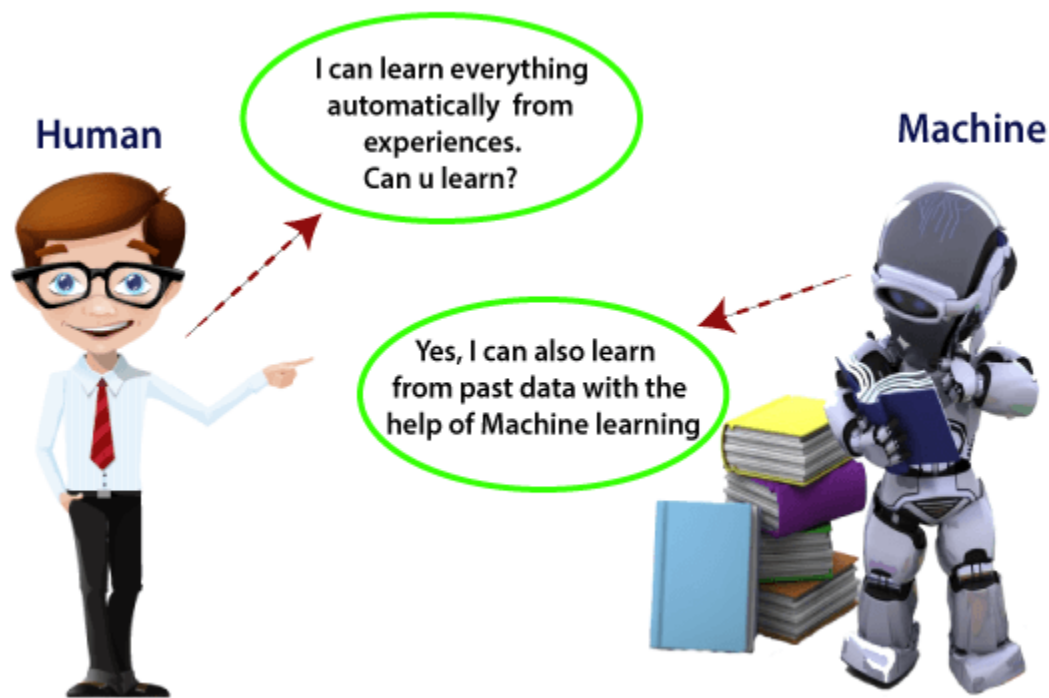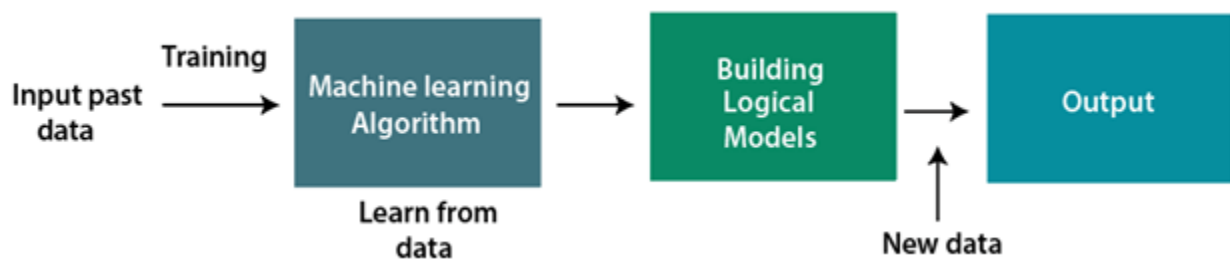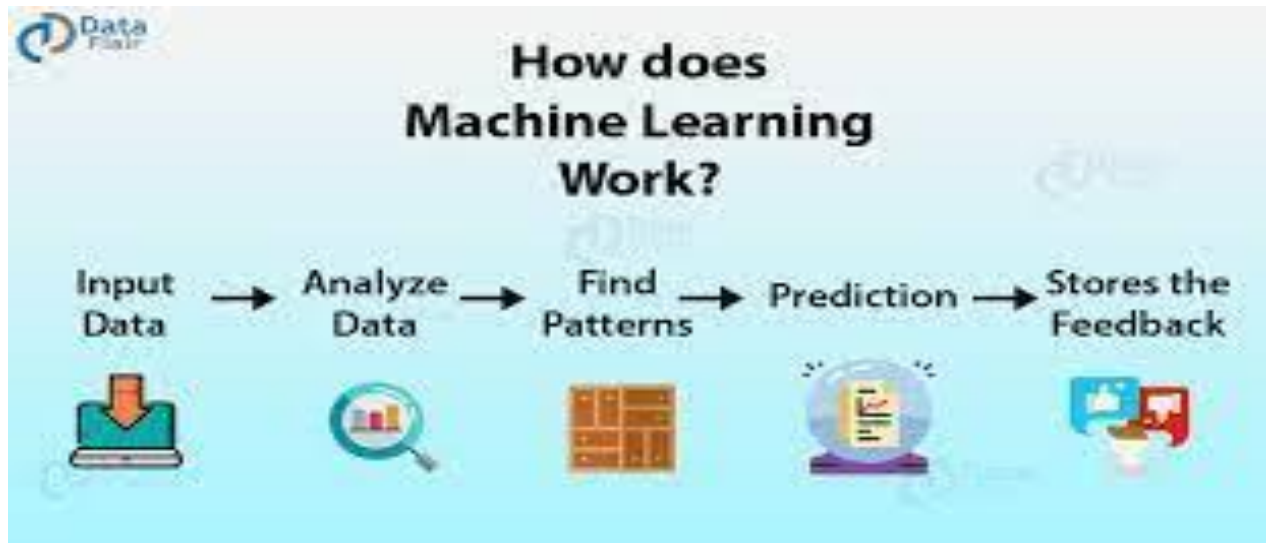# MACHINE LEARNING

