MACHINE LEARNING
Introduction

Trainer : Sujata Mohite
Email: sujata.mohite@sunbeaminfo.com

Trainer Introduction

- **Name**: Sujata Mohite

- **Designation** : Senior Technical Manager

- **Education** :

  - Masters of Engineering (ME) in Computers

  - Bachelors of Engineering in E&TC, From Pune University

- **Training Experience**

  - PreCAT Batches at Sunbeam

  - PG Courses: Advanced Analytics with Statistics, Networking, ML ,Aptitude

  - Modular Batches : Aptitude

  - Internship : Machine Learning

- **Professional Experience**

  - **10+ years**

- **Email** : sujata.mohite@sunbeaminfo.com

Course Contents

- Introduction to Machine Learning
- Preparing the Data
- Dimensionality Reduction
- Training Models
- Classification
- Support Vector Machine
- Decision Tree
- Ensemble Learning
- Case Study

80%
training

20%
test
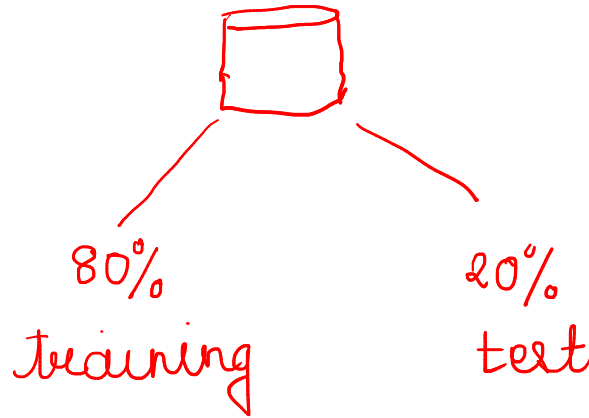
What we are going to Cover?

- **Fundamentals**
  - Data Science introduction ✓
  - Types ✓
  - Challenges ✓
- **Environmental Setup**
  - Python
  - Jupyter notebook ←
  - Pycharm installation
  - Numpy ✓
  - Pandas ✓
  - SciKit learn
- **Data Visualization**
  - Matplotlib ✓
  - Seaborn ✓

- **Supervised Learning**
  - Classification ←
  - Regression ←
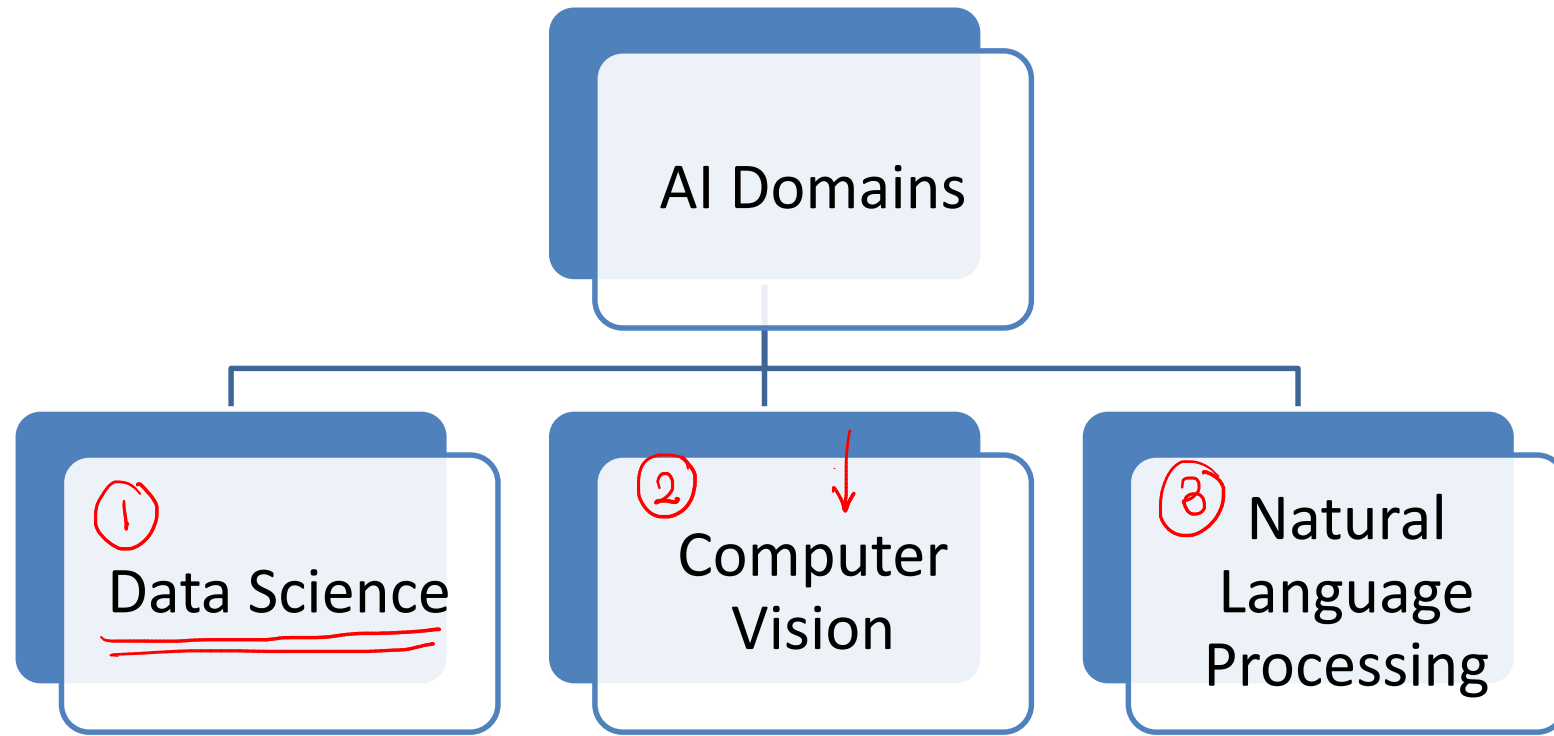
- **Deployment**
  - Case Study

