

Introducción a la Ciencia de Datos e Inteligencia Artificial

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingeniería Electrónica

Universidad Popular del Cesar

Contenido

- 1 [Datos e Información](#)
- 2 [Aclaremos conceptos](#)
- 3 [Por qué Python y Nube?](#)
- 4 [Tipos de aprendizaje](#)
- 5 [La clave del éxito](#)
- 6 [Conclusiones](#)

- 1 [Datos e Información](#)
- 2 [Aclaremos conceptos](#)
- 3 [Por qué Python y Nube?](#)
- 4 [Tipos de aprendizaje](#)
- 5 [La clave del éxito](#)
- 6 [Conclusiones](#)



Contenido

- 1 [Datos e Información](#)
- 2 [Aclaremos conceptos](#)
- 3 [Por qué Python y Nube?](#)
- 4 [Tipos de aprendizaje](#)
- 5 [La clave del éxito](#)
- 6 [Conclusiones](#)

- Los datos se pueden encontrar “fácilmente” en todos lados
 - Evolución del precio de las acciones de una empresa en bolsa
 - Estadísticas de resultados deportivos
 - Históricos de consumo de ciertos productos
 - Precios de mercado de bienes y/o servicios
 - ...

- Información: cómo y dónde buscarla?:
 - Normalmente subyace escondida detrás de los datos
 - Requiere del procesamiento y análisis de datos



- *Data*: Tener las cifras en crudo de un determinado fenómeno
- *Information*: Poder extraer de esas cifras relaciones, dependencias, influencias, causas y posibles consecuencias
- *Knowledge*: Saber cómo hacer frente a la información
- *Wisdom*: Tener el poder para hacerlo

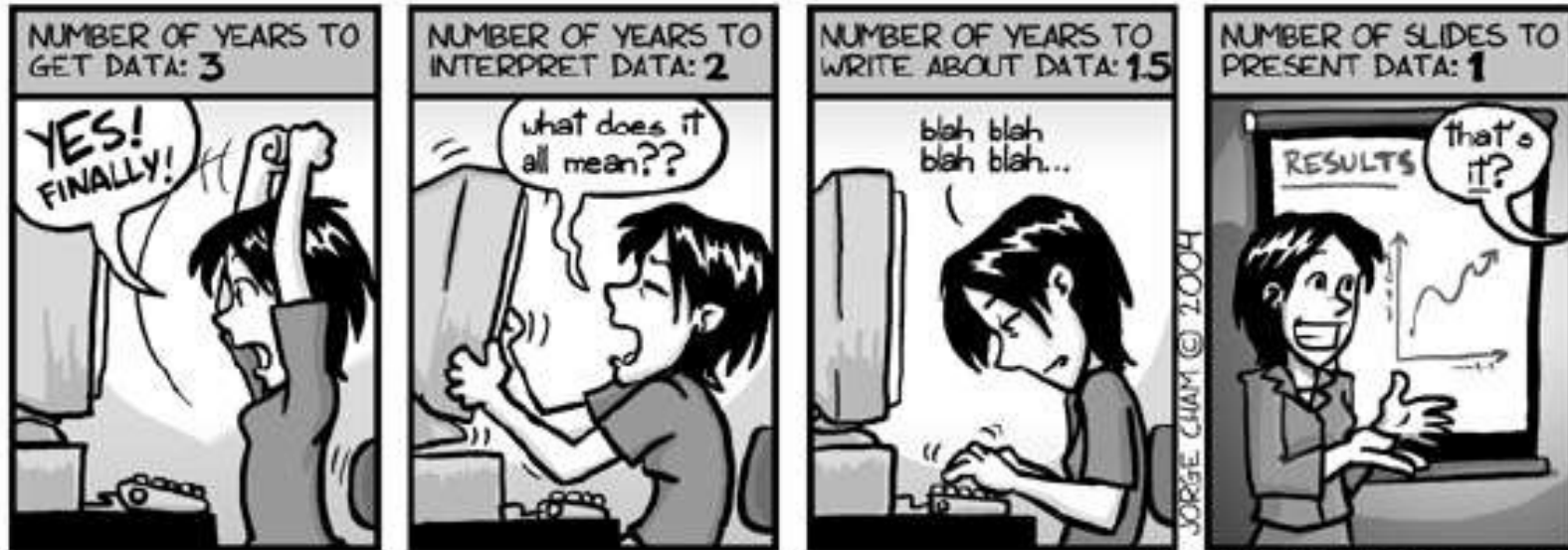
El caso (o mito) de la cerveza y los pañales

En una cadena de almacenes (Wal-Mart o Costco) analizaron los datos de compras de sus clientes

- *Data:* Registros de los artículos comprados, junto con la hora, el género del comprador y la edad.
- *Information:* Se descubrió una alta correlación entre: **compradores hombres**, **compras entre 5pm y 7pm**, **pañales** y **cervezas**
- *Knowledge:* Saber que los padres, después de salir del trabajo, suelen comprar pañales y también cervezas.
- *Wisdom:* Implementar nuevas estrategias de mercadeo.

Básicamente...²

DATA: BY THE NUMBERS



www.phdcomics.com

²<http://phdcomics.com/comics.php>

Data science

From Wikipedia, the free encyclopedia

Not to be confused with information science.

Data science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured,^{[1][2]} which is a continuation of some of the data analysis fields such as statistics, data mining, and predictive analytics,^[3] similar to Knowledge Discovery in Databases (KDD).

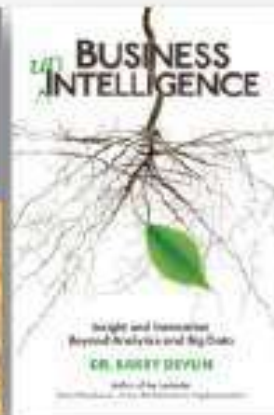
Overview [edit]

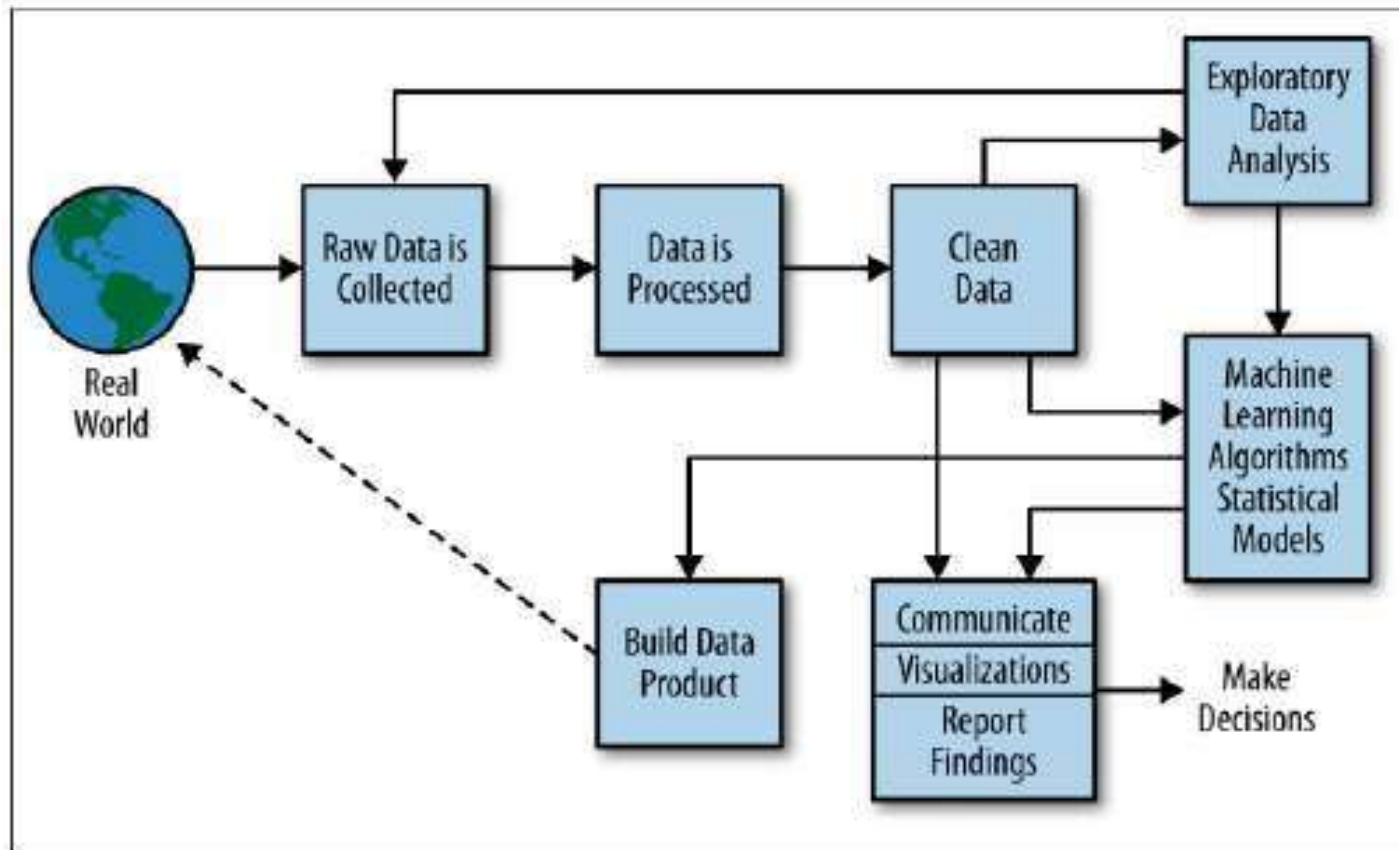
Data science employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, operations research,^[4] information science, and computer science, including signal processing, probability models, machine learning, statistical learning, data mining, database, data engineering, pattern recognition and learning, visualization, predictive analytics, uncertainty modeling, data warehousing, data compression, computer programming, artificial intelligence, and high performance computing. Methods that scale to big data are of particular interest in data science, although the

