

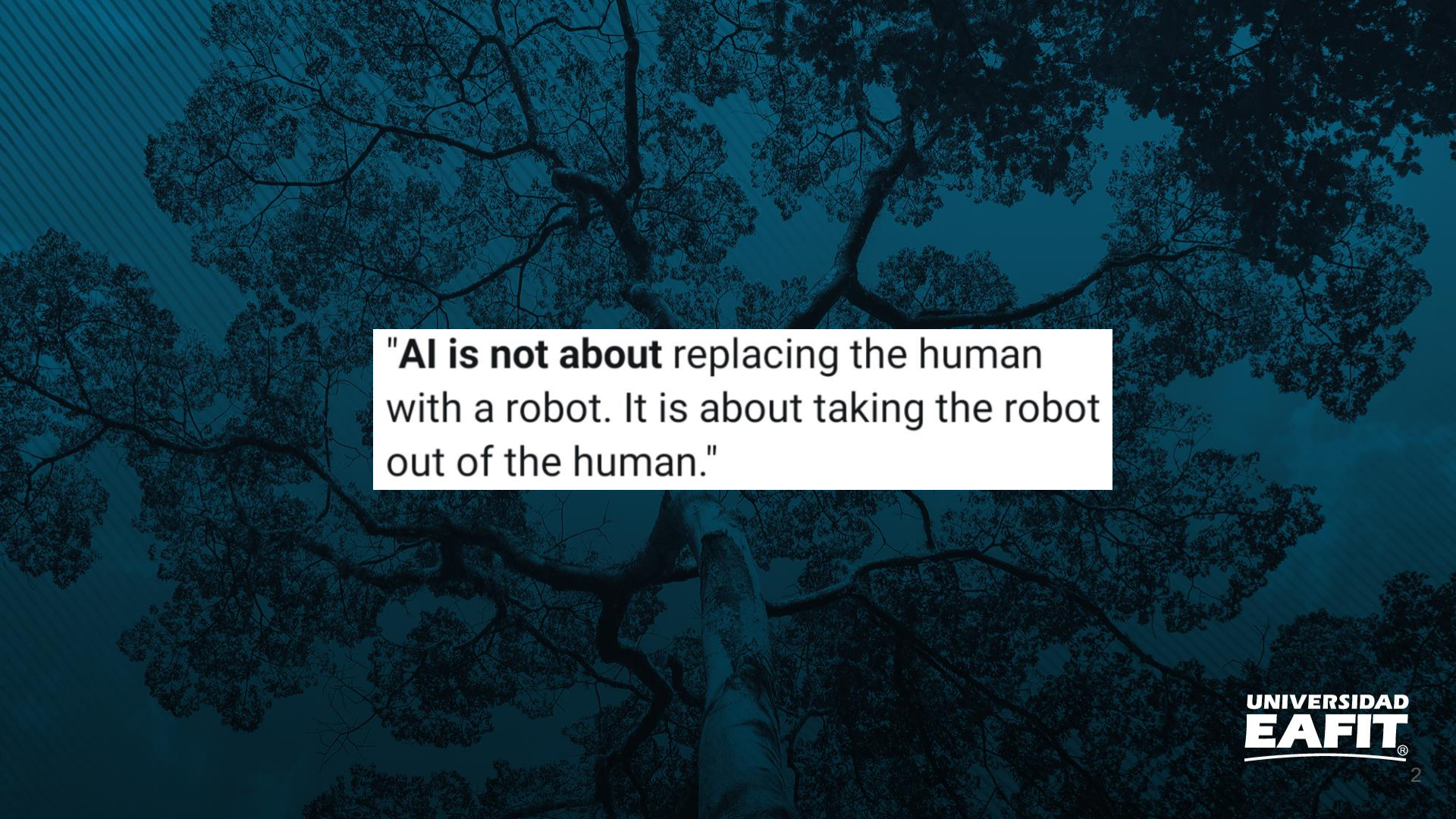
Introducción al aprendizaje de máquina

Aprendizaje de Máquina Aplicado

Juan David Martínez
jdmartinev@eafit.edu.co

César Leandro Higueta
clhiguitap@eafit.edu.co

2023

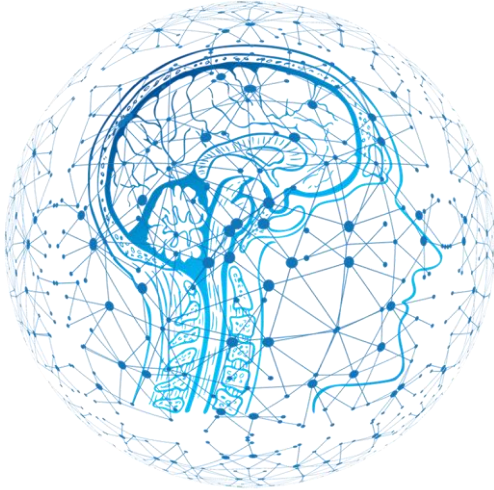


"AI is not about replacing the human with a robot. It is about taking the robot out of the human."

Agenda

- Introducción a inteligencia artificial y aprendizaje de máquina
- Tipos de aprendizaje de máquina
- Diferencia entre aprendizaje de máquina y programación tradicional
- Ciclo de vida de un proyecto de ML
- Ejemplo práctico: Churn prediction with k-nn

¿Qué es inteligencia artificial?



“The theory and development of computer systems able to perform tasks that normally require human intelligence” - Merriam Webster

Understanding language, reasoning, speech recognition, decision-making, navigating the visual world, manipulating physical objects, etc.

El objetivo de la IA es lograr que los computadores hagan cosas que requieren inteligencia humana.



