Training Day-51 Report:

Introduction to Machine Learning :-

- What Is Machine Learning?
  Machine learning (ML) is a discipline of artificial intelligence (AI) that provides machines with the ability to automatically learn from data and past experiences while identifying patterns to make predictions with minimal human intervention.
- Machine learning methods enable computers to operate autonomously without explicit programming. ML applications are fed with new data, and they can independently learn, grow, develop, and adapt.
- Machine learning derives insightful information from large volumes of data by leveraging algorithms to identify patterns and learn in an iterative process. ML algorithms use computation methods to learn directly from data instead of relying on any predetermined equation that may serve as a model.

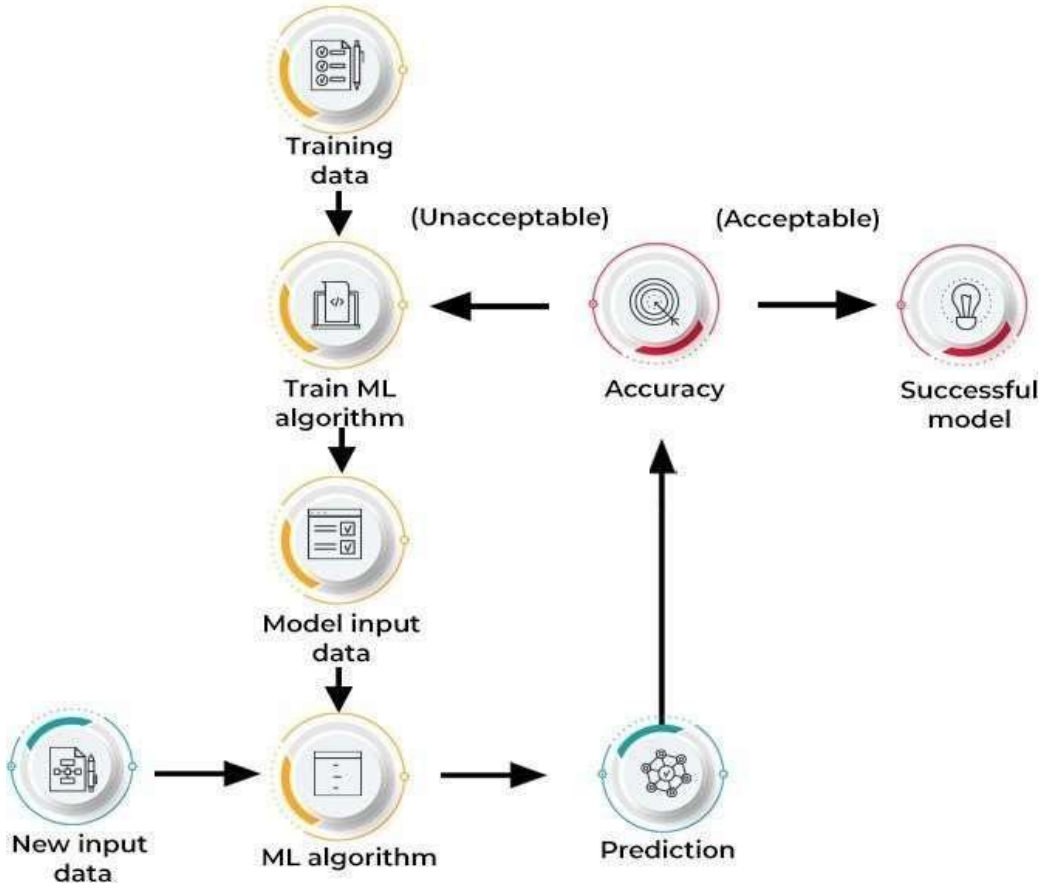
Today, with the rise of big data, IoT, and ubiquitous computing, machine learning has become essential for solving problems across numerous areas, such as:

- Computational finance (credit scoring, algorithmic trading)
- Computer vision (facial recognition, motion tracking, object detection)
- Computational biology (DNA sequencing, brain tumor detection, drug discovery)
- Automotive, aerospace, and manufacturing (predictive maintenance)
- Natural language processing (voice recognition)

How does machine learning work?

Machine learning algorithms are molded on a training dataset to create a model. As new input data is introduced to the trained ML algorithm, it uses the developed model to make a prediction.

HOW DOES MACHINE LEARNING WORK?



