



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

Diplomado en Programación y Aplicaciones de Python

Aplicaciones en Ciencia de Datos e
Inteligencia Artificial

Profesor: Francisco Pérez Galarce





Objetivos del curso

- Conocer las problemáticas y técnicas asociadas a la Ciencia de Datos e Inteligencia Artificial.
- Identificar las principales librerías de Python orientadas al análisis y visualización de datos.
- Aplicar las principales librerías Python orientadas al aprendizaje de máquina en conjuntos de datos reales.
- Implementar programas en Python basados en técnicas de aprendizaje de máquina.



Equipo



Francisco Pérez Galarce
Profesor encargado



Yesenia H. Salinas
Ayudante



Metodología y evaluaciones



Metodología de enseñanza y aprendizaje

Clases expositivas, donde se presentan conceptos, aspectos teóricos y aplicaciones de ejemplo.



Ejercicios prácticos aplicando conceptos vistos en clases

40% Clases expositivas

60% Clases prácticas



Evaluación de los aprendizajes:

Evaluación escrita de conceptos 20%

Desarrollo de tareas de programación 80%



Evaluaciones

Evaluación escrita de conceptos **20%**

- 2 controles escritos (contenido teórico e implementación) 10%
- Prueba final de finalización del curso 10%

Desarrollo de tareas de programación **80%**

- 2 actividades de implementación en clases 20%
- 2 Mini proyectos 40%
- Repositorio en Github 20%



Fechas de evaluaciones

Fecha	Actividad/Evaluación
29-10-24	Introducción al aprendizaje de máquina: exploración y procesamiento de datos con Python Actividad 1 (No evaluada)
05-11-24	Aprendizaje supervisado con Python : regresiones Actividad 2 (Evaluada)
12-11-24	Aprendizaje supervisado con Python : naive Bayes y métricas de evaluación Control 1
19-11-24	Aprendizaje supervisado con Python : decision tree, random forest Mini Proyecto 1
26-11-24	Aprendizaje supervisado con Python : KNN, SVM Actividad 3 (No Evaluada)
03-12-24	Aprendizaje no supervisado con Python: K-means y reducción de dimensionalidad Actividad 4 (Evaluada) – Control 2
10-12-24	Redes Neuronales I Mini Proyecto 2
17-12-24	Redes Neuronales II Prueba Final

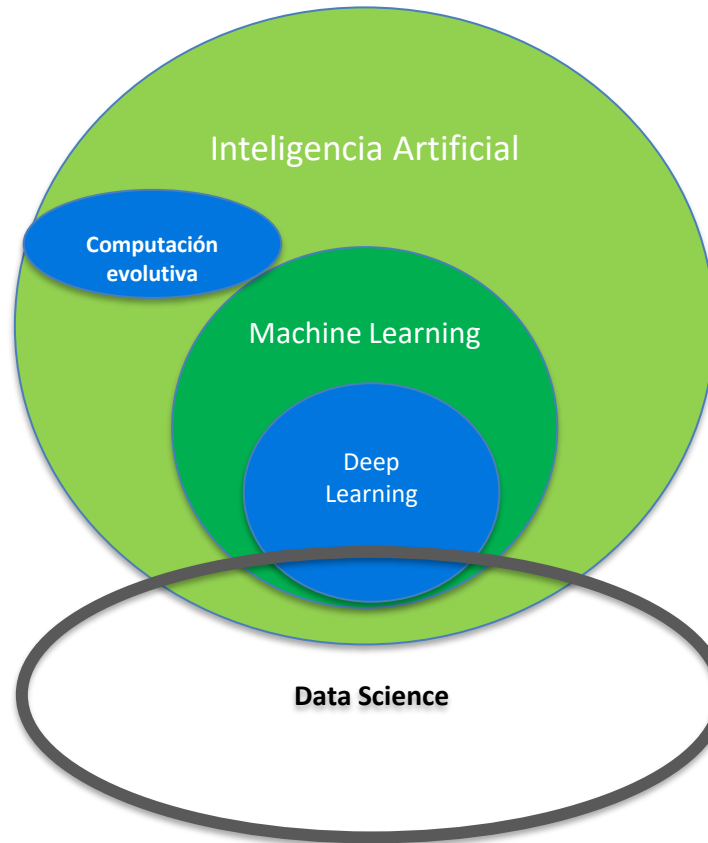


Introducción

- Aprendizaje de Máquina / Inteligencia artificial / Ciencia de Datos
- ¿Por qué estudiar estos temas?
- ¿Dónde se está utilizando y dónde podemos aplicarlo?



IA, ML, DL, DS,...





Aprendizaje de Máquina

Conjunto de métodos que detectan automáticamente patrones complejos en los datos, estos patrones pueden ser utilizados para predecir el futuro.

Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT press.

