UNIT-1

➢ What is Machine Learning?

Machine learning is a branch of artificial intelligence that involves developing algorithms and statistical models that allow computers to learn from data and make predictions or decisions without being explicitly programmed. Machine learning algorithms can be used to identify patterns in large datasets and use those patterns to make predictions or decisions about new, unseen data.

There are three main types of machine learning –

1. **Supervised Learning**

In supervised learning, the algorithm is trained on labeled data, meaning that the correct answer or output is provided for each input. The algorithm then uses this labeled data to make predictions about new, unseen data.

2. **Unsupervised Learning**

In unsupervised learning, the algorithm is trained on unlabeled data, meaning that the correct output or answer is not provided for each input. Instead, the algorithm must identify patterns and structures in the data on its own.

3. **Reinforcement Learning**

In reinforcement learning, the algorithm learns by receiving feedback in the form of rewards or punishments based on its actions. The algorithm then uses this feedback to adjust its behavior and improve its performance.

➢ Machine Learning Model

Before discussing the machine learning model, we must need to understand the following formal definition of ML given by Professor Mitchell –

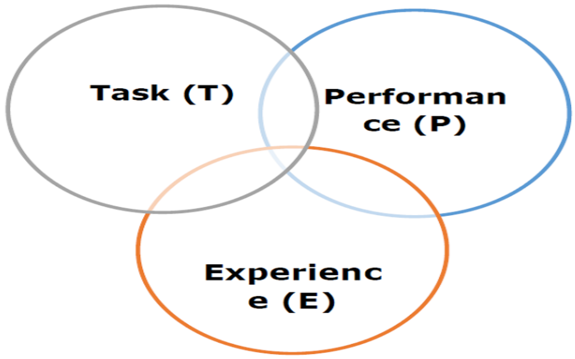
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

The above definition is basically focusing on three parameters, also the main components of any learning algorithm, namely Task (T), Performance (P) and experience (E). In this context, we can simplify this definition as –

ML is a field of AI consisting of learning algorithms that –

- Improve their performance (P)
- At executing some task (T)
- Over time with experience (E)

Based on the above, the following diagram represents a Machine Learning Model –



Let us discuss them more in detail now –

- **Task(T)**

From the perspective of problem, we may define the task T as the real-world problem to be solved. The problem can be anything like finding best house price in a specific location or to find best marketing strategy etc. On the other hand, if we talk about machine learning, the definition of task is different because it is difficult to solve ML based tasks by conventional programming approach.

A task T is said to be a ML based task when it is based on the process and the system must follow for operating on data points. The examples of ML based tasks are Classification, Regression, Structured annotation, Clustering, Transcription etc.

- **Experience (E)**

As name suggests, it is the knowledge gained from data points provided to the algorithm or model. Once provided with the dataset, the model will run iteratively and will learn some inherent pattern. The learning thus acquired is called experience (E). Making an analogy with human learning, we can think of this situation as in which a human being is learning or gaining some experience from various attributes like situation, relationships etc. Supervised, unsupervised and reinforcement learning are some ways to learn or gain experience. The experience gained by out ML model or algorithm will be used to solve the task T.

- **Performance (P)**

An ML algorithm is supposed to perform task and gain experience with the passage of time. The measure which tells whether ML algorithm is performing as per expectation or not is its performance (P). P is basically a quantitative metric that tells how a model is performing the task, T, using its experience, E. There are many metrics that help to understand the ML performance, such as accuracy score, F1 score, confusion matrix, precision, recall, sensitivity etc.

➢ Need of Machine Learning

Here are some reasons why it’s so essential in the modern world:

- **Data processing.** One of the primary reasons machine learning is so important is its ability to handle and make sense of large volumes of data. With the explosion of digital data from social media, sensors, and other sources, traditional data analysis methods have become inadequate. Machine learning algorithms can process these vast amounts of data, uncover hidden patterns, and provide valuable insights that can drive decision-making.
- **Driving innovation.** Machine learning is driving innovation and efficiency across various sectors. Here are a few examples:
- **Healthcare**. Algorithms are used to predict disease outbreaks, personalize patient treatment plans, and improve medical imaging accuracy.
- **Finance**. Machine learning is used for credit scoring, algorithmic trading, and fraud detection.
- **Retail**. Recommendation systems, supply chains, and **customer service** can all benefit from machine learning.
- The techniques used also find applications in sectors as diverse as agriculture, education, and entertainment.

