



# Machine Learning



# Big Data

---

# • INTRODUCCIÓN

# Principales divisiones

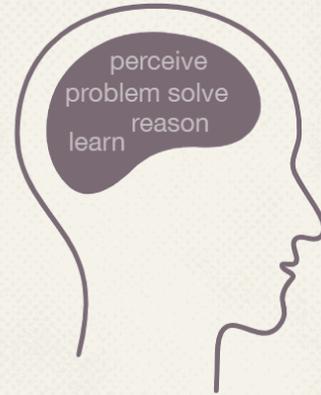


# ¿Por qué ahora?

- Convergencia
  - Algoritmos avanzados
  - Explosión de Información
  - Crecimiento exponencial en el poder computacional

# Inteligencia Artificial

- Típicamente definido como la habilidad de una maquina para desarrollar funciones cognitivas que asociamos con la mente humana.
  - Percibir
  - Razonar
  - Aprender
  - Interactuar con el ambiente
  - Solución de Problemas



# Machine Learning

- Los avances más recientes en el ámbito de la Inteligencia Artificial han sido obtenidos gracias a la aplicación de Machine Learning a grandes repositorios de información.
- Los algoritmos de Machine Learning detectan patrones y aprenden de como hacer predicciones y recomendaciones procesando información y experiencia en lugar de recibir instrucciones precisas de programación.
- Los algoritmos se adaptan en respuesta a nueva data con el objetivo de incrementar su eficiencia.

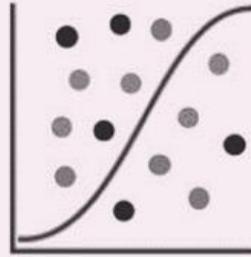
# Tipos de análisis

## Descriptivo



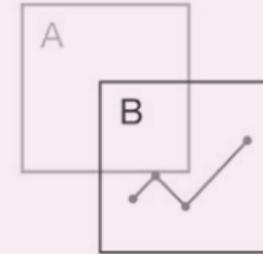
- Describe que paso
- Empleado ampliamente en muchas industrias

## Predictivo



- Anticipa lo que sucederá (probabilísticamente)
- Empleado por empresas que usan datos como fuente para generar insight

## Prescriptivo

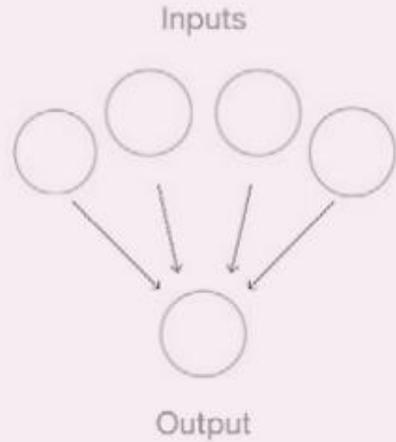


- Provee recomendaciones en que hacer para alcanzar metas
- Empleado ampliamente por empresas lideres en data e internet

