Fundamentos de Aprendizaje de Máquina

PhD Jorge Rudas

INFOTEP
INSTITUCIÓN DE EDUCACIÓN SUPERIOR

Reglas del curso

- Asistencia
- Celular
- Plagio
- Uso de herramientas de IA Generativa
- Fechas de entrega
- Calificaciones

Pre requisitos del curso

- Fundamentos sólidos de programación orientada a objetos
- Fundamentos de uso de lenguaje de programación Python 3.x
- Fundamentos de Álgebra Lineal y Cálculo Diferencial
- Fundamentos de manejo de datos en Python
  - Pandas
  - Numpy
- **Herramientas:**
  - GitHub
  - CodeLabs
  - Latex → OverLeaf

Que es la inteligencia artificial?

Que es el aprendizaje de máquina?

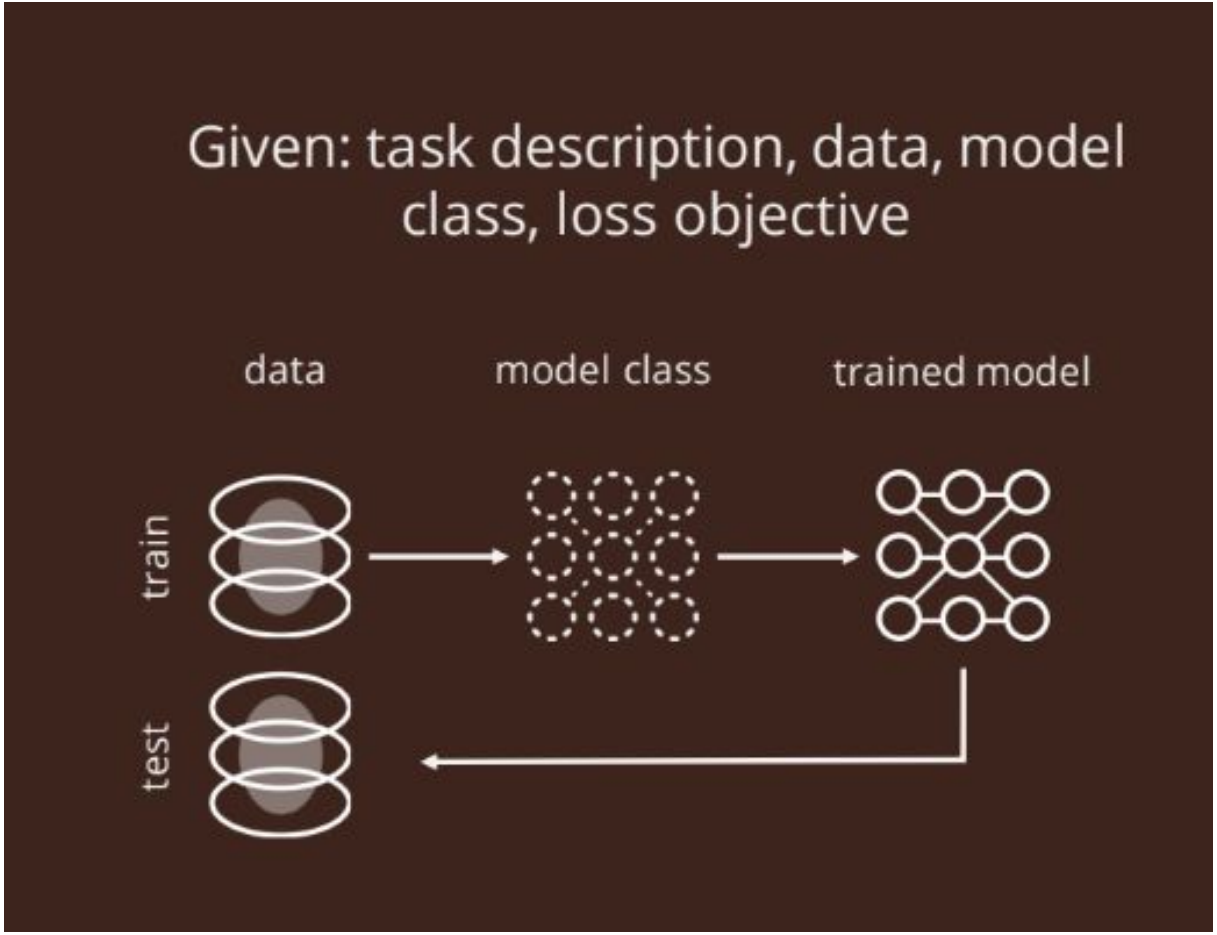
“Artificial intelligence (AI) is a field of study that involves creating machines that can learn, reason, and act independently. AI systems can perform tasks that usually require human intelligence, such as problem-solving, decision-making, and learning.”

