

Data Scientist



UNIDAD DIDACTICA 2

2.3 Data Mining

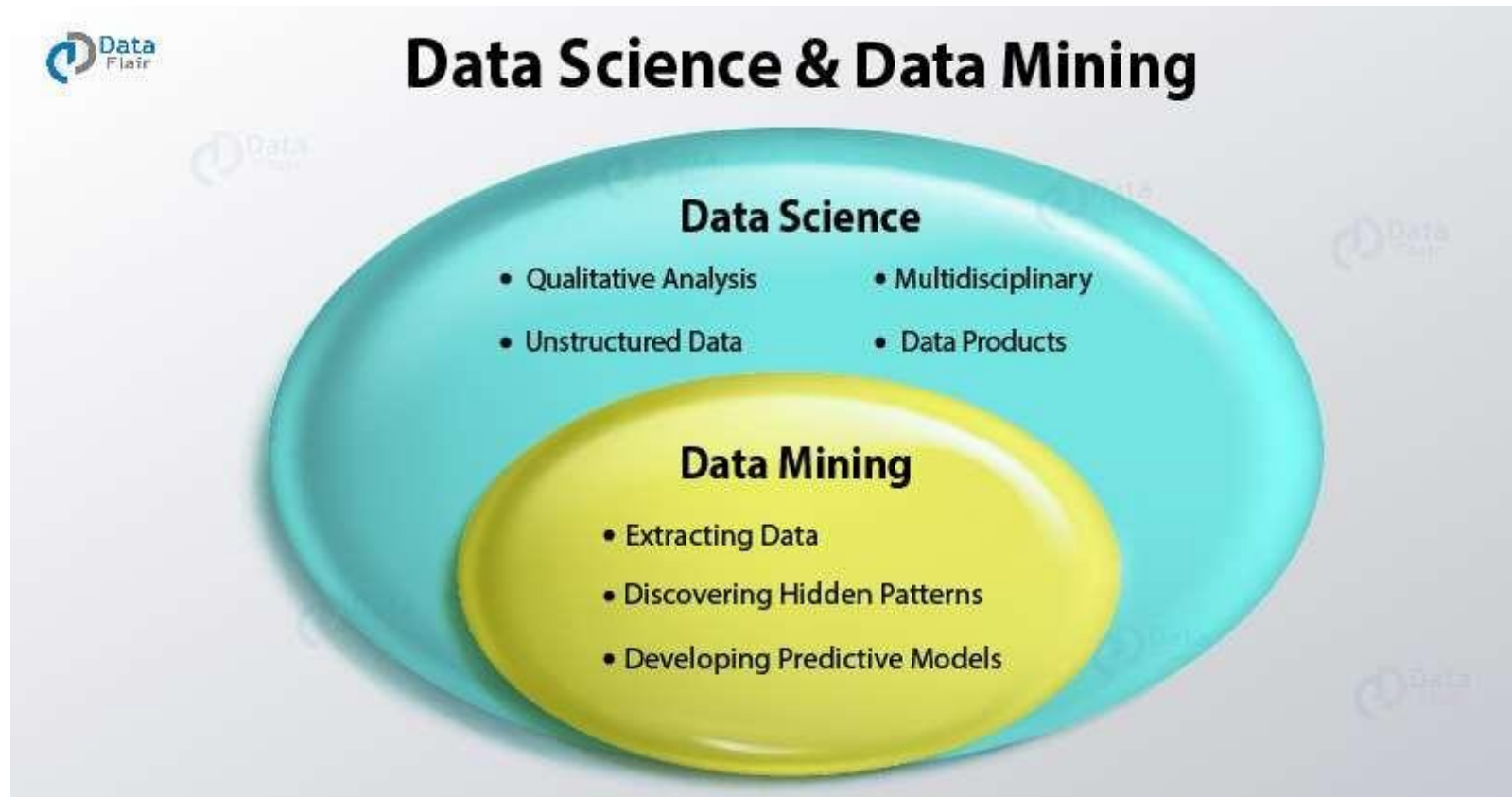
Ing. Gary David Guzmán Muñoz

Curso 2025

PREGUNTA

**¿Qué es Data
Mining?**

¿Qué es Data Mining?



