

ORIGINAL

Model for Curriculum Analysis Based on Process and Data Mining

Modelo para el análisis curricular basado en la minería de procesos y datos

Juan Camilo Herrera Franco¹  , Darío Enrique Soto Duran²  , Aixa Eileen Villamizar Jaimes²  

¹Instituto Tecnológico Metropolitano ITM, Institución Universitaria, Facultad de Ciencias Económicas y Administrativas. Medellín, Colombia.

²Tecnológico de Antioquia, Institución Universitaria, Facultad de Ingeniería. Medellín, Colombia.

Citar como: Herrera Franco JC, Soto Duran DE, Villamizar Jaimes AE. Model for Curriculum Analysis Based on Process and Data Mining. Salud, Ciencia y Tecnología. 2025; 5:1494. <https://doi.org/10.56294/saludcyt20251494>

Enviado: 07-08-2024

Revisado: 20-11-2024

Aceptado: 05-03-2025

Publicado: 06-03-2025

Editor: Prof. Dr. William Castillo-González 

Autor para la correspondencia: Darío Enrique Soto Duran 

ABSTRACT

The article addressed the importance of curricular aspects in academic management, highlighting its role in learning and optimizing the curriculum. Student performance was analyzed by studying enrollment, academic progress and risk factors such as dropout, poor performance and failure of subjects. The application of process mining and data analysis made it possible to identify hidden patterns, facilitating pedagogical interventions to improve educational quality. Key concepts were defined, differentiating between curricular management and data analysis. At the curricular level, successful graduation, student dropout, and strategic curriculum planning were detailed. Regarding data management, predictive analytics was highlighted as a tool for decision-making in higher education. The study used a quantitative approach, structured in three phases: exploration, design and validation. In the exploratory phase, previous studies on graduation, curricular management and dropout were reviewed. During the design phase, process mining techniques were integrated to evaluate the student trajectory and detect bottlenecks. Finally, in the validation phase, the model was applied in an academic program, demonstrating its effectiveness in detecting curricular problems. The results indicated that the CRISP-DM methodology was the most used in educational data mining, providing a structured framework for the identification and resolution of academic problems.

Keywords: Process Mining; Curriculum Management; Data Analytics.

RESUMEN

El artículo abordó la importancia de los aspectos curriculares en la gestión académica, destacando su papel en el aprendizaje y la optimización del plan de estudios. Se analizó el desempeño estudiantil mediante el estudio de la matrícula, el avance académico y factores de riesgo como la deserción, el bajo rendimiento y la reprobación de asignaturas. La aplicación de minería de procesos y análisis de datos permitió identificar patrones ocultos, facilitando intervenciones pedagógicas para mejorar la calidad educativa. Se definieron conceptos clave, diferenciando entre gestión curricular y análisis de datos. En el ámbito curricular, se detalló la graduación exitosa, la deserción estudiantil y la planificación estratégica del plan de estudios. En cuanto a la gestión de datos, se resaltó la analítica predictiva como herramienta para la toma de decisiones en educación superior. El estudio utilizó un enfoque cuantitativo, estructurado en tres fases: exploración, diseño y validación. En la fase exploratoria, se revisaron estudios previos sobre graduación, gestión curricular y deserción. Durante la fase de diseño, se integraron técnicas de minería de procesos para evaluar la trayectoria estudiantil y detectar cuellos de botella. Finalmente, en la fase de validación, se aplicó el modelo en un programa académico, demostrando su efectividad en la detección de problemas curriculares. Los resultados indicaron que la metodología CRISP-DM fue la más utilizada en la minería de datos educativa, proporcionando un marco estructurado para la identificación y resolución de problemas académicos.

Palabras clave: Minería de Procesos; Gestión Curricular; Analítica de Datos.

INTRODUCCIÓN

Los aspectos curriculares son elementos fundamentales que soportan el aprendizaje en un proceso formativo. Por ello, es indispensable analizar el desempeño académico de los estudiantes a partir del comportamiento que presentan en la matrícula y avance del plan de estudios.

El análisis del plan de estudios permite identificar patrones que facilitan la detección temprana de factores de riesgo asociados a diferentes variables propias de la gestión curricular de un programa tales como: la deserción estudiantil, el bajo rendimiento académico graduación exitosa, asignaturas de mayor reprobación entre otros. Estos factores, como lo menciona ⁽¹⁾ propician ajustes a las rutas curriculares, cambios en las intervenciones pedagógicas y mejoras en la eficacia de los procesos de enseñanza y aprendizaje. Esto contribuye a la retención de estudiantes y optimiza los recursos educativos, elevando la calidad académica.

Por otro lado, el la minería de procesos y el análisis de datos juegan un papel crucial identifican patrones ocultos y desarrollar planes de mejora asertivos dado que se analiza el comportamiento de procesos y los datos que genera. Siendo la minería de procesos una disciplina que combina herramientas y técnicas basadas en la minería de datos para analizar los procesos de negocio, cuyo registro de eventos de ejecución real se encuentra disponible en sistemas de información.⁽²⁾

Ahora bien, la minería de procesos aplicada en el ámbito curricular, tomando bajo análisis el plan de estudios responde interrogantes claves: ¿cómo se están ejecutando los procesos académicos en la realidad? y ¿dónde se encuentran los cuellos de botella? Entre otras. Estas preguntas son cruciales para identificar los puntos críticos dentro de la trayectoria académica que afectan el avance de los estudiantes en el programa con el propósito de alcanzar una graduación exitosa.

