



Actividad 1:

Metodología CRISP-DM

CE Inteligencia Artificial y Big Data

Sistemas de Big Data

2024/2025

Daniel Marín López

Índice

1. Enunciado.....	3
2. Metodología de CRISP-DM.....	3

1. Enunciado

Dado este supuesto:

“Eres parte del equipo de datos de una empresa de e-commerce. El gerente de operaciones está preocupado porque muchos pedidos llegan tarde a los clientes. Quiere entender por qué ocurre esto y qué se puede hacer para prevenirlo.”

Se debe aplicar la metodología CRISP-DM explicando qué realizarías en cada una de las fases.

2. Metodología de CRISP-DM

Podemos realizar la siguiente tabla para implementar la metodología sobre nuestro negocio.

Fase	Descripción	Aplicación
1. Comprensión del negocio	Identificamos los objetivos de nuestro proyecto desde una perspectiva empresarial	<p>Objetivo (Tarea 1.1): Reducir el retraso de los pedidos.</p> <p>Solución: Crear un modelo que detecte retrasos en los pedidos y active un control para reducirlo al mínimo posible.</p>
2. Comprensión de los datos	Nos familiarizamos con los datos disponibles una vez tenemos los objetivos claros.	<p>Datos disponibles:</p> <ul style="list-style-type: none"> • Fecha del pedido • Ciudad de destino • Tiempo de entrega (días) • Medio de envío (normal / urgente) • ¿Llegó tarde? (Sí / No) <p>Se recogen los datos en una tabla de excel (Tarea 2.1).</p> <p>Exploración (Tarea 2.2): En muchos de los pedidos, los retrasos se ven relacionados con el nº de días de entrega.</p>
3. Preparación de los datos	Preparamos los datos para la fase de modelado.	<p>Transformaciones (Tarea 3.1): Se agrega la columna “velocidad” en función del tiempo de entrega. Y quitar la ciudad de destino.</p> <p>Datos listos: Tenemos los datos con una cronología, y un tiempo de entrega que puede ser de lento a rápido.</p>

4. Modelado	Se aplican modelos de minería de datos que pueden ser predictivos o descriptivos.	<p>Modelo a usar (Tarea 4.1): Random Forest: Podemos usar este modelo para que en base a la fecha del pedido, el tiempo de entrega, el medio de envío y su velocidad se pueda estimar si va a llegar tarde o no.</p> <p>Entrenamiento: Usamos un registro de pedidos con un montón de pedidos que han sufrido retrasos.</p> <p>Resultado: El modelo predice un 80% de precisión cuando un pedido con velocidad lenta llegará tarde.</p>
5. Evaluación	Se realiza una evaluación del proyecto antes de su implementación.	<p>Evaluación (Tarea 5.1): Nuestro modelo responde de manera correcta a la mayoría de retrasos.</p> <p>Impacto: Se mira cómo solucionarlo: agregando sucursales más cercanas, mejor planificación del envío.</p> <p>Decisión: Se aprueba el modelo para su uso en la empresa.</p>
6. Despliegue	Se pone el proyecto en producción para que usuarios o sistemas tomen decisiones a partir de los resultados.	<p>Despliegue (Tarea 6.1): El gerente, tras ver los resultados, instala el modelo dentro del sistema de gestión de pedidos. Si se detecta retraso, se envía una notificación a la empresa para buscar reducirlo lo máximo posible.</p> <p>Mantenimiento: El modelo se actualiza cada mes de manera automática con plataformas de automatización como make.</p> <p>Monitoreo: También, mediante make, se puede recalibrar el modelo automáticamente de forma semanal.</p>

fecha_pedido	ciudad_destino	tiempo_entrega	medio_envio	tarde	velocidad
07/03/2024	Córdoba	50	normal	No	normal
12/09/2024	Almeria	24	normal	No	rápido
25/01/2025	Espejo	105	normal	Sí	lento
22/11/2024	Madrid	10	normal	Sí	rápido
10/10/2024	Barcelona	45	urgente	No	normal
31/10/2024	Alcoy	60	urgente	Sí	lento
23/12/2024	Granada	25	urgente	No	rápido
06/03/2025	Pozoblanco	5	urgente	No	rápido
31/08/2024	Málaga	12	normal	Sí	rápido
04/03/2025	Santiago de Compostela	75	urgente	Sí	lento