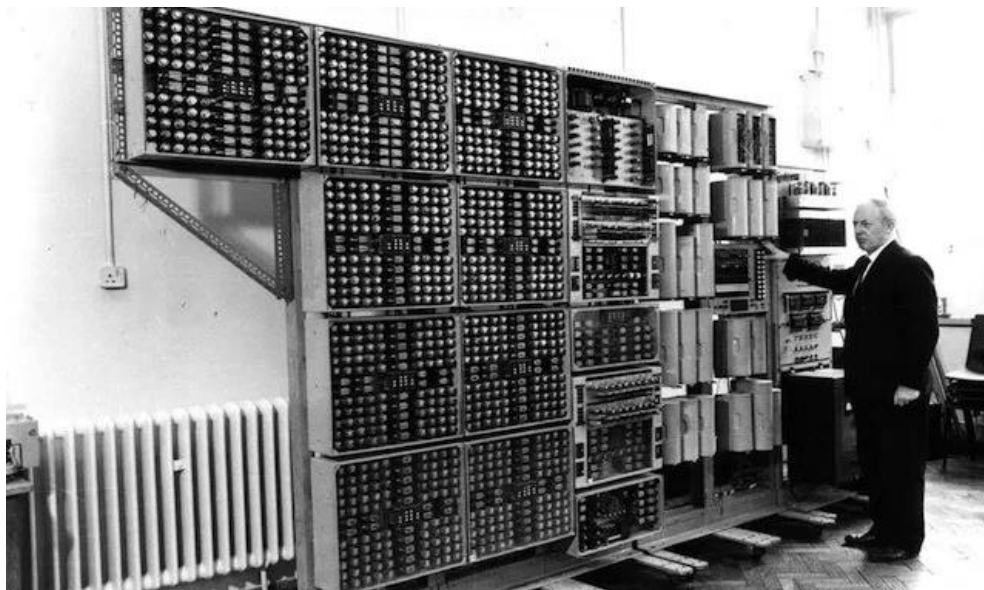


¿Qué es Big Data?

¿Donde se generan todos los tipos de datos?



¿Qué es Big Data?

¿Donde se generan todos los tipos de datos?



Redes sociales y multimedia
(todos generamos datos)



Instrumentos científicos
(colección de toda clase de datos)



Dispositivos móviles
(seguimiento de objetos)



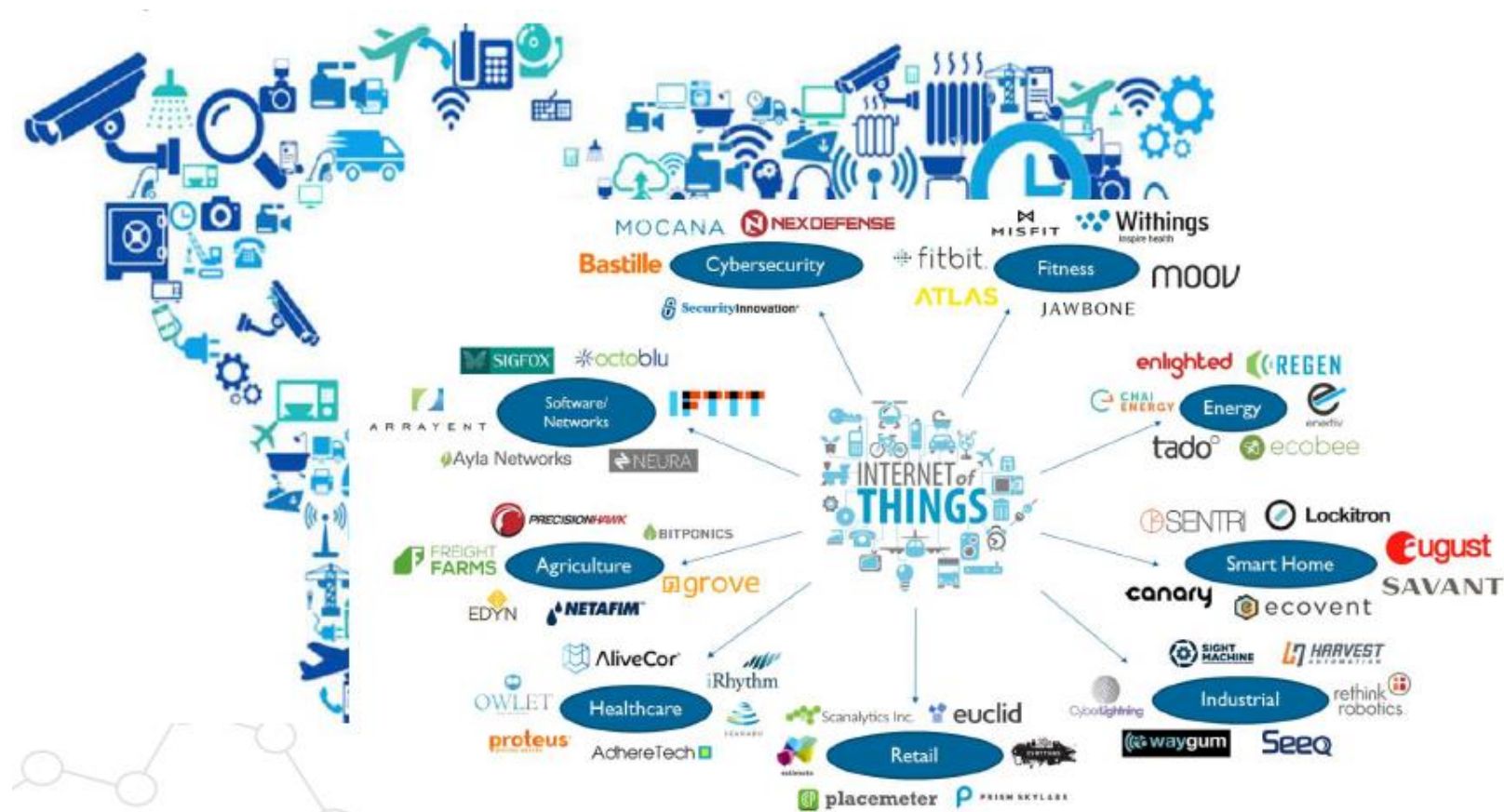
Redes de sensores
(se miden toda clase de datos)

