

## Study Unit 2

# Introduction to Theoretical Machine Learning



### Theoretical Machine Learning Outline

- Definition of machine learning (ML)
- Teams used in ML
- Type of ML
- Motivating examples
- Machine Learning Process
- Evaluation metrics for classification models
- Evaluation metrics for regression models

### Study Unit Duration

This Study Unit requires a minimum of 3 hours' formal study time.

You may spend an additional 2-3 hours on revision.

### Preamble

Over the past years, we all wondered whether a computer might be made to learn and improve with experience - the impact would be dramatic. Imagine a world where a computer could be made to learn about the treatments that are most effective for new diseases from the medical records or a piece of knowledge about a client that can default a loan when given.

### Learning Outcomes of Study Unit 2



Upon completion of this study unit, you should be able to:

- 2.1 Explain the Concept of Machine learning and identify the differences between Machine learning, artificial intelligent and deep learning.
- 2.2 Identify different types of machine learning methods, step involved in using the method and classification of problem solved
- 2.3 Discuss concept of Confusion Matrix and other performance evaluation metrics not in Confusion Matrix

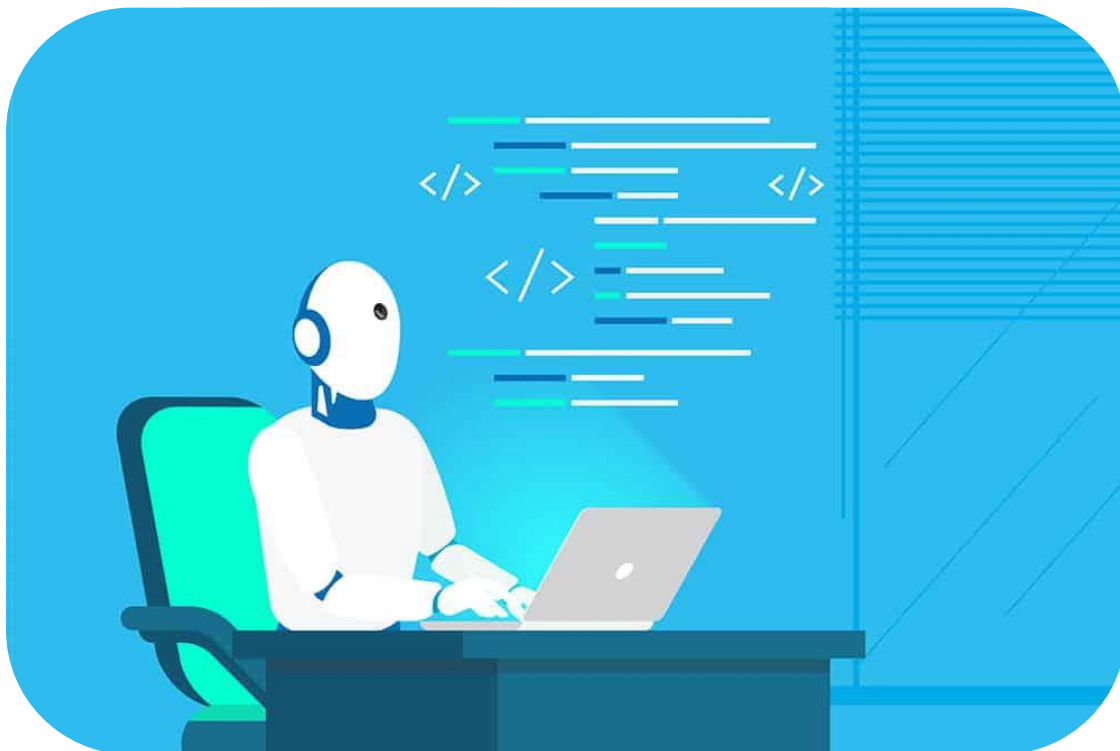
## Terminologies, Acronyms and their Meaning

<b>AI</b>	Artificial Intelligence
<b>ML</b>	Machine Learning
<b>RL</b>	Reinforcement Learning
<b>DL</b>	Deep learning
<b>EDA</b>	Exploratory Data Analysis
<b>np</b>	NumPy
<b>sns</b>	Seaborn

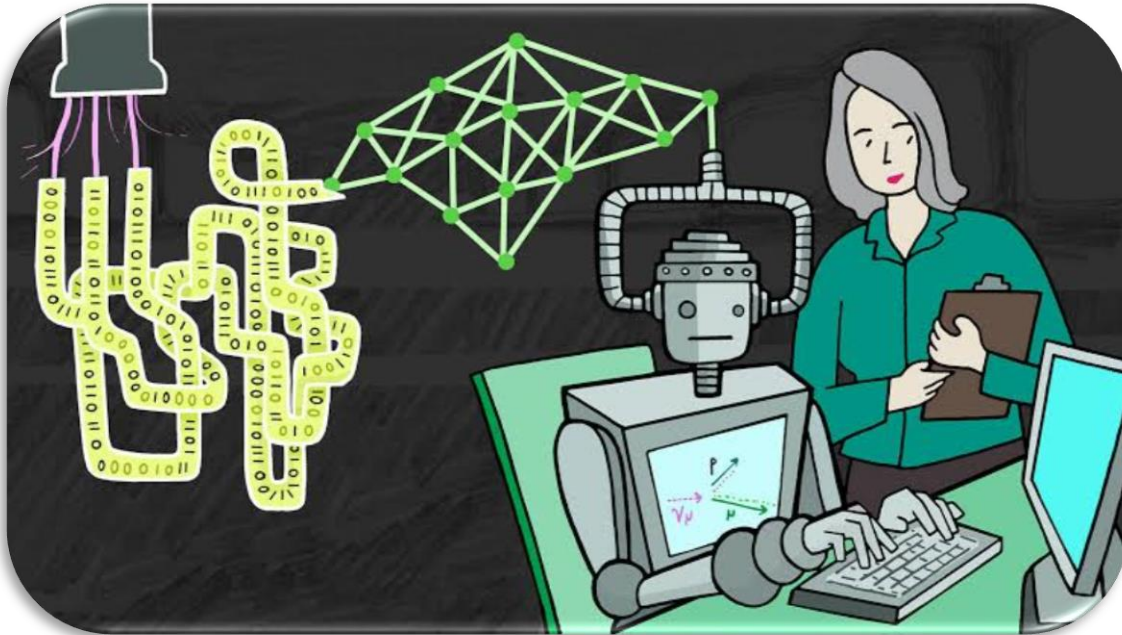
<b>pd</b>	Pandas
<b>TP</b>	True Positive
<b>FP</b>	False Positive
<b>FN</b>	False Negative
<b>TN</b>	True Negative
<b>RMSE</b>	Root Mean Squared Error