Asimismo, la analítica de datos se centra en hacer las preguntas correctas para generar previsiones y modelos predictivos sobre el rendimiento académico de los estudiantes. Según, autores⁽³⁾ la analítica de datos es el proceso mediante el cual las instituciones pueden analizar no solo el presente y el pasado, sino también construir modelos que permitan predecir el futuro y tomar medidas proactivas que mejoren la calidad del programa.

Por lo tanto, el artículo plantea el diseño y validación de un modelo integral que combina técnicas de analítica de datos y minería de procesos para ofrecer una solución práctica y efectiva en el monitoreo del plan de estudios y así facilitar acciones de mejora en la gestión curricular que sirva de insumo para realizar cambios significativos en el currículo, respondiendo a las necesidades del estudiante y a los desafíos del entorno profesional. Entre los factores analizados se encuentran las rutas curriculares, análisis de conformidad y cuellos de botella que permiten establecer análisis a variables como la graduación exitosa, la deserción estudiantil, asignaturas de mayor reprobación entre otras.

Conceptos Relacionados

El artículo identifica dos dominios temáticos que interactúan para el desarrollo del Modelo de gestión curricular basado en el análisis de datos y procesos; Para ello, se definen conceptos en asociados en el plano curricular y en segunda instancia los relacionados con la disciplina de la gestión de datos.

Curricular

- Gestión Curricular: se refiere al conjunto de decisiones, procesos y acciones que permiten la planificación, implementación, evaluación y ajuste de los programas educativos con el fin de asegurar la calidad en la enseñanza y el aprendizaje.⁽⁴⁾
- Graduación Exitosa: se refiere a la culminación de un programa académico por parte de los estudiantes dentro de un tiempo determinado y con el cumplimiento de los requisitos establecidos por la institución educativa.⁽⁵⁾
- Deserción estudiantil: es un fenómeno complejo que abarca todo abandono parcial o total de un programa académico.⁽⁶⁾
- Plan de estudios: es el resultado de la planificación estratégica del currículo que integra la distribución de las asignaturas que tienen como base la lógica para el desarrollo de las competencias, estableciendo prerrequisitos que aseguren que los estudiantes cuentan con los conocimientos previos necesarios para afrontar asignaturas más avanzadas.⁽³⁾

Gestión de datos

- Análisis de datos: es el proceso mediante el cual las instituciones pueden analizar no solo el presente y el pasado, sino también construir modelos que permitan predecir el futuro.⁽³⁾

- Análisis de procesos: La minería de procesos es una disciplina que combina herramientas y técnicas basadas en la minería de datos para analizar los procesos de negocio, cuyo registro de eventos de ejecución real se encuentra disponible en sistemas de información.⁽²⁾

MÉTODO

El enfoque metodológico del estudio se desarrolla bajo un paradigma cuantitativo,⁽⁷⁾ permitiendo la recolección y análisis de grandes volúmenes de datos con el objetivo de identificar patrones en la deserción y graduación de estudiantes. Se utiliza un método deductivo para la formulación y verificación de hipótesis.

Se inició el estudio mediante una estrategia de extracción de reportes del Sistema de Información Académico de la institución, obteniendo datos históricos de cohortes estudiantiles, registros académicos y estados de matrícula. Paralelamente, se realizó una revisión sistemática de la literatura para identificar los factores clave que inciden en la deserción y graduación en programas de ingeniería, estableciendo las bases teóricas que sustentaron la revisión de investigaciones previas en el contexto universitario. Con esta información, se procedió a la fase de diseño del Modelo de Gestión Curricular, el cual integra técnicas de minería de procesos y análisis de datos para visualizar las trayectorias académicas y detectar patrones de comportamiento estudiantil. Para ello, se utilizaron herramientas como Bizagi para el modelado de procesos, Power BI para el análisis visual y Python para la generación de modelos predictivos, permitiendo abordar de manera integral las problemáticas de deserción y graduación identificadas. Posteriormente, en la fase de validación se implementó el modelo mediante un estudio de caso en el programa académico de Ingeniería de Software del TdeA, utilizando una base de datos inicial de 32 224 registros que, tras un riguroso proceso de limpieza y depuración, se redujo a un conjunto final de 2 234 registros confiables. Este análisis, que abarcó los semestres 2018-1 y 2018-2, facilitó una comparación longitudinal de los indicadores académicos, mediante la gestión, análisis y visualización de datos con herramientas como Python, Power Automate y Power BI. Durante esta fase se filtraron los registros relevantes para identificar estados académicos clave (inscripciones, retiros y aprobaciones) y se incorporaron variables adicionales, como las tasas de retención y deserción, proporcionando una visión completa del comportamiento académico. Finalmente, la integración de estos métodos permitió validar el modelo de forma exhaustiva, demostrando su capacidad para detectar patrones y optimizar el plan de estudios a lo largo de las diferentes fases del análisis curricular.

Para garantizar la replicabilidad del presente estudio, se han adoptado diversas estrategias alineadas con las mejores prácticas recomendadas en la literatura⁽⁸⁾ y⁽⁹⁾. En primer lugar, el análisis de datos se ha documentado de manera detallada, evitando la selección de variables únicamente con base en umbrales de significación arbitrarios, y, en su lugar, fundamentando las decisiones en criterios estadísticos robustos y en el contexto educativo. Además, se han desarrollado múltiples modelos predictivos alternativos para evaluar diferentes escenarios, en este caso, se creó 1 modelo para mejorar la validez de los hallazgos, siguiendo enfoques como el criterio de información. Asimismo, se han generado scripts en Python para el procesamiento y análisis de datos, los cuales pueden ser reutilizados por otros investigadores interesados en replicar el estudio en distintos contextos educativos.

