

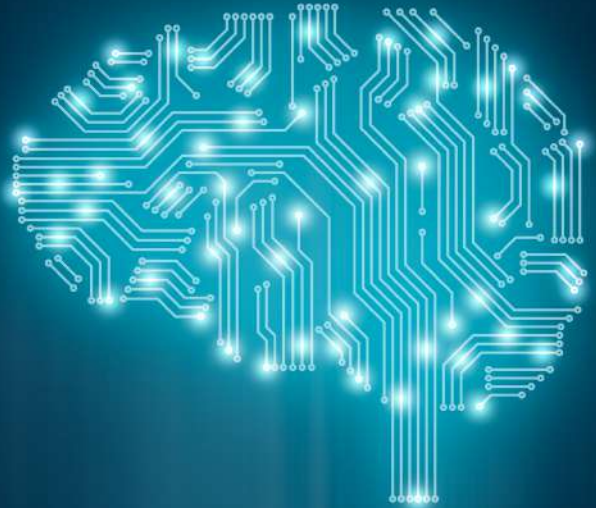
7

# DATA SCIENCE

E-BOOKS



## MACHINE LEARNING CON



ASÍ APRENDEN LAS MÁQUINAS

# Contenidos

MACHINE LEARNING CON 

QUÉ ES EL MACHINE LEARNING

¿QUÉ PROBLEMAS RESUELVE EL MACHINE LEARNING?

APLICACIONES PRÁCTICAS DEL MACHINE LEARNING

CRONOLOGÍA DEL MACHINE LEARNING

¿CÓMO APRENDEN LAS MÁQUINAS?

LOS RETOS DEL MACHINE LEARNING

PROCESO DE RESOLUCIÓN EN MACHINE LEARNING

PANORAMA PROFESIONAL DEL MACHINE LEARNING



MACHINE LEARNING CON R SOFTWARE



## Qué es el Machine Learning

El Machine Learning, conocido en español como **aprendizaje automático** o **aprendizaje de máquina**, estudia y diseña herramientas informáticas que se basan en la experiencia para la toma de decisiones futuras. Fundamenta su operatividad en programas capaces de aprender de los datos, como R Software.

### ¿Subdisciplina o campo de conocimiento autónomo?

Hagamos un poco de historia.

El Machine Learning surge en la década de los 60 del siglo XX como una **subdisciplina de la Inteligencia Artificial (IA)** que tiene su origen en las ciencias de la computación y las neurociencias.

Entendida como subdisciplina, es un método que permite otorgar a los ordenadores la capacidad de reconocer patrones, establecer relaciones y aprender basándose en datos sin que el ser humano tenga que escribir instrucciones o códigos para lograrlo.

Tres décadas después, en los años 90, los avances del Machine Learning comenzaron a aplicarse a diferentes áreas de conocimiento como: el razonamiento probabilístico, la investigación estadística, la recuperación de información y el reconocimiento de patrones. Esta evolución propició que se disgregara de la IA para convertirse en una **disciplina autónoma**, teniendo como principal objetivo el dar solución a problemas prácticos en los distintos ámbitos de aplicación: ingeniería, matemáticas, computación, etc.

*El Machine Learning es un campo de las ciencias computacionales que proporciona a las máquinas la capacidad de aprender sin ser explícitamente programadas.*



# Qué problemas resuelve el Machine Learning

Dependiendo del tipo de datos de los que dispongamos, el Machine Learning nos permitirá resolver los siguientes problemas:

## Aprendizaje supervisado

El algoritmo aprende de unos datos previos en los cuales se da una variable etiquetada que es la variable a predecir. Su objetivo es aprender de estos datos etiquetados para identificar los patrones y reglas que le permitan predecir la etiqueta al enfrentarse a nuevos casos.

Dentro del aprendizaje supervisado nos podemos encontrar con problemas de regresión y problemas de clasificación.

## Aprendizaje no supervisado

En estos problemas **no existen etiquetas con las que el algoritmo aprende**, es decir, no requiere de ningún etiquetado previo de los datos.

El fin de este tipo de aprendizaje es obtener relaciones, diferencias o asociaciones entre las distintas observaciones.

## Aprendizaje semiestructurado

Es una combinación de los dos aprendizajes anteriores. **El algoritmo aprende tanto de datos etiquetados como no etiquetados.**

*"Duda de los datos hasta que los datos no dejen lugar a dudas"*  
Henry Poincaré



## Aplicaciones prácticas del Machine Learning

El Machine Learning supone un enorme avance en el terreno de la computación. Gracias a esta disciplina hoy son posibles **aplicaciones prácticas** como:



### E-COMMERCE

La activación de sistemas de recomendación personalizada en e-commerce.



### COMUNICACIÓN

El filtrado de los correos indeseados, basura o spam.