Google IA team

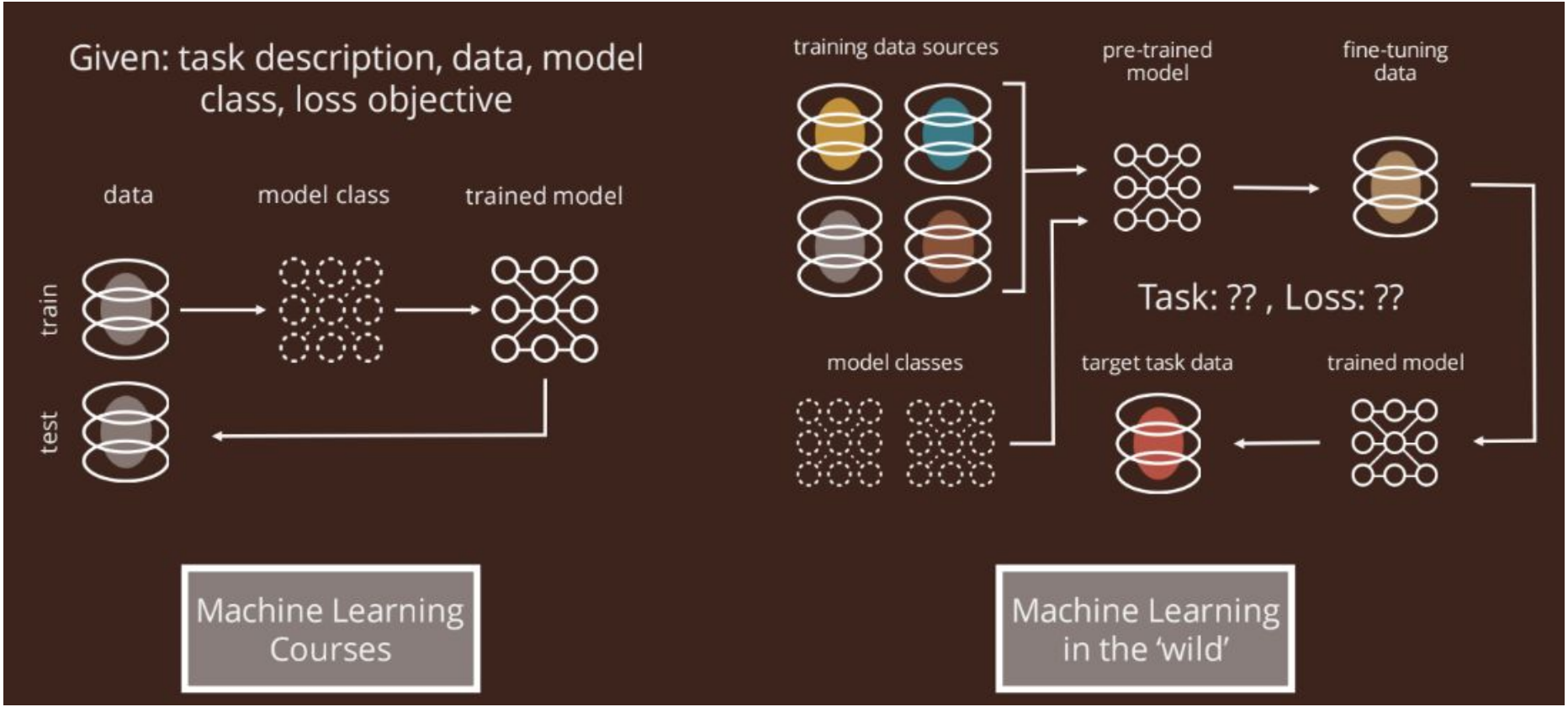
“Machine Learning is a system that gradually learns how to make useful predictions by studying lots of data to discover connections and correlations among them”

Google IA team

ML is the process which a set of **data** elements are used for **training** a piece of software, called a **model**, to make useful **predictions** or generate content from data.


Given: task description, data, model class, loss objective
data
model class
trained model
train
test

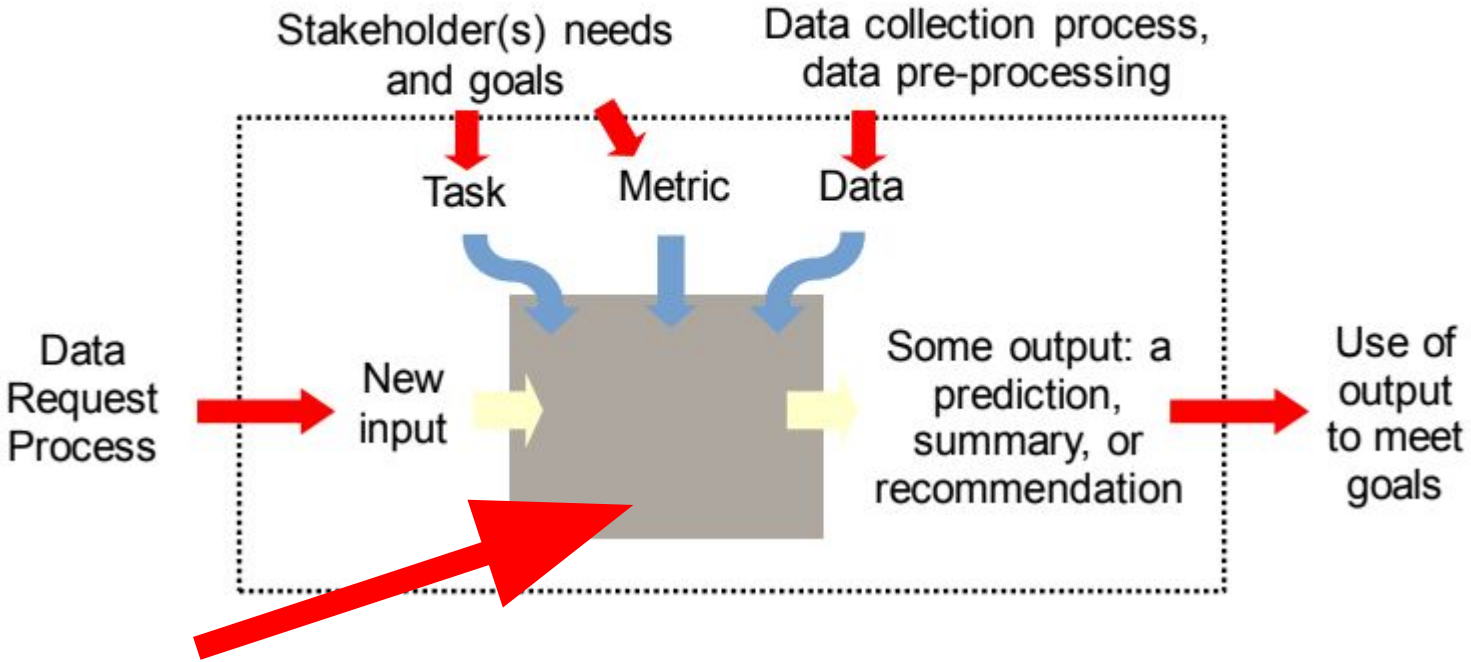
Una mala noticia!



Una buena noticia!

