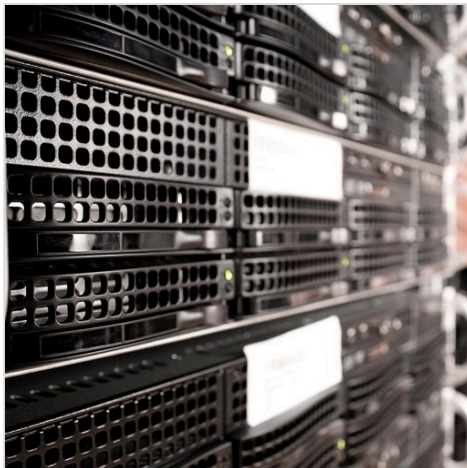


L U M B R E



Qué es la ciencia de datos (y en qué no lo es)

Grado: 1

Materia: Fundamentos de Ciencia de Datos · Análisis de Datos

11/10/2026

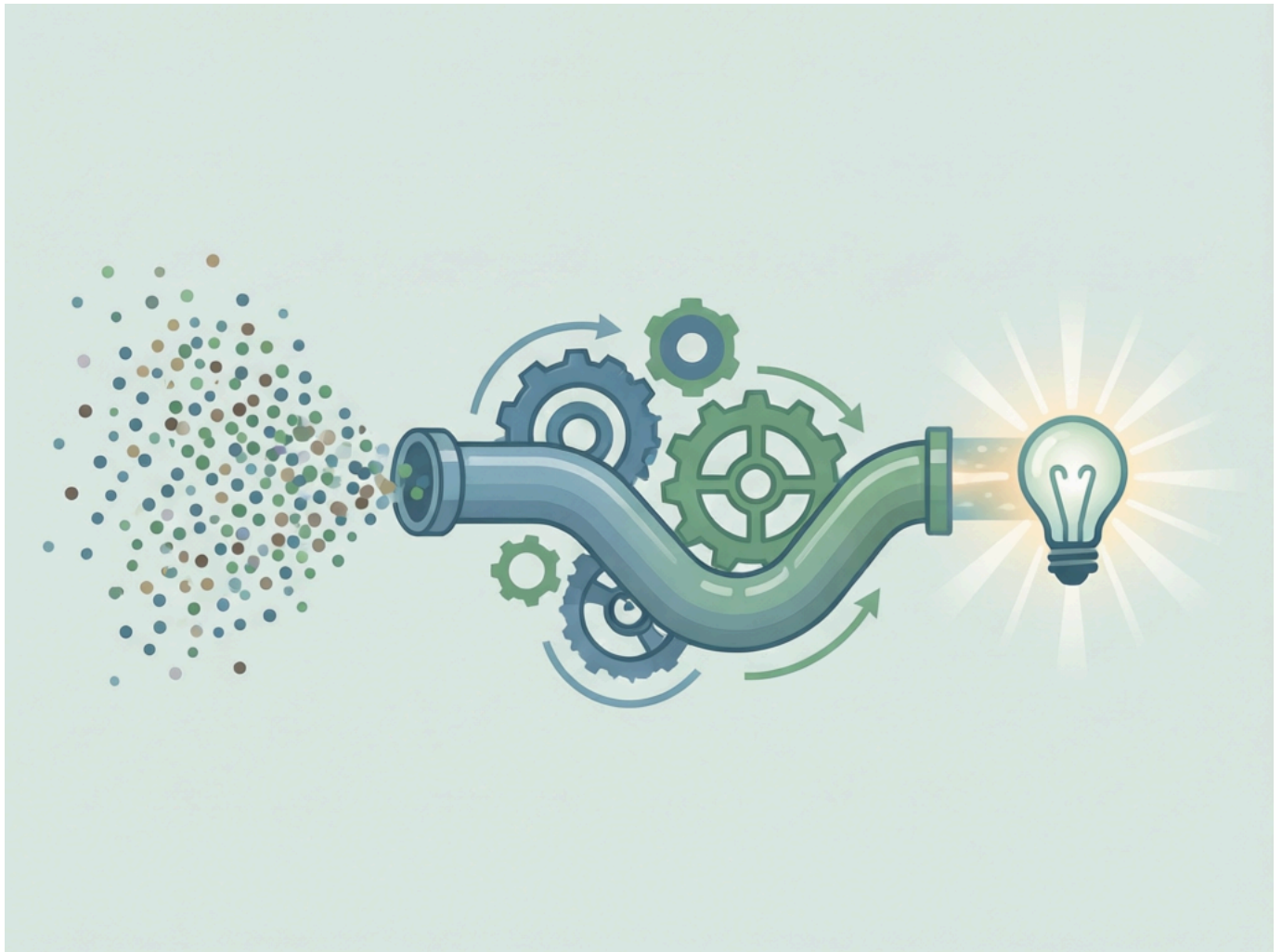
Antes de empezar

Un mapa honesto de la disciplina: qué problema resuelve cada papel, qué herramientas usan y por qué confundirlos es la primera fuente de proyectos que no llegan a producción.

QUÉ APRENDERÁS

- Describe el ciclo de trabajo de la ciencia de datos y distingue, con criterios, la analítica descriptiva, el aprendizaje automático y la ingeniería de datos.

Qué es la ciencia de datos (y en qué no lo es)



El ecosistema de la ciencia de datos: desde los datos hasta la ética.

La meta de hoy

La ciencia de datos no es «hacer gráficos con Python». Es un método para convertir preguntas difusas en evidencia decisional, usando datos, modelos y juicio sobre el contexto. Hoy trazamos ese mapa y sus fronteras.

Cuando alguien dice «necesito un modelo de IA», casi nunca está pidiendo una sola cosa. Detrás puede haber una pregunta de reporte (¿cuánto vendimos?), una predictiva (¿quién va a cancelar el servicio?) o una de infraestructura (¿cómo juntamos datos que viven en diez sistemas distintos?). Son tres disciplinas emparentadas —analítica, *machine learning* e ingeniería de datos— que se confunden todo el tiempo. Aprender a separarlas es el primer acto de rigor profesional.

