Introduction to Machine Learning

Focus *Always be on Focus → Master Slide*

CS-464

Evolutionary Computation

- W4-5: Genetic Algorithm
- W6: Particle Swarm

Fuzzy Logic

- W2: Introduction to Fuzzy Logic
- W3: Fuzzy Logic Control

Machine Learning

- W8: Introduction to ML
- W9: Clustering Technique
- W10: Linear Regression
- W11: Logistic Regression
- W13: Support Vector Machine
- W14: Neural Network

What is Machine Learning?

A good start at a Machine Learning definition is that it is a core sub-area of Artificial Intelligence (AI). ML applications learn from experience (well data) like humans without direct programming. When exposed to new data, these applications learn, grow, change, and develop by themselves. In other words, with Machine Learning, computers find insightful information without being told where to look. Instead, they do this by leveraging algorithms that learn from data in an iterative process.

While the concept of Machine Learning(ML) has been around for a long time, the ability to automate the application of complex mathematical calculations to Big Data has been gaining momentum over the last several years.

At a high level, Machine Learning is the ability to adapt to new data independently and through iterations. Basically, applications learn from previous computations and transactions and use "pattern recognition" to produce reliable and informed results.

How Machine Learning Works?

Machine Learning is, undoubtedly, one of the most exciting subsets of Artificial Intelligence. It completes the task of learning from data with specific inputs to the machine. It's important to understand what makes Machine Learning work and, thus, how it can be used in the future.

The Machine Learning process starts with inputting training data into the selected algorithm. Training data being known or unknown data to develop the final Machine Learning algorithm. The type of training data input does impact the algorithm, and that concept will be covered further momentarily.

To test whether this algorithm works correctly, new input data is fed into the Machine Learning algorithm. The prediction and results are then checked.

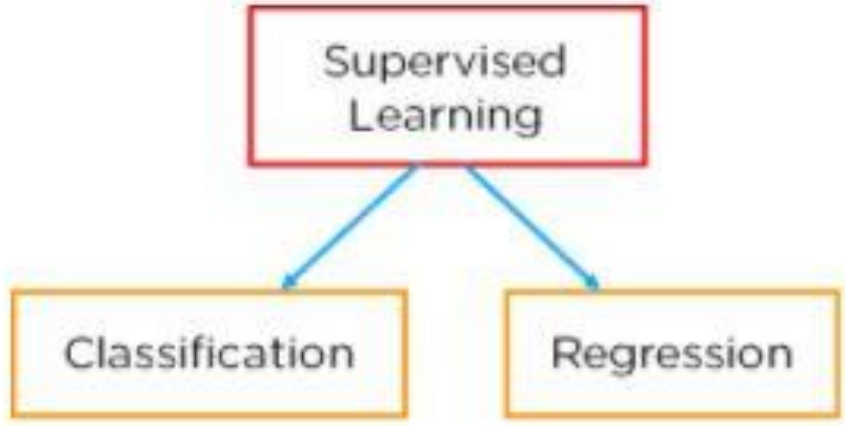
If the prediction is not as expected, the algorithm is re-trained multiple numbers of times until the desired output is found. This enables the Machine Learning algorithm to continually learn on its own and produce the most optimal answer that will gradually increase in accuracy over time.

What are the Different Types of Machine Learning?

Machine Learning is complex in itself, which is why it has been divided into two main areas, supervised learning and unsupervised learning. Each one has a specific purpose and action within Machine Learning, yielding particular results, and utilizing various forms of data. Approximately 70 percent of Machine Learning is supervised learning, while unsupervised learning ranges from 10 – 20 percent. Another method that is used less often is reinforcement learning.

1. Supervised Learning

In supervised learning, we use known or labeled data for the training data. Since the data is known, the learning is, therefore, supervised, i.e., directed into successful execution. The input data goes through the Machine Learning algorithm and is used to train the model. Once the model is trained based on the known data, you can use unknown data into the model and get a new response.