Day1 Agenda

- Introduction to data science
- AI, ML, DL, ML Applications
- ML Program
- Types of ML
    - Supervised
    - Unsupervised

Data Science

Main Domains of AI Technology

AI Domains

① Data Science

② Computer Vision

③ Natural Language Processing

1. What is Data Science ?

- Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data.

- Data science is related to data mining, machine learning and big data   *images/videos, --*

- It processes the numeric /alphanumeric data with data systems to extract information to make decision or predict the outcome for the problem statement.

- Example:
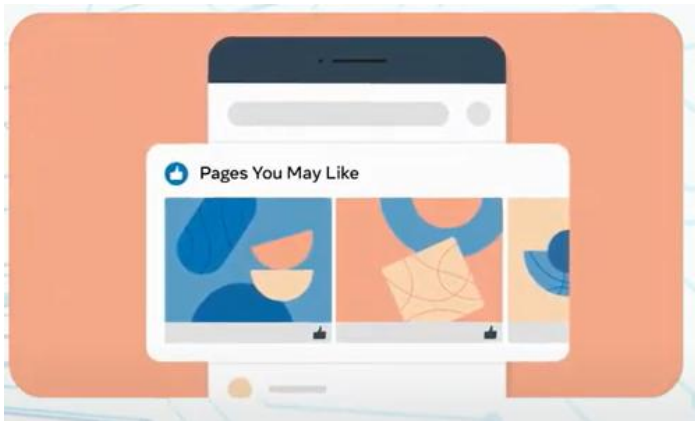
1. Recommendations   2. Price Predicting   3. Weather Forecasting
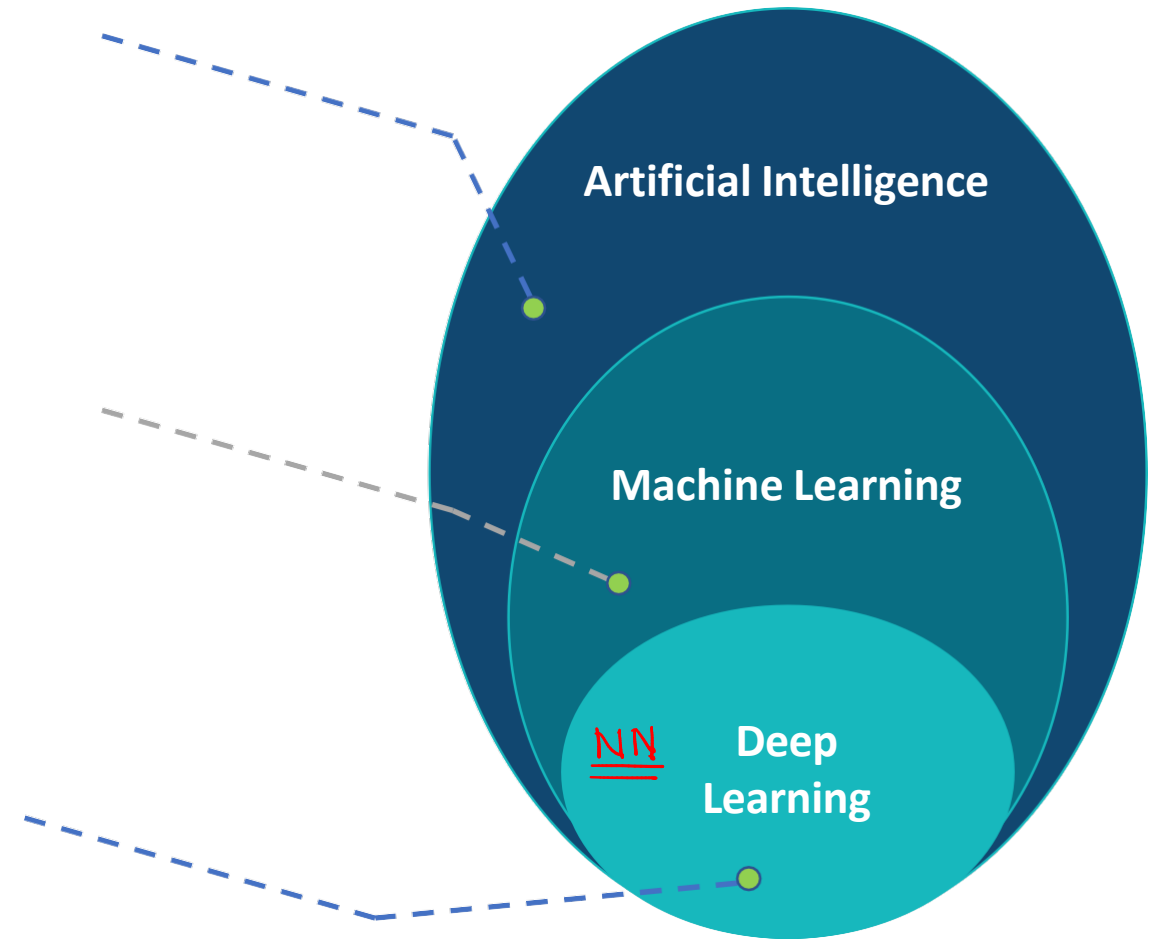
AI, ML, DL

Artificial Intelligence:
- A technique which enables machine to mimic human behavior

Machine Learning:
- Subset of AI which uses statistical methods to enable machines to improve the experience

Deep Learning:
- Subset of ML which makes the computation of multi-layer neural network feasible

Artificial Intelligence

Machine Learning

NN   Deep Learning

Artificial Intelligence

What is AI?

- Artificial Intelligence is an attempt to make a computer, a robot, or other piece of technology 'think' and process data in the same way as we humans do.

- AI is a branch of science which deals with helping machines finds solutions to complex problems in a more human-like fashion.

- AI therefore has to study how the human brain 'thinks', learns, and makes decisions when it tries to solve problems or execute a task.

- The aim of AI is to improve technology by adding functionality related to the human acts of reasoning, learning, and problem-solving.

alexa, siri

- Example : Home Automation Systems, Cortana is example of a voice controlled intelligent system

AI applications

- Google's search engine

- JPMorgan Chase's Contract Intelligence (COiN) platform uses AI, machine learning and image
recognition software to analyse legal documents

- IBM Watson: Healthcare organizations use IBM AI (Watson) technology for medial diagnosis

- Google's AI Eye Doctor can examine retina scans and identify a condition called as diabetic retinopathy which can cause blindness

- Facebook uses ML and DL to detect facial features and tag your friends

- Twitter uses AI to identify hate speech and terroristic language in the tweets

- Smart Assistants: Siri, Google Assistant, Alexa, Cortana

- Tesla automated cars

- Netflix uses AI for movie recommendations

- Spam filtering

Machine Learning

What is machine learning ?

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.
Machine Learning is the science (and art) of programming computers so they can learn from data.

Machine Learning (ML): Involves training algorithms to recognize patterns in data and make decisions or predictions without being explicitly programmed.

Deep Learning(DL) : A subset of machine learning that uses neural networks with many layers to analyze large amounts of data for highly accurate predictions.