## What do you think about machine learning?

When you think about Machine Learning (ML) what comes to your mind. This

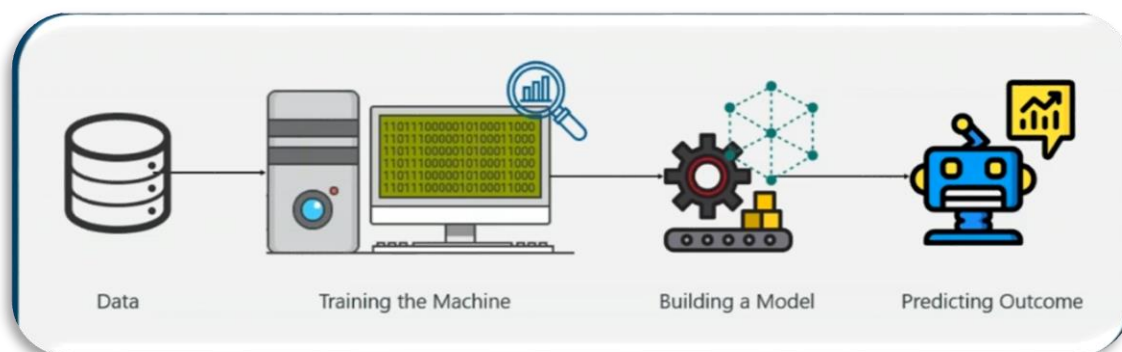


or

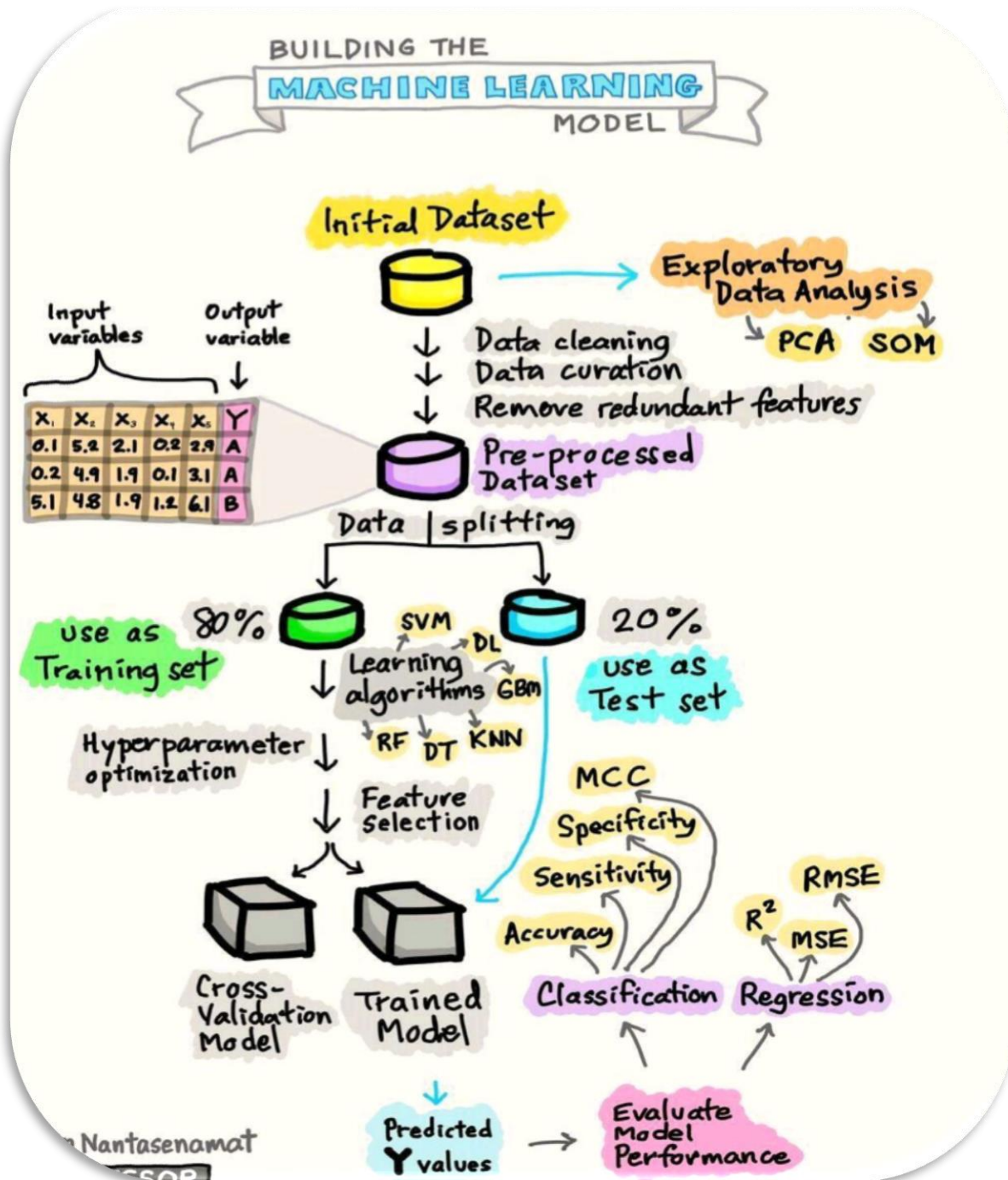


## 2.1 Definition of Machine learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. That is, it uses mathematical algorithms which end goal is to learn from data and make prediction about similar instance in a new data point.

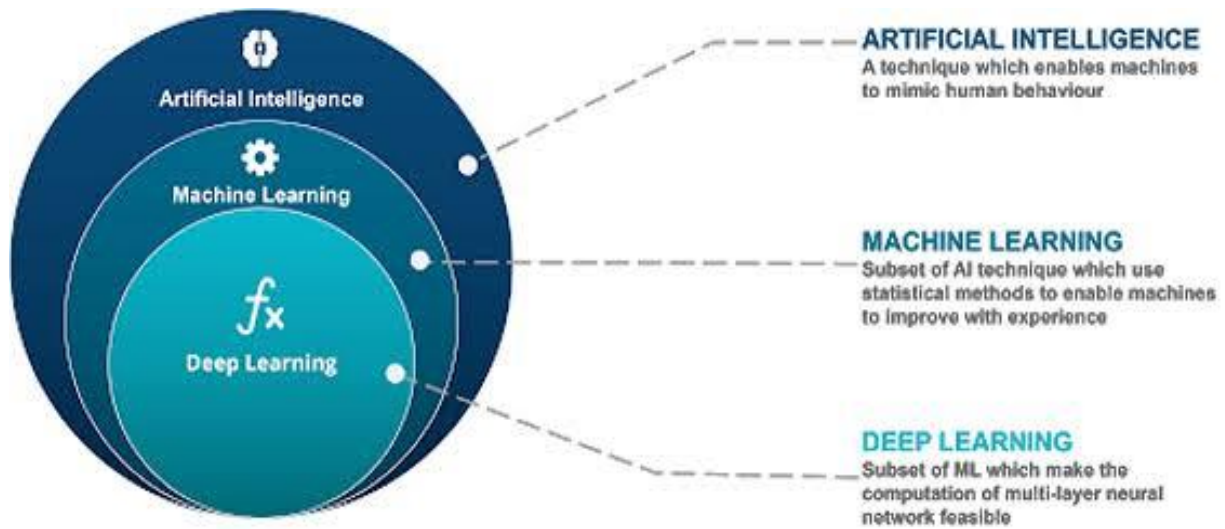


When else you hear about machine learning, think about this:

