

DATA SCIENCE



Federico Baiocco
baioccofed@gmail.com
3512075440



Clase 16 - Agenda

INTRODUCCIÓN A MACHINE LEARNING

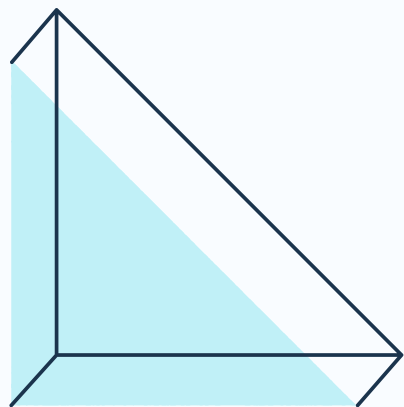
- Machine learning
- Supervised learning
- Unsupervised learning
- Decision tree

¿ Dudas de la clase pasada ?

¿ Todos pudieron terminar ?



Hasta ahora



Python

Estuvimos aprendiendo python, la herramienta principal que vamos a utilizar como data scientists

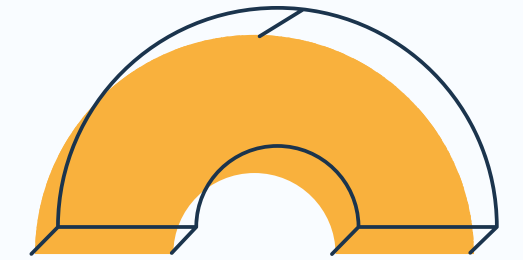
Librerías

Aprendimos a utilizar librerías para procesar y visualizar datos en python. Trabajando siempre en entornos de jupyter notebooks

Conceptos

Vimos conceptos como medidas de resumen, correlación entre variables y detección de outliers



Machine learning - Ejemplo




Un ejemplo típico.. Queremos clasificar emails como spam o no-spam.

Hey


Spam x



 **Amber Zimmerman** <lawerence.kamps@nokiamail.com>
para japalex

21 feb (hace 5 días) ☆  

 ¿Por qué este mensaje está en la carpeta de spam? Hemos visto que muchos mensajes de nokiamail.com son spam. [Más información](#)

 inglés ▾ > español ▾ [Traducir mensaje](#) [Desactivar para: inglés](#) x

hi , I'm Marina, from Russia. I would like to find a buddy for serious relations.
I'm seriously interested in learning more about you. Dating is not a joke for me, I don't play games.
I wish that you are serious too. I'm a lonely woman who wants to find someone who will love and understand me.
I'll tell you more about myself. I assume that I am a strong woman with goals, principles and ambitions.
If my message interested you, please reply
Have a happy day!
I'll be waiting for your letter!

¿ Cómo distinguen si
el mail es spam o no?



Machine learning - Ejemplo



¿ Cómo haríamos para decir si el mail es spam o no?

Detectar un mail como spam o no, no es una tarea fácil ya que requiere saber leer, relacionar conceptos, etc. Por todo esto, es casi imposible hacer un programa con reglas "duras" para este tipo de tareas.

```
if mail contains "serious relations":
```

```
    SPAM
```

```
elif mail contains "buddy":
```

```
    SPAM
```

```
else:
```

