

Machine Learning Aplicado

Fundamentos de MLA

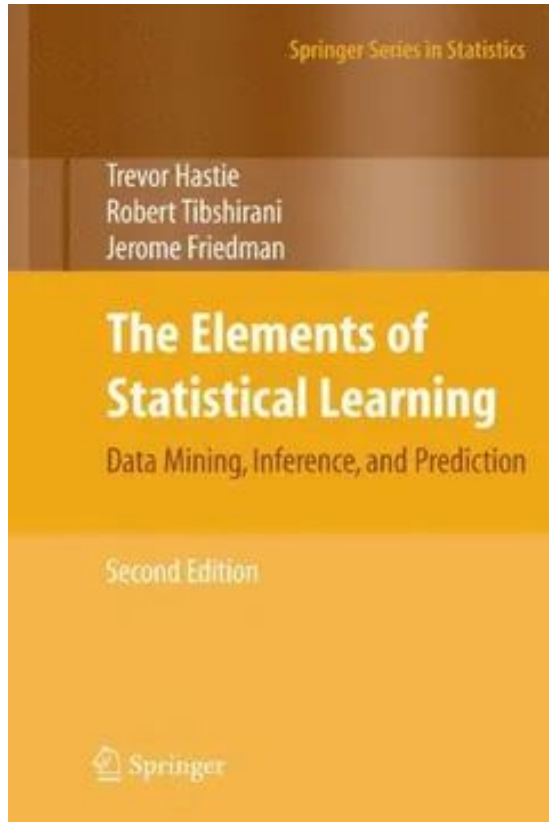
Ingeniería Civil Informática
Escuela de Ingeniería Informática
Universidad de Valparaíso

Profesor: Aaron Ponce Sandoval
correo: aaron.ponce@uv.cl

“El aprendizaje automático es el campo de estudio que brinda a las computadoras la **capacidad de aprender** sin ser programadas explícitamente.”

—Arthur Samuel, 1959





“Se están generando grandes cantidades de datos en muchos campos, y el trabajo de los estadísticos es darles sentido a todos: **extraer patrones y tendencias importantes y comprender** "lo que dicen los datos". A esto lo llamamos **aprender de los datos.**”

- Libro: The Elements of Statistical Learning: Data Mining, Inference, and Prediction

“Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna tarea T y alguna medida de desempeño P , si su desempeño en T , medido por P , mejora con la experiencia E .”

—Tom Mitchell, 1997



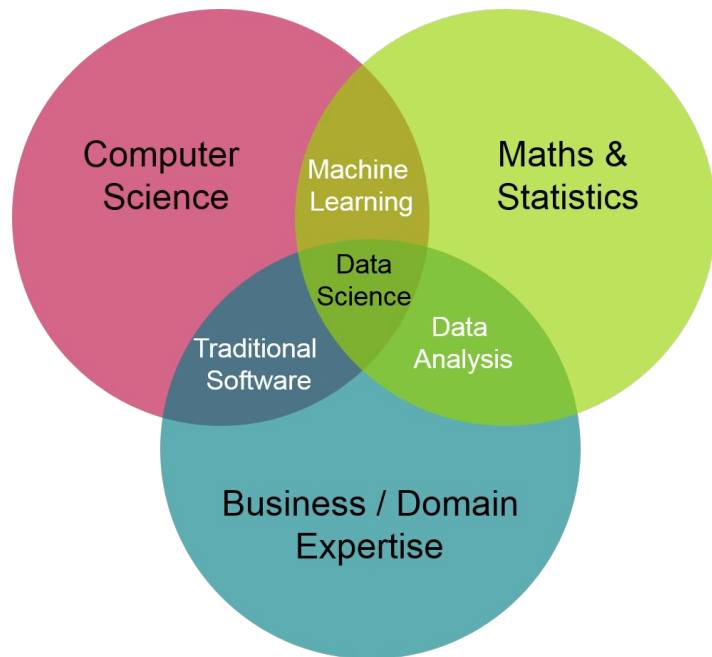
“El aprendizaje automático es un subcampo de la informática en el que las máquinas **aprenden a realizar tareas** para las que **no fueron programadas explícitamente**.

En resumen, las máquinas **observan un patrón e intentan imitarlo** de alguna manera que puede ser una imitación directa o indirecta.”