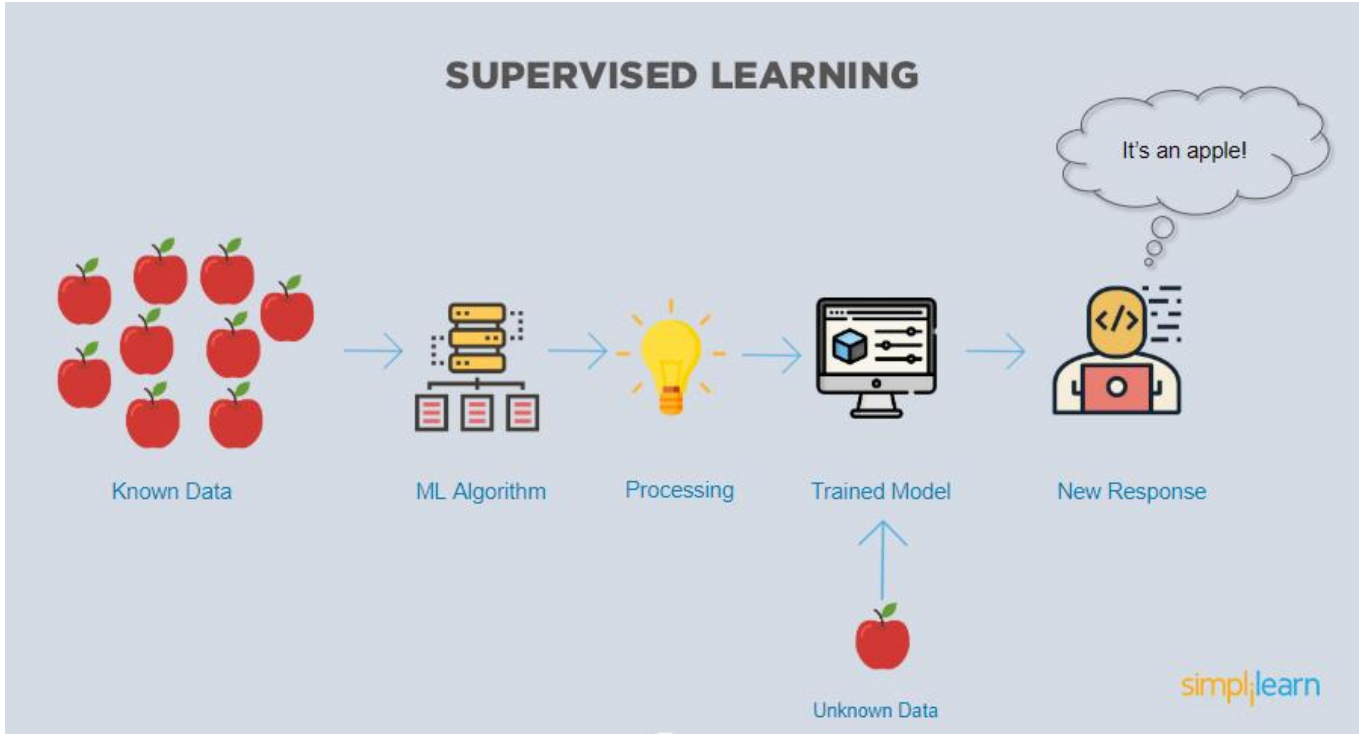
Classification

The first method is classification, and it falls under supervised learning. Classification is used when the output you are looking for is a "yes" or "no," or in the form of "a" or "b" or "true" or "false." For instance, if a shopkeeper wants to predict that a particular customer will come back to his shop or not, he will use a classification algorithm.

Regression

This method is used when the predicted data is numerical. If the shopkeeper wants to predict the price of a product based on its demand, he will choose regression.

Example: In this case, the model tries to figure out whether the data is an apple or another fruit. Once the model has been trained well, it will identify that the data is an apple and give the desired response.



