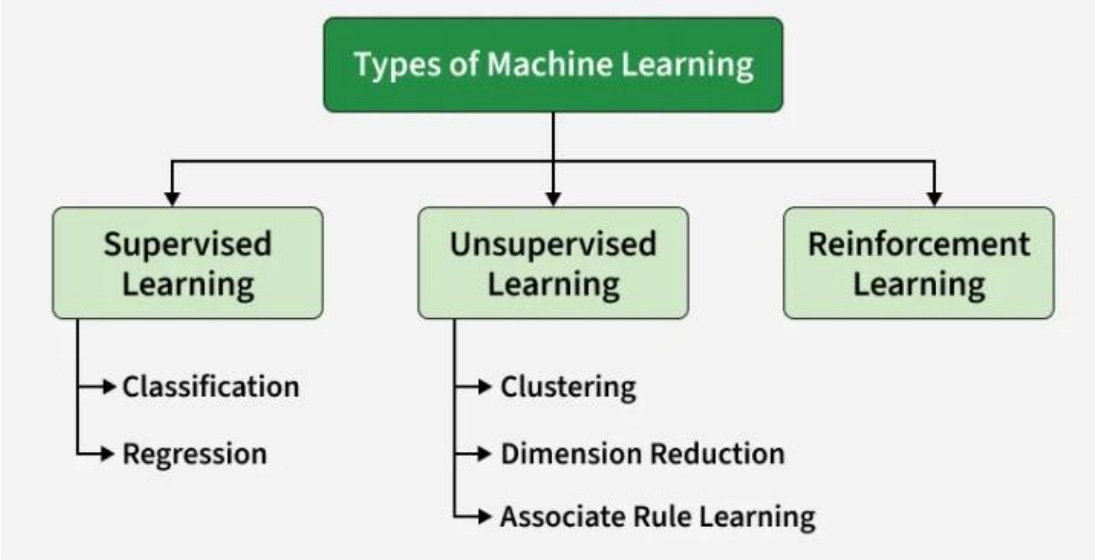
TYPES OF MACHINE LEARNING

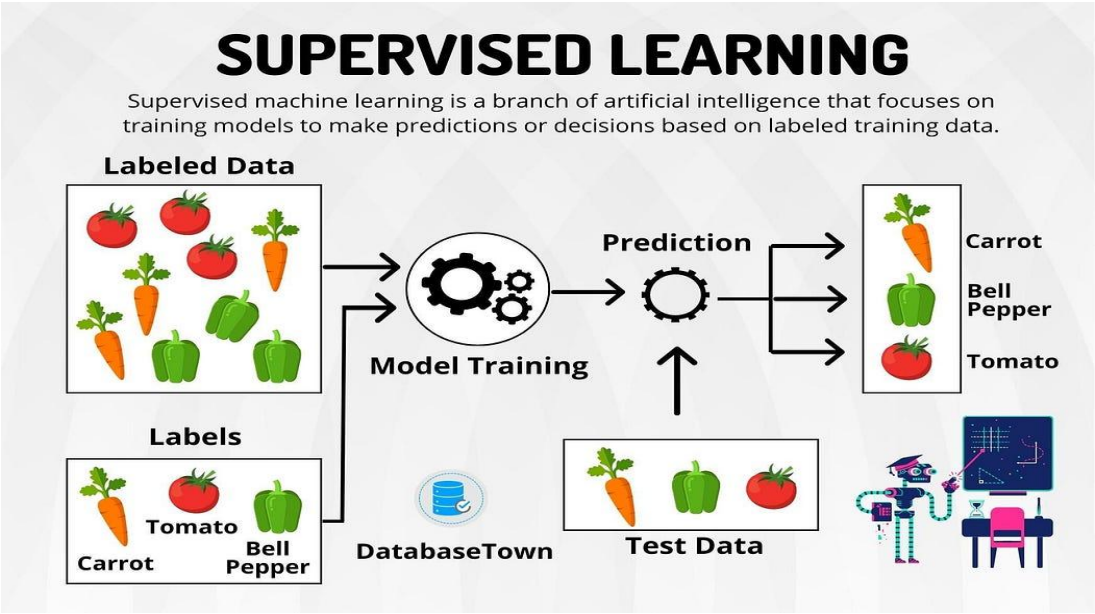
- Machine Learning teaches systems to learn patterns and make decisions like humans by analyzing and learning from data.
- There are several types of machine learning, each with special characteristics and applications.
- Some of the main types of machine learning algorithms are as follows:
  - Supervised Machine Learning
  - Unsupervised Machine Learning
  - Reinforcement Learning
  - Semi-Supervised Learning
  - Self-Supervised Learning