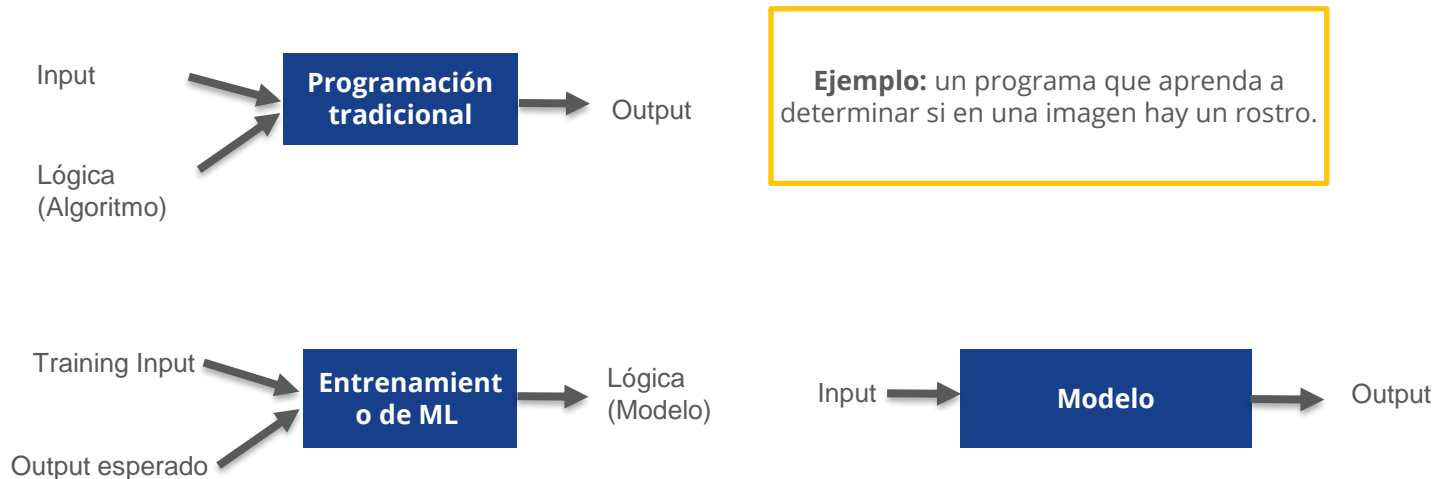


Aprendizaje de Máquina

Un programa de computadora aprende de la experiencia E con respecto a alguna tarea T y alguna medida de rendimiento P , si su rendimiento en T , medido por P , mejora con la experiencia E .

Mitchell, T. (1997). Machine Learning.

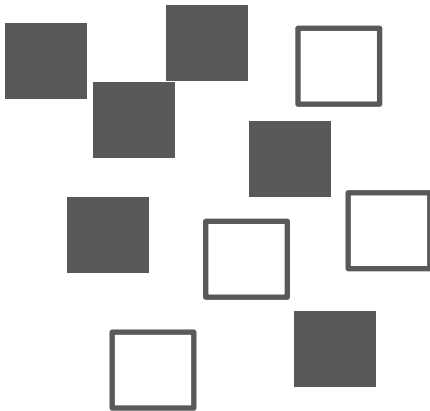
Definición: Campo de estudios que entrega a los computadores la habilidad de aprender sin ser explícitamente programados.



Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development*. **3** (3): 210–229.

Aprendizaje de máquina

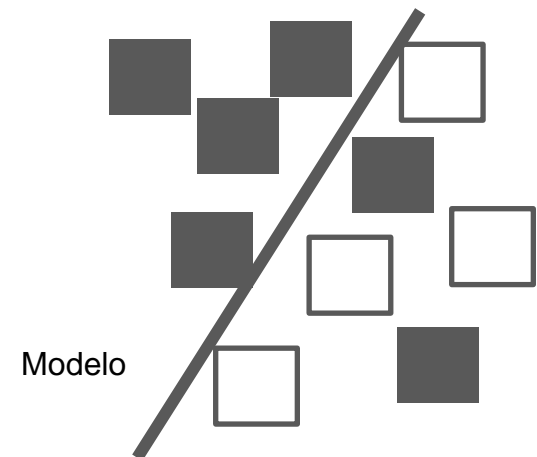
Experiencia (Datos)



Tarea

Separar los cuadrados negros de los blancos.

Rendimiento



Rendimiento = $8/10 = 80\%$

Machine Learning: ejemplos

Experiencia (Datos)

- Datos de donadores y receptores de órganos.
- Datos de compatibilidad.
- Datos de de la operación y post operatorios.
- Seguimiento del paciente.

Tarea

- Determinar el tiempo de vida del órgano trasplantado.

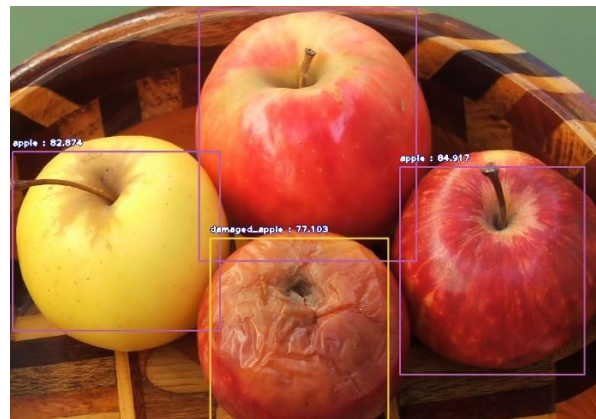
Métrica de rendimiento

- Error en expectativa de vida del organo.



Machine Learning: ejemplos

Experiencia (Datos)	Tarea	Métrica de rendimiento
<ul style="list-style-type: none">Imágenes de manzanas en una línea de proceso con etiqueta de calidad (defectuosa o no defectuosa).	<ul style="list-style-type: none">Identificar manzanas con desperfecto (ej: golpe de sol o polilla).	<ul style="list-style-type: none">Error de selección.



<https://medium.com/deepquestai/ai-in-agriculture-detecting-defects-in-apples-b246799b329c>



Machine Learning: ejemplos

Experiencia (Datos)	Tarea	Métrica de rendimiento
<ul style="list-style-type: none">▪ Mensajes en redes sociales.	<ul style="list-style-type: none">▪ Detectar mensajes con contenido agresivo.	<ul style="list-style-type: none">▪ Mensajes agresivos correctamente clasificados.



Machine Learning: ejemplos

Experiencia (Datos)	Tarea	Métrica de rendimiento
<ul style="list-style-type: none">Imágenes de olivas con mediciones del contenido de aceite de cada una.	<ul style="list-style-type: none">Predecir el contenido de aceite dada una nueva imagen.	<ul style="list-style-type: none">Error de predicción.