¿La Ciencia de los Datos es “eso” que hacen Google y Facebook?
Antes de profundizar en *¿qué es la **Ciencia de los Datos**?*,
entendamos primero un poco los conceptos que la acompañan

La Ciencia de los Datos está relacionada con áreas tan diversas (y a la vez tan afines) como son:

- *Big data*
- *Machine learning*
- *Data mining*
- *Business intelligence*
- *Data analytics*
- ...





El boom de la Ciencia de Datos

- En los últimos años ha habido un *boom* relacionado con el *big data* y la **Ciencia de los Datos**
- Las fuentes de datos se han multiplicado y diversificado (Internet, dispositivos móviles, sensores, transacciones comerciales, etc.)
- Se han reducido los costos en la obtención de los datos
- Estamos experimentando un cambio de paradigma en la forma como se analizan los datos y se extrae información de ellos
- La **Ciencia de los Datos** es un área aún por explorar y con grandísimas capacidades de expansión y desarrollo

El boom de la Ciencia de Datos

De acuerdo al Harvard Business Review³



³<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

Perfil de la científica de datos



Perfil del científico de datos



En resumen...

Científico de datos: "Persona que sabe más de **estadística** que cualquier programador y que a la vez sabe más de **programación** que cualquier estadístico". Necesitamos:

- Álgebra lineal
- Teoría de probabilidades
- Optimización
- Programación (Matlab, R, **Python, Cloud computing**)
- En conclusión necesitamos del aprendizaje estadístico (aprendizaje de máquina/automático - Machine Learning)
- ¿Necesitamos ser expertos en programación?

Contenido

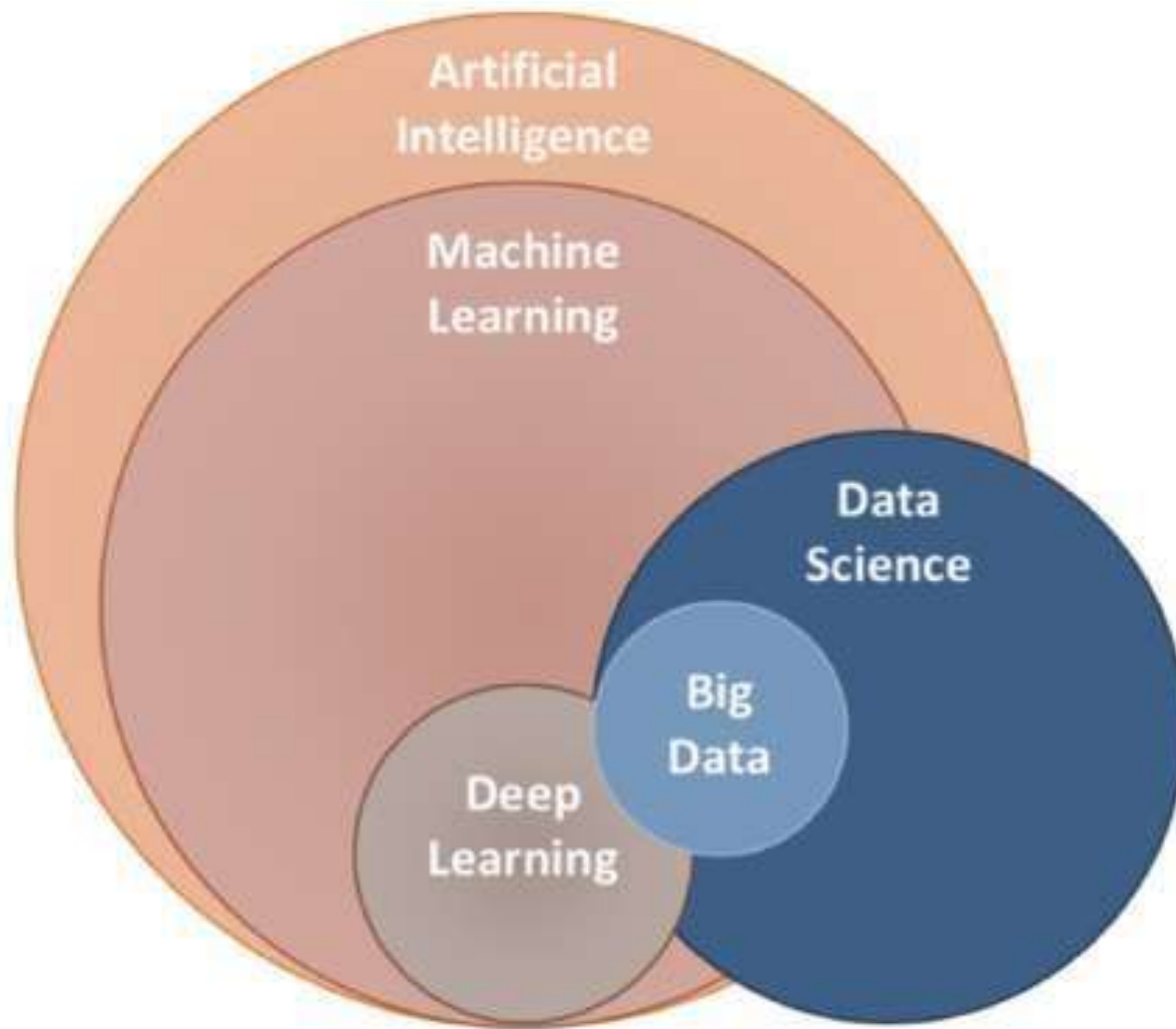
- 1 [Datos e Información](#)
- 2 [Aclaremos conceptos](#)
- 3 [Por qué Python y Nube?](#)
- 4 [Tipos de aprendizaje](#)
- 5 [La clave del éxito](#)
- 6 [Conclusiones](#)

IA vs CD vs ML

Aspecto	Inteligencia Artificial (IA)	Ciencia de Datos (CD)	Machine Learning (ML)
Definición	Busca simular inteligencia humana en las máquinas	Análisis y procesamiento de datos para obtener conocimiento	Subcampo de la IA que permite a las máquinas aprender de datos
Enfoque	Crear sistemas que tomen decisiones inteligentes	Extracción de información a partir de datos	Crear modelos que mejoren su rendimiento con el tiempo
Herramientas	Redes neuronales, sistemas expertos, algoritmos genéticos	Python, R, SQL, pandas, Matplotlib	Algoritmos de clasificación, regresión, redes neuronales
Relación	Utiliza machine learning para el aprendizaje de datos y toma de decisiones	Puede incluir IA y ML como herramientas	Parte de la IA, clave para analizar datos en CD

Necesitamos entender algo de estadística y programación!

