

L U M B R E



Flujo de trabajo CRISP-DM

Grado: 1

Materia: Fundamentos de Ciencia de Datos · Análisis de Datos

11/10/2026

Antes de empezar

CRISP-DM es el mapa de rutas del proyecto de datos: seis fases que no avanzan en línea recta sino en bucles. Conocerlo evita el error de saltar directamente a modelar sin entender el problema ni los datos.

QUÉ APRENDERÁS

- Describir las fases del proceso de exploración de datos CRISP-DM y aplicar su lógica iterativa a un proyecto real.

Flujo de trabajo CRISP-DM

Primero entender, luego modelar

El error más caro del principiante es abrir un cuaderno de notas y ponerse a entrenar modelos sin saber qué pregunta responde ni si los datos la soportan. CRISP-DM existe para impedir ese salto al vacío.

CRISP-DM (Cross-Industry Standard Process for Data Mining) organiza el trabajo de datos en seis fases encadenadas pero sobre todo iterativas. No es una receta rígida: es una disciplina que reconoce que cada vuelta afina la pregunta, los datos y el modelo.

VIDEO

Entender el proceso CRISP-DM

Las seis fases iterativas de un proyecto de minería de datos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=0kEWGCIQSPk>

Las seis fases de CRISP-DM

1. **1Comprensión del negocio** Traducir la necesidad en una pregunta medible con criterios de éxito y un plan.
2. **2Comprensión de los datos** Recolectar lo disponible, explorar su calidad y descubrir si de verdad hay señal.

3. **3Preparación** Seleccionar, limpiar y transformar hasta obtener una tabla analizable y sin fugas.
4. **4Modelado** Elegir técnicas, entrenar y ajustar parámetros con validación honesta.
5. **5Evaluación** Juzgar el resultado contra el objetivo de negocio, no solo contra una métrica técnica.
6. **6Despliegue** Poner el resultado a producir, monitorizar y documentar para repetirlo.

Fase 1 · Comprensión del negocio

Antes de los datos, los objetivos. Se habla con quien decide, se formula la pregunta, se fijan los criterios de éxito y se identifica qué significa un buen resultado. Aquí se define cuándo el proyecto habrá merecido la pena.

Fase 2 · Comprensión de los datos

- Reunir las fuentes disponibles y registrar su procedencia.
- Explorar distribuciones, vacíos y anomalías.
- Verificar que el dato existe y guarda relación con la pregunta.

Fase 3 · Preparación de datos

La fase más larga. Seleccionar filas y columnas relevantes, tratar ausentes, corregir errores, derivar variables y partir el conjunto para validación. Un error aquí se propaga a todo lo demás.

Fase 4 · Modelado

Elegir familia de modelos, entrenar y afinar. La técnica depende del tipo de pregunta: descriptiva, predictiva o de segmentación. Buenas prácticas exigen comparar contra una línea base sencilla.