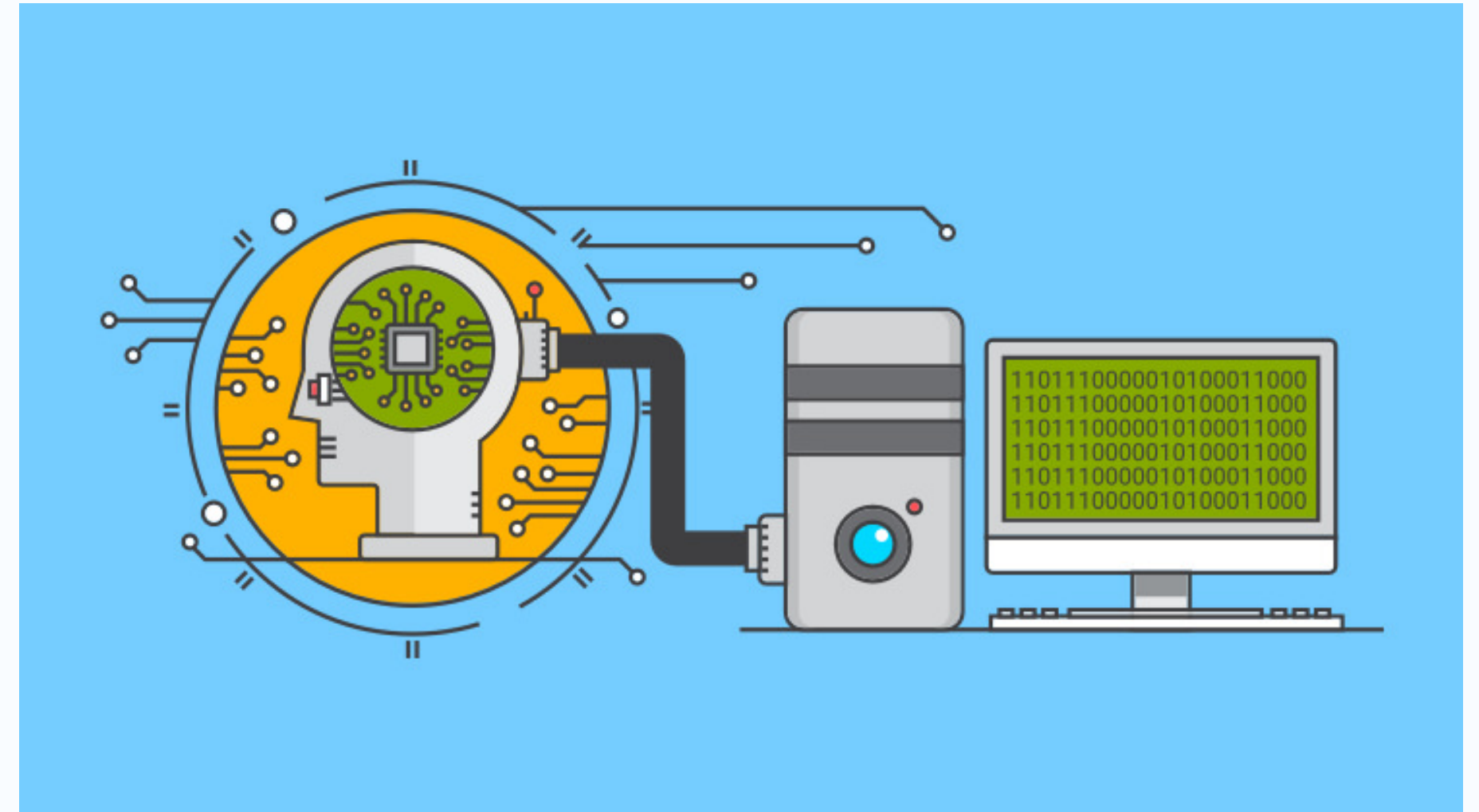
```
    NO-SPAM
```

Se imaginan? ...

