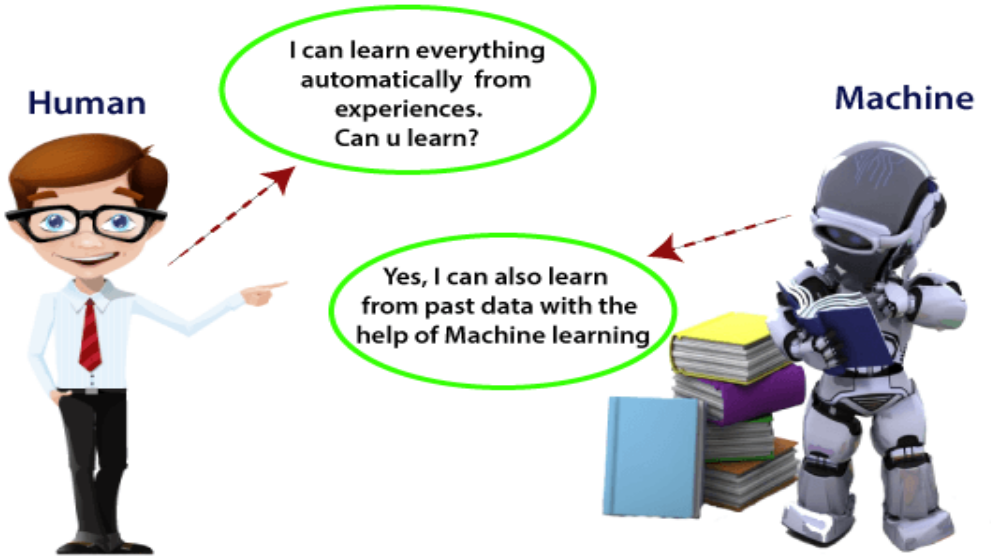
CHAPTER 1

INTRODUCTION TO MACHINE LEARNING

1.1. Introduction

- 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.
- In the real world, we are surrounded by humans who can learn everything from their experiences with their learning capability, and we have computers or machines which work on our instructions.
- But can a machine also learn from experiences or past data like a human does? So here comes the role of **Machine Learning**.

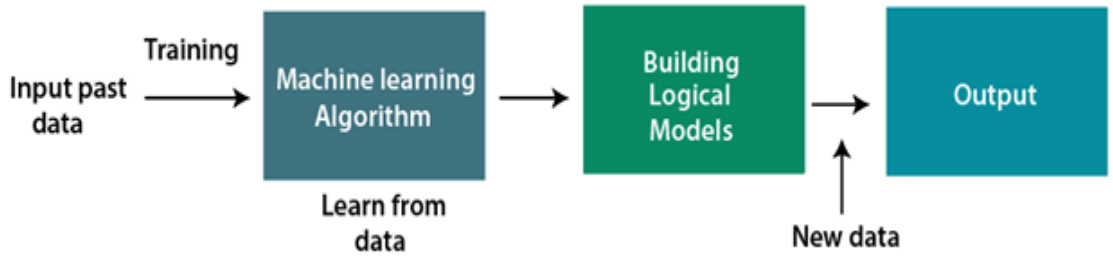


- Machine Learning is said as a subset of **artificial intelligence** that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own.
- The term machine learning was first introduced by **Arthur Samuel** in **1959**.
- We can 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.*"
- With the help of sample historical data, which is known as **training data**, machine learning algorithms build a **mathematical model** that helps in making predictions or decisions without being explicitly programmed.
- Machine learning brings computer science and statistics together for creating predictive models.
- Machine learning constructs or uses the algorithms that learn from historical data.
- The more we will provide the information, the higher will be the performance.

- Tom Mitchell defined Machine Learning as:
  *Machine Learning is the study of algorithms that improve their **performance P** at some **task T** with **experience E**.*
- A well-defined learning task is given by <P, T, E>.
- Consider the Robot Driving Problem. Here,
  Task (T): Driving on public 4 lane highway using vision sensors.
  Performance Measure (P): Average distance travelled before an error occurs.
  Experience (E): A sequence of images and steering commands recorded while observing a human driver.

1.2. How does Machine Learning work?

- A Machine Learning system **learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it**.
- The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.
- Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output.
- Machine learning has changed our way of thinking about the problem.
- The below block diagram explains the working of Machine Learning algorithm:



1.3. Feature 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.

1.4.Need of Machine Learning

- The need for machine learning is increasing day by day. The reason behind the need for machine learning is that it is capable of doing tasks that are too complex for a person to implement directly.
- As a human, we have some limitations as we cannot access the huge amount of data manually, so for this, we need some computer systems and here comes the machine learning to make things easy for us.

