

Clustering educativo: aplicación del algoritmo *K-means* para la mejora de estrategias docentes

Educational Clustering: Application of the K-Means
Algorithm to Improve Teaching Strategies

*Mario Sarián González**

* Doctor (C) en Innovación, IA y Transformación Digital por la Universidad Pontificia de Salamanca, España. Doctor en Empresas de la Universidad Central, Master of Business Administration por la Universidad de Sevilla, España. Magíster en Administración y Dirección de Empresas con especialización en Comunicación Estratégica por Universidad Pompeu Fabra, España. Ingeniero en Informática y Gestión por la Universidad Diego Portales, Chile. Profesor de la Universidad Autónoma de Chile y European Institute for international Studies (EIIIS), Estocolmo, Suecia. Correo electrónico: mario.sarian@uautonoma.cl; mgsariango.chs@upsa.es. Orcid: <https://orcid.org/0000-0003-2271-0532>.

Resumen

La creciente disponibilidad de datos educativos ha abierto nuevas oportunidades para optimizar los procesos de enseñanza y aprendizaje. En un escenario educativo cada vez más diverso y complejo, la inteligencia artificial emerge como una aliada poderosa para comprender mejor a los estudiantes y tomar decisiones pedagógicas más informadas. Este capítulo propone la aplicación del algoritmo de agrupamiento *K-means* como herramienta para analizar datos educativos y generar estrategias docentes diferenciadas. A partir del estudio de variables como el rendimiento académico, la participación en plataformas digitales y los estilos de aprendizaje, es posible identificar patrones que visibilizan la heterogeneidad del aula desde una perspectiva más profunda y basada en evidencia.

El enfoque propuesto va más allá del análisis técnico: busca construir una mirada pedagógica que conecte los datos con las historias reales de quienes aprenden. Se desarrolla una metodología ética y contextualizada que abarca desde la clarificación del propósito educativo hasta la interpretación de los resultados y el diseño de intervenciones adaptadas. Lejos de reemplazar al profesorado, el uso de *K-means* en educación se presenta como un complemento que enriquece el juicio docente, promueve una cultura de mejora continua y favorece una enseñanza más inclusiva, humana y situada.

Esta propuesta invita a docentes e investigadores a integrar la tecnología con sentido crítico, reconociendo que detrás de cada dato hay una persona y que el verdadero potencial de la inteligencia artificial en educación reside en su capacidad para acercarnos, no para estandarizarnos.

Palabras clave: algoritmos de aprendizaje automático, estrategias pedagógicas, inteligencia artificial en educación, *K-means*, personalización del aprendizaje.

Abstract

The increasing availability of educational data has opened new opportunities to optimize teaching and learning processes. In an increasingly diverse and complex educational landscape, artificial intelligence emerges as a powerful ally to better understand students and make more informed pedagogical decisions. This chapter proposes the application of the K-means clustering algorithm as a tool to analyze educational data and generate differentiated teaching strategies. By studying

variables such as academic performance, participation in digital platforms, and learning styles, it is possible to identify patterns that highlight classroom heterogeneity from a deeper, evidence-based perspective.

The proposed approach goes beyond technical analysis: it seeks to construct a pedagogical perspective that connects data with the real stories of learners. An ethical and contextualized methodology is developed, encompassing everything from clarifying the educational purpose to interpreting results and designing tailored interventions. Far from replacing teachers, the use of K-means in education is presented as a complement that enriches teachers' judgment, promotes a culture of continuous improvement, and fosters more inclusive, humane, and situated teaching.

This proposal invites teachers and researchers to integrate technology with a critical perspective, recognizing that behind every data point there is a person, and that the true potential of artificial intelligence in education lies in its ability to bring us closer together, not to standardize us.

Keywords: artificial intelligence in education, K-means, learning personalization, machine learning algorithms, pedagogical strategies.

Introducción

La educación contemporánea atraviesa una transformación profunda y sostenida, impulsada en gran medida por la expansión de las tecnologías digitales y el uso masivo de plataformas virtuales de aprendizaje. Estas herramientas no solo han modificado las formas de acceso al conocimiento, sino también los procesos mediante los cuales se produce, se comparte y se evalúa. Los entornos virtuales, los espacios colaborativos, los repositorios digitales y los sistemas automatizados de evaluación han ampliado significativamente las posibilidades pedagógicas; sin embargo, también han generado grandes volúmenes de datos que, en muchos casos, permanecen subutilizados desde el punto de vista educativo (Zawacki-Richter, 2019). En este contexto, la inteligencia artificial (IA) emerge como una herramienta estratégica para analizar e interpretar esa información, con el fin de comprender con mayor profundidad los procesos de enseñanza y aprendizaje (Chen, L. *et al.*, 2020).

Más que una moda tecnológica, la IA representa un nuevo enfoque para observar la realidad educativa desde perspectivas más profundas y basadas en evidencia. Su capacidad para procesar grandes volúmenes de información y detectar

patrones ocultos en los datos permite identificar comportamientos, trayectorias y necesidades específicas de los estudiantes. A diferencia del análisis estadístico tradicional, que suele centrarse en indicadores numéricos agregados, la IA invita a mirar la diversidad del estudiantado en toda su complejidad y riqueza.

Dentro del amplio abanico de herramientas que ofrece la IA, los algoritmos de aprendizaje automático (*machine learning*) ocupan un lugar central. Estos algoritmos se caracterizan por no requerir instrucciones explícitas para operar; aprenden directamente a partir de los datos, reconociendo regularidades y adaptándose a nuevas condiciones (Tuomi, 2018). Aplicados al campo educativo, permiten anticipar riesgos de deserción, identificar estudiantes con dificultades e incluso recomendar estrategias de enseñanza personalizadas basadas en el comportamiento histórico del alumnado.

Una de las técnicas más relevantes en este contexto es el *clustering*, también conocido como análisis de agrupamiento. Esta técnica consiste en reunir datos según su similitud, sin necesidad de contar con etiquetas o clasificaciones previas. En otras palabras, permite que los datos se organicen por sí mismos en grupos naturales y revelen estructuras y relaciones que no siempre resultan evidentes a simple vista (Zheliznyak *et al.*, 2017).

Esta lógica se adapta de manera extraordinaria a la realidad del aula, donde cada estudiante es diferente y trae consigo un conjunto único de experiencias, habilidades, intereses y dificultades. Mientras algunos requieren mayor acompañamiento y guía, otros prosperan en entornos más autónomos o colaborativos (Frades y Matthiesen, 2010). El *clustering* permite reconocer esas diferencias de manera sistemática y ayuda al docente a tomar decisiones informadas sobre cómo abordar la diversidad dentro del aula.

Entre los distintos algoritmos de *clustering* existentes, el más utilizado en educación es el *K-means*. Su funcionamiento parte de una premisa sencilla pero poderosa: dividir un conjunto de datos en k grupos, de modo que los elementos dentro de cada grupo sean lo más parecidos posible entre sí y, al mismo tiempo, distintos de los elementos de los demás grupos (Jain, 2008). El algoritmo identifica los “centros” o centroides de estos grupos, calcula la distancia de cada dato (estudiante) a esos centros y ajusta las agrupaciones hasta encontrar una configuración óptima (Barbakh *et al.*, 2009).

En el ámbito educativo, esto significa que no solo se agrupa a los estudiantes por calificaciones, sino también por múltiples variables que reflejan su comportamiento

y forma de aprender: desde la participación en plataformas virtuales hasta el estilo de aprendizaje predominante (Liu *et al.*, 2017). La tabla 1 muestra algunas de las aplicaciones comunes de la IA en el ámbito educativo:

Tabla 1. Aplicaciones comunes de la IA en educación

Elemento	Aplicación en el aula
Aprendizaje automático (Romero y Ventura, 2020)	Detectar patrones de rendimiento, sugerir rutas personalizadas
Procesamiento de lenguaje natural	Analizar mensajes en foros, generar retroalimentación automática
Sistemas expertos	Sugerir actividades según nivel del estudiante
Minería de datos educativos (Xiong <i>et al.</i> , 2024)	Identificar tendencias y necesidades de aprendizaje

Dentro del amplio espectro de técnicas que ofrece el aprendizaje automático, el agrupamiento o *clustering* ha adquirido una relevancia creciente, especialmente en el campo educativo. Esta técnica se distingue por su capacidad para organizar conjuntos de datos sin necesidad de contar con etiquetas o clasificaciones previas (Dutt *et al.*, 2015). En otras palabras, permite descubrir estructuras ocultas dentro de los datos a partir de similitudes internas entre los elementos, lo cual resulta sumamente valioso en contextos donde la diversidad del estudiantado es la norma y no la excepción.

En las aulas actuales, los estudiantes difieren en múltiples dimensiones: ritmo de aprendizaje, estilos cognitivos, niveles de motivación y acceso a recursos, entre muchos otros factores. Frente a esta complejidad, las herramientas tradicionales de evaluación suelen quedarse cortas. Es en este punto donde el *clustering* ofrece una alternativa poderosa, al permitir identificar grupos con características comunes de manera automática y basada en evidencia. Lejos de imponer categorías rígidas, esta metodología abre la posibilidad de comprender mejor la heterogeneidad presente en cada curso o cohorte (Srivastava y Kumar, 2021).

Uno de los algoritmos más populares y aplicados en esta área es el *K-means*, debido a su equilibrio entre simplicidad y efectividad. Este algoritmo parte de una premisa sencilla: dividir un conjunto de datos en “k” grupos, de forma que

los elementos dentro de cada grupo sean lo más parecidos posible entre sí y, a la vez, distintos de los elementos de los demás grupos. Para lograrlo, *K-means* analiza la distancia entre los puntos de datos (por ejemplo, estudiantes) y los “centros” o núcleos de cada grupo, lo que permite reajustar continuamente las agrupaciones hasta lograr una distribución óptima (Anuradha *et al.*, 2015).