- **Enabling automation**. Machine learning is a key enabler of automation. By learning from data and improving over time, machine learning algorithms can perform previously manual tasks, freeing humans to focus on more complex and creative tasks. This not only increases efficiency but also opens up new possibilities for innovation.

➢ History of Machine Learning

Machine Learning has gone through many phases of development since the inception of computers. The early History of Machine Learning, Timeline 1943-1979

- **1943: The First Neutral Network with Electric Circuit**

The first neutral network with electric circuit was developed by Warren McCulloch and Walter Pitts in 1943. The goal of the network was to solve a problem that had been posed by John von Neumann and others: how could computers be made to communicate with each other?

This early model showed that it was possible for two computers to communicate without any human interaction. This event is important because it paved the way for machine learning development.

- **1950: Turing Test**

The Turing Test is a test of artificial intelligence proposed by mathematician Alan Turing. It involves determining whether a machine can act like a human, or if humans can't tell the difference between human and machine given answers.

The goal of the test is to determine whether machines can think intelligently and demonstrate some form of emotional capability. It does not matter whether the answer is true or false but whether it is considered human or not by the questioner. There have been several attempts to create an AI that passes the Turing Test, but no machine has yet successfully done so.

The Turing Test has been criticized because it measures how much a machine can imitate a human rather than proving their true intelligence.

- **1952: Computer Checkers**

Arthur Samuel was a pioneer in machine learning and is credited with creating the first computer program to play championship-level checkers. His program, which he developed in 1952, used a technique called alpha-beta pruning to measure the chances of winning a game. This method is still widely used in games today. In addition, Samuel also developed the minimax algorithm, which is a technique for minimizing losses in games.

- **1957: Frank Rosenblatt – The Perceptron**

Frank Rosenblatt was a psychologist who is most famous for his work on machine learning. In 1957, he developed the perceptron, which is a machine learning algorithm. The Perceptron was one of the first algorithms to use artificial neural networks, widely used in machine learning.

It was designed to improve the accuracy of computer predictions. The goal of the Perceptron was to learn from data by adjusting its parameters until it reached an optimal solution. Perceptron's purpose was to make it easier for computers to learn from data and to improve upon previous methods that had limited success.

- **1967: The Nearest Neighbor Algorithm**

The Nearest Neighbor Algorithm was developed as a way to automatically identify patterns within large datasets. The goal of this algorithm is to find similarities between two items and determine which one is closer to the pattern found in the other item. This can be used for things like finding relationships between different pieces of data or predicting future events based on past events.

In 1967, Cover and Hart published an article on "Nearest neighbor pattern classification." It is a method of inductive logic used in machine learning to classify an input object into one of two categories. The pattern classifies the same items that are classified in the same categories as its nearest neighbors. This method is used to classify objects with a number of attributes, many of which are categorical or numerical and may have overlapping values.

- **1974: The Backpropagation**

Backpropagation was initially designed to help neural networks learn how to recognize patterns. However, it has also been used in other areas of machine learning, such as boosting performance and generalizing from data sets to new instances. The goal of backpropagation is to improve the accuracy of a model by adjusting its weights so that it can more accurately predict future outputs.

- ## 1979: The Stanford Cart

The Stanford Cart is a remote-controlled robot that can move independently in space. It was first developed in the 1960s and reached an important milestone in its development in 1979. The purpose of the Stanford Cart is to avoid obstacles and reach a specific destination: In 1979, "The Cart" succeeded for the first time in traversing a room filled with chairs in 5 hours without human intervention.

- ## 1997: A Machine Defeats a Man in Chess

In 1997, the IBM supercomputer Deep Blue defeated chess grandmaster Garry Kasparov in a match. It was the first time a machine had beaten an expert player at chess and it caused great concern for humans in the chess community. This was a landmark event as it showed that AI systems could surpass human understanding in complex tasks.

This marked a magical turning point in machine learning because the world now knew that mankind had created its own opponent- an artificial intelligence that could learn and evolve on its own.

- ## 2002: Software Library Torch

Torch is a software library for machine learning and data science. Torch was created by **Geoffrey Hinton**, Pedro Domingos, and Andrew Ng to develop the first large-scale free machine learning platform. In 2002, the founders of Torch created it as an alternative to other libraries because they believed that their specific needs were not met by other libraries. As of 2018, it has over 1 million downloads on Github and is one of the most popular machine learning libraries available today.

- ## 2006: Geoffrey Hinton, the father of Deep Learning

In 2006, Geoffrey Hinton published his "A Fast Learning Algorithm for Deep Belief Nets." This paper was the birth of deep learning. He showed that by using a deep belief network, a computer could be trained to recognize patterns in images.

