Nishant Raj

University Roll No. - 13030823129

Registration No. – 231300121019

Subject- Data Mining

Subject Code - PCCAIML603

Dept.- CSE AIML (B)

YEAR : 3rd

Semester : 6th

Introduction to Data Mining

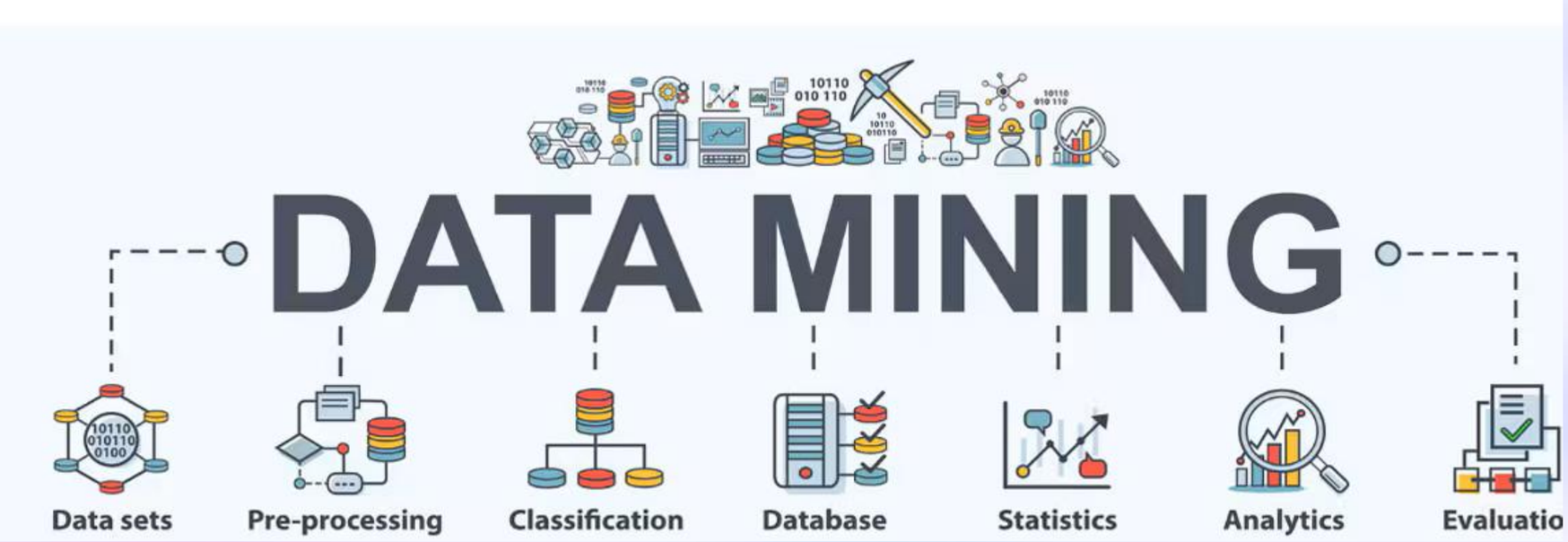
- Data mining is the process of discovering patterns and useful insights from large datasets.
- It uses techniques from statistics, machine learning, and database systems.

•**Importance of Data Mining:**

- Helps in decision-making by identifying trends and patterns.
- Used in various domains such as healthcare, finance, marketing, and security.