IA vs CD vs ML



Aprendizaje de máquina es la clave

- En una frase: *aprendizaje de máquina* es el conjunto de los **algoritmos** y las **técnicas** que se usan para diseñar sistemas que aprendan a partir de datos.
- Los fundamentos del *aprendizaje de máquina* se basan en las **matemáticas** y la **estadística**.
- De forma general, no tienen en cuenta el conocimiento del dominio y el pre-procesamiento de los datos.
- El aprendizaje de máquina es el eje central de la ciencia de datos y la inteligencia artificial.

El renacer de la inteligencia artificial (Premio Turing 2019)

'Godfathers of AI' honored with Turing Award, the Nobel Prize of computing

Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI

By James Vincent | Mar 27, 2019, 6:02am EDT

f t  SHARE

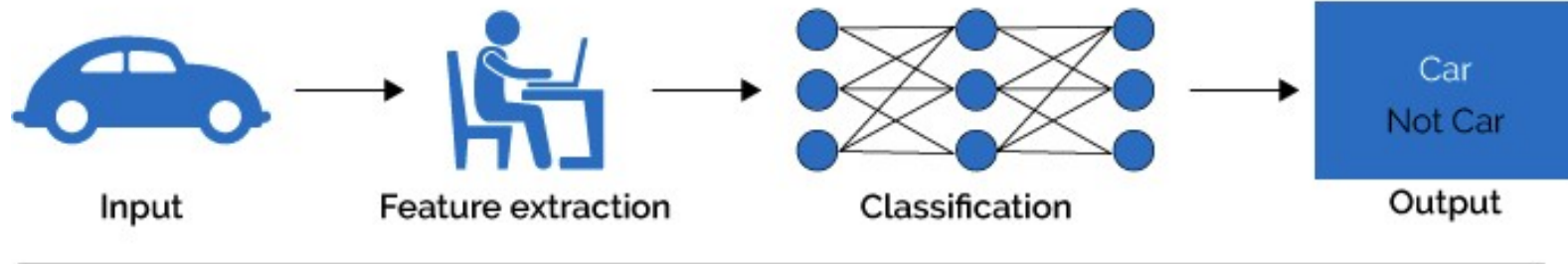


En 2006, Geoffrey Hinton et al. publicaron un artículo ⁴ que mostraba como el aprendizaje profundo podría reconocer dígitos a mano con una precisión > 98%, llamándolo Deep Learning.

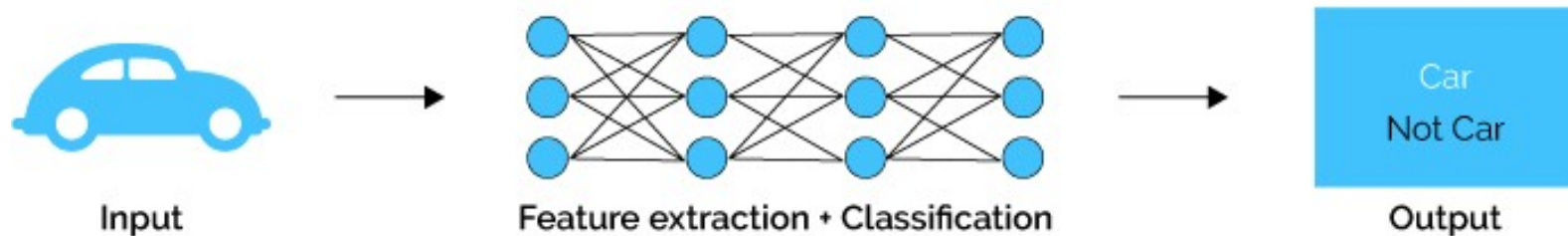
⁴ver <http://www.cs.toronto.edu/~hinton/>

Machine Learning vs Deep Learning

Machine Learning



Deep Learning



- [Ejemplo IA avanzada: piloto automático Tesla](#)

El renacer de la inteligencia artificial

- Entrenar un modelo de deep learning era considerado imposible en los 90s.
- Hinton y los demás investigadores en redes neuronales empezaron a destronar a los algoritmos clásicos de ML.
- En la actualidad: IA como corazón de muchos productos de tecnología de punta (búsqueda web, teléfonos inteligentes, reconocimiento de habla, autos que se conducen solos, chatbots inteligentes, etc...)
- La clave: mucho poder de cómputo y muchos datos!

Qué es aprendizaje de máquina? (Competencias básicas)

Básicamente...programar computadores para **aprender desde datos!**

Después de entender la importancia de la ciencia de los datos y su conexión con el aprendizaje de máquina, se busca entonces:

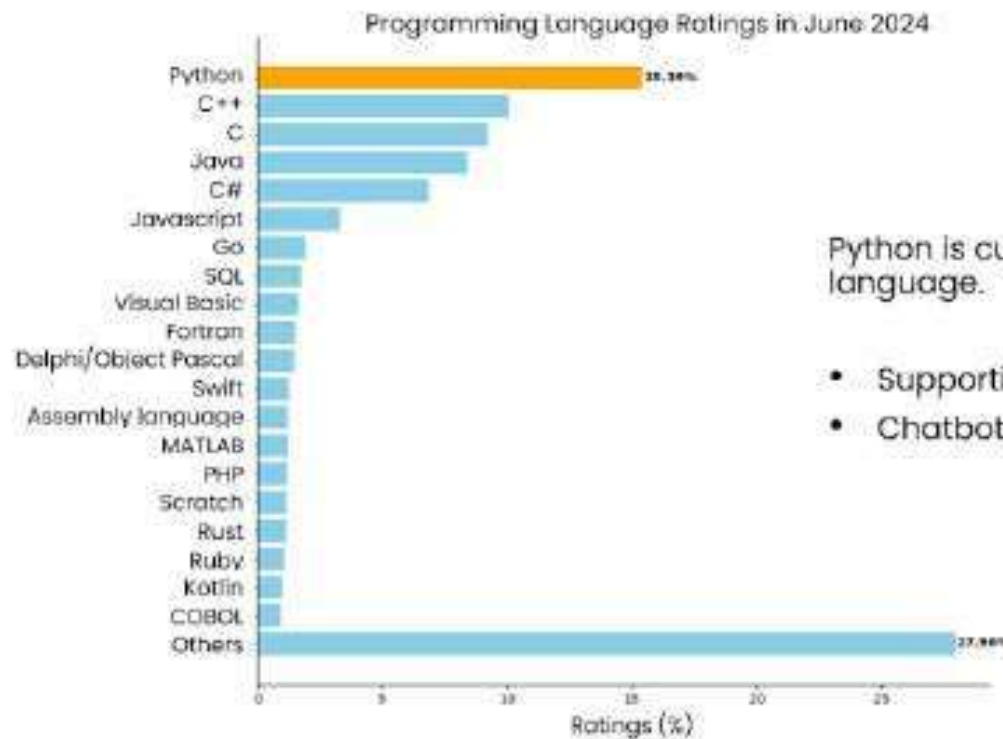
- Entender los modelos básicos de aprendizaje de máquina.
- Comprender modelos más avanzados (Deep learning).
- Fortalecer las competencias en estadística y programación.
- Utilizar herramientas libres y reconocidas en Python (Pandas, SciKitlearn, TensorFlow, Keras, PyTorch).

Contenido

- 1 [Datos e Información](#)
- 2 [Aclaremos conceptos](#)
- 3 [Por qué Python y Nube?](#)
- 4 [Tipos de aprendizaje](#)
- 5 [La clave del éxito](#)
- 6 [Conclusiones](#)

Por qué Python?

Why Python?



Python is currently **the most popular** programming language.

