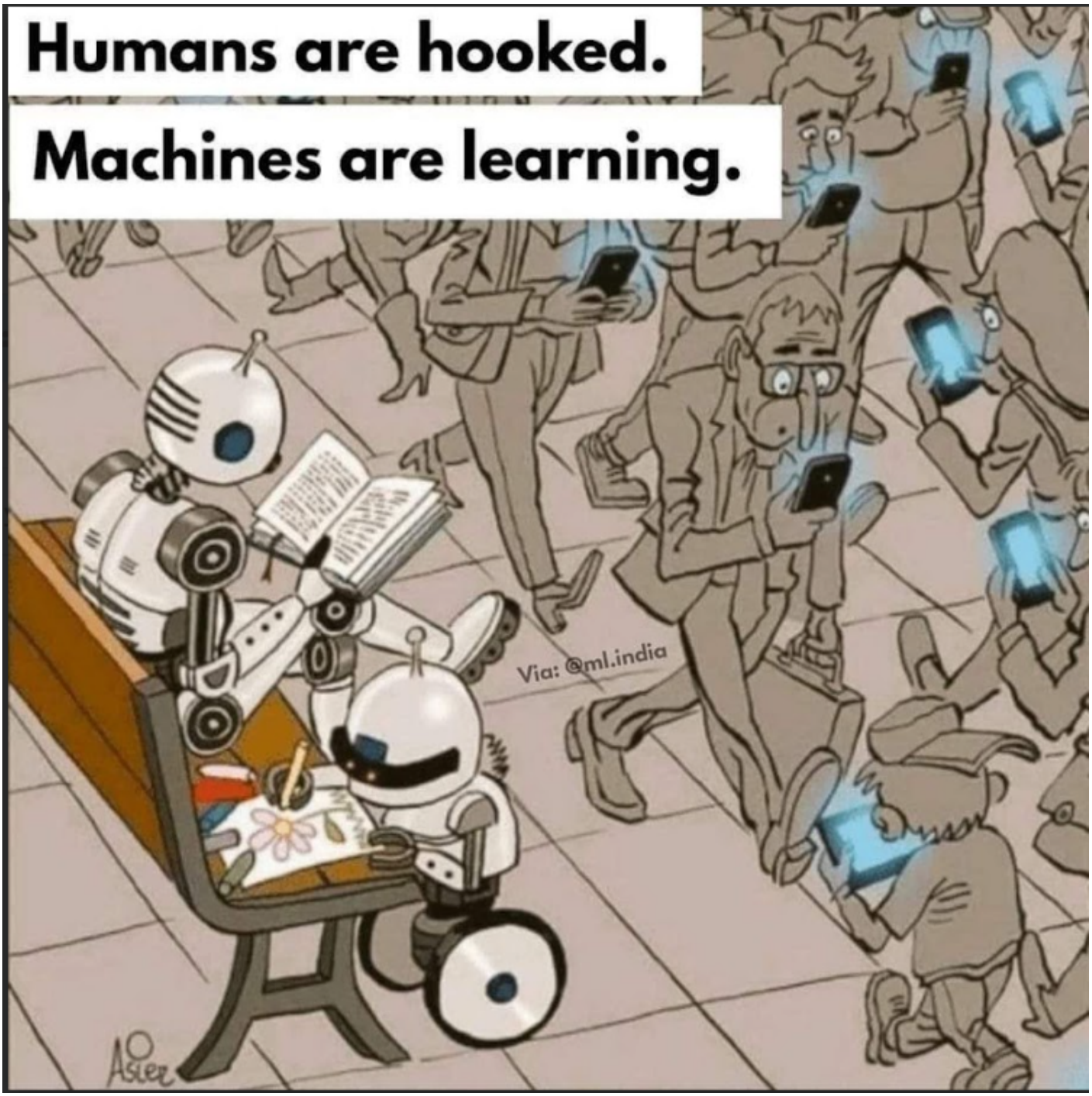
Unidad 2.2. Machine Learning. Como aprender sin instrucciones

Machine Learning

> **Concepto**
>
> "El aprendizaje automático es el arte y la ciencia de permitir a las máquinas aprender y mejorar automáticamente a partir de los datos, sin necesidad de ser programadas explícitamente para cada tarea específica".

