

Programación del “entrenamiento”

- Introducción a Machine Learning (Hoy!) → Teórico.
- Introducción a python → Práctico.
- Modelo de regression → Teórico.
- Modelo de clasificación → Teórico.
- Recomendaciones para implementar modelos de Machine Learning → Teórico.
- Preparación de datos en python → Práctico.
- Práctica de Machine Learning I → Práctico.
- Práctica de Machine Learning II → Práctico.



Introducción a Machine Learning

Junio 14 de 2019
Medellín

Contenido general

- Esquema general de lo qué se desea resolver.
- Definición de Machine Learning.
- Tipos de aprendizaje.
- ML vs IA vs Deep Learning vs Big Data.
- Ciclo de vida de un Proyecto de ML.

El reto

La mayoría de las veces el tema está enmarcado en un problema de **predicción**:

- Dada una entrada x
- Predecir una salida y



Entrada: Email.

Salida: “Spam” o “No Spam”.

Problema de clasificación binario.



Entrada: Síntomas (fiebre, tos, náuseas, temblores, ...).

Salida: Diagnostico (neumonía, gripa, resfriado, bronquitis, ...)

Problema de clasificación multiclase.

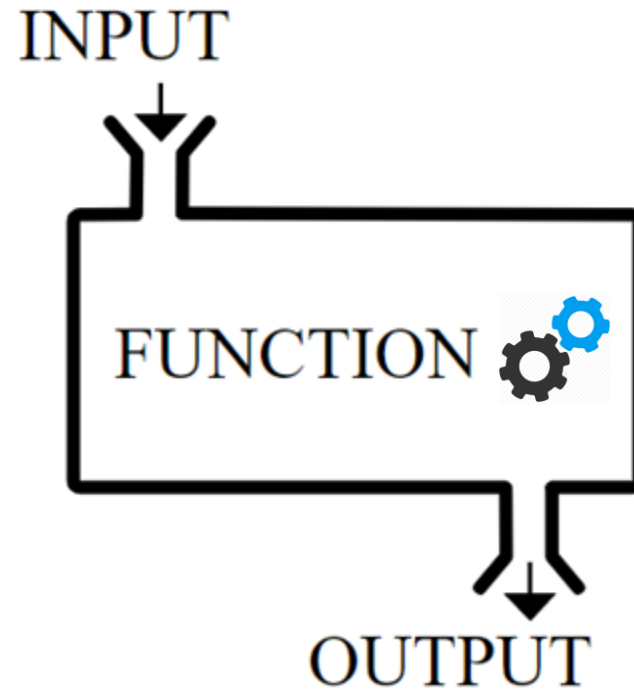


Entrada: Histórico del Gross mensual.

Salida: Predicción del Gross al mes siguiente

Problema de regresión.

La función de predicción

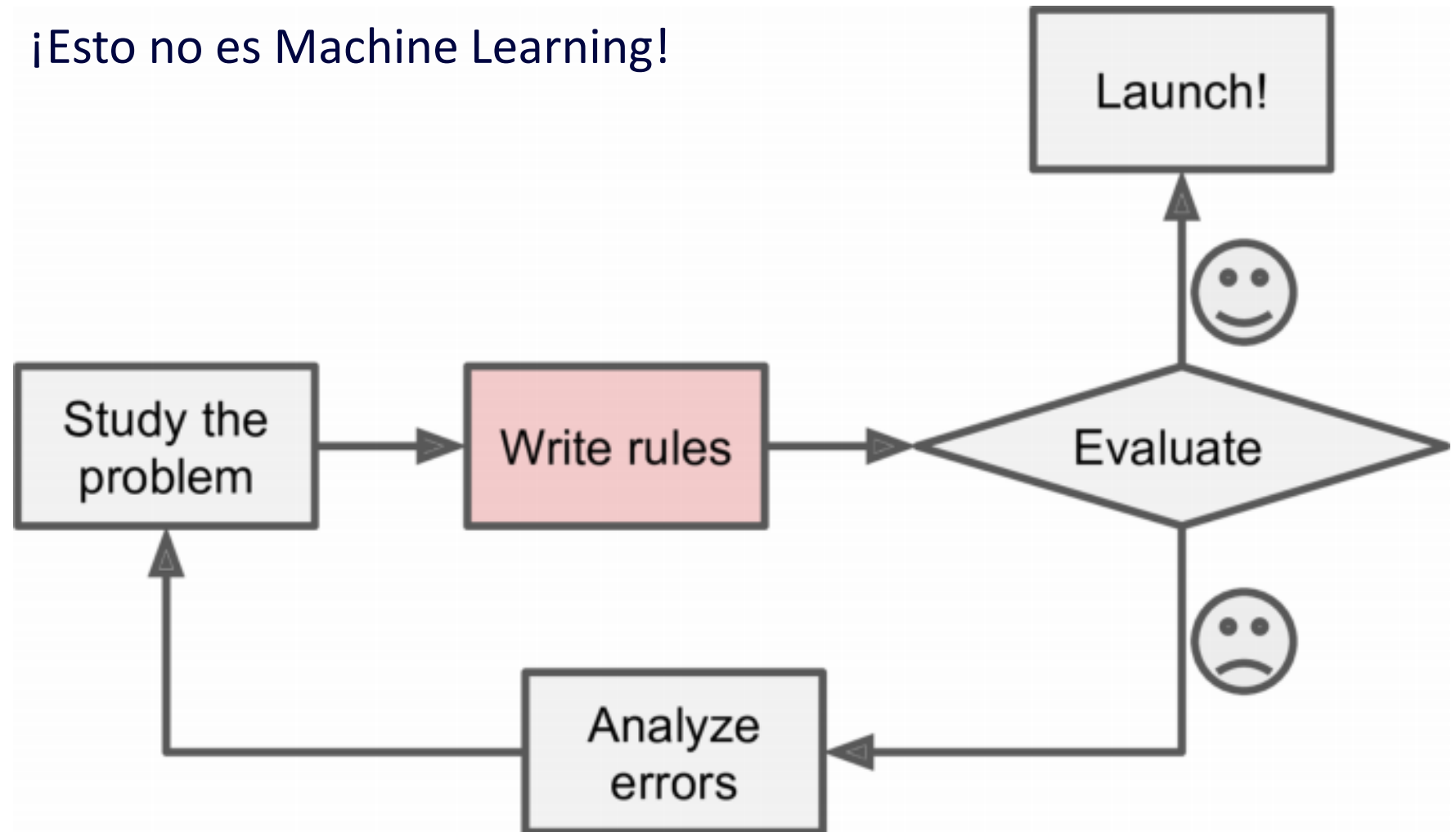


La idea es buscar una función de predicción que resuelva un problema particular.
El Machine Learning ayuda a encontrar la “mejor” función de predicción.

