retroalimentación. Al analizar datos de interacción, participación y desempeño, la analítica puede generar alertas tempranas, sugerencias de recursos, o perfiles de progreso que complementan las calificaciones tradicionales. Estudios recientes muestran que gran parte de las aplicaciones de analítica se orientan a enriquecer las prácticas de feedback, por ejemplo, mediante dashboards que ayudan a los docentes a identificar patrones de participación y proveer retroalimentación más específica y oportuna (Banihashem et al., 2022; Barana et al., 2019).

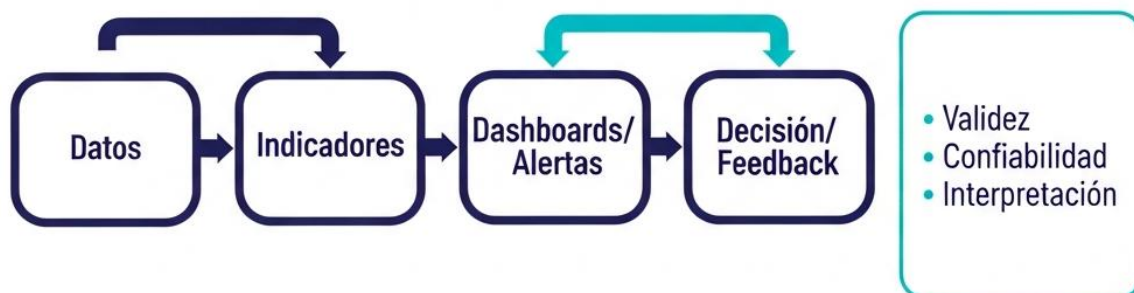
No obstante, el uso de datos para evaluación plantea preguntas críticas sobre validez y confiabilidad. A eso, Winne (2020) subraya que las trazas de actividad en entornos digitales no son indicadores neutros: requieren ser interpretadas con cautela, considerando la calidad de los datos, los modelos de inferencia y las consecuencias de las decisiones basadas en ellos.

La analítica del aprendizaje puede fortalecer la evaluación formativa al ofrecer evidencia rica y continua, pero también puede distorsionarla si se emplea de manera acrítica, reduciendo el aprendizaje a métricas fácilmente cuantificables (ver Figura. 1). De ahí la necesidad de marcos

que integren teoría del aprendizaje, diseño de evaluación y criterios de validez psicológica y pedagógica.

Figura 1.

Analítica del aprendizaje y evaluación educativa: del dato a la retroalimentación.



Nota. Diagrama que muestra cómo los datos de interacción/participación/desempeño pueden alimentar alertas, perfiles de progreso y retroalimentación, destacando la necesidad de interpretación cuidadosa por validez y consecuencias. Fuente. Elaboración propia.

La analítica del aprendizaje se ha desarrollado desde la minería de datos educativos hacia un campo interdisciplinar que combina técnicas avanzadas de análisis con preguntas sobre cómo mejorar la experiencia de estudiantes y docentes. Los hitos históricos evidencian una rápida expansión y una creciente sofisticación metodológica, a la vez que una fuerte preocupación por el impacto real en el aprendizaje.

La estrecha relación con la evaluación educativa posiciona a la analítica como un puente entre los datos y la toma de decisiones pedagógicas, pero también hace visible la importancia de discutir la validez, la ética y el sentido educativo de las métricas que se generan. Estos elementos serán fundamentales para comprender los tipos de analítica y los modelos de IA revisados en las siguientes secciones.

1.2. Conceptos clave de IA aplicada al proceso de aprendizaje

El despliegue de la analítica del aprendizaje en educación superior se apoya cada vez más en técnicas de inteligencia artificial. La IA no solo aporta algoritmos para clasificar, agrupar o predecir, sino que permite construir sistemas que toman decisiones en tiempo real, adaptan itinerarios de aprendizaje y generan retroalimentación personalizada. Para comprender este potencial, es necesario clarificar primero qué se entiende por IA, qué enfoques son más relevantes para la educación y cómo estos se traducen en experiencias concretas de personalización del aprendizaje (Russell & Norvig, 2020).

Esta sección comienza con definiciones básicas de inteligencia artificial, enfatizando la noción de agentes racionales y la diferencia entre IA general y IA específica de tareas. Luego presenta los enfoques de IA más influyentes en educación (sistemas tutoriales inteligentes, el aprendizaje automático y el procesamiento del lenguaje natural), y sus implicaciones para el análisis del proceso de aprendizaje. Finalmente, se discute el rol de la IA en la personalización del aprendizaje, mostrando tanto sus promesas como sus límites y riesgos.

1.2.1. Definiciones básicas de inteligencia artificial

Desde una perspectiva ampliamente aceptada, la inteligencia artificial puede definirse como el estudio y diseño de agentes racionales, es decir, sistemas que perciben su entorno y actúan de manera autónoma para maximizar alguna medida de desempeño o utilidad (Russell & Norvig, 2020).

En el ámbito educativo, estos agentes pueden ir desde algoritmos que recomiendan recursos hasta tutores inteligentes que adaptan problemas y pistas en función del progreso del estudiante.

Los organismos internacionales suelen adoptar definiciones complementarias. UNESCO entiende la IA como un conjunto de tecnologías capaces de procesar grandes volúmenes de datos, reconocer patrones, tomar decisiones y aprender a partir de la experiencia, con potencial para transformar procesos educativos y de gestión (Miao & Holmes, 2021).

En la práctica, la mayor parte de las aplicaciones educativas corresponden a IA estrecha o específica de tareas, como por ejemplo, sistemas de recomendación, clasificadores de riesgo o modelos de lenguaje, antes que a una IA general. Esta precisión conceptual es importante para evitar expectativas desmedidas y orientar la discusión hacia capacidades concretas, limitaciones y condiciones de uso responsable.

1.2.2. Enfoques de IA relevantes para la educación

Los enfoques de IA que más han impactado la educación pueden agruparse en tres grandes familias.

1. En primer lugar, los sistemas simbólicos o basados en reglas, que representan explícitamente conocimientos del dominio y modelos del estudiante para decidir qué contenido presentar a continuación. Este enfoque ha dado lugar a los clásicos sistemas tutoriales inteligentes, que han mostrado efectos positivos moderados sobre el rendimiento en diversas áreas (Holmes et al., 2019; Lin et al., 2023).

2. En segundo lugar, el aprendizaje automático permite entrenar modelos a partir de datos históricos para predecir rendimiento, detectar patrones de participación o recomendar recursos. Técnicas como los árboles de decisión, las redes neuronales y los modelos de aprendizaje profundo se utilizan tanto en analítica del aprendizaje como en sistemas adaptativos.

3. En tercer lugar, el procesamiento del lenguaje natural y, más recientemente, los modelos de lenguaje de gran tamaño, posibilitan chatbots educativos, generación automática de ejercicios y análisis de producciones escritas.

Este tipo de familias de técnicas se combinan cada vez más en aplicaciones híbridas, donde la IA no solo analiza datos, sino que también genera interacciones pedagógicas personalizadas (Zawacki-Richter et al., 2019).

1.2.3. Rol de la IA en la personalización del aprendizaje

La personalización del aprendizaje es uno de los argumentos más recurrentes para justificar el uso de IA en educación. En términos generales, se trata de adaptar el contenido, el ritmo, las actividades y la retroalimentación a las características, necesidades y preferencias de cada estudiante. Los sistemas de IA construyen modelos del aprendiz que integran información sobre dominio de contenidos, estrategias, motivación y contexto, y a partir de ellos toman decisiones sobre qué ofrecer a cada estudiante en un momento dado (Luckin et al., 2016).

Meta-análisis y revisiones sistemáticas muestran que los sistemas tutoriales inteligentes y plataformas adaptativas basadas en IA pueden mejorar el rendimiento y la eficiencia del aprendizaje, especialmente cuando se alinean con buenas prácticas pedagógicas y se integran en el currículo en lugar de ser añadidos de forma aislada (Holmes et al., 2019; Lin et al., 2023).

Sin embargo, la personalización también plantea desafíos: puede reforzar trayectorias desiguales si se entrenan modelos con datos sesgados, invisibilizar dimensiones no observables del aprendizaje como la creatividad o la colaboración, y reducir la agencia estudiantil si las recomendaciones se presentan como prescripciones inflexibles. De ahí la importancia de combinar IA con enfoques de diseño centrado en el humano y con oportunidades para que los estudiantes comprendan y cuestionen las decisiones algorítmicas.

Hasta este punto, se ha mostrado que la inteligencia artificial, entendida como el diseño de agentes capaces de tomar decisiones informadas y aprender de los datos, ofrece un conjunto de enfoques particularmente relevantes para la educación: sistemas tutoriales inteligentes, aprendizaje automático y procesamiento del lenguaje natural. Estos enfoques permiten modelar el progreso y las necesidades de los estudiantes y, sobre esa base, personalizar la experiencia de aprendizaje.

No obstante, la personalización basada en IA no es neutral: amplifica tanto las fortalezas como las debilidades de los modelos subyacentes y de los datos que los alimentan. Por ello, la integración de IA en analítica del aprendizaje requiere articular definiciones claras, fundamentos

pedagógicos y criterios éticos, cuestiones que reaparecerán al discutir los tipos y niveles de analítica en la siguiente sección.

1.3. Tipos y niveles de analítica del aprendizaje

La analítica del aprendizaje opera en múltiples escalas y con distintos fines. No es lo mismo generar un informe descriptivo sobre la participación en un curso que entrenar un modelo para predecir el riesgo de abandono o recomendar intervenciones específicas a un grupo de estudiantes. Además, los análisis pueden centrarse en individuos, asignaturas o la institución en su conjunto, lo que implica actores, decisiones y riesgos diferentes (Long & Siemens, 2011; Pan et al., 2024).

Esta sección aborda, en primer lugar, los tipos de analítica según el tipo de pregunta que responden: descriptiva, diagnóstica, predictiva y prescriptiva. Luego examina los niveles de aplicación (estudiante, curso e institución), ilustrando cómo cambian las métricas y decisiones en cada uno. Finalmente, se discute la distinción entre analítica del aprendizaje, analítica académica e analítica institucional, resaltando sus puntos de encuentro y tensión.

1.3.1. Analítica descriptiva, diagnóstica, predictiva y prescriptiva

De manera general, estudios en ciencia de datos como el de ciencia de datos de Ngulube (2025), diferencia a cuatro tipos de analíticas: Descriptiva (¿qué ocurrió?), Diagnóstica (¿por qué ocurrió?), Predictiva (¿qué es probable que ocurra?) y Prescriptiva (¿qué deberíamos hacer al respecto?), tal como se evidencia en la Figura. 2.

En el contexto educativo, la analítica descriptiva incluye, por ejemplo, dashboards que muestran tasas de acceso al aula virtual, entregas puntuales o distribución de calificaciones. Estos análisis permiten a docentes y coordinadores obtener una visión panorámica del curso, pero no explican necesariamente las causas subyacentes.

La analítica diagnóstica va un paso más allá al explorar relaciones entre variables: por ejemplo, cómo se relacionan patrones de participación en foros con el rendimiento final, o qué secuencias de recursos se asocian a mejores aprendizajes (Papamitsiou & Economides, 2014).

La analítica predictiva emplea modelos como regresiones o redes neuronales para estimar la probabilidad de que un estudiante abandone o apruebe, generando alertas tempranas.

Figura 2.

Comparación de analítica descriptiva, diagnóstica, predictiva y prescriptiva en educación.

Tipo de analítica	Pregunta	Ejemplo en educación
 Descriptiva	 ¿qué ocurrió?	 dashboards: accesos, entregas, calificaciones
 Diagnóstica	 ¿por qué ocurrió?	 relaciones: foros y rendimiento; secuencias de recursos
 Predictiva	 ¿qué es probable que ocurra?	 modelos: abandono/ aprobación; alertas tempranas
 Prescriptivos	 ¿qué deberíamos hacer?	 acciones: refuerzo, cambio de examen, priorizar tutorías

Nota. Tabla comparativa que resume de los tipos de analíticas, características y funciones.

Finalmente, la analítica prescriptiva propone acciones concretas: recomendar actividades de refuerzo, cambiar la fecha de un examen o priorizar tutorías para ciertos grupos. Pan et al. (2024) muestran que la mayoría de trabajos siguen concentrándose en la analítica descriptiva y predictiva, mientras que los enfoques prescriptivos, más complejos, están en desarrollo.

1.3.2. Analítica a nivel de estudiante, curso e institución

La analítica del aprendizaje puede situarse en distintos niveles de agregación. A nivel de estudiante, se enfatizan los perfiles individuales: trayectorias de interacción, progresos en competencias, patrones de autorregulación. Ejemplos típicos son los dashboards personales que muestran a cada estudiante su avance respecto a los resultados de aprendizaje y ofrecen recomendaciones de estudio.

A nivel de curso, el foco se desplaza hacia el diseño instruccional: distribución de actividades, alineamiento con resultados de aprendizaje, y evaluación de qué recursos o estrategias se asocian

con mejores desempeños. Estudios contextualizados han sistematizado el uso de analítica para la evaluación de cursos, evidenciando su potencial para informar decisiones de rediseño (Wong et al., 2025).

En el nivel institucional, la analítica se aproxima a lo que se denomina “academic analytics”, mismas que son modelos que combinan datos académicos, administrativos y demográficos para apoyar la planificación, la retención estudiantil y la asignación de recursos (Long & Siemens, 2011). Aquí, la analítica suele integrarse en sistemas de gestión estratégica y paneles para directivos, sin embargo, se advierte que las decisiones tomadas en este nivel pueden tener fuertes implicaciones para la equidad y la autonomía de las unidades académicas, por lo que se requieren marcos de gobernanza de datos y participación de múltiples actores (Márquez et al., 2024).

1.3.3. Analítica del aprendizaje vs. analítica académica e institucional

Long y Siemens (2011) propusieron una distinción influyente entre learning analytics y academic analytics. Mientras la primera se centra en el proceso de aprendizaje, incluyendo la interacción entre estudiantes, contenido y docentes, la segunda se orienta a la mejora de procesos y resultados organizacionales, como la eficiencia, la acreditación o la competitividad institucional.

En la práctica, esta distinción no es absoluta, ya que los mismos datos pueden alimentar tanto intervenciones pedagógicas como decisiones estratégicas. Reyes (2015) resume esta relación señalando que la analítica del aprendizaje se ubica en el cruce entre “big data” educativo y pedagogía, mientras que la analítica académica se aproxima más a las lógicas de business intelligence aplicadas a la educación superior.

Pan et al. (2024) y Viberg et al. (2018) muestran que muchos proyectos de analítica del aprendizaje comienzan como pilotos centrados en cursos o programas específicos y, si tienen éxito, escalan hacia estrategias institucionales más amplias.

Esta traslación plantea preguntas críticas:

- ¿Cómo asegurar que la ampliación del uso de datos no desplace la atención desde el aprendizaje hacia indicadores puramente administrativos?

- ¿Qué mecanismos de participación y transparencia se requieren para equilibrar intereses pedagógicos e institucionales?

Es así, que se ha argumentado como la analítica del aprendizaje puede entenderse según el tipo de pregunta que formula (descriptiva, diagnóstica, predictiva o prescriptiva) y el nivel en que opera (estudiante, curso, institución). Esta doble clasificación ayuda a clarificar los objetivos de cada proyecto analítico, los actores involucrados y los riesgos asociados.

Asimismo, la distinción entre analítica del aprendizaje y analítica académica resalta la tensión permanente entre la mejora del aprendizaje y las metas organizacionales. De modo que, más que campos separados, se trata de dimensiones que deben articularse cuidadosamente para que las decisiones guiadas por datos mantengan el foco en el aprendizaje, propósito central de las instituciones educativas.

1.4. Tendencias y desafíos actuales

En los últimos años, la analítica del aprendizaje y la IA en educación se han visto transformadas por tres grandes dinámicas: la expansión de los entornos híbridos y virtuales, la irrupción de la IA generativa y la creciente preocupación por los retos de adopción institucional y éticos. La pandemia de COVID-19 aceleró la migración hacia modalidades en línea y mixtas, generando nuevas fuentes de datos y necesidades de seguimiento.

Al mismo tiempo, el surgimiento de modelos de lenguaje como ChatGPT ha abierto posibilidades inéditas para la interacción educativa, la generación de contenidos y la automatización de retroalimentación, pero también controversias sobre plagio, dependencia y desinformación (Kasneci et al., 2023).

Esta sección examina, en primer lugar, el papel de la analítica del aprendizaje en entornos híbridos y virtuales. Luego analiza las oportunidades y riesgos de la IA generativa en la educación superior. Finalmente, discute los retos que enfrentan las instituciones para adoptar estas tecnologías de forma sostenible y centrada en las personas, a la luz de revisiones recientes sobre adopción y marcos humanocéntricos (Márquez et al., 2024; Alfredo et al., 2024).

1.4.1. Analítica del aprendizaje en entornos híbridos y virtuales

Los entornos híbridos y virtuales, en los que los estudiantes alternan entre actividades presenciales, síncronas en línea y asincrónicas en plataformas digitales generan trazas ricas de datos que pueden alimentar la analítica del aprendizaje. Revisiones sistemáticas recientes muestran que la mayoría de las implementaciones se han centrado en sistemas de gestión del aprendizaje (LMS), donde se monitorizan accesos, envío de tareas, visualización de recursos y participación en foros (Viberg et al., 2018; Pan et al., 2024).

En estos contextos, la analítica se utiliza para detectar estudiantes en riesgo, ajustar la carga de trabajo y apoyar la autorregulación mediante dashboards, sin embargo, los entornos híbridos plantean desafíos adicionales. Primero, combinan datos de múltiples fuentes (aulas físicas, videoconferencias, herramientas colaborativas externas, etc.), las cuales no siempre son interoperables ni de calidad homogénea. Segundo, las interpretaciones de los datos pueden ser más complejas, porque el no acceso a un recurso digital no implica necesariamente falta de estudio si el estudiante recurre a materiales alternativos.

Investigaciones en educación superior latinoamericanas también han señalado brechas de infraestructura y conectividad que condicionan el uso de analítica y pueden invisibilizar a los estudiantes con menor acceso (Quinteros et al., 2022). Por ello, el diseño de analítica en entornos híbridos debe considerar tanto la integración técnica como las desigualdades socio-tecnológicas.

1.4.2. IA generativa y nuevas formas de interacción educativa

La IA generativa, especialmente los modelos de lenguaje de gran tamaño, está reconfigurando rápidamente las formas de interacción educativa. Herramientas como los chatbots conversacionales pueden responder preguntas, explicar conceptos, sugerir ejemplos y ofrecer retroalimentación preliminar sobre textos, lo que abre posibilidades para el apoyo fuera del aula y la práctica autónoma. Kasneci et al. (2023) argumentan que estos modelos ofrecen oportunidades significativas para promover la personalización, la práctica guiada y el aprendizaje de idiomas, pero advierten que su uso debe acompañarse de alfabetización crítica y supervisión docente.

La UNESCO ha publicado orientaciones específicas sobre IA (incluida la IA generativa) en educación e investigación, enfatizando la necesidad de marcos éticos, transparencia algorítmica y equidad en el acceso (Miao & Holmes, 2021).