Cuando este tipo de algoritmo se traslada al ámbito educativo, sus implicaciones son notables. No se trata simplemente de agrupar estudiantes por sus calificaciones, sino de considerar un abanico más amplio de variables que capturan comportamientos, actitudes y formas de aprender. Así, es posible identificar perfiles estudiantiles basados en patrones reales de interacción con los contenidos, participación en plataformas digitales, tiempos de conexión, cumplimiento de tareas e incluso estilos de aprendizaje predominantes.

A partir de esta información, los docentes no solo pueden comprender mejor a su alumnado, sino también diseñar estrategias pedagógicas diferenciadas, adecuadas a las necesidades particulares de cada grupo. Por ejemplo, ciertos estudiantes podrían beneficiarse de un acompañamiento más cercano, mientras que otros aprovecharían propuestas más autónomas o colaborativas. En última instancia, aplicar *K-means* en educación representa una herramienta valiosa para avanzar hacia una enseñanza más equitativa, informada y centrada en el estudiante.

En la tabla 2 se presentan algunas de las variables comúnmente utilizadas para realizar este tipo de análisis con el algoritmo *K-means*. Estas variables han sido seleccionadas por su relevancia en la caracterización de comportamientos de aprendizaje y por su disponibilidad en entornos educativos digitalizados (Liu, 2022).

Tabla 2. Variables educativas para el análisis de agrupamiento con *K-means*

Variable	Tipo de dato	Posible fuente de información
Promedio de calificaciones	Numérico	Sistema de notas del centro educativo
Participación en foros virtuales	Numérico/categorico	Plataforma LMS (Moodle, Google Classroom, etc.)
Estilo de aprendizaje	Categorico	Cuestionarios de diagnóstico (VARK, CHAEA)

Tiempo conectado a la semana	Númérico	Registro de actividad en plataforma digital
Entrega de tareas a tiempo	Binario (Sí/No)	Reporte de actividades del docente o LMS

La posibilidad de trabajar con diversas variables que reflejan el comportamiento, la participación, el rendimiento y los estilos de aprendizaje de los estudiantes abre un abanico de oportunidades para construir perfiles más completos e integrales de cada uno. Este enfoque permite ir mucho más allá de la simple evaluación basada en calificaciones, que tradicionalmente ha dominado la mayoría de los sistemas educativos. En cambio, invita a reconocer que detrás de cada número existe una persona con un conjunto particular de habilidades, necesidades, motivaciones y contextos.

Al utilizar estas variables de manera articulada mediante herramientas como el algoritmo *K-means*, se puede superar la lógica de la enseñanza homogénea y estandarizada. En lugar de asumir que todos los estudiantes aprenden al mismo ritmo, de la misma manera y con los mismos recursos, este enfoque permite identificar patrones de comportamiento y estilos de aprendizaje diferenciados, lo que facilita el diseño de experiencias formativas más significativas, personalizadas y acordes con la realidad del grupo (Kurniawan y Hermawan, 2023).

Este trabajo se plantea como una contribución concreta a ese objetivo. En particular, la investigación se orienta hacia la aplicación del algoritmo *K-means* sobre un conjunto real de datos estudiantiles, con el propósito de detectar agrupamientos internos que reflejen similitudes entre estudiantes no evidentes a simple vista. A través de este análisis, se busca comprender con mayor profundidad cómo se configuran los distintos perfiles presentes en el aula y cómo esas configuraciones pueden informar y enriquecer las decisiones pedagógicas que toma el profesorado.

El centro de esta propuesta no reside en el uso de la tecnología por sí misma, sino en su potencial para transformar la mirada pedagógica. Una vez identificados los grupos naturales de estudiantes —basados en variables como la participación, el estilo cognitivo, el nivel de compromiso o el desempeño académico— será posible diseñar estrategias de enseñanza más pertinentes, ajustadas a las características y necesidades de cada clúster.

La finalidad última de este capítulo, por tanto, es ofrecer recomendaciones prácticas y fundamentadas que orienten al profesorado en la construcción de

ambientes de aprendizaje más inclusivos, efectivos y justos. Se trata de avanzar hacia un modelo educativo donde las decisiones no se tomen únicamente desde la intuición o la experiencia —que siguen siendo valiosas—, sino también desde el análisis riguroso de datos que reflejan la realidad cotidiana del aula.

En este sentido, el uso de algoritmos como *K-means* en contextos educativos no debe entenderse como una sustitución del juicio docente, sino como un recurso complementario que enriquece la toma de decisiones. La educación del siglo XXI exige nuevas formas de pensar, planificar y actuar en el aula, y este tipo de herramientas puede ser un apoyo valioso para responder con mayor precisión y sensibilidad a la diversidad de los estudiantes (Omar *et al.*, 2020).

Tabla 3. *Aportes del uso K-means a la mejora de la enseñanza*

Beneficio potencial	¿Cómo mejora la enseñanza?
Crear grupos con características comunes	Permite adaptar materiales y actividades a cada grupo
Identificar estudiantes en riesgo	Facilita intervenciones más oportunas
Personalizar la enseñanza	Mejora la motivación y el rendimiento individual
Tomar decisiones basadas en datos	Reduce la intuición y aumenta la efectividad pedagógica

En definitiva, la incorporación de herramientas como el algoritmo *K-means* en el ámbito educativo va mucho más allá del simple uso de nuevas tecnologías. Su verdadero valor radica en la posibilidad de generar una mirada más profunda y comprensiva sobre los estudiantes, sus trayectorias y sus necesidades particulares. Al analizar los datos que los propios procesos de enseñanza y aprendizaje generan —ya sea en plataformas virtuales, evaluaciones, participación u otros indicadores— se abre una oportunidad concreta para repensar las prácticas docentes al visibilizar dimensiones que antes quedaban fuera del radar pedagógico.

Este capítulo se propone justamente eso: mostrar cómo la IA, cuando se utiliza con criterio profesional, sentido ético y una clara orientación pedagógica, puede convertirse en una aliada para el profesorado. Lejos de buscar reemplazar el rol docente, estas tecnologías pueden actuar como un complemento que enriquece la

toma de decisiones en el aula al aportar información significativa que orienta mejor las estrategias de enseñanza y permite adaptarse con mayor precisión a los distintos ritmos y estilos de aprendizaje.

En este sentido, el uso del *clustering* educativo y, en particular, del algoritmo *K-means* se presenta como una herramienta concreta para avanzar hacia una educación más personalizada, equitativa y centrada en el estudiante. Esta metodología permite visibilizar patrones de comportamiento o desempeño que, de otro modo, podrían pasar inadvertidos, lo que facilita intervenciones pedagógicas más oportunas y efectivas. Innovar en educación, por tanto, no solo significa adoptar tecnologías emergentes, sino también generar una cultura docente basada en la reflexión, la evidencia y la mejora continua. Es en ese cruce entre tecnología y compromiso pedagógico donde surgen las verdaderas transformaciones con sentido (Ibrahim, 2020).

Este capítulo se concibe como una propuesta que entrelaza la reflexión pedagógica con la aplicación práctica de herramientas digitales avanzadas, particularmente aquellas derivadas de la IA. En un escenario educativo donde la abundancia de datos convive con una creciente complejidad en los perfiles del estudiantado, se vuelve necesario repensar las formas en que observamos, comprendemos y acompañamos los procesos de aprendizaje. Más allá de lo técnico, este trabajo se enmarca en la búsqueda de una enseñanza más cercana, que ponga en valor la diversidad y se nutra de nuevas formas de mirar el aula. La aplicación del algoritmo de agrupamiento *K-means* se plantea aquí como una vía para acercarse con mayor precisión y empatía a las trayectorias de los estudiantes y generar conocimiento útil para la acción docente (Peña-Ayala, 2014). A partir de esta intención, se despliegan los siguientes objetivos:

1. Comprender el lugar que puede ocupar la IA en la educación del presente al reconocer su potencial transformador en la forma de leer y acompañar el aprendizaje. Este primer objetivo busca situar la IA no como un conjunto de herramientas ajenas al quehacer docente, sino como una aliada que permite observar con nuevos ojos los procesos educativos. Se trata de ir más allá de los promedios o las calificaciones y adentrarse en una comprensión más rica, dinámica y contextual del aprendizaje (Holmes y Baker, 2022). Al acceder a datos generados en plataformas digitales o sistemas de gestión educativa, se abre la posibilidad de detectar patrones, anticipar dificultades y diseñar estrategias más ajustadas a lo que ocurre realmente en el aula. Esta mirada

implica, al mismo tiempo, asumir con responsabilidad los límites éticos de su uso y mantener el foco en las personas detrás de los datos (Williamson y Eynon, 2020).

