Unit I: Machine Learning Fundamentals

CTE309 Machine Learning

AS2024: BE Information Technology

College of Science and Technology

Royal University of Bhutan

Overview

- Introduction
- Machine learning
- ML vs Deep learning
- History of ML
- Importance of ML
- ML Applications
- Types of Machine learning
- ML pipeline
- Data preprocessing

Introduction

- Machines have come a long way since the Industrial Revolution.
  - Capabilities extend beyond manual activities to cognitive tasks
  - Eg. Self-driving car
- Triggers fear among observers
  - Eg. Threat to job security
  - BBC's online resource: *bar worker (77%), waiter (90%), chartered accountant (95%), receptionist (96%), and taxi driver (57%) each have a high chance of becoming automated by the year 2035.*
- AI technology is moving fast but adoption is still unchartered path

Introduction Cont…

- Machine learning (ML) is far from *out-of-the-box* solutions
- Machine operates based on statistical algorithms and overseen by skilled individuals – *data scientists and machine learning engineers*
- Probable labor markets – job opportunities are on growth
  - Supply is struggling to meet the demand
- Classical statistics – at the heart of ML
  - Drive cognitive abilities of machine
- Computer programming – another indispensable part of ML
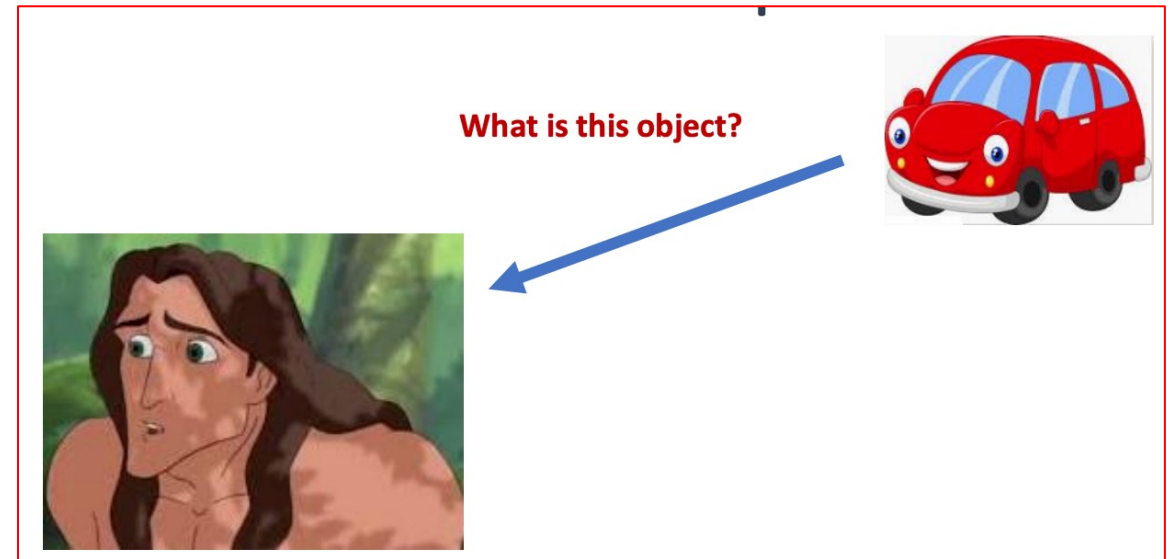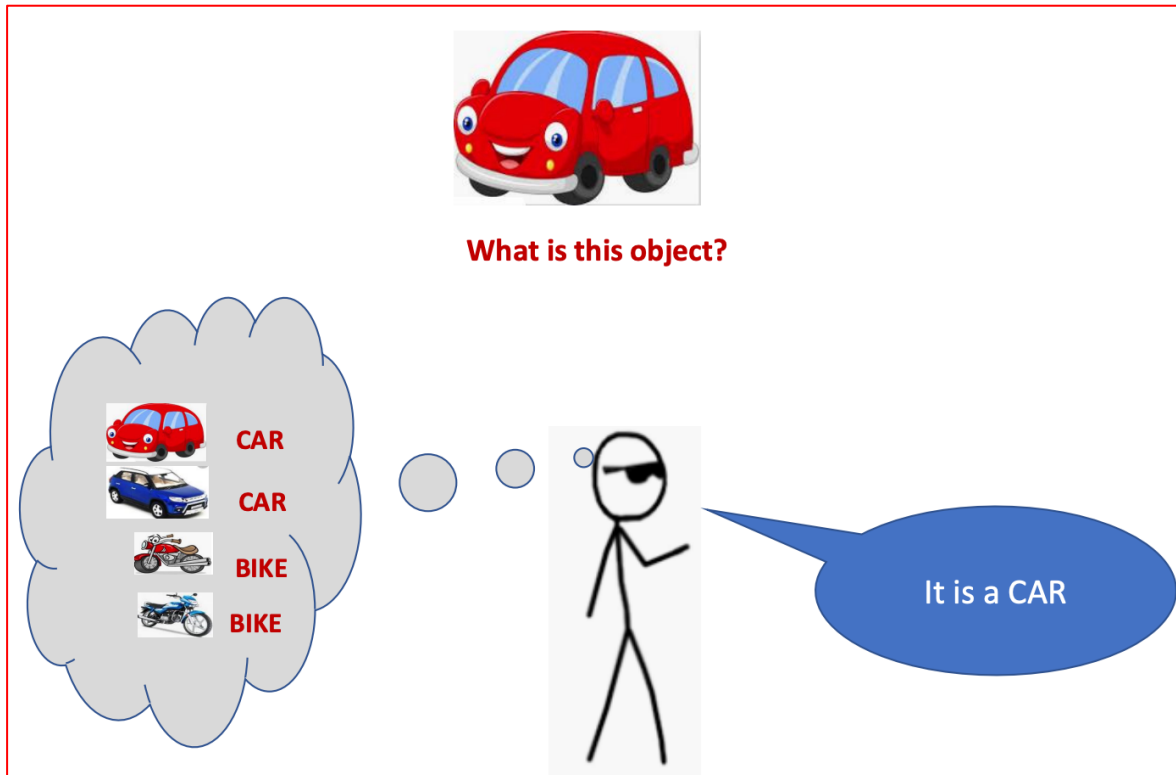
What is Machine Learning (ML)?

