

CURSO ESPECIALIZACIÓN IA Y BIG DATA

MÓDULO: SISTEMAS DE APRENDIZAJE AUTOMÁTICO

U1.4: FASES DEL APRENDIZAJE AUTOMÁTICO



L'FSE inverteix en el teu futur

Fons Social Europeu - FSE



UNIO EUROPEA



**GENERALITAT
VALENCIANA**

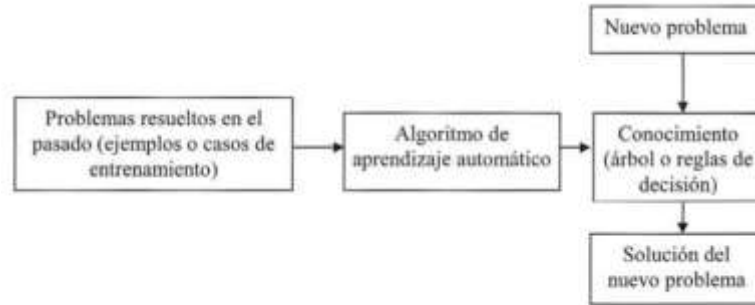
Conselleria d'Educació,
Cultura i Esport

Índice

- Machine learning
- El proceso del análisis predictivo
- Metodologías

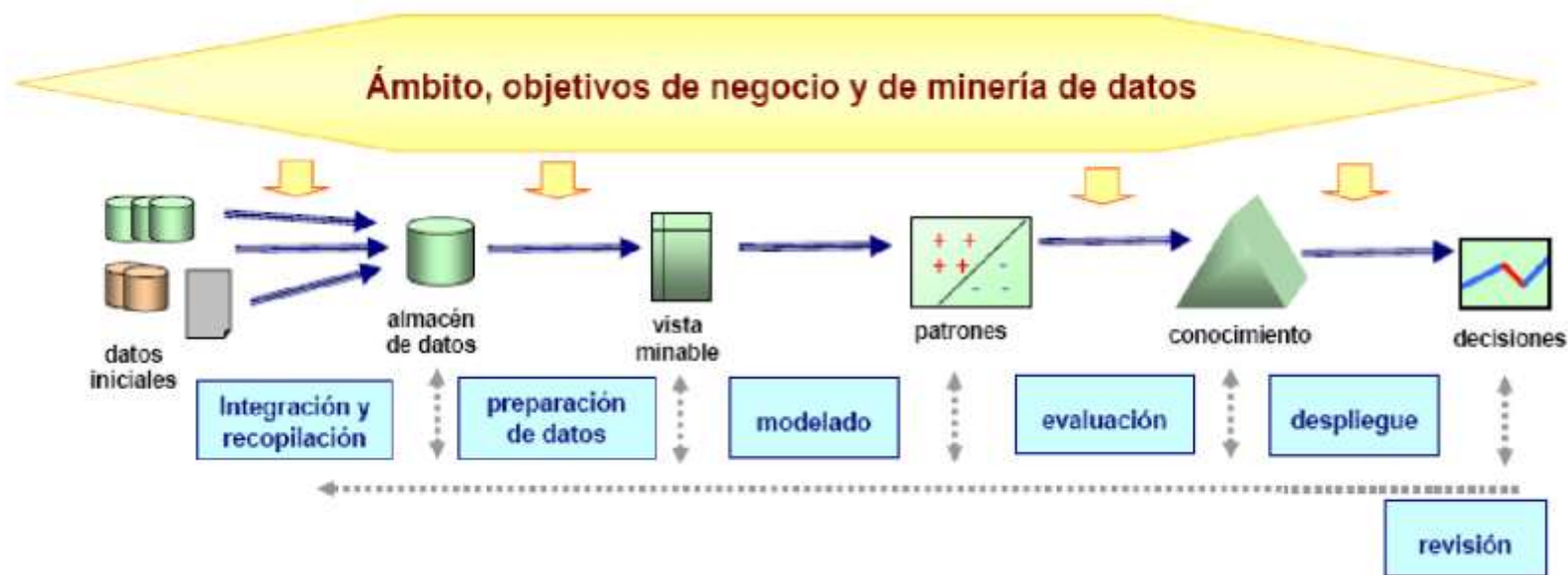
➤ Machine learning

- Machine learning es un subcampo de las ciencias de la computación y la inteligencia artificial que se sirve de algoritmos que permiten a las máquinas aprender imitando la forma en que lo hacen los seres humanos.
- Los juegos de datos encierran estructuras, patrones y reglas de los que es posible extraer conocimiento sobre los eventos que los han generado.



