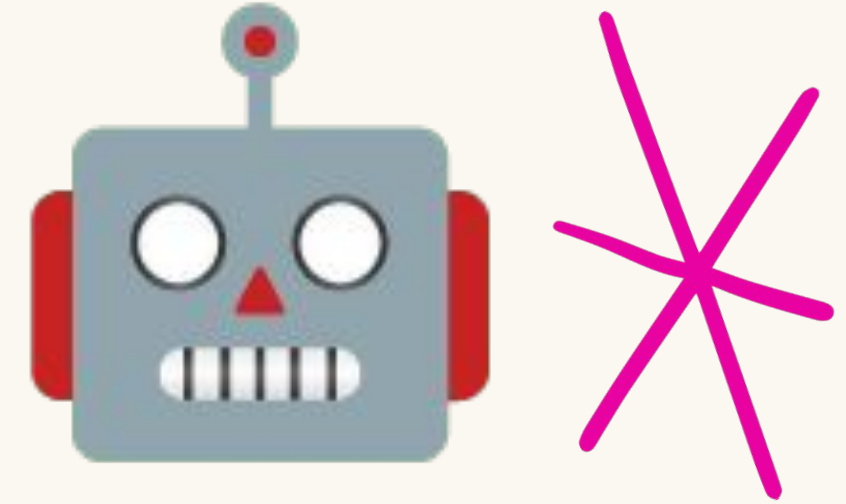


Desbloqueando el Poder de los Datos



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 🎵💬



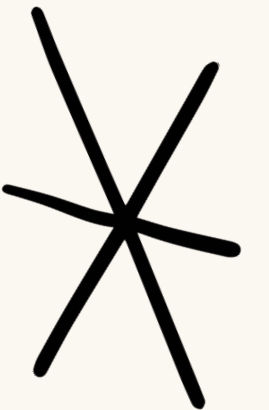
Adquisición de datos

Septiembre 10, 2024





1. Repaso de la última clase
2. Tema de hoy:
 - Datos estructurados vs no estructurados
 - ¿Cómo conseguir datos?
 - Tus propios datos
 - Datos de código abierto
 - APIs
 - Web scraping
3. Logística de la clase





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

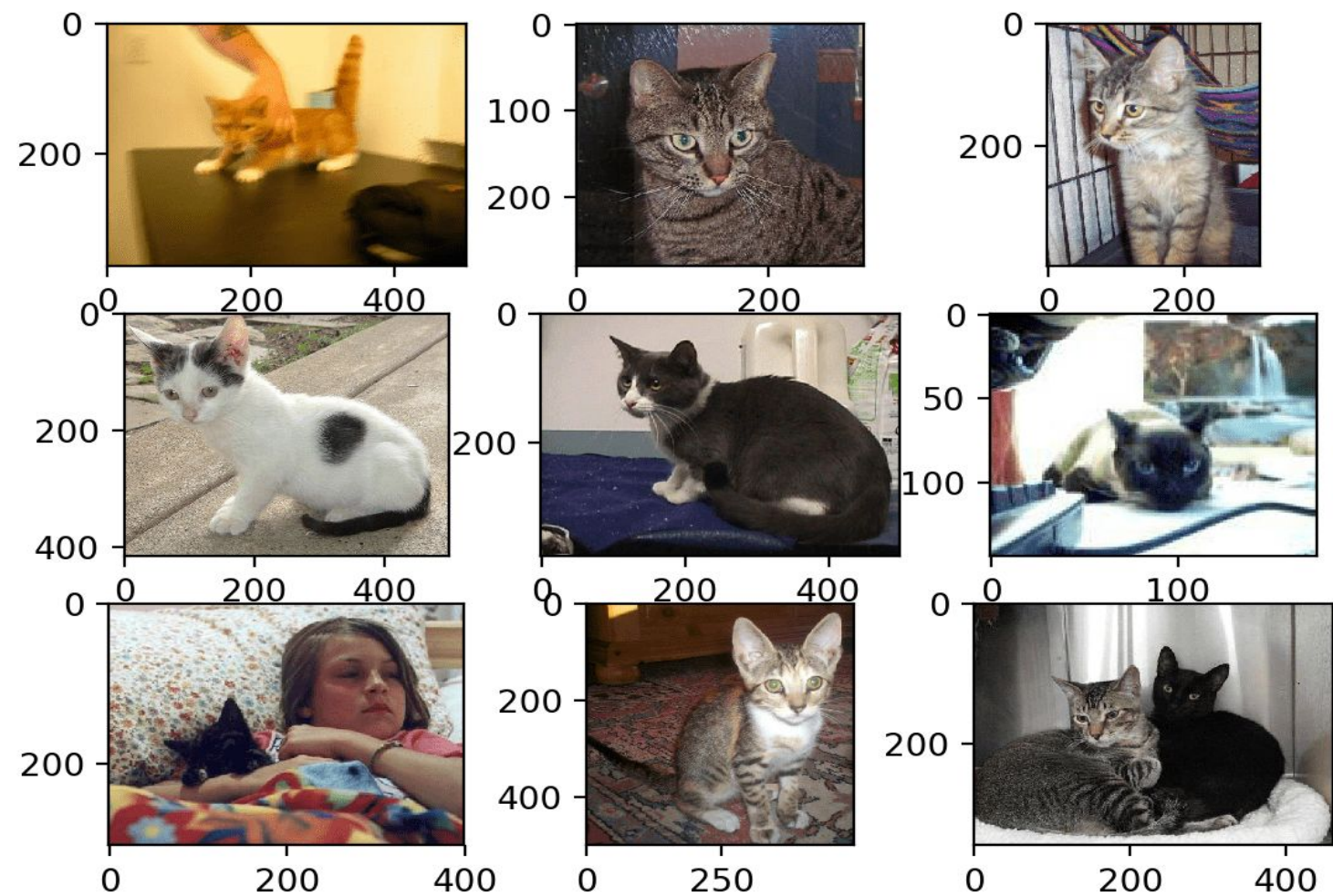


¿Escribes una lista de reglas como "los gatos tienen cuatro patas" y "los gatos tienen dos orejas"?

