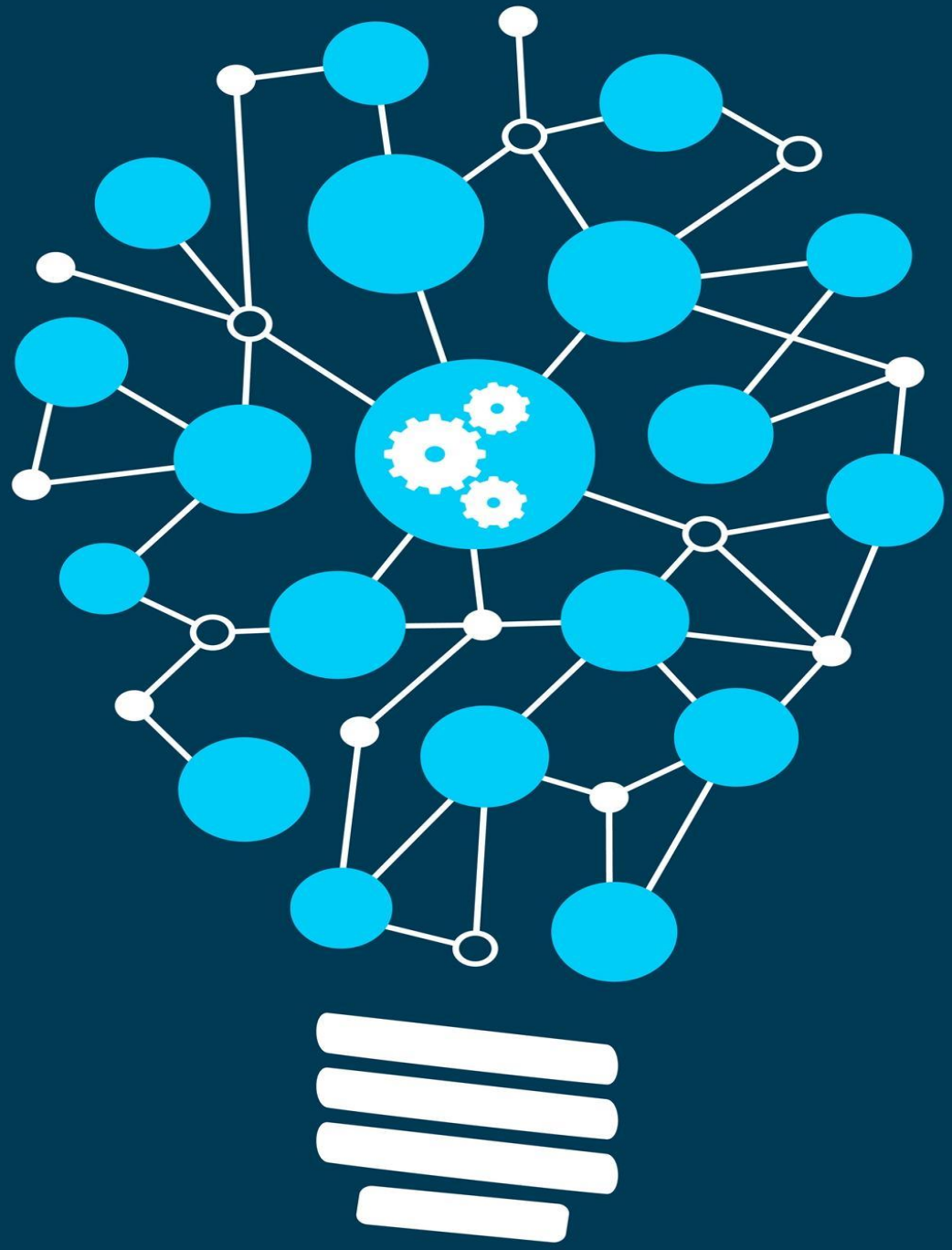
MACHINE LEARNING



Machine Learning

Credit Hours: (2+1)

Presenter
Ms. Qurat-ul-Ain Raja

Types of Machine Learning

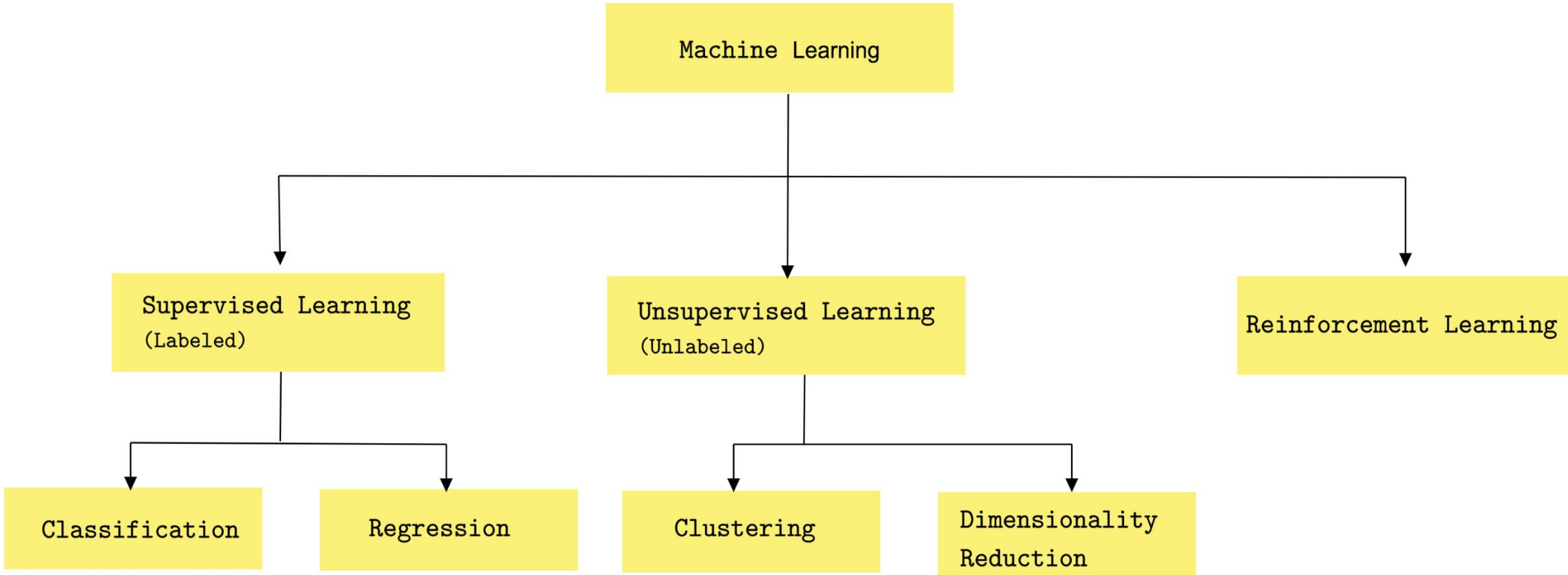
TYPES OF MACHINE LEARNING

Supervised Machine Learning

Unsupervised Machine Learning

Semi-Supervised Learning

Reinforcement Learning


Machine Learning
Supervised Learning
(Labeled)
Unsupervised Learning
(Unlabeled)
Reinforcement Learning
Classification
Regression
Clustering
Dimensionality
Reduction

Types of Machine Learning

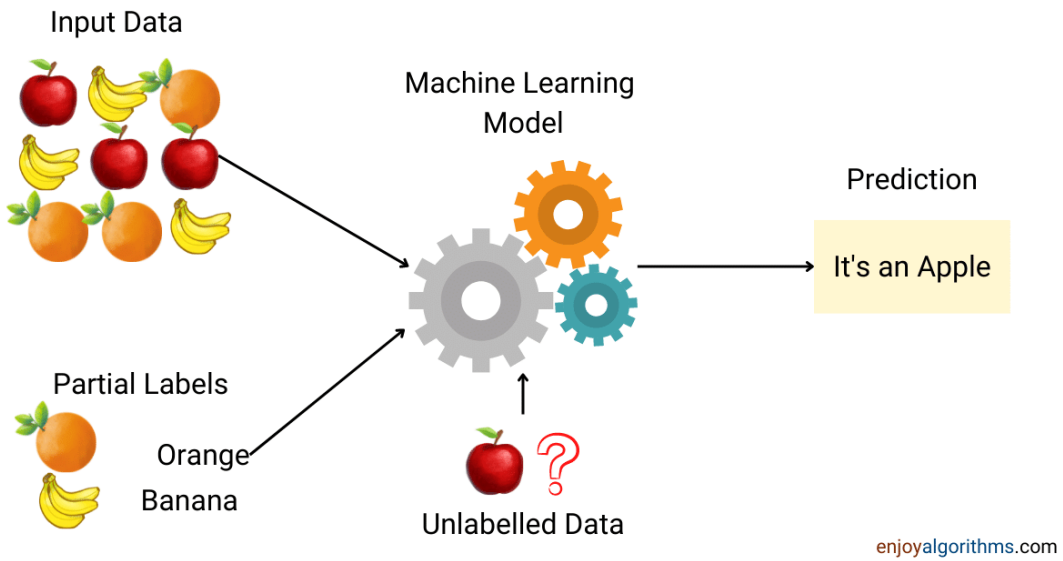
Types of Machine Learning

At a broad level, machine learning can be classified into three types:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Supervised Learning

- Supervised learning is a technique where the machine is given labelled input data and the expected output data. It gets the data from training data containing sets of examples.



Supervised Learning

- Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.
- The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.