Machine Learning: ejemplos

Experiencia (Datos)	Tarea	Métrica de rendimiento
<ul style="list-style-type: none">Registro de rendimiento académico, variable sociodemográficas y variable asociada a deserción académica.	<ul style="list-style-type: none">Identificar tempranamente alumnos con alta probabilidad de desertar.	<ul style="list-style-type: none">Porcentaje de alumnos desertores no identificados.





Machine Learning: ejemplos

Experiencia (Datos)	Tarea	Métrica de rendimiento
<ul style="list-style-type: none">▪ Histórico de transacciones y pagos de clientes.▪ Variables climáticas y número de esporas asociados a un hongo.▪ Registro de la intensidad de luz en el tiempo de estrellas y su clase.	<ul style="list-style-type: none">▪ Identificar clientes con alta probabilidad de fuga.▪ Alertar respecto a la presencia del hongo.▪ Asignar clase a nuevas estrellas dado su patrón de intensidad de luz en el tiempo.	<ul style="list-style-type: none">▪ Proporción de clientes fugados no identificados por el modelo.▪ Error de predicción.▪ Error de clasificación.



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

¿Por qué estudiar IA, DS o ML?



Situación actual

Disponibilidad de
datos

Oferta de modelos
y librerías

Tareas con
buen rendimiento



Disponibilidad de datos

Recolección de datos en diversos contextos



Smart cities



Ciencia y tecnología



Agricultura

<https://www.bbva.com/es/cielos-chilenos-atraen-la-observacion-astronomica-mundial/>

Herath, H. M. K. K. M. B., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, 2(1), 100076.

Bhat, S. A., & Huang, N. F. (2021). Big data and ai revolution in precision agriculture: Survey and challenges. *IEEE Access*, 9, 110209-110222.



Disponibilidad de datos

Recolección de datos en diversos contextos

¿En qué otros campos han visto potencial del machine learning y la inteligencia artificial?



Situación actual

Disponibilidad de
datos

Oferta de modelos
y librerías

Tareas con
buen rendimiento



Acceso a modelos y librerías



vs





Acceso a modelos y librerías

¿Por qué Python?



Simple y consistente

Oferta de librerías

Multiplataforma y multiparadigma

Gran comunidad


Flexibilidad para múltiples tipos de proyectos

Acceso a modelos y librerías



Acceso a modelos y librerías



 [Install](#) [User Guide](#) [API](#) [Examples](#) [Community](#) [More](#)

scikit-learn

Machine Learning in Python

[Getting Started](#) [Release Highlights for 1.0](#) [GitHub](#)

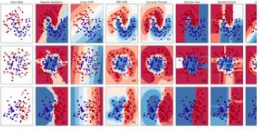
- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying which categoría an object belongs to.

Applications: Spam detection, image recognition.

Algorithms: SVM, nearest neighbors, random forest, and more...



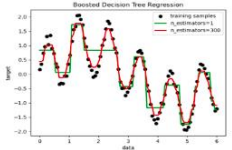
Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.

Algorithms: SVR, nearest neighbors, random forest, and more...




Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: k-Means, spectral clustering, mean-shift, and more...



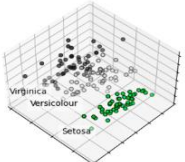
Examples

Dimensionality reducción

Reducing the number of random variables considerar.

Applications: Visualization, Increased efficiency

Algorithms: k-Means, feature selection, non-negative matrix factorization, and more...



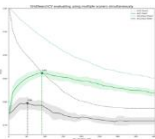
Examples

Model selection

Comparing, validating and choosing parameters and models.

Applications: Improved accuracy via parameter tuning

Algorithms: grid search, cross validation, metrics, and more...



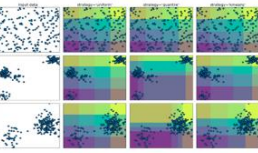
Examples

Preprocessing

Feature extraction and normalization.

Applications: Transforming input data such as text for use with máquina learning algorithms.

Algorithms: preprocessing, feature extraction, and more...



Examples



Situación actual

Disponibilidad de
datos

Oferta de modelos
y librerías

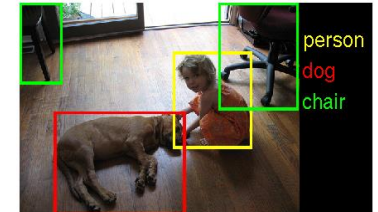
Tareas con
buen rendimiento

Situación actual

Tareas con
buen rendimiento

Detección de imágenes

imagen: <http://www.image-net.org/challenges/LSVRC/2014/>



Traducción de textos

imagen: https://es.wikipedia.org/wiki/Traductor_de_Google



Reconocimiento de voz

imagen: <https://cheekymunkey.co.uk/voice-activated-computers-future/>



www.educacionprofesional.ing.uc.cl

Chat-GPT (LLM, large language models)



Análisis de texto

Búsqueda de información de negocio

Clasificación

Generador de contenido

Asistente de programación

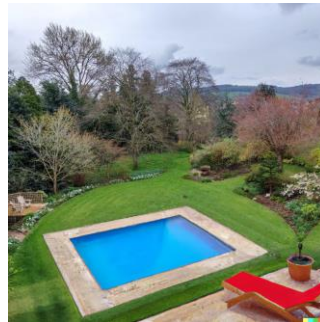
Automatización de venta

<https://www.leewayhertz.com/chatgpt-enterprise-usecases-and-solutions/>

¿Ustedes han usado chat gpt o algún LLM similar?



