**DAY 4 :BASICS OF MACHINE LEARNING**

**WHY IS ML USED IN CURRENT INDUSTRIES AND HOW CAN IT BE USED?**

**Healthcare**

Machine learning has revolutionized the healthcare sector by enabling early and accurate disease diagnosis, such as for cancer, COVID-19, and diabetes. It plays a major role in analyzing medical images like X-rays, MRIs, and CT scans with high precision. ML is also instrumental in accelerating drug discovery and understanding genetic data. Furthermore, it supports predictive modeling to forecast patient health outcomes and aids in designing personalized treatment plans tailored to individual patients.

**Banking and Finance**

In the financial industry, machine learning is extensively used to assess credit risk and improve fraud detection systems. It enables banks to automate decision-making for loan approvals and provides the foundation for high-speed algorithmic trading. Customer service is also enhanced through AI-powered chatbots. Moreover, machine learning models help in predicting stock market trends, optimizing investment strategies, and ensuring regulatory compliance through intelligent data analysis.

**Retail and E-commerce**

Retailers and e-commerce platforms leverage machine learning to enhance customer experience through recommendation systems, such as those seen on Amazon and Flipkart. By analyzing customer behavior, ML helps segment users and predict purchasing patterns. It also supports dynamic pricing models and ensures efficient supply chain and inventory management. Additionally, it can extract valuable insights from customer reviews using sentiment analysis.

**Automotive Industry**

Machine learning is a driving force behind innovations in the automotive field. It powers self-driving vehicles by helping them navigate autonomously and make real-time decisions. ML algorithms are used for predictive maintenance, enabling early detection of mechanical issues. It also enhances driver assistance technologies and facilitates traffic forecasting and route optimization. Voice-controlled AI systems in cars further improve user interaction and convenience.

**Technology and IT**

The tech industry is heavily reliant on machine learning for various applications such as natural language processing, which enables tasks like language translation and intelligent chatbots. It’s also used in image and speech recognition systems, spam email filtering, and virtual assistants like Siri and Alexa. Cybersecurity has greatly benefited from ML with models capable of detecting threats and anomalies in real time.

**Manufacturing**

In manufacturing, machine learning optimizes operations by predicting equipment failures and scheduling preventive maintenance. It helps maintain product quality through computer vision-based defect detection and supports accurate demand forecasting. ML models also streamline production processes by identifying inefficiencies and automating decision-making on the assembly line.

**Education**

The education sector uses machine learning to personalize learning experiences by recommending content tailored to each student's learning style. It can predict student performance and identify those at risk of falling behind. ML also enables automated grading systems and AI tutors that offer interactive and engaging learning. Plagiarism detection tools further maintain academic integrity using natural language analysis.

**Media and Entertainment**

Machine learning enhances user engagement in the media industry through content recommendation systems on platforms like Netflix and YouTube. It performs sentiment and emotion analysis to understand viewer reactions and deliver targeted content. Advertisers use ML to create personalized marketing campaigns, while other applications include fake news detection and tagging of audio and video content for better accessibility.

**Travel and Transportation**

The travel industry applies machine learning for dynamic pricing of flight fares and hotel rates. It powers recommendation engines that suggest destinations and travel itineraries. ML-based chatbots and virtual agents improve customer support, while predictive models forecast delays and optimize traffic flow. Additionally, autonomous vehicles and drones rely heavily on ML to operate safely and efficiently.

**Agriculture**

Machine learning aids modern agriculture by predicting crop yields based on various environmental factors. It helps detect diseases and pests using drones equipped with image analysis systems. Farmers can monitor soil health and automate irrigation with smart systems powered by ML. It also supports market price forecasting, helping farmers make informed decisions on crop sales.

**Telecommunications**

Telecom companies use machine learning for optimizing their networks and providing personalized plans to users. ML helps predict customer churn, allowing companies to take preventive actions. It also assists in fault detection and automatic resolution of technical issues. Speech recognition technologies in telecom applications enable natural voice interaction, enhancing customer service.

**TYPES OF MACHINE LEARNING TECHNIQUES SO FAR:**

1. **SUPERVISED ML**
2. **UNSUPERVISED ML**
3. **REINFORCEMENT ML**

* **SUPERVISED ML:**

Learn to predict target values from labelled data

It can be

1)Classification(here, the target values are discrete values)like a yes or no answer eg.spam and non-spam email

2)Regression(here, the target values are continuous values)like house prices,stock prices etc.

It involves training a model using labeled data, where each input comes with a corresponding correct output.

A fundamental concept in supervised machine learning is learning a class from examples. This involves providing the model with examples where the correct label is known, such as learning to classify images of cats and dogs by being shown labeled examples of both. The model then learns the distinguishing features of each class and applies this knowledge to classify new images.