- Supportive community
- Chatbots like ChatGPT are reliable

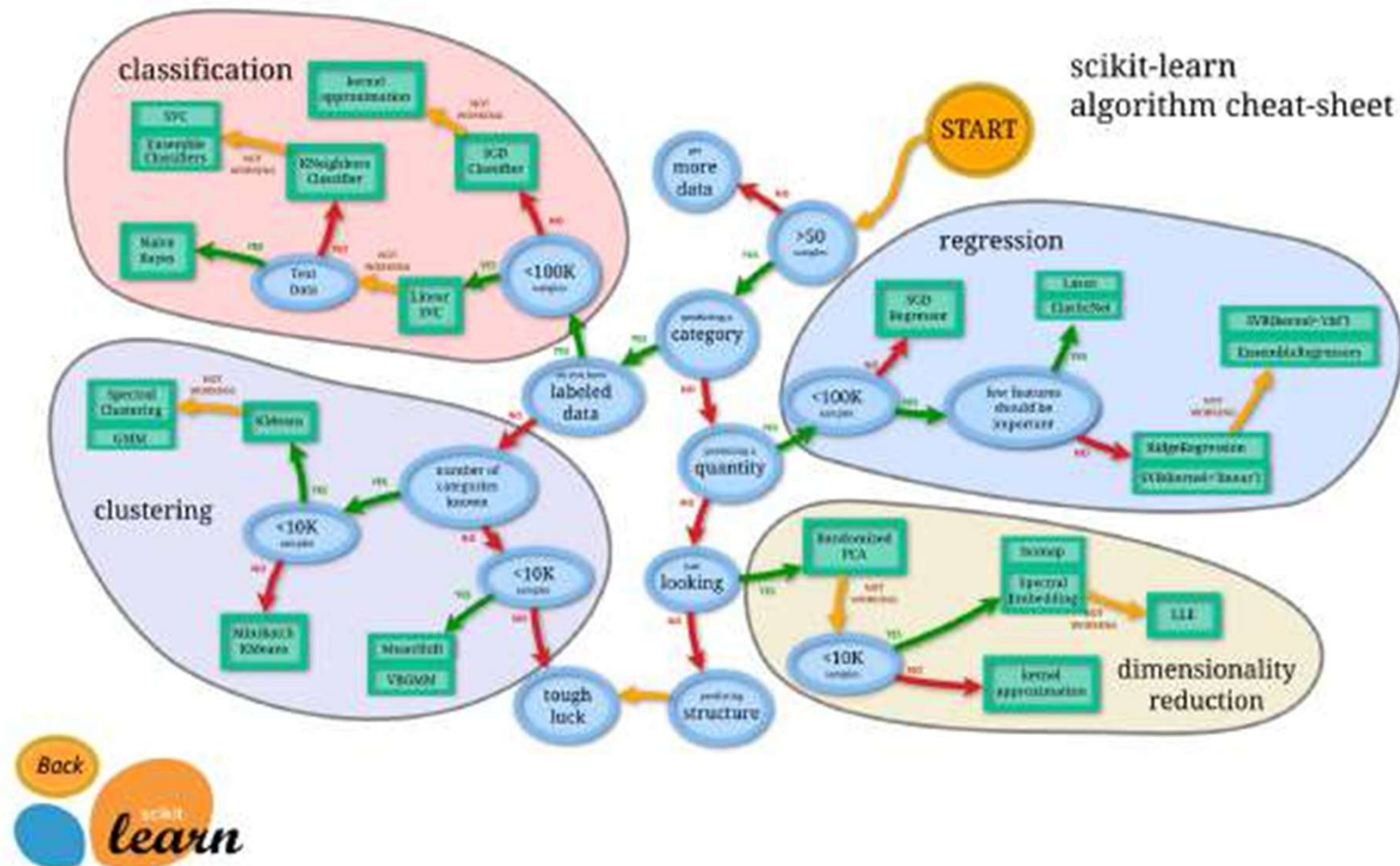
Nuestras librerías amigas Python - Pandas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$


	BandName	WavelengthMax	WavelengthMin
0	CoastalAerosol	450	430
1	Blue	510	450
2	Green	590	530
3	Red	670	640
4	NearInfrared	890	850
5	ShortWaveInfrared_1	1650	1570
6	ShortWaveInfrared_2	2290	2110
7	Cirrus	1380	1300

Nuestras librerías amigas Python - Scikit-learn



Nuestras librerías amigas

Python - TensorFlow, Keras, PyTorch



Nube con alto desempeño Gratis!
No quemes más tu PC!



Otras alternativas: Microsoft Azure, IBM Cloud,
Amazon SageMaker...

Fernando Pérez - IPython-Jupyter

Fernando Pérez ([Medellín, Colombia](#)) es un [físico](#), desarrollador de software y promotor del software libre. Es conocido como el creador de IPython.^{1 2 3 4 5 6}

En el año 2012 recibió el premio [por el Avance del Software Libre](#) de la [Free Software Foundation](#).^{7 8 9}

Es un [miembro investigador](#) de la [Python Software Foundation](#),¹⁰ y un [miembro fundador](#) de la organización [NumFOCUS](#).^{11 12}

Vida y carrera [\[editar\]](#)

Fernando Pérez nació en [Medellín, Colombia](#). Realizó su pregrado en [física](#) en la [Universidad de Antioquia](#) y su maestría también en física en la

Fernando Pérez



Información personal

Nacimiento 1972 ✓
[Medellín \(Colombia\)](#) ✓

Nacionalidad Colombiana

Educación

Educado en [Universidad de Antioquia](#)
[Universidad de Colorado en Boulder](#) ✓



**Universidad
Popular del Cesar**

Contenido

- 1 [Datos e Información](#)
- 2 [Aclaremos conceptos](#)
- 3 [Por qué Python y Nube?](#)
- 4 [Tipos de aprendizaje](#)
- 5 [La clave del éxito](#)
- 6 [Conclusiones](#)

Aprendiendo por reglas impuestas (rule by hand-handcraft)

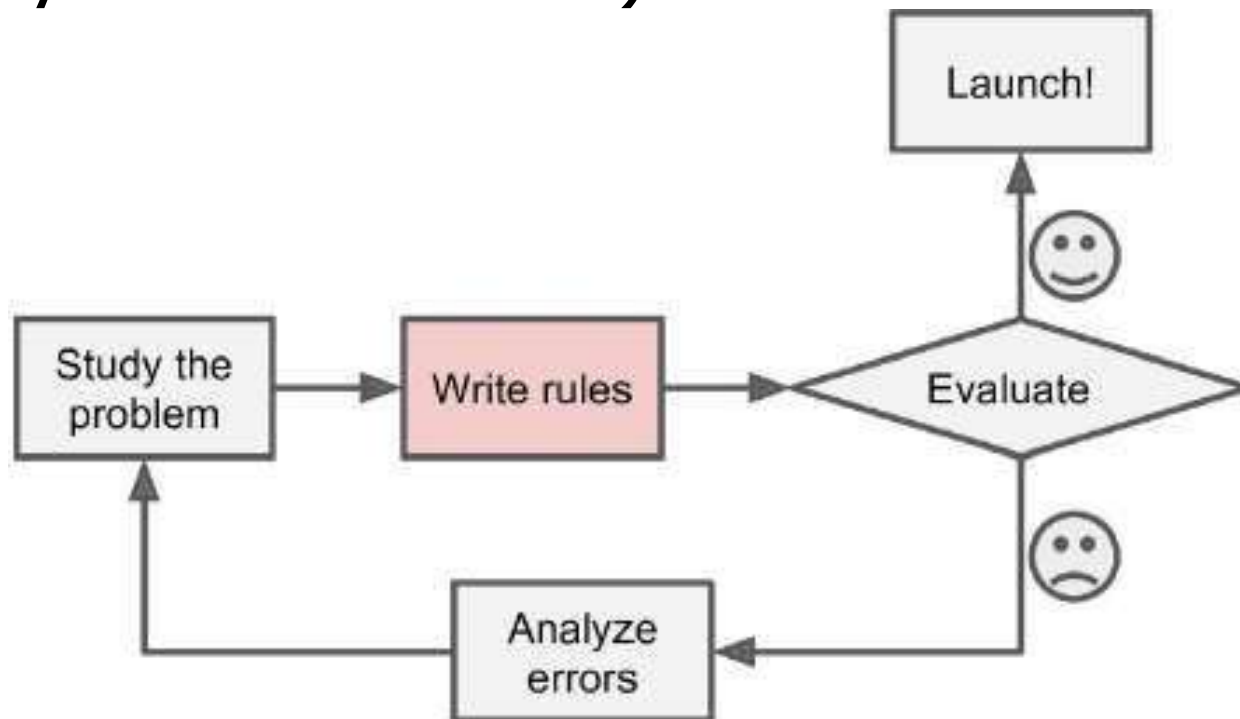


Figure: Aprendizaje por reglas impuestas. fuente: Hands on machine learning book.

Larga lista de reglas, difíciles de mantener y definir. Ejemplo: análisis clásicos desde modelos.