La disponibilidad de los datos originales, en la medida en que la confidencialidad institucional lo permita, y la documentación de cada paso metodológico contribuyen a la transparencia del estudio. Finalmente, se fomenta el acceso abierto a los resultados y se promueve la discusión académica sobre la replicabilidad en la investigación educativa, fortaleciendo así la fiabilidad y aplicabilidad del modelo propuesto en otros programas académicos y universidades.

RESULTADOS

A continuación, se presentan los resultados obtenidos a partir de cada una de las fases del modelo.

A. Fase de exploración

Se presentan trabajos que instancian el análisis de datos aplicada al contexto educativo. Encontrando que los trabajos se enfocan en el ámbito curricular en tres⁽³⁾ aspectos: a) la graduación exitosa, b) gestión curricular y c) la deserción estudiantil. Los cuales se describen a continuación:

Categoría	Descripción del Trabajo
Graduación exitosa	Según ⁽⁵⁾ desarrolla el Modelo de Integración Académica. De acuerdo con ⁽¹⁰⁾ analiza los factores individuales y contextuales que influyen en la graduación exitosa.
Gestión curricular	Acorde con lo planteado por los autores ⁽⁴⁾ Examina la importancia de la planificación estratégica del currículo. Los autores ⁽¹¹⁾ Discuten la evaluación y el ajuste continuo del currículo como un proceso colaborativo.

Deserción estudiantil Según ⁽¹⁰⁾ proponen una visión multidimensional de la deserción. El autor ⁽⁵⁾ detalla las causas académicas, sociales y económicas de la deserción. Acorde con el trabajo realizado por ⁽¹²⁾, explican el análisis de los factores que contribuye al diseño de estrategias para mejorar la retención estudiantil.

En el análisis se evidencia que el marco metodológico más utilizados en proyectos de minería de datos es el CRISP-DM (Cross Industry Standard Process for Data Mining), que provee seis fases para su desarrollo: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación, como lo detalla.⁽¹³⁾

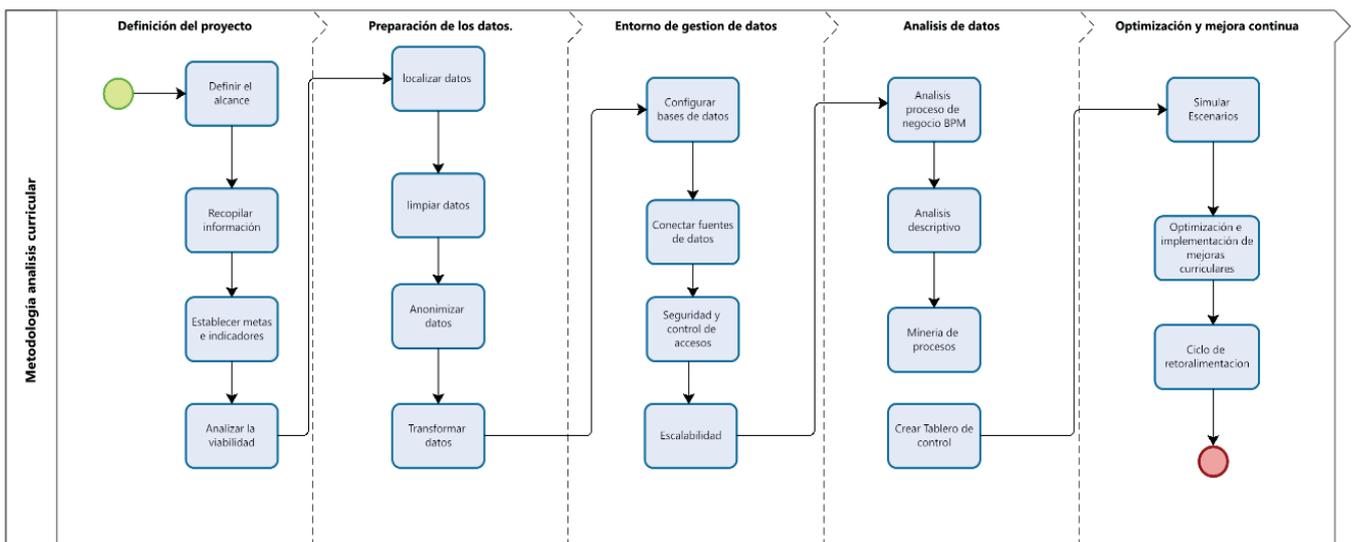
Este modelo proporciona una estructura sistemática que facilita la aplicación de técnicas de minería de procesos y analítica de datos, permitiendo identificar con precisión los puntos críticos en contexto académico, así como aplicar medidas correctivas.

A. Fase de Diseño

El modelo realiza el análisis del comportamiento del plan de estudios en el programa académico de educación superior para establecer a través del avance estudiantil las causas y situaciones que inciden en el desarrollo curricular del programa. Entre las variables que se analizan se tienen en cuenta: la deserción estudiantil y la graduación exitosa, a través de técnicas de analítica de procesos, que como lo menciona ⁽²⁾, son indispensables para el análisis de procesos curriculares. Asimismo, se implementan análisis de conformidad y cuellos de botella que se presentan en las rutas curriculares que presenta el plan de estudios de un programa académico.