Supervised Learning

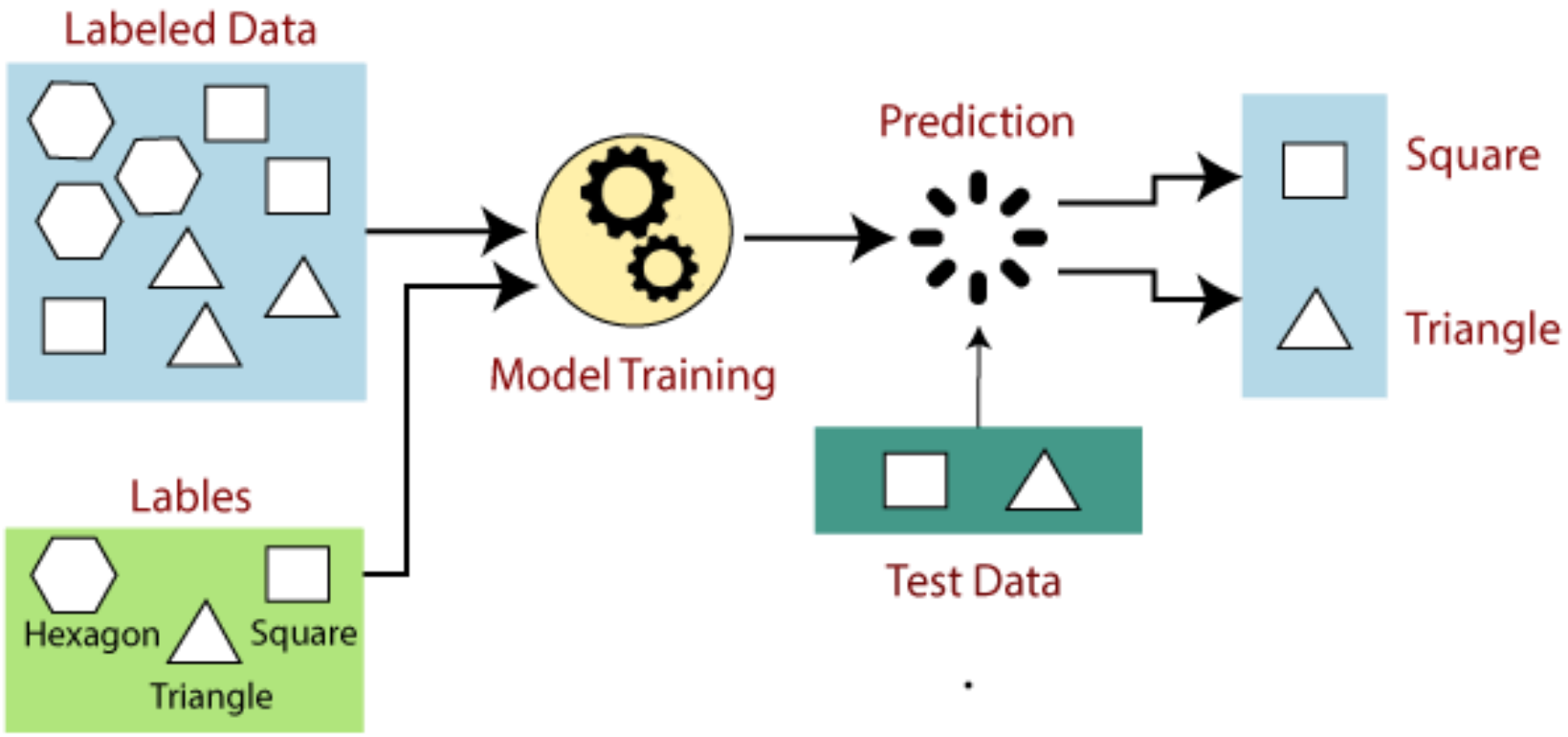
- Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.

- The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.

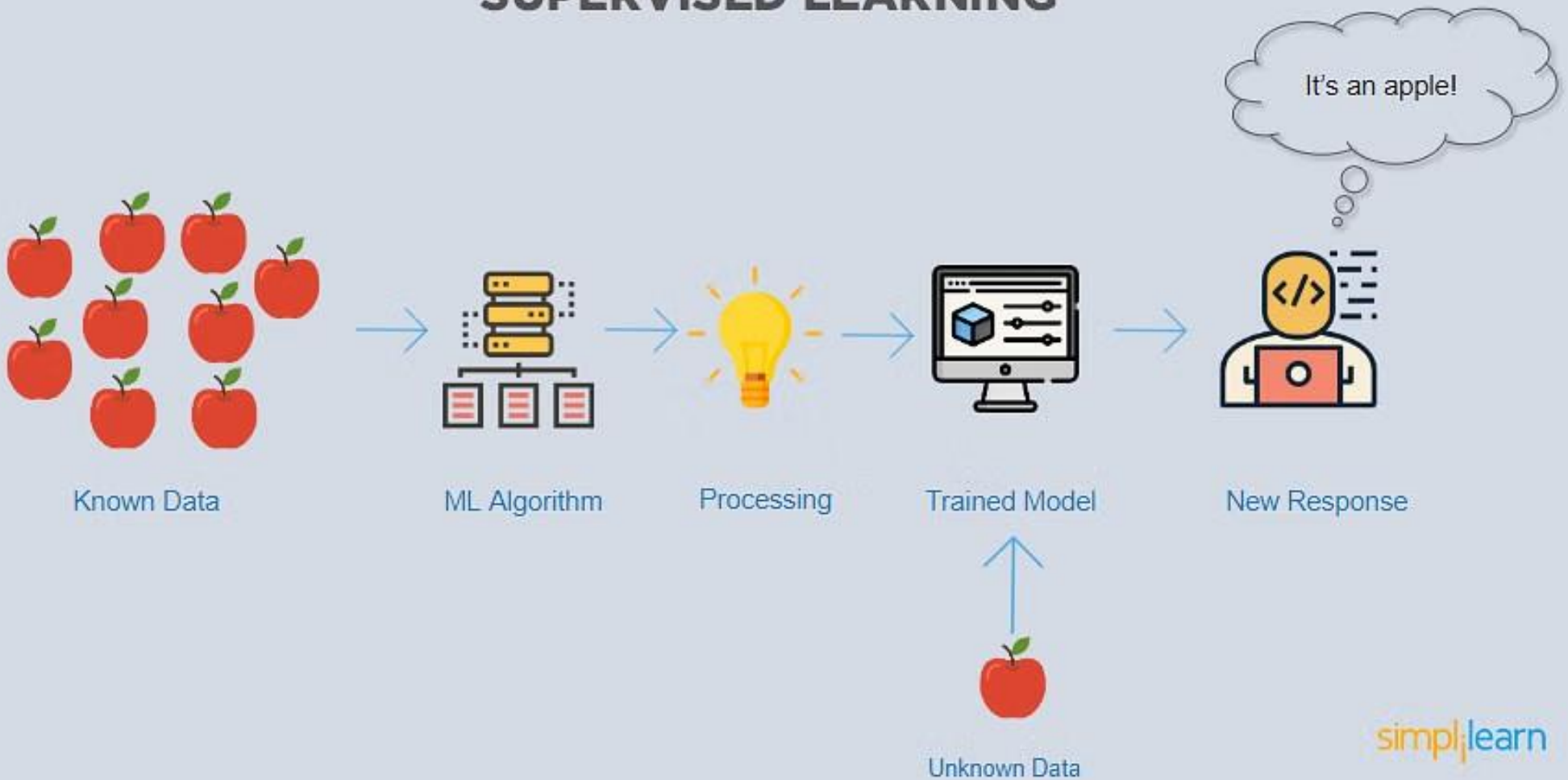
Supervised Learning

- Suppose we have an input dataset of cats and dog images. So, first, we will provide the training to the machine to understand the images, such as the *shape & size of the tail of cat and dog, Shape of eyes, colour, height (dogs are taller, cats are smaller), etc.*
- After completion of training, we input the picture of a cat and ask the machine to identify the object and predict the output.
- Now, the machine is well trained, so it will check all the features of the object, such as height, shape, colour, eyes, ears, tail, etc., and find that it's a cat. So, it will put it in the Cat category. This is the process of how the machine identifies the objects in Supervised Learning

Examples of Supervised Learning



SUPERVISED LEARNING



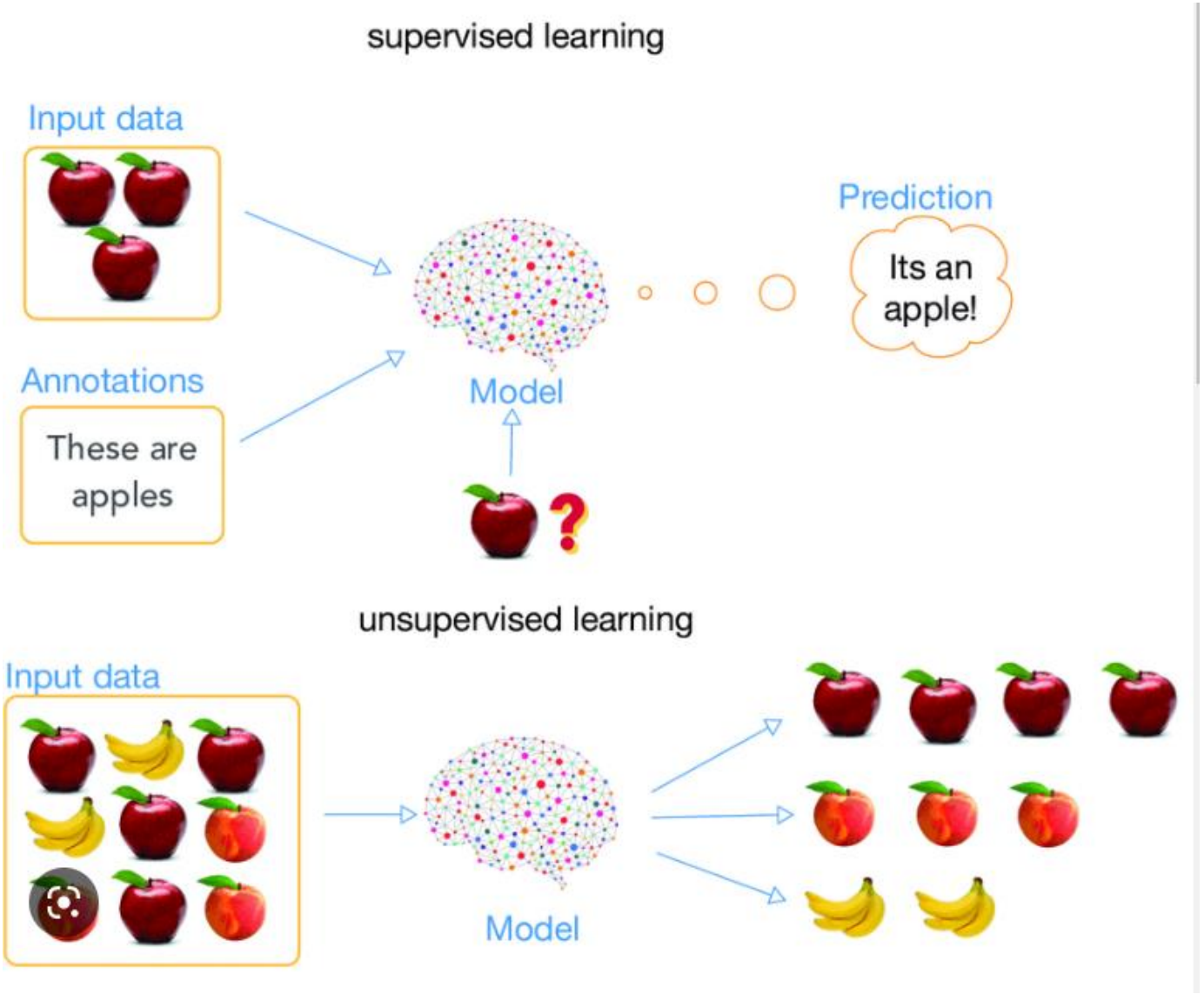
Advantages and Disadvantages of Supervised Learning

- **Advantages:**
  - Since supervised learning work with the labelled dataset so we can have an exact idea about the classes of objects.
  - These algorithms are helpful in predicting the output on the basis of prior experience.
- **Disadvantages:**
  - These algorithms are not able to solve complex tasks.
  - It may predict the wrong output if the test data is different from the training data.
  - It requires lots of computational time to train the algorithm.

