

# Machine Learning en Economía

## Sesión 1: Aproximaciones en Perú

Luis Chávez



Lima, 2025

# Table of Contents

- 1 Introducción
  - Conceptualización
  - Generalidades
- 2 Caso Peruano
- 3 Herramientas
- 4 Iniciando en Python
  - Acerca de
  - Instalación
  - Cuadernos

# Introducción

# Conceptos básicos

Machine Learning  
Deep Learning  
Inteligencia Artificial  
Minería de Datos  
Big Data  
Análisis de Datos  
Ciencia de Datos

# Conceptos básicos

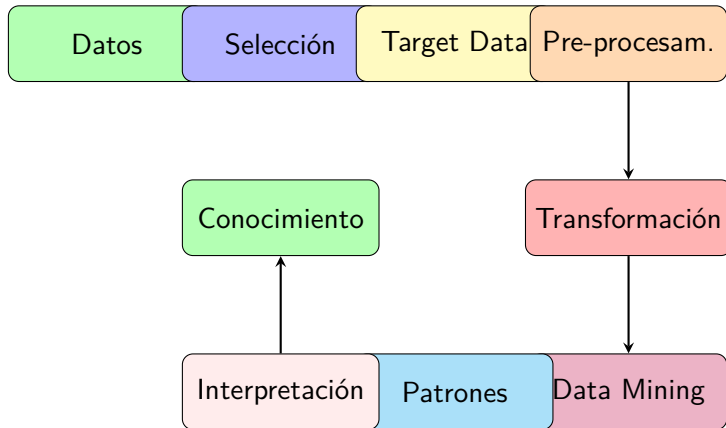
## Definición (Ciencia de datos)

Enfoque multidisciplinario que combina las matemáticas, la estadística, la programación y la IA para convertir los datos en conocimiento (IBM).

Hall Varian → Chief Economist Google.

# Conceptos básicos

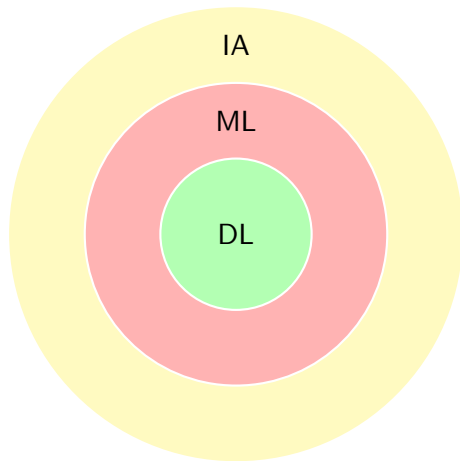
En resumen, el proceso D2K:



# Conceptos básicos



# Conceptos básicos





# Conceptos básicos

## Definición (IA)

Toda tecnología que desarrolla la inteligencia de las máquinas para crear sistemas con la finalidad de imitar algunas habilidades humanas.

## Definición (ML)

Ciencia que permite a los ordenadores, vía algoritmos y modelos estadísticos, aprender de los datos y realizar tareas automáticamente (sin ser programados previamente).

## Definición (DL)

Ciencia que utiliza redes neuronales artificiales para filtrar observaciones (imágenes, textos o sonidos) en capas para aprender a predecir o clasificar información (como el cerebro humano).

# Conceptos básicos

Small data:

- 1 Maximizar de **pocos** datos.
- 2 Solución: estructurar los datos.
- 3 Enfoque: muestreo complejo aproxima muestreo al azar (lento y caro, pero bueno).

Big data:

- 1 **Muchos** datos no estructurados.
- 2 Extracción inmediata.
- 3 Relativamente baratos.

# Conceptos básicos

$$y \leftarrow f(x, u)$$

Estadística clásica:

- 1 Interés en estimar el modelo  $f(\cdot)$ . Efecto causal.
- 2 Modelo: teoría o experimento (**out of trouble**).

Machine Learning:

- 1 Interés en  $y$ . Predecir, medir, clasificar.
- 2 Modelo: construirlo y aprenderlo.
- 3 Predicción puntual (no inferencia).