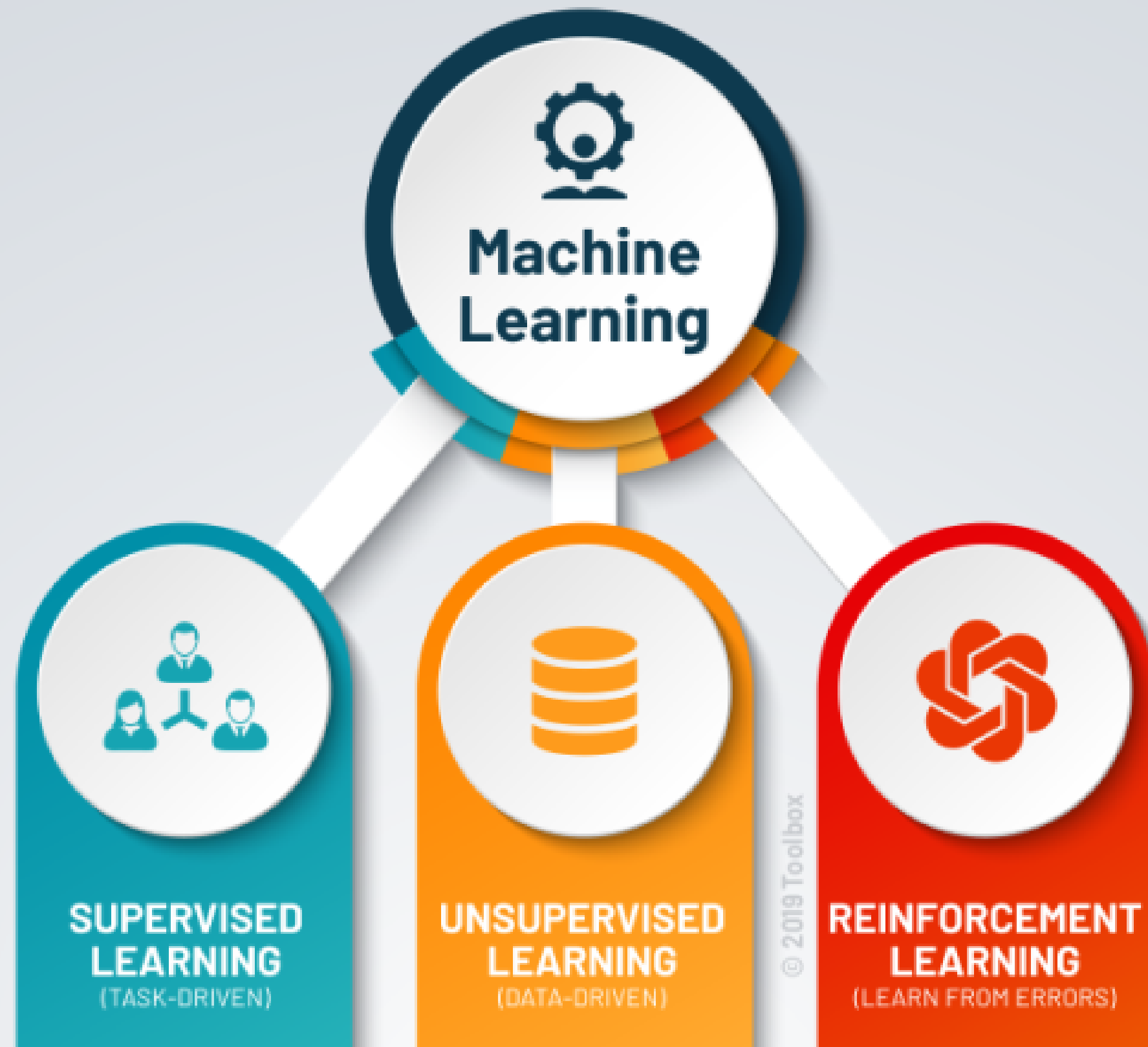
Machine learning

Un algoritmo con reglas "duras" como el anterior sería muy poco útil ya que es casi imposible definir todas las reglas de forma manual. Eso definitivamente NO es machine learning.

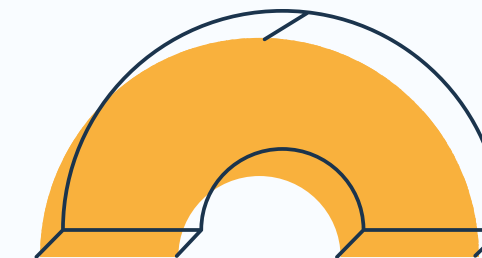
Un modelo de machine learning aprende a predecir a partir de datos.