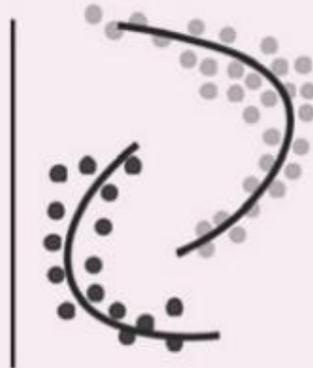
Focus of machine learning

# Tipos de Machine Learning

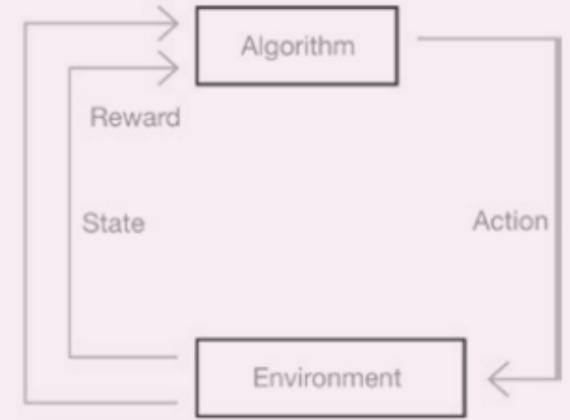
## Supervised learning



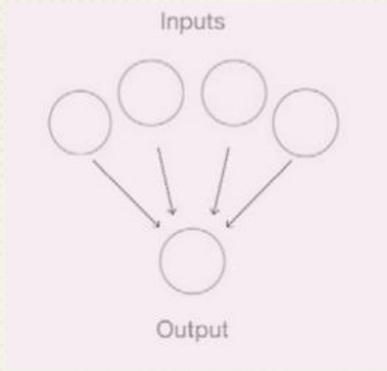
## Unsupervised learning



## Reinforcement learning

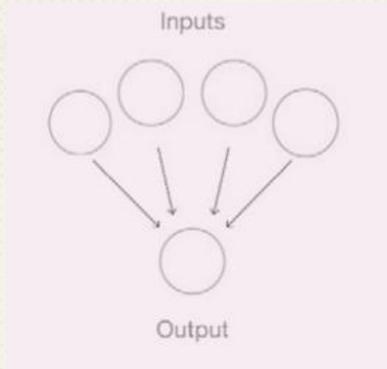


# Método Supervisado



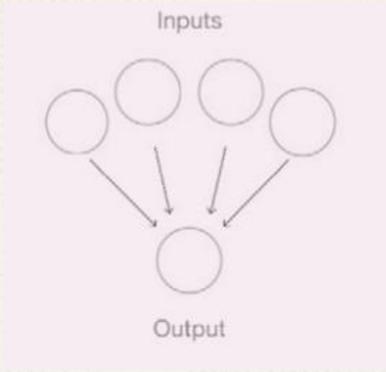
- Qué es
  - Un algoritmo que utiliza información de entrenamiento y retroalimentación del ser humano para aprender la relación entre determinadas variables independientes y la variable dependiente
  - Por ejemplo como las variables año y tasa de interés predicen el precio de las casas

# Método Supervisado



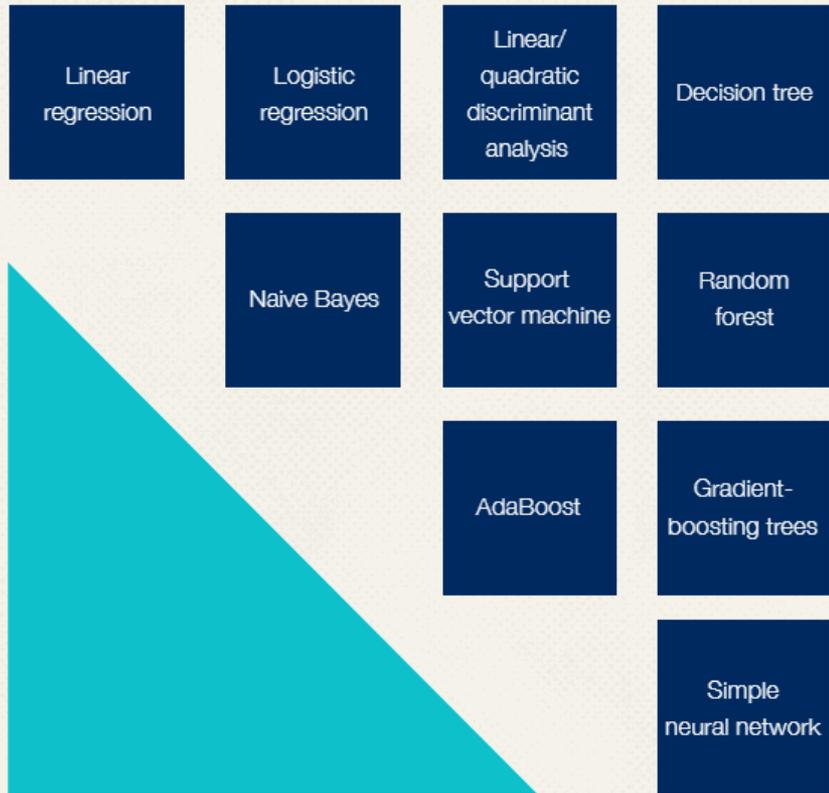
- Cuándo se utiliza
  - Cuando se tiene conocimiento de como clasificar las variables independientes y el tipo de comportamiento que se quiere predecir
  - Al final el algoritmo es el que calcula dicho comportamiento en el nuevo set de datos

# Método Supervisado



- Cómo trabaja
  - Un ser humano nombra cada variable independiente, por ejemplo: año, tasa de interés y define la variable dependiente que puede ser el precio de las casas
  - El algoritmo es entrenado en el set de datos con la respuesta incluida para encontrar una relación entre las variable dependiente y las independientes
  - Una vez finalizado el entrenamiento y se considera que el algoritmo es lo suficientemente certero, es aplicado en un nuevo set de datos sin la variable dependiente para que este la prediga.