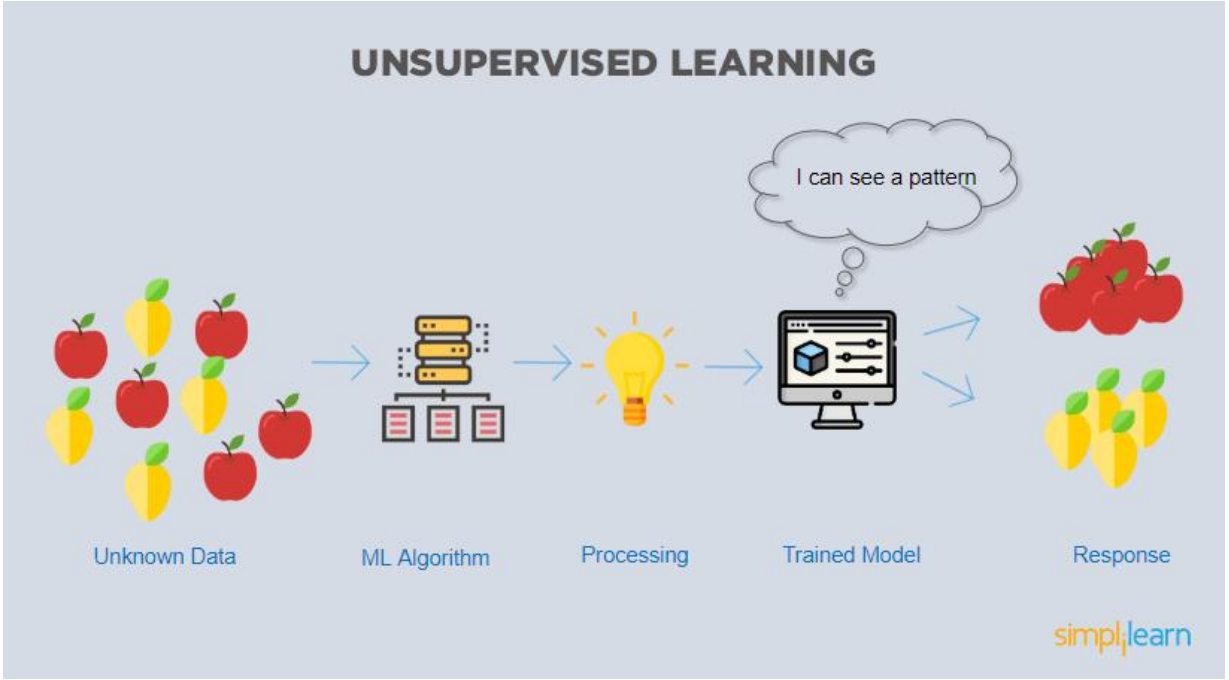
Here is the list of top algorithms currently being used for supervised learning are:

- **Polynomial regression**
- **Random forest**
- **Linear regression**
- **Logistic regression**
- **Decision trees**
- **K-nearest neighbors**
- **Naive Bayes**
- **Neural Network**

2. Unsupervised Learning

In unsupervised learning, the training data is unknown and unlabeled - meaning that no one has looked at the data before. Without the aspect of known data, the input cannot be guided to the algorithm, which is where the unsupervised term originates from. This data is fed to the Machine Learning algorithm and is used to train the model. The trained model tries to search for a pattern and give the desired response. In this case, it is often like the algorithm is trying to break code like the Enigma machine but without the human mind directly involved but rather a machine.