Copyright 2026



Humanos 'enganchados' mientras las computadoras aprenden

La forma tradicional de hacer que una computadora logre algo es darle instrucciones explícitas (si sucede esto, haz esto, de lo contrario haz aquello, etc.) que se escriben a mano en un lenguaje de programación determinado. Este método de programación de computadoras es extremadamente exitoso y se ha utilizado para desarrollar prácticamente todo el software que se ejecuta en nuestras computadoras, teléfonos e incluso automóviles. Sin embargo, este método no siempre es el más práctico.

Copyright 2026

Por ejemplo, consideremos el desarrollo de un programa para identificar imágenes. Habría que considerar todas las posibles combinaciones de píxeles para que en función de sus características podamos instruir al algoritmo o computadora acerca del objeto de que está 'visualizando'.

Sólo el juego GO ofrece 10 elevado a 700 posibilidades en cuanto a mposibles movivimientos lo que da una idea de lo incorrecto de este enfoque, al menos para el objetivo de que las máquinas aprendan y sean capaces de usar y 'entender' el lenguaje natural, además de distinguir objetos y otras tareas propias de nuestra inteligencia.

Sucede que se nos da muy bien programar cosas que hemos aprendido a hacer como una suma o una multiplicación pero nos resulta extremadamente difiicil enseñar lo que nos es innato, por ejemplo andar, mover un barzo etc...

El aprendizaje automático es un enfoque que permite a la computadora aprender patrones y relaciones en los datos a través de algoritmos y modelos matemáticos. En el caso de la identificación de imágenes, en lugar de programar reglas explícitas, se alimenta a la computadora con un conjunto de imágenes etiquetadas y se le permite aprender por sí misma cómo reconocer diferentes objetos o patrones en las imágenes.

El proceso de aprendizaje automático implica entrenar a la computadora utilizando ejemplos y ajustar sus conexiones internas, de modo que pueda hacer predicciones precisas sobre nuevas imágenes que nunca ha visto antes. Esto se logra mediante el uso de algoritmos de aprendizaje automático, como las redes neuronales, que son capaces de capturar características complejas y extraer información útil de los datos de entrada.

A medida que se alimenta a la computadora con más datos y se perfeccionan los algoritmos de aprendizaje, la precisión y la capacidad de reconocimiento de imágenes de la computadora mejoran con el tiempo. Este enfoque es especialmente útil en casos en los que las reglas explícitas serían demasiado difíciles o tediosas de programar, o cuando los patrones en los datos no son fácilmente discernibles para los humanos.

En resumen, el enfoque tradicional de programación mediante instrucciones explícitas está siendo complementado y en algunos casos reemplazado por el aprendizaje automático, que permite a las computadoras aprender a reconocer patrones y tomar decisiones basadas en datos en lugar de reglas predefinidas. Esta técnica está transformando la forma en que desarrollamos programas y abre un amplio abanico de posibilidades para aplicaciones más inteligentes y adaptativas.

Sería casi imposible escribir un programa así "a mano" debido al gran número de posibilidades a considerar (diferentes orientaciones de objetos, condiciones de iluminación, obstrucciones, etc.). En cambio, sería más fácil utilizar el método de aprendizaje automático, que es un método y una disciplina científica utilizada para programar computadoras utilizando datos en lugar de instrucciones explícitas. En el caso de la identificación de imágenes, el método de aprendizaje

Copyright 2026

automático consistiría en proporcionar a la computadora un conjunto de imágenes etiquetadas según lo que representan y permitir que la computadora descubra por sí misma cómo identificar nuevas imágenes.

Aprendizaje con datos no estructurados

Entendemos por datos no estructurados aquellos que se presentan sin un formato ni estructura específico, como los que provienen del lenguaje natural o en forma de imagen o sonido. Por contra datos estructurados serían aquellos que se presentan con algún tipo de organización como una tabla o un fichero con determinadas características.

En esta sección hablaremos de cómo hacer que las máquinas entiendan datos no estructurados.

Siguiendo con lo comentado al principio de la sección vamos a ver los datos de un modelo para clasificar vehículos distinguiendo entre coches y camiones.

El paradigma de aprendizaje automático permite crear clasificadores que a partir de un conjunto de datos de entrada (en este caso en forma de imágenes etiquetadas) pueda entrenarse y despues al recibir una imagen nueva podrá identificarla correctamente.

