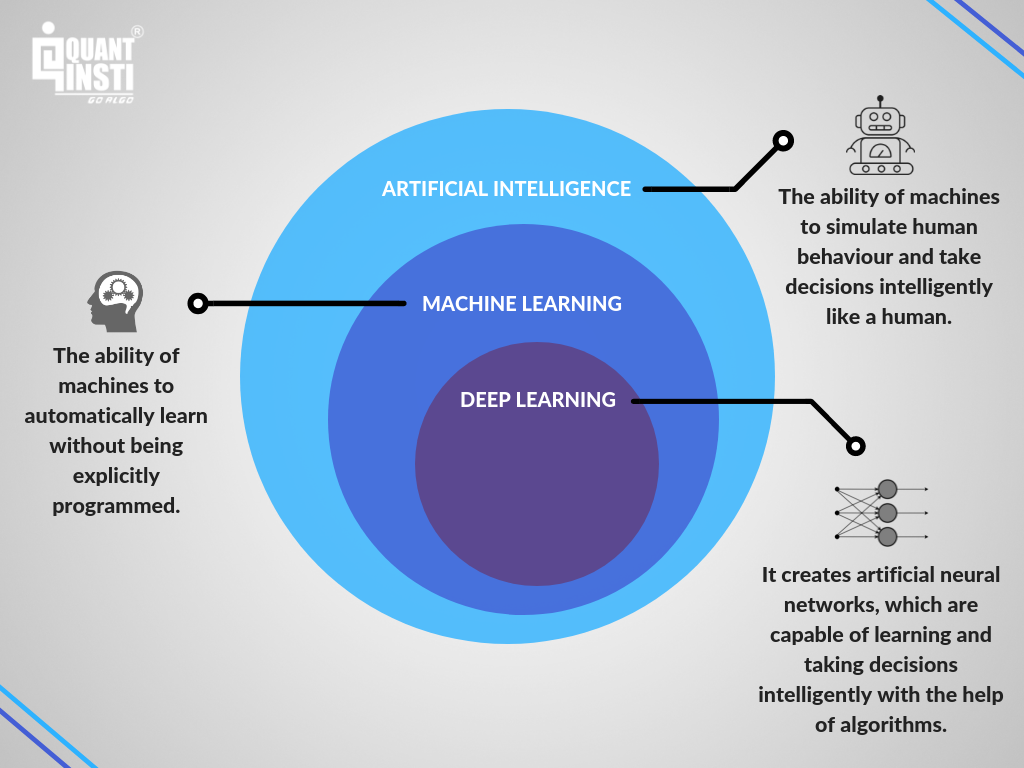
**Machine Learning**

**What is Machine Learning?**

Machine Learning, as the name suggests, provides machines with the ability to learn autonomously based on experiences, observations and analysing patterns within a given data set without explicitly programming. When we write a program or a code for some specific purpose, we are actually writing a definite set of instructions which the machine will follow. Whereas in machine learning, we input a data set through which the machine will learn by identifying and analysing the patterns in the data set and learn to take decisions autonomously based on its observations and learnings from the dataset.

### ****The difference between Machine Learning, Artificial Intelligence and Deep Learning****

While learning about machine learning basics, one often confuses Machine Learning, Artificial Intelligence and Deep Learning. the below diagram clears the concept of machine learning.



**Classification of Machine Learning Algorithms**

Machine Learning algorithms can be classified into:

1. Supervised Algorithms – Linear Regression, Logistic Regression, Support Vector Machine (SVM), Decision Trees, Random Forest
2. Unsupervised Algorithms – K Means Clustering.
3. Reinforcement Algorithm

#### ****1. Supervised Machine Learning Algorithms****

In this type of algorithm, the data set on which the machine is trained consists of labelled data or simply said, consists both the input parameters as well as the required output. For example, classifying whether a person is a male or a female. Here male and female will be our labels and our training dataset will already be classified into the given labels based on certain parameters through which the machine will learn these features and patterns and classify some new input data based on the learning from this training data.

Let us look at some of the examples of Supervised Learning Algorithms:

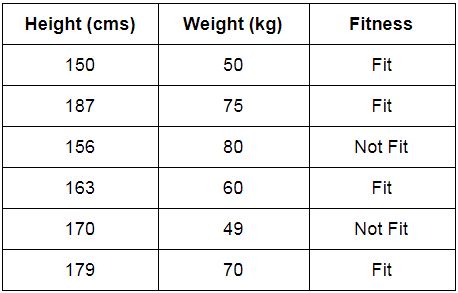
Supervised Learning Algorithms can be broadly divided into two types of algorithms, Classification and Regression.

##### **Classification Algorithms**

Just as the name suggests, these algorithms are used to classify data into predefined classes or labels. We will discuss one of the most used classification algorithm known as the K-Nearest Neighbor (KNN) Classification Algorithm.