- ## 2011: Google Brain

Google Brain is a research group of Google devoted to artificial intelligence and machine learning. The group was founded in 2011 by Google X and is located in Mountain View, California. The team works closely with other AI research groups within Google such as the DeepMind group that has developed AlphaGo, an AI that defeated the world champion at Go. Their goal is to build machines that can learn from data, understand language, answer questions in natural language, and have common sense reasoning.

The group is, as of 2021, led by Geoffrey Hinton, Jeff Dean and Zoubin Ghahramani and focuses on deep learning, a model of artificial neural networks that is capable to learn complex patterns from data automatically without being explicitly programmed.

- ## 2014: DeepFace

DeepFace is a deep learning algorithm which was originally developed in 2014 and is part of the company "Meta". The project received significant media attention after it outperformed human performance on the well-known "Faces in the Wild" test.

DeepFace is based on a deep neural network, which consists of many layers of artificial neurons and weights that connect each layer to its neighboring ones. The algorithm takes as input a training data set of photographs, with each photo annotated with the identity and age of its subject. The team has been very successful in recent years and published many papers on their research results. They have also trained several deep neural networks that have achieved significant success in pattern recognition and machine learning tasks.

- ## 2017: ImageNet Challenge – Milestone in the History of Machine Learning

  The **ImageNet Challenge** is a competition in computer vision that has been running since 2010. The challenge focuses on the abilities of programs to process patterns in images and recognize objects with varying degrees of detail.

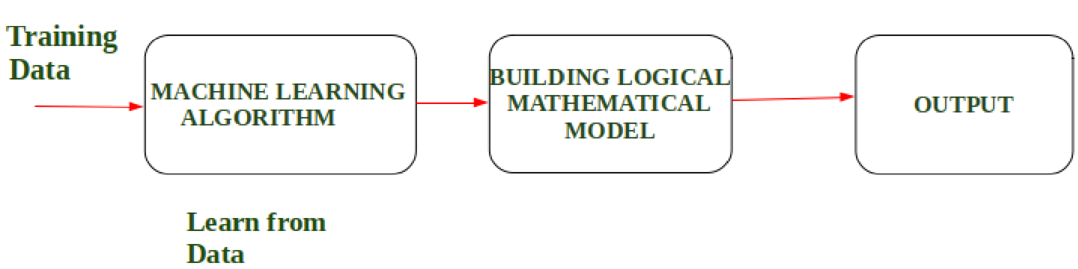
- In 2017, a milestone was reached. **29 out of 38 teams achieved 95%** accuracy with their computer vision models. The improvement in image recognition is immense.

➢ Features of Machine learning

- Machine learning is data driven technology. Large amount of data generated by organizations on daily bases. So, by notable relationships in data, organizations makes better decisions.
- Machine can learn itself from past data and automatically improve.
- From the given dataset it detects various patterns on data.
- For the big organizations branding is important and it will become more easy to target relatable customer base.
- It is similar to data mining because it is also deals with the huge amount of data.

➢ Mathematical model of Machine Learning

When we fed the Training Data to Machine Learning Algorithm, this algorithm will produce a mathematical model and with the help of the mathematical model, the machine will make a prediction and take a decision without being explicitly programmed.



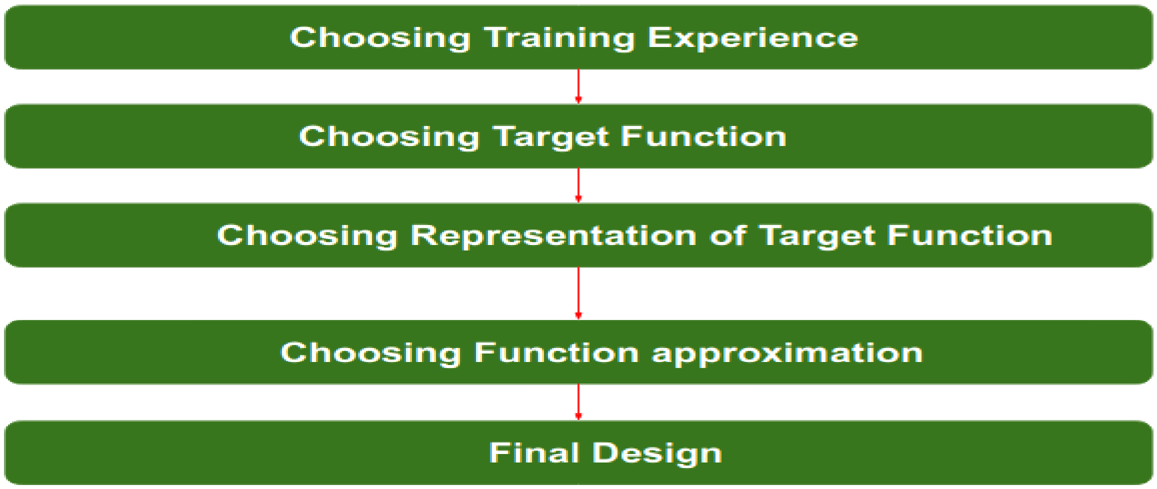
Example : In Driverless Car, the training data is fed to Algorithm like how to Drive Car in Highway, Busy and Narrow Street with factors like speed limit, parking, stop at signal etc. After that, a Logical and Mathematical model is created on the basis of that and after that, the car will work according to the logical model. Also, the more data the data is fed the more efficient output is produced.

