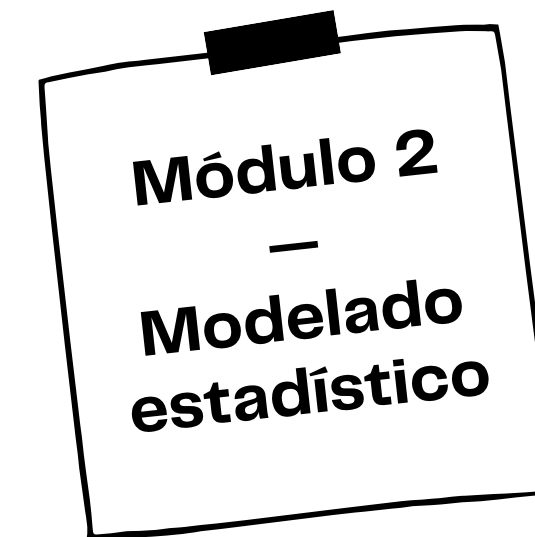


Minería de datos

Documento CRISP-DM



Q Agenda de hoy

- 1 ¿Qué es la minería de datos?
- 2 ¿Por qué surge la minería de datos?
- 3 CRISP-DM: Modelo para minería de datos
- 4 Fases del modelo de minería de datos

¿Qué es la minería de datos?

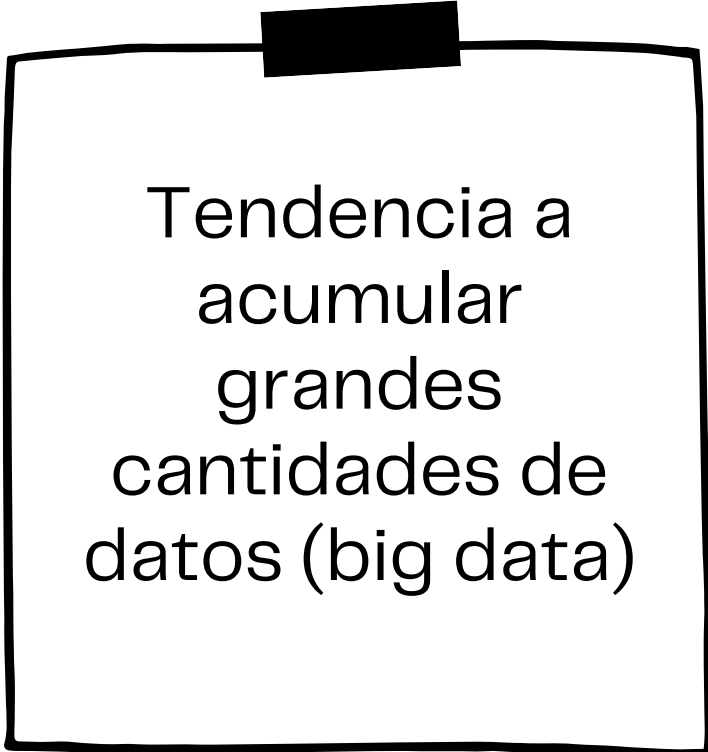
La minería de datos o data mining es el análisis automático de grandes cantidades de datos con el objetivo de descubrir y extraer patrones, relaciones y anomalías