The same fundamental principles we will learn in this class, can be re-used and extended to tackle ”ML in the Wild” too

High-Level Setup



Nunca pierdas de vista esta visión general

(por muy ingeniero que te vas a convertir!)

Example # 1

Input → Data?

Training?
Model?

Output → Prediction

Android Glide Typing

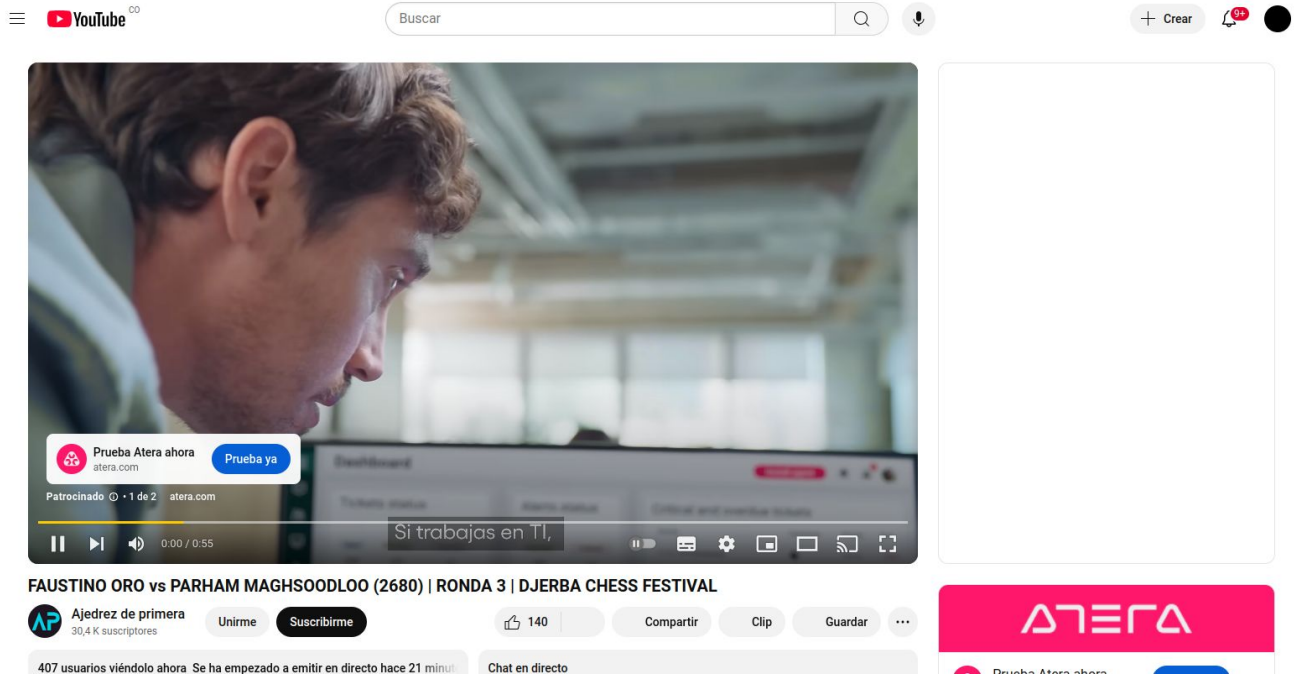
iOS QuickPath

Example # 2

Input → Data?

Training?
Model?

Output → Prediction

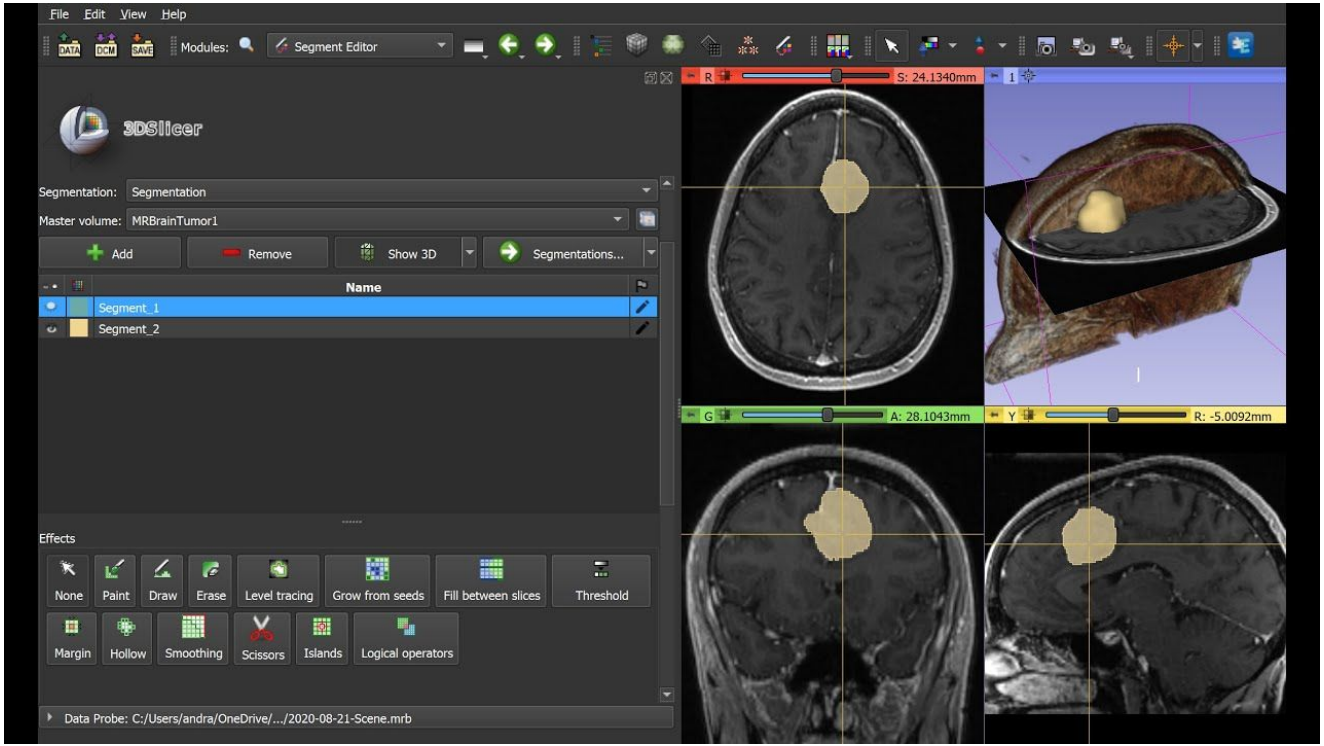


Example # 3

Input → Data?

