



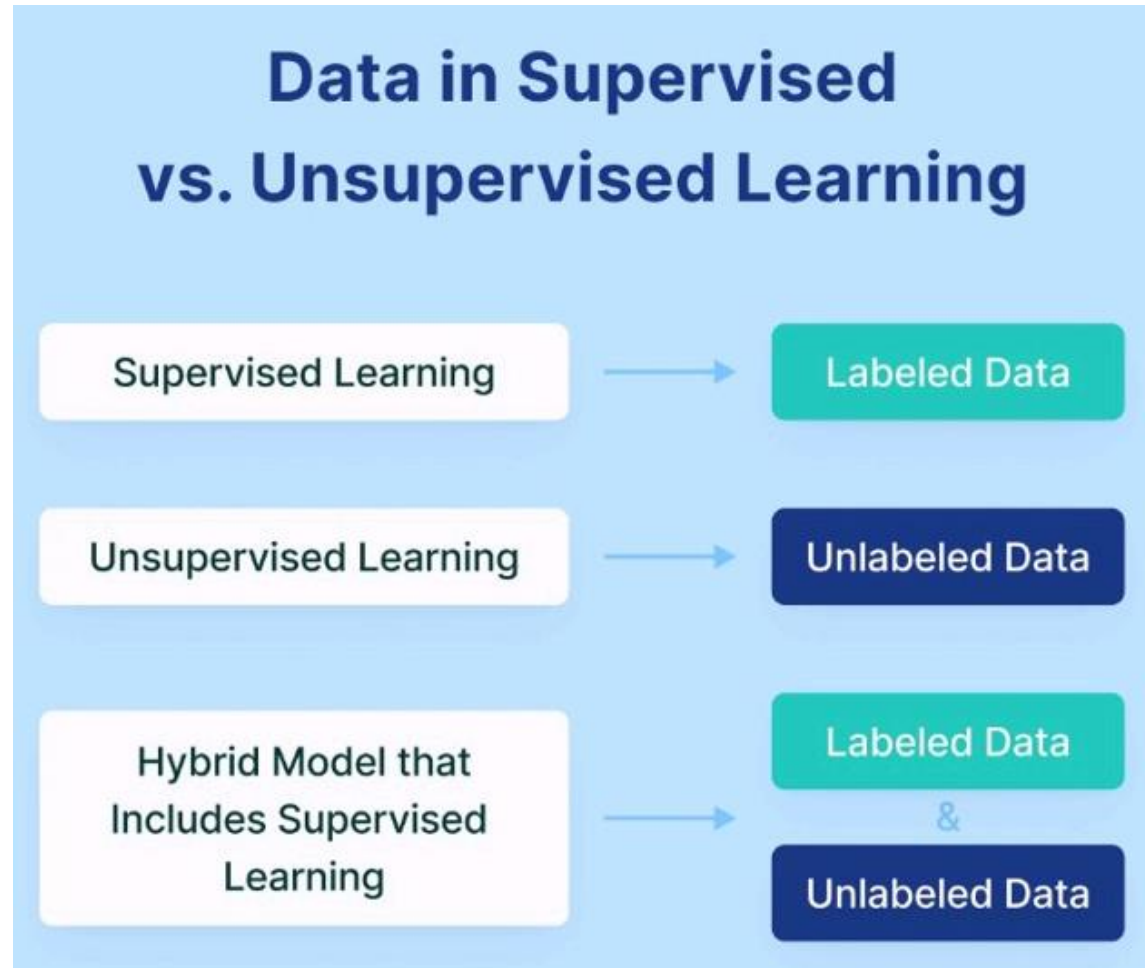
# DATA SCIENCE

By: Michel Samir Zaki



# Supervised vs Unsupervised

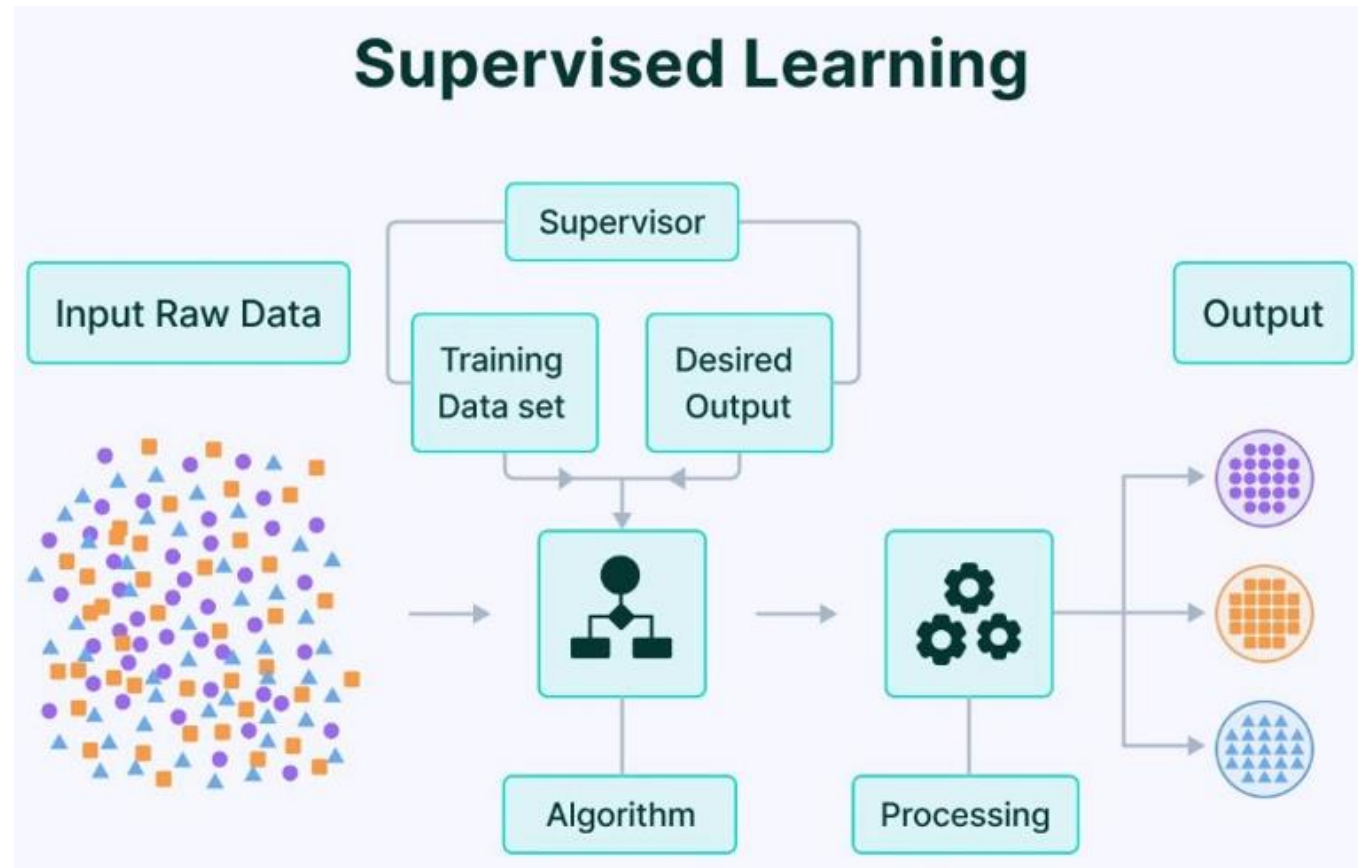
- one of them uses labeled data to predict outcomes, while the other does not.





# Supervised Learning

- **Supervised Learning** is the machine learning approach defined by its use of labeled datasets to train algorithms to classify data and predict outcomes.