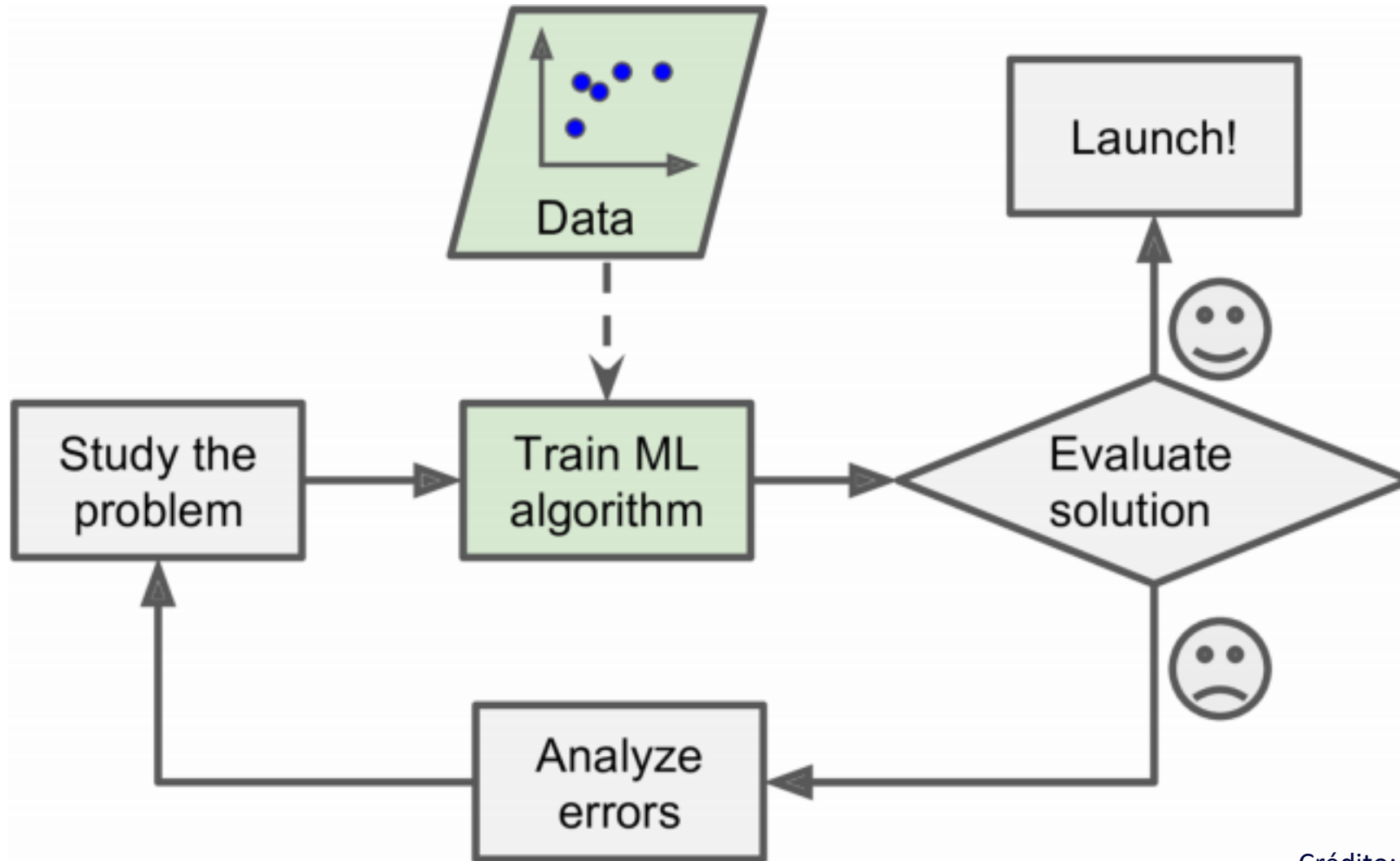
Sistema experto o (esquema basado en reglas)

- Mucho esfuerzo en la construcción.
- Trabajan muy bien en las áreas que cubren pero no pueden ser generalizados a casos no anticipados.
- No manejan de manera natural la incertidumbre.
- Son tomados como “frágiles”.

¡Esto no es Machine Learning!

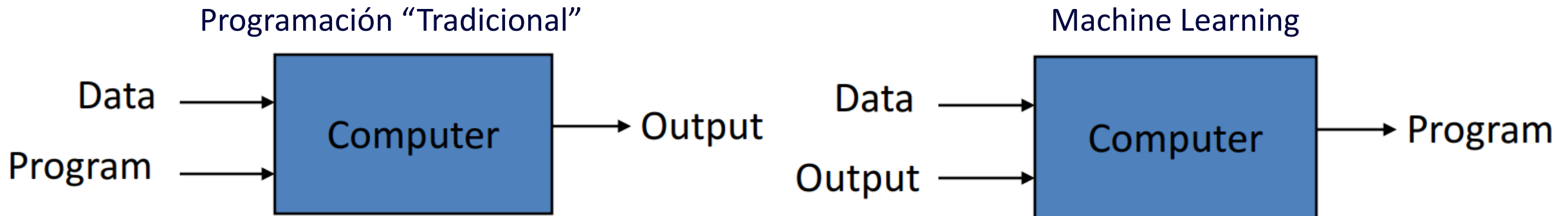


La estrategia del Machine Learning



¿Qué es Machine Learning?

