

**Ciencia de datos y metodologías pedagógicas en educación superior: Revisión sistemática
de la literatura 2019–2025**

Juan Manuel Oviedo Gómez

Asesor

Nidia Danigza Lugo López

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica
2025

Resumen

Esta investigación analiza cómo la Ciencia de Datos está transformando las metodologías pedagógicas en la educación superior, mediante una Revisión Sistemática de Literatura Científica publicada principalmente entre 2019 y 2025, aunque se consideraron algunos referentes fundamentales anteriores. Se aplicó un enfoque primordialmente cualitativo para la síntesis de los hallazgos, con el objetivo general de sintetizar la integración de herramientas de Ciencia de Datos en las metodologías pedagógicas de la educación superior, identificando tendencias, vacíos y oportunidades de innovación educativa basada en datos.

El estudio identificó que la implementación de la Ciencia de Datos en la educación superior está influenciada por factores contextuales críticos que van más allá de la infraestructura tecnológica. Las barreras éticas, como la falta de protocolos claros para el consentimiento informado y la gestión de la privacidad, junto con el riesgo de sesgos algorítmicos, emergen como obstáculos significativos. Asimismo, la insuficiente capacitación docente en la interpretación pedagógica y el uso ético de los datos, y la rigidez de algunas estructuras institucionales, limitan la adopción efectiva y equitativa de estas tecnologías. Se observó también una concentración geográfica y temática en la producción científica, subrayando la necesidad de desarrollos más contextualizados y diversificados.

La evaluación de beneficios y riesgos reveló que la Ciencia de Datos ofrece un potencial considerable para personalizar el aprendizaje, predecir el rendimiento estudiantil e intervenir tempranamente para mejorar la retención y el éxito académico. Sin embargo, estos beneficios conllevan riesgos como la posible estigmatización de estudiantes, la simplificación de perfiles de aprendizaje y la perpetuación de desigualdades si los algoritmos no se diseñan y auditan con criterios de equidad y transparencia.

Se concluye que, para una implementación ética, contextualizada e inclusiva, es fundamental que las instituciones inviertan en el desarrollo profesional docente, establezcan marcos robustos de gobernanza de datos y fomenten una cultura de innovación responsable. A partir de estos hallazgos, la investigación propone recomendaciones estratégicas para instituciones de educación superior, formuladores de políticas, docentes e investigadores, orientadas a superar las barreras identificadas, asegurar una implementación alineada con los marcos normativos vigentes, potenciar los beneficios de forma equitativa y guiar futuras investigaciones hacia metodologías que capturen el impacto cualitativo de estas herramientas en el aprendizaje y exploren su aplicación en una mayor diversidad de disciplinas y contextos culturales.

En definitiva, la Ciencia de Datos representa una oportunidad transformadora para la educación superior, siempre que su integración se guíe por una visión pedagógica clara, un compromiso ético firme y un enfoque centrado en el desarrollo integral y equitativo del estudiante.

Palabras clave: Ciencia de Datos, Metodologías Pedagógicas, Educación Superior, Revisión Sistemática de Literatura, Analítica del Aprendizaje, Personalización del Aprendizaje, Ética de Datos, Innovación Educativa.

Abstract

This research analyzes how Data Science is transforming pedagogical methodologies in higher education through a Systematic Literature Review of scientific publications, primarily from 2019 to 2025, although some foundational earlier references were considered. A predominantly qualitative approach was applied for the synthesis of findings, with the general objective of synthesizing the integration of Data Science tools into pedagogical methodologies in higher education, identifying trends, gaps, and opportunities for data-driven educational innovation.

The study identified that the implementation of Data Science in higher education is influenced by critical contextual factors that extend beyond technological infrastructure. Ethical barriers, such as the lack of clear protocols for informed consent and student data privacy management, along with the risk of unaudited algorithmic biases, emerge as significant obstacles. Likewise, insufficient teacher training in the pedagogical interpretation and ethical use of data, and the rigidity of some institutional structures, limit the effective and equitable adoption of these technologies. A geographical and thematic concentration in scientific production was also observed, underscoring the need for more contextualized and diversified developments.

The assessment of benefits and risks revealed that Data Science offers considerable potential for personalizing learning, predicting student performance, and intervening early to improve retention and academic success. However, these benefits entail risks such as the possible stigmatization of students, oversimplification of learning profiles, and the perpetuation of inequalities if algorithms are not designed and audited with criteria of equity and transparency.

It is concluded that for an ethical, contextualized, and inclusive implementation, it is fundamental for institutions to invest in teacher professional development, establish robust data governance frameworks, and foster a culture of responsible innovation. Based on these findings, the research proposes strategic recommendations for higher education institutions, policymakers, educators, and researchers, aimed at overcoming the identified barriers, ensuring implementation aligned with current regulatory frameworks, equitably enhancing the benefits, and guiding future research towards methodologies that capture the qualitative impact of these tools on learning and explore their application in a wider diversity of disciplines and cultural contexts.

Ultimately, Data Science represents a transformative opportunity for higher education, provided that its integration is guided by a clear pedagogical vision, a firm ethical commitment, and an approach centered on the integral and equitable development of the student.

Keywords: Data Science, Pedagogical Methodologies, Higher Education, Systematic Literature Review, Learning Analytics, Personalized Learning, Data Ethics, Educational Innovation.

Tabla de Contenido

Introducción	9
Justificación	12
Planteamiento del Problema	18
Objetivos	22
Objetivo General.....	22
Objetivos Específicos	22
Marco Teórico.....	23
Metodología	28
Tipo y Enfoque de Investigación.....	28
Método: Revisión Sistemática de Literatura	28
Pregunta de Investigación y Estrategia de Búsqueda	29
Criterios de Selección y Proceso de Revisión	29
Extracción, Análisis y Síntesis de Datos	30
Consideraciones Éticas	31
Resultados	32
Factores Contextuales en la Implementación de la Ciencia de Datos en Educación Superior..	32
Distribución Geográfica y Foco Temático de la Investigación	34
Hacia una Implementación Pedagógica, Ética e Inclusiva: Consideraciones Emergentes de la Literatura	39
Conclusiones	42
Recomendaciones	45
Referencias Bibliográficas	57

Lista de Tablas

Tabla 1 *Aplicaciones de la Ciencia de Datos en la Educación Superior y Objetivos*

Correspondientes 26

Tabla 2 *Aplicaciones Clave de la Ciencia de Datos en Educación Superior y Autores Relevantes*

..... 37

Lista de Figuras

Figura 1 *Distribución Geográfica de Estudios en IA en Educación Superior* 35

Figura 2 *Áreas del Conocimiento en IA en Educación Superior Porcentaje de Publicaciones* .. 36

Introducción

La presente investigación busca contribuir a la generación de conocimiento sobre la Ciencia de Datos y su aplicación, aporte e impacto en las metodologías pedagógicas dentro de contextos universitarios, enfocándose específicamente en los espacios de formación en educación superior. En un escenario donde la tecnología educativa ha demostrado ser un componente crucial para el logro de los fines de la educación (Torres Cañizález & Cobo Beltrán, 2017) y cuya evolución ha sido una constante (García-Valcárcel, A. 2002), la Ciencia de Datos emerge como una fuerza transformadora. El objeto de estudio de esta investigación se centra en la educación superior y el proceso de transformación generado por este fenómeno global que está reconfigurando la sociedad y múltiples procesos, tanto en el ámbito público como privado, con un interés particular en la prestación de servicios educativos. La innovación educativa, aunque enfrenta oportunidades y barreras (Michavila, 2009), encuentra en la Ciencia de Datos un campo fértil para su desarrollo.

La Ciencia de Datos es un fenómeno ocasionado por la evolución de los avances tecnológicos y científicos, desarrollados progresivamente gracias a grandes compañías de tecnología y organizaciones gubernamentales. Los cambios resultantes en la sociedad, especialmente en entornos organizacionales, se manifiestan en procesos de sistematización, automatización y desarrollo organizacional. De igual manera, la Ciencia de Datos ha cobrado un protagonismo creciente en diversos entornos, desde hogares hasta la operación de sistemas complejos y la optimización de decisiones individuales mediante la inteligencia artificial. Como disciplina en consolidación que investiga y desarrolla sistemas como la inteligencia artificial y la automatización, la Ciencia de Datos impulsa la transición hacia una sociedad eminentemente

digital, donde incluso conceptos como el metaverso comienzan a explorarse en contextos educativos (Chen et al., 2023).

Uno de los ámbitos más relevantes donde la Ciencia de Datos ya ejerce una influencia notable es el de la educación superior. En centros de conocimiento como universidades, observatorios, laboratorios de vigilancia estratégica y de tecnología, entre otros, se está estudiando, analizando y desarrollando el impacto de esta sobre diversas disciplinas y su aplicación en procesos multidisciplinares. Estrategias como la personalización del aprendizaje mediante la aplicación de esta ciencia, se perfilan como usos concretos en este nivel (Pinzón, 2025). Esto subraya la necesidad de generar información verídica respecto a la relación e impacto de la Ciencia de Datos en los procesos de enseñanza y aprendizaje, y su vinculación con los modelos pedagógicos universitarios, considerando el papel fundamental de la didáctica en estos procesos (Wilmer Casasola Rivera, 2020). La integración de estos avances tecnológicos debe considerar sus fundamentos y epistemología para una teoría tecnológica de la educación coherente (Cañellas, 2002).

El presente estudio se justifica por la necesidad de comprender a fondo cómo la Ciencia de Datos puede transformar las metodologías pedagógicas, afectando directamente la calidad de la educación en instituciones de educación superior y los modelos educativos que optan por su implementación (Martínez Herrera et al., 2024). Los hallazgos y conclusiones que pueda arrojar la revisión de literatura especializada permitirán evidenciar no solo los vacíos en la generación de conocimiento y la posible ausencia de modelos pedagógicos innovadores, sino también el grado de implementación en prácticas pedagógicas universitarias.

Para ello, es fundamenta analizar los factores contextuales, con lo anterior se hace referencia, específicamente, a infraestructura, formación docente, políticas, que influyen en la

implementación de estas herramientas en educación superior. (Objetivo Específico 1).

Asimismo, se buscará evaluar críticamente los beneficios y riesgos asociados al uso de Ciencia de Datos en la personalización del aprendizaje y predicción del rendimiento. (Objetivo Específico 2), que pueden nutrirse de recursos y estrategias ya existentes para la enseñanza de la estadística y la analítica de datos (Medina-Hernández et al., 2022). Finalmente, se proponen recomendaciones para futuras investigaciones o intervenciones institucionales que favorezcan una implementación ética, contextualizada e inclusiva de la Ciencia de Datos en la educación superior (Objetivo Específico 3), estableciendo así correlaciones teóricas y prácticas entre la Ciencia de Datos y las metodologías pedagógicas en la educación superior.

Justificación

La presente investigación se justifica por la imperiosa necesidad de optimizar las prácticas pedagógicas en la educación superior, un ámbito donde la Ciencia de Datos emerge como una fuerza transformadora con un potencial considerable. Esta necesidad de optimización no es meramente un anhelo institucional interno, sino una respuesta a las crecientes demandas sociales de rendición de cuentas, eficacia y pertinencia, aspiraciones que encuentran eco y sustento en el marco normativo que rige la educación superior en Colombia. De hecho, como señala Pinzón (2025), esta disciplina ofrece "herramientas analíticas capaces de optimizar los procesos de enseñanza y aprendizaje, permitiendo la personalización de la educación y la mejora de la gestión académica". Por tanto, un beneficio central de este estudio es la identificación y divulgación de estrategias efectivas y modelos de implementación exitosos que aprovechen este potencial, contribuyendo así al cumplimiento de los estándares de calidad educativa exigidos legalmente.

