



CORHUILA

CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación



EduMetrics

DATA ANALYSIS SOLUTIONS

- ▶ Cindy Liliana Vargas Duque
- ▶ Jesús Ariel González Bonilla
- ▶ Luis Ángel Vargas Narváez
- ▶ Óscar Eduardo Chávarro Arias



Propuesta EduMetrics

Empresa ficticia: *EduMetrics*, una plataforma web educativa enfocada en el análisis de datos para mejorar los procesos y estrategias en la formación para la educación superior.

El problema: En el contexto de la educación superior en Colombia, donde el Ministerio de Educación Nacional, a través de sistemas como SPADIES, realiza seguimiento a la trayectoria y los resultados de los estudiantes, se identifica la problemática central de mejorar o elevar el rendimiento académico. Esto se refiere específicamente a asegurar que los estudiantes alcancen altos niveles de desempeño, logren la aprobación de sus asignaturas y desarrollen efectivamente las competencias definidas en sus programas formativos de educación superior.





1. ¿Qué es el análisis de datos?

Corresponde al proceso metodológico que permite abordar el entendimiento de los datos mediante funciones y algoritmos matemáticos. Entre las características principales, corresponde a los momentos claves para la toma de decisiones a partir de los datos. Para el análisis de datos se requiere abordar las etapas de inspeccionar, limpiar, transformar y modelar información.

Su propósito fundamental es descubrir patrones útiles, identificar tendencias significativas y extraer conclusiones relevantes que, de otro modo, permanecen ocultas en la masa de datos sin generar valor. Este conocimiento es crucial para fundamentar la toma de decisiones estratégicas, permitiendo a las organizaciones optimizar procesos, anticipar resultados y actuar con mayor eficacia.

Propuesta EduMetrics

Esta disciplina, es ampliamente utilizada en las diferentes industrias, pero su principal valor, corresponde en lograr predecir parte del comportamiento de un proceso, al igual que realizar una retrospectiva, esto ayudaría a tomar decisiones más acertadas.

Importancia para EduMetrics:

EduMetrics entiende la optimización del rendimiento académico como un factor crucial en la educación superior. Por tal motivo mediante una estrategia de análisis de datos, habilita a las instituciones para identificar con precisión las variables que inciden directamente en el desempeño estudiantil. Esto permite el diseño de intervenciones pedagógicas focalizadas y la implementación de planes de apoyo personalizados, fundamentados en la evidencia que los propios datos proporcionan, con el fin de potenciar el éxito académico y mejorar la efectividad de las estrategias formativas.





2. Roles en el análisis de datos

Para el desarrollo efectivo de la estrategia de análisis de datos en EduMetrics, se han definido los siguientes roles fundamentales, cada uno con responsabilidades específicas que aseguran la correcta implementación y el aprovechamiento de la información.

Tabla 1. Asignación de roles en el proyecto

Integrante	Rol asignado	Función principal
Jesús Ariel González Bonilla	<i>Data Scientist</i>	Diseña modelos de análisis predictivo y prescriptivo; interpreta patrones relevantes para el rendimiento académico.
Luis Ángel Vargas Narváez	<i>Data Engineer</i>	Responsable de la recolección, organización y limpieza de los datos simulados; asegura su calidad y estructura.
Oscar Eduardo Chavarro Arias	<i>Project Manager</i>	Coordina el equipo, define el cronograma, comunica avances y asegura que el proyecto siga la metodología acordada.
Cindy Liliana Vargas Duque	<i>Analista de datos y visualización</i>	Genera análisis descriptivos y diagnósticos; crea dashboards y visualizaciones para facilitar la toma de decisiones.