- Una definición “soft”:
 - Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
- Una definición más formal:
 - Tom Mitchell (1998). A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.



¿Cuándo es buena idea usar Machine Learning?

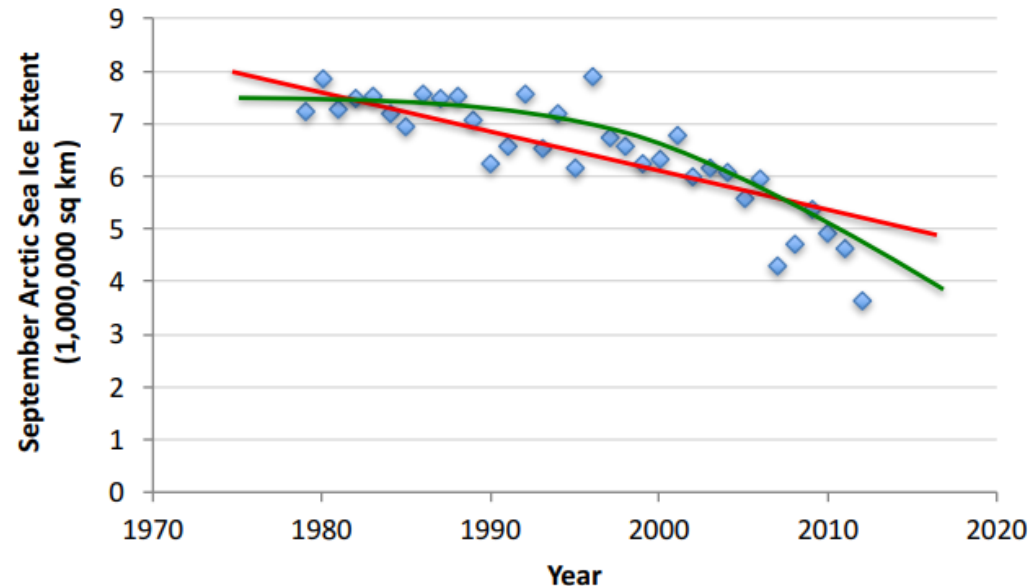
- No se tiene un conocimiento experto.
- No se puede explicar la experiencia en un tema.
- Los modelos deben ser constantemente personalizados.
- Los modelos son basados en grandes cantidades de datos.

Tipos de aprendizaje

- **Aprendizaje supervisado:** Datos de entrenamiento + salida deseada (labels).
 - Modelos de regresión.
 - Modelos de clasificación.
- **Aprendizaje no supervisado:** Datos de entrenamiento (sin labels).
 - Algoritmos de clustering.
 - Análisis de componentes principales.
- **Aprendizaje semisupervisado:** Datos de entrenamiento + algunos labels.
- **Aprendizaje por refuerzo:** Recompensas a partir de una secuencia de acciones.

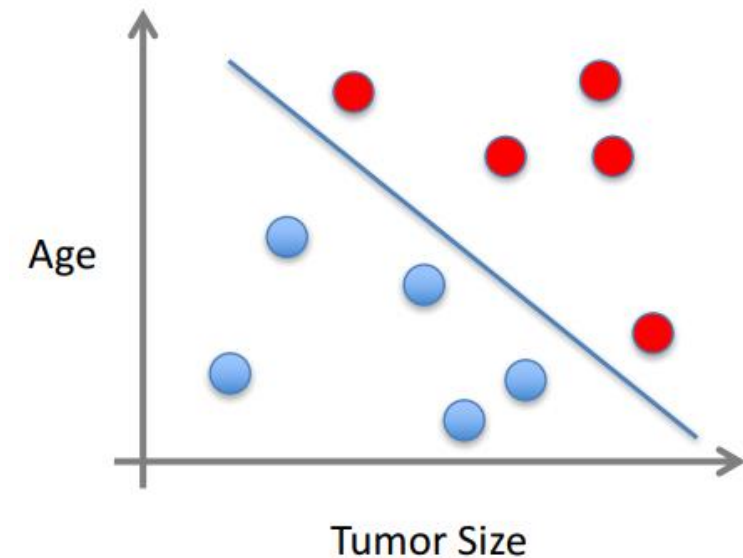
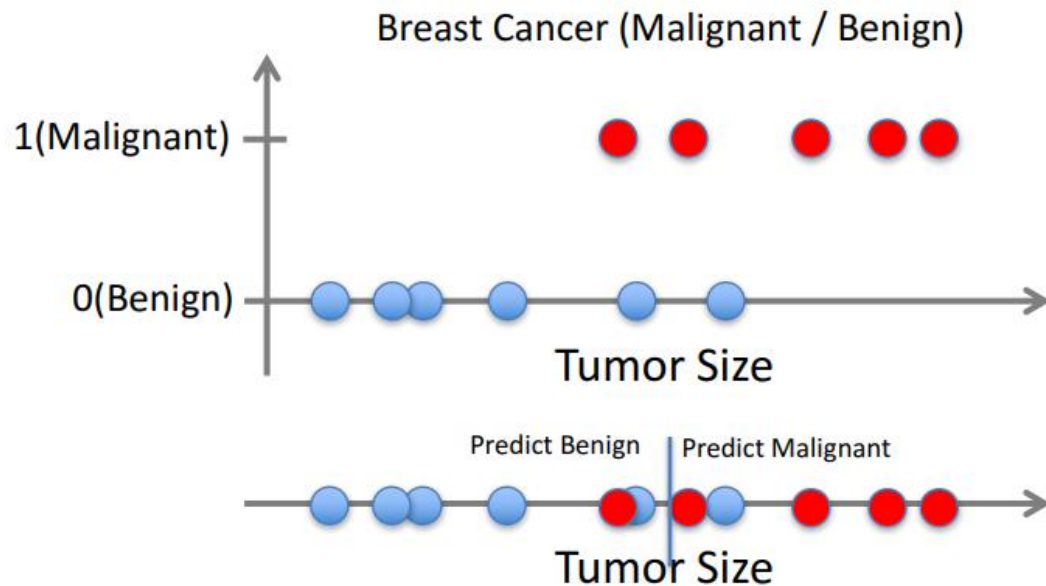
Regresión