El análisis sistemático de la literatura permitirá comprender cómo la aplicación de modelos predictivos y la personalización pueden conducir a mejoras demostrables en la reducción de la deserción estudiantil y el incremento del rendimiento académico (Núñez Villalobos, 2025), elevando así la calidad general de la educación superior (Martínez Herrera et al., 2024). Esta búsqueda de optimización y transformación se alinea directamente con los postulados fundamentales del marco legal colombiano. La Ley 30 de 1992 establece que la educación superior debe "despertará en los educandos un espíritu reflexivo, orientado al logro de la autonomía personal" (Art. 4) y tiene entre sus objetivos "Profundizar en la formación integral de los colombianos" (Art. 6a) y "Prestar a la comunidad un servicio con calidad" (Art. 6c). De manera complementaria, la Ley 115 de 1994, Ley General de Educación, consagra como uno de

los fines de la educación "El pleno desarrollo de la personalidad" (Art. 5, numeral 1) y "El acceso al conocimiento, la ciencia, la técnica y demás bienes y valores de la cultura" (Art. 5, numeral 7). La aplicación de la Ciencia de Datos para potenciar el rendimiento académico y disminuir la deserción estudiantil representa una vía concreta para materializar estos fines legalmente definidos. Adicionalmente, el rol del Estado en la promoción de la calidad educativa, enfatizado tanto en la Ley 115 de 1994 (Art. 4) como en las competencias nacionales delineadas en la Ley 715 de 2001, que incluyen "Definir, diseñar y establecer instrumentos y mecanismos para la calidad de la educación" (Art. 5.6), genera una demanda normativa por enfoques innovadores como los que explora esta investigación. Si bien el marco legal colombiano establece un claro mandato de calidad y formación integral, no prescribe metodologías pedagógicas específicas, especialmente aquellas que aprovechan tecnologías avanzadas. Por ello, este estudio no solo reviste interés académico, sino que también aporta una inteligencia operativa crucial para las Instituciones de Educación Superior (IES) que buscan cumplir con estos mandatos legales en un entorno contemporáneo saturado de datos. Las "mejoras demostrables" que la Ciencia de Datos puede ofrecer, y que esta revisión sistemática busca documentar, sustentan la toma de decisiones basada en evidencia, un principio implícitamente fomentado por normativas orientadas al aseguramiento de la calidad, como la Ley 1188 de 2008, que regula el registro calificado de programas y verifica sus condiciones de calidad.

Los beneficiarios de esta investigación son diversos y se sitúan en múltiples niveles del ecosistema educativo. Primordialmente, las instituciones de educación superior (IES) obtendrán un marco de referencia basado en evidencia, crucial para orientar la toma de decisiones estratégicas respecto a la adopción y la integración de estas tecnologías analíticas. Este marco es vital para el desarrollo estratégico del sistema de educación superior colombiano, especialmente

en el contexto de las metas nacionales para el avance científico y tecnológico. La Ley 29 de 1990 subraya que "Corresponde al Estado promover y orientar el adelanto científico y tecnológico y, por lo mismo, está obligado a incorporar la ciencia y la tecnología a los planes y programas de desarrollo económico y social del país" (Art. 1). Esta investigación contribuye directamente a este imperativo al explorar la aplicación de una tecnología clave como la Ciencia de Datos dentro de un sector crítico como es la educación superior. La identificación de estrategias y modelos de implementación efectivos también coadyuva al cumplimiento de las condiciones de calidad estipuladas en la Ley 1188 de 2008, tales como "La adecuada formación en investigación", "El fortalecimiento del número y calidad del personal docente" y "La garantía de una infraestructura física... que permitan la formación integral" (Art. 2). Así, esta investigación es importante para el marco normativo al ofrecer perspectivas prácticas que pueden informar cómo las IES innovan y mejoran su calidad, fortaleciendo la competitividad del sistema educativo colombiano. El marco de referencia basado en evidencia que se busca proveer no solo sirve a las IES individualmente, sino que también puede informar al Ministerio de Educación Nacional y a los organismos de acreditación en el refinamiento de estándares de calidad y mecanismos de apoyo, en consonancia con las funciones de supervisión establecidas en la Ley 30 de 1992 (Art. 3) y la Ley 715 de 2001 (Art. 5).

A su vez, los docentes y diseñadores curriculares se verán fortalecidos al contar con una comprensión más profunda de cómo la analítica de datos puede enriquecer sus prácticas pedagógicas, permitiéndoles una adaptación más precisa de los contenidos y estrategias a las necesidades individuales y a los diversos estilos de aprendizaje de los estudiantes (Torres Cañizález & Cobo Beltrán, 2017). Este empoderamiento del personal docente y el diseño de currículos pertinentes son pilares de una educación de calidad, tal como se refleja en la Ley 1188

de 2008 (Art. 2) y en el objetivo de la Ley 30 de 1992 de "Fomentar el desarrollo del pensamiento científico y pedagógico en Directivos y docentes" (Art. 31.i). Esta aproximación informada por datos redonda directamente en que los estudiantes accedan a experiencias educativas más personalizadas, equitativas y significativas, mejorando su trayectoria formativa (Aparicio, n.d.; Niño et al., 2024), lo cual apoya directamente el objetivo legal de "el pleno desarrollo de los alumnos" (Ley 30 de 1992, Art. 1) y de "formar la personalidad" (Ley 115 de 1994, Art. 13a). Finalmente, la comunidad científica y académica se beneficiará de una síntesis actualizada que no solo identifique vacíos de conocimiento, sino que también señale oportunidades para futuras investigaciones, un aspecto particularmente relevante para contextos regionales específicos como el colombiano, donde se percibe una menor concentración de estudios (López-Sevilla & Medina-Chicaiza, 2024). Esta contribución es especialmente pertinente dado el objetivo de la Ley 30 de 1992 de "Promover la unidad nacional, la descentralización, la integración regional y la cooperación interinstitucional con miras a que las diversas zonas del país dispongan de los recursos humanos y de las tecnologías apropiadas que les permitan atender adecuadamente sus necesidades" (Art. 6g). Al abordar la "menor concentración de estudios" en Colombia, esta investigación se posiciona como un elemento crítico para el desarrollo regional informado por la evidencia dentro del sector de la educación superior, respondiendo a un vacío donde las leyes nacionales promueven la descentralización y la pertinencia regional, pero la evidencia local para guiar tales iniciativas puede ser limitada.

Al considerar la integración de estas tecnologías, es crucial una reflexión pedagógica que guíe el proceso. Como pertinentemente advierten Torres Cañizález y Cobo Beltrán (2017), "la implementación tecnológica en educación debe alinearse con los fines formativos, evitando su reducción a un mero instrumento técnico". Esta reflexión y la priorización de consideraciones

éticas son imperativos que encuentran un sólido respaldo en la legislación educativa colombiana. La Ley 30 de 1992 establece de manera explícita que "La formación ética profesional debe ser elemento fundamental obligatorio de todos los programas de formación en las instituciones de Educación Superior" (Art. 129). El manejo ético de los datos estudiantiles y de los algoritmos en las aplicaciones de Ciencia de Datos se inscribe directamente bajo este mandato.

Adicionalmente, la Ley 115 de 1994 define como fines de la educación "El pleno desarrollo de la personalidad sin más limitaciones que las que le imponen los derechos de los demás y el orden jurídico, dentro de un proceso de formación integral, física, psíquica, intelectual, moral, espiritual, social, afectiva, ética, cívica y demás valores humanos" (Art. 5, numeral 1) y "La formación en el respeto a la vida y a los demás derechos humanos, a la paz, a los principios democráticos, de convivencia, pluralismo, justicia, solidaridad y equidad" (Art. 5, numeral 2). Una implementación ética e inclusiva de la Ciencia de Datos es, por lo tanto, fundamental para sostener estos principios, prevenir sesgos y asegurar la equidad. En este sentido, lo que se espera lograr con esta revisión sistemática es ir más allá de un simple catálogo de herramientas. Se busca proporcionar un diagnóstico comprensivo y actualizado que oriente una integración de la Ciencia de Datos que sea pedagógicamente sólida y reflexiva (Figaredo et al., 2020). Con ello, se espera generar recomendaciones concretas y contextualizadas que faciliten a las IES una implementación no solo efectiva, sino también ética e inclusiva. Este enfoque contribuirá a cerrar la brecha existente entre el vasto potencial teórico de la Ciencia de Datos y su aplicación práctica y transformadora en el aula, una brecha que no es solo técnica o pedagógica, sino que también puede reflejar un vacío en la gobernanza y la regulación específica de estas nuevas tecnologías.

El marco legal provee la brújula ética, y esta investigación, al enfatizar una implementación ética e inclusiva, actúa como una herramienta de navegación que ayuda a las IES a aplicar esa brújula al nuevo territorio de la pedagogía basada en datos.

En última instancia, se aspira a que los hallazgos de esta investigación sirvan como un insumo valioso para la evolución de los modelos educativos universitarios, fomentando una cultura de innovación educativa (Michavila, 2009) que sea adaptativa, eficiente y, sobre todo, verdaderamente centrada en el estudiante y sus necesidades de aprendizaje en el siglo XXI. Esta evolución e innovación, orientada por la presente investigación, no constituye un ejercicio académico aislado, sino una contribución sustantiva al progreso de la educación superior colombiana en consonancia con su marco normativo. La búsqueda de una educación “adaptativa, eficiente y, sobre todo, verdaderamente centrada en el estudiante” es plenamente coherente con los más altos propósitos de la legislación educativa colombiana, tales como asegurar la calidad (Ley 30 de 1992, Art. 3; Ley 115 de 1994, Art. 4), promover el desarrollo integral de los estudiantes (Ley 30 de 1992, Art. 1; Ley 115 de 1994, Art. 5.1) y fomentar una innovación que sirva al desarrollo nacional (Ley 29 de 1990, Art. 1). Por consiguiente, esta investigación se justifica por su potencial para dotar al sector de educación superior colombiano del conocimiento y las estrategias necesarias para aprovechar la Ciencia de Datos de manera responsable y efectiva, fortaleciendo así su capacidad para cumplir con sus obligaciones sociales y legales en el siglo XXI y materializando en la práctica los principios legales que definen la misión de la educación superior en el país.

Planteamiento del Problema

La educación superior contemporánea enfrenta el desafío crucial de adaptar sus metodologías pedagógicas a un entorno cada vez más tecnológico y caracterizado por la diversidad estudiantil. Con frecuencia, el predominio de enfoques tradicionales limita la capacidad de las instituciones para responder con agilidad a las diferencias en los estilos de aprendizaje, implementar una evaluación dinámica del desempeño y optimizar las estrategias de enseñanza. Esta situación evidencia una subutilización del vasto potencial que ofrece la Ciencia de Datos, disciplina que, mediante el análisis avanzado de información académica, podría transformar cualitativamente la toma de decisiones pedagógicas y la personalización de los itinerarios formativos (Pinzón, 2025).

Si bien la Ciencia de Datos ha evolucionado significativamente desde sus orígenes en la estadística tradicional, convirtiéndose en una herramienta capaz de procesar grandes volúmenes de información mediante técnicas como el aprendizaje automático, su incorporación en el ámbito de la educación superior aún es incipiente en muchos contextos. De acuerdo con Medina-Hernández et al. (2022), “la falta de capacitación docente, las brechas tecnológicas y la resistencia institucional dificultan el aprovechamiento de sus capacidades para diseñar estrategias pedagógicas basadas en evidencia, anticipar dificultades académicas y gestionar recursos de manera eficiente”. Estas barreras son fundamentales para entender los desafíos actuales en la adopción de estas tecnologías.

A pesar de que se reconocen ampliamente las ventajas de la Ciencia de Datos en la educación, incluyendo la creación de modelos predictivos para la intervención temprana, la facilitación de retroalimentación personalizada y la optimización de recursos institucionales (Norambuena et al., 2022), persisten obstáculos críticos para su adopción generalizada. Entre

ellos, destacan las dificultades para traducir los hallazgos analíticos en acciones pedagógicas concretas y efectivas, así como las limitaciones técnicas y económicas que restringen la implementación de sistemas robustos para la recopilación y el procesamiento de datos educativos a gran escala (Chen et al., 2023). Adicionalmente, la necesaria colaboración interdisciplinaria entre expertos en pedagogía y especialistas en Ciencia de Datos no siempre se materializa, lo que dificulta el diseño de soluciones tecnológicas que sean verdaderamente contextualizadas y pedagógicamente pertinentes (Becerra & López Alurralde, 2021).