- Existen gran cantidad de algoritmos en minería de datos que se pueden organizar en función del tipo de aprendizaje que usan: SUPERVISADOS y NO SUPERVISADOS.

➤ El proceso del análisis predictivo





El proceso del análisis predictivo

Las fases del proceso del KDD (Knowledge Discovery in Databases) o análisis predictivo son las siguientes:

- **Integración y recopilación**

Resultado: Almacén de datos

- **Preparación de datos**

Resultado: Vista minable (Vista preparada para la aplicación del método de minería de datos)

- **Minería de datos:** Seleccionar y aplicar el método de minería de datos apropiado.

Resultado: Patrones o modelos

- **Evaluación/Interpretación/Visualización:** Consiste en la evaluación, interpretación, transformación y representación de los patrones extraídos.

Resultado: Conocimiento

- **Difusión y uso del nuevo conocimiento:** El conocimiento generado debe ser aplicado.

Resultado: Decisiones



El proceso del análisis predictivo

Integración y recopilación:

1. Determinar las fuentes de información que pueden ser útiles y dónde conseguir las.
2. Diseñar el esquema de un almacén de datos que consiga unificar de manera operativa toda la información recogida.
3. Implantación del almacén de datos que permita la “navegación” y visualización previa de sus datos, para discernir qué aspectos pueden interesar en el estudio.

Resultado: Almacén de datos



El proceso del análisis predictivo

Preparación de datos:

1. Estudiar y comprender los datos que necesitamos
2. Selección, limpieza y transformación de los datos que se van a analizar. La selección incluye tanto una criba o fusión horizontal (filas) como vertical (atributos). Esta preparación está condicionada por el método de minería de datos que se aplicará.

Resultado: Vista minable (Vista preparada para la aplicación del método de minería de datos)



El proceso del análisis predictivo

Minería de datos: Seleccionar y aplicar el método de minería de datos apropiado.

Resultado: Patrones o modelos

Evaluación/Interpretación/Visualización: Consiste en la evaluación, interpretación, transformación y representación de los patrones extraídos.

Resultado: Conocimiento

Difusión y uso del nuevo conocimiento: El conocimiento generado debe ser aplicado.

Resultado: Decisiones

En todo momento el proceso es dirigido por su ámbito y sus objetivos. El resultado del KDD debe ser COMPENSIBLE, de otro modo no podrá usarse.

Las técnicas de Data Science surgen en la década de los 90.

En un intento de normalizar este proceso se crean a finales de los 90 dos metodologías principales:

- CRISP-DM (Cross Industry Standard Process for Data Mining)
- SEMMA (Sample, Explore, Modify, Model and Asses)

Ambas especifican las tareas a realizar en cada fase descrita por el proceso, y definiendo la salida de cada fase.

CRISP-DM es más completo porque tiene en cuenta la aplicación al entorno de negocio de los resultados y la más usada.

CRISP-DM

Fases:

Comprensión del negocio: Objetivos del Proyecto y comprensión de requerimientos. Definición del problema de Minería de Datos