- Dado un conjunto de datos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- Aprender una función $f(x)$ (una maquinita) que prediga y dado x
 - Si y es un número real, es decir, con “infinitos” posibles valores \rightarrow Regresión



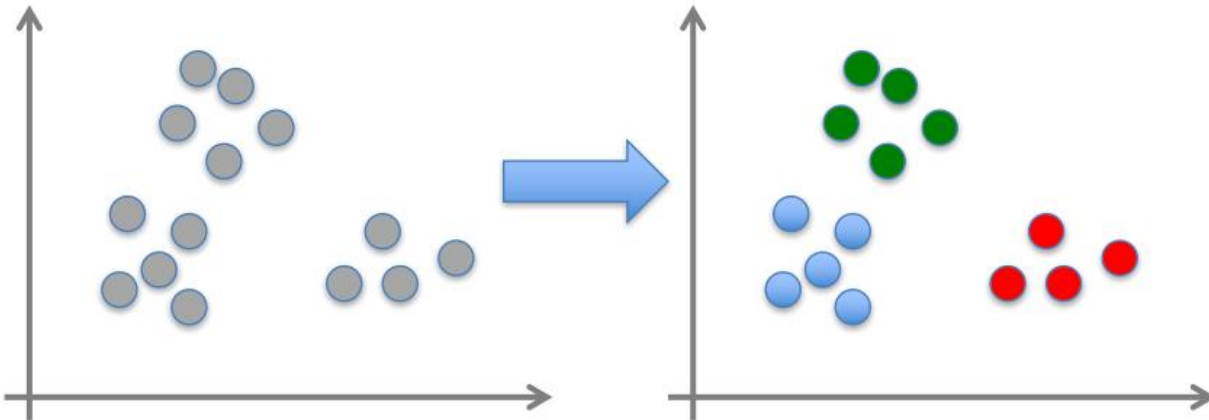
Clasificación

- Dado un conjunto de datos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- Aprender una función $f(x)$ (una maquinita) que prediga y dado x
 - Si y es categórico, es decir, un número finito de opciones \rightarrow Clasificación

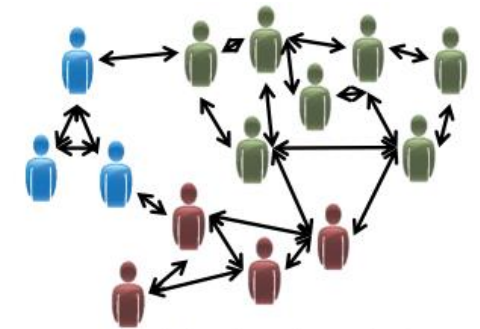


Clustering

- Dado un conjunto de datos x_1, x_2, \dots, x_n (sin etiquetas)
- Encontrar estructuras detrás de las x 's