Types of Machine Learning

Supervised Machine Learning

- Supervised learning is a type of machine learning where a model is trained to make predictions based on labeled data.
- The model is given input data and the corresponding correct output, and it learns to predict the output for new data based on this training.
- This allows the model to perform tasks such as classification and regression.



- Consider a scenario where a computer system is provided with a dataset containing images of various fruits.
- The system, through supervised learning, aims to categorize the fruits based on their respective characteristics such as color, shape, and size.
- Once the system is trained, it can predict the category of new, unseen fruit images by comparing them with the training data.
- Supervised learning can be bifurcated into two subcategories: classification and regression

Two main categories of supervised learning

1. Classification

- During the training phase, a classification algorithm is provided with data points each associated with a specific category.
- The task of the classification algorithm is to assign an input value to a class or category based on the training data.
- For example, consider an email classification system. The system is trained on a dataset containing spam and non-spam emails. The model identifies the features within the data that correlate to either class and creates the mapping function. When a new email arrives, the model uses this function to determine if the email is spam or not.

2.Regression

- Regression models, on the other hand, are used when the output is a real or continuous value.
- For example, predicting the price of a house based on its size, location, and other factors employs a regression model.
- The labels in this case are numerical values, and the model predicts these values based on the input features.

How does Supervised Learning work?

- Gather a dataset of labeled examples: The first step in supervised learning is to gather a dataset of labeled examples that the model will be trained on. The examples should be representative of the problem at hand and should include both the input data and the correct output labels.
- Preprocess the data: Next, the data may need to be preprocessed in order to prepare it for use in a machine learning model. This may include tasks such as cleaning the data, filling in missing values, or normalizing the data.
- Train the model: Once the data is prepared, the model can be trained on it. This involves using an optimization algorithm to adjust the parameters of the model in order to minimize the error between the model's predictions and the correct output labels.
- Test the model: After the model is trained, it can be tested on a separate dataset to evaluate its performance. The model's accuracy can be measured using metrics such as precision, recall, and F1 score.
- Use the model to make predictions: Once the model has been trained and tested, it can be used to make predictions on new, unseen data. The model will use the patterns and relationships learned during training to make predictions based on the input data.

Where to Use Supervised Learning

- When you have labeled data and want to predict outcomes.
- Ideal for classification (like spam detection) or regression tasks (like price forecasting).
- Best used in domains where historical data with outcomes is already available.

Applications

- Supervised learning is used in a wide variety of applications, including:
- Image, speech and text processing: For tasks like image classification, speech recognition and sentiment analysis.
- Predictive analytics: To forecast sales, customer churn, stock prices and weather conditions.
- Recommendation and personalization: Powering systems that suggest products, movies or content.
- Healthcare and finance: Used for medical diagnosis, fraud detection and credit scoring.
- Automation and control: In autonomous vehicles, manufacturing quality checks and gaming AI.