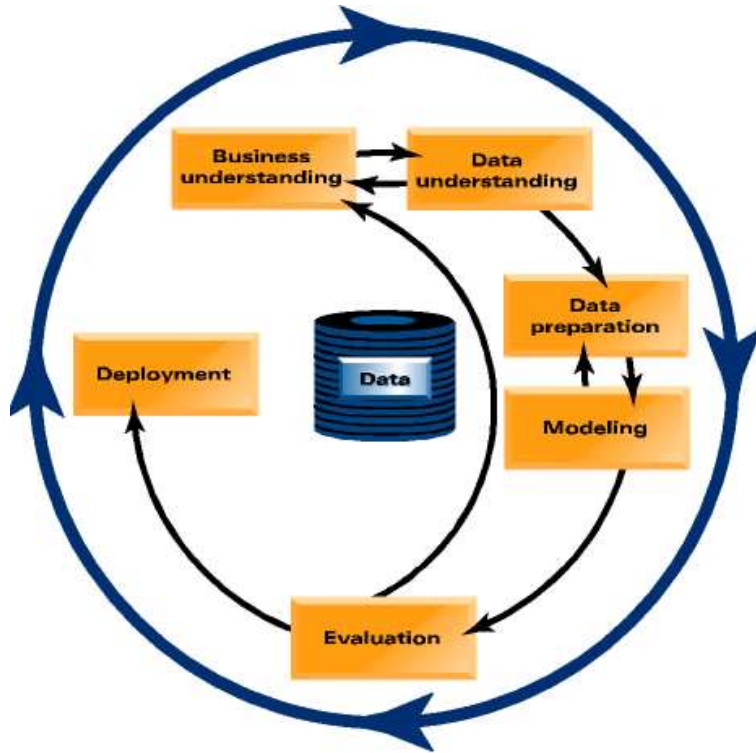
Comprensión de los datos: Recolección y recopiación de datos inicial. Identificación de problemas en la calidad de los datos.

Preparación de datos: Selección de tablas, registros y atributos. Limpieza y transformación de datos

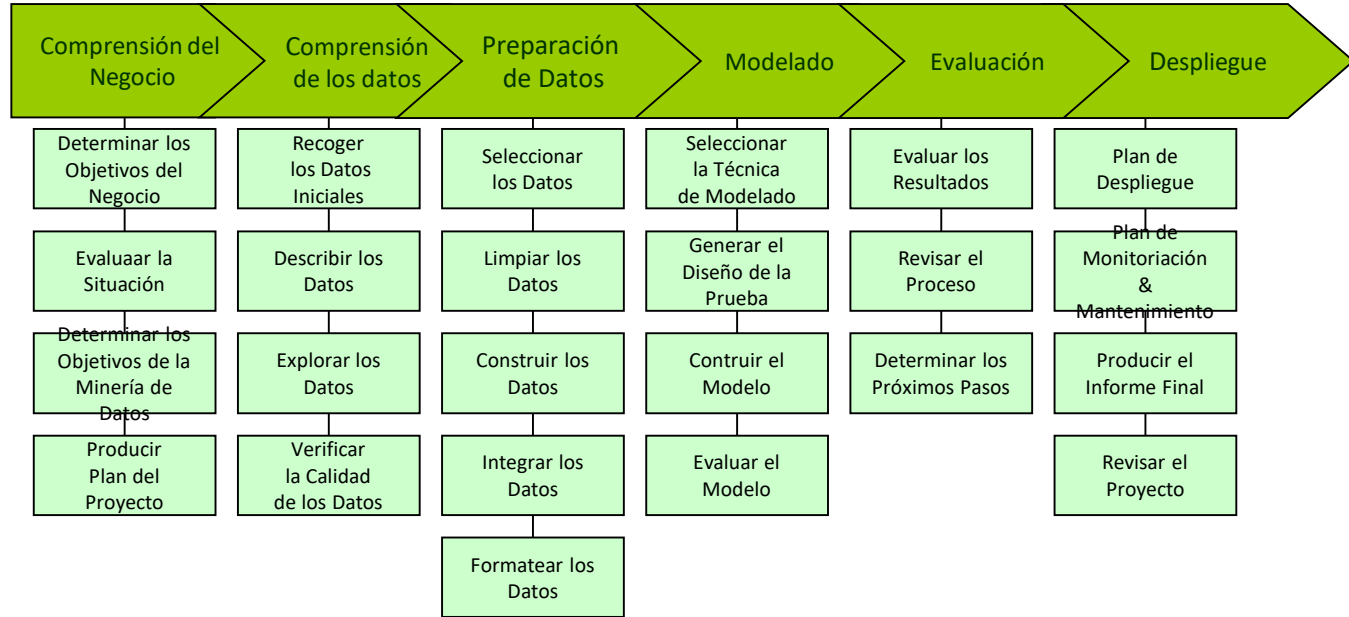
Modelado: Selección de técnicas de modelado y su aplicación. Calibración de parámetros

Evaluación: Evaluación de los objetivos y aspectos de negocio alcanzados.

Despliegue/Desarrollo: Desarrollo del modelo de resultados Implementación de la repetición del proceso de data mining



CRISP-DM



Gracias