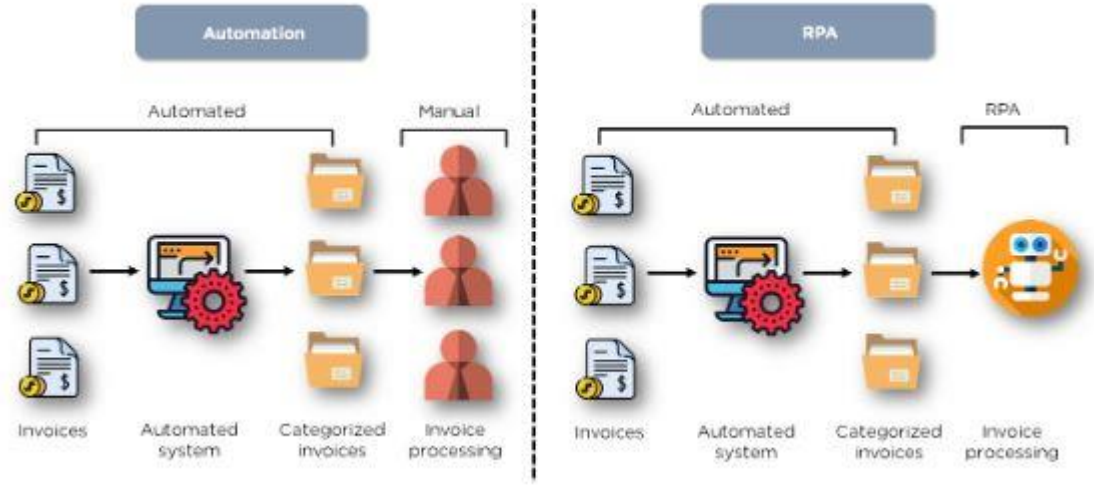
Example: In this case, the unknown data consists of apples and pears which look similar to each other. The trained model tries to put them all together so that you get the same things in similar groups.



The top 7 algorithms currently being used for unsupervised learning are:

- Partial least squares
- Fuzzy means
- Singular value decomposition
- K-means clustering
- Apriori
- Hierarchical clustering
- Principal component analysis
- Self Organizing Map (SOM)

Comparison Between Supervised and Unsupervised Learning



Supervised and unsupervised learning differ in several ways:

- First, the data used in supervised learning is labeled. In the examples shown above, you provide the system with a photo of an apple and let the system know that this is an apple. That is called labeled data. The system learns from the labeled data and makes future predictions. On the other hand, unsupervised learning does not require any labeled data because its job is to look for patterns in the input data and organize it.

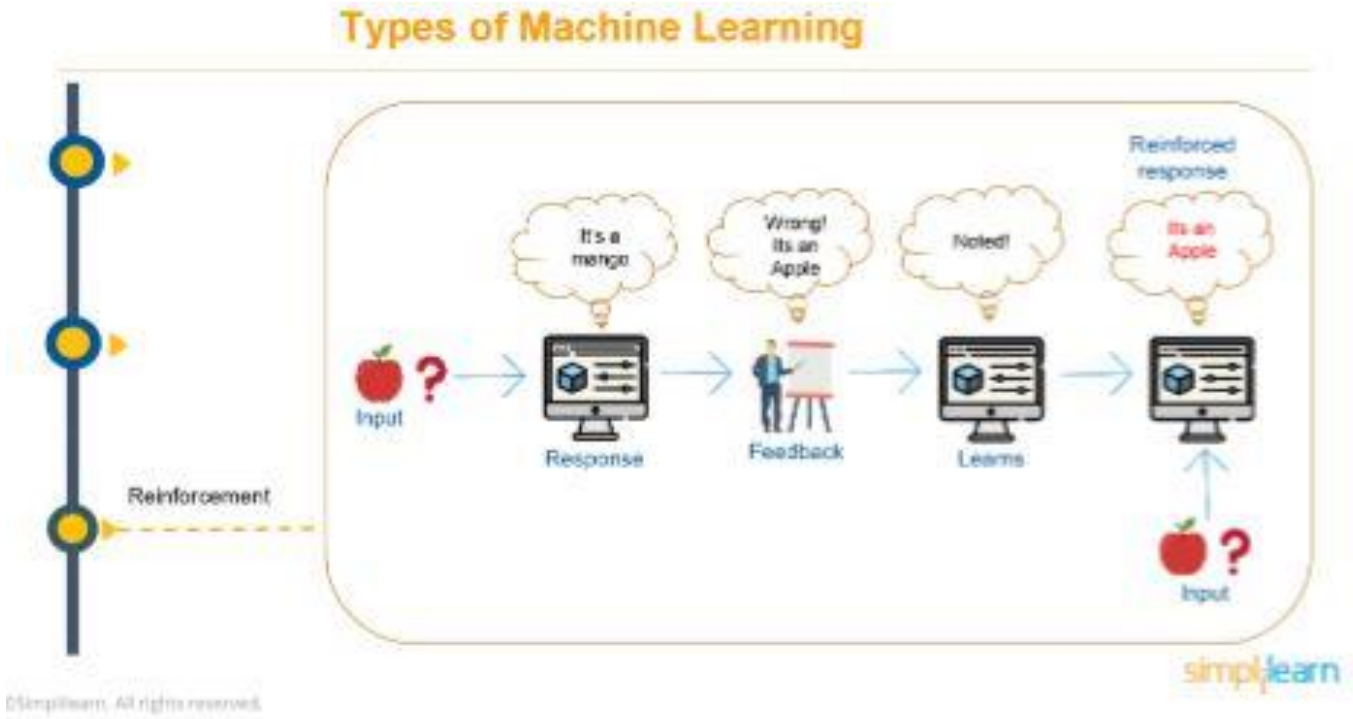
- Second, you get feedback in the case of supervised learning. That is, once you receive the output, the system remembers it and uses it for the next operation. That does not happen with unsupervised learning.