# Supervised Learning

- **Labelled Dataset: Feature columns + label (class) column**

Lead Location	Gender	Lead Company size	Lead Revenues	Number od emails		Client
UK	F	5	500000	3		NO
Ireland	F	6	600000	2		NO
UK	M	6	600000	5		YES
USA	F	10	1000000	12		YES
USA	F	3	300000	4		YES
USA	M	5	500000	7		YES
Germany	M	6	600000	1		NO



# Supervised Learning

## Supervised Machine Learning Methods

### 1. **Classification**

In classification problems, our output typically consists of classes or categories

### 2. **Regression**

In regression problems, the predicted output values are real numbers



# Supervised Learning

- **Classification**

## Binary Classification



- Spam
- Not spam

## Multiclass Classification



- Dog
- Cat
- Horse
- Fish
- Bird
- ...

## Multi-label Classification



- Dog
- Cat
- Horse
- Fish
- Bird
- ...





# Supervised Learning

- **Regression**

- **Simple Regression**

- Used to predict a continuous dependent variable based on a single independent variable.
    - Simple linear regression should be used when there is only a single independent variable.

- **Multiple Regression**

- Used to predict a continuous dependent variable based on multiple independent variables.
    - Multiple linear regression should be used when there are multiple independent variables.

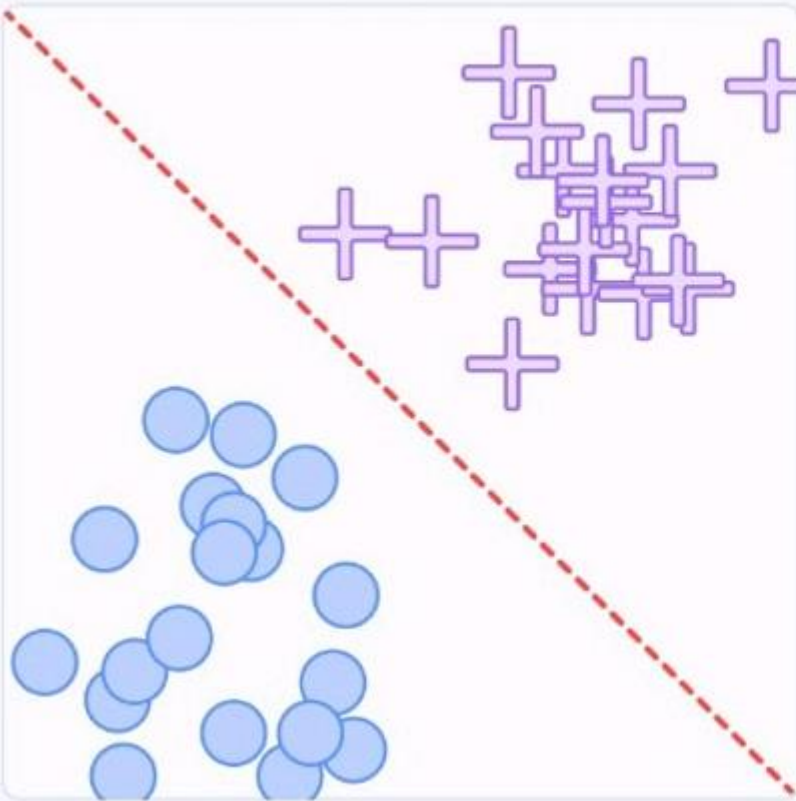
- **NonLinear Regression**

- Relationship between the dependent variable and independent variable(s) follows a nonlinear pattern.
    - Provides flexibility in modeling a wide range of functional forms.