•**Two Main Types of Learning in Data Mining:**

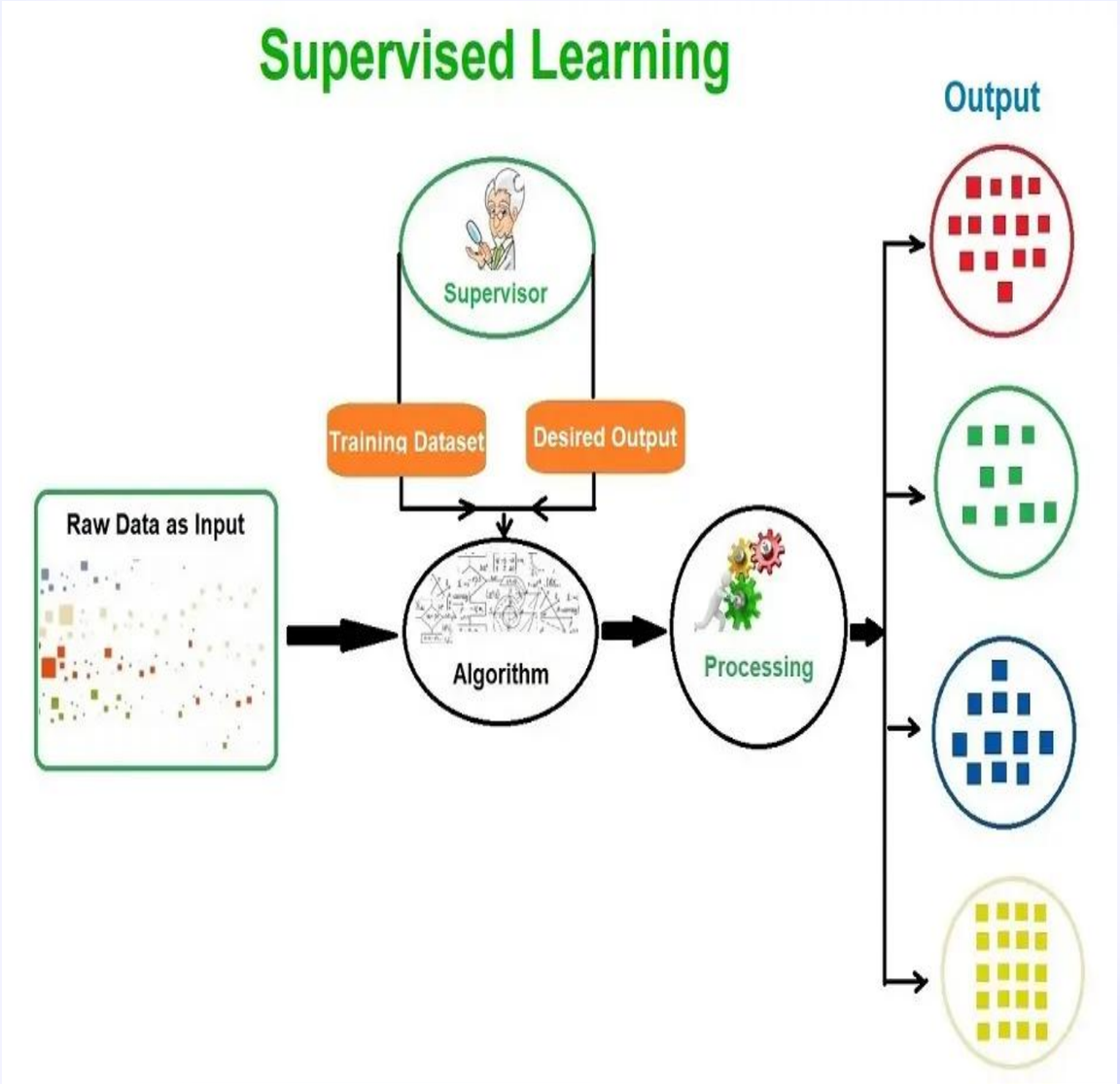
- Supervised Learning
- Unsupervised Learning



Supervised Learning

- **Definition:**
  - Supervised learning is a type of machine learning where the model is trained on labeled data.
  - The system learns from input-output pairs and makes predictions based on past observations.

- **Example:**
  - Predicting house prices based on square footage and location.
  - Email spam classification (spam vs. non-spam emails).



