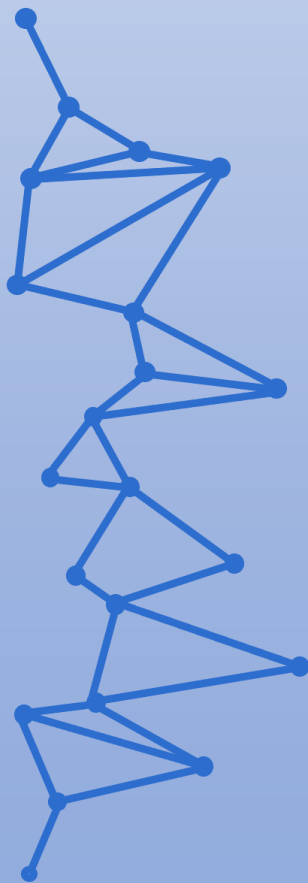




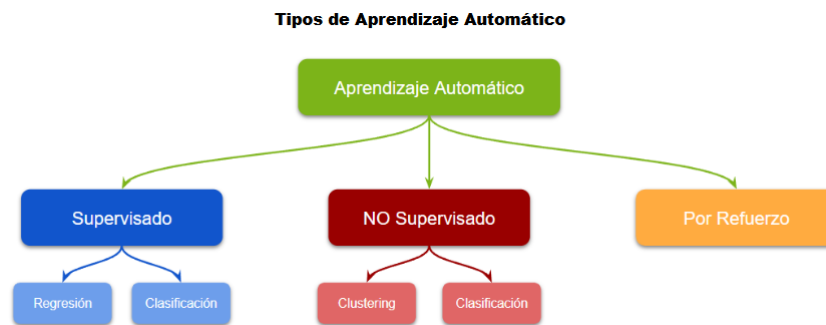
Curso de Especialización de Inteligencia Artificial y Big Data (IABD)



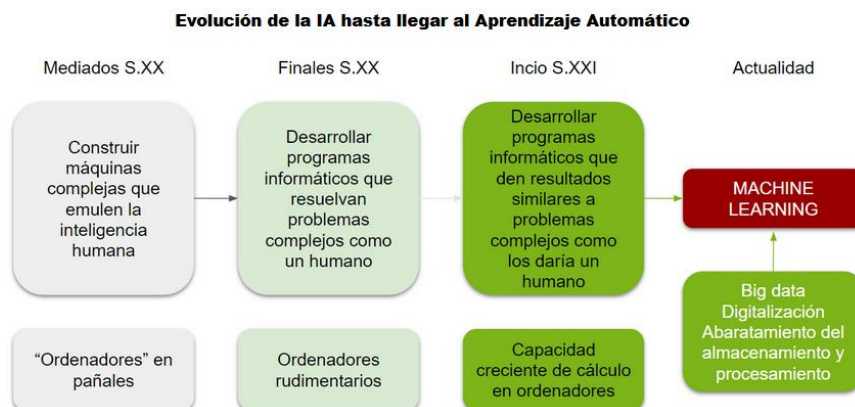
Modelos de Inteligencia Artificial

UD04. Modelos de Aprendizaje Automático.
Resumen.

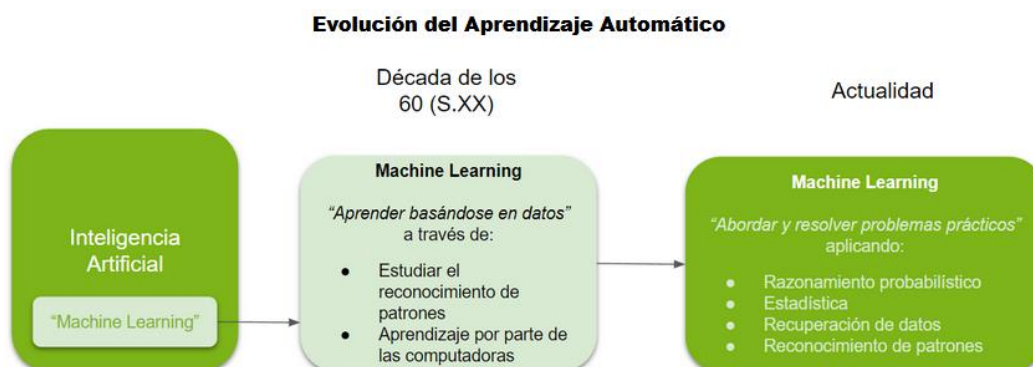
JUAN ANTONIO GARCIA MUELAS



Es importante entender que el **Aprendizaje Automático es una rama de la IA**, aunque en la actualidad se utiliza en todos los proyectos de IA.