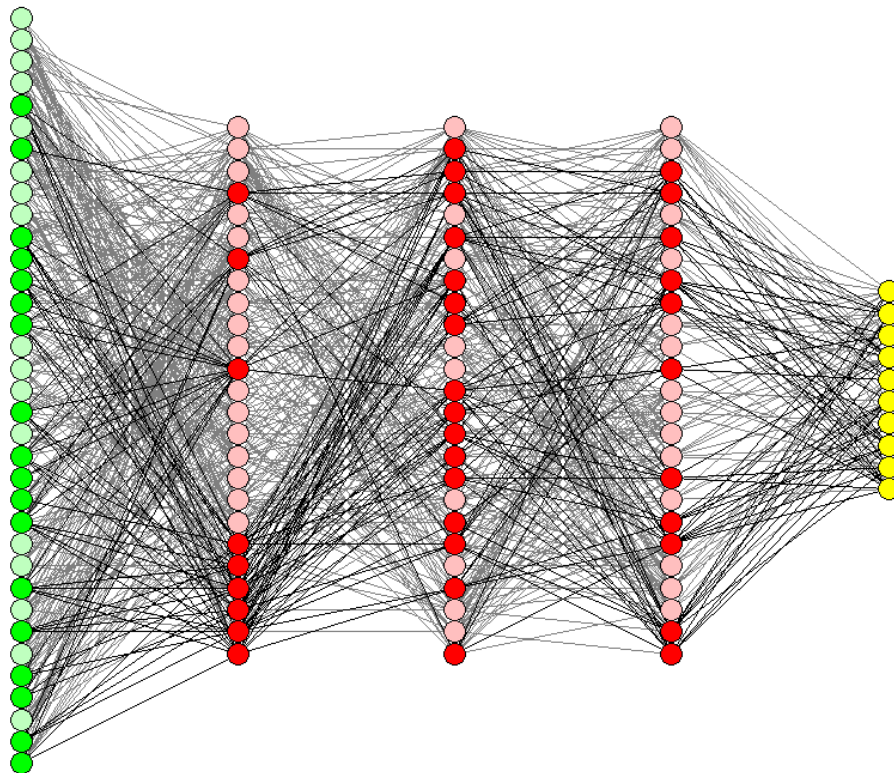


### 2.1.1 Differences between Machine Learning, Artificial Intelligence and Deep Learning

Artificial intelligence is imparting a cognitive ability to a machine. It is basically giving machine a form of intelligent.



Deep Learning is a computer algorithm that simulates the network of neurons in a brain. It is a subset of Machine Learning



## 2.1.2 Terms in Machine Learning (ML)

The following are some basic concepts which must be defined in machine learning:

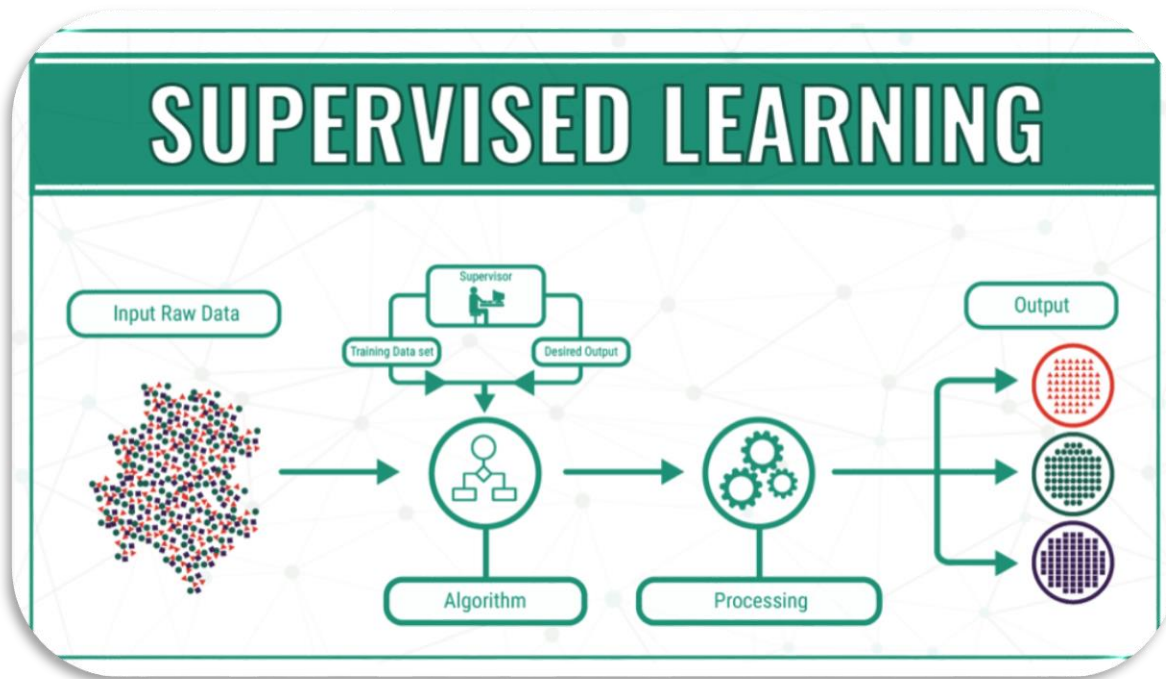
- ✚ Data: Data can be defined as any form of information that can be inputted to a computer
- ✚ Instance: rows/observations in the dataset
- ✚ Algorithm: A set of rules and statistical techniques used to learn patterns from data
- ✚ Model: A model is trained on data by using a machine learning algorithm
- ✚ Label: It is the output variable that needs to be predicted (i.e. whether the animal is a Dog or a Cat) by using the set of predictors variables known as features
- ✚ Feature: Attributes of the dataset that can predict the label
- ✚ Training data: This is the set of dataset that is set apart to train our models. This always contains 80% of the original dataset
- ✚ Testing data: The trained model(s) needs to be evaluated on the unseen data called test data. This shows the performance of the model(s). It is always 20% of the original dataset.

## 2.2 Types of Machine Learning

Machine learning techniques can be implemented in four ways and these include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.

### 2.2.1 Supervised learning

Supervised learning algorithms consist of a label which is to be predicted from a given set of features. The task is to learn a function that maps an input to an output based on example input-output pairs.



A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping test data or new input (X) that can predict the output or label (Y) for that data.

Common examples of supervised learning include:

- ✚ classifying e-mail as spam or not spam (inbox)
- ✚ predicting the housing price based on the attributes of the house

Supervised learning is the most studied and common type of machine learning. Learning with supervision is much easier than learning without supervision.

### **Type of supervise learning**

Supervised learning can be classification or regression.

#### **1. Regression**

A Regression model is the type of model in which the out variable is continuous e.g. predicting housing price at Mogadishu city in Somalia. Other example includes predicting highest amount of rainfall in Uganda in May 2021.