Examples of Applications

- Analyzing images of products on a production line to automatically classify them
    - This is image classification, typically performed using convolutional neural networks

- Detecting tumors in brain scans
    - This is semantic segmentation, where each pixel in the image is classified (typically use CNNs)

- Automatically classifying news articles *sports politics education*
    - This is natural language processing (NLP), and more specifically text classification

- Automatically flagging offensive comments on discussion forums
    - This is also text classification, using the same NLP tools

- Forecasting your company's revenue next year, based on many performance metrics
    - This is a regression task (i.e., predicting values) that may be tackled using any regression model

Examples of Applications

- Making your app react to voice commands
  - This is speech recognition, which requires processing audio samples: since they are long and complex sequences, they are typically processed using RNNs, CNNs, or Transformers

- Detecting credit card fraud
  - This is anomaly detection example

- Segmenting clients based on their purchases so that you can design a different marketing strategy for each segment
  - This is clustering example

- Representing a complex, high-dimensional dataset in a clear and insightful diagram
  - This is data visualization, often involving dimensionality reduction techniques

Examples of Applications

- Recommending a product that a client may be interested in, based on past purchases
    - This is a recommender system

- Building an intelligent bot for a game
    - This is often tackled using Reinforcement Learning
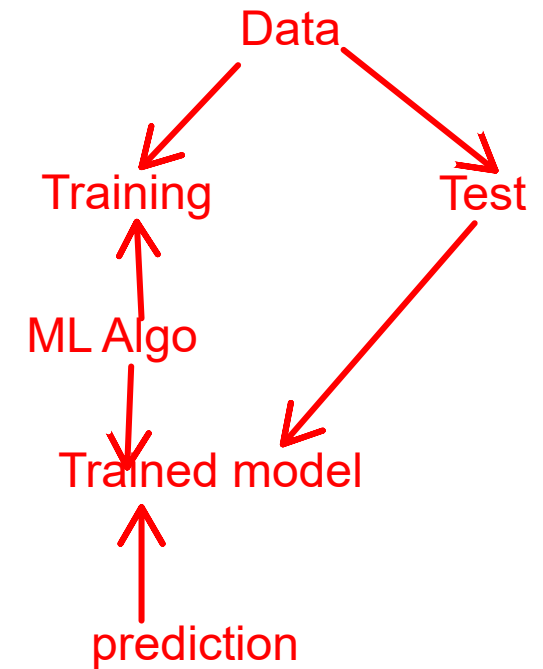
Take an Example

x	Y
1	1
2	4
3	9
4	16
5	25
8	??

test

$$y = x^2$$

Formula /
Model /
Function

$$= 8^2$$

$$= 64$$

Data

Training   Test

ML Algo

Trained model

prediction

input/independent   output/dependent variable

X	Y
1	1
2	4
3	9
4	16
5	25

$Y = X^2$ formula/model/function

dataset

accuracy = good

X = 6   Y=36
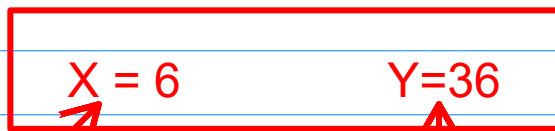
unseen
data   prediction

ML Program

Non ML Programming

Input → Program Implement Formula → Output

ML Programming

Existing Data For Understanding → ML Program → Formula → Output
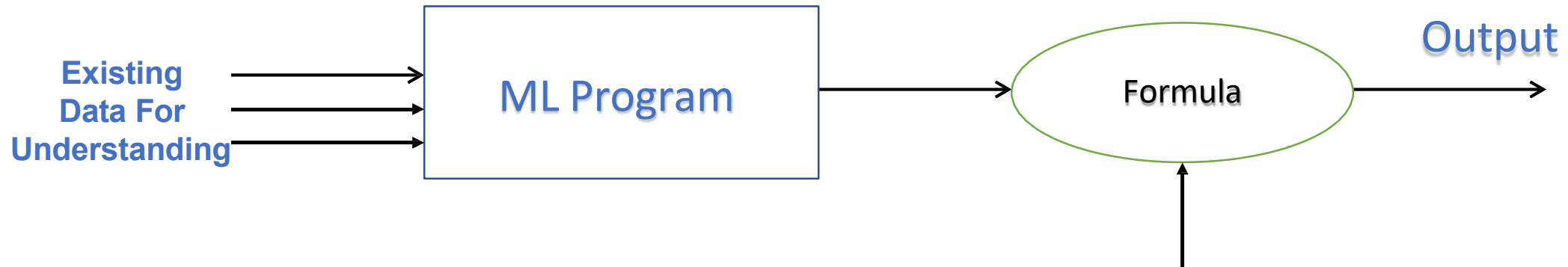
Where to use Machine Learning ?

