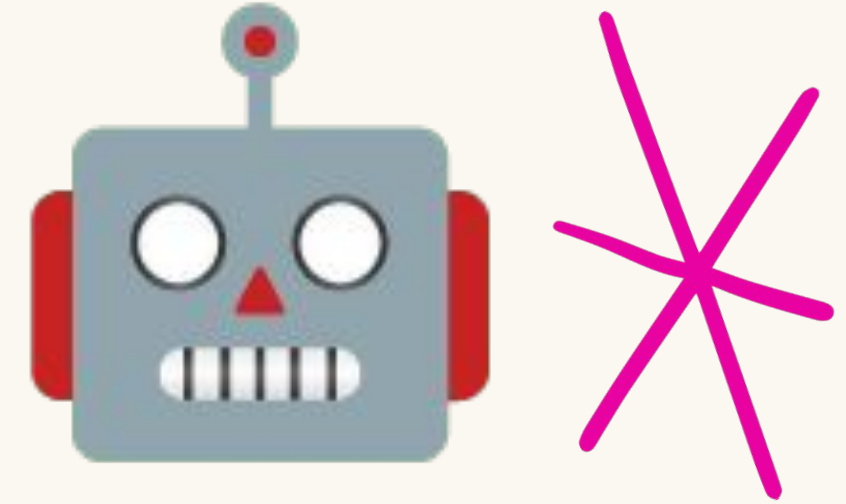


Desbloqueando el Poder de los Datos

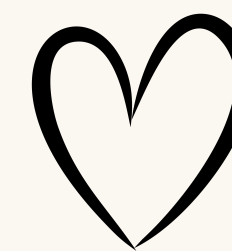


Inteligencia Artificial & Ciencia de Datos para todos

Comenzamos a las 7:05 a.m. en punto.

¿Te gustaría comenzar el día con alguna canción en específico?

Coméntala en el chat 🎵💬



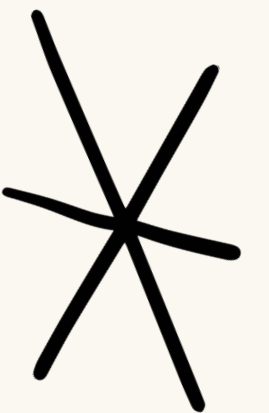
Pasos en un proyecto de machine learning

Septiembre 3, 2024

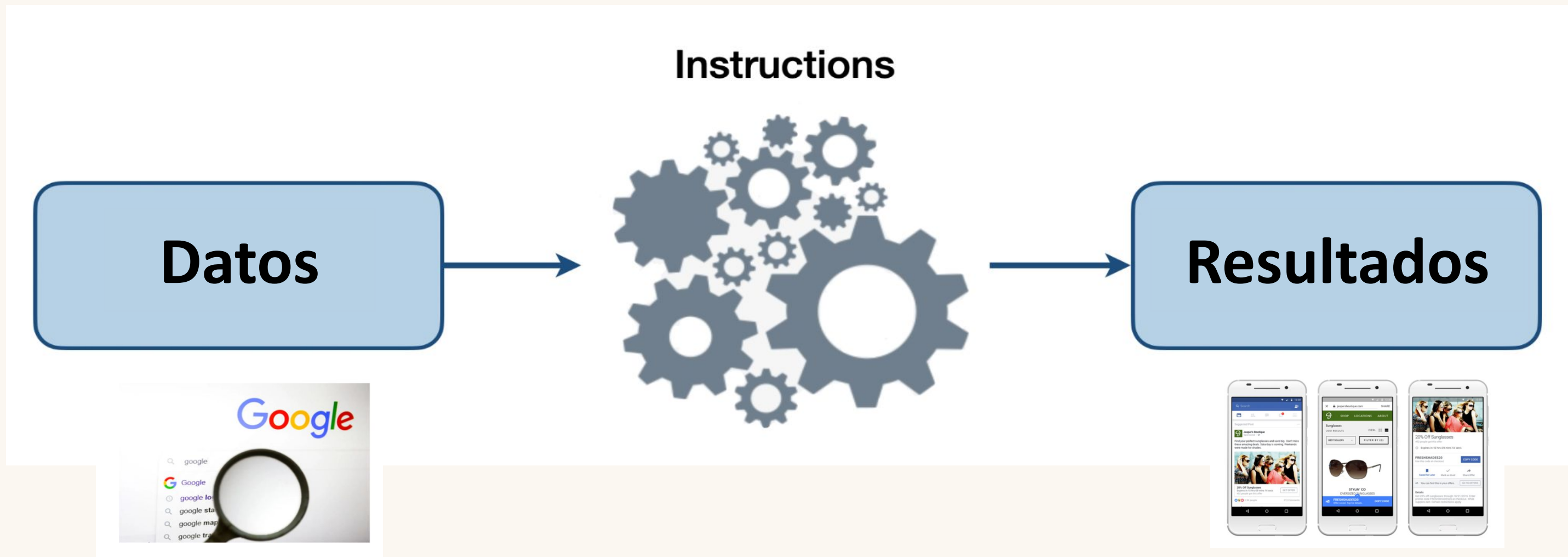




1. Repaso de la última clase
2. Tema de hoy:
 - ¿Qué es machine learning?
 - Diferencia entre ML vs DL vs AI
 - Partes de un modelo de ML
 - Familias de modelos de ML
 - Pasos de un proyecto de ML
 - Ejemplos
3. Logística de la clase



Un **algoritmo informático** es un conjunto de instrucciones que permiten que el computador genere resultados



Los **lenguajes de programación** nos permiten comunicarnos con el computador



Programación en Python parte I

https://colab.research.google.com/drive/14i3bmnxjpdEHmjWWNo_iSKF336Dv3XtC

- ❑ ¡Hola Mundo!
- ❑ Variables: Nombre y lugar de almacenamiento de un dato
- ❑ Tipos de datos básicos en Python
 - ❑ `int`: Números enteros
 - ❑ `float`: Números decimales
 - ❑ `boolean`: Valor verdadero (`True`) o falso (`False`)
 - ❑ `string`: Texto
- ❑ Aritmética
- ❑ Comparaciones lógicas
- ❑ Operadores lógicos

Programación en Python parte 2

<https://colab.research.google.com/drive/1Jlv5nKz6sDaSVsbyjosM14eo14ObNoXn>

- ❑ Cadenas de texto

- ❑ Estructuras de datos

 - ❑ `list`: Listas `[]` Versátil

 - ❑ `tuple`: Tuplas `()` Una vez creadas, no pueden ser modificadas.

 - ❑ `set`: Conjuntos `{ }` Cada elemento es único. No hay orden.

 - ❑ `dict`: Diccionarios `{ k:v, k:v, ..., k:v }` Parejas de llaves y valores. Las llaves son únicas.

- ❑ Condicionales

 - ❑ `if`: si

 - ❑ `elif`: sino, si

 - ❑ `else`: sino

Programación en Python parte 3

<https://colab.research.google.com/drive/1wjEPKGQWJMrZ2r-lm-KbnOCm3ZXBxZlR>

❑ Bucles

- ❑ For (**Para** – Se utiliza para repetir código un **número concreto** de veces.)
- ❑ While (**Mientras que** – Se utiliza para repetir código **hasta** que una condición se cumpla.)
- ❑ Compresión de listas

❑ Funciones

❑ Librerías

- ❑ Se instalan con ! **pip install nombre_libreria**

El nombre que le quieras dar a tu función

Datos de entrada
Separados por comas, dentro del parentesis

Así le decimos a Python que esto es una función

Para crear la función

```
def mi_funcion(dato1, dato2, dato3):
```

```
# código con instrucciones
```

```
# más instrucciones
```

```
return resultado
```

Una tabulación para indicar que esas líneas de código pertenecen a la función.