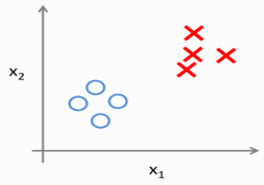
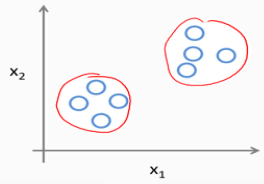
Applications of Supervised Learning

- **Image Segmentation:**
  Supervised Learning algorithms are used in image segmentation. In this process, image classification is performed on different image data with pre-defined labels.
- **Medical Diagnosis:**
  Supervised algorithms are also used in the medical field for diagnosis purposes. It is done by using medical images and past labelled data with labels for disease conditions. With such a process, the machine can identify a disease for the new patients.
- **Fraud Detection -** Supervised Learning classification algorithms are used for identifying fraud transactions, fraud customers, etc. It is done by using historic data to identify the patterns that can lead to possible fraud.
- **Spam detection -** In spam detection & filtering, classification algorithms are used. These algorithms classify an email as spam or not spam. The spam emails are sent to the spam folder.
- **Speech Recognition -** Supervised learning algorithms are also used in speech recognition. The algorithm is trained with voice data, and various identifications can be done using the same, such as voice-activated passwords, voice commands, etc.

Supervised & Unsupervised Learning



Difference between Supervised and Unsupervised Learning

Supervised	Unsupervised
Input Data is labelled	Input Data is Unlabelled
Uses training Dataset	Uses just input dataset
Data is classified based on training dataset	Uses properties of given data to classify it.
Used for prediction	Used for Analysis
Divided into two types Regression & Classification	Divided into two types Clustering & Association
Known number of classes	Unknown number of classes
	
Use off-line analysis of data	Use Real-Time analysis of data

Categories of Supervised Learning

- **Classification**
  - Classification refers to taking an input value and mapping it to a discrete value. In classification problems, our output typically consists of classes or categories. This could be things like trying to predict what objects are present in an image (a cat/ a dog) or whether it is going to rain today or not.
- **Regression**
  - Regression is related to continuous data (value functions). In Regression, the predicted output values are real numbers. It deals with problems such as predicting the price of a house or the trend in the stock price at a given time, etc.

Classification

- Classification algorithms are used to solve the classification problems in which the output variable is categorical, such as "**Yes" or No, Male or Female, Red or Blue, etc**.
- Classification refers to taking an input value and mapping it to a discrete value. In classification problems, our output typically consists of classes or categories. This could be things like trying to predict what objects are present in an image (a cat/ a dog) or whether it is going to rain today
- The classification algorithms predict the categories present in the dataset. Some real-world examples of classification algorithms are **Spam Detection, Email filtering, etc.**

Regression

- Regression is related to continuous data (value functions). In Regression, the predicted output values are real numbers. It deals with problems such as predicting the price of a house or the trend in the stock price at a given time, etc.
- It predicts the continuous output variables based on the independent input variable. like the prediction of house prices based on different parameters like house age, distance from the main road, location, area, etc.

enjoy algorithms

Regression

What will be the temperature tomorrow?

Fahrenheit

Classification

Will it be hot or cold tomorrow?

Fahrenheit

Unsupervised Learning

- Unsupervised learning is different from the Supervised learning technique; as its name suggests, there is no need for supervision. It means, in unsupervised machine learning, the machine is trained using the unlabeled dataset, and the machine predicts the output without any supervision.
- The main aim of the unsupervised learning algorithm is to group or categories the unsorted dataset according to the similarities, patterns, and differences. Machines are instructed to find the hidden patterns from the input dataset.

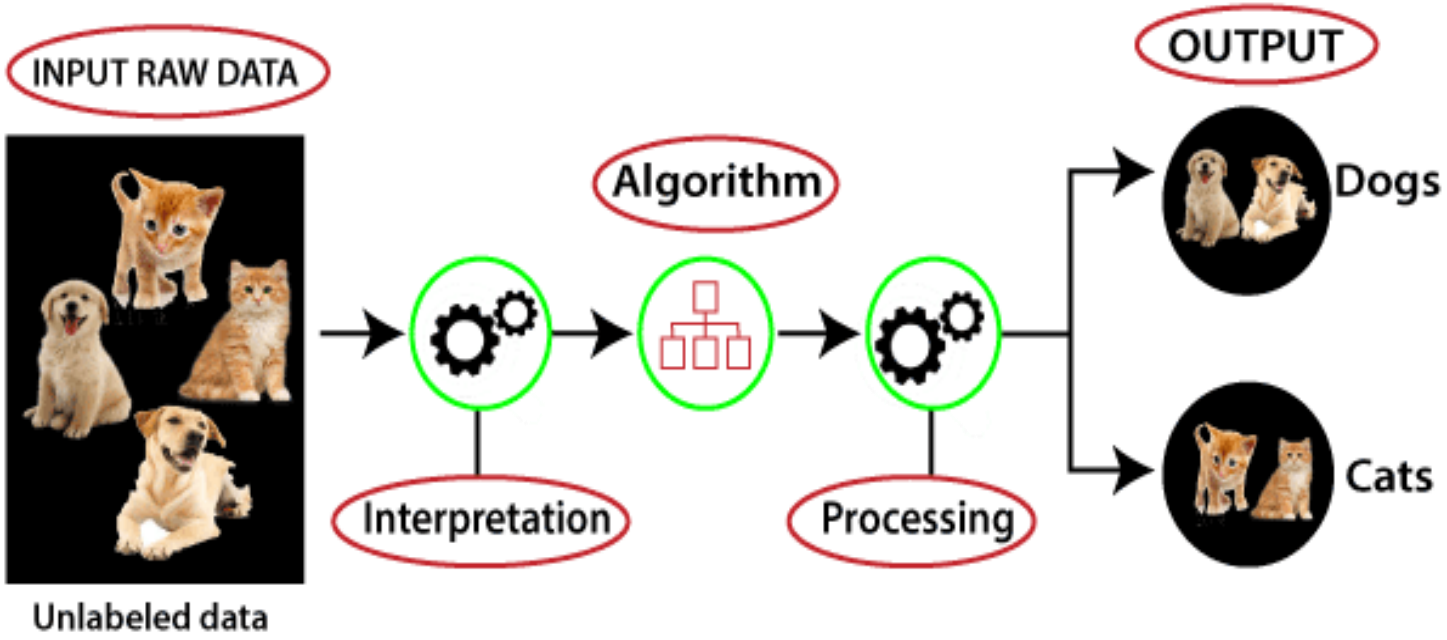
    - suppose there is a basket of fruit images, and we input it into the machine learning model. The images are totally unknown to the model, and the task of the machine is to find the patterns and categories of the objects.
    - So, now the machine will discover its patterns and differences, such as colour difference, shape difference, and predict the output when it is tested with the test dataset.

Categories of Unsupervised Machine Learning

1. Clustering
2. Association

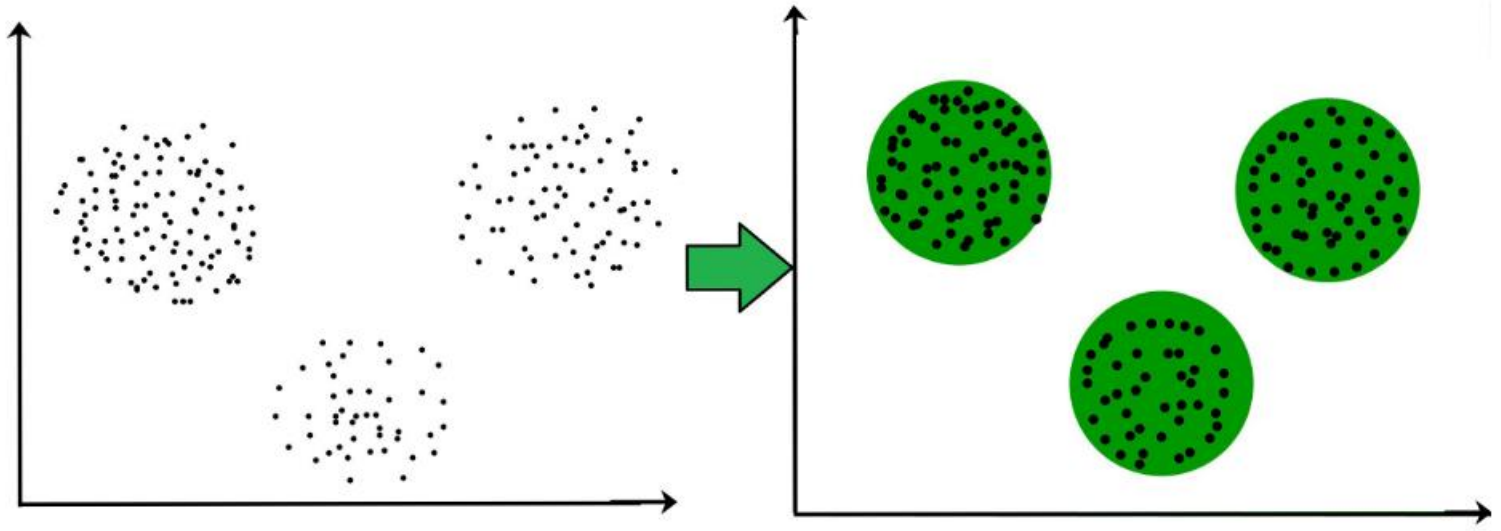
Unsupervised Learning

- This type of algorithm consists of input data without labelled response. There will not be any pre existing labels and human intervention is also less. It is mostly used in exploratory analysis as it can automatically identify the structure in data.