TYPES OF MACHINE LEARNING



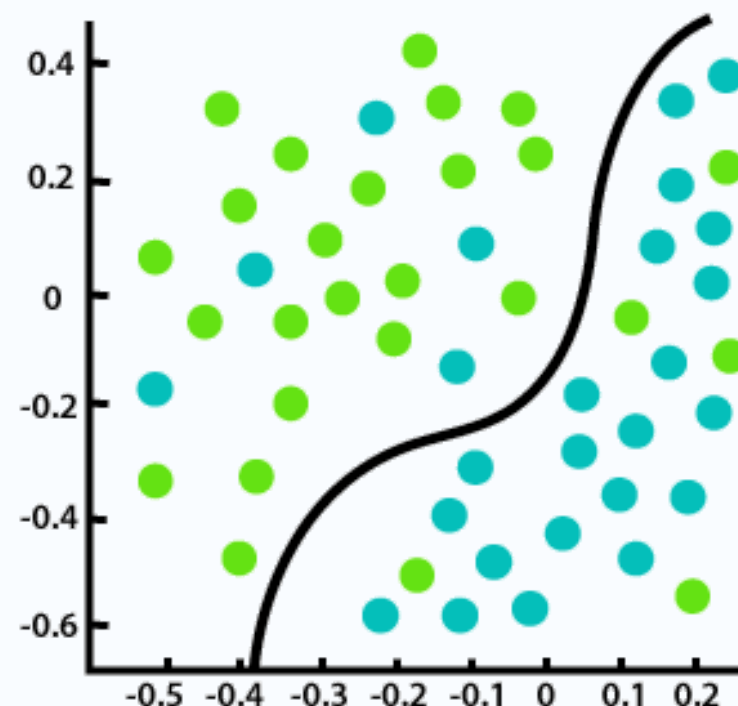
Aprendizaje supervisado



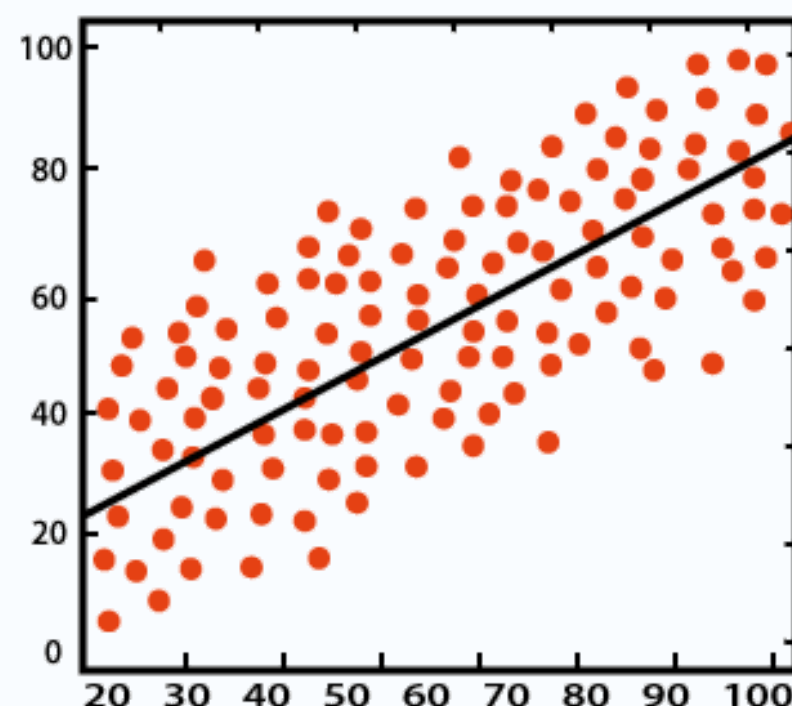
Vamos a empezar por aprendizaje supervisado.

En aprendizaje supervisado, tenemos X (features), y (variable a predecir) y buscamos conseguir un modelo ($f(X) = y$) que nos sirva para predecir sobre nuevos datos de X de los cuales no conocemos la salida.

Tenemos 2 tipos de problemas:



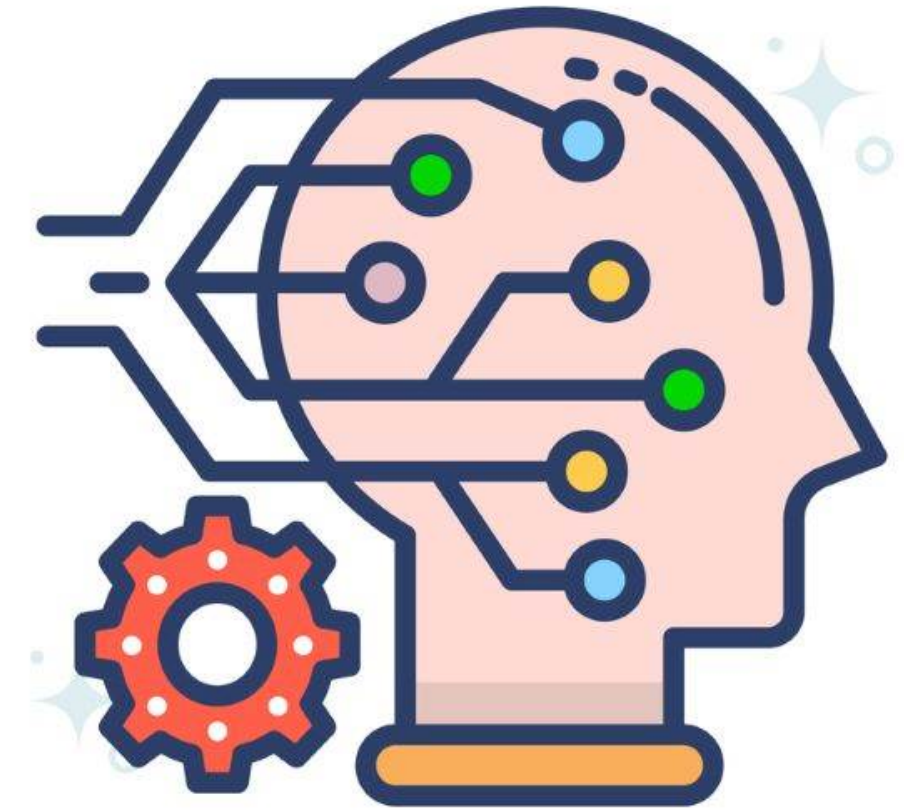
Classification



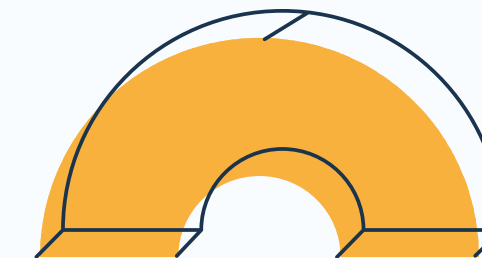
Regression

Aprendizaje supervisado

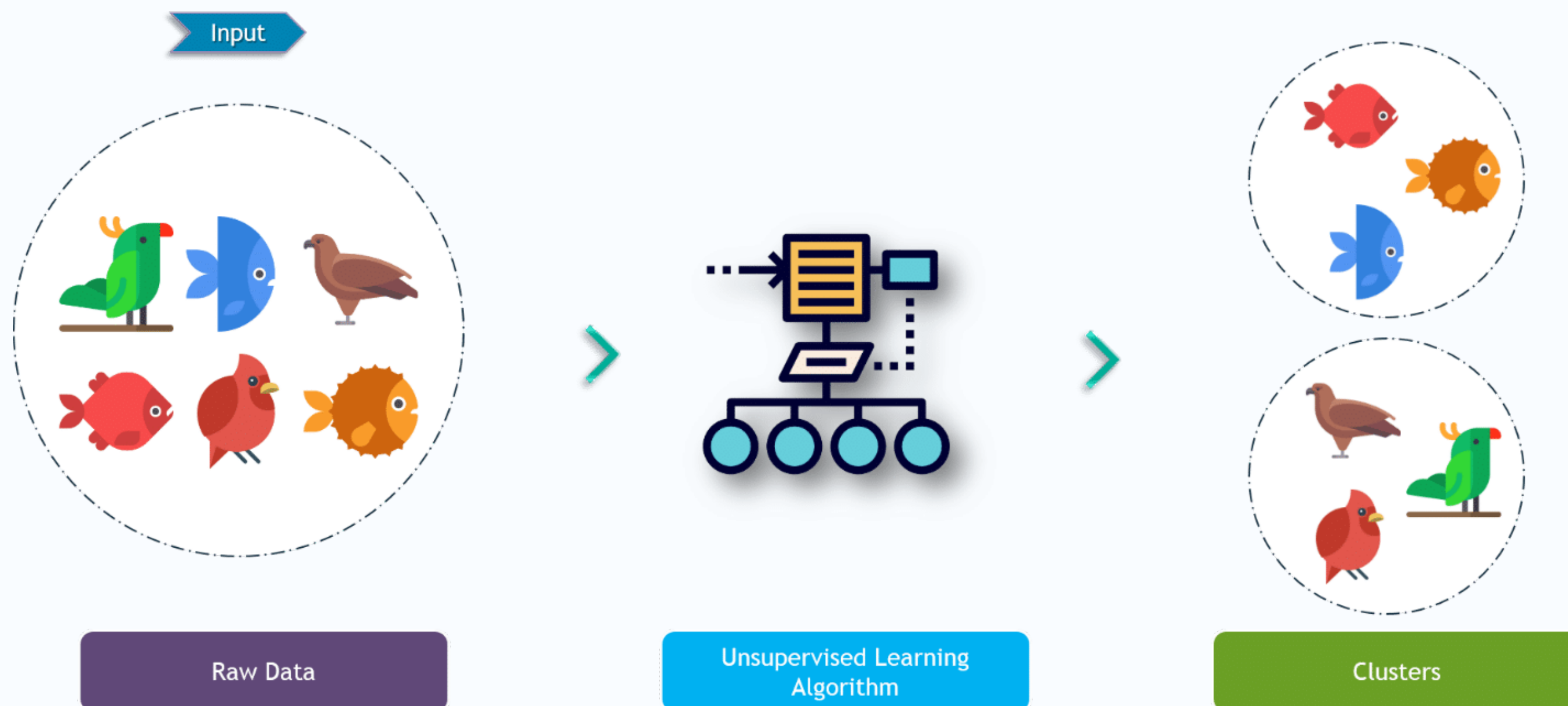
- Decision trees
- SVM
- KNN
- Random forest
- Redes neuronales
-



