

ACÀMICA

# 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, etc.
- ✓ Aprendimos algunas técnicas de preprocesamiento de datos con Pandas
- ✓ Aplicamos estas herramientas al **Análisis Exploratorio de Datos**

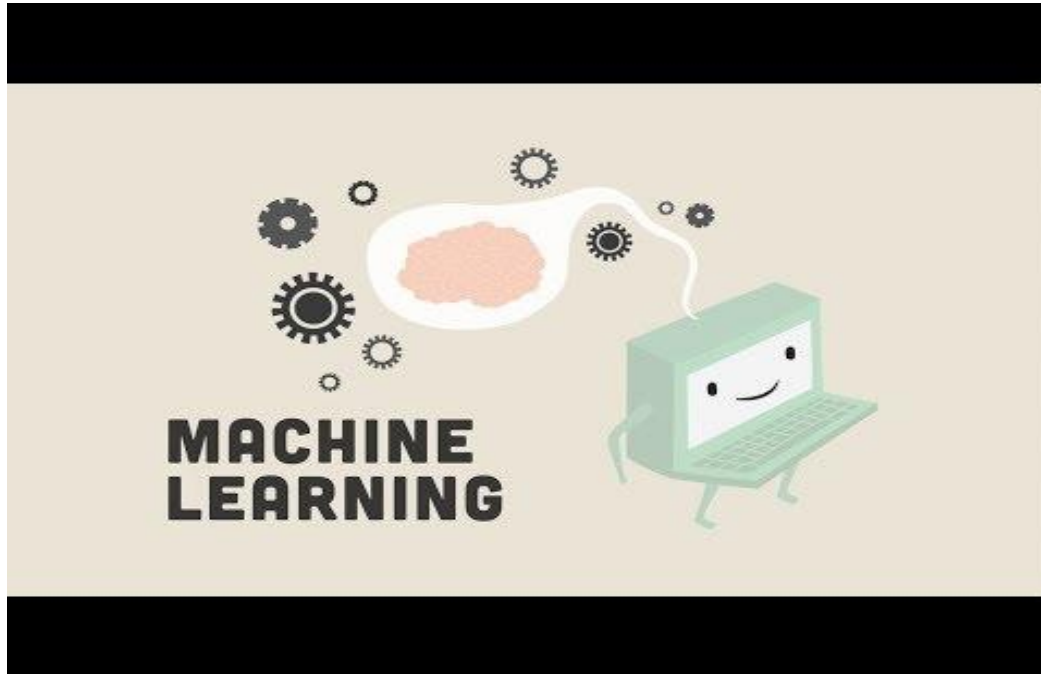
Vamos a seguir profundizando en herramientas (programación y librerías) y en estadística a lo largo de las clases.

**¡Ahora vamos a ver cómo hace la computadora para aprender de los datos!**



TEMA DEL DÍA

# Machine Learning



# Agenda

---

Daily

Explicación: Árbol de decisión

**Break.**

Hands-on training

Automatización y futuro del trabajo

Cierre.



# Daily



Daily



## Sincronizando...

### Toolbox



¿Cómo te ha ido?  
¿Obstáculos?  
¿Cómo seguimos?

