

L U M B R E



Capstone resuelto: el predictor de abandono, de la pregunta al despliegue

Materia: Ciencia de Datos e IA

11/10/2026

Antes de empezar

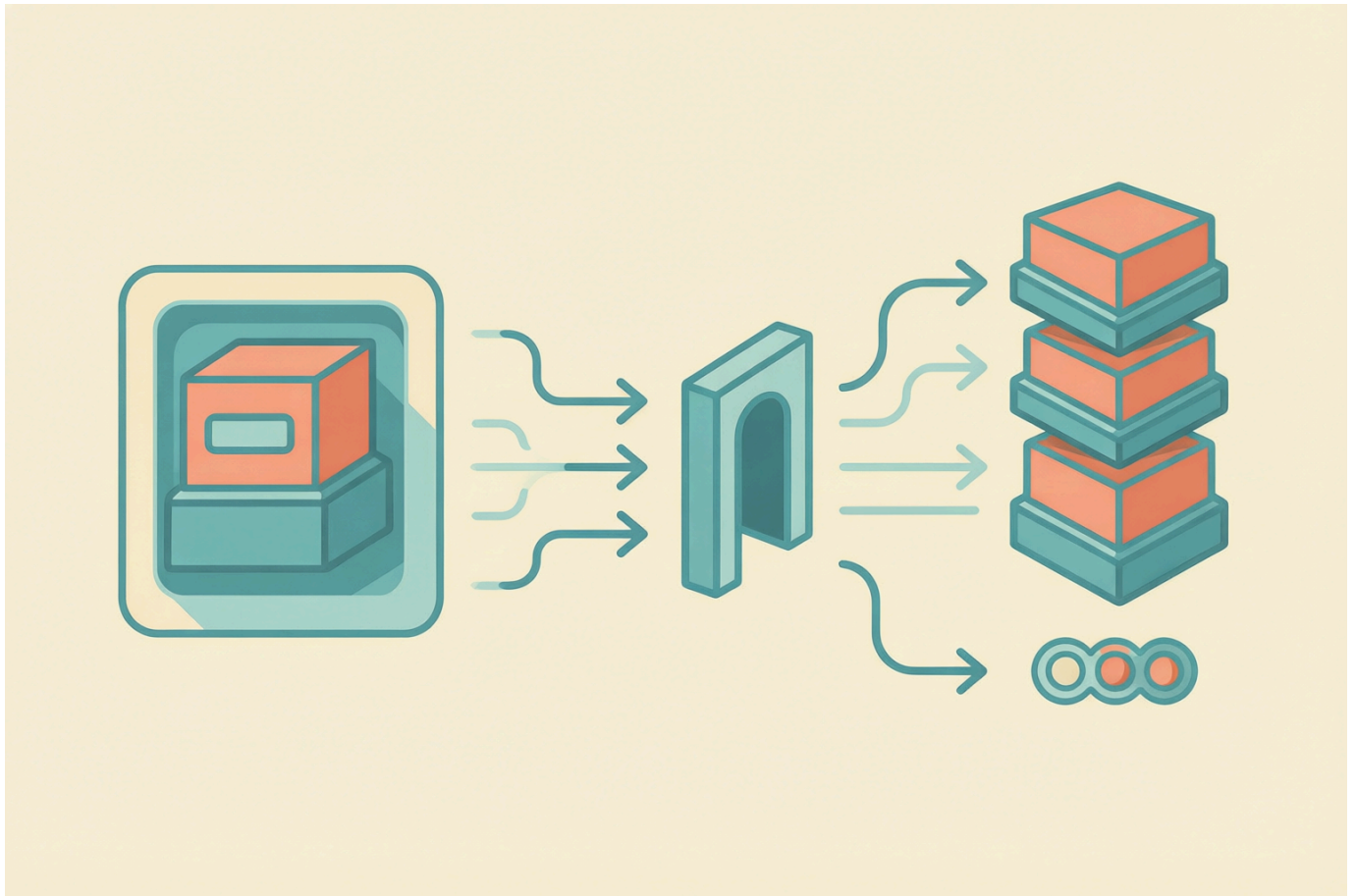
Capstone completo y resuelto de churn: pregunta con costos, EDA con hallazgos, limpieza anti-fuga, baseline batido, confusion en dolares, barrido de umbral con mejora del 41%, despliegue con drift y autoexamen.

QUÉ APRENDERÁS

- Integration de toda la carrera en un proyecto end-to-end con las siete etapas, metricas de negocio en dolares, umbral optimo, explicabilidad y despliegue monitoreado.

Capstone resuelto: el predictor de abandono, de la pregunta al despliegue

Este es el proyecto que ata TODA la carrera: un caso completo de abandono de clientes resuelto en siete etapas, con los numeros, las decisiones y los errores que se cometen de verdad. Cada etapa muestra su entregable concreto y el codigo que la ejecuta, citando el OA especifico donde se fundo la tecnica. Correr este capstone entero (una tarde con un dataset tuyo) es la mejor preparacion para una defensa de proyecto que cualquier rubrica.



Un pipeline no es solo código: cada etapa tiene una pregunta, un entregable y un criterio de paso a la siguiente.

Las siete etapas

Definición del problema → obtención y EDA → limpieza → baseline → evaluación con métrica de NEGOCIO → análisis de error y mejora → despliegue y monitoreo. El salto cualitativo del estudiante al profesional: cerrar cada etapa con su entregable ANTES de abrir la siguiente.