## Motivating Examples on Regression Problem

### Problem 1: Predicting medical insurance charge from the health provider in Ethiopia

The objective of this regression task is to predict the medical insurance cost given the age, sex, body mass index (BMI), number of children, and region of the main beneficiary from a health insurance provider in Ethiopia.

age	sex	bmi	children	smoker	regions	charges
19	female	27.9	0	yes	Addis Ababa	16884.924
18	male	33.77	1	no	Afar	1725.552
28	male	33	3	no	Amhara	4449.462
33	male	22.705	0	no	Benishangul-Gumuz	21984.471
32	male	28.88	0	no	Dire Dawa	3866.855
50	male	30.97	3	no	Gambela	10600.548
18	female	31.92	0	no	Harari	2205.981
18	female	36.85	0	no	Oromia	1629.833
21	female	25.8	0	no	Sidama	2007.945
61	female	29.07	0	yes	Somali	29141.36
19	female	27.9	0	yes	Southern Nations	10969.185
18	male	33.77	1	no	Tigray	11245.456
28	male	33	3	no	Benishangul-Gumuz	11521.727
33	male	22.705	0	no	Oromia	11797.999
32	male	28.88	0	no	Dire Dawa	12074.27
50	male	30.97	3	no	Afar	12350.541
18	female	31.92	0	no	Gambela	12626.812
18	female	36.85	0	no	Harari	12903.083
21	female	25.8	0	no	Afar	13179.355
61	female	29.07	0	yes	Amhara	13455.626

**Label:** The label for this problem is medical insurance charges.

**Features:** There are 6 features and they are age, sex, bmi, children, smoker and, regions.

### Problem 2: Predicting amount of tips given to a food server after a party in the restaurant

The objective of the regression task is to predict the amount of tip (gratuity in Kenya Shilling) given to a food server based on total\_bill, gender, smoker (whether they smoke in the party or not), day(day of the week for the party), time(time of the day whether for lunch or dinner), and size(size of the party)

	A	B	C	D	E	F	G
1	total_bill	tip	gender	smoker	day	time	size
2	2125.5	360.79	Male	No	Thur	Lunch	1
3	2727.18	259.42	Female	No	Sun	Dinner	5
4	1066.02	274.68	Female	Yes	Thur	Dinner	4
5	3493.45	337.9	Female	No	Sun	Dinner	1
6	3470.56	567.89	Male	Yes	Sun	Lunch	6
7	2411.08	296.48	Female	Yes	Thur	Lunch	2
8	4607.43	374.96	Female	No	Thur	Dinner	4
9	1165.21	700.87	Female	No	Mon	Dinner	2
10	2895.04	347.71	Male	No	Sat	Dinner	5
11	2622.54	253.97	Male	Yes	Thur	Lunch	6
12	2572.4	499.22	Male	No	Wed	Lunch	2
13	1602.3	336.81	Male	Yes	Wed	Lunch	1
14	1888.97	451.26	Female	No	Tues	Lunch	3
15	2643.25	541.73	Male	No	Sun	Dinner	6
16	1899.87	213.64	Male	Yes	Fri	Dinner	4
17	2358.76	156.96	Female	Yes	Mon	Lunch	1
18	3374.64	401.12	Male	Yes	Tues	Lunch	1
19	2353.31	555.9	Male	Yes	Sun	Lunch	4
20	2655.24	276.86	Female	Yes	Mon	Dinner	1

**Label:** The label for this problem is tip.

**Features:** There are 6 features and they are total bill, gender, smoker, day, time and, size.

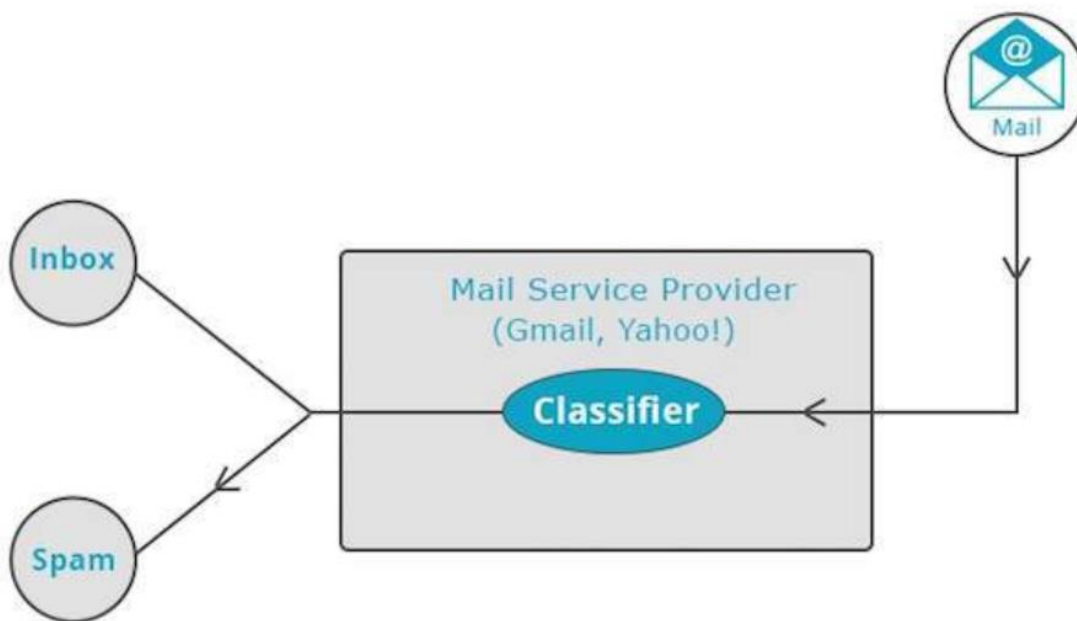
There are many regression algorithms and the most common ones are:

- ✚ Ordinary Least Square (OLS)
- ✚ Support Vector Machine (SVM)
- ✚ K- Nearest Neighbour (KNN)

- Decision Tree
- Random Forest

## 2. Classification

A classification (model) is the type of model in which the output variable (i.e. the label) is discrete/categorical. e.g. predict if the patient has cancer or not, if an employee will leave or stay etc. Classification problem is an example of pattern recognition. Machine learning method that implements classification is known as a classifier.

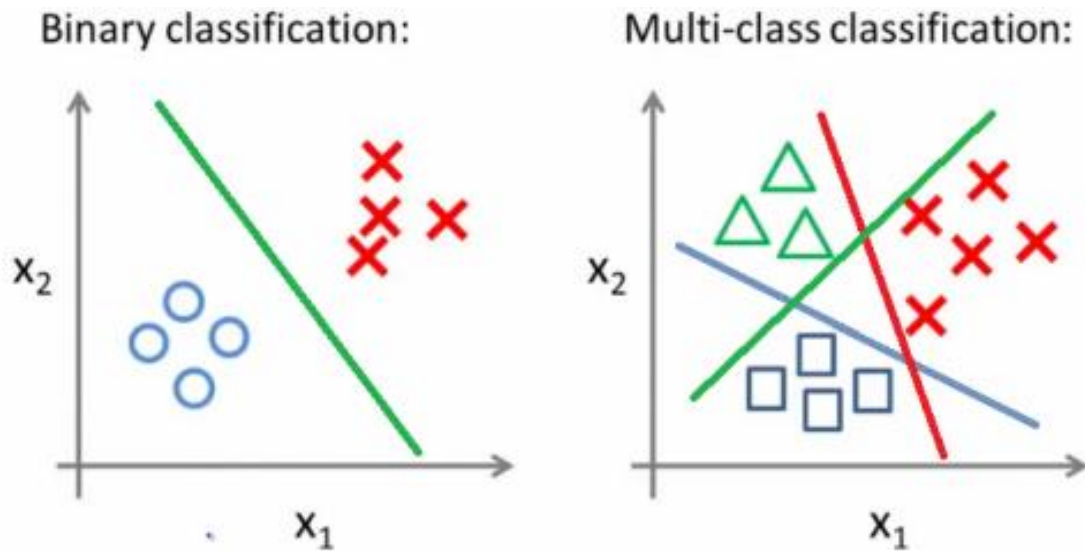


### E-mail filtering binary classification problem

#### Types of classification task

- Binary classification: When two classes are involved in the label (Y), it is called binary classification problem. For example, e-mail spam detection (1: spam; or 0: not spam), disease modelling whether a patient has disease or not.
- Multi-class classification: When more than two classes are involved in the label (Y), it is called multi classification problem. For example, digit recognition (where classes

go from 0 to 9), predicting a party that wins the election (there may be more than two parties in the election).