# Ejemplos Supervisados



# Método Supervisado – Regresión Lineal

## Algorithms

### **Linear regression**

Highly interpretable, standard method for modeling the past relationship between independent input variables and dependent output variables (which can have an infinite number of values) to help predict future values of the output variables

## Sample business use cases

- Understand product-sales drivers such as competition prices, distribution, advertisement, etc
- Optimize price points and estimate product-price elasticities

# Método Supervisado – Regresión Logística

## Algorithms

### **Logistic regression**

Extension of linear regression that's used for classification tasks, meaning the output variable is binary (eg, only black or white) rather than continuous (eg, an infinite list of potential colors)

## Sample business use cases

- Classify customers based on how likely they are to repay a loan
- Predict if a skin lesion is benign or malignant based on its characteristics (size, shape, color, etc)

# Método Supervisado – Árbol de Decisión

## Algorithms

### **Decision tree**

Highly interpretable classification or regression model that splits data-feature values into branches at decision nodes (eg, if a feature is a color, each possible color becomes a new branch) until a final decision output is made

## Sample business use cases

- Provide a decision framework for hiring new employees
- Understand product attributes that make a product most likely to be purchased

# Método Supervisado – Bosques Aleatorios

## Algorithms

### **Random forest**

Classification or regression model that improves the accuracy of a simple decision tree by generating multiple decision trees and taking a majority vote of them to predict the output, which is a continuous variable (eg, age) for a regression problem and a discrete variable (eg, either black, white, or red) for classification

## Sample business use cases

- Predict call volume in call centers for staffing decisions
- Predict power usage in an electrical-distribution grid

# Método Supervisado – Redes Neuronales

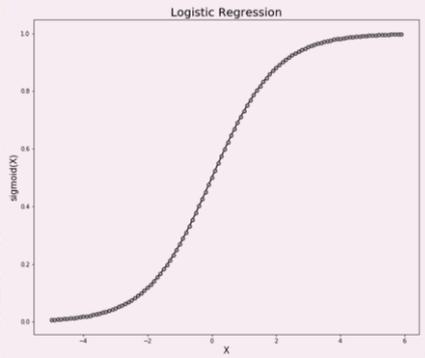
## Algorithms

### **Simple neural network**

Model in which artificial neurons (software-based calculators) make up three layers (an input layer, a hidden layer where calculations take place, and an output layer) that can be used to classify data or find the relationship between variables in regression problems

## Sample business use cases

- Predict the probability that a patient joins a healthcare program
- Predict whether registered users will be willing or not to pay a particular price for a product



Regresión logística

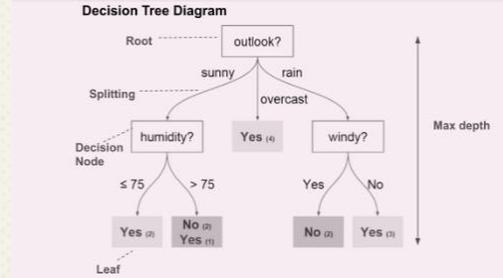
Probabilidad de que alguien pague un crédito

Predice si una lesión en la piel es maligna o no

Árbol de decisión

Rango de confianza para contratar nuevos empleados

Entender los atributos que vuelven a un producto mas deseable en al compra

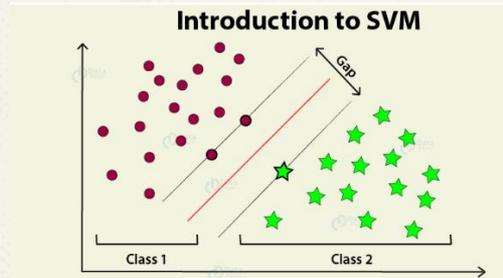
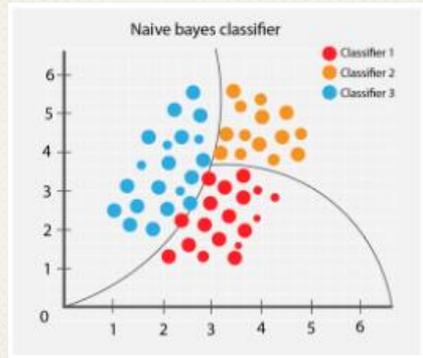


