



Nov 2025

## Master Class 2025-2 Machine Learning: **Aprendizaje Supervisado**

### Maestría en Matemáticas Aplicadas y Ciencias de la Computación

*Fabián Sánchez Salazar*

## Sobre mi ☺

Matemático, Magister y Doctor en Matemáticas de la Universidad Nacional. Con intereses de investigación en ecuaciones diferenciales parciales, machine learning, procesamiento de lenguaje natural e Inteligencia Artificial. Con más de 17 años de experiencia docente en diferentes universidades de Bogotá y más de 7 años en cargos de dirección de programas académicos. Actualmente se desempeña como profesor principal de carrera en la Escuela de Ingeniería, Ciencia y Tecnología de la Universidad del Rosario, donde también es director de la Maestría en Matemáticas Aplicadas y Ciencias de la Computación.



FABIAN SANCHEZ SALAZAR  
[fabian.sanchez@urosario.edu.co](mailto:fabian.sanchez@urosario.edu.co)

1

# INTRODUCCIÓN



# Conceptos IA

## INTELIGENCIA ARTIFICIAL

Visión por  
Computador

Vehículos  
Autónomos

Robótica

Internet de las  
Cosas

Reconocimiento de  
voz

### MACHINE LEARNING

Aprendizaje  
Supervisado

Aprendizaje  
No  
Supervisado

Aprendizaje Por  
Refuerzo

### DEEP LEARNING

Redes  
Neuronales  
(ANN)

Redes  
Convolucionales  
(CNN)

Redes  
Recurrentes  
(RNN)

LSTM

GANS

GRU

### PROCESAMIENTO DE LENGUAJE NATURAL

IA  
GENERATIVA

LLM

# Conceptos IA

## INTELIGENCIA ARTIFICIAL

Campo de las ciencias de la computación que se enfoca en crear y entrenar máquinas para realizar tareas inteligentes. Imitar lo que hace el humano

## MACHINE LEARNING

Se basa en la creación de algoritmos por medio de los cuales se busca hacer predicciones o reconocer patrones en un conjunto de datos

## DEEP LEARNING

Es un subcampo del machine learning en los cuales los algoritmos están inspirados en el funcionamiento de la neurona humana (redes neuronales)

## PROCESAMIENTO DE LENGUAJE NATURAL

Subcampo de la IA que permite la comunicación entre el humano y la máquina por medio de lenguaje natural



DATOS



# Conceptos IA



## INTELIGENCIA ARTIFICIAL GENERATIVA

Es una rama de la inteligencia artificial que puede crear contenido como texto, audio, vídeo, imágenes, etc.



## LLM LARGE LANGUAGE MODEL

Es un tipo específico de modelo de IA entrenado con grandes volúmenes de texto para procesar y generar lenguaje natural.  
Gpt4-Claude. Llama

# Tipos de Datos

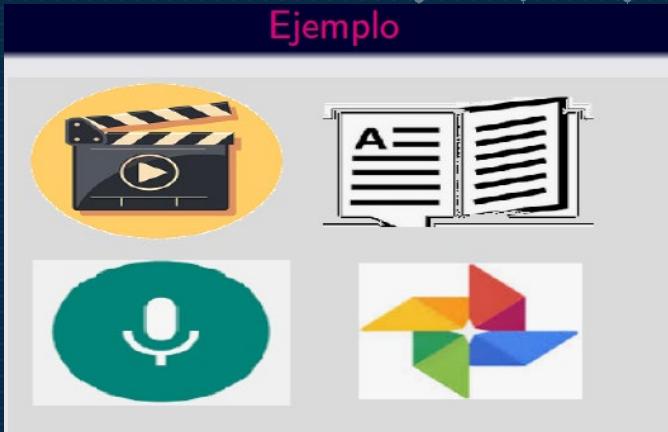
## DATOS ESTRUCTURADOS

Los datos estructurados son los que se pueden representar en una distribución fila-columna

EDAD	SEXO	ESTATURA	NIVEL ESCOLAR
21	M	1,54	MAESTRÍA
26	F	1,55	PROFESIONAL
30	F	1,6	DOCTORADO
31	F	1,7	PROFESIONAL
35	M	1,71	MAESTRÍA

## DATOS NO ESTRUCTURADOS

Es cualquier forma de dato que no encaja en el modelo relacional o conjunto de tablas de bases de datos estructurados.



# Machine Learning

¿Cuándo se debe usar  
algoritmos de Machine  
Learning?



Depende de lo que queremos  
resolver y lo que estamos  
buscando en los datos

## TIPOS DE APRENDIZAJE

Aprendizaje  
Supervisado

Aprendizaje NO  
Supervisado

Aprendizaje por  
Refuerzo

# Aprendizaje Supervisado

Los algoritmos de aprendizaje supervisado intentan modelar la relación y las dependencias entre la salida de predicción objetivo y las características de entrada, de manera que podemos predecir los valores de salida para los nuevos datos basados en esas relaciones, que ha aprendido de conjuntos de datos anteriores alimentados.

## VARIABLES PREDICTORAS

	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes PedigreeFunction	Age	DIABETES
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

VARIABLE  
OBJETIVO

CATEGÓRICA-  
DISCRETA



ALGORITMOS  
DE  
CLASIFICACIÓN

¿Será posible predecir que una paciente tenga diabetes en función de un conjunto de variables predictoras?