La ciencia de datos es un conjunto de operaciones de datos que también implica minería de datos.



¿Qué es Data Mining?

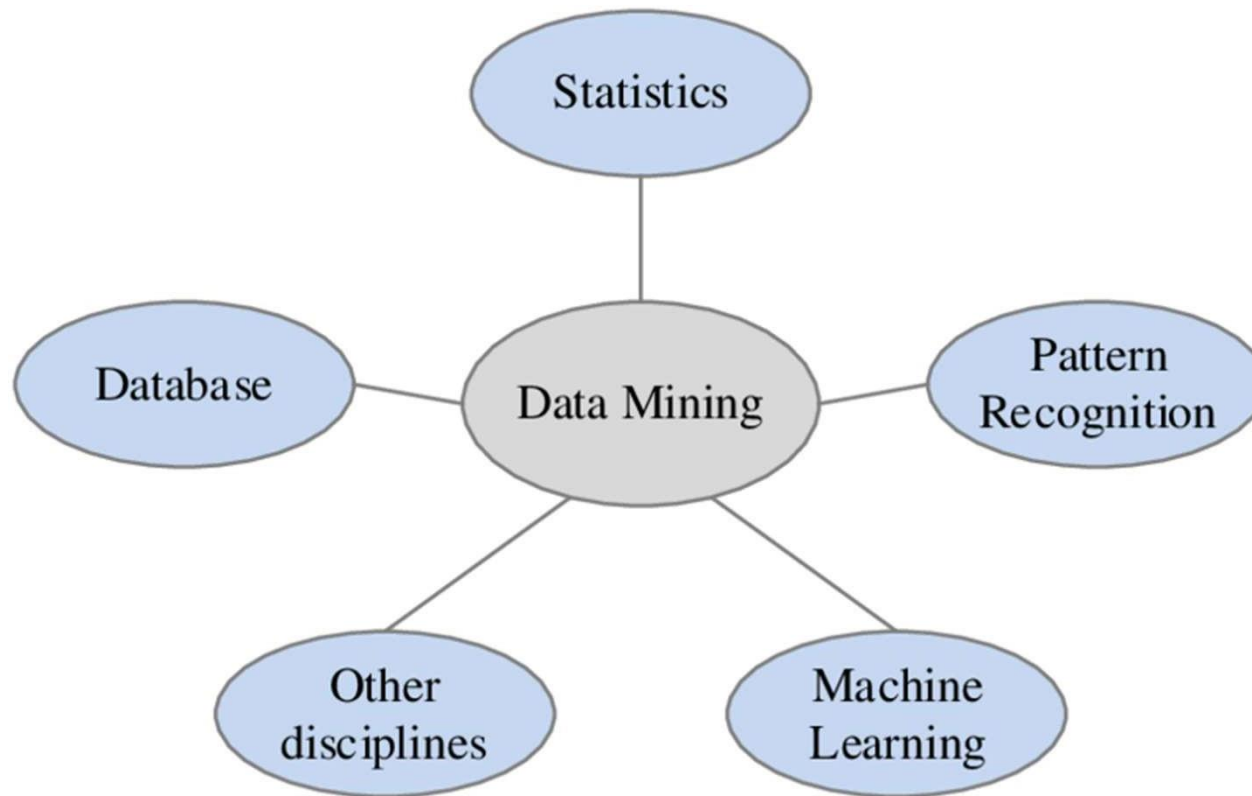
Es una disciplina que consiste en descubrir patrones en grandes conjuntos de datos mediante el uso de algoritmos y modelos estadísticos.

El objetivo de la minería de datos es extraer información útil de los datos que pueda utilizarse para mejorar la toma de decisiones.



¿Qué es Data Mining?

Engloba áreas desde la estadística pura al machine learning, pasando por gestión de datos.