Fuente: Equipo EduMetrics



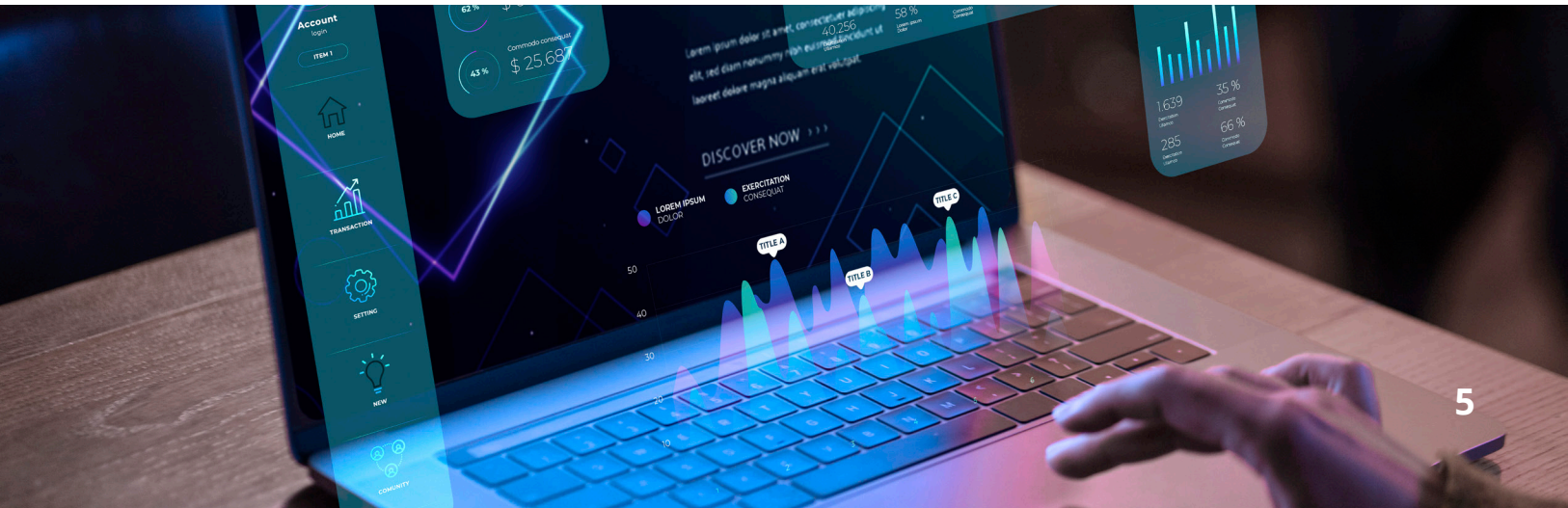


3. Tipos de análisis de datos aplicados

Tabla 2. Tipos de análisis aplicados al proyecto

Tipo de Análisis	Definición	Dato en EduMetrics
Análisis Descriptivo	Examina datos históricos para resumir y visualizar las características de un conjunto de información, mostrando que ha sucedido.	Promedio de calificaciones de los estudiantes en el programa Ingeniería de Sistemas en el periodo 2025A.
Análisis Diagnóstico	Profundiza en los datos para comprender ciertos eventos o patrones y las posibles causas que lo generan.	Correlación entre el número de horas dedicadas al estudio (auto-reportadas y registradas en plataforma) y las calificaciones obtenidas.
Análisis Predictivo	Utiliza datos históricos y algoritmos estadísticos o de machine learning para estimar la probabilidad de resultados futuros.	Puntuación de riesgo individual que indica la probabilidad de que un estudiante repruebe una asignatura basándose en su rendimiento previo y patrones de interacción con la plataforma.
Análisis Prescriptivo	Va más allá de la predicción, recomendando acciones específicas y óptimas que se pueden tomar para alcanzar un objetivo deseado o mitigar un riesgo.	Generación automática de un plan de acompañamiento personalizado, este incluye la asignación de tutorías específicas para los estudiantes identificados con alta probabilidad de reprobación una asignatura.