ó

¿Le muestras al niño muchas fotos de diferentes gatos

¿Qué es machine learning? (Aprendizaje automático)

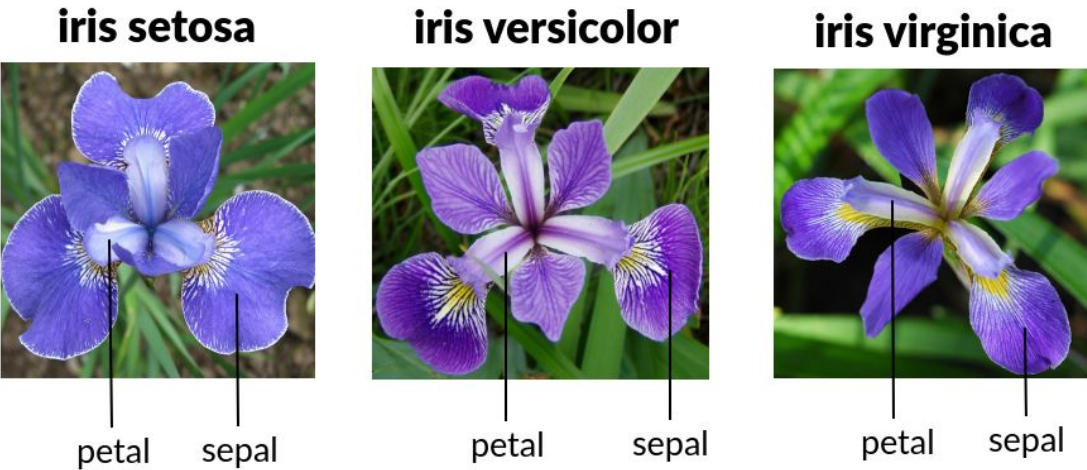


El **machine learning** es la misma idea.

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 a **generalizar** a partir de esos datos.

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

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:

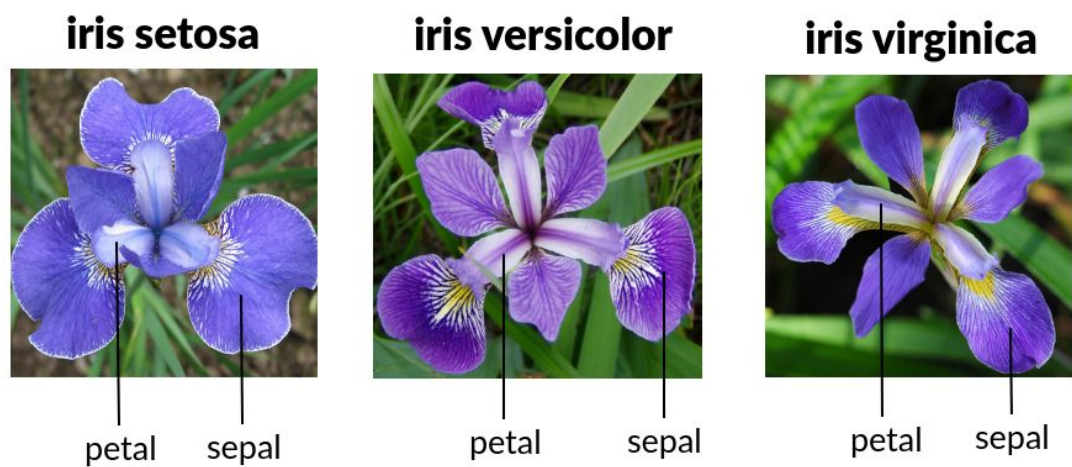
- Setosa
- Versicolor
- Virginica

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



1. ¿Tenemos etiquetas?

**MACHINE
LEARNING**

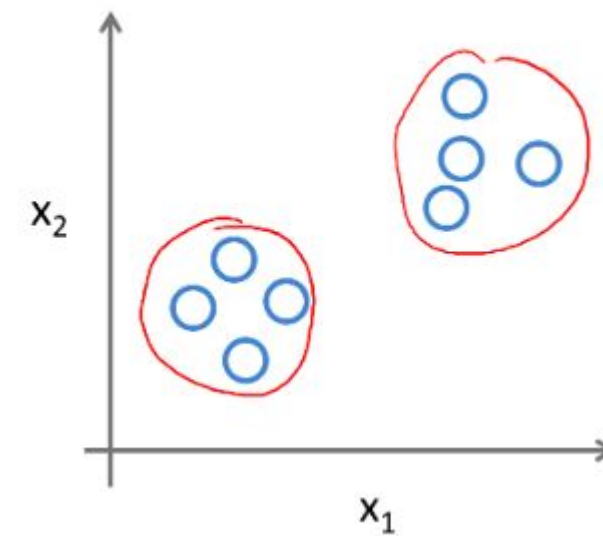
Aprendizaje no supervisado

Encontrar patrones o estructura en un conjunto de datos sin etiquetas

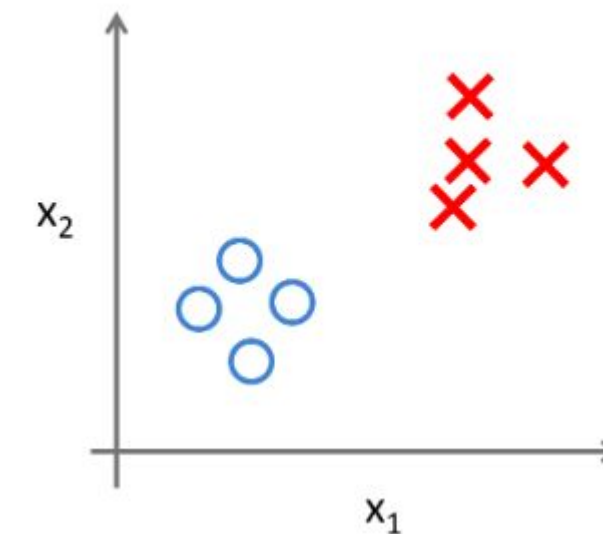
Aprendizaje supervisado

Aprender de un conjunto de datos que tiene etiquetas para hacer predicciones en nuevos datos nunca antes vistos