### TRANSPORTES

La optimización del funcionamiento de las redes ferroviarias.



### MOVILIDAD

La implementación a gran escala del coche autónomo.

Pero ¿qué avances tecnológicos nos han llevado hasta el momento presente en el que creamos máquinas capaces de aprender sin estar programadas para ello?



Solo necesitas mirar a tu alrededor para ver **el poder del Machine Learning en la vida cotidiana.**

Cuando realizamos una búsqueda en Google, cuando Facebook reconoce los rostros en nuestras fotos, cuando nuestro correo filtra correos sospechosos de SPAM, los asistentes virtuales como Siri o Alexa, cuando los navegadores GPS predicen el tráfico, cuando se nos recomienda contenido similar en Youtube, Netflix o Pinterest, gente que conocemos en Twitter o Instagram, cuando nos ofrecen productos de Amazon, cuando Paypal o las aseguradoras realizan análisis de detección de fraude, etc.

Todo esto es Machine Learning.



## Cronología del Machine Learning

Estos son las principales **contribuciones científicas contemporáneas** que hicieron posible el origen y la evolución del Machine Learning:

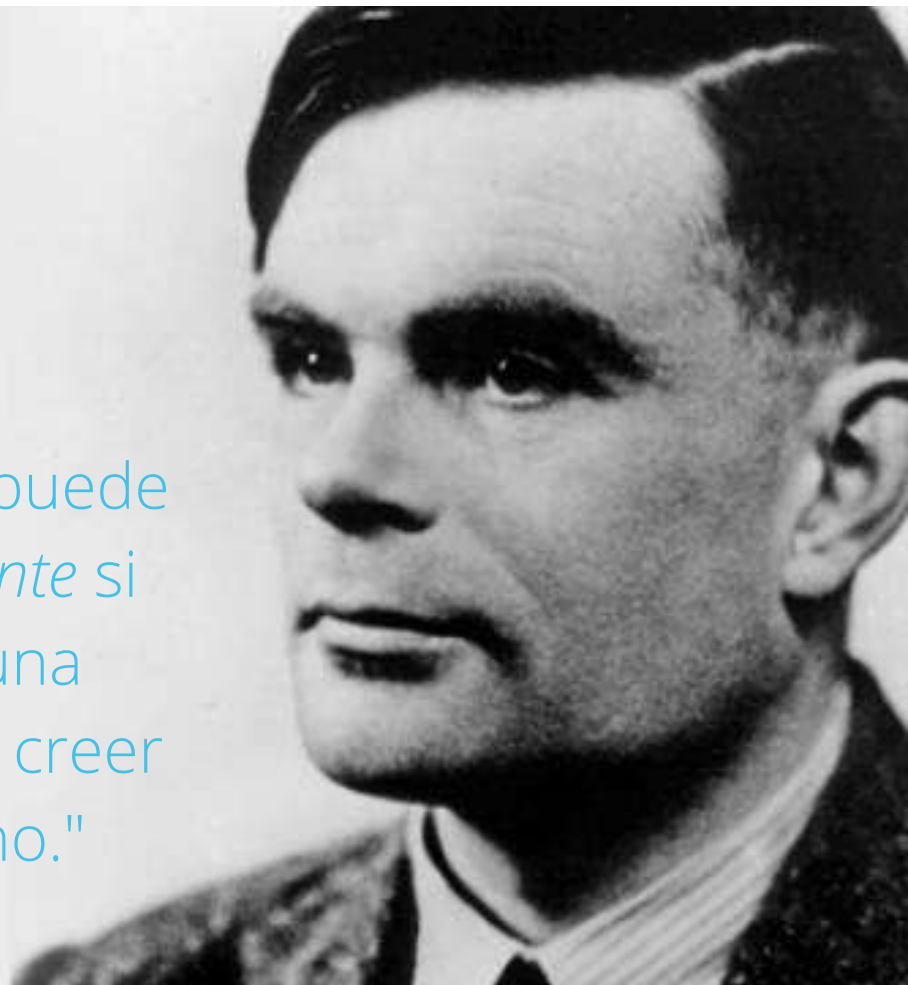
### 1950. Maquinaria computacional e inteligencia

En 1950 el matemático inglés Alan Turing publicó su artículo académico “Computing machinery and intelligence” (“Maquinaria computacional e inteligencia”).

Dentro del primer epígrafe de este texto “**el padre**” de la **Inteligencia Artificial contemporánea** se planteaba la siguiente cuestión “¿Pueden pensar las máquinas?”



"Una computadora puede ser llamada *inteligente* si logra engañar a una persona haciéndole creer que es un humano."