Cognitive Computing



Computer Vision



Machine Learning



Neural Networks



Deep Learning



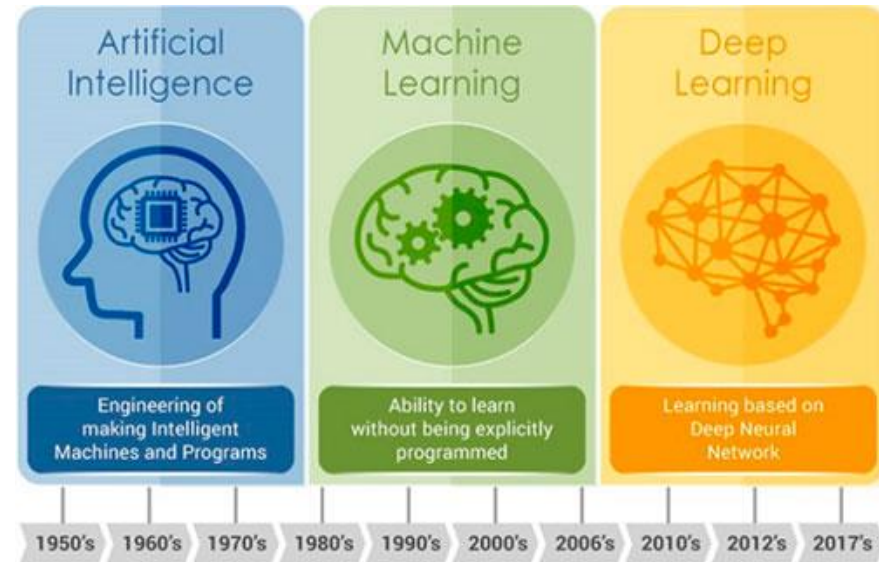
Natural Language Processing

¿Qué es aprendizaje de máquina?

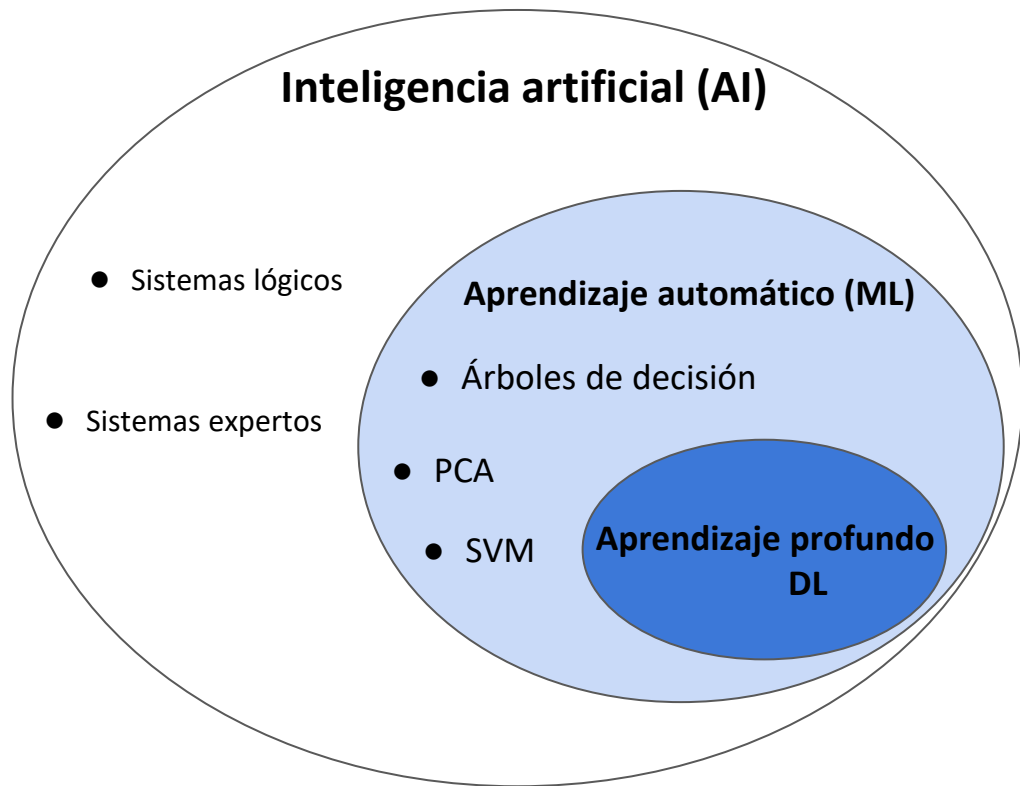
"Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy"

Los algoritmos de ML se caracterizan por su habilidad para **aprender de los datos** sin estar explícitamente programados.

¡ML se utiliza comúnmente para hacer predicciones!



AI vs ML vs DL



Inteligencia artificial: estudia los comportamientos inteligentes en dispositivos.

Machine learning: estudia los algoritmos que mejoran su rendimiento incorporando nuevos datos a un modelo existente.