- 1 Lenguaje supervisado: el conjunto de datos contiene tanto entradas (observables) como resultados deseados (target), cuyo objetivo es la predicción.
- 2 Lenguaje no supervisado: el conjunto de datos contiene solo entradas, cuyo objetivo es explorar los datos y encontrar patrones o mecanismos de organización.
- 3 Lenguaje por refuerzo: determina automáticamente el comportamiento ideal en un contexto específico.

# Taxonomía



# Taxonomía

- 1 Tareas: predicción o descripción.
- 2 Técnicas: algoritmos o procedimientos estadísticos específicos.
- 3 Herramientas: software, paquetes, librerías.

# Antecedentes

- 1 1950: Alan Turing crea el “Test de Turing”.
- 2 1952: primer programa capaz de aprender.
- 3 1967: se crea el algoritmo “Nearest Neighbor”.
- 4 1997: el ordenador Deep Blue (IBM) vence al campeón mundial de ajedrez Gary Kaspárov.
- 5 2006: se acuña deep learning.
- 6 2015: Elon Musk y Sam Altman fundan OpenAI.
- 7 2022: nace chat GTP.

## Ejemplo 1 (filtros spam)

El término *spam* se remonta a la serie cómica británica Monty Python, en la cual una pareja en un restaurante recibía sin razón alguna spam (Spiced Pork And Meat). Los correos spam han sido solucionado con filtros (Bayes o Markov).



## Ejemplo 2 (reconocimiento facial)

Algoritmos utilizan un gran conjunto de datos de fotografías etiquetadas según tienen un rostro o no para estimar una función  $f(x)$  que predice la presencia  $y$  de un rostro a partir de los píxeles  $x$ .

¿Estos algoritmos sólo aplican técnicas estándar a conjuntos de datos nuevos y grandes?  
¿cómo encajan con lo que sabemos? Como economistas, ¿cómo podemos utilizarlos?  
**Resuelve más problemas, entre ellos el volumen de datos: big data.**

## ¿Porqué utilizar ML?

- 1 Logra descubrir patrones generalizables (inferencia).
- 2 Ajusta formas funcionales complejas (underfitting).
- 3 Reduce sesgo de selección en evaluación de intervenciones.
- 4 Permite tratar grandes volúmenes de datos.
- 5 ...

## ¿Impacto de ML?

- 1 Nuevos modelos econométricos.
- 2 Nuevos métodos de identificación de efecto causal.
- 3 Investigaciones interdisciplinarias.
- 4 Automatización de procesos.
- 5 ...

Adds: <https://n9.cl/dfa62>

## Definición (entrenamiento de modelo)

Proceso que consiste en proporcionar datos a un algoritmo para que aprenda determinados patrones. Puede realizar predicciones, evidenciar estructuras entre otros.

# Implicancias

## Underfitting y overfitting:

### Underfitting

Entreno al modelo con  
1 sola raza de perro



Muestra nueva:  
¿Es perro?



NO  
FALLO

La máquina fallará en reconocer al perro por falta de suficientes muestras. No puede generalizar el conocimiento.

### Overfitting

Entreno al modelo con  
10 razas de perro color marrón



Muestra nueva:  
¿Es perro?



NO  
FALLO

La máquina fallará en reconocer un perro nuevo porque no tiene estrictamente los mismos valores de las muestras de entrenamiento.

Figure: Extraído de Aprende ML

## Caso Peruano

# Introducción



Figure: Lab-IA UP

<https://ia.up.edu.pe/>

## Caso peruano

Nº	Institución	Aplicación	ML
1	IGP	Pronóstico metereológico de precipitación	Downscaling
2	MIDIS y UPCH	Detección de anemia en niños	Deep Learning
3	Banca Comercial	Modelos de riesgo crediticio y consumo	Deep Learning
4	ESSALUD	Predicción de sobrepeso y obesidad	Redes neuronales
5	BCRP	Proyecciones de Inflación	Random Forest

Table: Algunos tratados en Perú



## Herramientas

# Plataformas

- Microsoft Azure → Azure Machine Learning.
- Anaconda Navigator → Anaconda, Jupiter, PyCharm, R.
- IBM Watson → watsonx.
- Amazon Web Services (AWS) → Amazon Sagemaker.
- Apache spark.
- Vertex AI Platform (Google).