Aprendizaje no supervisado

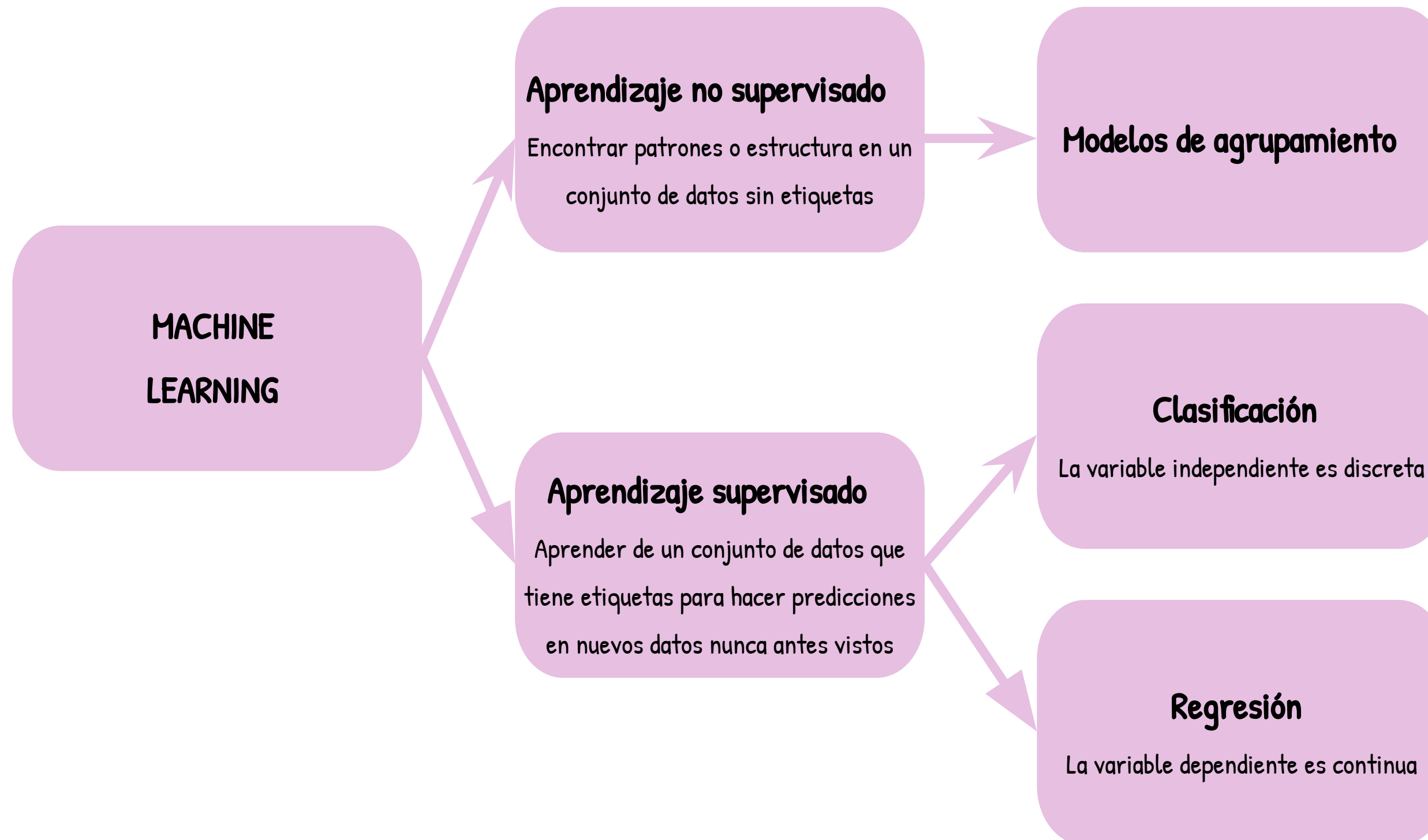


Aprendizaje supervisado

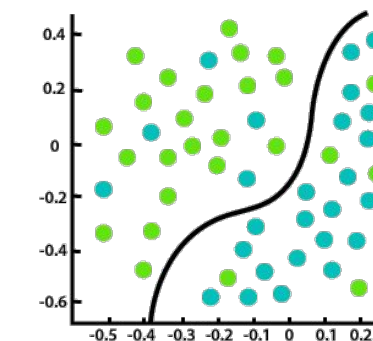
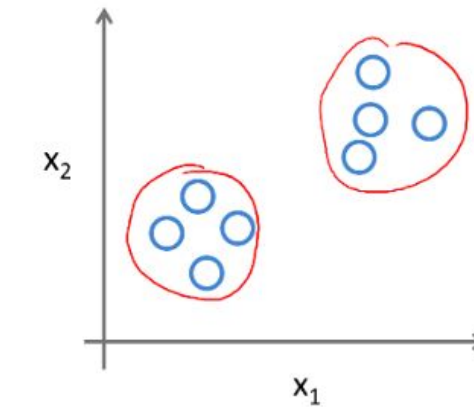


1. ¿Tenemos etiquetas?

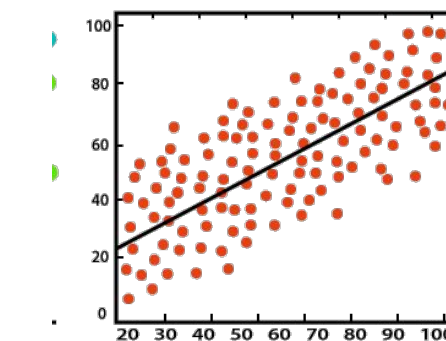
2. ¿De qué tipo son nuestras etiquetas?