### Motivating Examples on Classification Problem

#### Problem 1: Predicting whether patient has liver disease or not

The goal of this classification problem is to build a binary classifier that divides the patients into groups (liver patient or non-liver patient) using characteristics such as Age, Gender, TB (Total Bilirubin), DB (Direct Bilirubin), Alkphos (Alkaline Phosphatase), Sgpt (Alamine Aminotransferase), Sgot (Aspartate Aminotransferase), TP (Total Protiens), ALB (Albumin), and A\_G (Ratio Albumin and Globulin Ratio) of the patients in Uganda.

	Age	Gender	TB	DB	Alkphos	Sgpt	Sgot	TP	ALB	A_G	Selector
1:	65	1	0.7	0.1	187	16	18	6.8	3.3	0.90	Yes
2:	62	0	10.9	5.5	699	64	100	7.5	3.2	0.74	Yes
3:	62	0	7.3	4.1	490	60	68	7.0	3.3	0.89	Yes
4:	58	0	1.0	0.4	182	14	20	6.8	3.4	1.00	Yes
5:	72	0	3.9	2.0	195	27	59	7.3	2.4	0.40	Yes
---											
579:	60	0	0.5	0.1	500	20	34	5.9	1.6	0.37	No
580:	40	0	0.6	0.1	98	35	31	6.0	3.2	1.10	Yes
581:	52	0	0.8	0.2	245	48	49	6.4	3.2	1.00	Yes
582:	31	0	1.3	0.5	184	29	32	6.8	3.4	1.00	Yes
583:	38	0	1.0	0.3	216	21	24	7.3	4.4	1.50	No

**Label:** The label for this problem is Selector and it has two classes namely yes or no.

**Features:** There are 10 features and they are Age, Gender, TB (Total Bilirubin), DB (Direct Bilirubin), Alkphos (Alkaline Phosphotase), Sgpt (Alamine Aminotransferase), Sgot(Aspartate Aminotransferase), TP(Total Protiens), ALB(Albumin), A\_G(Ratio Albumin and Globulin Ratio).

## Problem 2: Predicting Contraceptive Method Choice in Somalia







The problem is to build a multi-class classifier that predicts the current contraceptive method choice (no use, long-term methods, or short-term methods) of Somalia woman based on her demographic and socio-economic characteristics such as wife age, wife education, husband education, children, wife religion, wife working, husband occupation, standard of living index and media exposure.

wife_age	wife_education	husband_education	children	wife_religion	wife_working	husband_occupation	standard_of_living_index	media_exposure	contraceptive_method_used
40	4	4	6	Islam	No	1	3	Good	Long-term
44	4	2	8	Islam	No	3	3	Good	No-use
48	1	1	8	Islam	Yes	2	2	Not good	No-use
26	2	2	4	Islam	No	3	4	Good	Long-term
28	2	3	3	Islam	No	3	4	Good	Short-term
28	3	4	3	Islam	No	2	1	Good	No-use
41	4	4	8	Islam	No	1	4	Good	Long-term
34	2	4	7	Islam	No	2	3	Good	Short-term
21	2	1	3	Non-Islam	Yes	3	1	Not good	Short-term
25	4	4	1	Islam	No	1	4	Good	No-use
49	1	2	5	Islam	Yes	2	2	Not good	No-use
42	1	3	6	Islam	No	3	3	Good	No-use
35	2	2	6	Islam	Yes	3	1	Not good	No-use
38	3	4	7	Islam	No	3	2	Good	Short-term
42	4	4	4	Islam	No	3	4	Good	Long-term
43	1	2	8	Islam	No	2	4	Good	No-use
34	3	3	6	Islam	No	2	3	Good	No-use
41	1	3	5	Islam	No	2	3	Not good	No-use
32	3	2	4	Islam	No	2	2	Not good	No-use
45	3	3	8	Islam	No	2	3	Good	Long-term
45	1	3	7	Islam	Yes	1	3	Good	No-use
45	1	3	10	Islam	No	3	4	Good	No-use

**Label:** The label for this problem is contraceptive method used and it has three classes namely long-term, short-term, or no-use.

**Features:** There are 9 features and they are wife\_age, wife\_education, husband\_education, children, wife\_religion, wife\_working, husband\_occupation, standard\_of\_living\_index and media\_exposure.

There are many classification learning algorithms and they include:

-  Support Vector Machine (SVM)
-  K- Nearest Neighbour (KNN)
-  Logistic Regression
-  Decision Trees
-  Random Forest
-  Naive Bayes

### Additional resources

For more information about the difference between classification and regression please visit this [source](#).

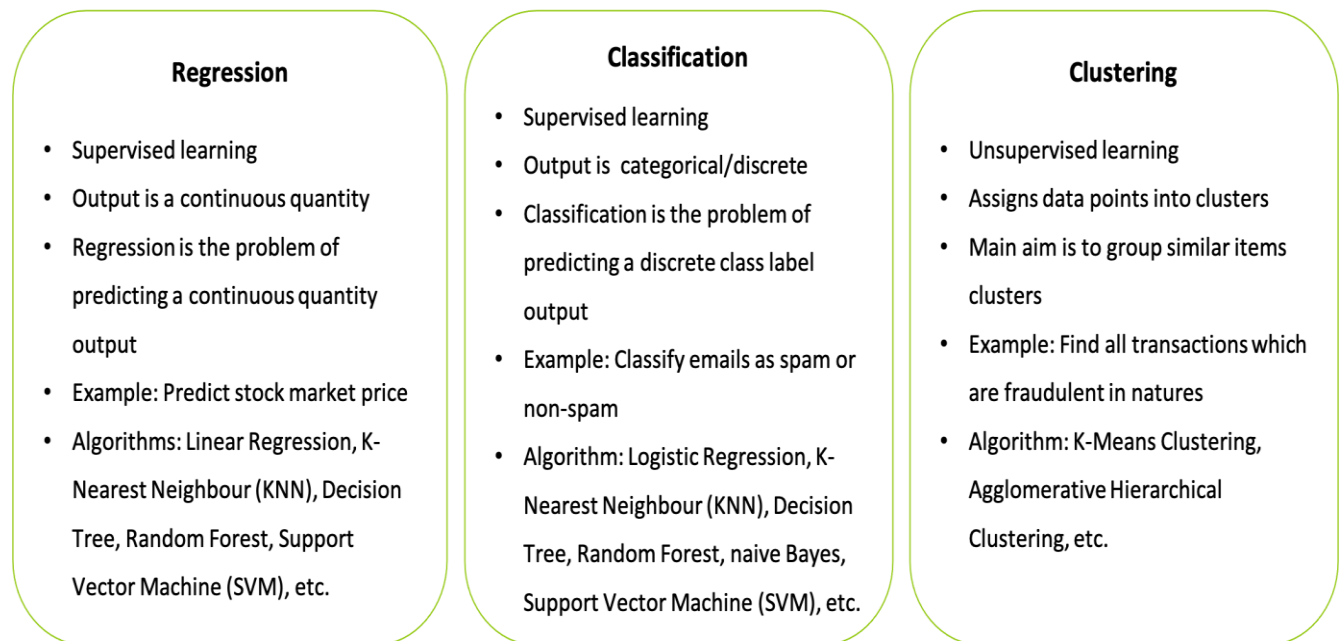
### 2.2.2 Unsupervised learning

Unsupervised learning is the training of the model on an unlabeled dataset. That is, there is no information regarding the label in the training dataset. We build a mathematical model of a set of data by finding similarities in the observations. Unsupervised learning is the training of the model on an unlabeled dataset.

