



# **Actividad 5.3:**

## **Comparación de métodos**

CE Inteligencia Artificial y Big Data

Sistemas de Big Data

2024/2025

Daniel Marín López

José Domingo Carrillo García

# Índice

|                                  |   |
|----------------------------------|---|
| 1. Enunciado y tabla.....        | 3 |
| 2. Comparación y Conclusión..... | 4 |

## 1. Enunciado y tabla

Una cadena de supermercados quiere mejorar su estrategia de fidelización de clientes. Para ello, cuenta con una gran base de datos de transacciones de compra, con información sobre productos adquiridos, fechas, horarios, ventas, y códigos de cliente.

El objetivo del proyecto es identificar patrones de comportamiento de los clientes para lanzar campañas personalizadas y mejorar la retención.

Se pide lo siguiente:

1. Divide la actividad en tres secciones: KDD, SEMMA y P<sup>3</sup>TQ (como muestra la tabla más abajo)
2. Para cada metodología, explica paso a paso cómo se abordaría el problema utilizando sus fases. No hace falta hacer el análisis real, sino describir el proceso que seguirías.
3. Al final, redacta una breve comparación: ¿qué metodología te parece más útil o clara para este problema? ¿Cuál te resultó más práctica de aplicar para el caso propuesto? Incluye ventajas y desventajas al responder a estas preguntas en tu conclusión.

La tabla de las metodologías sería la siguiente:

| Metodología | Pasos                     | Acción   |
|-------------|---------------------------|--|
| KDD         | Selección                 | Se seleccionan los datos relevantes: transacciones, productos, fechas, horarios y códigos de cliente.  |
|             | Preproceso                | Se limpian y depuran los datos: corrección de errores, eliminación de duplicados e imputación de valores faltantes.                              |
|             | Transformación            | Se transforman los datos para facilitar el análisis: agregación por cliente, normalización de cantidades, codificación de variables categóricas. |
|             | Minería de Datos          | Se aplican algoritmos de clustering o asociación para detectar patrones de compra y segmentar clientes.  |
|             | Interpretación/Evaluación | Se interpretan los resultados para identificar perfiles de clientes y definir estrategias de campañas personalizadas.                            |
| SEMMA       | Sample                    | Se extrae una muestra representativa de la base de datos de clientes y transacciones.  |
|             | Explore                   | Se analiza visual y estadísticamente la muestra para identificar relaciones, outliers y comportamientos generales.                               |

|                   |          |   |
|-------------------|----------|---|
|                   | Modify   | Se crean nuevas variables derivadas (frecuencia de compra, valor promedio, etc.) y se transforman los datos.              |
|                   | Model    | Se construyen modelos predictivos o descriptivos, como segmentación de clientes o reglas de asociación.                   |
|                   | Assess   | Se evalúa la precisión, utilidad y aplicabilidad de los modelos para definir su implementación en campañas.               |
| P <sup>3</sup> TQ | Product  | Se identifican los productos que más influyen en la fidelización y cuáles podrían formar parte de ofertas personalizadas. |
|                   | Place    | Se analiza en qué tiendas o canales los clientes fieles realizan sus compras con más frecuencia.                          |
|                   | Price    | Se evalúan los precios que los clientes están dispuestos a pagar según su perfil y comportamiento.                        |
|                   | Time     | Se identifican los días y horarios con mayor actividad por tipo de cliente para enviar promociones oportunas.             |
|                   | Quantity | Se analiza cuánto compran los clientes para ajustar campañas según volumen de compra y hábitos.                           |

## 2. Comparación y Conclusión

**KDD** es una metodología robusta y detallada que encaja con proyectos que usan grandes volúmenes de datos como en este caso. Su enfoque completo permite desde la preparación hasta la interpretación de resultados. Es clara, aunque puede parecer más académica o compleja para quienes no tienen experiencia en minería de datos.

**SEMMA** es más práctica desde una perspectiva de implementación técnica. Es perfecta para científicos de datos y analistas porque se centra en la preparación y modelado, con pasos bien definidos. Sin embargo, se centra más en el proceso técnico que en la comprensión del negocio.

**P<sup>3</sup>TQ** es útil para traducir el análisis en decisiones comerciales concretas y no llega a ser una metodología de minería de datos como tal, sino más bien una guía estratégica para alinear el análisis con el marketing y la gestión operativa. Es la más intuitiva para equipos de negocio, pero requiere apoyo técnico para generar los datos que alimenten su análisis.

La conclusión a la que hemos llegado en base a lo visto anteriormente es que la **metodología KDD** es la más útil y clara para este problema porque abarca todo el proceso desde la selección de datos hasta la interpretación final. Permite una visión completa y ordenada del análisis. Sin embargo, SEMMA fue la más práctica de aplicar en este caso, especialmente en la fase técnica de análisis de datos. P<sup>3</sup>TQ, aunque complementaria, necesita de los resultados de KDD o SEMMA para que sea efectiva.