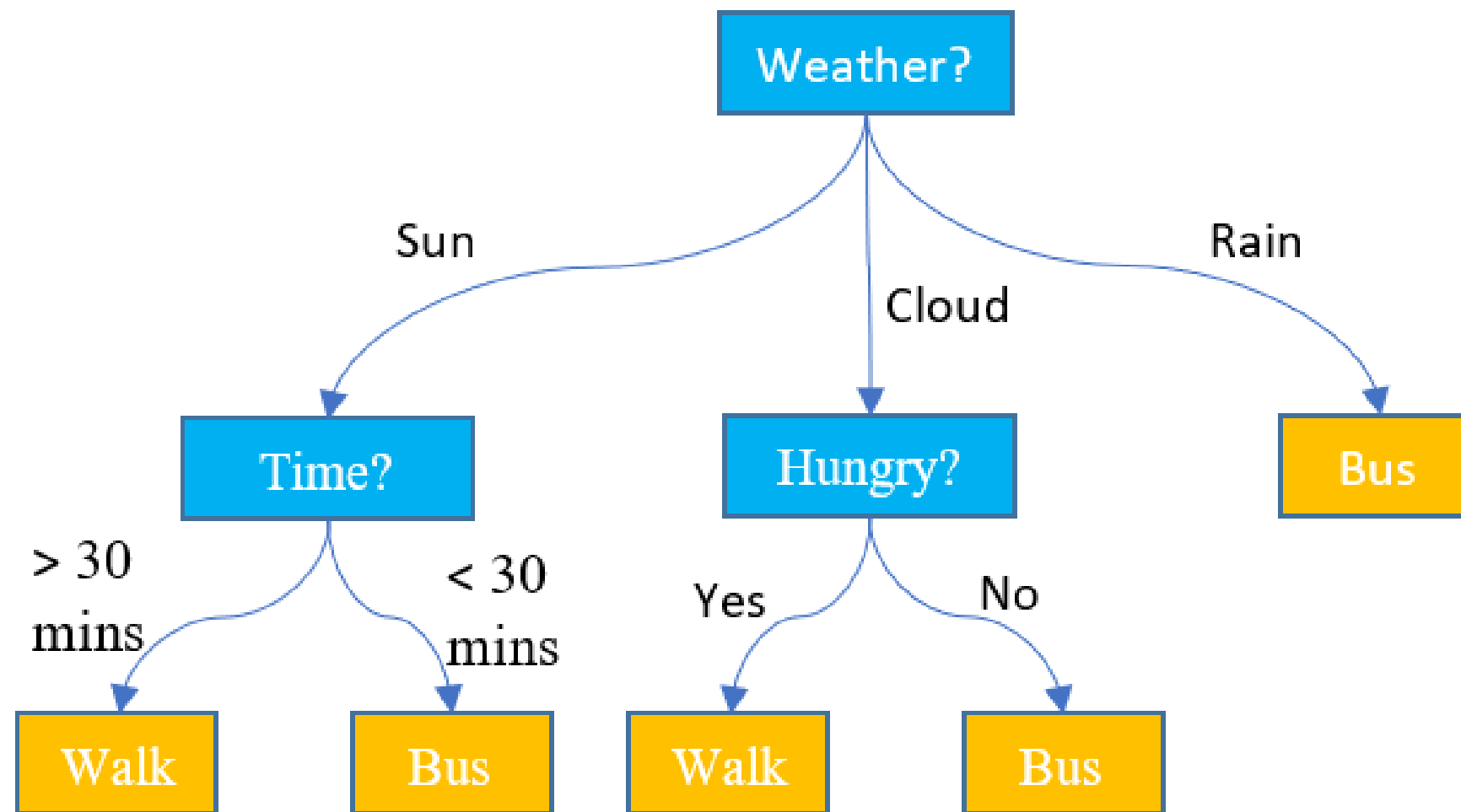
Aprendizaje no supervisado



En este caso, solo tenemos "features" X pero no tenemos ningún "y" que queramos predecir. Lo que se busca, es encontrar patrones en los datos:



Decision tree



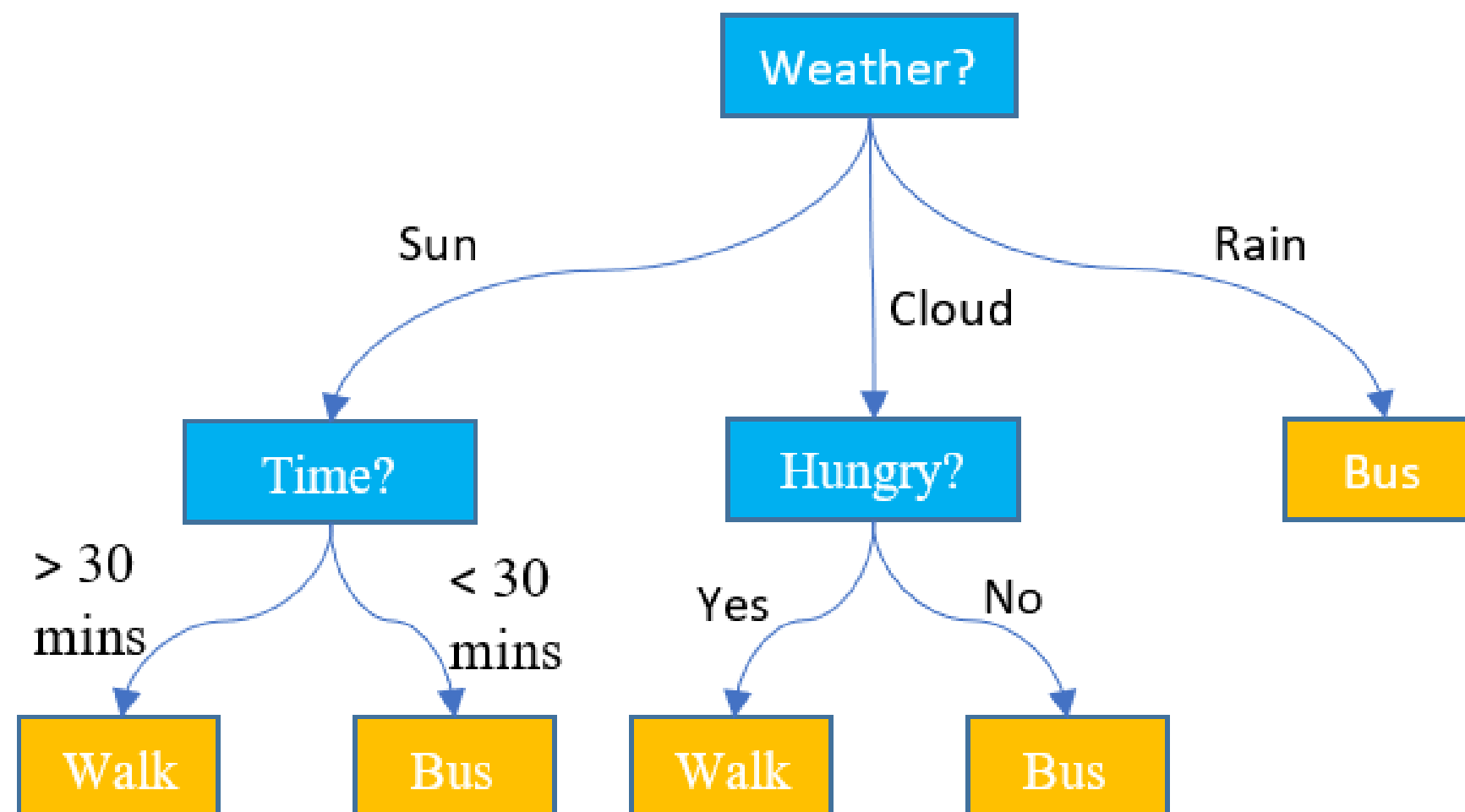
Decision trees

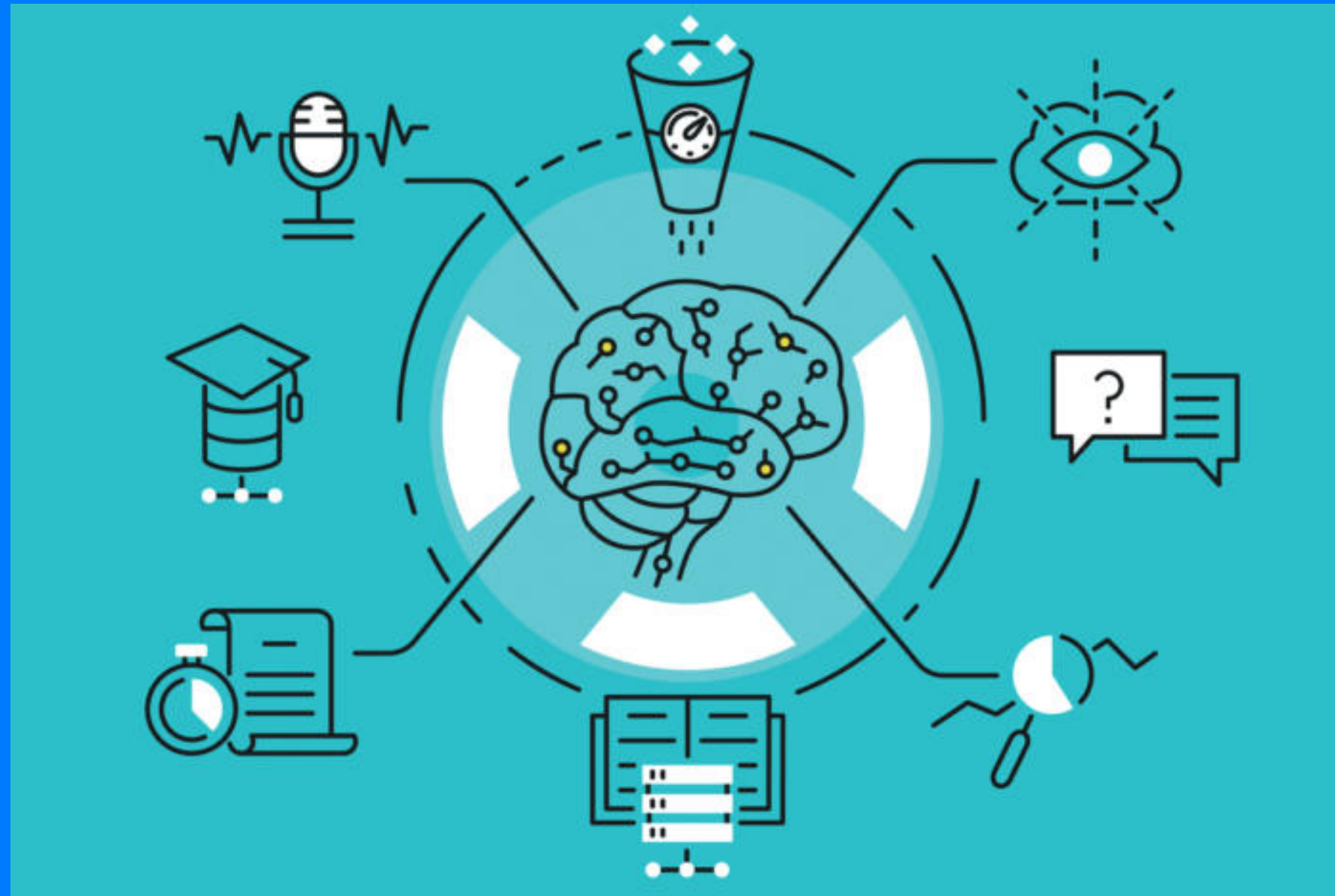


Vemos que son un conjunto de IFs encadenados



¿ Por qué decimos que esto SI es machine learning ?





**Abrimos notebook
"intro machine learning "**