

L U M B R E



Modelado del KDD

Grado: 5

Materia: Minería de Datos · Machine Learning

03/10/2026

Antes de empezar

Minar datos no es correr un algoritmo: es un proceso con cabeza. El descubrimiento de conocimiento en bases de datos ordena las fases — seleccionar, limpiar, transformar, minar, interpretar— para convertir montañas de registros en conocimiento fiable y útil.

QUÉ APRENDERÁS

- Diseñar un proceso completo de descubrimiento de conocimiento en bases de datos, de la selección del problema a la interpretación y despliegue del hallazgo.

Modelado del KDD

Proceso, no golpe de suerte

El hallazgo brillante sobre un proceso desordenado sigue siendo un accidente. El descubrimiento de conocimiento disciplina la minería en fases encadenadas, cada una con sus decisiones, sus revisiones y su vuelta atrás cuando toca.

VIDEO

CRISP-DM: de los datos a las decisiones

Vídeo en español que recorre el proceso estándar de proyectos de minería de datos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=28H0pYR21jA>

Del dato al conocimiento

KDD es el viaje completo: partir de datos en bruto, enormes y sucios, y terminar en conocimiento accionable y comprensible. La minería de patrones es solo una fase; el proceso la envuelve de contexto, limpieza y validación que evitan conclusiones falsas.

Comprensión del problema

Todo empieza por entender el objetivo de negocio y el dominio: qué se quiere saber, contra qué línea base, con qué coste del error. Sin esta brújula, se mina con entusiasmo en la dirección equivocada.

Selección de datos

Identificar y extraer los datos relevantes entre el arsenal disponible. Elegir mal el conjunto —viejo, parcial, no representativo— condena todo lo demás, pues ningún algoritmo salva una materia prima inadecuada.



El proceso KDD convierte la materia prima en bruto en conocimiento valioso pasando por fases encadenadas: seleccionar, limpiar, transformar, minar e interpretar.

Preprocesado y limpieza

Eliminar ruido, duplicados e inconsistencias, y decidir cómo tratar valores faltantes. Es la fase que más tiempo devora y la que más determina la calidad final: basura dentro, basura fuera, con apariencia científica.

Transformación y reducción

Normalizar, agregar, seleccionar características y proyectar a dimensiones manejables. Aquí el ingeniero de características y el analista exploratorio preparan el terreno para que el minado sea factible y significativo.

Minado de patrones

La fase estrella: elegir y aplicar algoritmos —asociación, clustering, clasificación— para extraer patrones. Pero la magia depende de todo lo previo y de una validación honesta; sin ellas, solo produce espejismos cuantificados.

Interpretación y evaluación

Filtrar los patrones por interés y validez, descartar espurios y redundantes, y traducirlos a un lenguaje que el negocio entienda. Aquí la paradoja de Simpson y el confusor acechan: sin escepticismo, se publica el error.

Iteración constante

El KDD no es una línea recta sino un bucle: cada hallazgo reformula la pregunta, revela datos que faltaban o limpia supuestos erróneos, y devuelve a fases anteriores. La madurez es abrazar esa recursividad en vez de fingir un avance ordenado.

De KDD a CRISP-DM

La metodología CRISP-DM popularizó estas fases con una vista de negocio y un despliegue explícitos: comprensión del negocio y los datos, modelado, evaluación y despliegue. Es el esqueleto de facto de todo proyecto analítico serio.

Errores frecuentes

- Saltar a minar sin limpiar ni entender los datos, confiando en el algoritmo.
- Dar el proceso por lineal cuando exige iterar hacia atrás al aparecer evidencia.
- Entregar patrones sin interpretarlos ni validar su interés para el negocio.

Ejemplo resuelto: un proyecto de retención con KDD

Una telefónica quiere predecir qué clientes cancelarán y sigue el proceso completo. Las fases:

1. Comprender el problema: qué decisión apoya, con qué línea base y qué coste tiene el error.

2. Seleccionar y limpiar los datos, dedicando aquí el mayor esfuerzo porque manda sobre la calidad.
3. Transformar y reducir: codificar, imputar y seleccionar características dentro del pipeline.
4. Minar patrones con un modelo validado de forma honesta, sin fuga.
5. Interpretar y comunicar el hallazgo al negocio, con su incertidumbre y condiciones.

¿Para qué sirve en la realidad?

El KDD y su pariente CRISP-DM son el esqueleto de todo proyecto analítico serio: evitan el error de saltar al algoritmo, obligan a limpiar y validar, y aseguran que el resultado termine en una decisión comprensible y no en un número brillante e inútil.

1. ¿Qué fase del KDD, aunque poco glamorosa, suele determinar en mayor medida la calidad del resultado?

- ☐ La visualización final
- ☐ El preprocesado y limpieza de datos
- ☐ La elección del algoritmo de minado
- ☐ El despliegue en producción

2. El minado de patrones es todo el proceso KDD.

- ☐ Verdadero
- ☐ Falso

Comunicar el hallazgo

El conocimiento que no se comunica no existe. La última milla del KDD es traducir patrones a decisiones, con su incertidumbre y sus condiciones, en el lenguaje de quien debe actuar. Un hallazgo correcto pero ininteligible equivale a no haberlo encontrado.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. KDD



4. Ordena las fases canónicas del proceso KDD:

Elementos:

- comprender el problema de negocio y el dominio
- extraer y seleccionar los datos relevantes
- limpiar, transformar y reducir
- aplicar algoritmos para hallar patrones
- interpretar, validar y comunicar el conocimiento

Problema	. . .
Selección	. . .
Limpiar	. . .
Minar	. . .
Interpretar	. . .

5. Une cada fase del KDD con su tarea.

5.	extraer los datos relevantes	-----
5.	quitar ruido, duplicados y faltantes	-----
5.	normalizar, reducir y preparar	-----
5.	aplicar algoritmos para hallar patrones	-----

Opciones: seleccion · preprocesado · transformacion · minado

Proceso, no golpe de suerte

Proceso KDD

Negocio

entender el problema

línea base

Datos

seleccionar

limpiar

transformar

Minado

elegir algoritmo

validar honesto

Saber

interpretar

comunicar
desplegar

El mapa mental interactivo es interactivo en la versión online.



Conceptos básicos de CRISP-DM <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview> Resumen oficial de las fases de la metodología de minería de datos.

6. Tu equipo saltó directamente a entrenar un modelo y obtuvo precisión muy alta. Al volver al inicio del KDD, ¿qué dos fases olvidadas sospechas y qué comprobarías primero?

Referencia curricular

Grado:	5
Materia:	Minería de Datos · Machine Learning

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



<https://lumbre.lat/object/cdia-mine-003-modelado-del-kdd>

Material educativo generado desde Lumbre · lumbre.lat