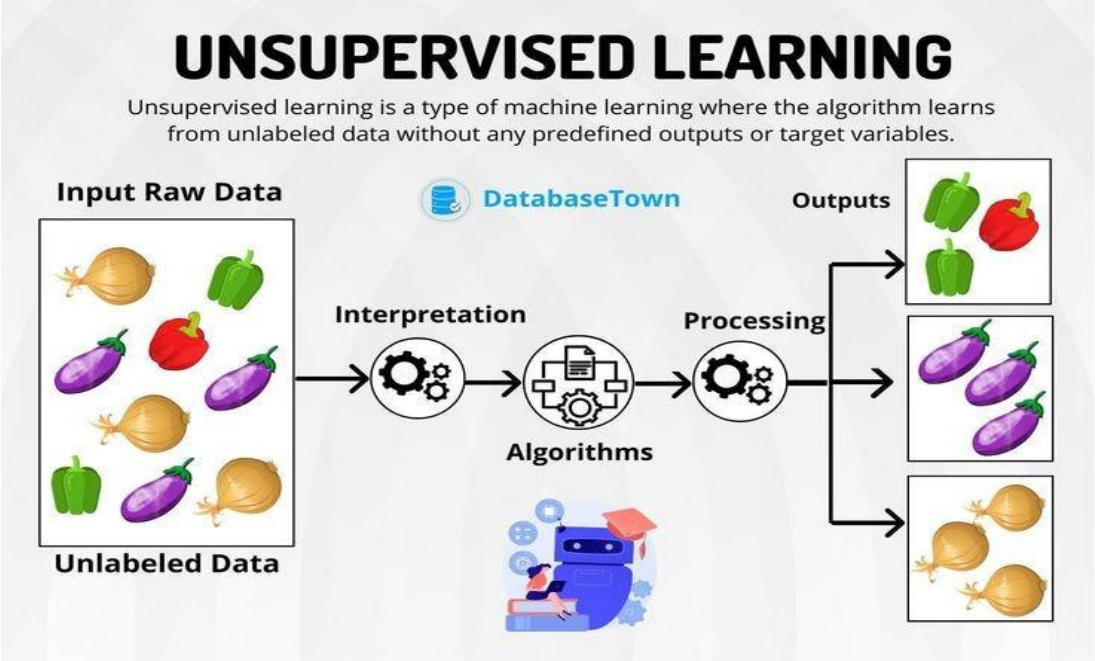
2. Unsupervised Machine Learning

- Imagine walking into a library where there are no labels on the books, and you have no idea where to find the book you need.

- Now, imagine if there was a system that could automatically group similar books together based on their content, without any prior knowledge about the genres.
- This is the magic of unsupervised learning in machine learning

Unsupervised Machine Learning

- Unsupervised learning is a type of machine learning where the algorithm learns patterns from unlabelled data.
- unsupervised learning works with data that has no labels
- The goal is to identify hidden structures and patterns within the data.
- In unsupervised learning, you give the model a bunch of images and it figures out on its own , without being explicitly told what they are.



Example: If you have customer data without labels, the algorithm can group similar customers based on purchase behavior useful for segmentation and marketing.

Main categories of unsupervised learning

1. Clustering
- Clustering is about grouping data points so that those in the same group (cluster) are more similar to each other than to those in other groups.
- Example: Imagine you have a basket of fruits and you want to group them based on their types without any prior knowledge. The algorithm might group apples, oranges, and bananas separately based on their features like color, size, and shape.

2. Dimensionality Reduction
- Dimensionality reduction reduces the number of random variables under consideration by obtaining a set of principal variables. This is especially useful for visualizing data.

- Example: Think of a dataset with 100 features. Dimensionality reduction techniques like Principal Component Analysis (PCA) can reduce this to 2 or 3 features, making it easier to visualize and analyze.

Common Algorithms in Unsupervised Learning

1. K-Means Clustering
- How it Works: K-Means clustering aims to partition n observations into k clusters where each observation belongs to the cluster with the nearest mean. The algorithm iteratively assigns each data point to the nearest cluster center and then updates the cluster centers based on the assigned points.
- Example: If you have data on customer purchases, K-Means can group customers into clusters with similar purchasing behaviours, helping businesses tailor their marketing strategies.

2. Hierarchical Clustering
- How it Works: This algorithm builds a hierarchy of clusters either from top to bottom (divisive) or bottom to top (agglomerative). Each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
- Example: Creating a family tree is similar to hierarchical clustering, where individuals are grouped based on their relationships.