Es fundamental considerar que, como advierten Torres Cañizález y Cobo Beltrán (2017), “la implementación tecnológica en educación debe alinearse con los fines formativos, evitando su reducción a un mero instrumento técnico”. En este sentido, la irrupción del Big Data y la inteligencia artificial en el sector educativo ha generado percepciones ambivalentes: mientras que para algunos representan una oportunidad sin precedentes para personalizar el aprendizaje y democratizar el acceso al conocimiento, para otros plantean serios riesgos asociados a la privacidad de los datos y a los sesgos algorítmicos, los cuales podrían exacerbar las desigualdades existentes en lugar de mitigarlas (Becerra & López Alurralde, 2021). Ello exige que su integración se aborde desde un enfoque eminentemente crítico, ético y adaptativo.

El panorama educativo actual refleja una transición hacia modelos formativos híbridos y más flexibles. No obstante, como subraya Casasola Rivera (2020), “la innovación educativa enfrenta barreras como la falta de formación docente y la rigidez institucional”, elementos que, coincidentemente, también han limitado históricamente la adopción de enfoques basados en datos. A pesar de estas dificultades, el potencial de la Ciencia de Datos para abordar desafíos centrales de la educación superior –como el análisis de patrones de deserción, la adaptación de

contenidos curriculares y la optimización de recursos– es innegable y se alinea con las necesidades contemporáneas del sector.

La propia disciplina educativa ha experimentado una constante evolución, incorporando progresivamente herramientas tecnológicas que han redefinido sus fundamentos teóricos y metodológicos. La tecnología educativa, tal como la describe García-Valcárcel Muñoz, A. (2002), ha transitado desde su concepción inicial vinculada a los medios audiovisuales hasta los sofisticados sistemas computacionales que hoy transforman la interacción y los procesos en el aula. En la actualidad, el avance hacia algoritmos de aprendizaje automático y análisis predictivos representa un nuevo paradigma, con profundas implicaciones para la educación superior.

La integración de la Ciencia de Datos en las metodologías pedagógicas no es una mera actualización tecnológica, sino una evolución que se inscribe en esta tradición de la disciplina educativa por asimilar y potenciar los avances de su tiempo. El éxito de esta empresa dependerá crucialmente de la capacidad para superar las brechas formativas existentes en el personal docente y para diseñar e implementar sistemas analíticos que sean éticos, transparentes y que prioricen la equidad y el desarrollo integral del estudiante. En esta línea, la transformación de la cultura institucional se erige como un factor clave para construir una educación superior que sea verdaderamente inclusiva y esté adaptada a las complejas demandas del siglo XXI.

En este contexto, la presente investigación busca responder a interrogantes cruciales, tales como: ¿cuáles son los factores contextuales (infraestructura tecnológica, formación docente y políticas institucionales) que han facilitado o limitado la integración de herramientas de Ciencia de Datos en la educación superior, según la literatura científica actual?; ¿qué beneficios y riesgos reporta la literatura académica sobre el uso de la Ciencia de Datos para la

personalización del aprendizaje y la predicción del rendimiento estudiantil en contextos de educación superior?; y ¿qué recomendaciones emergen en la literatura para una implementación ética, inclusiva y contextualizada de herramientas de Ciencia de Datos en la educación superior? A partir de estas interrogantes, la investigación tiene como propósito proporcionar un marco teórico-práctico que oriente a las instituciones educativas en la transición hacia modelos pedagógicos innovadores, más inclusivos, eficientes y adaptados a las demandas contemporáneas, aprovechando de manera responsable el potencial de la Ciencia de Datos.

Objetivos

Objetivo General

Sintetizar la integración de herramientas de Ciencia de Datos en las metodologías pedagógicas de la educación superior mediante una Revisión Sistemática de Literatura, identificando tendencias, vacíos y oportunidades de innovación educativa basada en datos.

Objetivos Específicos

Analizar los factores contextuales (infraestructura, formación docente, políticas) que influyen en la implementación de estas herramientas en educación superior.

Evaluar críticamente los beneficios y riesgos asociados al uso de Ciencia de Datos en la personalización del aprendizaje y predicción del rendimiento.

Proponer recomendaciones para futuras investigaciones o intervenciones institucionales que favorezcan una implementación ética, contextualizada e inclusiva de la Ciencia de Datos en la educación superior.

Marco Teórico

La educación superior se encuentra en un proceso constante de adaptación y búsqueda de enfoques innovadores para optimizar la enseñanza y el aprendizaje (Michavila, 2009). En este contexto dinámico, la integración de tecnologías emerge como un factor crucial. Particularmente, el conjunto de herramientas y técnicas analíticas avanzadas provenientes del análisis de datos presenta una oportunidad significativa para la transformación de las metodologías pedagógicas en estas instituciones (Lemus-Delgado & Pérez Navarro, 2020). Este marco teórico explora la convergencia entre dichas herramientas y la pedagogía en la educación superior, fundamentándose en la literatura para investigar su impacto.

La adopción e integración de nuevas tecnologías es una progresión natural en la innovación educativa. Históricamente, la tecnología ha sido fundamental para los objetivos formativos, y la introducción de enfoques basados en datos es una manifestación contemporánea de esta tendencia. La investigación propuesta se alinea con este interés generalizado, evidenciado por la atención que estas nuevas aproximaciones analíticas reciben en diversos campos académicos, incluyendo su aplicación en la formación profesional de áreas como la bibliotecología y sus contribuciones en estudios globales (Estrada Cuzcano & Alhuay-Quispe, 2025; Lemus-Delgado & Pérez Navarro, 2020).

Las metodologías pedagógicas tradicionales están siendo revisadas profundamente ante los avances tecnológicos. La era digital exige una transición hacia experiencias de aprendizaje más dinámicas y adaptadas a las necesidades individuales (Aparicio, 2025; Niño et al., 2024). Las teorías de la tecnología educativa ofrecen un marco para comprender cómo estas herramientas analíticas pueden integrarse efectivamente, enfatizando la alineación entre tecnología y objetivos pedagógicos. En este sentido, el aprendizaje personalizado, facilitado por

el análisis de datos, adquiere creciente importancia para atender la diversidad estudiantil y mejorar los resultados (Aparicio, 2025; Niño et al., 2024). Este cambio hacia una instrucción más individualizada podría transformar fundamentalmente la pedagogía en la educación superior, siempre que su integración sea reflexiva y se base en principios pedagógicos sólidos.

Diversas herramientas y técnicas, incluyendo el aprendizaje automático, la analítica del aprendizaje y el modelado predictivo, presentan un potencial notable para la educación superior (Figaredo et al., 2020). La analítica del aprendizaje, por ejemplo, permite obtener información valiosa sobre los patrones de los estudiantes y adaptar las intervenciones pedagógicas. Los modelos predictivos pueden identificar a estudiantes en riesgo, posibilitando estrategias de apoyo proactivas y mejorando la retención (Núñez Villalobos, 2025). En relación con la personalización, Niño et al. (2024) sostienen que "el uso del Big Data en la educación facilita la personalización del aprendizaje en tiempo real, ajustando el contenido y la forma de presentación en función del progreso y las necesidades individuales de cada estudiante". La creciente sofisticación de estas herramientas ofrece amplias posibilidades, aunque su aplicación efectiva requiere que los educadores desarrollen nuevas competencias en alfabetización y análisis de datos.

La implementación del aprendizaje automático y otros métodos avanzados en la educación superior enfrenta desafíos. Medina-Hernández et al. (2022) señalan que "la falta de infraestructura adecuada, de experiencia técnica especializada y de formación para el profesorado pueden constituir barreras significativas para la adopción generalizada de estas tecnologías". Las consideraciones éticas, especialmente en cuanto a la privacidad de los datos estudiantiles y los posibles sesgos algorítmicos, son de suma importancia y requieren atención meticulosa. Es crucial garantizar la transparencia de los modelos de aprendizaje automático para

fomentar la confianza y facilitar su uso efectivo. Los beneficios potenciales deben sopesarse cuidadosamente con las responsabilidades éticas inherentes al manejo de información sensible, pues no abordar estas preocupaciones podría mermar el impacto positivo de estas tecnologías. La integración exitosa requiere un enfoque que considere tanto los aspectos tecnológicos como los factores humanos y organizacionales.

El análisis de casos exitosos sobre la integración de herramientas analíticas puede proporcionar información valiosa. Estos ejemplos podrían incluir el uso de paneles de analítica para ofrecer datos relevantes sobre la participación y el rendimiento estudiantil (Figaredo et al., 2020) o la aplicación del aprendizaje automático para personalizar rutas de aprendizaje y recomendar recursos (Aparicio, 2025; Pinzón, 2025). Los enfoques basados en datos también pueden optimizar el diseño curricular. El éxito a menudo implica proporcionar a los educadores herramientas intuitivas e información práctica y suele depender de una sólida colaboración entre científicos de datos, tecnólogos educativos y docentes (Medina-Hernández et al., 2022).

Estas aproximaciones analíticas tienen el potencial de optimizar los resultados del aprendizaje. De acuerdo con Norambuena et al. (2022) y Figaredo et al. (2020), el análisis de datos en educación "tiene el potencial de optimizar significativamente los resultados del aprendizaje en la educación superior al proporcionar información basada en datos sobre los procesos de aprendizaje de los estudiantes". Al identificar y abordar disparidades en los resultados entre diferentes grupos, estas herramientas pueden contribuir a una mayor equidad (Nuñez Villalobos, 2025). Las capacidades de personalización en tiempo real pueden adaptar las experiencias de aprendizaje dinámicamente (Niño et al., 2024), permitiendo intervenciones más específicas para estudiantes que podrían tener dificultades. No obstante, es fundamental

implementar estos enfoques respetando la autonomía del estudiante y evitando limitar la exposición a diversas perspectivas.

Garantizar la privacidad y seguridad de los datos estudiantiles es primordial. Abordar posibles sesgos en algoritmos y datos es crucial para no perpetuar desigualdades. La transparencia y la rendición de cuentas son esenciales para generar confianza. Estas consideraciones éticas están profundamente entrelazadas con valores de justicia y equidad, y su abordaje debe implicar un diálogo continuo entre todas las partes interesadas.

La siguiente tabla resume las aplicaciones clave de estas técnicas analíticas en la educación superior y sus objetivos correspondientes dentro de esta investigación, ilustrando cómo se alinean con la evaluación crítica de su impacto en la optimización y personalización del aprendizaje, así como en la transformación general de las metodologías pedagógicas.

Tabla 1

Aplicaciones de la Ciencia de Datos en la Educación Superior y Objetivos Correspondientes

Herramienta/Técnica de Ciencia de Datos	Aplicación en Educación Superior	Objetivo de la Investigación.
Aprendizaje Automático	Personalización de rutas de aprendizaje, recomendación de recursos	Evaluar críticamente el impacto de la Ciencia de Datos en la optimización del aprendizaje y la personalización.
Analítica del Aprendizaje	Obtención de información sobre patrones de aprendizaje, identificación de áreas de dificultad	Examinar el impacto de la Ciencia de Datos en la transformación de metodologías pedagógicas y evaluar su impacto en la optimización del aprendizaje.

Herramienta/Técnica de Ciencia de Datos	Aplicación en Educación Superior	Objetivo de la Investigación.
Modelado Predictivo	Identificación de estudiantes en riesgo de abandono o bajo rendimiento	Evaluar críticamente el impacto de la Ciencia de Datos en la optimización del aprendizaje y considerar su influencia en la equidad.
Big Data	Personalización del aprendizaje en tiempo real, adaptación dinámica del contenido.	Evaluar críticamente el impacto de la Ciencia de Datos en la optimización del aprendizaje y la personalización.