Para estructura el modelo se toma como referencia la metodología CRISP-DM, estructurando el modelo en cinco fases que permiten comprender integralmente el análisis de proceso aplicado al programa académico a partir del plan de estudios.

En la siguiente figura, se puede observar la metodología del análisis curricular.



Powered by Bizagi Modeler

Figura 1. Modelo para analizar procesos curriculares

Definición del proyecto: Se establecen los objetivos, el alcance del análisis y los indicadores clave de desempeño (KPI) alineados con la institución. El currículo académico fue modelado en Bizagi BPM para visualizar rutas académicas, interdependencias entre asignaturas y posibles cuellos de botella que afectan el avance de los estudiantes. Primero, se extrajeron, transformaron y cargaron los datos utilizando Power Query y Power Automate, alojando la base de datos académica en Azure o SharePoint. A continuación, con Power Automate se analizaron las rutas de procesos que seguían los estudiantes a lo largo de sus trayectorias académicas, identificando las secuencias de eventos más frecuentes y las desviaciones significativas. Posteriormente, se realizó un análisis de causa raíz para detectar los problemas y desviaciones más críticos, examinando de forma detallada los datos y eventos asociados. Por último, se presentaron los resultados a través de informes generados en la misma plataforma, resaltando las mejoras propuestas y los indicadores clave de rendimiento. Dichos indicadores se midieron mediante métricas cuantitativas (por ejemplo, tasas de retención, tiempos

de finalización de cursos) obtenidas directamente de la base de datos académica y contrastadas con fuentes institucionales oficiales, garantizando así la fiabilidad de la información.

Los indicadores KPI fueron medidos mediante métricas cuantitativas basadas en fuentes de datos institucionales, asegurando un análisis preciso y ajustes fundamentados en los hallazgos obtenidos, como se muestra en la siguiente tabla:

Tabla 2. Indicadores KPI para el análisis del éxito de un programa académico

Ítem	Indicador	Descripción	Cálculo	Importancia
1	Tasa de aprobación	Mide el porcentaje de estudiantes que aprueban una asignatura	(Número de estudiantes que aprueban una asignatura / Número total de estudiantes inscritos) * 100	Evalúa si los estudiantes están comprendiendo el contenido de las asignaturas y están cumpliendo con los requisitos académicos.
2	Índice de deserción	Mide el porcentaje de estudiantes que abandonan el programa académico antes de finalizar sus estudios.	(Número de estudiantes que abandonan el programa / Número total de estudiantes matriculados) * 100	Identifica causas que impulsan a los estudiantes a abandonar el programa.
3	Tiempo promedio de graduación	Calcula el tiempo que tarda un estudiante en completar el programa académico desde su ingreso hasta la graduación.	Σ (Tiempo que cada estudiante tardó en graduarse) / Número total de estudiantes graduados	Evalúa la eficiencia del proceso curricular y si los estudiantes logran completar el programa en el tiempo esperado.
4	Satisfacción estudiantil	Determina el nivel de satisfacción de los estudiantes con aspectos del programa académico.	Promedio ponderado de respuestas en encuestas de satisfacción estudiantil.	Refleja si el programa cumple con las expectativas y necesidades de los estudiantes.
5	Tasa de retención	Mide el porcentaje de estudiantes que continúan sus estudios en el programa después de cada año.	(Número de estudiantes que continúan sus estudios en el próximo año / Número total de estudiantes matriculados) * 100	Indica si los estudiantes están comprometidos con el programa y si reciben el apoyo adecuado.
6	Tasa de repetición de asignaturas	Calcula el porcentaje de estudiantes que tienen que repetir una asignatura.	(Número de estudiantes que repiten una asignatura / Número total de estudiantes inscritos) * 100	Revela si existen problemas en la enseñanza o dificultad excesiva en los contenidos.
7	Utilización de recursos educativos	Mide la frecuencia con la que los estudiantes acceden y usan recursos educativos.	Registro de acceso a bibliotecas, plataformas de aprendizaje o laboratorios.	Permite evaluar el acceso a recursos educativos y su relación con el rendimiento académico.
8	Tasa de éxito en requisitos previos	Determina el porcentaje de estudiantes que aprueban con éxito los cursos que son requisitos previos para otras asignaturas.	(Número de estudiantes que aprueban los requisitos previos / Número total de estudiantes inscritos) * 100	Evalúa si los cursos previos están diseñados adecuadamente y si los estudiantes están bien preparados para asignaturas más avanzadas.
9	Progreso académico acumulado	Comprueba el porcentaje de créditos o asignaturas completadas en relación con el total requerido para graduarse.	(Número de créditos completados / Número total de créditos necesarios para graduarse) * 100	Monitorea el avance general de los estudiantes en el programa, identificando si están progresando según lo planificado.
10	Tasa de abandono de asignaturas	Mide el porcentaje de estudiantes que abandonan una asignatura antes de finalizarla.	(Número de estudiantes que cancelan una asignatura / Número total de estudiantes inscritos) * 100	Identifica asignaturas con alta tasa de abandono para ajustar su estructura o apoyo académico.
11	Participación en actividades extracurriculares	Valora el porcentaje de estudiantes que participan en actividades extracurriculares relacionadas con el desarrollo académico	(Número de estudiantes participantes en actividades extracurriculares / Número total de estudiantes matriculados) * 100	Evalúa la relación entre la participación en actividades complementarias y el rendimiento académico.
12	Acceso a tutorías y mentorías	Mide la frecuencia con la que los estudiantes acceden a servicios de mentoría académica.	Número de estudiantes que acceden a tutorías / Número total de estudiantes matriculados en el programa	Determina si los estudiantes están utilizando los recursos de apoyo disponibles y si estos impactan positivamente en su rendimiento.
13	Tasa de estudiantes con créditos atrasados	Determina el porcentaje de estudiantes que no han completado los créditos correspondientes a su semestre académico	(Número de estudiantes con créditos pendientes de semestres anteriores / Número total de estudiantes en ese nivel) * 100	Detecta problemas relacionados con el retraso en la acumulación de créditos, lo que puede afectar el tiempo de Graduación