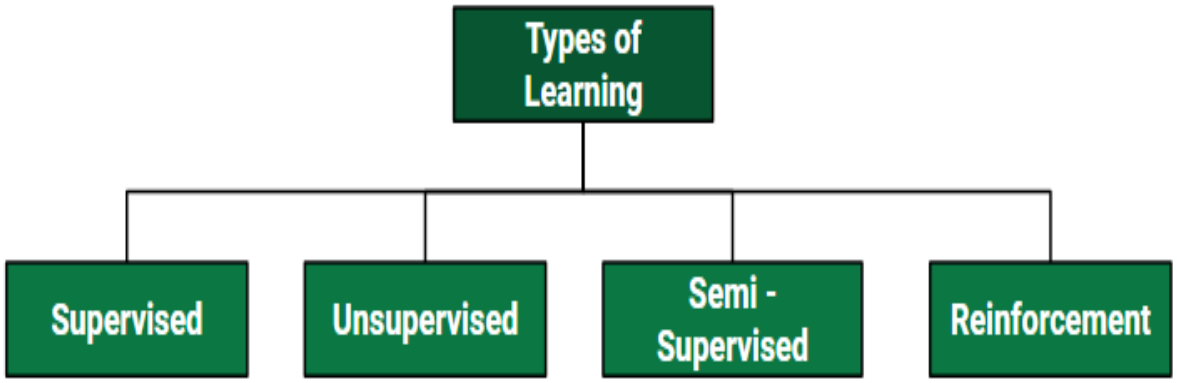
- We can train machine learning algorithms by providing them the huge amount of data and let them explore the data, construct the models, and predict the required output automatically.
- The performance of the machine learning algorithm depends on the amount of data, and it can be determined by the cost function.
- With the help of machine learning, we can save both time and money.
- The importance of machine learning can be easily understood by its uses cases, Currently, machine learning is used in self-driving cars, cyber fraud detection, face recognition, and friend suggestion by Facebook, etc.
- Various top companies such as Netflix and Amazon have built machine learning models that are using a vast amount of data to analyze the user interest and recommend product accordingly.
- Following are some key points which show the importance of Machine Learning:
  1. Rapid increment in the production of data
  2. Solving complex problems, which are difficult for a human
  3. Decision making in various sector including finance
  4. Finding hidden patterns and extracting useful information from data.

1.5.Data Mining Vs Machine Learning

Basis for comparison	**Data mining**	**Machine learning**
Meaning	Extracting knowledge from a large amount of data	Introduce new algorithm from data as well as past experience
History	Introduced in 1930, initially referred as knowledge discovery in databases	introduced in near 1950, the first program was Samuel's checker-playing program
Responsibility	Data mining is used to get the rules from the existing data.	Machine learning teaches the computer to learn and understand the given rules.
Origin	Traditional databases with unstructured data	Existing data as well as algorithms.
Implementation	We can develop our own models where we can use data mining techniques for	We can use machine learning algorithm in the decision tree, neural networks and some other area of artificial intelligence.
Nature	Involves human interference more towards manual.	Automated, once design self-implemented, no human effort
Application	Used in cluster analysis	Used in web search, spam filter, credit scoring, fraud detection, computer design
Abstraction	Data mining abstract from the data warehouse	Machine learning reads machine

Techniques involve	Data mining is more of a research using methods like machine learning	Self-learned and trains system to do the intelligent task.
Scope	Applied in the limited area	Can be used in a vast area.

1.6. Classification of Machine Learning



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

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

1. Supervised Learning

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

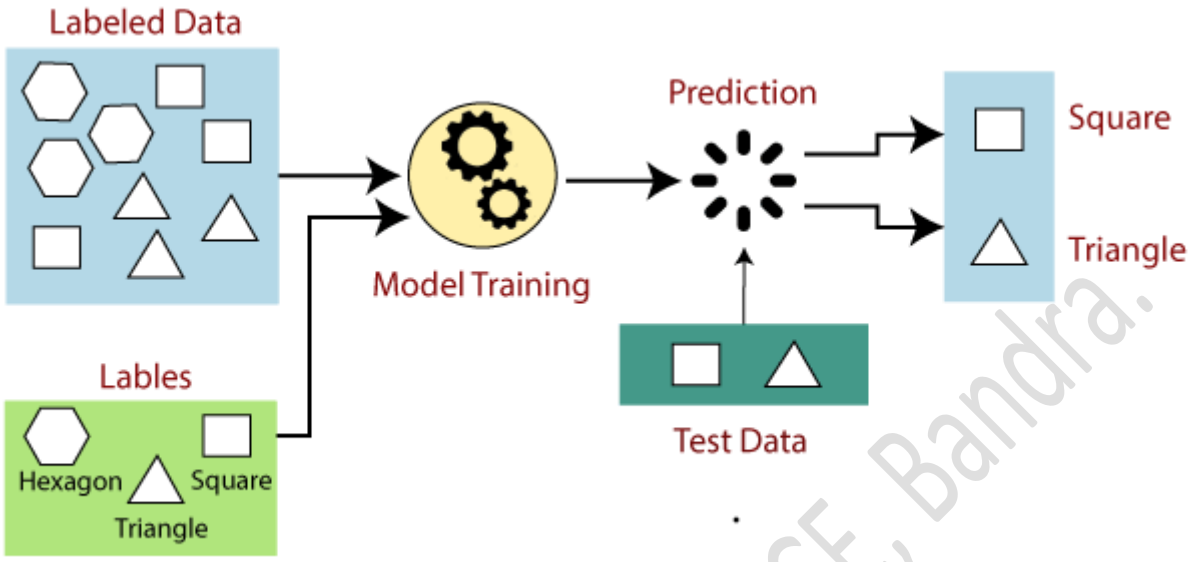
The system creates a model using labelled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.