Deep learning: usa algoritmos de redes neuronales artificiales para extraer características cada vez más complejas a partir de los datos de entrada, buscando mejorar su rendimiento

Ejemplos de IA, ML, DL

Voice assistants



Google news

Armed man who broke into Trudeau residence charged with threatening to kill or injure PM

The Guardian · 1 hour ago

- Corey Hurren, alleged Rideau Hall intruder, threatened Trudeau: RCMP officer
Global News · 4 hours ago
- Corey Hurren had multiple firearms, uttered threat against Trudeau, court documents allege
CBC.ca · 2 hours ago
- Man arrested near Rideau Hall had several weapons, threatened PM Trudeau: RCMP
CTV News · 22 minutes ago



Recommendation systems



Face recognition



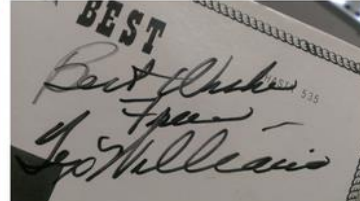
Auto-completion



Stock market prediction



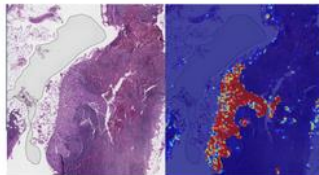
Character recognition



Self-driving car



Cancer diagnosis



Drug discovery



AlphaGo



[Voice assistants](#)

[Google news](#)

[Recommendation systems](#)

[Face recognition](#)

[Auto-completion](#)

[Stock market prediction](#)

[Character recognition](#)

[Self-driving car](#)

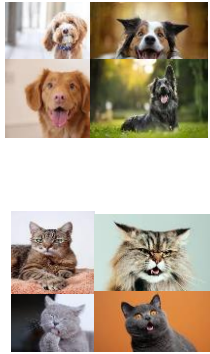
[Cancer diagnosis](#)

[Drug Discovery](#)

[AlphaGo](#)

Tipos de ML: Aprendizaje supervisado

Training



Output: Model that maps images to a labels

Deployment



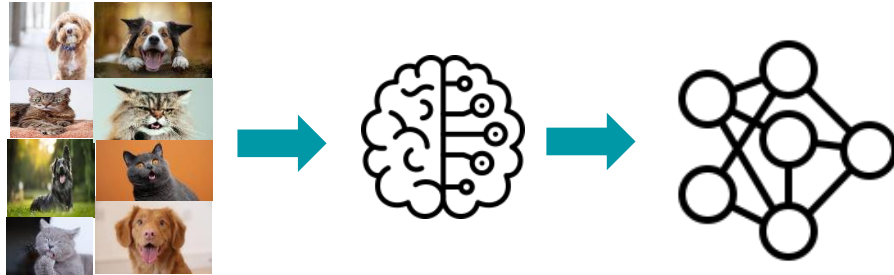
$$P(dog) = 0,99$$

$$P(cat) = 0,01$$

Output: Probability for each label

Tipos de ML: Aprendizaje no supervisado

Training



Input: Images (without labels)

Output: Model that finds similar groups (clusters)



Deployment

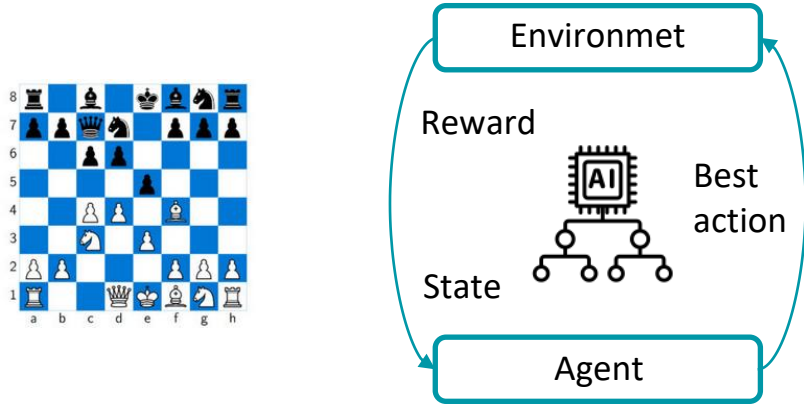


Input: Image

Output: group

Tipos de ML: Aprendizaje por refuerzo

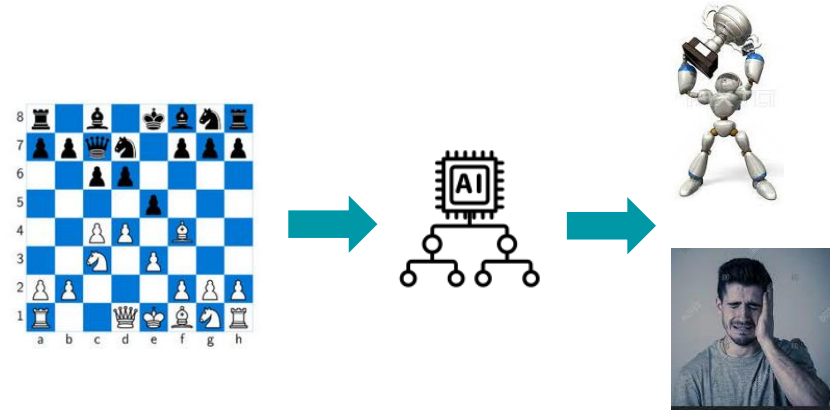
Training



Input: Images (without labels)

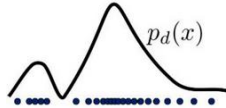
Output: best action

Deployment

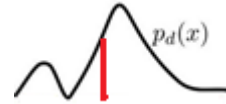


Tipos de ML: Modelos generativos

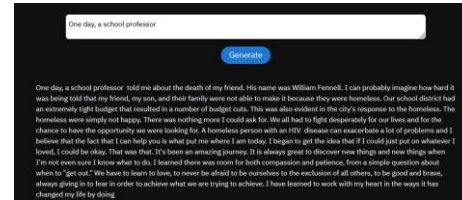
Training



Deployment



<https://thispersondoesnotexist.com/>



<https://hyperwriteai.com>

ML vs programación tradicional

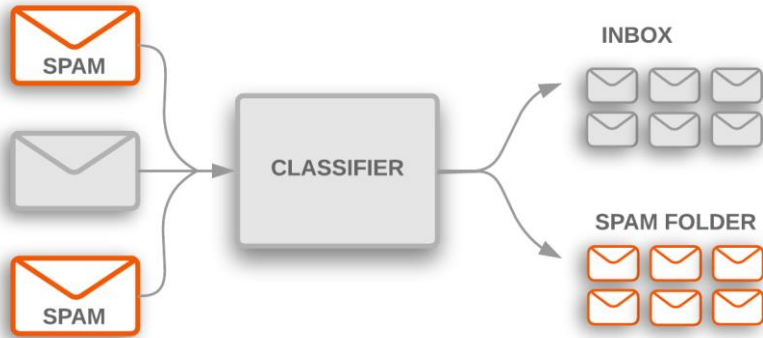


Se definen las reglas y el programador las implementa.