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for **building mathematical models and making predictions using historical data or information**. Currently, it is being used for various tasks such as **image recognition**, **speech recognition**, **email filtering**, **Facebook auto-tagging**, **recommender system**, and many more.

Machine Learning is said as a subset of **artificial intelligence** that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. The term machine learning was first introduced by **Arthur Samuel** in **1959**. We can define it in a summarized way as

"Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed."

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.
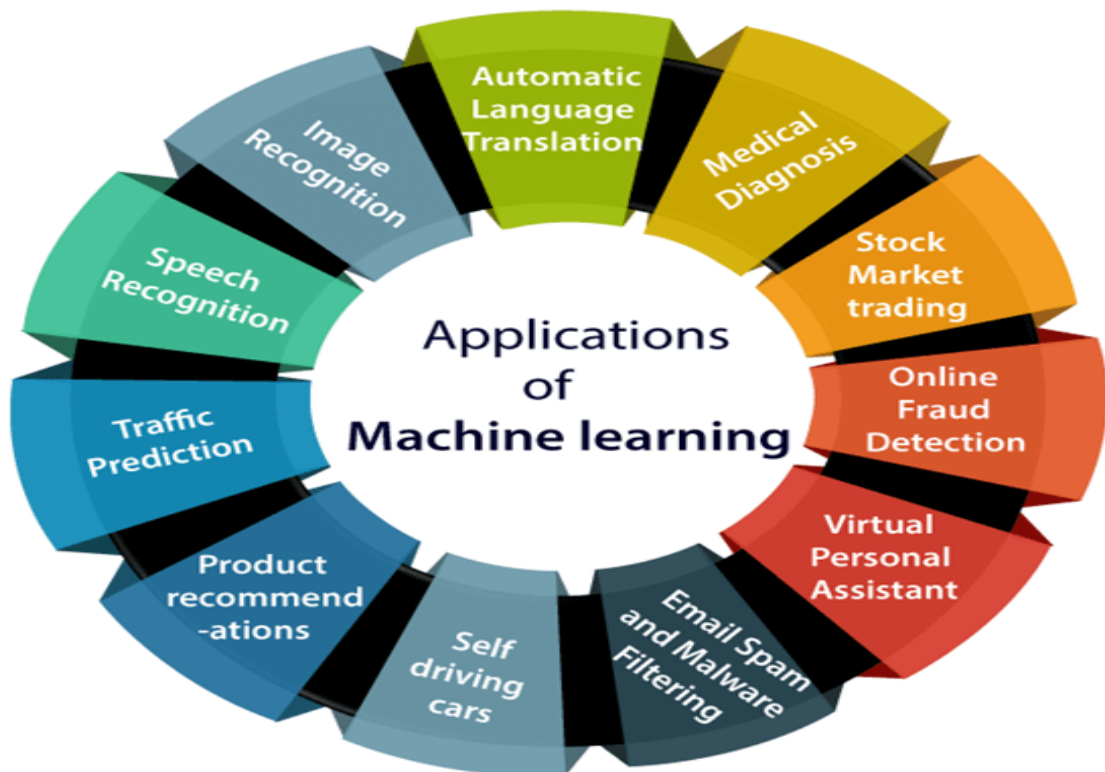
Importance of Machine Learning:

➢ Rapid increment in the production of data

➢ Solving complex problems, which are difficult for a human

➢ Decision making in various sector including finance

➢ Finding hidden patterns and extracting useful information from data.