Dall-e



<https://www.seotraininglondon.org/business-applications-dalle2/>



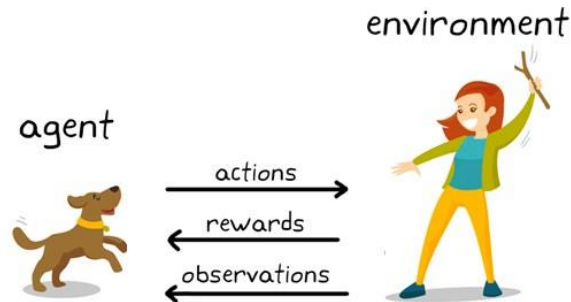
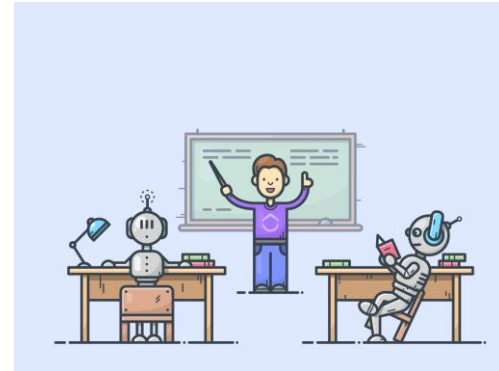
ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

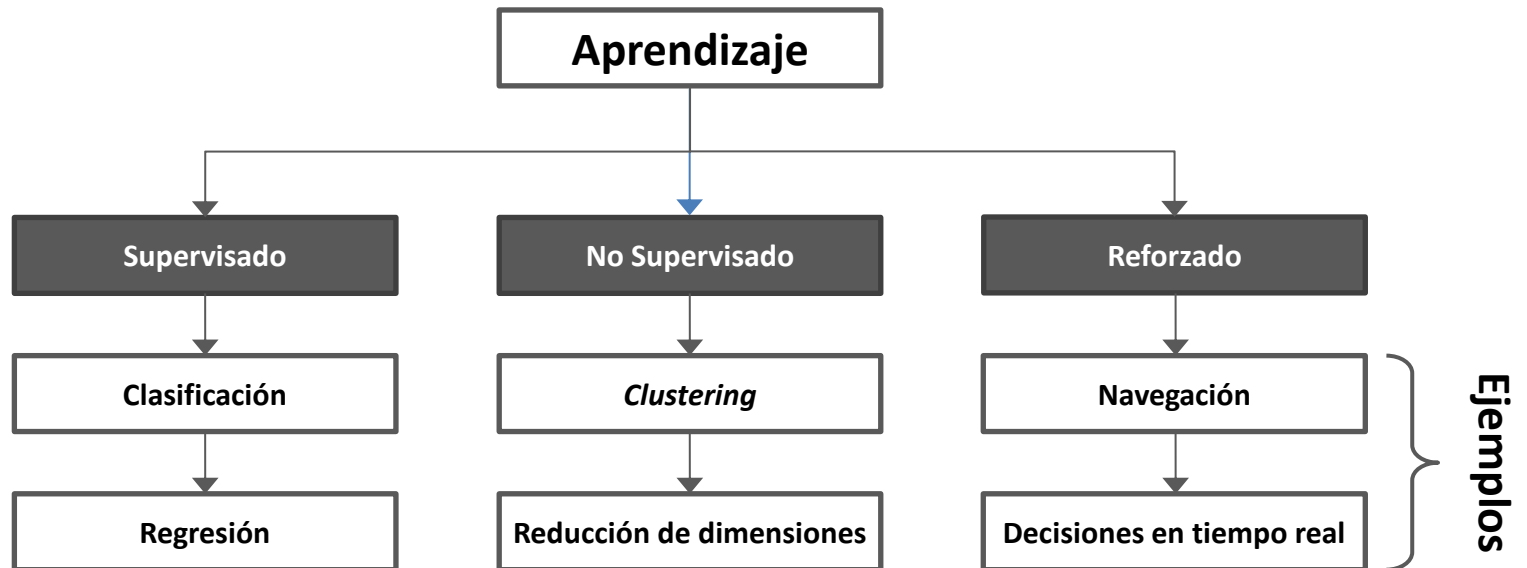
EDUCACIÓN
PROFESIONAL

¿Cómo aprendemos los humanos?

**¿Cómo Podemos replicar ese proceso de
aprendizaje a las máquinas?**

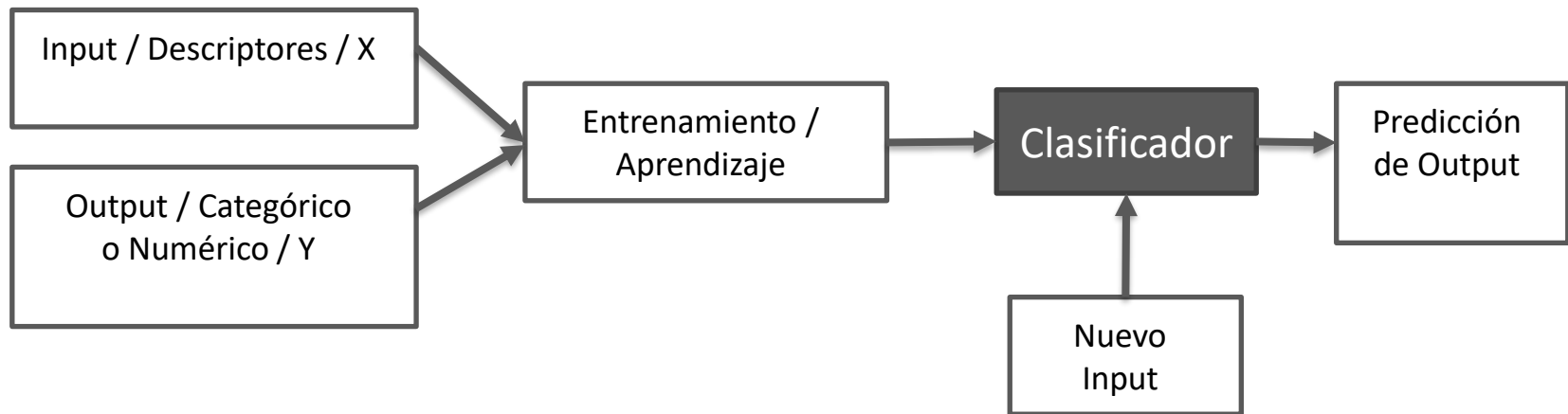
¿Cómo aprendemos?