El algoritmo aprende las reglas en lugar de ser implementadas por un programador.

ML vs programación tradicional



Subject: Get 50% off now
From: promotions@online.com

When
spam
you
click
Pay
with

Subject: URGENT: tax review
From: tax@online.com

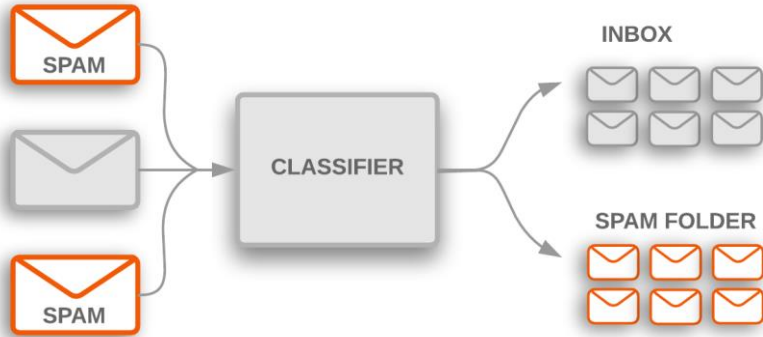
Your tax review is pending acceptance. Review within 24 hours:
<https://taxes.we-are-legit.com>

Tax office.

Reglas:

- If sender = promotions@online.com -> spam
- If title contains “tax review” and sender domain is “online.com” -> spam
- Otherwise -> Good email

ML vs programación tradicional



Subject: Waiting for your reply
From: prince1@test.com

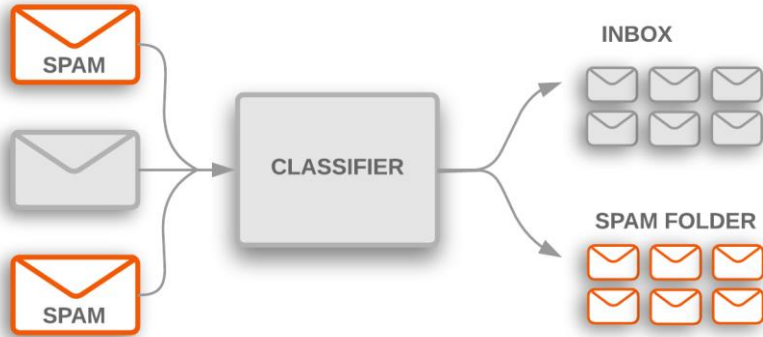
We are delighted to inform you that you won 1.000.000 (one million) US Dollars. To claim the prize, you need to pay a small processing fee. Please deposit \$10 to our PayPal account at prince@test.com. Once we receive the money, we will start the transfer.

Congratulations again!

Reglas:

- If sender = promotions@online.com -> spam
- If title contains “tax review” and sender domain is “online.com” -> spam
- **If body contains a word “deposit” -> spam**
- Otherwise -> Good email

ML vs programación tradicional



Subject: Totally legit email
From: pedro@gmail.com

I transferred \$50 to you one year ago, and now I'm moving out.
Please refund my deposit.

Pedro.

Reglas:

- If sender = promotions@online.com -> spam
- If title contains "tax review" and sender domain is "online.com" -> spam

If body contains a word "deposit":

If se

nder domain is "test.com" -> spam

- **If body >= 100 words -> spam**
- **Otherwise -> Good email**

ML vs programación tradicional



Machine Learning approach

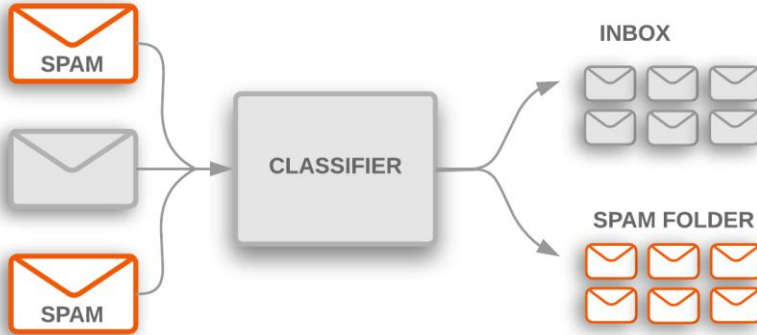
- Ejemplos
- Etiquetas

Características

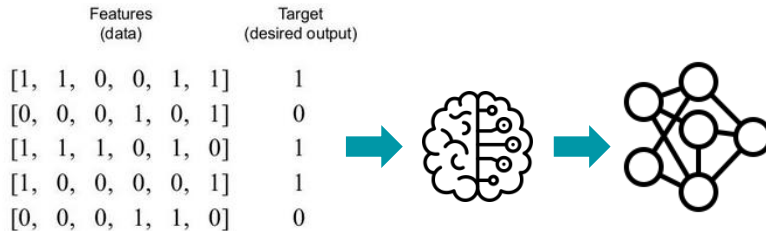
- Length of title > 10? true/false
- Length of body > 10? true/false
- Sender promotions@online.com? true/false
- Sender hpYOSKml@test.com? true/false
- Description contains "deposit"? True/false