El (los) resultado(s) que queremos retornar. Debe ir antecedido por la palabra `return`

Para usar la función ya creada

`mi_funcion("hola", 3141592, [23,7])`

El nombre que le
diste a tu función

Datos de entrada
Separados por comas, dentro del parentesis



Notebook de hoy

Respuestas del taller anterior:

https://colab.research.google.com/drive/1rm91WGUJNv_hRn77BE3qj2heKA2t4zLt?usp=sharing



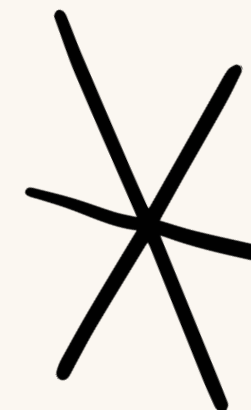


1. Repaso de la última clase 

2. Tema de hoy:

- ¿Qué es machine learning?
- Diferencia entre ML vs DL vs AI
- Partes de un modelo de ML
- Familias de modelos de ML
- Pasos de un proyecto de ML
- Ejemplos

3. Logística de la clase



🤔 ¿Cómo le enseñamos a un niño lo que es un gato?





🤔 ¿Cómo le enseñamos a un niño lo que es un gato?

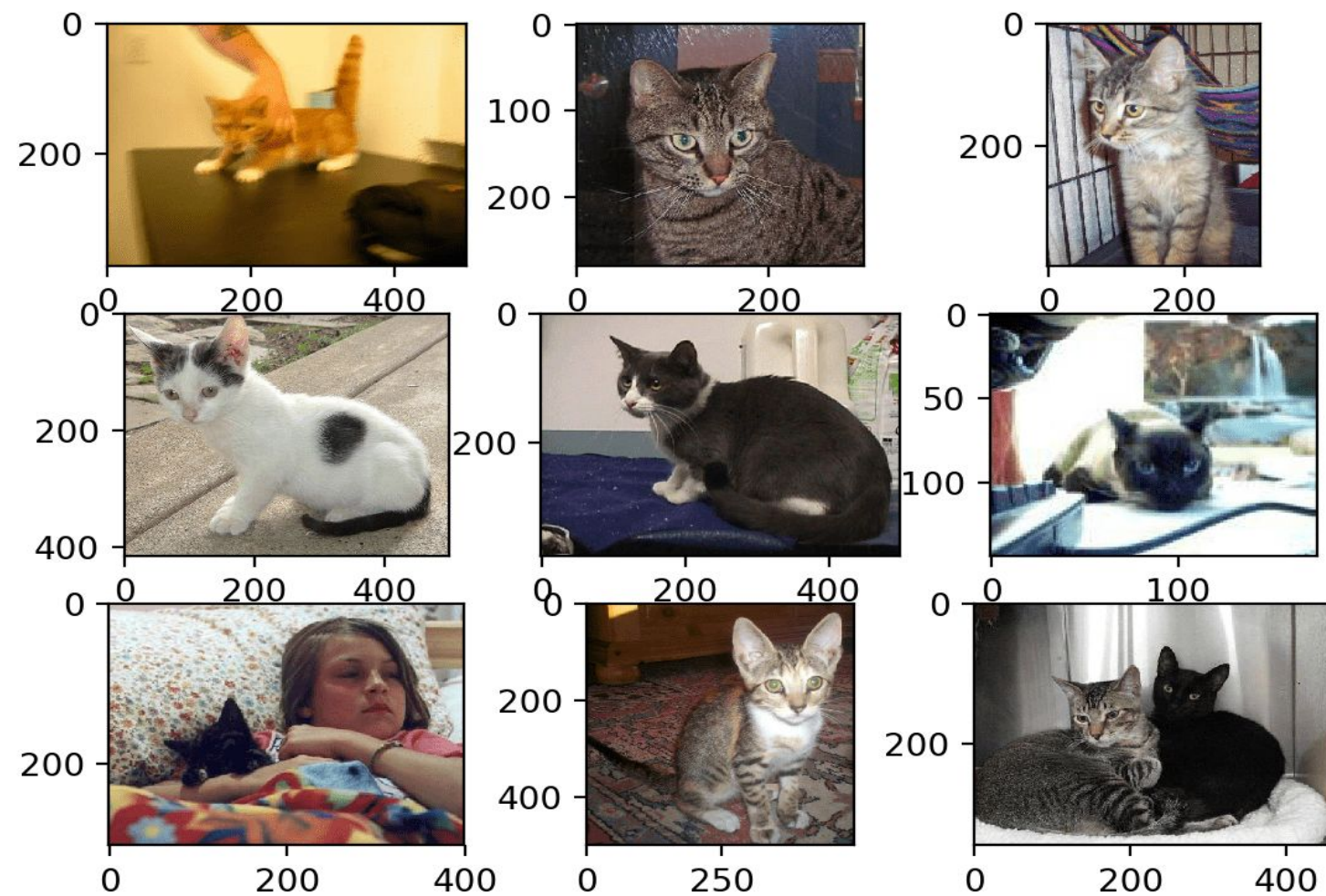


¿Escribes una lista de reglas como "los gatos tienen cuatro patas" y "los gatos tienen dos orejas"?

ó

¿Le muestras al niño muchas fotos de diferentes gatos

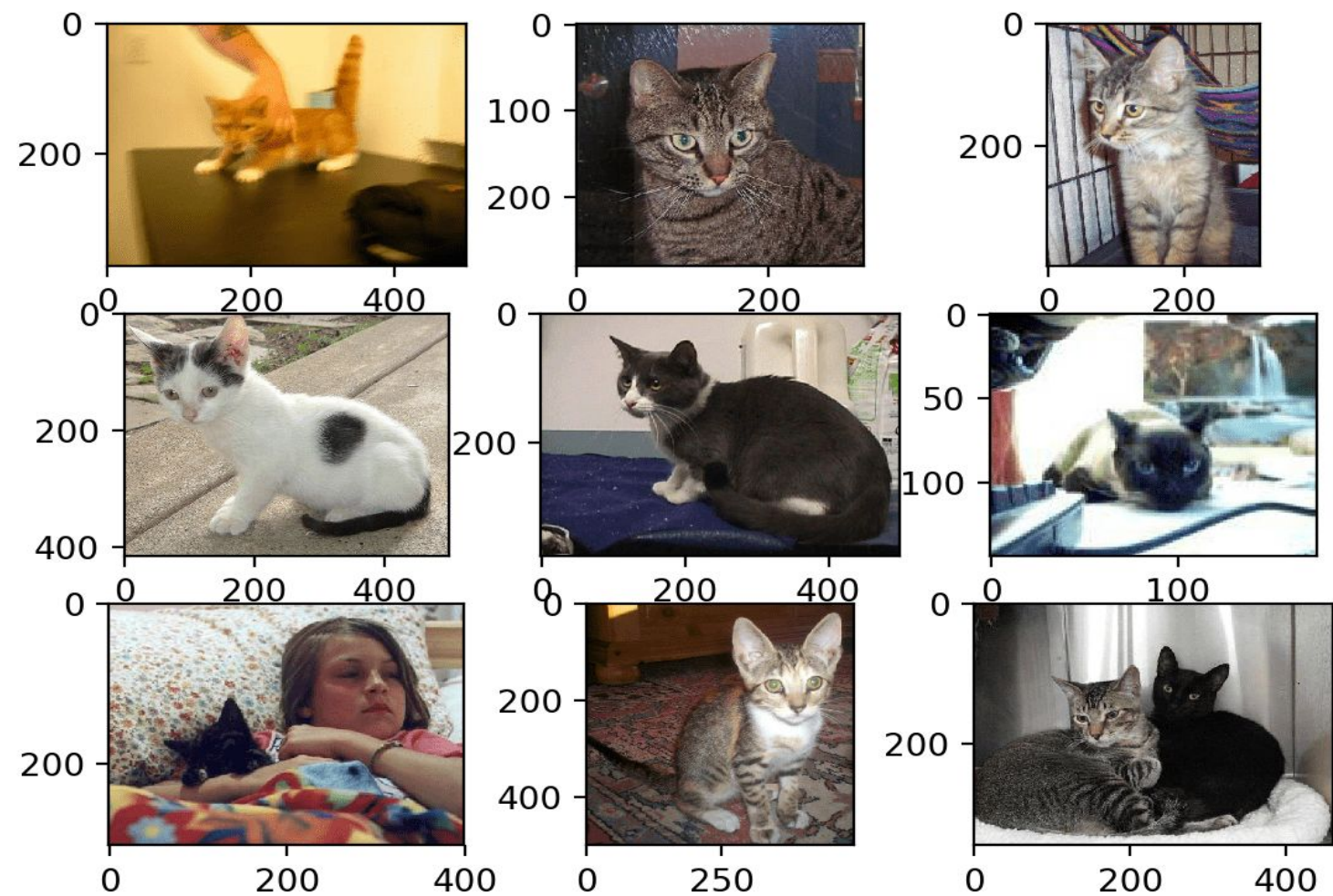
¿Qué es machine learning? (Aprendizaje automático)



El **machine learning** es la misma idea.