Desde la perspectiva de la analítica del aprendizaje, la IA generativa introduce nuevos tipos de datos (por ejemplo, diálogos estudiante–agente) y plantea retos para distinguir entre producción humana y asistida por IA en actividades de evaluación. Además, existe el riesgo de que la delegación de tareas cognitivas a sistemas generativos reduzca el desarrollo de habilidades fundamentales si no se diseñan actividades que promuevan reflexión, contraste de fuentes y construcción activa del conocimiento.

En consecuencia, la integración de IA generativa requiere rediseñar tareas, criterios de evaluación y mecanismos de acompañamiento para evitar relaciones de dependencia acrítica con las herramientas.

1.4.3. Retos para la adopción en instituciones de educación superior

La adopción de analítica del aprendizaje y IA en instituciones de educación superior no es solo un problema técnico, sino principalmente organizacional, cultural y ético. La revisión sistemática de Márquez et al. (2024) identifica factores clave que influyen en la adopción: liderazgo institucional, claridad de propósito, infraestructura tecnológica, competencias de los actores, marcos de política y consideraciones éticas sobre privacidad y consentimiento.

Sin una visión estratégica compartida, muchos proyectos de analítica quedan atrapados en fases piloto y no logran escalamiento sostenible. En términos de política, Tsai y Gašević (2017) muestran que las instituciones que avanzan en la implementación de analítica suelen desarrollar políticas explícitas que abordan gobernanza de datos, roles y responsabilidades, formación de usuarios y evaluación de impacto.

Más recientemente, Alfredo et al. (2024) proponen un enfoque de analítica e IA centradas en lo humano, que implica involucrar activamente a estudiantes y docentes en el diseño de

herramientas, priorizar la transparencia, y evaluar los efectos sobre la agencia y el bienestar de los distintos actores.

Entre los retos persistentes destacan la resistencia al cambio, la sobrecarga de trabajo docente, la falta de alfabetización de datos e IA, y la necesidad de asegurar que los beneficios de estas tecnologías se distribuyan de manera equitativa entre programas y grupos estudiantiles.

Las tendencias actuales muestran un campo de analítica del aprendizaje y IA en educación en plena expansión, pero atravesado por tensiones significativas. Los entornos híbridos y virtuales ofrecen nuevas oportunidades para monitorizar y apoyar el aprendizaje, a costa de una mayor complejidad técnica y de interpretación. La IA generativa introduce formas potentes de interacción y apoyo, pero también riesgos para la integridad académica y la autonomía cognitiva de los estudiantes.

Finalmente, los procesos de adopción institucional evidencian que la sostenibilidad de estas iniciativas depende de marcos de gobernanza, liderazgo y participación que sitúen a las personas, y no solo a los datos o a los algoritmos en el centro de las decisiones.

Cierre del capítulo

Este primer capítulo ha presentado los fundamentos de la analítica del aprendizaje y la inteligencia artificial en educación, ofreciendo un marco desde el cual comprender los desarrollos posteriores. En primer lugar, se revisó la evolución histórica de la analítica del aprendizaje, desde la minería de datos educativos hasta las actuales iniciativas de analítica e IA centradas en el humano. Esta trayectoria ha mostrado un desplazamiento desde el énfasis en la capacidad técnica de los modelos hacia preguntas sobre su alineamiento con el aprendizaje, la evaluación formativa y la equidad (Gašević et al., 2015; Ferguson, 2019).

En segundo lugar, se introdujeron conceptos clave de inteligencia artificial relevantes para la educación, destacando los enfoques que permiten construir sistemas adaptativos, analizar grandes volúmenes de datos educativos y generar nuevas formas de interacción, incluidas las basadas en IA generativa (Holmes et al., 2019; Kasneci et al., 2023).

En tercer lugar, se discutieron los tipos y niveles de analítica del aprendizaje, lo que permite articular con mayor precisión los objetivos de cada proyecto, los actores involucrados y las decisiones pedagógicas o institucionales que se pretende apoyar. Finalmente, se analizaron tendencias y desafíos actuales, evidenciando que la expansión de la analítica y la IA en educación superior está condicionada por factores de adopción, marcos de política y consideraciones éticas sobre privacidad, transparencia y agencia estudiantil (Márquez et al., 2024; Alfredo et al., 2024).

Los próximos capítulos profundizarán en aspectos más específicos como lo son: las fuentes y calidad de los datos educativos, las técnicas de análisis y modelado, el diseño de dashboards y sistemas de apoyo a la decisión, así como los modelos de IA aplicados a la predicción, recomendación y generación de contenido.

Sobre la base conceptual presentada aquí, se buscará no solo describir herramientas y métodos, sino también desarrollar una mirada crítica y situada sobre su uso en contextos reales de enseñanza y aprendizaje. El objetivo final es que los lectores (estudiantes y maestros de pre y pos grado, docentes de primaria y secundaria, investigadores y profesionales de la educación, etc.) cuenten con criterios sólidos para participar activamente en el diseño, implementación y evaluación de proyectos de analítica del aprendizaje e inteligencia artificial en sus propias instituciones.

CAPÍTULO II: INTELIGENCIA ARTIFICIAL APLICADA AL ANÁLISIS DEL PROCESO DE APRENDIZAJE

“El aprendizaje automático es el campo de estudio que da a los ordenadores la capacidad de aprender sin ser programados de manera explícita.” Arthur Samuel.

Introducción del capítulo

El primer capítulo de este libro presentó la analítica del aprendizaje como un campo que aprovecha datos educativos para comprender y mejorar los procesos formativos. En este segundo capítulo el foco se desplaza hacia la inteligencia artificial (IA) como motor computacional que permite automatizar, ampliar y sofisticar ese análisis, especialmente cuando el volumen, la variedad y la velocidad de los datos superan las capacidades humanas o de la estadística tradicional.

Comprender los fundamentos de la IA no es, por tanto, un lujo técnico, sino una condición para que docentes, diseñadores instruccionales y gestores educativos puedan participar críticamente en su diseño, adopción y evaluación. De modo que el objetivo central de este capítulo es explicar cómo la IA se aplica al análisis del proceso de aprendizaje, conectando conceptos técnicos básicos con situaciones educativas reales (tutorías automatizadas, plataformas adaptativas, sistemas de alerta temprana y chatbots conversacionales).

Primero, en la sección 2.1 se presentan los fundamentos de la IA: qué se entiende operativamente por IA, cuáles son sus principales enfoques y por qué el aprendizaje automático se ha convertido en el eje de la IA actual. La sección 2.2 analiza la relevancia específica de la IA en educación, especialmente en la personalización de rutas y ritmos, la automatización de la evaluación y el seguimiento, y el apoyo a la toma de decisiones pedagógicas basadas en datos.

Por su parte, la sección 2.3 se examinan aplicaciones concretas orientadas al análisis del proceso de aprendizaje: sistemas de tutoría inteligente, plataformas adaptativas con sistemas de alerta temprana y chatbots educativos. Y finalmente, la sección 2.4 profundiza en la relación entre IA y analítica del aprendizaje, clarificando diferencias conceptuales, puntos de convergencia en el uso de datos y escenarios de integración.

En conjunto, el capítulo busca ofrecer un lenguaje compartido entre especialistas en tecnología y profesionales de la educación, de modo que puedan interpretar con criterio las promesas y límites de la IA cuando se utiliza para observar, modelar y acompañar el aprendizaje humano.

2.1. Fundamentos de la inteligencia artificial

Esta sección introduce los elementos mínimos de la IA que un lector del ámbito educativo necesita para comprender los capítulos posteriores, no se pretende formar expertos técnicos, sino aclarar qué significa hablar de IA en términos operativos, qué enfoques y técnicas son más relevantes y por qué el aprendizaje automático se ha vuelto central en la práctica contemporánea.

Las subsecciones abordan, respectivamente, una definición de trabajo de IA, un panorama sintético de técnicas y la posición del aprendizaje automático como eje de la IA actual.

2.1.1. Definición operativa de inteligencia artificial

En términos operativos, la IA puede entenderse como el conjunto de métodos computacionales que permiten a sistemas digitales percibir, razonar, aprender y actuar sobre la base de datos, con distintos grados de autonomía (Russell & Norvig, 2020). Desde una perspectiva educativa, esta definición se traduce en sistemas capaces de interpretar evidencias del desempeño estudiantil y generar recomendaciones, retroalimentación o decisiones de forma parcialmente automatizada.

Un sistema que analiza patrones de intentos en cuestionarios y sugiere actividades de refuerzo está “actuando inteligentemente” según esta definición. En el ámbito de la política educativa, organismos como la UNESCO insisten en que la IA debe entenderse no solo como una tecnología, sino como un ecosistema socio-técnico que cambia currículos, formas de evaluación, formación docente y equidad en el acceso (Miao et al., 2021).

Esta perspectiva, permite que la definición operativa incluya dimensiones éticas y de gobernanza: no basta con que el sistema sea técnicamente inteligente, debe ser también pedagógicamente significativo y socialmente responsable.

2.1.2. Principales enfoques y técnicas de IA

Los enfoques de IA pueden agruparse, de forma simplificada, en tres familias:

1. Simbólica, basada en reglas explícitas y conocimiento representado en forma lógica.
2. Aprendizaje automático, que induce patrones a partir de datos.
3. Híbrida, que combina ambos enfoques (Russell & Norvig, 2020).

En educación, la IA simbólica se ha utilizado en sistemas expertos y ontologías curriculares, mientras que el aprendizaje automático domina hoy las tareas de predicción de deserción, recomendación de recursos y clasificación automática de producciones estudiantiles (Wang et al., 2024).

En la práctica, muchas aplicaciones educativas integran varias técnicas como: sistemas de recomendación en un LMS pueden usar filtrado colaborativo, mientras un sistema de corrección automática de preguntas abiertas recurre a modelos de lenguaje y redes neuronales, a los cuales se los representa de manera general en la subsiguiente imagen.

Figura 3.

Enfoques de IA en educación: simbólica, aprendizaje automático e híbrida.



Nota. Presenta las tres familias de enfoques (simbólica, aprendizaje automático, híbrida) y refuerza el papel del aprendizaje automático como eje/puente en la IA actual.

Por su parte, los sistemas de tutoría inteligente combinan modelos de dominio simbólicos con algoritmos de aprendizaje para adaptar problemas y pistas al progreso del estudiante (Lin et al.,

2023). Entender esta diversidad ayuda a los educadores a distinguir entre soluciones “opacas” basadas en modelos complejos y aquellas más interpretables, cuestión clave para la transparencia y la confianza pedagógica.

2.1.3. Aprendizaje automático como eje de la IA actual

El aprendizaje automático (Machine Learning), actualmente se ha convertido en el eje de la IA contemporánea porque desplaza el foco desde programar reglas a aprender directamente de los datos. En el contexto educativo esto implica, por ejemplo, entrenar modelos que predicen el riesgo de abandono a partir de la actividad en la plataforma, o modelos que estiman el dominio de un concepto según las respuestas del estudiante en diferentes tareas (Wang et al., 2024).

En los últimos años, los estudios sistemáticos muestran que la gran mayoría de aplicaciones de IA en educación utilizan variantes de *aprendizaje supervisado*, *no supervisado* o *profundo*, y que muchas se apoyan en datos recogidos por sistemas de gestión del aprendizaje, plataformas adaptativas o tutores inteligentes (Garzón et al., 2025; Lin et al., 2023).

Desde la perspectiva anterior, estos modelos no solo permiten automatizar tareas, sino también construir representaciones dinámicas del proceso de aprendizaje, que luego sirven de base para intervenciones de analítica del aprendizaje más sofisticadas. De allí que, en este libro, el aprendizaje automático se aborde como puente técnico entre los datos educativos y las decisiones pedagógicas mediadas por IA.

A manera de resumen, esta sección ha presentado la IA como un conjunto de métodos que permiten a los sistemas educativos “leer” y responder a los datos de maneras cada vez más complejas. La distinción entre enfoques simbólicos, de aprendizaje automático e híbridos ofrece un marco para interpretar la diversidad de aplicaciones descritas en la literatura reciente.

Reconocer el papel central del aprendizaje automático prepara el terreno para comprender, en las secciones siguientes, cómo estas técnicas se concretan en personalización, evaluación automatizada y apoyo a la toma de decisiones pedagógicas.

2.2. Relevancia de la IA en educación

Esta sección sitúa la IA en el corazón de las preocupaciones pedagógicas: cómo aprender mejor, con quién, a qué ritmo y con qué apoyos. La IA no se limita a automatizar procesos administrativos; interviene directamente en la organización de rutas de aprendizaje, en la evaluación continua y en la toma de decisiones docentes. Las subsecciones siguientes, permiten desarrollar tres dimensiones claves: personalización, automatización de evaluación y seguimiento, y apoyo a decisiones basadas en datos.

2.2.1. Personalización de rutas y ritmos de aprendizaje

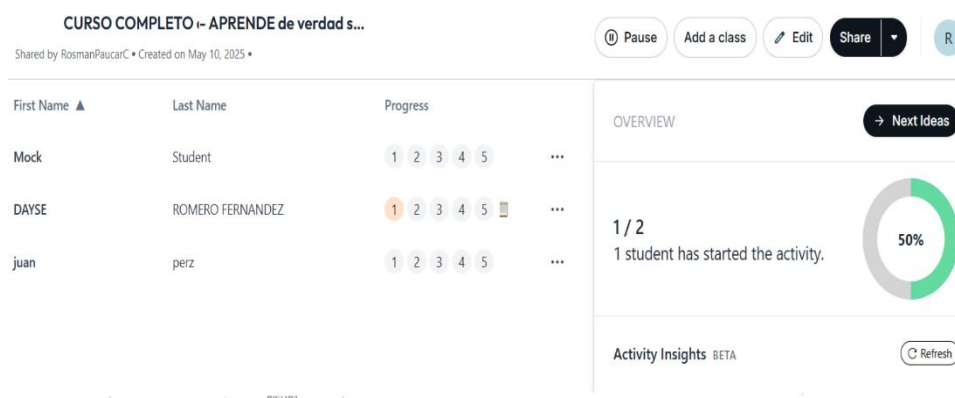
La personalización ha sido uno de los argumentos más reiterados para justificar el uso de IA en educación. En lugar de ofrecer la misma secuencia de contenidos a todo el grupo, los sistemas impulsados por IA ajustan actividades, recursos y niveles de dificultad según el desempeño y las preferencias del estudiante.

Estudios recientes, muestran que las plataformas y tutores inteligentes que combinan modelos de estudiante con algoritmos de recomendación logran mejoras en logro, retención y compromiso, especialmente cuando se integran con buenas prácticas didácticas (Hariyanto et al., 2025; Vorobyeva et al., 2025).

En entornos concretos, esto se traduce en indicadores donde el estudiante visualiza su progreso y recibe sugerencias como “repasar problemas de razonamiento algebraico” o “ver un video breve antes de intentar el cuestionario”. Para el claustro docente, la IA puede ayudar a detectar subgrupos con necesidades similares, facilitando estrategias de diferenciación pedagógica. Dichos procesos se ven representados gráficamente en herramientas web con trabajo en vivo como el de la Figura 4.

Figura 4.

Proceso y sugerencias de aprendizaje con herramientas de IA automático e híbrido.



Nota. Presenta el número y progreso en tiempo real de estudiantes al momento del desarrollo de actividades mediante extensión de IA app.briskteaching en Google Chrome. Fuente: https://app.briskteaching.com/teacher_activity_monitor?id=2911793.

En este aspecto, se advierte que la personalización no es solo técnica: requiere marcos pedagógicos claros para evitar que el modelo de IA reduzca el aprendizaje a repetir patrones pasados y limite oportunidades de desafío y exploración (Garzón et al., 2025). Un claro ejemplo, se observa en la figura anterior, donde se observa como los docentes pueden observar en tiempo real la participación de los educandos.

2.2.2. Automatización de procesos de evaluación y seguimiento

La IA también ha sido utilizada para automatizar componentes de la evaluación, desde la corrección de pruebas objetivas hasta la retroalimentación sobre producciones abiertas. Una revisión reciente muestra que los sistemas de evaluación basados en IA pueden reducir carga de trabajo docente, acelerar la entrega de retroalimentación y aumentar la frecuencia de evaluaciones formativas, siempre que se combinen con revisión humana (Mpolomoka, 2025).

En educación superior, por ejemplo, se han implementado modelos que clasifican respuestas cortas según rúbricas, detectan plagio complejo mediante análisis semántico o monitorizan patrones de participación en foros para valorar trabajo colaborativo. Estas herramientas no solo generan calificaciones, sino datos ricos sobre el proceso, como el tiempo dedicado, los intentos previos o las estrategias de revisión.

Dichos datos alimentan sistemas de analítica del aprendizaje que permiten identificar estudiantes que necesitan apoyo adicional o ajustar el diseño de actividades (Ngulube & Ncube, 2025). La evidencia también subraya riesgos: sesgos en los datos de entrenamiento, opacidad de los criterios de corrección y posible dependencia excesiva de la automatización, lo que exige mantener siempre un rol supervisado del docente.

2.2.3. Apoyo a la toma de decisiones pedagógicas basada en datos

Más allá de personalizar y evaluar, la IA puede actuar como sistema de apoyo a la decisión para docentes y gestores. Integrada con analítica del aprendizaje, permite convertir datos dispersos (accesos, entregas, resultados, interacción) en indicadores sintetizados y recomendaciones accionables. Los estudios sobre IA en educación muestran que, cuando los docentes reciben alertas interpretables y contextualizadas, son más propensos a intervenir tempranamente con estudiantes en riesgo y a rediseñar actividades poco efectivas (Alfredo et al., 2024; Weidlich et al., 2025).

Por ejemplo, un panel docente puede indicar que un conjunto de estudiantes combina baja participación en foros con intentos repetidos fallidos en un tipo específico de problema, sugiriendo revisar la explicación inicial o proponer actividades guiadas adicionales. A nivel institucional, las recomendaciones pueden orientar decisiones sobre apoyo académico, rediseño de asignaturas o desarrollo profesional docente (Miao et al., 2021; Garzón et al., 2025).

La clave es que la IA no reemplaza el juicio pedagógico, sino que lo amplifica, siempre que las recomendaciones sean transparentes, negociables y alineadas con propósitos formativos.

En conjunto, la IA en educación adquiere relevancia cuando se articula con objetivos pedagógicos claros: apoyar itinerarios personalizados, enriquecer la evaluación formativa y ofrecer información procesable para la toma de decisiones. La evidencia disponible sugiere efectos positivos, pero también insiste en que la calidad de los resultados depende tanto del diseño algorítmico como de las condiciones institucionales y didácticas en que se despliegan estas herramientas.