3. Principal Component Analysis (PCA)
- How it Works: PCA reduces the dimensionality of the data by transforming it into a new set of variables, the principal components, which are orthogonal and capture the maximum variance in the data.
- Example: PCA can be used in image compression, where a large image dataset is reduced to a smaller set of variables while preserving most of the important information.

Where to Use Unsupervised Learning

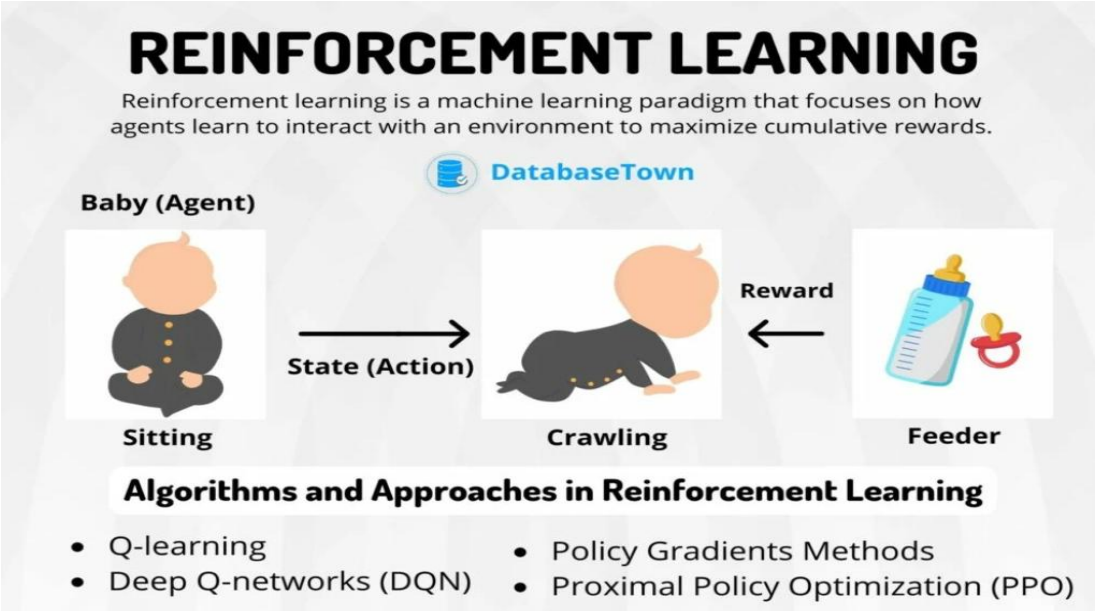
- When data is unlabeled or unstructured.
- Useful for exploratory analysis, clustering or feature extraction.
- Common in marketing, recommendation systems and fraud detection where patterns matter more than labels.