Training?
Model?

Output → Prediction

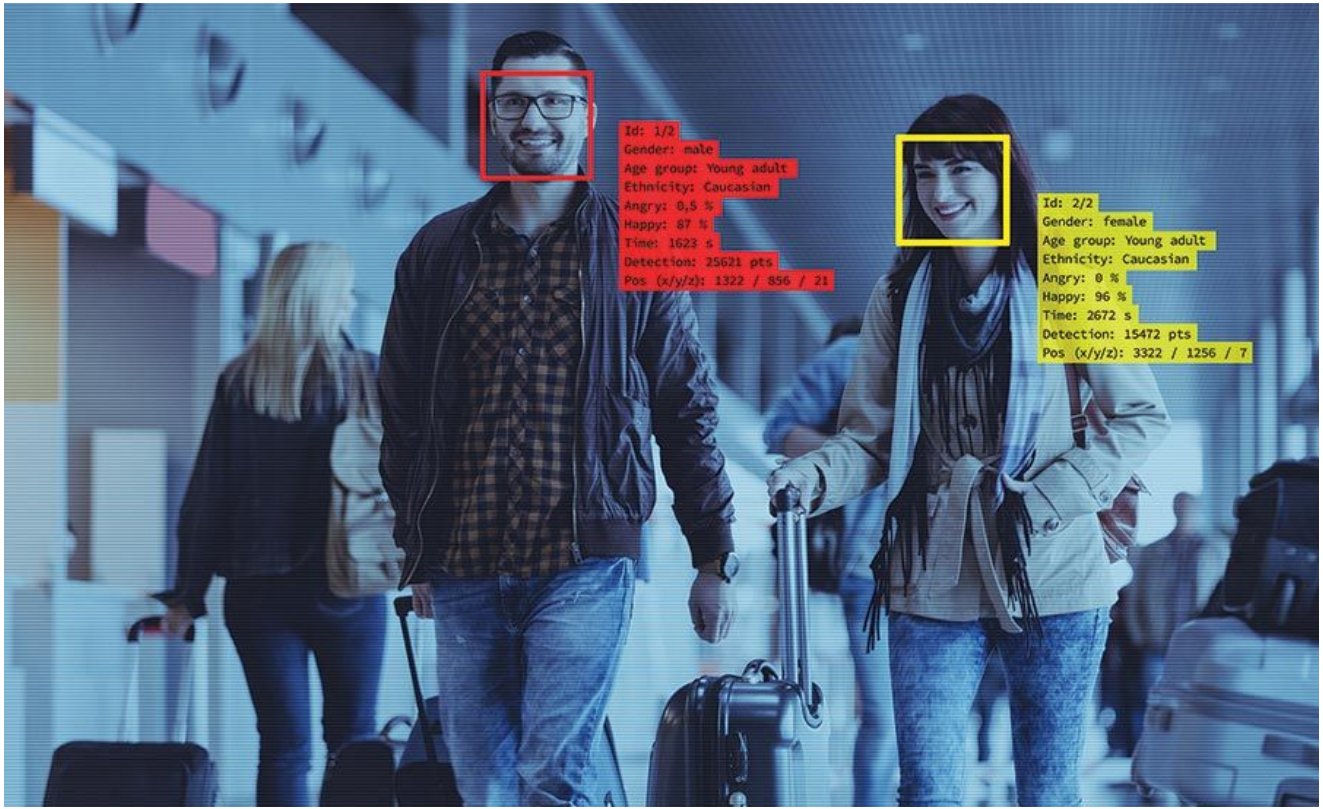


Example # 4

Input → Data?

Training?
Model?

Output → Prediction



Example # 5

Input → Data?

Training?
Model?

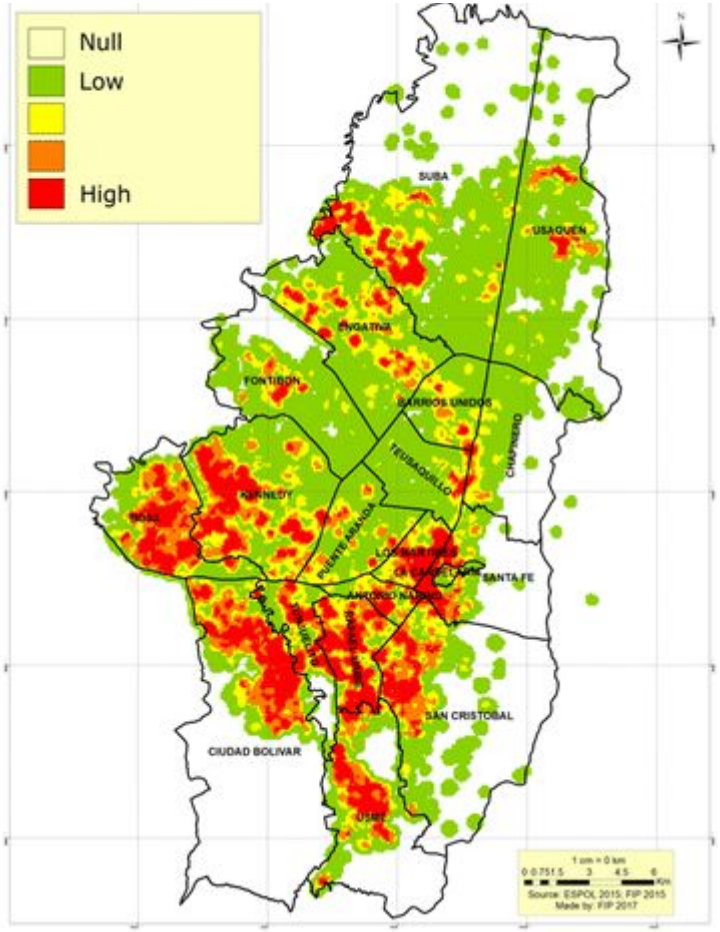
Output → Prediction

Example # 6

Input → Data?