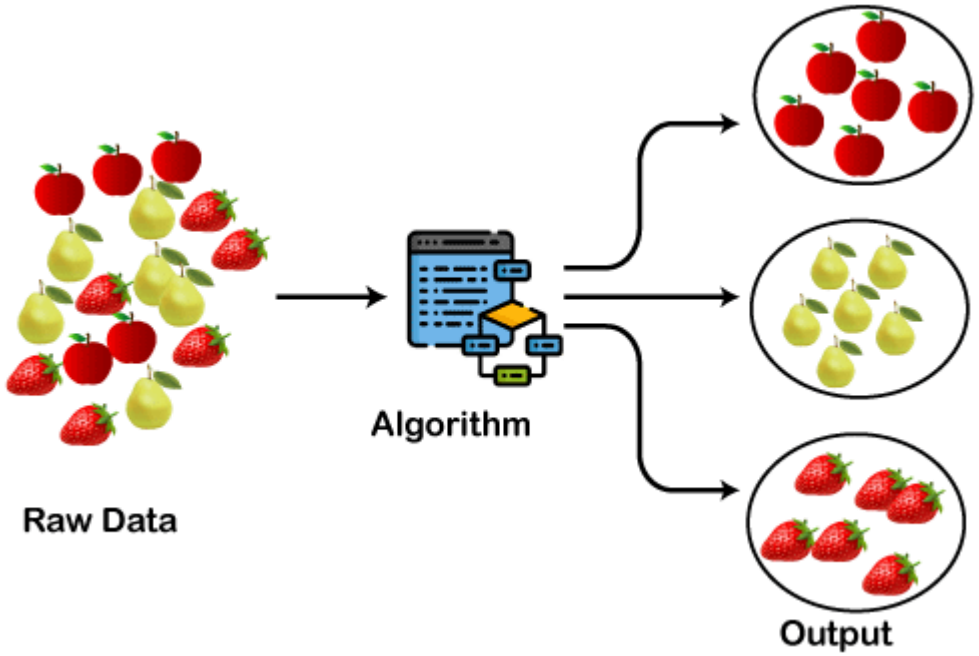


Clustering

- Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.



Clustering



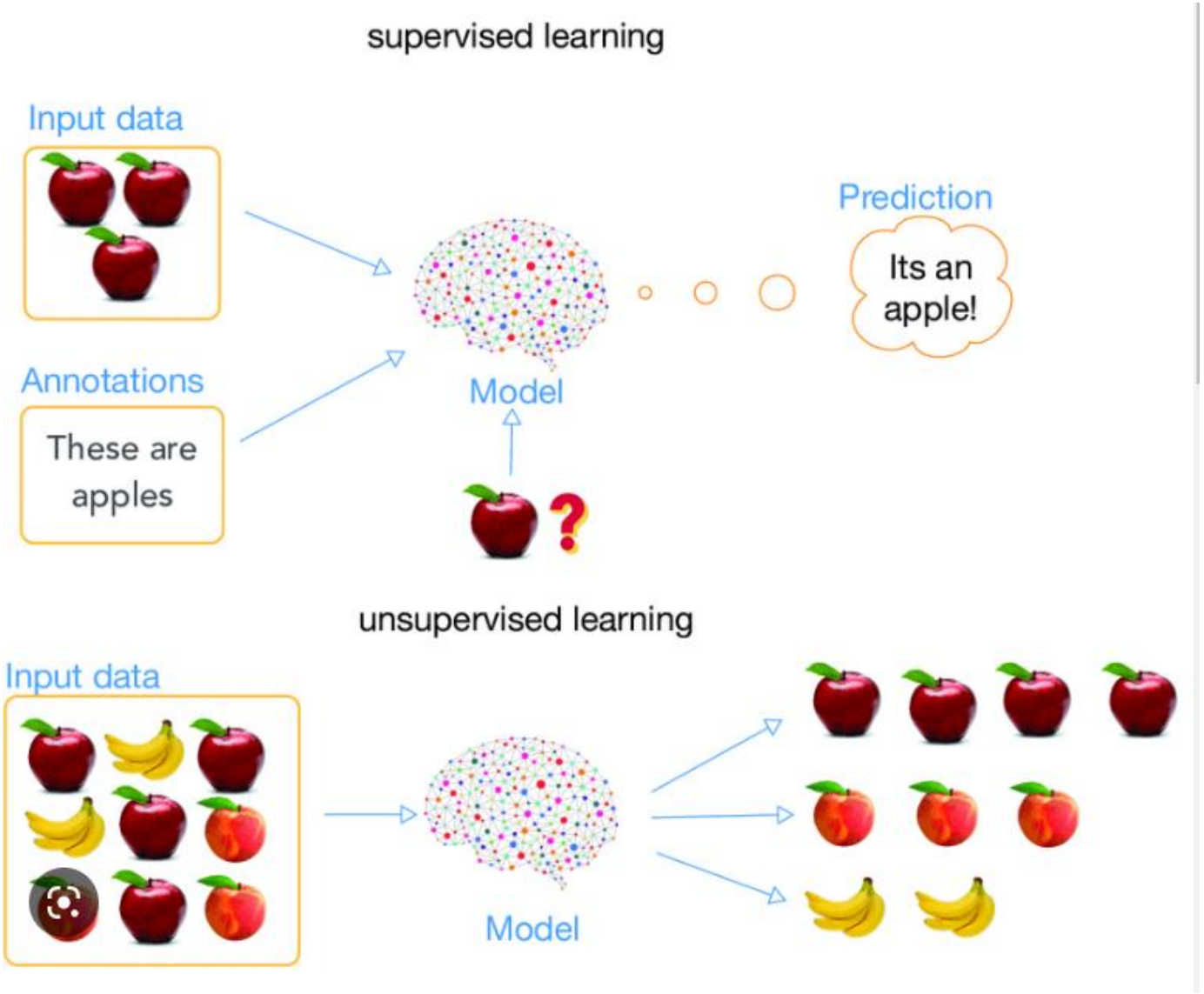
Advantages and Disadvantages of Unsupervised Learning

- **Advantages:**
  - These algorithms can be used for complicated tasks compared to the supervised ones because these algorithms work on the unlabeled dataset.
  - Unsupervised algorithms are preferable for various tasks as getting the unlabeled dataset is easier as compared to the labelled dataset.
- **Disadvantages:**
  - The output of an unsupervised algorithm can be less accurate as the dataset is not labelled, and algorithms are not trained with the exact output in prior.
  - Working with Unsupervised learning is more difficult as it works with the unlabelled dataset that does not map with the output.

Applications of Unsupervised Learning

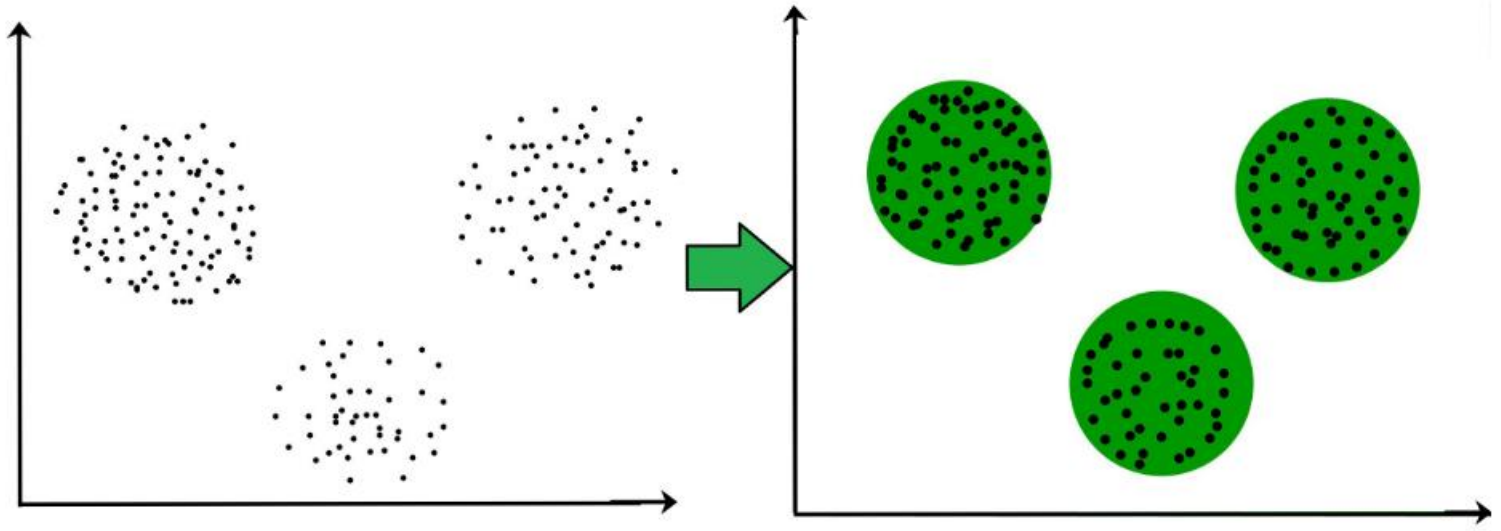
- **Network Analysis:** Unsupervised learning is used for identifying plagiarism and copyright in document network analysis of text data for scholarly articles.
- **Recommendation Systems:** Recommendation systems widely use unsupervised learning techniques for building recommendation applications for different web applications and e-commerce websites.
- **Anomaly Detection:** Anomaly detection is a popular application of unsupervised learning, which can identify unusual data points within the dataset. It is used to discover fraudulent transactions.
- **Singular Value Decomposition:** Singular Value Decomposition or SVD is used to extract particular information from the database. For example, extracting information of each user located at a particular location.

