

Report Title

REPORT SUBTITLE

# Name | Course Title | Date**Machine Learning: Key Differences**

**Clustering vs Classification**

**Clustering**

**Definition**: Clustering is an unsupervised learning technique used to group similar data points into clusters without pre-defined labels.

**Key Points**:

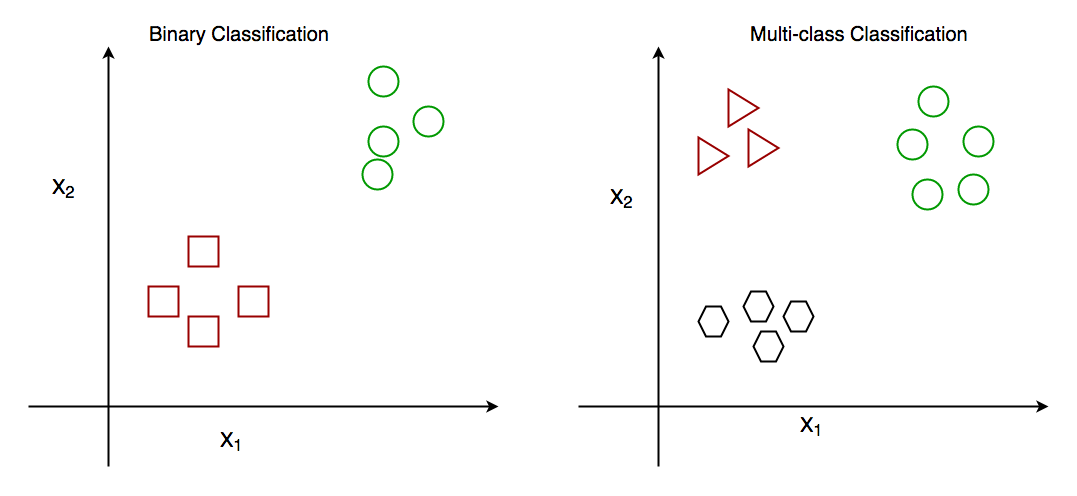
* **Unsupervised Learning**: No labeled data is required.
* **Objective**: To find natural groupings in data.
* **Examples**: Customer segmentation, image compression.
* **Algorithms**: K-means, Hierarchical Clustering, DBSCAN.
* **Output**: A set of clusters, each containing data points that are similar to each other.

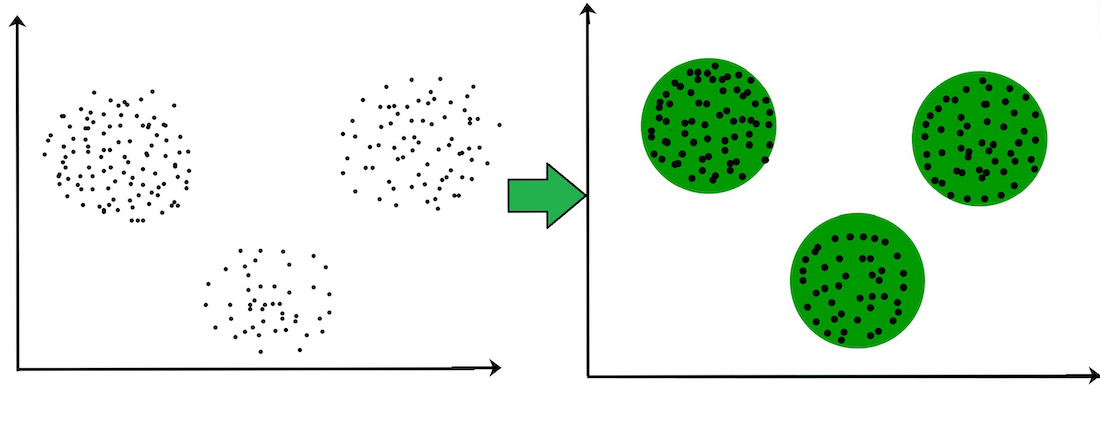
**Classification**

**Definition**: Classification is a supervised learning technique used to assign predefined labels to new data points based on training data.

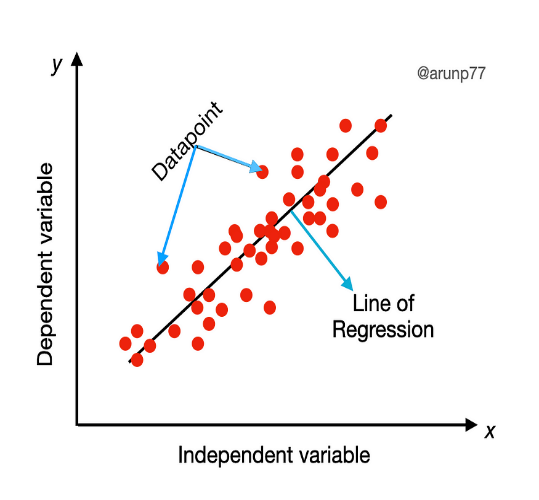
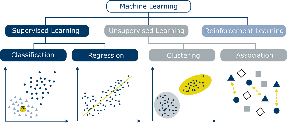
**Key Points**:

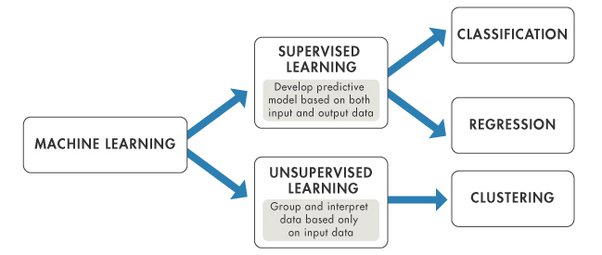
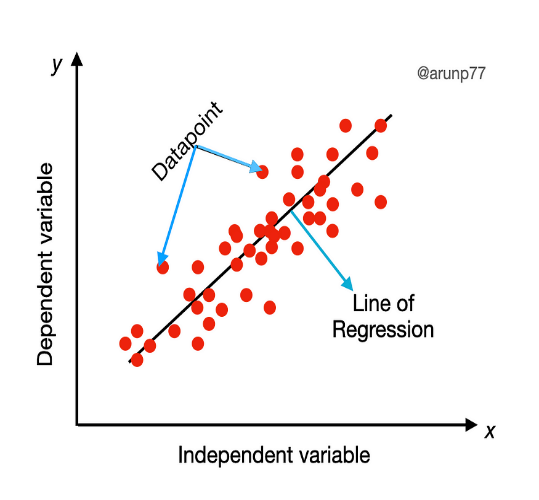
* **Supervised Learning**: Requires labeled training data.
* **Objective**: To predict the category or class of a given data point.
* **Examples**: Email spam detection, image recognition.
* **Algorithms**: Decision Trees, Random Forest, Support Vector Machines (SVM), Neural Networks.
* **Output**: A label or class assigned to each data point.

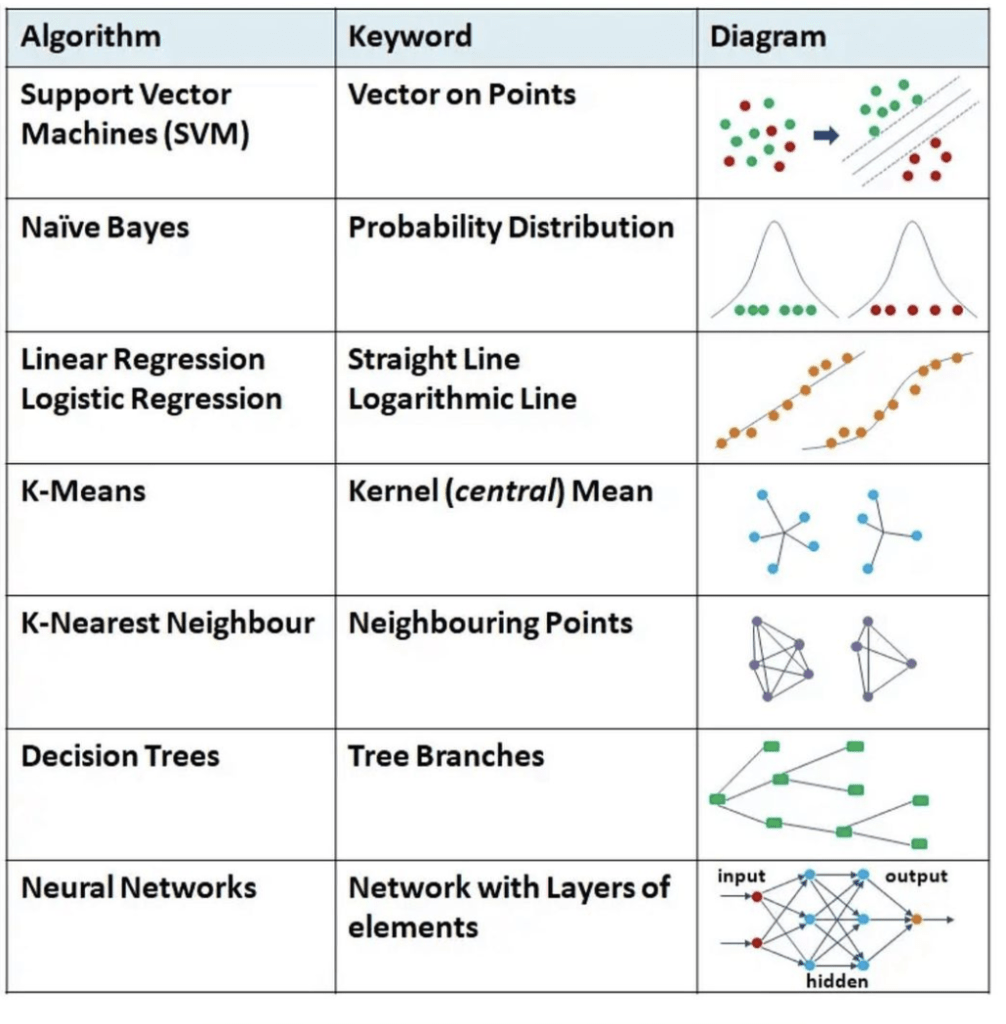




**Diagram: Clustering vs Classification**







Clustering is an unsupervised learning method that groups data based on similarities without predefined labels, while classification is a supervised learning method that assigns predefined labels to data points.

Regression predicts continuous values and focuses on modeling the relationship between variables, whereas classification predicts discrete labels and categorizes data points into predefined classes.

Data Mining: Concepts and Techniques" by Jiawei Han, Micheline Kamber, and Jian Pei, clustering techniques like K-means and hierarchical clustering are used to discover natural groupings in data, whereas "Pattern Recognition and Machine Learning" by Christopher M. Bishop explains how classification techniques such as Decision Trees and SVM require labeled training data to predict categories.

"The Elements of Statistical Learning" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman details regression methods like linear regression for predicting continuous outcomes, while "An Introduction to Statistical Learning" by Gareth James et al. explains classification techniques for assigning data points to categories, such as in email spam detection and handwritten digit recognition.