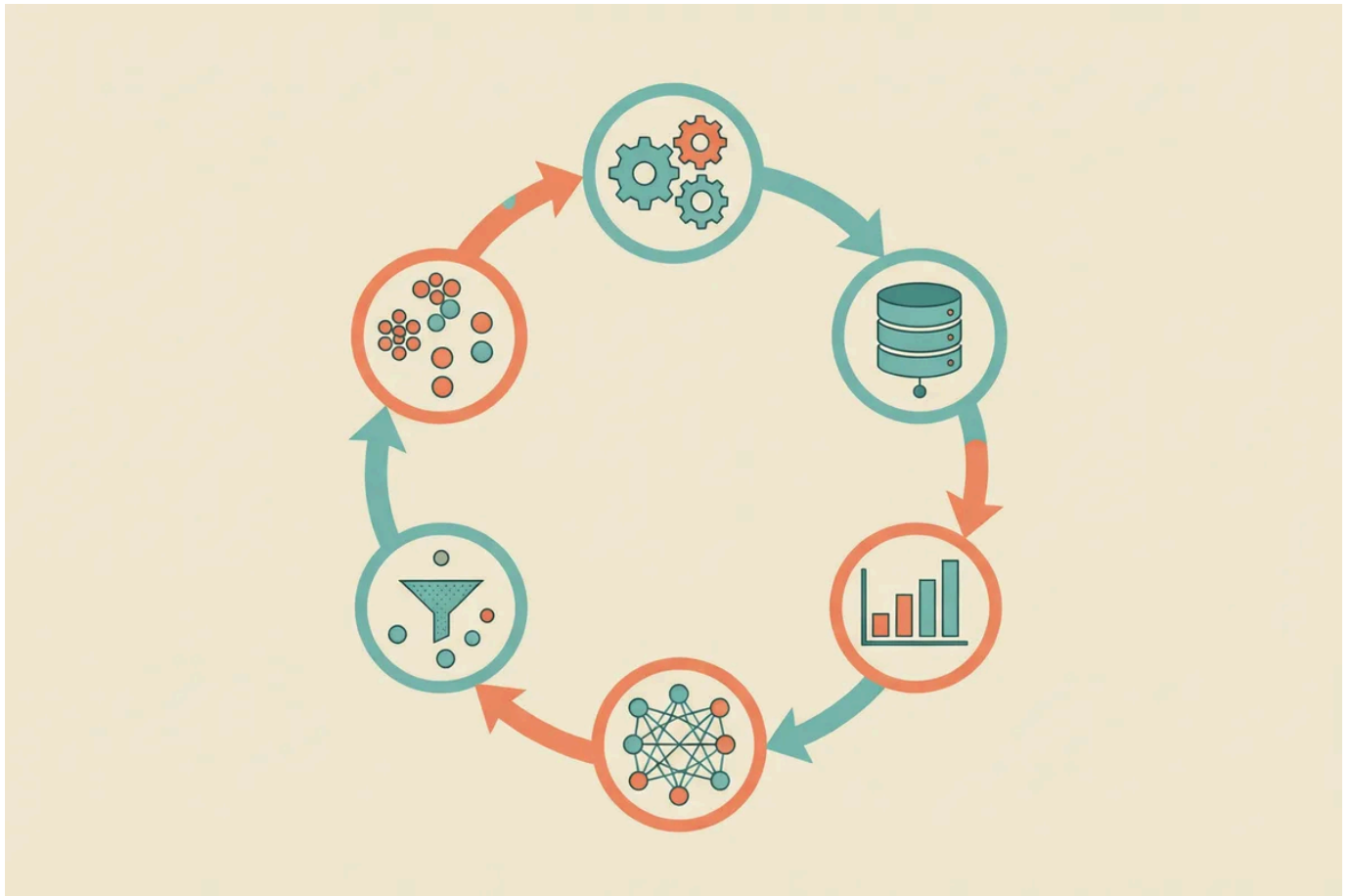
Fase 5 · Evaluación

Se comprueba que el modelo responde a la pregunta de negocio y no solo a la métrica de validación. Se revisan los supuestos, los casos límite y si el beneficio esperado justifica el coste.

Fase 6 · Despliegue

- Integrar el resultado donde se tomará la decisión.
- Monitorizar la degradación y la deriva de datos.

- Documentar y dejar un plan de mantenimiento y reentrenamiento.



CRISP-DM avanza en bucles: cada vuelta afina pregunta, datos y modelo.

La iteración como corazón del método

Las flechas del diagrama vuelven hacia atrás: a menudo preparar los datos revela que la pregunta era otra, y evaluar obliga a reconsiderar el modelo. Esa retroalimentación es la fortaleza, no un defecto del proceso.

Quién hace qué en cada fase

Negocio y datos conversan en las dos primeras; el científico de datos domina el centro; ingeniería y operaciones cierran el despliegue. CRISP-DM hace visible que un proyecto no es cosa de una sola persona.

Alternativas y complementos

- KDD y SEMMA enfatizan más el descubrimiento y la minería.
- Kanban y métodos ágiles aportan cadencia a las iteraciones.
- La elección del marco importa menos que no saltarse entender problema y datos.

Errores frecuentes

- Saltar a modelar sin entender el negocio ni los datos.
- Tratar el proceso como una línea recta en vez de un bucle.
- Evaluar solo con la métrica técnica y olvidar el objetivo real.

Ejemplo resuelto: aplicar CRISP-DM a la rotación de personal

Un equipo quiere reducir las bajas. Recorrer las fases evita el salto al vacío:

1. Negocio: definir qué cuenta como baja indeseada y el ahorro esperado.
2. Datos: reunir historial y encuestas y comprobar que hay señal.
3. Preparación: limpiar y crear variables sin filtrar el futuro.
4. Modelado: probar un modelo simple frente a la línea base.
5. Evaluación y despliegue: juzgar el impacto real y monitorizar.

¿Para qué sirve en la realidad?

CRISP-DM da un lenguaje común al equipo y obliga a validar supuestos antes de invertir: saber en qué fase estás evita los errores más caros del proyecto.

1. En CRISP-DM, la fase en la que normalmente se invierte más tiempo real es:

- ☐ Comprensión del negocio
- ☐ Preparación de datos
- ☐ Modelado
- ☐ Despliegue

2. CRISP-DM es un proceso estrictamente lineal que avanza una sola vez por cada fase.

- ☐ Verdadero
- ☐ Falso

3. Relaciona cada fase de CRISP-DM con la tarea que le corresponde.

3.	Comprensión del negocio	-----
3.	Comprensión de los datos	-----
3.	Preparación	-----
3.	Modelado	-----
3.	Evaluación	-----
3.	Despliegue	-----

Opciones: Definir la pregunta y los criterios de éxito · Recolectar y explorar las fuentes disponibles · Limpiar y transformar hasta poder analizar · Entrenar y ajustar los modelos · Juzgar el resultado contra el objetivo de negocio · Poner a producir y monitorizar

CRISP-DM https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining
Estándar de proceso para proyectos de minería de datos.

Seis fases en bucle

CRISP-DM

Negocio

pregunta medible

criterios de éxito

Datos

recolectar y explorar

verificar señal

Preparación

limpiar y transformar

sin fugas

Modelo, evaluación y despliegue

entrenar y validar

desplegar y monitorizar

El mapa mental interactivo es interactivo en la versión online.



KDD https://en.wikipedia.org/wiki/Knowledge_discovery_in_databases Descubrimiento de conocimiento en bases de datos, proceso emparentado con CRISP-DM.

4. Un compañero quiere «empezar metiendo los datos en una red neuronal». Con las palabras de CRISP-DM explícale qué dos fases se está saltando y qué riesgo tiene cada salto.

Referencia curricular

Grado:	1
Materia:	Fundamentos de Ciencia de Datos · Análisis de Datos

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-fcd-006-flujo-de-trabajo-crisp-dm>

Material educativo generado desde Lumbre · lumbre.lat