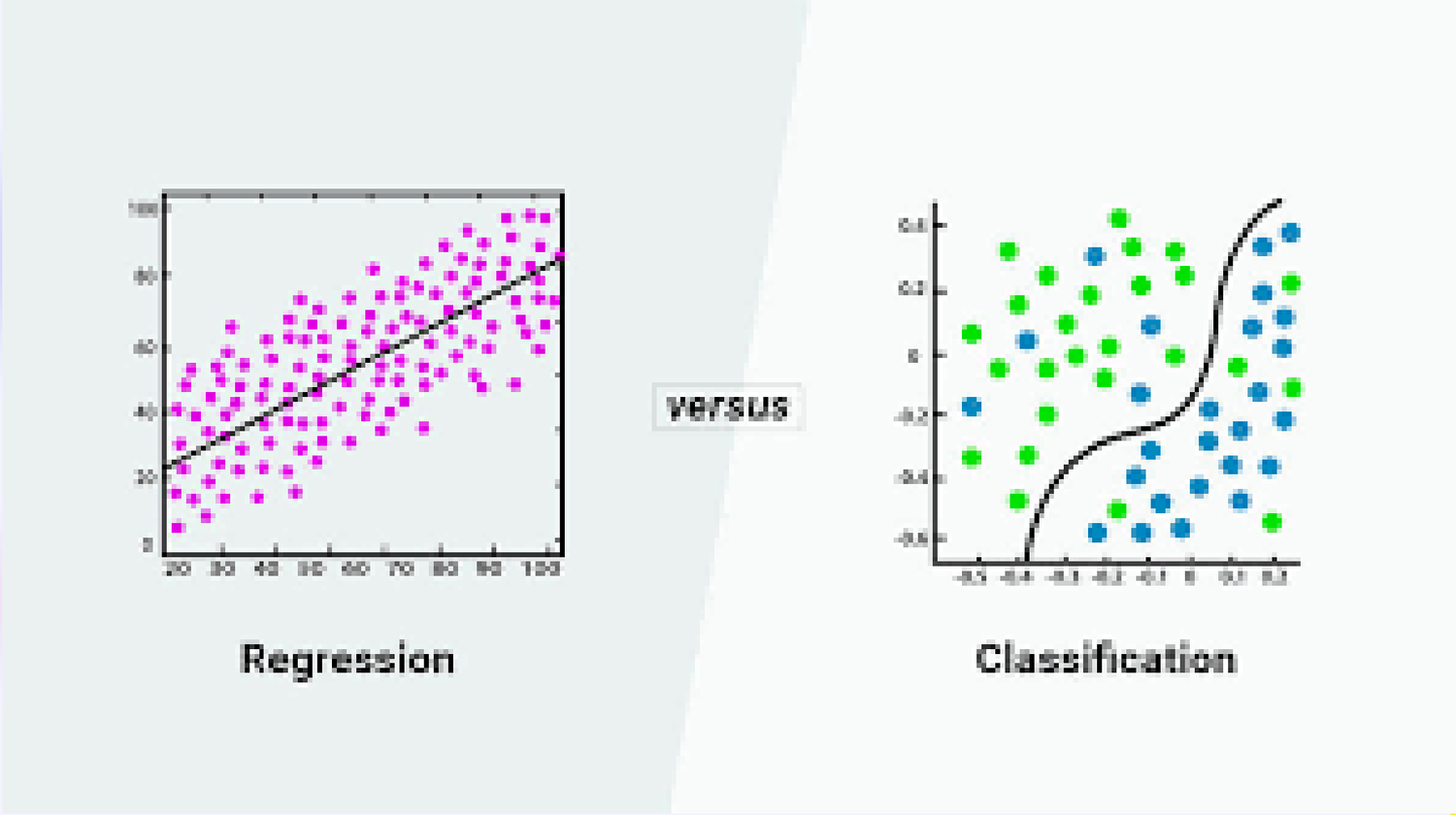
Types of Supervised Learning

1.Classification:

1. The output is a category or label.
2. Example: Identifying whether an email is spam or not.

2.Regression:

1. The output is a continuous value.
2. Example: Predicting stock prices based on historical data.



Unsupervised Learning

- **Definition:**
  - Unsupervised learning is a type of learning where the model finds patterns in data without labeled outputs.
  - It groups similar data points without explicit supervision.
- **Example:**
  - Customer segmentation in marketing.
  - Detecting fraudulent transactions based on unusual spending behavior.

Types of Unsupervised Learning

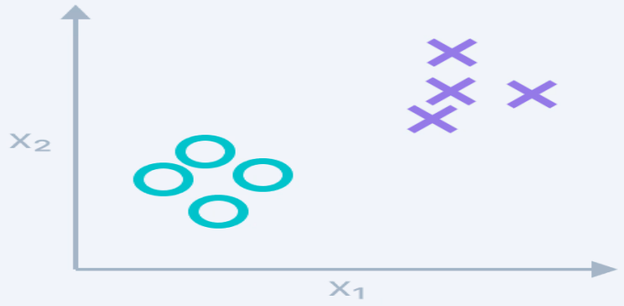
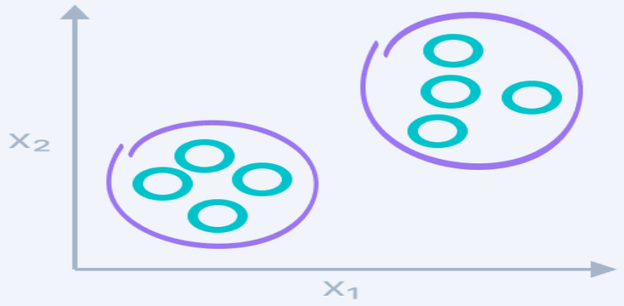
1.Clustering:

1. Groups data into clusters based on similarity.
2. Example: Grouping customers with similar purchasing behavior.

2.Association Rule Learning:

1. Finds relationships between variables in a dataset.
2. Example: Market basket analysis (customers who buy bread often buy butter).

Comparison of Supervised and Unsupervised Learning:

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
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
	

Applications of Supervised and Unsupervised Learning

- **Supervised Learning Applications:**
  - Fraud detection in banking.
  - Speech recognition systems.
  - Medical diagnosis (predicting diseases based on symptoms).

- **Unsupervised Learning Applications:**
  - Customer segmentation for targeted marketing.
  - Anomaly detection in cybersecurity.
  - Recommender systems (Netflix movie recommendations).

Conclusion

- Supervised learning is useful when labeled data is available and is effective for prediction tasks.
- Unsupervised learning helps in discovering hidden patterns in data where labels are not present.
- Both learning approaches have their own significance and are widely used in data mining applications.
- Supervised learning is commonly used in fraud detection, medical diagnosis, and speech recognition.
- Unsupervised learning is crucial for clustering, anomaly detection, and recommendation systems.