Aprendizaje estadístico (Aprendizaje de máquina)

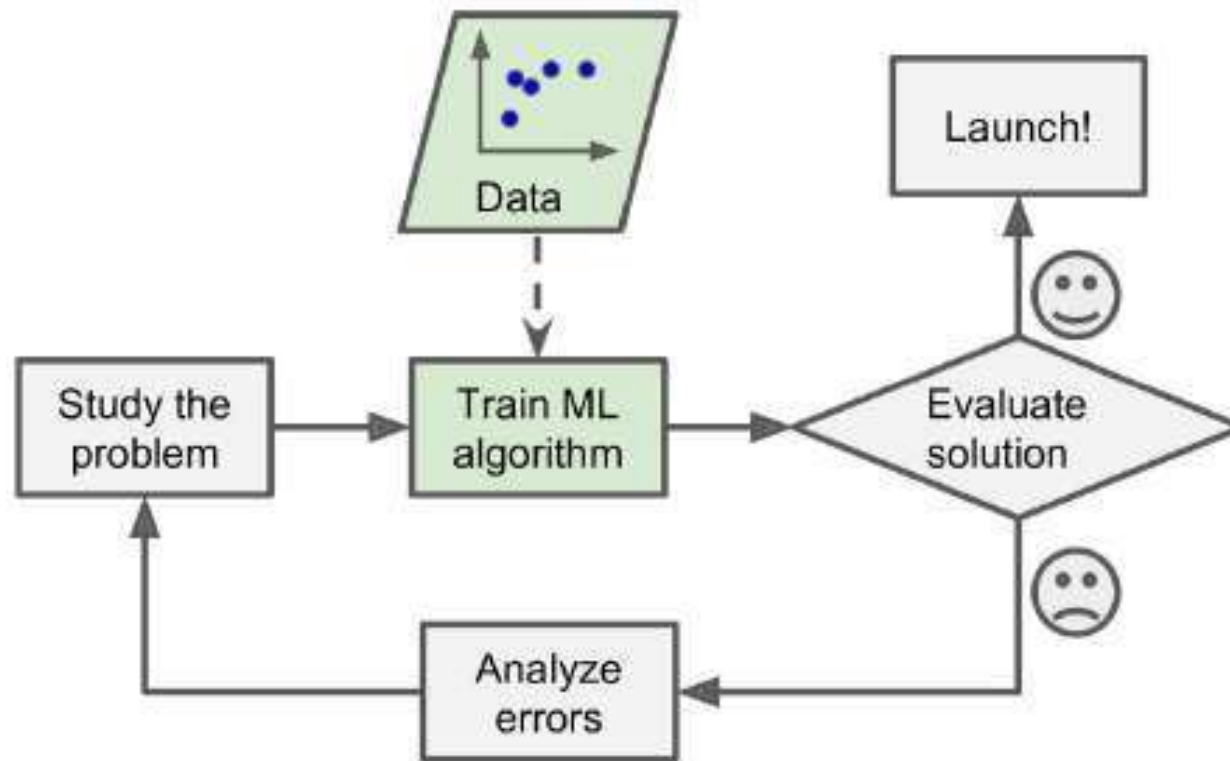


Figure: Aprendizaje de máquina. fuente: Hands on machine learning book.

Aprendiendo desde los datos!

Con supervisión humana: clasificación

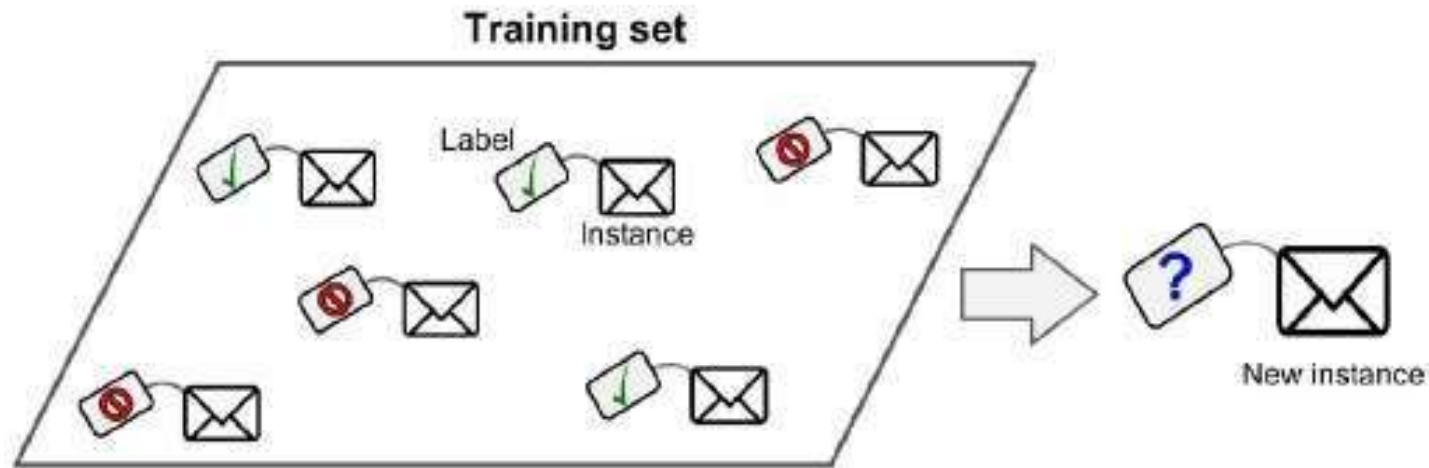


Figure: Aprendizaje supervisado en clasificación. fuente: Hands on machine learning.

- Instancia u observación: muestra del fenómeno en estudio.
- Atributo: propiedad que codifica la instancia.
- Característica: atributo con valor (cardinal o nominal).
- Etiqueta (nominal): membresia de grupo
- **Ejemplo: reconocimiento correo spam vs no spam.**

Con supervisión humana: regresión

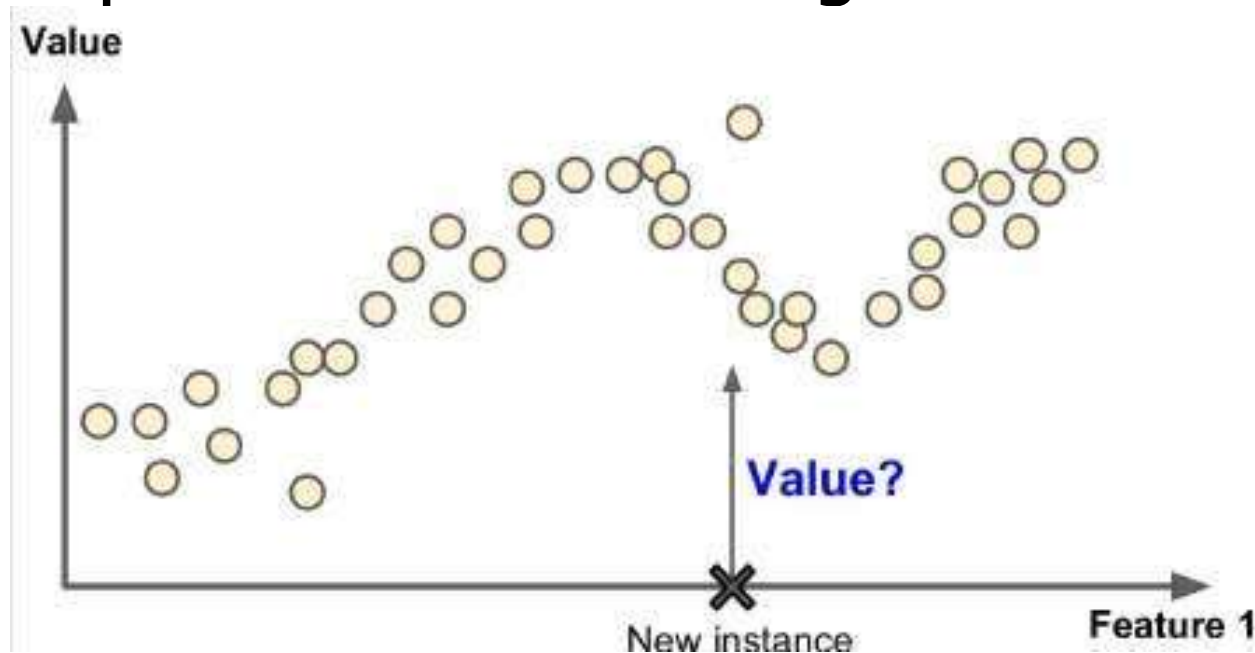


Figure: Aprendizaje supervisado en regresión. fuente: Hands on machine learning.

- Se mantiene el mismo concepto que en clasificación, cambiando el tipo de variable etiqueta por variable continua.
- Ejemplo: predicción valor del dólar en COP.