What is this object?

Human can learn from past experience and make decision of its own

What is Machine Learning (ML)?
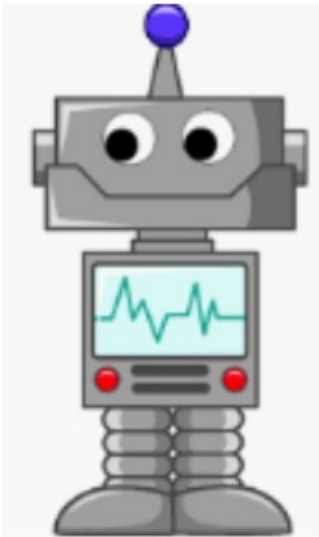
What is Machine Learning (ML)?

What is this object?
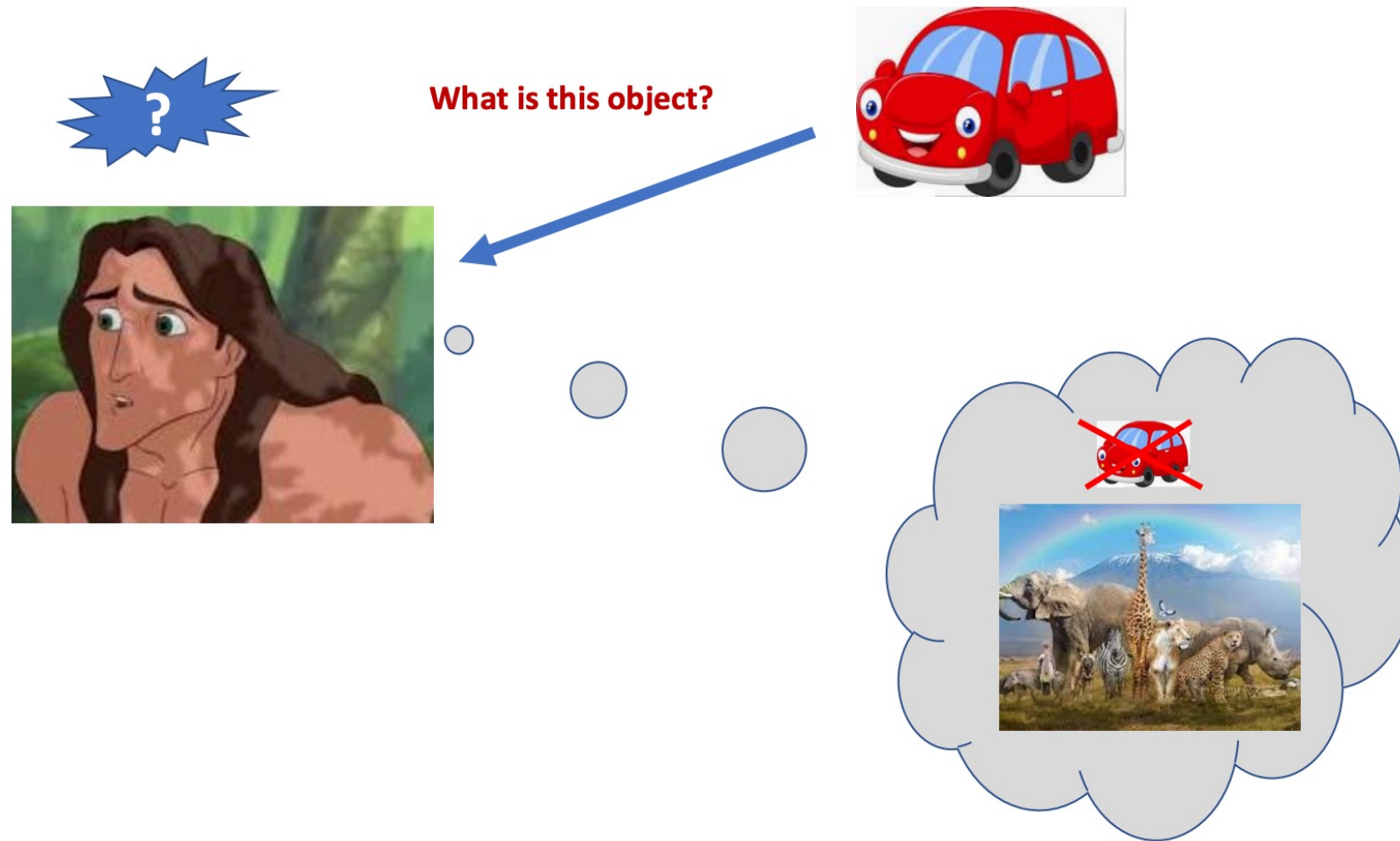
But he is a human being. He can observe and Learn.