# Aprendizaje Supervisado

## VARIABLES PREDICTORAS

index	TV	Radio	Newspaper	Web	Sales
0	230.1	37.8	69.2	306,634	22.1
1	44.5	39.3	45.1	302,653	10.4
2	17.2	45.9	69.3	49,490	9.3
3	151.5	41.3	58.5	257,816	18.5
4	180.8	10.8	58.4	195,660	12.9

VARIABLE  
OBJETIVO

NUMÉRICA  
CONTINUA



ALGORITMOS  
DE REGRESIÓN

¿Será posible predecir las ventas en función de la inversión destinada a diferentes medios de pauta publicitaria?



# Flujo general de ML en aprendizaje supervisado



## Aprendizaje NO Supervisado

Se busca encontrar patrones en el conjunto de datos. Pero en este caso no tenemos una variable objetivo definida (datos sin etiquetar)

Nombre	Sexo	Estrato	Fuma	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes PedigreeFunction	Age	Diabetes
Juan	M	3	1	148	72	35	0	33.6	0.627	50	1
Carlos	M	2	0	85	66	29	0	26.6	0.351	31	0
Ana	F	3	1	183	64	0	0	23.3	0.672	32	1
Maria	F	4	1	89	66	23	94	28.1	0.167	21	0
Hugo	M	3	0	137	40	35	168	43.1	2.288	33	1

¿Se pueden agrupar las personas de acuerdo con las características que no se pueden observar a simple vista?



CLUSTERING

## Aprendizaje NO Supervisado

Nombre	Sexo	Estrato	Fuma	Glucose	Blood Pressure
Juan	M	3	1	148	72
Carlos	M	2	0	85	66
Ana	F	3	1	183	64
María	F	4	1	89	66
Hugo	M	3	0	137	40

...

Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age
35	0	33.6	0.627	50
29	0	26.6	0.351	31
0	0	23.3	0.672	32
23	94	28.1	0.167	21
35	168	43.1	2.288	33

100 VARIABLES



¿Se puede reducir el número de variables tratando de conservar la mayor cantidad de información posible?



**REDUCCIÓN DE  
DIMENSIONALIDAD**

2

# EDA (Exploratory Data Analysis)



# Data Frames

**VARIABLES**

**COLUMNA  $C_j$**

PERSONA	EDAD	SEXO	ESTATURA	NIVEL ESCOLAR	MARCA DE AUTO	NÚMERO DE HIJOS	SALARIO
PERSONA 1	21	M	1,54	MAESTRÍA	AUDI	0	1200000
PERSONA 2	26	F	1,55	PROFESIONAL	RENAULT	5	1250000
<i>FILA <math>F_i</math></i>	<i>PERSONA 3</i>	30	F	1,6	DOCTORADO	BMW	2
	PERSONA 4	31	F	1,7	PROFESIONAL	RENAULT	2
	PERSONA 5	35	M	1,71	MAESTRÍA	AUDI	1
	PERSONA 6	65	M	1,8	MAESTRÍA	AUDI	1
	PERSONA 7	45	M	1,54	MAESTRÍA	BMW	1