Estadística

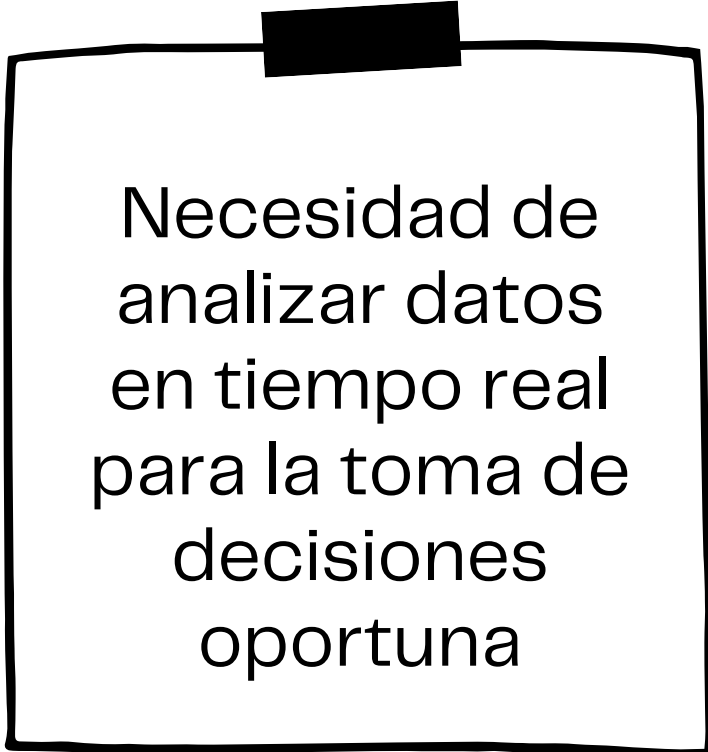
Inteligencia
Artificial

Aprendizaje
automático

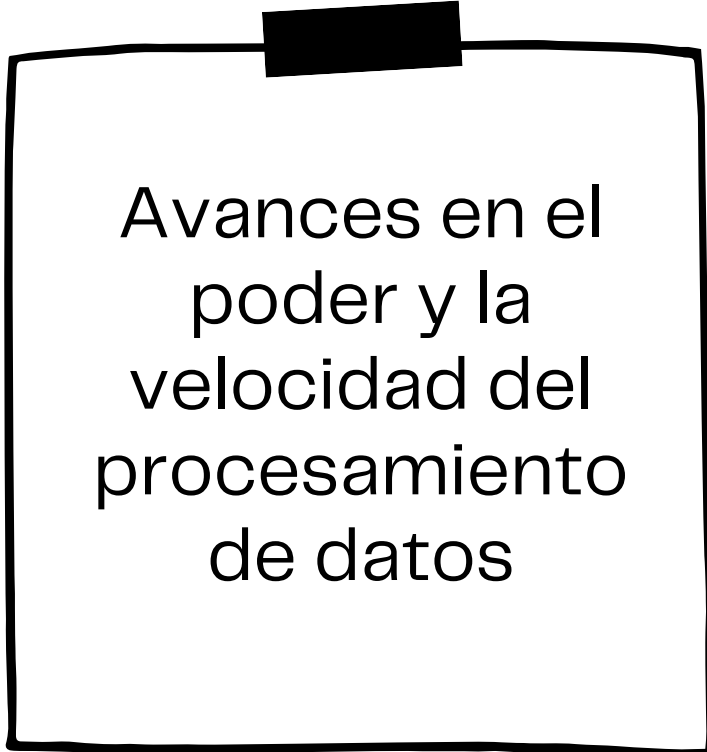
¿Por qué surge la minería de datos?



Tendencia a
acumular
grandes
cantidades de
datos (big data)



