

ACÀMICA

¡Bienvenidos/as a Data Science!



Agenda

Proyecto 2 y Hasta Ahora

Explicación: Machine Learning

Break

Explicación: Aprendizaje Supervisado, Árbol de Decisión

Hands-on training

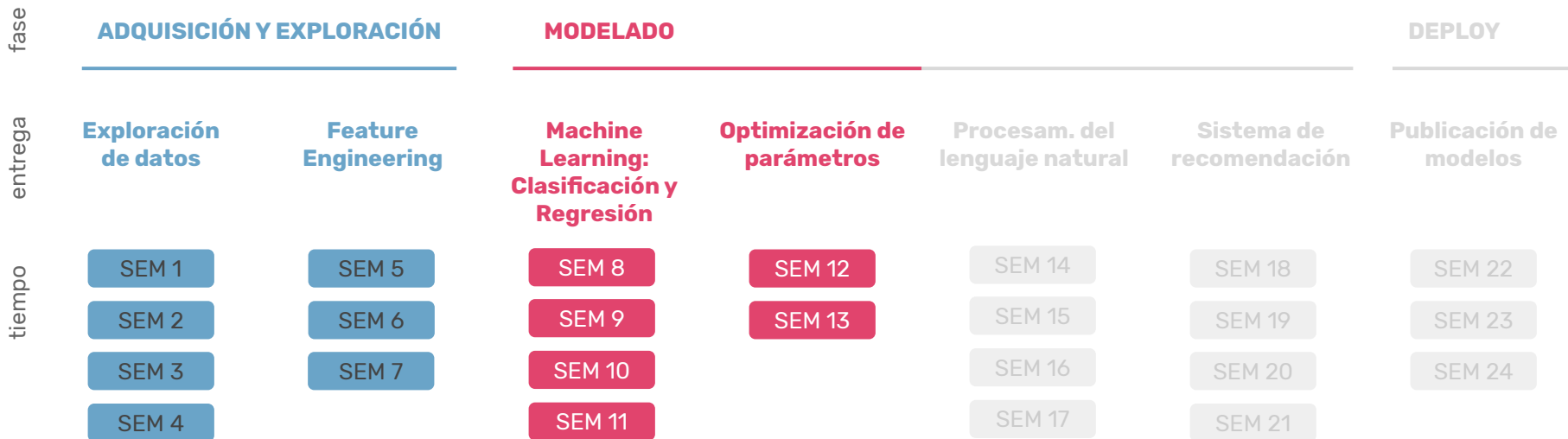
Cierre



Proyecto 2: Modelado



Análisis Exploratorio de Datos (EDA)



Proyecto EDA: Hoja de ruta

Entregas 1 y 2

SEM 1 - 7

EDA:

- Python
- Numpy
- Pandas
- Visualización de datos: Matplotlib y Seaborn
- Estadística
- Transformación de Datos
- Outliers

Usted
Está Aquí

SEM 8

- Intro a Machine Learning
- Aprendizaje Supervisado: Clasificación
- Árboles de Decisión
- Overfitting y Underfitting, Train/Test Split

SEM 9

- k-Vecinos más cercanos
- Métricas de Evaluación para Clasificación

Entrega 3

SEM 10

- Aprendizaje Supervisado: Regresión
- Métrica de Evaluación para Regresión

SEM 11



Hasta ahora...



Hasta ahora

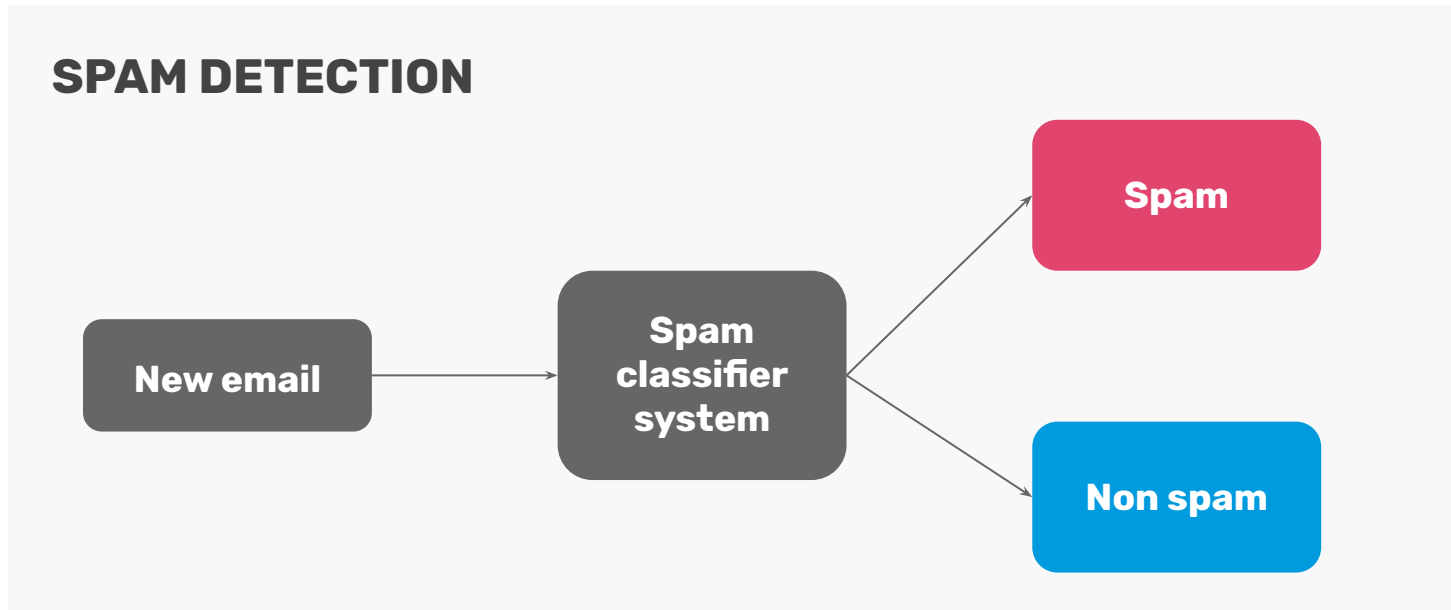
- ✓ Vimos un poco de programación con Python en un entorno particular, Jupyter, y aprendimos a utilizar muchas de las librerías típicas del ambiente de Data Science (Numpy, Pandas, Matplotlib, Seaborn, etc.)
- ✓ Repasamos varios conceptos de estadística: variables aleatorias, distribuciones, correlación, outliers, etc.
- ✓ Aprendimos algunas técnicas de preprocesamiento de datos con Pandas y con Scikit-Learn
- ✓ Aplicamos estas herramientas al **Análisis Exploratorio de Datos**

Vamos a seguir profundizando en herramientas (programación y librerías) y en estadística a lo largo de las clases. **¡Pero ahora vamos a ver cómo hace la computadora para aprender de los datos!**

Machine Learning



Machine Learning - Ejemplo clásico



¿Cual de estos mails parece ser spam?

Hola Juan,

Soy Pedro, el socio del proyecto inmobiliario. Quería avisarte que la reunión del jueves se pasó para el viernes.

Saludos,
Pedro.

Hola juan_86,

Soy Namubi, príncipe de Nigeria. Preciso que mande su numero de cuenta bancaria y contraseña para transferir herencia millonaria.

Caricias significativas,
Namubi

¿Cual de estos mails parece ser spam?

Hola Juan,

Soy Pedro, el socio del proyecto inmobiliario. Quería avisarte que la reunión del jueves se pasó para el viernes.

Saludos,
Pedro.

Hola juan_86,

Soy Namubi, príncipe de Nigeria. Preciso que mande su numero de cuenta bancaria y contraseña para transferir herencia millonaria.

Caricias significativas,
Namubi



**¿Cómo distinguieron
Spam de no Spam?**

¿Como distinguieron Spam de no Spam?

No es una tarea sencilla de realizar, de hecho hoy en día mucha gente es víctima de publicidad engañosa o estafas por medio de mails.