- Problems for which existing solutions require a lot of fine-tuning or long lists of rules:
  - one Machine Learning algorithm can often simplify code and perform better than the traditional approach

- Complex problems for which using a traditional approach yields no good solution:
  - the best Machine Learning techniques can perhaps find a solution

- Fluctuating environments:
  - a Machine Learning system can adapt to new data

- Getting insights about complex problems and large amounts of data

Types

Types of Machine Learning

- There are so many different types of Machine Learning systems that it is useful to classify them in broad categories, based on the following criteria

    - Whether or not they are trained with human supervision
        - supervised, unsupervised, and Reinforcement Learning

Variable / Column / Feature

50 L

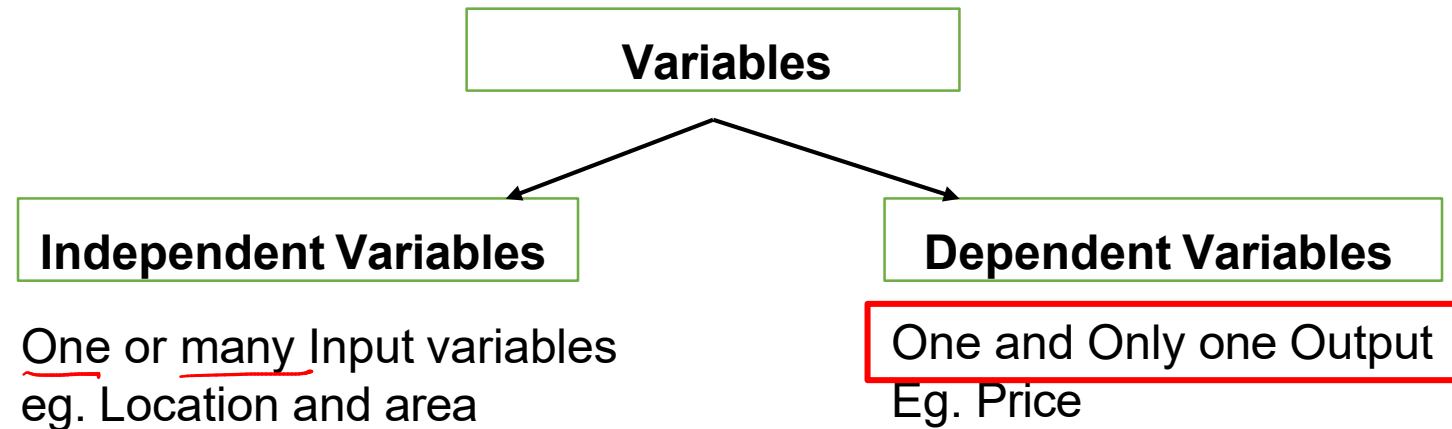| No.	Location	Area	Price
1	Pune ✓	Abc (Hinjewadi)	70 L
2	Mumbai ✓	Pqr (Navi Mumbai) ↑ Panvel	90 L

Known Data
Or
Labelled Data

No. column not required
Location and Area are Independent variables
Price is Dependent Variable

Variables

↓ ↓

Independent Variables

One or many Input variables
eg. Location and area

Dependent Variables

One and Only one Output
Eg. Price

Supervised
Unsupervised
Reinforcement Learning

Supervised Learning

- The majority of practical machine learning uses supervised learning

- Supervised learning is where you have **input variables (x)** and an **output variable (Y)** and you use an algorithm to learn the mapping function from the input to the output

$$Y = f(X) \implies \text{model / formula} \quad ,$$

X | Y
1 | 1
2 | 4
3 | 9
4 | 16

- The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

- It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process

- We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher

- Learning stops when the algorithm achieves an acceptable level of performance (measured in terms of **accuracy**)