# Análisis Exploratorio de Datos

## Paso 1

Preguntas a responder



## Paso 3

Tipos de variables



## Paso 5

Visualización-Gráficas



## Paso 7

Reporte de conclusiones y resultados



Generalidades del conjunto de datos

## Paso 2

Estadística Descriptiva

## Paso 4

Relación entre variables

## Paso 6

Datos faltantes



ANÁLISIS

Datos Atípicos



MEDIA

MEDIANA

MODA

VARIANZA

MÁXIMO

MÍNIMO

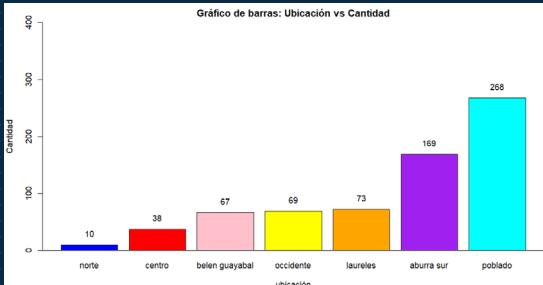
CUARTILES

DESVIACIÓN ESTANDAR

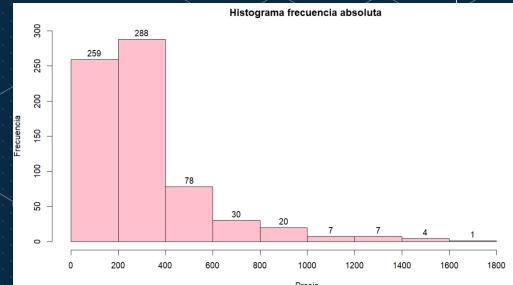
CORRELACIÓN

# EDA: Medidas Estadísticas

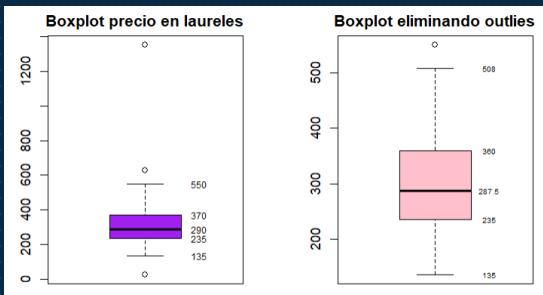
## GRÁFICO DE BARRAS



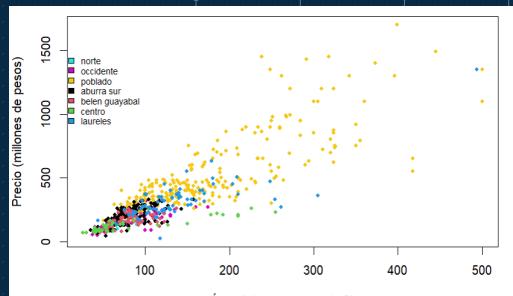
## HISTOGRAMA



## BOXPLOT



## DIAGRAMA DE DISPERSIÓN



3

# Proceso de Machine Learning Supervisado





# Identificación de Variables

1

VARIABLE  
OBJETIVO

VARIABLES  
PREDICTORAS

$y$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	...	$X_n$

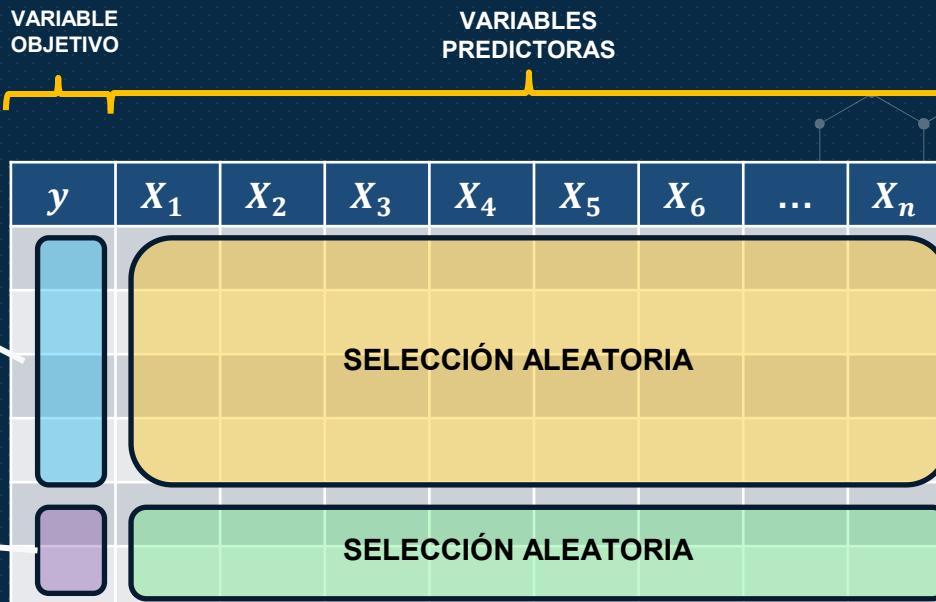


2

# Conjunto de Entrenamiento y Prueba

Conjunto de  
entrenamiento en y  
 $y_{train}$

Conjunto de Prueba  
en y  
 $y_{test}$



*La proporciones recomendadas para entreamiento y prueba*

70% – 30%

80% – 20%

75% – 25%



# Escalar las Variables

3

VARIABLE  
OBJETIVOVARIABLES  
PREDICTORAS

$y$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$\dots$	$X_n$
	$X_{train}$							
	$X_{test}$							

transform

$$\frac{X_{test} - \mu}{\sigma}$$

$$\frac{\mu}{\sigma}$$

fit transform

$$\frac{X_{train} - \mu}{\sigma}$$

# Selección de Modelos

4

## Clasificación

1. K-Veinos más cercanos (kNN)
2. Regresión Logística
3. Árboles de Decisión (classifier)
4. Random Forest
5. Naive Bayes
6. Support vector machine (SVC)
7. Gradient Boosting
8. Multilayer Perceptron (MLP)
9. Redes Neuronales

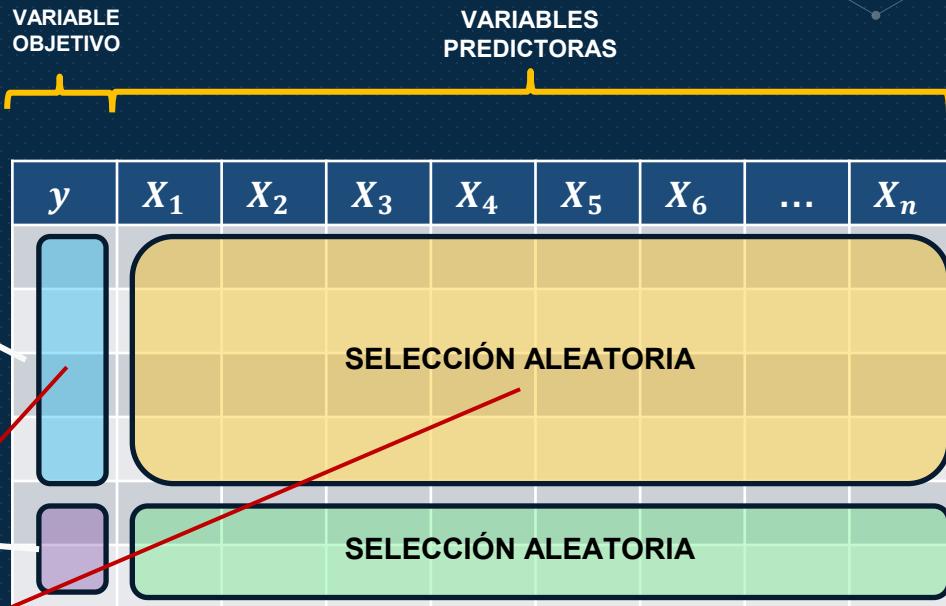
## Regresión

1. k-Veinos más cercanos (kNN)
2. Regresión Lineal (Simple-Múltiple)
3. Árboles de Decisión (regressor)
4. Random Forest
5. Support vector machine (SVR)
6. Gradient Boosting
7. Multilayer Perceptron (MLP)
8. Redes Neuronales