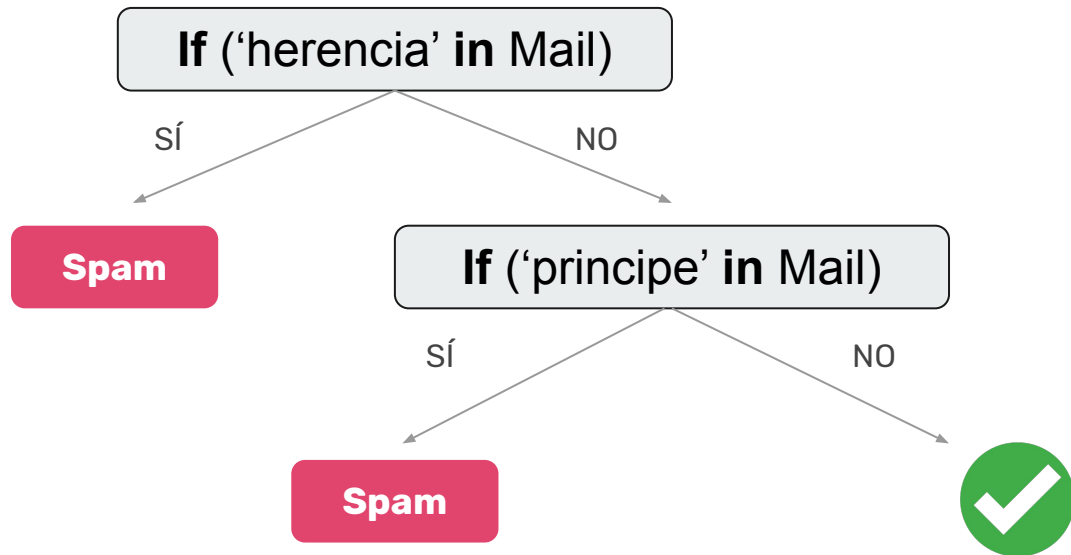
¿Como distinguieron Spam de no Spam?

No es una tarea sencilla de realizar, de hecho hoy en día mucha gente es víctima de publicidad engañosa o estafas por medio de mails.

La tarea implica un procesamiento de alto nivel de abstracción (saber leer, relacionar conceptos, etc...), por lo cual resulta difícil (casi imposible) programar explícitamente un algoritmo que la realice.

**¿Cómo se imaginan un
algoritmo (programa)
que realice esta tarea?**

Algoritmo de detección de spam



Algoritmo de detección de spam



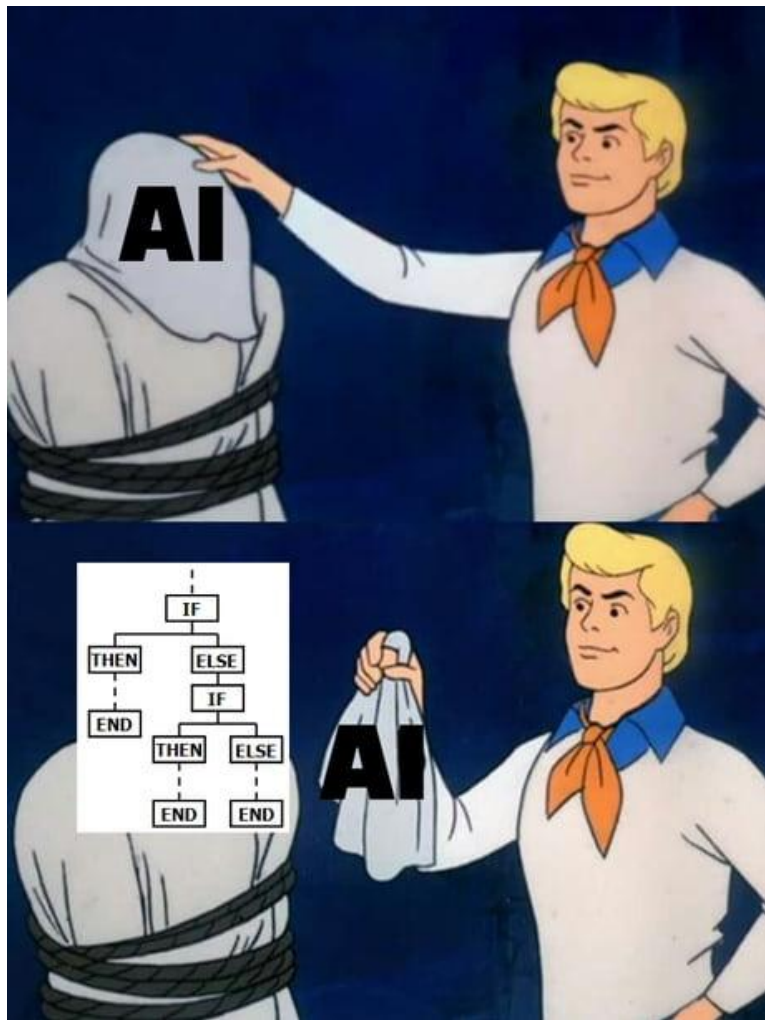


Un algoritmo construido de esta forma 'manual' NO es lo que entendemos por Machine Learning.



Un algoritmo construido de esta forma 'manual' NO es lo que entendemos por Machine Learning.

¿POR QUÉ?



Un algoritmo construido de esta forma 'manual' NO es lo que entendemos por Machine Learning.

¿POR QUÉ?

En Machine Learning el modelo (algoritmo) debe aprender a predecir a partir de los datos.

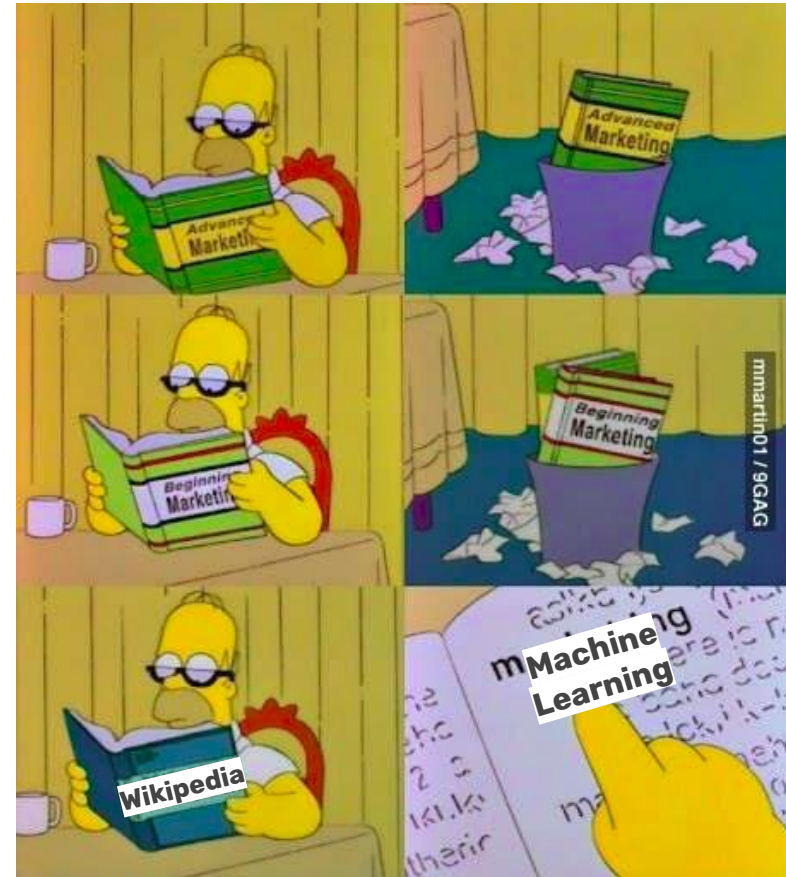
¿Cuál es la definición de MACHINE LEARNING?



Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence.

Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.^{[1][2]:2}

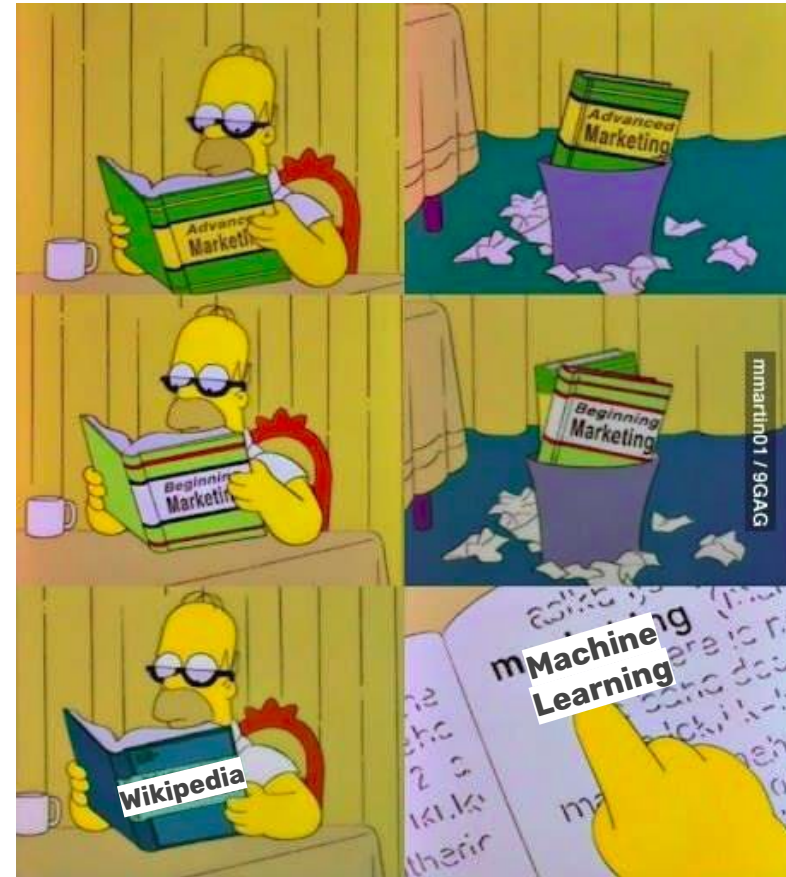
Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task.



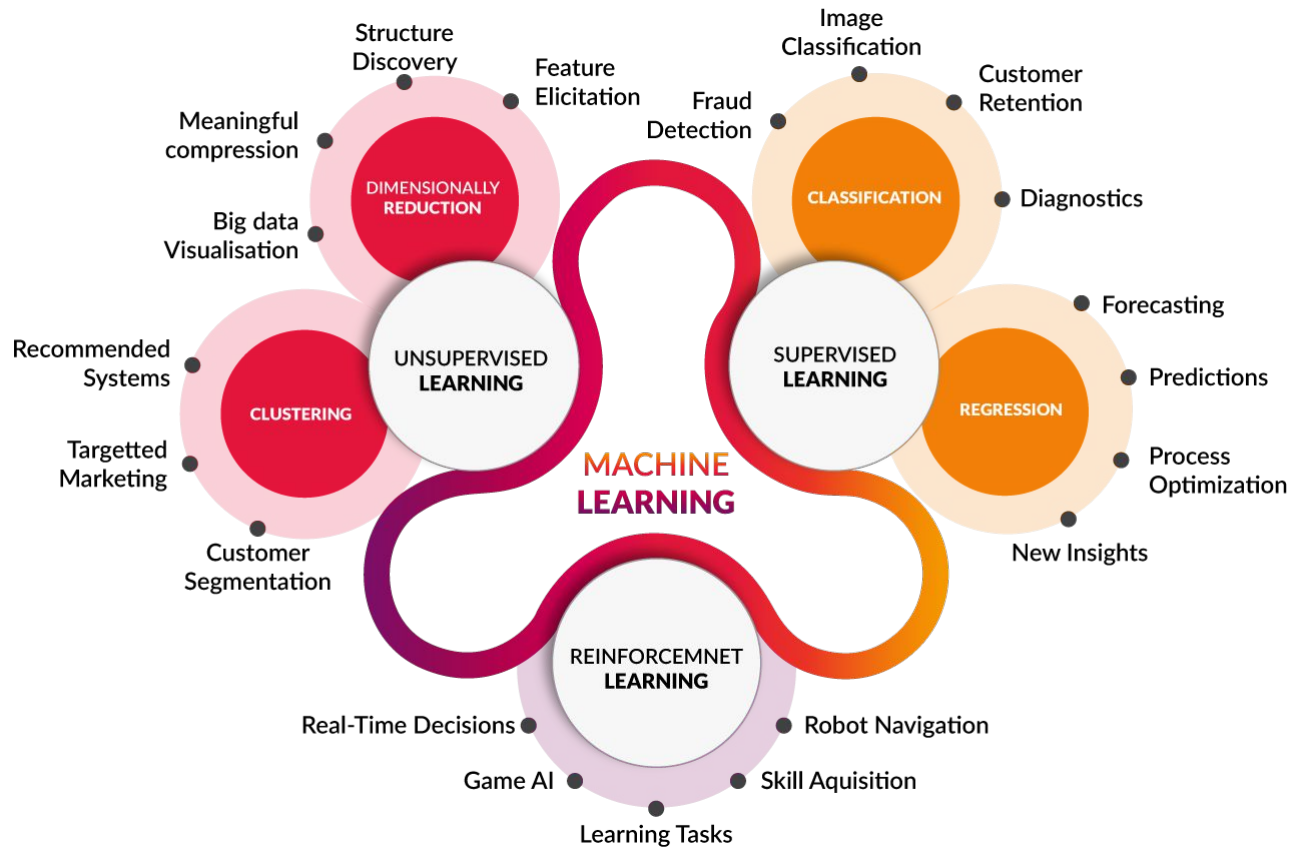
Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence.

Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.^{[1][2]:2}

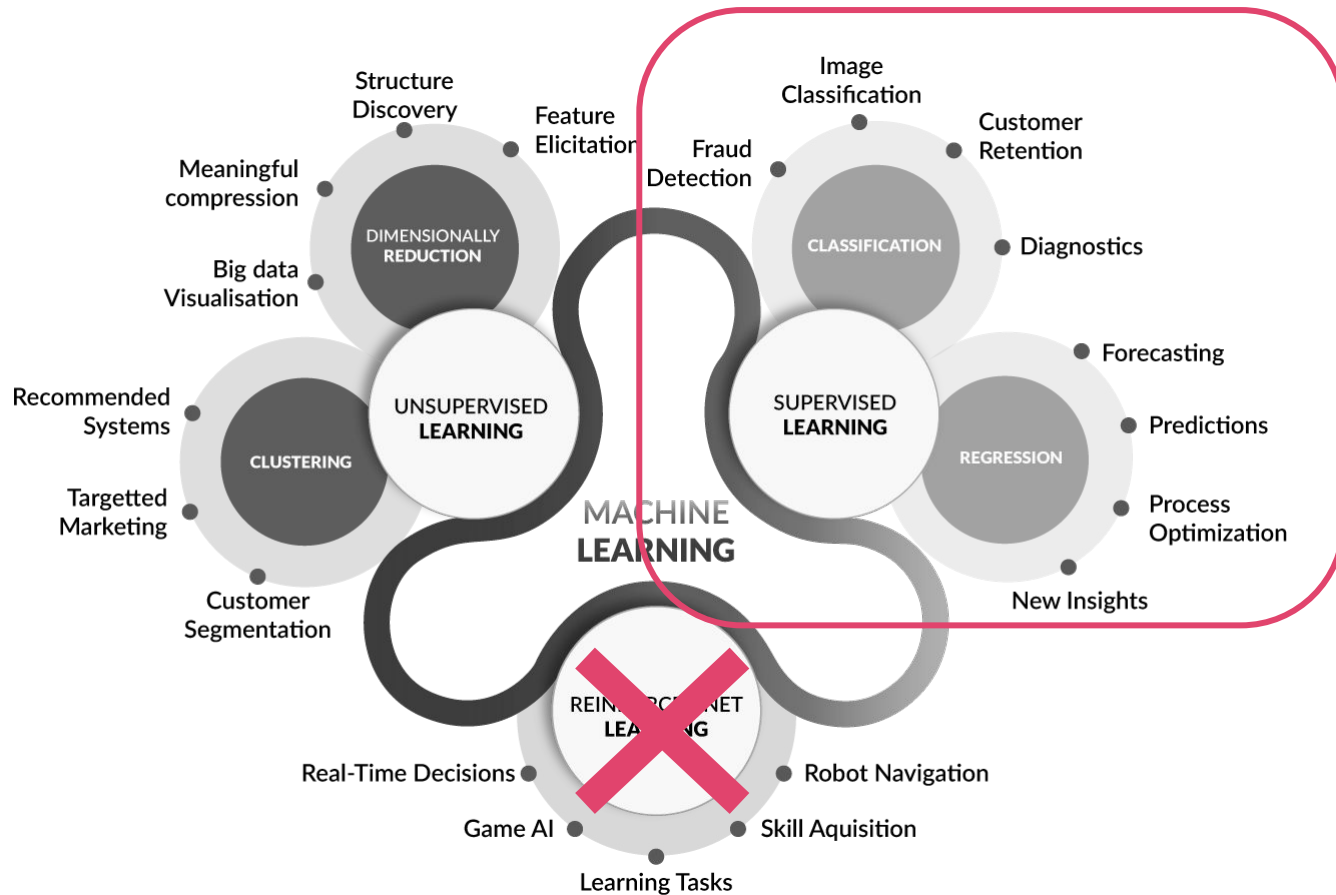
Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task.