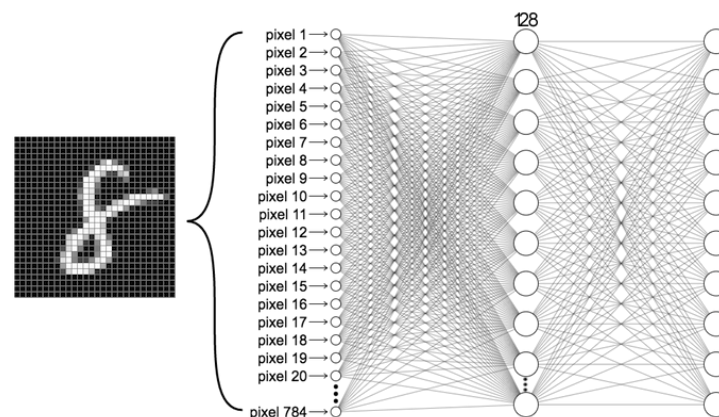
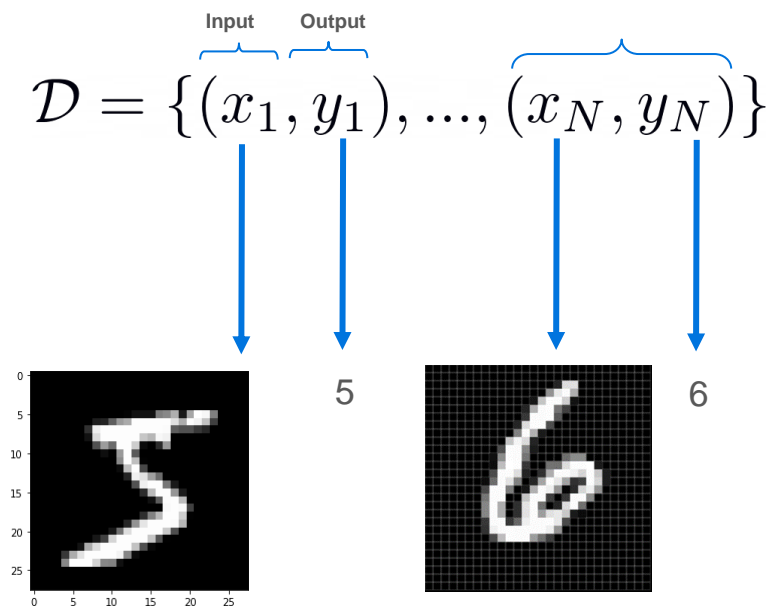


Aprendizaje supervisado

Aprender una función (Clasificador/Regresor) que permita obtener los valores del **output** por medio del **input**.



Aprendizaje supervisado





Ejemplos de aprendizaje supervisado

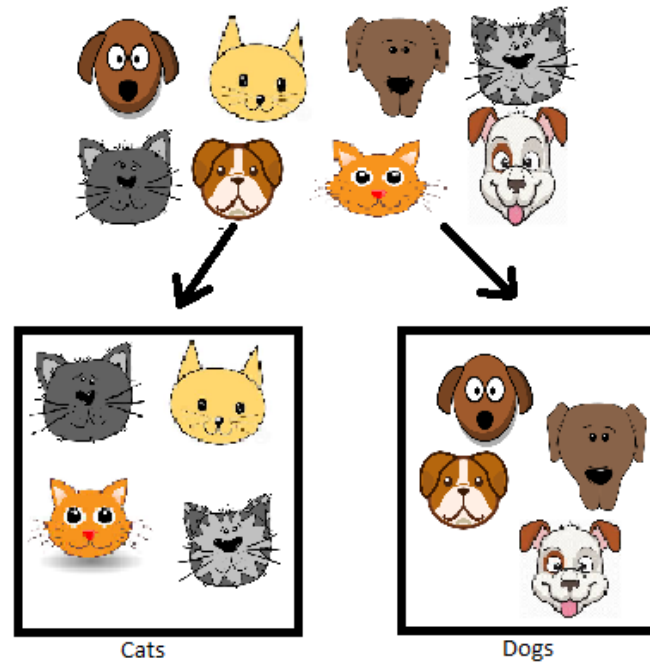
**Reconocimiento
de caras**

**Pronosticar nivel de
contaminación**

**Predecir comportamiento
de pago**

**Clasificar estrellas
en el espacio**

Aprendizaje no supervisado



Aprendizaje no supervisado

Datos



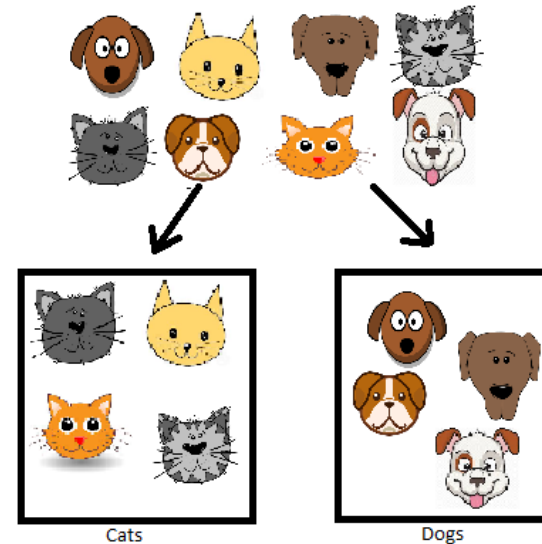
No tienen un output definido

Aprendizaje no supervisado



Entender patrones ocultos en ellos.