7 Steps of Machine Learning

1. Gathering Data
2. Preparing that Data
3. Choosing a Model
4. Training
5. Evaluation
6. Hyperparameter Tuning
7. Prediction

This slide is 100% editable. Adapt it to your needs and capture your audience's attention

Applications of Machine learning

- Automatic Language Translation
- Medical Diagnosis
- Stock Market trading
- Online Fraud Detection
- Virtual Personal Assistant
- Email Spam and Malware Filtering
- Self driving cars
- Product recommend-ations
- Traffic Prediction
- Speech Recognition
- Image Recognition

Types of Machine Learning:


```
                        ┌─────────────────┐
                        │ Machine Learning │
                        └─────────────────┘
```

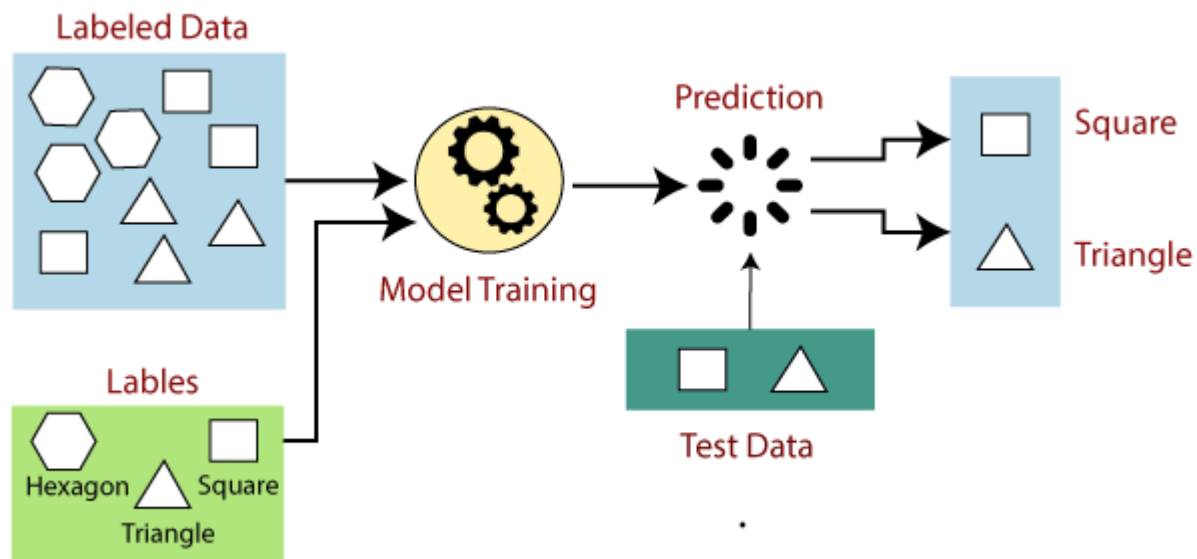| Supervised Learning	Unsupervised Learning	Semi-supervised Learning	Reinforcement Learning
Input data is called traning data and has a known label.	Input data is not labeled. Extract general structures.	Input data is mixture of laeled and unlabeled datal.	Reward feedback is required to learn its behavior.
K-nearest neighbor Decision tree Neural network Support vector machine	K-means clustering Self-organizing map Principal Component Analysis	Maximum likelilhood learning Generative models	Q-learning Multi-armed bandits Markov decision process
Traffic classfication Intrusion detection Routing optimization	Traffic clustering Filtering algorithms	Resource management Network behavior analysis	Network reconfiguration Resource management Decision-making

1.Supervised learning:

Supervised machine learning is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output. Here, the labelled data specifies that some of the inputs are already mapped to the output. More preciously, we can say; first, we train the machine with the input and corresponding output, and then we ask the machine to predict the output using the test dataset.