El *Machine Learning* o **Aprendizaje Autónomo** (Automático) inicialmente se focalizaba en lograr que la máquina aprendiera **basándose en datos**, estudiando el **reconocimiento de patrones** (casos similares entre el total de elementos del *data set* o base de datos). Ahora **se centra más en "resolver" problemas prácticos que en "aprender"**, aunque evidentemente "aprende". Al reconocimiento de patrones se añade ahora el **razonamiento probabilístico**, la **estadística** y la **recuperación de datos**.



En un proyecto de Aprendizaje Automático suele requerir mucho más tiempo y esfuerzo la fase de recogida de datos (RAW), limpieza de los mismos y su transformación para ser "comprendidos" por la herramienta de entrenamiento, que la fase de entrenamiento y aprendizaje propiamente dicha.

DEFINICIONES.

Arthur Samuel, lo definía como el campo del estudio que da a las computadoras la capacidad de **aprender sin ser programadas explícitamente**.

Tom Mitchell: Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna clase de tareas T y medida de rendimiento P , si su desempeño en las tareas en T medido por P mejora con la experiencia E .

Por ejemplo: jugar a las damas.

E es la **experiencia** de jugar muchas partidas de damas.

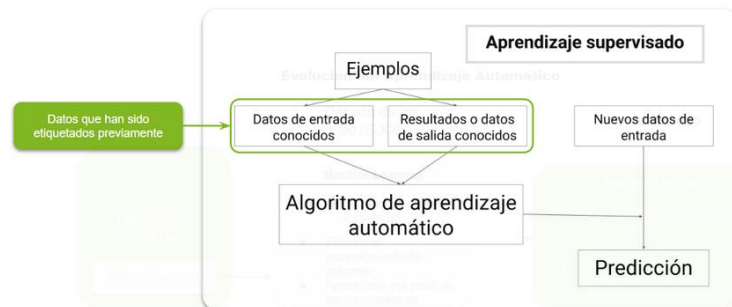
T es la **tarea** de jugar a las damas.

P es la **probabilidad** de que el programa gane la partida actual.

La máquina observa, ganando **experiencia**, hace la **tarea** y comprueba su **rendimiento**.

El **Aprendizaje Automático** (*Machine Learning*) es un proceso de **adquisición** de conocimiento de manera **automática** mediante la utilización de **ejemplos** (experiencia) de entrenamiento.

La característica fundamental del **Aprendizaje Automático Supervisado** es que dicho aprendizaje se realiza a partir de **datos que ya han sido etiquetados previamente**, o sea, **aprende sobre datos** (de entrada, o de salida) que le proporcionamos **con sus características**.



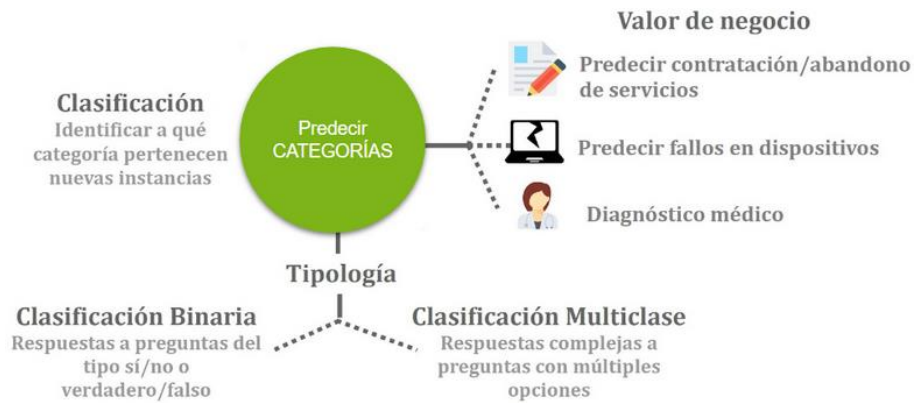
Los **problemas de Aprendizaje Supervisado** se dividen en dos categorías: **Regresión** (donde el **campo objetivo o respuesta es de tipo numérico**) y **Clasificación** (donde el campo objetivo o **respuesta es de tipo categórico**).

En los problemas de **Regresión** se busca predecir **qué valor tendrá** el campo objetivo para una nueva instancia.



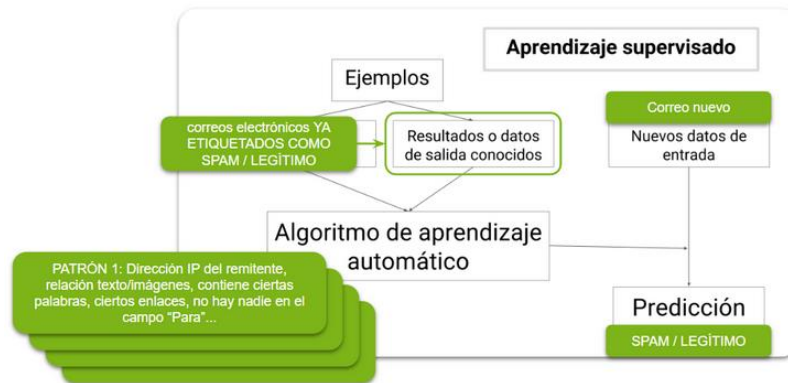
Los problemas de **Clasificación** buscan predecir **a qué categoría pertenece el campo objetivo** de cada instancia a partir de una lista de posibles categorías (si sólo hay dos categorías posibles, será un problema de **Clasificación Binaria**; si requieren respuestas complejas y predicciones entre múltiples categorías **Clasificación Multiclase**).