Nota. Las aplicaciones detalladas en la Tabla 1 ilustran cómo las diversas herramientas y técnicas de la Ciencia de Datos se alinean con los objetivos de esta investigación, buscando evaluar críticamente su impacto en la optimización y personalización del aprendizaje, así como en la transformación general de las metodologías pedagógicas. *Fuente:* Adaptada de (Figaredo, et al., 2020. Pp. 33-43)

Esta conexión directa entre las capacidades técnicas y los fines investigativos subraya el potencial de la Ciencia de Datos no solo para generar intervenciones efectivas, sino también para informar sobre cómo estas intervenciones pueden ser diseñadas y aplicadas de manera reflexiva y orientada a objetivos pedagógicos claros. No obstante, la implementación del aprendizaje automático y otros métodos de la Ciencia de Datos en la educación superior no está exenta de desafíos.

Metodología

La presente investigación adoptó un enfoque descriptivo y no experimental, centrado en el análisis e interpretación de la información existente en la literatura académica especializada. Dada la naturaleza del problema planteado y los objetivos establecidos, que buscaron examinar, identificar, analizar y evaluar los aportes, problemáticas, experiencias y el impacto de la Ciencia de Datos en las metodologías pedagógicas universitarias, la metodología más pertinente fue la Revisión Sistemática de Literatura.

Tipo y Enfoque de Investigación

El estudio se enmarcó en un tipo de investigación documental y descriptiva. Fue documental porque la fuente primaria de datos fue la literatura académica existente. Fue descriptiva porque buscó caracterizar el estado actual del conocimiento respecto a la integración de la Ciencia de Datos en la pedagogía universitaria, detallando problemáticas, experiencias exitosas e impactos, sin manipular variables ni establecer relaciones causales directas a través de experimentación. El enfoque fue primordialmente cualitativo, ya que la recopilación y análisis de datos se basó en la interpretación y síntesis de contenidos textuales de los estudios seleccionados para responder a las preguntas de investigación.

Método: Revisión Sistemática de Literatura

La Revisión Sistemática de Literatura es un método riguroso y transparente que permite identificar, evaluar y sintetizar la evidencia disponible sobre una pregunta de investigación específica. A diferencia de las revisiones narrativas tradicionales, este enfoque sigue un protocolo estructurado para minimizar sesgos y asegurar que la síntesis de conocimiento sea exhaustiva y reproducible. Este método fue idóneo para establecer las correlaciones teóricas y

prácticas entre la Ciencia de Datos y las metodologías pedagógicas en la educación superior, tal como lo definió el objetivo general.

Pregunta de Investigación y Estrategia de Búsqueda

Todo el proceso de búsqueda y análisis se guio por el Planteamiento del Problema y los Objetivos General y Específicos. La pregunta central de investigación, derivada del objetivo general, fue: ¿De qué manera la integración de herramientas de Ciencia de Datos en las metodologías pedagógicas de la educación superior puede ser sintetizada mediante una Revisión Sistemática de Literatura para identificar tendencias, vacíos y oportunidades de innovación educativa basada en datos?

Se diseñó una estrategia de búsqueda exhaustiva utilizando una combinación de palabras clave y términos relacionados, derivados de los objetivos y el planteamiento del problema. Los términos clave en español incluyeron, entre otros: "Ciencia de Datos", "aprendizaje automático", "inteligencia artificial", "educación superior", "pedagogía", "metodologías educativas", "optimización aprendizaje", "experiencias exitosas", "problemáticas", y "barreras". Se utilizaron operadores booleanos (AND, OR) para refinar las búsquedas. Las bases de datos académicas exploradas fueron Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ERIC y Google Scholar avanzado.

Criterios de Selección y Proceso de Revisión

Para asegurar la pertinencia y calidad de los estudios seleccionados, se establecieron criterios claros de inclusión y exclusión.

Los Criterios de Inclusión fueron: estudios primarios (artículos de revista, actas de conferencias, tesis doctorales, reportes técnicos) que abordaron la aplicación de la Ciencia de Datos (incluyendo aprendizaje automático, análisis predictivo, analítica de aprendizaje) en el

contexto de la educación superior y su impacto o relación con las metodologías pedagógicas. Se priorizaron estudios publicados en español o inglés, dentro del rango de fechas entre 2002 y 2025, para asegurar la actualidad del conocimiento.

Los Criterios de Exclusión fueron: editoriales, cartas al editor, capítulos de libro (a menos que fueran parte de un compendio temático muy relevante), estudios no enfocados específicamente en educación superior o que trataron la Ciencia de Datos de forma genérica sin aplicarla a la pedagogía, estudios duplicados o de baja calidad metodológica tras una evaluación inicial.

El proceso de selección se realizó en fases. Inicialmente, se revisaron títulos y resúmenes según los criterios de inclusión y exclusión. Los estudios potencialmente relevantes pasaron a una revisión de texto completo. Se mantuvo un registro detallado del proceso (número de estudios identificados, excluidos y razones) para garantizar transparencia. Finalmente, 29 estudios fueron seleccionados para la revisión completa, cuyas referencias se listan al final de este documento.

Extracción, Análisis y Síntesis de Datos

De los estudios seleccionados para la revisión completa, se extrajo información relevante para responder a los objetivos de la investigación utilizando una ficha de extracción de datos. Esta ficha incluyó: datos bibliográficos (autores, año, publicación), diseño del estudio (si aplicó), contexto geográfico y educativo, la aplicación específica de la Ciencia de Datos o herramienta analítica, las metodologías pedagógicas involucradas, e información relevante para cada uno de los objetivos específicos. Para el Objetivo Específico 1, se extrajeron datos sobre factores contextuales (infraestructura, formación docente, políticas). Para el Objetivo Específico 2, se registraron beneficios y riesgos asociados al uso de estas herramientas en la personalización del

aprendizaje y predicción del rendimiento. Para el Objetivo Específico 3, se identificaron las recomendaciones propuestas por los autores, además de cualquier otra información pertinente.

La información extraída fue organizada y sintetizada para identificar patrones, tendencias y hallazgos clave relacionados con cada objetivo específico. Para el Objetivo 1, se agruparon y describieron los principales factores contextuales. Para el Objetivo 2, se analizaron y categorizaron los beneficios y riesgos reportados. Para el Objetivo 3, se sintetizaron las recomendaciones propuestas. Se buscaron correlaciones teóricas y prácticas entre la Ciencia de Datos y las metodologías pedagógicas, como se planteó en el objetivo general, sintetizando cómo los hallazgos conectaban estos campos. Se identificaron brechas de conocimiento y áreas poco exploradas. Los resultados de la síntesis y el análisis se presentaron de manera organizada, generalmente siguiendo la estructura de los objetivos de investigación, utilizando tablas, gráficos y descripciones narrativas para ilustrar los hallazgos.

Consideraciones Éticas

Dado que la investigación se basó en la revisión de literatura pública, las consideraciones éticas principales se centraron en asegurar la rigurosidad y la transparencia del proceso. Se garantizó la correcta atribución de las ideas y hallazgos a sus autores originales mediante un sistema de citación y referenciación consistente (estilo APA 7ma edición). Se evitó el plagio en todas sus formas, y el proceso de selección y análisis se realizó de manera objetiva y reproducible.

Resultados

Este capítulo presenta los hallazgos derivados de la Revisión Sistemática de Literatura, enfocándose en examinar los aportes de la Ciencia de Datos a la transformación de las metodologías pedagógicas en la educación superior. El análisis se estructura en tres secciones principales que corresponden a los objetivos específicos de la investigación, abordando los factores contextuales en la implementación, las aplicaciones, beneficios y riesgos de su utilización, y las consideraciones emergentes de la literatura para una integración efectiva, pedagógica y ética.

Factores Contextuales en la Implementación de la Ciencia de Datos en Educación Superior

La integración efectiva de herramientas de Ciencia de Datos en la educación superior está supeditada a una compleja interacción de factores contextuales. La literatura revisada identifica consistentemente que las barreras para la adopción de estas tecnologías no son meramente técnicas, sino que abarcan dimensiones pedagógicas, organizacionales, de capacitación y, crucialmente, éticas.

Un desafío fundamental radica en la infraestructura tecnológica y la preparación del personal. Medina-Hernández et al. (2022) señalan que "la falta de capacitación docente, las brechas tecnológicas y la resistencia institucional dificultan el aprovechamiento de sus capacidades para diseñar estrategias pedagógicas basadas en evidencia, anticipar dificultades académicas y gestionar recursos de manera eficiente" (p. 63). Esta observación es crucial, ya que la carencia de infraestructura robusta, sistemas de gestión de datos interoperables y herramientas analíticas adecuadas limita la capacidad institucional para un manejo eficiente de los datos educativos. Desde la perspectiva pedagógica, la rigidez curricular y metodológica prevalente en numerosas instituciones constituye un obstáculo (Martínez Herrera et al., 2024, p. 2.); pues los

modelos de enseñanza tradicionales no se adaptan con facilidad a la personalización y flexibilidad que la Ciencia de Datos puede propiciar. Casasola Rivera (2020) indica que en la docencia universitaria, a menudo "El recurso didáctico fundamental de muchos docentes es la clase magistral" (p. 46), un enfoque que puede presentar resistencia a la innovación basada en datos.

La resistencia a la innovación por parte de algunos sectores docentes y administrativos también se identifica como un freno. Esta situación se ve agravada por la falta de capacitación docente específica en Ciencia de Datos y analítica del aprendizaje. Michavila (2009) advierte que "son los profesores los actores decisivos de los procesos de cambio educativo y las innovaciones, tanto pedagógicas como tecnológicas, sólo son viables si cuentan con la complicidad y el protagonismo de los docentes" (p. 4).

Los desafíos éticos representan otro conjunto de factores contextuales de gran peso. La gestión de datos sensibles de los estudiantes y la garantía de su privacidad emergen como preocupaciones centrales. Marín y Tur (2023), en su revisión sobre la privacidad de datos en Tecnología Educativa, encontraron que "muchas políticas de privacidad son poco claras para la ciudadanía, que siente una falta de control sobre sus datos personales" (p. 8) . Esta falta de transparencia y la ausencia de protocolos claros de consentimiento informado, como reporta la literatura sobre el tema (Vásquez y Ruiz, 2021, citado en Oviedo Gómez, 2025), no solo genera vulnerabilidades, sino que también puede erosionar la confianza. El estudio de Aravena Nina y Tapia Quispe (2022) sobre la responsabilidad en la privacidad de datos, refuerza esta idea al indicar que "los propietarios deben tener conocimiento de la aplicación de sus datos y no deben dar su información detallada sin su conocimiento" (p. 5), lo que implica la necesidad de un consentimiento informado y consciente.