Por ejemplo aqui podemos ver un conjunto de imágenes de coches y camiones


COCHE

COCHE

CAMIÓN

CAMIÓN

COCHE

COCHE

CAMIÓN

A la computadora se le han proporcionado varias imágenes de camiones y coches, cada una etiquetada según lo que representan. Estas imágenes etiquetadas se llaman ejemplos (también conocidos como puntos de datos u observaciones) y forman un conjunto de datos. El resultado es un programa llamado modelo que es capaz de identificar nuevos casos y clasificarlos en un tipo de vehículo en este caso.

Es preciso señalar la difucltad de este proceso derivada del contexto que rodea a los casos de prueba. en este caso vemos como según que ejemplos no identifican claramente a un coche o camión debido a factores como el ambiente, luces, orientación, distancia, objetos circundantes etc... Es por ello que se reuieren ingentes cantidades de datos para facilitar y mejorar el modelo final de aprendizaje.

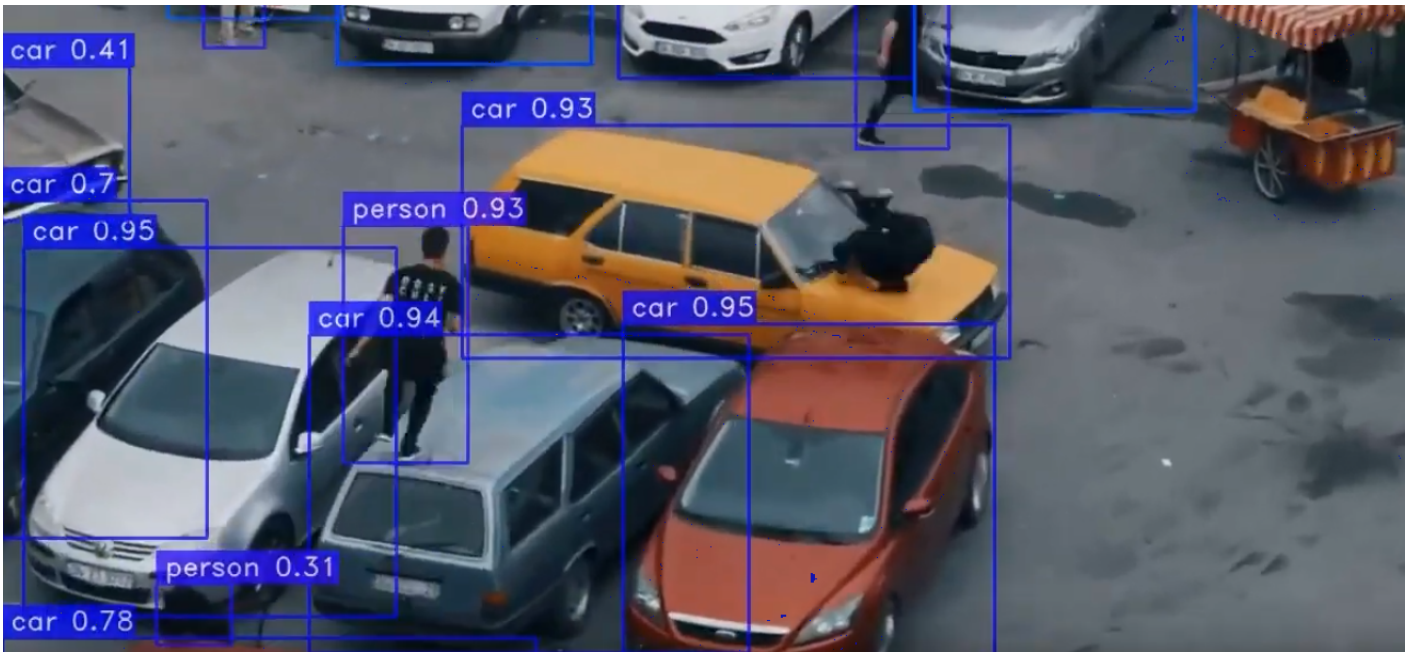
En cierto sentido, el aprendizaje automático es programación a través de ejemplos. La computadora (es decir, la máquina) aprende a realizar una tarea a partir de ejemplos de esa tarea. En este contexto, "aprender" significa ampliamente que se utiliza información de ciertos datos para crear el programa.