¿Qué es Big Data?

¿Donde se almacenan todos los datos que se generan en nuestra vida diaria?



¿Qué es Big Data?



Big data es la tecnología clave para procesar grandes cantidades de datos

¿Qué es un Data Science?

Datos de distintas fuentes



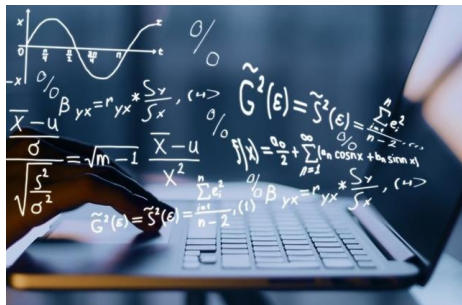
La ciencia que nos permite sacar conclusiones a partir de ello es:



Dominio de Matemática, Estadística, Sistemas, Informática, Marketing, optimización, visualización, entre otros.

¿Por qué la necesidad de un Data Science?

Matemática y Estadística



Conocimiento y dominio de habilidades blandas



Programación y base de datos



Comunicación y visualización

¿Ejemplos de predicción?

facebook

Perfiles de tus amigos

Linked in

Habilidades laborales

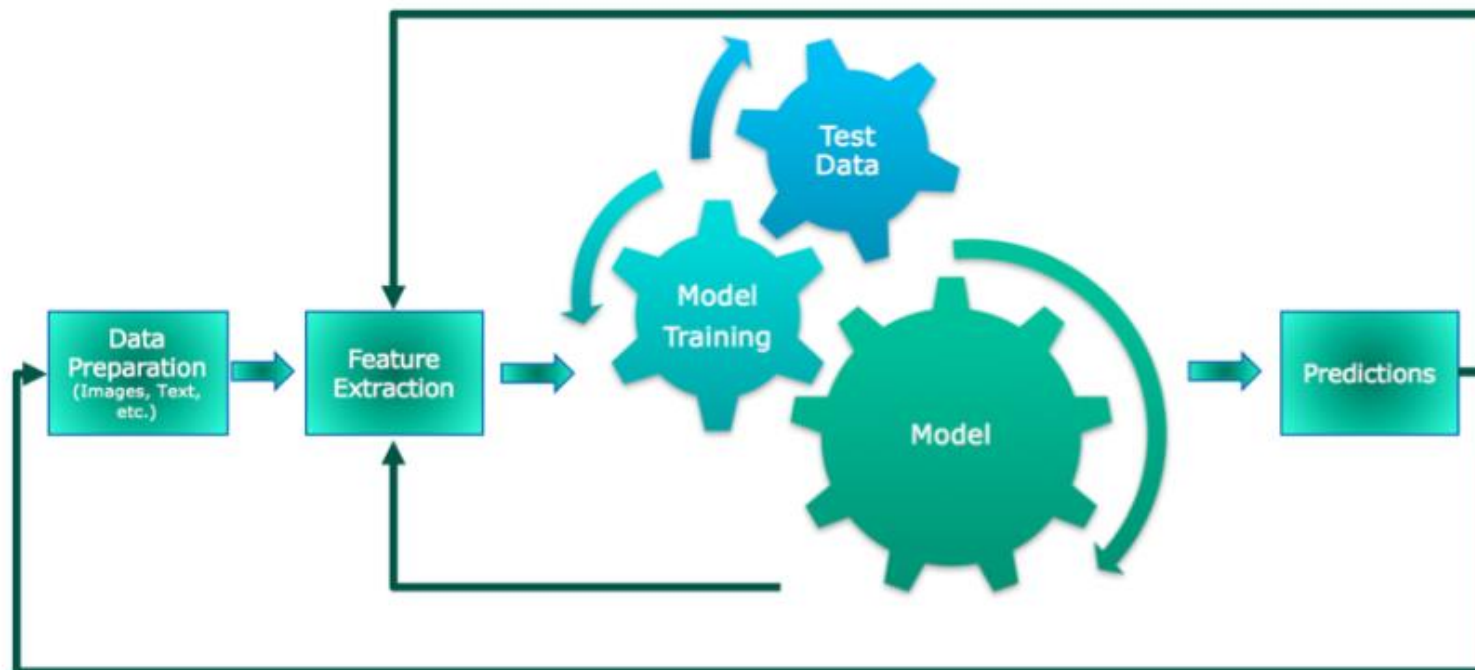
NETFLIX

Las películas que te gustan



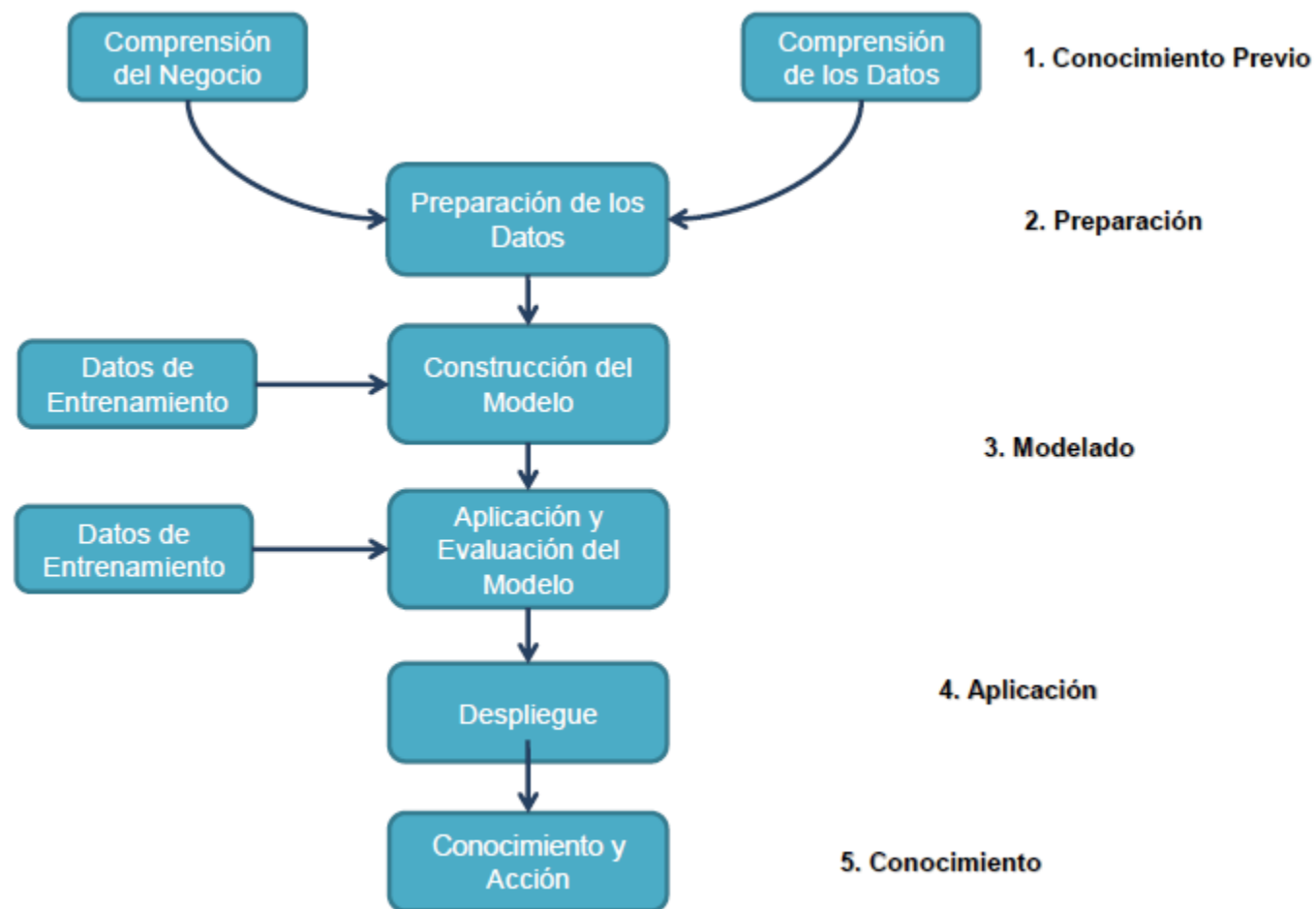
Reconocimiento facial

Introducción al Machine Learning



Machine learning es la técnica que permite entrenar a las computadoras para que aprendan a descubrir por si solos los patrones que se encuentran dentro de los datos