2.3. Aplicaciones de IA al análisis del proceso de aprendizaje

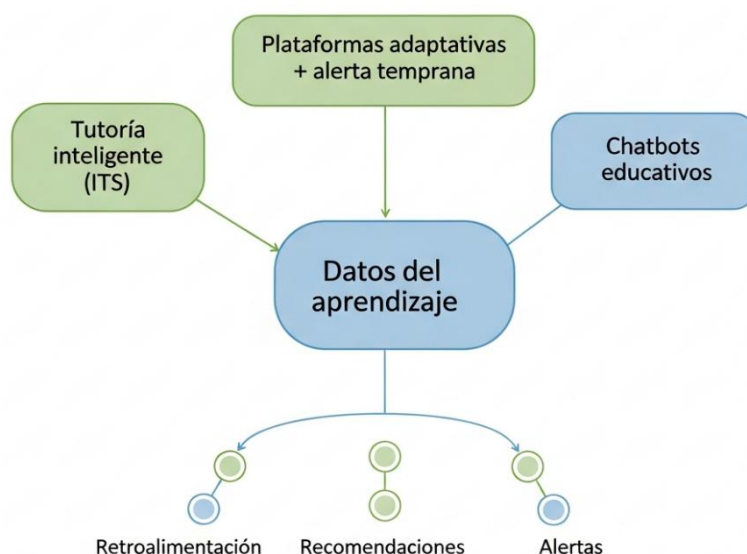
Esta sección desciende al nivel de las aplicaciones concretas que permiten observar y modelar el proceso de aprendizaje en tiempo real. Se abordan tres familias de sistemas que combinan IA y analítica del aprendizaje:

- ✓ **Sistemas de tutoría inteligente:** programas que actúan como un “tutor digital”: detectan en qué te equivocas o qué te falta aprender y te dan explicaciones, ejercicios y retroalimentación personalizada para mejorar.
- ✓ **Plataformas adaptativas con alerta temprana:** plataformas de aprendizaje que ajustan el contenido y la dificultad según tu desempeño, y además avisan a tiempo (al estudiante o al docente) cuando detectan señales de riesgo, como baja participación, caídas en notas o tareas atrasadas.
- ✓ **Chatbots educativos:** asistentes conversacionales (chat) que responden preguntas, explican temas, guían actividades, recuerdan tareas o dan apoyo básico de estudio, de forma inmediata y en lenguaje sencillo.

En todos los casos, el énfasis recae en cómo estos sistemas recogen, interpretan y utilizan datos para apoyar la actividad de estudiantes y docentes.

Figura 5.

Aplicaciones de IA para observar y acompañar el aprendizaje.



Nota. Mapa que conecta tutoría inteligente, plataformas adaptativas con alerta temprana y chatbots con el uso de datos para apoyar actividad de estudiantes y docentes.

En la figura anterior, se ve gráficamente como el aprendizaje hace el empleo de aplicación de IA para por ejemplo personalizar el aprendizaje, guiarlo y fortalecerlo. De manera seguida, se desarrollaran estos puntos de manera más amplia, junto a estos tipos de IA.

2.3.1. Sistemas de tutoría inteligente

Los sistemas de tutoría inteligente (Intelligent Tutoring Systems, “ITS”) son quizá la tradición más consolidada de IA en educación. Estos sistemas modelan el conocimiento del dominio, el estado del estudiante y las estrategias didácticas, para ofrecer problemas, pistas y explicaciones adaptadas en tiempo casi real. Los ITS bien diseñados pueden mejorar el rendimiento y favorecer estrategias de autorregulación, especialmente en matemáticas, ciencias y programación (Lin et al., 2023).

Desde la perspectiva del análisis del proceso de aprendizaje, los ITS generan un registro granular de interacciones como pasos intermedios en la resolución de problemas, tiempos de respuesta, errores recurrentes, uso de pistas. Estos datos permiten construir modelos de dominio más precisos y detectar patrones de aprendizaje productivo o superficial.

Además, los ITS emergentes integran técnicas de aprendizaje profundo y procesamiento del lenguaje natural para analizar respuestas abiertas y diálogos tutoriales, lo que amplía las posibilidades de proporcionar retroalimentación cualitativa y no solo correctiva (Garzón et al., 2025).

2.3.2. Plataformas adaptativas y sistemas de alerta temprana

Las plataformas adaptativas integran algoritmos de IA que ajustan el contenido, el ritmo y la secuencia de actividades según el desempeño continuo del estudiante. En paralelo, los sistemas de alerta temprana emplean modelos predictivos para identificar estudiantes con riesgo de abandono, fracaso o desenganche antes de que la situación sea irreversible. Los estudios en estas tecnologías muestran que los enfoques que combinan análisis de trazas de actividad (clics,

tiempos, entregas) con indicadores de contexto (historial académico, perfil) pueden predecir con razonable precisión quiénes requieren intervención (Hariyanto et al., 2025; Ngulube & Ncube, 2025).

En un curso en línea, por ejemplo, un modelo entrenado con cohortes anteriores puede detectar que un patrón de “ingreso tardío al aula virtual + no completar las primeras actividades” se asocia con alto riesgo de abandono. El sistema puede entonces enviar mensajes personalizados al estudiante y alertas al docente, integrando la predicción con recomendaciones concretas (reprogramar plazos, ofrecer tutoría, simplificar instrucciones).

Estas soluciones se consideran particularmente valiosas en programas con alta masificación, pero su efectividad depende de la calidad de los datos, del diseño de las intervenciones y del respeto a principios éticos como la no estigmatización y la transparencia en el uso de los modelos (Alfredo et al., 2024).

2.3.3. Chatbots educativos y asistentes conversacionales

Los chatbots educativos y asistentes conversacionales, basados cada vez más en grandes modelos de lenguaje, se han difundido rápidamente como recursos para responder dudas, explicar conceptos, ofrecer ejemplos o acompañar el trabajo autónomo. Una revisión sistemática reciente muestra que estos sistemas mejoran especialmente el acceso a apoyo inmediato, la práctica de habilidades y la personalización percibida, aunque los resultados sobre motivación y pensamiento crítico son más heterogéneos (Labadze et al., 2023).

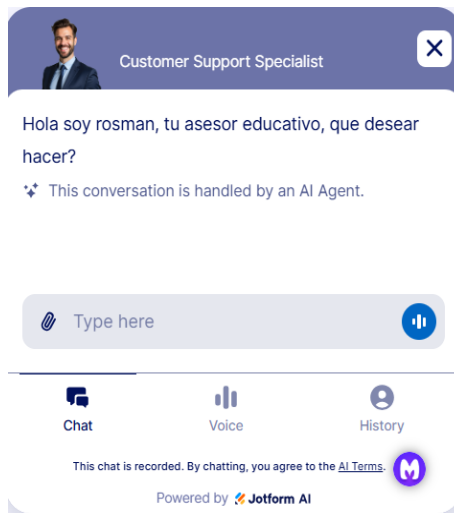
Desde el punto de vista del análisis del proceso de aprendizaje, los chatbots generan diálogos que revelan dificultades conceptuales, estrategias de estudio y emociones (ver Figura 6). Estos registros pueden alimentar modelos que identifiquen patrones de demanda de ayuda o zonas del currículo particularmente problemáticas, informando ajustes en el diseño instruccional (Garzón et al., 2025).

Asimismo, los lineamientos recientes de UNESCO subrayan la necesidad de que el diseño de chatbots considere criterios de transparencia, protección de datos y promoción de habilidades

críticas, para evitar que la interacción se limite a respuestas automáticas sin reflexión (Miao & Holmes, 2023).

Figura 6.

Aplicación de chatbot de asesoría educativa y del aprendizaje.



Nota. Chatbot de asesoría educativa y conversacional diseñado en Jotform AI. Fuente: <https://www.jotform.com/agent>

Las aplicaciones revisadas ilustran cómo la IA permite observar el aprendizaje en proceso y no solo sus resultados finales. ITS, plataformas adaptativas, sistemas de alerta temprana y chatbots producen un flujo continuo de datos que, si se analiza con criterio pedagógico, puede traducirse en intervenciones más oportunas y personalizadas. Sin embargo, la utilidad real de estas herramientas depende de integrar su diseño con marcos de analítica del aprendizaje y políticas claras sobre su uso.

2.4. Relación entre IA y analítica del aprendizaje

Tras revisar aplicaciones concretas, esta sección aborda la relación conceptual y práctica entre IA y analítica del aprendizaje. Aunque ambos campos comparten el uso intensivo de datos, sus orígenes, objetivos y métodos no son idénticos.

Las subsecciones distinguen primero las diferencias de enfoque, luego los puntos de convergencia en el uso de datos educativos y, finalmente, describen escenarios de integración IA–analítica del aprendizaje que se observan en la investigación reciente.

2.4.1. Diferencias conceptuales y operativas entre IA y analítica del aprendizaje

Conceptualmente, la analítica del aprendizaje se define como la medición, recopilación, análisis y presentación de datos sobre estudiantes y contextos con el propósito de comprender y optimizar el aprendizaje y los entornos en que ocurre. Su foco histórico ha sido apoyar procesos de mejora educativa, combinando métodos estadísticos, visualización y técnicas de ciencia de datos. La IA, en cambio, se centra en desarrollar sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como reconocer patrones complejos, tomar decisiones o generar lenguaje (Russell & Norvig, 2020; Pan et al., 2024).

Operativamente, muchos proyectos de analítica del aprendizaje emplean IA, pero no todos. Existen paneles descriptivos o modelos estadísticos simples que se consideran analítica del aprendizaje sin incorporar algoritmos avanzados. A la inversa, hay aplicaciones de IA educativa, por ejemplo, sistemas de corrección automática que no se articulan con ciclos explícitos de interpretación pedagógica ni con prácticas de mejora institucional típicas de la analítica del aprendizaje (Alfredo et al., 2024). Reconocer estas diferencias ayuda a evitar que ambos conceptos se usen como sinónimos y permite planificar proyectos en los que la IA es un componente técnico al servicio de preguntas propias de la analítica del aprendizaje.

2.4.2. Puntos de convergencia en el uso de datos educativos

Pese a las diferencias, IA y analítica del aprendizaje convergen en la explotación de grandes volúmenes de datos educativos para generar evidencias accionables. Las revisiones recientes muestran que muchas intervenciones exitosas combinan técnicas de IA (clasificación, predicción, procesamiento del lenguaje) con marcos de analítica del aprendizaje centrados en feedback, autorregulación y mejora del diseño instruccional (Banihashem et al., 2022; Weidlich et al., 2025).

Por ejemplo, un sistema puede usar modelos de lenguaje para analizar comentarios en foros y clasificar intervenciones según niveles de pensamiento crítico. Los resultados se integran luego en paneles de analítica del aprendizaje que muestran al docente qué tipos de participación predominan y qué estudiantes requieren apoyo para argumentar mejor.

De manera similar, los registros de interacción con un chatbot educativo pueden alimentar modelos de aprendizaje automático que estimen estilos de ayuda preferidos, mientras la analítica del aprendizaje traduce esos hallazgos en recomendaciones de rediseño curricular (Garzón et al., 2025; Ngulube & Ncube, 2025). En ambos casos, la convergencia reside en que la IA potencia el análisis, pero la analítica del aprendizaje orienta las preguntas pedagógicas y las formas de uso.

2.4.3. Escenarios de integración IA y analítica del aprendizaje

Estudios identifican varios escenarios de integración entre IA y analítica del aprendizaje. Un primer escenario es el de los sistemas híbridos, donde algoritmos de IA generan predicciones o clasificaciones (por ejemplo, riesgo de abandono) que se visualizan en paneles de analítica del aprendizaje diseñados para docentes y gestores.

Un segundo escenario corresponde a las intervenciones adaptativas, en las que la analítica del aprendizaje monitoriza el efecto de cambios introducidos por sistemas inteligentes (tutores, chatbots y plataformas adaptativas), y ajusta iterativamente los modelos (Alfredo et al., 2024; Hariyanto et al., 2025).

Un tercer escenario emergente es el de entornos guiados por principios de “human-centred learning analytics”, donde la integración IA–analítica se diseña explícitamente para mantener al estudiante y al docente en el centro de la toma de decisiones.

En estos casos, las recomendaciones generadas por IA se explican, se discuten y pueden ser aceptadas o rechazadas por los usuarios, alineándose con lineamientos internacionales sobre IA generativa y educación que llaman a garantizar transparencia, participación y responsabilidad compartida (Miao & Holmes, 2023; UNESCO, 2023).

Estos escenarios anticipan el tipo de ecosistemas sociotécnicos que se examinarán en capítulos posteriores, donde los datos, los modelos y las prácticas pedagógicas se analizan de forma integrada.

La IA y la analítica del aprendizaje mantienen identidades conceptuales distintas, pero se encuentran cada vez más entrelazadas en la práctica. La IA aporta técnicas potentes para modelar y predecir, mientras la analítica del aprendizaje ofrece marcos para interpretar esos modelos y convertirlos en mejoras concretas del aprendizaje y la enseñanza. Entender esta relación es clave para diseñar proyectos educativos en los que los algoritmos se pongan al servicio de preguntas pedagógicas y no al revés.

Cierre del capítulo

Este capítulo ha profundizado en la inteligencia artificial aplicada al análisis del proceso de aprendizaje, articulando fundamentos técnicos mínimos con aplicaciones educativas concretas. En primer lugar, se ha presentado una definición operativa de IA y se han diferenciado sus principales enfoques, destacando el papel central del aprendizaje automático como puente entre los datos educativos y las decisiones pedagógicas. En segundo lugar, se ha enfatizado la relevancia de la IA en educación para personalizar rutas y ritmos de aprendizaje, automatizar componentes de evaluación y seguimiento, y apoyar la toma de decisiones basada en evidencias.

En el plano práctico, se han examinado aplicaciones como los sistemas de tutoría inteligente, las plataformas adaptativas con sistemas de alerta temprana y los chatbots educativos, subrayando que su valor radica en la capacidad de hacer visibles procesos de aprendizaje antes invisibles y de generar intervenciones más oportunas. Al mismo tiempo, se ha insistido en la importancia de enmarcar estas tecnologías en principios pedagógicos, éticos y de gobernanza, asumiendo que la IA no reemplaza sino reconfigura el trabajo docente y la experiencia estudiantil.

Finalmente, se ha clarificado la relación entre IA y analítica del aprendizaje, mostrando que la primera aporta capacidades algorítmicas y la segunda proporciona marcos para interpretar y usar los datos con sentido educativo.

En los capítulos siguientes se profundizará precisamente en esos componentes, tipos de datos educativos, procesos de análisis y modelos, e indicadores de seguimiento del aprendizaje, de modo que el lector pueda comprender cómo se construyen, validan y utilizan los modelos de IA y de analítica del aprendizaje en contextos reales. Con ello se busca fortalecer una mirada crítica y propositiva sobre el uso de la IA en educación, centrada en el aprendizaje y no en la tecnología por sí misma.

CAPÍTULO III: RECOPIACIÓN, TIPOS Y CALIDAD DE LOS DATOS EDUCATIVOS

“Parece no haber dudas de que la AI también influenciará de manera significativa lo que enseñamos y aprendemos y cómo lo hacemos.” Barber.

Introducción del capítulo

En los capítulos anteriores se detalló en que la analítica del aprendizaje y la inteligencia artificial solo son tan útiles como los datos que las alimentan. La propuesta de modelos predictivos precisos, de sistemas adaptativos y de decisiones pedagógicas informadas se sostiene, en último término, sobre la calidad de los datos educativos disponibles. Cuando los registros son incompletos, incoherentes o poco representativos, las visualizaciones engañan, los modelos se sesgan y las intervenciones dejan de ser fiables.

Comprender qué tipo de datos se generan en educación, cómo se estructuran, de dónde provienen y con qué nivel de calidad llegan al analista es, por tanto, un requisito previo para cualquier proyecto serio de analítica del aprendizaje con IA.

Este capítulo se sitúa en continuidad con el Capítulo 1, donde se definieron la analítica del aprendizaje y sus fines, y con el Capítulo 2, en el que se describieron enfoques de IA aplicados al proceso de aprendizaje. Aquí desplazamos el foco hacia la “materia prima”: los datos educativos, mismos que son el insumo principal, y sobre los cuales gira el interés por estudiarlos, tratando de evitar el proceso plasmado en la Figura 7. Nos interesa tanto su naturaleza conceptual como sus características técnicas y su inserción en contextos reales, como aulas virtuales, sistemas de gestión del aprendizaje (LMS), evaluaciones en línea o repositorios institucionales.

Figura 7.

Por qué la calidad de datos importa: de la recolección a decisiones fiables.



Nota. Cadena causa-efecto: baja calidad (incompletos/incoherentes) → visualizaciones engañosas → modelos sesgados → intervenciones poco fiables.

En la sección 3.1 se aborda la naturaleza de los datos en educación, distinguiendo entre tipos de datos (numéricos, categóricos, texto, imagen, biométricos) y entre enfoques cuantitativos, cualitativos y mixtos. La sección 3.2 analiza cómo la estructura (datos tabulares frente a texto libre, por ejemplo) y la dinámica temporal condicionan las posibilidades analíticas y los modelos de IA aplicables. La sección 3.3 se centra en las principales fuentes de datos en entornos educativos digitales, desde LMS hasta encuestas y otras plataformas.

Finalmente, la sección 3.4 introduce el problema de la calidad de datos y los retos habituales en su recolección, discutiendo estrategias para construir conjuntos de datos educativos confiables. A partir de este capítulo, los desarrollos posteriores sobre preparación, modelado y visualización de datos se apoyarán en estas bases conceptuales.

3.1. Naturaleza de los datos en educación

Esta sección define qué se entiende por “dato educativo” y explora su diversidad tipológica. En analítica del aprendizaje, los datos no son solo calificaciones, incluyen trazas de interacción, producciones de los estudiantes, registros institucionales y, cada vez más, señales multimodales.

Figura 8.

Diversidad del dato educativo: trazas, producciones, registros y señales multimodales.



Nota. Mapa radial que sintetiza la diversidad de datos educativos más allá de calificaciones.

Reconocer esta diversidad de los datos (Figura 8), es fundamental para seleccionar apropiadamente métricas, algoritmos y decisiones pedagógicas, mismas que son quienes nos permitirán a futuro la toma correcta de decisiones.

3.1.1. Concepto de dato en el contexto educativo

En el contexto educativo, un dato puede entenderse como una representación digital de algún aspecto del proceso de enseñanza–aprendizaje, suficientemente estructurada como para ser almacenada, recuperada y analizada. Esta definición incluye desde la nota de un examen hasta el registro de un clic en un recurso del LMS, el número de mensajes publicados en un foro o el tiempo dedicado a un vídeo instruccional (Papadogiannis et al., 2024).

A diferencia de otros dominios, los datos educativos adquieren significado solo cuando se interpretan en relación con objetivos de aprendizaje, contextos curriculares y características del estudiantado.

Un mismo valor, por ejemplo “tres accesos por semana al curso”, puede indicar compromiso o desinterés según el tipo de actividad, la duración del curso o la modalidad de evaluación. La literatura reciente en analítica del aprendizaje insiste en que estos datos son “huellas” o “trazas” de actividad que requieren una lectura pedagógica informada, más allá de su dimensión técnica (Lang et al., 2022).