Necesidad de
analizar datos
en tiempo real
para la toma de
decisiones
oportuna



Avances en el
poder y la
velocidad del
procesamiento
de datos

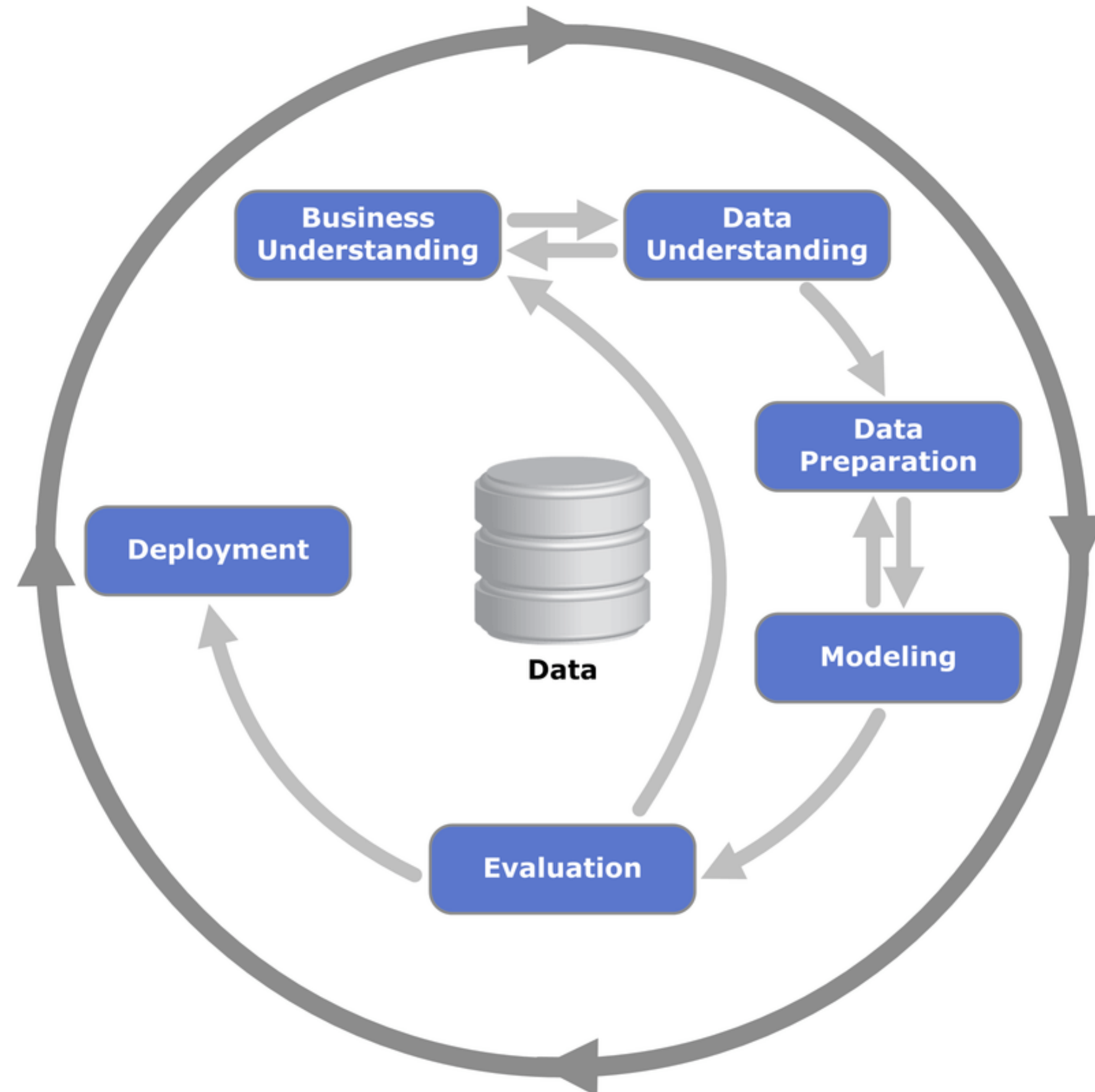
CRISP-DM: Modelo para minería de datos

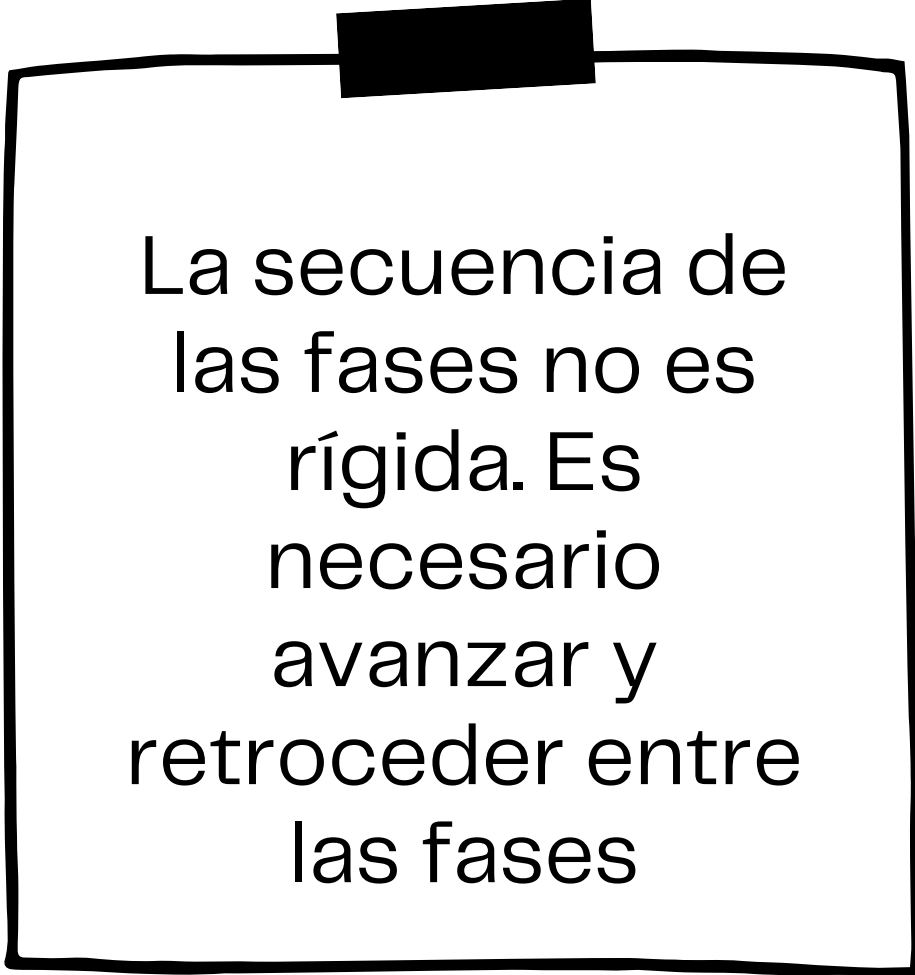
CRISP-DM es un modelo que proporciona una descripción general del ciclo de vida de un proyecto de minería de datos. Contiene las fases de un proyecto, sus respectivas tareas