- The goal of supervised learning is to map input data with the output data.
- The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher.
- The example of supervised learning is **spam filtering**.
- Supervised learning can be grouped further in two categories of algorithms:
  a. Classification
  b. Regression

How Supervised Learning Works?

- In supervised learning, models are trained using labelled dataset, where the model learns about each type of data.
- Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output.

- The working of Supervised learning can be easily understood by the below example and diagram:



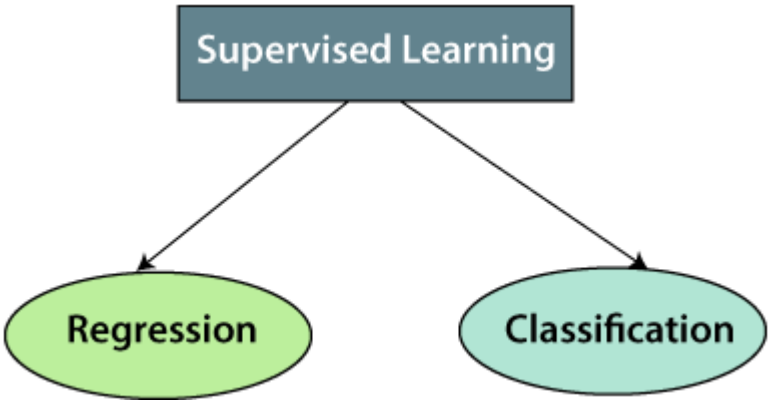
- Suppose we have a dataset of different types of shapes which includes square, rectangle, triangle, and Polygon.
- Now the first step is that we need to train the model for each shape.
  1. If the given shape has four sides, and all the sides are equal, then it will be labelled as a **Square**.
  2. If the given shape has three sides, then it will be labelled as a **triangle**.
  3. If the given shape has six equal sides, then it will be labelled as **hexagon**.
- Now, after training, we test our model using the test set, and the task of the model is to identify the shape.
- The machine is already trained on all types of shapes, and when it finds a new shape, it classifies the shape on the bases of a number of sides, and predicts the output.

Steps Involved in Supervised Learning

1. Determine the type of training dataset.
2. Collect/Gather the labelled training data.
3. Split the training dataset into training dataset, test dataset, and validation dataset.
4. Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
5. Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
6. Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
7. Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

Types of supervised Machine Learning Algorithms

Supervised learning can be further divided into two types of problems:



1. Regression

- Regression algorithms are used if there is a relationship between the input variable and the output variable.
- It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.
- Below are some popular Regression algorithms which come under supervised learning:
  - Linear Regression
  - Regression Trees
  - Non-Linear Regression
  - Bayesian Linear Regression
  - Polynomial Regression

2. Classification

- Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.
- It is used in Spam Filtering, classifying size of tumour as small, medium, or large.
- Below are some popular Classification algorithms which come under supervised learning:
  - Random Forest
  - Decision Trees
  - Logistic Regression
  - Support vector Machines

Advantages of Supervised learning

- With the help of supervised learning, the model can predict the output on the basis of prior experiences.
- In supervised learning, we can have an exact idea about the classes of objects.
- Supervised learning model helps us to solve various real-world problems such as **fraud detection, spam filtering**, etc.

Disadvantages of supervised learning

- Supervised learning models are not suitable for handling the complex tasks.
- Supervised learning cannot predict the correct output if the test data is different from the training dataset.

- Training required lots of computation times.
- In supervised learning, we need enough knowledge about the classes of object.