Organize computing clusters



Social network analysis

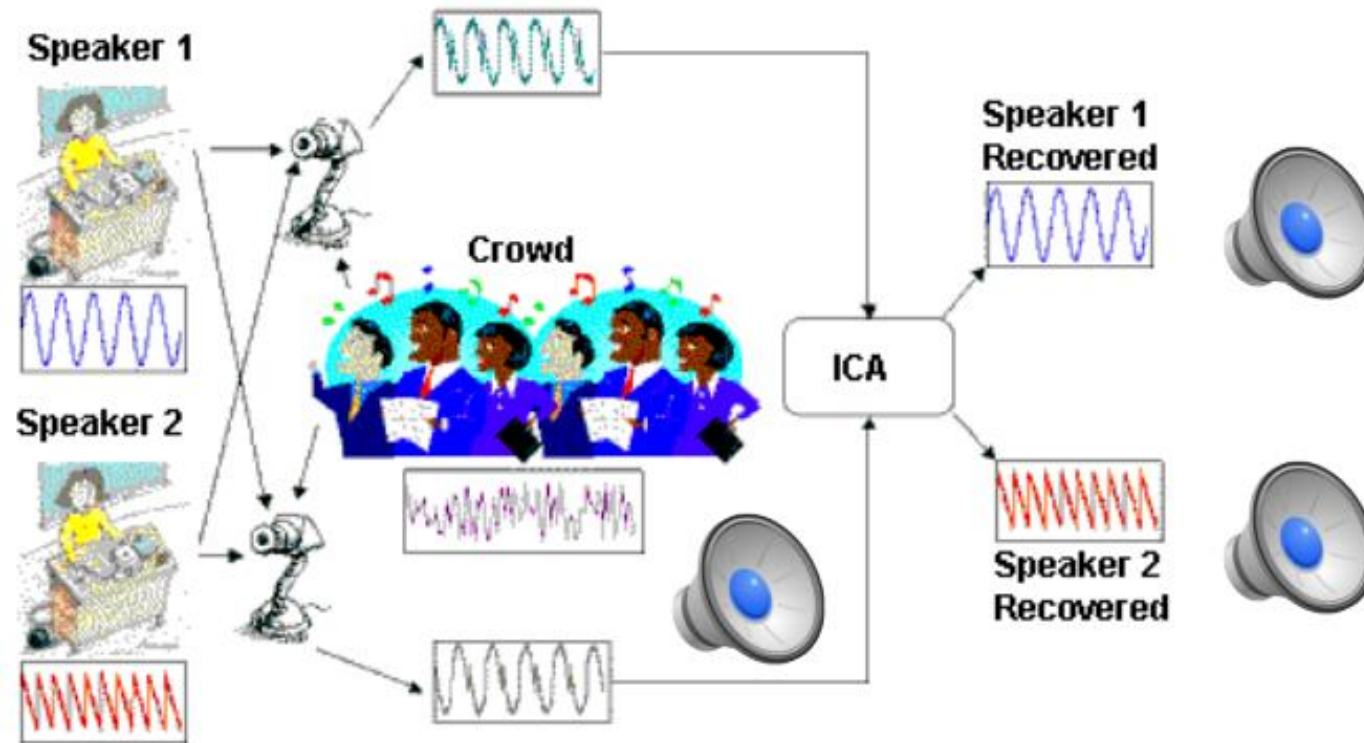


Market segmentation



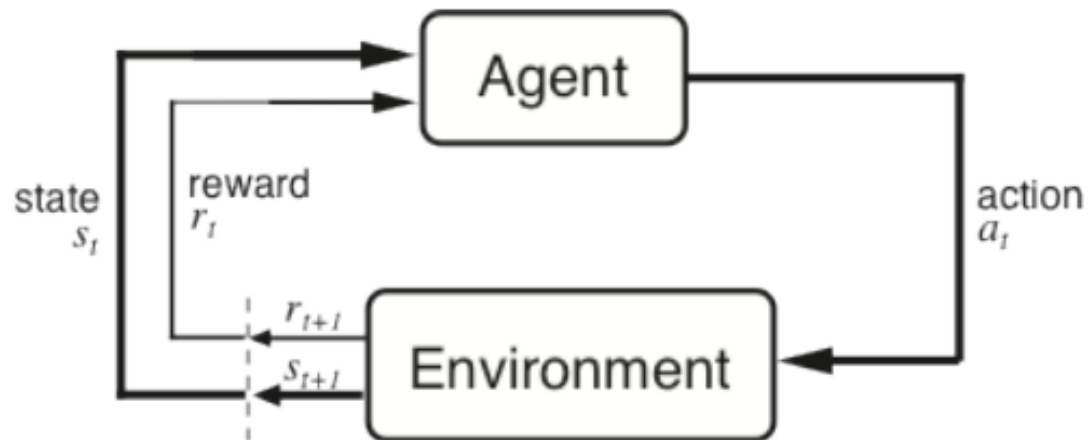
Astronomical data analysis

Análisis de componentes independientes



Aprendizaje por refuerzo

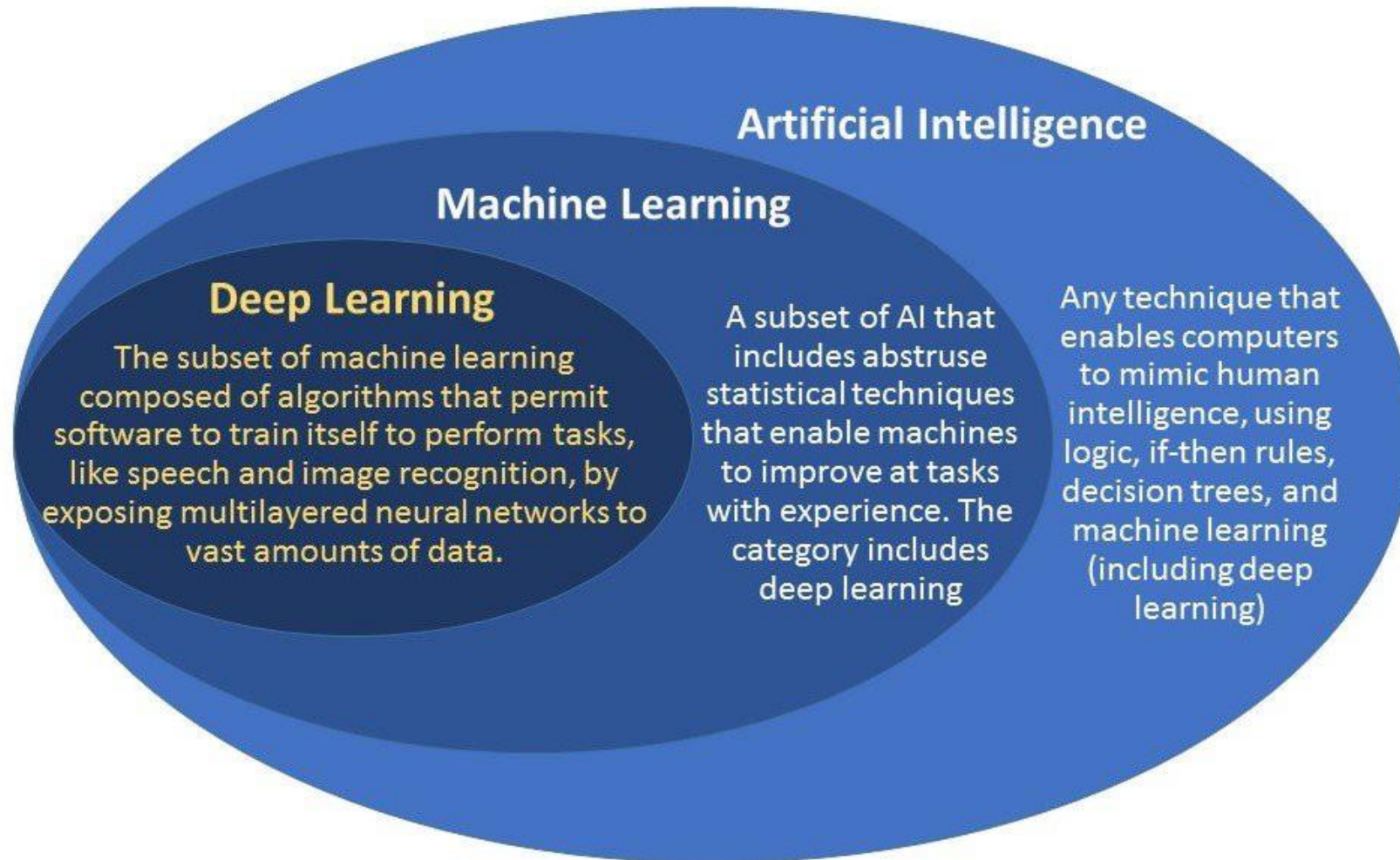
- Dada una secuencia de estados y acciones con recompensas establecer una política, es decir, las acciones a realizar en un estado dado.