Supervised & Unsupervised Learning

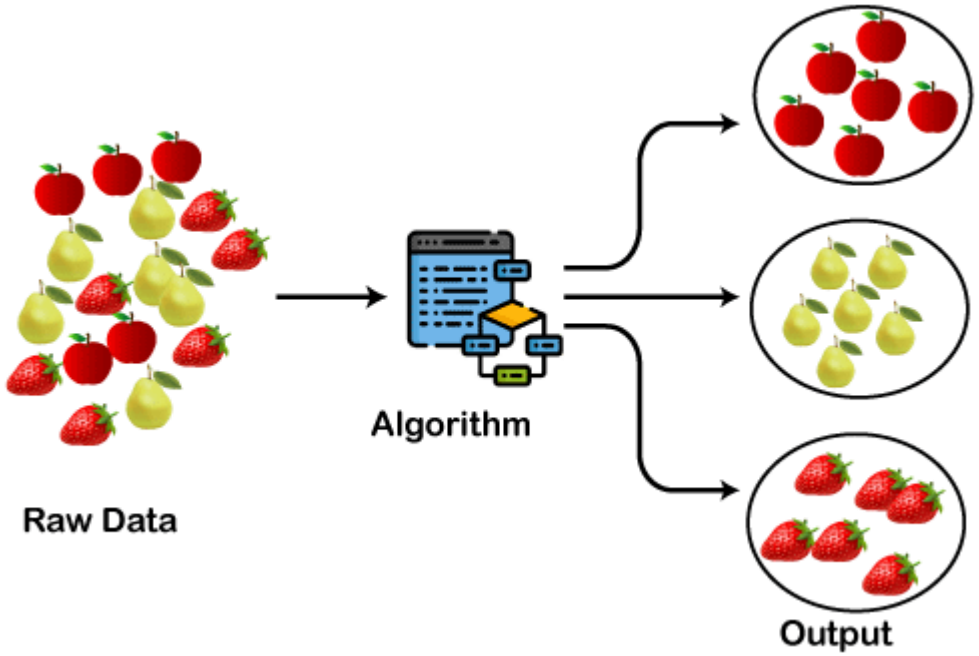


Clustering

- Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

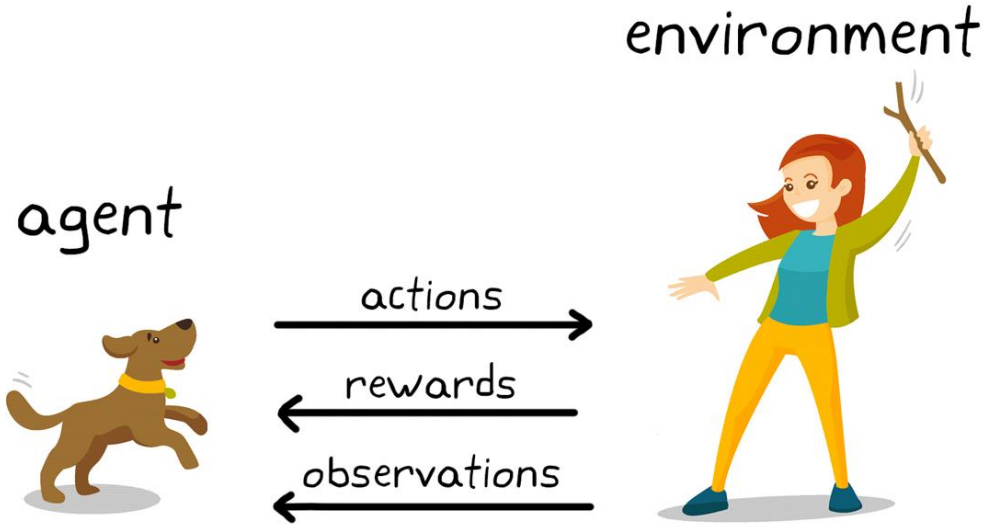


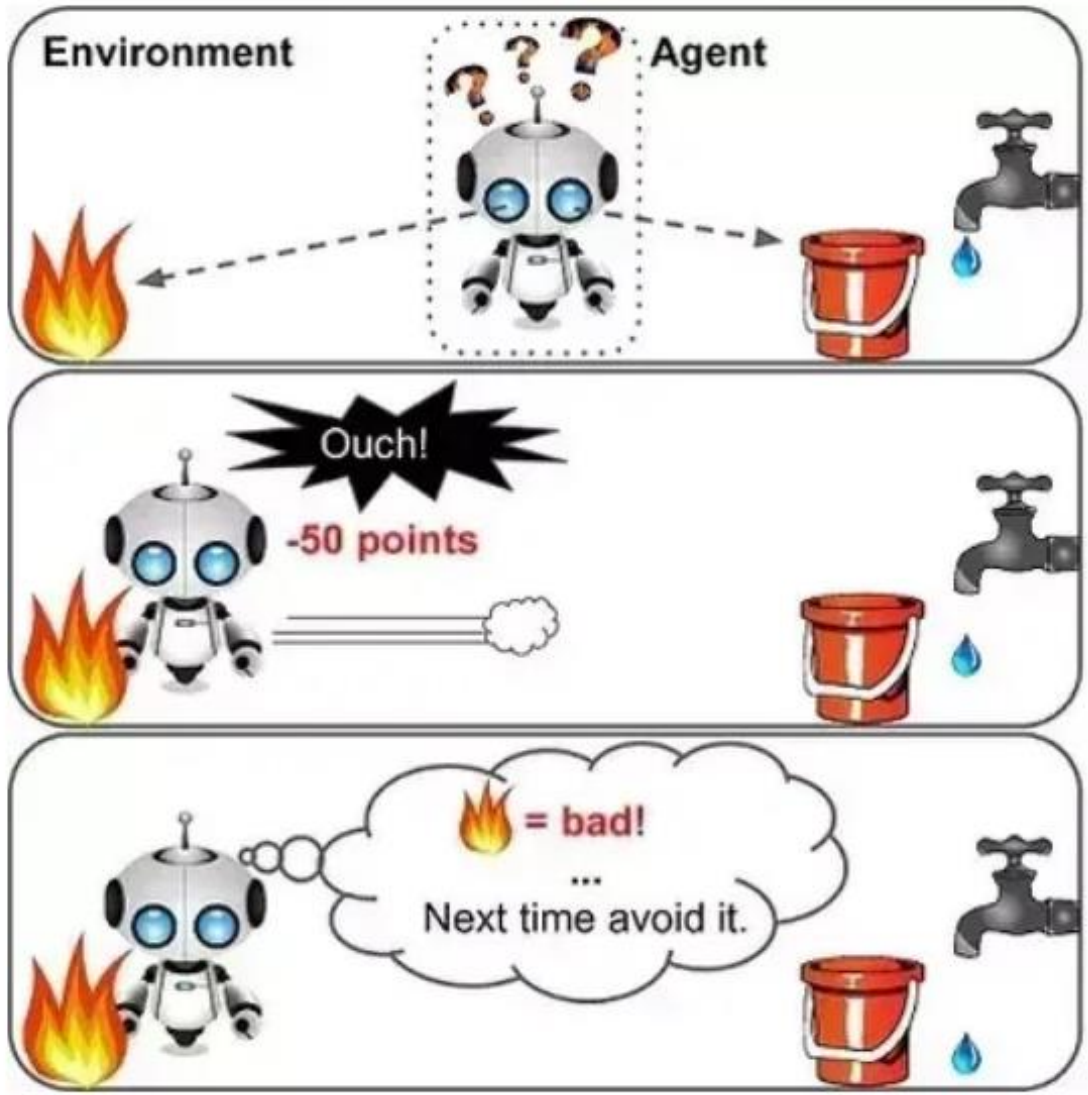
Clustering



Reinforcement Learning

- This model is used in making a sequence of decisions. It is an learning by interacting with the environment. It is based on the observation that intelligent agents tend to repeat the action that are rewarded for and refrain from action that are punished for. It can be said that it is an trail and error method in finding the best outcome based on experience.