[1, 1, 0, 0, 1, 1]

ML vs programación tradicional

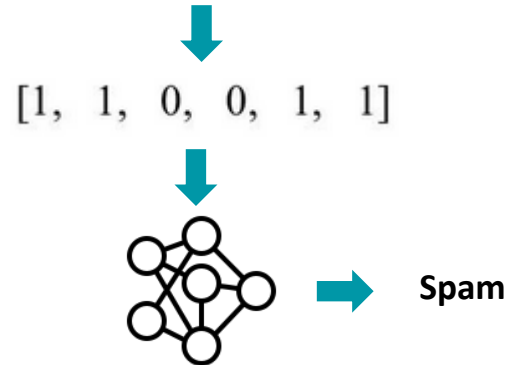


Entrenamiento



Despliegue

Subject: URGENT: tax review
From: tax@online.com
Your tax review is pending acceptance. Review within 24 hours:
<https://taxes.we-are-legit.com>
Tax office.

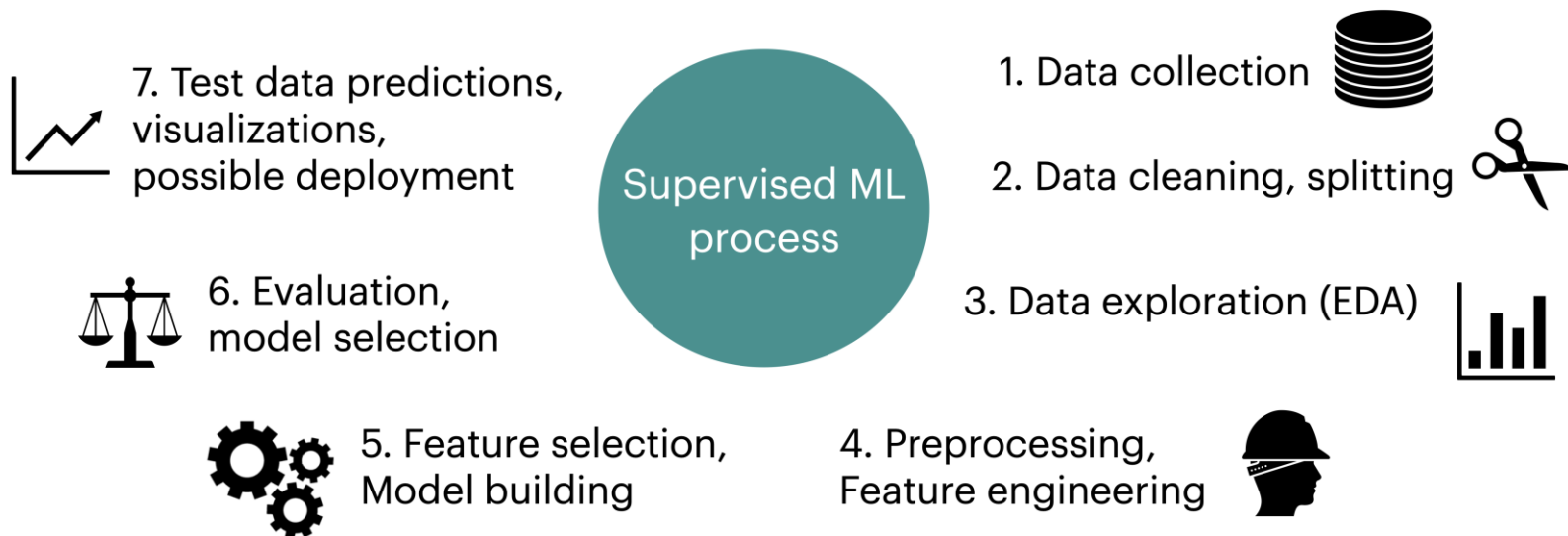


Pasos para entrenar un modelo de ML

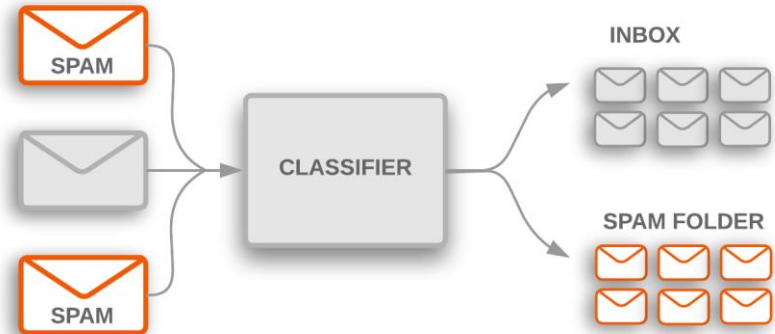
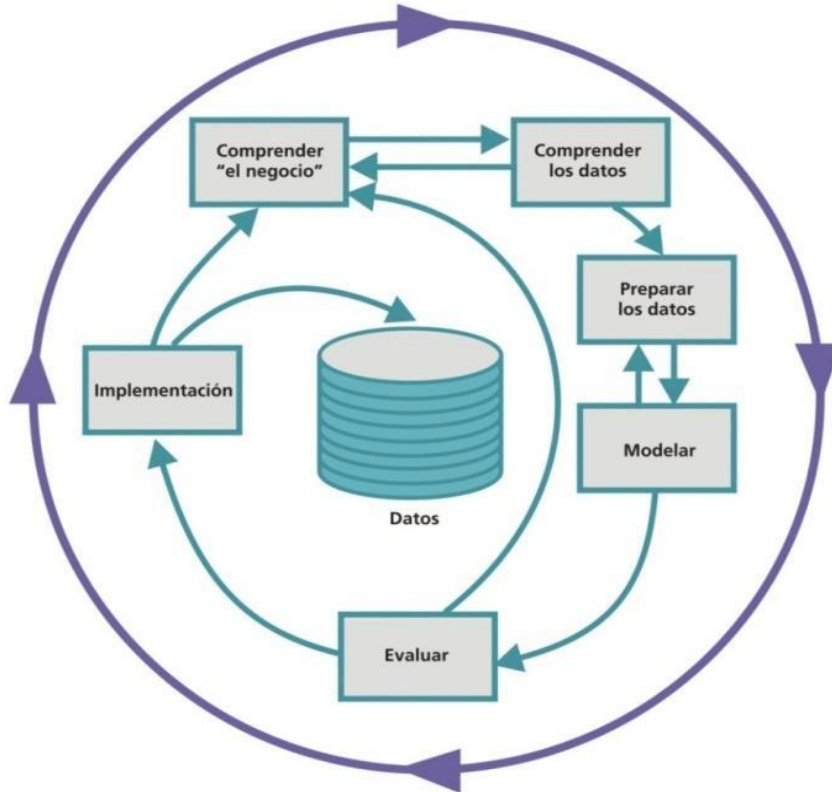
What question do I want to answer?