En lugar de programar una computadora con un conjunto específico de instrucciones explícitas para realizar una tarea, le proporcionamos a la computadora una gran cantidad de datos y dejamos que aprenda las reglas a partir de esos datos.

¿Qué es machine learning? (Aprendizaje automático)



El **machine learning** es la misma idea.

En lugar de programar una computadora con un conjunto específico de instrucciones explícitas para realizar una tarea, le proporcionamos a la computadora una gran cantidad de datos y dejamos que aprenda las reglas a partir de esos datos.

Al igual que un niño, cuanto más ejemplos tenga la computadora para aprender, ¡mejor será en esa tarea!

¿Qué es machine learning? (Aprendizaje automático)

Otro ejemplo



- Imaginemos que vamos a comprar un café grande en Juan Valdez por \$5.000
- Sabemos que si compramos uno para un amigo también, nos costará \$10.000
- ¿Cómo podrías "predecir" el costo total para n número de amigos?

¿Qué es machine learning? (Aprendizaje automático)

Otro ejemplo



- Imaginemos que vamos a comprar un café grande en Juan Valdez por \$5.000
- Sabemos que si compramos uno para un amigo también, nos costará \$10.000
- ¿Cómo podrías "predecir" el costo total para **n** número de amigos?

total = 5.000 + 5.000 x (número de amigos)

$$y = \beta_0 + \beta_1 X_1$$

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



Artificial Intelligence

The diagram consists of three overlapping circles. The largest circle is pink and contains the text 'Artificial Intelligence'. Inside it is a blue circle containing the text 'Machine Learning'. Inside the blue circle is a yellow circle containing the text 'Deep Learning'. Each circle has a corresponding label in a white box: 'Inteligencia Artificial' for the pink circle, 'Aprendizaje automático' for the blue circle, and 'Aprendizaje profundo' for the yellow circle.


Inteligencia Artificial

Machine Learning

Aprendizaje automático

Deep Learning

Aprendizaje profundo



Inteligencia Artificial

Definición: La inteligencia artificial (IA) se refiere al desarrollo de sistemas informáticos que pueden realizar tareas que normalmente requieren inteligencia humana.

Alcance: Amplio. La IA abarca todo lo que permite a las computadoras imitar la inteligencia humana, incluyendo la robótica, el procesamiento del lenguaje natural y la resolución de problemas.



Artificial Intelligence

The diagram consists of three overlapping circles. The largest circle is pink and contains the text 'Artificial Intelligence' and 'Inteligencia Artificial'. Inside it is a blue circle containing 'Machine Learning' and 'Aprendizaje automático'. Inside the blue circle is a yellow circle containing 'Deep Learning' and 'Aprendizaje profundo'.

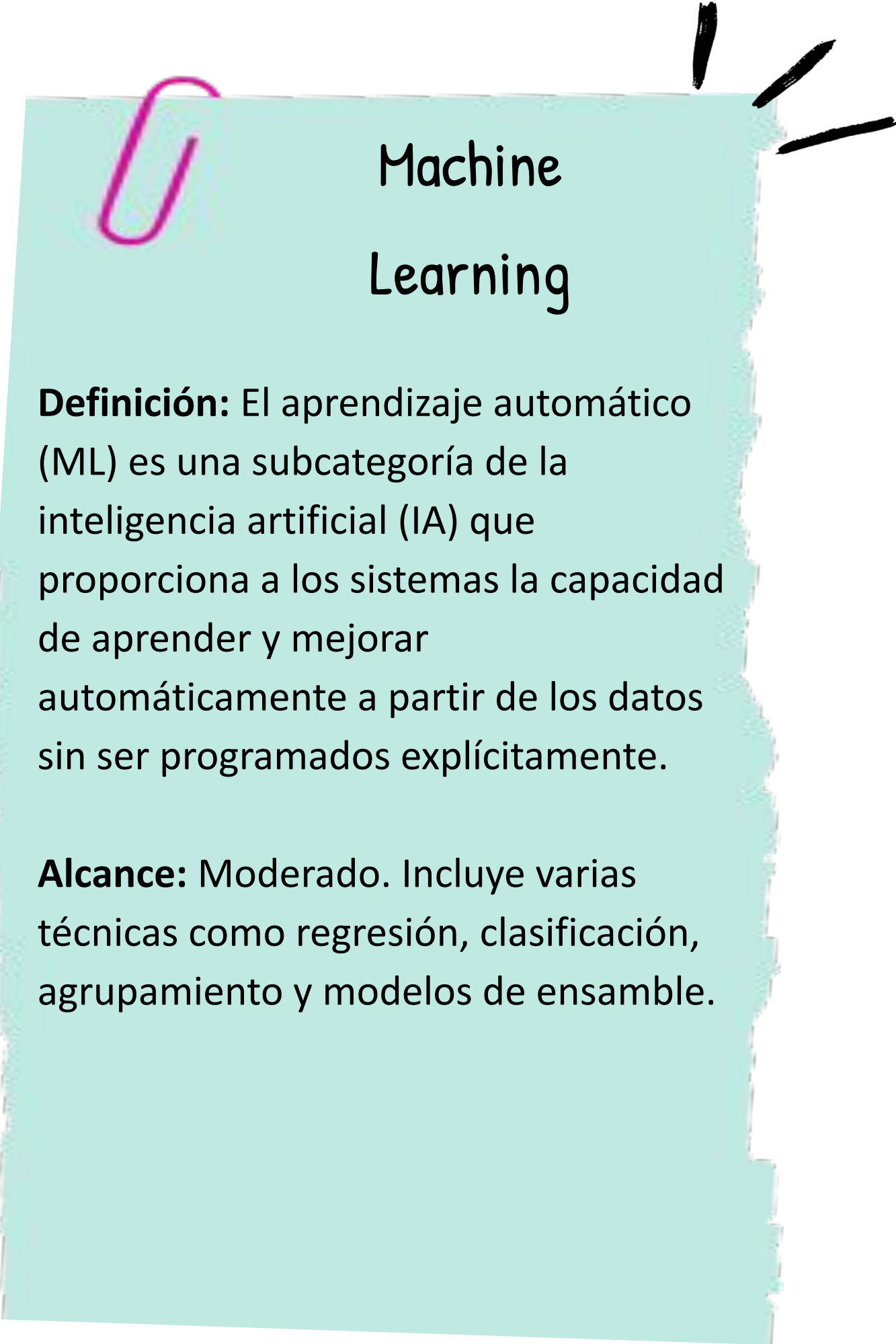
Inteligencia Artificial

Machine Learning

Aprendizaje automático

Deep Learning

Aprendizaje profundo



Machine Learning

Definición: El aprendizaje automático (ML) es una subcategoría de la inteligencia artificial (IA) que proporciona a los sistemas la capacidad de aprender y mejorar automáticamente a partir de los datos sin ser programados explícitamente.

Alcance: Moderado. Incluye varias técnicas como regresión, clasificación, agrupamiento y modelos de ensamble.



Artificial Intelligence

The diagram consists of three overlapping circles. The largest circle is pink and contains the text 'Artificial Intelligence' and 'Inteligencia Artificial'. Inside it is a blue circle containing 'Machine Learning' and 'Aprendizaje automático'. Inside the blue circle is a yellow circle containing 'Deep Learning' and 'Aprendizaje profundo'.