Mapa



Mapa



Aprendizaje Supervisado

$$f(X) = Y$$

tenemos datos X

tenemos datos Y

Aprendizaje Supervisado

$$f(X) = Y$$

¿Qué buscamos
con "f"?

tenemos datos Y

tenemos datos X

Aprendizaje Supervisado

$$f(X) = Y$$

Un modelo **f** que
permita determinar
la salida a partir de la
entrada

tenemos datos X

tenemos datos Y

Aprendizaje Supervisado

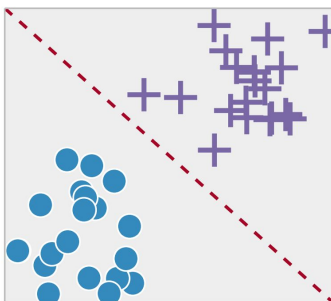
$$f(X) = Y$$



Con este modelo podremos predecir **Y**, para nuevos datos **X** de los cuales no conocamos la salida.

Aprendizaje Supervisado

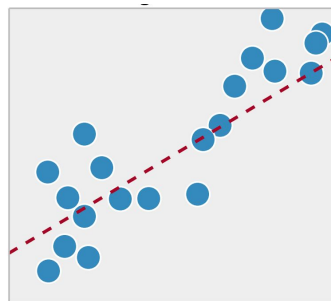
Clasificación



La variable de salida es una categoría:

- Enfermo / Sano
- Gato / Perro / Pájaro
- **Spam / no Spam**

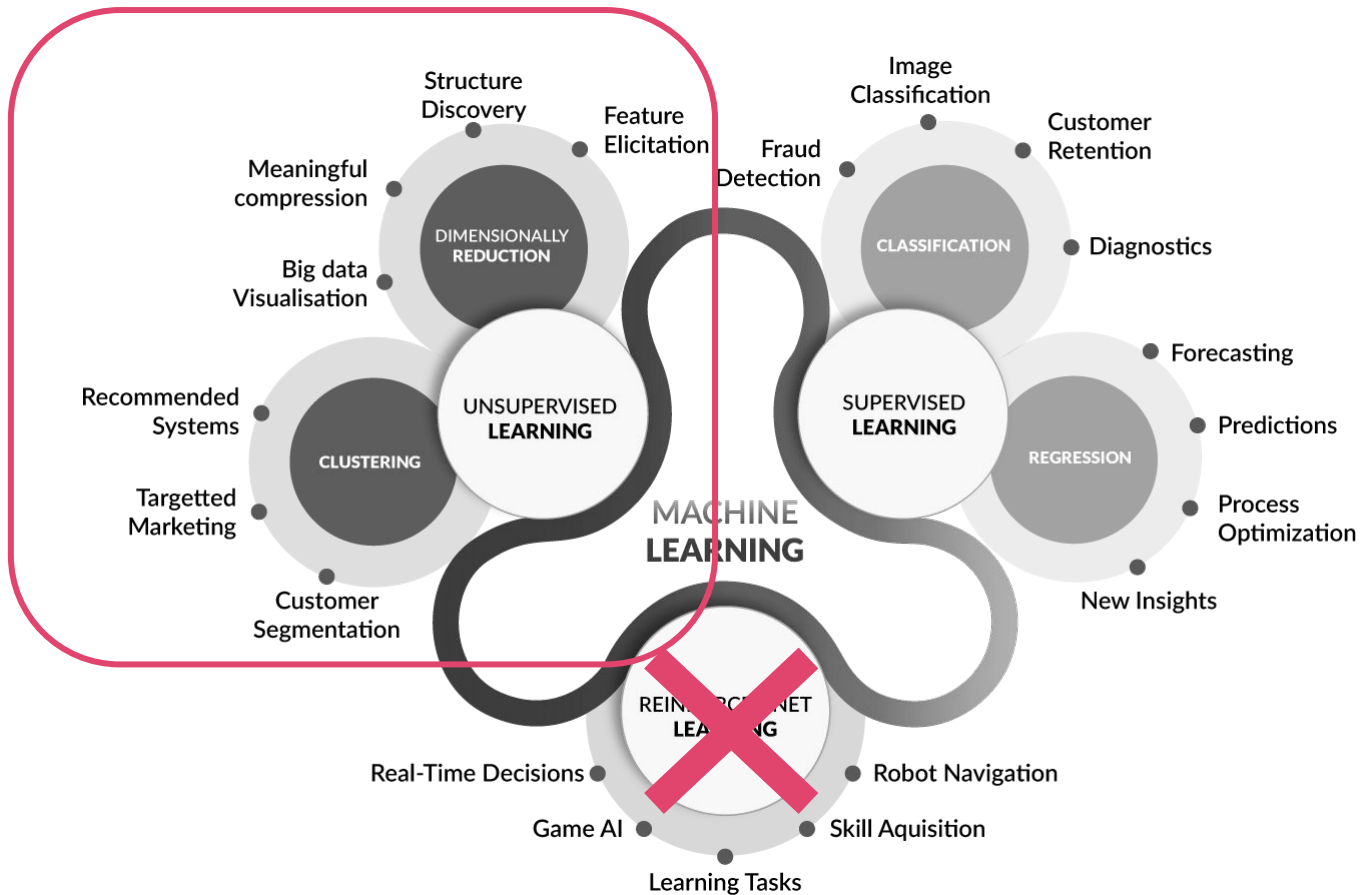
Regresión



La variable de salida es un valor:

- Precio
- Cantidad

Mapa



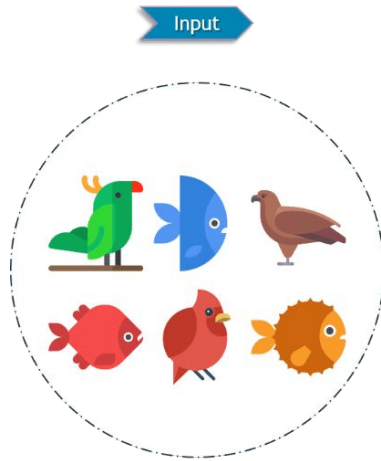
Aprendizaje **NO** Supervisado

En este tipo de algoritmos **solo** tenemos los datos de entrada **X**, no hay una salida deseada **Y**.

Lo que se buscan son patrones de similaridad en los datos de entrada.