- Andrew W. Trask, DeepMind

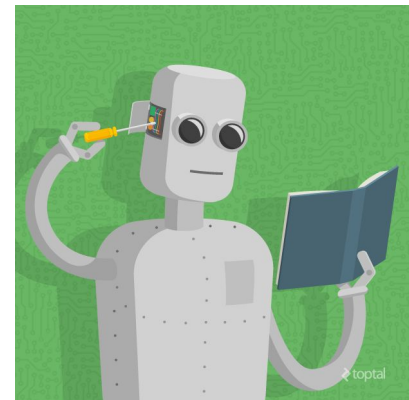


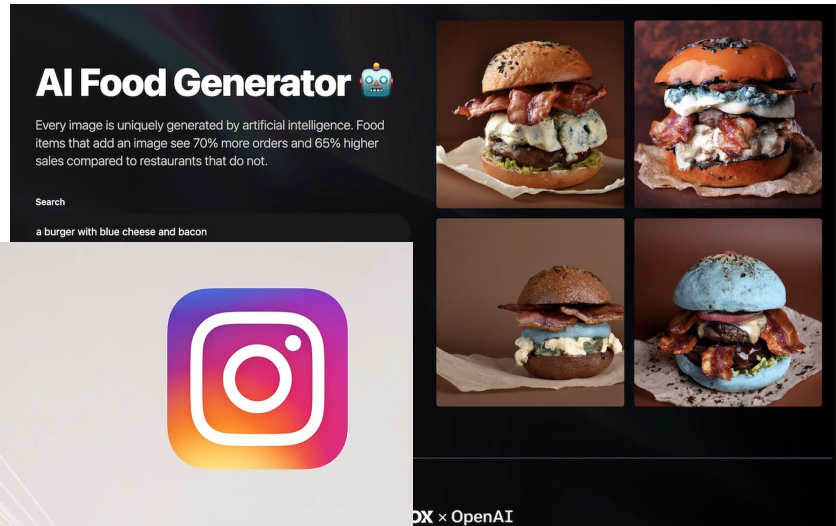
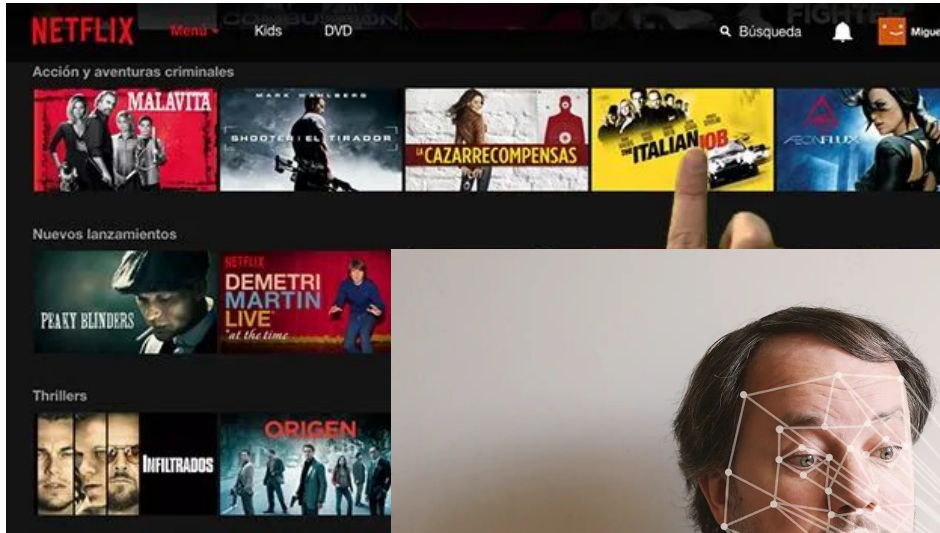
¿Qué es Machine Learning?



El aprendizaje automático trata de extraer conocimiento de los datos. Para crear máquinas con capacidad de aprender por medio de la observación de un patrón.

Es un campo de investigación en la intersección de la estadística, la inteligencia artificial y la informática, que también se conoce como análisis predictivo o aprendizaje estadístico.





Four Industrial Revolutions



1st Industrial Revolution
(Late 18th – early 19th Century)

Mechanization and the
introduction of steam and
water power



2nd Industrial Revolution
(Late 19th – mid 20th Century)

Emergence of new sources of
energy (electricity, oil, and
gas) and mass production.