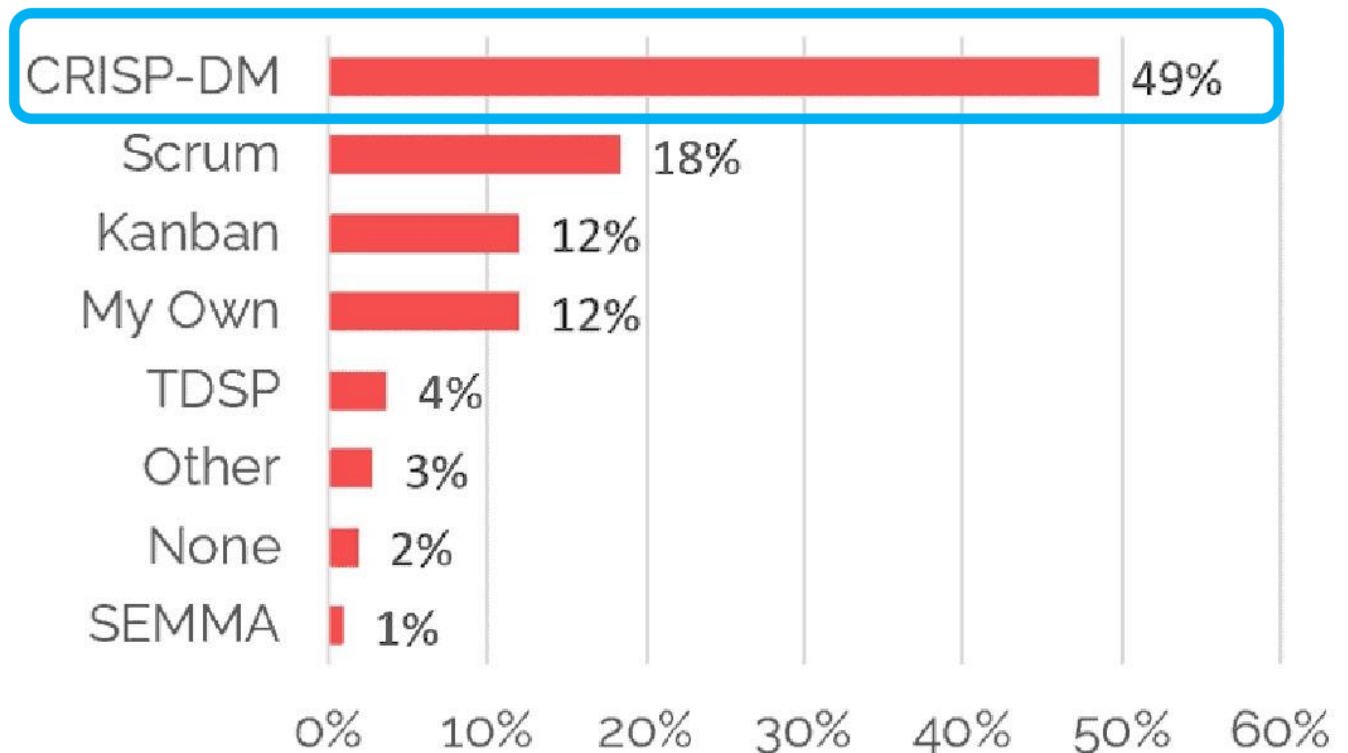


Data Mining y KDD

KDD es un **modelo iterativo e interactivo** para obtener conocimiento a partir de los datos

Es un **modelo genérico**, que requiere muchas **decisiones no automatizadas** y solo las **establece a grandes rasgos** (no profundiza en la descripción de tareas).