2. Explicar el algoritmo *K-means* desde una lógica comprensible que conecte su estructura matemática con su sentido educativo. Lejos de ser un procedimiento meramente técnico, el *K-means* puede entenderse como un puente entre el análisis riguroso y la práctica pedagógica (Jain, 2010). En este capítulo se desglosa su funcionamiento paso a paso, mostrando cómo parte de datos concretos para organizar la diversidad en grupos significativos. Se aborda la noción de centroides, la forma en que se mide la similitud entre estudiantes y cómo el algoritmo se ajusta hasta encontrar agrupamientos coherentes. Más allá de la fórmula, el interés se centra en cómo esta herramienta puede ayudarnos a visibilizar realidades que muchas veces permanecen ocultas: estudiantes con trayectorias similares, desafíos comunes o estilos de aprendizaje afines (Papamitsiou y Economides, 2014).
3. Proponer una metodología que pueda ser utilizada por docentes e investigadores para implementar el *clustering* educativo de forma ética, rigurosa y contextualizada. Este objetivo busca traducir la propuesta en una hoja de ruta clara que oriente paso a paso la aplicación del análisis. Desde la formulación de una pregunta pedagógica significativa hasta la recolección, depuración y análisis de los datos, se plantea un modelo centrado en las decisiones que deben tomarse para que el proceso sea coherente con la realidad del aula (Slade y Prinsloo, 2013). No se trata de aplicar un algoritmo de forma automática, sino de seleccionar las variables adecuadas —como la participación, los tiempos de conexión o los estilos de aprendizaje— con una intención clara: comprender mejor a quienes aprenden (Ferguson, 2012). Esta propuesta metodológica está pensada tanto para investigaciones formales como para espacios educativos cotidianos donde el análisis de datos puede enriquecer la práctica docente.
4. Traducir los agrupamientos obtenidos en propuestas concretas de enseñanza diferenciada, capaces de responder a la diversidad real del aula. El valor del análisis no se agota en la segmentación de estudiantes, sino que comienza allí. Una vez reconocidos los patrones, lo esencial es preguntarse qué hacer con esa información. Este objetivo se orienta a diseñar respuestas pedagógicas específicas, que vayan desde el apoyo personalizado para estudiantes con

mayores dificultades hasta itinerarios de aprendizaje más autónomos para quienes muestran mayor capacidad de autorregulación (van Leeuwen *et al.*, 2014). También se propone el uso del trabajo colaborativo como vía para potenciar las fortalezas distribuidas en los grupos. En todos los casos, el foco está en adaptar la enseñanza a los distintos ritmos, intereses y formas de aprender, reconociendo que una propuesta homogénea ya no basta para responder a la complejidad del aula actual (Ifenthaler y Yau, 2020).

5. Promover una cultura de análisis crítico de datos en educación, que complemente la experiencia y el juicio docente sin reemplazarlos. Finalmente, el capítulo plantea que el uso de herramientas basadas en IA debe estar guiado por una ética del cuidado, la equidad y la mejora continua (Mandinach *et al.*, 2015). Lejos de promover una mirada reduccionista o mecanicista, se defiende la idea de que los datos, cuando se utilizan correctamente, pueden ofrecer pistas valiosas para actuar con mayor sensibilidad y justicia. En este sentido, el análisis no reemplaza al docente, sino que lo acompaña en su tarea de observar, interpretar y decidir. Fomentar una cultura pedagógica basada en evidencia, pero profundamente humana, es uno de los desafíos más urgentes en este tiempo de transformación educativa (Dwivedi, 2023).

En conjunto, estos objetivos no solo orientan el desarrollo del capítulo, sino que invitan a una forma distinta de pensar la tecnología en la educación. No se trata únicamente de incorporar algoritmos, sino de utilizarlos con propósito, criterio pedagógico y compromiso ético. Se espera que esta propuesta inspire a docentes, investigadores y tomadores de decisiones a explorar nuevas formas de conocer y acompañar a sus estudiantes al hacer del análisis de datos un acto profundamente humano y transformador.

Desarrollo

Los escenarios educativos actuales están marcados por una creciente diversidad de estudiantes, por ritmos de aprendizaje muy distintos y por entornos digitales que generan grandes volúmenes de información sobre cómo se enseña y cómo se aprende. Para los docentes, esta realidad supone una complejidad que no siempre puede abordarse con las herramientas tradicionales. En este contexto desafiante, la IA se perfila como un recurso capaz de ofrecer nuevas formas de comprender lo que ocurre en el aula y apoyar la toma de decisiones pedagógicas a partir de datos reales, más allá de la intuición o la experiencia acumulada (Aguilar, 2024).

Entre las múltiples aplicaciones de la IA en educación, los algoritmos de agrupamiento —en especial, el *K-means*— ofrecen una alternativa concreta para observar con mayor claridad los distintos perfiles de estudiantes dentro de un mismo grupo. Gracias a su capacidad para procesar grandes volúmenes de información y encontrar patrones que no son evidentes a simple vista, esta herramienta permite identificar conjuntos de estudiantes con características similares en aspectos clave como el rendimiento académico, la participación digital, la autonomía o la forma en que prefieren aprender (Araka *et al.*, 2022). Lejos de ser una clasificación rígida, estos agrupamientos se convierten en una fuente valiosa para diseñar estrategias que respondan realmente a lo que cada grupo necesita.

Lo interesante del algoritmo *K-means* es que no parte de etiquetas predefinidas ni de categorías impuestas desde fuera. Por el contrario, su lógica permite que los propios datos revelen cómo se organizan naturalmente los estudiantes según sus similitudes (Arslan *et al.*, 2023). En otras palabras, permite ver aquello que a menudo escapa a la observación cotidiana: que no todos aprenden de la misma forma, que los niveles de compromiso varían y que las trayectorias escolares pueden ser muy distintas, aunque pertenezcan al mismo curso.

Aplicado al aula, el algoritmo actúa como una especie de lente que ayuda a ordenar el caos aparente. A través de un proceso de ensayo y ajuste, define grupos internos que representan perfiles reales y observables dentro de la clase. Desde el punto de vista pedagógico, esta información es sumamente útil: permite al profesorado diseñar intervenciones específicas, distribuir mejor los apoyos, organizar el trabajo colaborativo con mayor coherencia y, en general, lograr que la enseñanza esté más alineada con quienes la reciben.

Ahora bien, para que este tipo de análisis tenga verdadero sentido, no basta con aplicar un procedimiento automático. Es fundamental que cada etapa del proceso responda a una intención educativa. Por ello, este capítulo propone una metodología concreta que va más allá del algoritmo: comienza por definir qué se quiere observar, continúa con la selección de variables que representen la experiencia de aprendizaje, prosigue con la preparación cuidadosa de los datos y culmina con una interpretación pedagógica que permita transformar los resultados en acciones concretas dentro del aula.

En este camino, también se abordan elementos clave como la ética en el uso de datos educativos, la necesidad de contextualizar las decisiones y la importancia

de mantener siempre el foco en la persona que aprende. Porque la IA, por sí sola, no mejora la educación: es el propósito con que se incorpora y la sensibilidad con que se interpretan sus resultados lo que marca la diferencia.

En definitiva, esta propuesta no busca reemplazar el juicio docente, sino enriquecerlo. Aporta herramientas para observar con mayor profundidad, actuar con más información y construir una enseñanza más consciente, justa y atenta a la diversidad real de cada grupo de estudiantes. La IA, bien utilizada, puede ser una aliada poderosa para quienes enseñan con vocación y sentido, y también para quienes aprenden desde realidades muy distintas, pero con el mismo derecho a ser comprendidos y acompañados en su proceso.

Metodología

Método estático: modelo matemático del Algoritmo *K-means*

Fundamento matemático del algoritmo *K-means*

Si bien el algoritmo *K-means* suele destacarse por su aparente simplicidad y por la rapidez con que puede procesar grandes volúmenes de información, su verdadero valor radica en la lógica matemática que sustenta su funcionamiento. Esta lógica no es arbitraria, sino que responde a un principio fundamental del análisis de datos: encontrar estructuras significativas agrupando elementos que compartan características similares (MacQueen, 1967). En otras palabras, busca organizar los datos de manera que los elementos dentro de un mismo grupo o clúster tengan más aspectos en común entre sí que con los pertenecientes a otros grupos (Pratama *et al.*, 2023).

Desde una perspectiva educativa, puede imaginarse que cada estudiante se representa como un punto en un espacio de varias dimensiones. Cada una de esas dimensiones corresponde a una variable relevante: el promedio de sus calificaciones, su nivel de participación en clases virtuales, la puntualidad en la entrega de trabajos o incluso su estilo de aprendizaje preferido (Huang, 2008). De este modo, la realidad académica de cada alumno se convierte en una coordenada única dentro de un espacio de análisis.

El propósito del algoritmo *K-means* es precisamente encontrar patrones en ese espacio: identificar conjuntos de estudiantes que, por sus comportamientos o características, tienden a agruparse de manera natural (Blömer *et al.*, 2016). Para ello, el algoritmo define un número predeterminado de grupos llamados clústeres y busca minimizar la distancia entre cada estudiante y el centro de su grupo,

también conocido como centroide (Vattani, 2011). Cuanto más cerca se encuentra un estudiante del centro de su grupo, más representativo resulta ese clúster de su realidad (Chaturvedi y Rajavat, 2013).

Este enfoque no solo permite clasificar a los estudiantes según sus similitudes, sino también descubrir formas más efectivas de acompañar sus procesos de aprendizaje. Así, el algoritmo deja de ser una herramienta puramente técnica para convertirse en un recurso valioso al servicio de la comprensión pedagógica.

Desde una perspectiva matemática, el algoritmo busca minimizar la suma de las distancias cuadráticas entre los puntos de datos y sus respectivos centroides. Esta suma se conoce como la función de coste o función objetivo, y se expresa así:

$$J = \sum_{i=1}^k \sum_{x \in S_i} \|x - c_i\|^2$$

Donde:

- J es el valor total de la función objetivo, que el algoritmo intenta minimizar.
- k es el número de clústeres definidos.
- S_i representa el conjunto de puntos asignados al clúster.
- x es un punto de datos (por ejemplo, un estudiante).
- C_i es el centroide del clúster i .
- $\|x - c_i\|^2$ representa la distancia euclidiana al cuadrado entre el punto x y su centroide.

El algoritmo *K-means* se basa en un procedimiento cíclico y progresivo mediante el cual ajusta sus cálculos hasta lograr una organización óptima de los datos (Tan *et al.*, 2019). Esta metodología no se ejecuta en un único paso, sino que sigue una serie de etapas repetidas que permiten refinar progresivamente los resultados. En términos simples, el algoritmo aprende y mejora su agrupamiento a medida que avanza en su procesamiento (Morissette y Chartier, 2013).

El proceso comienza con una fase conocida como inicialización, en la cual el sistema selecciona aleatoriamente un número determinado de puntos dentro del conjunto de datos que actuarán como centroides iniciales (Wang, 2023). En el contexto educativo, estos puntos podrían representar configuraciones hipotéticas de estudiantes que comparten ciertas características comunes.