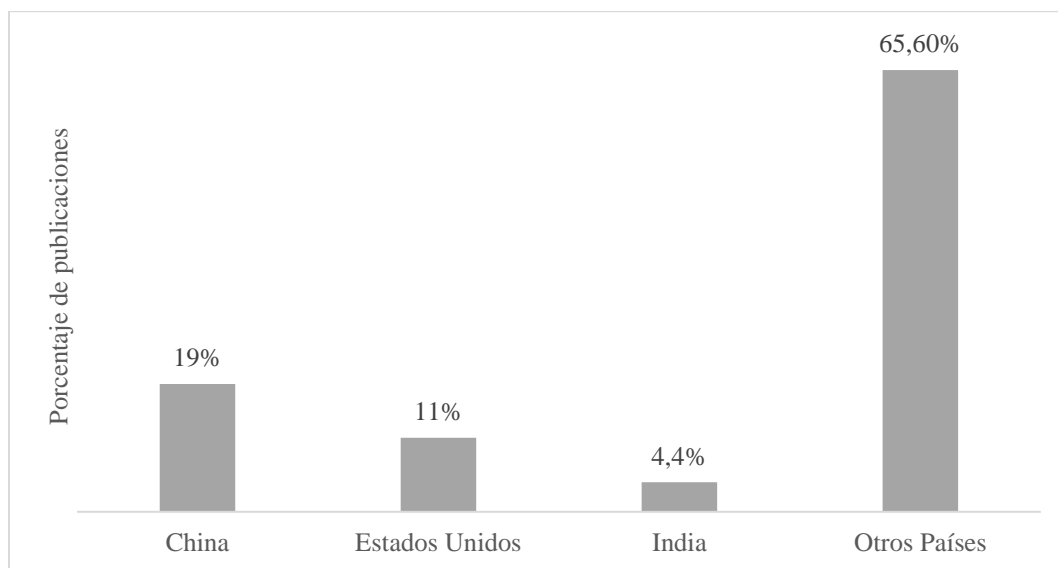
Adicionalmente, la problemática de los sesgos algorítmicos es un factor contextual ineludible. La tesis de Ponte Ramírez (2024), una revisión sistemática sobre el uso de Machine Learning (ML) en educación superior, destaca entre los desafíos y barreras la "infraestructura inadecuada", la "resistencia al cambio" y la "falta de capacitación" (p. 33-34). Si bien la tesis menciona la importancia de la ética al considerar, por ejemplo, el uso de ChatGPT (Ponte Ramírez, 2024, p. 35), no presenta un hallazgo cuantitativo específico sobre el porcentaje de estudios que evalúan métricas de equidad en sus resultados. La ausencia de una evaluación crítica de la equidad en las aplicaciones de Ciencia de Datos, como se ha indicado en diversos foros, puede llevar a que los algoritmos perpetúen o incluso amplifiquen desigualdades existentes.

Distribución Geográfica y Foco Temático de la Investigación

Un factor contextual importante que emerge de la revisión es la marcada desigualdad en la producción científica y la implementación de la Ciencia de Datos en la educación superior a nivel global.

Figura 1

Distribución Geográfica de Estudios en IA en Educación Superior

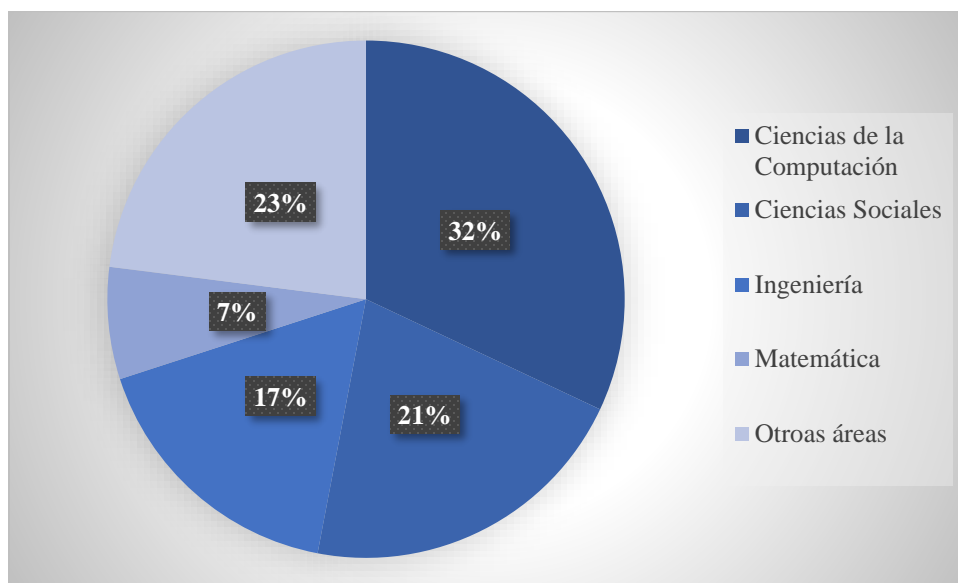


Nota. Carrión-Barco et al. (2025) identificaron en su análisis bibliométrico que, en el campo de la inteligencia artificial aplicada a la enseñanza-aprendizaje en educación superior, los países con mayor producción son "China (19%), Estados Unidos (11%) y la India (4.4%)" (p. 60).

El estudio de López-Sevilla y Medina-Chicaiza (2024) sobre la producción científica de Data Lakes –un concepto clave para la infraestructura de la Ciencia de Datos– identifica en su muestra que "Italia destaca como el líder en investigaciones sobre lago de datos, contribuyendo con el 8% del total de artículos publicados en el mencionado período", seguido por "Francia y España, (...) contribuyendo con un 5% de los artículos publicados cada uno" y "China e India (...) cada una con una contribución del 4%" (p. 49) . Estos patrones de concentración sugieren la necesidad de impulsar la investigación y aplicación de estas herramientas en regiones como Latinoamérica, y específicamente en Colombia, así como en una gama más amplia de disciplinas.

Figura 2

Áreas del Conocimiento en IA en Educación Superior Porcentaje de Publicaciones



Nota. Esta concentración se refleja también en las áreas del conocimiento abordadas, donde el mismo estudio señala un predominio de "ciencias de la computación (32%), ciencias sociales (21%), ingeniería (17%) y matemáticas (7%)" *Fuente:* (Carrión-Barco et al., 2025, p. 60).

La literatura científica revisada converge en señalar el notable potencial de la Ciencia de Datos para transformar la educación superior. Este protagonismo se manifiesta en su capacidad para optimizar la toma de decisiones, personalizar las experiencias de aprendizaje y mejorar tanto el rendimiento académico como las tasas de retención estudiantil. Pinzón Aparicio (2025) resume este potencial al afirmar que la Ciencia de Datos se orienta a "explorar el impacto de las herramientas como la minería de datos, el aprendizaje automático y las analíticas de aprendizaje en las necesidades individuales, anticipar desafíos y optimizar procesos educativos en un marco ético que garantice la privacidad" (p. 55) . La personalización del aprendizaje, en particular, es

identificada como una de sus aplicaciones más concretas y prometedoras (Aparicio, s.f.; Pinzón Aparicio, 2025; Niño et al., 2024).

Las principales herramientas y técnicas identificadas, sus aplicaciones y autores relevantes se resumen en la Tabla 2.

Tabla 2

Aplicaciones Clave de la Ciencia de Datos en Educación Superior y Autores Relevantes

Herramienta/Técnica de Ciencia de Datos	Aplicación en Educación Superior	Autores Principales (Ejemplos de la revisión)
Aprendizaje Automático (Machine Learning)	Personalización de rutas de aprendizaje, recomendación de recursos, predicción de rendimiento.	Pinzón (2025); Núñez Villalobos (2025); Ponte Ramirez (2024); Menacho Ángeles et al. (2024)
Analítica del Aprendizaje (Learning Analytics)	Obtención de información sobre patrones de aprendizaje, identificación de áreas de dificultad, retroalimentación.	Figaredo et al. (2020); Rojas-Castro (2017); Norambuena et al. (2022); Solano-Barliza et al. (2023)
Modelado Predictivo	Identificación temprana de estudiantes en riesgo de abandono o bajo rendimiento académico.	Núñez Villalobos (2025); Norambuena et al. (2022)
Big Data	Personalización del aprendizaje en tiempo real, adaptación dinámica del contenido y estrategias.	Niño et al. (2024); Becerra & López Alurralde (2021)

Nota. Estas aplicaciones buscan abordar problemáticas educativas clave, en la revisión sistemática sobre modelos predictivos basados en analíticas de aprendizaje, encontraron que si bien estos "mejoran la precisión en la identificación de estudiantes en riesgo en comparación con

los métodos tradicionales" (p. 8), también se concluye que "la construcción actual de modelos de predicción de abandono universitario posee importantes limitaciones" (p. 1), señalando la dependencia de la calidad de los datos y el contexto. *Fuente:* Norambuena et al. (2022)

Experiencias específicas de implementación ilustran beneficios cuantificables. El trabajo de Núñez Villalobos (2025) sobre un modelo predictivo para la retención estudiantil, destaca que "los datos conseguidos, permiten una visión más precisa de cómo se comporta el modelo a diferentes umbrales, mostrando su capacidad para diferenciar entre los estudiantes que desertan y los que continúan" (p. 2). Su estudio busca ofrecer, mediante el uso de regresión logística y redes neuronales, "intervenciones efectivas y adaptables a las necesidades de los estudiantes" (Núñez Villalobos, 2025, p. 2). De forma similar, Menacho Ángeles et al. (2024), al investigar la IA como herramienta para el aprendizaje autónomo, resaltan que "la IA representa una herramienta crucial para el aprendizaje autónomo en la educación superior para estudiantes y docentes, siempre y cuando se utilice con ética y responsabilidad" (p. 1).

El estudio de Salas-Rueda et al. (2025) se centró en analizar "el uso de la aplicación educativa sobre las matemáticas para la Licenciatura en Geografía Aplicada considerando la Ciencia de datos", la cual "utiliza un simulador web para facilitar el aprendizaje sobre el despeje de la variable 'tiempo' en la función de crecimiento de la población" (p. 293). Los resultados del análisis de regresión lineal indicaron que "el simulador de la aplicación afecta positivamente la motivación ($p=0.0459$) y el entusiasmo ($p=0.0174$) en la Unidad 3 'Relaciones y funciones'" (Salas-Rueda et al., 2025, pp. 297, 299). Así, concluyen que "la aplicación web sobre las matemáticas es una opción tecnológica para facilitar el proceso educativo debido a que esta herramienta web está disponible los 365 días del año y fomenta el aprendizaje personalizado por

medio del simulador" (Salas-Rueda et al., 2025, p. 302). Este estudio, por lo tanto, evidencia cómo una herramienta tecnológica específica, cuya efectividad en aspectos afectivos y de facilitación del aprendizaje personalizado se evalúa mediante técnicas de Ciencia de Datos, puede constituir un apoyo valioso en la enseñanza de temas complejos.

A pesar de estos beneficios, la literatura también es clara respecto a los riesgos. Becerra y López Alurralde (2021), al explorar las representaciones sociales del Big Data y la inteligencia artificial, señalan que, junto a nociones positivamente valoradas como el conocimiento, conviven "referencias a problemas sociales, negativamente valorados, como el desempleo y el control social" (p. 89). Esta ambivalencia refleja la tensión entre el potencial optimizador y los temores asociados al uso de la información y la equidad educativa.

Hacia una Implementación Pedagógica, Ética e Inclusiva: Consideraciones Emergentes de la Literatura

La integración exitosa de la Ciencia de Datos en la educación superior, según la literatura revisada, trasciende la mera adopción técnica y exige una profunda alineación con principios pedagógicos sólidos, así como una atención prioritaria a las dimensiones éticas y de inclusión.

La tecnología debe servir como un facilitador del aprendizaje activo y significativo. Torres Cañizález y Cobo Beltrán (2017) recalcan que la tecnología educativa, para lograr sus finalidades, "emplea diversos medios y recursos para el aprendizaje escolar, ya sean los tradicionales (libros, pizarra, entre otros), o las herramientas que ofrecen las tecnologías de información y comunicación (TIC)" (p. 31). Subrayan, además, que la tecnología educativa "implica una reflexión pedagógica, de la cual subyace una teoría, una metodología y una práctica formativa en contextos educativos determinados" (Torres Cañizález & Cobo Beltrán, 2017, p. 33, parafraseando la distinción que hacen).

La preparación docente es un factor crítico. Casasola Rivera (2020) argumenta que "la investigación didáctica es de suma importancia porque ningún aprendizaje está determinado, así como, ningún método de enseñanza garantiza verdaderamente un óptimo aprendizaje en el estudiante" (pp. 40-41). Esto subraya que los docentes deben ser investigadores de su práctica, adaptando las herramientas tecnológicas. En esta línea, Michavila (2009) enfatiza que "el cambio de paradigma educativo se debe centrar en la sustitución de una enseñanza excesivamente teórica por una educación activa, apoyándose en una formación más práctica y en el uso de nuevas tecnologías" (p. 4).

