

# Unsupervised Learning

## Introduction

Unsupervised Learning is a type of machine learning where the model is trained using unlabeled data. The system tries to discover hidden patterns, structures, or relationships within the data without any predefined output labels.

## What is Unsupervised Learning?

In unsupervised learning, the algorithm works on input data without knowing the correct output. It learns by identifying similarities, differences, and patterns directly from the data.

Unlabeled Data → Algorithm → Patterns / Groups

## When to Use Unsupervised Learning

- 1 When labeled data is not available
- 2 To explore hidden patterns in data
- 3 For data exploration and grouping

## Main Tasks in Unsupervised Learning

- 1 Clustering
- 2 Association Rule Mining
- 3 Dimensionality Reduction

### ***Clustering***

Clustering groups similar data points together based on their characteristics. Each group is called a cluster.

Example: Customer segmentation in marketing.

Data Points → Similarity Measure → Clusters

### ***Association Rule Mining***

Association rule mining identifies relationships between variables in large datasets. It is commonly used in market basket analysis.

Example: Customers who buy bread also buy butter.

### ***Dimensionality Reduction***

Dimensionality reduction reduces the number of input features while preserving important information.

Example: Principal Component Analysis (PCA).

## Common Unsupervised Learning Algorithms

- 1 K-Means Clustering
- 2 Hierarchical Clustering
- 3 DBSCAN
- 4 Apriori Algorithm

## 5 Principal Component Analysis (PCA)

### Advantages

- 1 Does not require labeled data
- 2 Discovers hidden patterns
- 3 Useful for exploratory analysis

### Disadvantages

- 1 Results can be difficult to interpret
- 2 No direct accuracy measure
- 3 May produce meaningless clusters

### Real-Life Example

In e-commerce platforms, unsupervised learning is used to group customers based on browsing behavior and purchase history for targeted marketing.

### Summary

Unsupervised learning is useful when labeled data is unavailable. It helps uncover patterns and structures that are not immediately visible.