Designing a Learning System in Machine Learning :
According to Tom Mitchell, "A computer program is said to be learning from experience (E), with respect to some task (T). Thus, the performance measure (P) is the performance at task T, which is measured by P, and it improves with experience E."

Example: In Spam E-Mail detection,

- **Task, T:** To classify mails into Spam or Not Spam.
- **Performance measure, P:** Total percent of mails being correctly classified as being "Spam" or "Not Spam".
- **Experience, E:** Set of Mails with label "Spam"

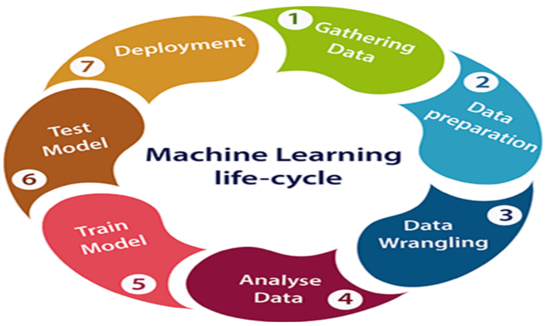
Steps for Designing Learning System are:



- **Machine learning life cycle**

Machine learning life cycle involves seven major steps, which are given below:

- **Gathering Data**
- **Data preparation**
- **Data Wrangling**
- **Analyse Data**
- **Train the model**
- **Test the model**
- **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 preprocessing 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

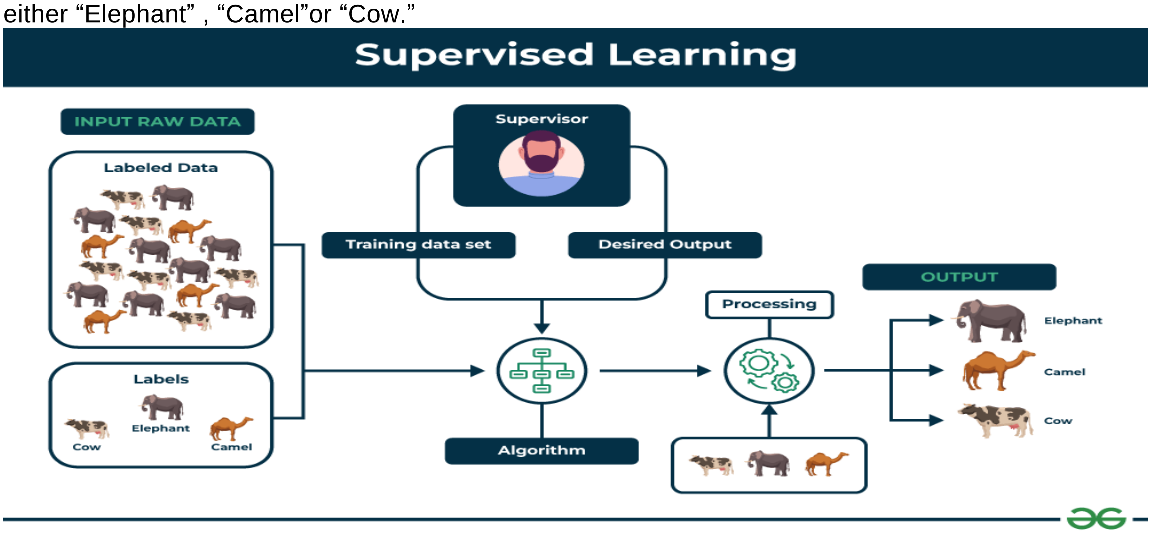
The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system. If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will 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.