**How Supervised Machine Learning Works?**

Where **supervised learning algorithm** consists of input features and corresponding output labels. The process works through:

* **Training Data:**The model is provided with a training dataset that includes input data (features) and corresponding output data (labels or target variables).
* **Learning Process:**The algorithm processes the training data, learning the relationships between the input features and the output labels. This is achieved by adjusting the model's parameters to minimize the difference between its predictions and the actual labels.

After training, the model is evaluated using atest dataset to measure its accuracy and performance. Then the model's performance is optimized by adjusting parameters and using techniques like cross-validation to balance bias and variance. This ensures the model generalizes well to new, unseen data.

While training the model, data is usually split in the ratio of 80:20 i.e. 80% as training data and the rest as testing data. In training data, we feed input as well as output for 80% of data. The model learns from training data only.

**Practical Examples of Supervised learning**

Few practical examples of supervised machine learning across various industries:

* [**Fraud Detection in Banking**](https://www.geeksforgeeks.org/online-payment-fraud-detection-using-machine-learning-in-python/)**:** Utilizes supervised learning algorithms on historical transaction data, training models with labeled datasets of legitimate and fraudulent transactions to accurately predict fraud patterns.
* [**Parkinson Disease Prediction:**](https://www.geeksforgeeks.org/parkinson-disease-prediction-using-machine-learning-python/)Parkinson’s disease is a progressive disorder that affects the nervous system and the parts of the body controlled by the nerves.
* [**Cancer cell classification:**](https://www.geeksforgeeks.org/ml-cancer-cell-classification-using-scikit-learn/)Implements supervised learning for cancer cells based on their features, and identifying them if they are ‘malignant’ or ‘benign

Supervised ML can be further categorized into various algorithms depending on their characteristics and applications

1)Linear Regression model is used to predict a continuous value

2)Logistic Regression model is used to predict a binary output variable

3)Decision Tree model is used to model decisions and their possible consequences. Each internal node in the tree represents a decision, while each leaf node represents a possible outcome.

4)Random forests again are made up of multiple decision trees that work together to make predictions. Each tree in the forest is trained on a different subset of the input features and data. The final prediction is made by aggregating the predictions of all the trees in the forest.

5)Gradient Boosting combines weak learners, like decision trees, to create a strong model. It iteratively builds new models that correct errors made by previous ones.

6)The SVM algorithm creates a hyperplane to segregate n-dimensional space into classes and identify the correct category of new data points.

Training a Supervised model:

**1. Data Collection and Preprocessing**  
Start by collecting a labeled dataset that includes input features and corresponding output labels. Clean the data to address missing values and apply feature scaling if necessary to maintain data quality for supervised learning.

**2. Splitting the Data**  
Divide the dataset into two parts: 80% for training and 20% for testing to evaluate performance on unseen data.

**3. Model Selection**  
Choose the most suitable machine learning algorithm depending on the nature of the problem. This decision greatly impacts the effectiveness of supervised learning.

**4. Model Training**  
Train the selected model by providing it with the training data, enabling it to learn and adjust internal parameters based on the patterns it identifies.

**5. Model Evaluation**  
Assess how well the model performs using the test set, applying relevant evaluation metrics such as accuracy, precision, recall, or F1-score.

**6. Hyperparameter Optimization**  
Fine-tune the model’s hyperparameters (like learning rate or tree depth) using methods such as grid search or cross-validation to improve performance.

**7. Final Training and Testing**  
Retrain the model using the full training dataset and optimized hyperparameters, then test it once more to confirm it performs well before moving to production.

**8. Model Deployment**  
Deploy the final, validated model to begin making predictions on real-world or new, unseen data.

Unsupervised learning is a type of machine learning where algorithms work with **unlabeled data**—meaning the data doesn’t come with predefined categories or outcomes. Unlike supervised learning, where output labels are given, unsupervised learning models aim to discover hidden structures, groupings, or relationships within the input data entirely on their own, without human guidance.

In this approach, the model receives only input features and tries to uncover patterns or natural groupings in the data.

**Example Illustration**

Imagine a dataset of various animals—such as elephants, camels, and cows. This raw, unlabeled data is processed by an unsupervised algorithm.

* The **interpretation stage** shows that the algorithm has no prior knowledge about animal types.
* During the **processing phase**, it uses techniques like clustering to analyze the data.
* The **output stage** might reveal that the algorithm successfully grouped the animals into distinct categories (species) based on visual or numerical features—even though it wasn’t told what a cow or elephant is.

**How Does Unsupervised Learning Work?**

The algorithm analyzes unlabelled input data and tries to extract meaningful patterns, similarities, or structures from it. This process can be complex, but it helps discover valuable insights that might not be visible with labeled data.

