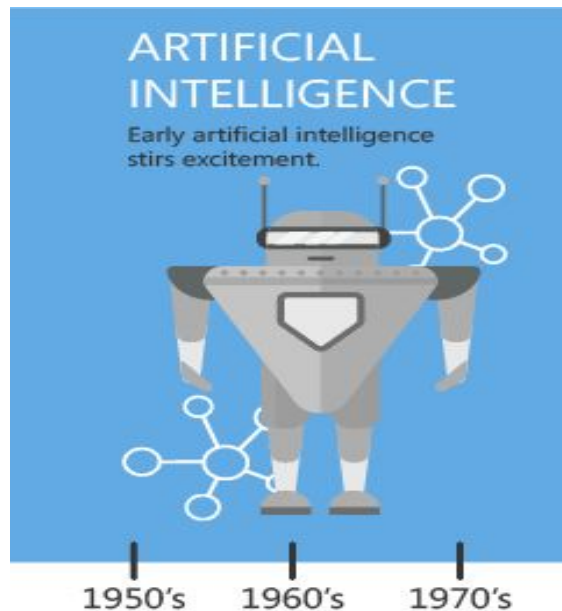




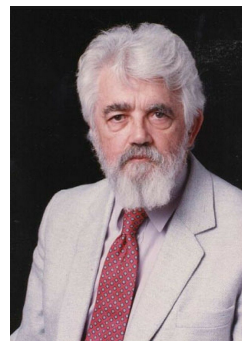
Python para ML

Introducción a Machine Learning

Historia

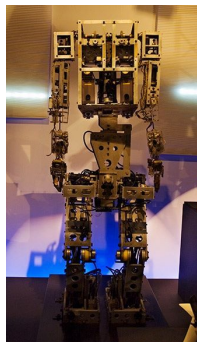


En 1954 John McCarthy acuña el término “inteligencia artificial”



“Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.”
-Prof. John McCarthy

Nacimiento de las primeras máquinas “inteligentes”

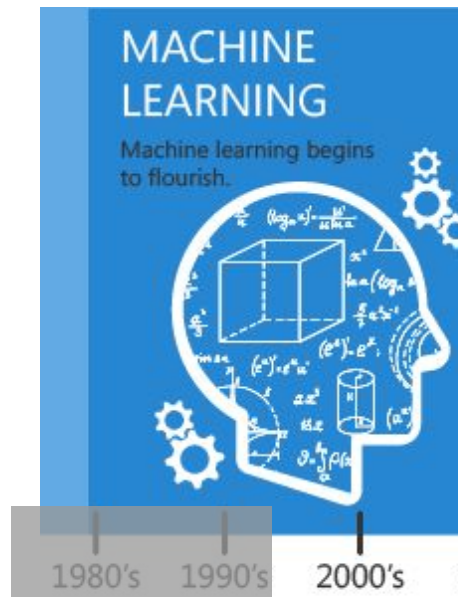


Ferranti Mark 1 (1951):
jugador de damas

General Problem Solver (1959):
motor solucionador genérico

WABOT-1 (1972):
primer robot humanoide inteligente

Historia



↓
winter years

Entre finales de 1970 e inicios de 1990, los científicos informáticos encontraron dificultades en conseguir financiamiento para investigar sobre inteligencia artificial

poco poder
computacional

limitados usos
prácticos

experimentaciones
fallidas



1982: Japón pone en marcha el “Proyecto Informático de Quinta Generación” (FGCP). Invirtieron \$400 millones revolucionando el procesamiento informático, implementando programación lógica y mejorando la inteligencia artificial.

1997:

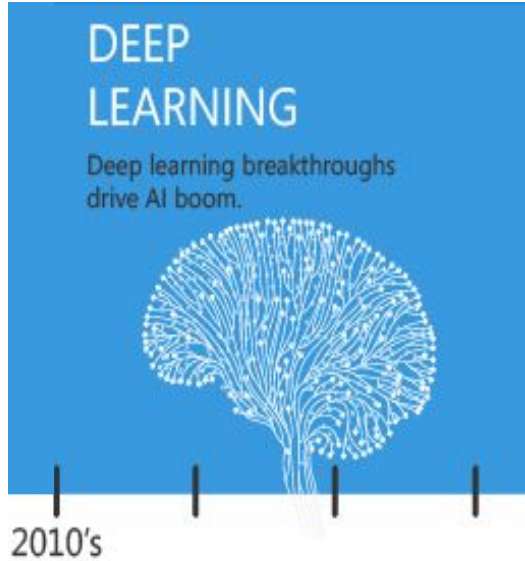
Deep Blue (IBM) vence al campeón mundial de ajedrez