### Challenge

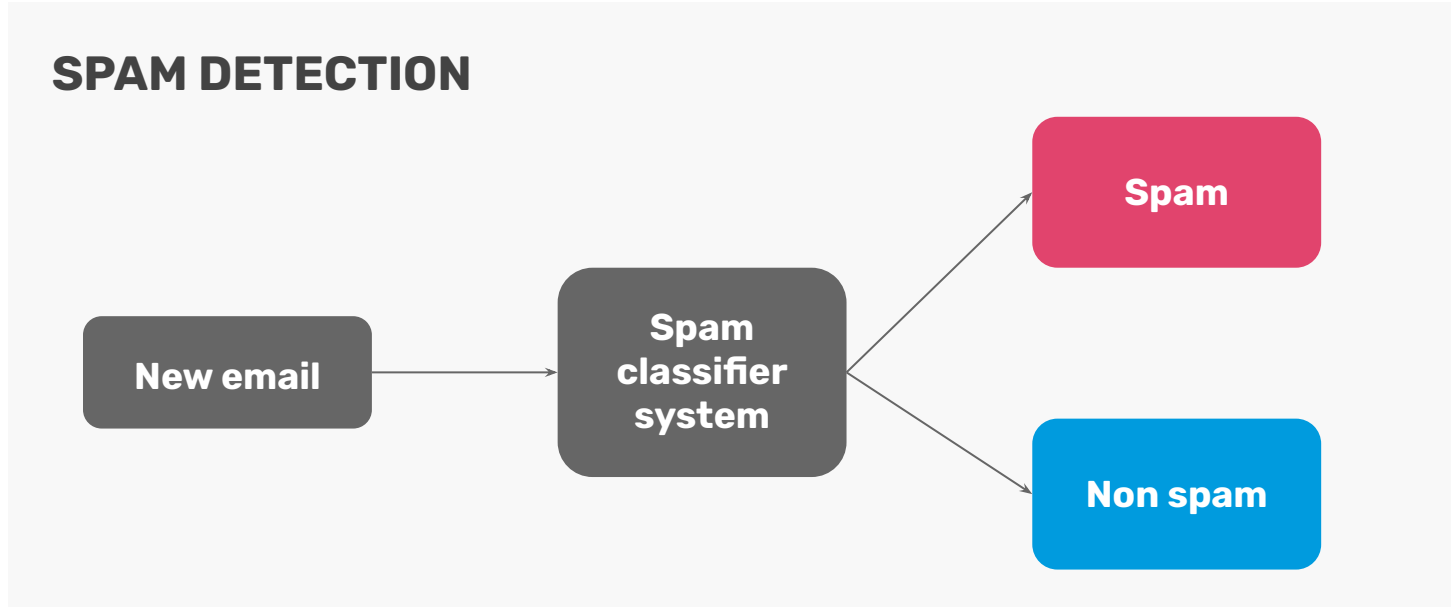


¿Cómo te ha ido?  
¿Obstáculos?  
¿Cómo seguimos?

# 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.

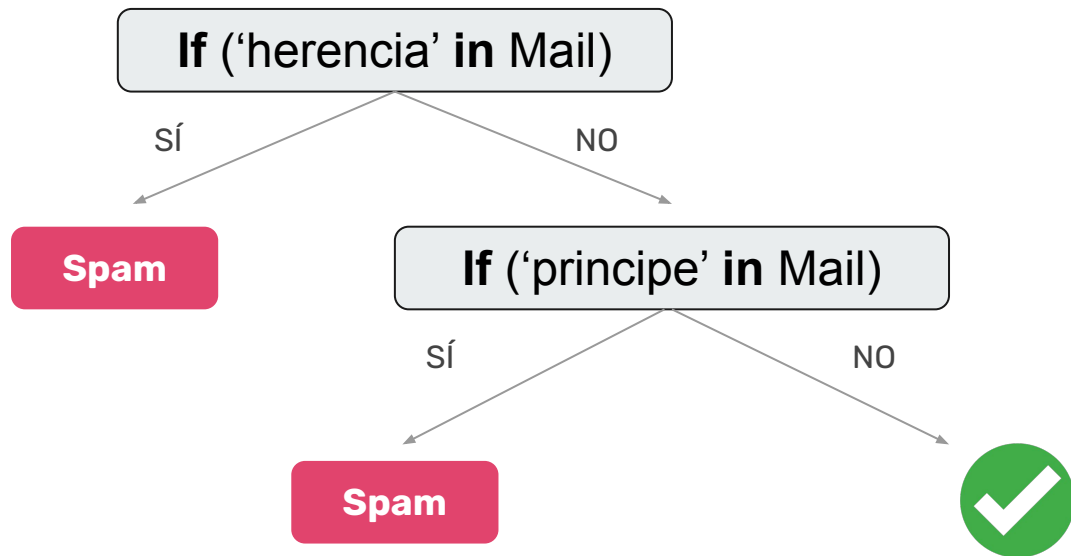
# ¿Como distinguieron Spam de no Spam?

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



Entonces, ¿era ésto  
nomás Machine  
Learning?

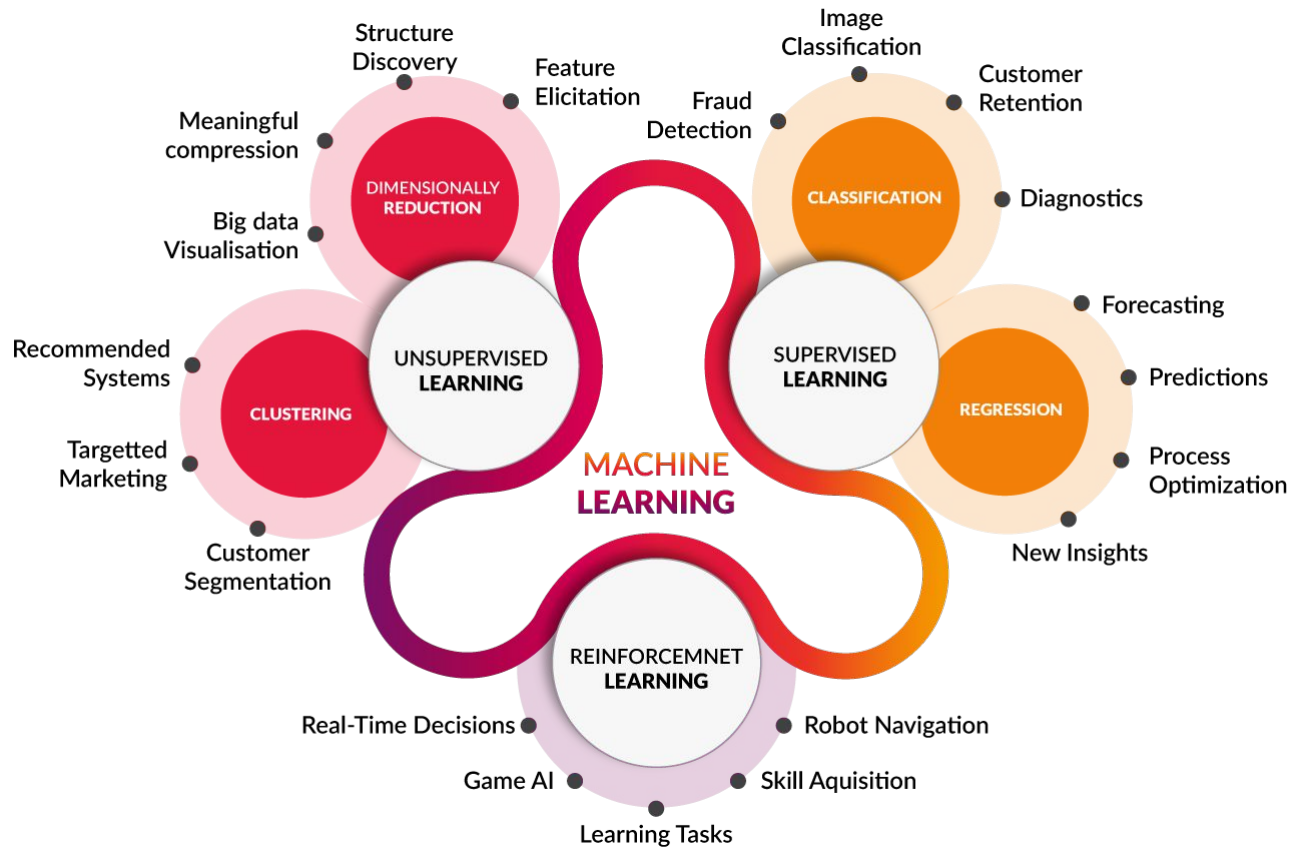
Tratemos de ver si es  
realmente así.