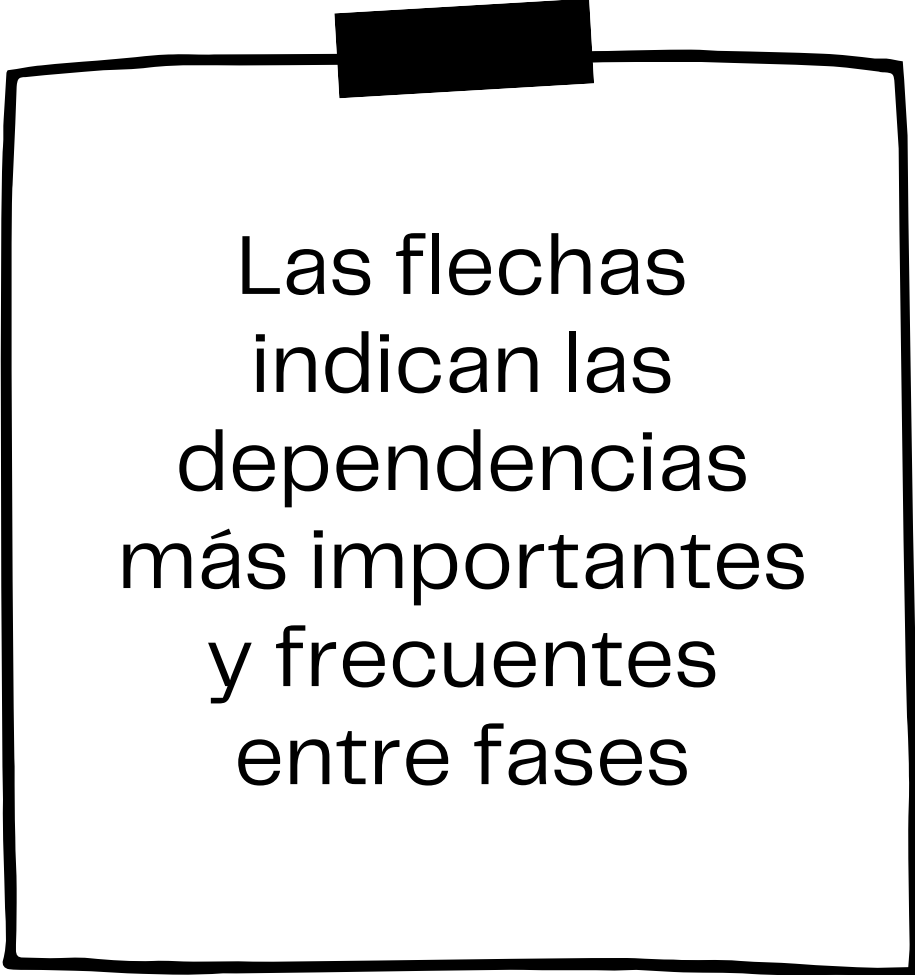
CRISP-DM:
Cross Industry
Standard
Process for
Data Mining