- Lastly, supervised learning is mostly used to predict data, whereas unsupervised learning is used to find hidden patterns or structures in data.

3. Reinforcement Learning

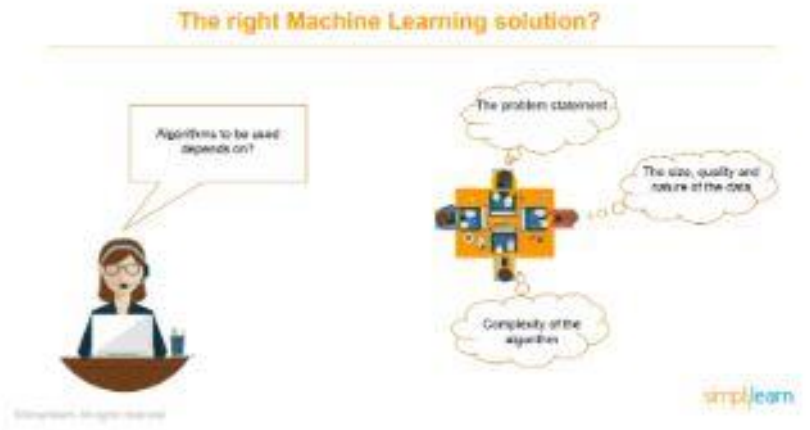
Like traditional types of data analysis, here, the algorithm discovers data through a process of trial and error and then decides what action results in higher rewards. Three major components make up reinforcement learning: the agent, the environment, and the actions. The agent is the learner or decision-maker, the environment includes everything that the agent interacts with, and the actions are what the agent does.

Reinforcement learning occurs when the agent chooses actions that maximize the expected reward over a given time. This is easiest to achieve when the agent is working within a sound policy framework.



How do you Choose the Right Machine Learning Solution to Use?

This is a question that you need to answer before building a machine learning model. Selecting the right kind of solution for your model is essential to avoid losing a lot of time, energy, and processing costs.



The following are factors that will help you select the right kind of machine learning solution based on supervised, unsupervised, and reinforcement learning:

1. Imagine that you'd like to predict the future stock market prices. If you are new to machine learning, you would have trouble figuring out the right solution. But with time and practice, you will begin to understand that for a problem statement like this, solution-based supervised learning will work the best for obvious reasons.