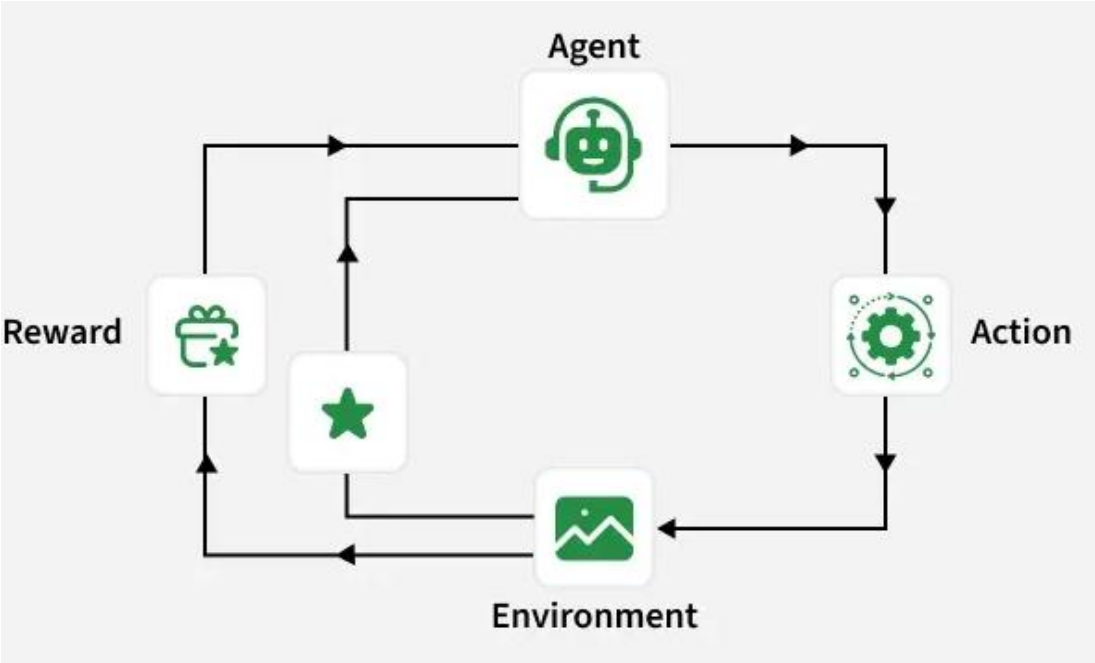
Applications of Unsupervised Learning

- Here are some common applications of unsupervised learning:
- Clustering and segmentation: Group similar data points, customers or images.
- Anomaly detection: Spot unusual patterns or outliers in data.
- Dimensionality reduction: Simplify large datasets while retaining key information.
- Recommendation and marketing: Identify user preferences and improve product suggestions.
- Data preprocessing and analysis: Clean data, detect patterns and support exploratory data analysis (EDA).

3. Reinforcement Learning

- Reinforcement learning trains an agent to make a sequence of decisions through trial and error.
- The agent interacts with the environment, receives feedback in the form of rewards or penalties and learns optimal actions over time.





- Example: An AI agent learning to play chess gets positive feedback for good moves and negative for poor ones. Over time, it learns strategies to win more often.
- Here are some of most common reinforcement learning algorithms:
- Q-learning: Learns the best action for each state based on expected rewards.
- SARSA (State-Action-Reward-State-Action): Similar to Q-learning but updates values for the action actually taken.
- Deep Q-learning: Uses neural networks to handle complex state-action relationships

Types of Reinforcement Learning

- Positive Reinforcement: Rewards desired behavior (e.g., giving points for correct answers).
- Negative Reinforcement: Removes negative outcomes to encourage good actions (e.g., turning off a buzzer after the right move).

Where to Use Reinforcement Learning

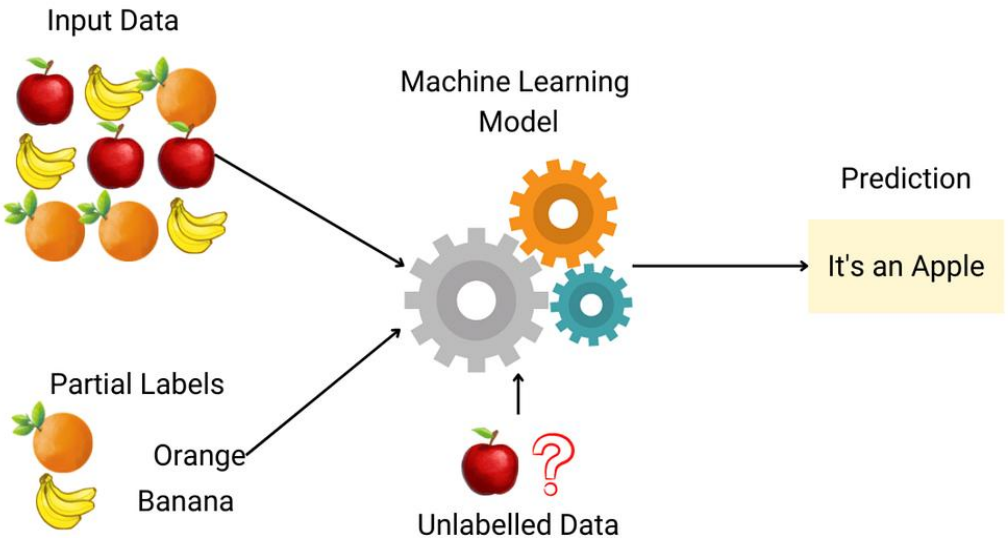
- When you need an agent to learn by interacting with an environment.
- Best for decision-making or optimization tasks involving trial and feedback loops.
- Used when long-term performance or adaptive behavior is more important than immediate accuracy.