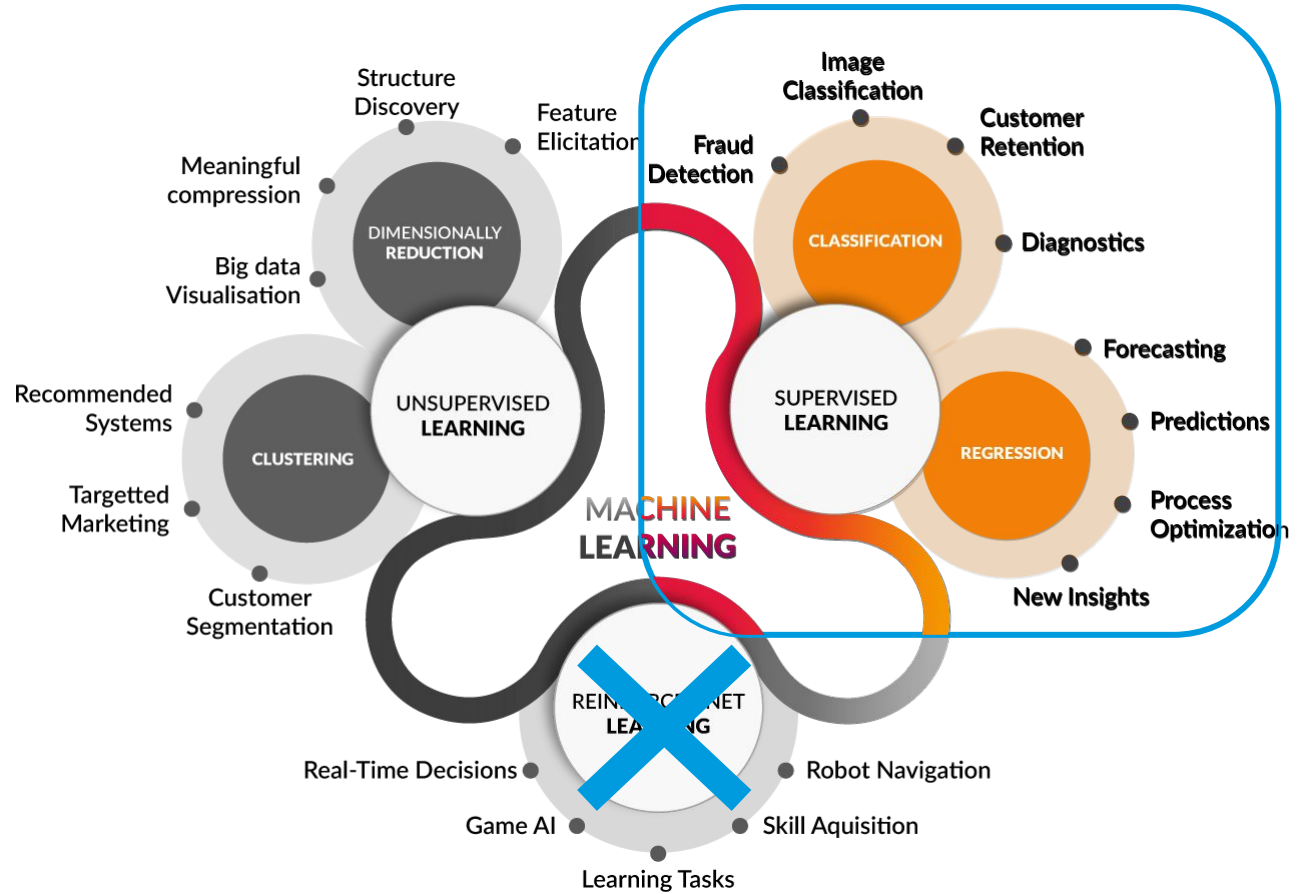


# Aprendizaje supervisado



# Mapa





# Aprendizaje Supervisado

$$f(X) = Y$$

tenemos datos/atributos X

tenemos datos/etiquetas Y

# Aprendizaje Supervisado

$$f(X) = Y$$

¿Qué buscamos  
con "f"?