El plan de estudios consta de 10 niveles con un sistema de prerrequisitos progresivamente complejo, donde la dependencia entre asignaturas se intensifica a partir del tercer semestre.

Preparación de datos: Se localizan, limpian y transforman los datos para su análisis, asegurando calidad, anonimización y compatibilidad con herramientas analíticas.

Gestión de datos: Se configura una base de datos segura e integrada en plataformas como Azure o SharePoint, garantizando escalabilidad y protección de la información.

Análisis y visualización: Se identifican patrones de rendimiento académico con herramientas como Power BI y Bizagi, optimizando rutas curriculares mediante dashboards interactivos y minería de procesos.

Modelos predictivos y optimización: Se implementa un modelo de aprendizaje automático con Python para predecir la deserción académica y proponer mejoras en el plan de estudios. Un ciclo de retroalimentación basado en un diagrama de Gantt permite ajustes continuos.

B. Fase de Validación:

Los resultados del estudio se basan en la implementación de un modelo de gestión curricular basado en el análisis de datos y la minería de procesos, presenta un caso de estudio en el que se analiza el programa de Ingeniería de Software del Tecnológico de Antioquia - Institución Universitaria (TdeA). A continuación, se presentan los principales hallazgos:

Tabla 3. Análisis por categorías

Categoría de análisis	Descripción	Hallazgos principales	Implicaciones
Deserción estudiantil	Factores que contribuyen al abandono del programa académico por parte de los estudiantes.	- Alta tasa de reprobación en asignaturas clave. - Cuellos de botella en segundo y tercer semestre.	- Necesidad de rediseñar el currículo para reducir la complejidad en etapas críticas. - Implementación de estrategias como tutorías personalizadas para mejorar la retención.
Graduación exitosa	Estrategias y condiciones que permiten a los estudiantes completar su programa académico a tiempo.	- Reducción del tiempo promedio de graduación en un semestre.	- Optimización de las trayectorias académicas para asegurar un progreso fluido. - Importancia de flexibilizar los prerrequisitos para adaptarse a las necesidades de los estudiantes.
Gestión del currículo académico	Planificación y ajuste continuo del plan de estudios para mejorar el desempeño y la retención.	- Mejora en la secuenciación de asignaturas. - Identificación de rutas curriculares eficientes. - Integración de herramientas de visualización y análisis de datos.	- Facilitar la toma de decisiones basadas en datos por parte de administradores y docentes. - Promover la actualización periódica del currículo para adaptarse a las tendencias académicas y laborales.
Minería de procesos aplicada al currículo	Análisis de rutas académicas y trayectorias estudiantiles mediante herramientas avanzadas como	- Identificación de puntos críticos en el progreso de los estudiantes.	- Uso de tecnologías avanzadas como Power Automate, para modelar y analizar datos. - Aplicación de simulaciones para predecir y mitigar problemas antes de que afecten a los estudiantes.
Análisis descriptivo con Python	Análisis predictivo de factores de riesgo relacionados con la deserción.	Identificación de estudiantes en riesgo mediante patrones detectados en datos históricos.	Aplicando Machine Learning, se creó un modelo predictivo utilizando Python que permite planificar estrategias preventivas.
Indicadores clave de desempeño (KPIs)	Métricas utilizadas para evaluar la efectividad del modelo de gestión curricular.	- Incremento en la tasa de retención y aprobación. - Reducción en los índices de repetición de asignaturas.	- Los KPIs permiten un monitoreo continuo del progreso institucional. - Su integración en tableros de control facilita la comunicación y la acción coordinada entre los diferentes actores del programa.
Impacto de las estrategias de intervención	Efectividad de las acciones implementadas para mitigar la deserción y mejorar el rendimiento.	- Reducción de la deserción en un 10 %. - Mayor aceptación de los estudiantes hacia las estrategias implementadas.	- Las intervenciones personalizadas y específicas son clave para abordar los factores multidimensionales de la deserción. - Importancia de realizar pruebas piloto y ajustar estrategias en función de los resultados obtenidos.
Pensum como proceso empresarial	Análisis del pensum académico como un proceso empresarial enfocado en la eficiencia y sostenibilidad.	Optimización de contenidos, eliminación de redundancias y alineación con las demandas del mercado laboral.	Implementación de un enfoque sistémico para diseñar el pensum considerando los objetivos institucionales y el perfil de egreso.