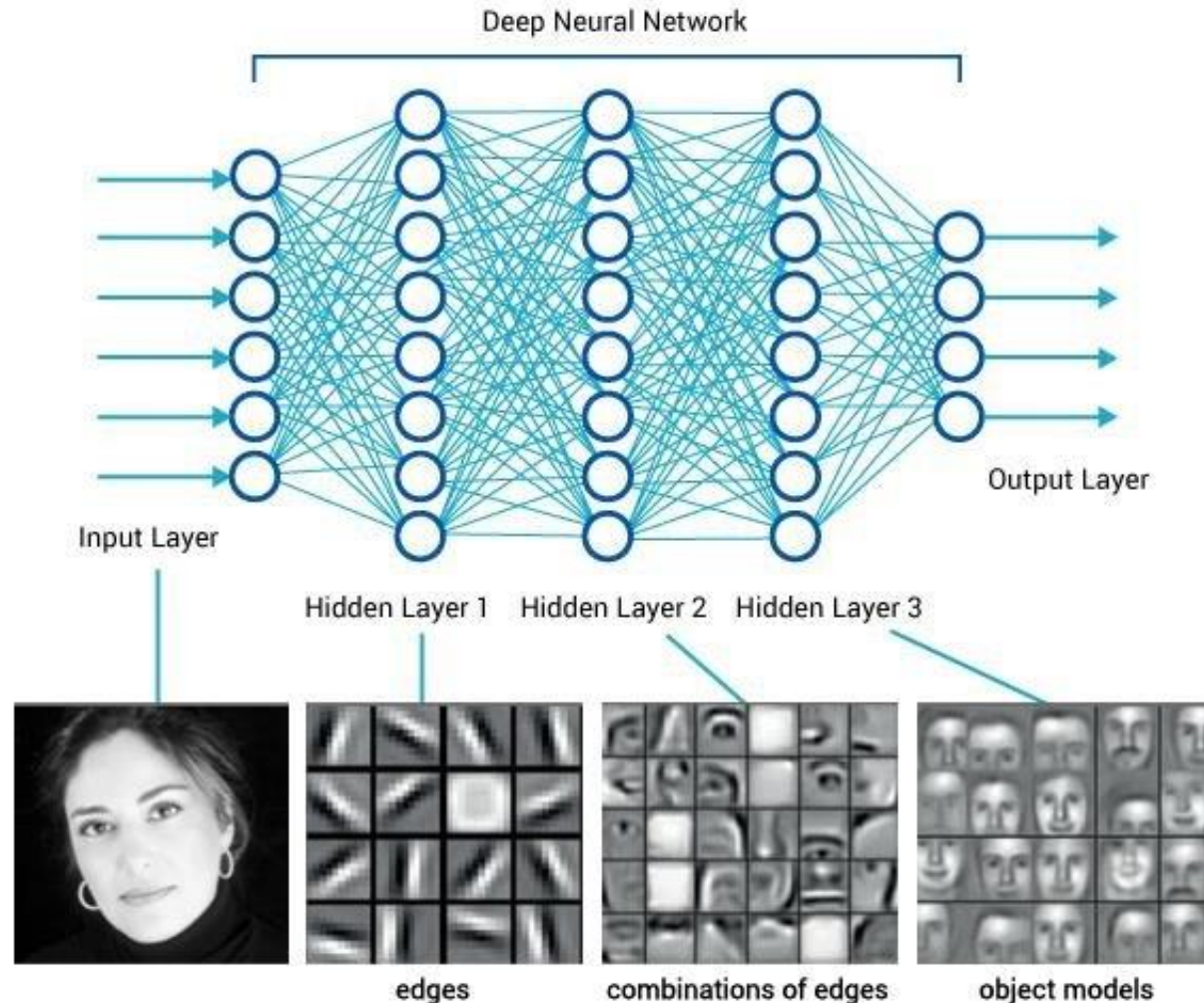
<https://www.youtube.com/watch?v=4cgWya-wjgY>