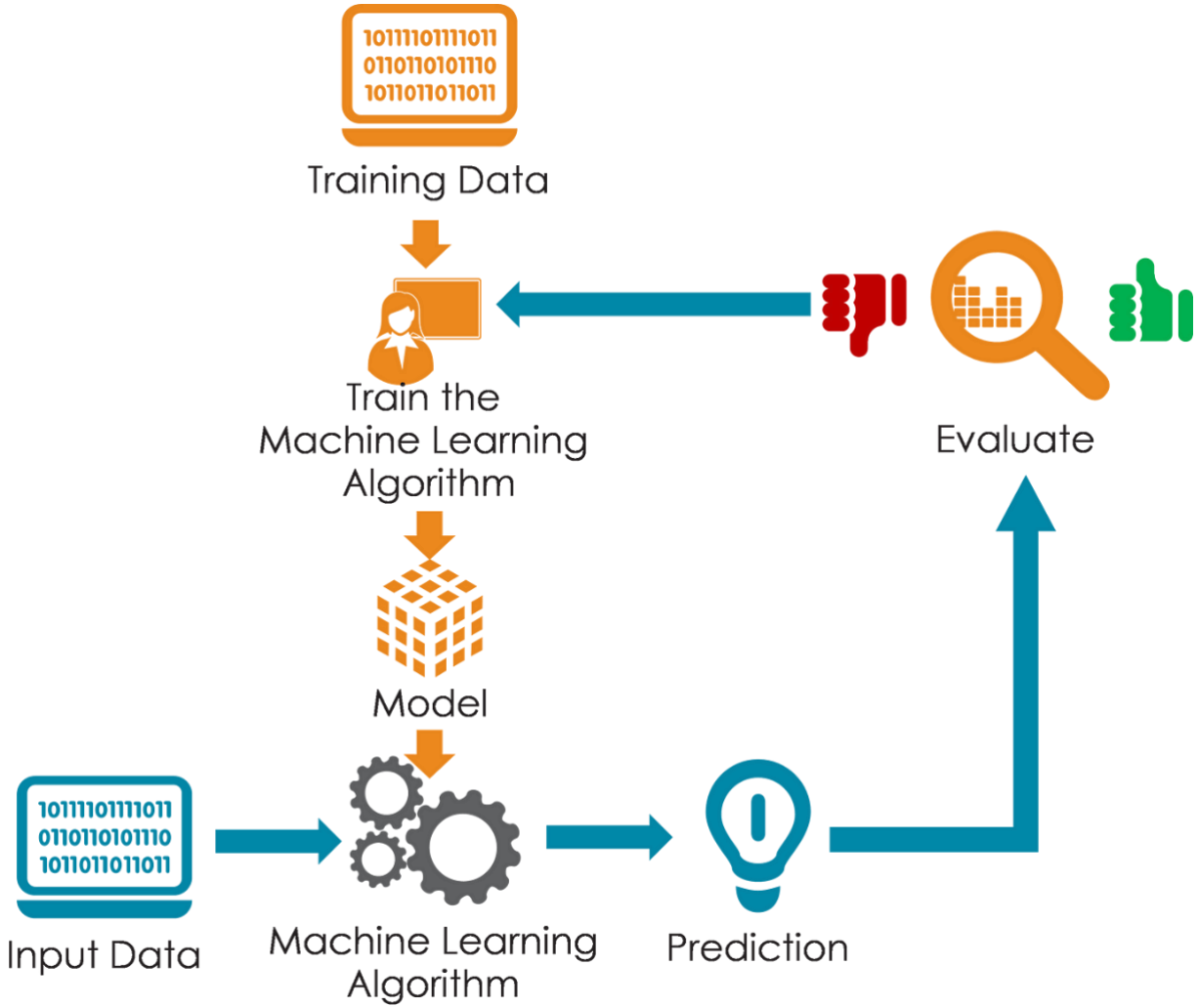


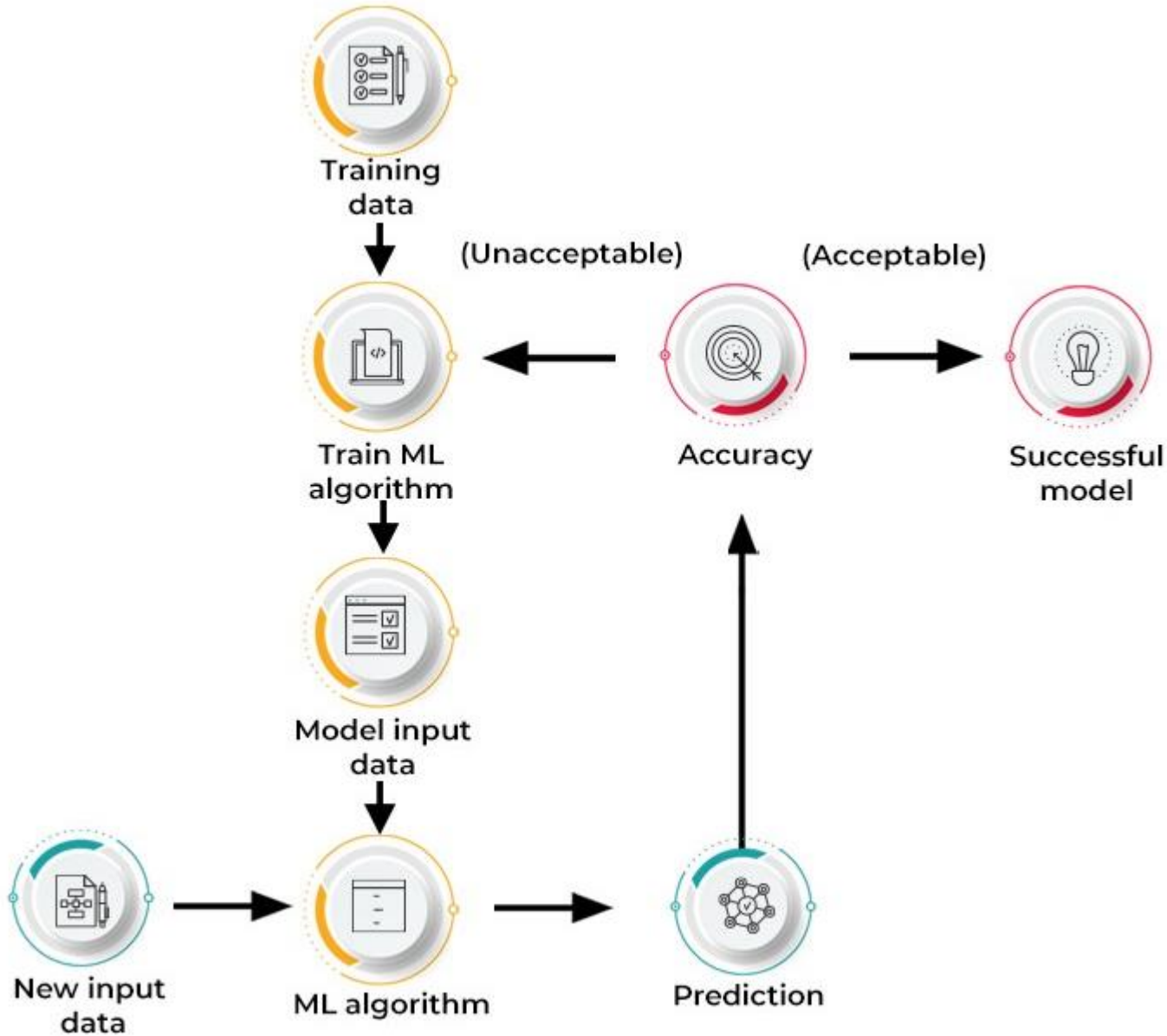
1. Observe
2. Select action using policy
3. Action!
4. Get reward or penalty
5. Update policy (learning step)
6. Iterate until an optimal policy is found

Reinforcement Learning

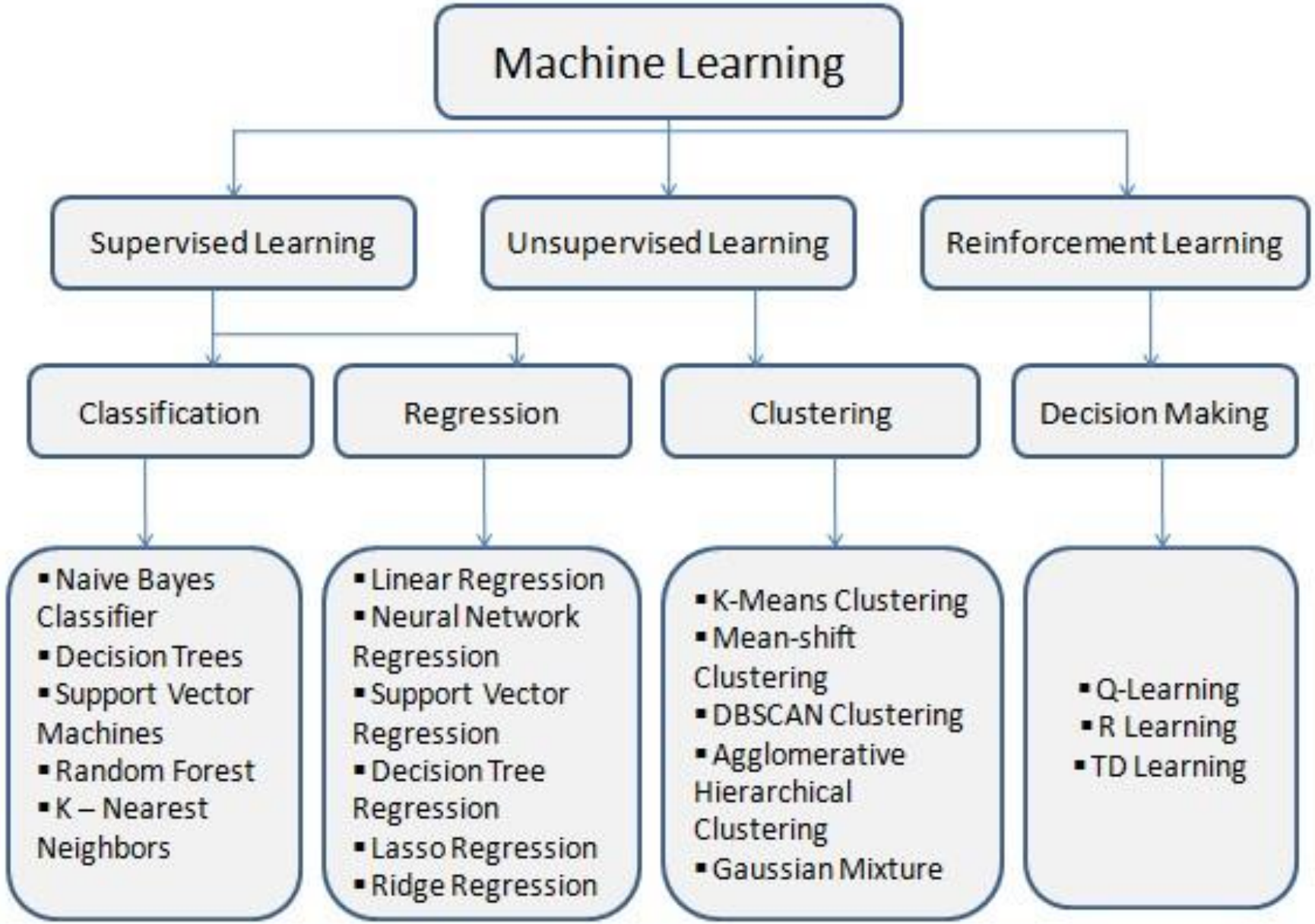
- Reinforcement learning is a feedback-based learning method, in which a learning agent gets a reward for each right action and gets a penalty for each wrong action.
- The agent learns automatically with these feedbacks and improves its performance. In reinforcement learning, the agent interacts with the environment and explores it. The goal of an agent is to get the most reward points, and hence, it improves its performance.
- The robotic dog, which automatically learns the movement of his arms, is an example of Reinforcement learning.

How does Machine Learning Work?



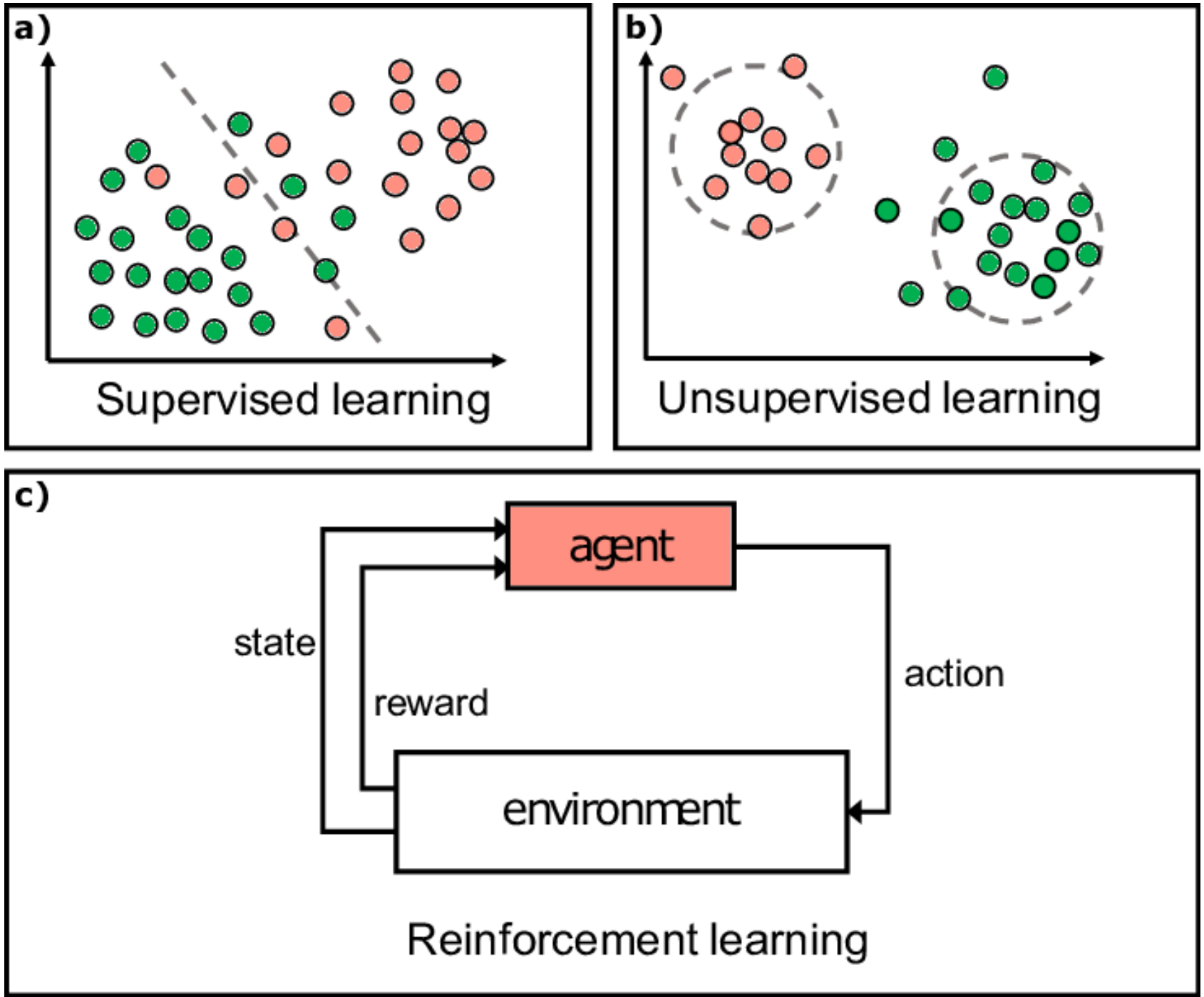

Training data
(Unacceptable)
(Acceptable)
Train ML algorithm
Accuracy
Successful model
Model input data
New input data
ML algorithm
Prediction