Cada categoría destaca un aspecto clave del problema abordado, los resultados obtenidos en este concuerdan con la literatura científica que resalta la importancia de una gestión curricular flexible y basada en datos para mitigar la deserción y promover la graduación exitosa en educación superior. Diversas investigaciones han evidenciado que factores como la integración académica y social, el acompañamiento institucional, la personalización de las trayectorias curriculares y el uso de herramientas analíticas avanzadas juegan un papel fundamental en la permanencia estudiantil.

⁽¹⁰⁾ Subraya que la integración académica y social del estudiante es un factor esencial para su permanencia en la universidad. Su teoría del desarrollo estudiantil sugiere que la participación activa en experiencias académicas y extracurriculares fomenta un sentido de pertenencia que disminuye el riesgo de abandono. En sintonía con este planteamiento, mientras que, ⁽⁵⁾ destaca la necesidad de repensar las acciones institucionales para incrementar la finalización de estudios, sugiriendo que el abandono estudiantil no solo responde a factores individuales, sino también a fallas en la estructura y el acompañamiento institucional. En este sentido, las universidades deben diseñar estrategias que no solo respondan a problemas académicos, sino también a factores psicosociales y económicos que afectan la continuidad de los estudiantes.

El uso de herramientas de analítica predictiva ha demostrado ser una estrategia efectiva para mejorar el éxito estudiantil, al permitir la identificación temprana de alumnos con mayor riesgo de abandono. ⁽¹⁴⁾, Enfatiza que la aplicación de modelos predictivos basados en machine Learning y minería de datos puede proporcionar información clave para diseñar intervenciones oportunas y personalizadas. En esta línea, mientras tanto, ⁽⁷⁾ sostiene que el análisis cuantitativo robusto es vital para tomar decisiones fundamentadas en la gestión académica. Mediante el procesamiento de datos históricos de los estudiantes, las instituciones pueden anticipar patrones de deserción y ajustar sus estrategias de retención en consecuencia.

Además, ⁽¹⁵⁾ y ⁽¹²⁾ señalan que el desarrollo de estrategias personalizadas y el seguimiento continuo a lo largo de la trayectoria académica constituyen prácticas efectivas para incrementar la retención. La implementación de tutorías, programas de mentoría y asesoramiento académico han demostrado tener un impacto significativo en la reducción de la deserción, ya que permiten a los estudiantes recibir apoyo específico según sus necesidades. En este contexto, la automatización de sistemas de alerta temprana puede facilitar la detección de dificultades académicas y emocionales, posibilitando la intervención proactiva de docentes y orientadores.

Por otro lado, la propuesta de optimizar el currículo con base en la minería de procesos y la analítica de datos coincide con estudios como el de ⁽¹⁶⁾, que demostraron cómo las técnicas de aprendizaje automático pueden guiar la creación de rutas curriculares más eficientes. Al analizar la progresión académica de los estudiantes y su relación con el desempeño en determinadas asignaturas, es posible ajustar la estructura de los planes de estudio para minimizar la deserción y mejorar la tasa de graduación. Asimismo, ⁽¹⁷⁾ resaltan el uso de modelos de predicción para anticipar la cantidad de graduados, lo que permite a las instituciones planificar mejor los recursos y optimizar la oferta académica.

La integración de la analítica académica con la toma de decisiones en educación superior es un aspecto clave señalado por diversos investigadores. ⁽¹⁸⁾ sostiene que la incorporación de herramientas de análisis de datos en la gestión universitaria puede proporcionar información precisa sobre las tendencias de retención y abandono, facilitando el diseño de estrategias más eficaces. Estas herramientas pueden incluir desde dashboards interactivos hasta sistemas de inteligencia artificial que monitorean el desempeño estudiantil en tiempo real.

Finalmente, la mejora de los planes de estudio y la implementación de estrategias de intervención oportunas pueden reducir significativamente la deserción en las instituciones de educación superior. Una gestión curricular basada en datos permite una mayor flexibilidad en las trayectorias académicas, adaptando la oferta de asignaturas según las necesidades y perfiles de los estudiantes. En este sentido, es fundamental que las universidades adopten un enfoque centrado en el estudiante, donde la innovación educativa y el soporte institucional sean ejes clave para garantizar la permanencia y el éxito académico.

Limitaciones del Estudio

A pesar de los hallazgos relevantes obtenidos en esta investigación, es necesario considerar una serie de limitaciones que pueden influir en la interpretación y generalización de los resultados:

Una de las principales limitaciones radica en la falta de normalización de los datos. Debido a la necesidad de realizar un proceso de limpieza y depuración, existe la posibilidad de que se hayan producido errores o sesgos al eliminar registros duplicados o inconsistentes. Esta situación puede afectar la exactitud de los resultados y, por ende, la validez de las conclusiones.

Variaciones en el plan de estudios (pensum), el cambio periódico de los planes de estudio dificulta la comparación longitudinal de los resultados, ya que las asignaturas, prerrequisitos y contenidos pueden variar significativamente entre una cohorte y otra. Esto limita la posibilidad de establecer patrones de deserción o graduación de manera homogénea a lo largo del tiempo.

Fuentes de datos limitadas, en algunos casos, la información no proviene de múltiples fuentes de datos, sino

de una sola base académica o de un repositorio institucional. Esta restricción reduce la capacidad de contrastar la información con otras fuentes que podrían enriquecer el análisis y disminuir posibles sesgos.