# Plataformas

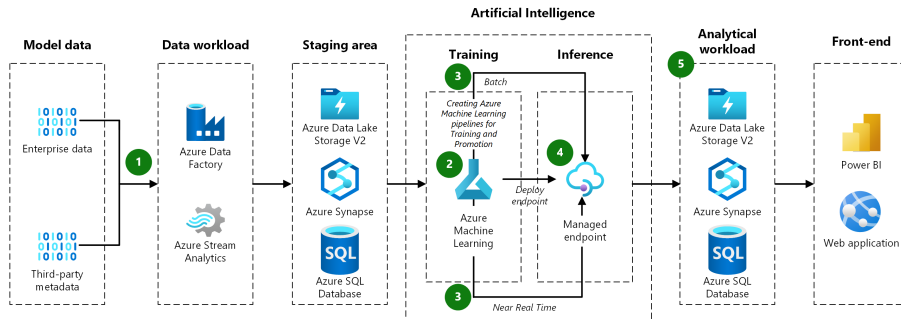


Figure: Azure



Figure: IBM

# Plataformas

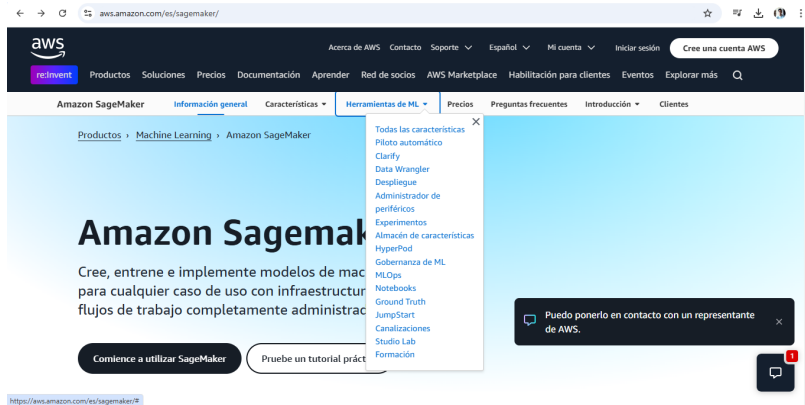


Figure: AWS

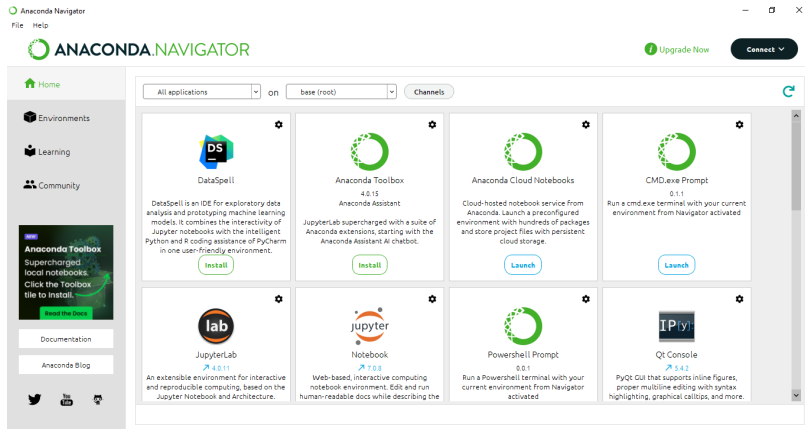


Figure: Anaconda Navigator

# Lenguajes de Programación

- Python.
- R.
- Julia.
- Scala.
- MATLAB ML.
- C++.
- ...