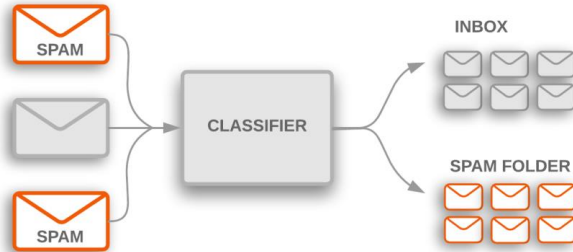
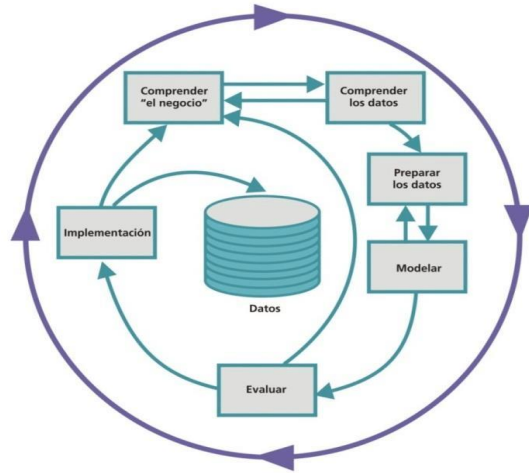
Formulation to supervised machine learning problem



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



CRISP-DM – Comprender el negocio



Problema:

- Los clientes se quejan del spam
- Analizar qué tan grave es el problema
- ¿Hay una solución simple?
- ¿Podemos solucionarlo con ML?

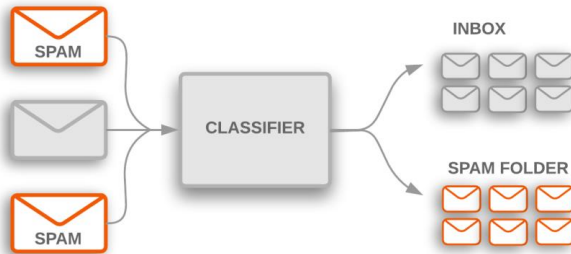
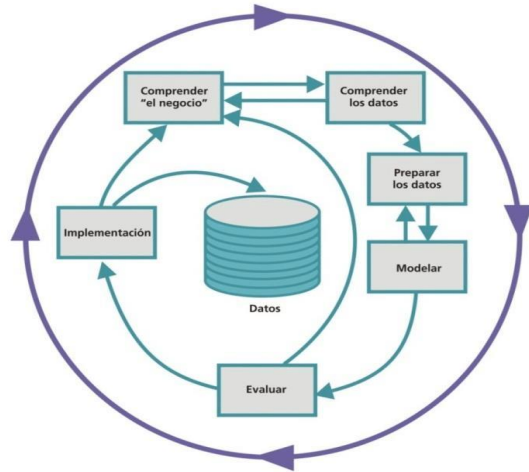
Meta:

- Reducir la cantidad de mensajes spam
- Reducir la cantidad de quejas recibidas por los usuarios

Meta medible:

- Reducir la cantidad de spam en un 50%

CRISP-DM – Comprender los datos



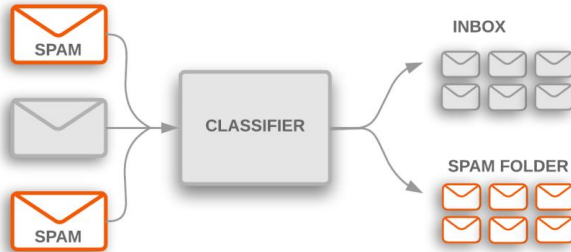
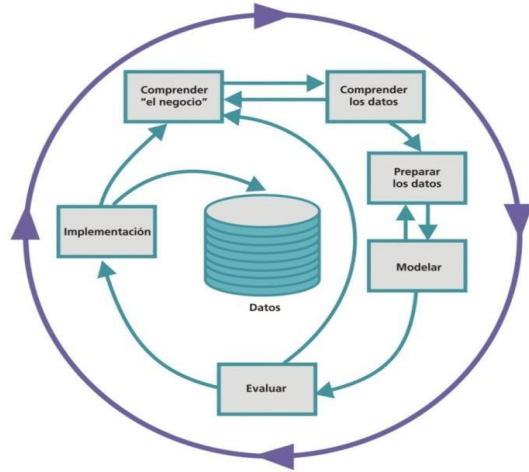
Identificar las fuentes de datos

- Botón de spam
- ¿Los datos son lo suficientemente buenos?
- ¿Los datos son confiables?
- ¿Los almacenamos de forma correcta?
- ¿La base de datos es lo suficientemente grande?
- ¿Necesitamos recolectar más datos?