Naive bayes

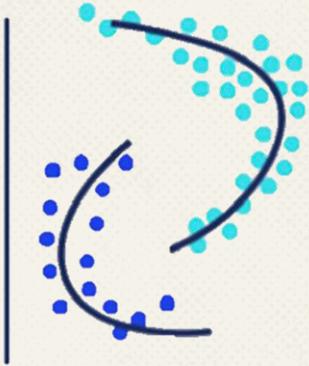
Crear filtros para poder filtrar correo spam

SVM

Probabilidad de que un cliente de clic en una publicidad

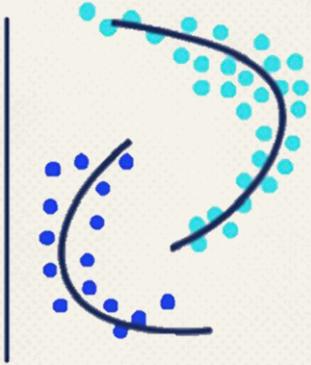


# Método No Supervisado

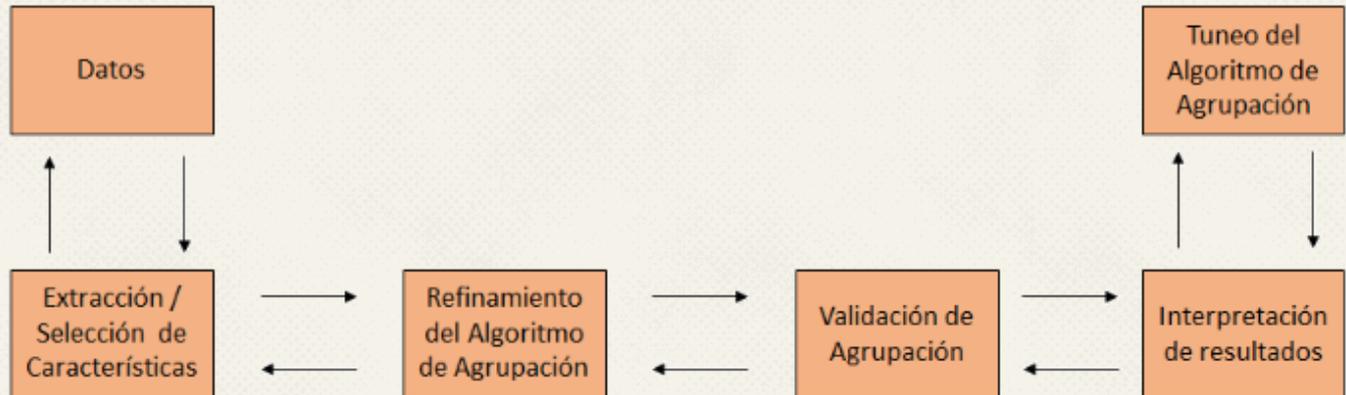


- Qué es
  - Un algoritmo explora un set de información sin otorgarle una variable objetivo definida
  - El objetivo es estudiar la estructura de una base de datos que se encuentra comúnmente oculta
  - Por ejemplo: explorar la información demográfica de los clientes y con esto identificar patrones

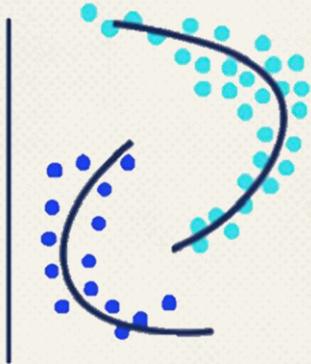
# Método No Supervisado



- Cuándo se utiliza
  - Cuando no se sabe como se debe clasificar la información y se desea que el algoritmo encuentre patrones y realice clasificaciones por nosotros



# Método No Supervisado



- Cómo trabaja
  - Un algoritmo recibe información que no esta necesariamente nombrada
  - Por ejemplo un set de información demográfico de los clientes
  - El algoritmo infiere una estructura en base a la información
  - Luego identifica grupos de información que muestran el mismo comportamiento