Luego se desarrolla la etapa de asignación, en la que cada dato —en este caso, cada estudiante descrito por un conjunto de variables— se vincula al centroide más cercano (Pérez-Ortega *et al.*, 2019). Esta cercanía se calcula mediante una métrica matemática llamada *distancia euclidiana*, que permite establecer qué tan similar es un estudiante respecto al centro de cada grupo (Mirkin, 2011). De este modo, los estudiantes se organizan en clústeres según su afinidad con determinados perfiles.

La tercera etapa corresponde a la actualización. En ella, el algoritmo recalcula el centroide de cada grupo considerando ahora el promedio de todos los datos que han sido asignados a ese clúster (Cui *et al.*, 2013). Es decir, ajusta la posición del centro de cada grupo para que represente mejor a sus miembros actuales.

Este ciclo de asignación y actualización se repite sucesivamente hasta alcanzar una condición de estabilidad. Esto ocurre cuando los centroides ya no cambian de manera significativa entre una iteración y otra, o bien cuando se cumple un número máximo de repeticiones definido por el analista (Liu y Lu, 2015). El resultado final de este proceso iterativo son agrupaciones estables de estudiantes, en las que cada grupo refleja patrones de comportamiento, participación o rendimiento que comparten sus integrantes.

Uno de los aspectos más importantes al utilizar el algoritmo *K-means* es determinar cuántos grupos (*k*) deben formarse. Si se elige un número demasiado bajo, los grupos serán excesivamente generales y no reflejarán de manera adecuada la diversidad del alumnado; por el contrario, si se elige un número demasiado alto, se corre el riesgo de sobreajustar los datos y perder claridad en el análisis (Goyal y Aggarwal, 2017).

Para resolver esta cuestión, se emplea una técnica llamada método del codo (*elbow method*), que consiste en observar cómo varía la función objetivo del algoritmo —generalmente representada por la suma de las distancias cuadradas dentro de los grupos— a medida que se incrementa el número de clústeres (Lakshmi *et al.*, 2018).

Cuando se grafica esta relación, es común observar una curva decreciente. Sin embargo, llega un punto en el que las mejoras en la agrupación se vuelven cada vez menos significativas. Ese punto de inflexión, donde la pendiente de la curva cambia bruscamente, se asemeja a un codo en la gráfica (Basso *et al.*, 2024); de ahí proviene su nombre. Identificar ese punto es clave, ya que suele indicar el número óptimo

de agrupaciones que mejor equilibran la precisión del modelo con su simplicidad interpretativa.

Esta combinación entre matemática, lógica algorítmica y análisis visual hace del *K-means* una herramienta no solo potente, sino también accesible para el análisis educativo. Permite transformar conjuntos complejos de información en representaciones claras y accionables, lo que aporta un valor real a quienes buscan entender y mejorar los procesos de enseñanza y aprendizaje.

Método lógico: algoritmo *K-means*

Más que un simple procedimiento técnico, lo que aquí se propone es una forma de mirar el aula con otros ojos. Este algoritmo se plantea como una herramienta con propósito educativo, una guía que combina el análisis de datos con una intención pedagógica clara. La idea no es solo procesar información, sino utilizarla como un medio para comprender mejor a quienes aprenden, reconociendo que detrás de cada número hay historias, trayectorias y maneras distintas de participar en el proceso formativo (Sharma *et al.*, 2012).

Este enfoque busca acompañar tanto a docentes como a investigadores en la tarea de encontrar patrones o agrupaciones dentro de un grupo de estudiantes. Pero no cualquier agrupación, sino aquellas que realmente tengan sentido para orientar decisiones más conscientes en la enseñanza. La intención es que, a partir de esta mirada más afinada, se puedan diseñar estrategias que se ajusten mejor a los distintos perfiles del aula, haciéndola más inclusiva, más cercana y más humana.

Lo valioso de este método es que no se limita a ver los datos como algo distante. Más bien, invita a observar con atención las diferencias que existen entre los estudiantes al entender que esa diversidad no es un problema por resolver, sino una oportunidad para hacer de la educación un espacio más justo y significativo. Así, lejos de pretender sustituir el juicio pedagógico, este enfoque busca enriquecerlo y aportar nuevas herramientas para tomar decisiones que respondan mejor a las realidades que se viven en cada sala de clase.

Paso 1: clarificar el propósito pedagógico

Cualquier intento de aplicar IA en el ámbito educativo debería partir, ante todo, de una inquietud genuina. Mucho antes de pensar en herramientas o bases de

datos, lo fundamental es detenerse a reflexionar y formular las preguntas correctas (Chansa *et al.*, 2024): preguntas que surjan desde la experiencia cotidiana en el aula, desde las dudas que permanecen después de una clase, desde esa sensación de que algo importante aún no está del todo claro.

- ¿Qué aspectos de mis estudiantes todavía no logro percibir con claridad?
- ¿De qué manera podría agruparlos para comprender mejor sus formas de aprender y así apoyar más efectivamente su proceso?
- ¿Estoy buscando identificar necesidades particulares, estilos de aprendizaje o niveles de participación que no siempre se hacen visibles a simple vista?

Estas preguntas no tienen un origen técnico; nacen del ejercicio pedagógico, de la observación atenta y del deseo de mejorar lo que hacemos cada día. En este sentido, el punto de partida no está en el algoritmo ni en los datos, sino en la intención: en el propósito de usar la tecnología como una aliada para mirar con mayor profundidad y sentido a quienes aprenden con nosotros (Guo y Dacanay, 2024). Solo cuando existe claridad sobre el para qué, el análisis de datos cobra verdadero valor en la educación.

Paso 2: seleccionar las variables que representen la experiencia de aprendizaje

No todos los datos tienen el mismo valor ni aportan por igual a la comprensión profunda del proceso educativo. En este punto, la experiencia docente y el conocimiento del entorno en el que se desarrolla el aprendizaje se vuelven elementos esenciales (Dash y Kumar, 2024). No se trata de recolectar información de forma indiscriminada, sino de elegir con intención aquellas variables que realmente permitan captar lo que ocurre con los estudiantes, más allá de lo evidente.

Es precisamente esta mirada pedagógica la que orienta qué aspectos observar. Por ejemplo, los promedios o calificaciones pueden ofrecer una primera aproximación al rendimiento académico, pero no bastan por sí solos para comprender la complejidad del aprendizaje. También resulta clave considerar otras dimensiones menos tradicionales, como la frecuencia con que un estudiante participa en foros o actividades colaborativas, su nivel de interacción en clases virtuales o el tiempo que dedica a explorar recursos en línea (Mahamuni *et al.*, 2024).

Asimismo, reconocer los estilos de aprendizaje predominantes puede ser una puerta de entrada para adaptar mejor las estrategias de enseñanza. Otros indicadores, como las horas de conexión a la plataforma o la puntualidad en la entrega de tareas, aportan pistas sobre el compromiso, la autonomía y la manera en que cada estudiante se organiza frente a las demandas del entorno virtual.

Seleccionar estas variables con criterio no solo enriquece el análisis, sino que permite construir una representación más amplia, justa y plural de quienes están al otro lado de la pantalla o del pupitre. Se pasa así de una visión limitada, centrada exclusivamente en números, a una comprensión más matizada y significativa de la experiencia educativa (Quiroz *et al.*, 2024).

Paso 3: recolectar, depurar y preparar los datos

Una vez que se ha definido con claridad qué aspectos del aprendizaje se desean observar mediante la elección de variables significativas, llega el momento de dar un paso crucial: recopilar los datos. Esta tarea, que podría parecer meramente operativa, es en realidad una etapa estratégica, ya que de la calidad y pertinencia de los datos dependerá, en gran parte, la utilidad de todo el análisis posterior.

Las fuentes de información pueden ser diversas. En los entornos educativos actuales, es común recurrir a plataformas de gestión del aprendizaje (LMS), donde se almacenan registros de acceso, participación en foros, entregas de tareas y resultados de evaluaciones (Rachovski *et al.*, 2024). También es posible complementar esta información con datos provenientes de sistemas institucionales, como el historial de notas, o incluso con instrumentos creados por el propio docente, tales como encuestas diagnósticas o formularios de autoevaluación (Dzhanegizova *et al.*, 2023).

Sin embargo, obtener los datos es solo el comienzo. Para que estos tengan sentido y puedan utilizarse con fines analíticos, es necesario realizar un proceso de depuración y preparación. Esta etapa, conocida como limpieza y transformación de datos, implica revisar cuidadosamente la información recopilada y aplicar distintos criterios técnicos, siempre con una lógica pedagógica detrás. Algunas de las acciones más comunes incluyen:

- Detectar y eliminar registros incompletos o con errores evidentes que puedan distorsionar los resultados.

- Unificar escalas de medición, por ejemplo, cuando distintas plataformas usan criterios diferentes para evaluar el rendimiento.
- Codificar variables cualitativas, como estilos de aprendizaje o niveles de participación, para que puedan ser interpretadas por los algoritmos.
- Normalizar los datos, en caso de que existan valores desproporcionados que afecten la equidad del análisis.

Este trabajo, aunque muchas veces invisible, es fundamental. Garantiza que el conjunto de datos refleje de forma fiel lo que ocurre en el aula y no una imagen parcial o sesgada. Así, se sientan las bases para un análisis más riguroso, pero también más justo, que permita tomar decisiones pedagógicas informadas, coherentes y centradas en la diversidad real del estudiantado.

Paso 4: determinar el número óptimo de clústeres (k)

Uno de los momentos más importantes dentro del proceso de análisis es decidir cuántos grupos o clústeres conviene identificar dentro del conjunto de datos. Esta no es una elección menor ni puramente técnica; por el contrario, tiene profundas implicaciones en cómo se interpretarán las diferencias y similitudes entre los estudiantes y, por tanto, en cómo se diseñarán posteriormente las estrategias pedagógicas que respondan a esa diversidad (Baisova, 2021).