➢ Classification of Machine Learning:

1. Supervised learning

Supervised learning is a type of machine learning algorithm that learns from labeled data. Labeled data is data that has been tagged with a correct answer or classification.
Supervised learning, as the name indicates, has the presence of a supervisor as a teacher. Supervised learning is when we teach or train the machine using data that is well-labeled. Which means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples (data) so that the supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

For example, a labeled dataset of images of Elephant, Camel and Cow would have each image tagged with either "Elephant" , "Camel"or "Cow."



Key Points:

- Supervised learning involves training a machine from labeled data.
- Labeled data consists of examples with the correct answer or classification.
- The machine learns the relationship between inputs (fruit images) and outputs (fruit labels).
- The trained machine can then make predictions on new, unlabeled data.

Example:

Let's say you have a fruit basket that you want to identify. The machine would first analyze the image to extract features such as its shape, color, and texture. Then, it would compare these features to the features of the fruits it has already learned about. If the new image's features are most similar to those of an apple, the machine would predict that the fruit is an apple.
For instance, suppose you are given a basket filled with different kinds of fruits. Now the first step is to train the machine with all the different fruits one by one like this:

- If the shape of the object is rounded and has a depression at the top, is red in color, then it will be labeled as –**Apple**.
- If the shape of the object is a long curving cylinder having Green-Yellow color, then it will be labeled as –**Banana**.

Now suppose after training the data, you have given a new separate fruit, say Banana from the basket, and asked to identify it.
Since the machine has already learned the things from previous data and this time has to use it wisely. It will first classify the fruit with its shape and color and would confirm the fruit name as BANANA and put it in the Banana category. Thus the machine learns the things from training data(basket containing fruits) and then applies the knowledge to test data(new fruit).

Types of Supervised Learning

Supervised learning is classified into two categories of algorithms:

- **Regression**: A regression problem is when the output variable is a real value, such as "dollars" or "weight".
- **Classification**: A classification problem is when the output variable is a category, such as "Red" or "blue" , "disease" or "no disease".

Supervised learning deals with or learns with "labeled" data. This implies that some data is already tagged with the correct answer.

1- Regression

Regression is a type of supervised learning that is used to predict continuous values, such as house prices, stock prices, or customer churn. Regression algorithms learn a function that maps from the input features to the output value.

Some common regression algorithms include:

- Linear Regression
- Polynomial Regression
- Support Vector Machine Regression
- Decision Tree Regression
- Random Forest Regression

2- Classification

Classification is a type of supervised learning that is used to predict categorical values, such as whether a customer will churn or not, whether an email is spam or not, or whether a medical image shows a tumor or not. Classification algorithms learn a function that maps from the input features to a probability distribution over the output classes.

Some common classification algorithms include:

- Logistic Regression
- Support Vector Machines
- Decision Trees
- Random Forests
- Naive Baye

Evaluating Supervised Learning Models

Evaluating supervised learning models is an important step in ensuring that the model is accurate and generalizable. There are a number of different metrics that can be used to evaluate supervised learning models, but some of the most common ones include:

For Regression

- **Mean Squared Error (MSE):** MSE measures the average squared difference between the predicted values and the actual values. Lower MSE values indicate better model performance.
- **Root Mean Squared Error (RMSE):** RMSE is the square root of MSE, representing the standard deviation of the prediction errors. Similar to MSE, lower RMSE values indicate better model performance.
- **Mean Absolute Error (MAE):** MAE measures the average absolute difference between the predicted values and the actual values. It is less sensitive to outliers compared to MSE or RMSE.
- **R-squared (Coefficient of Determination):** R-squared measures the proportion of the variance in the target variable that is explained by the model. Higher R-squared values indicate better model fit.

For Classification

- **Accuracy:** Accuracy is the percentage of predictions that the model makes correctly. It is calculated by dividing the number of correct predictions by the total number of predictions.
- **Precision:** Precision is the percentage of positive predictions that the model makes that are actually correct. It is calculated by dividing the number of true positives by the total number of positive predictions.
- **Recall:** Recall is the percentage of all positive examples that the model correctly identifies. It is calculated by dividing the number of true positives by the total number of positive examples.
- **F1 score:** The F1 score is a weighted average of precision and recall. It is calculated by taking the harmonic mean of precision and recall.
- **Confusion matrix:** A confusion matrix is a table that shows the number of predictions for each class, along with the actual class labels. It can be used to visualize the performance of the model and identify areas where the model is struggling.