Training?
Model?

Output → Prediction



- **Unsupervised learning**
- **Supervised learning**
- Probabilistic
- Non-Probabilistic
- Reinforcement learning
- Generative AI

*“**Supervised learning** models can make predictions after seeing lots of data with the correct answers and then discovering the connections between the elements in the data that produce the correct answers. This is like a student learning new material by studying old exams that contain both questions and answers. Once the student has trained on enough old exams, the student is well prepared to take a new exam. These ML systems are "supervised" in the sense that a human gives the ML system data with the known correct results”*

Google AI Team

Taller individual # 1a

https://profecaevp.wordpress.com/wp-content/uploads/2015/04/mancha_png_by_ona_smile-d5tmk7x.png

Cree un Jupyter Notebook con una rutina **abierta** en python para calcular el área de la mancha

Taller individual # 1b

https://profecaevp.wordpress.com/wp-content/uploads/2015/04/mancha_png_by_ona_smile-d5tmk7x.png

Cree un Jupyter Notebook con una rutina en python para calcular el área de la mancha siguiendo la guía a continuación

Algoritmo

Paso 1 → Defina ***n*** puntos bidimensionales aleatorios distribuidos de forma uniforme sobre la imagen. ***w*** será el ancho y ***h*** el alto de la imagen

Paso 2 → Determine la variable ***p_dentro*** (número de puntos aleatorios que están dentro de la mancha)

Paso 3 → $$Área = \lim_{n \to \infty} \frac{p_{dentro}}{n} wh$$

Taller grupal # 2

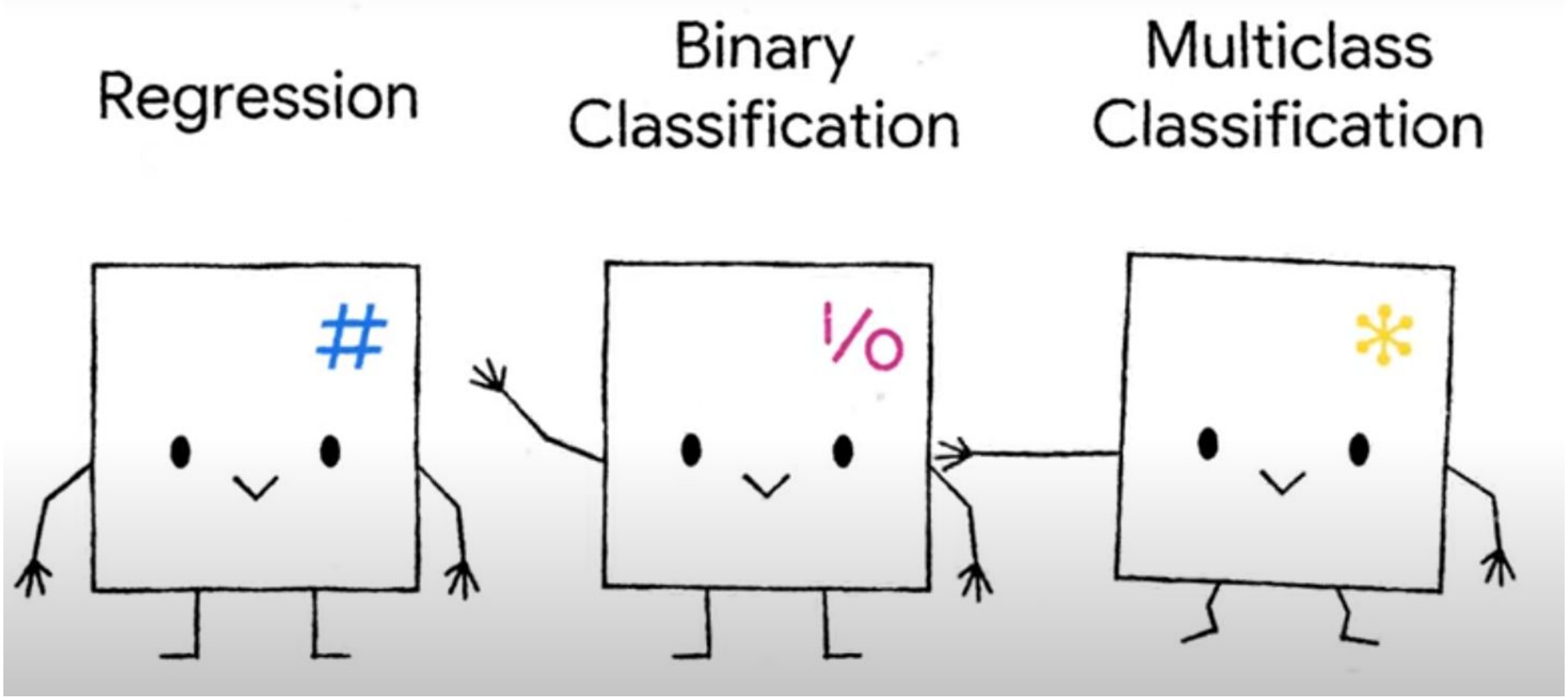
Dataset

https://www.datos.gov.co/Comercio-Industria-y-Turismo/Personas-Naturales-Jur-dicas-y-Establecimientos-en/jxfg-er2i/about_data