After the model is trained, each new observation is assigned to a cluster of observations with similar characteristics. The goal of unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data or segment into different groups or clusters based on their attributes.

Most common application of unsupervised machine learning techniques include Clustering, Factor Analysis (FA), Anomaly detection, Association mining, and Latent variable model (PCA).

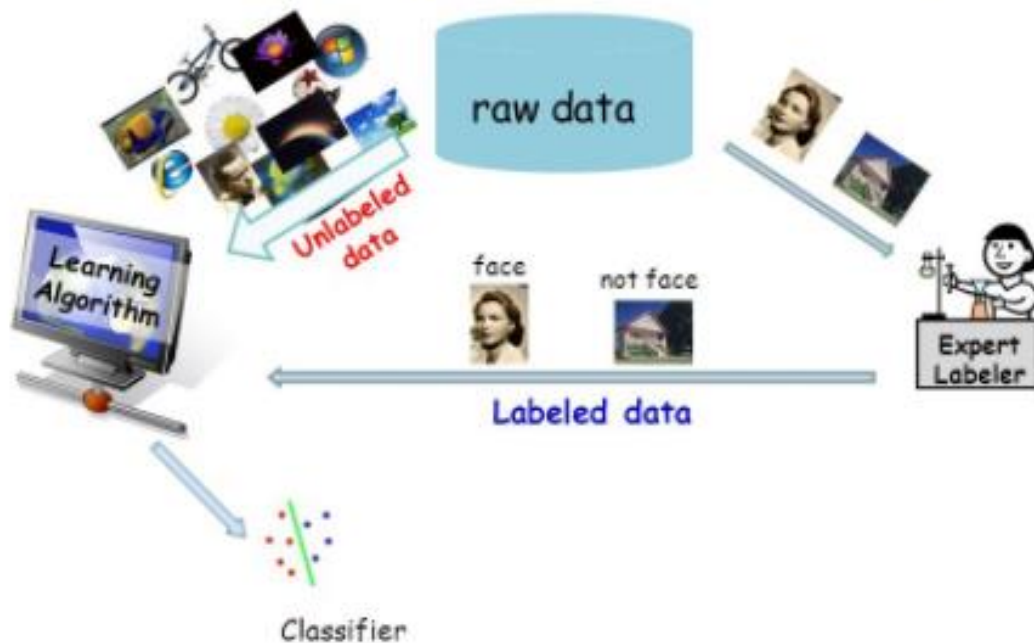
### 2.2.3 Classification of problems solved by supervised and unsupervised machine learning (ML)



#### Semi-supervised learning

Semi-supervised machine learning algorithm lies between supervised and unsupervised learning. It is a learning paradigm concerned with the study of how computers and natural systems such as humans learn in the presence of both labeled and unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Semi-supervised learning is usually chosen when the acquired labeled data requires skilled and relevant resources in order to train it or learn from it.

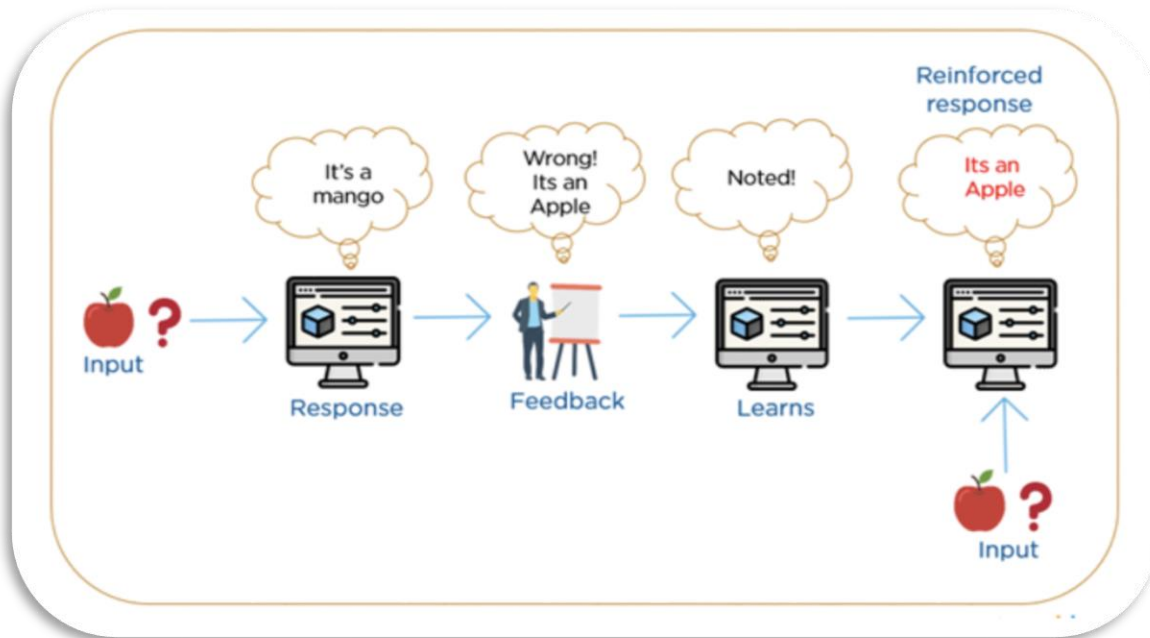
## Semi-Supervised Learning



Popular semi-supervised learning models include self-training, mixture models, co-training and multiview learning, graph-based methods, and semi-supervised support vector machines.

### Reinforcement learning

Reinforcement machine learning algorithm is a learning method that interacts with its environment by producing actions and discovers errors or rewards. The idea involved in reinforcement learning is that the machine trains itself on a continual basis based on the environment it is exposed to, and applies its enriched knowledge to solve business problems.



Source: [Shanika Perera](#).

Reinforcement learning algorithms are used in learning to play a game i.e. chess against a human opponent, in self-driving cars or a robot that learns from a fall.