Preparar los datos

- Transformar los datos para que se puedan entregar a un algoritmo de ML
- Limpieza, imputación de datos faltantes, transformación de variables
- Feature engineering

CRISP-DM – Modelar



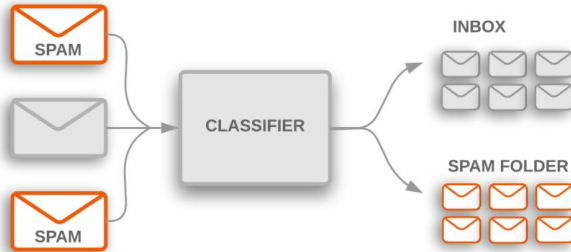
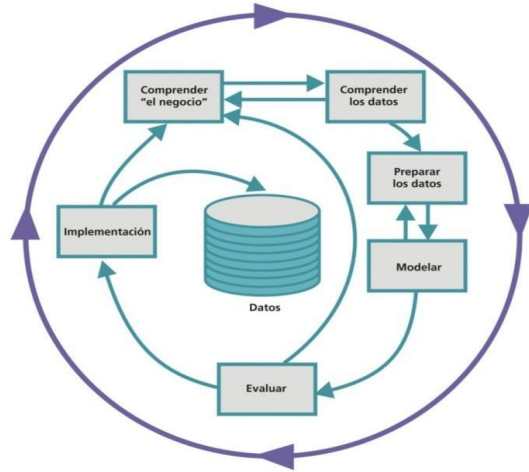
Entrenamiento del modelo

- Intentar diferentes modelos que sean adecuados para los datos con los que está trabajando
- Seleccionar el mejor
- Si el modelo no es lo suficientemente bueno, devolverse al paso anterior

Preparar los datos

- Incluir más datos y/o más características

CRISP-DM – Evaluar



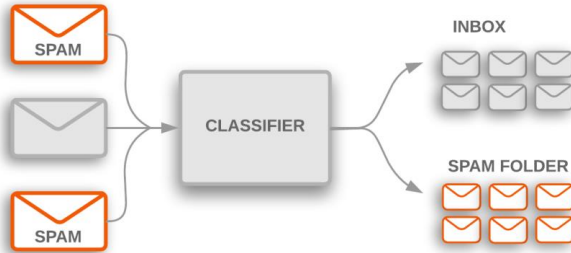
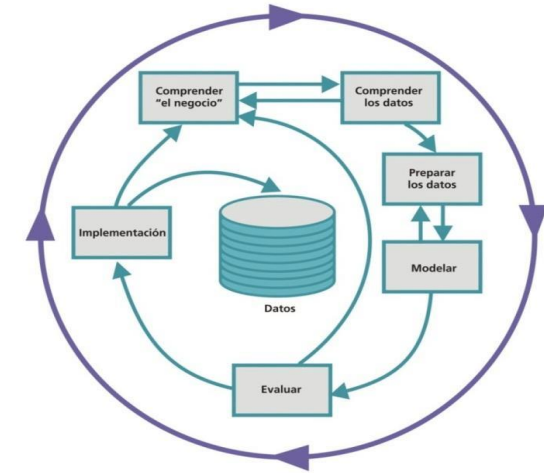
Evaluación del modelo

- ¿El modelo es lo suficientemente bueno?
- ¿Alcanzamos la meta propuesta?

Análisis retrospectivo

- Re-evaluar la meta: reducir la cantidad de spam en un 30%
- Dejar de trabajar en el proyecto ☹️

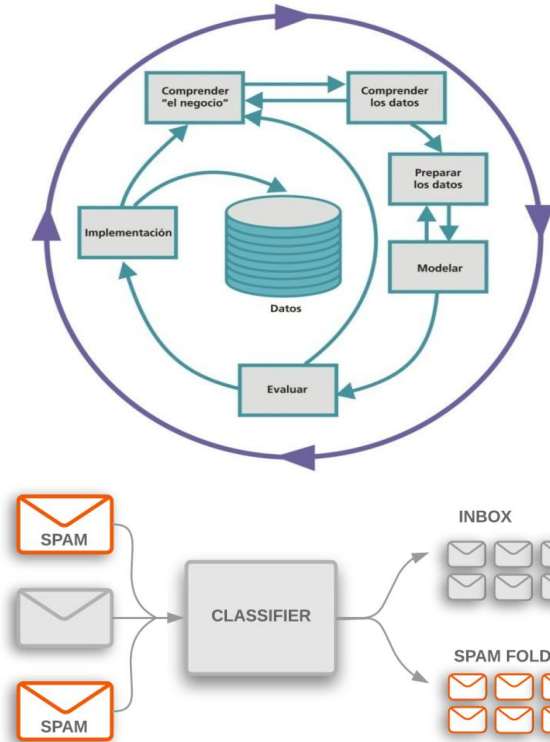
CRISP-DM – Desplegar



Despliegue del modelo

- Poner el modelo disponible para todos los usuarios
- Monitorear el rendimiento del modelo en el tiempo (MLOps)
- Asegurar la calidad u mantenibilidad del proyecto
- Re-entrenar cuando sea necesario

CRISP-DM – Iterar



Iterar

- ¡Los proyectos con ML requieren varias iteraciones!