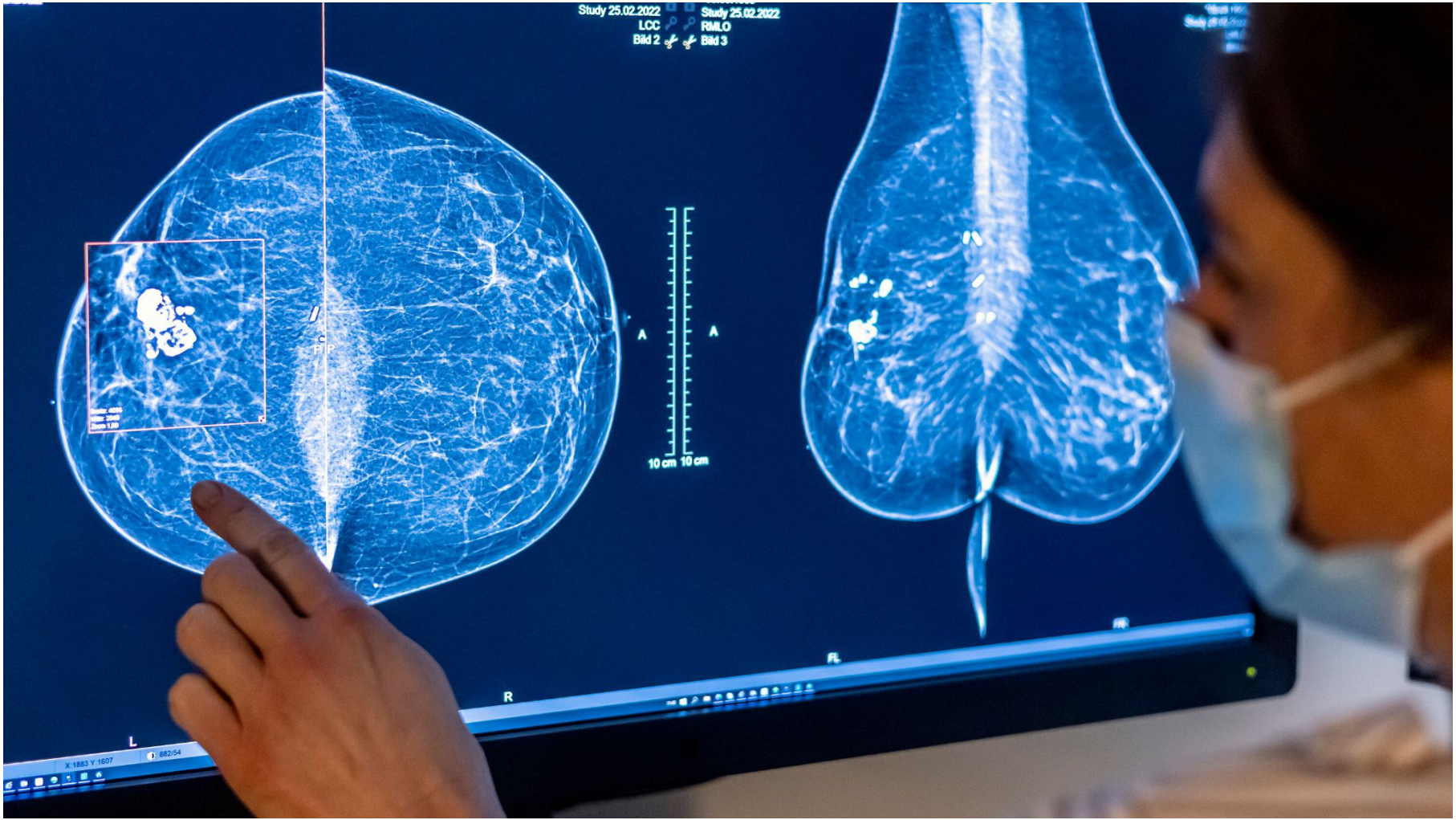
- Defina una métrica de calidad de datos
- Defina un algoritmo para evaluar la calidad de sus datos
- Defina un algoritmo para preprocesar sus datos
    - Datos vacíos
    - Datos duplicados
    - Datos atípicos o anómalos
    - Normalización de datos
    - Gestión de datos categóricos

Referencia de apoyo

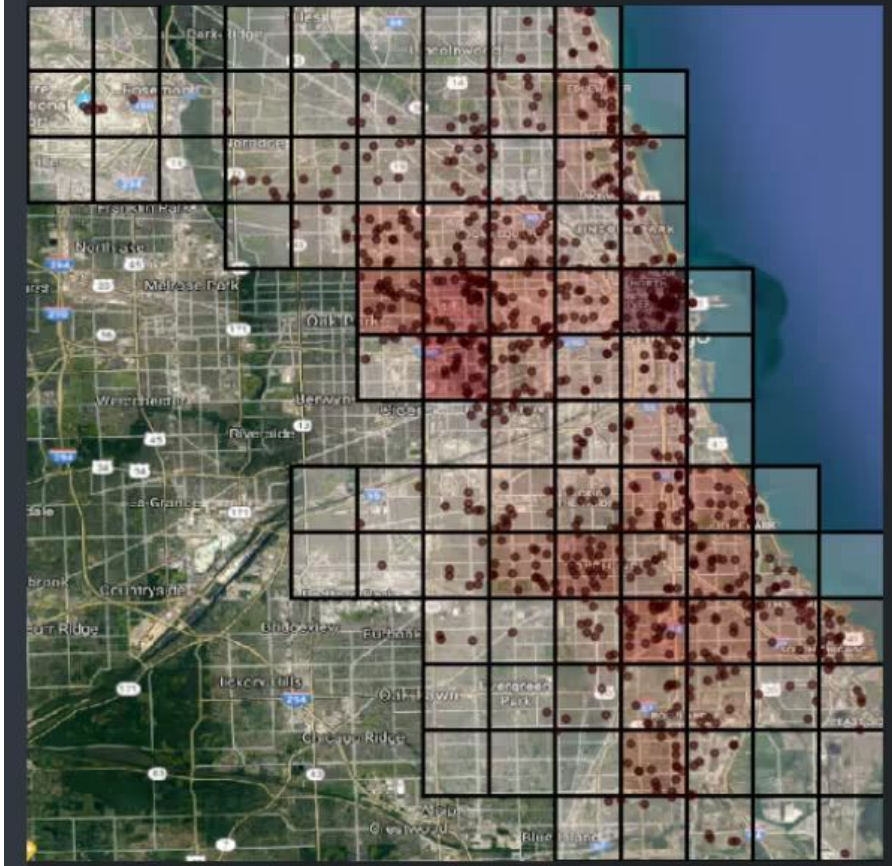
https://www.purestorage.com/es/knowledge/what-is-data-preprocessing.html#:~:text=El%20preprocesamiento%20de%20datos%20para%20el%20aprendizaje%20autom%C3%A1tico%20(ML%20se,la%20precisi%C3%B3n%20de%20los%20modelos.