Inteligencia Artificial

Machine Learning

Aprendizaje automático

Deep Learning

Aprendizaje profundo



Deep Learning

Definición: El aprendizaje profundo (DL) es una subcategoría del aprendizaje automático (ML) que utiliza redes neuronales. Estas redes intentan simular el comportamiento del cerebro humano—aunque aún lejos de igualar su capacidad—para "aprender" a partir de grandes cantidades de datos.

Alcance: Reducido. DL es una forma específica, pero poderosa, de aprendizaje automático.



Artificial Intelligence

The diagram consists of three overlapping circles. The largest circle is pink and contains the text 'Artificial Intelligence' and 'Inteligencia Artificial'. Inside it is a blue circle containing 'Machine Learning' and 'Aprendizaje automático'. Inside the blue circle is a yellow circle containing 'Deep Learning' and 'Aprendizaje profundo'.

Inteligencia Artificial

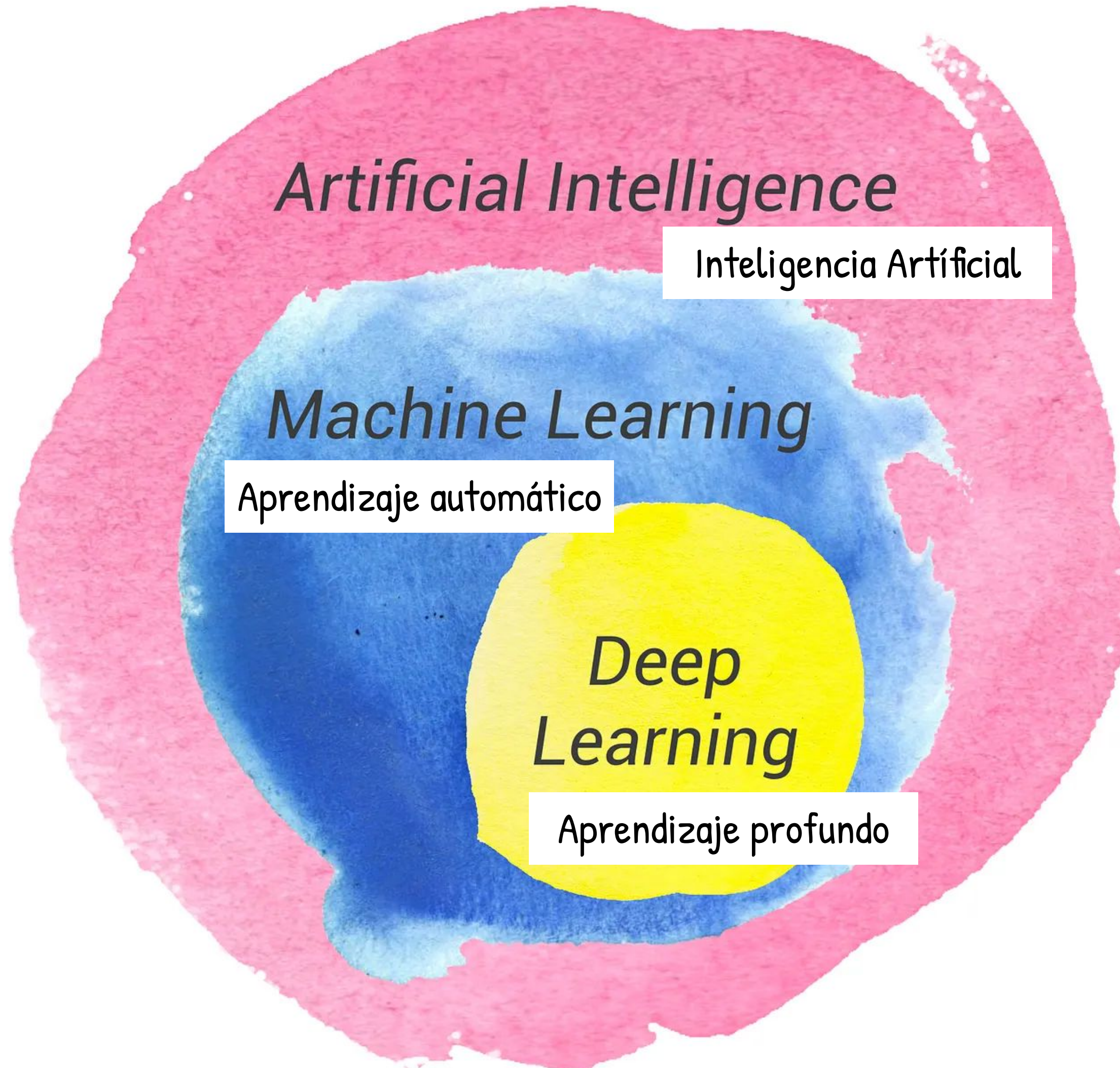
Machine Learning

Aprendizaje automático

*Deep
Learning*

Aprendizaje profundo

En este curso nuestro
principal enfoque será
machine learning
(aprendizaje automático)



En este curso nuestro principal enfoque será **machine learning** (aprendizaje automático)

Ventas				
Fecha	Producto	Potencia	Unidades	Ganancias
enero	Bicicletas	Eléctrica	476	\$751.604
enero	Bicicletas	Manual	302	\$581.350
enero	Motonetas	Eléctrica	387	\$427.248
enero	Motonetas	Manual	309	\$48.513
enero	Patinetas	Eléctrica	251	\$135.791
febrero	Bicicletas	Eléctrica	354	\$558.966
febrero	Bicicletas	Manual	219	\$336.165
febrero	Motonetas	Eléctrica	312	\$583.128
febrero	Motonetas	Manual	419	\$396.793

Partes de un modelo de Machine Learning

iris setosa



petal sepal

iris versicolor



petal sepal

iris virginica



petal sepal

Partes de un modelo de Machine Learning

In [4]:

```
import seaborn as sns
df = sns.load_dataset('iris')
df.head()
```

Out[4]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Partes de un modelo de Machine Learning

Entradas del Modelo

Son las variables o datos de entrada que se utilizan para hacer predicciones

También conocidas como:

- Input
- Características (Features)
- Atributos
- Predictores
- Entradas
- Variables independientes
- Dimensiones
- X
- Probablemente más...

In [4]:

```
import seaborn as sns
df = sns.load_dataset('iris')
df.head()
```

Out [4]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Partes de un modelo de Machine Learning

Salidas del Modelo

Son los valores o resultados que el modelo intenta predecir a partir de los datos de entrada

También conocidas como:

- Output
- Objetivo
- Respuesta
- Target
- Salida
- Variable dependiente
- Etiquetas
- Y
- Probablemente más...

In [4]:

```
import seaborn as sns
df = sns.load_dataset('iris')
df.head()
```

Out [4]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Partes de un modelo de Machine Learning

Fila de datos (Input + Output)

Cada fila representa una observación o un caso específico dentro del conjunto de datos

También conocida como:

- Observación
- Punto de datos
- Registro
- Fila
- Probablemente más...

In [4]:

```
import seaborn as sns
df = sns.load_dataset('iris')
df.head()
```

Out [4]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Partes de un modelo de Machine Learning

Etiquetas (en el contexto del aprendizaje supervisado)
Son los valores de las variables objetivo que el modelo intenta predecir

En este caso específico
las etiquetas son:

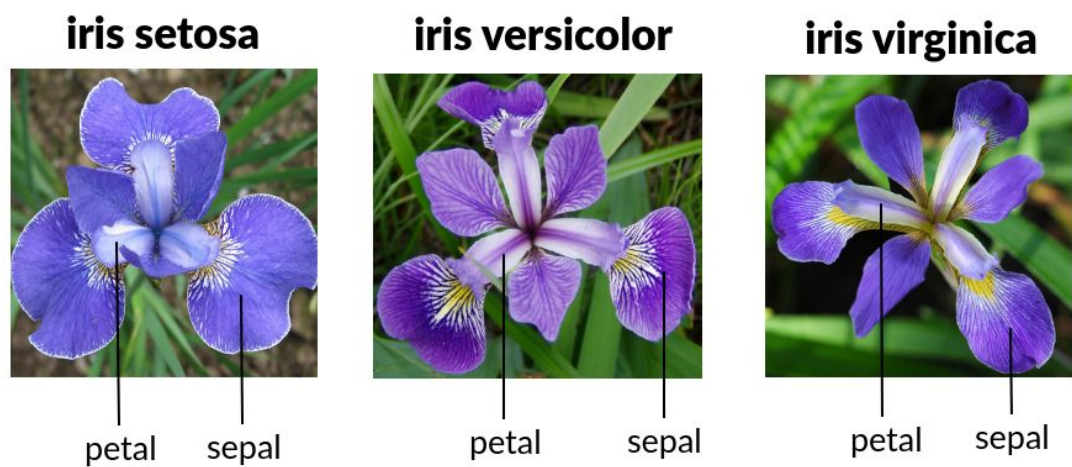
- Setosa
- Versicolor
- Virginica

In [4]:

```
import seaborn as sns
df = sns.load_dataset('iris')
df.head()
```

Out [4]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa



Familia de modelos de Machine Learning

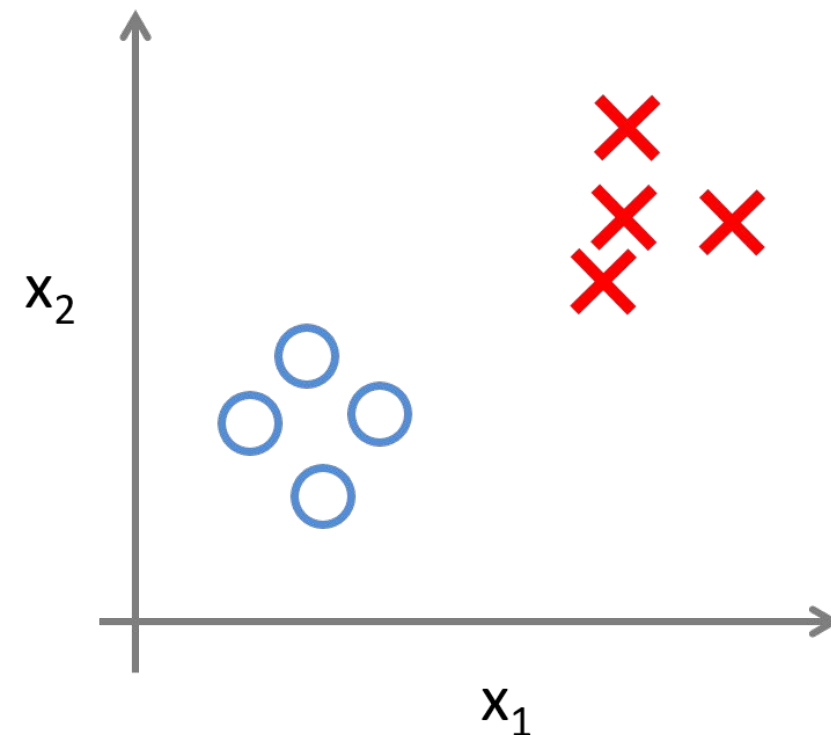
¿Qué herramienta usarías para clavar el clavo?



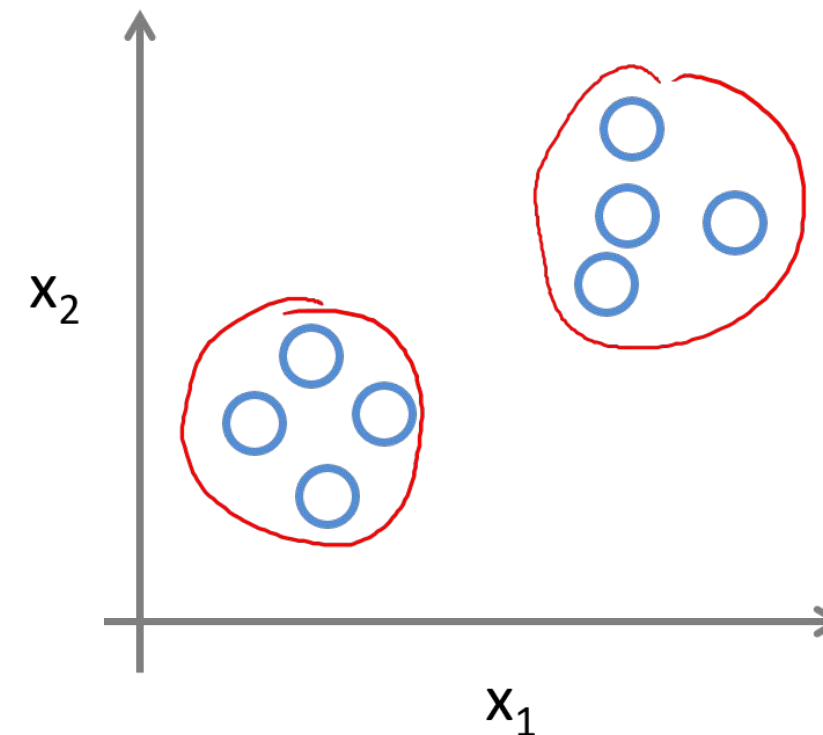
Familia de modelos de Machine Learning

Aprendizaje supervisado vs Aprendizaje no supervisado

Aprendizaje supervisado



Aprendizaje no supervisado



- **Aprendizaje supervisado:** Problemas con etiquetas
- **Aprendizaje no supervisado:** Problemas sin etiquetas



Punto de control

¿Supervisado ó no supervisado?

1. Predecir los precios de las casas en función de características como la ubicación, el tamaño y el número de habitaciones.
1. Agrupar a los clientes con comportamientos de compra similares para realizar campañas de mercadeo dirigidas.



Punto de control

¿Supervisado ó no supervisado?

1. Predecir los precios de las casas en función de características como la ubicación, el tamaño y el número de habitaciones.