Las implementaciones más efectivas logran una sinergia entre tecnología y pedagogía. Sin embargo, Figaredo et al. (2020) advierten que, si bien "en la última década el análisis de los datos educativos ha crecido de manera decidida, la transferencia de esa investigación a las prácticas de los educadores/as en las clases o en las políticas institucionales ha sido bastante limitada" (p. 35). Esta brecha es un desafío central. Para ello, es crucial la perspectiva que ofrece Colom Cañellas (2002), quien concibe que la tecnología educativa aporta "una nueva concepción del aprendizaje, una propuesta basada en los medios y una alternativa de carácter metacognitivo" (p. 11, Resumen), lo que implica que las herramientas de Ciencia de Datos deben integrarse para fomentar el constructivismo y la personalización.

La dimensión ética es un eje central y transversal. La preocupación por "la privacidad de los datos en el uso de la tecnología en educación es un aspecto ético de interés creciente", especialmente con "el uso de sistemas de analíticas de aprendizaje y de inteligencia artificial, que permiten la recogida y el análisis de grandes cantidades de datos" (Marín & Tur, 2023, p. 7). La necesidad de "garantizar la privacidad y el manejo responsable de la información" es un consenso (Martínez et al., 2020, citado en Pinzón Aparicio, 2025, p. 56). Chen et al. (2023), al

analizar los desafíos del Edu-Metaverse, identifican como cruciales: "1) data security and privacy protection; 2) balance between the real world and virtual world identities; 3) preparing instructors for Edu-Metaverse" . Estos desafíos son directamente aplicables a la Ciencia de Datos en educación, resaltando la urgencia de desarrollar marcos éticos robustos.

Conclusiones

Esta revisión sistemática de literatura ha culminado en una síntesis comprensiva de la intrincada relación entre la Ciencia de Datos y las metodologías pedagógicas en la educación superior, revelando un panorama de notables oportunidades, pero también de considerables desafíos que deben ser abordados con diligencia y una perspectiva crítica. Los hallazgos han permitido dar respuesta a cada uno de los objetivos planteados, delineando un camino hacia una integración más efectiva, ética y pedagógicamente fundamentada de estas poderosas herramientas analíticas.

En respuesta al primer objetivo, centrado en el análisis de los factores contextuales que influyen en la implementación de herramientas de Ciencia de Datos, se concluye que las barreras más significativas no residen únicamente en la disponibilidad de infraestructura tecnológica o en la resistencia inicial a la innovación. Más profundamente, son los factores de índole ética, como la ausencia de protocolos robustos para el consentimiento informado y la gestión de la privacidad de los datos estudiantiles, junto con el riesgo latente de sesgos algorítmicos no auditados, los que emergen como los obstáculos más críticos. A esto se suma, de manera determinante, la insuficiente capacitación del personal docente, no tanto en el manejo técnico de las herramientas, sino en la capacidad para interpretar los datos con una perspectiva pedagógica y aplicarlos de forma ética y efectiva en el diseño de experiencias de aprendizaje. Se concluye, por tanto, que la superación de estos factores contextuales, especialmente los éticos y formativos, así como la consideración de la distribución geográfica y temática desigual de la investigación, son prerequisites indispensables para una adopción exitosa y equitativa de la Ciencia de Datos en las instituciones de educación superior.

Respecto al segundo objetivo, que buscaba evaluar críticamente los beneficios y riesgos asociados al uso de la Ciencia de Datos para la personalización del aprendizaje y la predicción del rendimiento, se concluye que esta disciplina ofrece un potencial transformador considerable. La literatura evidencia beneficios tangibles como la mejora en la identificación temprana de estudiantes en riesgo, la optimización de la retención estudiantil y la posibilidad de adaptar las trayectorias de aprendizaje a las necesidades individuales, lo cual puede conducir a mejoras en el rendimiento académico. Sin embargo, estos beneficios no están exentos de riesgos significativos. La implementación acrítica de modelos predictivos puede conducir a la estigmatización o a la simplificación excesiva de las necesidades estudiantiles, mientras que la personalización algorítmica puede, paradójicamente, limitar la exposición a la diversidad de pensamiento si no se diseña cuidadosamente. Se concluye que la materialización de los beneficios y la mitigación de los riesgos dependen de un delicado equilibrio, una constante evaluación crítica del impacto de estas herramientas y una priorización de los principios pedagógicos y éticos sobre la mera eficiencia tecnológica.

Finalmente, en relación con el tercer objetivo, orientado a proponer recomendaciones para futuras investigaciones e intervenciones institucionales, se concluye que es imperativo avanzar hacia una visión más holística e integrada. La investigación actual, si bien cuantifica impactos positivos, a menudo omite la evaluación de dimensiones cualitativas del aprendizaje, como el desarrollo del pensamiento crítico, la creatividad o el bienestar socioemocional del estudiante, y la forma en que la equidad se ve afectada en la práctica. Por ello, se recomienda enfáticamente que las futuras investigaciones adopten metodologías mixtas que capturen esta complejidad. Para las instituciones, se concluye que es esencial desarrollar e implementar marcos éticos sólidos y políticas claras de gobernanza de datos; invertir sostenidamente en programas de

desarrollo profesional docente que enfatizan la pedagogía basada en datos y la ética digital; fomentar una cultura de experimentación e innovación responsable; y, crucialmente, promover la adaptación y validación local de las herramientas analíticas, especialmente en contextos subrepresentados en la literatura actual, como el latinoamericano y el colombiano. Es necesario también impulsar estudios que exploren la aplicación de estas herramientas más allá de las áreas STEM, abarcando las humanidades y las ciencias sociales.

En síntesis, esta travesía por el conocimiento existente nos revela que la Ciencia de Datos no es una panacea, pero sí una aliada formidable con un potencial disruptivo y profundamente transformador para las metodologías pedagógicas en la educación superior. Abre puertas a soluciones innovadoras para problemáticas educativas persistentes y ofrece caminos hacia una optimización del aprendizaje que puede ser medible y significativa. El recorrido por la literatura nos llena de entusiasmo por las posibilidades vislumbradas: una educación más personalizada, más proactiva y, en última instancia, más humana, si se guía con sabiduría. No obstante, la realización plena de este potencial no es automática; depende intrínsecamente de nuestra capacidad colectiva –como educadores, investigadores, directivos y diseñadores de políticas– para navegar sus complejidades con una visión clara, un compromiso ético inquebrantable y una pasión constante por el aprendizaje y la mejora continua. El futuro de la educación superior, enriquecido por la Ciencia de Datos, nos convoca a ser arquitectos de un cambio reflexivo, donde la tecnología sirva para empoderar y emancipar, asegurando que cada avance contribuya a una formación más justa, equitativa y profundamente significativa para todos los estudiantes que cruzan nuestras aulas, preparándolos no solo para un mundo de datos, sino para un mundo de sentido.

Recomendaciones

El presente capítulo tiene como finalidad proponer una serie de recomendaciones estratégicas y líneas de acción futuras, las cuales se derivan directamente de las conclusiones y hallazgos de la investigación sobre Ciencia de Datos y Metodologías Pedagógicas en Educación Superior. Se busca ofrecer orientaciones concretas para las instituciones de educación superior (IES), los formuladores de políticas públicas, los docentes y los investigadores, con el objetivo primordial de promover una integración de la Ciencia de Datos en las metodologías pedagógicas que sea efectiva, ética, contextualizada e inclusiva dentro del sistema de educación superior en Colombia.

Las recomendaciones aquí presentadas se articulan de manera directa con el objetivo general de esta investigación, que consiste en: "Sintetizar la integración de herramientas de Ciencia de Datos en las metodologías pedagógicas de la educación superior mediante una Revisión Sistemática de Literatura, identificando tendencias, vacíos y oportunidades de innovación educativa basada en datos". De igual forma, estas recomendaciones responden a las principales conclusiones del estudio, las cuales señalan la existencia de significativas barreras éticas, una apremiante necesidad de capacitación docente, la notable influencia de factores contextuales diversos, y el crucial balance que debe mantenerse entre los beneficios potenciales y los riesgos inherentes al uso de la Ciencia de Datos en el ámbito de la educación superior. Este capítulo se enfoca, de manera particular, en la materialización del Objetivo Específico 3 de la investigación: "Proponer recomendaciones para futuras investigaciones o intervenciones institucionales que favorezcan una implementación ética, contextualizada e inclusiva de la Ciencia de Datos en la educación superior".

La implementación de la Ciencia de Datos en el contexto educativo trasciende la mera adopción de nuevas tecnologías; representa una transformación profunda que exige una visión estratégica clara, un compromiso institucional sólido y una alineación rigurosa con los principios fundamentales que rigen la educación superior. Las siguientes recomendaciones se centran en la creación de las condiciones indispensables para que esta implementación no solo sea exitosa, sino también genuinamente beneficiosa para todos los actores del ecosistema educativo.

Una de las conclusiones más relevantes de la investigación de base es la identificación de "barreras éticas, como la falta de protocolos claros para el consentimiento informado y la gestión de la privacidad, junto con el riesgo de sesgos algorítmicos" como obstáculos significativos para la adopción efectiva de la Ciencia de Datos. Este hallazgo resuena profundamente con el marco legal colombiano. La Ley 30 de 1992, en su Artículo 4°, establece que la Educación Superior "despertará en los educandos un espíritu reflexivo, orientado al logro de la autonomía personal, en un marco de libertad de pensamiento". Complementariamente, el Artículo 129 de la misma ley subraya que "La formación ética profesional debe ser elemento fundamental obligatorio de todos los programas de formación en las instituciones de Educación Superior". En una línea similar, la Ley 115 de 1994, en su Artículo 5°, numeral 2, promueve "La formación en el respeto a la vida y a los demás derechos humanos, a la paz, a los principios democráticos, de convivencia, pluralismo, justicia, solidaridad y equidad". Estos principios constitucionales y legales deben ser la piedra angular sobre la cual se construyan los marcos de gobernanza de datos en las IES. La ausencia de tales marcos no constituye únicamente una barrera técnica o procedimental, sino que representa una contradicción fundamental con los fines esenciales de la educación superior, como son la formación integral, el desarrollo ético y el respeto irrestricto a

los derechos humanos. Implementar la Ciencia de Datos sin una sólida base ética podría, en consecuencia, socavar la confianza pública en las IES y desvirtuar su misión fundamental.

Se recomienda que las IES desarrollen e implementen políticas explícitas, detalladas y de fácil acceso que regulen de manera integral todo el ciclo de vida de los datos de estudiantes y docentes, incluyendo su recolección, almacenamiento, procesamiento, uso, compartición y disposición final. Estas políticas deben estar en estricta consonancia con la legislación nacional vigente en materia de protección de datos personales.

Asimismo, es imperativo garantizar el consentimiento informado, explícito, libre y revocable de cada individuo antes de utilizar sus datos personales para cualquier fin relacionado con la Ciencia de Datos. Este proceso de consentimiento debe incluir una explicación clara y comprensible de los propósitos específicos del tratamiento de los datos, los potenciales riesgos y beneficios asociados, las medidas de seguridad implementadas y los derechos del titular de los datos. Este enfoque se alinea con el "pleno desarrollo de la personalidad" y la "autonomía" del individuo, consagrados en la Ley 115 de 1994, Artículos 1 y 5 numeral 1.

Se aconseja el desarrollo y la aplicación de protocolos rigurosos para la auditoría periódica y sistemática de los algoritmos y modelos de Ciencia de Datos que se utilicen en la institución. El objetivo de estas auditorías es identificar, analizar y corregir posibles sesgos (ya sean inherentes a los datos de entrenamiento o introducidos durante el modelado) que puedan perpetuar desigualdades existentes o conducir a decisiones injustas o discriminatorias, tal como se advierte en la investigación. Esta práctica es fundamental para asegurar la "equidad", uno de los principios rectores de la educación según la Ley 115 de 1994, Artículo 5 numeral 2.