What is Machine Learning (ML)?
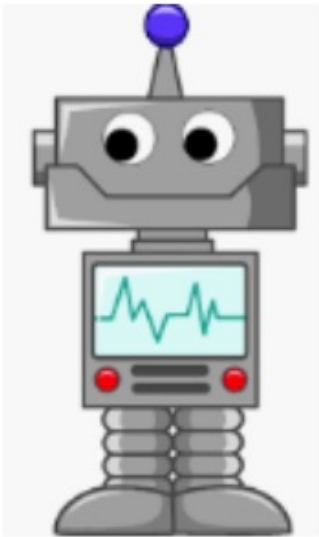

Past experiences
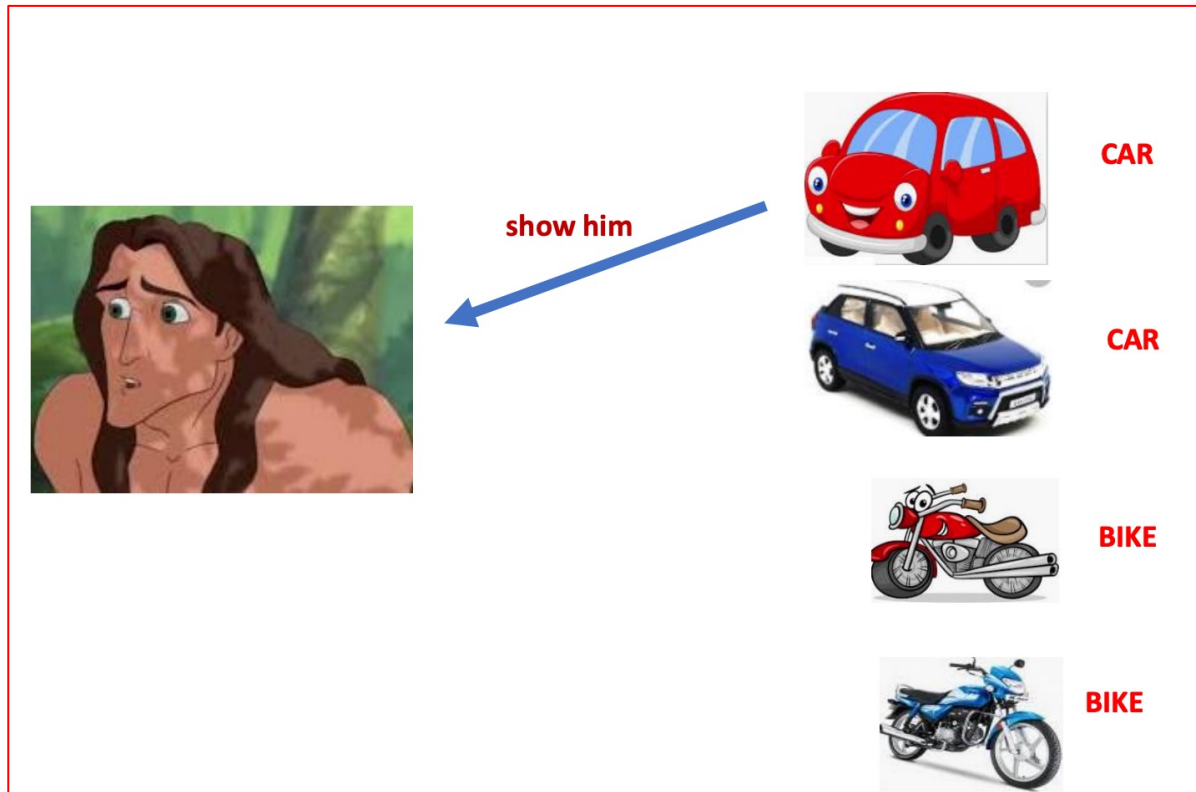
What is Machine Learning (ML)?

- Perform arithmetic operations
- Comparison
- Print
- Plot the graph

Machines follow instructions

Lack decision making capability

What is Machine Learning (ML)?

We want Machine to act like us!!

Machines follow instructions

Identify the object

Price in 2025?
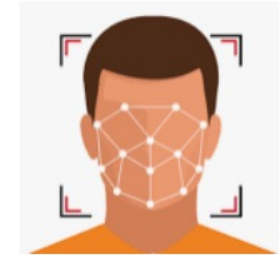
recognize face

I ~~made~~ met him yesterday

Understand Natural Language and correct grammar

Lack decision making capability

What is Machine Learning (ML)?