Aprendizaje no supervisado

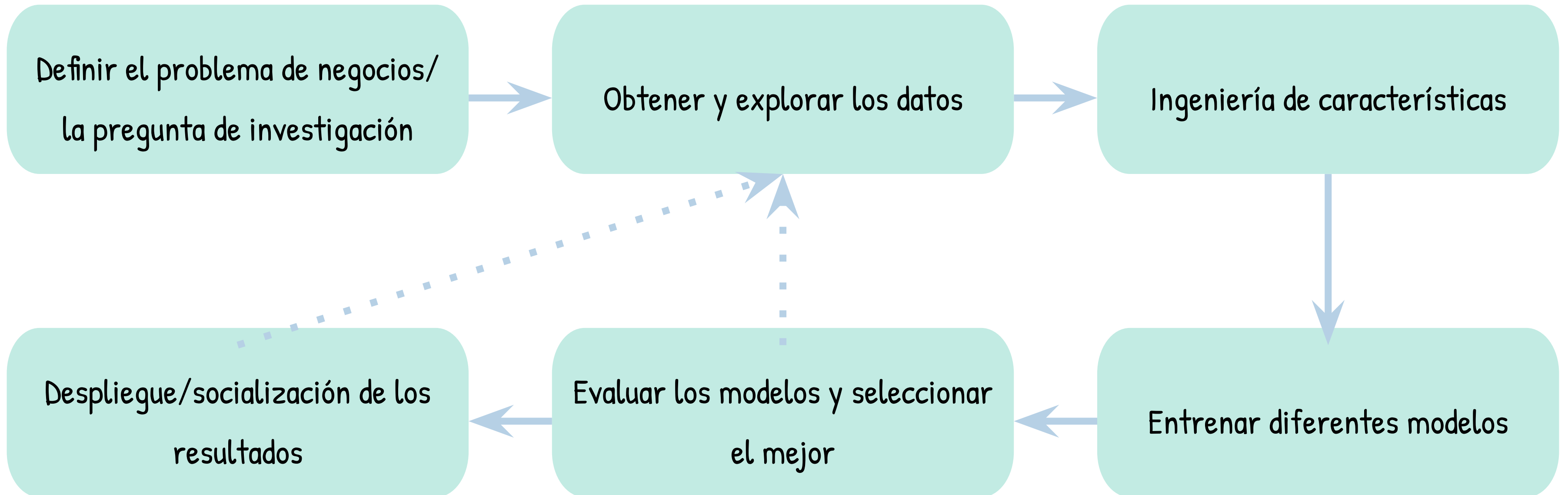


Classification

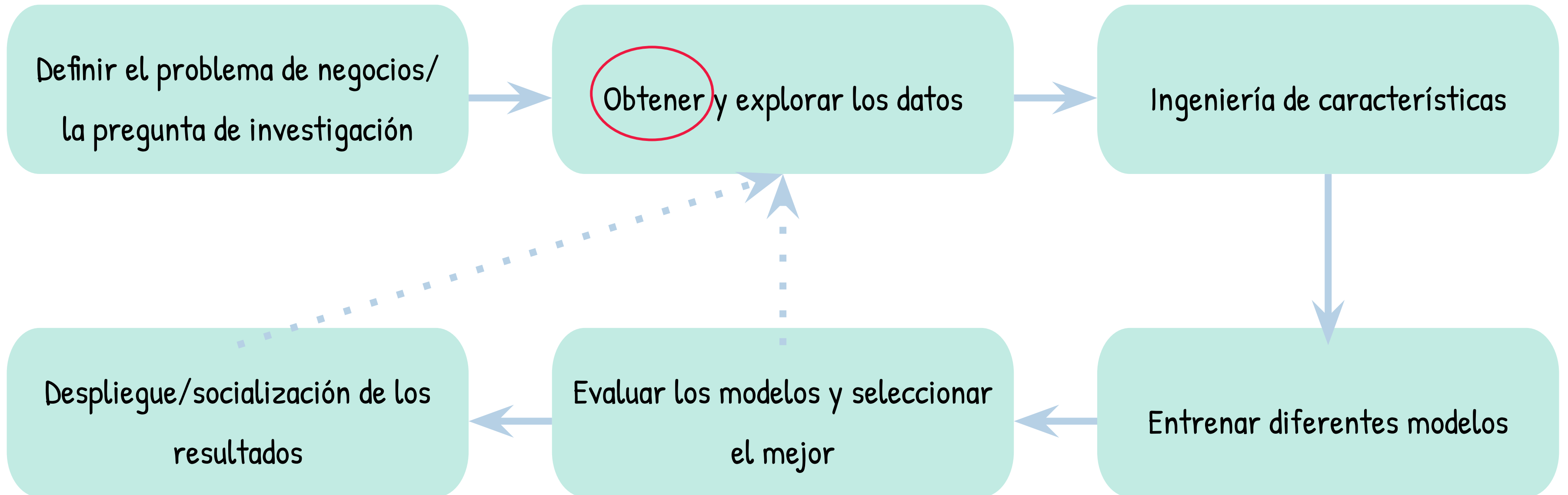


Regression

Pasos en un proyecto de Machine Learning

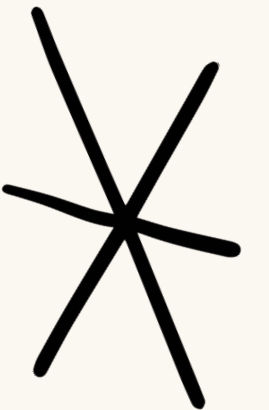


Pasos en un proyecto de Machine Learning



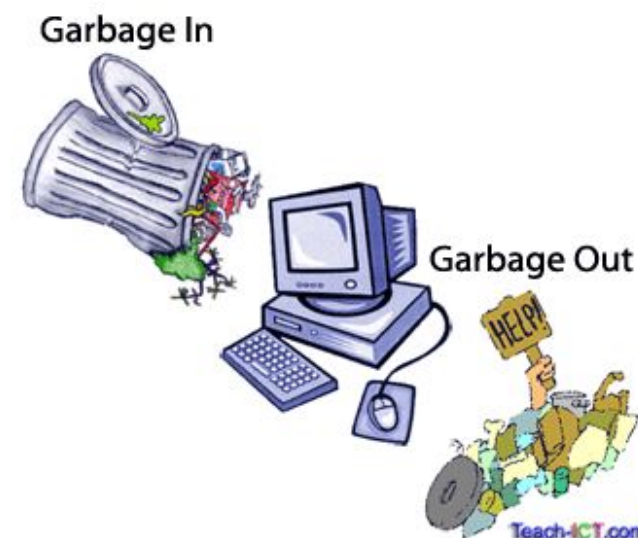


1. Repaso de la última clase ✓
2. Tema de hoy:
 - Datos estructurados vs no estructurados
 - ¿Cómo conseguir datos?
 - Tus propios datos
 - Datos de código abierto
 - APIs
 - Web scraping
3. Logística de la clase