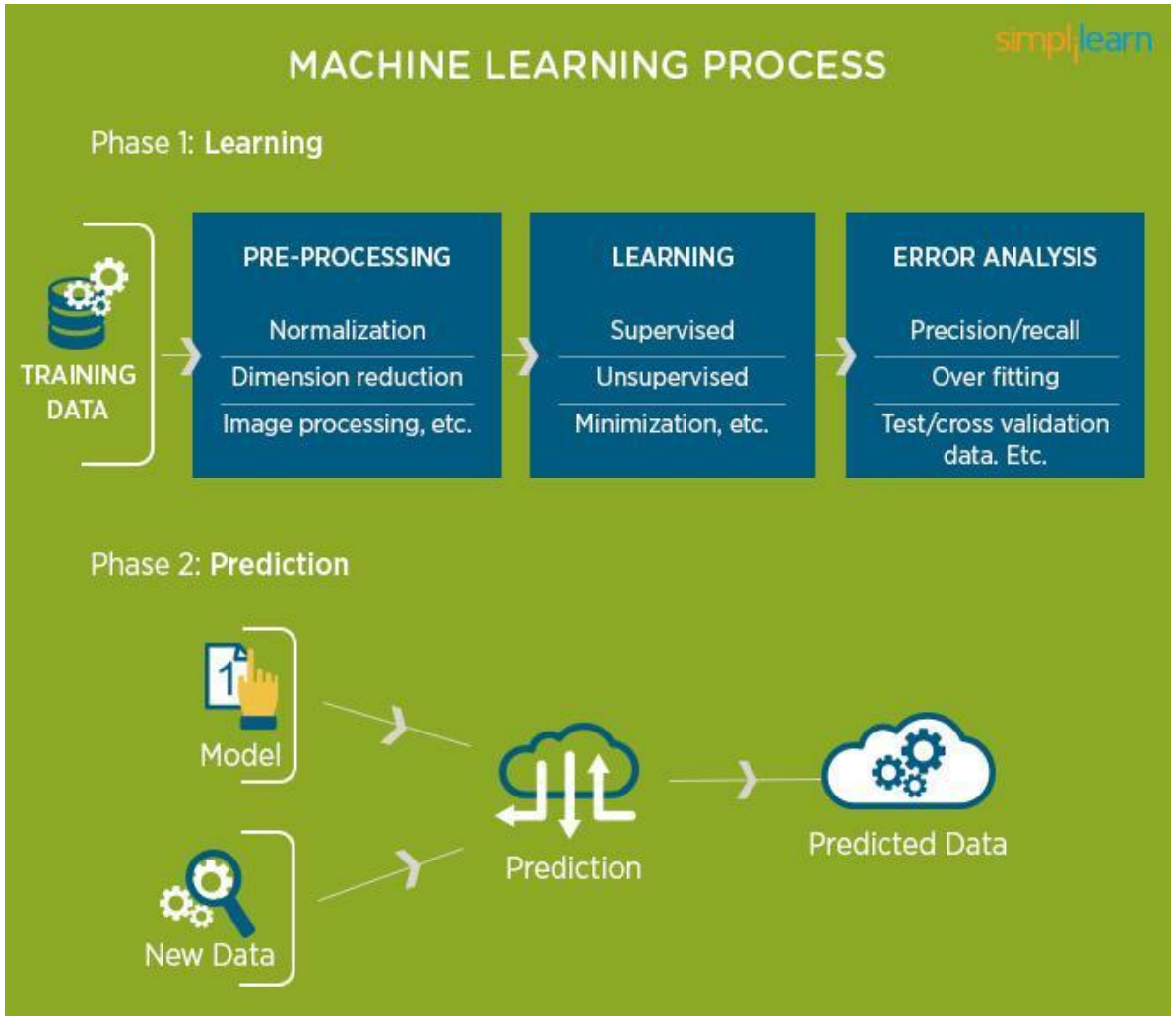
2. The size, quality, and nature of the data are also essential factors. If the data is cluttered, you will choose unsupervised. If the data set is extensive and categorical, choose supervised learning solutions.