3.1.2. Tipos de datos

Los datos educativos pueden clasificarse según su forma de representación como se explica a continuación:

- ✓ Los datos lógicos, se expresan como valores binarios o booleanos (por ejemplo, “entrega realizada: sí/no”), muy frecuentes en formularios y registros administrativos.
- ✓ Los datos numéricos incluyen calificaciones, tiempos de conexión, número de intervenciones o puntuaciones de cuestionarios, y suelen ser la base de análisis estadísticos y modelos predictivos clásicos (Papadogiannis et al., 2024).
- ✓ Los datos categóricos representan atributos como carrera, modalidad de estudio o tipo de actividad (tareas, cuestionarios, foros).
- ✓ Datos textuales, procedentes de mensajes en foros, ensayos, chats con docentes o respuestas abiertas en encuestas.

En los últimos años se han incorporado datos de imagen y vídeo (por ejemplo, grabaciones de pizarras digitales, capturas de pantallas compartidas o videoconferencias), así como datos de redes, utilizados para modelar interacciones entre estudiantes y recursos.

Finalmente, los datos biométricos como ritmo cardíaco, seguimiento ocular o expresiones faciales, cobran protagonismo en la analítica multimodal centrada en el compromiso y las emociones (Guerrero-Sosa et al., 2025).

3.1.3. Datos cuantitativos, cualitativos y mixtos

Otra forma habitual de clasificar los datos en el ámbito educativo, se distingue entre cuantitativos, cualitativos y mixtos. Los datos cuantitativos se expresan numéricamente y permiten operaciones aritméticas directas, como promedios, correlaciones o modelos de regresión. Ejemplos típicos son las calificaciones, el número de recursos consultados o el tiempo dedicado a una actividad. La literatura de analítica del aprendizaje ha mostrado cómo estos datos permiten construir modelos

de predicción temprana de riesgo académico o patrones de participación (Papadogiannis et al., 2024).

Los datos cualitativos incluyen narrativas, descripciones de experiencias o interacciones complejas, como intervenciones en foros, portafolios reflexivos o diarios de aprendizaje. Estos datos capturan matices de percepción, motivación y estrategias de estudio que los indicadores puramente numéricos no reflejan.

Entre ambos polos pueden situarse los datos mixtos (mixtos), como escalas Likert o rúbricas, que codifican juicios cualitativos en categorías ordenadas (por ejemplo, “insuficiente–aceptable–sobresaliente”). Este tipo de codificación facilita la integración de valoraciones docentes o de autoevaluación en modelos cuantitativos, sin perder completamente la riqueza interpretativa de los juicios educativos (Lang et al., 2022).

En conjunto, el “dato educativo” es un concepto amplio que incluye múltiples formas de representación y enfoques de medición, que van desde valores booleanos hasta señales biométricas, los datos educativos capturan dimensiones muy distintas del aprendizaje.

Esta diversidad obliga a que los proyectos de analítica del aprendizaje con IA definan con cuidado qué aspectos del fenómeno se pretenden modelar y cómo se traducirán en variables observables. En los siguientes apartados se analizará cómo la estructura y la dinámica de estos datos condicionan los métodos analíticos disponibles y los resultados que pueden obtenerse.

3.2. Estructura y dinámica de los datos educativos

Tras caracterizar la naturaleza básica de los datos, es necesario considerar su organización y comportamiento en el tiempo. La estructura (por ejemplo, tablas bien definidas frente a texto libre) y la dinámica (datos estáticos frente a secuencias en evolución) influyen directamente en la selección de técnicas de análisis y de IA.

3.2.1. Datos estructurados y no estructurados

Los datos estructurados se almacenan siguiendo esquemas formales, como tablas con columnas y tipos fijos (número de accesos, nota del examen, programa académico, etc.). Este formato facilita

operaciones de filtrado, agregación y modelado estadístico, y es el predominante en sistemas de información académica y bases de datos institucionales (Papadogiannis et al., 2024). En un LMS, por ejemplo, los registros de calificaciones o de matriculación se almacenan en tablas relacionales, mientras que las trazas de acceso a recursos se registran en tablas de logs con campos como usuario, recurso, marca temporal y acción.

En contraste, los datos no estructurados carecen de un esquema tabular predefinido, este es el caso de textos de foros, correos electrónicos entre docentes y estudiantes, vídeos de clases sin segmentar o grabaciones de audio de debates de grupo.

El auge de la analítica multimodal ha puesto de relieve la necesidad de integrar ambos tipos de datos: estructurados (p. ej., notas y tiempos) y no estructurados (p. ej., lenguaje natural, gestos, expresiones faciales) para obtener una visión más rica del aprendizaje (Guerrero-Sosa et al., 2025).

3.2.2. Datos estáticos y dinámicos en el proceso educativo

Los datos estáticos representan información que cambia poco a lo largo del tiempo, como características sociodemográficas, historial de estudios previos o plan de estudios de un programa. En la analítica del aprendizaje, estas variables suelen utilizarse como contexto o covariables de control al construir modelos de rendimiento o abandono (Lang et al., 2022).

Por su parte, los datos dinámicos reflejan procesos en evolución: secuencias de accesos a la plataforma, sucesión de respuestas en un cuestionario adaptativo, cambios en el rendimiento a lo largo del semestre o transformaciones en las redes de interacción entre estudiantes.

En la práctica, esto se traduce en series temporales o registros de eventos con marcas de tiempo, donde el orden y la distancia entre eventos son informativos. Estudios en este ámbito, muestran que tener en cuenta la naturaleza dinámica de las interacciones con el LMS permite explicar mejor la relación entre patrones de estudio y desempeño académico que si solo se consideran agregados simples, como la suma de accesos (Hernández-García et al., 2024).

3.2.3. Implicaciones de la estructura y dinámica para el análisis

La estructura y la dinámica de los datos condicionan las técnicas analíticas y de IA que pueden aplicarse. Los datos estructurados y estáticos se ajustan bien a modelos de regresión, árboles de decisión o algoritmos clásicos de minería de datos; sin embargo, cuando la información clave reside en el orden de los eventos (por ejemplo, el modo en que un estudiante alterna entre recursos y actividades), se requieren enfoques secuenciales, como modelos de series temporales, redes recurrentes o análisis de secuencias (Lang et al., 2022).

Con los datos no estructurados se necesitan fases adicionales de preprocesamiento, como técnicas de procesamiento del lenguaje natural para texto, o de visión por computador para imagen y vídeo. La multimodalidad exige, además, sincronizar fuentes con distintas frecuencias de muestreo y granularidades, reto ampliamente documentado en revisiones recientes (Guerrero-Sosa et al., 2025).

Para el diseño de proyectos de analítica del aprendizaje con IA es crucial decidir qué nivel de complejidad es viable dada la infraestructura disponible y si los beneficios pedagógicos justifican el esfuerzo adicional de integrar múltiples fuentes y formatos.

En síntesis, no basta con identificar “qué” datos se tienen, sino también “cómo” se organizan y “cómo” evolucionan. La distinción entre datos estructurados/no estructurados y estáticos/dinámicos ayuda a seleccionar herramientas analíticas y arquitecturas de IA adecuadas, así como a anticipar los retos de integración y preprocesamiento. Estos elementos preparan el terreno para el siguiente apartado, en el que se examinan las principales fuentes digitales desde las que se generan y concentran estos datos en los entornos educativos actuales.

3.3. Fuentes de datos en entornos educativos digitales

Hasta aquí se han descrito la naturaleza y organización de los datos educativos. Esta sección se centra en “dónde” se originan dichos datos en la práctica. En entornos digitales, la mayor parte

de las trazas de aprendizaje se concentran en unos pocos sistemas clave, que conviene conocer para planificar cualquier iniciativa de analítica del aprendizaje.

3.3.1. LMS y plataformas institucionales

Los sistemas de gestión del aprendizaje (LMS), como Moodle, Canvas o Blackboard, son la fuente central de datos en muchas instituciones de educación superior. Estos sistemas almacenan información sobre matriculación, calificaciones, entregas de tareas, intentos de cuestionarios, participación en foros y registros de acceso a recursos.

Varios estudios han mostrado que las interacciones en el LMS, frecuencia de ingreso, visualización de contenidos o uso de cuestionarios de práctica, se relacionan de forma significativa con el rendimiento académico (Hernández-García et al., 2024).

Además de los LMS, las instituciones cuentan con sistemas de información académica (SIA), plataformas de gestión de bibliotecas, repositorios de trabajos de grado y sistemas de gestión de exámenes en línea. En educación a distancia o virtual, los estudios indican que, en muchos contextos, los datos provenientes de LMS constituyen la base de la mayoría de estudios de analítica del aprendizaje, especialmente en programas en línea o híbridos (Palanci et al., 2024). Sin embargo, estos datos suelen centrarse en la actividad dentro de una plataforma específica y pueden no reflejar la totalidad de la experiencia de aprendizaje de los estudiantes, lo que debe considerarse al interpretar resultados.

3.3.2. Formularios, encuestas y registros docentes

Otra fuente relevante de datos educativos proviene de formularios y encuestas aplicadas a estudiantes y docentes. Estos instrumentos recogen percepciones sobre la calidad de la docencia, la carga de trabajo, la satisfacción con el curso o el uso de recursos. A diferencia de los trazes digitales, estas respuestas suelen ser autorreportadas y están más próximas a dimensiones afectivas y motivacionales, necesarias para comprender el proceso de aprendizaje en su complejidad (Lang et al., 2022).

Los registros docentes como planillas de seguimiento, rúbricas de evaluación, notas de observación de clase o reportes de tutoría, añaden una capa cualitativa o mixta que complementa las evidencias derivadas de plataformas digitales.

Revisiones sobre analítica del aprendizaje en educación a distancia o en línea destacan la utilidad de combinar logs del LMS con encuestas y registros docentes para interpretar por qué ciertos patrones de interacción se asocian con mejores o peores resultados (Palanci et al., 2024). Esta triangulación fortalece las conclusiones y reduce el riesgo de atribuir causalidad a simples correlaciones.

3.3.3. Otras fuentes digitales de datos de aprendizaje

Más allá de los LMS y los instrumentos institucionales, existen numerosas fuentes complementarias de datos de aprendizaje. En cursos masivos abiertos (MOOC), por ejemplo, plataformas específicas generan trazas detalladas de navegación, visualización de vídeos y participación en foros globales. Aplicaciones de respuesta en clase (clickers), simuladores en línea, laboratorios virtuales y herramientas de codificación colaborativa producen registros muy finos de las acciones del estudiante.

En contextos más avanzados, las investigaciones multimodales incorporan sensores, dispositivos portables y sistemas de videoconferencia para capturar gestos, expresiones faciales o indicadores fisiológicos, con el fin de estudiar compromiso, colaboración o desarrollo de habilidades blandas (Guerrero-Sosa et al., 2025).

Asimismo, estudios recientes sobre paisajes de datos educativos en América Latina han mostrado la relevancia de observatorios nacionales y repositorios abiertos para la analítica a nivel de sistema, subrayando al mismo tiempo importantes brechas de granularidad para usos de learning analytics centrados en el estudiante (Ruidias et al., 2025).

Esta sección, evidencia que los datos educativos surgen de un ecosistema distribuido de sistemas como LMS, plataformas institucionales, encuestas, registros docentes, repositorios abiertos y herramientas especializadas.

Cada fuente ofrece una ventana distinta sobre el aprendizaje y, por tanto, sobre diferentes tipos de preguntas de investigación y decisión pedagógica. En los capítulos posteriores, cuando se aborden técnicas de preparación y análisis de datos, será crucial tener en cuenta el origen de la información para valorar su alcance, limitaciones y posibles sesgos.

3.4. Calidad de los datos y problemas de recolección

Disponer de grandes volúmenes de datos no garantiza análisis confiables, la calidad de los datos en términos de completitud, precisión, coherencia, actualidad y relevancia condiciona la validez de cualquier modelo de analítica del aprendizaje o de IA educativa.

Esta sección describe dimensiones de calidad de datos, problemas comunes en su captura y almacenamiento, y pautas para construir conjuntos de datos confiables.

3.4.1. Dimensiones de calidad: completitud, precisión, coherencia, actualidad y relevancia

Los marcos generales de calidad de datos, ampliamente desarrollados en el campo de los sistemas de información, proponen dimensiones como completitud, precisión, coherencia, actualidad y relevancia, entre otras (Wang & Strong, 1996; Wang et al., 2024).

- *La completitud*, se refiere a la medida en que los valores esperados están presentes, por ejemplo, cuando faltan respuestas en una encuesta o notas en el historial académico.
- *La precisión*, alude a la corrección de los datos frente a la realidad, como errores de digitación en la calificación de un examen.
- *La coherencia*, implica que los datos no se contradicen entre sí, tanto dentro de un mismo sistema como entre sistemas (por ejemplo, cuando el número de créditos de una asignatura difiere entre el LMS y el sistema académico).
- *La actualidad (o vigencia)*, se relaciona con que los datos estén actualizados respecto del momento del análisis; en educación, esto es clave cuando se usan datos históricos para decisiones de intervención temprana.

- *La relevancia*, se refiere al grado en que los datos son pertinentes para las preguntas educativas planteadas.

De modo tal, que disponer de muchos indicadores irrelevantes puede complicar los análisis sin mejorar la toma de decisiones, y hacer que esta tarea se vuelva infructuosa (Wang et al., 2024).

3.4.2. Problemas frecuentes en la captura y almacenamiento de datos educativos

En la práctica, los conjuntos de datos educativos presentan múltiples problemas. Uno de los más habituales es la presencia de valores perdidos o incompletos, ya sea porque los estudiantes no responden encuestas, porque ciertos eventos no se registran (por ejemplo, actividades realizadas fuera del LMS) o por errores en la configuración de los sistemas.

Asimismo, son frecuentes los errores de codificación, como categorías inconsistentes para la misma variable entre cursos o años académicos, lo que limita la comparabilidad longitudinal (Palanci et al., 2024).

Otra dificultad reside en la fragmentación de datos entre sistemas no integrados es el rendimiento que puede registrarse en un sistema, la asistencia en otro y la participación en un tercero, sin identificadores comunes claros. Estudios sobre datos educativos en América Latina, han señalado que muchos observatorios y repositorios fueron diseñados inicialmente para fines de rendición de cuentas y no para analítica fina del aprendizaje, lo que genera brechas de granularidad y problemas de interoperabilidad (Ruidias et al., 2025).

Además, la ausencia de documentación adecuada (metadatos) dificulta conocer cómo se generaron las variables y bajo qué supuestos, lo que reduce la transparencia y la reproducibilidad.

3.4.3. Construcción de datasets educativos confiables

Construir conjuntos de datos educativos confiables implica mucho más que aplicar algoritmos de limpieza automática. El punto de partida es una definición clara de los objetivos analíticos y de las decisiones pedagógicas que se esperan informar, lo que orienta la selección de variables relevantes.

A continuación, es necesario diseñar un esquema de datos explícito y documentado, especificando fuentes, transformaciones y criterios de codificación. Este tipo de “alfabetización en datos educativos” ha sido señalada como clave para la colaboración entre especialistas en datos y expertos pedagógicos (Papadogiannis et al., 2024).

Sobre esta base, se deben implementar procedimientos sistemáticos de limpieza (detección de valores atípicos, imputación de valores perdidos, armonización de categorías), así como estrategias de integración entre sistemas, asegurando identificadores consistentes y metadatos ricos. En trabajos recientes se ha insistido en la importancia de adoptar principios FAIR (localizables, accesibles, interoperables y reutilizables) y marcos de calidad adaptados al ámbito educativo para garantizar que los datos no solo estén disponibles, sino que sean efectivamente utilizables para analítica del aprendizaje y analítica académica (Ruidias et al., 2025; Wang et al., 2024).

La participación de docentes y responsables académicos en la validación de los conjuntos de datos es un componente crítico para asegurar que las variables reflejan adecuadamente la realidad educativa que se pretende modelar, y es así, donde resulta el valor real de procesar y limpiar los datos, y de manera especial, donde se encuentran almacenados, conforme se evidencia en la siguiente imagen.

Figura 9.

Preparación de dataset académico.

	A	B	C	D	E	F	G	H	I
	cedula_estudiante	nombres_estudiante	descripcion	tipo	dominio_valores	origen	periodicidad	responsable	calificacion
1	7012039405	Juan Pérez Andrade	Participación en foro Semana 1	Indicador	Participación	Primaria	Semanal	Docente	10
2	7011223344	Ana Lúcia Torres	Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	8
3	7043020282	Pepito Luis Andrade	Participo en debate	Indicador	Participación	Universidad	Mensual	Docente	9
4			Participación en foro Semana 2	Indicador	Participación	Primaria	Semanal	Docente	8
5			Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	10
6			Participo en debate	Indicador	Participación	Universidad	Mensual	Docente	7
7			Participación en foro Semana 3	Indicador	Participación	Primaria	Semanal	Docente	6,5
8			Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	9
9			Participo en debate	Indicador	Participación	Universidad	Mensual	Docente	8
10			Participación en foro Semana 4	Indicador	Participación	Primaria	Semanal	Docente	7,5
11			Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	7
12			Participo en debate	Indicador	Participación	Universidad	Mensual	Docente	6,5
13			Participación en foro Semana 5	Indicador	Participación	Primaria	Semanal	Docente	9
14			Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	8
15			Participo en debate	Indicador	Participación	Universidad	Mensual	Docente	7,5
16			Participación en foro Semana 6	Indicador	Participación	Primaria	Semanal	Docente	7
17			Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	6,5
18			Participo en debate	Indicador	Participación	Universidad	Mensual	Docente	9
19			Participación en foro Semana 2	Indicador	Participación	Primaria	Semanal	Docente	8
20			Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	9
21			Participo en debate	Indicador	Participación	Universidad	Mensual	Docente	8
22			Participación en foro Semana 3	Indicador	Participación	Primaria	Semanal	Docente	10
23			Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	9
24			Participo en debate	Indicador	Participación	Universidad	Mensual	Docente	8
25			Participación en foro Semana 4	Indicador	Participación	Primaria	Semanal	Docente	7,5
26			Entregó Tarea 1 y recibió retroalimentación	Indicador	Entregas	Secundaria	Semanal	Tutor	7
27			Participo en foro Semana 2	Indicador	Participación	Universidad	Mensual	Docente	6,5
28			Entregó Tarea 1 y recibió retroalimentación	Indicador	Participación	Primaria	Semanal	Docente	10
29									

Nota. Dataset académico de registros de maestrantes de UNEMI de maestría en IA para la educación.

De este modo, en estos puntos se pone de relieve que la calidad de los datos no es un detalle técnico menor, sino un determinante de la validez de los análisis y de la confiabilidad de los modelos de IA en educación. Dimensiones como completitud, precisión, coherencia, actualidad y relevancia proporcionan un vocabulario común para evaluar y mejorar conjuntos de datos educativos. El reconocimiento de problemas frecuentes: fragmentación, documentación insuficiente, datos “débiles” para learning analytics, y las estrategias para mitigarlos preparan el terreno para los capítulos siguientes, dedicados a la preparación, análisis y visualización de datos orientados a la toma de decisiones pedagógicas.

Cierre del capítulo

Este capítulo ha situado los datos educativos en el centro de la analítica del aprendizaje y de la IA aplicada a la educación. Tras discutir su naturaleza conceptual, se ha mostrado que los datos van mucho más allá de las calificaciones, incluyendo trazas de interacción, producciones textuales, registros multimedia, redes de relación y, en contextos avanzados, señales biométricas. También se ha subrayado la coexistencia de enfoques cuantitativos, cualitativos y mixtos, y la necesidad de interpretarlos siempre dentro de marcos pedagógicos claros.