tenemos datos/etiquetas Y

tenemos datos/atributos X

# Aprendizaje Supervisado

$$f(X) = Y$$

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

tenemos datos/atributos X

tenemos datos/etiquetas Y

# Aprendizaje Supervisado

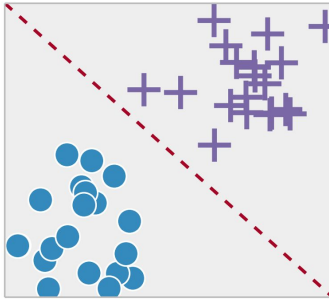
$$f(X) = Y$$



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

# Aprendizaje Supervisado

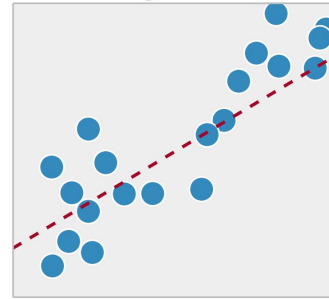
## Clasificación



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



# Árbol de decisión



Machine Learning



Aprendizaje Supervisado



**Clasificación**



**Modelos**

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

# Árbol de decisión

**Un árbol de decisión, al igual que cualquier modelo de clasificación, te permite generar un modelo predictivo en base a un Dataset de entrenamiento.**

CC

CRASH COURSE COMPUTER SCIENCE

34

# MACHINE LEARNING

MACHINE LEARNING



# Árbol de decisión

## 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**

Imaginemos que tenemos este Dataset de polillas Emperador y Luna.



