

# Introducción

Minería de Datos - Grado en Matemáticas

DSLAB

2025-09-24



*We are drowning in information and starving for knowledge. — John Naisbitt*

*This deluge of data calls for automated methods of data analysis, which is what machine learning provides. In particular, we define **machine learning** as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!)*

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Sistemas que piensan como humanos	Sistemas que piensan racionalmente
<p>«El nuevo y excitante esfuerzo de hacer que los computadores piensen... máquinas con mentes, en el más amplio sentido literal». (Haugeland, 1985)</p> <p>«[La automatización de] actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje...» (Bellman, 1978)</p>	<p>«El estudio de las facultades mentales mediante el uso de modelos computacionales». (Charniak y McDermott, 1985)</p> <p>«El estudio de los cálculos que hacen posible percibir, razonar y actuar». (Winston, 1992)</p>
Sistemas que actúan como humanos	Sistemas que actúan racionalmente
<p>«El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia». (Kurzweil, 1990)</p> <p>«El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor». (Rich y Knight, 1991)</p>	<p>«La Inteligencia Computacional es el estudio del diseño de agentes inteligentes». (Poole <i>et al.</i>, 1998)</p> <p>«IA... está relacionada con conductas inteligentes en artefactos». (Nilsson, 1998)</p>

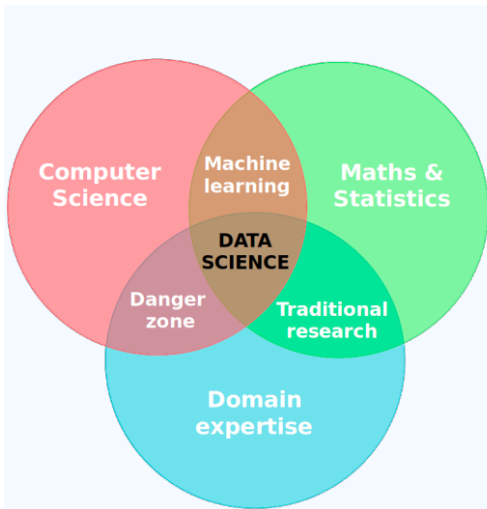
**Figura 1.1** Algunas definiciones de inteligencia artificial, organizadas en cuatro categorías.

**Figure 1:** Russell, S. J., & Norvig, P. (2010). Artificial intelligence a modern approach. London.

La **inteligencia artificial** es un campo de las ciencias de la computación que se enfoca en el desarrollo de sistemas y programas informáticos capaces de realizar tareas que, cuando se ejecutan por parte de seres humanos, requieren inteligencia y aprendizaje. Estos sistemas de IA pueden aprender de datos, adaptarse a nuevas situaciones, tomar decisiones, resolver problemas y realizar tareas específicas sin intervención humana directa. La IA busca imitar y replicar procesos cognitivos y de toma de decisiones humanas, permitiendo a las máquinas realizar actividades que normalmente requerirían la inteligencia humana.

Multitud de disciplinas interconectadas:

- Robótica
- **Aprendizaje Automático**
- Procesamiento del Lenguaje Natural (NLP)
- Visión por computador
- Lógica difusa
- Redes neuronales artificiales → Deep Learning
- Reconocimiento de patrones
- Computación evolutiva