Adicionalmente, se debe fomentar activamente la diversidad (de género, etnia, disciplina, perspectiva, etc.) en los equipos encargados del diseño, desarrollo y evaluación de los sistemas

de Ciencia de Datos, con el fin de asegurar una multiplicidad de miradas que contribuyan a la detección temprana y mitigación de sesgos.

Se propone el establecimiento de comités de ética de datos y gobernanza institucional de carácter multidisciplinario. Estos comités estarían encargados de supervisar la implementación ética de la Ciencia de Datos en la IES, revisar y aprobar proyectos que involucren el uso intensivo de datos, desarrollar y actualizar las políticas institucionales, y actuar como instancia de consulta y resolución de dilemas éticos que puedan surgir. Es crucial que estos comités cuenten con representación de docentes de diversas áreas, estudiantes, expertos en ética, juristas especializados en protección de datos y personal técnico con conocimiento en Ciencia de Datos. La autonomía universitaria, reconocida en el Artículo 28 de la Ley 30 de 1992, faculta a las IES para "darse y modificar sus estatutos" y "adoptar sus correspondientes regímenes", lo que provee el marco legal para la creación y funcionamiento de dichos comités.

Más allá de ser una simple formalidad, una gobernanza de datos efectiva y transparente puede transformarse en una ventaja competitiva significativa para las IES. En una era donde la conciencia sobre la privacidad y el uso ético de los datos está en aumento, aquellas instituciones que demuestren un compromiso genuino y visible con estos principios podrían atraer a estudiantes, docentes e investigadores que valoran la responsabilidad y la protección de sus derechos digitales, posicionándose, así como líderes en innovación educativa responsable.

La investigación de base identifica como limitaciones cruciales la "insuficiente capacitación docente en la interpretación pedagógica y el uso ético de los datos", así como la "rigidez de algunas estructuras institucionales". Estos hallazgos demandan acciones concretas para el desarrollo de capacidades. La Ley 115 de 1994, en su Artículo 4, subraya la importancia de la "cualificación y formación de los educadores". De manera complementaria, la Ley 30 de

1992, en su Artículo 31, literal i), orienta a "Fomentar el desarrollo del pensamiento científico y pedagógico en Directivos y docentes", y su Artículo 123 establece que el régimen docente debe contemplar "sistemas de evaluación y capacitación". Estos mandatos legales refuerzan la necesidad imperante de invertir en el capital humano y en la modernización institucional.

Se recomienda enfáticamente que las IES diseñen y ofrezcan programas de desarrollo profesional continuo, especializado y pertinente para sus cuerpos docentes. Estos programas deben abordar, como mínimo:

- Los fundamentos conceptuales y técnicos de la Ciencia de Datos y sus herramientas, con un enfoque claro en su aplicación práctica al contexto educativo.
- Estrategias para la interpretación pedagógica de los datos generados por los estudiantes y los sistemas de aprendizaje, con el fin de informar la toma de decisiones en el aula, personalizar la enseñanza y mejorar el diseño curricular.
- Principios y prácticas para el uso ético de los datos en la educación, incluyendo un profundo entendimiento de la privacidad, el consentimiento informado, la prevención de sesgos algorítmicos y la equidad.
- El desarrollo de la alfabetización digital y competencias avanzadas para integrar de manera efectiva las tecnologías de datos en las prácticas de enseñanza, aprendizaje y evaluación. Estos programas deben ser coherentes con lo dispuesto en el Artículo 111 de la Ley 115 de 1994, que orienta la formación de los educadores hacia su "profesionalización, actualización, especialización y perfeccionamiento". Es fundamental que la capacitación docente no se perciba como un evento aislado, sino como un proceso continuo, integrado de manera significativa en la trayectoria y carrera profesional del docente. Este proceso debería estar

vinculado a sistemas de incentivos y al reconocimiento formal de la innovación pedagógica y la adquisición de nuevas competencias en Ciencia de Datos.

Se debe promover activamente la experimentación pedagógica con herramientas y metodologías de Ciencia de Datos, siempre dentro de un marco de responsabilidad, supervisión ética y evaluación continua de los resultados e impactos. Este enfoque se alinea con la Ley 1188 de 2008, Artículo 2, condición de carácter institucional 3, que exige "El desarrollo de una cultura de la autoevaluación, que genere un espíritu crítico y constructivo de mejoramiento continuo".

Paralelamente, es crucial adaptar las estructuras administrativas y académicas de las IES para que sean más flexibles, ágiles y propicias para la colaboración interdisciplinaria y la adopción de innovaciones basadas en datos, tal como lo sugiere la Ley 1188 de 2008, Artículo 2, condición de carácter institucional 2. La autonomía universitaria, consagrada en el Artículo 28 de la Ley 30 de 1992, otorga a las IES la potestad de "crear, organizar y desarrollar sus programas académicos" y "definir y organizar sus labores formativas", lo cual incluye la capacidad de liderar y gestionar la innovación pedagógica y tecnológica. La rigidez institucional, a menudo, puede ser un síntoma de una aversión al riesgo inherente o de una falta de visión estratégica a largo plazo. Fomentar una cultura de "innovación responsable", como se deriva de las conclusiones del estudio, implica necesariamente la creación de espacios seguros y controlados para la experimentación, donde el aprendizaje a partir de los errores sea valorado como parte del proceso de adopción de tecnologías potencialmente disruptivas como la Ciencia de Datos.

Las IES deben realizar inversiones estratégicas en la infraestructura tecnológica y en los sistemas de información necesarios para la recolección, gestión, procesamiento, análisis y visualización de datos educativos de manera segura, eficiente y escalable. Esta necesidad se alinea con la competencia nacional de "Definir, diseñar, reglamentar y mantener un sistema de

información del sector educativo", estipulada en la Ley 715 de 2001, Artículo 5.4, y con la obligación de las IES de suministrar información relevante, como se indica en la Ley 30 de 1992, Artículo 38.b.

Es fundamental asegurar la interoperabilidad entre los diversos sistemas institucionales (académicos, administrativos, financieros, de bienestar, etc.) y garantizar la calidad, integridad y consistencia de los datos para facilitar análisis significativos y la generación de conocimiento útil para la toma de decisiones.

La investigación de base pone de manifiesto una "concentración geográfica y temática en la producción científica" relacionada con la Ciencia de Datos en educación, subrayando la necesidad imperante de "desarrollos más contextualizados y diversificados". Este llamado a la contextualización encuentra un fuerte respaldo en el marco legal colombiano. La Ley 30 de 1992, en su Artículo 4, enfatiza la importancia de tener en cuenta "la particularidad de las formas culturales existentes en el país". Más específicamente, su Artículo 6, literal g, exhorta a que "las diversas zonas del país dispongan de los recursos humanos y de las tecnologías apropiadas que les permitan atender adecuadamente sus necesidades". Adicionalmente, la Ley 1188 de 2008, en el Artículo 2, como condición de calidad de los programas, exige la "adecuada justificación del programa para que sea pertinente frente a las necesidades del país".

Se recomienda promover activamente la investigación y el desarrollo de herramientas y aplicaciones de Ciencia de Datos que respondan de manera efectiva a los contextos específicos de Colombia y sus diversas regiones, considerando las particularidades culturales, sociales, económicas y educativas de cada una. La contextualización no debe entenderse meramente como una adaptación superficial de herramientas o modelos desarrollados en otros entornos, sino como un proceso de co-creación de soluciones que involucre a las comunidades y los actores locales.

Esto asegura que la Ciencia de Datos responda a problemas sentidos y relevantes localmente, en lugar de a prioridades impuestas desde el exterior.

Asimismo, es crucial fomentar la aplicación de la Ciencia de Datos no solo en las áreas tradicionalmente asociadas a ella, como las STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), sino también explorar y potenciar su uso en las humanidades, las artes, las ciencias sociales y otras áreas del conocimiento, tal como se sugiere en las conclusiones del estudio. Esto permitirá una comprensión más rica y diversificada de los fenómenos educativos y sociales.

Es imperativo diseñar e implementar estrategias que aseguren que los beneficios derivados de la aplicación de la Ciencia de Datos en la educación –tales como la personalización del aprendizaje, la identificación temprana de estudiantes en riesgo y la optimización de recursos– lleguen a todos los estudiantes, con especial atención a aquellos que se encuentran en situaciones de vulnerabilidad socioeconómica, geográfica o de cualquier otra índole. El objetivo debe ser reducir las brechas educativas existentes, no ampliarlas, como se advierte sobre los riesgos en la investigación.

Se debe garantizar que el acceso tanto a la formación en competencias de Ciencia de Datos como a las herramientas tecnológicas necesarias sea equitativo para todos los estudiantes y docentes. Este principio se alinea con el Artículo 5 de la Ley 30 de 1992, que establece que la Educación Superior "será accesible a quienes demuestren poseer las capacidades requeridas", y con el Artículo 31, literal e, de la misma ley, que busca "Facilitar a las personas aptas el acceso al conocimiento... así como los mecanismos financieros que lo hagan viable". Una implementación inclusiva de la Ciencia de Datos tiene el potencial de transformar a las IES en verdaderos motores de desarrollo regional más equitativo, al formar capital humano local con la

capacidad de utilizar datos para identificar y resolver problemas específicos de su entorno, democratizando así el conocimiento y empoderando a las regiones para su propio desarrollo sostenible.

Las conclusiones de la revisión sistemática de literatura enfatizan de manera contundente que "Las futuras investigaciones deben enfocarse en metodologías que capturen el impacto cualitativo de estas herramientas en el aprendizaje y exploren su aplicación en una mayor diversidad de disciplinas y contextos culturales." Esta directriz se alinea con el espíritu de la Ley 29 de 1990, que en su Artículo 1º establece que "Corresponde al Estado promover y orientar el adelanto científico y tecnológico", y en su Artículo 2º, la importancia de "crear condiciones favorables para la generación de conocimiento científico y tecnología nacionales". De igual manera, la Ley 30 de 1992, en su Artículo 6, literal b, promueve "la creación, el desarrollo y la transmisión del conocimiento". Estas disposiciones legales ofrecen un marco robusto para impulsar una agenda de investigación que profundice en la comprensión y optimización del uso de la Ciencia de Datos en la educación superior colombiana.

Se recomienda priorizar el desarrollo de estudios que trasciendan las métricas puramente cuantitativas (como tasas de rendimiento o retención) y se enfoquen en explorar cómo las intervenciones pedagógicas basadas en Ciencia de Datos afectan las dimensiones cualitativas del aprendizaje. Esto incluye investigar su influencia en el desarrollo del pensamiento crítico, la creatividad, la motivación intrínseca, la autonomía del estudiante, las habilidades de colaboración y el bienestar socioemocional.

Para ello, es fundamental fomentar el uso de metodologías de investigación cualitativa (estudios de caso, etnografías, entrevistas en profundidad, análisis de discurso) y enfoques mixtos que permitan capturar la riqueza y complejidad de las experiencias de aprendizaje

mediadas por la Ciencia de Datos, así como las percepciones y vivencias de estudiantes y docentes.

Se debe impulsar la realización de investigaciones que analicen cómo las herramientas, técnicas y metodologías de Ciencia de Datos pueden ser adaptadas, implementadas y evaluadas de manera efectiva y éticamente responsable en los diversos contextos culturales presentes en Colombia. Esto implica considerar la rica diversidad regional, étnica y socioeconómica del país, asegurando que las soluciones sean pertinentes y respetuosas.

Asimismo, es crucial ampliar el foco de la investigación más allá de las disciplinas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), explorando sistemáticamente el potencial y los desafíos de la aplicación de la Ciencia de Datos en las humanidades, las ciencias sociales, las artes, las ciencias de la salud y otras áreas del conocimiento. La naturaleza compleja de los fenómenos abordados en estas disciplinas requiere un enfoque investigativo inherentemente interdisciplinario, que combine la experticia de científicos de datos, pedagogos, sociólogos, antropólogos, éticistas y especialistas en los diversos campos del saber, para así generar comprensiones profundas y soluciones robustas y contextualizadas.