El análisis de la estructura y dinámica de los datos ha evidenciado que la forma en que se organizan (estructurados o no estructurados), y cómo evolucionan en el tiempo (estáticos o

dinámicos), condiciona las técnicas de análisis y los modelos de IA aplicables. Las secciones dedicadas a las fuentes de datos han mostrado que los LMS, las plataformas institucionales, las encuestas, los registros docentes y otros sistemas especializados conforman un ecosistema de datos cuya integración es tanto un reto técnico como organizacional (Hernández-García et al., 2024; Palanci et al., 2024).

Finalmente, la discusión sobre calidad de datos ha dejado claro que ningún modelo, por sofisticado que sea, puede compensar datos incompletos, imprecisos o poco relevantes. Los marcos de calidad y los principios FAIR ofrecen guías concretas para mejorar la fiabilidad de los conjuntos de datos educativos, mientras que las contribuciones recientes sobre paisajes de datos en educación alertan sobre las brechas existentes, especialmente en contextos donde los sistemas fueron diseñados para la rendición de cuentas y no para el apoyo fino al aprendizaje (Ruidias et al., 2025; Wang et al., 2024).

En los capítulos siguientes, este andamiaje conceptual se traducirá en procesos operativos: selección de variables, integración de fuentes, técnicas de preprocesamiento, construcción de indicadores e implementación de modelos de analítica del aprendizaje e IA.

La comprensión profunda de qué datos están disponibles, cómo se generan y qué grado de calidad presentan permitirá que dichas técnicas se apliquen con mayor responsabilidad y con un impacto más significativo en la mejora de los procesos de enseñanza y aprendizaje.

CAPÍTULO IV: PREPARACIÓN, ANÁLISIS Y VISUALIZACIÓN DE DATOS EDUCATIVOS

“Considerar el sesgo un mero problema de datos representa una visión demasiado limitada.”

UNICEF.

Introducción del capítulo

La analítica del aprendizaje y los modelos de inteligencia artificial solo son tan sólidos como los datos y el tratamiento que reciben. Tras abordar en el capítulo anterior la naturaleza, las fuentes y la calidad de los datos educativos, este capítulo se centra en el “puente” entre la recolección de datos y la construcción de modelos de IA e indicadores: la preparación, el análisis y la visualización de los datos. Sin una etapa rigurosa de limpieza, exploración y representación visual, los resultados analíticos pueden ser engañosos o llevar a decisiones pedagógicas equivocadas.

En la literatura de learning analytics se reconoce que las decisiones tomadas en la fase de preparación de datos afectan directamente la validez de los análisis y la utilidad de las intervenciones posteriores.

Esto incluye cómo se tratan los datos perdidos, los valores atípicos, las inconsistencias y los registros duplicados, así como la elección de técnicas descriptivas y predictivas y la forma de presentarlas a docentes y gestores. Trabajos recientes muestran, además, que el uso de visualizaciones y dashboards influye en la capacidad de los usuarios educativos para interpretar los datos y transformar la información en acción pedagógica.

En este contexto, el capítulo se organiza en cuatro bloques. La sección 4.1 aborda la preparación y limpieza de datos, centrada en problemas típicos de conjuntos de datos educativos. La sección 4.2 se ocupa de los métodos de análisis, tanto descriptivos como predictivos, y de su interpretación pedagógica. La sección 4.3 profundiza en la visualización de datos, discutiendo criterios para elegir gráficos y principios de diseño comprensibles para actores educativos. Finalmente, la sección 4.4 presenta herramientas concretas, desde hojas de cálculo hasta entornos de

programación, que permiten implementar estas prácticas en proyectos reales de analítica del aprendizaje e IA en educación.

4.1. Preparación y limpieza de datos

La preparación y limpieza de datos constituye la primera fase operativa después de la recolección y antes de cualquier modelado avanzado o construcción de indicadores. En el contexto educativo, esta fase implica revisar los registros provenientes de LMS, sistemas académicos, encuestas y otras fuentes para identificar datos faltantes, valores atípicos, inconsistencias y duplicados.

Dado que los datos educativos suelen ser “desordenados” con errores de registro, entradas incompletas y cambios en los sistemas, su depuración es esencial para evitar sesgos en los análisis y en los modelos de IA que se construyan posteriormente.

En las subsecciones siguientes se presentan tres problemas centrales de esta etapa: el tratamiento de datos perdidos, el manejo de valores atípicos e inconsistencias, y la detección de instancias duplicadas. En conjunto, ofrecen un marco práctico para que equipos docentes, investigadores y unidades de analítica institucional preparen conjuntos de datos confiables que alimentarán los análisis del capítulo 4.2, las visualizaciones del 4.3 y, más adelante, los modelos de IA y los indicadores del capítulo 5.

4.1.1. Tratamiento de datos perdidos (missing data)

En educación, los datos perdidos pueden surgir por inasistencias a evaluaciones, errores de registro en el LMS o respuestas omitidas en encuestas. Ignorar estas ausencias puede distorsionar promedios, correlaciones y modelos predictivos, en especial cuando la no respuesta se relaciona con el desempeño o el nivel de participación del estudiantado.

La literatura distingue entre datos perdidos completamente al azar, al azar y no al azar; aunque esta tipología es técnica, ayuda a justificar si es razonable eliminar casos o si conviene imputar valores.

En learning analytics se observan tres estrategias frecuentes.

- ✓ La primera es la eliminación de registros incompletos, práctica sencilla pero riesgosa cuando el porcentaje de pérdida es alto o sistemático.
- ✓ La segunda son las imputaciones simples, como reemplazar valores faltantes por la media de la clase o por una categoría “no responde”.
- ✓ La tercera incluye técnicas más avanzadas, como imputación múltiple o modelos de aprendizaje automático que estiman el valor faltante a partir de otros atributos del estudiante.

En proyectos educativos, la elección debe explicitarse y documentarse, pues afecta directamente la identificación de estudiantes en riesgo, la evaluación de intervenciones y la equidad de las decisiones que se tomen sobre la base de los datos.

4.1.2. Manejo de valores atípicos e inconsistencias

Los valores atípicos (outliers) son observaciones que se alejan notablemente del patrón general de los datos, como tiempos de respuesta extremadamente breves en un cuestionario o calificaciones imposibles en un registro de notas. En contextos educativos, estos outliers pueden reflejar errores de captura, conductas de “adivinanza rápida” o situaciones excepcionales que conviene examinar más de cerca.

Detectarlos implica combinar técnicas estadísticas, como diagramas de caja, puntajes z o reglas basadas en cuartiles, con conocimiento del contexto pedagógico, por ejemplo, la duración habitual de una actividad o la escala real de calificaciones.

Las inconsistencias surgen cuando la misma variable aparece registrada con reglas distintas a lo largo del tiempo o entre sistemas, como diferencias en la codificación de estados de matrícula o en escalas de evaluación. La revisión de rangos válidos (tal como se observa en la Figura 10), tipos de datos y reglas de negocio es indispensable para evitar interpretaciones erróneas.

Figura 10.

Valores atípicos en dataset educativos.

Casos atípicos



Filtra estudiantes con valores marcados como atípicos en asistencia, tiempo en plataforma o nota final

	A ▼	B ▼	C ▼	D ▼	E ▼	F ▼	G ▼	+
1	ID_Estudiante	Porcentaje_Asistencia	Tiempo_Plataforma_Minutos	Nota_Final(Rendimiento)	Atípico_Asistencia	Atípico_Tiempo	Atípico_Nota	
2	15	10	12	10			Atípico	
3	17	72	1000	75		Atípico		
4	22	40	5000	55		Atípico		
+	New row							

Nota. Cálculo de valores atípicos de dataset educativo utilizando herramienta de IA Rows IA.

Fuente: <https://rows.com>.

Revisiones recientes subrayan que el manejo de outliers e inconsistencias es parte integral de los flujos de trabajo de analítica del aprendizaje y minería de datos educativos, ya que afecta la estabilidad de los modelos predictivos y la confianza de docentes y gestores en los resultados.

En lugar de eliminar automáticamente los valores extremos, es recomendable documentar las decisiones y, cuando corresponda, discutir casos con los equipos pedagógicos.

4.1.3. Detección y eliminación de instancias duplicadas

Los duplicados aparecen cuando se registra más de una vez el mismo evento o estudiante, por ejemplo, por migraciones de datos entre plataformas, errores de integración de sistemas o reenvío de formularios.

En un LMS, esto puede manifestarse como múltiples filas idénticas de acceso al mismo recurso en el mismo instante, o como duplicidad de identificadores de estudiante tras cambios administrativos. Este problema afecta métricas básicas como recuentos de accesos, número de estudiantes activos o tasa de aprobación, y puede inflar artificialmente la participación o el rendimiento.

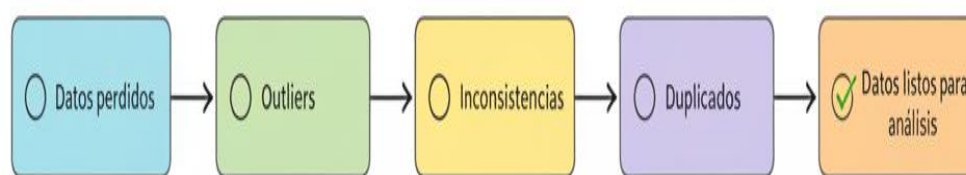
En la literatura de calidad de datos se recomienda definir reglas claras de unicidad, apoyadas en claves compuestas (por ejemplo, identificador de estudiante, curso, fecha y tipo de evento), así como utilizar procedimientos automáticos para identificar potenciales duplicados y revisar manualmente los casos dudosos.

En proyectos de learning analytics, los duplicados también pueden detectarse al comparar resúmenes agregados con reportes oficiales institucionales, de modo que las discrepancias disparen alertas de revisión. Eliminar duplicados o marcarlos explícitamente, es fundamental para que los análisis posteriores representen fielmente el comportamiento real del estudiantado y sustenten indicadores confiables en etapas de modelado.

La preparación y limpieza de datos educativos implica tratar con rigor los datos perdidos, los valores atípicos, las inconsistencias y los duplicados. Estas tareas, aunque a menudo consideradas “preliminares”, tienen un impacto directo en la validez de cualquier análisis descriptivo o predictivo realizado posteriormente, donde dicho proceso, se evidencia en la Figura 11, para llegar a la depuración resultado de los datos deseados.

Figura 11.

Preparación y limpieza: checklist operativo para datos educativos.



Nota. Flujo de limpieza con foco en datos perdidos, valores atípicos, inconsistencias y duplicados como condición para validez posterior.

Para quienes diseñan proyectos de analítica del aprendizaje o modelos de IA educativa, es clave documentar las decisiones tomadas en esta etapa y comunicarlas a los equipos pedagógicos, dado que influyen en la identificación de estudiantes en riesgo, en la evaluación de cursos y en la formulación de políticas institucionales.

En el capítulo 5, se mostrará cómo estas decisiones de limpieza condicionan la calidad de los modelos de IA y de los indicadores que se construyan sobre los datos preparados.

4.2. Métodos de análisis de datos educativos

Una vez que los datos han sido limpiados y estructurados, el siguiente paso consiste en analizarlos para comprender patrones de rendimiento, participación y trayectoria estudiantil. Los métodos de análisis en educación abarcan desde estadísticas descriptivas básicas hasta modelos de aprendizaje automático orientados a la predicción o clasificación de resultados.

Sin embargo, más allá de la sofisticación técnica, el desafío central es traducir los resultados en decisiones pedagógicas concretas (ver Figura 12), como rediseñar actividades, ajustar la carga evaluativa o implementar apoyos focalizados.

Figura 12.

De resultados a acción: interpretación pedagógica del análisis.



Nota. Modelo simple que convierte hallazgos de rendimiento/participación/trayectoria en decisiones (rediseño, apoyo focalizado), enfatizando interpretación pedagógica.

En las subsecciones siguientes se presentan tres dimensiones complementarias: el análisis descriptivo del rendimiento y la participación, los enfoques predictivos para anticipar resultados de aprendizaje y la interpretación pedagógica de los hallazgos. Juntas permiten avanzar desde la mera descripción de lo que ya ocurrió hacia el uso de datos para orientar intervenciones formativas y políticas institucionales.

4.2.1. Análisis descriptivo del rendimiento y la participación

El análisis descriptivo busca resumir y explorar los datos disponibles sin realizar inferencias causales ni predicciones. En el ámbito educativo, esto incluye calcular tasas de aprobación,

promedios de calificaciones, distribución de intentos en cuestionarios y patrones de acceso a recursos digitales. Estudios recientes en educación superior muestran que, a través de estadísticas descriptivas y visualizaciones simples, es posible identificar asignaturas con alta tasa de reprobación, periodos críticos del semestre o actividades con baja participación que requieren rediseño.

A nivel de interacción con LMS, el análisis descriptivo suele centrarse en conteos de clics, tiempo de conexión, participación en foros y uso de materiales multimedia. Estas métricas permiten comparar grupos, detectar brechas de participación y explorar si existen relaciones preliminares entre compromiso y rendimiento. Textos introductorios para docentes y directivos resaltan que este tipo de análisis, cuando se presenta de forma clara y contextualizada, se convierte en una herramienta potente para el monitoreo continuo de cursos y programas, sin requerir conocimientos avanzados de estadística.

4.2.2. Análisis predictivo para anticipar resultados de aprendizaje

El análisis predictivo utiliza modelos estadísticos o de aprendizaje automático para anticipar resultados futuros, como la probabilidad de aprobar una asignatura, el riesgo de abandono o el logro de determinadas competencias. Revisión sistemática reciente muestra que técnicas como árboles de decisión, bosques aleatorios, máquinas de soporte vectorial y redes neuronales son ampliamente utilizadas para predecir el rendimiento estudiantil a partir de datos de interacción, evaluaciones previas y características demográficas.

Trabajos empíricos en entornos virtuales han demostrado que, combinando datos de participación en el LMS con información de evaluaciones, los modelos predictivos pueden identificar tempranamente estudiantes en riesgo, permitiendo intervenciones oportunas.

Al mismo tiempo, se advierte sobre la necesidad de considerar la calidad de los datos de entrada, el equilibrio de clases (por ejemplo, pocos casos de abandono frente a muchos de éxito) y la transparencia de los modelos. En contextos educativos, no basta con una precisión alta: es

fundamental que los modelos sean interpretables y se utilicen como apoyo a la acción docente, no como mecanismos automáticos de decisión.

4.2.3. Interpretación pedagógica de los resultados

La interpretación pedagógica constituye el eslabón que conecta los resultados del análisis con decisiones concretas de enseñanza, evaluación y apoyo al estudiantado. Investigaciones recientes sobre learning analytics enfatizan que los datos, por sí solos, no “hablan”: requieren ser contextualizados en relación con los objetivos de aprendizaje, el diseño instruccional y las condiciones de los estudiantes.

Por ejemplo, una correlación entre número de accesos al LMS y calificación final no implica que “más clics causen mejores notas”; puede reflejar que el uso intensivo del entorno es un indicador de compromiso general, o que los estudiantes con dificultades consultan más los materiales.

Revisiones sobre dashboards y feedback destacan que las analíticas son más efectivas cuando se acompañan de apoyo interpretativo, guías para la acción y espacios de reflexión docente, en lugar de presentarse como rankings o juicios definitivos.

La interpretación pedagógica también exige considerar cuestiones éticas y de justicia educativa: cómo se utiliza la información sobre “riesgo” o “bajo rendimiento”, qué oportunidades de mejora se ofrecen y cómo se evita estigmatizar a ciertos grupos de estudiantes. De este modo, el análisis de datos se alinea con una visión formativa y no meramente clasificatoria del proceso educativo.

Los métodos de análisis de datos educativos abarcan desde exploraciones descriptivas hasta modelos predictivos complejos, pero su valor real reside en la capacidad de informar decisiones pedagógicas y de política académica. Las estadísticas descriptivas permiten monitorear rendimiento y participación, mientras que los modelos predictivos contribuyen a anticipar problemas y planificar intervenciones.

Sin embargo, sin una interpretación pedagógica crítica y contextualizada, incluso las analíticas más sofisticadas pueden llevar a conclusiones erróneas o injustas. La siguiente sección mostrará

cómo la visualización adecuada de los resultados facilita este proceso de interpretación y comunicación con distintos actores educativos.

4.3. Visualización de datos educativos

La visualización de datos educativos tiene como finalidad transformar métricas y resultados de análisis en representaciones gráficas comprensibles para docentes, estudiantes y gestores. En learning analytics, las visualizaciones, incluidos dashboards y paneles personalizados son un componente central para apoyar la toma de decisiones basada en datos.

Sin embargo, diseñar buenas visualizaciones requiere atender tanto a criterios técnicos (tipo de variable, escala, distribución) como a criterios comunicativos (claridad, carga cognitiva, alineación con las preguntas pedagógicas).

En las subsecciones siguientes se abordan tres aspectos clave: cómo seleccionar el tipo de gráfico según el objetivo analítico y la naturaleza de las variables, el uso de gráficos básicos (barras, histogramas, diagramas de caja y otros en educación), y los principios de diseño que hacen que estas visualizaciones sean comprensibles y accionables para la comunidad educativa.

4.3.1. Selección de gráficos según tipo de variable y objetivo

La elección del gráfico adecuado depende de lo que se quiere mostrar y del tipo de variable. Para comparar frecuencias entre categorías (por ejemplo, estados de aprobación, grupos de riesgo, modalidades de estudio), suelen utilizarse gráficos de barras. Para representar tendencias temporales como la evolución semanal de accesos al LMS o el progreso en evaluaciones, se recomiendan gráficos de líneas. Variables continuas, como calificaciones o tiempos de conexión, pueden explorarse mediante histogramas y diagramas de caja.

Las revisiones sobre visual learning analytics destacan que una buena selección de gráficos facilita la identificación de patrones relevantes, evita interpretaciones erróneas y reduce la saturación visual.

En la práctica, esto implica cuestionarse el siguiente listado de interrogantes:

1. ¿El gráfico ayuda a responder una pregunta pedagógica concreta?
2. ¿Permite comparar grupos de interés (por ejemplo, estudiantes en riesgo frente a los que no están en riesgo)?
3. ¿Muestra adecuadamente la variabilidad de los datos?

Responder estas preguntas antes de diseñar visualizaciones contribuye a que los dashboards sean herramientas de reflexión, en lugar de colecciones de gráficos decorativos.