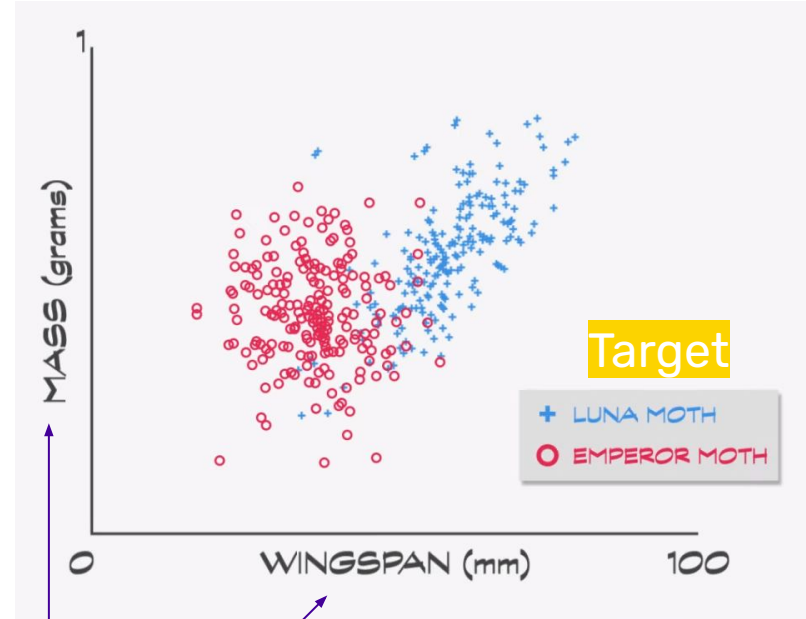
# Árbol de decisión

## 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
...	...	...

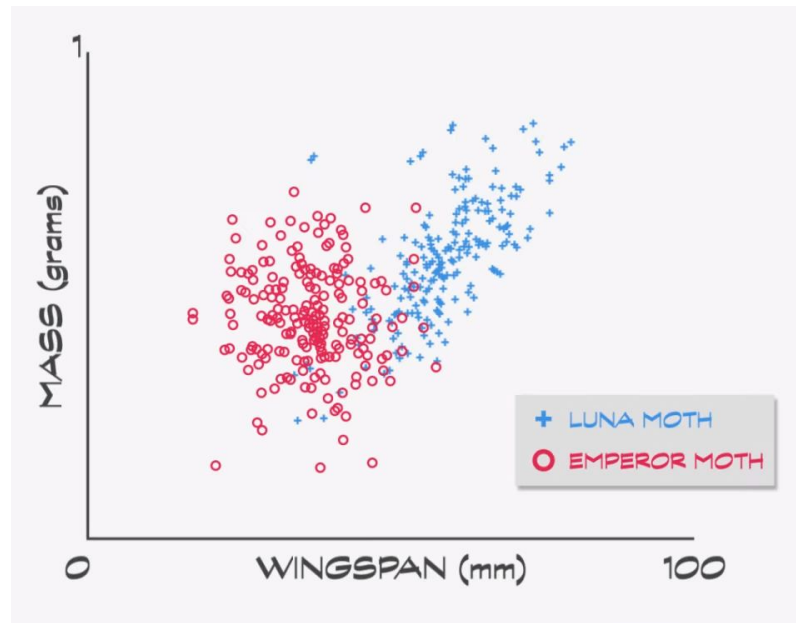
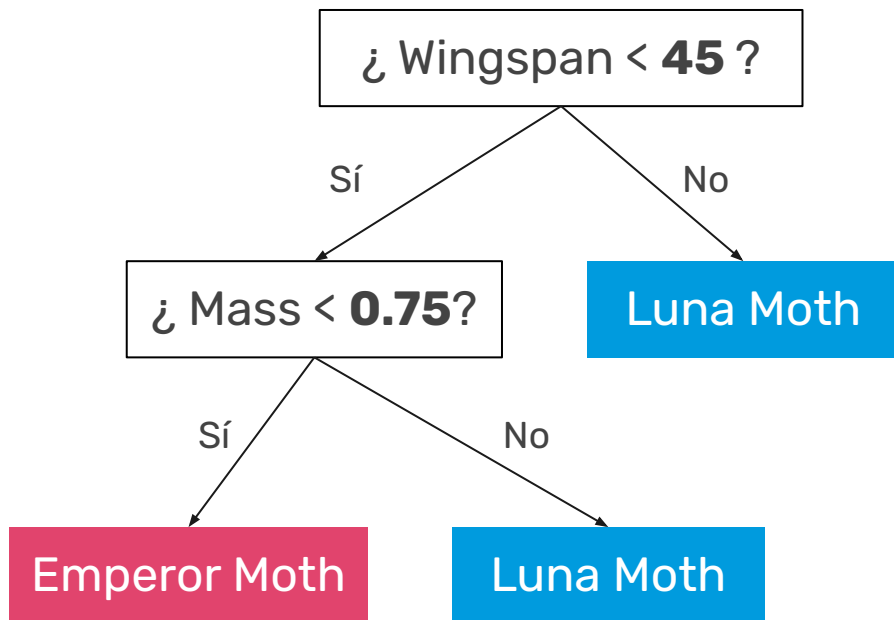
Features **X**