$$\mathcal{D} = \{x_1, x_2, \dots, x_{N-1}, x_N, \}$$



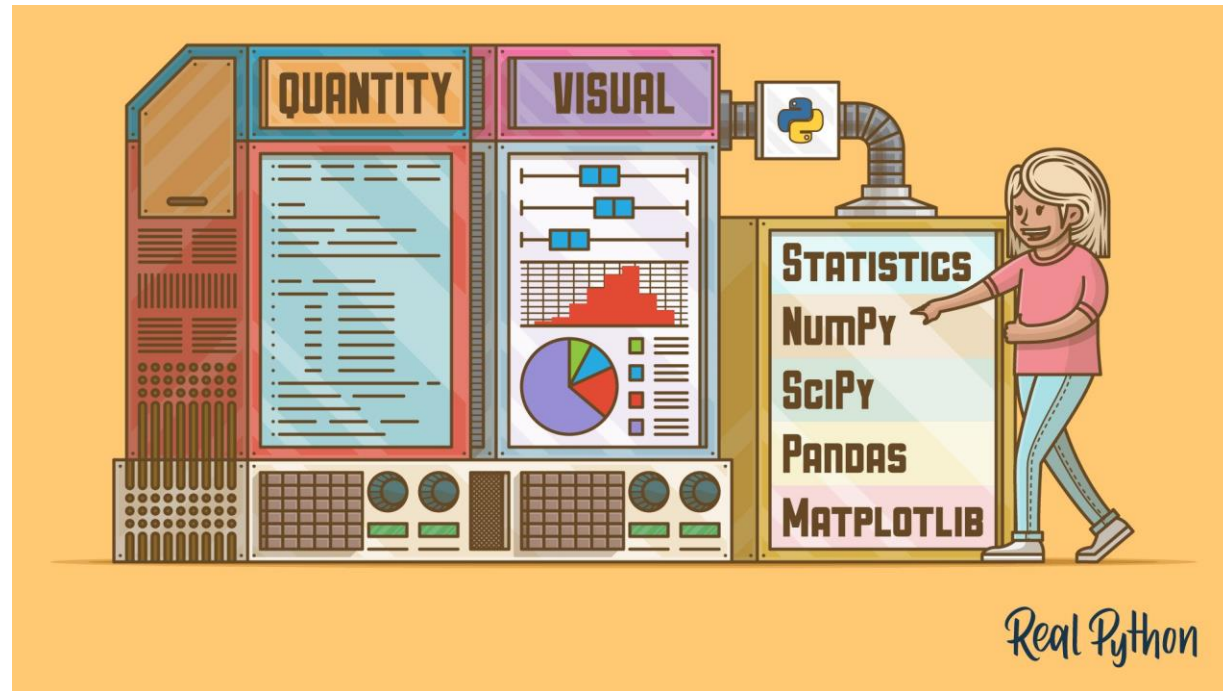
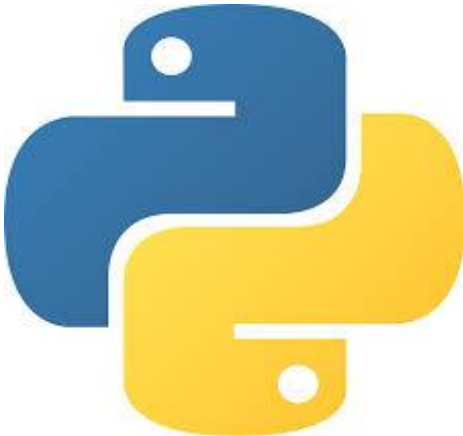


ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

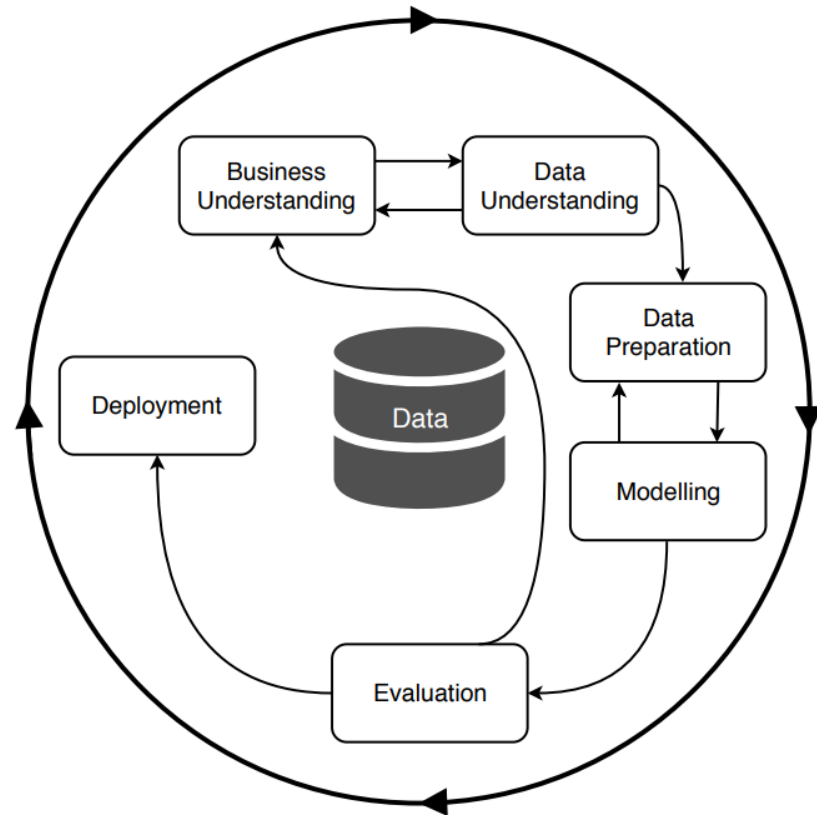
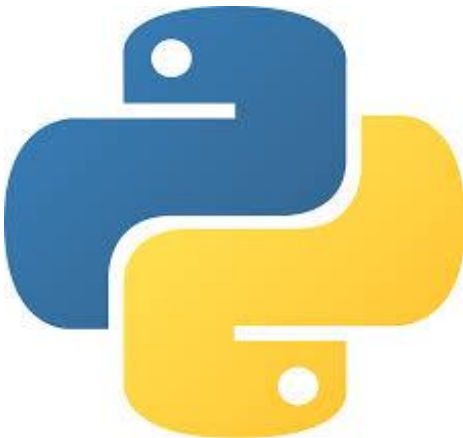
EDUCACIÓN
PROFESIONAL