Etapa 1 · La pregunta (la que nadie quiere trabajar)

Pedido vago: "queremos predecir churn". Buenas preguntas tienen cuatro componentes — decisión, sujeto, horizonte y métrica: "Que clientes con suscripción mensual cancelaran en los próximos 30 días, para que el equipo de retención les ofrezca una llamada ANTES, priorizando llamadas donde el costo de la llamada (3 USD) es menor que el margen recuperado (25 USD)". Ese número de márgenes NO es decoración: define el umbral operativo de la etapa 6 y se acuerda CON el negocio en la etapa 1. Entregable: una página escrita y firmada: pregunta, métrica de éxito, costos de error, límite temporal.

Etapa 2 · Datos y primera mirada

ACTIVIDAD INTERACTIVA
DISPONIBLE EN LUMBRE

**1. EDA de 20 minutos que
cambia el proyecto**



Etapa 3 · Limpieza con decision

ACTIVIDAD INTERACTIVA DISPONIBLE
EN LUMBRE

**2. Cuatro movidas, cuatro
justificaciones**



Etapa 4 · Baseline y modelo

ACTIVIDAD INTERACTIVA DISPONIBLE
EN LUMBRE

**3. La humilde regresion
logistica primero**



Etapa 5 · Metrica de NEGOCIO (el giro de timon)

ACTIVIDAD INTERACTIVA DISPONIBLE EN
LUMBRE

**4. Traducir la confusion a
dolares**



Etapa 6 · Mejora: modelo y umbral

ACTIVIDAD INTERACTIVA DISPONIBLE EN
LUMBRE

**5. Donde el trabajo fino
paga**



Etaapa 7 · Desplegar y no irse

- Artefactos: modelo serializado + requirements.txt + version DVC/git del dataset exacto (el OA de Git te dice por que).
- Servir: un microservicio Flask con /predict que reciba el perfil y devuelva probabilidad; el negocio consume la LISTA priorizada diaria, no el endpoint crudo.
- Monitoreo: drift de entradas (KS o PSI por columna) y caída del recall real (el ground truth llega 30 días después): alertas antes del desastre silencioso.
- Feedback loop: cada llamada de retención registrada (salvo/perdido) alimenta la siguiente versión con etiqueta real — el proyecto vive en ciclo, no termina en el informe.

El capstone en un vistazo

- **Churn end-to-end**
 - Etapa 1
 - Pregunta con costos acordados
 - Etapas 2-3
 - EDA con decisiones etiquetadas
 - Etapas 4-5
 - Baseline batido y métrica en dólares
 - Etapas 6-7
 - Umbral óptimo, SHAP, equidad
 - Servir, monitorear, realimentar

El mapa mental interactivo es interactivo en la versión online.



Autoexamen cronometrado (20 minutos)

6. Del baseline: confusion $[[620,80],[120,180]]$. Calcula el RECALL de churn (180 entre 300), en decimal.

Respuesta:

7. Con costos 25 USD por FN y 3 USD por FP, el costo del estado baseline (120 FN, 80 FP) es de cuantos USD.

Respuesta:

8. La clase mayoritaria da accuracy 0.78. Un modelo nuevo saca 0.79 y recall de churn 0.55. ¿Lo despliegas?

- ☐ Si: mayor accuracy siempre gana
- ☐ No: no bate el baseline en la metrica que importa (costo)
- ☐ Si: era el mejor modelo en CV
- ☐ No: los modelos simples no sirven

9. El escalado con z-score se ajusta (fit) sobre todo el dataset antes del split para aprovechar mas datos.

- ☐ Verdadero
- ☐ Falso

10. El _____ de decision (p. ej. 0.30 en vez de 0.50) se elige minimizando el costo del negocio, no la accuracy.**11. Etapa con su entregable.**

11.	Definicion	-----
11.	EDA	-----
11.	Evaluacion	-----
11.	Despliegue	-----

Opciones: Una pagina con pregunta y costos · Cinco hallazgos graficados · Confusion traducida a dolares · Lista priorizada diaria + alertas de drift

Video de repaso

VIDEO

Proyecto de ML end-to-end: video complementario

Un proyecto completo ejecutado ante tus ojos, para comparar con el tuyo.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=o6vbe5G7xNo>

Para profundizar

MLOps: pipelines de ML de Google Cloud <https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning> Guía gratuita; ordena por niveles de madurez todo lo que acabas de recorrer.

MLflow: ciclo de vida del experimento <https://mlflow.org/docs/latest/> Registro de experimentos, versiones de modelo y línea de promoción a producción: el inventario del capstone desplegado.

12. Recorre tus últimos 3 proyectos con las siete etapas: ¿en cuál faltaron las etapas 1 (pregunta con costo) o 5 (métrica de negocio)? ¿Cuál fue el número que no preguntaste al negocio? Ese es el párrafo que más impresionara en una defensa: "mi métrica de éxito es X porque su costo es Y".

Referencia curricular

Materia:	Ciencia de Datos e IA
-----------------	-----------------------

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



<https://univ.lumbre.lat/object/cdia-proy-005-capstone-resuelto-predictor-de-abandono>

Material educativo generado desde Lumbre · lumbre.lat