Target **Y**

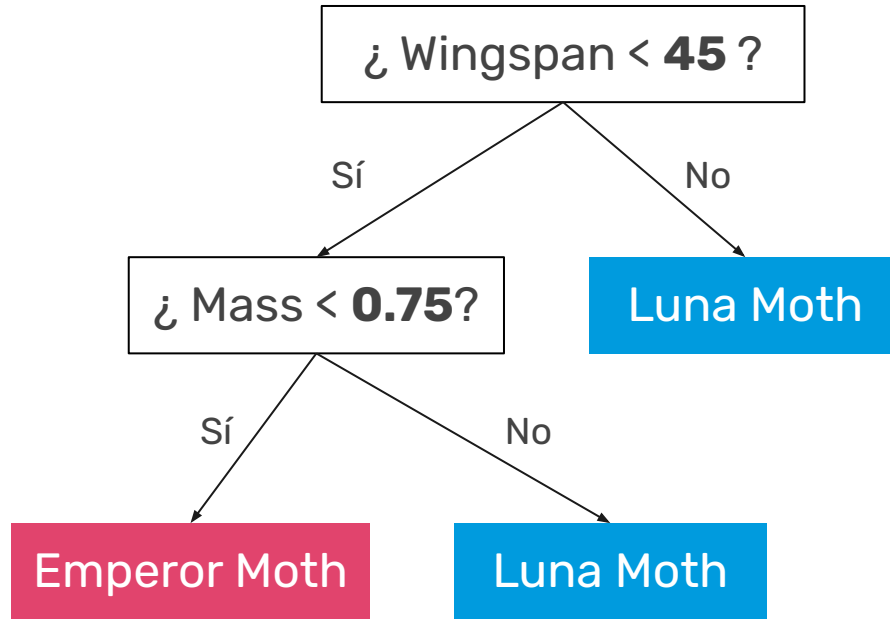


Features/Atributos

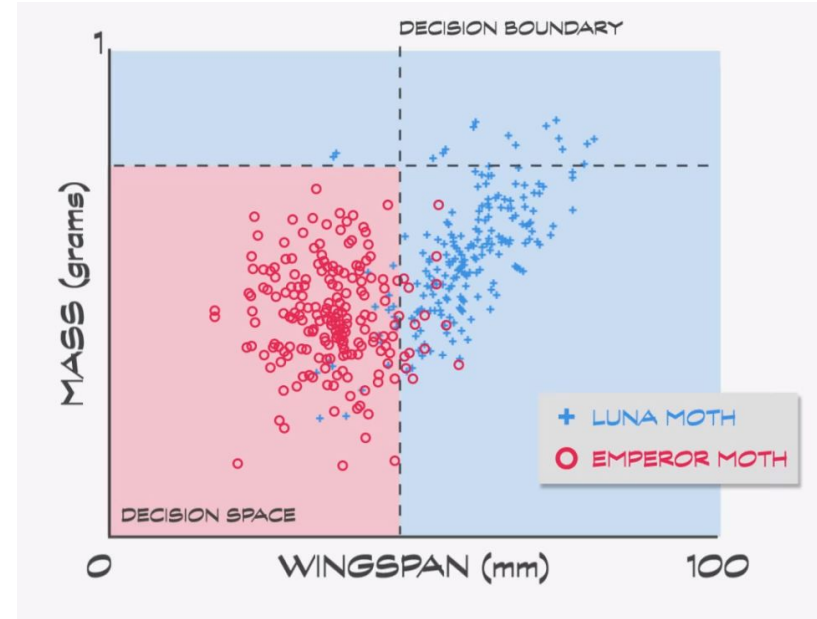
# Árbol de decisión



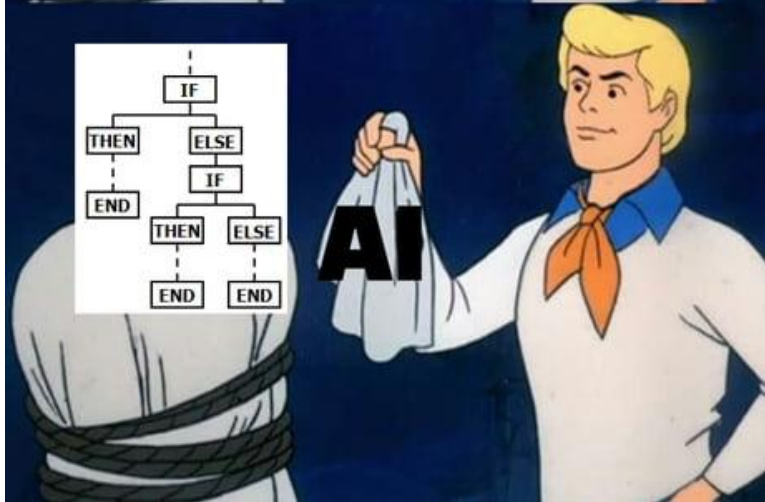
# Árbol de decisión

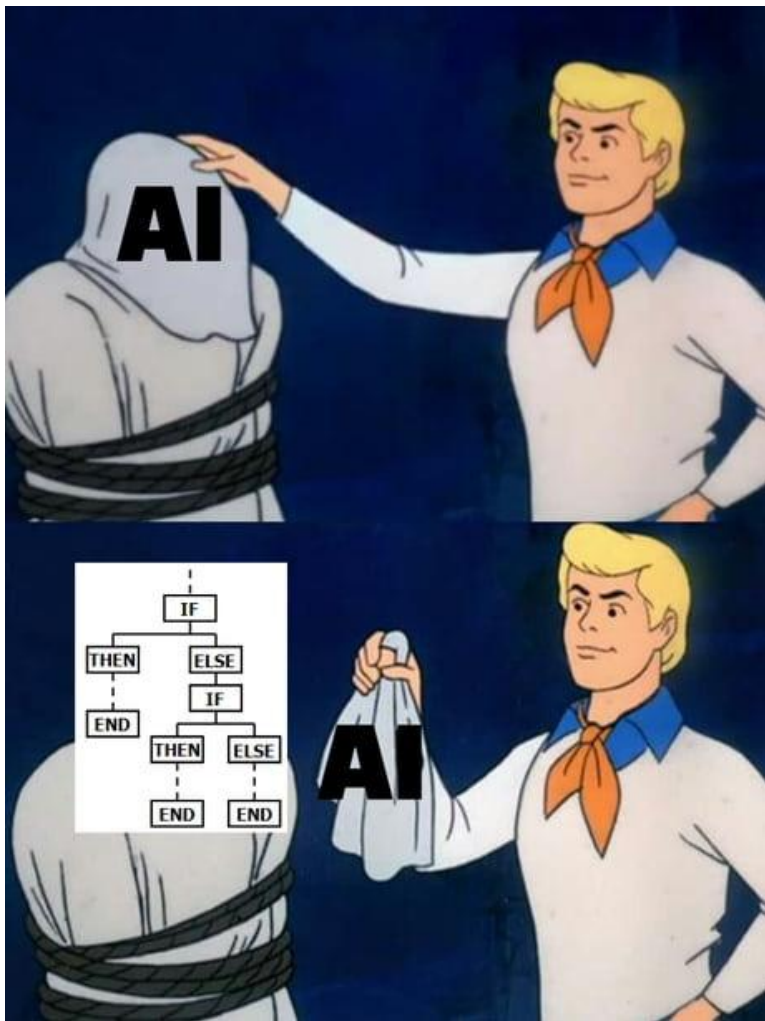


# Gráfico de decisión



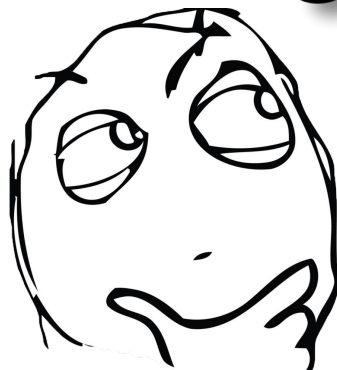




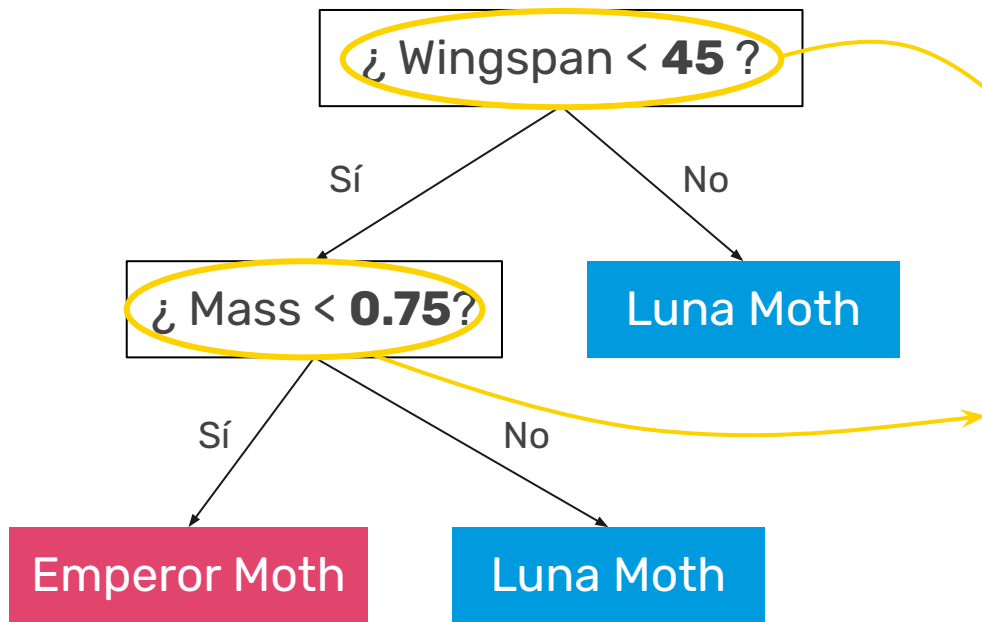