Aprendizaje **NO** Supervisado

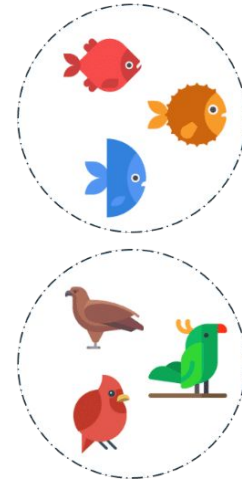
Datos de
entrada **X**



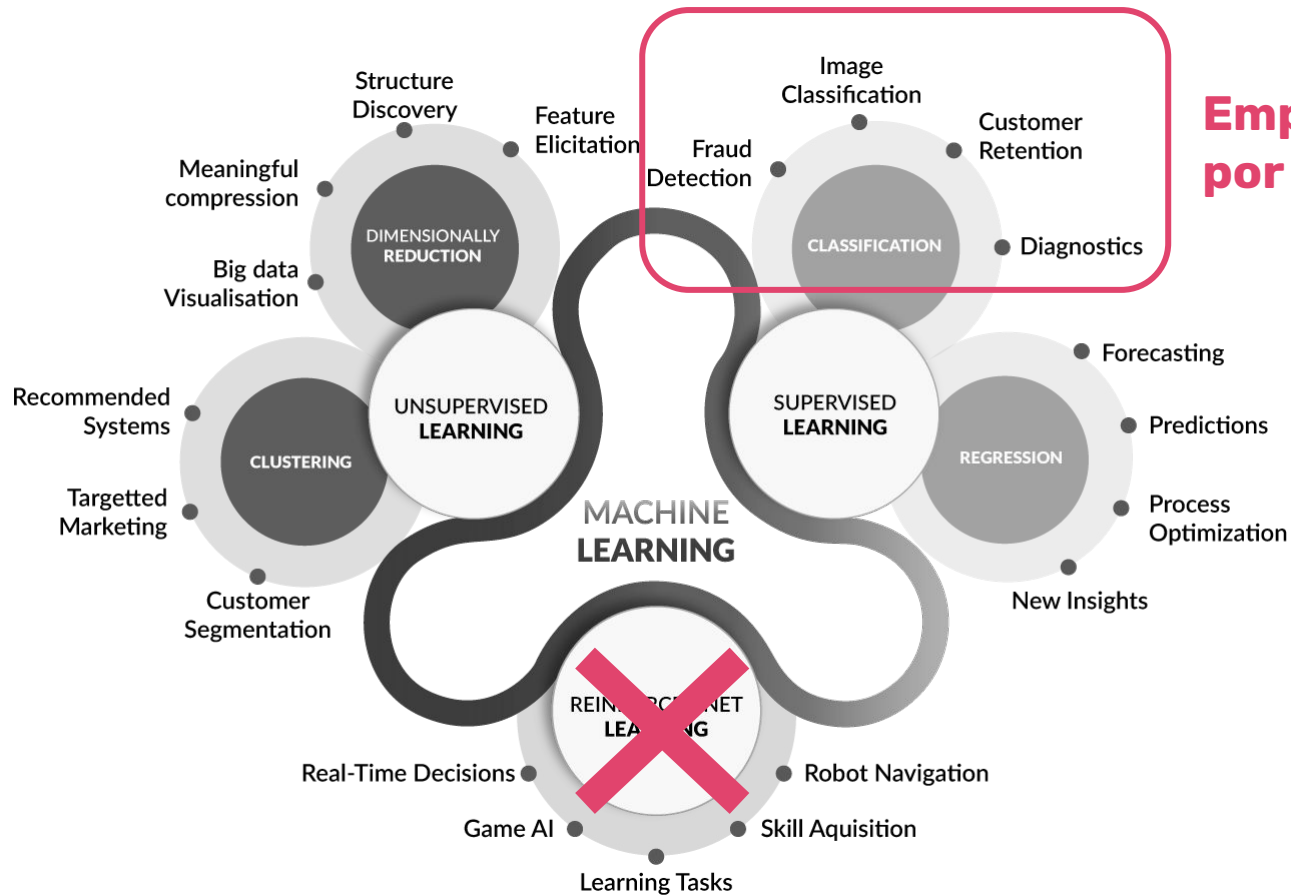
Algoritmo no
Supervisado



Patrones
encontrados
(grupos)



Mapa



**Empezamos
por acá**

A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver spoon are visible, though they are out of focus. The overall lighting is soft and even, highlighting the textures of the coffee and the smooth surface of the cup.

¡BREAK!



Aprendizaje Supervisado: Árbol de decisión



Machine Learning



Aprendizaje Supervisado



Clasificación



Modelos

- **Árbol de Decisión**
- Support Vector Machines
- k-nearest neighbors
- Random Forest
- Perceptrón
- etc...

CC

CRASH COURSE COMPUTER SCIENCE

34

MACHINE LEARNING

MACHINE LEARNING



Árbol de decisión - Ejemplo del video

LABELED DATA

MASS	WINGSPAN	MOTH SPECIES (LABEL)
2.5	41	EMPEROR
3.7	36	EMPEROR
4.2	72	LUNA
3.2	27	EMPEROR
5.5	60	LUNA
5.1	51	EMPEROR
4.0	43	EMPEROR
3.6	40	EMPEROR
8.2	75	LUNA
7.6	82	LUNA
...

Árbol de decisión - Ejemplo del video

LABELED DATA

MASS	WINGSPAN	MOTH SPECIES (LABEL)
2.5	41	EMPEROR
3.7	36	EMPEROR
4.2	72	LUNA
3.2	27	EMPEROR
5.5	60	LUNA
5.1	51	EMPEROR
4.0	43	EMPEROR
3.6	40	EMPEROR
8.2	75	LUNA
7.6	82	LUNA
...