Supervised Learning

- Output variable is already known for each input variable
- Algorithm learns to map input and output
- Model learns to associate features in the images with these predefine categories.
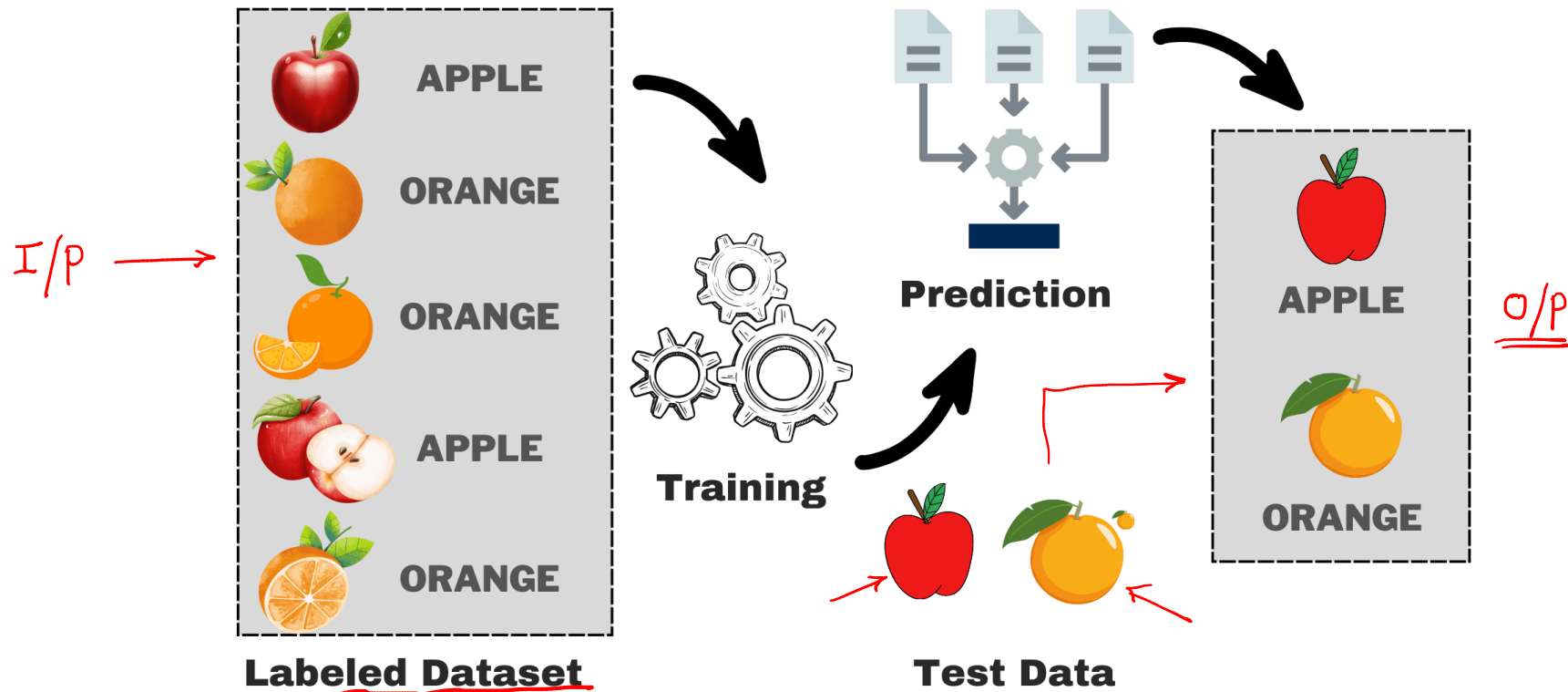
Supervised Learning – Problems

- **Regression**
    - Related to predicting future values
    - E.g.
        - Population growth prediction
        - Expecting life expectancy
        - Market forecasting/prediction
        - Advertising Popularity prediction
        - Stock prediction
    - Algorithms
        - Linear and multi-linear regression
        - Logistic regression
        - Naïve Bayes
        - Support Vector Machine

Supervised Learning – Problems

- **Classification**
  - Related to classify the records
  - Based on class / labels (eg. Email : Spam / Ham , Gender : Male / Female , Loan : Yes / No)

    decide based

    discount/ promotn   *bank statement*

  - E.g.
    - Find whether an email received is a spam or ham
    - Identify customer segments
    - Find if a bank loan is granted
    - Identify if a kid will pass or fail in an examination
  - Algorithms
    - Logistic Regression
    - Decision Tree
    - Random Forest
   - Support Vector Machine
    - K-nearest neighbor

Unsupervised Learning

- Unsupervised learning is where you only have input data (X) and no corresponding output variables

- The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data

- These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher

- Algorithms are left to their own devises to discover and present the interesting **structure** in the data

- Structure in the form of GROUPS / CLUSTERS / ASSOCIATION

- Mostly used for EDA (Exploratory Data Analysis)