2. Unsupervised Learning

- Unsupervised learning is a learning method in which a machine learns without any supervision.
- The training is provided to the machine with the set of data that has not been labelled, classified, or categorized, and the algorithm needs to act on that data without any supervision.
- The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.
- In unsupervised learning, we don't have a predetermined result.
- The machine tries to find useful insights from the huge amount of data.
- It can be further classifieds into two categories of algorithms:
  a. Clustering
  b. Association

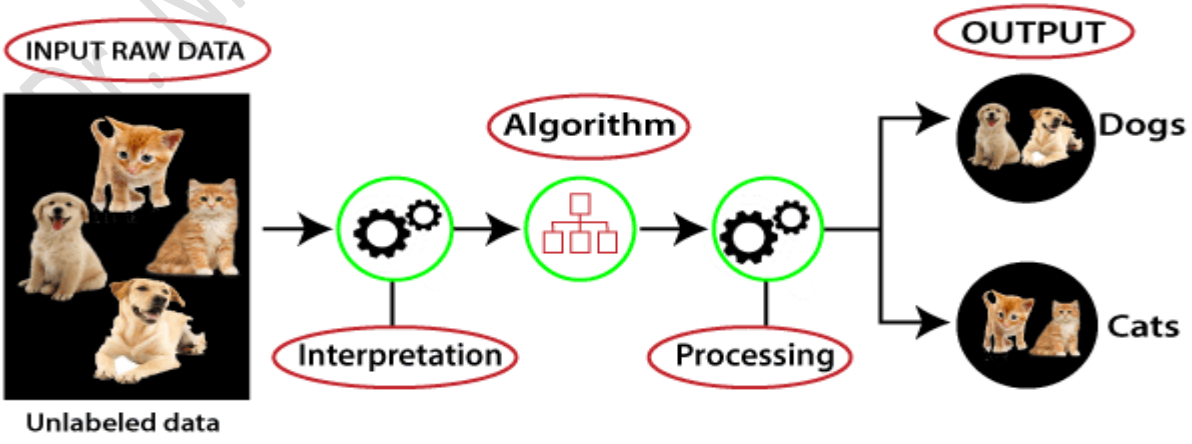
Why use Unsupervised Learning?

Below are some main reasons which describe the importance of Unsupervised Learning:

- Unsupervised learning is helpful for finding useful insights from the data.
- Unsupervised learning is much similar as a human learns to think by their own experiences, which makes it closer to the real AI.
- Unsupervised learning works on unlabelled and uncategorized data which make unsupervised learning more important.
- In real-world, we do not always have input data with the corresponding output so to solve such cases, we need unsupervised learning.

How Unsupervised Learning Works?

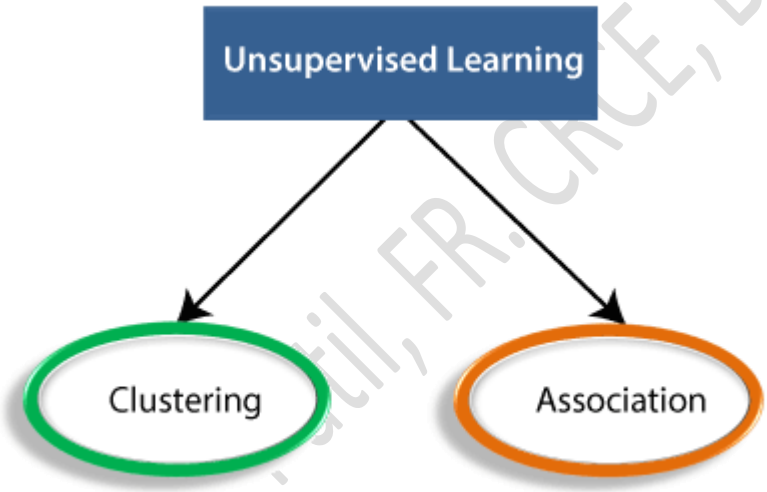
- Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs.



- The algorithm is never trained upon the given dataset, which means it does not have any idea about the features of the dataset.
- The task of the unsupervised learning algorithm is to identify the image features on their own.
- Unsupervised learning algorithm will perform this task by clustering the image dataset into the groups according to similarities between images.
- Firstly, it will interpret the raw data to find the hidden patterns from the data and then will apply suitable algorithms such as k-means clustering, Decision tree, etc.
- Once it applies the suitable algorithm, the algorithm divides the data objects into groups according to the similarities and difference between the objects.

Types of Unsupervised Learning Algorithm

The unsupervised learning algorithm can be further categorized into two types of problems:



- **Clustering**:
  - Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group.
  - Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.
  - Below are some popular Clustering algorithms which come under unsupervised learning:
    a. k-means clustering
    b. k-nearest neighbors (KNN)
    c. Hierarchical Clustering
- **Association**
  - An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database.
  - It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective.