Dataset

[So that machine can derive inferences from the data]
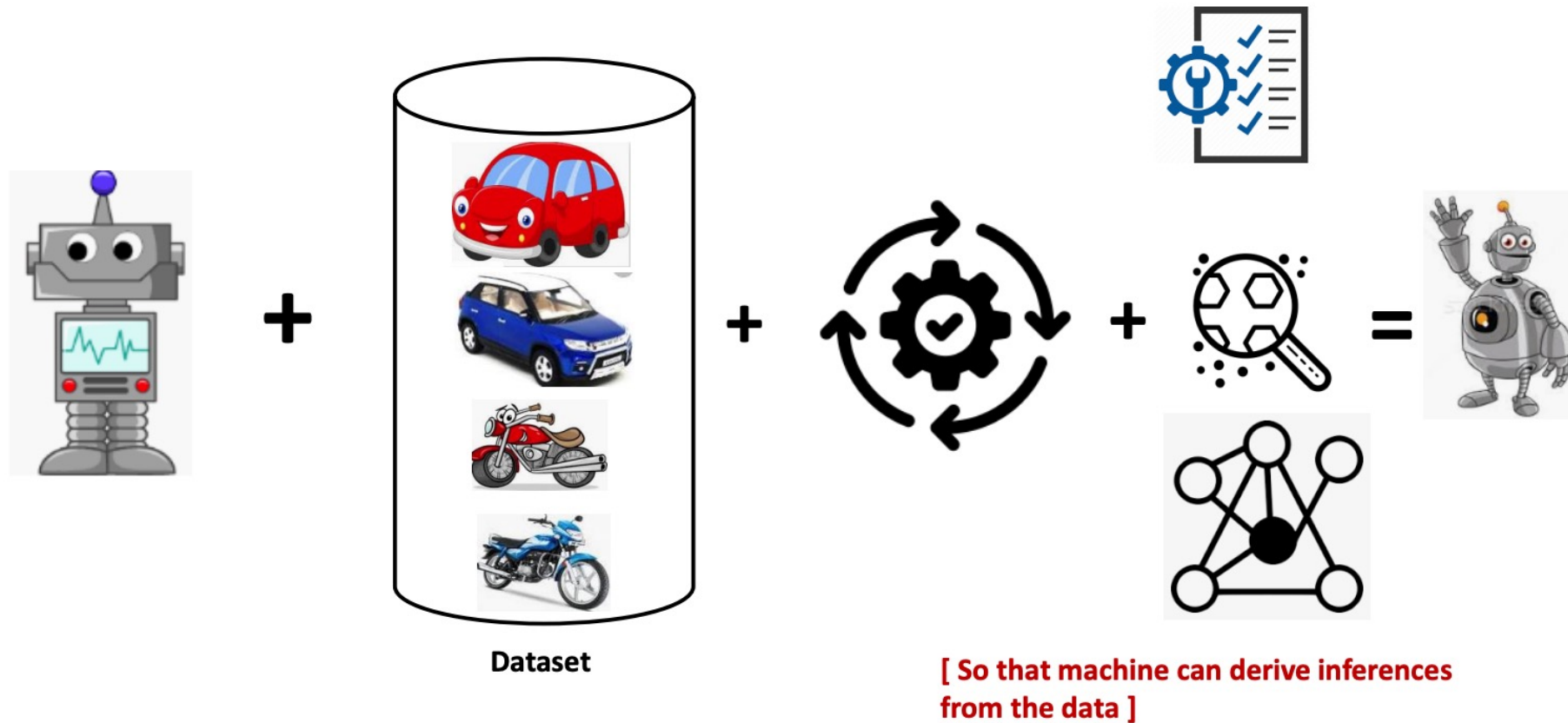
What is Machine Learning (ML)?

- In its simplest form, machine learning is a set of algorithms learned from data and/or experiences, rather than being explicitly programmed.

- Building machines that automatically learn from experience.

- a subfield of computer science that gives computers the ability to learn without being explicitly programmed.

ML vs Deep Learning
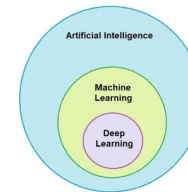
Deep Learning:

- Based on neural networks
- Uses deep architectures
- Very successful in many applications
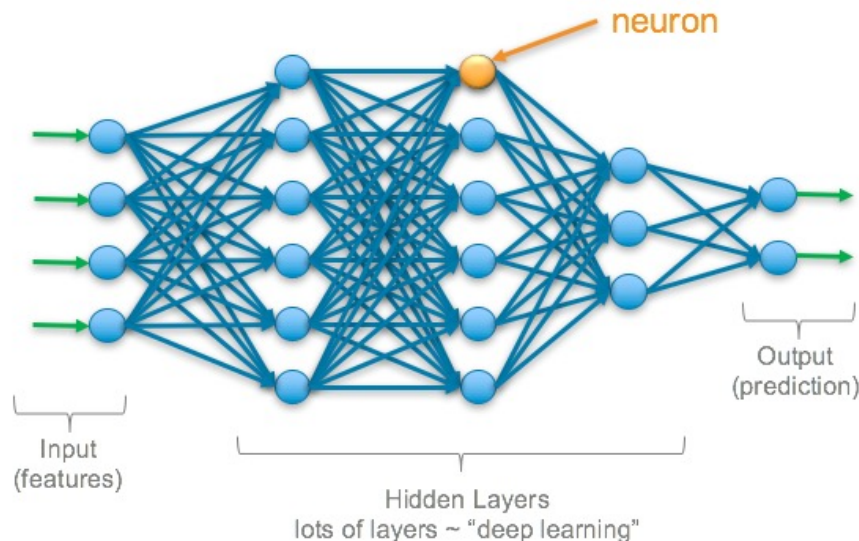
Both are hot topics and buzzwords in tech industries

deep learning and machine learning are both subsets of artificial intelligence.