La identificación de imágenes no es la única aplicación en la que el aprendizaje automático es útil; se utiliza en la actualidad para realizar una variedad de tareas, desde identificar correos electrónicos no deseados, hasta predecir los precios de las acciones, o jugar videojuegos. Sin embargo, el aprendizaje automático no es un reemplazo directo de la programación tradicional. Para comprender mejor para qué se puede utilizar el aprendizaje automático, repasemos algunos de sus ámbitos de aplicación actuales.

Un ámbito general de aplicación tiene que ver con la imitación de las habilidades humanas. Esto incluye tareas de percepción, como comprender datos visuales y auditivos; tareas intuitivas, como jugar videojuegos; y la tarea muy importante de comprender texto. A continuación, se presentan ejemplos de aplicaciones que se encuadran en este ámbito.

- **Detectar objetos de un vídeo en tiempo real**

Copyright 2026



- **Mejorar la calidad imagen**



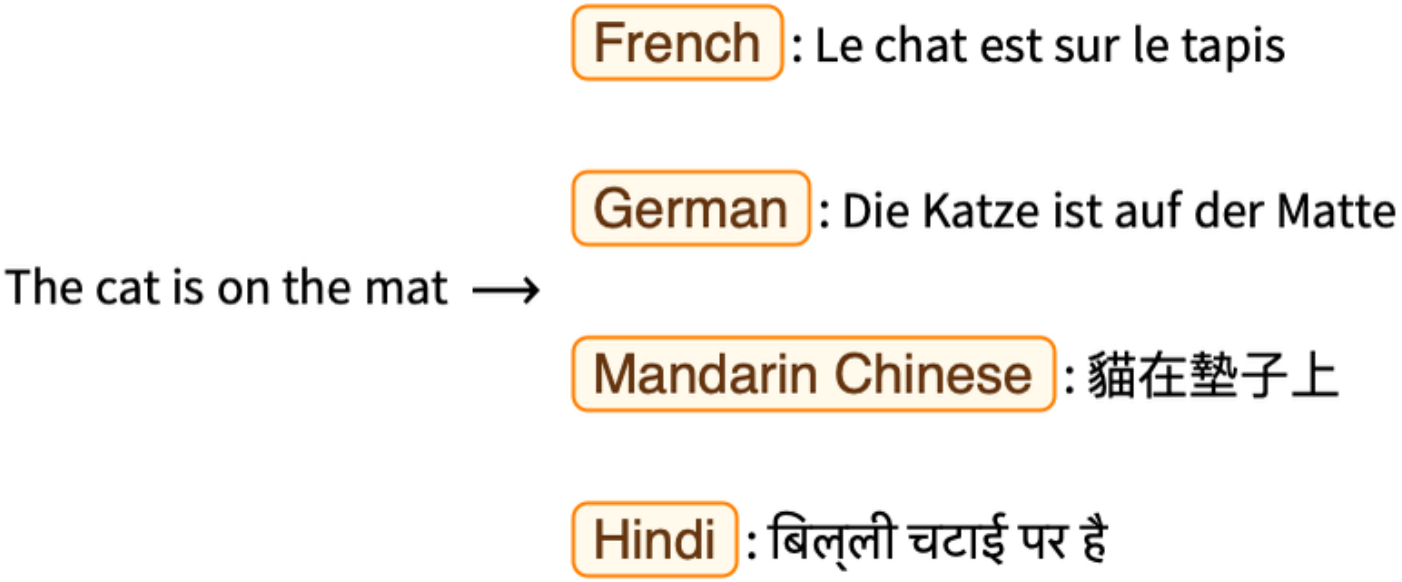
54 × 40

→

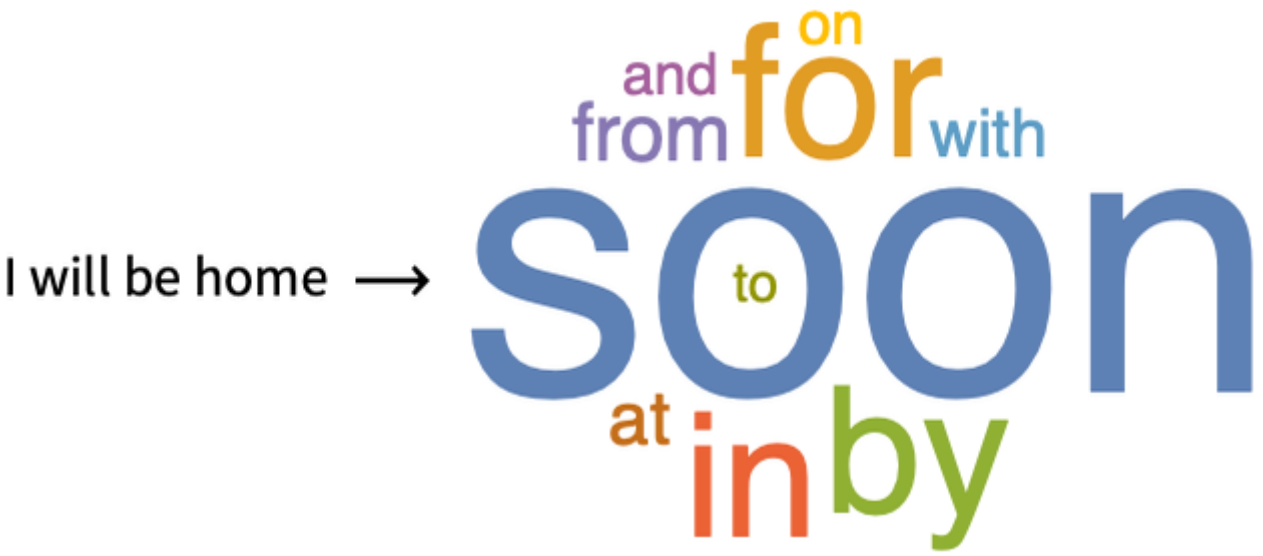


216 × 160

- **Traducción**



- **Predicción (generación) de texto**



Son solo algunos ejemplos, que además se van ampliando con la mejora de tecnologías y algoritmos.

En general estamos hablando de convertir información en múltiples formatos (sonido, vídeo, texto e imagen).

Actualmente podemos incluir tareas como conversión de texto a música, texto a imagen, limpiezas de audios o incluso conversión de pensamiento a imagen, algo que de momento se ha conseguido de forma rudimentaria en laboratorio.

Copyright 2026

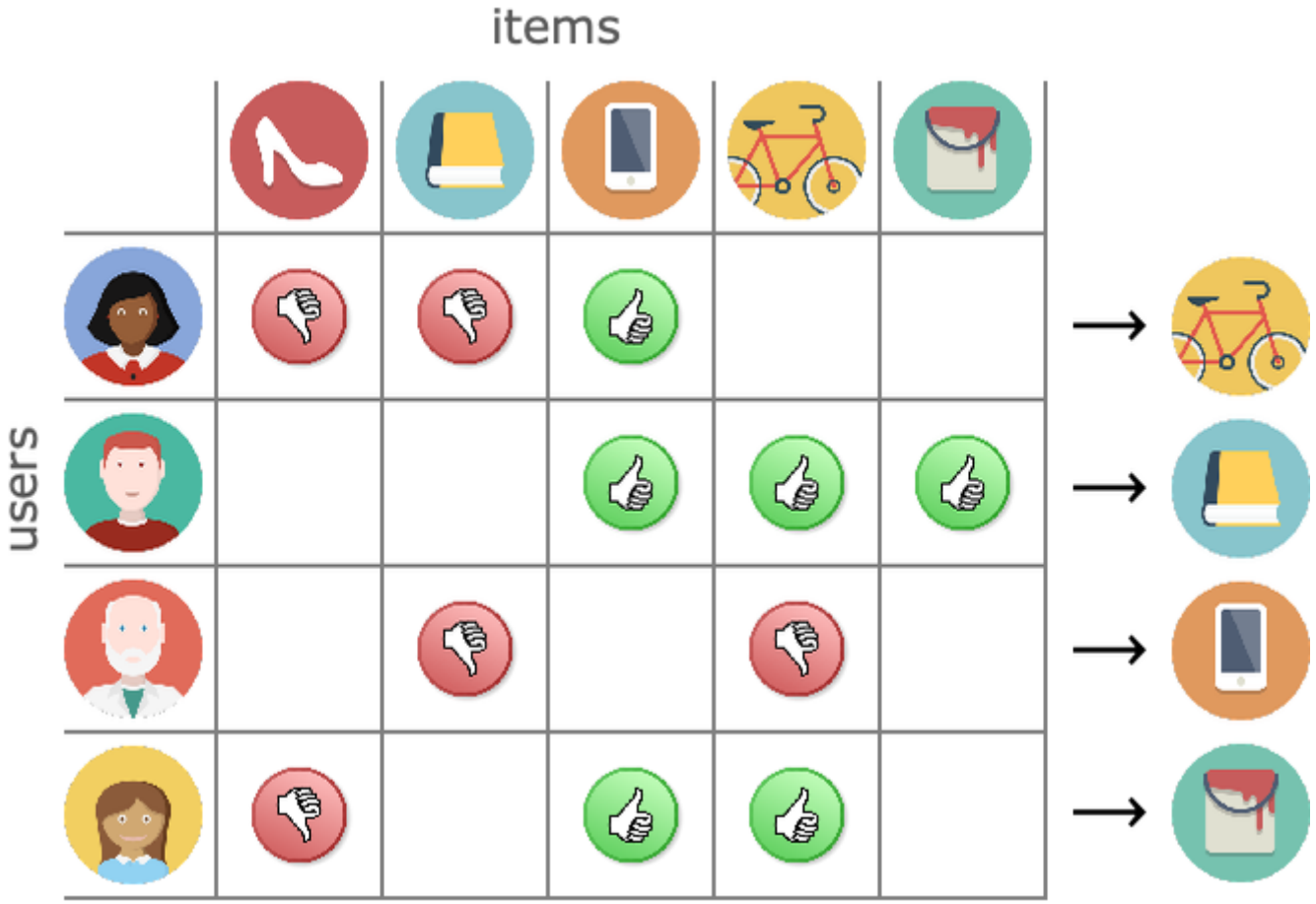
En general han surgido decenas de aplicaciones llamadas multimodales que permiten la conversión/generación de cualquier tipo de contenido, incluyendo vídeos, imagen y voz.