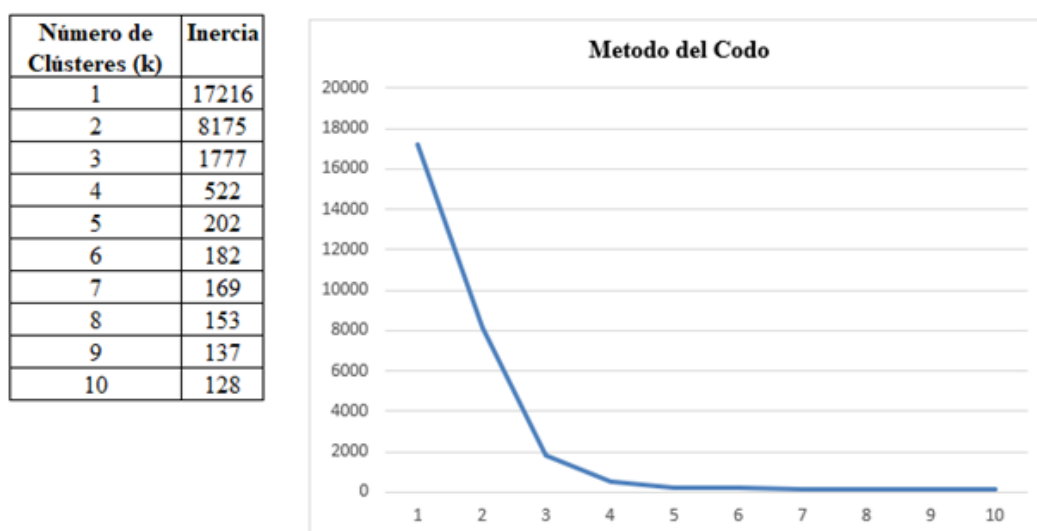
Para apoyar esta decisión, existe una técnica ampliamente utilizada conocida como el método del codo. Esta herramienta gráfica ofrece una forma visual y accesible de observar el comportamiento del modelo a medida que se incrementa el número de grupos. En términos simples, lo que se hace es calcular cuánto mejora la agrupación al añadir más clústeres (Wu, 2021). En las primeras etapas, cada nuevo grupo genera una mejora evidente en la precisión del modelo; sin embargo, llega un punto en que esa mejora comienza a ser marginal y deja de justificarse el aumento de complejidad.

Ese momento de quiebre o doblez en la curva —que suele adoptar la forma de un codo, de ahí su nombre— se interpreta como una señal natural del número óptimo de agrupamientos. Elegir ese punto permite alcanzar un equilibrio adecuado entre la simplicidad del modelo y su capacidad para representar la realidad del aula de manera significativa (Kodinariya y Makwana, 2013).

Desde una mirada pedagógica, optar por la cantidad justa de grupos no solo facilita el análisis técnico, sino que evita la fragmentación excesiva del estudiantado

y permite trabajar con categorías que realmente tengan sentido para el quehacer docente. No se trata de agrupar por agrupar, sino de descubrir patrones que aporten valor, que permitan ver lo que antes no se veía y actuar en consecuencia.

Figura 1. Método del codo para determinar número óptimo de clústeres



Aquí se presenta el gráfico del método del codo generado con datos simulados. Como puede observarse, la curva muestra una caída pronunciada en los primeros valores de k ; sin embargo, a partir de cierto punto —alrededor de $k = 5$ — las mejoras en la inercia se vuelven marginales. Ese punto de inflexión es lo que comúnmente se interpreta como el número óptimo de clústeres.

Paso 5: aplicar el algoritmo *K-means*

Una vez que se han definido cuidadosamente las variables que se van a analizar y se ha determinado el número adecuado de grupos a identificar —lo que, en términos técnicos, corresponde al valor de k —, se procede a aplicar el algoritmo de agrupamiento (Oyelade *et al.*, 2019). En este caso, el algoritmo *K-means* funciona como una herramienta que no solo organiza datos, sino que ayuda a revelar estructuras invisibles a simple vista dentro de un conjunto de estudiantes.

El procedimiento se basa en un proceso de ajuste progresivo que busca encontrar agrupaciones naturales dentro de los datos. Para ello, comienza seleccionando al azar k puntos que funcionarán inicialmente como centroides, es decir, representantes provisionales de cada grupo.

A continuación, cada estudiante —representado por sus variables previamente seleccionadas y normalizadas— se asocia al centroide más cercano. Esta cercanía se mide utilizando una fórmula matemática conocida como distancia euclidiana, que calcula qué tan similar es un estudiante respecto a cada uno de los centroides.

Luego, con los grupos ya formados, se realiza una actualización: cada centroide se ajusta en función del promedio de las características de los estudiantes que se han asignado a ese grupo. Este paso es clave, porque permite que los centroides se desplacen hacia posiciones más representativas dentro del espacio de datos.

Este proceso no ocurre una sola vez, sino que se repite varias veces de manera iterativa, hasta que las asignaciones dejan de cambiar de forma significativa. En otras palabras, se busca que el sistema llegue a un punto de estabilidad, en el cual los estudiantes ya no “migran” de un grupo a otro porque el agrupamiento ha alcanzado una estructura coherente y consistente.

El resultado final es una organización del conjunto estudiantil en clústeres estables que reflejan ciertos patrones comunes, ya sea en el rendimiento académico, la participación en entornos virtuales, los estilos de aprendizaje o cualquier otro conjunto de dimensiones previamente consideradas (Chen, J. *et al.*, 2020). Estos grupos no son artificiales ni forzados: emergen de los propios datos y permiten al docente o investigador mirar el aula desde una perspectiva más amplia y estructurada, basada en evidencia.

Paso 6: interpretar los resultados con una mirada pedagógica

Una vez que el algoritmo ha hecho su trabajo y los estudiantes han sido agrupados según los patrones encontrados en los datos, comienza quizás la etapa más valiosa del proceso: la interpretación con sentido humano y educativo (Trivedi, 2023). Porque más allá de los gráficos, las métricas o el rigor técnico, el verdadero potencial de esta herramienta se manifiesta cuando se conecta con el aula, con las personas reales que dan vida al proceso de aprendizaje.

El desafío ahora no es matemático, sino pedagógico: detenerse a observar con atención qué caracteriza a cada grupo, qué historias se ocultan tras esos patrones y, sobre todo, qué implicaciones tiene todo ello para la práctica docente. ¿Existen estudiantes que destacan por su constancia, pero no necesariamente por

su rendimiento? ¿Se observa un grupo que participa poco, pero entrega todas sus tareas con puntualidad? ¿Hay perfiles que requieren mayor acompañamiento o ajustes metodológicos específicos?

Estas preguntas no se responden con números, sino con mirada crítica, sensibilidad y disposición a leer entre líneas. Es en este punto donde el análisis cobra vida y se transforma en una herramienta poderosa para la toma de decisiones educativas más informadas, inclusivas y pertinentes.

Este momento invita al docente a salir del plano técnico y adentrarse en el terreno de lo humano, preguntándose con honestidad: ¿Qué me está diciendo esta información sobre mis estudiantes? ¿Qué realidades están detrás de estos grupos? ¿Cómo puedo responder mejor a sus necesidades desde mi rol?

La interpretación contextualizada implica reconocer que cada grupo identificado por el algoritmo representa una diversidad que ya estaba presente en el aula, pero que ahora se vuelve más visible. Y esa visibilidad puede convertirse en acción: en ajustes de estrategias didácticas, en el rediseño de actividades, en nuevas formas de evaluar o, simplemente, en la creación de espacios de diálogo más cercanos y personalizados (Vashishth *et al.*, 2024).

En definitiva, esta etapa final no busca cerrar el análisis, sino abrir nuevas preguntas y posibilidades para transformar la enseñanza desde una comprensión más profunda y respetuosa de quienes aprenden. Porque, en el fondo, los datos nos hablan, pero es el docente quien decide cómo escuchar y qué hacer con lo que ha descubierto.

Paso 7: diseñar estrategias diferenciadas

Cada grupo identificado por el algoritmo no es solo una categoría técnica o una segmentación abstracta. Muy por el contrario, cada clúster representa una puerta abierta hacia nuevas formas de entender y acompañar el proceso educativo. En lugar de ver los grupos como etiquetas fijas, se los puede concebir como puntos de partida para diseñar estrategias más sensibles, más cercanas y alineadas con las necesidades reales de quienes aprenden (Ozbey *et al.*, 2016).

Cuando se examinan con atención los perfiles que surgen del análisis, se comienza a visualizar con mayor claridad qué tipo de apoyos, desafíos o recursos podría beneficiar a cada conjunto de estudiantes. Por ejemplo, si un grupo

muestra un rendimiento por debajo del promedio y escasa participación en actividades colaborativas, esto puede ser una señal para diseñar intervenciones de acompañamiento más frecuente, que combinen refuerzo de contenidos con instancias de motivación y escucha activa.

En el caso de los estudiantes que se agrupan por su desempeño sobresaliente o por un alto nivel de autonomía, es posible pensar en rutas de aprendizaje más flexibles y desafiantes, que les permitan avanzar a su propio ritmo y explorar contenidos más allá del currículo básico, respetando sus intereses y capacidad de autorregulación (Borna *et al.*, 2024).

Por otro lado, puede presentarse el caso de clústeres formados por estudiantes que, aunque no destacan individualmente, muestran fortalezas complementarias. Allí surge la posibilidad de promover estrategias de trabajo colaborativo y fomentar el desarrollo de habilidades sociales y académicas en un entorno de apoyo mutuo, donde cada uno pueda aportar desde lo que sabe o desde su manera particular de aprender. Incluso la retroalimentación docente puede enriquecerse a partir de este enfoque: en lugar de emitir comentarios generales, el análisis de datos permite ofrecer retroalimentación personalizada y más significativa, basada en el nivel de participación, el estilo de aprendizaje predominante o las áreas que requieren mejora (Grasse *et al.*, 2023).