Types of supervised Machine learning Algorithms:
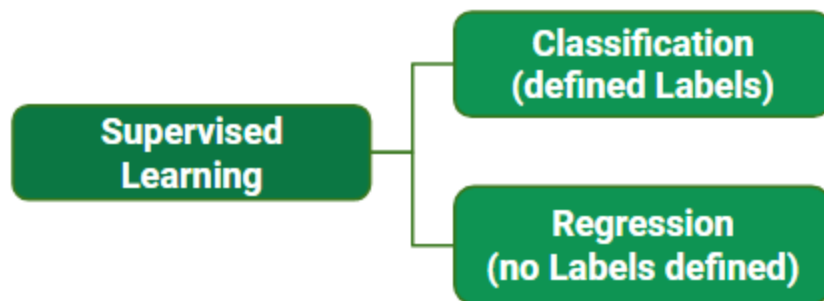
a) Classification

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. The classification algorithms predict the categories present in the dataset. Some real-world examples of classification algorithms are Spam Detection, Email filtering, etc.

Some popular classification algorithms are given below:

➢ **Random Forest Algorithm**
➢ **Decision Tree Algorithm**

- ➢ **Logistic Regression Algorithm**

- ➢ **Support Vector Machine Algorithm**

b) Regression

Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc.

Some popular Regression algorithms are given below:

- ➢ **Simple Linear Regression Algorithm**

- ➢ **Multivariate Regression Algorithm**

- ➢ **Decision Tree Algorithm**

- ➢ **Lasso Regression**

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

Some common applications of Supervised Learning are given below:

- ➤ **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.

## 2. Unsupervised Machine 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.

In unsupervised learning, the models are trained with the data that is neither classified nor labelled, and the model acts on that data 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.

Unlabeled data
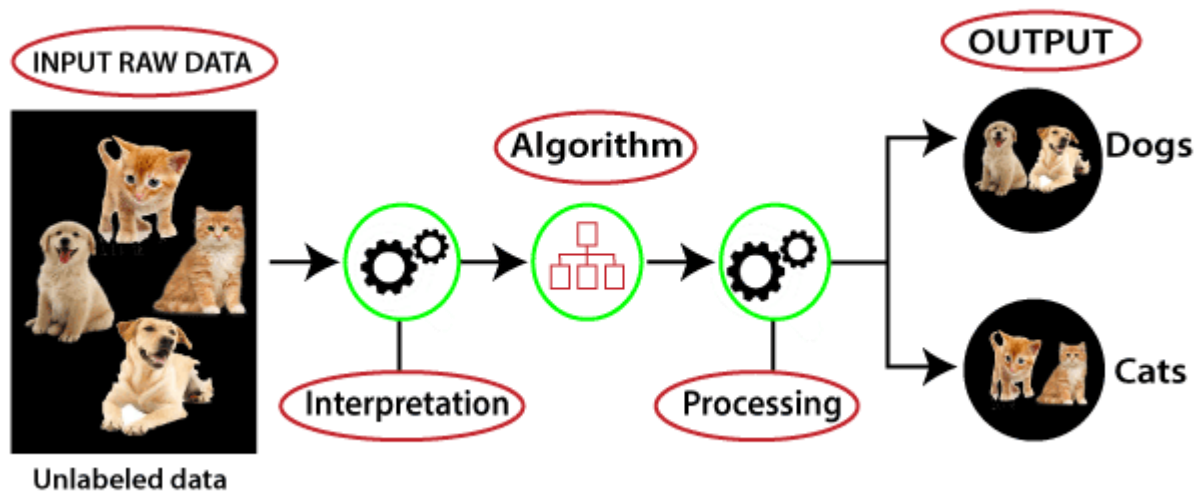
Types of unsupervised algorithm
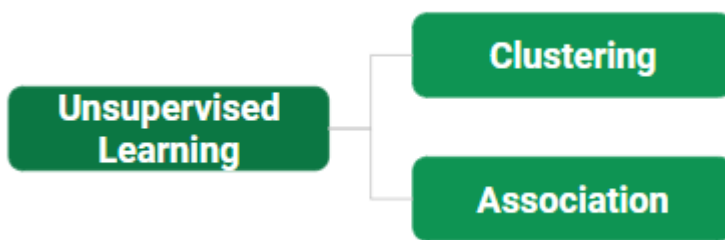
1) Clustering

The clustering technique is used when we want to find the inherent groups from the data. It is a way to group the objects into a cluster such that the objects with the most similarities remain in one group and have fewer or no similarities with the objects of other groups. An example of the clustering algorithm is grouping the customers by their purchasing behaviour.