Estas tareas generalmente se consideran parte de la inteligencia artificial (aunque "inteligencia humana" sería más apropiado) y generalmente se abordan utilizando redes neuronales artificiales, un campo conocido como aprendizaje profundo que veremos en la próxima sección. Estas tareas solían ser difíciles o incluso imposibles de resolver en el pasado, pero las cosas han ido cambiando desde la década de 2010 gracias a computadoras más rápidas y al renovado interés en las redes neuronales. En la actualidad, el aprendizaje automático se utiliza ampliamente para resolver estas tareas; por ejemplo, las redes sociales utilizan el aprendizaje automático para analizar grandes cantidades de imágenes y textos con el fin de seleccionar contenido relevante para los usuarios. Cabe destacar que el tipo de datos involucrados en estas tareas (imágenes, audio, texto, etc.) es más complejo y "difuso" que los números organizados en una hoja de cálculo, por lo que se llama datos no estructurados.

Aprendizaje con datos estructurados

Otro ámbito de aplicación importante se refiere a la utilización de grandes cantidades de datos estructurados. Los conjuntos de datos estructurados son los que generalmente se nos vienen a la mente cuando hablamos de datos: números y etiquetas almacenados en hojas de cálculo o bases de datos. Los datos estructurados podrían ser, por ejemplo, datos de ventas recopilados por una empresa minorista: tipo de producto, fechas de venta, precio, etc. La tarea más común al trabajar con datos estructurados es predecir el valor de una variable (también conocida como atributo) de interés, como las futuras cifras de ventas, pero también puede tratarse de comprender los datos, como identificar grupos. A continuación, se presentan ejemplos de tales tareas.