Se sugiere promover la investigación de carácter metodológico orientada al diseño, la implementación y la validación de enfoques de evaluación innovadores y rigurosos. Estos enfoques deben ser capaces de integrar de manera coherente y sinérgica métodos cuantitativos y cualitativos para valorar el impacto integral (académico, social, ético, pedagógico) de la Ciencia de Datos en la educación superior.

El desarrollo de estas metodologías mixtas es crucial para generar evidencia sólida y matizada que pueda informar de manera efectiva tanto las prácticas pedagógicas cotidianas como la formulación de políticas institucionales y nacionales. Este tipo de investigación se alinea

directamente con el fomento de la investigación científica y tecnológica que busca la excelencia académica y la prioridad social, como lo establece la Ley 30 de 1992 en su Artículo 126. La validación local de herramientas y metodologías de Ciencia de Datos, especialmente a través de investigación contextualizada y participativa, es fundamental para evitar una "colonización digital" o la adopción acrítica de modelos desarrollados en otros contextos. Esto permitirá asegurar que las tecnologías y los enfoques analíticos sirvan genuinamente a los propósitos educativos y culturales de Colombia, fomentando una Ciencia de Datos propia, pertinente y soberana para la educación superior.

La transformación de las metodologías pedagógicas en la educación superior mediante la integración de la Ciencia de Datos es una empresa compleja que requiere una visión estratégica y un compromiso sostenido. La puesta en marcha efectiva de las recomendaciones aquí presentadas no es tarea de actores aislados, sino que demanda un esfuerzo concertado y una colaboración activa y continua entre el gobierno nacional y territorial, las instituciones de educación superior, los cuerpos docentes, los estudiantes, el sector productivo y la sociedad civil en su conjunto. Este llamado a la acción colaborativa encuentra eco en la Ley 29 de 1990, cuyo Artículo 1° ya vislumbraba la necesidad de establecer "mecanismos de relación entre sus actividades de desarrollo científico y tecnológico y las que, en los mismos campos, adelanten la universidad, la comunidad científica y el sector privado colombianos".

Es crucial entender que la integración de la Ciencia de Datos no es un proyecto con un inicio y un fin definidos, sino un proceso evolutivo y dinámico. Requiere una visión a largo plazo, una inversión sostenida en capital humano e infraestructura tecnológica, y, fundamentalmente, el fomento de una cultura organizacional que valore la adaptación, el aprendizaje permanente y la innovación responsable.

Finalmente, se reafirma el inmenso potencial transformador que la Ciencia de Datos posee para la educación superior. No obstante, este potencial solo podrá materializarse plenamente si su integración se guía por una visión pedagógica clara y bien definida, un compromiso ético firme e inquebrantable con la protección de los derechos y la dignidad de las personas, y un enfoque persistentemente centrado en el desarrollo integral, equitativo y significativo de cada estudiante que transita por las aulas universitarias, preparándolos no solo para un mundo saturado de datos, sino para un mundo cargado de sentido y propósito.

Referencias Bibliográficas

- Aparicio, L. A. P. (s. f.). *Personalización del aprendizaje mediante ciencia de datos: Estrategias y aplicaciones en la educación superior* [Manuscrito no publicado].
- Aravena Nina, D. K., & Tapia Quispe, H. E. (2022). La privacidad en el uso de los datos en la ciencia de datos [Tesis de licenciatura, Universidad Peruana de Ciencias Aplicadas]. Repositorio Académico UPC. <http://hdl.handle.net/10757/667215>
- Becerra, G., & López Alurralde, J. P. (2021). Representaciones sociales del big data y la inteligencia artificial: Una exploración estructural. *Cultura y Representaciones Sociales*, 16(31), 89–115. <https://www.scielo.org.mx/pdf/crs/v16n31/2007-8110-crs-16-31-00009.pdf>
- Carrión-Barco, G., Del Castillo Castro, C. I., Chayan Coloma, A., & Lecca Orrego, G. F. (2025). Mapeo de la investigación en inteligencia artificial aplicada a la enseñanza-aprendizaje en educación superior: Un análisis bibliométrico. *Revista Reflexiones de la Sociedad y Economía*, 2(1), 51–72. <https://doi.org/10.62776/rse.v2i1.19>
- Casasola Rivera, W. (2020). El papel de la didáctica en los procesos de enseñanza y aprendizaje universitarios. *Comunicación*, 29(1), 38–51.
- Chen, X., Zou, D., Xie, H., & Wang, F. L. (2023). Metaverse in education: Contributors, cooperations, and research themes. *IEEE Transactions on Learning Technologies*, 16(6), 1111–1129. <https://doi.org/10.1109/TLT.2023.3277952>
- Colom Cañellas, A. J. (2002). Para una teoría tecnológica de la educación: Fundamentos y epistemología. *Educación y Pedagogía*, 14(33), 13–27.

- Ékora. (s. f.). *Transformando la educación: Datos, analítica y personalización del aprendizaje en Colombia*. Ecosistema digital CINDE. <https://ekora.com.co/transformando-la-educacion-datos-analitica-y-personalizacion-del-aprendizaje-en-colombia/>
- Estrada Cuzcano, A., & Alhuay-Quispe, J. (2025). Ciencia de datos y la formación profesional en bibliotecología: Un análisis textual y revisión de los currículos en Iberoamérica. *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*, 39(102), 203–220. <https://doi.org/10.22201/iibi.24488321xe.2025.102.58943>
- Figaredo, D. D., Reich, J., & Ruipérez-Valiente, J. A. (2020). Analítica del aprendizaje y educación basada en datos: Un campo en expansión. *RIED-Revista Iberoamericana de Educación a Distancia*, 23(2), 33–43. <https://doi.org/10.5944/ried.23.2.27105>
- García-Valcárcel Muñoz, A. (2002). Tecnología educativa: Características y evolución de una disciplina. *Educación y Pedagogía*, 14(33), 67–87.
- Lemus-Delgado, D., & Pérez Navarro, R. (2020). Ciencias de datos y estudios globales: Aportaciones y desafíos metodológicos. *Colombia Internacional*, (102), 41–62. <https://doi.org/10.7440/colombiaint102.2020.03>
- Ley 29 de 1990. (1990, 7 de febrero). Por la cual se dictan disposiciones para el fomento de la investigación científica y el desarrollo tecnológico y se otorgan facultades extraordinarias. *Diario Oficial No. 39.164*.
- Ley 30 de 1992. (1992, 28 de diciembre). Por la cual se organiza el servicio público de la educación superior. *Diario Oficial No. 40.700*.
- Ley 115 de 1994. (1994, 8 de febrero). Por la cual se expide la ley general de educación. *Diario Oficial No. 41.214*.

Ley 715 de 2001. (2001, 21 de diciembre). Por la cual se dictan normas orgánicas en materia de recursos y competencias. *Diario Oficial No. 44.654*.

Ley 1188 de 2008. (2008, 25 de abril). Por la cual se regula el registro calificado de programas de educación superior. *Diario Oficial No. 46.971*.

López-Sevilla, G. M., & Medina-Chicaiza, R. P. (2024). Análisis bibliométrico de la producción científica sobre lago de datos. *INNOVA Research Journal*, 9(2), 40–57.

<https://doi.org/10.33890/innova.v9.n2.2024.2426>

Marín, V. I., & Tur, G. (2023). La privacidad de los datos en tecnología educativa: Resultados de una revisión de alcance. *EduTec. Revista Electrónica de Tecnología Educativa*, (83), 7–23. <https://doi.org/10.21556/edutec.2023.83.2701>

Martínez Herrera, G., Chávez Escobedo, J. M., & Martínez Macías, J. G. (2024). Modelo educativo y calidad de la educación en instituciones de educación superior. *Daena: International Journal of Good Conscience*, 19(2), 1–11.

Medina-Hernández, E. J., Muñiz, J. L., Guzmán-Aguilar, D. S., & Holguín-Higueta, A. (2022). Recursos y estrategias para la enseñanza de la estadística y la analítica de datos en la educación superior. *Formación Universitaria*, 15(3), 61–68.

<https://doi.org/10.4067/S0718-50062022000300061>

Menacho Ángeles, M. R., Pizarro Arancibia, L. M., Osorio Menacho, J. A., & León Pizarro, B. L. (2024). Inteligencia artificial como herramienta en el aprendizaje autónomo de los estudiantes de educación superior. *INVECOM: Estudios Transdisciplinarios en Comunicación y Sociedad*, 4(2), e040258.

https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S2739-00632024000200158

- Michavila, F. (2009). La innovación educativa: Oportunidades y barreras. *Arbor*, 185(Extra_1), 3–8. <https://doi.org/10.3989/arbor.2009.extran1201>
- Niño, J. D. M., Gonzalez Roys, G. A., & Aguilar Vázquez, B. (2024). Big data in education: Real-time personalization of learning. *Science Advanced*, 3(1), 6–15. <http://scienceadvanced.org/sa/index.php/Revista/article/view/12>
- Norambuena, J. M., Badilla-Quintana, M. G., & Angulo, Y. L. (2022). Modelos predictivos basados en uso de analíticas de aprendizaje en educación superior: Una revisión sistemática. *Texto Livre: Linguagem e Tecnologia*, 15, e36310. <https://doi.org/10.35699/1983-3652.2022.36310>
- Núñez Villalobos, D. A. (2025). Modelo predictivo basado en aprendizaje automático para la retención estudiantil en educación superior. *European Public & Social Innovation Review*, 10, 1–21. <https://doi.org/10.31637/epsir-2025-1307>
- Pinzón Aparicio, L. A. (2025). Personalización del aprendizaje mediante ciencia de datos: Estrategias y aplicaciones en la educación superior. *Revista Latinoamericana de Calidad Educativa*, 2(1), 55–60. <https://alumnieditora.com/index.php/ojs/article/view/71/116>
- Ponte Ramírez, C. A. (2024). Uso de machine learning en la educación superior entre 2019 y 2024: Una revisión sistemática [Tesis de maestría, Universidad César Vallejo]. Repositorio Institucional UCV. <https://repositorio.ucv.edu.pe/handle/20.500.12692/150161>
- Rojas-Castro, P. (2017). Learning analytics: Una revisión de la literatura. *Educación y Educadores*, 20(1), 106–128. <https://doi.org/10.5294/edu.2017.20.1.6>
- Salas-Rueda, R. A., González-García, H., & Becerra-Torres, E. (2025). Análisis de la aplicación educativa sobre las matemáticas para la licenciatura en Geografía Aplicada considerando

- la ciencia de datos. *Sociedad & Tecnología*, 8(2), 293–303.
<https://doi.org/10.51247/st.v8i2.511>
- Salmerón Moreira, Y. M., Luna Alvarez, H. E., Murillo Encarnación, W. G., & Pacheco Gómez, V. A. (2023). El futuro de la inteligencia artificial para la educación en las instituciones de educación superior. *Conrado*, 19(93), 27–34.
http://scielo.sld.cu/scielo.php?pid=S1990-86442023000400027&script=sci_arttext
- Solano-Barliza, A. D., Ojeda, A. D., & Aarón-Gonzálvez, M. (2023). Enseñanza de la analítica de datos usando aprendizaje basado en proyectos colaborativos. *Formación Universitaria*, 16(6), 23–32. <https://doi.org/10.4067/S0718-50062023000600023>
- Torres Cañizález, P. C., & Cobo Beltrán, J. K. (2017). Tecnología educativa y su papel en el logro de los fines de la educación. *Educere*, 21(68), 31–40.
<http://www.redalyc.org/articulo.oa?id=35652744004>
- Vásquez, A., & Ruiz, R. (2021). Ética y privacidad en la ciencia de datos aplicada a la educación. *Journal of Educational Ethics*, 6(4), 89–101.