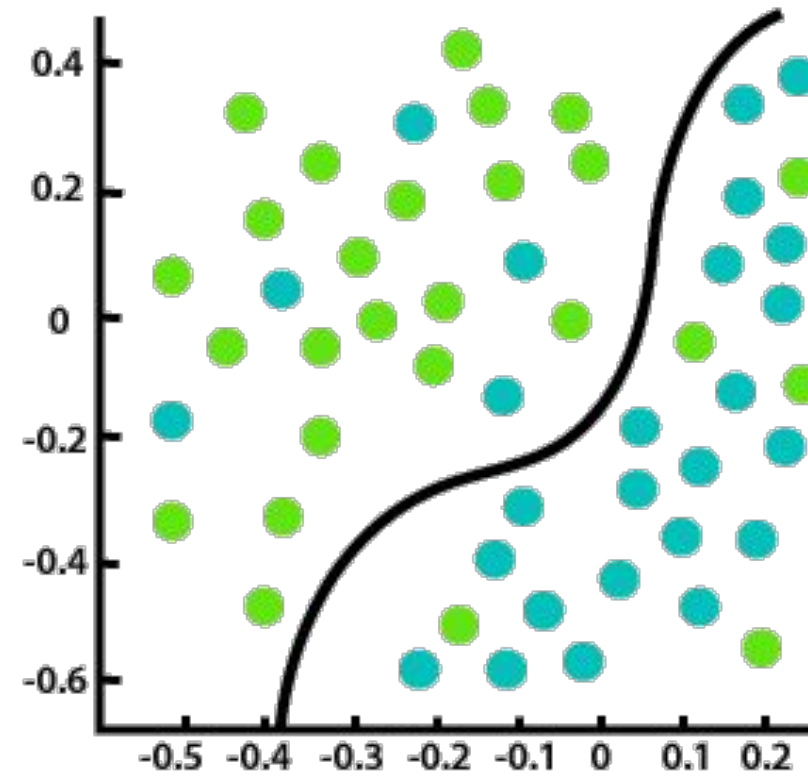
SUPERVISADO

1. Agrupar a los clientes con comportamientos de compra similares para realizar campañas de mercadeo dirigidas.

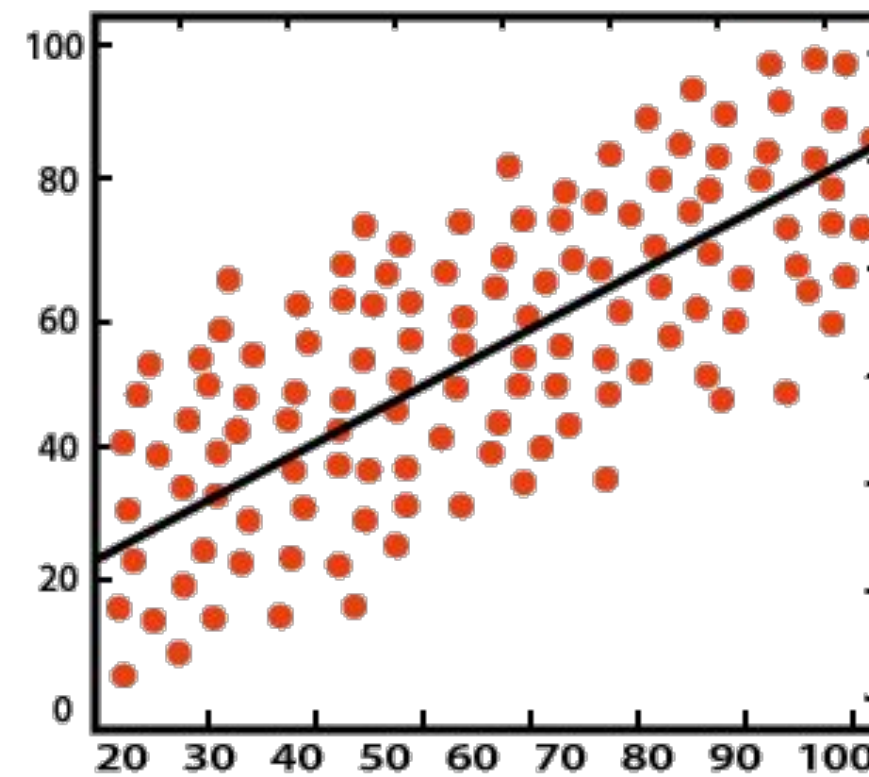
NO SUPERVISADO

Familia de modelos de Machine Learning

Regresión vs Clasificación



Clasificación



Regresión

- **Regresión:** Variable objetivo cuantitativa
- **Clasificación:** Variable objetivo cualitativa



Punto de control

¿Regresión ó clasificación?

1. Predecir si un paciente tiene una enfermedad particular o no en función de sus síntomas.
1. Estimar el salario de una persona en función de su nivel educativo y su experiencia laboral.



Punto de control

¿Regresión ó clasificación?