Restricciones de seguridad y acceso a la información, en el proceso de obtención de los datos, pueden presentarse medidas de seguridad que impidan el libre acceso a ciertos repositorios o bases de datos. Esto puede obstaculizar la recolección completa de la información, lo cual, a su vez, dificulta la realización de análisis más exhaustivos y detallados. Además, la presencia de políticas de privacidad puede restringir la publicación de ciertos datos sensibles, limitando la amplitud del estudio.

Recursos tecnológicos y conocimiento especializado, no todas las instituciones o investigadores cuentan con los recursos tecnológicos necesarios, entre esos, software de análisis, servidores, herramientas de minería de procesos; ni con el conocimiento especializado para manejar grandes volúmenes de información. Esto puede impedir la aplicación de la metodología propuesta en otros contextos educativos o profesionales.

Posibles sesgos en el manejo de datos, si bien los datos son reales, el proceso de limpieza y selección de variables puede introducir sesgos involuntarios. Sin embargo, estos se reducen significativamente al trabajar con información institucional oficial y registros académicos directos, lo que otorga un mayor grado de confiabilidad a los resultados.

Sugerencias para futuras investigaciones

La evidencia recopilada a lo largo de este estudio indica que la combinación de minería de procesos y análisis de datos puede ofrecer soluciones efectivas para reducir la deserción estudiantil y optimizar la gestión curricular. No obstante, con el fin de maximizar el potencial de este enfoque, se plantean varias líneas de investigación y mejora, una de ellas es la replicabilidad en otros programas y contextos; sería valioso replicar el modelo propuesto en programas académicos de otras áreas del conocimiento, por ejemplo, ciencias de la salud, ciencias sociales o artes, para determinar la solidez y generalizabilidad de los resultados. Asimismo, extender la aplicación a instituciones de distintos países permitiría evaluar cómo factores culturales, normativos y económicos influyen en la eficacia del modelo, ofreciendo una perspectiva comparativa más amplia.

Una sugerencia importante sería la integración de variables socioeconómicas y demográficas, entre ellos los factores económicos y familiares, incluyendo información relacionada con la situación financiera del estudiante, el nivel educativo de sus padres y la disponibilidad de becas o ayudas podría brindar un panorama más completo sobre las razones que inciden en la deserción o en el avance académico.

Las modalidades de estudio, podría ser un análisis que marcaría diferencias entre estudiantes presenciales, virtuales o híbridos y ayudaría a determinar si la modalidad de enseñanza afecta la efectividad del modelo y los índices de deserción.

Otro aspecto, sería la profundización en metodologías mixtas, incluyendo un enfoque cualitativo combinando el análisis de datos con metodologías cualitativas, como entrevistas en profundidad o grupos focales, puede revelar aspectos subjetivos que no se reflejan en las métricas numéricas, por ejemplo, motivación personal, expectativas o percepciones sobre la calidad de la enseñanza.

Por su parte, la triangulación de la información, podría contrastar los resultados cuantitativos con la evidencia cualitativa podría fortalecer la validez de las conclusiones y ofrecer una comprensión más integral de los procesos de deserción y graduación.

la optimización y evolución de la analítica predictiva con modelos más avanzados, explorando algoritmos de aprendizaje profundo o técnicas de ensamble para mejorar la precisión de la predicción de deserción, así como la identificación de perfiles de riesgo académico.

Actualización continua de los datos: Dado que los planes de estudio y las cohortes evolucionan con el tiempo, resultaría útil implementar un sistema de monitoreo constante que ajuste los modelos predictivos de forma dinámica a medida que cambian las condiciones académicas o administrativas.

CONCLUSIONES

La discusión de los resultados obtenidos en este estudio revela que la implementación del modelo puede generar un impacto en la gestión curricular basado en análisis de datos y minería de procesos. Uno de los hallazgos más significativos es que, a través de la implementación, es posible la identificación de los factores que contribuyen a la deserción estudiantil, entre los que destacan las altas tasas de reprobación en asignaturas clave, los cuellos de botella en semestres iniciales y las limitaciones en el seguimiento académico de los estudiantes. Estos problemas subrayan la necesidad de intervenciones específicas que mitiguen las barreras académicas y promuevan un progreso más fluido a lo largo del programa.

El análisis de las trayectorias académicas permitió mapear con precisión las rutas seguidas por los estudiantes, identificando puntos críticos en el currículo donde se concentra la mayor parte de las dificultades. Por ejemplo, el segundo y tercer semestre presentan altos índices de repetición y abandono, lo que genera acumulación de estudiantes y saturación de recursos en estas etapas.

La gestión curricular desempeña un papel fundamental en el éxito de este modelo. La planificación

estratégica del currículo, basada en datos y en la comprensión de las dinámicas académicas, puede permitir una mejor secuenciación de las asignaturas y la alineación de los objetivos pedagógicos con las necesidades del entorno. La minería de procesos aplicada a las rutas académicas ha sido una herramienta clave para poder entender cómo los estudiantes interactúan con el currículo y qué modificaciones pueden garantizar una experiencia de aprendizaje más efectiva.

El impacto del modelo de gestión no se limita al rendimiento académico; también puede abarcar aspectos emocionales y sociales que afectan la experiencia del estudiante. La integración de intervenciones personalizadas, como tutorías y programas de apoyo psicosocial, ha contribuido a un ambiente educativo más inclusivo y equitativo.