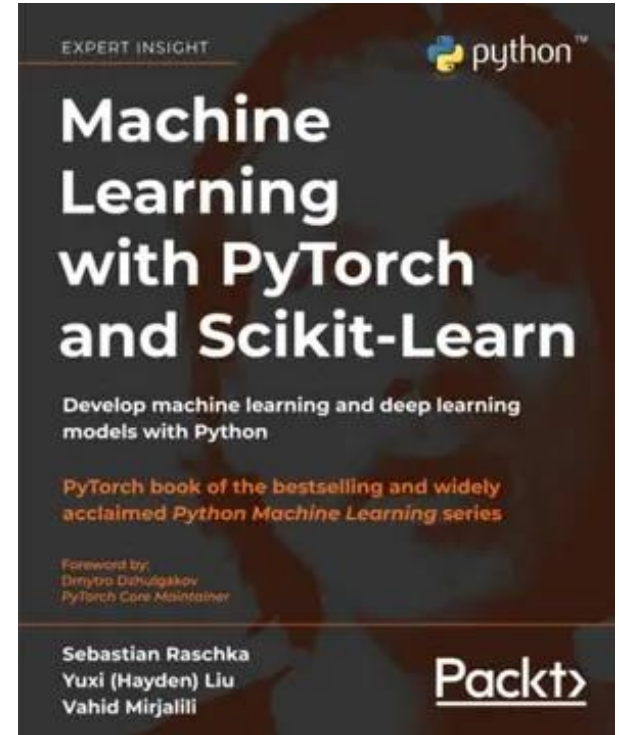
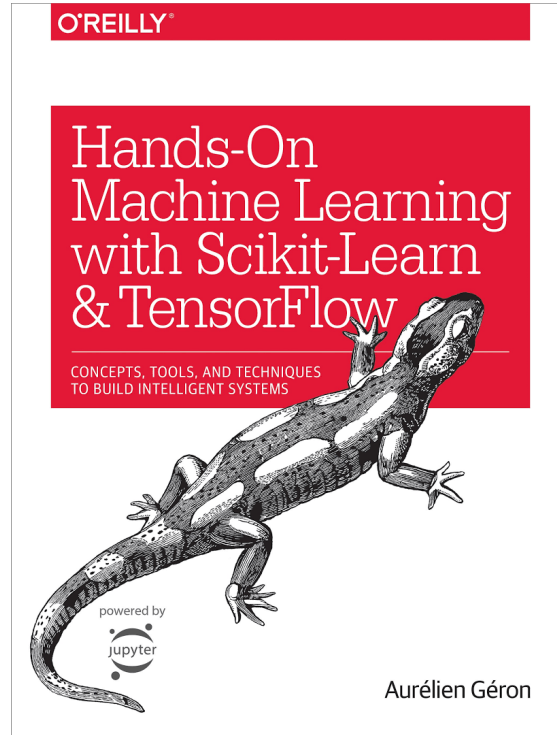
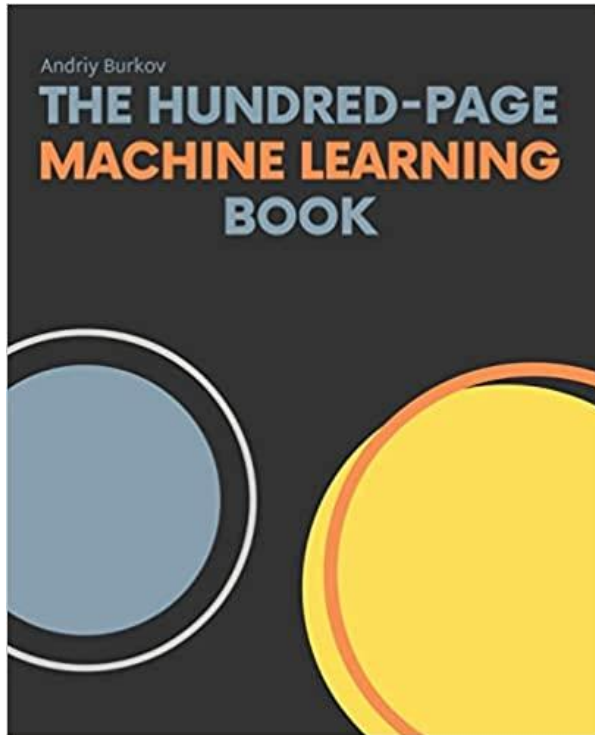


Start simple
Learn from feedback
Improve

CRISP-DM – Resumen

- **Entendimiento del problema:** Definir una meta medible. Preguntarse, ¿necesitamos ML?
- **Entendimiento de los datos:** ¿Tenemos los datos, son lo suficientemente buenos?
- **Preparación de los datos:** Transformar los datos en una tabla para entregar a un algoritmo de ML
- **Modelado:** Entrenar y seleccionar el mejor modelo
- **Evaluación:** Evaluar si se cumplió la meta
- **Despliegue:** Poner disponible el modelo para todos los usuarios
- **Iterar:** Empezar simple, aprender de la retroalimentación, mejorar

Libros recomendados



Aprendizaje continuo

YouTube



Two Minute Papers ✓

259,158 suscriptores



sentedx ✓

701,576 suscriptores



Siraj Raval ✓

663,099 suscriptores

Cursos en línea

The Coursera logo, consisting of the word "coursera" in a blue, lowercase, sans-serif font.



deeplearning.ai



CS231n:
Convolutional
Neural Networks for Visual
Recognition