Features **X**

Target **Y**

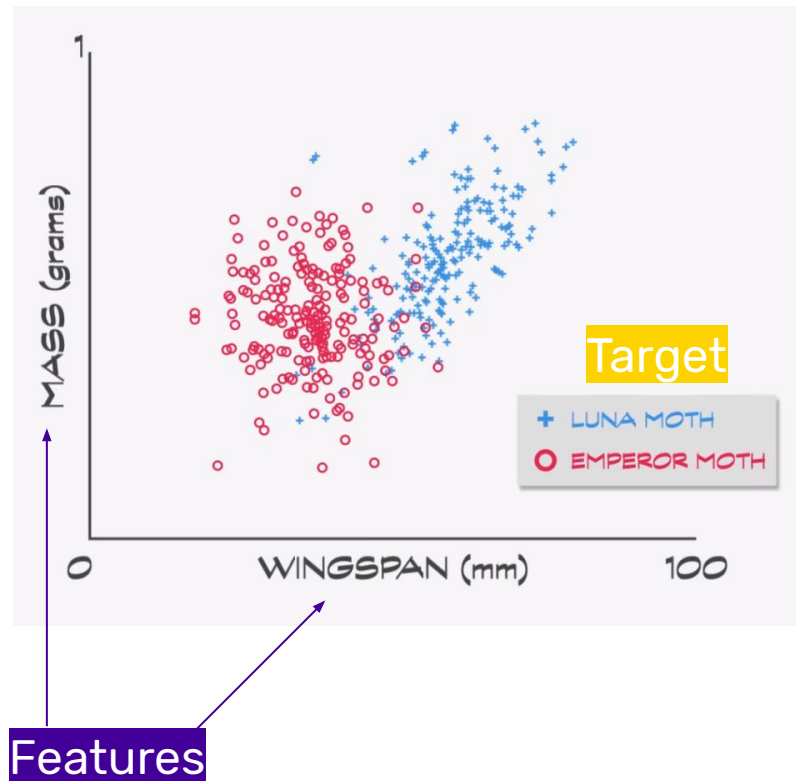
Árbol de decisión - Ejemplo del video

LABELED DATA

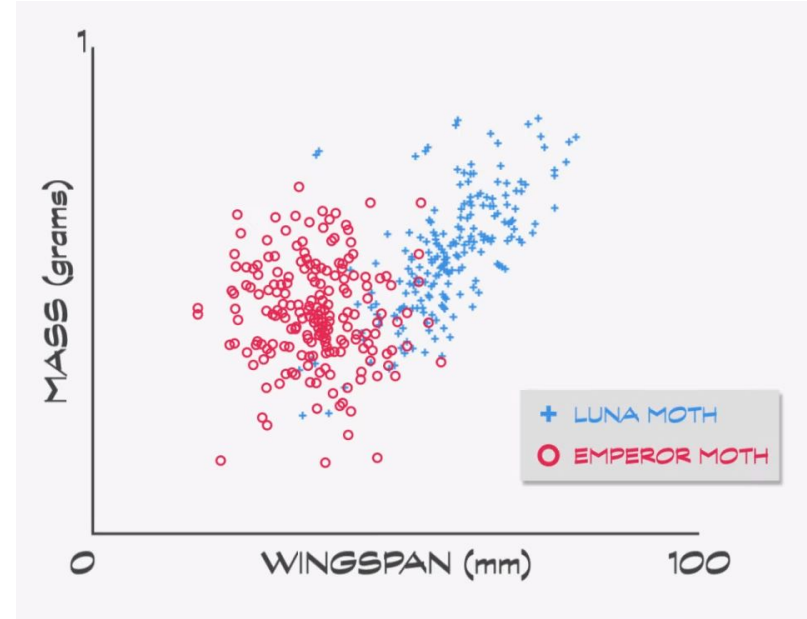
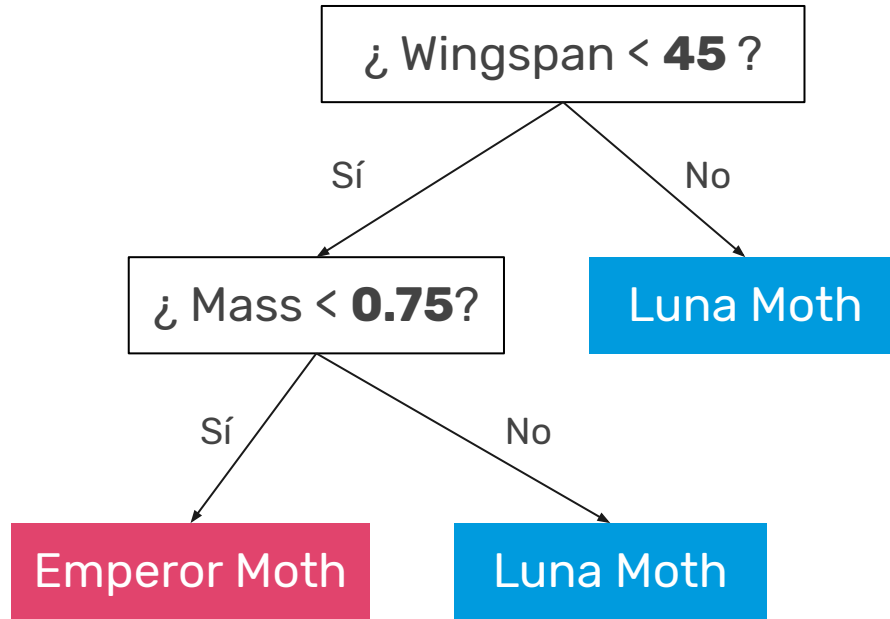
MASS	WINGSPAN	MOTH SPECIES (LABEL)
2.5	41	EMPEROR
3.7	36	EMPEROR
4.2	72	LUNA
3.2	27	EMPEROR
5.5	60	LUNA
5.1	51	EMPEROR
4.0	43	EMPEROR
3.6	40	EMPEROR
8.2	75	LUNA
7.6	82	LUNA
...

Features **X**

Target **Y**



Árbol de decisión - Ejemplo del video



Árbol de decisión

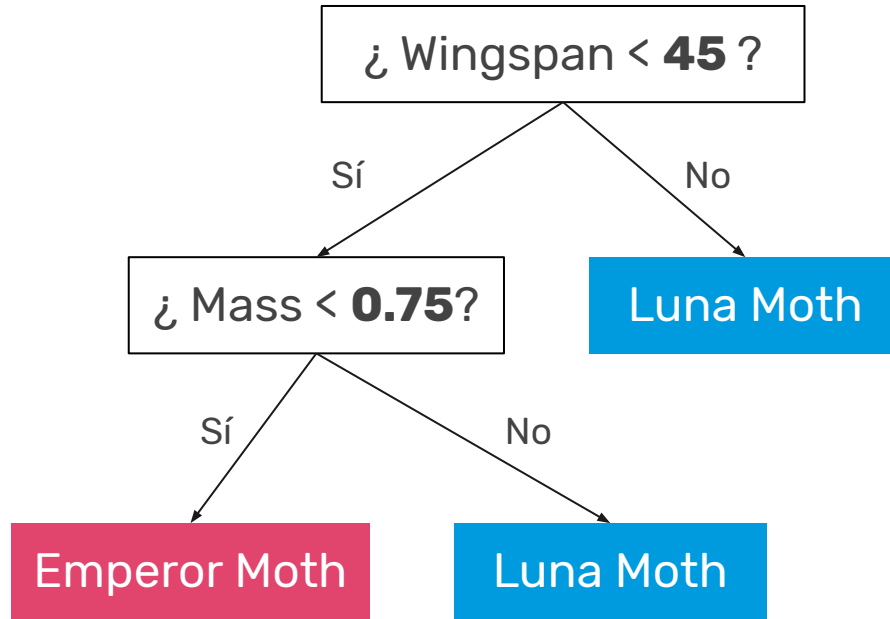
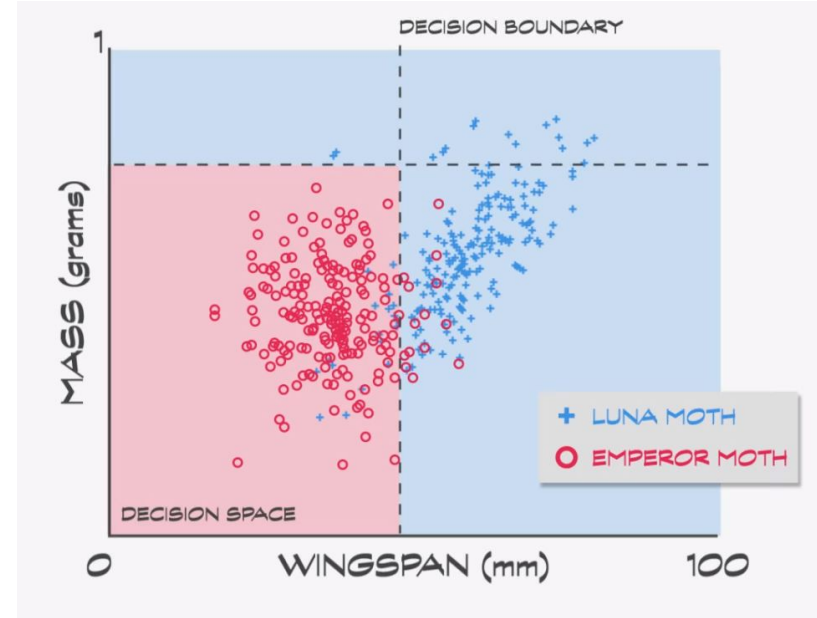
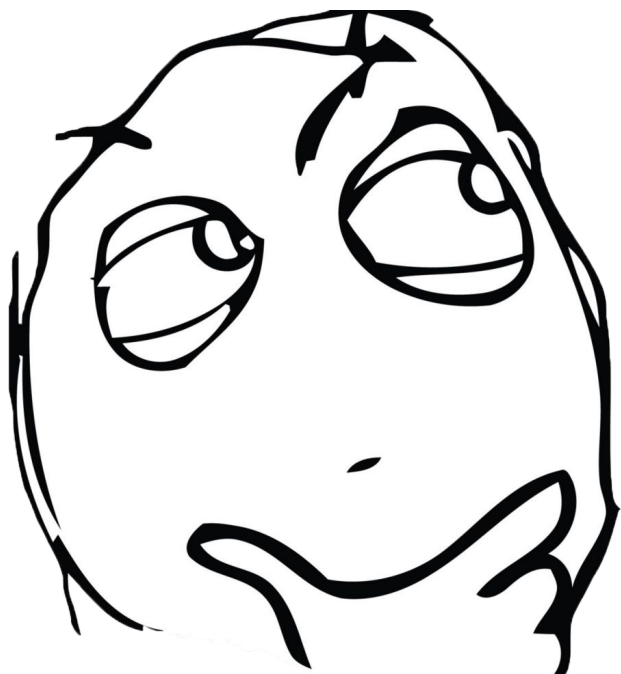
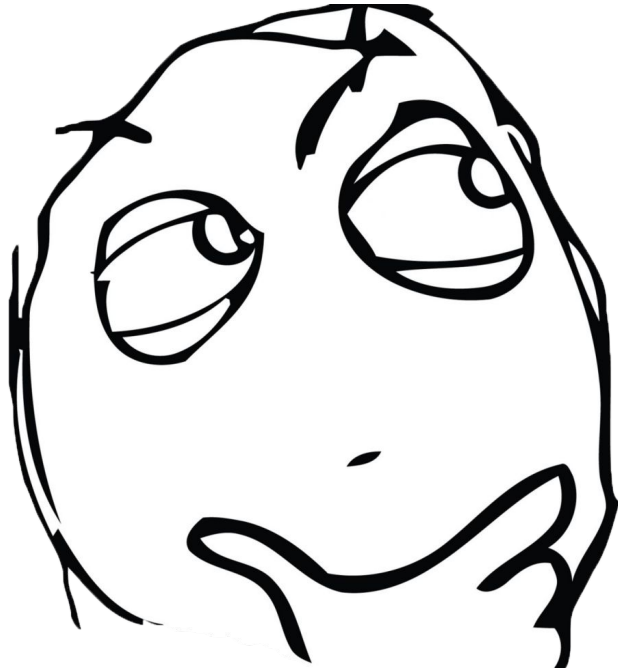