Uno de los hallazgos más relevantes es la capacidad del modelo para adaptarse y escalarse a otros programas académicos. La flexibilidad del enfoque y su base en tecnologías universales permiten su replicación en diferentes contextos, lo que lo convierte en una herramienta valiosa para instituciones que enfrentan desafíos similares. Además, la evaluación constante de indicadores clave de desempeño (KPIs) asegura que el modelo pueda ajustarse dinámicamente según las necesidades emergentes.

Es importante señalar que la implementación de este modelo también enfrenta desafíos. Entre ellos se encuentran la resistencia al cambio por parte de algunos actores, las limitaciones tecnológicas iniciales y la necesidad de formación continua del personal. Superar estas barreras requiere un enfoque integral que combine capacitación, comunicación efectiva y un liderazgo comprometido.

Finalmente, entre los hallazgos se destaca la eficacia del modelo de gestión curricular basado en el análisis de datos y procesos mejora el desempeño académico, reducir la deserción y optimizar la experiencia educativa. Este modelo no solo aborda problemas específicos del programa de ingeniería de software del Tecnológico de Antioquia, sino que también ofrece una hoja de ruta replicable para otras instituciones.

REFERENCIAS

1. Quiñones Huatangari L, Jara DM, Alvarado N, Milla ME, Gamarra OA. Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos. *Rev Cienc Tecnol*. 27 de noviembre de 2020;(34):45-50.
2. Aguirre Mayorga HS, Rincón García N. Minería de procesos: desarrollo, aplicaciones y factores críticos. *Cuad Adm [Internet]*. 26 de noviembre de 2015 [citado 4 de marzo de 2025];28(50). Disponible en: http://revistas.javeriana.edu.co/index.php/cuadernos_admon/article/view/14233
3. Treviño-Reyes R, Rivera-Rodríguez FS, Garza-Alonso JA. La analítica de datos como ventaja competitiva en las organizaciones. *Vinculatégica EFAN*. 18 de diciembre de 2020;6(2):1063-74.
4. 11 ideas clave: cómo aprender y enseñar competencias. 1a edición: septiembre 2007. Barcelona: Graó; 2007.
5. *Completing College: Rethinking Institutional Action* by Vincent Tinto.
6. Villegas BR, Núñez Lira LA. Factores asociados a la deserción estudiantil en el ámbito universitario. Una revisión sistemática 2018-2023. *RIDE Rev Iberoam Para Investig El Desarro Educ [Internet]*. 27 de mayo de 2024 [citado 4 de marzo de 2025];14(28). Disponible en: <https://www.ride.org.mx/index.php/RIDE/article/view/1923>
7. Novoa CAB. Investigación cuantitativa.
8. Aguilera Arévalo HE, Ramírez Contreras MG. Reproducibilidad, Replicación y Prácticas de Investigación Cuestionables en Estudios de Repositorios de Tesis de Maestría y Doctorado en México y Guatemala. *Rev Académica CUNZAC*. 5 de febrero de 2022;5(1):71-80.
9. Gorla DE. Reproducibilidad y replicabilidad en la investigación en ciencias naturales: ¿Hay una crisis?
10. Astin AW. *Student Involvement: A Developmental Theory for Higher Education*. octubre de 1999;VOL 40 NO 5:VOL 40 NO 5.
11. Fullan DM. INVESTIGACIÓN SOBRE EL CAMBIO EDUCATIVO: PRESENTE Y FUTURO. *Revista Mexicana de Investigación Educativa*. diciembre de 2010;15(núm. 47):1145.

12. Chiarino N, Rodríguez Enríquez C, Curione K, Machado A, Bonilla M, Aspirot L, et al. Abandono y permanencia estudiantil en universidades de Latinoamérica y el Caribe: Una revisión sistemática mixta. *Actual Investig En Educ.* 1 de mayo de 2024;24(2):1-37.

13. Schröer C, Kruse F, Gómez JM. A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Comput Sci.* 2021;181:526-34.

14. Bird K. Predictive Analytics in Higher Education: The Promises and Challenges of Using Machine Learning to Improve Student Success. *AIR Prof File [Internet].* 20 de diciembre de 2023 [citado 4 de marzo de 2025];(Fall 2023). Disponible en: <https://www.airweb.org/publication/Article-161>

15. Abdullah MA. SISTEM PEMBETUNGAN TERPUSAT: STRATEGI MELESTARIKAN PENGURUSAN KUMBAHAN. *Environ Manage.* 2001;

16. Kuz A, Morales R. Ciencia de Datos Educativos y aprendizaje automático: un caso de estudio sobre la deserción estudiantil universitaria en México. *Educ Knowl Soc EKS.* 26 de junio de 2023;24:e30080.

17. Ríos ZGD, Carrillo LCM, Rodríguez NCR. Territorios nómadas: nuevos lenguajes y subjetividades. Una experiencia sobre el lenguaje juvenil e infantil en la escuela. P.

18. Hurtado S, Suaza C, Aguilar J, Suescún E. Ciclo Autonomo de Analisis de Datos aplicado en la.

FINANCIACIÓN

Ninguna.

CONFLICTO DE INTERESES

Ninguno.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Curación de datos: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Análisis formal: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Investigación: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Metodología: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Administración del proyecto: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Recursos: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Software: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Supervisión: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Validación: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Visualización: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Redacción - borrador original: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.

Redacción - revisión y edición: Juan Camilo Herrera Franco, Darío Enrique Soto Duran, Aixa Eileen Villamizar Jaimes.