Obtener datos para un proyecto de Machine Learning

- La adquisición de datos es el proceso de identificar, recopilar y extraer información útil de diversas fuentes para su uso en proyectos de ciencia de datos y aprendizaje automático
- Dado que los datos se han convertido en un recurso tan valioso como el petróleo en la economía digital, una adecuada adquisición garantiza que los modelos de aprendizaje automático tengan una base sólida para su entrenamiento.
- La calidad, relevancia y variedad de los datos obtenidos influyen directamente en la efectividad, precisión y desempeño del modelo, haciendo que los datos sean un componente esencial para el éxito del proyecto.



Datos estructurados vs datos no estructurados

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

- Los **datos estructurados** están altamente organizados y son fácilmente legibles por máquinas. Normalmente se almacenan en formatos tabulares, como hojas de cálculo (CSV, Excel) o bases de datos relacionales (SQL).

Cada observación está en un fila y sus características en columnas predefinidas, lo que facilita su procesamiento y análisis.



- Los **datos no estructurados** no siguen un formato o estructura específica, lo que los hace más difíciles de organizar y analizar. Este tipo de datos incluye texto libre, imágenes, videos, audios y otros formatos multimedia.

Debido a su naturaleza, los datos no estructurados a menudo requieren técnicas avanzadas, como procesamiento de lenguaje natural (NLP) o redes neuronales convolucionales (CNN).

¿Cómo conseguir datos?

1. Tus propios datos

- Datos que generas o recopilas tú mismo, como encuestas, formularios, experimentos o datos de tu empresa/universidad.
- Puedes cargarlos desde archivos locales (CSV, Excel, SQL, JSON) o desde la nube (Google Drive, AWS S3, etc.).

2. Datos de código abierto

- Los conjuntos de datos de código abierto son colecciones de datos disponibles de manera gratuita, que cualquier persona puede usar, modificar y compartir.
- Universidades, gobiernos y organizaciones de investigación también publican a menudo conjuntos de datos abiertos.

¿Cómo conseguir datos?

3. APIs

- Una API (Interfaz de Programación de Aplicaciones) es una interfaz que permite acceder y recopilar datos de diversas fuentes de manera automatizada, facilitando la obtención de grandes volúmenes de información en ciencia de datos para su análisis.
- Ejemplos incluyen la API de Twitter, Google Maps, Spotify, o APIs financieras para datos de mercado.

4. Web Scraping

- Extraer datos de sitios web que no tienen una API disponible, pero permiten el acceso público a sus datos. Herramientas como BeautifulSoup, Scrapy, o Selenium te permiten automatizar este proceso.

¿Cómo conseguir datos?

5. Bases de datos

- **SQL:** Obtener datos de bases de datos relacionales como MySQL, PostgreSQL, o SQLite.
- **NoSQL:** Obtener datos de bases de datos NoSQL como MongoDB o Firebase.

6. Comprar datos

- Plataformas donde puedes comprar o descargar conjuntos de datos, como Quandl o AWS Data Exchange.

7. Simulación de datos

- Si no tienes acceso a datos reales, puedes generar datos sintéticos o simulados usando herramientas como scikit-learn o Faker.

Datos abiertos

1. Repositorios de datos abiertos

- OpenML.org <https://openml.org>
- Kaggle.com <https://kaggle.com/datasets>
- PapersWithCode.com <https://paperswithcode.com/datasets>
- UC Irvine Machine Learning Repository <https://archive.ics.uci.edu/ml>
- Amazon's AWS datasets <https://registry.opendata.aws>
- TensorFlow datasets <https://tensorflow.org/datasets>
- Google's data search engine <https://datasetsearch.research.google.com/>

2. Portales que tienen un listado de datos

- DataPortals.org <https://dataportals.org/>
- Listado de Wikipedia
https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research
- Quora's list <https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>
- Reddit's dataset <https://www.reddit.com/r/datasets>
- GitHub * <https://github.com/>

Datos abiertos


3. Específicos por lugar

- San Francisco Open Data <https://datasf.org/opendata/>
- NYC Open Data <https://opendata.cityofnewyork.us/>
- Datos Abiertos Londres <https://opendata.london.ca/>
- Datos Abiertos Colombia <https://www.datos.gov.co/>
- Datos Abiertos Bogotá <https://datosabiertos.bogota.gov.co/about>
- Burundi Open Data for Africa: <https://burundi.opendataforafrica.org/>
- Datos Abiertos Popayán
<https://www.popayan.gov.co/Ciudadanos/Paginas/Datos-Abiertos-Alcaldia-de-Popayan.aspx#gsc.tab=0>
- Datos Abiertos Cauca
<https://www.cauca.gov.co/NuestraGestion/Paginas/Datos-Abiertos.aspx>



Notebook de hoy

<https://colab.research.google.com/drive/1SD0uy67f7RAxAocdK4DXW3e2lNliAWA2?usp=sharing>