Es una serie de **IFs**...

¿Por qué ahora  
decimos que es  
Machine Learning?

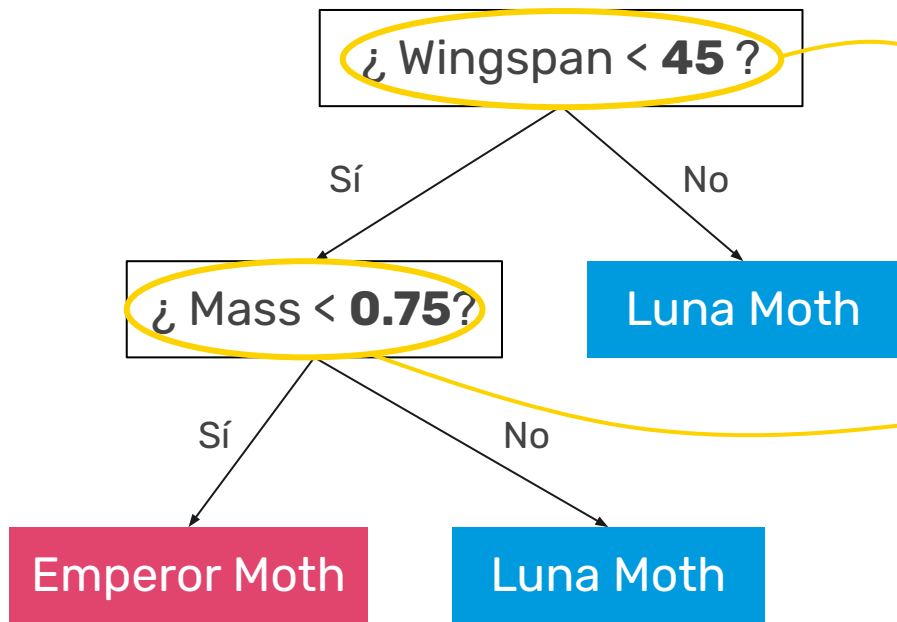


# ¿Por qué decimos que es Machine Learning?



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

# ¿Por qué decimos que es Machine Learning?



En el próximo encuentro vamos a ver **cómo** es que se eligen estos valores a partir de un proceso matemático.