Windows implementa **Dragon Systems**, reconocimiento de lenguaje

Se inicia **Kismet**, robot que reconoce y demuestra emociones humanas

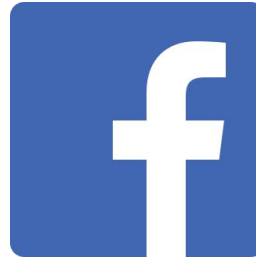
Historia

Ley de Moore: memoria y capacidad de computadoras se duplica cada año



Big Data:

procesamiento y extracción de información de data masiva (grandes cantidades) y compleja (distintos formatos)



Motivación



Desde hace décadas se habla de la capacidad de las máquinas para emular la inteligencia humana.

Las posibilidades de una máquina superan ampliamente a las de los humanos a nivel de procesamiento, memoria y físico.

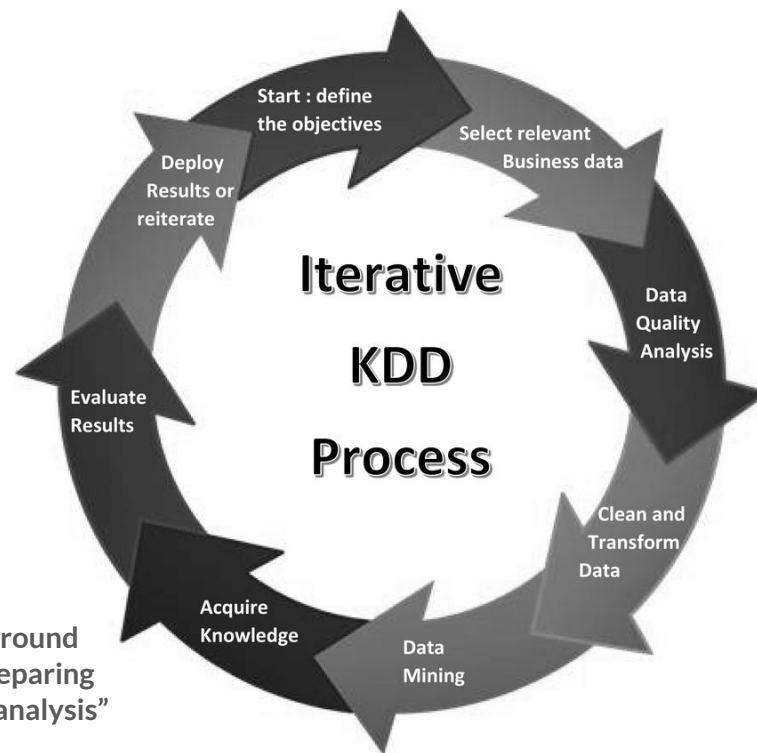
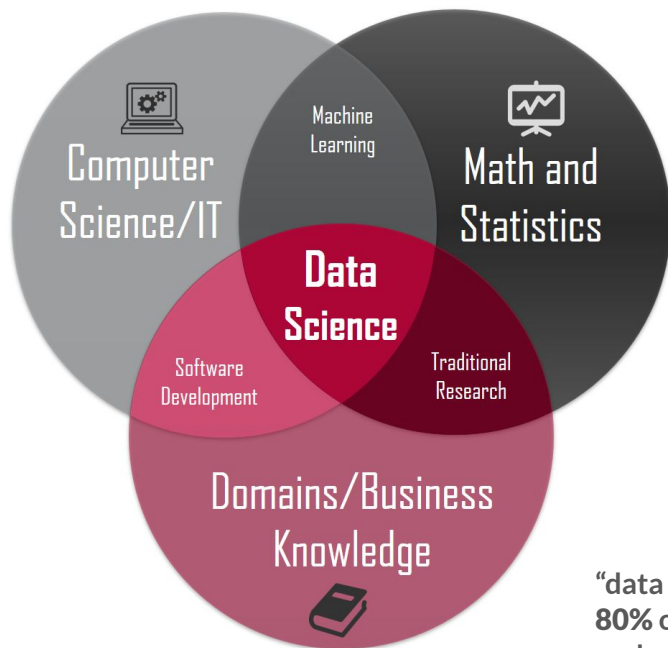
Además, hoy contamos con cantidades innumerables de data por procesar, es decir, conocimiento por extraer.

\$107,801 es el salario anual para un científico de datos en Estados Unidos.