- Machine learning is a subset of AI, which enables the machine to automatically learn from data, improve performance from past experiences, and make predictions.
- Machine learning contains a set of algorithms that work on a huge amount of data. Data is fed to these algorithms to train them, and on the basis of training, they build the model & perform a specific task.
- ML algorithms help to solve different business problems like Regression, Classification, Forecasting, Clustering, and Associations, etc.


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Classification
Regression
Clustering
Decision Making
▪ Naive Bayes Classifier
▪ Decision Trees
▪ Support Vector Machines
▪ Random Forest
▪ K – Nearest Neighbors
▪ Linear Regression
▪ Neural Network Regression
▪ Support Vector Regression
▪ Decision Tree Regression
▪ Lasso Regression
▪ Ridge Regression
▪ K-Means Clustering
▪ Mean-shift Clustering
▪ DBSCAN Clustering
▪ Agglomerative Hierarchical Clustering
▪ Gaussian Mixture
▪ Q-Learning
▪ R Learning
▪ TD Learning

Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.


a)
Supervised learning
b)
Unsupervised learning
c)
agent
state
reward
action
environment
Reinforcement learning

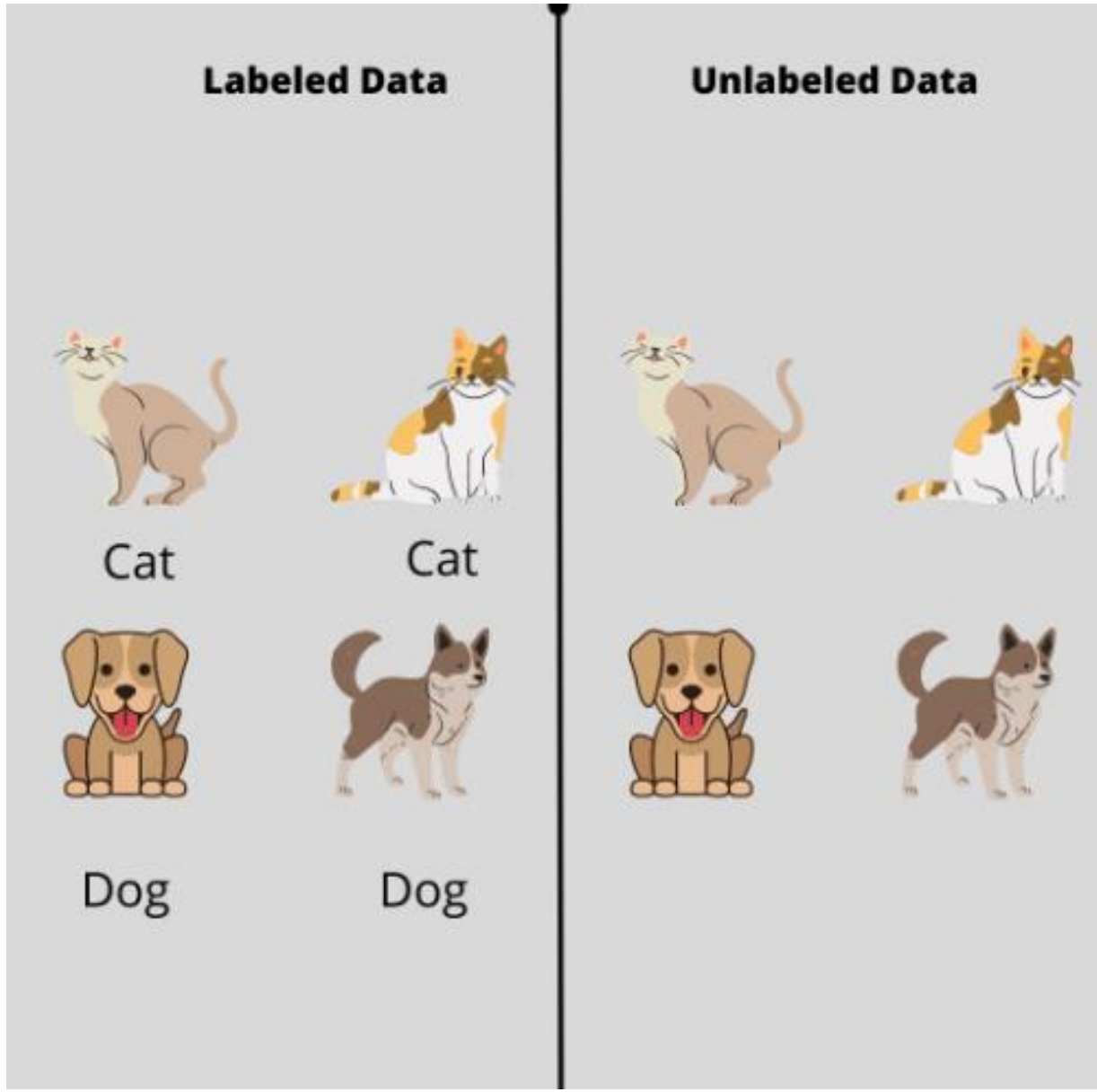
Labelled and Unlabeled data

Labeled data

- Labeled data is data that has some predefined tags such as name, type, or number. For example, an image has an apple or banana. At the same time, unlabelled data contains no tags or no specified name. Labeled data is used in Supervised Learning techniques, whereas Unlabelled data is used in Unsupervised Learning.
- Any data which has a characteristic, category, or attributes assigned to it can be referred to as labeled data. For example, a photo of a cat, the height of a human, price of a product is some examples of labeled data.

Unlabeled Data

- Any data that does not have any labels specifying its characteristics, identity, classification, or properties can be considered unlabeled data.
- For example photos, videos, or text that do not have any category or classification assigned to it can be referred to as unlabeled data.


Labeled Data
Unlabeled Data
Cat
Cat
Dog
Dog

How to efficiently apply data labeling?

- The key to building successful and efficient Machine Learning models is to continuously feed them with a massive amount of high-quality data. Over time, the model will get better and better at making accurate predictions.
- To begin with, Data scientists need to train the model that is labeled by humans. The model will start applying labels automatically to all the data it understands and will pass back the rest of the data, that it does not understand back to humans for annotation.
- The returned data is once again fed into the model to retrain and improve its capability to automatically assign labels to new data. Over time, the machine will become proficient in labeling most of the data on its own without requiring much supervision.

Training and Testing Data in Machine Learning

Training and Testing Data

- Training data and test data sets are two different but important parts in machine learning. While training data is necessary to teach an ML algorithm, testing data, as the name suggests, helps you to validate the progress of the algorithm's training and adjust or optimize it for improved results.

- Machine learning algorithms learn from data, so having the correct data is critical to building successful models. Training and test data sets help us evaluate our models' performance.

Training Data

- it’s a portion of our actual dataset that is fed into the machine learning model to discover and learn patterns. In this way, it trains our model.
- Training data is typically larger than testing data. This is because we want to feed the model with as much data as possible to find and learn meaningful patterns. Once data from our datasets are fed to a machine learning algorithm, it learns patterns from the data and makes decisions.
- Algorithms enable machines to solve problems based on past observations. Kind of like learning from example, just like humans. The only difference is that machines require a lot more examples in order to be able to see patterns and learn. As machine learning models are exposed to more relevant training data, the more they improve over time.

Testing Data

- Once your machine learning model is built (with your training data), you need unseen data to test your model. This data is called testing data, and you can use it to evaluate the performance and progress of your algorithms' training and adjust or optimize it for improved results.
- Testing data has two main criteria. It should:
    - Represent the actual dataset
    - Be large enough to generate meaningful predictions
- This dataset needs to be new, "unseen" data. This is because your model already "knows" the training data. How it performs on new test data will let you know if it's working accurately or if it requires more training data to perform to your specifications.
- Test data provides a final, real-world check of an unseen dataset to confirm that the machine learning algorithm was trained effectively.
- In data science, it's typical to see your data split into 80% for training and 20% for testing.

How Much Training Data do You Need

- The complexity of the problem
- The complexity of the learning algorithm

Features in Machine Learning

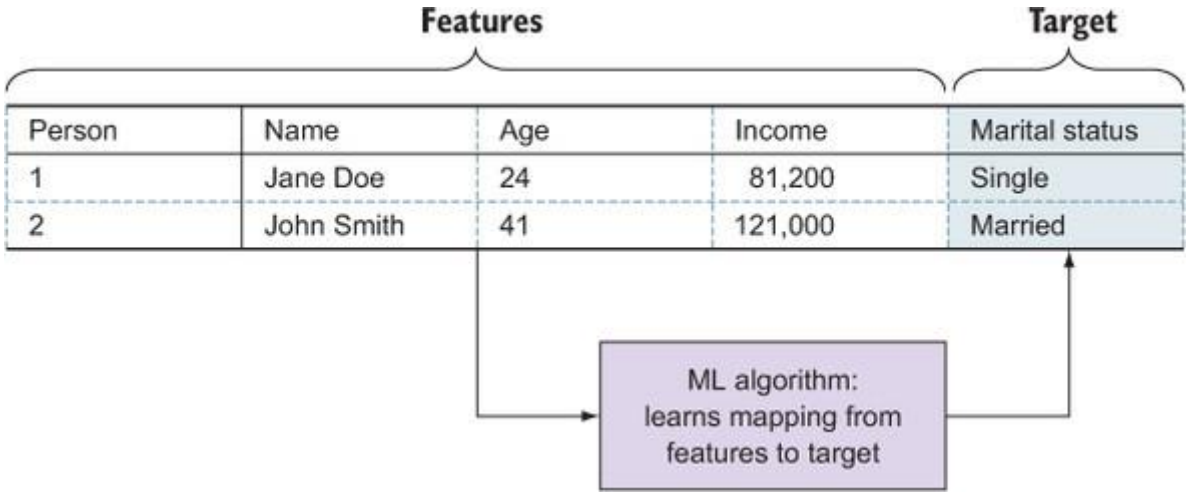
Features

- In machine learning, features are individual independent variables that act like a input in your system. While making the predictions, models use such features to make the predictions. And using the feature engineering process, new features can also be obtained from old features in machine learning.
- Features in machine learning is very important, being building a blocks of datasets, the quality of the features in your dataset has major impact on the quality of the insights you will get while using the dataset for machine learning.
- Feature engineering is the process of using the domain knowledge of the data to create features that makes machine learning algorithms work properly. If feature engineering is performed properly, it helps to improve the power of prediction of machine learning algorithms by creating the features using the raw data that facilitate the machine learning process.