Regression
Binary Classification
Multiclass Classification
#
1/0
*

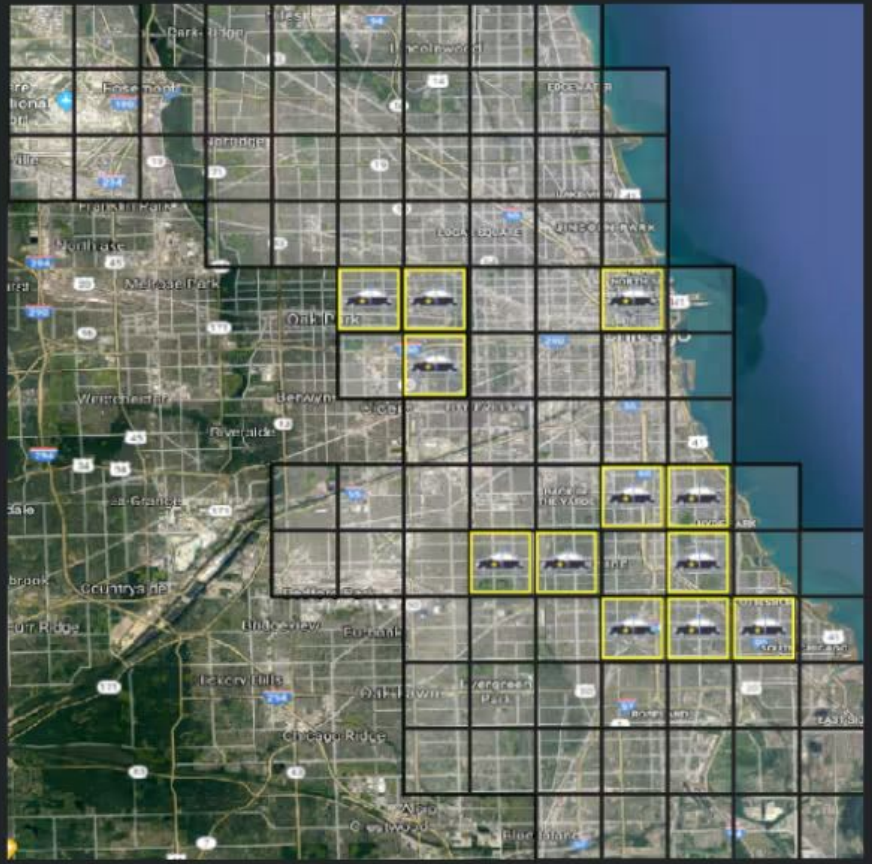
Thinkstock


Study 25.02.2022
LCC
Bild 2
Study 25.02.2022
RMLO
Bild 3
10 cm
10 cm
R
L

Yesterday's hotspots (08/26/2018)



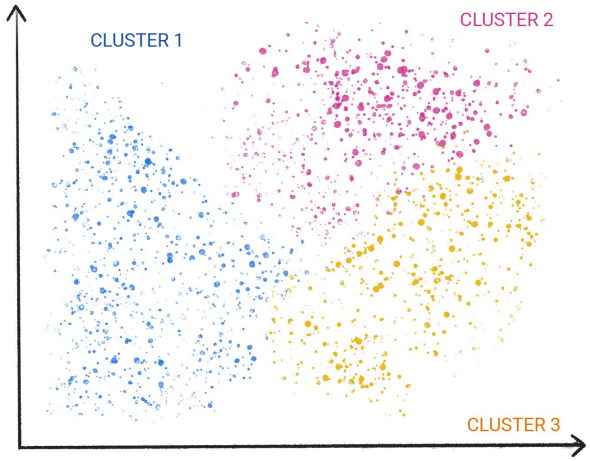
Plan for today (08/27/2018)