Sin supervisión humana: agrupamiento

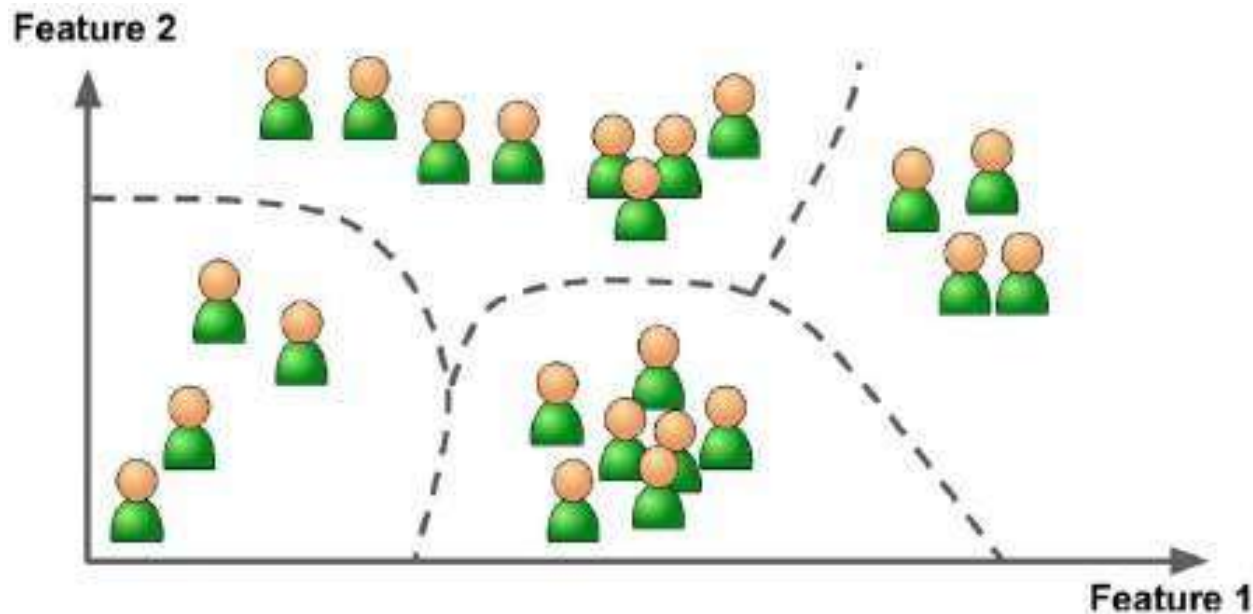


Figure: Aprendizaje no supervisado - agrupamiento (conglomerados).

fuelle: Hands on machine learning.

- Se buscan grupos a partir de las relaciones entre las instancias (regularidades entre datos).
- Ejemplo: perfilamiento de clientes en bancos.

Sin supervisión humana: reducción de dimensión

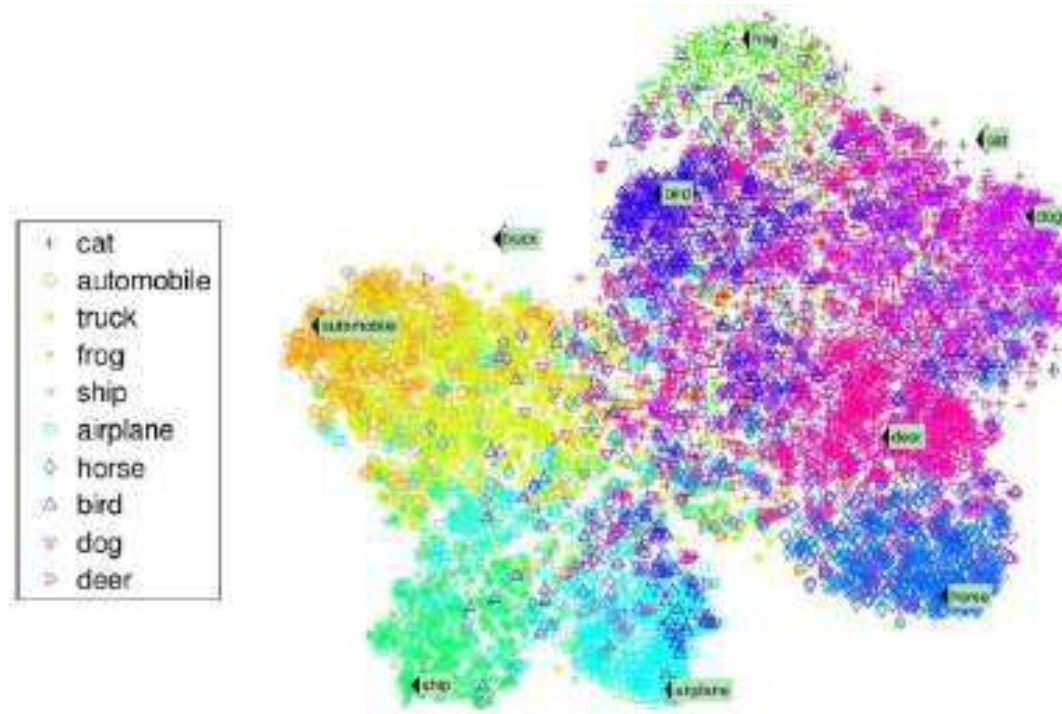


Figure: Aprendizaje no supervisado - visualización de datos. fuente: Hands on machine learning.

- Preservar relaciones de alta dimensión (espacio original de instancias) en un espacio de baja dimensión.
- Ejemplo: Visualización de datos con Dashboards

Sin supervisión humana: detección de anómalos



Figure: Aprendizaje no supervisado - detección de anómalos. fuente: Hands on machine learning.

- La nueva instancia sigue las regularidades encontradas en el espacio de entrenamiento?
- Ejemplo: detección de ataques o fraudes bancarios.