# Entrenar el Modelo

5



Entrenamos el Modelo  
con estos conjuntos

$Modelo(X_{train}, y_{train})$

Tomamos el modelo  
entrenado y lo  
evaluamos en  $X_{test}$

$y_{pred} = \text{Modelo\_entrenado}(X_{test})$



# Evaluación del Modelo

6

Tomamos el modelo y  
lo evaluamos en  $X_{test}$

$$y_{pred} = \text{Modelo\_entrenado}(X_{test})$$



Vamos a comparar los valores  
que nos da el modelo en  $X_{test}$   
y los valores reales  $y_{test}$

## Clasificación

Matriz de Confusión  
+  
Métricas

- Accuracy
- Recall
- Specificity
- Precision
- F1-Score

## Métricas

- Mse (error cuadrático medio)
- $R^2$  coeficiente de determinación
- RMSE

## Regresión

# Validación Cruzada

7

ITERACIÓN 1



EVALUACIÓN DEL MODELO

Accuracy 1

ITERACIÓN 2



Accuracy 2

ITERACIÓN 3



Accuracy 3

ITERACIÓN 4



Accuracy 4

ITERACIÓN 5



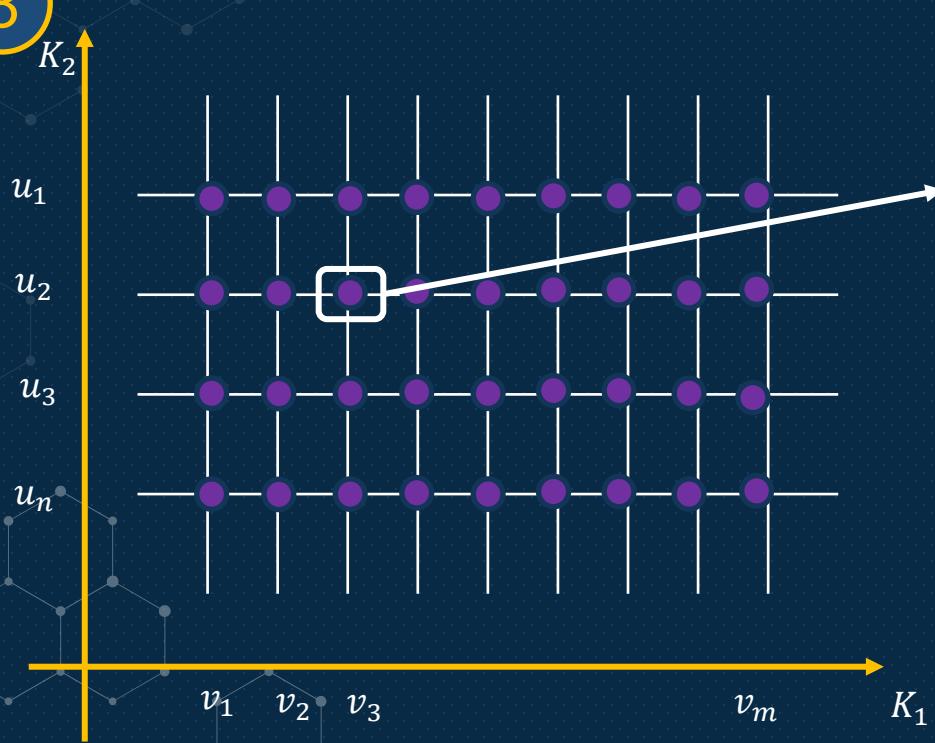
Accuracy 5

Se promedian los accuracy obtenidos