Some of the popular clustering algorithms are given below:

- ➢ **K-Means Clustering algorithm**
- ➢ **Mean-shift algorithm**
- ➢ **DBSCAN Algorithm**

- ➢ **Principal Component Analysis**
- ➢ **Independent Component Analysis**

2) Association

Association rule learning is an unsupervised learning technique, which finds interesting relations among variables within a large dataset. The main aim of this learning algorithm is to find the dependency of one data item on another data item and map those variables accordingly so that it can generate maximum profit. This algorithm is mainly applied in Market Basket analysis, Web usage mining, continuous production**,** etc.

Some popular algorithms of Association rule learning are **Apriori Algorithm, Eclat, FP-growth algorithm.**

Advantages and Disadvantages of Unsupervised Learning Algorithm

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.

3. Semi-Supervised Learning

Semi-Supervised learning is a type of Machine Learning algorithm that lies between Supervised and Unsupervised machine learning. It represents the intermediate ground between Supervised (With Labelled training data) and Unsupervised learning (with no labelled training data) algorithms and uses the combination of labelled and unlabeled datasets during the training period.

Although Semi-supervised learning is the middle ground between supervised and unsupervised learning and operates on the data that consists of a few labels, it mostly consists of unlabeled data. As labels are costly, but for corporate purposes, they may have few labels. It is completely different from supervised and unsupervised learning as they are based on the presence & absence of labels.
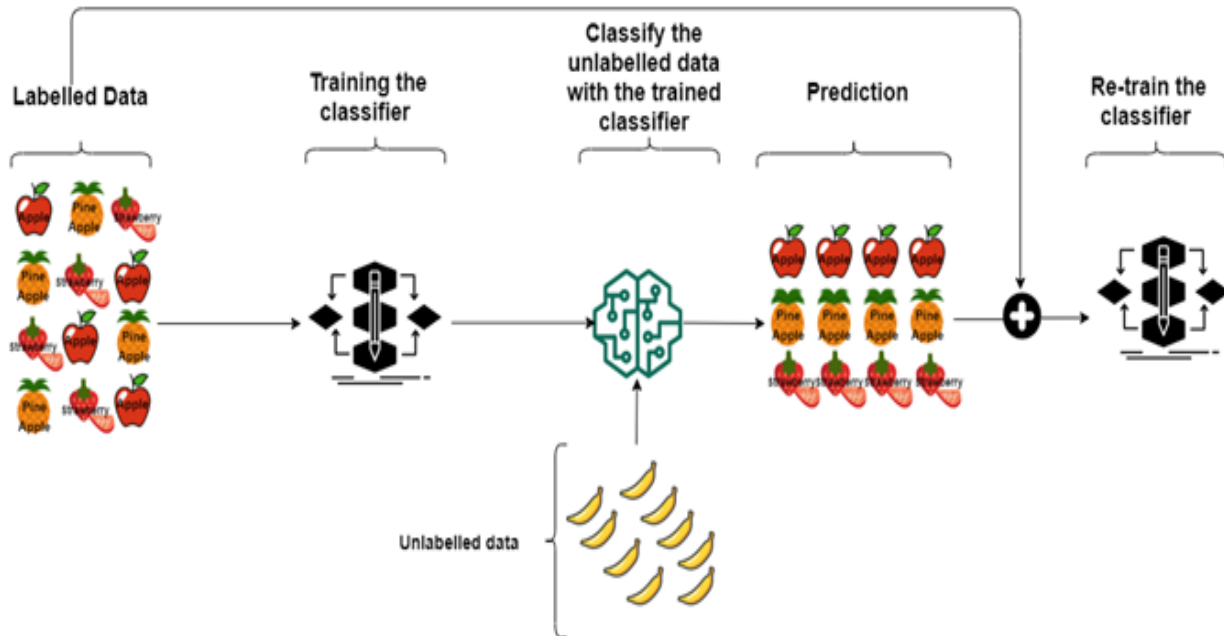
To overcome the drawbacks of supervised learning and unsupervised learning algorithms, the concept of Semi-supervised learning is introduced. The main aim of semi-supervised learning is to effectively use all the available data, rather than only labelled data like in supervised learning. Initially, similar data is clustered along with an unsupervised learning algorithm, and further, it helps to label the unlabeled data into labelled data. It is because labelled data is a comparatively more expensive acquisition than unlabeled data.