And deep learning is a subset of machine learning.

Machine learning is an AI technique, and deep learning is a machine learning technique.

ML vs Deep Learning

Deep Learning Vs Machine Learning

Factors	Deep Learning	Machine Learning
Data Requirement	Requires large data	Can train on lesser data
Accuracy	Provides high accuracy	Gives lesser accuracy
Training Time	Takes longer to train	Takes less time to train
Hardware Dependency	Requires GPU to train properly	Trains on CPU
Hyperparameter Tuning	Can be tuned in various different ways.	Limited tuning capabilities

Artificial Intelligence, ML and Deep Learning

ARTIFICIAL INTELLIGENCE

The ability of machines to simulate human behaviour and take decisions intelligently like a human.

MACHINE LEARNING

The ability of machines to automatically learn without being explicitly programmed.

DEEP LEARNING

It creates artificial neural networks, which are capable of learning and taking decisions intelligently with the help of algorithms.

ML with its parent fields and Sister Disciples

History of ML

1950
Alan Turing created a test to check if a machine could fool a human being into believing it was talking to a machine.

1957
First neural network for computers (the perceptron) was invented by Frank Rosenblatt, which simulated the thought processes of the human brain.

1979
Students of Stanford University, California, invented the Stanford Cart which could navigate and avoid obstacles on its own.

2002
A software library for Machine Learning, named Torch is first released.

1952
The first computer learning program, a game of checkers, was written by Arthur Samuel.

1967
The Nearest Neighbor Algorithm was written.

1997
IBM's Deep Blue beats the world champion at Chess.

2016
AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Chinese Board Game Go competition.
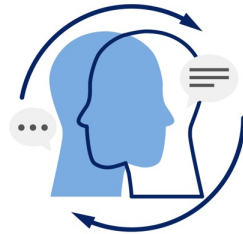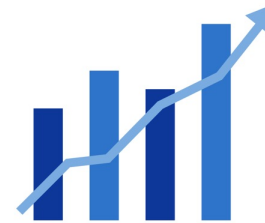
Importance of ML