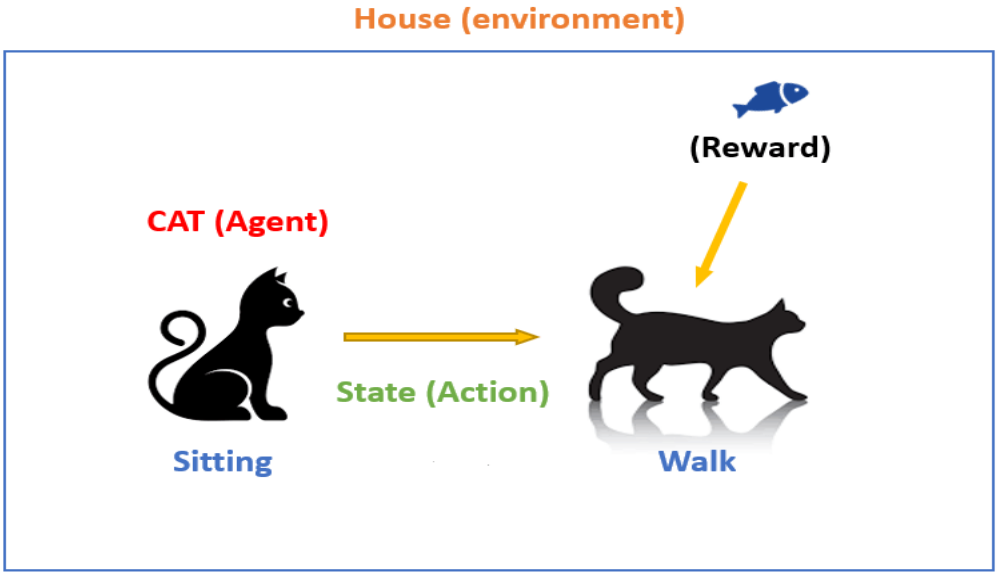
- Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item.
- A typical example of Association rule is **Market Basket Analysis**.
- Below are some popular Association algorithms which come under unsupervised learning:
  a. Apriori algorithm
  b. FP-Growth algorithm

3. 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 Reinforcement Learning works?

- Let's see some simple example which helps you to illustrate the reinforcement learning mechanism.
- Consider the scenario of teaching new tricks to your cat.
- As cat doesn't understand English or any other human language, we can't tell her directly what to do. Instead, we follow a different strategy.
- We emulate a situation, and the cat tries to respond in many different ways. If the cat's response is the desired way, we will give her fish.
- Now whenever the cat is exposed to the same situation, the cat executes a similar action with even more enthusiastically in expectation of getting more reward(food).
- That's like learning that cat gets from "what to do" from positive experiences.
- At the same time, the cat also learns what not do when faced with negative experiences.



Explanation about the example:

In this case,

- Your cat is an agent that is exposed to the environment. In this case, it is your house. An example of a state could be your cat sitting, and you use a specific word in for cat to walk.
- Our agent reacts by performing an action transition from one "state" to another "state."
- For example, your cat goes from sitting to walking.
- The reaction of an agent is an action, and the policy is a method of selecting an action given a state in expectation of better outcomes.
- After the transition, they may get a reward or penalty in return.

Approaches to implement Reinforcement Learning

There are mainly three ways to implement reinforcement-learning in ML, which are:

Value-based:

- The value-based approach is about to find the optimal value function, which is the maximum value at a state under any policy.
- Therefore, the agent expects the long-term return at any state(s) under policy π.

Policy-based:

- Policy-based approach is to find the optimal policy for the maximum future rewards without using the value function.
- In this approach, the agent tries to apply such a policy that the action performed in each step helps to maximize the future reward.
- The policy-based approach has mainly two types of policy:
  a. Deterministic: The same action is produced by the policy (π) at any state.
  b. Stochastic: In this policy, probability determines the produced action.

Model-based:

- In the model-based approach, a virtual model is created for the environment, and the agent explores that environment to learn it.
- There is no particular solution or algorithm for this approach because the model representation is different for each environment.

Types of Reinforcement Learning

Two kinds of reinforcement learning methods are:

1. **Positive**:
   - It is defined as an event, that occurs because of specific behavior.
   - It increases the strength and the frequency of the behavior and impacts positively on the action taken by the agent.
   - This type of Reinforcement helps you to maximize performance and sustain change for a more extended period.
   - However, too much Reinforcement may lead to over-optimization of state, which can affect the results.

2. **Negative**:
    - Negative Reinforcement is defined as strengthening of behavior that occurs because of a negative condition which should have stopped or avoided.
    - It helps you to define the minimum stand of performance.
    - However, the drawback of this method is that it provides enough to meet up the minimum behavior.