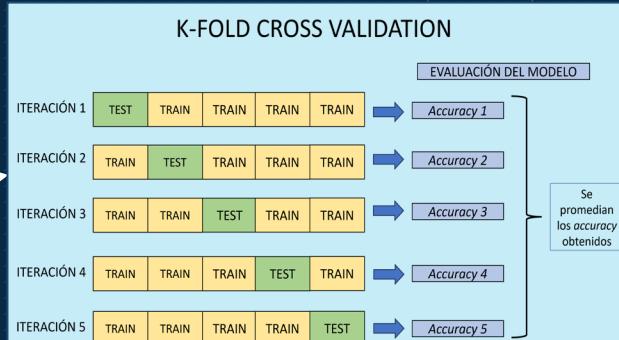
# Optimización de Hiperparámetros

## Grid Search

8



Se selecciona el mejor modelo



Se hace validación cruzada en cada combinación de parámetros

4

# Ideas Intuitivas de Algunos Modelos de Clasificación

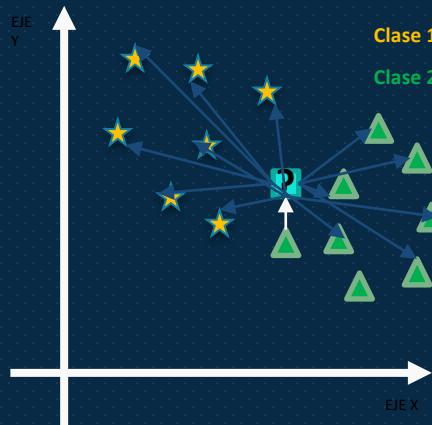


# K-vecinos más cercanos

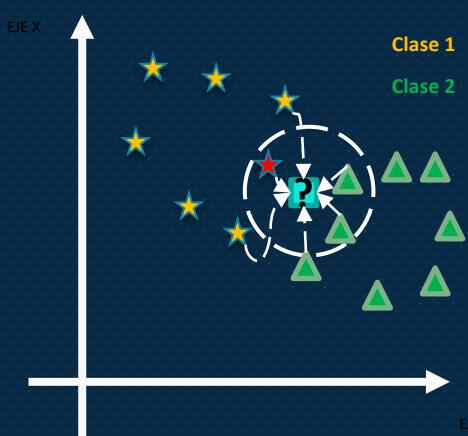
## DATO INICIAL



## CALCULAR LAS DISTANCIAS



## ENCONTRANDO LOS VECINOS Y VOTANDO



Se calculan todas las distancias del cuadro a todos los otros puntos y se ordenan las distancias de menor a mayor

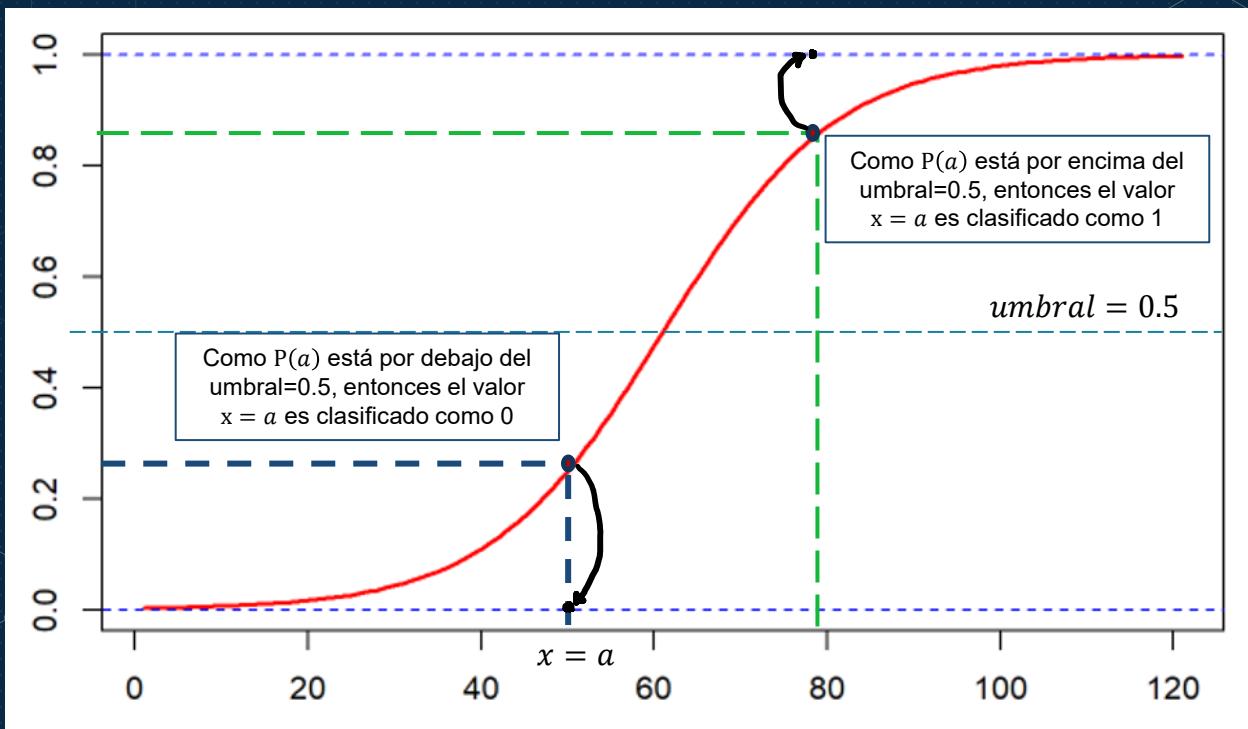
Los vecinos más cercanos serán aquellos que tengan distancias más pequeñas.

Se elige un número determinado de vecinos ( $k$ ).

La clasificación se hace de acuerdo con la clase mayoritaria (mayor número de elementos)

# Regresión Logística

$$P(a) = P(y = 1|x = a) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 a)}}$$