# Lenguajes de Programación

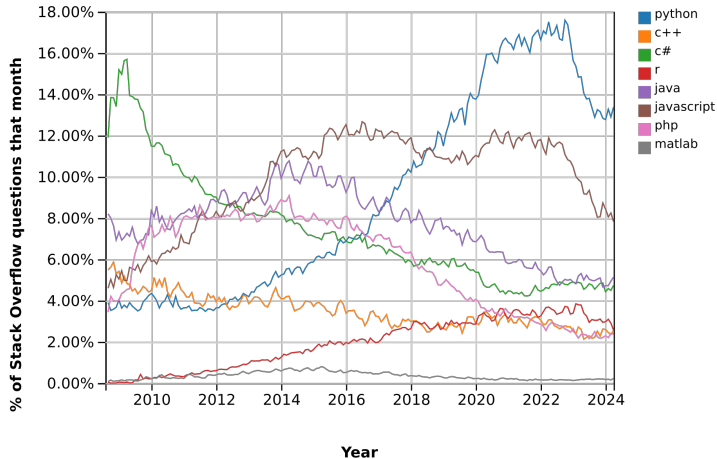


Figure: Tendencia actual (ver)



# Paquetes

- TensorFlow.
- Keras.
- Scikit-learn.
- MLib
- PyTorch
- ...

# Iniciando en Python

# Acerca de

Fundado por Guido Van Rossum en 1989. Inspirado en el programa de televisión de la BBC “Monty Python’s Flying Circus”.

## Definición (Python)

Es un lenguaje de programación orientado al desarrollo de software, ciencia de datos y ML.

# Características

- Lenguaje interpretado.
- Lenguaje sencillo.
- Orientado a objetos (POO).
- Bibliotecas dinámicas.

- Jupiter: <https://jupyter.org/>.
- PyCharm: <https://www.jetbrains.com/es-es/pycharm/>.
- IDLE: madre.
- Sublime Text: <https://www.sublimetext.com/>.
- Spyder: <https://www.spyder-ide.org/>.
- Atom: <https://atom-editor.cc/>.
- Google Colab: <https://colab.research.google.com/>.