Taller # 6

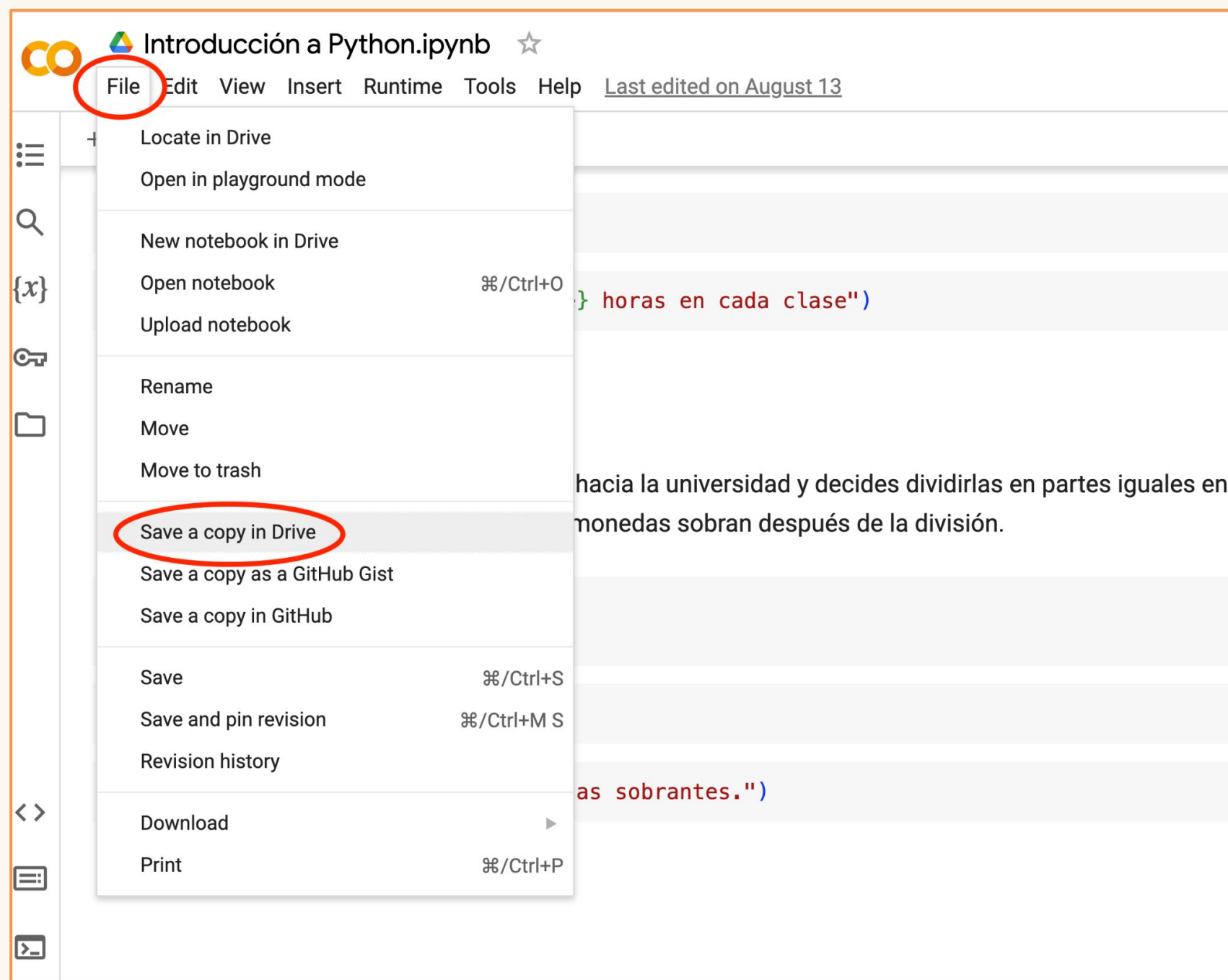
Adquisición de datos

Fecha de entrega: Septiembre 16, 2024

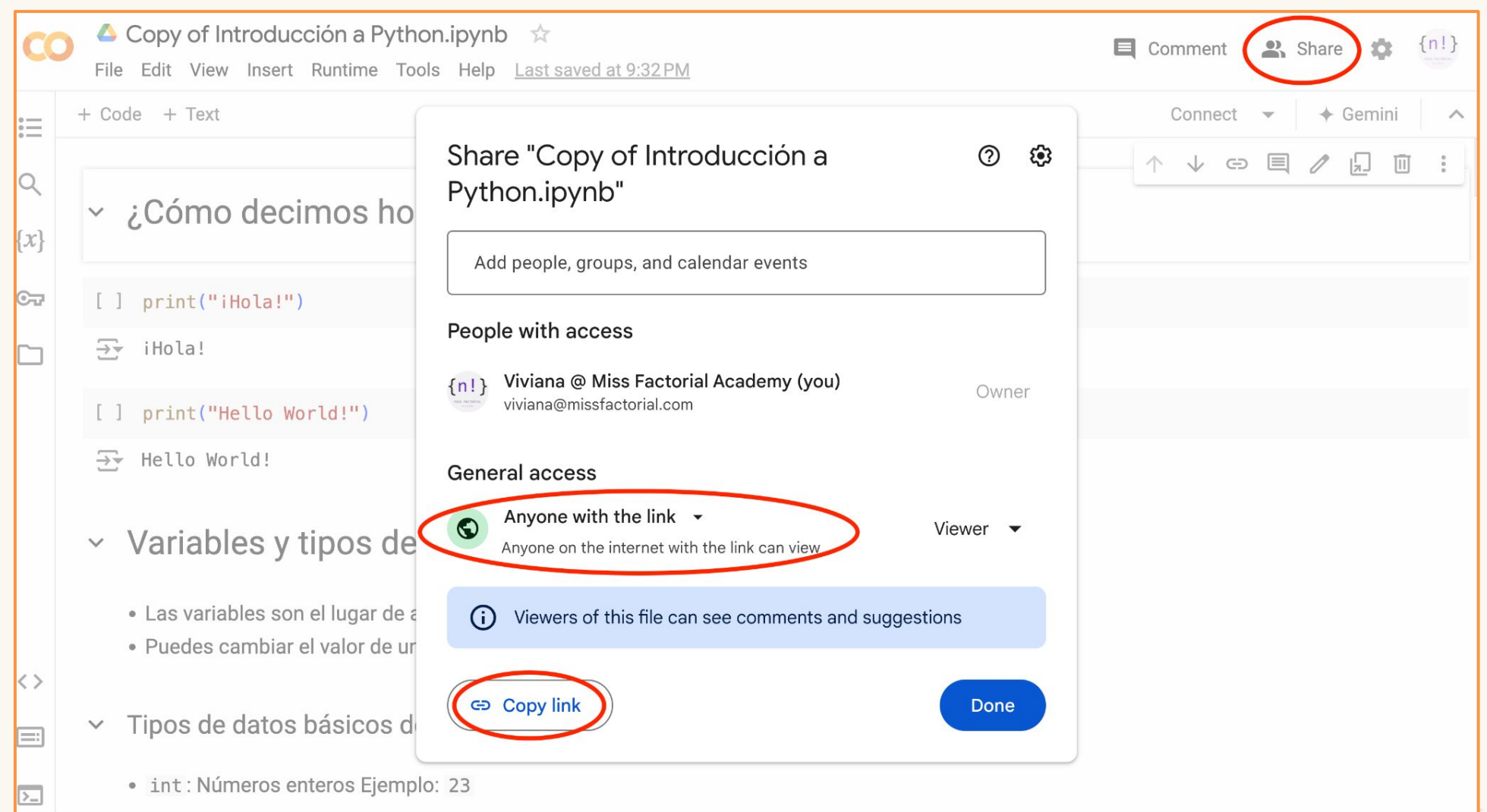
<https://colab.research.google.com/drive/16ne-V9DwTVEfhA6AAPcFxZDrp1SV-14M?usp=sharing>