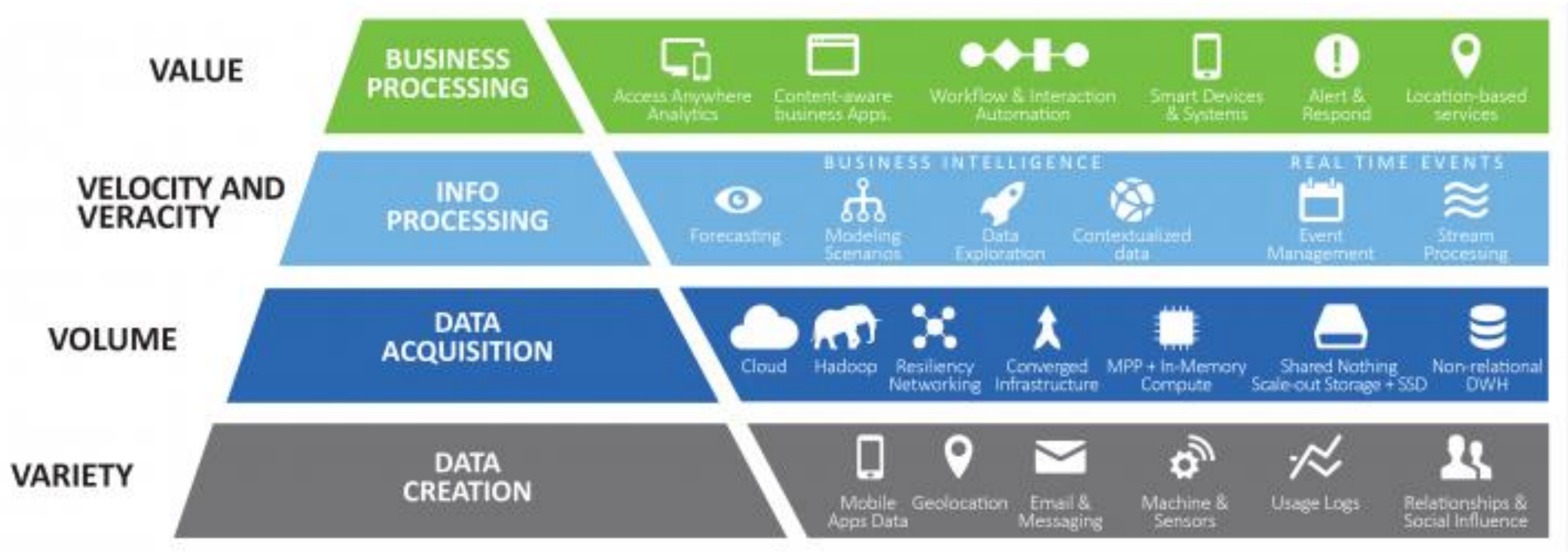
- Es común que se usen como si fueran la misma cosa, pero **no** son los mismo!!!!
- La IA se relaciona a máquinas que pueden realizar tareas que son características de la inteligencia humana (planear, entender el lenguaje, reconocer objetos y sonidos, resolver problemas nuevos).
- En principio uno podría codificar un IA sin usar Machine Learning pero puede llegar a requerir de millones de líneas de código para realizar una tarea sencilla para un humano y con las claras desventajas de los sistemas expertos.
- El Machine Learning es una forma de llegar a una IA.
- Deep Learning es un tipo de Machine Learning y usa las llamadas redes neuronales profundas (Artificial Neural Network; ANN).
- Big Data hace relación al gran volumen de datos, tanto estructurados (e.g. SQL) como no estructurados (e.g. json), que inundan los negocios cada día y el uso que le dan las organizaciones para darles valor.



Inteligencia Artificial vs Machine Learning vs Deep Learning vs Big Data



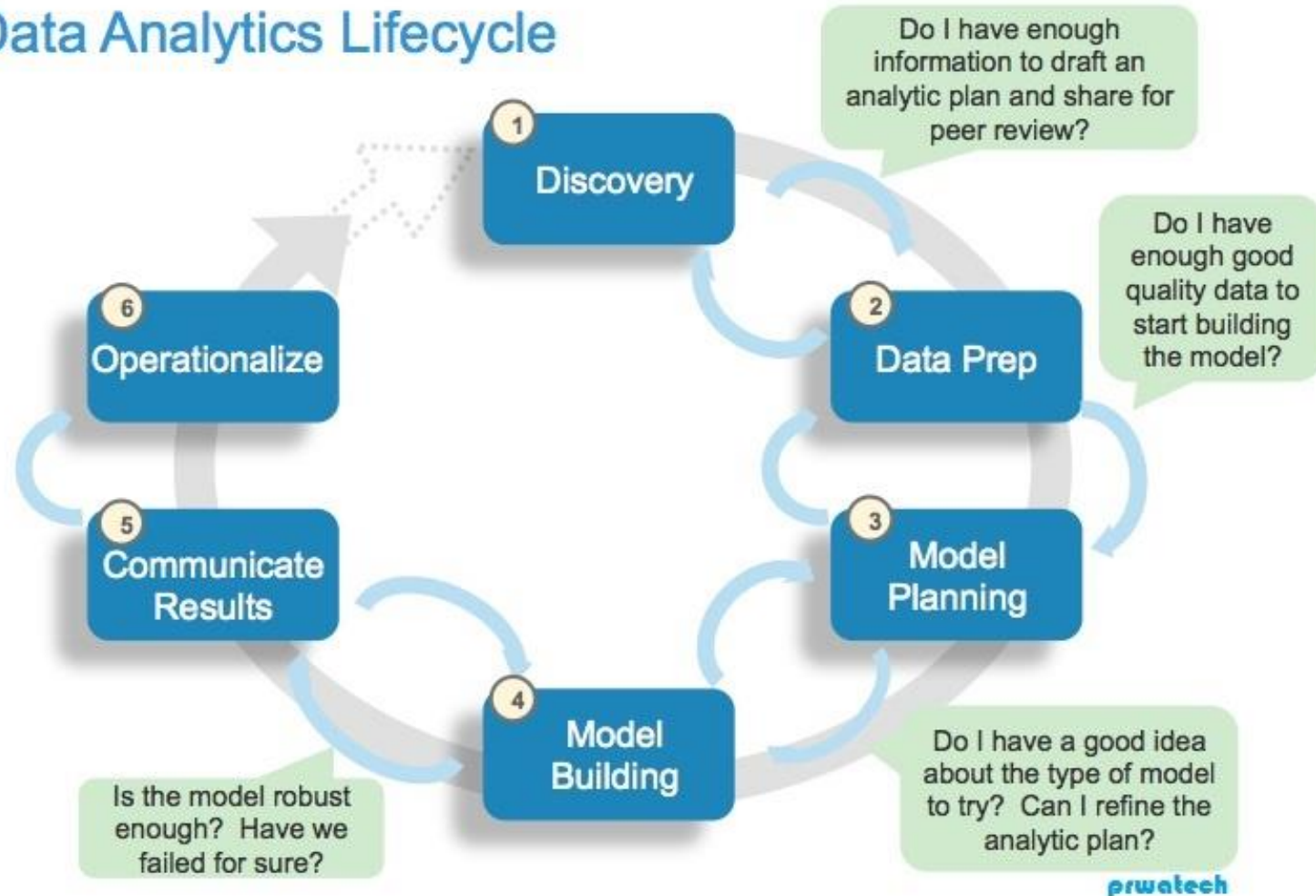
Inteligencia Artificial vs Machine Learning vs Deep Learning vs Big Data