4. Semi-Supervised Learning

- In this type of learning, the algorithm is trained upon a combination of labeled and unlabeled data.
- Typically, this combination will contain a very small amount of labeled data and a very large amount of unlabeled data.
- The basic procedure involved is that first, the programmer will cluster similar data using an unsupervised learning algorithm and then use the existing labeled data to label the rest of the unlabeled data.
- Applications
    a. Speech Analysis: Labelling of audio files
    b. Internet Content Classification: Labelling each web page

1.7. Supervised Vs Unsupervised Learning

Supervised Learning	**Unsupervised Learning**
Supervised learning algorithms are trained using labelled data.	Unsupervised learning algorithms are trained using unlabelled data.
Supervised learning model takes direct feedback to check if it is predicting correct output or not.	Unsupervised learning model does not take any feedback.
Supervised learning model predicts the output.	Unsupervised learning model finds the hidden patterns in data.
In supervised learning, input data is provided to the model along with the output.	In unsupervised learning, only input data is provided to the model.
The goal of supervised learning is to train the model so that it can predict the output when it is given new data.	The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset.
Supervised learning needs supervision to train the model.	Unsupervised learning does not need any supervision to train the model.
Supervised learning can be categorized in **Classification** and **Regression** problems.	Unsupervised Learning can be classified in **Clustering** and **Associations** problems.
Supervised learning can be used for those cases where we know the input as well as corresponding outputs.	Unsupervised learning can be used for those cases where we have only input data and no corresponding output data.

Supervised learning model produces an accurate result.	Unsupervised learning model may give less accurate result as compared to supervised learning.
Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output.	Unsupervised learning is more close to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences.
It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc.	It includes various algorithms such as Clustering, KNN, and Apriori algorithm.

1.8. Supervised Vs Unsupervised Vs Reinforcement Learning

Criteria	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	The machine learns by using labelled data	The machine is trained on unlabelled data without any guidance	An agent interacts with its environment by performing actions and learning from errors or rewards.
Type of Problem	Regression and Classification	Association and Clustering	Reward-based
Type of Data	Labelled data	Unlabelled data	No predefined data
Training	External Supervision	No Supervision	No Supervision
Approach	Maps the labelled inputs to the known outputs	Understands patterns and discovers the output	Follows the trial-and-error method

1.9.Machine Learning Applications

Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. Below are some most trending real-world applications of Machine Learning:

1. **Image Recognition**:
   - Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc.
   - The popular use case of image recognition and face detection is, **Automatic friend tagging suggestion**: Facebook provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's **face detection** and **recognition algorithm**. It is based on the

Facebook project named "**Deep Face**," which is responsible for face recognition and person identification in the picture.

2. **Speech Recognition**
   - While using Google, we get an option of "**Search by voice**," it comes under speech recognition, and it's a popular application of machine learning.
   - Speech recognition is a process of converting voice instructions into text, and it is also known as "**Speech to text**", or "**Computer speech recognition**."
   - At present, machine learning algorithms are widely used by various applications of speech recognition.
   - **Google assistant**, **Siri**, **Cortana**, and **Alexa** are using speech recognition technology to follow the voice instructions.

3. **Traffic prediction:**
   - If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.
   - It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:
     - **Real Time location** of the vehicle form Google Map app and sensors
     - **Average time has taken** on past days at the same time.
   - Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

4. **Product recommendations:**
   - Machine learning is widely used by various e-commerce and entertainment companies such as **Amazon, Netflix**, etc., for product recommendation to the user.
   - Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.
   - Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.
   - As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

5. **Self-driving cars:**
   - One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars.
   - **Tesla**, the most popular car manufacturing company is working on self-driving car.
   - It is using unsupervised learning method to train the car models to detect people and objects while driving.

6. **Email Spam and Malware Filtering:**
   - Whenever we receive a new email, it is filtered automatically as important, normal, and spam.
   - We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning.
   - Below are some spam filters used by Gmail:

- Content Filter
- Header filter
- General blacklists filter
- Rules-based filters
- Permission filters

- Some machine learning algorithms such as **Multi-Layer Perceptron**, **Decision tree**, and **Naïve Bayes classifier** are used for email spam filtering and malware detection.

7. **Virtual Personal Assistant**:
   - We have various virtual personal assistants such as **Google assistant, Alexa, Cortana, Siri.**
   - As the name suggests, they help us in finding the information using our voice instruction.
   - These assistants can help us in various ways just by our voice instructions such as play music, call someone, open an email, scheduling an appointment, etc.
   - These virtual assistants use machine learning algorithms as an important part.
   - These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly.