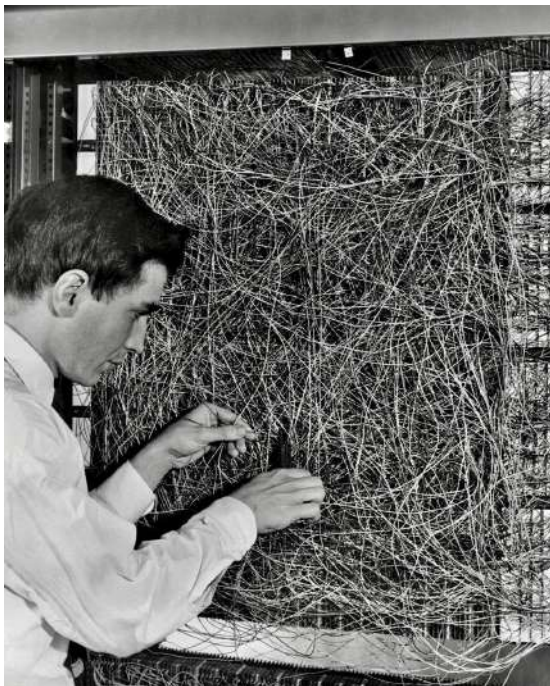


## 1957. El algoritmo de Perceptron

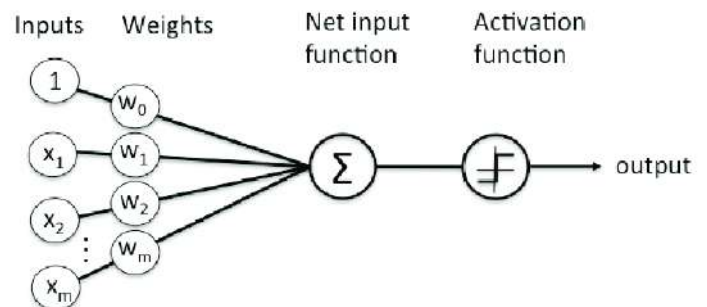
En 1957 el psicólogo norteamericano, Frank Rosenblatt, diseña el algoritmo Perceptron para el aprendizaje supervisado de clasificadores binarios.

Este algoritmo se utilizó para el reconocimiento de imágenes y es el **origen de las redes neuronales**. El esquema del perceptrón funciona como el de una neurona artificial.

Los avances de Rosenblatt en el campo de la Inteligencia Artificial culminaron con la construcción del hardware del súper ordenador Mark I en 1960, el primero capaz de aprender nuevas habilidades basadas en acciones de prueba y error, utilizando un tipo de red neuronal que simulaba el proceso de pensamiento humano.



Frank Rosenblatt trabajando en el proyecto Mark I







## 1990. La era del data driven

En esta década el científico de la computación, Peter J. Denning publicó un artículo en American Scientist en el que planteaba lo siguiente:

"¿Qué máquinas podemos construir para monitorizar un flujo de datos, o filtrar una base de datos grabada y proponernos un resumen estadístico de lo que hay? [...] es posible construir máquinas que puedan reconocer o predecir patrones en flujos de datos sin entender el significado de los patrones."



A partir de los 90 y hasta la actualidad, el progreso del Machine Learning se basará en dos premisas:

- **DATA DRIVEN**

El avance de esta disciplina depende directamente de la **capacidad de análisis de grandes cantidades de datos**.

- **ESPECIALIZACIÓN**

El conocimiento se diversifica y adquiere sentido en su aplicación práctica.





## ¿Cómo aprenden las máquinas?

Según el científico de la computación estadounidense, Tom M. Mitchell:



"Una máquina aprende si es capaz de experimentar y utilizar dicha experiencia para mejorar su rendimiento en experiencias similares en el futuro".

El aprendizaje automático de las máquinas se da a través de tres componentes que se producen de forma sucesiva:

### 1. ENTRADA DE DATOS

Entran en juego la observación, el almacenamiento en la memoria y la recuperación de la información para proporcionar una base fáctica de razonamiento.

### 2. ABSTRACCIÓN

Implica la traducción de datos en representaciones más amplias.

### 3. GENERALIZACIÓN

Supone el uso de datos abstractos para formar una base para la acción.





## Los retos del Machine Learning

El aprendizaje automático surgido en la era del Big Data precisa dar solución a **tres retos fundamentales** que surgen al analizar grandes volúmenes de datos, como :

- **VARIEDAD DE LOS DATOS**

Han surgido nuevos tipos de datos, los datos no estructurados, datos binarios que no tienen estructura interna identificable.

- **ESCALABILIDAD**

Se busca la rapidez en el rendimiento y en el procesamiento de los datos, por lo que se escala en horizontal.

- **VELOCIDAD**

La velocidad de generación de datos precisa cada vez más velocidad de procesamiento.





# Proceso de resolución en Machine Learning

La resolución de problemas de Machine Learning precisa la **aplicación de un método sistemático**, paso a paso.

