**Machine Learning. [[1]](#footnote-1)**

Data contains a lot of wealth information and traditional data analysis practices or programs could not deal with it.

* Machine learning uses algorithms (also called “models”) **to identify patterns in data.** 
  + The process by which this model learns these patterns is called “model training”.
  + Once a model is trained, it could be used to make predictions
    - The next time new data appears, the model will try to make predictions based on the patterns of the past data.
* Machine learning models can be divided into three basic classes:
  + **Supervised learning:** 
    - One of the most commonly used type of machine learning.
    - Human provides input data and the correct outputs. The model tries to identify relationships and dependencies between the input data and the correct output.
    - **It is used to solve regression and classification problems.** 
      * Regression models:
        + Used to predict a numeric or real value. (for instance, the price of a house)
      * Classification models:
        + Used to predict whether something belongs to a category or class.
  + **Unsupervised learning**
    - Here the data is not labelled by a human
    - The models must analyze the data and try to identify patterns and structure within the data **based only on the characteristics of the data itself.** (The researcher does not have the outputs or he/she does not introduce they in order to use the logic of the machine)
    - **Clustering and anomaly detection are two examples of this learning style.**
      * Clustering models.
        + Used to divide each record of a data set into one of a small number of similar groups. (For instance, customers that have the same interests)
      * Anomaly detection
        + Identifies outliers in a data set
  + **Reinforcement learning.** 
    - It is loosely based on the way human beings and other organisms learn.
    - Takes the best set of actions given it is current environment in order to get the most reward over time

**Deep learning**

It is an especial type of machine learning.

It refers to a general set of models and techniques that tries to loosely emulate the way the human brain solves a wide range of problems.

It is commonly used to analyze natural language, both spoken and text, as well as images, audio, etc.

Requires very large data sets of labelled data to train a model. (it is very expensive)

1. https://www.youtube.com/watch?v=R8rtbuOi72Q [↑](#footnote-ref-1)