En síntesis, el análisis no termina en la segmentación: empieza allí. Lo verdaderamente transformador ocurre cuando esos datos se traducen en decisiones pedagógicas concretas. Cuando la información deja de ser solo un diagnóstico para convertirse en una herramienta que inspira la acción. Una acción que no busca homogeneizar, sino responder con sentido, intención y empatía a la diversidad que siempre ha estado presente en el aula, pero que ahora puede verse con más claridad (Piedra-Castro, 2024).

Paso 8: retroalimentar el proceso y repetir con mejoras

Finalmente, es importante comprender que este tipo de análisis no debe concebirse como una fotografía fija ni como un ejercicio puntual sin continuidad. Muy por el contrario, debe entenderse como parte de un proceso dinámico, abierto y en constante evolución, en el cual la información obtenida se convierte en una herramienta viva que dialoga con la realidad cambiante del aula (Janaki y Mariyappan, 2024).

El verdadero valor del *clustering* no reside únicamente en su capacidad para clasificar estudiantes en un momento determinado, sino en su potencial para alimentar un ciclo de mejora continua (Rosales-García, 2024). Una vez que se han diseñado e implementado estrategias pedagógicas basadas en los agrupamientos, se vuelve esencial detenerse a observar qué efectos han tenido esas decisiones: ¿se han reducido las brechas de participación? ¿Ha mejorado el rendimiento en los grupos más rezagados? ¿Se ha fortalecido la autonomía en los estudiantes más avanzados?

Este tipo de preguntas da lugar a una nueva etapa: la recopilación de nuevas evidencias, ya sea mediante los mismos indicadores utilizados en el análisis inicial o incorporando nuevas variables que amplíen la mirada. En este contexto, volver a aplicar el algoritmo —ya con datos actualizados— permite identificar si se han producido cambios significativos en los perfiles estudiantiles y, en consecuencia, ajustar nuevamente las estrategias (Xiao *et al.*, 2025).

En otras palabras, el *clustering* puede y debe repetirse en distintos momentos del proceso educativo, no como una rutina mecánica, sino como un gesto consciente de escucha pedagógica. De este modo, el aula deja de verse como un sistema estático y pasa a reconocerse en toda su complejidad: un espacio vivo, cambiante, donde cada estudiante transita su propio camino de transformación.

Aceptar esta dinámica invita a abandonar la idea de certezas absolutas y a adoptar un enfoque más flexible, donde las decisiones se construyen con base en datos, pero también con intuición docente, empatía y apertura al cambio. En este sentido, el análisis se transforma en una herramienta al servicio de una práctica educativa más reflexiva, más informada y, sobre todo, más humana.

Resultado esperado

La aplicación de este algoritmo va mucho más allá de la simple generación de agrupamientos técnicos. Su verdadero valor radica en cómo permite enriquecer la mirada del docente sobre su propio grupo de estudiantes, ofreciéndole herramientas concretas para comprender con mayor profundidad las múltiples formas en que se manifiesta el aprendizaje en el aula.

Al utilizar este enfoque, se pasa de observar a los estudiantes como un conjunto uniforme, regido únicamente por promedios o métricas tradicionales, a reconocerlos como parte de un ecosistema complejo y diverso, donde cada

individuo aporta una experiencia, un ritmo, una necesidad y un potencial distintos. El algoritmo no solo revela patrones ocultos, sino que también invita a reflexionar pedagógicamente sobre lo que esos patrones significan en la práctica cotidiana.

El resultado de este proceso no se traduce únicamente en una mayor organización de la información, sino en una enseñanza más consciente y situada, construida desde la evidencia, pero también desde la empatía. Una enseñanza que se adapta, que escucha, que responde con sentido a quienes están aprendiendo y que se atreve a replantear sus métodos cuando la realidad lo requiere.

En este contexto, la IA no sustituye el criterio docente, sino que lo complementa. Se convierte en una aliada que aporta claridad en medio de la complejidad y permite tomar decisiones más acertadas, más justas y, sobre todo, más humanas. Porque, en última instancia, lo que se espera no es simplemente clasificar a los estudiantes, sino crear condiciones más favorables para que cada uno de ellos pueda aprender, crecer y desarrollarse con plenitud.

Conclusiones

A lo largo de este capítulo, se ha planteado un recorrido que trasciende la dimensión técnica del análisis de datos, adentrándose en una perspectiva pedagógica que busca comprender con mayor profundidad la diversidad de los estudiantes en contextos educativos reales. El uso del algoritmo *K-means*, lejos de ser un ejercicio meramente computacional, ha demostrado tener un potencial significativo cuando se articula con una intención educativa clara y con un enfoque ético y humano.

Uno de los principales aprendizajes de esta experiencia es que los estudiantes no pueden ni deben ser entendidos como un grupo uniforme. Cada uno transita el proceso de aprendizaje con un ritmo distinto, con niveles de motivación variados, con condiciones personales únicas y con formas particulares de relacionarse con los contenidos y con sus pares (Ahmad *et al.*, 2024). Esta heterogeneidad, aunque siempre ha estado presente en las aulas, a menudo se diluye cuando las decisiones pedagógicas se basan únicamente en promedios, calificaciones o criterios estandarizados. En este sentido, el uso del *clustering* ha permitido rescatar y visibilizar esa diversidad desde un enfoque más justo, empático y contextualizado.

Al trabajar con variables como la participación en entornos virtuales, la entrega de tareas, los estilos de aprendizaje o la interacción en espacios colaborativos, el

algoritmo ha permitido identificar agrupamientos que no solo reflejan datos, sino trayectorias de vida y formas de habitar el aula. Esta organización basada en afinidades reales, y no en supuestos, ofrece a los docentes una nueva manera de mirar a sus estudiantes: no como individuos aislados, sino como miembros de colectivos con características comunes que pueden ser comprendidas y acompañadas de forma más efectiva.

Otro aporte fundamental de esta investigación ha sido confirmar que la IA, cuando se integra con criterio y propósito, puede potenciar —y no sustituir— el rol del profesorado (Ward *et al.*, 2024). El análisis de datos no debe entenderse como una amenaza a la dimensión humana de la educación, sino como un complemento que ofrece información valiosa para tomar mejores decisiones. No se trata de automatizar la enseñanza, sino de enriquecerla. No se trata de reemplazar el juicio docente, sino de fortalecerlo con evidencias que surgen de la realidad cotidiana del aula.

Desde lo metodológico, la propuesta presentada aquí se construye sobre una secuencia clara y aplicable, que va desde la definición de un propósito educativo hasta la retroalimentación de los resultados obtenidos. Esta lógica de mejora continua permite no solo identificar problemáticas o brechas, sino también implementar ajustes y volver a analizar, lo que genera un ciclo de aprendizaje institucional que se sostiene en el tiempo (Kujur *et al.*, 2022). Lo más relevante es que este ciclo no depende exclusivamente de grandes infraestructuras tecnológicas, sino de una disposición a observar, interpretar y actuar con base en los datos que ya están disponibles en muchas prácticas escolares.

Tabla 4. *Principales hallazgos del análisis con el algoritmo K-means y sus implicancias pedagógicas*

Hallazgos	Descripción	Implicancia pedagógica
Visibilización de la diversidad del estudiantado	El análisis permitió identificar que los estudiantes no constituyen un grupo homogéneo, sino que presentan trayectorias y necesidades diferenciadas (Nurhidayati y Marzuki, 2023).	Promueve el diseño de estrategias pedagógicas que reconocen y valoran la diversidad, lo que favorece una enseñanza más inclusiva.

Agrupamientos basados en patrones reales	El algoritmo <i>K-means</i> generó clústeres a partir de variables significativas como la participación digital, el estilo de aprendizaje o la entrega de tareas (Torcate et al., 2020).	Permite planificar intervenciones más ajustadas y contextualizadas, orientadas a grupos con características compartidas.
Apoyo a la toma de decisiones docentes	Los resultados ofrecieron al profesorado evidencia concreta para entender mejor el comportamiento y el desempeño de sus estudiantes.	Fortalece el juicio docente con datos al brindar mayor precisión en la planificación y seguimiento de los aprendizajes.
Potencial para retroalimentar procesos pedagógicos	La metodología puede repetirse en distintos momentos del curso, lo que permite observar cambios y ajustar las estrategias aplicadas (Chi, 2021).	Favorece una cultura de mejora continua basada en ciclos de análisis, acción y reflexión.
Articulación entre tecnología y propósito educativo	El uso de IA no se presenta como una solución técnica aislada, sino como un recurso al servicio del acompañamiento pedagógico (Gu et al., 2022).	Refuerza la necesidad de integrar herramientas digitales con sentido crítico y ético, lo que mantiene siempre el foco en el bienestar estudiantil.

En cuanto a las implicaciones prácticas, el uso del *clustering* permite una mejor distribución de recursos, tiempos y esfuerzos. Agrupar a los estudiantes a partir de evidencias concretas abre la posibilidad de diseñar experiencias de aprendizaje más pertinentes y significativas, que respeten los distintos niveles de autonomía, las necesidades de apoyo y las potencialidades de cada grupo. Esto contribuye a la construcción de aulas más equitativas, donde todos puedan avanzar desde su propio punto de partida y no desde un ideal homogéneo que no refleja la realidad (Restuono et al., 2024).

De cara al futuro, los caminos que se abren son múltiples. Por un lado, es posible seguir explorando nuevas variables que enriquezcan el análisis, incluyendo dimensiones socioemocionales, contextos familiares o niveles de acceso tecnológico. Por otro, pueden combinarse estas técnicas con otras herramientas de análisis predictivo que permitan anticipar posibles dificultades o detectar a tiempo señales de desconexión escolar (Bah, 2024). También será clave ampliar esta metodología a otros niveles educativos y contextos geográficos, evaluando su aplicabilidad en distintas realidades.