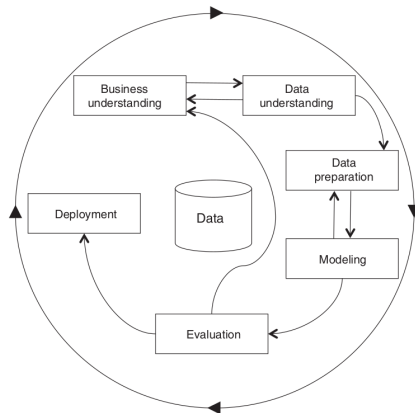


(a) Foundations



(b) Applications

## Cross Industry Standard Process for Data Mining (CRISP-DM)



**Figure 4** The CRISP-DM life cycle (based on figure 2 in Chapman, Clinton, Kerber, et al. 1999).

**Figure 2:** Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.

- **Business Understanding:** Comprender plenamente el problema empresarial que se aborda y diseñar una solución de análisis de datos para el mismo
- **Data Understanding:** Comprender las diferentes fuentes de datos disponibles en la entidad y los diferentes tipos de datos que contienen dichas fuentes
- **Data Preparation:** Poner las distintas fuentes de datos disponibles en un formato adecuado a partir del cual puedan inducirse modelos de aprendizaje automático
- **Modeling:** Crear distintos modelos de aprendizaje automático y seleccionar el mejor para su implantación
- **Evaluation:** Estudiar y validar el rendimiento del modelo para confirmar que es capaz de hacer predicciones precisas antes de ser desplegado
- **Deployment:** Integrar con éxito el modelo de aprendizaje automático en el proceso de la empresa/organización



- **Aprendizaje supervisado:** Datos  $\mathbf{x}_i$  con etiquetas, La etiqueta  $y_i$  refleja el valor de la variable objetivo (respuesta). El objetivo es “aprender” una función  $f$  de los inputs  $\mathbf{x}_i$  para predecir los outputs  $y_i$  dado un conjunto  $\mathcal{D}$  de datos formado por  $n$  pares de observaciones:

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$$

- **Clasificación:** variable objetivo es cualitativa  $y_i \in \{1, 2, \dots, n\}$ . Ej: predecir si un paciente tiene una enfermedad o no
- **Regresión:** variable objetivo es cuantitativa  $y_i \in \mathbb{R}$ . Ej: predecir la resistencia de un material en base a sus características

- **Aprendizaje no supervisado:** Datos no etiquetados. Se parte de un conjunto de datos  $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^n$  y el objetivo es descubrir patrones de interés entre los datos, describir los datos encontrando grupos de observaciones similares entre sí (Clustering). Ej: Encontrar grupos de clientes similares para una campaña de marketing.

Algunos libros añaden la siguiente categoría:

- **Aprendizaje por refuerzo.** Para aprender cómo actuar a través de recompensas o castigos (un bebé aprendiendo a caminar)

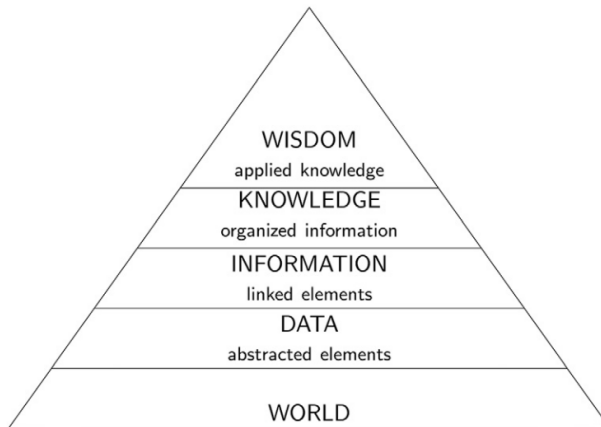


Figure 2 The DIKW pyramid (adapted from Kitchin 2014a).

**Figure 3:** Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.

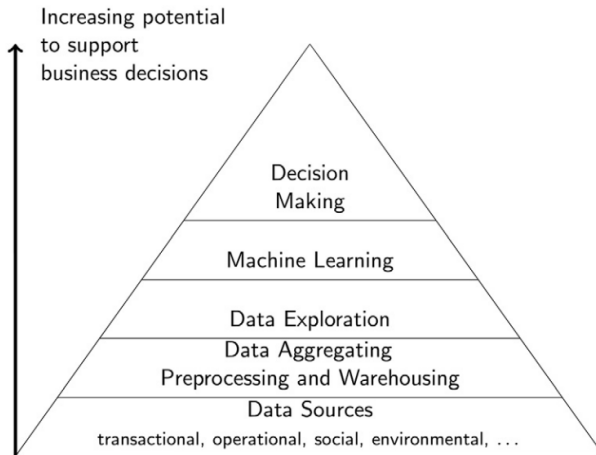


Figure 3 Data science pyramid (adapted from Han, Kamber, and Pei 2011).

**Figure 4:** Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence a modern approach*. London.

Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.