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.

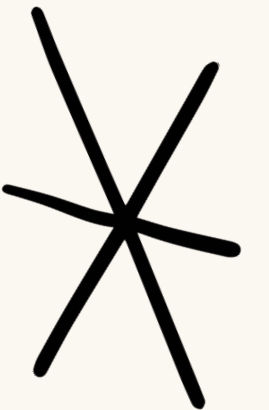


vroberta@unicomfaucauca.edu.co





1. Repaso de la última clase ✓
2. Tema de hoy:
 - Datos estructurados vs no estructurados ✓
 - ¿Cómo conseguir datos? ✓
 - Tus propios datos
 - Datos de código abierto
 - APIs
 - Web scraping
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.
- Mañana, 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 44	Septiembre 9, 2024	30%	9%



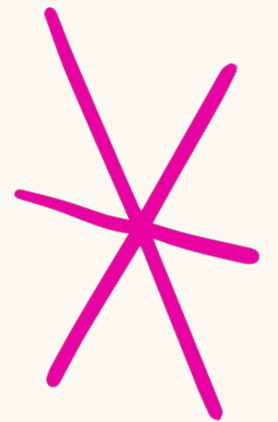
(Opcional)

Encuesta anónima: ¿Cómo va el curso?

<https://forms.gle/dYPNuZosMmgvjKUU9>



¡Gracias!



¿Dudas? Email de la profe:

vroberta@unicomfauca.edu.co

Página web del curso con toda la info:

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

Google Cloud

Documentación

Áreas de tecnología

Herramientas para productos cruzados

Sitios relacionados

Buscar

Español - A...

Consola

Comunicate con nosotros

Comenzar gratis

Filtrar

Página principal de documentación

Conceptos básicos

Comienza a usar Google Cloud

Guías generales de IA, IA generativa y AA

Guías de seguridad generales

Authentication

Métodos de autenticación de Google

Casos de uso de autenticación

Formas de autenticación

Autentica para usar bibliotecas cliente

Autentica para usar la CLI de gcloud

Autentica para usar REST

Autentica mediante la identidad temporal como cuenta de servicio

Autentícate mediante claves de API

Credenciales predeterminadas de la aplicación

Obtén un token de ID

Tipos de tokens

Productos de administración de identidades

Notas de la versión

Cambios recientes en el producto

Crea una clave de API

Para crear una clave de API, usa una de las siguientes opciones:

ConsolegcloudJavaPythonREST

1. En la consola de Google Cloud, ve a la página **Credenciales**.

Ir a Credenciales

2. Haz clic en **Crear credenciales** y, luego, selecciona **Clave de API** en el menú.

Se mostrará la string a la clave recién creada en el cuadro de diálogo **Se creó la clave de API**.

Copia tu string de clave y mantenla segura. A menos que uses una clave de prueba que pretendas borrar más adelante, agrega [restricciones de aplicación y de clave de API](#).

Usa una clave de API

Si una API admite el uso de claves de API, puedes usar claves de API para autenticarla. Puedes usar claves de API con [solicitudes de REST](#) y [bibliotecas cliente](#) que las admitan.

Usa una clave de API con REST

Puedes pasar la clave de API a una llamada a la API de REST como un parámetro de consulta con el siguiente formato. Reemplaza **API_KEY** por la string de clave de tu clave de API.

Por ejemplo, si deseas pasar una clave de API para solicitar `documents.analyzeEntities` a la API de Cloud Natural Language, usa lo siguiente:

En esta página

[Introducción a las claves de API](#)

Antes de comenzar

cloud.google.com usa cookies de Google para brindar sus servicios, mejorar su calidad y analizar el tráfico. [Más información.](#)

Entendido

Comienza tu prueba gratuita con un crédito de \$300. No te preocupes, no se te cobrará si se acaban los créditos. [Más información](#)

DESCARTAR

COMENZAR GRATIS

Google Cloud

Selecciona un proyecto

Buscar (/) recursos, documentos, productos y más

Buscar

V

API APIs y servicios

APIs y servicios habilitados

Biblioteca

Credenciales

Pantalla de consentimiento ...

Acuerdos de uso de páginas

Credenciales

Para ver esta página, selecciona un proyecto.

CREAR PROYECTO

Google Cloud

Te damos la bienvenida, Viviana Roberta Márquez

Crea y administra tus instancias, discos, redes y otros recursos de Google Cloud desde un solo lugar.

V

Viviana Roberta Márquez

[CAMBIAR DE CUENTA](#)

vroberta@unicomfauca.edu.co

País

Colombia

Condiciones del Servicio

☒ Acepto las [Condiciones del Servicio de Google Cloud Platform](#) y las Condiciones del Servicio de [cualquier servicio y APIs aplicables](#).