Fases de un problema de Machine Learning



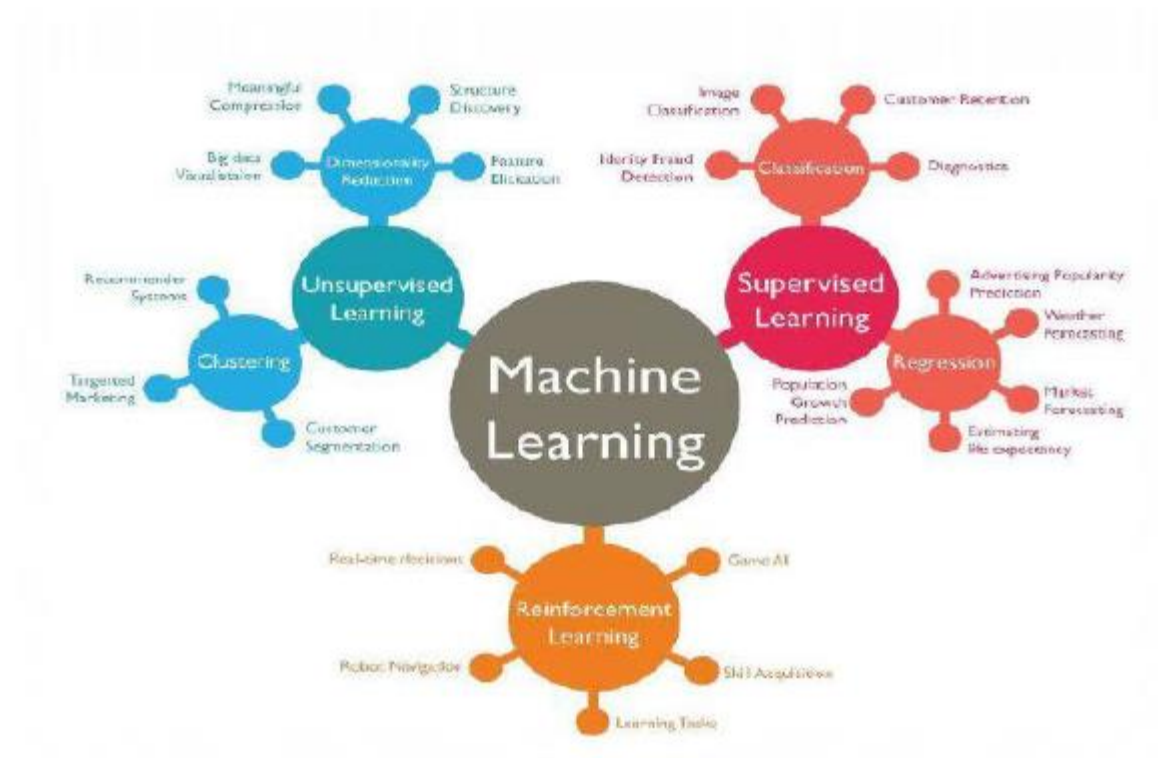
Definiciones básicas

- **Conjunto de Datos (Data Set):** El total del conjunto de datos sobre los que queremos desarrollar un algoritmo de Machine Learning con el fin de obtener un modelo que lo represente lo mejor posible. Contendrá variables independientes y dependientes.
- **Variables Independientes (Features), (VI):** Aquellas columnas del Data Set que serán usadas por el algoritmo para generar un modelo que prediga lo mejor posible las variables dependientes.

