*Data science encompasses the entire process of obtaining knowledge from data by integrating methods from statistics, computer science, and other fields to gain insight from data. In practice, data science encompasses an iterative process of data harvesting, cleaning, analysis and visualization, and deployment.*

"Data scientist is a person who is better at statistics than any software engineer and better at software engineering than any statistician".

*Machine learning, on the other hand, is mainly concerned with fairly generic algorithms and techniques that are used in analysis and modeling phases of data science process.*

"Machine Learning relates with the study, design and development of the algorithms that give computers the capability to learn without being explicitly programmed."

***What kind of problems can machine learning solve?***

*Among the different machine learning approaches, there are three main ways of learning, as shown in the following list:*

## *Supervised Learning*

*In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output.*

*Supervised learning problems are categorized into "regression" and "classification" problems. In a regression problem, we are trying to predict results within a continuous output, meaning that we are trying to map input variables to some continuous function. In a classification problem, we are instead trying to predict results in a discrete output. In other words, we are trying to map input variables into discrete categories.*

* *Given a set of example inputs, X, and their outcomes, Y, supervised learning aims to learn a general mapping function, f, that transforms inputs to outputs, as f: X Y*
* *An example of supervised learning is credit card fraud detection, where the learning algorithm is presented with credit card transactions (matrix X) marked as normal or suspicious. (vector Y). The learning algorithm produces a decision model that marks unseen transactions as normal or suspicious (that is the f function).*

*Example:*

*(a) Regression - Given a picture of a person, we have to predict their age on the basis of the given picture*

*(b) Classification - Given a patient with a tumor, we have to predict whether the tumor is malignant or benign.*

## *Unsupervised Learning*

*Unsupervised learning allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables.*

*We can derive this structure by clustering the data based on relationships among the variables in the data.*

*With unsupervised learning there is no feedback based on the prediction results.*

***Example:***

*Clustering: Take a collection of 1,000,000 different genes, and find a way to automatically group these genes into groups that are somehow similar or related by different variables, such as lifespan, location, roles, and so on.*

*Non-clustering: The "Cocktail Party Algorithm", allows you to find structure in a chaotic environment. (i.e. identifying individual voices and music from a mesh of sounds at a*[*cocktail party*](https://en.wikipedia.org/wiki/Cocktail_party_effect)*).*

* *In contrast, unsupervised learning algorithms do not assume given outcome labels, Y as they focus on learning the structure of the data, such as grouping similar inputs into clusters. Unsupervised learning can, hence, discover hidden patterns in the data. An example of unsupervised learning is an item-based recommendation system, where the learning algorithm discovers similar items bought together, for example, people who bought book A also bought book B.*

# *Data and problem definition*

*Data is simply a collection of measurements in the form of numbers, words, measurements, observations, descriptions of things, images, and so on.*

*The most common way to represent the data is using a set of attribute-value pairs.*

*Bob = {*

*height: 185cm,*

*eye color: blue,*

*hobbies: climbing, sky diving*

*}*