Actualizaciones por correo electrónico

☐ Quiero recibir correos electrónicos periódicos sobre novedades, actualizaciones de productos y ofertas especiales de Google Cloud y los socios de Google Cloud.

[ACEPTAR Y CONTINUAR](#)

Proyecto nuevo



Tienes 12 projects restantes en tu cuota. Solicita un incremento o borra algunos proyectos. [Más información](#)

[MANAGE QUOTAS](#)

Nombre del proyecto *

ejemplo-clase-09092024



ID del proyecto: ejemplo-clase-09092024. No se puede cambiar más adelante.

[EDITAR](#)

Organización *

unicomfauca.edu.co



Selecciona una organización para vincularla a un proyecto. No podrás cambiar esta selección más adelante.

Ubicación *

 unicomfauca.edu.co

[EXPLORAR](#)

Organización o carpeta superior

[CREAR](#)

[CANCELAR](#)

APIs y servicios

APIs y servicios habilitados

Biblioteca

Credenciales

Pantalla de consentimiento ...

Acuerdos de uso de páginas

Credenciales

+ CREAR CREDENCIALES

BORRAR

RESTABLECER CREDENCIALES BORRADAS

Clave de API

Identifica tu proyecto con una clave de API simple para verificar la cuota y el acceso

ID de cliente de OAuth

Solicita el consentimiento del usuario para que tu app pueda acceder a sus datos

Cuenta de servicio

Habilita la autenticación de servidor a servidor en el nivel de la app mediante cuentas robot

Ayúdame a elegir

Responde algunas preguntas para decidir qué tipo de credencial usar

Recuerda configurar la pantalla de consentimiento

CONFIGURAR PANTALLA DE CONSENTIMIENTO

Claves de API

Nombre

No hay claves de API para mostrar

Restricciones

Acciones

ID de clientes OAuth 2.0

Nombre

Fecha de creación ↓

Tipo

ID de cliente

Acciones

No hay clientes de OAuth para mostrar

Cuentas de servicio

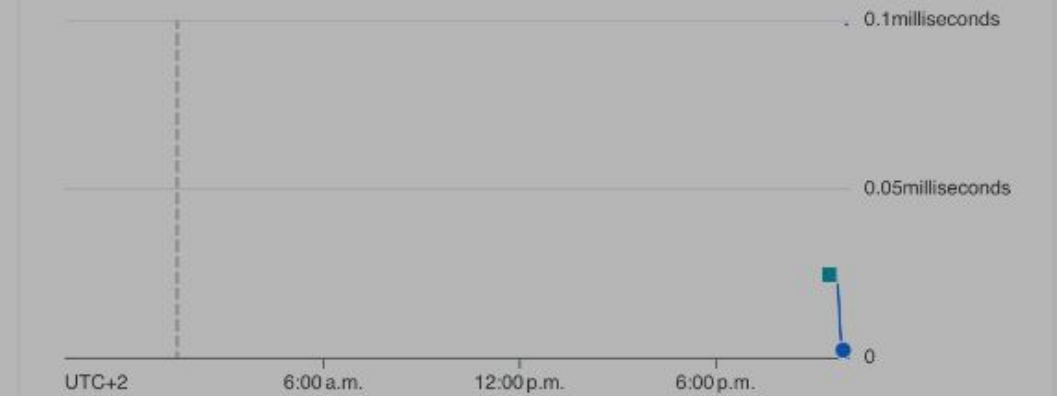
Correo electrónico

Nombre ↑

Acciones

No hay cuentas de servicio para mostrar

Administrar cuentas de servicio



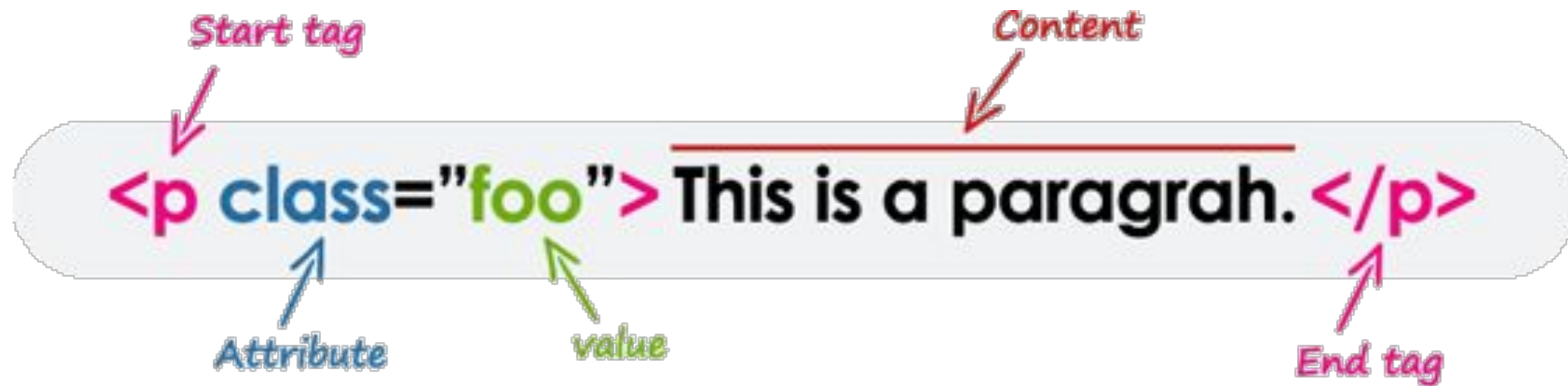
Filtro

Filtro

?

Nombre	↓ Solicitudes	Errores (%)	Latencia mediana (ms)	95% de latencia (ms)
Distance Matrix API	7	100	21	31
Analytics Hub API				
BigQuery API				
BigQuery Connection API				
BigQuery Data Policy API				
BigQuery Migration API				
BigQuery Reservation API				
BigQuery Storage API				
Cloud Dataplex API				
Cloud Datastore API				
Cloud Logging API				
Cloud Monitoring API				
Cloud SQL				

Partes de una etiqueta de HTML



Ejemplo:

```
<a href="http://www.google.com/" id="buscador">Google</a>
```