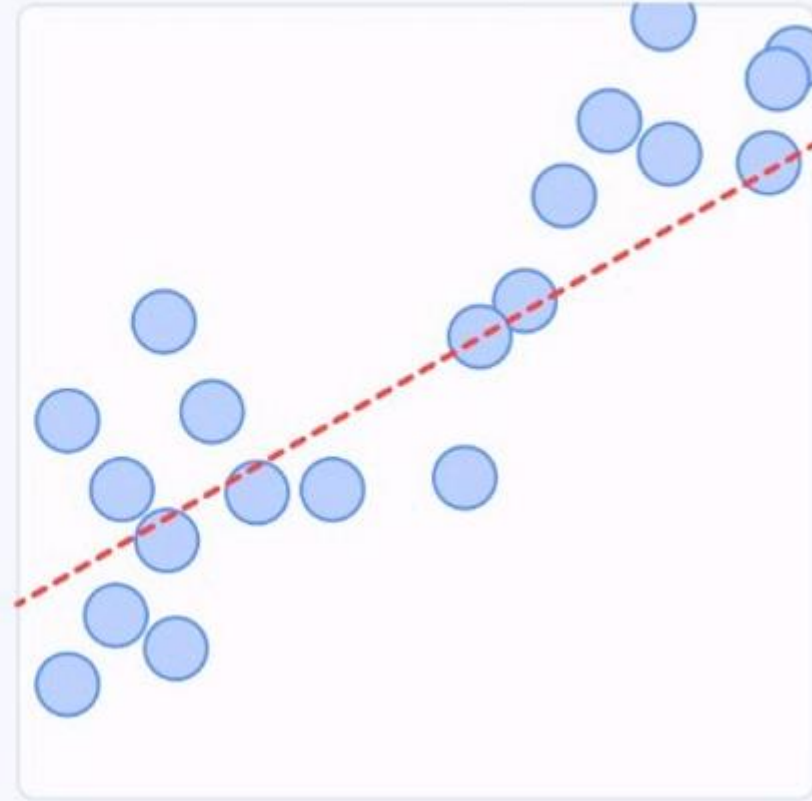


# Supervised Learning

Classification



Regression







# Supervised Learning

## Supervised Machine Learning Applications

Support Vector Machines (SVM), Logistic Regression, Naive Bayes, Neural Networks, K-nearest neighbor (KNN), and Random Forest.



# Unsupervised Learning

- **Unsupervised Learning** is a type of machine learning in which the algorithms are provided with data that does not contain any labels.
- It is a kind of self-learning where the algorithm can **find previously hidden patterns** in the unlabeled datasets.



# Unsupervised Learning

## Unsupervised Machine Learning Methods

### 1. **Clustering**

We find hidden patterns in the data based on their similarities or differences and are used to group data items or create clusters.

### 2. **Association**

We can find the relationship of one data item to another data item.

### 3. **Dimensionality reduction**

The algorithm works to reduce the dimensions of the data. It is used for feature extraction



# Unsupervised Learning

## Uses of unsupervised learning algorithms



Clustering



Association



Dimensionality  
reduction


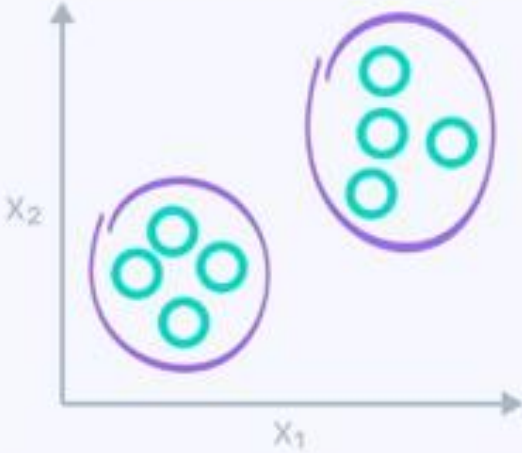


# Supervised vs Unsupervised

Supervised learning	Unsupervised learning
Input data is labeled	Input data is unlabeled
Has a feedback mechanism	Has no feedback mechanism
Data is classified based on the training dataset	Assigns properties of given data to classify it
Divided into Regression & Classification	Divided into Clustering & Association



# Supervised vs Unsupervised

Used for prediction	Used for analysis
Algorithms include: decision trees, logistic regressions, support vector machine	Algorithms include: k-means clustering, hierarchical clustering, apriori algorithm
A known number of classes	A unknown number of classes
	



The image features a blue-tinted background showing silhouettes of several groups of business professionals in a modern office environment. They are standing on a reflective floor, and a city skyline is visible through the windows in the background. The text "Thank You" is centered in the middle of the image.

Thank You