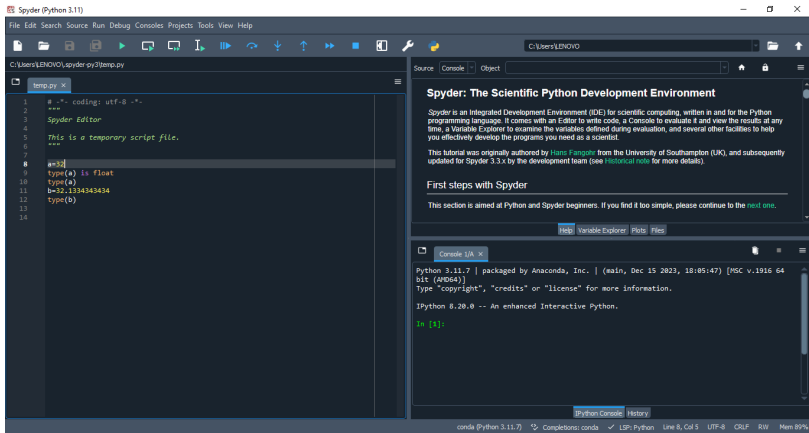


Figure: Spyder

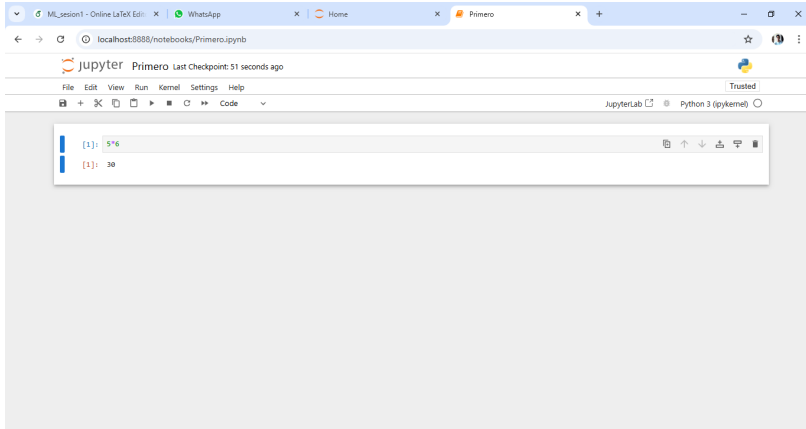


Figure: Jupyter

# Instalando Colab

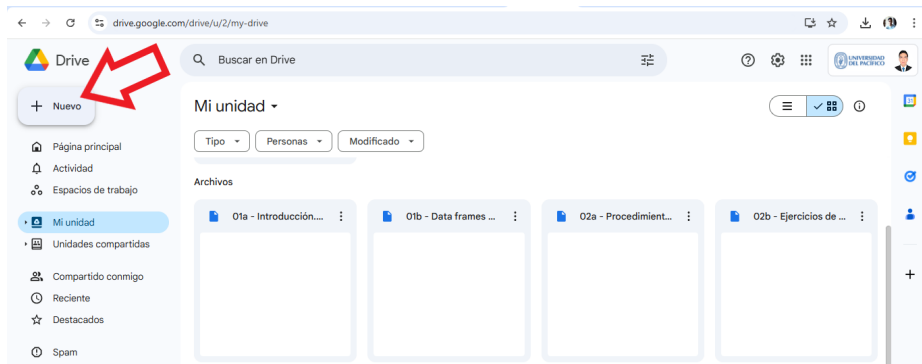


Figure: Paso 1



# Instalando Colab

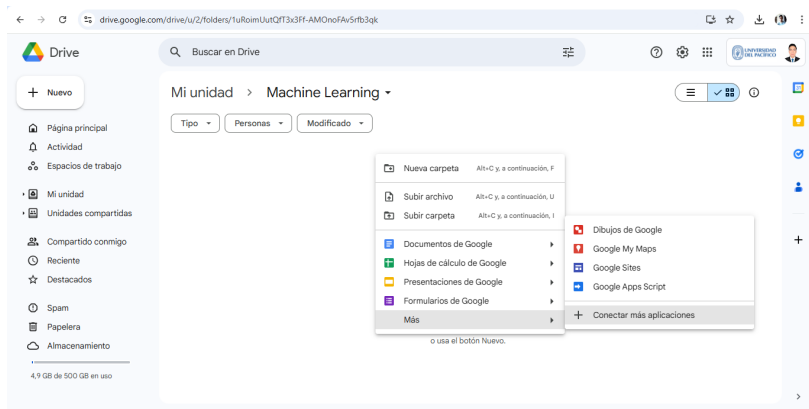


Figure: Paso 2

# Instalando Colab

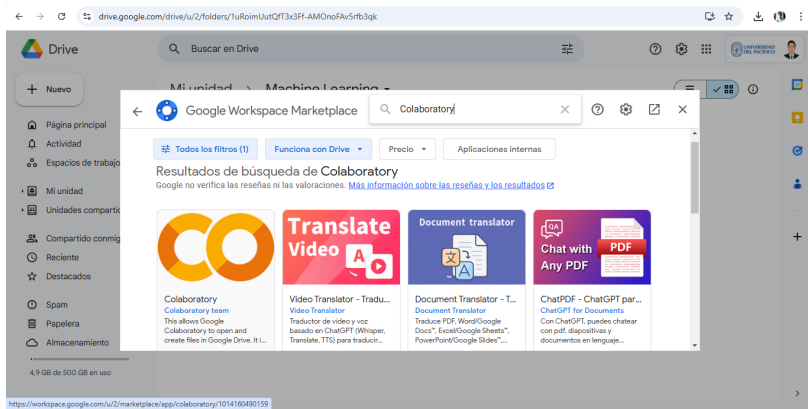


Figure: Paso 3

# Instalando Colab

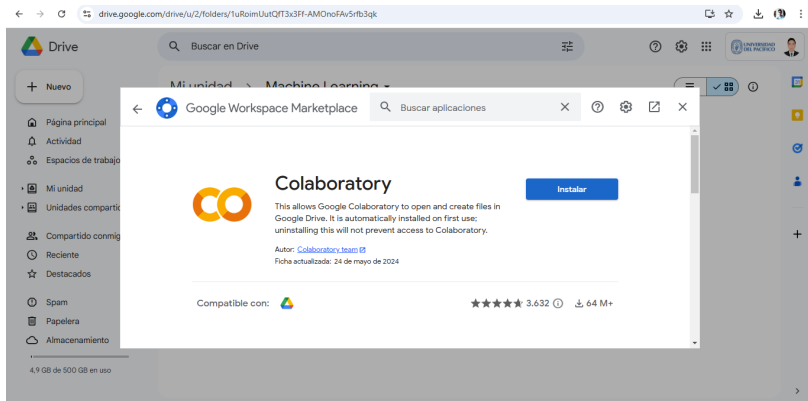


Figure: Paso 4

# Instalando Colab

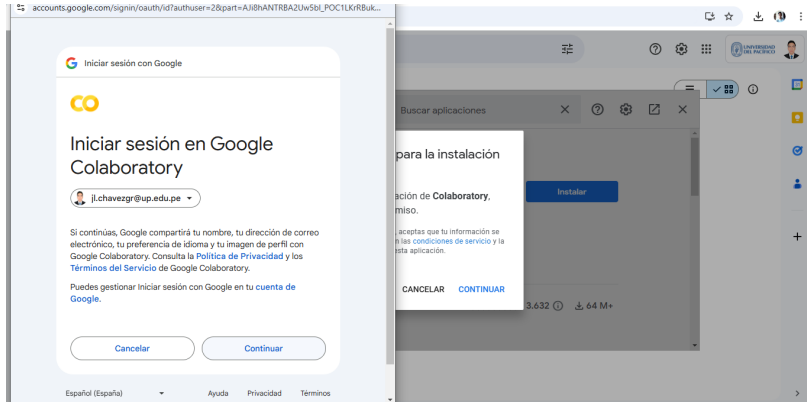


Figure: Paso 5

# Instalando Colab

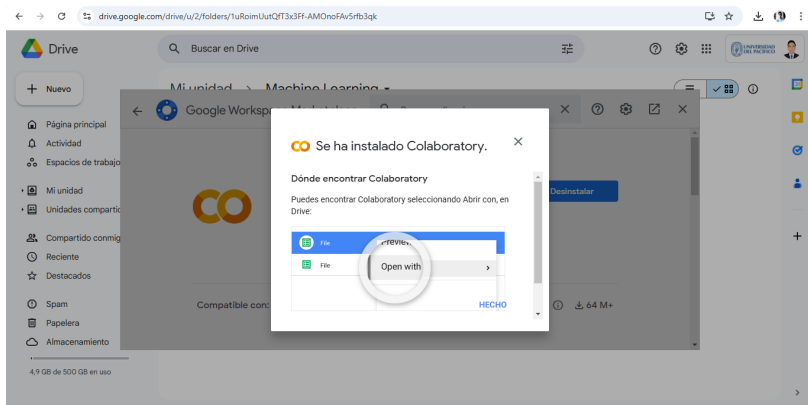


Figure: Paso 6

# Cuadernos

Los ficheros con extensión *.py* corresponden al script de python escritos en formato de texto plano (código fuente), mientras que *.ipynb* corresponden a un **python notebook** escrito en un IDE (Integrated Development Environment).

---

```
# Primer código en Python
```

```
print((2+5)*3)
```

```
21
```

---

A trabajar...

# Referencias

- Géron, A. (2023). Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow. O'Reilly.
- Sosa, W. (2019). Big Data. Siglo XXI.
- <https://quantecon.org/>.

¡Gracias!