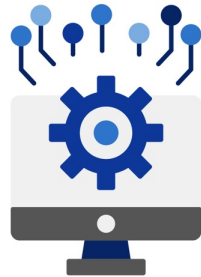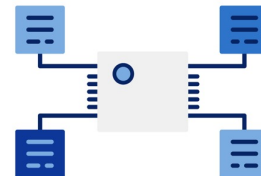
Automation
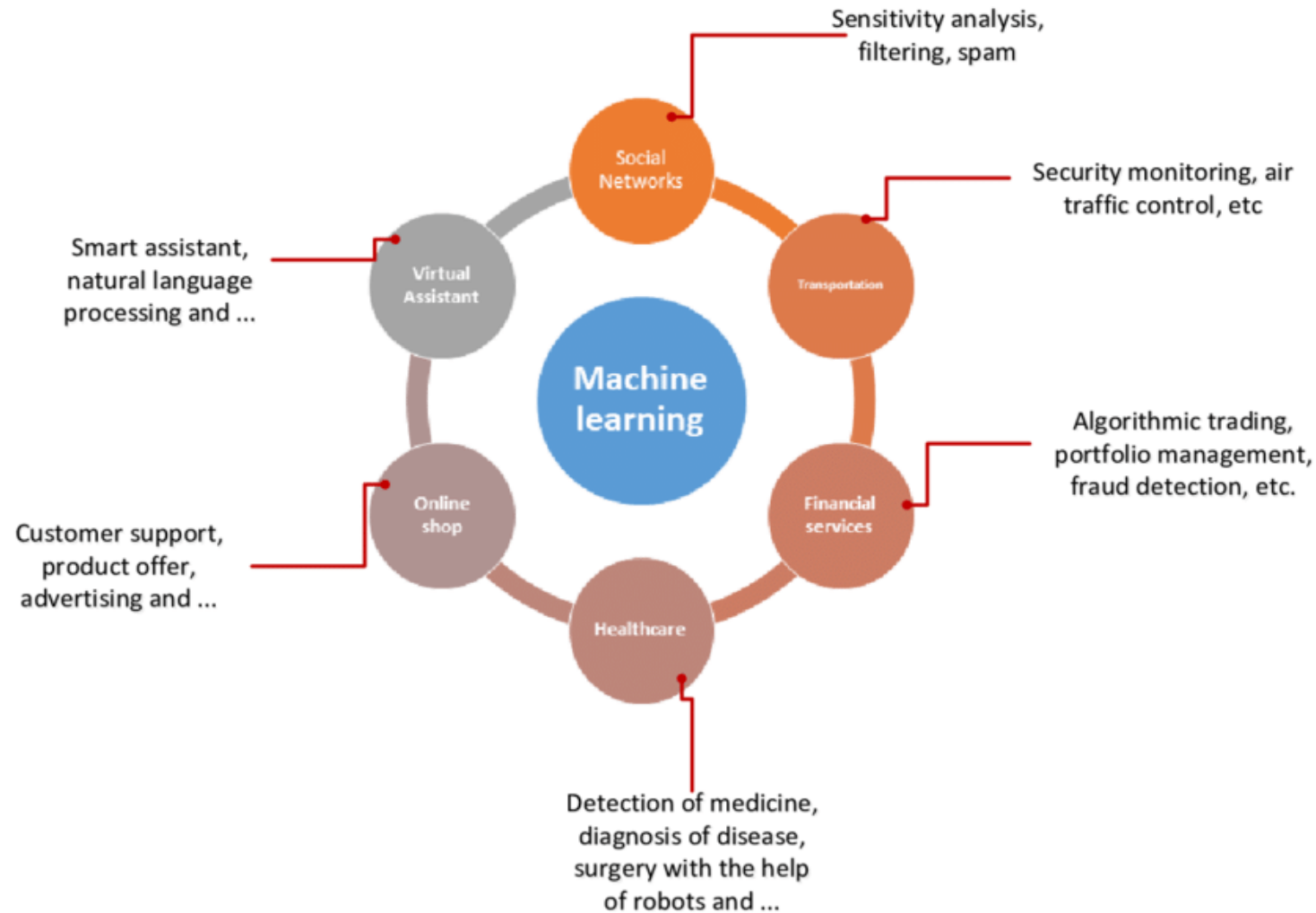
Less reliance on