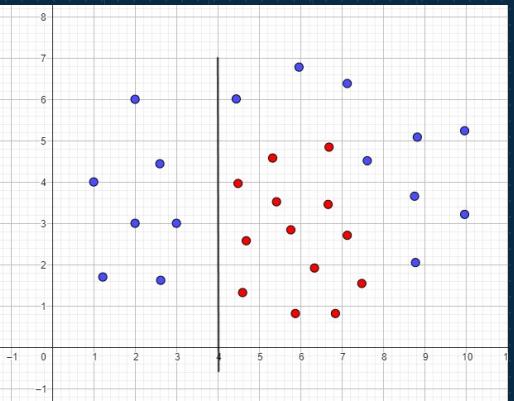


Función Sigmoidal

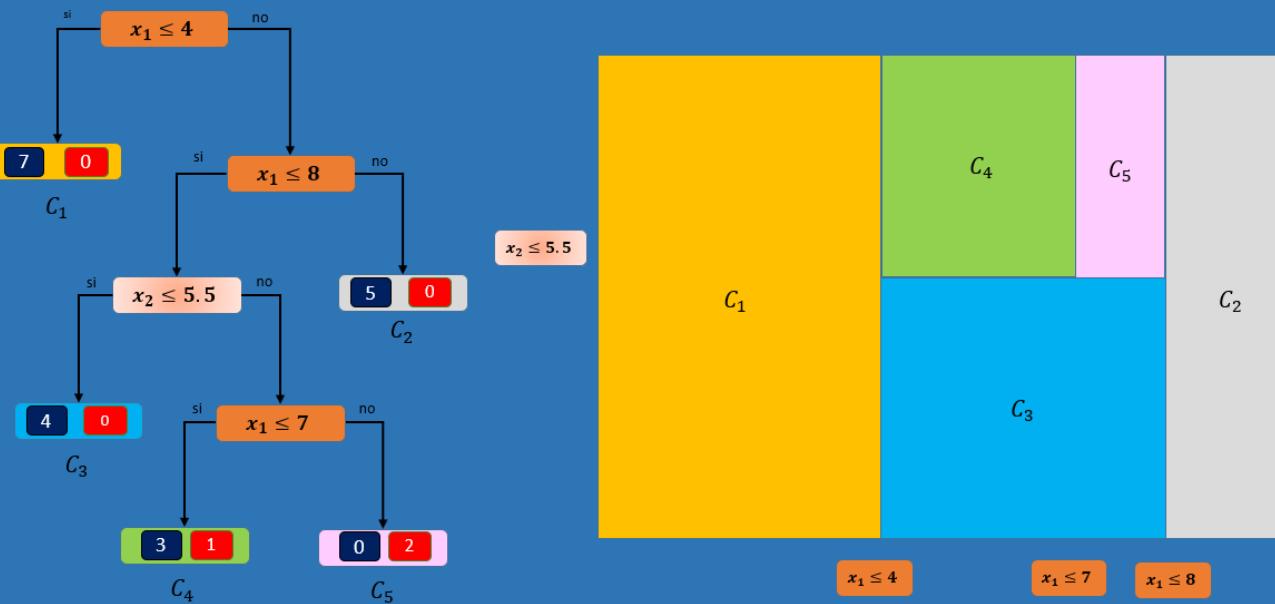
$$f(x) = \frac{1}{1 + e^{-x}}$$



# Árboles de Decisión



¿Cómo podemos separar estos puntos con líneas verticales u horizontales?



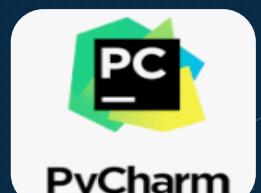
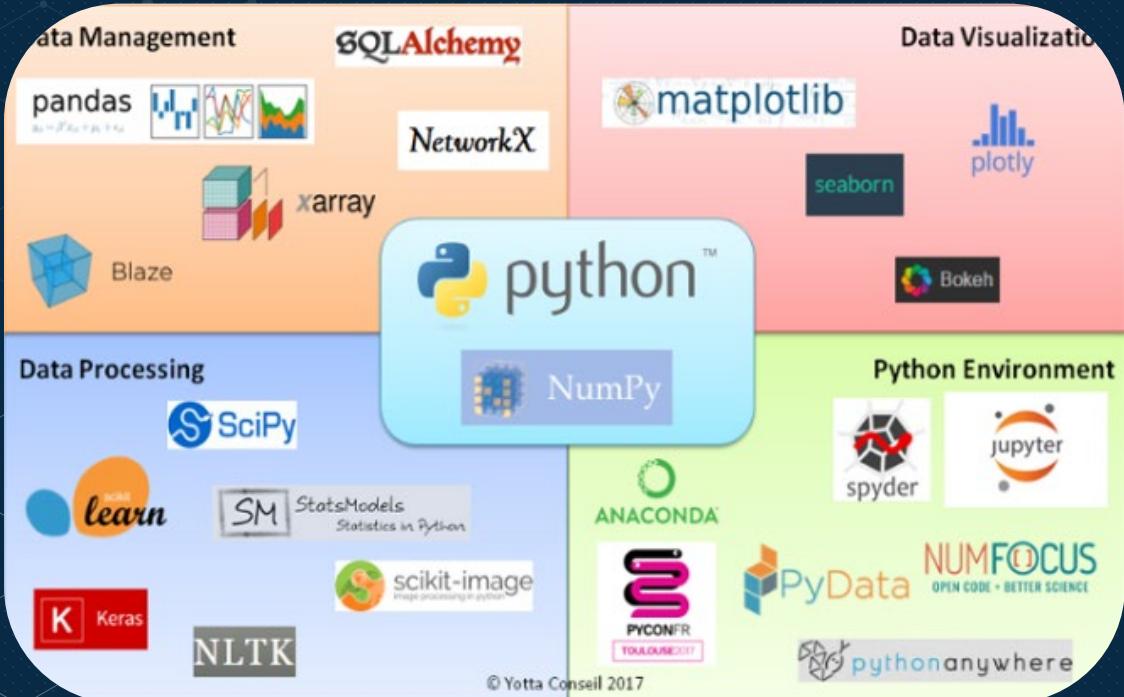
5