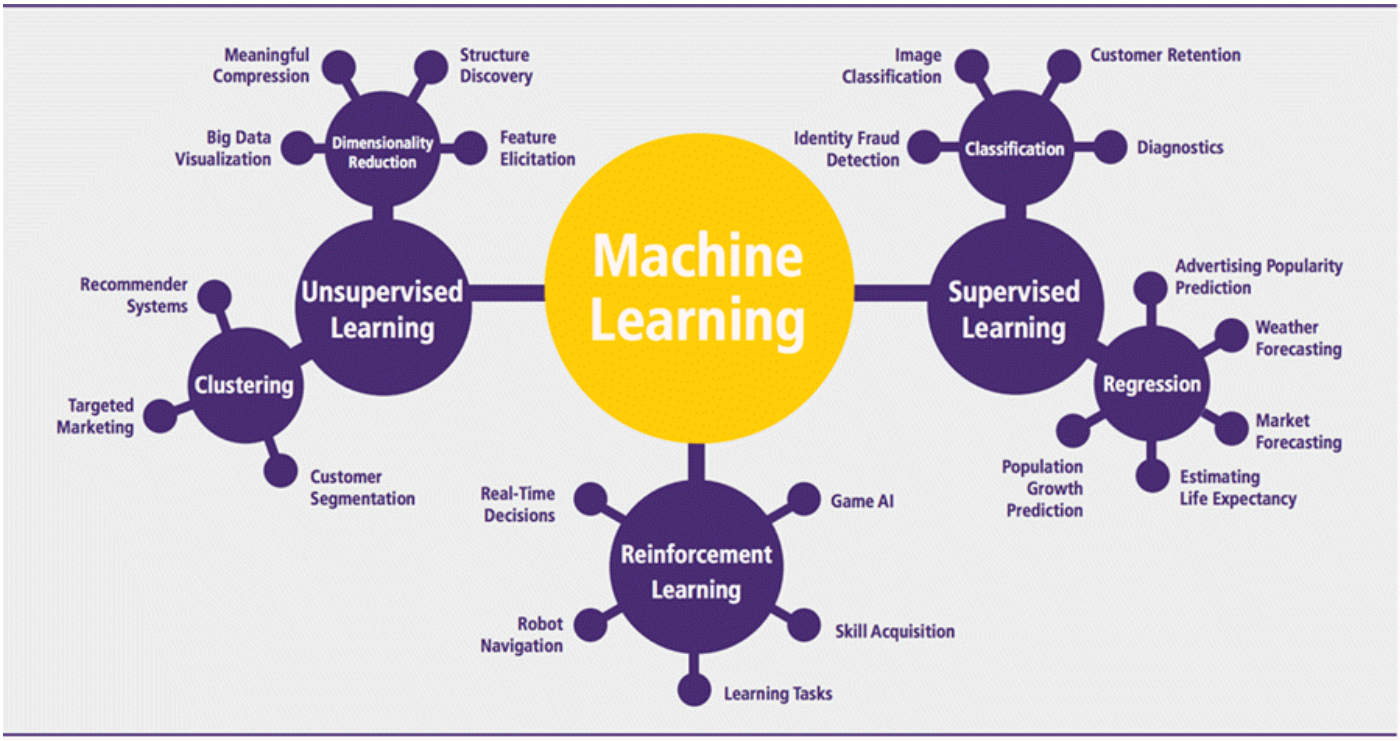
- **Predecir** valores como el precio de una casa o el valor de un stock partiendo de datos históricos
- **Clasificar** los clientes en categorías según sus patrones de comportamiento y su perfil
- **Recomendar** productos en función de los gustos o compras de otros usuarios



- **Detección de Fraude**

Location	Date	Time	Amount		Fraud probabi
California, United States	Thu 26 Nov 2020	14:09:59	$80.00	→	0.1%
California, United States	Fri 27 Nov 2020	15:29:39	$40.00		0.1%
France	Sat 28 Nov 2020	03:04:15	€1400.00		4%

Para finalizar incluimos una imagen que resume perfectamente las aplicaciones de *Machine Learning*



Aclaración El aprendizaje automático a menudo se confunde con la estadística.

Existen muchas similitudes entre estas dos disciplinas, ya que los modelos de aprendizaje automático son modelos estadísticos. En cierto sentido, el aprendizaje automático podría considerarse un subcampo de la estadística. Sin embargo, estas dos áreas difieren en sus objetivos y prácticas. El objetivo del aprendizaje automático generalmente es predecir algo, mientras que el objetivo de la estadística generalmente es comprender algo (por ejemplo, "¿Esta droga ayuda a curar esta enfermedad?"). Como consecuencia, los modelos de aprendizaje automático suelen ser complejos (como conjuntos de árboles, redes neuronales) y "cajas negras", lo que significa que es difícil interpretar lo que hacen. Por otro lado, los modelos en estadística suelen ser simples para poder ser interpretables (como regresión logística, modelos lineales generalizados). Finalmente, el aprendizaje automático a menudo maneja grandes cantidades de datos y de diversos tipos (datos estructurados, imágenes, textos, etc.), mientras que la estadística generalmente trabaja con conjuntos de datos más pequeños y "más simples".

Revision #19
Created 2023-06-23 09:30:42 CEST by Luis Hueso
Updated 2024-03-16 12:11:54 CET by Luis Hueso

Copyright 2026