Fuente: Equipo EduMetrics





4. Metodología

Metodología seleccionada: Ciclo de vida del análisis de datos (CRISP-DM (Cross Industry Standard Process for Data Mining))

La metodología CRISP-DM (Cross Industry Standard Process for Data Mining) ha sido seleccionada para el desarrollo del proyecto EduMetrics por el equipo de trabajo, debido a su enfoque estructurado, flexible y ampliamente validado para resolver problemas mediante análisis de datos. Esta metodología permite abordar el proceso desde la comprensión del contexto educativo hasta la implementación de soluciones basadas en datos, facilitando una visión integral del rendimiento académico y su mejora continua.

Además, CRISP-DM se adapta eficientemente a distintos dominios, lo que la convierte en una opción idónea para contextos educativos donde intervienen múltiples fuentes de datos y actores (Chapman et al., 2000). Su carácter iterativo y su orientación hacia la toma de decisiones basadas en evidencia permiten que las instituciones de educación superior puedan transformar datos complejos en acciones estratégicas (Wirth & Hipp, 2000), lo que la hace especialmente pertinente para los objetivos de EduMetrics.

En el desarrollo del proyecto, se aplican las siguientes etapas:

Comprensión del problema:

Determina cómo el análisis de datos puede mejorar el rendimiento académico y disminuir la deserción.

Recolección de datos:

Simulación de un conjunto de datos con variables como: nombre del programa, promedio de notas, número de ausencias, participación en tutorías, y situación socioeconómica, entre otros, detallados en la siguiente tabla.

Análisis e interpretación:

- Se aplican métodos estadísticos y de machine learning para detectar patrones.
.....
- Se generan visualizaciones y reportes que evidencian los hallazgos.
.....
- Se plantean acciones a partir de las predicciones.
.....



Tabla 3. Fases detalladas aplicadas al proyecto en metodología CRISP-DM

CRISP-DM	Aplicación en el Proyecto EduMetrics
Entendimiento del negocio	<ul style="list-style-type: none"> - Identificar los principales factores que afectan el rendimiento académico en el programa académico seleccionado: Socioeconómicos, académicos (pedagógico, instrumentos de evaluación, calificaciones), asistencia, motivación. - Definir los objetivos del análisis: predecir bajo rendimiento, identificar patrones de éxito, diseñar alertas tempranas. - Involucrar a actores clave (docentes, estudiantes, psicólogas, administrativos) para comprender el contexto educativo.
Entendimiento de los datos	<ul style="list-style-type: none"> - Recopilar datos históricos académicos: notas, revisión de los microdiseños curriculares, asistencia, motivación (encuestas), caracterización de estudiante (datos socioeconómicos), interacción de los estudiantes en las plataformas de aprendizaje, entre otros. - Verificar la disponibilidad de los datos, generar el diccionario de datos y validar calidad de los datos institucionales. - Analizar la estructura, formatos, y posibles inconsistencias o valores faltantes en los datos.
Preparación de los datos	<ul style="list-style-type: none"> - Extracción de los datos - Limpieza de datos: tratamiento de valores faltantes, duplicados y errores. - Transformación de variables: normalización, codificación categórica. - Integración de múltiples fuentes de datos (académicos, socioeconómicos, encuestas, LMS, etc.). - Dar formato a las variables relevantes para el modelado.
Modelado	<ul style="list-style-type: none"> - Selección y aplicación de técnicas de minería de datos (regresión, árboles de decisión, clustering, redes neuronales, etc.). - Construcción de modelos para predecir el rendimiento académico o clasificar a estudiantes en riesgo. - Ajuste de parámetros y comparación de modelos.
Evaluación	<ul style="list-style-type: none"> - Evaluación de la precisión, interpretabilidad y utilidad de los modelos generados. - Revisión del proceso y análisis de errores. - Interpretación de resultados por parte del equipo y validación con el cliente para comprobar que los modelos aporten valor real.
Despliegue	<ul style="list-style-type: none"> - Implementación de dashboards o sistemas de alerta temprana para el cliente. - Integración de los modelos en el sistema de gestión académico institucional o plataformas de seguimiento estudiantil. - Entrega de informe final (Documentación del proceso, hallazgos, modelos, conclusiones y recomendaciones) - Plan de seguimiento y mejora continua.