El ciclo: cinco fases que no son una línea recta

Casi todos los marcos de trabajo (CRISP-DM es el clásico) describen el mismo recorrido con distintos nombres. Lo importante no son las etiquetas, sino dos ideas: que se **itera** y que el ciclo **se cierra** cuando el hallazgo vuelve al negocio en forma de decisión.

El ciclo de la ciencia de datos

1. 1. Entender **Pregunta y objetivo** Traducir un problema del mundo real en una pregunta medible, con hipótesis y criterios de éxito.
2. 2. Datos **Recolectar y limpiar** Localizar fuentes, evaluar calidad y construir un conjunto analysable; aquí se va la mayor parte del tiempo real.
3. 3. Explorar **EDA y características** Describir, visualizar y crear variables que revelen estructura antes de modelar.
4. 4. Modelar **Ajustar y validar** Entrenar modelos y estimar su comportamiento con datos no vistos, sin fuga de información.
5. 5. Comunicar **Desplegar y decidir** Convertir el resultado en una recomendación accionable y, si corresponde, ponerla en producción.

Nota la asimetría: la fase 2 consume con diferencia más esfuerzo que las demás. Un modelo sofisticado sobre datos sucios sigue dando basura; de ahí el dicho «basura entra, basura sale».

Tres papeles, tres preguntas distintas

Un proyecto real suele necesitar a los tres; lo que cambia es cuál domina.