Semi supervisado

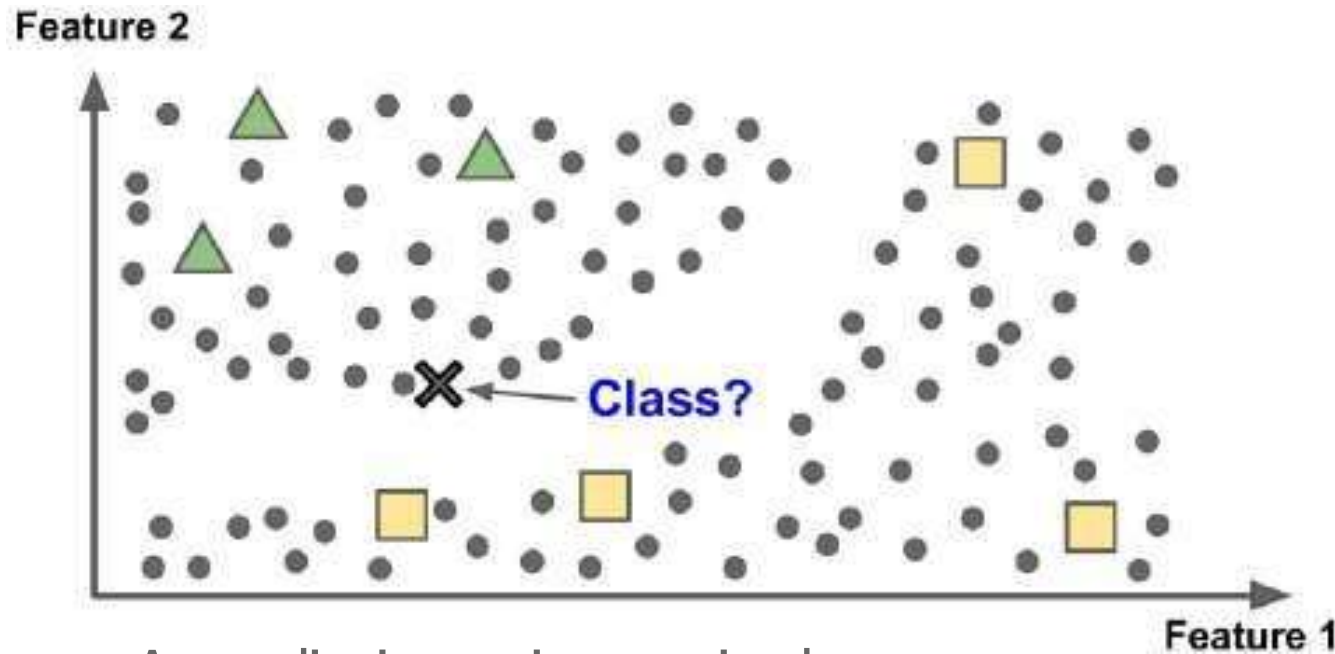
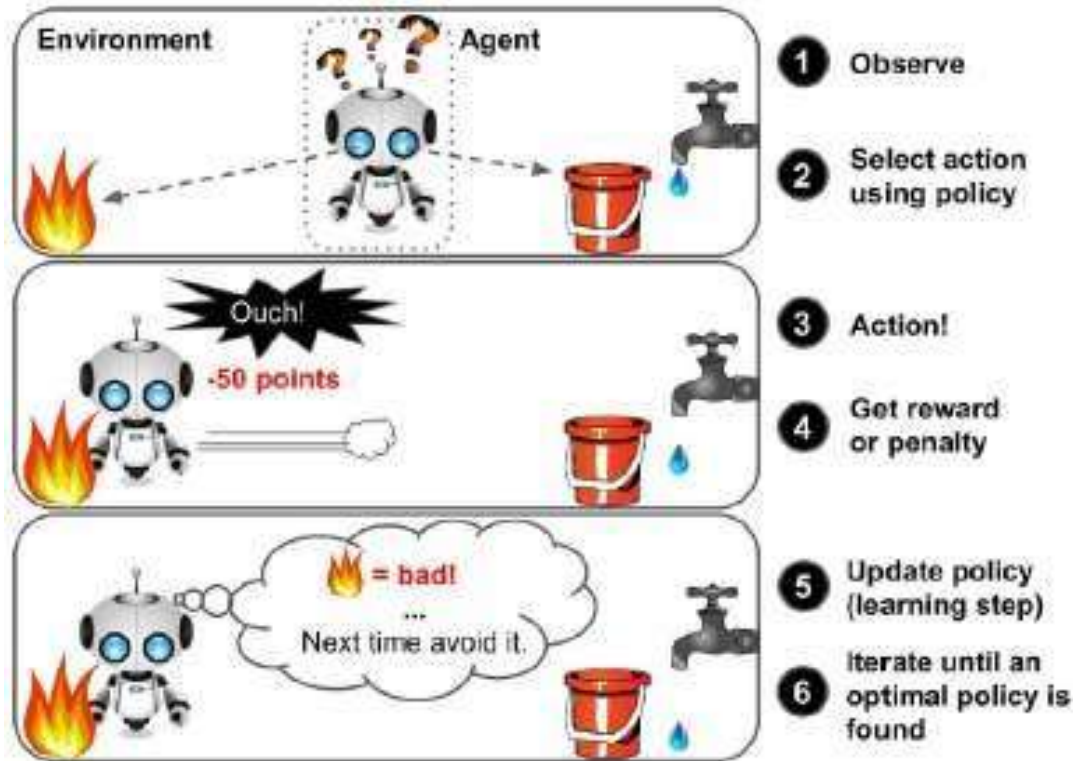


Figure: Aprendizaje semi supervisado. fuente: Hands on machine learning.

- Algunas instancias poseen etiqueta (con supervisión humana) pero la mayoría no (sin supervisión humana).
- Ejemplo: etiquetado de imágenes médicas



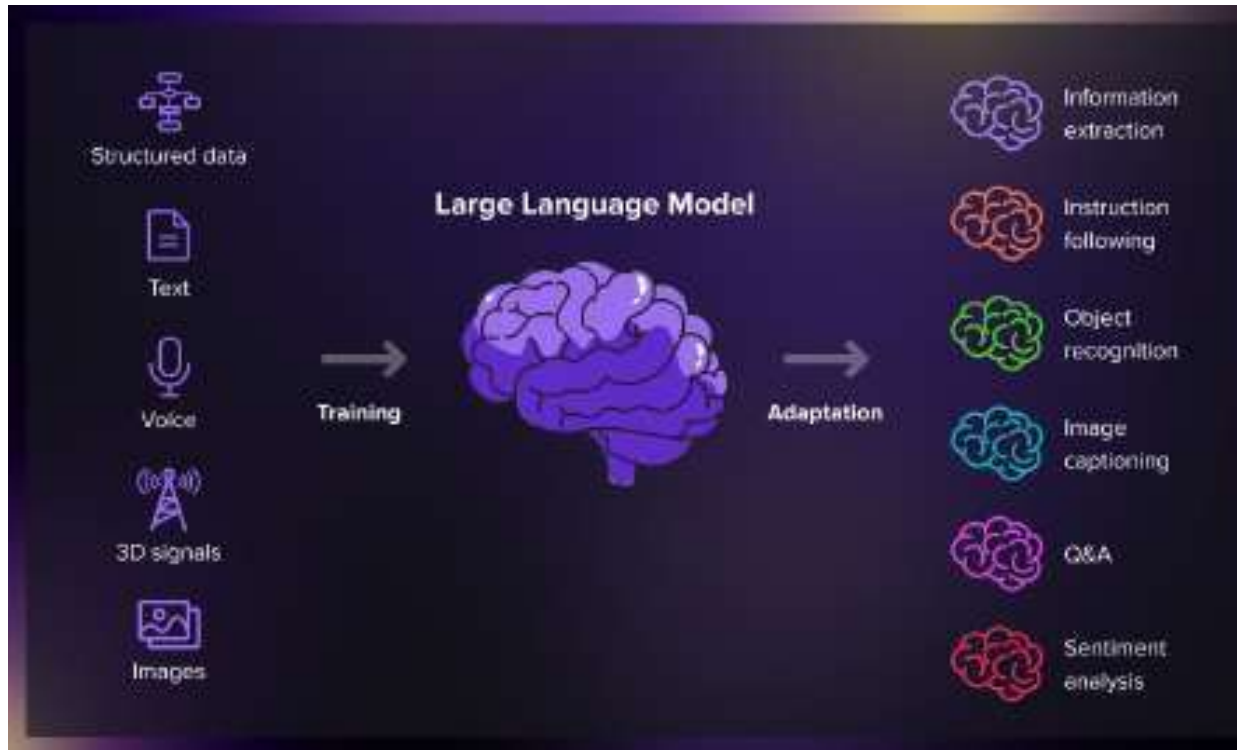
Aprendizaje por refuerzo



- El sistema (agente), observa el ambiente y toma decisiones obteniendo recompensas o penalizaciones.
- Ejemplo: [Control de videojuegos](#)

Figure: Aprendizaje por refuerzo. fuente: Hands on machine learning.

Modelos generativos: LLM



- Procesamiento de Lenguaje Natural
- Large Language Models
- Ejemplo: [Deep Fakes,](#) [ChatGPT-](#) [OpenAI o1](#)