4.3.2. Uso de gráficos de barras, histogramas, diagramas de caja y otros

En contextos educativos, un conjunto reducido de gráficos bien diseñados cubre la mayoría de necesidades. Dichas variantes se describen de manera consecutiva:

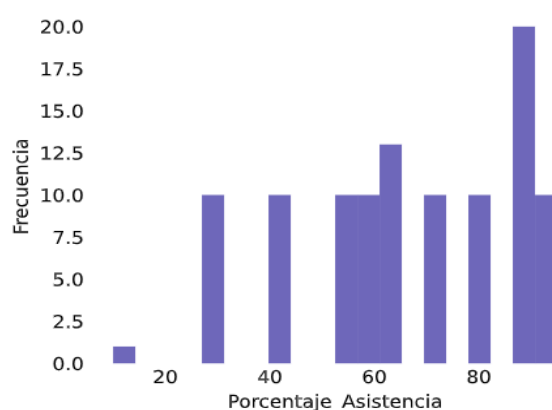
- Los gráficos de barras son útiles para mostrar la distribución de calificaciones por rangos, la proporción de estudiantes que completan actividades o el uso relativo de diferentes recursos.
- Los histogramas permiten observar la forma de la distribución de una variable, por ejemplo, si las calificaciones se concentran en valores altos o bajos, y detectar asimetrías o agrupamientos (observar Figura 13).
- Los diagramas de caja, por su parte, resumen la mediana, los cuartiles y los valores atípicos, resultando especialmente apropiados para comparar grupos o cursos.

Investigaciones sobre visual learning analytics, muestran que estos gráficos básicos, combinados con filtros interactivos (por ejemplo, por cohorte, asignatura o género), ayudan a docentes y estudiantes a explorar los datos desde distintas perspectivas sin perder claridad.

Figura 13.

Visualización de resultados en dataset educativos.

Histogramas: Asistencia, Tiempo y Nota



Nota. Cálculo de valores mediante histograma entre la asistencia, tiempo en plataforma y nota de los estudiantes de maestría utilizando herramienta de IA Rows IA. Fuente: <https://rows.com>.

Además, gráficos de dispersión pueden utilizarse para visualizar relaciones entre dos variables, como participación y rendimiento, mientras que mapas de calor resultan útiles para representar intensidades de uso de recursos o momentos críticos del calendario académico. La clave es mantener un equilibrio entre detalle y simplicidad para evitar sobrecargar a los usuarios.

4.3.3. Principios de diseño de visualizaciones comprensibles para docentes y gestores

El diseño de visualizaciones para docentes y gestores requiere tener en cuenta la diversidad de alfabetización de datos y de tiempo disponible para interpretar gráficos. Recomendaciones derivadas de la investigación en dashboards educativos subrayan la importancia de utilizar escalas claras, leyendas explícitas, colores con significado consistente y textos breves que expliquen qué representa cada elemento.

También se sugiere evitar el uso excesivo de efectos tridimensionales o elementos decorativos que no aportan información y pueden distraer del mensaje principal. Los tableros de analítica del aprendizaje, según investigaciones señalan que las visualizaciones son más útiles cuando están alineadas con metas de aprendizaje explícitas y ofrecen recomendaciones o explicaciones sobre posibles acciones, en lugar de limitarse a mostrar datos “en bruto”.

Asimismo, se ha documentado que la co-diseñación de dashboards con docentes y estudiantes favorece que las visualizaciones resulten significativas y accionables para quienes las usarán en

la práctica. Estos principios de diseño serán fundamentales cuando, en el capítulo 5, se aborden indicadores específicos y modelos de IA que deban comunicarse a diferentes públicos dentro de la institución.

La visualización de datos educativos traduce métricas y resultados analíticos en representaciones comprensibles que apoyan la reflexión y la toma de decisiones. La elección adecuada de gráficos, el uso cuidadoso de barras, histogramas y diagramas de caja, y la aplicación de principios de diseño centrados en el usuario son factores decisivos para que dashboards y paneles de analítica resulten realmente útiles.

Estas visualizaciones constituyen, además, la interfaz principal entre los análisis descritos en la sección 4.2 y los usuarios finales que deberán interpretar los resultados. En la siguiente sección se presentarán herramientas concretas para implementar estas prácticas de análisis y visualización en proyectos de analítica del aprendizaje e IA en educación.

4.4. Herramientas para el análisis de datos educativos

La implementación práctica de la preparación, análisis y visualización de datos educativos requiere seleccionar herramientas adecuadas a los recursos y competencias disponibles en cada institución. Estas herramientas van desde hojas de cálculo de uso cotidiano hasta plataformas de inteligencia de análisis y de negocios (por ejemplo, Rows, Julius, R, Power BI, etc.) y entornos de programación que permiten desarrollar modelos avanzados de IA.

La elección no es solamente técnica, debido a que condiciona quién puede participar en el proceso analítico, cómo se comparten los resultados y qué tan sostenible es la iniciativa en el tiempo. En el gráfico siguiente, se muestra cómo con una herramienta de IA se puede ingresar y analizar datos de pequeño, mediano y gran volumen de información.

Figura 14.

Análisis de datos educativos con herramientas de IA.

A few quick notes from the summary stats:

- There are **104 students** (count = 104 across all columns).
- Average `Nota_Final(Rendimieneto)` is about **72.06** with a fairly wide spread (std about **15.91**).
- `Tiempo_Plataforma_Minutos` ranges from **10** to **210**.

(Full descriptive summary is available here:)

	ID_Estudiante	Porcentaje_Asistencia	Tiempo_Plataforma_Minutos	Intentos_Evaluacion_Final(...	Nota_Final(R
count	104.0	104.0	104.0	104.0	104.0
mean	50.91346153846154	66.6826923076923	134.5096153846154	1.9807692307692308	72.05769230
std	28.910343691617342	20.948510350022232	60.089576368952294	1.0238018735700847	15.90556288
min	1.0	10.0	10.0	0.0	10.0
25%	26.75	55.0	90.0	1.0	62.0
50%	50.5	65.0	140.0	2.0	70.0
75%	75.25	88.0	195.0	3.0	85.0
max	100.0	95.0	210.0	4.0	100.0

Nota. Uso de herramienta de IA Julius ai para el Análisis de datos educativos con de los estudiantes de maestría. Fuente: <https://julius.ai/chat?id=314737f4>.

En las subsecciones siguientes se describen tres tipos de herramientas: hojas de cálculo para limpieza y análisis básico, plataformas de BI y dashboards educativos, y entornos de programación como Python en Google Colab. El objetivo es ofrecer un panorama realista de opciones que pueden combinarse en proyectos de distinta escala y madurez.

4.4.1. Hojas de cálculo para limpieza y análisis básico

Las hojas de cálculo, como Microsoft Excel o Google Sheets, siguen siendo la puerta de entrada más común al análisis de datos para docentes y equipos de gestión. En proyectos educativos permiten realizar tareas de limpieza básicas: filtrar registros, detectar valores faltantes, corregir errores tipográficos, y cálculos descriptivos, como promedios, porcentajes de aprobación o tablas dinámicas por grupo y asignatura. Manuales recientes sobre analítica educativa para docentes destacan el valor pedagógico de involucrar al profesorado en estas actividades, incluso cuando después se utilizarán herramientas más avanzadas.

En iniciativas de formación en ciencia de datos en educación se ha observado que trabajar con “datos desordenados” en hojas de cálculo ayuda a desarrollar competencias de pensamiento crítico sobre la calidad y las limitaciones de los datos disponibles.

Aunque las hojas de cálculo no sustituyen a entornos de programación o plataformas de BI, su accesibilidad y amplia difusión las convierten en un componente estratégico para la fase inicial de preparación y análisis descriptivo, así como para prototipar indicadores que luego se automatizarán en sistemas más robustos.

4.4.2. Plataformas de BI y dashboards educativos

Las plataformas de inteligencia de negocios, como Microsoft Power BI o Tableau, permiten conectar múltiples fuentes de datos, crear modelos semánticos, diseñar dashboards interactivos y compartirlos de forma segura con diferentes perfiles de usuario. Aunque surgieron en el ámbito empresarial, su uso en educación ha crecido para construir paneles de aprendizaje que integran información académica, de participación en LMS y de encuestas institucionales.

Investigaciones recientes proponen marcos que combinan BI y learning analytics para apoyar la gestión de programas en línea, evidenciando la capacidad de estas herramientas para facilitar análisis multidimensionales y monitoreo en tiempo real.

Además, estudios sobre dashboards en educación indican que, cuando se configuran adecuadamente, estas plataformas permiten incorporar analíticas predictivas básicas, filtros personalizados para docentes y alertas tempranas sobre estudiantes en riesgo.

Figura 15.

Representación y Análisis de datos en dashboards educativos.



Nota. Uso de herramienta de Poweer BI para la representación y el análisis de datos educativos.

Fuente: Power BI.

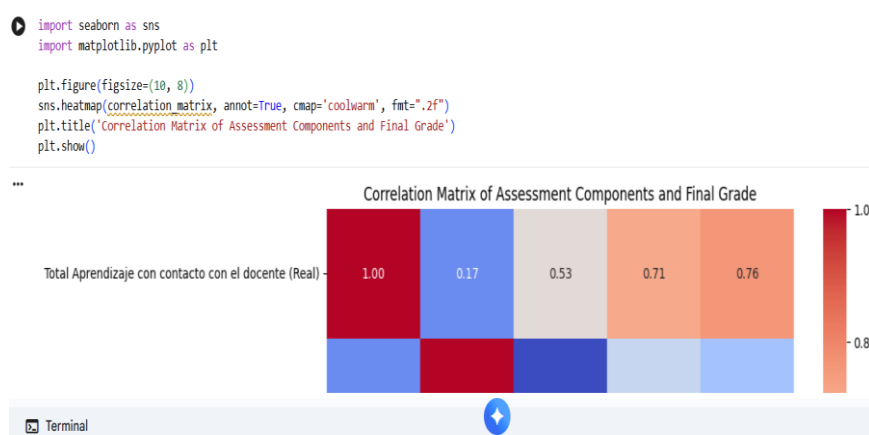
El desafío consiste en asegurar que el diseño de los dashboards (como el de la Figura 15), responda a preguntas pedagógicas concretas y que la comunidad educativa reciba formación para interpretar los paneles y utilizarlos de forma responsable.

4.4.3. Entornos de programación con IA

Para proyectos de mayor complejidad, especialmente aquellos que incluyen modelos de IA, análisis de texto o procesamiento de grandes volúmenes de datos, los entornos de programación resultan indispensables. Python en Google Colab (ver Figura 16), se ha consolidado como una opción accesible para educación, al permitir ejecutar código en la nube sin necesidad de instalar software local, compartir notebooks y documentar de forma integrada el flujo de análisis.

Figura 16.

Uso de herramienta de programación con IA para análisis de datos educativos.



Nota. Uso de Python con herramienta Google Colab para la representación y el análisis de datos educativos. Fuente: <https://colab.research.google.com/>.

La teoría y las Investigaciones recientes sobre técnicas y herramientas en learning analytics y minería de datos educativos señalan que lenguajes como Python y R ofrecen bibliotecas especializadas para preprocesamiento, modelado y visualización, además de integrarse con formatos abiertos y APIs de distintas plataformas educativas.

En la práctica, estos entornos son utilizados para desarrollar prototipos de modelos predictivos, experimentar con diferentes algoritmos y, posteriormente, integrarlos en dashboards o sistemas

institucionales. Dado que requieren mayor nivel de alfabetización en programación, su adopción suele concentrarse en equipos técnicos o grupos de investigación; sin embargo, constituyen un componente clave para avanzar hacia una analítica del aprendizaje basada en IA robusta y adaptable.

Las herramientas disponibles para la preparación, análisis y visualización de datos educativos conforman un ecosistema que va desde soluciones de entrada, como las hojas de cálculo, hasta entornos de programación que permiten implementar modelos avanzados de IA. La elección adecuada depende de los objetivos del proyecto, de las competencias del equipo y de la infraestructura institucional.

En todos los casos, resulta esencial que las herramientas seleccionadas apoyen procesos transparentes, reproducibles y alineados con las preguntas pedagógicas de interés. Estas consideraciones serán fundamentales cuando, en el capítulo 5, se aborden la construcción de modelos de IA específicos, indicadores e instrumentos de monitoreo continuo del aprendizaje.

Cierre del capítulo

Este capítulo ha mostrado que la preparación, el análisis y la visualización de datos educativos constituyen una fase intermedia crítica entre la recolección de datos y la construcción de modelos de IA e indicadores. La limpieza cuidadosa de los datos incluyendo el tratamiento de datos perdidos, outliers, inconsistencias y duplicados, no es simplemente un requisito técnico, sino una condición para producir diagnósticos justos y decisiones pedagógicas bien fundamentadas.

Del mismo modo, los métodos de análisis descriptivo y predictivo adquieren sentido en la medida en que sus resultados se interpretan a la luz de los objetivos formativos, de la diversidad del estudiantado y de las condiciones institucionales.

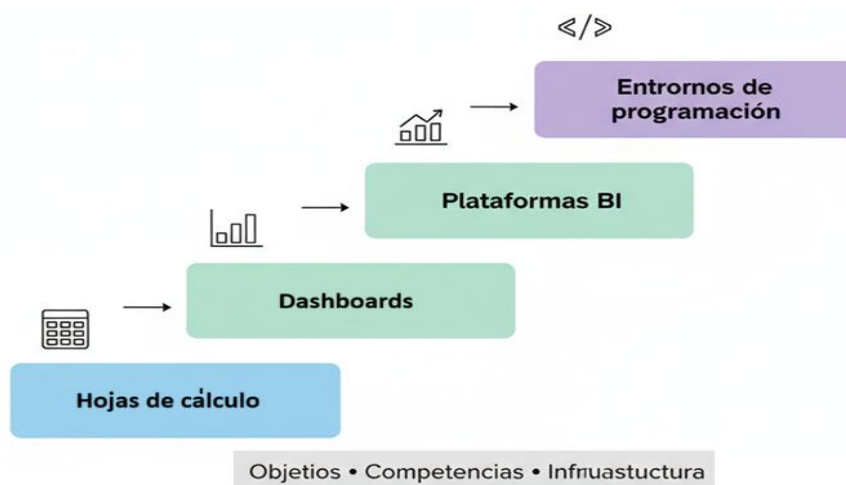
La visualización desempeña un papel central al hacer visibles patrones de rendimiento y participación que, de otro modo, permanecerían ocultos en bases de datos complejas. Los dashboards y paneles de analítica, diseñados con criterios claros y alineados con las necesidades

de docentes y gestores, se convierten en instrumentos de reflexión y acción educativa, más que en meros resúmenes numéricos.

Finalmente, la selección de herramientas, que van desde hojas de cálculo hasta plataformas de BI y entornos de programación, determina quién puede participar en el proceso analítico, qué tan reproducibles son los flujos de trabajo y cómo se integran los resultados en la gestión cotidiana de la enseñanza. Lo cual, se evidencia en la siguiente representación grafica que representa un entorno y espacio adecuado.

Figura 17.

Ecosistema de herramientas para analítica e IA: de hojas de cálculo a programación.



Nota. Escalera que ubica herramientas por complejidad/capacidad y por criterios de elección (objetivos, competencias, infraestructura).

De cara a los capítulos siguientes, este entramado de prácticas de limpieza, análisis y visualización constituye la base sobre la cual se construirán modelos de IA específicos, indicadores de seguimiento y métricas de desempeño a distintos niveles (estudiante, curso e institución). En particular, el capítulo 5 profundizará en cómo transformar los datos preparados en modelos predictivos e intervenciones automatizadas, así como en la definición de indicadores que permitan monitorear de forma continua la calidad del proceso de aprendizaje, siempre bajo criterios éticos y pedagógicos coherentes con el enfoque del libro.

CAPÍTULO V: MODELOS DE IA, INDICADORES Y MÉTRICAS PARA EL SEGUIMIENTO DEL APRENDIZAJE

“Lo que necesitamos es un grado de especificidad sobre lo que la AIEd nos permite evaluar, invertir, planificar, entregar y comprobar.” Barber.

Introducción del capítulo

Los capítulos 3 y 4 han mostrado que la analítica del aprendizaje depende de datos educativos de calidad y de técnicas sólidas de preparación, análisis y visualización, sin embargo, el ciclo analítico queda incompleto si los hallazgos no se transforman en modelos, indicadores y métricas que orienten decisiones pedagógicas concretas: qué estudiantes requieren apoyo, qué actividades deben ajustarse, qué evidencias sustentan la evaluación y qué tendencias informan la mejora de un curso o programa.

Este capítulo cierra ese ciclo al pasar de la descripción de “lo que ocurrió” a enfoques predictivos y prescriptivos, sin perder de vista la evaluación auténtica, la retroalimentación oportuna y el juicio profesional docente.

En la sección 5.1 se revisan modelos básicos de IA aplicados a datos educativos, con énfasis en clasificación (identificación de perfiles y riesgo) y regresión (predicción de rendimiento), incorporando una discusión crítica sobre sesgos y ética. En 5.2 se abordan sistemas de recomendación y apoyo a la decisión, entendidos como instrumentos para personalizar recursos y actividades, pero también como artefactos que deben ser interpretados y negociados pedagógicamente. La sección 5.3 profundiza en el diseño de indicadores de aprendizaje: qué hace “bueno” a un indicador, cómo distinguir indicadores de proceso, resultado e impacto y cómo alinearlos con competencias y resultados de aprendizaje.

Y Finalmente, 5.4 integra métricas, dashboards y evaluación continua, destacando cómo el seguimiento visual, la automatización de reportes y el análisis de tendencias pueden sostener ciclos de mejora a nivel de aula, curso y programa, con trazabilidad y responsabilidad.

5.1. Modelos básicos de IA aplicados a datos educativos

Los modelos básicos de IA permiten transformar registros educativos (LMS, evaluaciones, participación, trazas temporales) en señales interpretables para el seguimiento. Su valor no está en “predecir por predecir”, sino en activar decisiones: tutorías, rediseño de actividades, ajustes de evaluación y apoyos diferenciados.

En esta sección se presentan aplicaciones típicas de clasificación y regresión, y se discuten riesgos éticos y de sesgo que afectan la validez de las inferencias pedagógicas.

5.1.1. Modelos de clasificación para identificación de perfiles y riesgo

En el ámbito educativo, la clasificación consiste en asignar a cada estudiante una categoría operativa (por ejemplo, “en riesgo” o “no en riesgo”, “alta” o “baja autorregulación”) a partir de variables observables dentro y fuera del entorno virtual. Entre las más utilizadas se encuentran el cumplimiento de entregas, la actividad semanal en el aula virtual, los patrones de navegación (ingresos, tiempos de permanencia, recursos consultados) y el desempeño en tareas o cuestionarios.

En términos pedagógicos, estos modelos son especialmente valiosos cuando se integran en sistemas de alerta temprana, porque permiten detectar señales iniciales de rezago y actuar antes de que el problema se consolide. Sin embargo, su utilidad real depende de una condición clave: que el “riesgo” no se quede en una etiqueta, sino que se traduzca en acciones concretas de acompañamiento y enseñanza.

Para ilustrar mejor, por ejemplo, en un curso en línea, el sistema puede clasificar como riesgo alto a quienes no entregan actividades clave en las dos primeras semanas, no revisan las guías básicas o presentan resultados muy bajos en el primer cuestionario diagnóstico. Con esa información, el equipo docente puede activar intervenciones breves y realistas, como metas semanales pequeñas, recordatorios con sentido formativo (no punitivo), una tutoría corta de diez minutos para aclarar dudas, o la derivación a un recurso de apoyo específico. Un caso sencillo sería enviar a un estudiante un mensaje orientador del tipo: “Esta semana enfócate en completar la actividad 1 y

revisa el ejemplo resuelto; si te parece, coordinamos una tutoría breve para planificar tu calendario”.