1. Predecir si un paciente tiene una enfermedad particular o no en función de sus síntomas.

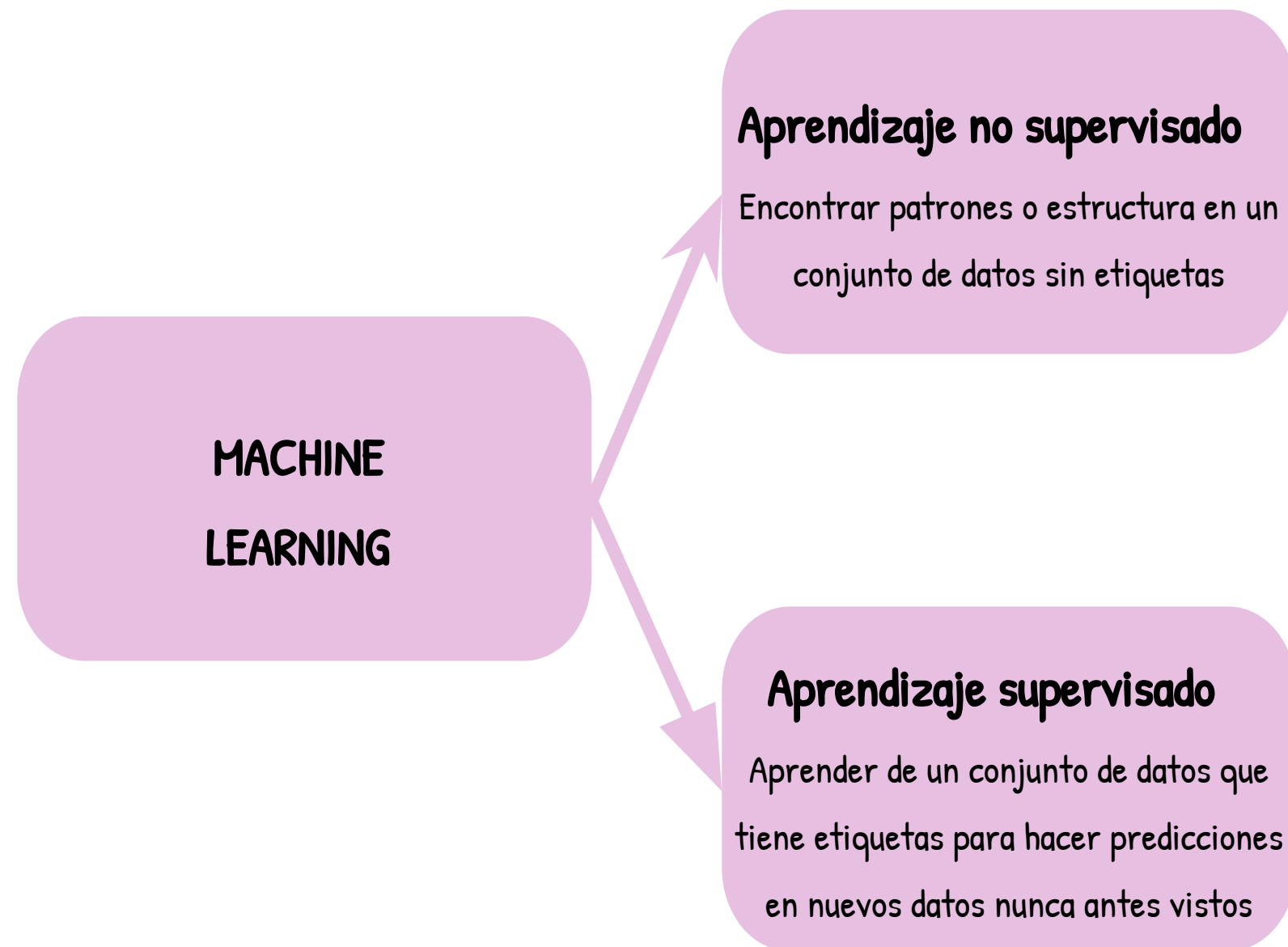
SUPERVISADO - CLASIFICACIÓN

1. Estimar el salario de una persona en función de su nivel educativo y su experiencia laboral.

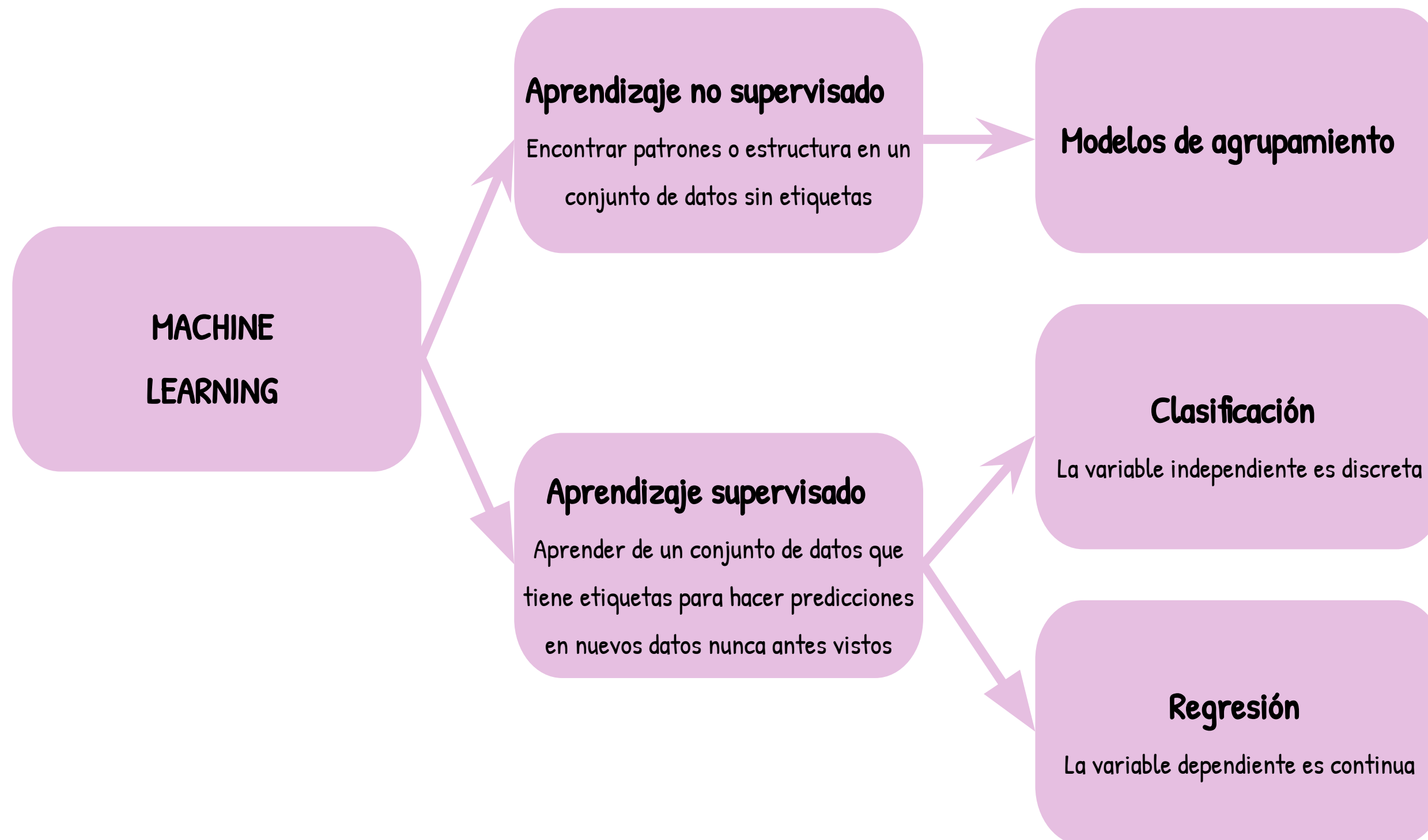
SUPERVISADO - REGRESIÓN



1. ¿Tenemos etiquetas?

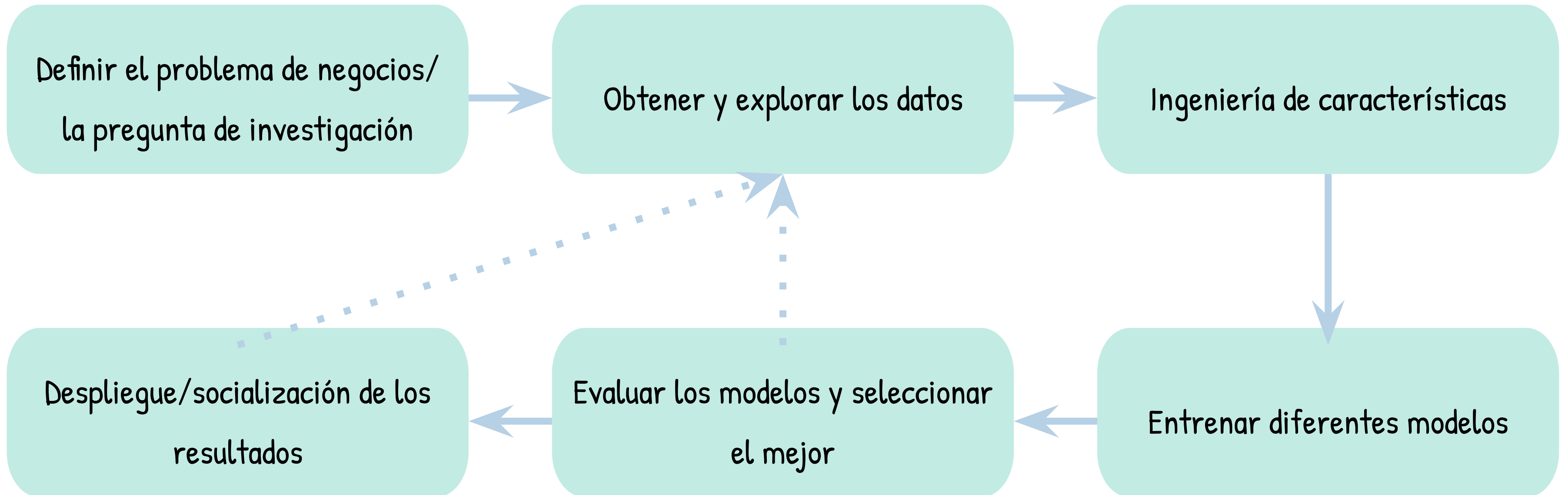


1. ¿Tenemos etiquetas?
2. ¿De qué tipo son nuestras etiquetas?





Pasos en un proyecto de Machine Learning



Ejemplos

1. ¿Cuáles son los atributos para que una canción se haga exitosa?
<https://github.com/marcoantoniogodoy/ucberkeley-mlai-capstone-project>
2. ¿De qué hablan las canciones que representan a Colombia?
<https://github.com/jfreddypuentes/Colombia-songs-word-cloud>
3. Detectar transacciones bancarias fraudulentas
https://github.com/KSpiliop/Fraud_Detection
4. Más proyectos: <https://www.kaggle.com/competitions/>



Ensayo

(Taller # 5)

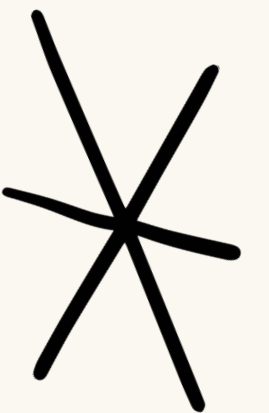
Fecha de entrega: Septiembre 9, 2024

1. Piensa en un problema que te gustaría resolver con machine learning y descríbelo
2. ¿Cómo puedes conseguir los datos que necesitas para este proyecto?
3. ¿Cuáles son sus datos de entradas? ¿Cuáles son sus datos de salida? ¿Cómo se vería cada una de las filas con las observaciones del conjunto de datos?
4. ¿Qué familia de modelos de Machine Learning utilizarías para este problema?

Crea un Google Doc con este ensayo y compártelo con vroberta@unicomfaucauca.edu.co



1. Repaso de la última clase ✓
2. Tema de hoy: ✓
 - ¿Qué es machine learning? ✓
 - Diferencia entre ML vs DL vs AI ✓
 - Partes de un modelo de ML ✓
 - Familias de modelos de ML ✓
 - Pasos de un proyecto de ML ✓
 - Ejemplos ✓
3. Logística de la clase