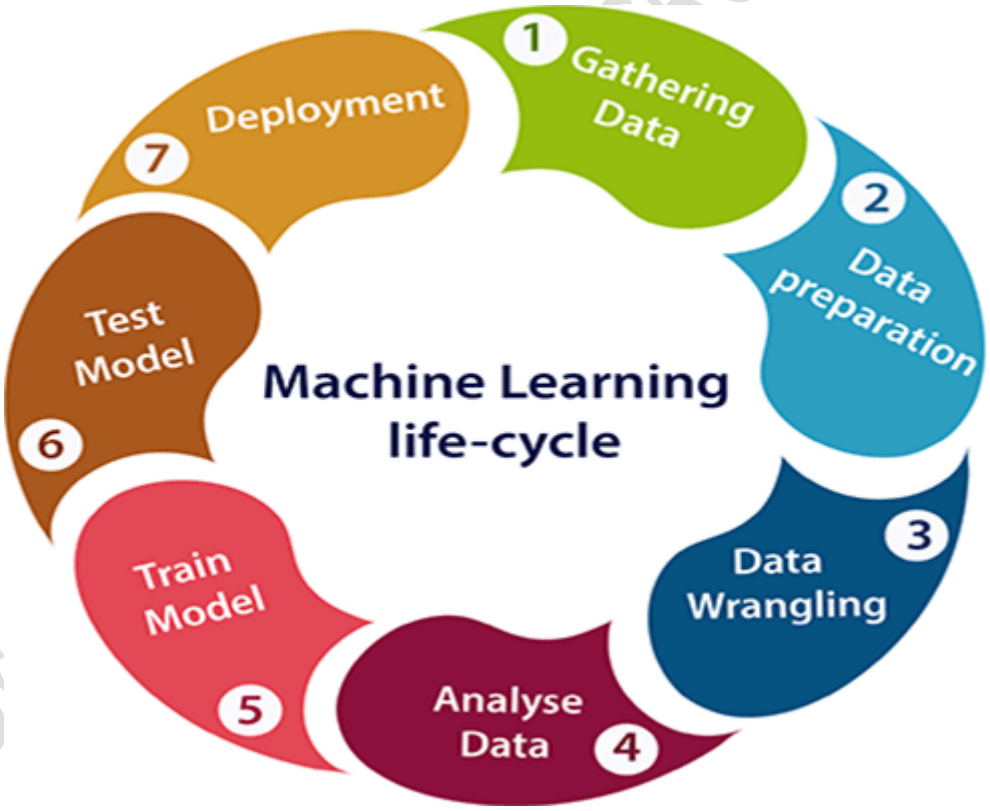
8. **Online Fraud Detection**:
   - Machine learning is making our online transaction safe and secure by detecting fraud transaction.
   - Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as fake accounts, fake ids, and steal money in the middle of a transaction.
   - So to detect this, Feed Forward Neural network helps us by checking whether it is a genuine transaction or a fraud transaction.
   - For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round.
   - For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions more secure.

9. **Stock Market trading:**
   - Machine learning is widely used in stock market trading.
   - In the stock market, there is always a risk of up and downs in shares, so for this machine learning's long short term memory neural network is used for the prediction of stock market trends.

10. **Medical Diagnosis**:
    - In medical science, machine learning is used for diseases diagnoses.
    - With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain.
    - It helps in finding brain tumors and other brain-related diseases easily.

1.10. Steps for Developing Machine Learning Applications (Life Cycle of Machine Learning)

- Machine learning has given the computer systems the abilities to automatically learn without being explicitly programmed. But how does a machine learning system work? So, it can be described using the life cycle of machine learning.
- Machine learning life cycle is a cyclic process to build an efficient machine learning project.
- The main purpose of the life cycle is to find a solution to the problem or project.
- Machine learning life cycle involves seven major steps, which are given below:
  1. Gathering Data
  2. Data preparation
  3. Data Wrangling
  4. Analyse Data
  5. Train the model
  6. Test the model
  7. Deployment



- The most important thing in the complete process is to understand the problem and to know the purpose of the problem.
- Therefore, before starting the life cycle, we need to understand the problem because the good result depends on the better understanding of the problem.
- In the complete life cycle process, to solve a problem, we create a machine learning system called "model", and this model is created by providing "training". But to train a model, we need data, hence, life cycle starts by collecting data.

1. **Gathering Data:**

- Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.
- In this step, we need to identify the different data sources, as data can be collected from various sources such as files, database, internet, or mobile devices.
- It is one of the most important steps of the life cycle.
- The quantity and quality of the collected data will determine the efficiency of the output.
- The more will be the data, the more accurate will be the prediction.
- This step includes the below tasks:
  - Identify various data sources
  - Collect data
  - Integrate the data obtained from different sources
- By performing the above task, we get a coherent set of data, also called as a dataset. It will be used in further steps.

2. Data preparation

- After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.
- In this step, first, we put all data together, and then randomize the ordering of data.
- This step can be further divided into two processes:

➢ **Data exploration**:
  - ✓ It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data.
  - ✓ A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.

➢ Data pre-processing:
  - ✓ Now the next step is pre-processing of data for its analysis.

3. Data Wrangling

- Data wrangling is the process of cleaning and converting raw data into a useable format.
- It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step.
- It is one of the most important steps of the complete process.
- Cleaning of data is required to address the quality issues.
- It is not necessary that data we have collected is always of our use as some of the data may not be useful.
- In real-world applications, collected data may have various issues, including:
  - **Missing Values**
  - **Duplicate data**
  - **Invalid data**
  - **Noise**
- So, we use various filtering techniques to clean the data.

- It is mandatory to detect and remove the above issues because it can negatively affect the quality of the outcome.