Fases del modelo

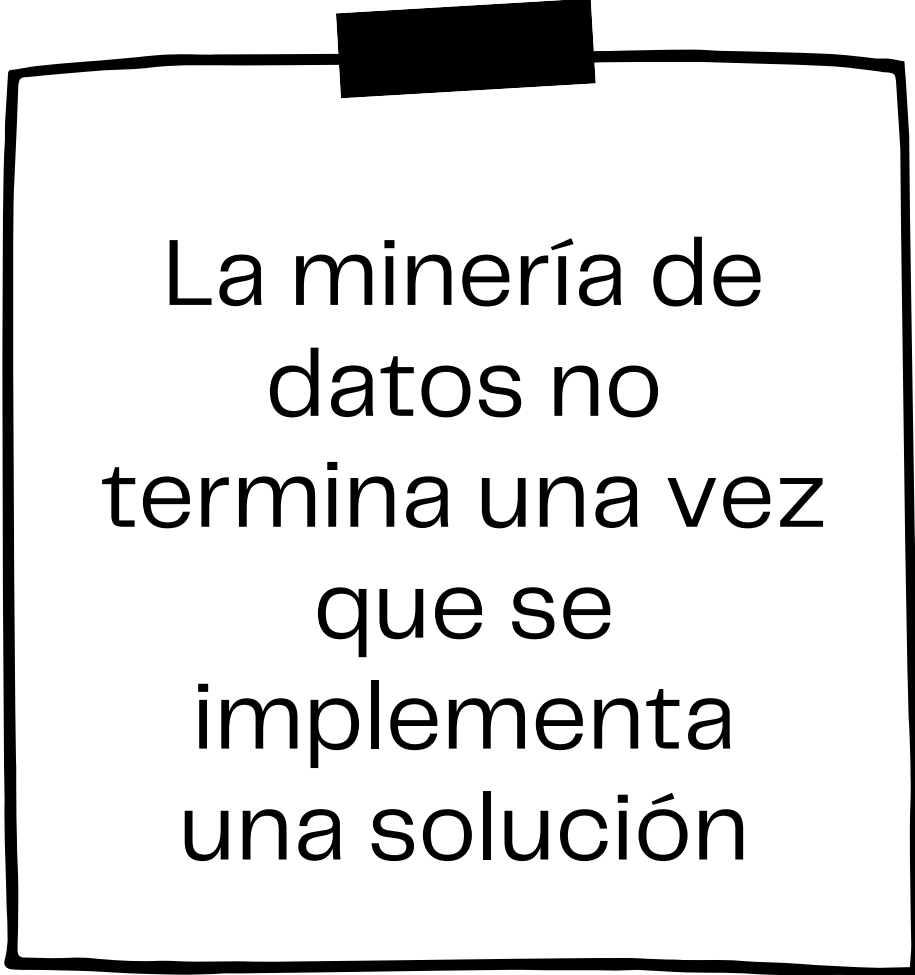


A hand-drawn rectangular box with a black border and rounded corners. A black rectangular clip is attached to the top center edge.

La secuencia de las fases no es rígida. Es necesario avanzar y retroceder entre las fases

A hand-drawn rectangular box with a black border and rounded corners. A black rectangular clip is attached to the top center edge.

Las flechas indican las dependencias más importantes y frecuentes entre fases

A hand-drawn rectangular box with a black border and rounded corners. A black rectangular clip is attached to the top center edge.

La minería de datos no termina una vez que se implementa una solución

Fase 1: Entendimiento del negocio

1

Determinar los objetivos del proyecto desde la perspectiva del negocio

2

Definir inventario de recursos y riesgos, costos y beneficios del proyecto

3

Determinar objetivos de la minería de datos (objetivos desde la perspectiva técnica)

4

Crear un plan de trabajo para alcanzar los objetivos planteados

Fase 2: Entendimiento de los datos

1

Recolectar base de datos inicial

2

Describir de forma muy general la base de datos

3

Explorar la base de datos (filtros, visualizaciones, reportes)

4

Verificar la calidad de los datos (identificar errores, valores faltantes)

Fase 3: Preparación de los datos

1

Seleccionar
los datos
que serán
utilizados
en el análisis

2

Limpiar la
base de
datos

3

Construir
nuevos
registros o
nuevos
atributos

4

Integrar
base de
datos con
otras tablas

5

Transformar el
formato de
algunos datos en
caso de ser
necesario

Fase 4: Modelación

1

Seleccionar técnica que utilizará para la modelación de los datos

2

Diseñar el proceso de entrenamiento, de pruebas y de evaluación del modelo

3

Ejecutar el o los modelos en la base de datos

4

Resumir los resultados entregados por los modelos y seleccionar los de mejor desempeño

Fase 5: Evaluación

1

Evaluar los resultados obtenidos con el mejor modelo y determinar si cumple con los objetivos planteados inicialmente

2

Hacer una revisión general del proceso de minería de datos y definir si hay tareas faltantes o que se deban repetir

3

Determinar los pasos siguientes dependiendo de los resultados obtenidos (se procede con el despliegue o se requieren acciones adicionales)

Fase 6: Despliegue

1

Diseñar una estrategia para el despliegue del modelo

2

Planear el monitoreo y mantenimiento del modelo

3

Crear reporte final del proceso de minería de datos

4

Identificar aspectos positivos, negativos y de mejora en el proceso