Papel	Pregunta que responde	Producto típico	Herramientas comunes
Analítica de datos	¿Qué pasó y por qué?	Dashboard, informe, KPI	SQL, hoja de cálculo, BI
Machine learning	¿Qué puede pasar?	Modelo predictivo o de segmentación	Python, scikit-learn, frameworks de DL
Ingeniería de datos	¿Cómo llegan los datos, limpios y a tiempo?	Pipeline confiable y gobernado	Spark, Airflow, warehouses, cloud

La trampa del «vamos a meter IA»

Que puedas entrenar una red neuronal no significa que el proyecto sea de machine learning. Si la decisión se toma mirando un agregado y una tendencia, es analítica, y esa es una respuesta perfectamente válida y más barata.

Mapa de la disciplina

Ciencia de datos

Analítica

descriptiva

diagnóstica

reporte y KPIs

Modelado

predictivo (ML)

exploratorio (no supervisado)

experimental (A/B)

Ingeniería

pipelines

calidad y gobernanza

despliegue (MLOps)

Comunicación

visualización honesta

decisión de negocio

ética del dato

El mapa mental interactivo es interactivo en la versión online.



1. Una empresa quiere saber cuántos clientes perdieron el trimestre pasado y en qué regiones. ¿Qué papel domina ese trabajo?

- ☐ Ingeniería de datos
- ☐ Analítica descriptiva
- ☐ Machine learning supervisado
- ☐ Deep learning

2. Toda pregunta de negocio que involucre datos debe resolverse con un modelo de machine learning.

- ☐ Verdadero
- ☐ Falso

3. Clasifica cada tarea según el papel que le corresponde.

Elementos:

- Desglosar ventas por tienda y mes
- Predecir qué clientes cancelarán en 30 días
- Construir la carga diaria desde la app al almacén
- Agrupar clientes por comportamiento sin etiquetas
- Montar el pipeline que valida y versiona los datos
- Publicar un semáforo de indicadores para dirección

Analítica	. . .
Machine learning	. . .
Ingeniería de datos	. . .

4. El marco CRISP-DM ordena el trabajo en fases que se _____ : tras comunicar, suele aparecer una nueva pregunta. La fase que más tiempo consume casi siempre es la de los _____ , porque los datos raramente llegan _____ .

Dónde se rompen los proyectos (y cómo se ve desde lejos)

- La pregunta llega difusa: nadie definió qué decisión cambiará el resultado, así que el análisis no tiene a dónde apuntar.
- Se empieza por el modelo y se buscan datos después: el orden correcto va del problema a los datos, no al revés.
- Se ignora la calidad del dato: duplicados, valores perdidos y fugas de tiempo inflan métricas que luego no se sostienen en producción.

Diagnosticar a cuál de estas fallas pertenece un proyecto es, en sí, una competencia central de la carrera. A lo largo del plan cada una tiene su materia: la pregunta se trabaja en metodología y comunicación, el dato en ingeniería y calidad, y el modelo en las asignaturas de aprendizaje automático.

Un criterio para no confundirse

1. Pregunta el tipo de afirmación: ¿describe el pasado, predice el futuro o mueve datos?
2. Revisa la decisión: si nadie va a actuar con el resultado, probablemente sobre-analizas.
3. Elige la herramienta más simple que responda con validez; sube de complejidad solo cuando la evidencia lo exija.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

5. Vocabulario mínimo



Ciencia de datos (visión general) https://es.wikipedia.org/wiki/Ciencia_de_datos Punto de partida para ver el panorama de la disciplina y sus relaciones.

VIDEO

Video complementario

Recurso audiovisual para reforzar los conceptos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=xzZSILlQ2QI>

6. Elige una decisión real que deba tomar tu escuela, tu trabajo o tu municipio. Escríbela como pregunta y decide a cuál de los tres papeles pertenece (¿analítica, ML o ingeniería?) y por qué. Justifica qué pasaría si la resolvieras con el papel equivocado.

UC Berkeley Data Science <https://datascience.berkeley.edu/Recursos> y definiciones de ciencia de datos.

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-001-que-es-la-ciencia-de-datos-y-en-que-no-lo-es>

Material educativo generado desde Lumbre · lumbre.lat