- Is to understand the data's structure, find patterns, identify anomalies or outliers, and check assumptions before formal modeling or hypothesis testing. EDA is a crucial first step in data analysis and machine learning projects to gain initial insights and ensure data quality
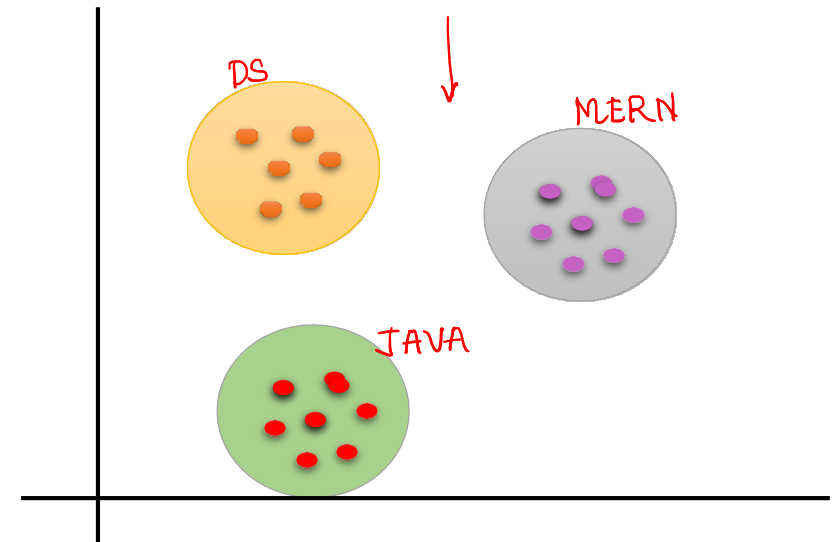
Unsupervised Learning

- Analyzes and clusters unlabeled datasets.
- These algorithms find hidden patterns and data without any human intervention.
- The training model has only input parameter values and discovers the groups or patterns on its own.
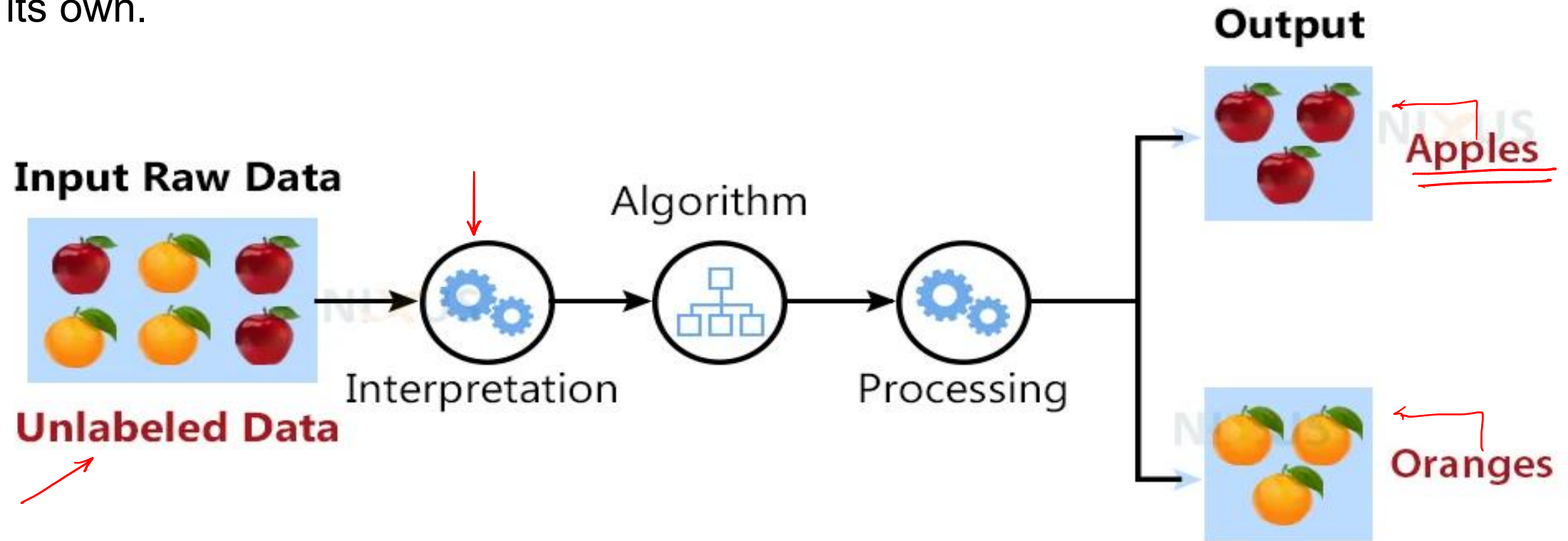
Unsupervised Learning - Problems