##### **KNN Classification Algorithm**

This algorithm is used to classify a set of data points into specific groups or classes based on the similarities between the data points.  
Let’s consider an example where we need to check whether a person is fit or not based on the height and weight of a person.  
Suppose we give the following table as the training data set:



Now consider a new person needs to be classified as fit/not fit. Let us consider the value of K=3, which means will consider 3 nearest neighbours. The nearest neighbours can be found out by determining the Euclidean difference between the height and weight of one person and the height and weight of the persons given in the table. The persons with the 3 least differences will be considered as the nearest neighbours. Now we will check how many out of these 3 are fit. If 2 or more out of the 3 are fit, then we will classify the new person as fit and vice versa. In case, we get an equal number of neighbours with different outcomes, then we can increase the value of K and check again.

##### **Regression Algorithms**

These algorithms are used to determine the mathematical relationship between two or more variables and the level of dependency between variables. These can be used for predicting an output based on the interdependency of two or more variables. For example, an increase in the price of a product will decrease its consumption, which means, in this case, the amount of consumption will depend on the price of the product. Here, the amount of consumption will be called as the dependent variable and price of the product will be called the independent variable. The level of dependency of the amount of consumption on the price of a product will help us predict the future value of the amount of consumption based on the change in prices of the product.

We have two types of regression algorithms: Linear Regression and Logistic Regression

**(a) Linear Regression**

Linear regression is used with continuously valued variables, like the previous example in which the price of the product and amount of consumption are continuous variables, which means that they can have an infinite number of possible values. Linear regression can also be represented as a graph known as scatter plot, where all the data points of the dependent and independent variables are plotted and a straight line is drawn through them such that the maximum number of points will lie on the line or at a smaller distance from the line. This line – also called the regression line, will then help us determine the relationship between the dependent and independent variables along with which the linear regression equation is formed.

You can learn about Linear Regression and how it can be used to predict the stock prices in detail in [this blog](https://www.quantinsti.com/blog/machine-learning-trading-predict-stock-prices-regression/).

**(b) Logistic Regression**

The difference between linear and logistic regression is that logistic regression is used with categorical dependent variables (eg: Yes/No, Male/Female, Sunny/Rainy/Cloudy, Red/Blue etc.), unlike the continuous valued variables used in linear regression. Logistic regression helps determine the probability of a certain variable to be in a certain group like whether it is night or day, or whether the colour is red or blue etc. The graph of logistic regression consists of a non-linear sigmoid function which demonstrates the probabilities of the variables.

If you wish to implement Logistic Regression in Python, then you can refer this article which includes the codes to do so.  
Apart from these, there are more Supervised algorithms like Support Vector Machine (SVM), Decision Trees, Random Forests.  
You can also learn how to use SVM and Decision Trees for Trading these articles, [Trading Using Machine Learning In Python](https://www.quantinsti.com/blog/trading-using-machine-learning-python-svm-support-vector-machine/) – SVM (Support Vector Machine) and [Decision Tree Classifier For Trading](https://www.quantinsti.com/blog/use-decision-trees-machine-learning-predict-stock-movements/) Part-1.

#### ****2. Unsupervised Machine Learning Algorithms****

Unlike supervised learning algorithms, where we deal with labelled data for training, the training data will be unlabelled for Unsupervised Machine Learning Algorithms. The clustering of data into a specific group will be done on the basis of the similarities between the variables. Some of the unsupervised machine learning algorithms are K-means clustering, neural networks. In this article, we will talk about the k-means clustering algorithm.

##### **K-means clustering**

Before we understand the working of K-means clustering algorithm, let us first break down the word K-means clustering to understand what it means.

Clustering: In this algorithm, we form clusters which are a collection of data points grouped together due to their similarities.

K refers to the number of centroids which will be considered for a specific problem whereas ‘means’ refers to a centroid which is considered as the central point of any cluster.

##### **Working of K-means Clustering Algorithm**

1. Define the value of K. For eg: if K= 2, then we will have two centroids.
2. Randomly select K data points as centroids.
3. Check the distance of each data point with the centroids.
4. Assign the data point to the centroid with which it has a minimum distance, thus forming a cluster of similar data points.
5. Recalculate the centroid of each newly formed cluster and reassign the data points to the cluster whose centroid is at a minimum distance from the data point.

You can decide the number of iterations for repeating step 5 to optimize the algorithm. When the centroid stops changing or remains same after some amount of iterations then that will be our stopping point and the algorithm will be fully optimized.

Another machine learning concept which is extensively used in the field is Neural Networks. You can read about the working of neural networks and how it can be used for stock price prediction in [this article](https://www.quantinsti.com/blog/working-neural-networks-stock-price-prediction/).

