# ***Clustering vs Classification***

# **Definition:**

* **Clustering**: An unsupervised learning technique that groups a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups.
* **Classification**: A supervised learning technique that assigns labels to a set of observations based on training data.

### Examples:

* **Clustering**: Grouping customers based on purchasing behavior. For example, a retailer may want to group customers into different segments (clusters) such as 'high spenders', 'frequent buyers', and 'seasonal shoppers'.
* **Classification**: Email filtering, where emails are classified into categories such as 'spam' or 'not spam' based on labeled data.

### Key Differences:

1. **Nature of Learning**:
   * **Clustering**: Unsupervised learning.
   * **Classification**: Supervised learning.
2. **Data Labeling**:
   * **Clustering**: Does not require labeled data.
   * **Classification**: Requires labeled data.
3. **Output**:
   * **Clustering**: Generates a set of clusters.
   * **Classification**: Assigns predefined labels to new data points.
4. **Goal**:
   * **Clustering**: To find inherent structures and patterns within the data.
   * **Classification**: To predict the category of a given data point.
5. **Examples of Algorithms**:
   * **Clustering**: K-means, DBSCAN, Hierarchical Clustering.
   * **Classification**: Decision Trees, Support Vector Machines, Neural Networks

## ***2. Regression vs Classification***

### Definition:

* **Regression**: A supervised learning technique used to predict a continuous output variable based on one or more input features.
* **Classification**: A supervised learning technique used to predict a categorical output variable.

### Examples:

* **Regression**: Predicting the price of a house based on features such as size, location, and number of rooms.
* **Classification**: Predicting whether a patient has a certain disease based on medical test results.

### Key Differences:

1. **Nature of Output**:
   * **Regression**: Predicts continuous values.
   * **Classification**: Predicts discrete class labels.
2. **Goal**:
   * **Regression**: To estimate the relationship between variables and predict outcomes.
   * **Classification**: To assign input data into predefined categories.
3. **Evaluation Metrics**:
   * **Regression**: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared.
   * **Classification**: Accuracy, Precision, Recall, F1 Score.
4. **Use Cases**:
   * **Regression**: Stock price prediction, weather forecasting.
   * **Classification**: Image recognition, spam detection.
5. **Examples of Algorithms**:
   * **Regression**: Linear Regression, Polynomial Regression, Ridge Regression.
   * **Classification**: Logistic Regression, Naive Bayes, k-Nearest Neighbors (k-NN).