#### Additional resources

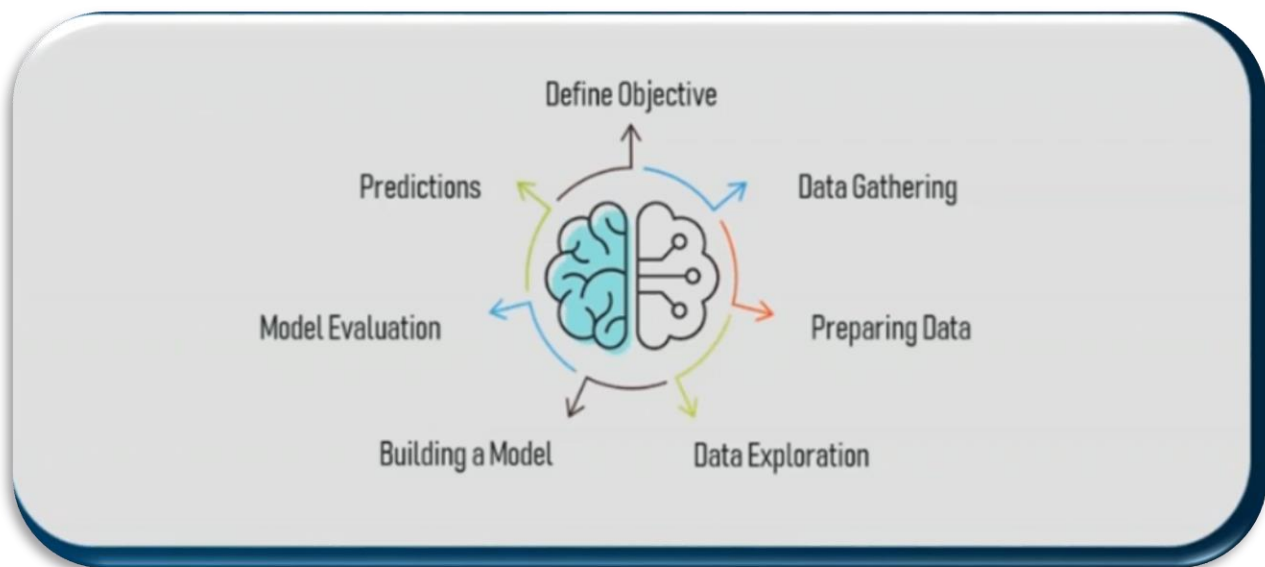
You can visit this [source](#) for more information about reinforcement learning.

## Machine learning summary

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
<b>Definition</b>	The machine learns by using labelled data	The machine is trained on unlabeled data without any guidance	An agent interacts with its environments by producing actions & discovers errors or rewards
<b>Type of problems</b>	Regression and Classification	Association & Clustering	Reward based
<b>Type of data</b>	Labelled data	Unlabeled data	No pre-defined data
<b>Training</b>	External supervision	No supervision	No supervision
<b>Approach</b>	Map labelled input to known output	Understand patterns and discover output	Follow trial and error method
<b>Popular algorithms</b>	Linear regression, Logistic regression, Support Vector Machine, KNN, etc.	K-means, C-means etc.	Q-Learning, State-action-reward-state-action (SARSA), etc.

### 2.2.3 Machine Learning Process

Machine Learning process involves building a **predictive model** that can be used to find a solution for a **problem statement**.



### Step Involved in Machine Learning

There are seven step involved in using learning machine

### Step 1: Define the objective of the problem

For example, to predict the possibility of rain by studying the weather conditions.



- What are we trying to predict?
- What are the features in the data?
- What is the label in the data?
- What kind of problem are we facing?  
Classification, regression, or clustering?

*Objective statement*

### Step 2: Data Gathering

Data such as weather conditions, humidity level, temperature, pressure, etc. are either collected manually or scraped from the web.

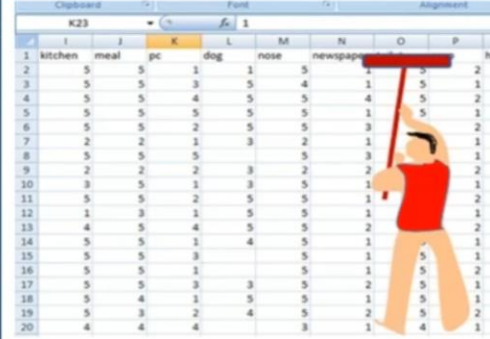


*Data gathering*

### Step 3: Data Preprocessing

Data cleaning involves getting rid of inconsistencies in data such as missing values or redundant variables.

- Transform data into desired format
- Data cleaning
  - Missing values
  - Corrupted data
  - Remove unnecessary data



	I	J	K	L	M	N	O	P	Q
1	kitchen	meal	pc	dog	nose	newspaper			house
2		5	5	1	5	5	1	5	2
3		5	5	3	5	4	1	5	1
4		5	5	4	5	5	4	5	2
5		5	5	5	5	5	1	5	1
6		5	5	2	5	5	3	5	2
7		2	2	1	3	2	1	5	1
8		5	5	5	5	5	3	5	1
9		2	2	2	3	2	2	5	2
10		3	5	1	3	5	1	5	1
11		5	5	2	5	5	1	5	2
12		1	3	1	5	5	1	5	1
13		4	5	4	5	5	2	5	2
14		5	5	1	4	5	1	5	1
15		5	5	3		5	1	5	1
16		5	5	1		5	1	5	2
17		5	5	3	3	5	2	5	1
18		5	4	1	5	5	1	5	1
19		5	3	2	4	5	2	5	2
20		4	4	4		3	1	4	1

#### Data cleaning

It also includes encoding all the features that are string data type into numeric by a method known as Label Encoder or one-hot encoding. You can learn more about this concept [here](#).

### Step 4: Exploratory Data Analysis (EDA)

Data Exploration involves understanding the patterns and trends in the data. At this stage all the useful insights are drawn and correlations between the variables are understood.



### *Exploratory Analysis*

#### **Step 5: Building a Machine Learning Model**

At this stage a predictive model is built by using machine learning algorithms such as Linear Regression, Decision Trees, etc.

- Machine Learning model is built by using the training data set
- The model is the Machine Learning algorithm that predicts the output by using the data fed to it



It is necessary when building supervised machine learning models to split our dataset into training and test data, the learning method should see only the training data to learn the relationship between the set of features (X) and label (Y). This learned information forms what is called a machine learning model. The training data is 80% of data for model training,

and the remainder 20% will be used for model evaluation or performance. The model built is then used to predict the label (Y) in test data by looking only the input (X) of test data.

### Step 6: Model Evaluation & Optimization

The efficiency of the model is evaluated by using metrics in **A or B** and any further improvement in the model are implemented.

- Machine Learning model is evaluated by using the testing data set
- The accuracy of the model is calculated
- Further improvement in the model are done by using techniques like Parameter tuning