human interaction
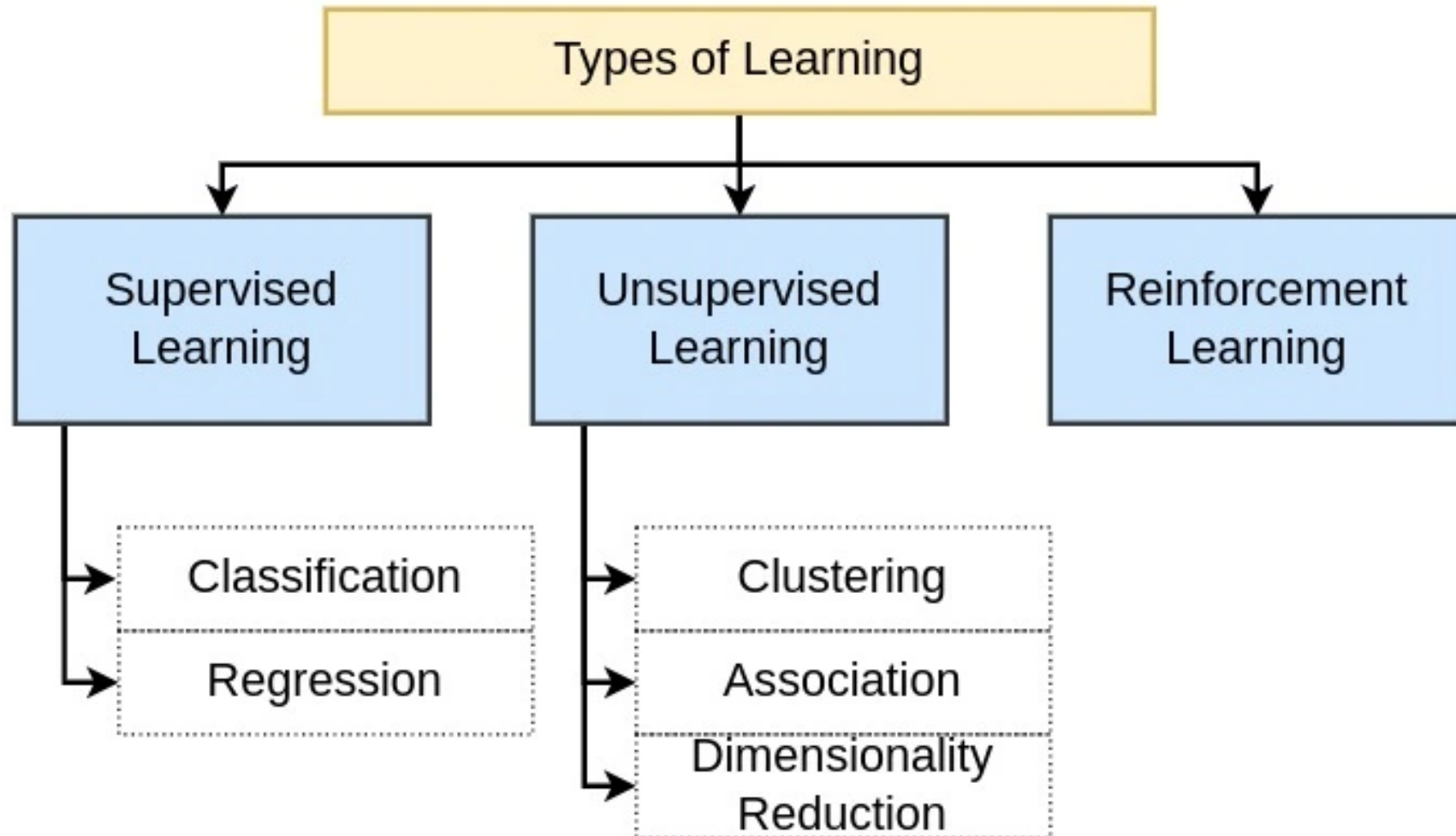
Scope of
improvement
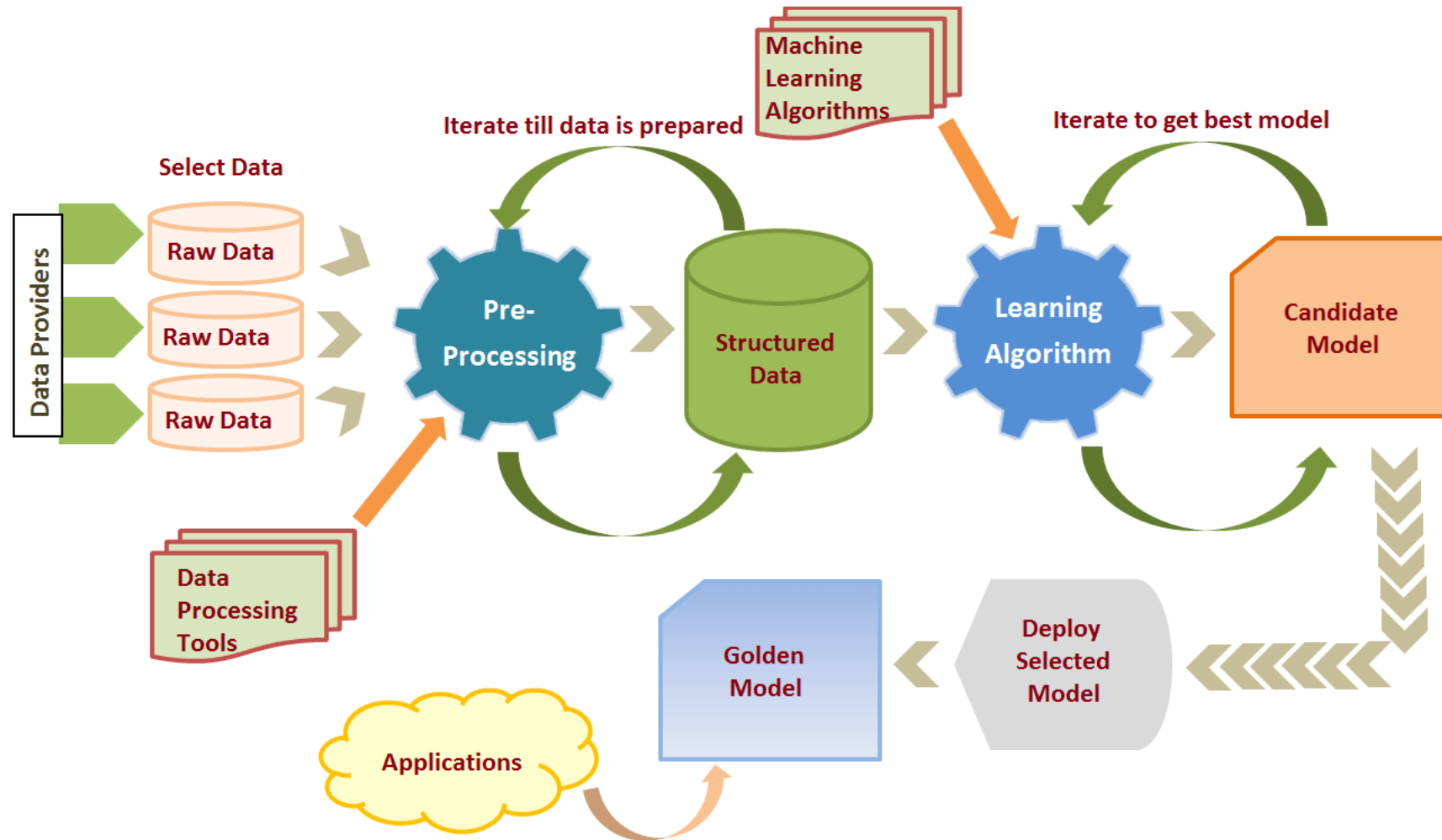
Efficient
data handling