# Ejemplos No Supervisados



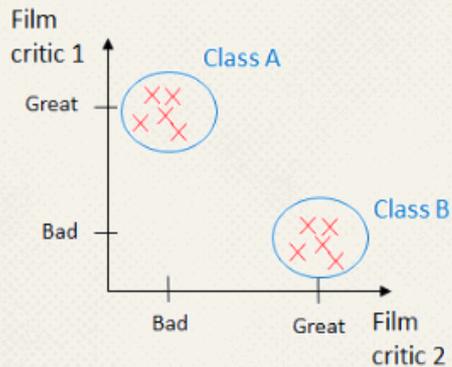
K-means  
clustering

Gaussian  
mixture model

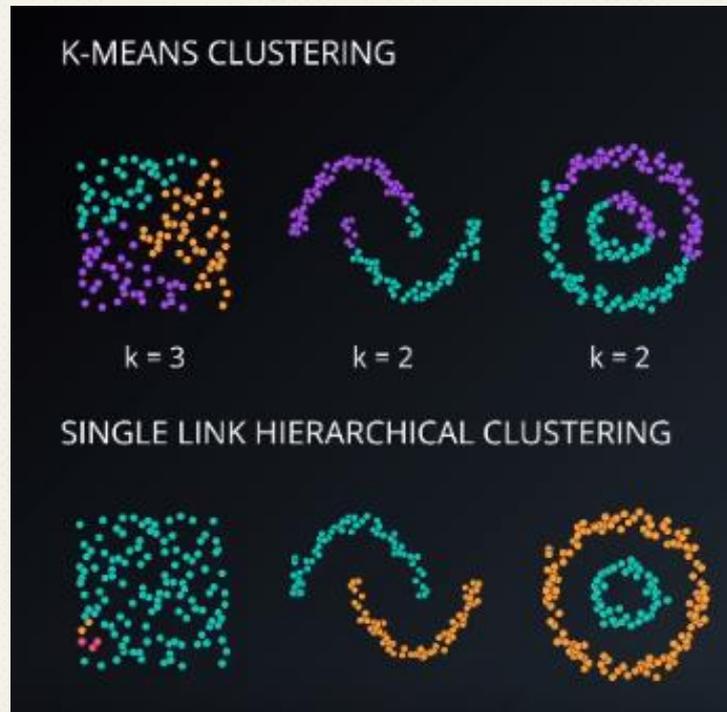
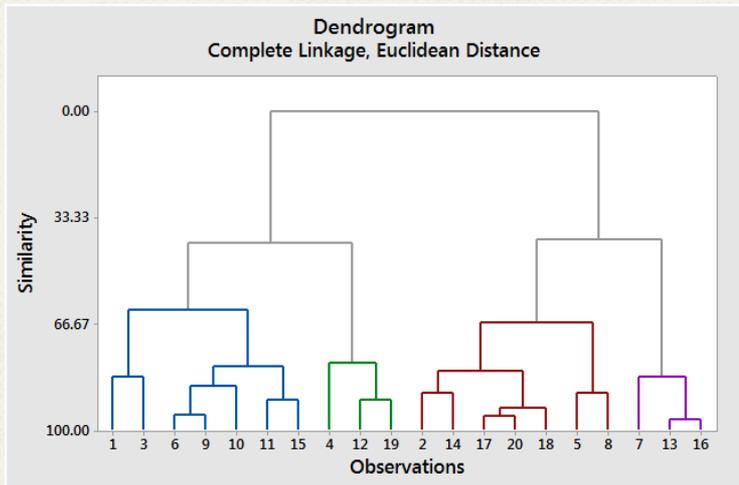
Hierarchical  
clustering