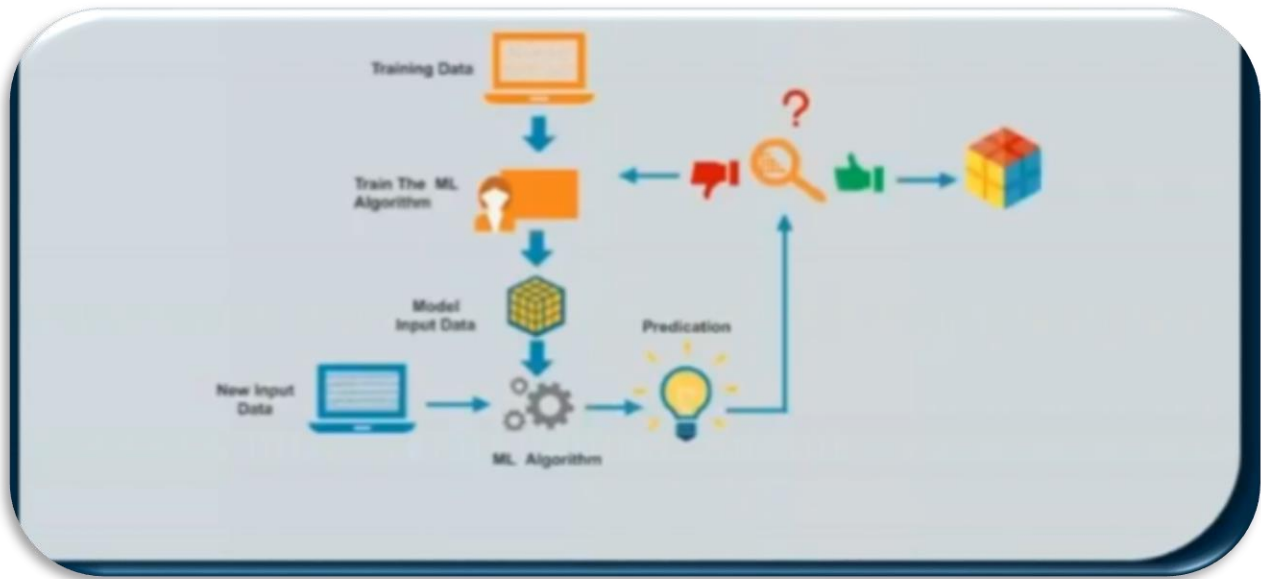
Machine Learning Model

### *Model Evaluation*

The predicted values of Y is then compared to the known label (Y) in the test dataset to evaluate the performance of the model.

### Step 7: Predictions






The final outcome is predicted after performing parameter tuning and improving the accuracy of the model.



### *Prediction*

## **2.3 Evaluation Metrics for Classification Models**

There are many different metrics for evaluating the performance of classification models and they include:

-  Accuracy
-  Sensitivity
-  Precision
-  F1 score
-  Area under Receiver Operating Characteristics curve (AUC)

Most of these metrics can be calculated from a confusion matrix.

### **2.3.1 Confusion Matrix and Other metrics not in Confusion Matrix**

#### **Confusion Matrix**

A confusion matrix is one of the metrics that evaluate the performance of a classifier on a test data or validation. A confusion matrix which is also known as an error matrix is a table that allows visualization of the performance of an algorithm. The rows of the confusion

matrix represent the actual labels that were contained in the test dataset while the columns represent what the classifier predicted and vice-versa.

		Predicted	
		Positive	Negative
Actual	Positive	True positive	False Negative
	Negative	False Positive	True Negative

### Concept in Confusion Matrix

The following are the terms commonly used in confusion matrix. Here we are referring to diabetics' scenario.

- ✚ **Positive (1)**: Presence of diabetics
- ✚ **Negative (0)**: Absence of diabetics
- ✚ **True Positive (TP)**: The patient is diabetic, and the classifier predicted it as diabetic
- ✚ **False Positive (FP)**: The patient is not diabetic, and the classifier predicted it as diabetic. This is known as Type I error.
- ✚ **True Negative (TN)**: The patient is not diabetic, and the classifier predicted it as not diabetic
- ✚ **False Negative (FN)**: The patient is diabetic, and the classifier predicted it as not diabetic

The confusion matrix shows the ways in which classifier is confused when making a prediction and the types of errors that are being made. A typical example of a confusion matrix using Logistic regression classifier on diabetics' dataset is shown below:

		Predicted	
		Diabetic	Non Diabetic
Actual	Diabetic	38	3
	Non Diabetic	2	70

### 2.3.2 Performance evaluation metrics of a classification model

The statistical measure of the performance of a given binary classifier can be evaluated using accuracy, sensitivity, and specificity.

**Accuracy or Classification Rate:** Accuracy of any given classifier is the ratio of what classifier predicted correctly over the total number of predictions.

Accuracy can be expressed as follows:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} = \frac{TP + TN}{TP + FN + FP + TN}$$

Accuracy often encounters problem when the classes in the data are not balanced. That is, there is no equal proportion in the class label. For example, an imbalanced data may have higher number of people without diabetes and fewer number of people with diabetes. In that case, we can use F1 to evaluate those model that we built.

**Misclassification error:** This is the fraction of points that the classifier misclassified the true label. It is also called classification error.

$$\text{Classification error} = \frac{FP + FN}{TP + FN + FP + TN}$$

or

$$\text{Classification error} = 1 - Accuracy$$

**Sensitivity:** Sensitivity or Recall, is the ratio of the total number of patients that are correctly classified as having the disease (diabetic) present divided by the total number of patients that have the diseases in the test data.

Sensitivity can be expressed as follows:

$$Sensitivity = \frac{TP}{TP + FN}$$

**Precision:** The precision measures how often a classifier correctly predicts the response of interest (disease present). For instance, when the classifier predicts diabetes to be present in a set of samples, how often is it correct?

Precision can be expressed as follows:

$$Precision = \frac{TP}{TP + FP}$$

**F-measure:** Also known as F-score or F1 score is a measure of binary classifier's accuracy. It uses both precision and recall to compute the score. F-measure is the harmonic mean of the precision and recall. For a given binary classifier, F1 score reaches its best value at 1 and worst at 0

$$F\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$




### 2.3.3 Other metrics not in Confusion Matrix

-  Area under a ROC curve (AUC)
-  Logloss

## Evaluation Metrics for Regression Learning Methods

Unlike classification that has the label in a categorical variable, output variable in regression learning model is continuous and therefore, we will have different evaluation metrics to assess the performance of such model.

The main metrics for model evaluation in regression models includes:

-  Mean Square Error (MSE)
-  Root Mean Square Error (RMSE)
-  Mean Absolute Error (MAE)

The lower the error, the better the model. All these metrics can range from 0 to  $\infty$

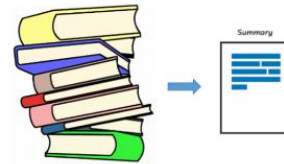
## Summary of Confusion matrix

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) <b>Type II Error</b>	<b>Sensitivity</b> $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) <b>Type I Error</b>	True Negative (TN)	<b>Specificity</b> $\frac{TN}{(TN + FP)}$
		<b>Precision</b> $\frac{TP}{(TP + FP)}$	<b>Negative Predictive Value</b> $\frac{TN}{(TN + FN)}$	<b>Accuracy</b> $\frac{TP + TN}{(TP + TN + FP + FN)}$

*Accuracy score*

You can also learn more by clicking this [reference](#).

## Summary of Study Unit 2



In this study unit, you have learnt that:

1. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
2. Machine learning methods can be classified into four (4) and they include Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, and Reinforcement Learning.
3. Supervised learning algorithms consist of a label which is to be predicted from a given set of features.
4. Supervised Learning task can be Regression or Classification
5. Unsupervised learning is the training of the model on an unlabeled dataset.

6. Machine learning processes include defining the objective of the problems, data gathering, data preprocessing, exploratory data analysis (EDA), building a machine learning model, model evaluation and optimization, and prediction.
7. The confusion matrix shows the ways in which classifier is confused when making a prediction and the types of errors that are being made
8. The statistical measure of the performance of a given binary classifier can be evaluated using accuracy, sensitivity, specificity, and F1 score.
9. The main metrics for model evaluation in regression models includes, Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The lower the error, the better the model