3rd Industrial Revolution
(Second half of 20th Century)

Rise of electronics,
telecommunications and
computers

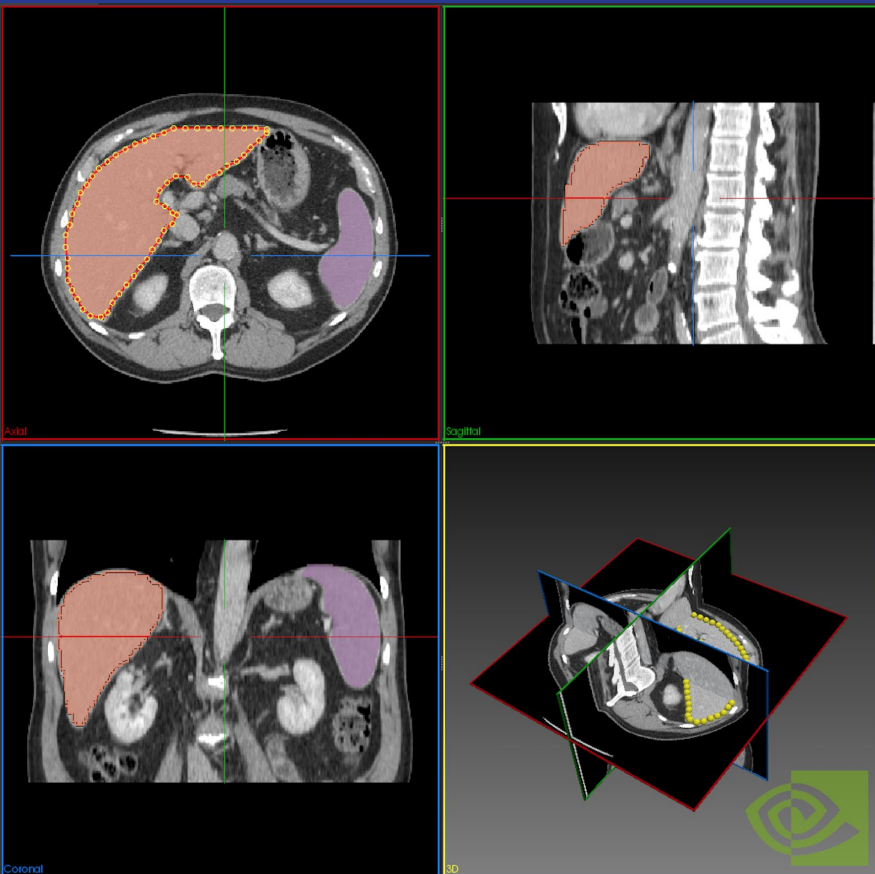


4th Industrial Revolution
(Early 21st Century)

Smart, inter-connected
cyber-physical systems
(IoT, AI, Big Data, ...)

Finanzas: El machine learning se ha utilizado para mejorar la precisión en la toma de decisiones de inversión, reducir el riesgo crediticio, prevenir el fraude y mejorar la personalización de los productos financieros.