Advantages and disadvantages of Semi-supervised Learning

Advantages:

➢ It is simple and easy to understand the algorithm.

➢ It is highly efficient.

➢ It is used to solve drawbacks of Supervised and Unsupervised Learning algorithms.

Disadvantages:

➢ Iterations results may not be stable.

➢ We cannot apply these algorithms to network-level data.

➢ Accuracy is low.

4. Reinforcement Learning

Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance. Agent gets rewarded for each good action and get punished for each bad action; hence the goal of reinforcement learning agent is to maximize the rewards.

➢ In reinforcement learning, there is no labelled data like supervised learning, and agents learn from their experiences only.
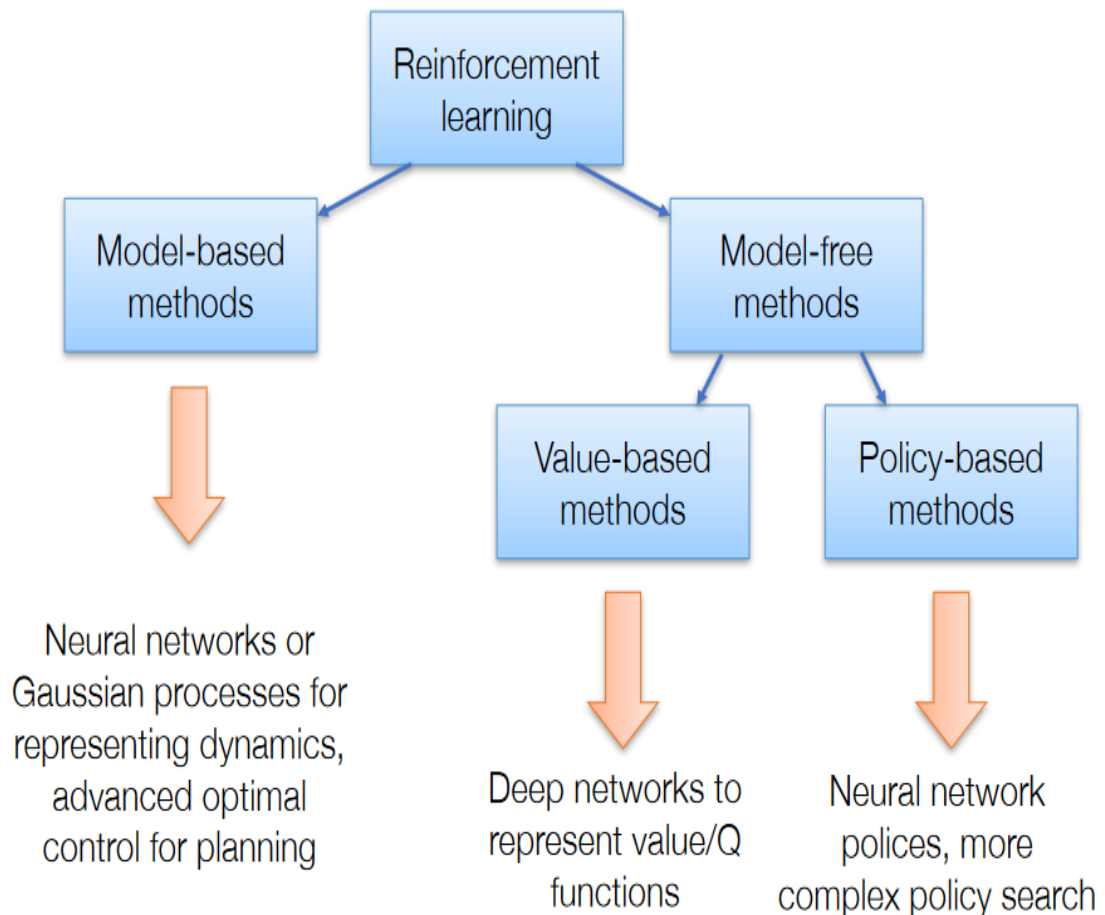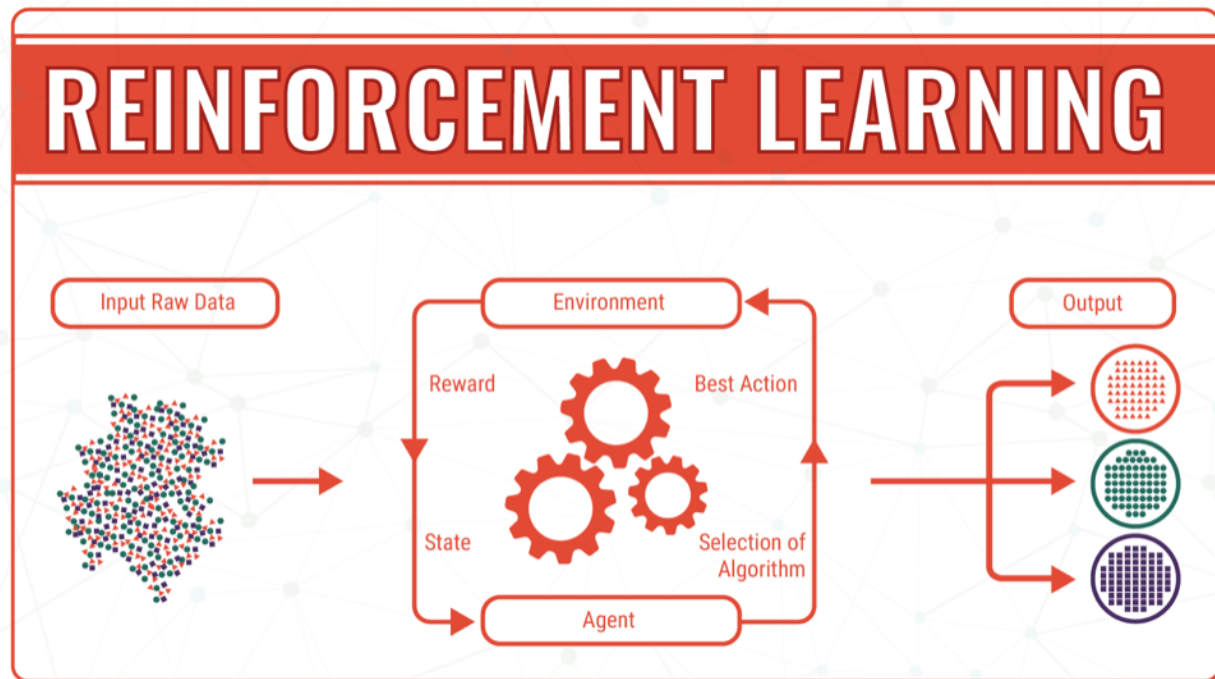
Categories of Reinforcement Learning