Wide range
of applications

Applications of ML

ML Categories/Paradigm

Machine Learning Pipeline

Dataset

- a set of data grouped into a collection
- gain insights, make informed decisions, or train algorithms.
- mostly require cleaning and preprocessing to ensure data quality and suitability for analysis or modeling.
- Datasets can be stored in multiple formats
  - CSV, Excel, JSON
- Types:
  - Numerical Dataset
  - Categorical Dataset
  - Time series Dataset
  - Image Dataset

```
{
  "DocumentType": 1,
  "No.": "S-ORD101001",
  "SellToCustNo": "10000",
  "PostingDate": "2023-04-02",
  "Lines": [
    {
      "LineNo": 10000,
      "Type": 2,
      "No": "1996-S",
      "Quantity": 12,
      "UnitPrice": 1397.3
    },
    {
      "LineNo": 20000,
      "Type": 2,
      "No": "1900-S",
      "Quantity": 4,
      "UnitPrice": 192.8
    }
  ]
}
```

Data Preprocessing

- an iterative process for the transformation of the raw data into understandable and useable forms.

- Characteristics of raw data
  - incompleteness, inconsistencies, lacking in behavior, and trends while containing errors

- handle the *missing values* and address *inconsistencies*.

- Transforming data into suitable formats for a particular machine-learning problem – *essential consideration*

Data Preprocessing

- Model outcomes and knowledge discovery is significantly affected
  - presence of **irrelevant**, **redundant information**, **noisy**, and **unreliable data**
  - making the training phase more difficult.
- take the most amounts of time spent on an ML project but worth it.
- The steps involved include cleaning, instance selection, normalization, transformation, feature extraction, and selection.
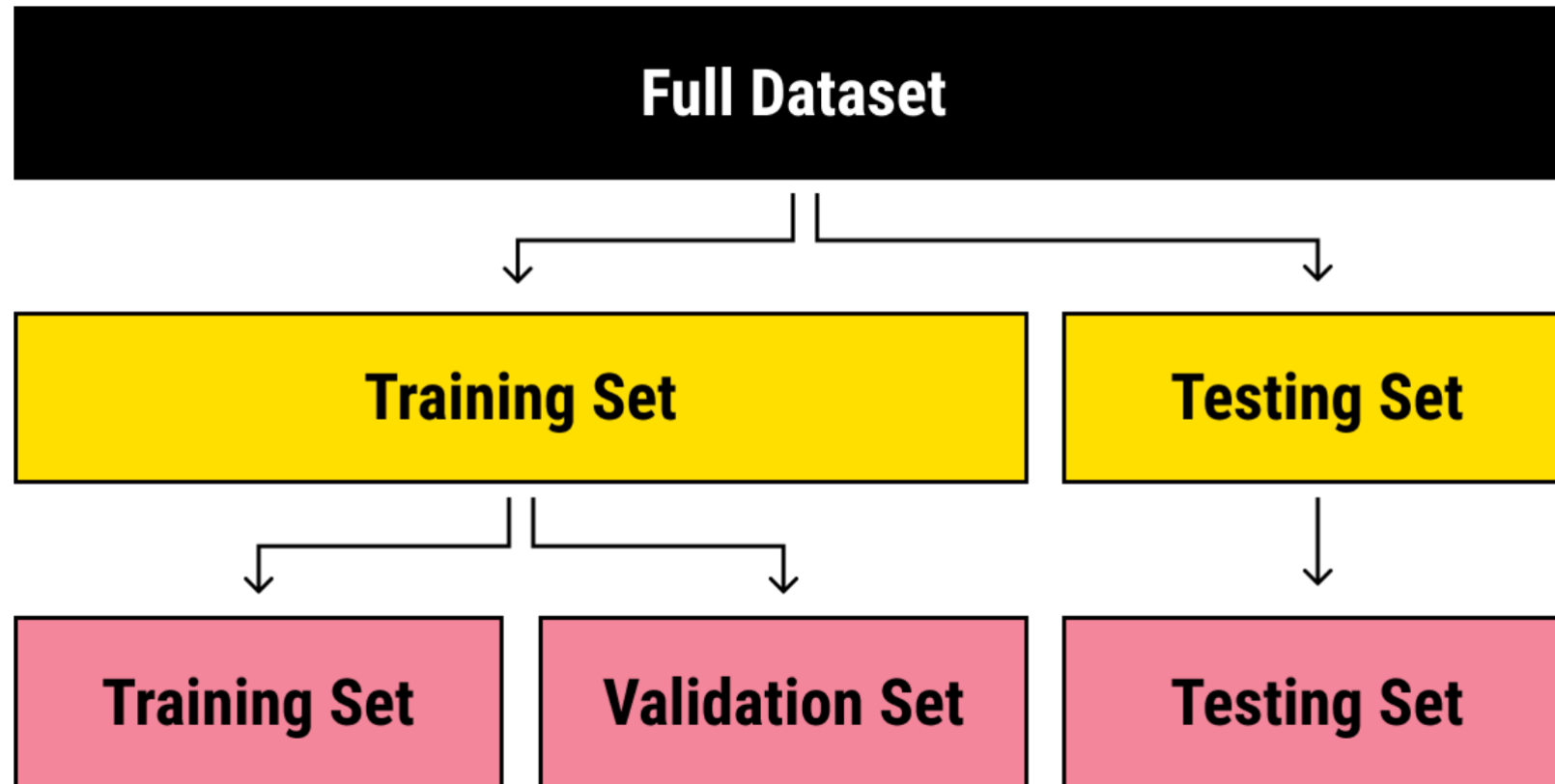
Data

Country	Age	Salary	Purchased
France	44	72000	No
Spain	27	48000	Yes
Germany	30	54000	No
Spain	38	61000	No
Germany	40		Yes
France	35	58000	Yes
Spain		52000	No
France	48	79000	Yes
Germany	50	83000	No
France	37	67000	Yes

Categorical Encoding

Column Names	One-Hot Encoding	Label Encoding
Description	Converts each unique category value into a new binary column.	Assigns each unique category value an integer.
Example	"India" -> [1, 0, 0] "Japan" -> [0, 1, 0] "USA" -> [0, 0, 1]	"India" -> 0 "Japan" -> 1 "USA" -> 2
When to Use	Non-ordinal categorical features. Manageable number of unique categories.	Ordinal categorical features. Large number of unique categories.
Advantages	Prevents the model from assuming any inherent order.	Simple and efficient for ordinal data. Does not increase dimensionality.
Disadvantages	Can lead to high dimensionality with many unique categories.	Imposes an arbitrary order on non-ordinal data. Model might assume false relationships.

Dataset Split

Thank you