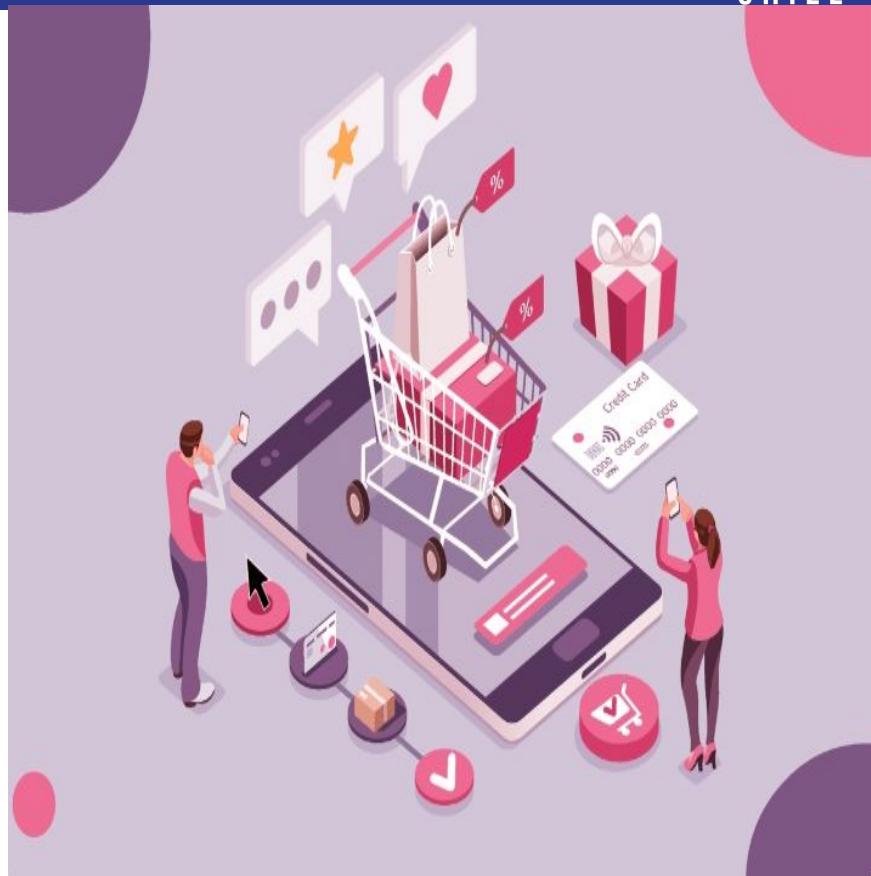




Salud: En el sector de la salud, el machine learning se ha utilizado para mejorar la precisión en el diagnóstico y tratamiento de enfermedades, para analizar grandes cantidades de datos de salud y para mejorar la eficiencia en la gestión de registros médicos.

Sector Retail

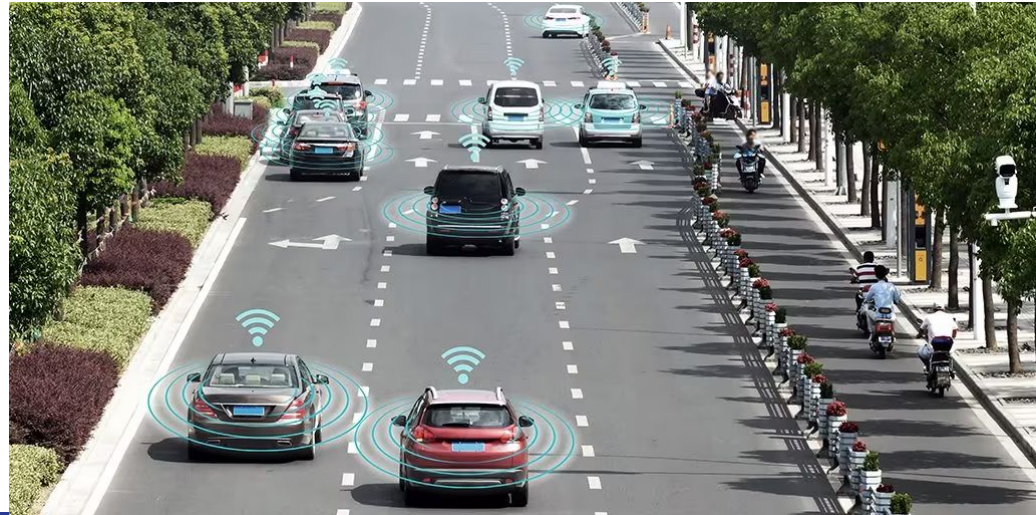
Retail: En el sector minorista, el machine learning se ha utilizado para personalizar la experiencia del cliente, predecir la demanda y optimizar la cadena de suministro.





Manufactura: En el sector manufacturero, el machine learning se ha utilizado para mejorar la eficiencia en la producción, predecir la calidad de los productos y mejorar la planificación de la producción.

Automoción: En el sector automotriz, el machine learning se ha utilizado para mejorar la seguridad de los vehículos, para predecir y prevenir fallos en los componentes, y para mejorar la eficiencia en la producción.



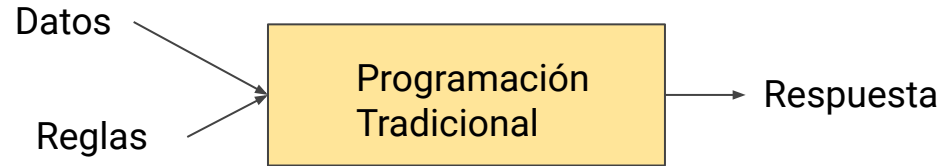
Sector Agrícola

Agricultura: En el sector agrícola, el machine learning se ha utilizado para mejorar la eficiencia en la producción, para predecir y prevenir plagas y enfermedades en las plantas, y para mejorar la gestión del riego.