Reinforcement learning is categorized mainly into two types of methods/algorithms:

➢ **Positive Reinforcement Learning:** Positive reinforcement learning specifies increasing the tendency that the required behaviour would occur again by adding something. It enhances the strength of the behaviour of the agent and positively impacts it.

➢ **Negative Reinforcement Learning:** Negative reinforcement learning works exactly opposite to the positive RL. It increases the tendency that the specific behaviour would occur again by avoiding the negative condition.

Advantages and Disadvantages of Reinforcement Learning

Advantages

➢ It helps in solving complex real-world problems which are difficult to be solved by general techniques.

➢ The learning model of RL is similar to the learning of human beings; hence most accurate results can be found.

➢ Helps in achieving long term results.

Disadvantage

➢ RL algorithms are not preferred for simple problems.

➢ RL algorithms require huge data and computations.

➢ Too much reinforcement learning can lead to an overload of states which can weaken the results.

Applications of Reinforcement Learning

➢ **Video Games:**RL algorithms are much popular in gaming applications. It is used to gain super-human performance. Some popular games that use RL algorithms are **AlphaGO** and **AlphaGO Zero**.

➢ **Resource Management:**

The "Resource Management with Deep Reinforcement Learning" paper showed that how to use RL in computer to automatically learn and schedule resources to wait for different jobs in order to minimize average job slowdown.

➢ **Robotics:**

RL is widely being used in Robotics applications. Robots are used in the industrial and manufacturing area, and these robots are made more powerful with reinforcement learning. There are different industries that have their vision of building intelligent robots using AI and Machine learning technology.

➢ **TextMining**

Text-mining, one of the great applications of NLP, is now being implemented with the help of Reinforcement Learning by Salesforce company.