3. Finally, you should choose a solution based on the complexity of the algorithm. As for the problem statement where you predict stock market prices, using reinforcement learning can be a solution, although that would be difficult and time-consuming, unlike supervised learning.

Why is Machine Learning Important?

To better understand the uses of Machine Learning, consider some instances where Machine Learning is applied: the self-driving Google car; cyber fraud detection; and, online recommendation engines from Facebook, Netflix, and Amazon. Machines can enable all of these things by filtering useful pieces of information and piecing them together based on patterns to get accurate results.

How Machine Learning works?



The rapid evolution in Machine Learning(ML) has caused a subsequent rise in the use cases, demands—and, the sheer importance of ML in modern life. Big Data has also become a well-used buzzword in the last few years. This is, in part, due to the increased sophistication of Machine Learning, which enables the analysis of large chunks of Big Data. Machine Learning has also changed the way data extraction and interpretation are done by automating generic methods/algorithms, thereby replacing traditional statistical techniques.

Main Uses of Machine Learning

Typical results from Machine Learning applications we either see or don't regularly include web search results, real-time ads on web pages and mobile devices, email spam filtering, network intrusion detection, and pattern and image recognition. All these are by-products of using Machine Learning to analyze massive volumes of data.

Traditionally, data analysis was trial and error-based, an approach that becomes impossible when data sets are large and heterogeneous. Machine Learning provides smart alternatives to analyzing vast volumes of data. By developing fast and efficient algorithms and data-driven models for real-time processing of data, Machine Learning can produce accurate results and analysis.

Pro Tip: For more on Big Data and how it's revolutionizing industries globally, check out our article about what Big Data is and why you should care.

Some Machine Learning Algorithms And Processes

If you're studying what is Machine Learning, you should familiarize yourself with standard Machine Learning algorithms and processes. These include neural networks, decision trees, random forests, associations, and sequence discovery, gradient boosting and bagging, support vector machines, self-organizing maps, k-means clustering, Bayesian networks, Gaussian mixture models, and more.

To get the most value out of Big Data, other Machine Learning tools and processes that leverage various algorithms include:

- Comprehensive data quality and management
- GUIs for building models and process flows
- Interactive data exploration and visualization of model results
- Comparisons of different Machine Learning models to quickly identify the best one
- Automated ensemble model evaluation to determine the best performers
- Easy model deployment so you can get repeatable, reliable results quickly
- An integrated end-to-end platform for the automation of the data-to-decision process

What Kind of Problem do I Need to Solve? How do I Solve it?

The Problem to Solve	The Category of Techniques	Covered in this Course
I want to group items by similarity. I want to find structure (commonalities) in the data	Clustering	K-means clustering
I want to discover relationships between actions or items	Association Rules	Apriori
I want to determine the relationship between the outcome and the input variables	Regression	Linear Regression Logistic Regression
I want to assign (known) labels to objects	Classification	Naïve Bayes Decision Trees
I want to find the structure in a temporal process I want to forecast the behavior of a temporal process	Time Series Analysis	ACF, PACF, ARIMA
I want to analyze my text data	Text Analysis	Regular expressions, Document representation (Bag of Words), TF-IDF

Prerequisites for Machine Learning (ML)

For those interested in learning beyond what is Machine Learning, a few requirements should be met to be successful in pursual of this field. These requirements include:

1.Basic knowledge of programming languages such as Python, R, Java, JavaScript, etc

2.Intermediate knowledge of statistics and probability

3.Basic knowledge of linear algebra. In the linear regression model, a line is drawn through all the data points, and that line is used to compute new values.

4.Understanding of calculus

5.Knowledge of how to clean and structure raw data to the desired format to reduce the time taken for decision making.