Data Science

Combinación de conocimiento experto,
matemáticas y estadísticas y programación
para extraer patrones de la data



“data scientists spend around
80% of their time on preparing
and managing data for analysis”

IA & ML



IA

Machines that mimic "cognitive" functions that humans associate with the **human mind**, such as "learning" and "problem solving".

Ability of a digital computer to **perform tasks commonly associated with intelligent beings**.

concepto de la existencia de máquinas inteligentes

ML

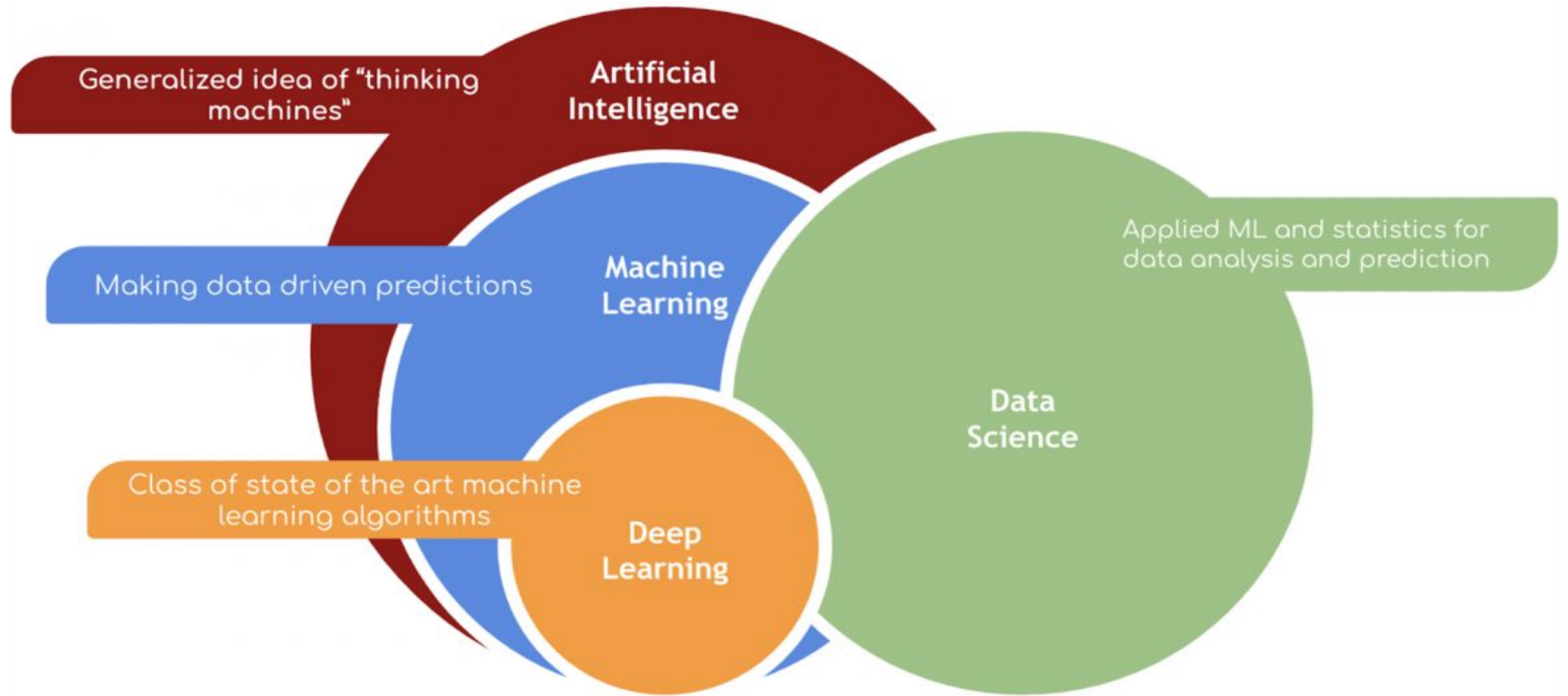
The study of computer **algorithms** that improve automatically through experience.

Algorithms that use statistics to find patterns in massive amounts of data.