Applications of Supervised learning

Supervised learning can be used to solve a wide variety of problems, including:

- **Spam filtering:** Supervised learning algorithms can be trained to identify and classify spam emails based on their content, helping users avoid unwanted messages.
- **Image classification:** Supervised learning can automatically classify images into different categories, such as animals, objects, or scenes, facilitating tasks like image search, content moderation, and image-based product recommendations.

- **Medical diagnosis:** Supervised learning can assist in medical diagnosis by analyzing patient data, such as medical images, test results, and patient history, to identify patterns that suggest specific diseases or conditions.
- **Fraud detection:** Supervised learning models can analyze financial transactions and identify patterns that indicate fraudulent activity, helping financial institutions prevent fraud and protect their customers.
- **Natural language processing (NLP):** Supervised learning plays a crucial role in NLP tasks, including sentiment analysis, machine translation, and text summarization, enabling machines to understand and process human language effectively.

Advantages of Supervised learning

- Supervised learning allows collecting data and produces data output from previous experiences.
- Helps to optimize performance criteria with the help of experience.
- Supervised machine learning helps to solve various types of real-world computation problems.
- It performs classification and regression tasks.
- It allows estimating or mapping the result to a new sample.
- We have complete control over choosing the number of classes we want in the training data.

Disadvantages of Supervised learning

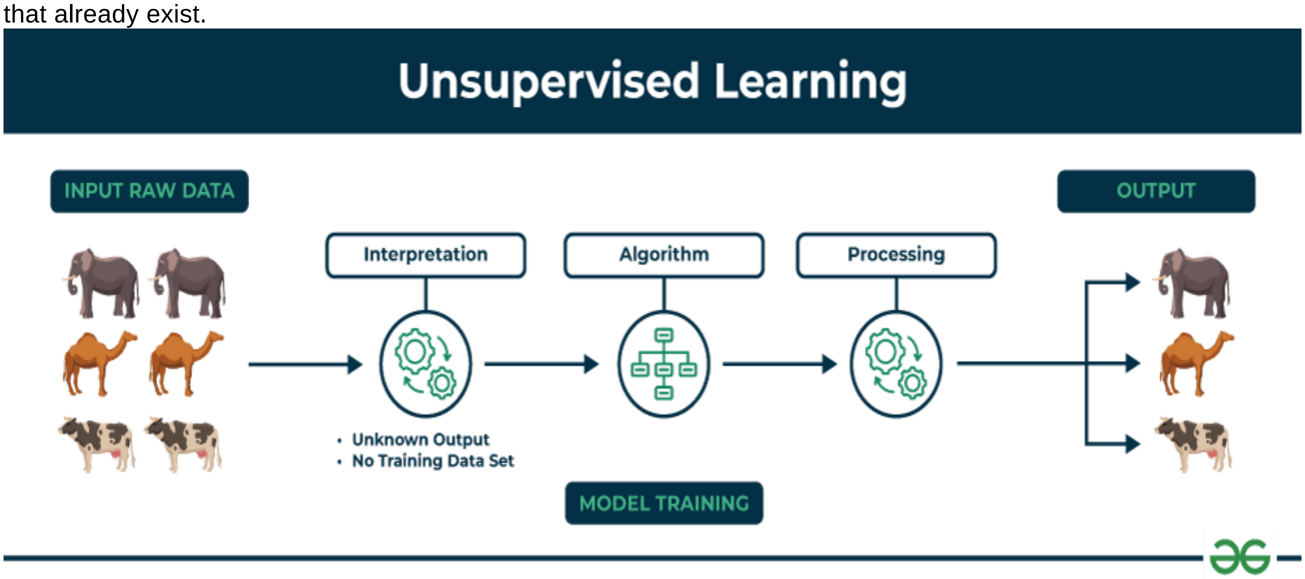
- Classifying big data can be challenging.
- Training for supervised learning needs a lot of computation time. So, it requires a lot of time.
- Supervised learning cannot handle all complex tasks in Machine Learning.
- Computation time is vast for supervised learning.
- It requires a labelled data set.
- It requires a training process.

➢ Unsupervised learning

Unsupervised learning is a type of machine learning that learns from unlabeled data. This means that the data does not have any pre-existing labels or categories. The goal of unsupervised learning is to discover patterns and relationships in the data without any explicit guidance.

Unsupervised learning is the training of a machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data.
Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself.