Python para Ciencia de datos

Ecosistema de Ciencia de datos con Python



<https://realpython.com/python-statistics/>



Cross-Industry Standard Process for Data Mining (CRISP-DM)

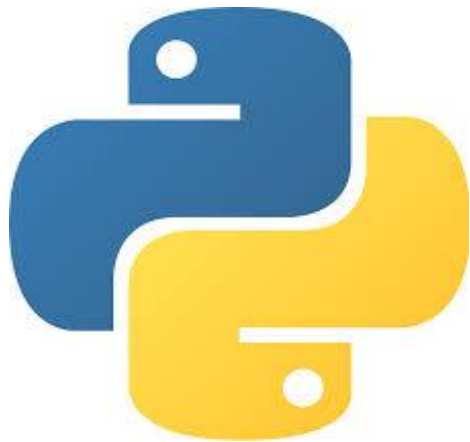
Martínez-Plumed, Fernando, et al. "CRISP-DM twenty years later: From data mining processes to data science trajectories." *IEEE transactions on knowledge and data engineering* 33.8 (2019): 3048-3061.

Acceso a modelos y librerías





¿Por qué Python?



Simple y consistente

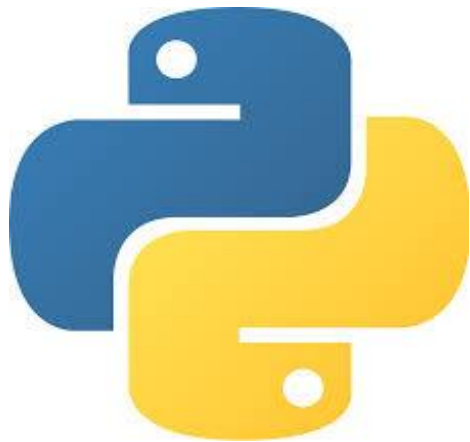
Oferta de librerías

Multi-plataforma

Gran comunidad y popularidad



¿Para qué se está usando Python?