Features

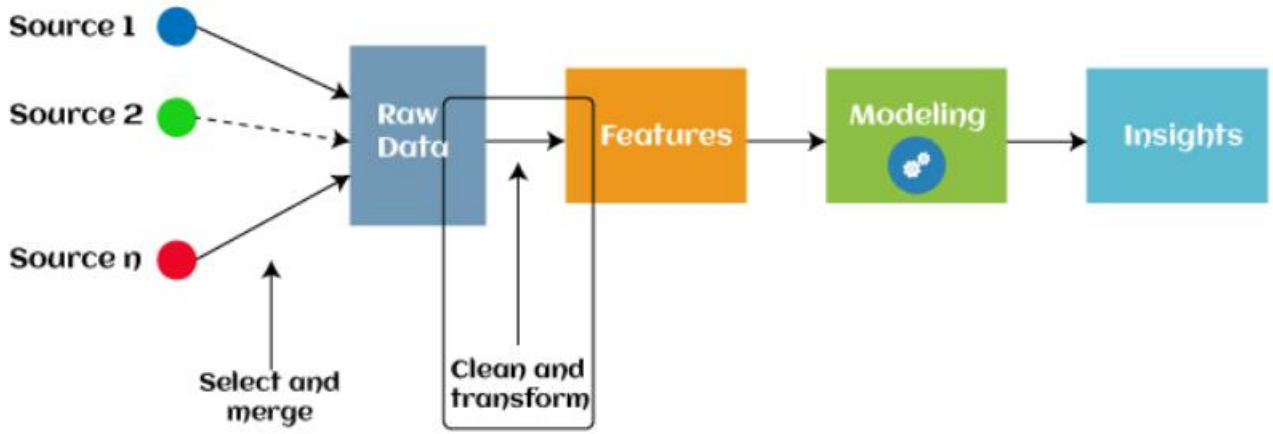
- Feature engineering is the most important part of machine leaning that makes difference between and good and bad model. And there are several steps involved in feature engineering (To be studied in data mining)

- Examples of numerical features include age, height, weight, and income. Numerical features can be used in machine learning algorithms directly. Categorical features are discrete values that can be grouped into categories. Examples of categorical features include gender, color, and zip code.



Feature Engineering

- Feature engineering is the pre-processing step of machine learning, which is used to transform raw data into features that can be used for creating a predictive model using Machine learning or statistical Modelling. Feature engineering in machine learning aims to improve the performance of models.

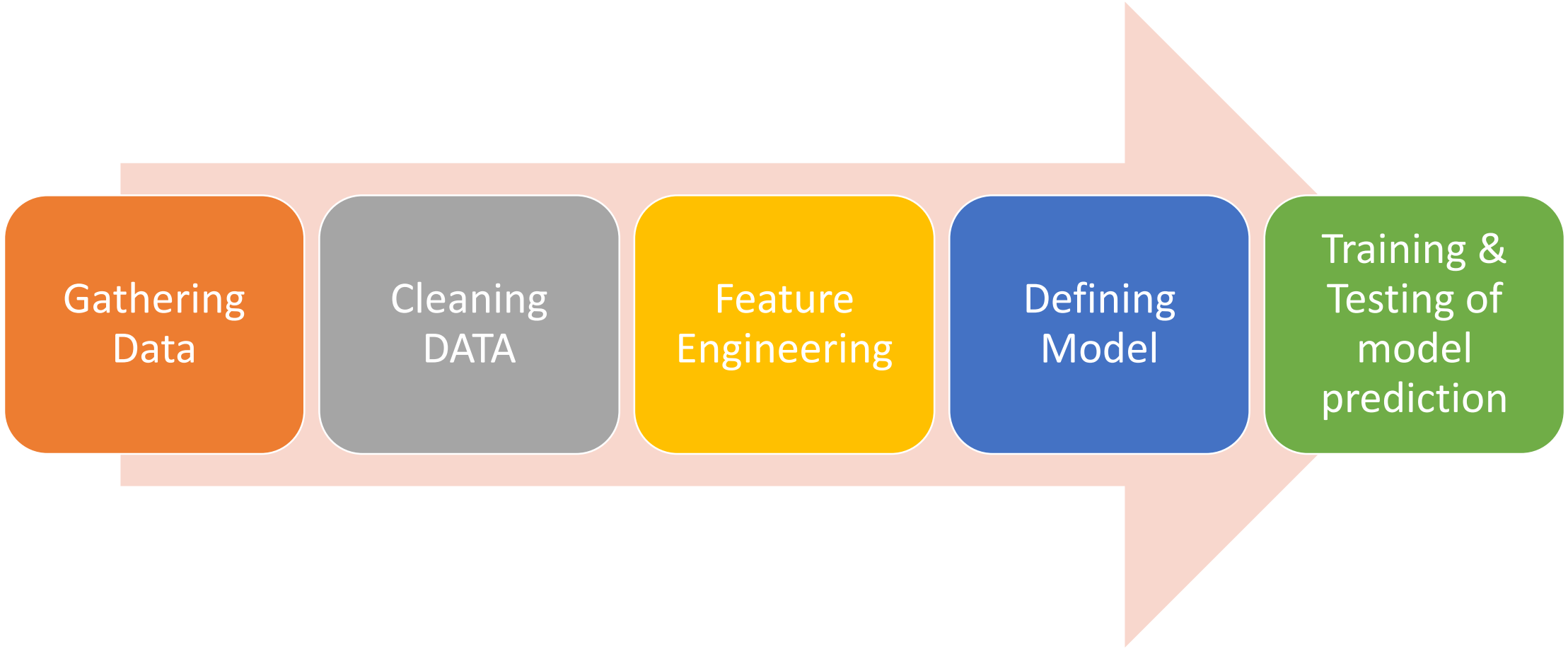


- It is very important to chose features that work well and ignore features that do not offer sufficient benefits. Adding more features to a machine learning model can have both positive and negative effects on the accuracy of the model.

- The accuracy of the prediction will enhance as we keep adding more and more features, however, we may need to take a conscious call on how far this has to be taken as every additional feature will make the computation expensive as we deal with a huge volume of data along with complex mathematical computations.

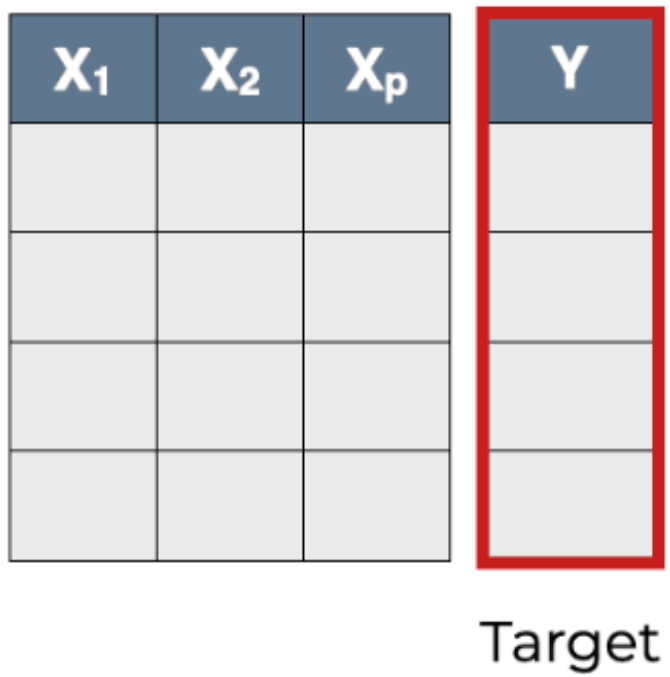
- To perform the feature engineering in machine learning you need data experts like data scientists or hire machine learning engineer who can understand and perform the feature engineering process with right instructions.

Steps To Do Feature Engineering in ML

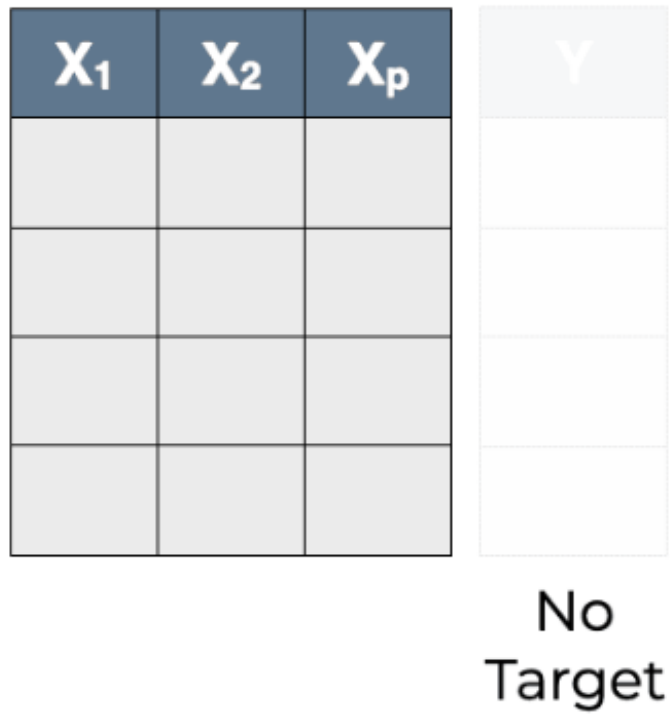


Supervised Vs Unsupervised Learning, Explained

Supervised



Un-Supervised



User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

Figure A: CLASSIFICATION

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

Figure B: REGRESSION

Thank You!