Recommender  
system

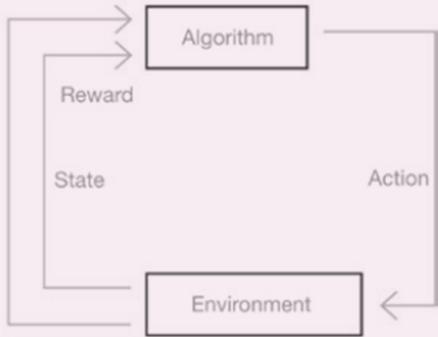
# K means



# Jerarquica

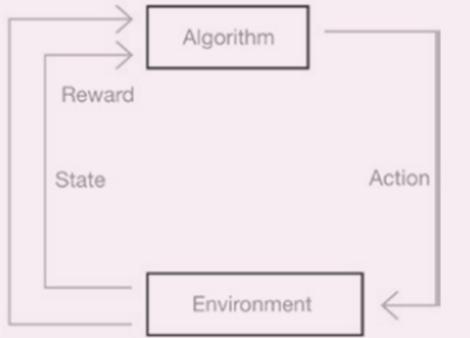


# Método Reforzado



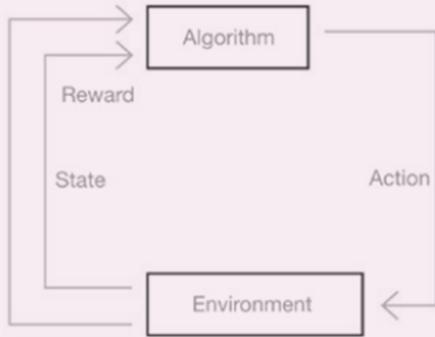
- Qué es
  - Un algoritmo aprende a realizar tareas simplemente con el objetivo de maximizar las recompensas que obtiene por sus iteraciones
  - Por ejemplo: Maximizar ganancias que recibe por inversiones de un determinado portafolio en la bolsa.

# Método Reforzado



- Cuándo se utiliza
  - Cuando no se tiene mucha información para entrenar un modelo y no se puede identificar claramente cuanta es la necesaria
  - O cuando la única manera de aprender de un determinado problema es interactuando con el.

# Método Reforzado



- Cómo trabaja
  - Un algoritmo toma acciones específicas en el ambiente que es desplegado
  - Por ejemplo: realizar transacciones en un portafolio de inversión
  - Recibe recompensas por lograr incrementos en objetivos, ejemplo mayor rentabilidad de la inversión
  - El algoritmo optimiza la mejor serie de acciones ya que se va corrigiendo a si mismo a medida pasa el tiempo

# Ejemplos Reforzado



Optimize the trading strategy for an options-trading portfolio



Balance the load of electricity grids in varying demand cycles



Stock and pick inventory using robots



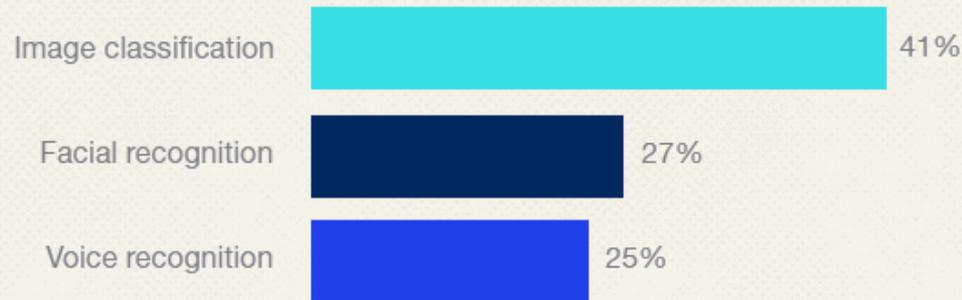
Optimize the driving behavior of self-driving cars



Optimize pricing in real time for an online auction of a product with limited supply

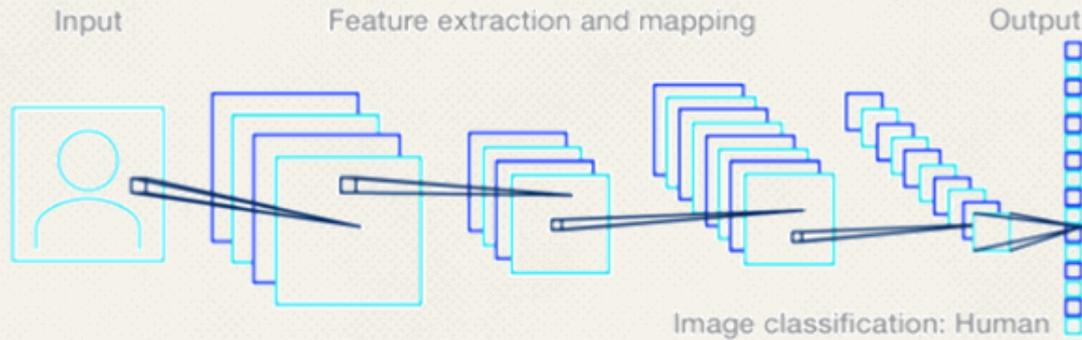
# Deep Learning

- Es un tipo de Machine Learning el cual puede procesar mayores cantidades de información, necesita menos intervención humana y usualmente produce resultados más certeros que modelos de ML tradicionales



# Tipos de Deep Learning

## Convolutional neural network



## Recurrent neural network