Además, se vuelve urgente seguir profundizando en las reflexiones éticas que rodean el uso de datos en educación: ¿qué información es pertinente recolectar?, ¿quién tiene acceso a esos datos?, ¿cómo se protege la privacidad de los estudiantes? y ¿de qué manera se asegura que las decisiones derivadas del análisis no reproduzcan desigualdades, sino que las corrijan?

En resumen, esta experiencia no propone una fórmula cerrada ni una solución definitiva, sino una invitación a mirar con otros ojos: a reconocer que los datos no son enemigos de la pedagogía, sino aliados potenciales cuando se los aborda con respeto, conciencia crítica y una clara orientación hacia el bienestar del estudiantado. En tiempos en los que la tecnología avanza velozmente, lo fundamental no es adaptarse ciegamente, sino decidir con intención cómo integrarla y para qué. Y si ese *para qué* tiene como centro a los estudiantes y sus procesos de aprendizaje, entonces estaremos en el camino correcto.

Referencias

- Aguilar, S. (2024). AI in the classroom: How teachers make ethical judgments. *USC Center for Generative AI and Society*.
- Ahmad, K., Iqbal, W., El-Hassan, A., Qadir, J., Benhaddou, D., Ayyash, M. y Al-Fuqaha, A. (2024). Data-driven artificial intelligence in education: A comprehensive review. *IEEE Transactions on Learning Technologies*, 17, 12–31. <https://doi.org/10.1109/TLT.2023.3314610>
- Anuradha, C., Velmurugan, T. y Anandavally, R. (2015). Clustering algorithms in educational data mining: A review. *International Journal of Power Control and Computation*, 7(1), 47–52. https://www.researchgate.net/publication/317166682_Clustering_Algorithms_in_Educational_Data_Mining_A_Review
- Araka, E., Oboko, R., Maina, E. y Gitonga, R. (2022). Using educational data mining techniques to identify profiles in self-regulated learning: An empirical evaluation. *The International Review of Research in Open and Distributed Learning*, 23(1), 131–162. <https://doi.org/10.19173/irrodl.v22i4.5401>
- Arslan, E., Özkaymak, A. y Nesrin, Ö. (2023). *Learning style detection using K-means clustering: Clustering students based on gamification user types and learning styles*. arXiv. <https://arxiv.org/abs/2310.14430>
- Bah, F. A. (2024). Artificial intelligence in education system. *International Journal of Advanced Research in Science, Communication and Technology*, 4(3), 439–443. <https://ijarsct.co.in/Paper22863.pdf>

- Baisova, G. (2021). Integration of artificial intelligence into educational programs to develop scientific analysis skills in a multidisciplinary environment. *Bulletin of Science and Practice*, 10(11), 410–416. <https://doi.org/10.33619/2414-2948/108/54>
- Barbakh, W.A., Wu, Y. y Fyfe, C. (2009). Review of clustering algorithms. En *Non-Standard Parameter Adaptation for Exploratory Data Analysis. Studies in Computational Intelligence*, (pp. 7–28). Springer. https://doi.org/10.1007/978-3-642-04005-4_2
- Basso, A., Kenmogne, E. B. y Frénay, B. (2024). Fast k-means with stable instance sets. En *2024 International Joint Conference on Neural Networks, IJCNN 2024 - Proceedings (Proceedings of the International Joint Conference on Neural Networks)*. IEEE. <https://doi.org/10.1109/IJCNN60899.2024.10650923>
- Blömer, J., Lammersen, C., Schmidt, M. y Sohler, C. (2016). *Theoretical analysis of the K-means algorithm: A survey*. arXiv. <https://arxiv.org/abs/1602.08254>
- Borna, M., Saadt, H., Tavassoli, H. y Akbari, E. (2024). Analyzing click data with AI: Implications for student performance prediction and learning assessment. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1421479>
- Chansa, C., Sain, Z. H., Shogbesan, Y. O., Phiri, E. V. y Akpan, W. M. (2024). Ethical implications of AI and machine learning in education: A systematic analysis. *International Journal of Instructional Technology*, 3(1), 1–13. <https://doi.org/10.33650/ijit.v3i1.9364>
- Chaturvedi, N. y Rajavat, A. (2013). An improvement in K-mean clustering algorithm using better time and accuracy. *International Journal of Programming Languages and Applications*, 3(4), 13–19. <https://doi.org/10.5121/ijpla.2013.3402>
- Chen, L., Chen, P. y Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75 264–75 278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Chen J., Zhang J., Wu J., Wu Y., Si H. y Lin K. (2020). Review on the research of K-means clustering algorithm in big data. *Proceedings of the 2020 International Conference on Electrical and Control Engineering (ICECE 2020)*. <https://doi.org/10.1109/ICECE51594.2020.9353036>
- Chi, D. (2021). Research on the application of K-means clustering algorithm in student achievement. En *Proceedings of the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 435–438). IEEE. <https://doi.org/10.1109/ICCECE51280.2021.9342164>
- Cui, J., Liu, J. y Liao, Z. (2013). Research on K-means clustering algorithm and its implementation. En *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)* (pp. 1804–1806). Atlantis Press. <https://doi.org/10.2991/iccsee.2013.452>
- Dash, S. y Kumar, P. (2024). Exploring the intersection of education and artificial intelligence: A comprehensive review. *International Journal of Multidisciplinary*

- Approach Research and Science*, 2(2), 601–610. <https://doi.org/10.59653/ijmars.v2i02.637>
- Dutt, A., Aghabozrgi, S., Ismail, M. A. B. y Mahroeian, H. (2015). Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics Engineering*, 5(2), 112–116. <https://ijee.org/index.php/ijee/article/view/396>
- Dwivedi, Y. K. (2023). Ethical AI for teaching and learning. *Cornell University Center for Teaching Innovation*.
- Dzhanegizova, A., Nurseit, A. M. y Vyborova, K. S. (2023). Artificial intelligence in education: Analysis of dynamics, perception, and prospects for integration. *Qainar Journal of Social Science*, 2(4), 34–49. <https://doi.org/10.58732/2958-7212-2023-4-34-49>
- Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges* (Informe técnico). Knowledge Media Institute, The Open University. <http://kmi.open.ac.uk/publications/pdf/kmi-12-01.pdf>
- Frades, I. y Matthiesen, R. (2010). Overview on techniques in cluster analysis. *Methods in Molecular Biology* (Clifton, N.J.), 593, 81–107. https://doi.org/10.1007/978-1-60327-194-3_5
- Goyal, M., y Aggarwal, S. (2017). A review on K-mode clustering algorithm. *International Journal of Advanced Research in Computer Science*, 8(7), 725–730. <https://doi.org/10.26483/ijarcs.v8i7.4301>
- Grasse, O., Mohr, A., Lange, A.-K. y Jahn, C. (2023). AI approaches in education based on individual learner characteristics: A review. En *Proceedings of the 2023 IEEE 12th International Conference on Engineering Education (ICEED)* (pp. 50–55). IEEE. <https://doi.org/10.1109/ICEED59801.2023.10264043>
- Gu, R., Jing, X., Zhao, D., Cai, L. y Gao, H. (2022). Research and application of improved *K-means* algorithm based on educational big data. En *Proceedings of the 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)* (pp. 746–749). IEEE. <https://doi.org/10.1109/ICCEAI55464.2022.00157>
- Guo, Z. y Dacanay, M. L. D. (2024). Development of an AI moral education assessment tool: Identifying and validating key factors through factor analysis. *International Journal of Education and Humanities*, 17(1), 17–23. <https://doi.org/10.54097/z5zpw272>
- Holmes, W. y Baker, T. (2022). *Artificial intelligence in education: Promises and implications for teaching and learning* (2.^a ed.). Center for Curriculum Redesign.
- Huang, X. (2008). Clustering analysis and algorithms. En V. Sugumaran (ed.), *Intelligent information technologies: Concepts, methodologies, tools, and applications*

- (pp. 389–396). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-59904-941-0.ch021>
- Ibrahim, M. S. (2020). Learning style detection using K-means clustering. *FUDMA Journal of Sciences*, 4(3), 375–381. <https://doi.org/10.33003/fjs-2020-0403-351>
- Ifenthaler, D. y Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68, 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>
- Jain A. K. (2008). Data clustering: 50 years beyond K-means. En W. Daelemans, B. Goethals y K. Morik (eds.), *Machine learning and knowledge discovery in databases. ECML PKDD 2008* (vol. 5211). Springer. https://doi.org/10.1007/978-3-540-87479-9_3
- Jain A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Janaki T. y Mariyappan M. (2024). Exploring the influence of AI-powered learning tools on student understanding and academic performance: A comprehensive analysis. *Shanlax International Journal of Management*, 18–26. <https://doi.org/10.34293/management.viisi-may.7834>
- Kodinariya T. M. y Makwana P. R. (2013). Review on determining number of cluster in K-means clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), 90–95. <https://n9.cl/hy754>
- Kujur A. G. P. y Tiwari R. Kr. (2022). Study and design of decision-making models for education system using artificial intelligence. *Proceedings of the 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, 434–439. <https://doi.org/10.1109/ICRTCST54752.2022.9781903>
- Kurniawan, R. W. D. y Hermawan, A. (2023). Personalized student learning mechanisms using K-means and K-medoids clustering algorithms based on individual preferences. *International Journal of Computer Applications*, 185(46), 13–19. <https://doi.org/10.5120/ijca2023923274>
- Lakshmi, K., Visalakshi, N. K. y Shanthi, S. (2018). Data clustering using K-means based on Crow Search Algorithm. *Sādhanā*, 43, 190. <https://doi.org/10.1007/s12046-018-0962-3>
- Liu, H. M. y Lu, J. G. (2015). Brief survey of K-means clustering algorithms. En *Applied Mechanics and Materials* (Vol. 740, pp. 624–628). Trans Tech Publications, Ltd. <https://doi.org/10.4028/www.scientific.net/amm.740.624>
- Liu, Q., Ba, S., Huang, J., Wu, L. y Lao, C. (2017). A study on grouping strategy of collaborative learning based on clustering algorithm. En S. Cheung, L. F. Kwok, W. Ma, L. K. Lee y H. Yang (Eds.), *Blended learning. New challenges and*