Exalecom, Inc: <https://www.exalecom.com/resources/blog/the-5-vs-of-big-data-predictions-for-2016/>

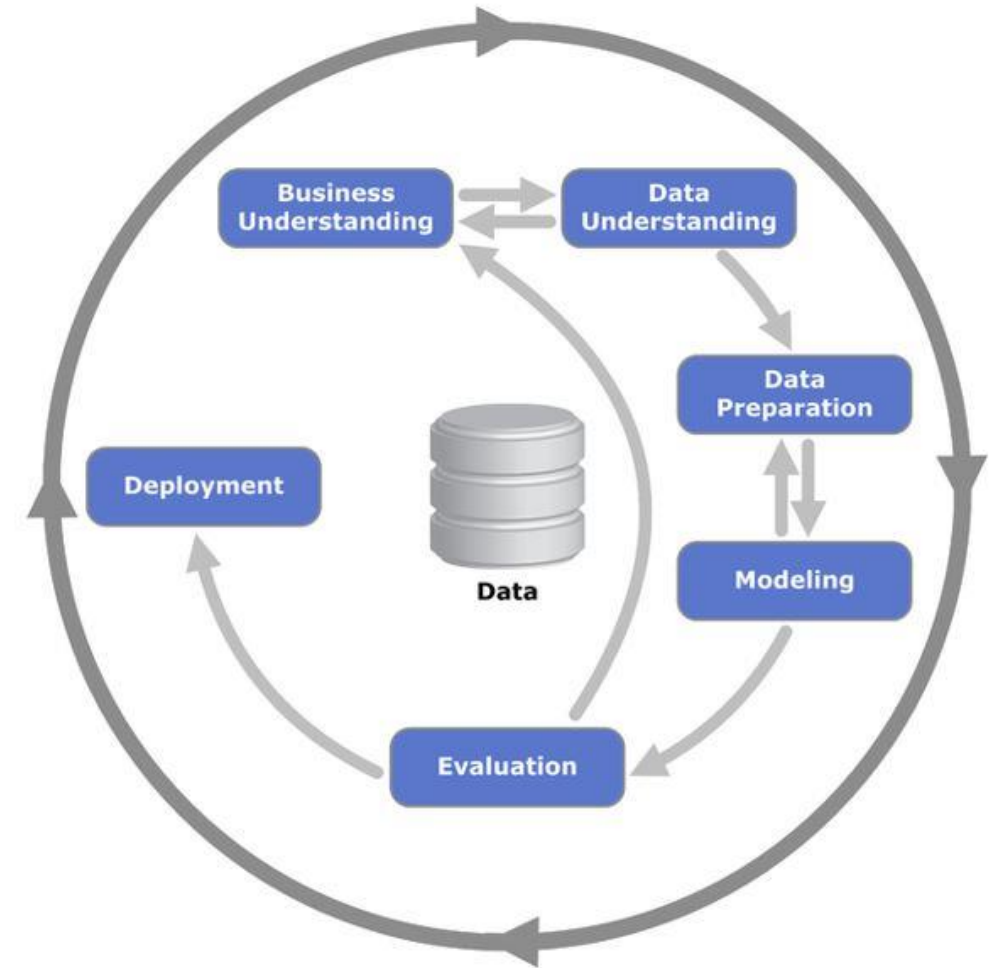
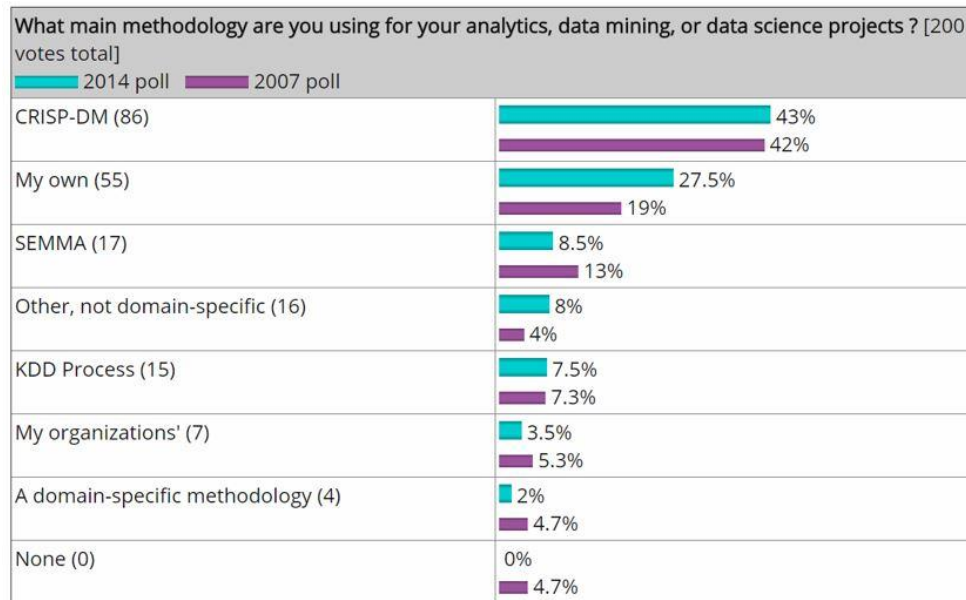
(El caso ideal del) Ciclo de vida de un Proyecto de ML

Data Analytics Lifecycle



Caso particular: Metodología CRISP-DM

Cross-industry standard process for data mining



<https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>

<https://commons.wikimedia.org/w/index.php?curid=24930610>

Referencias

- Machine Learning, Andrew Ng (Coursera)
- Foundations of Machine Learning, David S. Rosenberg <https://bloomberg.github.io/foml/>
- <https://medium.com/iotforall/the-difference-between-artificial-intelligence-machine-learning-and-deep-learning-3aa67bff5991>
- <https://medium.com/datadriveninvestor/ai-vs-machine-learning-vs-deep-learning-ba3b3c58c32>
- https://www.seas.upenn.edu/~cis519/fall2017/lectures/01_introduction.pdf

tigô