## Punto de partida

### 1. Definir el objetivo

Es fundamental **concretar la meta** que queremos alcanzar que debe adecuarse a los datos de los que disponemos.



Una vez definido el objetivo hemos de:

### 2. Obtener y comprender los datos

Obtener, describir y unificar los datos en un mismo formato puede ser una de las tareas más costosas del proceso.

### 3. Preparar los datos

En esta fase es importante limpiar los datos y prepararlos para entrenar el modelo.

### 4. Entrenar el Modelo

Que supone elegir el modelo más adecuado en función del problema, y también los parámetros que mejor se ajusten a nuestros datos sin caer en el sobreajuste.

### 5. Resultado del Modelo

Una vez entrenado nuestro modelo debemos probarlo sobre datos nuevos con los que no haya sido entrenado (conjunto de validación) para valorar como ajusta a nuevos datos y medir la precisión de la predicción.

### 6. Implantación

Como fase final instauramos el modelo y establecemos un seguimiento sobre posibles variaciones.

## Panorama profesional del Machine Learning

El auge del Machine Learning crea innumerables oportunidades laborales. Una carrera como experto en Machine Learning abre rápidamente las puertas a algunas de las oportunidades profesionales más emocionantes y de mayor demanda en la actualidad.

Aquí tienes algunos datos:

- Un informe reciente de la consultora Gartner proyecta que se espera que los campos de la Inteligencia Artificial y el Machine Learning creen **2.3 millones de nuevos empleos para 2020**.
- En los últimos años, las ofertas de trabajo relacionadas con la IA **han aumentado en más del 100%** en los principales sitios de carrera como Indeed.
- **Los roles de Data Scientist han aumentado en un 650%** desde 2012, y se prevé que haya 20 millones de puestos para expertos en Machine Learning en el año 2026.









## Machine Learning con R Software

### ¿Por qué es esencial para el Machine Learning?





R es una de las plataformas de aprendizaje automático más potentes y es utilizada por los mejores científicos de datos del mundo.

Si deseas dar el gran salto para convertirte en un experto en Machine Learning, **¡aprender y dominar R es una necesidad absoluta!**

Te damos 8 razones:

-  es **FLEXIBLE** porque cuenta con numerosas extensiones (paquetes o bibliotecas) de código abierto para la manipulación y visualización de datos, el aprendizaje automático y el procesamiento del lenguaje natural, por nombrar solo algunas.
-  es **POTENTE** debido a la variedad de técnicas que ofrece. Proporciona excelentes funciones de visualización, que es esencial para explorar los datos antes de enviarlos a cualquier aprendizaje automatizado, así como evaluar los resultados del algoritmo de aprendizaje.
-  es **GRATUITO** porque es un software de código abierto. En consecuencia, cualquiera se puede usarlo sin comprar una licencia y puede modificarlo para crear nuevas funciones.
-  es **MULTIPLATAFORMA**. Da la capacidad de ejecutarse en muchos sistemas operativos. R está diseñado para ejecutarse en cualquier entorno, sea Mac, Windows o Linux.



-  es **LO ÚLTIMO EN TECNOLOGÍA** porque lo usan los académicos. R es un lenguaje de programación creado para el análisis estadístico. Por eso es de vanguardia en ciencia de datos. R es utilizado por los mejores científicos de datos del mundo. Los académicos que desarrollan nuevos algoritmos los están desarrollando en R y lanzándolos como paquetes de R.
-  está **BIEN RESPALDADO** porque es uno de los principales idiomas para la ciencia de datos. Afortunadamente, R tiene muchos seguidores y se usa mucho en los círculos académicos e industriales, lo que significa que cuentas con numerosos paquetes de análisis útiles y actualizados.
-  también es utilizado en la actualidad por las **EMPRESAS COMERCIALES**. Existe una larga lista de empresas que declaran públicamente su adopción de la plataforma, desde Google, Facebook, Twitter, Microsoft, New York Times, Deloitte, NOAA, etc. Incluso las principales compañías vendedoras de software y servicios de análisis de datos, como Oracle, IBM, Mathematica, MATLAB, SPSS y SAS, proporcionan integración con R y sus plataformas.
-  es **FÁCIL DE APRENDER**. R es desarrollado por estadísticos para estadísticos. No es necesario tener experiencia en informática, ni años de experiencia en codificación, ni un doctorado en estadística para comenzar a aprender R. Cualquiera puede aprender la herramienta más poderosa para el análisis y visualización de datos.

*R Software es un lenguaje estadístico desarrollado por y para estadísticos.*



Inicio  
Marzo  
2020

# Máster de Estadística Aplicada con R Software



Especialízate en análisis de datos avanzado

X edición







# DATA SCIENCE

Máxima Formación e-books



¡Síguenos en redes sociales!