You can use unsupervised learning to examine the animal data that has been gathered and distinguish between several groups according to the traits and actions of the animals. These groupings might correspond to various animal species, providing you to categorize the creatures without depending on labels that already exist.



Key Points

- Unsupervised learning allows the model to discover patterns and relationships in unlabeled data.
- Clustering algorithms group similar data points together based on their inherent characteristics.
- Feature extraction captures essential information from the data, enabling the model to make meaningful distinctions.
- Label association assigns categories to the clusters based on the extracted patterns and characteristics.

Example

Imagine you have a machine learning model trained on a large dataset of unlabeled images, containing both dogs and cats. The model has never seen an image of a dog or cat before, and it has no pre-existing labels or categories for these animals. Your task is to use unsupervised learning to identify the dogs and cats in a new, unseen image.

For instance, suppose it is given an image having both dogs and cats which it has never seen.
Thus the machine has no idea about the features of dogs and cats so we can't categorize it as 'dogs and cats '. But it can categorize them according to their similarities, patterns, and differences, i.e., we can easily categorize the above picture into two parts. The first may contain all pics having **dogs** in them and the second part may contain all pics having **cats** in them. Here you didn't learn anything before, which means no training data or examples. It allows the model to work on its own to discover patterns and information that was previously undetected. It mainly deals with unlabelled data.

Types of Unsupervised Learning

Unsupervised learning is classified into two categories of algorithms:

- **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

1. Clustering

Clustering is a type of unsupervised learning that is used to group similar data points together. Clustering algorithms work by iteratively moving data points closer to their cluster centers and further away from data points in other clusters.

1. Exclusive (partitioning)
2. Agglomerative
3. Overlapping
4. Probabilistic

Clustering Types:-

1. Hierarchical clustering
2. K-means clustering
3. Principal Component Analysis
4. Singular Value Decomposition
5. Independent Component Analysis
6. Gaussian Mixture Models (GMMs)
7. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

2. Association rule learning

Association rule learning is a type of unsupervised learning that is used to identify patterns in a data. Association rule learning algorithms work by finding relationships between different items in a dataset.
Some common association rule learning algorithms include:

- Apriori Algorithm
- Eclat Algorithm
- FP-Growth Algorithm

Evaluating Non-Supervised Learning Models

Evaluating non-supervised learning models is an important step in ensuring that the model is effective and useful. However, it can be more challenging than evaluating supervised learning models, as there is no ground truth data to compare the model's predictions to.
There are a number of different metrics that can be used to evaluate non-supervised learning models, but some of the most common ones include:

- **Silhouette score:** The silhouette score measures how well each data point is clustered with its own cluster members and separated from other clusters. It ranges from -1 to 1, with higher scores indicating better clustering.
- **Calinski-Harabasz score:** The Calinski-Harabasz score measures the ratio between the variance between clusters and the variance within clusters. It ranges from 0 to infinity, with higher scores indicating better clustering.
- **Adjusted Rand index:** The adjusted Rand index measures the similarity between two clusterings. It ranges from -1 to 1, with higher scores indicating more similar clusterings.
- **Davies-Bouldin index:** The Davies-Bouldin index measures the average similarity between clusters. It ranges from 0 to infinity, with lower scores indicating better clustering.
- **F1 score:** The F1 score is a weighted average of precision and recall, which are two metrics that are commonly used in supervised learning to evaluate classification models. However, the F1 score can also be used to evaluate non-supervised learning models, such as clustering models.

Application of Unsupervised learning

Non-supervised learning can be used to solve a wide variety of problems, including:

- Anomaly detection: Unsupervised learning can identify unusual patterns or deviations from normal behavior in data, enabling the detection of fraud, intrusion, or system failures.
- Scientific discovery: Unsupervised learning can uncover hidden relationships and patterns in scientific data, leading to new hypotheses and insights in various scientific fields.
- Recommendation systems: Unsupervised learning can identify patterns and similarities in user behavior and preferences to recommend products, movies, or music that align with their interests.
- Customer segmentation: Unsupervised learning can identify groups of customers with similar characteristics, allowing businesses to target marketing campaigns and improve customer service more effectively.
- Image analysis: Unsupervised learning can group images based on their content, facilitating tasks such as image classification, object detection, and image retrieval.

Advantages of Unsupervised learning

- It does not require training data to be labeled.
- Dimensionality reduction can be easily accomplished using unsupervised learning.
- Capable of finding previously unknown patterns in data.
- Unsupervised learning can help you gain insights from unlabeled data that you might not have been able to get otherwise.
- Unsupervised learning is good at finding patterns and relationships in data without being told what to look for. This can help you learn new things about your data.