4. **Data Analysis**
   - Now the cleaned and prepared data is passed on to the analysis step. This step involves:
     - **Selection of analytical techniques**
     - **Building models**
     - **Review the result**
   - The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome.
   - It starts with the determination of the type of the problems, where we select the machine learning techniques such as Classification, Regression, Cluster analysis, Association, etc. then build the model using prepared data, and evaluate the model.
   - Hence, in this step, we take the data and use machine learning algorithms to build the model.

5. **Train Model**
   - Now the next step is to train the model, in this step we train our model to improve its performance for better outcome of the problem.
   - We use datasets to train the model using various machine learning algorithms.
   - Training a model is required so that it can understand the various patterns, rules, and, features.

6. **Test Model**
   - Once our machine learning model has been trained on a given dataset, then we test the model. In this step, we check for the accuracy of our model by providing a test dataset to it.
   - Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

7. **Deployment**
   - The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system.
   - If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system.
   - But before deploying the project, we will check whether it is improving its performance using available data or not.
   - The deployment phase is similar to making the final report for a project.

1.11. Issues in Machine Learning

The field of machine learning is concerned with answering questions such as the following:

1. What algorithms exist for learning general target functions from specific training examples?
2. In what settings will particular algorithms converge to the desired function, given sufficient training data?

3. Which algorithms perform best for which types of problems and representations?
4. How much training data is sufficient?
5. What general bounds can be found to relate the confidence in learned hypotheses to the amount of training experience and the character of the learner's hypothesis space?
6. When and how can prior knowledge held by the learner guide the process of generalizing from examples?
7. Can prior knowledge be helpful even when it is only approximately correct?
8. What is the best strategy for choosing a useful next training experience, and how does the choice of this strategy alter the complexity of the learning problem?
9. What is the best way to reduce the learning task to one or more function approximation problems? Put another way, what specific functions should the system attempt to learn? Can this process itself be automated?
10. How can the learner automatically alter its representation to improve its ability to represent and learn the target function?

10.11. Tasks in Machine Learning

Type of ML Task	Description	Example
Classification	Pick one of N labels	Cat, dog, horse, or bear
Regression	Predict numerical values	Click-through rate
Clustering	Group similar examples	Most relevant documents (unsupervised)
Association rule learning	Infer likely association patterns in data	If you buy hamburger buns, you're likely to buy hamburgers (unsupervised)
Structured output	Create complex output	Natural language parse trees, image recognition bounding boxes
Ranking	Identify position on a scale or status	Search result ranking

10.12. Selecting Right Machine Learning Algorithm

This is a generic, practical approach that can be applied to most machine learning problems:

1. **Categorize the problem**
   The first step is to categorize the problem.
   Categorize by the input:
   - If it is a labelled data, it's a supervised learning problem.
   - If it's unlabelled data with the purpose of finding structure, it's an unsupervised learning problem.
   - If the solution implies to optimize an objective function by interacting with an environment, it's a reinforcement learning problem.

   Categorize by output:
   - If the output of the model is a number, it's a regression problem.
   - If the output of the model is a class, it's a classification problem.
   - If the output of the model is a set of input groups, it's a clustering problem.

2. **Understand Your Data**
   - Data itself is not the end game, but rather the raw material in the whole analysis process.
   - The process of understanding the data plays a key role in the process of choosing the right algorithm for the right problem.
   - Some algorithms can work with smaller sample sets while others require tons and tons of samples.
   - Certain algorithms work with categorical data while others like to work with numerical input.

   Analyze the Data
   - In this step, there are two important tasks which are understand data with descriptive statistics and understand data with visualization and plots.

   Process the data
   - The components of data processing include pre-processing, profiling, cleansing, it often also involves pulling together data from different internal systems and external sources.

   Transform the data
   - The traditional idea of transforming data from a raw state to a state suitable for modeling is where feature engineering fits in.
   - Transform data and feature engineering may, in fact, be synonyms. And here is a definition of the latter concept. Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data.
3. **Find the available algorithms**
   - After categorizing the problem and understand the data, the next milestone is identifying the algorithms that are applicable and practical to implement in a reasonable time.
   - Some of the elements affecting the choice of a model are:
     - The accuracy of the model.
     - The interpretability of the model.
     - The complexity of the model.
     - The scalability of the model.
     - How long does it take to build, train, and test the model?
     - How long does it take to make predictions using the model?
     - Does the model meet the business goal?
4. **Implement machine learning algorithms.**
   - Set up a machine learning pipeline that compares the performance of each algorithm on the dataset using a set of carefully selected evaluation criteria.
   - Another approach is to use the same algorithm on different subgroups of datasets.
   - The best solution for this is to do it once or have a service running that does this in intervals when new data is added.
5. **Optimize hyper-parameters.**
   - There are three options for optimizing hyper-parameters, grid search, random search, and Bayesian optimization.