El **valor de negocio** de los problemas de Clasificación es muy amplio y **abarca**, entre otros, los campos de la **economía**, la **tecnología** y la **medicina**.



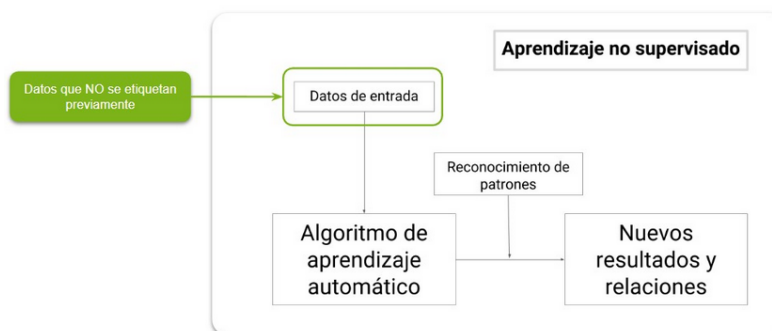
Por lo tanto, uno de los usos más extendidos del **Aprendizaje Supervisado** consiste en hacer predicciones a futuro basadas en comportamientos o características que se han visto en los datos ya almacenados (el histórico de datos).

El aprendizaje supervisado permite **buscar patrones** en datos históricos relacionando todos campos con un campo especial, llamado **campo objetivo**.



El **aprendizaje no supervisado** usa **datos históricos que no están etiquetados**. El fin es explorarlos para encontrar alguna estructura o forma de organizarlos.

Esquema Aprendizaje No Supervisado



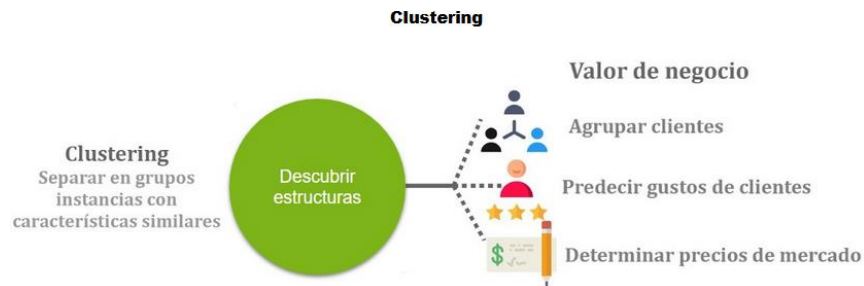
El **objetivo** es que la IA encuentre relaciones de tres tipos:

- ✓ Similitudes, Diferencias y Asociaciones.

Dependiendo de cuál sea dicho objetivo, los **problemas se clasificarán en** tres tipos diferentes:

- ✓ **Clustering**. El objetivo es generar **agrupaciones** o *clusters* buscando las **instancias** que son **similares entre sí** y predecir a que grupo pertenecerá la siguiente instancia.

Entendemos por instancia, cada uno de los elementos que forman el conjunto de datos que se proporcionan para dicho Aprendizaje Automático.



Permite hacer estudios de precios, publicaciones o series sugeridas...

- ✓ **Detección de Anomalías.** Busca las **instancias que se diferencian** de las demás (ejemplo: fraudes económicos).



- ✓ **Asociaciones.** El objetivo es **encontrar relaciones entre los diferentes valores** que toman los campos de una instancia. Por ejemplo, encontrar la relación en la venta de un producto e indicar si se adquiere con más probabilidad junto a otro.



En **Aprendizaje por Refuerzo**, el objetivo es aprender cómo **mapear situaciones o acciones** para maximizar una cierta recompensa. Se trata de **programar agentes mediante premio y castigo** sin necesidad de especificar cómo realizar la tarea. Lo más importante es definir y programar las condiciones que deben cumplirse.

Para saber más

Minería de datos o *Data mining*

No es raro ver cómo se usan indistintamente los conceptos minería de datos y *machine learning*. Son conceptos "primos hermanos", pero no son lo mismo. Desde nuestro punto de vista, la principal diferencia radica en **el objetivo** que tiene cada una de las disciplinas. Mientras que **la minería de datos descubre patrones anteriormente desconocidos**, **el machine learning se usa para reproducir patrones conocidos y hacer predicciones basadas en los patrones**.

En pocas palabras se podría decir que la minería de datos tiene una función exploratoria mientras que el machine learning se focaliza en la predicción.