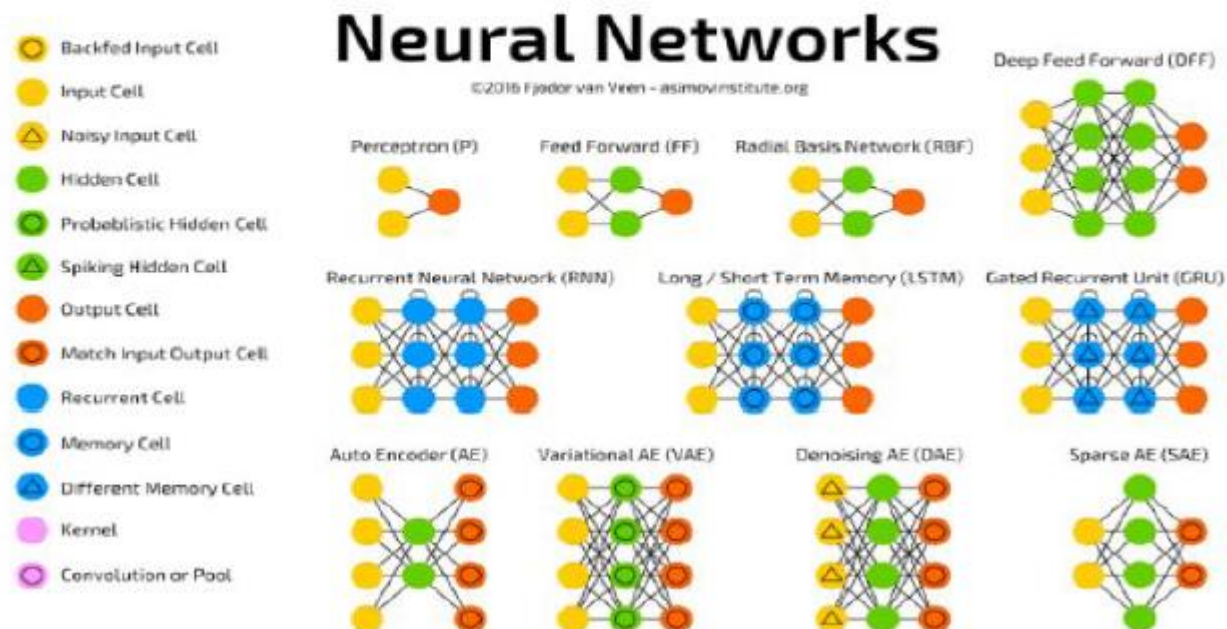
Definiciones básicas

- **Variables dependientes (Labels, Target), (VD):** Columna del data set que responde a una correlación de VI y que debe ser predicha por el futuro modelo
- **Conjunto de Datos de Entrenamiento (Training Set):** Subconjunto del Data Set que será utilizado para entrenar el modelo que se pretende generar.
- **Conjunto de Datos de Test (Test Set):** Subconjunto del data set que se le pasará al modelo una vez haya sido entrenado para comprobar, mediante el uso de diferentes métricas, sus indicadores más importantes de calidad.

Ramas de Machine Learning



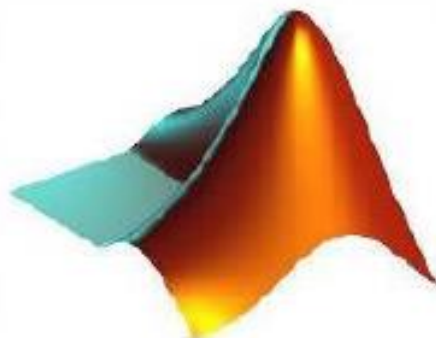
DEEP LEARNING (APRENDIZAJE PROFUNDO)



Panorama tecnológico: Software Machine Learning



SPSS Modeler



Estudio de caso

Caso aplicativo Machine Learning: West Airlines

West Airlines es una empresa de transporte aéreo que recientemente realizó una alianza con una empresa de telecomunicaciones para la venta de servicios de comunicación. El archivo EastWestAirlenesNN.xlsx contiene la información de clientes a los que se le ofreció el servicio y los que respondieron de forma favorable.

Se pide:

Elaborar un modelo predictivo que cubra las necesidades de la aerolínea, para potenciar la marca y realizar una venta más efectiva del servicio.