# SOFTWARE

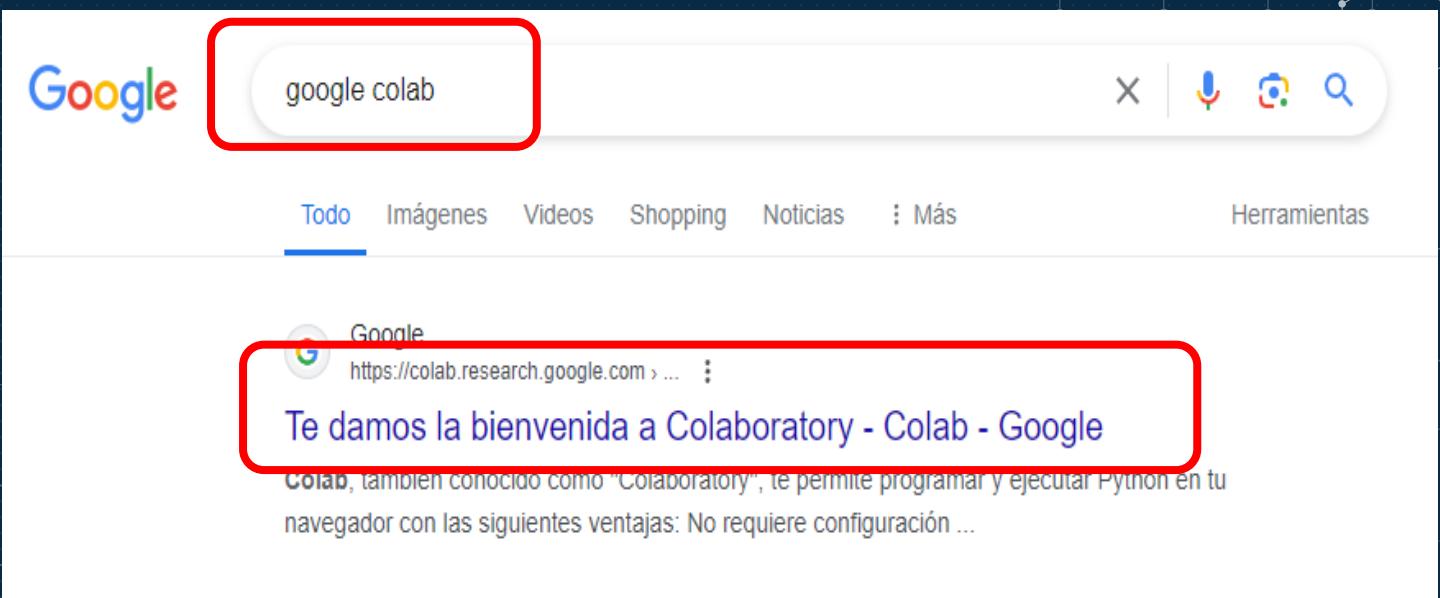




# PYTHON



# PYTHON+COLAB



The screenshot shows a Google search results page. The search bar at the top contains the query "google colab", which is highlighted with a red rectangle. Below the search bar, there are tabs for "Todo", "Imágenes", "Videos", "Shopping", "Noticias", and "Más". To the right of these tabs is a "Herramientas" section. The main search results area features a card from Google Colaboratory. This card has a red border and contains the following text:  
**Te damos la bienvenida a Colaboratory - Colab - Google**  
Colab, también conocido como "Colaboratory", te permite programar y ejecutar Python en tu navegador con las siguientes ventajas: No requiere configuración ...  
The URL https://colab.research.google.com is visible above the card.



Te damos la bienvenida a Colaboratory

Archivo Edar Ver Insertar Entorno de ejecución Herramientas Ayuda

Compartir Iniciar sesión

Índice

Primeros pasos

Ciencia de datos

Aprendizaje automático

Más recursos

Ejemplos destacados

+ Sección

+ Código + Texto Copiar en Drive

Conectar

## Te damos la bienvenida a Colab

### (Novedad) Prueba la API de Gemini

- [Generate a Gemini API key](#)
- [Talk to Gemini with the Speech-to-Text API](#)
- [Gemini API: Quickstart with Python](#)
- [Gemini API code sample](#)
- [Compare Gemini with ChatGPT](#)
- [More notebooks](#)

Si ya conoces Colab, echa un vistazo a este vídeo para obtener información sobre las tablas interactivas, la vista del historial de código ejecutado y la paleta de comandos.