IA predictiva vs generativa

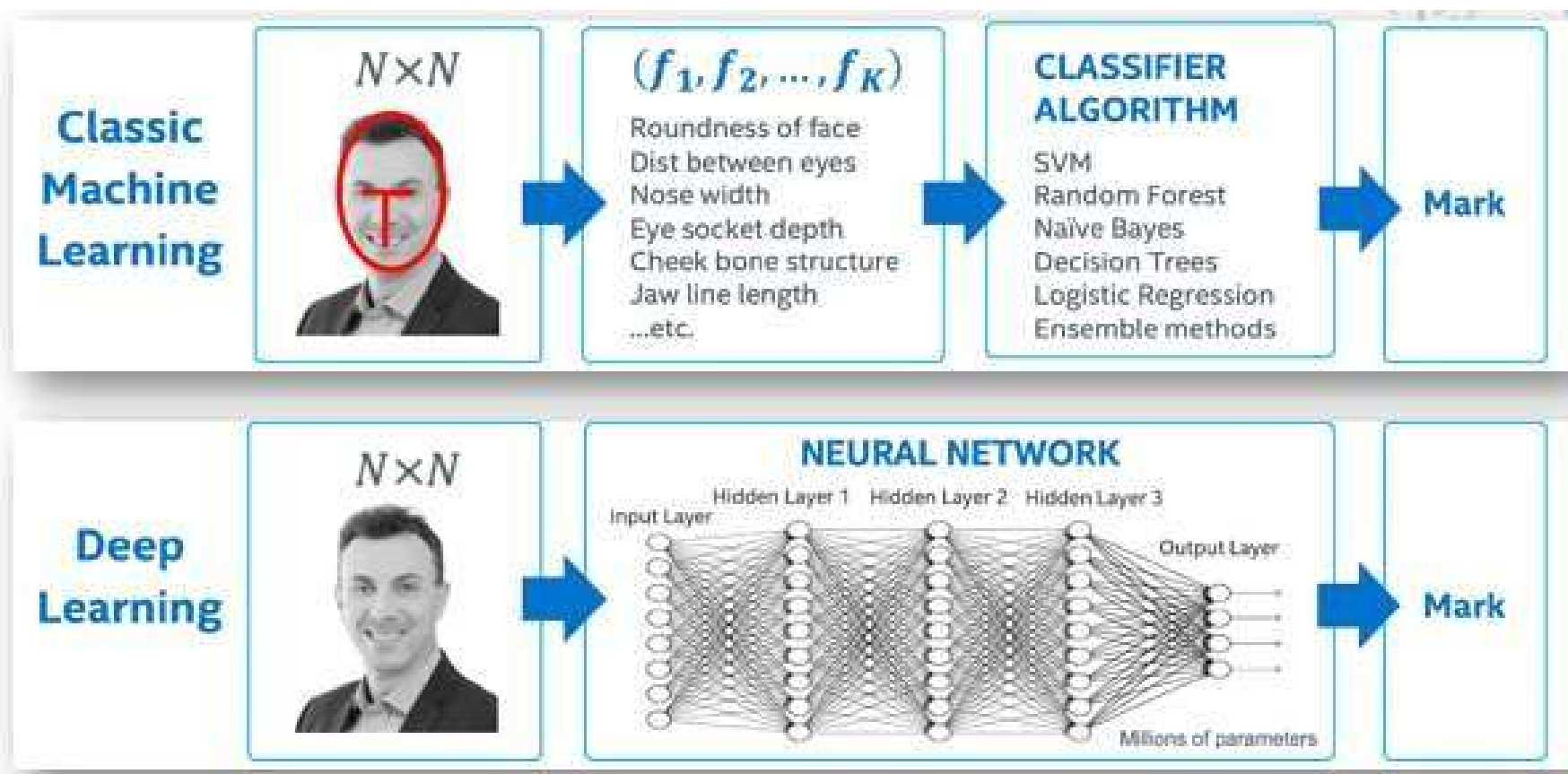


- Modelos fundacionales: entrenamiento en datos masivos + arquitectura avanzada = representaciones profundas y generalizadas.

Contenido

- 1 [Datos e Información](#)
- 2 [Aclaremos conceptos](#)
- 3 [Por qué Python y Nube?](#)
- 4 [Tipos de aprendizaje](#)
- 5 [La clave del éxito](#)
- 6 [Conclusiones](#)

Aprendizaje clásico vs. Aprendizaje profundo



- Herramientas de proceso idóneas
- Grandes cantidades de datos
- Capital humano capacitado (interdisciplinar)



**Universidad
Popular del Cesar**

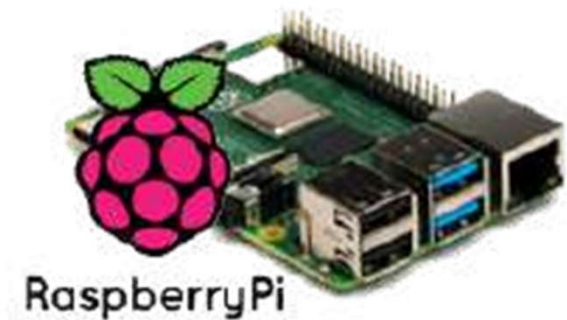
Aplicaciones a la medida: Webservices-Dashboards

**Facilidad de ejecución de modelos en ciencia de datos desde
servicios web y dispositivos móviles**



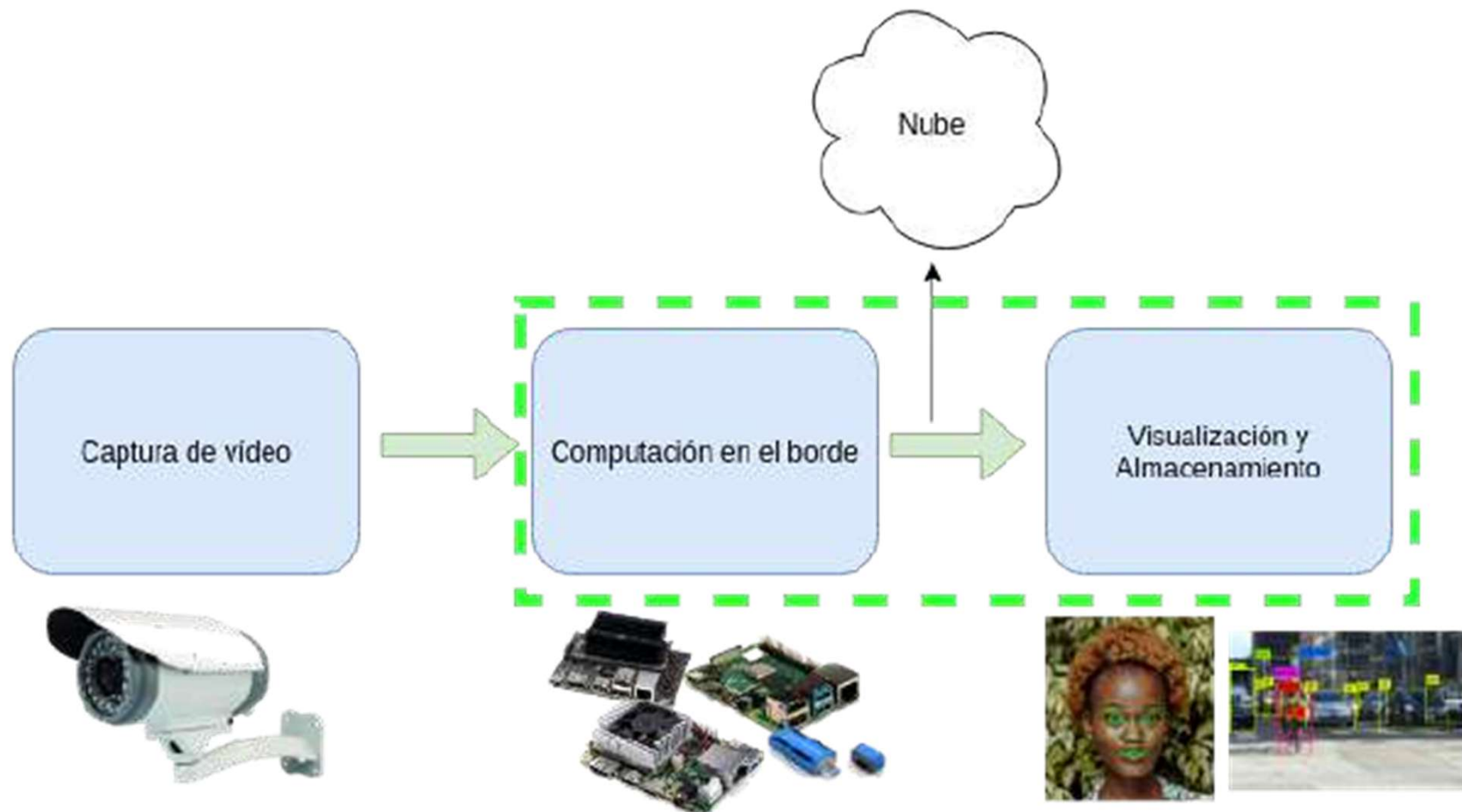
Implementación de IA: Sistemas Embebidos

- Económico
- Flexible
- Portable



Aplicaciones a la medida

A pesar de su bajo costo estos dispositivos pueden realizar tareas complejas en tiempo real





**Universidad
Popular del Cesar**

Contenido

- 1 [Datos e Información](#)
- 2 [Aclaremos conceptos](#)
- 3 [Por qué Python y Nube?](#)
- 4 [Tipos de aprendizaje](#)
- 5 [La clave del éxito](#)
- 6 [Conclusiones](#)

Conclusiones

- Se requiere de **capital humano inter-disciplinar** para **extraer información relevante**.
- **Buen uso de datos = buen uso de recursos = mayor competitividad = mayor seguridad.**
- **Mucho por hacer, investigar, e implementar!**

NO tienes que ser experto
en estadística y programación!



Nuevo lenguaje de programación: Lenguaje humano!

Gracias!

Prof. José Ramón Iglesias gamarra, Ph.D.

email: joseiglesias@unicesar.edu.co