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 fork 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!**

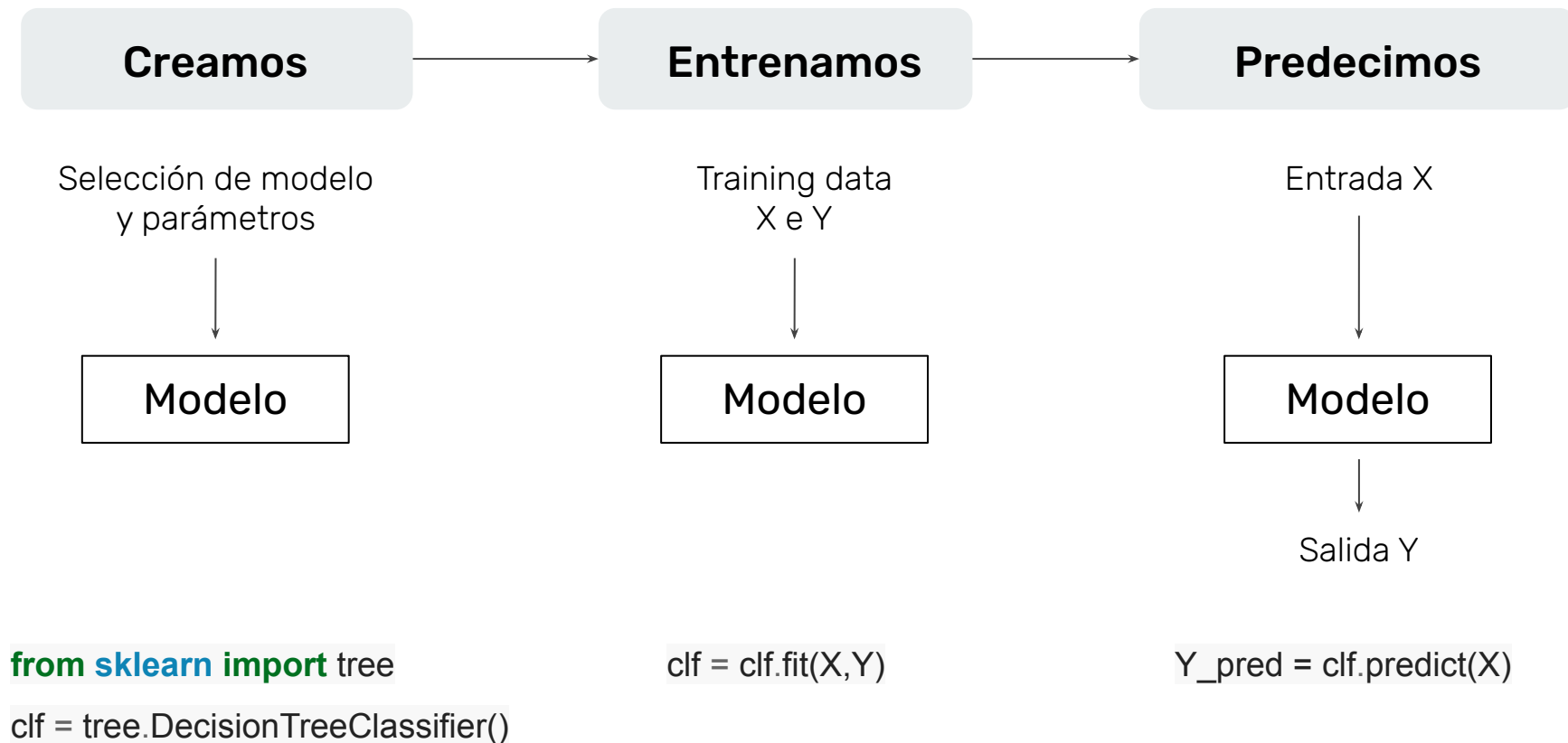
---



# Hands-on training



# Recordamos el flujo de trabajo **Scikit Learn**



**Hands-on  
training**



DS\_Toolbox\_11\_ML.ipynb



PARA DEBATIR

# Automatización y futuro del trabajo



# **¿Crees que hay empleos que se ven afectados por el desarrollo de Machine Learning?**

Elabora una lista con los 5 empleos que más riesgo corren y otros 5 con los que menos se ven afectados (o incluso se potenciados).

# Algunos empleos que ya están siendo reemplazados:

- Traductores
- Radiólogos
- Choferes
- Costureros
- Técnicos en matemáticas



## Las 10 ocupaciones menos amenazadas

1. Terapeutas recreacionales
2. Supervisores de trabajos mecánicos, de instalación y reparación
3. Directores de manejo de emergencias
4. Trabajadores sociales en salud mental y abuso de sustancias
5. Audiólogos
6. Terapistas ocupacionales
7. Expertos en órtesis y prótesis
8. Trabajadores sociales en salud
9. Cirujanos maxilofaciales
10. Supervisores de bomberos y trabajos de prevención

*Fuente: Carl Benedikt Frey y Michael A. Osborne, El futuro del empleo: ¿qué tan susceptibles son nuestros trabajos a la computarización?*

# Recursos



# Seaborn

- [Python Data Science Handbook](#) - Capítulo 5, "Machine Learning" tiene todo lo que necesitas para arrancar. ¡La recomendamos fuertemente!
- La [documentación de Scikit-Learn](#) es excelente. Además, ofrece una muy buena introducción a casi todos los temas de Machine Learning que veremos. En particular, la sección de [Tutoriales](#) puede ser muy útil para arrancar.
- [Video muy interesante](#) sobre el impacto de Machine Learning (Automatización moderna) en la sociedad.

# Para la próxima

---

1. Termina el notebook de hoy.
2. Lee la Toolbox 12.

ACÀMICA