Training Day-52 Report:

Types of Machine Learning:-

Machine learning algorithms can be trained in many ways, with each method having its pros and cons. Based on these methods and ways of learning, machine learning is broadly categorized into three main types:

1. Supervised Machine Learning:-

This type of ML involves supervision, where machines are trained on labeled datasets and enabled to predict outputs based on the provided training. The labeled dataset specifies that some input and output parameters are already mapped. Hence, the machine is trained with the input and corresponding output. A device is made to predict the outcome using the test dataset in subsequent phases.
For example, consider an input dataset of parrot and crow images. Initially, the machine is trained to understand the pictures, including the parrot and crow's color, eyes, shape, and size. Post-training, an input picture of a parrot is provided, and the machine is expected to identify the object and predict the output. The trained machine checks for the various features of the object, such as color, eyes, shape, etc., in the input picture, to make a final prediction. This is the process of object identification in supervised machine learning.
The primary objective of the supervised learning technique is to map the input variable
(a) with the output variable
(b). Supervised machine learning is further classified into two broad categories:

- Classification: These refer to algorithms that address classification problems where the output variable is categorical; for example, yes or no, true or false, male or female, etc. Real-world applications of this category are evident in spam detection and email filtering.
- Some known classification algorithms include the Random Forest Algorithm, Decision Tree Algorithm, Logistic Regression Algorithm, and Support Vector Machine Algorithm.

Regression: Regression algorithms handle regression problems where input and output variables have a linear relationship. These are known to predict continuous output variables. Examples include weather prediction, market trend analysis, etc.
Popular regression algorithms include the Simple Linear Regression Algorithm, Multivariate Regression Algorithm, Decision Tree Algorithm, and Lasso Regression.

2. **Unsupervised Machine Learning:-**

Unsupervised learning refers to a learning technique that's devoid of supervision. Here, the machine is trained using an unlabeled dataset and is enabled to predict the output without any supervision. An unsupervised learning algorithm aims to group the unsorted dataset based on the input's similarities, differences, and patterns.

For example, consider an input dataset of images of a fruit-filled container. Here, the images are not known to the machine learning model. When we input the dataset into the ML model, the task of the model is to identify the pattern of objects, such as color, shape, or differences seen in the input images and categorize them. Upon categorization, the machine then predicts the output as it gets tested with a test dataset. Unsupervised machine learning is further classified into two types:

- Clustering: The clustering technique refers to grouping objects into clusters based on parameters such as similarities or differences between objects. For example, grouping customers by the products they purchase.
  Some known clustering algorithms include the K-Means Clustering Algorithm, Mean-Shift Algorithm, DBSCAN Algorithm, Principal Component Analysis, and Independent Component Analysis.
- Association: Association learning refers to identifying typical relations between the variables of a large dataset. It determines the dependency of various data items and maps associated variables. Typical applications include web usage mining and market data analysis.

Popular algorithms obeying association rules include the Apriori Algorithm, Eclat Algorithm, and FP-Growth Algorithm.

3. **Reinforcement Learning:-**
   Reinforcement learning is a feedback-based process. Here, the AI component automatically takes stock of its surroundings by the hit & trial method, takes action, learns from experiences, and improves performance. The component is rewarded for each good action and penalized for every wrong move. Thus, the reinforcement learning component aims to maximize the rewards by performing good actions.

   Unlike supervised learning, reinforcement learning lacks labeled data, and the agents learn via experiences only. Consider video games. Here, the game specifies the environment, and each move of the reinforcement agent defines its state. The agent is entitled to receive feedback via punishment and rewards, thereby affecting the overall game score. The ultimate goal of the agent is to achieve a high score. Reinforcement learning is applied across different fields such as game theory, information theory, and multi-agent systems. Reinforcement learning is further divided into two types of methods or algorithms:

   - Positive reinforcement learning: This refers to adding a reinforcing stimulus after a specific behavior of the agent, which makes it more likely that the behavior may occur again in the future, e.g., adding a reward after a behavior.
   - Negative reinforcement learning: Negative reinforcement learning refers to strengthening a specific behavior that avoids a negative outcome.