**Example Use Case:**

A shopping mall collects customer data through membership cards, tracking each individual’s purchases. By applying unsupervised learning, the mall can **segment customers** based on their behavior or preferences—even though there are no explicit labels like "frequent buyer" or "budget shopper."

**Characteristics of Input Data in Unsupervised Learning:**

* **Unstructured Data**: May contain noise, missing values, or disorganized information.
* **Unlabeled Data**: Contains only input features (no output/target values), making it easier to gather but harder to analyze than labeled data.

**Unsupervised ML**

**Major Unsupervised Learning Techniques**

**1. Clustering**

Clustering involves grouping data points based on similarity. These groups, or **clusters**, are formed without prior labels. This technique is useful for identifying hidden structures in raw data.

**Popular Clustering Algorithms:**

* **K-Means**: Divides data into *K* distinct clusters by minimizing distance within each group.
* **Hierarchical Clustering**: Builds a tree of clusters by progressively merging or splitting groups.
* **DBSCAN (Density-Based Spatial Clustering)**: Identifies dense regions and treats scattered points as outliers.
* **Mean-Shift**: Moves data points toward the densest area to form clusters.
* **Spectral Clustering**: Uses graph theory to cluster based on relationships among points.

**2. Association Rule Learning**

Also known as **association rule mining**, this technique finds interesting relationships or correlations among variables in large datasets. It’s especially useful in **market basket analysis**—e.g., discovering that people who buy milk often buy bread too.

**Common Algorithms:**

* **Apriori**: Identifies frequent itemsets and derives rules step-by-step.
* **FP-Growth**: A faster alternative to Apriori, avoids generating unnecessary combinations.
* **Eclat**: Uses set intersections to find frequent patterns efficiently.
* **Tree-Based Methods**: Use tree structures to manage and mine very large datasets effectively.

**3. Dimensionality Reduction**

This technique simplifies data by reducing the number of input features while preserving essential information. It's useful for both improving model performance and visualizing complex data.

**Example:**

A student dataset with 100 features (height, age, grades, etc.) might be reduced to just a few key features like height and grades for easier analysis or visualization.

**Common Dimensionality Reduction Algorithms:**

* **Principal Component Analysis (PCA)**: Converts correlated features into a smaller set of uncorrelated variables.
* **Linear Discriminant Analysis (LDA)**: Reduces dimensions while enhancing class separation.
* **Non-negative Matrix Factorization (NMF)**: Breaks data into non-negative components for better interpretability.
* **Locally Linear Embedding (LLE)**: Preserves local relationships between data points during dimensionality reduction.
* **Isomap**: Maintains global structure by preserving distances along the data manifold.

**REINFORCEMENT LEARNING**

Reinforcement Learning is a type of machine learning where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of **rewards** or **penalties**. The goal is to **maximize cumulative rewards** over time.

**Core Concept**

An **agent** performs **actions** in an **environment**, transitioning through various **states** and receiving **rewards** or **penalties** as feedback. This helps the agent learn the best strategies to achieve a goal.

**Key Components**

* **Agent**: The decision-maker.
* **Environment**: The system where the agent operates.
* **State**: The current situation of the agent.
* **Action**: Choices the agent can make.
* **Reward**: Feedback from the environment based on the agent’s action.

**How RL Works**

1. The agent observes the current **state**.
2. It selects and performs an **action**.
3. The **environment** responds with a **reward** and a new **state**.
4. The agent updates its strategy to **maximize future rewards**.

**Important RL Elements**

* **Policy**: The agent's strategy to decide the next action.
* **Reward Function**: Provides immediate feedback (positive or negative).
* **Value Function**: Predicts long-term rewards from a state.
* **Model of the Environment**: (optional) Predicts next states and rewards, useful for planning.

**Example: Robot Navigating a Maze**

* A robot explores a maze to reach a **diamond** and avoid **fire hazards**.
* It receives **positive rewards** for correct moves and **penalties** for hazards.
* Through **trial and error**, the robot learns the **optimal path** by:
  + **Exploring** different paths,
  + **Receiving feedback** for each move,
  + **Adjusting its behavior** based on experience,
  + And **choosing the path** that maximizes cumulative reward.

**Types of Reinforcements**

1. **Positive Reinforcement**
   * Occurs when a behavior leads to a rewarding event.
   * **Pros**: Boosts performance, helps long-term learning.
   * **Cons**: Overuse may reduce effectiveness.
2. **Negative Reinforcement**
   * Happens when a behavior helps avoid a negative outcome.
   * **Pros**: Encourages necessary actions to avoid penalties.
   * **Cons**: Might lead to doing just enough to avoid punishment, not more.