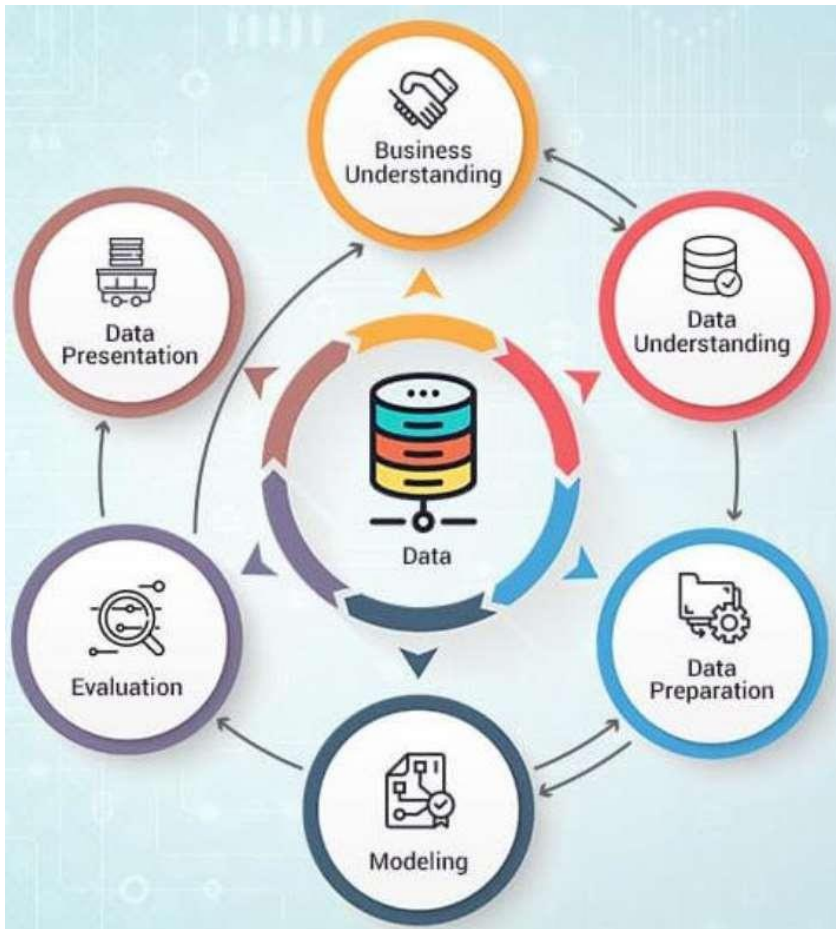
De este modelo surgen
modelos más específicos.



CRISP-DM

CRISP-DM

Cross-Industry Standard Process for Data Mining



Fases para el Data Mining:

1. **Entender el negocio:** Business Analyst. Objetivos de negocio a alcanzar (KPIs).
2. **Entender los datos:** Data Analyst. Analisis de fuentes de datos, cantidad, calidad, propiedades, cómo nos ayuda a entender negocio.
3. **Preparación de datos:** limpieza y almacenamiento.
4. **Modelado:** Data Scientist. Detección de patrones, estadísticos, predicciones (machine learning).
5. **Evaluación del modelo y resultados obtenidos.**
6. **Visualizacion** (técnicas business intelligence).

El 80% del tiempo

Fases para el Data Mining **CRISP-DM**:

1. **Entender el negocio:** Business Analist. Objetivos de negocio a alcanzar (KPIs).
2. **Entender los datos:** Data Analist. Analisis de fuentes de datos, cantidad, calidad, propiedades, cómo nos ayuda a entender negocio.
3. **Preparación de datos:** limpieza y almacenamiento.
4. **Modelado:** Data Scientist. Detección de patrones, estadísticos, predicciones (machine learning).
5. **Evaluación del modelo y resultados obtenidos.**
6. **Visualizacion** (técnicas business intelligence).

80%



Data Mining vs Machine Learning

Data Mining no es Machine Learning.

Son dos disciplinas interconectadas, valiéndose el Data Mining de los procesos de Machine Learning

Recursos y Bibliografía

- **lofullstack repository.**

Disponible en línea:

<https://github.com/iofullstack/data-scientist/tree/main/Web-Mining>

- **REF 1 - Data Mining Web Data Mining Techniques, Tools and Algorithms An Overview**