Adicionalmente, investigaciones en el área, sugieren que estos sistemas pueden incorporarse a la docencia para intervenir a tiempo, especialmente en entornos en línea donde el compromiso varía considerablemente entre estudiantes (Bañeres et al., 2023). Ahora bien, la clasificación debe dialogar con el contexto para evitar interpretaciones erróneas. Un “bajo acceso” a la plataforma puede reflejar barreras tecnológicas (conectividad limitada, uso compartido del dispositivo), condiciones laborales, o incluso estrategias de estudio fuera de línea, como descargar materiales una sola vez y trabajar en cuaderno.

También, puede ocurrir lo contrario, donde un estudiante con muchos clics y tiempo de navegación podría estar perdido y repetir recursos sin comprenderlos. Por ello, el uso responsable exige validar las categorías con docentes, revisar umbrales y posibles falsos positivos, y combinar señales cuantitativas con evidencias de evaluación auténtica, como portafolios, proyectos aplicados, reflexiones breves o productos que muestren comprensión.

En suma, la clasificación aporta valor cuando se entiende como una guía para tomar decisiones pedagógicas informadas, sensibles al contexto y orientadas a mejorar el aprendizaje, no como un mecanismo automático de etiquetado.

5.1.2. Modelos de regresión para predicción de rendimiento académico

La regresión es una técnica de modelado que permite estimar valores continuos en educación, como la nota final esperada, la probabilidad de aprobar con una calificación específica, el progreso previsto en una unidad o incluso el nivel de desempeño que podría alcanzarse si se mantienen determinadas condiciones de estudio. A diferencia de la clasificación (que asigna categorías), la regresión resulta especialmente útil cuando el objetivo pedagógico exige anticipar magnitudes y no solo etiquetas, porque ofrece una visión más clara del desempeño esperado y facilita decisiones didácticas más ajustadas.

Por ejemplo, si un curso utiliza evaluaciones formativas semanales como cuestionarios cortos, ejercicios de práctica y participación en foros, un modelo de regresión puede estimar la nota final a partir de esos indicadores. Si la proyección muestra que un grupo de estudiantes se mantiene alrededor de un 10/20 hacia la semana 4 (primer mes), el docente puede ajustar la secuencia didáctica sin perder coherencia evaluativa, puede reforzar conceptos umbral (aquellos que abren aprendizajes posteriores), proponer práctica adicional con retroalimentación inmediata, reorganizar ejemplos y actividades para reducir carga cognitiva, o incluso redistribuir el peso de ciertas tareas siempre que se mantenga la validez de la evaluación.

De manera que, un ejemplo sencillo sería detectar que el bajo rendimiento se concentra en problemas de proporcionalidad: entonces se añade una guía breve con ejercicios graduados y una sesión de repaso, y se ofrece una actividad de recuperación formativa que no “infla” la nota, pero sí mejora la comprensión.

Ahora bien, en el ámbito educativo, la predicción no debe leerse como destino fijo, sino como una proyección condicionada, la cual cambia si cambian las oportunidades de aprendizaje, la calidad de la retroalimentación y los apoyos disponibles. Dicho de otro modo o palabras, si un estudiante recibe tutoría, mejora su planificación o accede a recursos adecuados, la estimación puede variar de forma significativa. Por eso, el valor pedagógico de la regresión no está en “etiquetar” el futuro, sino en ayudar a tomar decisiones tempranas para transformarlo.

Un reto clave es que los modelos se equivocan, y esos errores pueden ser pedagógicamente relevantes. Es frecuente que fallen más en estudiantes con trayectorias “atípicas”, por ejemplo, quienes participan poco al inicio, pero se recuperan después o con participación intermitente, como lo es la conectividad irregular, responsabilidades laborales o familiares). En entornos de aprendizaje en línea, se ha observado que comprender los errores algorítmicos (por qué se predice “no entrega” cuando sí ocurre, o por qué se anticipa bajo rendimiento pese a un repunte real) ayuda a los docentes a usar la predicción como señal de indagación, no como sentencia (Hlosta et al., 2022). Siendo así que por ejemplo, si el sistema predice baja nota por poca actividad en

plataforma, el docente puede verificar si el estudiante descarga materiales y trabaja fuera de línea, o si necesita apoyo para organizar tiempos.

Complementariamente, la literatura reciente advierte que la heterogeneidad de datos y contextos limita la transferencia de modelos entre instituciones. Un modelo entrenado en una universidad, con un LMS específico, ciertos perfiles de ingreso y prácticas docentes particulares, puede rendir mal en otra con dinámicas distintas. Esto refuerza la necesidad de validación local, ajuste de variables y revisión de supuestos antes de utilizar predicciones para decisiones académicas relevantes (Vaarma & Li, 2024).

A manera de resumen, la regresión aporta valor cuando se integra como herramienta de apoyo a la docencia: orienta intervenciones oportunas, promueve una evaluación formativa más inteligente y sostiene decisiones basadas en evidencia, siempre bajo una lectura contextual, prudente y pedagógicamente responsable.

5.1.3. Riesgos de sesgo y consideraciones éticas en modelos educativos

El riesgo ético central no es solo la privacidad, sino la injusticia inferencial: modelos que amplifican desigualdades al asociar “riesgo” con variables que funcionan como proxy de condición socioeconómica, idioma o acceso tecnológico.

En analíticas del aprendizaje, existe un componente de agencia que sostienen que estudiantes y docentes deben poder comprender, cuestionar y corregir decisiones apoyadas por IA, especialmente cuando afectan evaluaciones, becas o permanencia.

En multimodal learning analytics, se enfatiza que la equidad requiere examinar cómo se capturan y traducen señales (voz, gestos, interacción) y cómo esas señales se interpretan culturalmente (Jin et al., 2024).

Un punto crítico actualmente, es el consentimiento y la gobernanza del dato, de forma que la propensión a consentir analíticas no es homogénea y puede generar sesgos en quién queda representado en los modelos, afectando su validez y legitimidad institucional (Li et al., 2022). Por ello, el seguimiento del aprendizaje debe incorporar auditorías de sesgo, documentación del

modelo (propósito, límites, población) y mecanismos de explicación orientados a la acción pedagógica (Khosravi et al., 2022).

La clasificación y regresión aportan señales para actuar a tiempo, pero su valor educativo depende de cómo se integran a tutorías, rediseño didáctico y evaluación formativa. El seguimiento basado en IA exige validar localmente, interpretar errores, y sostener prácticas éticas (consentimiento, transparencia, equidad y rendición de cuentas). Sin estas condiciones, se puede establecer que la “precisión” técnica puede convertirse en decisiones pedagógicas injustas, indebidas y/o poco útiles.

5.2. Sistemas de recomendación y apoyo a la decisión

Los sistemas de recomendación desplazan la analítica del aprendizaje desde un enfoque principalmente diagnóstico (identificar qué ocurre) hacia uno orientado a la acción (qué conviene hacer a continuación), como qué recurso sugerir, qué actividad proponer, qué ruta alternativa ofrecer o qué apoyo activar. En educación, este tipo de recomendación no puede basarse únicamente en patrones de consumo (por ejemplo, “quienes vieron X también vieron Y”), porque el aprendizaje no se reduce a preferencias de navegación.

La recomendación debe sustentarse en criterios pedagógicos explícitos, tales como la alineación con los resultados de aprendizaje, la pertinencia disciplinar, el nivel de dificultad, la carga de trabajo, la progresión de contenidos y las condiciones reales del estudiante.

Por ejemplo, si un estudiante obtiene un desempeño bajo en un cuestionario de conceptos básicos, el sistema podría recomendarle un recurso breve y focalizado (una lectura guiada o un video corto con ejemplos) y, después, una actividad de práctica con retroalimentación inmediata. En cambio, si el estudiante ya domina los contenidos iniciales, la recomendación podría orientarse a una ruta de profundización, como un caso aplicado o un reto con problemas auténticos.

Otro caso sencillo, puede evidenciarse en un curso de escritura académica, si el sistema detecta dificultades recurrentes en citación, podría sugerir una micro actividad de 15 minutos con ejercicios de corrección y un modelo comentado, en lugar de enviar “más lecturas” sin dirección.

Ahora bien, para que estas recomendaciones sean responsables y efectivas, deben ser interpretables y criticables por docentes y estudiantes. El claustro docente necesita comprender por qué se recomienda un recurso (criterio y evidencia), para decidir si se ajusta al momento formativo; el estudiante, por su parte, debe poder elegir con autonomía informada, evitando que la recomendación se perciba como una imposición. En este sentido, esta sección revisa enfoques de recomendación, su uso para la personalización y la interpretación crítica por docentes y estudiantes.

5.2.1. Enfoques de recomendación en contextos educativos

En términos simples, recomendar es “seleccionar lo más adecuado” para un estudiante o grupo de ellos, bajo un objetivo formativo y meta esperada. Los enfoques más comunes incluyen:

1. Recomendaciones basadas en contenido de relación entre recurso y necesidad conceptual.
2. Filtrado colaborativo de preferencias de estudiantes similares.
3. Enfoques híbridos que combinan desempeño, trazas de actividad y requisitos curriculares.

La recomendación cobra sentido en educación, cuando se ancla a evidencias, como puede ser el caso de un estudiante con errores persistentes en fracciones recibe práctica graduada y guiada, con retroalimentación, mientras otro con buen dominio recibe problemas de transferencia o aplicación.

Estudios actuales como el de Khosravi et al. (2022). subrayan que la recomendación educativa debe diseñarse con sensibilidad a partes interesadas (stakeholders), como lo son estudiantes, docentes y gestores requieren explicaciones diferentes, porque su “decisión” también es distinta, y que estas a su vez pueden ser aceptar, adaptar, intervenir, rediseñar, o similares.

Además, el alineamiento con indicadores derivados del diseño instruccional ayuda a evitar recomendaciones descontextualizadas (Ahmad et al., 2022).

5.2.2. Recomendación de recursos y actividades personalizadas

La recomendación puede operar como extensión de la evaluación formativa: identificar una brecha y sugerir un recurso o actividad que la cierre. En la práctica, esto se traduce en “rutas” de apoyo (videos breves, lecturas guiadas, cuestionarios adaptativos, ejercicios de recuperación) o en “andamiajes” para tareas auténticas (rúbricas comentadas, ejemplos anotados, listas de verificación).

En programas de pregrado o posgrado, una recomendación pertinente puede ser un artículo metodológico específico cuando el indicador de proceso muestra baja calidad en la formulación del problema o en la argumentación.

Un riesgo frecuente es recomendar “más de lo mismo”, con recursos similares que aumentan carga sin mejorar comprensión, y que, para mitigarlo, se requiere combinar señales de desempeño con indicadores de proceso (p. ej., tiempo productivo, revisiones del borrador, uso de retroalimentación).

Los estudios en educación mediada por tecnologías como el de López-Regalado et al. (2024), indican que la IA aplicada a educación superior se vuelve más valiosa cuando conecta personalización con decisiones didácticas y no solo con automatización. Así, la recomendación debe incluir criterio pedagógico explícito y condiciones de aplicación.

5.2.3. Interpretación de recomendaciones por parte de docentes y estudiantes

La potencia de los sistemas de recomendación reside tanto en sus algoritmos como en la manera en que docentes y estudiantes interpretan sus sugerencias. La literatura en ética de la IA educativa subraya la importancia de que las recomendaciones sean explicables, es decir, que el sistema ofrezca razones comprensibles de por qué sugiere un recurso o una intervención concreta. Sin esa transparencia, existe el riesgo de que las recomendaciones se perciban como órdenes opacas o, por el contrario, como simples “decorados” sin valor real para la toma de decisiones.

Una recomendación educativa no es una orden, sino es una hipótesis de apoyo que debe interpretarse según contexto, evidencias y propósito evaluativo. Para el docente, interpretar

implica decidir si la recomendación es coherente con el plan del curso, si respeta la evaluación auténtica y si no introduce inequidades como, por ejemplo, exigir recursos externos no accesibles). Para el estudiante, interpretar implica comprender el “por qué” y el “para qué” de la sugerencia, y cómo se vincula con metas y criterios de logro.

En este ámbito, la explicabilidad de las recomendaciones es clave, debido a que cuando el sistema muestra qué evidencias sustentan la recomendación (errores frecuentes, baja participación en actividad crítica, tendencia de mejora estancada), se fortalece la agencia y se reduce el uso mecánico (Khosravi et al., 2022).

Además, la experiencia docente con analíticas predictivas sugiere que el valor práctico aumenta cuando las señales se presentan como insumo para conversación pedagógica como tutorías, feedback, refuerzos, etc., y no como vigilancia (Herodotou et al., 2023). En síntesis, interpretar recomendaciones requiere alfabetización de datos y acuerdos institucionales sobre límites y responsabilidades (Li et al., 2022).

En conjunto, los sistemas de recomendación representan una capa intermedia entre los modelos de IA y la práctica educativa cotidiana; de modo tal que, transforman predicciones y patrones en propuestas concretas de acción para estudiantes y docentes. Su diseño y uso consciente permiten avanzar hacia experiencias de aprendizaje más personalizadas, siempre y cuando se mantenga el equilibrio entre automatización, apoyo humano y reflexión pedagógica, a razón de que este equilibrio será igualmente central en el diseño de indicadores de aprendizaje, tema de la sección siguiente.

5.3. Diseño de indicadores de aprendizaje

Los modelos de IA y los sistemas de recomendación aportan valor educativo solo cuando sus salidas se traducen en indicadores comprensibles y accionables para docentes y estudiantes. En analítica del aprendizaje, un indicador es una representación cuantitativa o cualitativa de aspectos relevantes del proceso o de los resultados educativos, construida a partir de datos recopilados de manera sistemática y con criterios explícitos (ver Figura 18).

En otras palabras, los indicadores convierten datos dispersos en información interpretable y, ayudan a observar patrones, monitorear avances y justificar decisiones pedagógicas con mayor consistencia.

Por ello, los indicadores funcionan como el puente entre datos y decisiones, ya que permiten “nombrar” lo que importa para el aprendizaje y seguirlo en el tiempo. Ahora bien, un indicador mal definido puede conducir a intervenciones equivocadas, incluso si el modelo subyacente es sofisticado. Por ejemplo, si se usa “tiempo de conexión” como sinónimo de “compromiso”, podría clasificarse como “altamente comprometido” a un estudiante que mantiene la plataforma abierta sin estudiar, o como “poco comprometido” a otro que descarga materiales y estudia fuera de línea.

Por el contrario, un indicador más robusto podría combinar evidencia de progreso, como “porcentaje de actividades clave completadas a tiempo” y “calidad de respuestas en tareas aplicadas”. Asimismo, los indicadores deben formularse con propiedades de calidad: claridad (que cualquiera pueda entenderlo), pertinencia (que mida algo relevante), validez (que refleje el constructo), confiabilidad (que sea estable), oportunidad (que llegue a tiempo) y accionabilidad (que sugiera qué hacer).

En esta sección se desarrollan tipologías de indicadores, distinguiendo entre indicadores de proceso (p. ej., regularidad de estudio semanal), de resultado (p. ej., logro en evaluaciones alineadas) y de impacto (p. ej., mejora sostenida en desempeño o permanencia). Finalmente, se enfatiza la alineación de indicadores con competencias y resultados de aprendizaje, de modo que cada métrica oriente intervenciones coherentes con lo que el curso declara formar.

5.3.1. Concepto y propiedades de calidad de los indicadores educativos

Un indicador educativo es una medida operacional que representa un constructo relevante (participación significativa, avance conceptual, calidad argumentativa) y que permite monitorear cambios a lo largo del tiempo.

La calidad de un indicador se evalúa por su claridad (definición no ambigua), pertinencia (relación con objetivos), viabilidad (datos disponibles y costo razonable), sensibilidad (detecta cambios

reales) y utilidad (sugiere decisiones posibles). Una revisión sobre indicadores desde la perspectiva del diseño instruccional enfatiza que los indicadores más útiles son los que se vinculan explícitamente a eventos del diseño (actividad, propósito, evidencia esperada) y no solo a “clicks” o conteos (Ahmad et al., 2022)

En el campo de la educación, un indicador educativo es una medida que sintetiza información sobre el aprendizaje o sobre las condiciones en que este ocurre. A diferencia de un dato aislado, por ejemplo, “12 accesos al aula virtual”. De modo que un indicador supone cierto procesamiento e interpretación, como “porcentaje de estudiantes que accede al menos tres veces por semana al aula virtual”. Los indicadores permiten resumir tendencias, establecer comparaciones y monitorear cambios a lo largo del tiempo.

Para ser útiles, los indicadores deben cumplir al menos cuatro propiedades de calidad.

1. Validez, el indicador debe representar de manera razonable el constructo que pretende medir (participación, logro, abandono).
2. Fiabilidad, producir resultados consistentes cuando se calcula en condiciones similares.
3. Interpretabilidad, ser comprensible para sus usuarios naturales (docentes, gestores, estudiantes).
4. Sensibilidad y viabilidad: ser capaz de detectar cambios relevantes sin exigir esfuerzos desproporcionados de recolección o cómputo.

Los marcos de indicadores para analítica del aprendizaje insisten, además, en la alineación con objetivos pedagógicos explícitos y en la consideración de efectos inesperados, como la sobre focalización en aquello que es fácil de medir. Además, la validez del indicador se fortalece cuando se integra con la evaluación, de forma que, si el curso declara aprendizaje centrado en el estudiante, entonces los indicadores deben reflejar evidencias auténticas y coherentes con la alineación constructiva (Divjak et al., 2023). En otras palabras, no basta medir, sino que hay que medir lo que respalda decisiones formativas justas.

5.3.2. Indicadores de proceso, resultado e impacto

Los indicadores pueden organizarse por niveles tal como se muestra en la gráfica 18.

Figura 18.

Indicadores de aprendizaje: proceso, resultado e impacto.



Nota. Taxonomía de indicadores: proceso (cómo se aprende), resultado (logros), impacto (efectos amplios) para evitar confusiones habituales.

Ahora bien, estos niveles se explican de manera siguiente:

- *Los indicadores de proceso*, describen cómo se aprende, ritmo de avance, uso de retroalimentación, participación en discusiones, revisiones de un borrador. Son útiles para intervenir a tiempo, porque apuntan a prácticas y estrategias.
- Los indicadores de resultado, capturan logros observables, dominio de un contenido, desempeño en una rúbrica, logro de una competencia).
- Los indicadores de impacto evalúan efectos más amplios y tardíos, retención, progresión, desempeño en prácticas profesionales, continuidad en el programa.

Esta tipología ayuda a evitar confusiones comunes, como por ejemplo: un “alto tiempo en plataforma” no es impacto, y una “nota final” no explica el proceso. En experiencias de analítica mediada por estrategias pedagógicas, se ha mostrado que integrar indicadores de proceso con capacidades docentes (competencias digitales y toma de decisiones basada en datos), favorece ciclos de mejora continua (Brito Romero et al., 2025).