Gráfico de decisión



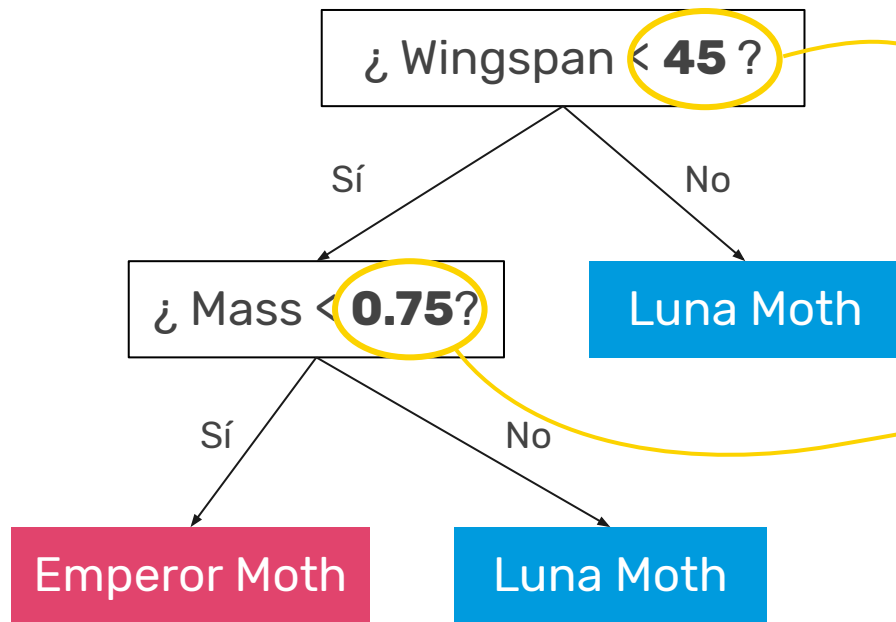




Es una serie de **IFs**...

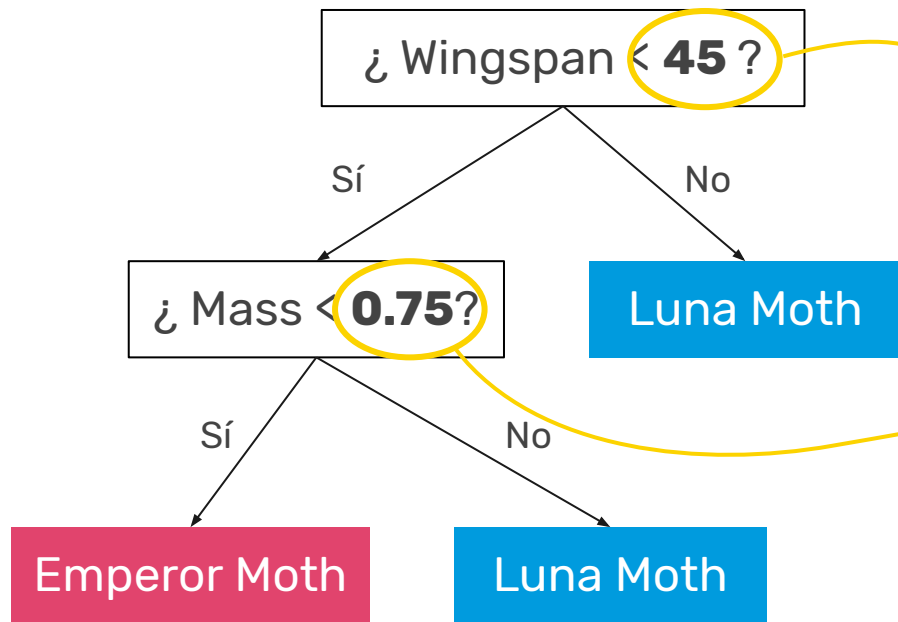
¿Por qué ahora decimos
que es Machine Learning?

¿Por qué decimos que es Machine Learning?



Es **Machine Learning** porque estos valores se eligen automáticamente al entrenar el modelo, a partir de los datos **X** e **Y**.

¿Por qué decimos que es Machine Learning?



La próxima clase vamos a ver **cómo** es que se eligen estos valores a partir de un proceso matemático.