Training Day-53 Report:

Supervised Learning:-

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

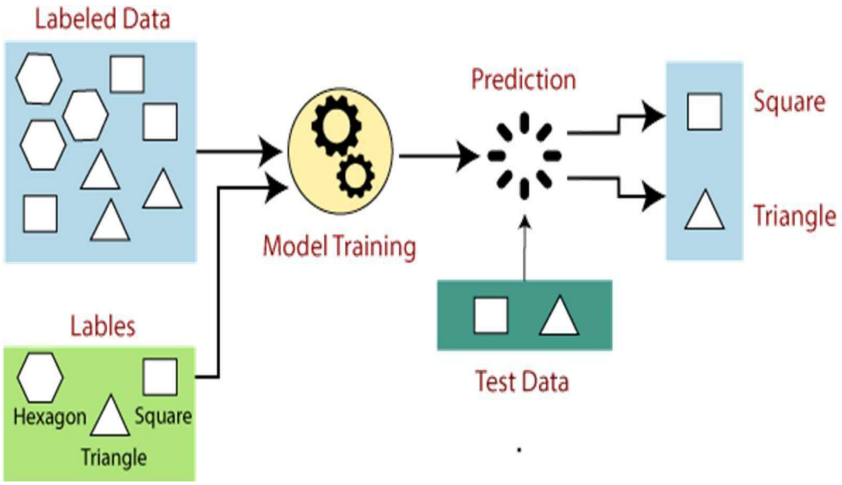
Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to **find a mapping function to map the input variable(x) with the output variable(y)**.

In the real-world, supervised learning can be used for **Risk Assessment, Image classification, Fraud Detection, spam filtering**, etc.

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.

- If the given shape has four sides, and all the sides are equal, then it will be labelled as a **Square**.
- If the given shape has three sides, then it will be labelled as a **triangle**.
- 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:

- First Determine the type of training dataset oCollect/Gather the labelled training data.
- Split the training dataset into training **dataset, test dataset, and validation dataset**.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.

- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

Training Day-54 Report:

Types of Supervised Learning:-

Supervised learning is typically divided into two main categories: regression and classification. In regression, the algorithm learns to predict a continuous output value, such as the price of a house or the temperature of a city. In classification, the algorithm learns to predict a categorical output variable or class label, such as whether a customer is likely to purchase a product or not.

- **Regression**

  Regression is a supervised learning technique used to predict continuous numerical values based on input features. It aims to establish a functional relationship between independent variables and a dependent variable, such as predicting house prices based on features like size, bedrooms, and location.

  The goal is to minimize the difference between predicted and actual values using algorithms like Linear Regression, Decision Trees, or Neural Networks, ensuring the model captures underlying patterns in the data.

- **Classification**

  Classification is a type of supervised learning that categorizes input data into predefined labels. It involves training a model on labeled examples to learn patterns between input features and output classes. In classification, the target variable is a categorical value. For example, classifying emails as spam or not.

  The model's goal is to generalize this learning to make accurate predictions on new, unseen data. Algorithms like Decision Trees, Support Vector Machines, and Neural Networks are commonly used for classification tasks.

Training Day-55 Report:

What is Regression?

Regression is a statistical approach used to analyze the relationship between a dependent variable (target variable) and one or more independent variables (predictor variables). The objective is to determine the most suitable function that characterizes the connection between these variables.
It seeks to find the best-fitting model, which can be utilized to make predictions or draw conclusions.

Regression in Machine Learning:-

It is a supervised machine learning technique, used to predict the value of the dependent variable for new, unseen data. It models the relationship between the input features and the target variable, allowing for the estimation or prediction of numerical values.
Regression analysis problem works with if output variable is a real or continuous value, such as “salary” or “weight”. Many different models can be used, the simplest is the linear regression. It tries to fit data with the best hyper- plane which goes through the points.

Terminologies Related to the Regression Analysis in Machine Learning:-

Terminologies Related to Regression Analysis:

- **Response Variable**: The primary factor to predict or understand in regression, also known as the dependent variable or target variable.
- **Predictor Variable**: Factors influencing the response variable, used to predict its values; also called independent variables.
- **Outliers**: Observations with significantly low or high values compared to others, potentially impacting results and best avoided.
- **Multicollinearity**: High correlation among independent variables, which can complicate the ranking of influential variables.
- **Underfitting and Overfitting**: Overfitting occurs when an algorithm performs well on training but poorly on testing, while underfitting indicates poor performance on both datasets.

Regression Types

There are two main types of regression:

- **Simple Regression**
  - Used to predict a continuous dependent variable based on a single independent variable.
  - Simple linear regression should be used when there is only a single independent variable. □

- **Multiple Regression**

    - Used to predict a continuous dependent variable based on multiple independent variables.
    - Multiple linear regression should be used when there are multiple independent variables.

- **Non-Linear Regression**
    - Relationship between the dependent variable and independent variable(s) follows a nonlinear pattern.

    - Provides flexibility in modeling a wide range of functional forms.