Science of getting **computers to act** without being explicitly programmed.

uso de estadística y data para que una computadora aprenda algo

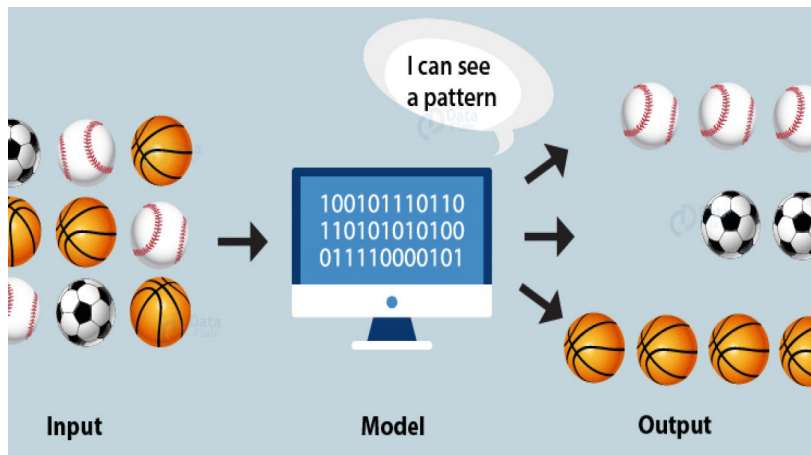
Definiciones



Machine Learning



Proceso en el que una computadora, a partir de mucha data, encuentra una función (o un patrón) que cumpla con un objetivo deseado



Machine Learning



Pasos

1. Juntar data
2. Preparar data
3. Elegir algoritmo(s)
4. Entrenar algoritmo(s)
5. Evaluar modelo
6. Ajustar parámetros
7. Hacer predicciones



Algoritmos de Machine Learning



Supervisados

En el aprendizaje supervisado, los algoritmos trabajan con **datos “etiquetados”** (labeled data), intentando encontrar una función que, dadas las variables de entrada (input data), les asigne la **etiqueta de salida adecuada**.

El algoritmo se **entrena** con un “histórico” de datos y así “aprende” a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, **predice el valor de salida**. (Simeone, 2018)

No supervisados

El aprendizaje no supervisado tiene lugar cuando no se dispone de datos “etiquetados” para el entrenamiento. **Sólo conocemos los datos de entrada**, pero no existen datos de salida que correspondan a un determinado input.

Por tanto, sólo podemos describir la estructura de los datos, para intentar encontrar algún tipo de **patrón**. Tienen un carácter **exploratorio**.

Algoritmos de Machine Learning



Clasificación (supervisado)

Algoritmos que buscan clasificar un conjunto de datos basados en una cantidad de clases limitadas.

¿Es perro o gato?

¿Está resfriado o no?

¿Está molesto, feliz o triste?

Principales usos

- Clasificación de imágenes
- Detección de fraude
- Retención de clientes
- Diagnósticos médicos

Principales algoritmos

- Logistic regression
- Support vector machines (SVM)
- Decision trees
- Random forests
- K-nearest neighbours (KNN)
- Boosting algorithms
- Artificial Neural Networks (ANN)

Algoritmos de Machine Learning



Regresión (supervisado)

Algoritmos que buscan predecir resultados numéricos en un rango infinito de valores.

¿Cuál será el PBI en el 2020?

¿Cuántos infectados habrá la próxima semana?

¿Cuál será la utilidad de la empresa?

Principales usos

- Pronóstico de resultados de marketing
- Pronóstico de clima
- Pronóstico de crecimiento poblacional
- Estimación de expectativa de vida
- Predicción de tiempo de permanencia

Principales algoritmos

- Linear and not linear regression
- Support vector machines (SVM)
- Decision trees
- Random forests
- K-nearest neighbours (KNN)
- Boosting algorithms
- Artificial Neural Networks (ANN)

Algoritmos de Machine Learning



Clusterización (no supervisado)

Algoritmos que agrupan observaciones con características (variables) similares.

¿Cuáles son mis segmentos de clientes?

¿Qué productos son similares?

¿Qué distritos son similares?

Principales usos

- Sistemas de recomendación
- Agrupamiento de clientes
- Agrupamiento poblacional
- Agrupamiento de células

Principales algoritmos

- K-Means
- Hierarchical clustering
- Density Based Scan Clustering (DBSCAN)
- Gaussian Clustering Model

Algoritmos de Machine Learning



Reducción de dimensionalidad (no supervisado)

Algoritmos que buscan agrupar variables con alta correlación.

Estos algoritmos se usan generalmente como fase previa a otros algoritmos cuando se cuenta con una gran cantidad de variables.

Si tenemos un dataset con 10 columnas, podemos utilizar un algoritmo de reducción de dimensionalidad para agrupar las columnas en 3 y poder graficar los datos.

Principales usos

- Preprocesamiento de datos
- Clusterización
- Visualización de datos

Principales algoritmos

- Análisis de componentes principales (PCA)
- Linear discriminant analysis (LDA)
- T-distributed Stochastic Neighbor Embedding (TSNE)
- Singular Value Decomposition (SVD)
- Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP)