Applications of Reinforcement Learning

- Gaming and simulation: Teaching agents or NPCs to play and adapt intelligently.
- Robotics and automation: Enabling robots to perform tasks autonomously.
- Autonomous vehicles: Helping self-driving cars make real-time decisions.
- Healthcare and finance: Optimizing treatment plans, trading and resource allocation.
- Recommendation and personalization: Improving user experience through adaptive suggestions.
- Industrial and energy management: Optimizing control systems and energy use.

4.Semi-Supervised Learning: Supervised + Unsupervised Learning

- Semi-supervised learning is the type of machine learning that uses a combination of a small amount of labeled data and a large amount of unlabeled data to train models.
- This approach to machine learning is a combination of supervised machine learning, which uses labeled training data, and unsupervised learning, which uses unlabeled training data.
- In other words, it is partially supervised and partially unsupervised learning.



- Let's take one example from the below image to make it clear. Suppose a bucket consists of three fruits , apple, banana and orange. Someone captured the image of all the three but labeled only the orange and banana images. Here, the model first will classify the new apple image as not a banana and not orange. Then someone will observe these predictions and label them as apples. Then retraining the model with that label will give it the ability to classify apple images as an apple

Popular Techniques

- Graph-based Learning: Spreads label information through data relationships.
- Label Propagation: Iteratively assigns labels to unlabeled data.
- Co-training: Uses two models to train and label each other's data.
- Self-training: Uses model predictions as pseudo-labels.
- Generative Adversarial Networks (GANs): Generates synthetic data to improve learning.

Where to Use Semi-Supervised Learning

- When you have limited labeled data but plenty of unlabeled data.
- Useful for domains with high labeling costs, such as medical, NLP or image datasets.
- Ideal when unlabeled data still holds valuable information that can improve learning performance.

Applications

- Image Classification: Combine small labeled and large unlabeled image datasets to improve accuracy.
- Natural Language Processing (NLP): Enhance language models by using a mix of labeled and vast unlabeled text data.
- Speech Recognition: Boost accuracy by leveraging limited transcribed audio and more unlabeled speech data.

- Recommendation Systems: Improve recommendations using sparse labeled data and abundant unlabeled user behavior.
- Healthcare & Medical Imaging: Improve medical image analysis with a mix of labeled and unlabeled images.

Conceptual Working of Semi-Supervised Learning

- SSL uses pseudo labeling to train the model with less labeled training data than supervised learning. The whole process of SSL can be explained as below:
- — Primarily, it trains the model with less amount of training data similar to the supervised learning. The training continues until the model gives accurate results.
- — The algorithm use the unlabeled dataset with pseudo labels in the next step, and now the result may not be accurate.
- — Now, the labels from labeled data and pseudo labels data are linked together.
- — The input data in labeled training data and pseudo labels data are linked together.
- — In the end, again train the model with the new combined input as did in the first step. It will reduce error rate and improve accuracy of the model.

5.Self-Supervised Learning

- Self-Supervised Learning (SSL) is a modern approach where models generate their own labels from raw data. It doesn't rely on manual annotation instead, the model learns by predicting parts of data from other parts.
- Example: In NLP, models like BERT or GPT learn by predicting masked words in sentences, using surrounding context as supervision. This helps them learn language understanding without human labeling.

Popular Techniques

- Masked Modeling (BERT)
- Contrastive Learning (SimCLR, MoCo)
- Autoencoders
- Predictive Coding

Applications

- Natural Language Processing
- Computer Vision and Speech Recognition
- Video understanding
- Pre-training for large AI models

Where to Use Self-Supervised Learning

- When manual labeling is impossible or expensive.
- Suitable for large-scale datasets like text, audio and images.
- Best for pre-training models that can later be fine-tuned for specific supervised tasks