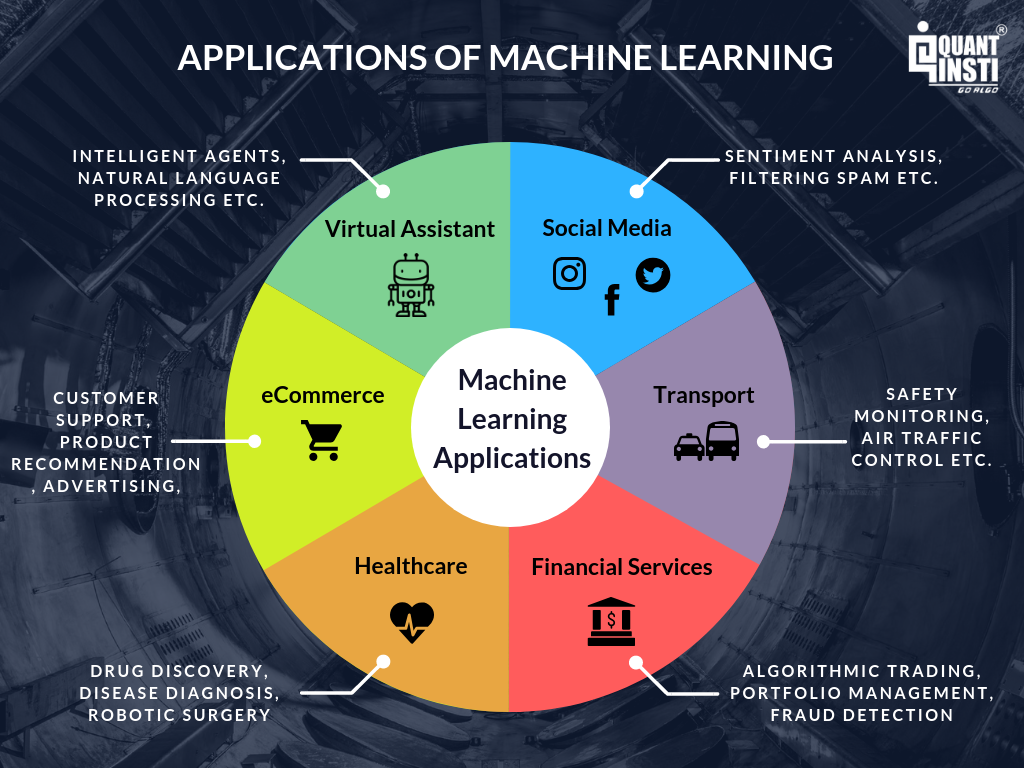
#### ****3. Reinforcement Machine Learning Algorithms****

Reinforcement Learning is a type of Machine Learning in which the machine is required to determine the ideal behaviour within a specific context, in order to maximize its rewards. It works on the rewards and punishment principle which means that for any decision which a machine takes, it will be either be rewarded or punished due to which it will understand whether or not the decision was correct. This is how the machine will learn to take the correct decisions to maximize the reward in the long run.

For reinforcement algorithm, a machine can be adjusted and programmed to focus more on either the long-term rewards or the short-term rewards. When the machine is in a particular state and has to be the action for the next state in order to achieve the reward, this process is called the Markov Decision Process.

### ****Applications of Machine Learning****

We have covered most about machine learning basics that would clear fundamentals of machine learning, the machine learning process, machine learning concepts and examples of machine learning that would be essential to a machine learning beginner.



### ****Growth and Future of Machine Learning****

Machine Learning is growing at a tremendous rate and we will soon be able to see its applications across all of the major domains. Various reports regarding machine learning have all pointed to an upward growth curve for this domain. According to [IFI Claims](https://www.ificlaims.com/rankings-8-fast-growing.htm) Patent Services (Patent Analytics), Machine Learning patents witnessed a growth of 34% CAGR between 2013 and 2017, of which the major patent producers included companies like IBM, Microsoft, Google, LinkedIn etc.

A [survey](https://s3.amazonaws.com/files.technologyreview.com/whitepapers/MITTR_GoogleforWork_Survey.pdf) by MIT and Google Cloud demonstrates that 60% of the organisations have already been using Machine Learning strategies and one-third of them are at an early stage of development.  
This [report](http://www.datascienceassn.org/sites/default/files/Predictive%20Analytics%20and%20Machine%20Learning%20Solutions%20Vendors%202017.pdf) by Forrester predicts huge growth for Machine Learning, which forecasts that the Predictive Analytics and Machine Learning (PAML) market will grow at 21% CAGR through 2021.

Businesses and other major domains are not just adopting new technologies but are adopting new machine learning technologies to automate many of processes which are helping them increase their productivity. We are now entering into the age of Artificial Intelligence and Machine Learning thus making it a domain impossible to ignore and a lot to explore!