- innovative practices. ICBL 2017. Lecture Notes in Computer Science* (Vol. 10 309, pp. 284-294). Springer, Cham. https://doi.org/10.1007/978-3-319-59360-9_25
- Liu, R. (2022). Data analysis of educational evaluation using K-means clustering method. *Computational Intelligence and Neuroscience*. <https://doi.org/10.1155/2022/3762431>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. En L. M. Le Cam y J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume I: Statistics* (Vol. 5.1, pp. 281-297). Berkeley Symposium on Mathematical Statistics and Probability.
- Mahamuni, A. J., Parminder y Tonpe, S. S. (2024). Enhancing educational assessment with artificial intelligence: Challenges and opportunities. En *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1-5). <https://doi.org/10.1109/ickecs61492.2024.10616620>
- Mandinach, E. B., Parton, B. M., Gummer, E. S. y Anderson, R. (2015). Ethical and appropriate data use requires data literacy. *Phi Delta Kappan*, 96(5), 25-28. <https://doi.org/10.1177/0031721715569465>
- Mirkin, B. (2011). K-means and related clustering methods. En *Clustering for data mining: A data recovery approach* (pp. 221-281). Springer.
- Morissette, L. y Chartier, S. (2013). The K-means clustering technique: General considerations and implementation in Mathematica. *The Quantitative Methods for Psychology*, 9(1), 15-24. <https://doi.org/10.20982/tqmp.09.1.p015>
- Nurhidayati y Marzuki, I. (2023). Perhitungan matematis klasterisasi nilai mata kuliah mahasiswa menggunakan algoritma K-means. *AL JABAR Jurnal Pendidikan dan Pembelajaran Matematika*, 2(2), 83-92. <https://doi.org/10.46773/aljabar.v2i2.778>
- Omar, T., Alzahrani, A. y Zohdy, M. (2020). Clustering approach for analyzing the student's efficiency and performance based on data. *Journal of Data Analysis and Information Processing*, 8, 171-182. <https://doi.org/10.4236/jdaip.2020.83010>
- Oyelade, J., Isewon, I., Oladipupo, O., Emebo, O., Omogbadegun, Z. y Aromolaran, O. (2019). Data clustering: Algorithms and its applications. En *2019 19th International Conference on Computational Science and Its Applications (ICCSA)* (pp. 71-81). <https://doi.org/10.1109/ICCSA.2019.000-1>
- Ozbey, N., Karakose, M. y Ucar, A. (2016). The determination and analysis of factors affecting to student learning by artificial intelligence in higher education. En *2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1-6). <https://doi.org/10.1109/ITHET.2016.7760740>
- Papamitsiou, Z. y Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence.

- Journal of Educational Technology & Society*, 17(4), 49–64. <http://www.jstor.org/stable/jeductechsoci.17.4.49>
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4-1), 1432–1462. <https://doi.org/10.1016/j.eswa.2013.08.042>
- Pérez-Ortega, J., Almanza-Ortega, N. N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C. y Martínez-Rebollar, A. (2020). The K-means algorithm evolution. En K. Sud, P. Erdogmus y S. Kadry (Eds.), *Introduction to Data Science and Machine Learning* (pp. 1-24). IntechOpen. <https://doi.org/10.5772/intechopen.85447>
- Piedra-Castro, W. I. (2024). Aplicación de la gestión educativa con inteligencia artificial en la enseñanza en educación superior y las ciencias sociales. *Código Científico Revista de Investigación*, 5(2), 52–70. <https://doi.org/10.55813/gaea/ccri/v5/n2/538>
- Pratama, M. A. Y., Hidayah, A. R. y Avini, T. (2023). Clustering K-means untuk analisis pola persebaran bencana alam di Indonesia. *Jurnal Informatika dan Teknologi Komputer (JITEK)*, 3(2), 108–114. <https://doi.org/10.55606/jitek.v3i2.1506>
- Quiroz, L. A., Alvites, M. G. y Aquino, M. (2024). Aplicaciones de la inteligencia artificial en el análisis del rendimiento académico en la educación superior: Una revisión sistemática. *Micaela Revista de Investigación - UNAMBA*, 5(2), 25-32. <https://doi.org/10.57166/micaela.v5.n2.2024.153>
- Rachovski, T., Petrova, D. y Ivanov, I. (2024). Automated creation of educational questions: Analysis of artificial intelligence technologies and their role in education. *ENVIRONMENT. TECHNOLOGY. RESOURCES. Proceedings of the International Scientific and Practical Conference*, 2, 465-467. <https://doi.org/10.17770/etr2024vol2.8101>
- Restuono, L., Siahaan, A. P. U., Wijaya, R. F., Sitorus, Z. e Iqbal, M. (2024). Analysis and exploration of clustering algorithms for new student segmentation. *International Journal of Computer Sciences and Mathematics Engineering*, 3(1), 17–28. <https://doi.org/10.61306/ijecom.v3i1.61>
- Romero, C. y Ventura, S. (2020). *Educational data mining and learning analytics: An updated survey*. arXiv. <https://doi.org/10.48550/arXiv.2402.07956>
- Rosales-García, J. C. (2024). Estrategias didácticas para la mejora continua en el uso de la red social Facebook. *Revista Docentes 2.0*, 17(1), 193–199. <https://doi.org/10.37843/rted.v17i1.457>
- Sharma, R., Alam, M. A. y Rani, A. (2012). K-means clustering in spatial data mining using Weka interface. En *International Conference on Advances in Communication and Computing Technologies (ICACACT) 2012. Proceedings published by International*

- Journal of Computer Applications (IJCA)*. <https://research.ijcaonline.org/icacact/number1/icacact1006.pdf>
- Silva, G., Godwin, G. y Jayanagara, O. (2024). The impact of AI on personalized learning and educational analytics. *International Transactions on Education Technology (ITEE)*, 3(1), 36–46. <https://doi.org/10.33050/itee.v3i1.669>
- Slade, S. y Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1509–1528. <https://doi.org/10.1177/0002764213479366>
- Srivastava, J. y Kumar, R. (2021). Data mining in education sector: A review. *International Journal of Advanced Computer Science and Applications*, 615–620.
- Tan, P.-N., Steinbach, M. y Kumar, V. (2019). *Introduction to data mining* (2.^a ed.). Pearson.
- Torcate, A., Barbosa, J. y Rodrigues, C. (2020). Utilizando o learning analytics com o K-means para análise de dificuldades de aprendizagem na educação básica. En *Anais do XXVI Workshop de Informática na Escola* (pp. 31-40). <https://doi.org/10.5753/cbie.wie.2020.31>
- Trivedi, N. B. (2023). AI in education—A transformative force. En *2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI)* (pp. 1-4). Wardha, India. <https://doi.org/10.1109/IDICAIEI58380.2023.10406541>
- Tuomi, I. (2018). *The impact of artificial intelligence on learning, teaching, and education: Policies for the future*. European Commission. <https://dx.doi.org/10.2760/12297>
- van Leeuwen, A., Janssen, J., Erkens, G. y Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. *Computers & Education*, 79, 28-39. <https://doi.org/10.1016/j.compedu.2014.07.007>
- Vashishth, T. K., Sharma, V., Sharma, K. K., Kumar, B., Panwar, R. y Chaudhary, S. (2024). AI-driven learning analytics for personalized feedback and assessment in higher education. En T. Nguyen y N. Vo (eds.), *Using traditional design methods to enhance AI-driven decision making* (pp. 206-230). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-0639-0.ch009>
- Vattani, A. (2011). K-means requires exponentially many iterations even in the plane. *Discrete & Computational Geometry*, 45, 596–616. <https://doi.org/10.1007/s00454-011-9340-1>
- Wang, Z. I. (2023). K-means clustering. En D. Ghai, S. L. Tripathi, S. Saxena, M. Chanda y M. Alazab (Eds.), *Machine learning algorithms for signal and image processing* (pp. 194–211). Wiley-IEEE Press.
- Ward, B., Bhati, D., Neha, F. y Guercio, A. (2024). *Analyzing the impact of AI tools on student study habits and academic performance*. arXiv. <https://arxiv.org/abs/2412.02166>

- Williamson, B. y Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. <https://doi.org/10.1080/17439884.2020.1798995>
- Wu, B. (2021). K-means clustering algorithm and Python implementation. En 2021 *IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE)* (pp. 55-59). SC, USA. <https://doi.org/10.1109/CSAIEE54046.2021.9543260>
- Xiao, N., Pei, Y., Yuan, C., Bu, Y. y Cai, Z. (2025). Transforming education with artificial intelligence: A comprehensive review of applications, challenges, and future directions. *International Theory and Practice in Humanities and Social Sciences*, 2(1), 337–356. <https://doi.org/10.70693/itphss.v2i1.211>
- Xiong, Z., Li, H., Liu, Z., Chen, Z., Zhou, H., Rong, W. y Ouyang, Y. (2024). *A review of data mining in personalized education: current trends and future prospects*. arXiv. <https://arxiv.org/abs/2402.17236>
- Zawacki-Richter, O., Marín, V. I., Bond, M. y Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zheliznyak, I., Rybchak, Z. y Zavuschak, I. (2017). Analysis of clustering algorithms. En N. Shakhovska (ed.), *Advances in Intelligent Systems and Computing* (vol. 512, pp. 305–314). Springer, Cham. https://doi.org/10.1007/978-3-319-45991-2_21