A nivel institucional, el desafío es construir paneles donde el proceso, resultado e impacto se articulen sin reducir el aprendizaje a un único número (López-Regalado et al., 2024).

5.3.3. Alineación de indicadores con resultados de aprendizaje y competencias

Diseñar indicadores sin una referencia clara a los resultados de aprendizaje y las competencias del currículo conduce con frecuencia a métricas interesantes, pero pedagógicamente irrelevantes. El principio de “alineación constructiva”, formulado originalmente en el ámbito del diseño curricular, cobra nueva fuerza en la analítica del aprendizaje, lo que se mide, lo que se enseña y lo que se espera que los estudiantes aprendan deben estar coherentemente articulados. En usos recientes de learning analytics centradas en la autorregulación se ha mostrado cómo es posible definir indicadores que mapean directamente a procesos como planificación, monitoreo o reflexión, evitando limitarse a contar clics o tiempos de conexión.

En la práctica, esto implica traducir cada resultado de aprendizaje en uno o varios indicadores observables. Alinear indicadores con resultados y competencias significa que cada métrica debe responder las siguientes interrogantes: ¿qué evidencia demuestra el logro? y ¿qué acción pedagógica habilita?. Por ejemplo, si una competencia exige argumentación con evidencia, el indicador no puede ser solo “número de intervenciones en foro”; debe capturar calidad (coherencia, citas, contraargumentos) y progreso a lo largo del curso. La alineación también reduce sesgos, porque obliga a justificar por qué un dato representa un aprendizaje y no una condición externa.

Un aporte reciente es integrar analítica con validez de la evaluación, que permitan comparar el peso planificado de resultados de aprendizaje con el peso real en tareas y calificaciones, y usar analítica para sostener ajustes iterativos del diseño (Divjak et al., 2023).

De manera paralela, revisiones sobre indicadores basados en diseño instruccional proponen mapear eventos bajo el siguiente orden: actividad → evidencia → indicador → decisión, para asegurar interpretabilidad y uso formativo (Ahmad et al., 2022).

Así, la analítica se vuelve un componente de la evaluación auténtica y no un sistema paralelo de control. En síntesis, diseñar indicadores es decidir qué “cuenta” como evidencia de aprendizaje. Indicadores de calidad son claros, viables, sensibles y accionables, y su tipología (proceso–

resultado–impacto) ayuda a intervenir sin confundir señales. La alineación con competencias y validez evaluativa convierte las métricas en soporte de decisiones justas y pedagógicamente significativas, sosteniendo el seguimiento del aprendizaje con trazabilidad.

La sección ha mostrado que los indicadores de aprendizaje actúan como puente entre los datos brutos y las decisiones pedagógicas. Su diseño requiere claridad conceptual, criterios de calidad y alineación con resultados y competencias. En los apartados siguientes, veremos cómo estos indicadores se concretan en métricas específicas, se representan mediante dashboards y se integran en esquemas de evaluación continua que buscan mejorar de forma sistemática la enseñanza y el aprendizaje.

5.4. Métricas, dashboards y evaluación continua

Si los indicadores definen “qué” se quiere observar, las métricas y los dashboards especifican “cómo” se mide y “cómo” se presenta la información a quienes deben actuar sobre ella. Las métricas y dashboards materializan el seguimiento en prácticas sostenibles, visualizan señales, estructuran conversaciones de retroalimentación y permiten observar tendencias.

En educación, su finalidad no es “vigilar”, sino apoyar evaluación continua, tutorías y mejora del diseño. Esta sección integra métricas cuantitativas y cualitativas, criterios de diseño de dashboards y automatización de reportes para ciclos de mejora.

5.4.1. Métricas cuantitativas y cualitativas para evaluar el aprendizaje

Las métricas cuantitativas como tasas de entrega, progresión, aprobaciones, regularidad, y predicciones, ofrecen comparabilidad y detección temprana, pero suelen requerir contexto para no generar interpretaciones simplistas. Las métricas cualitativas como la calidad de productos, reflexiones, análisis de portafolios, evidencia en rúbricas, son más cercanas a la evaluación auténtica, aunque demandan acuerdos de codificación y criterios consistentes.

Una recomendación práctica es diseñar paquetes de métricas que permitan combinar un indicador de proceso (uso de retroalimentación), uno de resultado (rúbrica) y uno de bienestar/condición (participación sostenida) para orientar intervenciones. Las investigaciones actuales sobre

dashboards, informan que la utilidad aumenta cuando las métricas se vinculan a decisiones pedagógicas explícitas como qué ajustar, a quién apoyar o cuándo intervenir, y cuando se evita la saturación de indicadores sin propósito (Paulsen & Lindsay, 2024).

En educación superior, además, se propone que la IA aporta valor cuando se integra a estrategias de evaluación y acompañamiento, no como reemplazo del criterio docente (López-Regalado et al., 2024).

5.4.2. Diseño y uso de dashboards para el seguimiento del aprendizaje

Un dashboard educativo efectivo y bien diseñado debe responder a las siguientes tres preguntas:

- ¿Qué está pasando?
- ¿Por qué podría estar pasando?
- ¿Qué podemos hacer ahora?

Esto implica elegir visualizaciones alineadas con decisiones que incluyen mapas de progreso por competencia, alertas interpretables, distribución de desempeño por actividad crítica, y trazas temporales para reconocer cambios.

La experiencia y los estudios en analíticas demuestran que muchos dashboards fracasan por centrarse en métricas fáciles (clics) y no en señales pedagógicamente accionables; además, el rol del docente se vuelve clave para traducir visualización en intervención (Paulsen & Lindsay, 2024).

Para sostener evaluación continua, es útil diseñar vistas diferenciadas, ejemplo de ella, una vista docente con riesgo, progreso por resultados, comparativos por cohorte) y una vista estudiantil (metas, evidencias, recomendaciones explicadas). La explicabilidad también aplica aquí: mostrar evidencias y límites del dato fortalece confianza y reduce decisiones automáticas (Khosravi et al., 2022).

A nivel de educación de posgrado, los dashboards pueden apoyar evaluación auténtica al mostrar progreso en portafolios, calidad de borradores y uso de feedback, facilitando tutorías focalizadas.

5.4.3. Automatización de reportes y análisis de tendencias para la mejora continua

La automatización de reportes (semanales, por unidad, por competencia), libera tiempo docente y mejora consistencia, pero solo es educativa si se orienta a acciones de mejora como por ejemplo ajustes de secuencia, rediseño de evaluación, apoyos diferenciados y retroalimentación oportuna.

Un enfoque útil es trabajar con “tendencias” que incluyen evolución del progreso por competencia, persistencia de errores clave, o cambios en participación tras una intervención. Aquí, el valor está en la comparación temporal y no en el ranking estático.

En contextos latinoamericanos, se proponen estrategias metodológicas donde la analítica se integra a la mejora continua del proceso de enseñanza-aprendizaje, destacando la necesidad de competencias docentes para interpretar reportes y convertirlos en decisiones (Brito Romero et al., 2025).

Complementariamente, estudios sobre validez y analítica sugiere que los reportes deben dialogar con la evaluación, si las tendencias muestran desalineación entre resultados y tareas, se ajusta el diseño con evidencia (Divjak et al., 2023). Así, la automatización no reemplaza el juicio profesional, y lo estructura con datos trazables y oportunos. Las métricas y los dashboards convierten datos en seguimiento visible, siempre que combinen evidencia cuantitativa y cualitativa y se vinculen a decisiones.

El diseño debe priorizar accionabilidad y explicabilidad, evitando paneles “decorativos”, y la automatización de reportes y análisis de tendencias, fortalecen ciclos de mejora continua, pero su impacto depende de competencias docentes y de la integración con la evaluación y el rediseño pedagógico.

Cierre del capítulo

Este capítulo mostró que el seguimiento del aprendizaje basado en IA requiere una arquitectura coherente de modelos, indicadores, métricas y dashboards que se articulen con evaluación y decisiones pedagógicas. Los modelos de clasificación y regresión aportan señales para intervenir

tempranamente, pero su valor real emerge cuando se validan localmente, se interpretan con cautela y se traducen en acciones de apoyo.

Los sistemas de recomendación amplían la analítica hacia decisiones prescriptivas, aunque su legitimidad educativa depende de explicaciones comprensibles, de criterios pedagógicos explícitos y de la agencia de docentes y estudiantes. En paralelo, el diseño de indicadores establece qué evidencias importan, debido a que, sin indicadores válidos y alineados con competencias, las métricas pueden convertirse en ruido o en control improductivo.

Figura 19.

Dashboards y mejora continua: preguntas clave, reportes y tendencias.



Nota. Representación que de automatización de reportes y análisis de tendencias como soporte de ciclos de mejora con trazabilidad.

Para finalizar, el capítulo recalcó un equilibrio imprescindible de la automatización con responsabilidad. La automatización de reportes y paneles puede mejorar oportunidad y consistencia (tal como se observa en la Figura 19), pero el juicio profesional de docente sigue siendo insustituible para contextualizar datos, reconocer excepciones, sostener evaluaciones y tratar con equidad.

En conclusión, las revisiones y estudios investigativos actuales, apuntan a sistemas más explicables, sensibles a contexto y capaces de integrar evidencias multimodales sin amplificar sesgos. Además, persisten desafíos de gobernanza y consentimiento, auditorías de equidad,

transferencia entre contextos, y formación docente en alfabetización de datos. Finalmente, la analítica del aprendizaje apoyada por IA solo “mejora” educación cuando se integra en ciclos reflexivos de enseñanza, evaluación y rediseño, con trazabilidad y un compromiso explícito con la justicia educativa.

REFERENCIAS

- Ahmad, A., Schneider, J., Griffiths, D., Biedermann, D., Schiffner, D., Greller, W., & Drachsler, H. (2022). Connecting the dots – A literature review on learning analytics indicators from a learning design perspective. *Journal of Computer Assisted Learning*. Advance online publication. <https://doi.org/10.1111/jcal.12716>
- Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
- Banihashem, S. K., Khalil, M., Hatala, M., & Echeverria, V. (2022). A systematic review of the role of learning analytics in supporting feedback. *Computers & Education*, 184, 104501. <https://doi.org/10.1016/j.compedu.2022.104501>
- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489. <https://doi.org/10.1016/j.edurev.2022.100489>
- Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A. E., & Cortadas, P. (2023). An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, 20, Article 3. <https://doi.org/10.1186/s41239-022-00371-5>
- Barana, A., Marchisio, M., & Sacchet, M. (2019). Learning analytics to improve formative assessment in an online environment. *Journal of e-Learning and Knowledge Society*, 15(3), 131–146.
- Brito Romero, H. P., Erraez Alvarado, J. L., López Fernández, R., & Sánchez Gálvez, S. (2025). Estrategia metodológica mediada por la analítica del aprendizaje para las competencias docentes en las tecnologías de la información y comunicación. *Revista Científica Internacional Arandu UTIC*, 12(2), 4090–4104. <https://doi.org/10.69639/arandu.v12i2.1216>

- Divjak, B., Svetec, B., Horvat, D., & Kadoić, N. (2023). Assessment validity and learning analytics as prerequisites for ensuring student-centred learning design. *British Journal of Educational Technology*, 54(1), 313–334. <https://doi.org/10.1111/bjet.13290>
- Ferguson, R., Clow, D., Griffiths, D., & Brasher, A. (2019). Moving forward with learning analytics: Expert views. *Journal of Learning Analytics*, 6(3), 43–59. <https://doi.org/10.18608/jla.2019.63.8>
- Reyes, J. A. (2015). The skinny on big data in education: Learning analytics simplified. *TechTrends*, 59(2), 75–80. <https://doi.org/10.1007/s11528-015-0842-1>
- Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic review of artificial intelligence in education: Trends, benefits, and challenges. *Multimodal Technologies and Interaction*, 9(8), 84. <https://doi.org/10.3390/mti9080084>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>
- Guerrero-Sosa, J. D. T., Romero, F. P., Menéndez-Domínguez, V. H., Serrano-Guerrero, J., Montoro-Montarroso, A., & Olivas, J. A. (2025). A comprehensive review of multimodal analysis in education. *Applied Sciences*, 15(11), 5896. <https://doi.org/10.3390/app15115896>
- Hariyanto, H., Kristianingsih, F. X. D., & Maharani, R. (2025). Artificial intelligence in adaptive education: A systematic review of techniques for personalized learning. *Discover Education*, 4, 458. <https://doi.org/10.1007/s44217-025-00908-6>
- Hernández-García, Á., Cuenca-Enrique, C., Del-Río-Carazo, L., & Iglesias-Pradas, S. (2024). Exploring the relationship between LMS interactions and academic performance: A learning cycle approach. *Computers in Human Behavior*, 155, 108183. <https://doi.org/10.1016/j.chb.2024.108183>
- Herodotou, C., Maguire, C., Hlosta, M., & Mulholland, P. (2023). Predictive learning analytics and university teachers: Usage and perceptions three years post implementation. In *Proceedings of the 13th International Learning Analytics and Knowledge Conference (LAK '23)* (pp. 68–78). ACM. <https://doi.org/10.1145/3576050.3576061>
- Hlosta, M., Herodotou, C., Papathoma, T., Gillespie, A., & Bergamin, P. (2022). Predictive learning analytics in online education: A deeper understanding through explaining algorithmic errors. *Computers and Education: Artificial Intelligence*, 3, 100108. <https://doi.org/10.1016/j.caeai.2022.100108>
- Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

- Jin, Y., Echeverria, V., Yan, L., Zhao, L., Alfredo, R., Tsai, Y.-S., Gašević, D., & Martinez-Maldonado, R. (2024). FATE in MMLA: A student-centred exploration of fairness, accountability, transparency, and ethics in multimodal learning analytics. *Journal of Learning Analytics*, 11(3), 6–23. <https://doi.org/10.18608/jla.2024.8351>
- Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Lang, C., Siemens, G., Wise, A. F., & Merceron, A. (Eds.). (2022). *Handbook of learning analytics* (2.^a ed.). SoLAR. <https://doi.org/10.18608/hla22>
- Li, W., Sun, K., Schaub, F., & Brooks, C. (2022). Disparities in students' propensity to consent to learning analytics. *International Journal of Artificial Intelligence in Education*, 32(3), 564–608. <https://doi.org/10.1007/s40593-021-00254-2>
- Lin, Y., Huang, C., & Lu, O. H. T. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments*, 10, 41. <https://doi.org/10.1186/s40561-023-00260-y>
- Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 31–40.
- López-Regalado, M. Á., Núñez-Rojas, N., López-Gil, M., & Sánchez-Rodríguez, J. (2024). Inteligencia artificial en la educación universitaria: Una revisión sistemática. *Pixel-Bit. Revista de Medios y Educación*, 70, 97–122. <https://doi.org/10.12795/pixelbit.106336>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- Márquez, L., Henriquez, V. A., Chevreux, H., & Guerra, J. (2024). Adoption of learning analytics in higher education institutions: A systematic literature review. *British Journal of Educational Technology*, 55(2), 439–459. <https://doi.org/10.1111/bjet.13385>

- Miao, F., & Holmes, W. (2021). AI and education: Guidance for policy-makers. UNESCO.
- Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). AI and education: Guidance for policy-makers. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000376709>
- Mpolomoka, D. L. (2025). Utilizing artificial intelligence for assessment in higher education. *Pedagogical Research*, 10(3), em0243. <https://doi.org/10.29333/pr/16677>
- Ngulube, P. (2025). Leveraging learning analytics to improve the user experience within learning management systems in higher education institutions: A systematic review. *Information*, 16(5), 419. <https://www.mdpi.com/2078-2489/16/5/419>
- Palanci, A., Yilmaz, R. M., & Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*, 29(17), 22629–22650. <https://doi.org/10.1007/s10639-024-12737-5>
- Pan, Z., Lin, C. H., & Huang, X. (2024). A systematic review of learning analytics–incorporated instructional interventions on learning management systems. *Journal of Learning Analytics*, 11(1), 1–26. <https://doi.org/10.18608/jla.2024.8093>
- Papadogiannis, I., Wallace, M., & Karountzou, G. (2024). Educational data mining: A foundational overview. *Encyclopedia*, 4(4), 1644–1664. <https://doi.org/10.3390/encyclopedia4040108>
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49–64.
- Paulsen, L., & Lindsay, E. (2024). Learning analytics dashboards are increasingly becoming about learning and not just analytics: A systematic review. *Education and Information Technologies*, 29(11), 14279–14308. <https://doi.org/10.1007/s10639-023-12401-4>
- Ruidias, R. R. S., Nunes, B. P., Manrique, R., & Siqueira, S. (2025). Assessing data landscapes for quality education in Latin America: A FAIRness perspective on Brazil, Colombia, and Peru. *Journal of Learning Analytics*, 12(2), 175–195. <https://doi.org/10.18608/jla.2025.8441>
- Russell, S. J., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.). Pearson.
- Tsai, Y.-S., & Gašević, D. (2017). Learning analytics in higher education—Challenges and policies: A review of eight learning analytics policies. In *Proceedings of the 7th International Conference on Learning Analytics and Knowledge* (pp. 233–242). ACM. <https://doi.org/10.1145/3027385.3027400>

- Vaarma, M., & Li, M. (2024). Using machine learning to predict student dropout: A systematic review. *Technology in Society*, 76, 102557. <https://doi.org/10.1016/j.techsoc.2024.102557>
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98–110. <https://doi.org/10.1016/j.chb.2018.07.027>
- Vorobyeva, K. I., Belous, S., Savchenko, N. V., Smirnova, L. M., Nikitina, S. A., & Zhdanov, S. P. (2025). Personalized learning through AI: Pedagogical approaches and critical insights. *Contemporary Educational Technology*, 17(2), ep574. <https://doi.org/10.30935/cedtech/16108>
- Wang, J., Lin, S., & Zhang, Y. (2023). Overview of data quality: Examining the dimensions, antecedents, and outcomes. *Journal of the American Medical Informatics Association*, 30(4), 810–822. <https://doi.org/10.1093/jamia/ocac230>
- Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S., & Aggarwal, S. (2024). Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality. *Journal of the Knowledge Economy*, 15(1), 1159–1178. <https://doi.org/10.1007/s13132-022-01096-6>
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252, 124167. <https://doi.org/10.1016/j.eswa.2024.124167>
- Weidlich, J., Fink, A., Frey, A., Jivet, I., Gombert, S., Menzel, L., Giorgashvili, T., Yau, J., & Drachsler, H. (2025). Highly informative feedback using learning analytics: How feedback literacy moderates student perceptions of feedback. *International Journal of Educational Technology in Higher Education*, 22(1), 43. <https://doi.org/10.1186/s41239-025-00539-9>
- Winne, P. H. (2020). Construct and consequential validity for learning analytics. *Computers in Human Behavior*, 112, 106457. <https://doi.org/10.1016/j.chb.2020.106457>
- Wong, B. T. M., Cheon, L. k., Liu, M. (2025). The role of learning analytics in evaluating course effectiveness in higher education: A systematic review. *Sustainability*, 17(2), 559. <https://doi.org/10.3390/su17020559>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International*

Journal of Educational Technology in Higher Education, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>



ISBN: 978-9907-809-00-8