- **Clustering**
    - discover the inherent groupings in the data, such as grouping customers by purchasing behaviour
    - E.g.
        - Batsman vs bowler
        - Customer spending more money vs less money
    - Algorithms
        - K-means clustering
        - Hierarchical clustering

Unsupervised Learning - Problems

- **Association**
  - An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y
  - E.g.
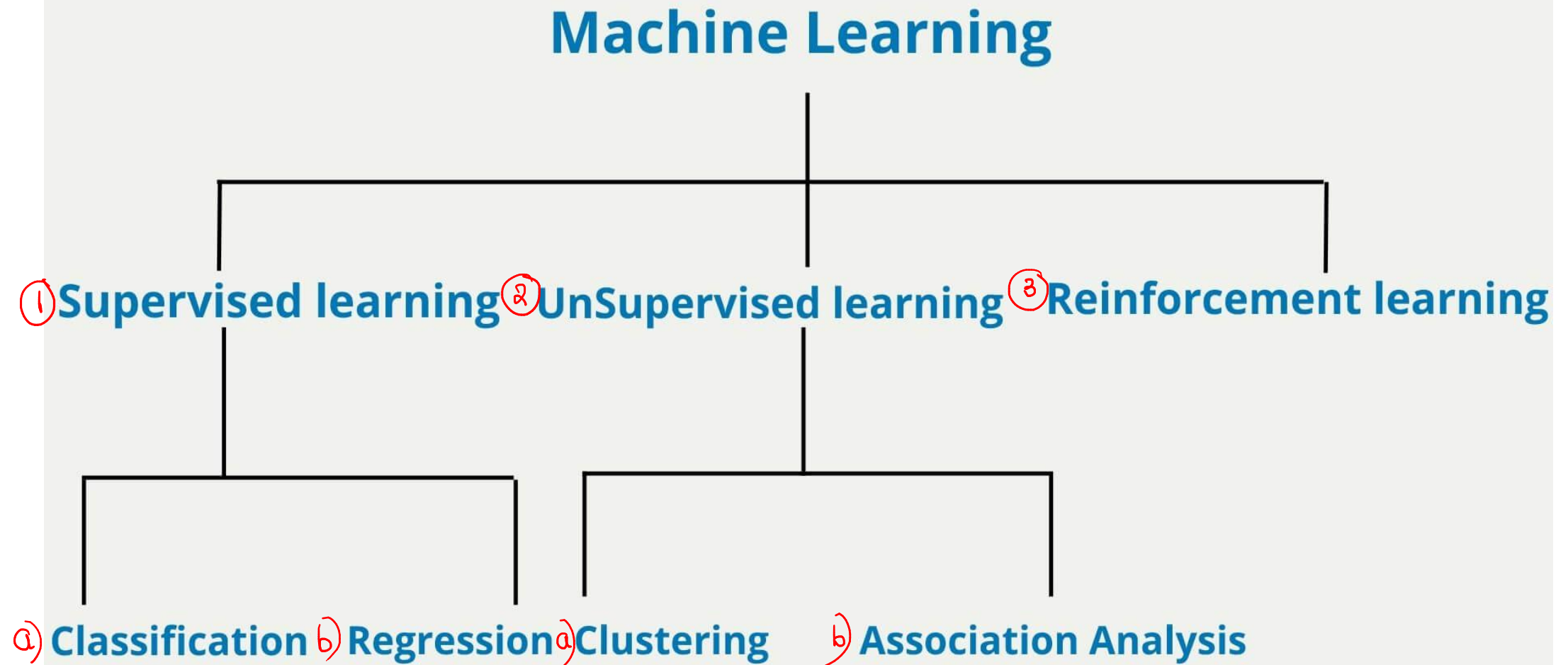   - Market basket analysis ✓
  - Algorithms
    - Apriori ✓
    - Eclat ✓

Stationary
Cloth store
Electronic gadgets

Machine Learning

Machine Learning

1. Supervised learning
   a. Classification
   b. Regression

2. UnSupervised learning
   a. Clustering
   b. Association Analysis

3. Reinforcement learning

Thank You!!