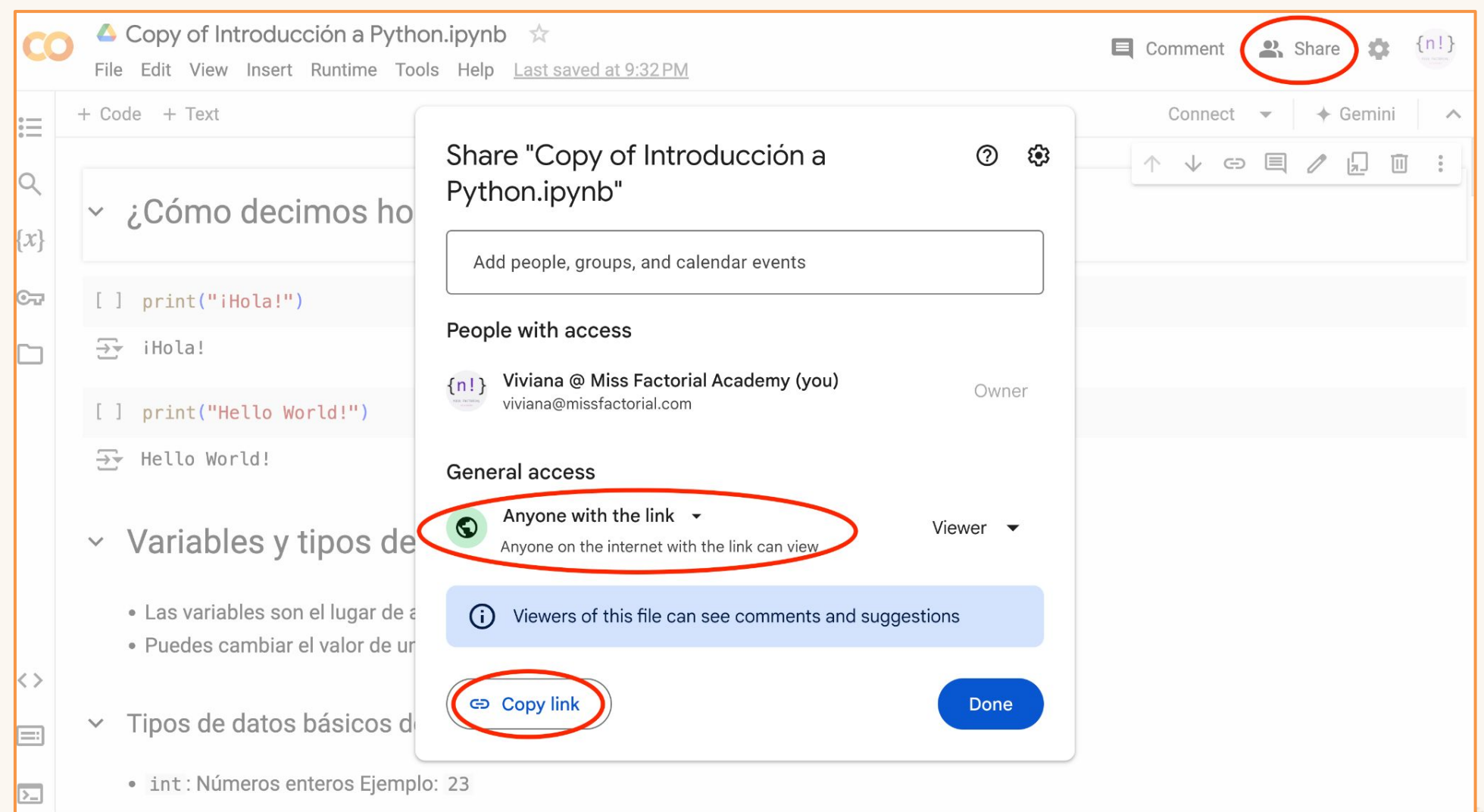
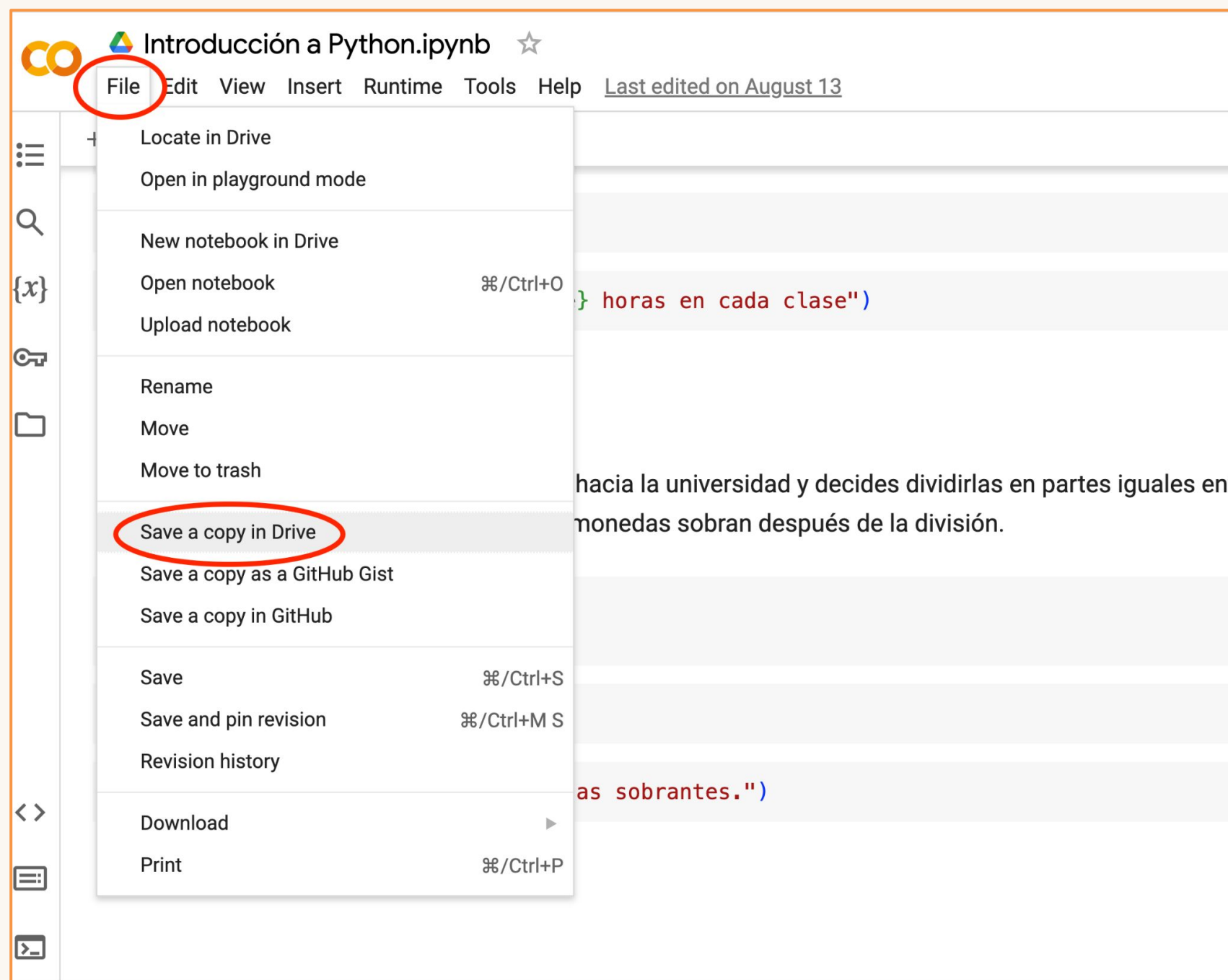
Socialización de notas primer corte:

- A lo largo del semestre les ido respondiendo a cada taller con su nota, si no han recibido esto, por favor contactárme.
- Al terminar la próxima clase, le enviare a cada uno de ustedes un email con sus notas del primer corte y el número de asistencias. Por favor revisar el correo institucional.
- El último día para enviar reclamos y/o aclaraciones es el 13 de septiembre.

Taller #	Descripción	Enlace	Fecha de entrega	Porcentaje en el primer corte	Porcentaje en el curso
Taller # 1 (Encuesta)	Encuesta Google Docs	Enlace	Agosto 19, 2024	10%	3%
Taller # 2	Operaciones aritméticas	Enlace	Agosto 19, 2024	20%	6%
Taller # 3	Estructura de datos y condicionales	Enlace	Agosto 26, 2024	20%	6%
Taller # 4	Bucles, funciones y librerías	Enlace	Septiembre 2, 2024	20%	6%
Taller # 5 (Ensayo)	Problema que te gustaría resolver con machine learning	Diapositiva 46	Septiembre 9, 2024	30%	9%

Para enviar los talleres de código

- ❑ Hacer click en **archivo** → **guardar copia en mi Drive** para que les quede una copia en su cuenta, de lo contrario, los resultados no serán guardados.
- ❑ En la copia creada, hacer click en **compartir**, asegurarse que el enlace sea visible a **cualquier persona**, copiar el enlace y enviarlo.



¡Gracias!



¿Dudas? Email de la profe:

vroberta@unicomfacauca.edu.co

Página web del curso con toda la info:

<https://github.com/vivianamarquez/unicomfacauca-ai-2024>