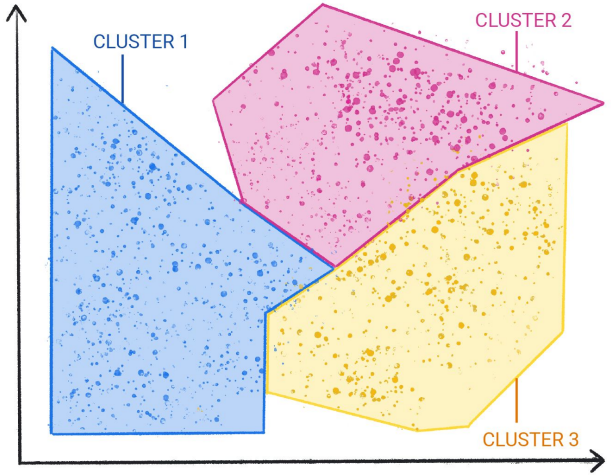


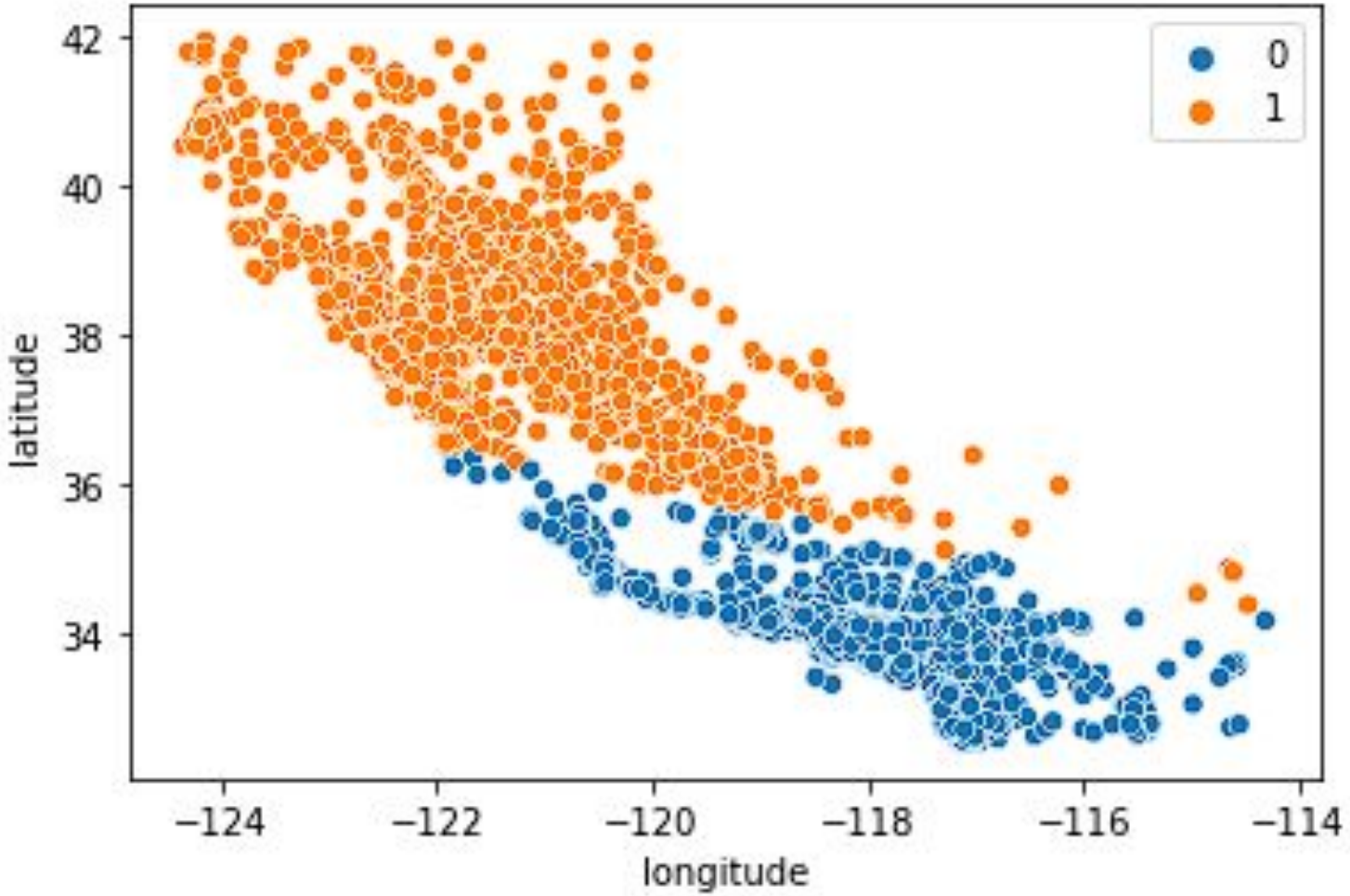
Defiled Sagnosis

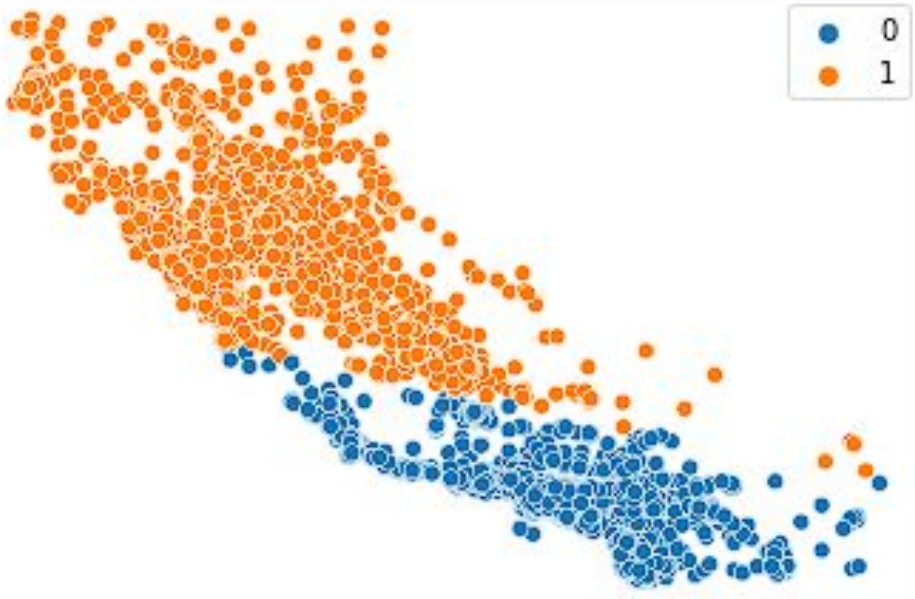
*“**Unsupervised learning** models make predictions by being given data that does not contain any correct answers. An unsupervised learning model's goal is to identify meaningful patterns among the data. In other words, the model has no hints on how to categorize each piece of data, but instead it must infer its own rules.”*

Google AI Team


CLUSTER 1
CLUSTER 2
CLUSTER 3


CLUSTER 1
CLUSTER 2
CLUSTER 3


42
40
38
36
34
latitude
−124
−122
−120
−118
−116
−114
longitude
0
1



1. x → Edad
   y → Calificaciones promedio en el programa de Ingeniería Informática
   Colegio Privado - **Colegio Público**

2. x → Edad
   y → Investigaciones abiertas en la procuraduría
   Partido Político A - **Partido Político B**

3. x → Longitud
   y → Latitud
   Reporte de homicidio - **Reporte de riñas**