Flujo de trabajo **Scikit Learn**

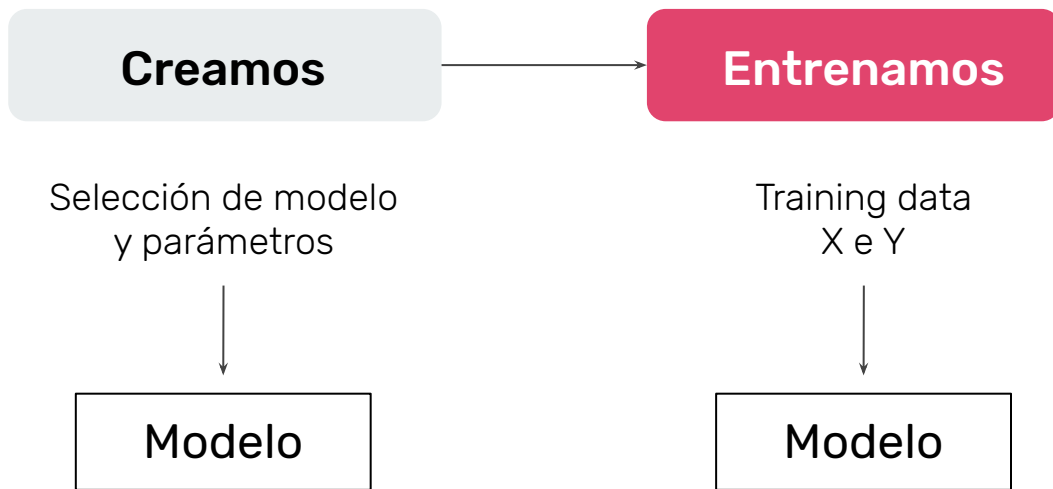
Creamos

Selección de modelo
y parámetros

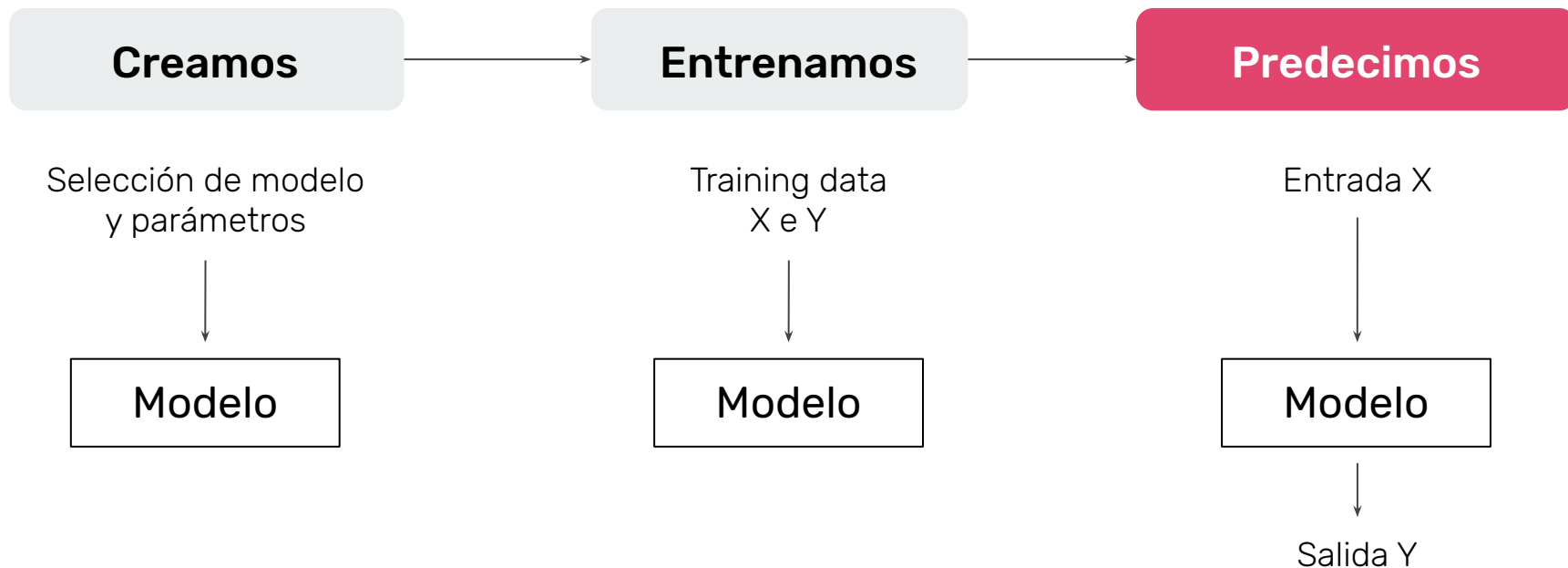


Modelo

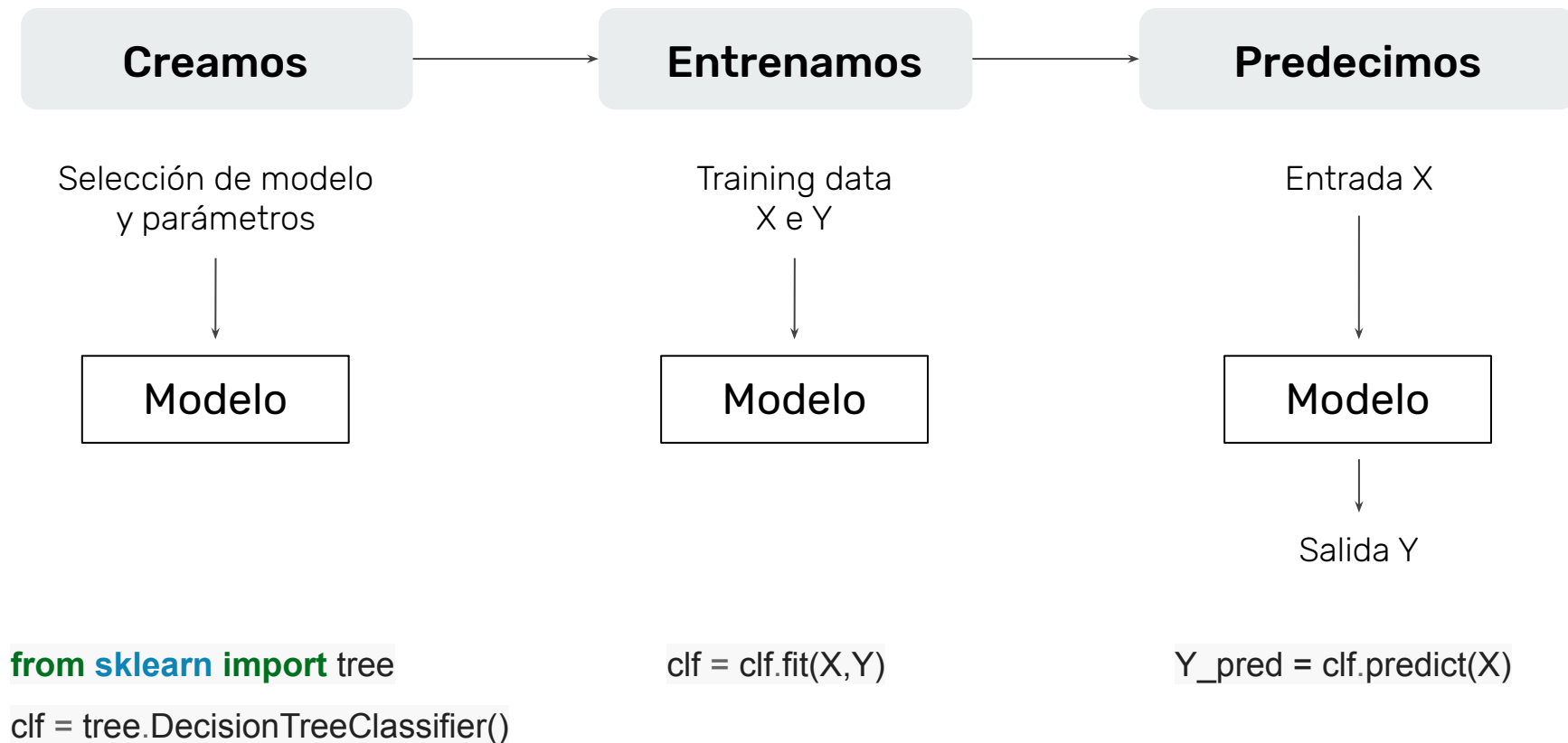
Flujo de trabajo **Scikit Learn**



Flujo de trabajo **Scikit Learn**



Flujo de trabajo **Scikit Learn**



Hands-on training



**Hands-on
training**



DS_Clase_15_ML.ipynb

Recursos



Recursos



1. [Scikit-Learn Decision Trees Explained](#): Buena (y completa) introducción a Árboles de decisión con Scikit-Learn.
2. **Capítulo 5, “Machine Learning”, de [Python Data Science Handbook](#)**. Acá van a encontrar una introducción general a ML.
3. **Capítulo 5.08, “In-Depth: Decision Trees and Random Forests”, de [Python Data Science Handbook](#)**. Acá van a encontrar árboles de decisión explicado con código funcional para copiar y pegar en sus proyectos.
4. [Video muy interesante](#) sobre el impacto de Machine Learning (Automatización moderna) en la sociedad.



Para la próxima

1. Ver los videos de la plataforma “Machine Learning: Árboles de Decisión” y “Validación y testeo de modelos: Validación y testeo” (¡nos salteamos algunos videos!)
2. Completar el notebook de hoy
3. ¡Terminar la entrega 02 si aún no lo hicieron!

ACÀMICA