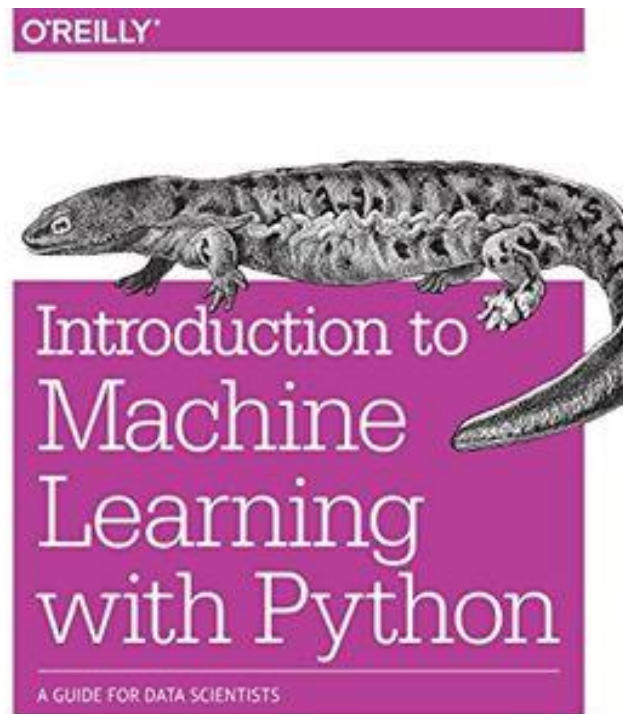
Desarrollo web

Big Data

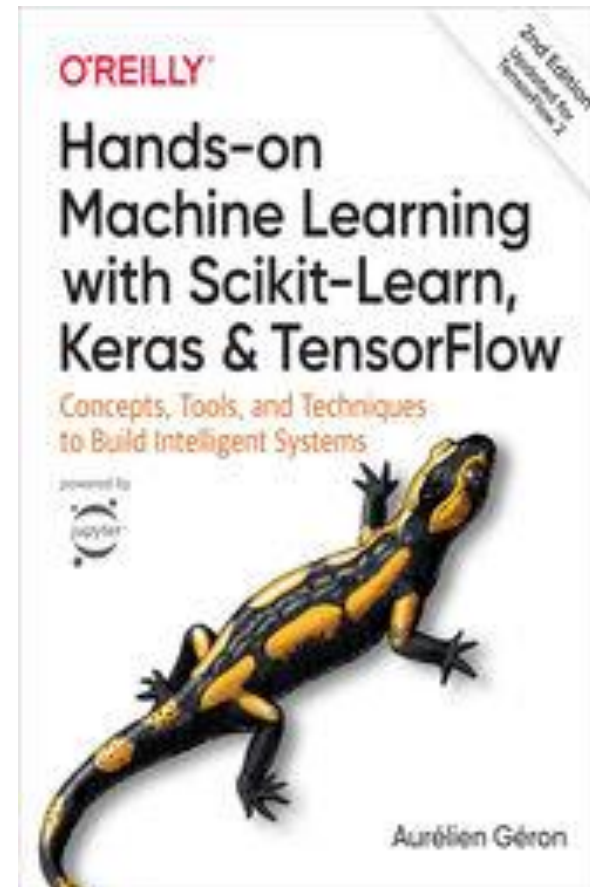
IoT (Raspberry pi – Arduino - Jetson Nano)

Ciencia de Datos / Aprendizaje de Máquina
/ Análisis geoespacial

Python para Machine Learning



Andreas C. Müller & Sarah Guido



Aurélien Géron



Rápida y eficiente alternativa para manipular y acceder a datos.

Simple manejo de datos perdidos

Amplia oferta de métodos de lectura y escritura de archivos.

Múltiples opciones para manejo de BBDD como por ejemplo unión y separación de datos.

<https://data-flair.training/blogs/python-pandas-features/>



Eficiente manejo de N-dimensional arrays

Métodos de álgebra lineal, estadística y optimización

Posibilidad de integración con lenguajes de bajo nivel.



Construida sobre Numpy

Integración

Optimización

Scientific Python (SciPy)



Gran variedad de gráficos y amplia
comunidad.

Construida sobre Numpy y Scipy

Interfaz familiar (similar a Matlab)

<https://jakevdp.github.io/blog/2013/03/23/matplotlib-and-the-future-of-visualization-in-python/>

<https://www.machinelearningplus.com/plots/matplotlib-tutorial-complete-guide-python-plot-examples/>



Construida sobre Numpy, Scipy, Pandas,
Matplotlib, etc.

Más de 30 colaboradores activos.

Pensada para puesta en producción de
modelos.

Gran variedad de modelos y funciones
utilizadas en las distintas etapas de un
proyecto de Aprendizaje de máquinas.



Mantenidas por grandes de la tecnología.

Código y modelos totalmente abierto.

Uso de GPUs optimizado

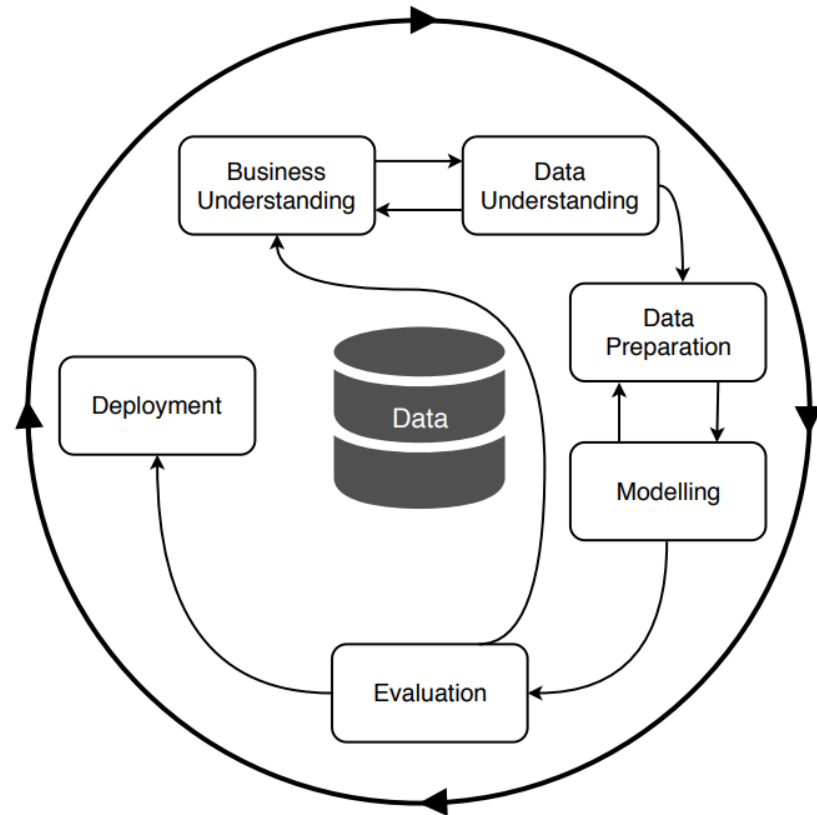
Construcción de complejos modelos
amigablemente.



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

Análisis exploratorio de datos con Python



Cross-Industry Standard Process for Data Mining (CRISP-DM)

Martínez-Plumed, Fernando, et al. "CRISP-DM twenty years later: From data mining processes to data science trajectories." *IEEE transactions on knowledge and data engineering* 33.8 (2019): 3048-3061.



Análisis exploratorio de datos

a

Análisis descriptivo

b

Análisis del tipo de dato

c

Datos perdidos y datos atípicos

d

Visualización de datos

e

Análisis de correlaciones



Bibliografía

- McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc."
- Muller, A. C., & Guido, S. (2017). Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media.
- Mitchell, T. (1997). Machine Learning. MacGraw-Hill Companies. Inc., Boston.



ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA

EDUCACIÓN
PROFESIONAL

Diplomado en Programación y Aplicaciones de Python

Aplicaciones en Ciencia de Datos e
Inteligencia Artificial

Profesor: Francisco Pérez Galarce