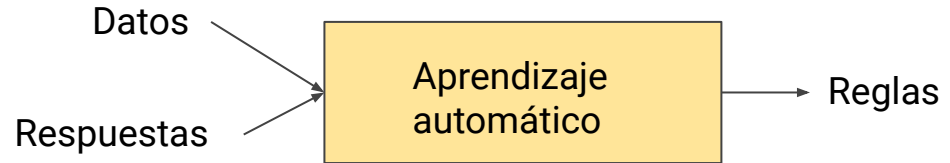


Enfoque de Solución de Problemas

Enfoque Tradicional



Enfoque de Aprendizaje Automático



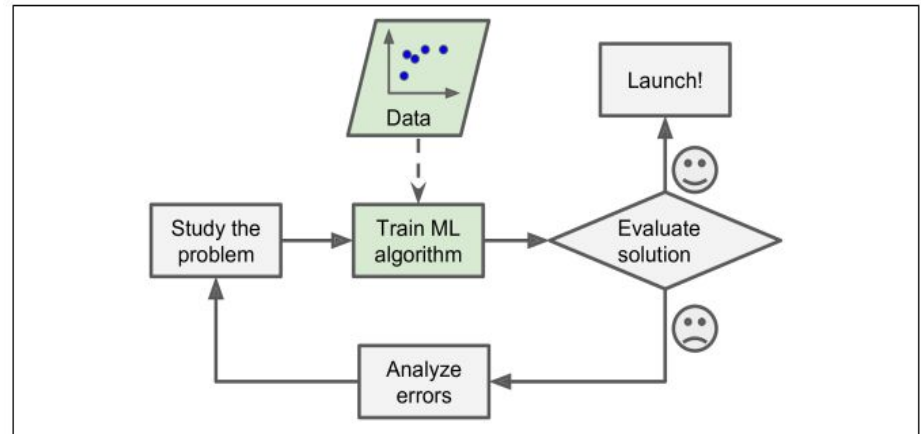
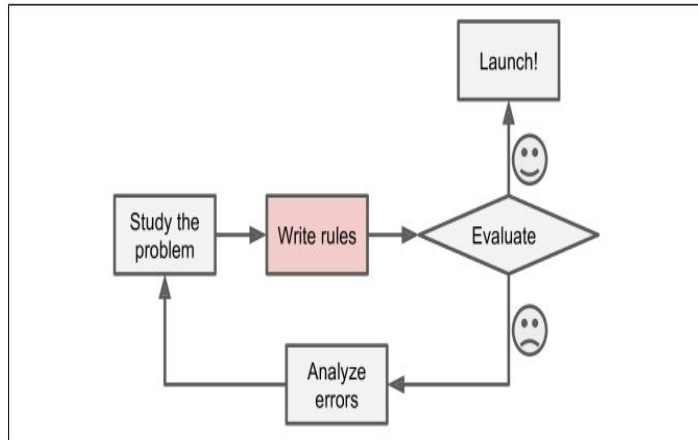
Ejemplo: Problema

La empresa Skynet desea filtrar correos no deseados (spam) de los correos deseados relacionados al departamento de finanzas. Debido a que han aumentado las estafas crediticias.



Enfoque de Solución (analítica)

1. Podríamos notar que algunas palabras en los correos como ("tarjeta de crédito", "gratis" y "increíble", "ganar", "imperdible") tienden a aparecer mucho en el tema. Tal vez también notaremos algunos otros patrones en el nombre del remitente, el cuerpo del correo electrónico, etc.
2. Escribiríamos un algoritmo de detección para cada uno de los patrones que notó, y su programa marcaría los correos electrónicos como spam si se detectan varios de estos patrones.



Definición de un problema de Aprendizaje

Función Objetivo $f: X \rightarrow Y$

Input space: X (conjunto de todos posibles input x)

Output Space: Y (conjunto de todos los output (si o no))

Conjunto de datos D : muestras de input-output (x_1, y_1)

Algoritmo de Aprendizaje: $g: X \rightarrow Y$ (se aproxima a f)

1. Diagnóstico médico: Entra un paciente con un historial médico y algunos síntomas, y se quiere identificar el problema.
2. Reconocimiento de dígitos escritos a mano (por ejemplo, reconocimiento de código postal para la clasificación del correo).
3. Determinar si un correo electrónico es spam o no.
4. Identificar un rostro en una zona de seguridad prohibida.