19

# PYTHON+COLAB

Te damos la bienvenida a Colaboratory

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

Nuevo cuaderno 1

Índice

Ctrl+O

+ Texto Copiar en Drive

mos la bienvenida a

dad) Prueba la API de G

erate a Gemini API key

2

Es necesario iniciar sesión en Google

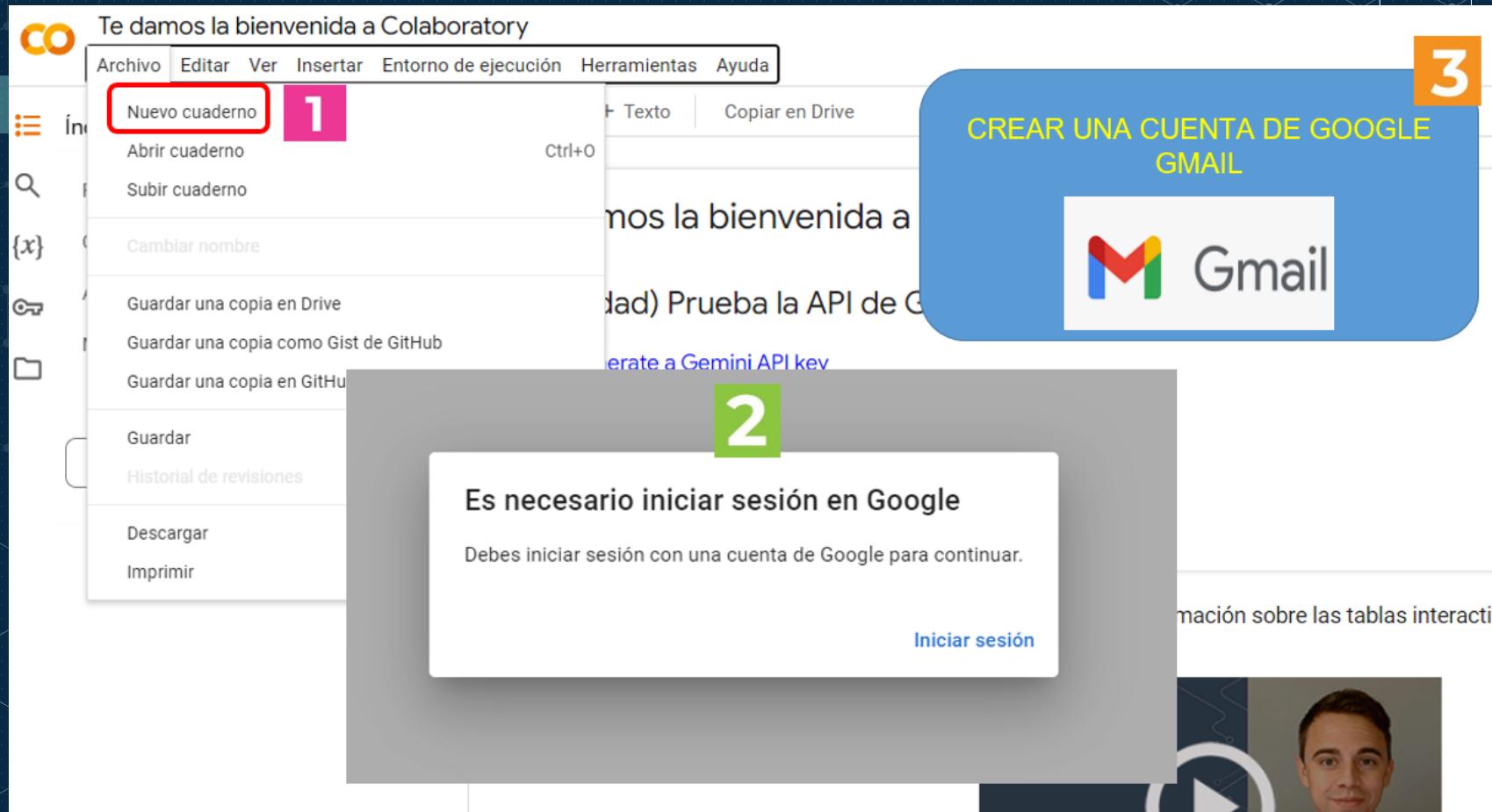
Debes iniciar sesión con una cuenta de Google para continuar.

Iniciar sesión

3

CREAR UNA CUENTA DE GOOGLE GMAIL

Gmail



Te damos la bienvenida a Colaboratory

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

Nuevo cuaderno

Abrir cuaderno Subir cuaderno Cambiar nombre

Guardar una copia en Drive Guardar una copia como Gist de GitHub Guardar una copia en GitHub

Guardar Historial de revisiones

Descargar Imprimir

Ctrl+O Ctrl+S Ctrl+P

Texto Copiar en Drive

mos la bienvenida a Colab

dad) Prueba la API de Gemini

[Generate a Gemini API key](#)  
[Talk to Gemini with the Speech-to-Text API](#)  
[Gemini API: Quickstart with Python](#)  
[Gemini API code sample](#)  
[Compare Gemini with ChatGPT](#)  
[Share notebooks](#)

Si ya conoces Colab, echa un vistazo a este vídeo para obtener información sobre las tablas interactivas ejecutado y la paleta de comandos.



# PYTHON+COLAB

The screenshot shows the Google Colab interface. At the top, there's a navigation bar with the title "Untitled0.ipynb" and icons for File, Edit, View, Insert, Runtime, Tools, and Help. To the right of the title are buttons for Comment, Share, and Settings, along with a user profile picture. A red box highlights the "Conectar" button in the toolbar. A large blue arrow points upwards from the bottom of the screen towards this button. On the left side, there are sidebar icons for Code (+ Código), Text (+ Texto), and various data types like {x}, Keys, and Folders. The main workspace has a message: "Empieza a programar o a crear código con IA." Below the workspace are standard Colab toolbar icons: up arrow, refresh, comment, settings, copy, delete, and more.



## PYTHON+COLAB

¡Ahora, a trabajar en Colab!



# Gracias!

[fabian.sanchez@urosario.edu.co](mailto:fabian.sanchez@urosario.edu.co)