Disadvantages of Unsupervised learning

- Difficult to measure accuracy or effectiveness due to lack of predefined answers during training.
- The results often have lesser accuracy.
- The user needs to spend time interpreting and label the classes which follow that classification.
- Unsupervised learning can be sensitive to data quality, including missing values, outliers, and noisy data.
- Without labeled data, it can be difficult to evaluate the performance of unsupervised learning models, making it challenging to assess their effectiveness.

➢ Supervised vs. Unsupervised Machine Learning

Parameters	**Supervised Machine learning**	**Unsupervised machine learning**
Input Data	Algorithms are trained using labeled data.	Algorithms are used against data that is not labeled
Computational Complexity	Simpler method	Computationally complex
Accuracy	Highly accurate	Less accurate

Parameters	Supervised Machine learning	Unsupervised machine learning
No. of classes	No. of classes is known	No. of classes is not known
Data Analysis	Uses offline analysis	Uses real-time analysis of data
Algorithms used	Linear and Logistics regression, Random forest, Support Vector Machine, Neural Network, etc.	K-Means clustering, Hierarchical clustering, Apriori algorithm, etc.
Output	Desired output is given.	Desired output is not given.
Training data	Use training data to infer model.	No training data is used.
Complex model	It is not possible to learn larger and more complex models than with supervised learning.	It is possible to learn larger and more complex models with unsupervised learning.
Model	We can test our model.	We can not test our model.
Called as	Supervised learning is also called classification.	Unsupervised learning is also called clustering.
Example	Example: Optical character recognition.	Example: Find a face in an image.

➢ Supervised Learning and Reinforcement Learning Comparison

BASIS FOR COMPARISON	Supervised Learning	Reinforcement learning
Definition	Works on existing or given sample data or examples	Works on interacting with the environment
Preference	Preferred in generalized working mechanisms where routine tasks are required to be done	Preferred in the area of Artificial Intelligence
Area	Comes under the area of Machine Learning	Comes under the area of Machine Learning
Platform	Operated with interactive software systems or applications	Supports and works better in Artificial Intelligence where Human Interaction is prevalent
Generality	Many open source projects are evolving development in this area	More useful in Artificial Intelligence
Algorithm	Many algorithms exist in using this learning	Neither supervised nor unsupervised algorithms are used
Integration	Runs on any platform or with any applications	Runs with any hardware or software devices

➢ Applications of Machine Learning

Machine learning has become a ubiquitous technology that has impacted many aspects of our lives, from business to healthcare to entertainment. Here are some popular applications of machine learning −

1. Image and Speech Recognition

Image and speech recognition are two areas where machine learning has made significant advancements. Machine learning algorithms are used in applications such as facial recognition, object detection, and speech recognition to accurately identify and classify images and speech.

2. Natural Language Processing

Natural Language Processing (NLP) is a field of computer science that deals with the interaction between computers and humans using natural language. NLP uses machine learning algorithms to analyze, understand, and generate human language, making it possible to build chatbots, language translators, and voice assistants.

3. Fraud Detection

Machine learning is widely used in the finance industry for fraud detection. Machine learning algorithms can analyze vast amounts of transactional data to detect patterns and anomalies that may indicate fraudulent activity, helping to prevent financial losses and protect customers.

4. Predictive Maintenance

Predictive maintenance is a process of using machine learning algorithms to predict when maintenance will be required on a machine, such as a piece of equipment in a factory. By analyzing data from sensors and other sources, machine learning algorithms can detect patterns that indicate when a machine is likely to fail, enabling maintenance to be performed before the machine breaks down.

5. Healthcare

Machine learning has also found many applications in the healthcare industry. For example, machine learning algorithms can be used to analyze medical images and detect diseases such as cancer, or to predict patient outcomes based on their medical history and other factors.

6. Recommendation Systems

Recommendation systems are used to provide personalized recommendations to users based on their past behavior and preferences. Machine learning algorithms are used to analyze user data and generate recommendations for products, services, and content.

An example of these services is very common for example YouTube. It recommends new videos and content based on the users past search patterns. Netflix recommends movies and series based on the interest provided by users when someone creates an account for the very first time.

7. Autonomous Vehicles

Machine learning is a critical technology for the development of autonomous vehicles. Machine learning algorithms are used to process data from sensors and cameras, allowing vehicles to detect and respond to their environment in real-time. The most common example of this use case is that of the Tesla cars which are well-tested and proven for autonomous driving.