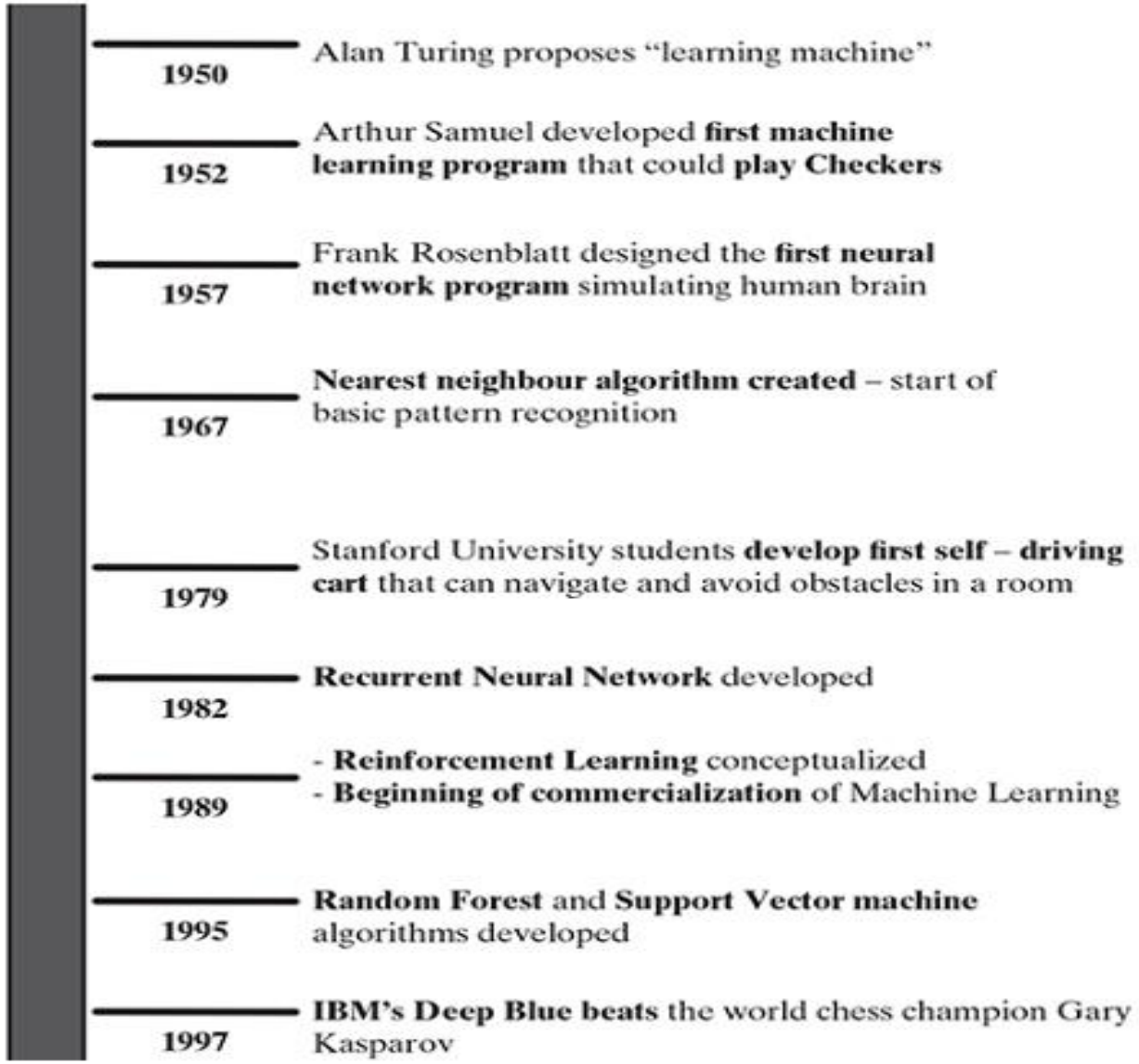
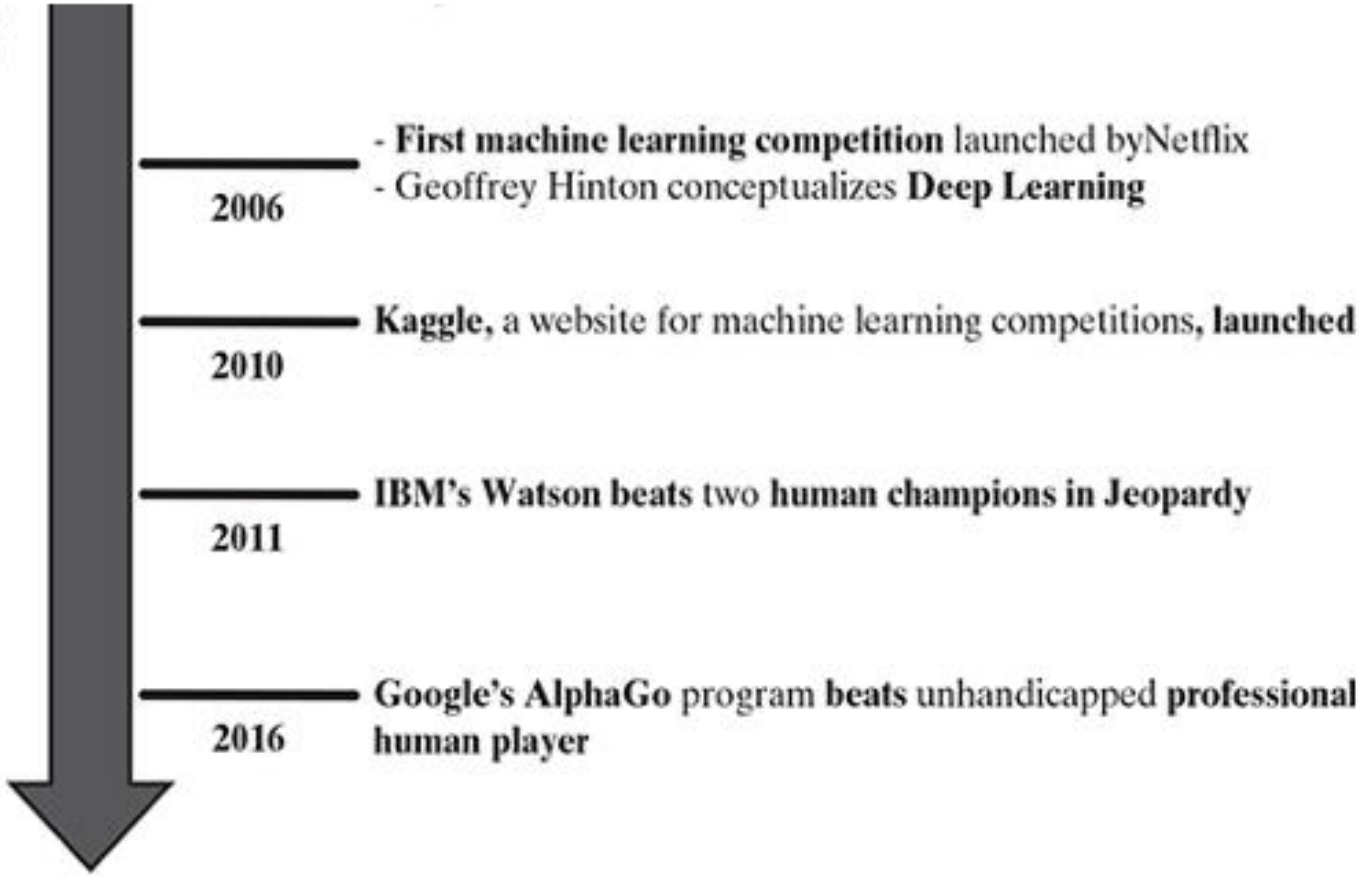
Unit 1 – Introduction to Machine Learning

Ms. Geetanjali R
Assistant Professor,
Department of MCA
Ramaiah Institute of Technology

Introduction

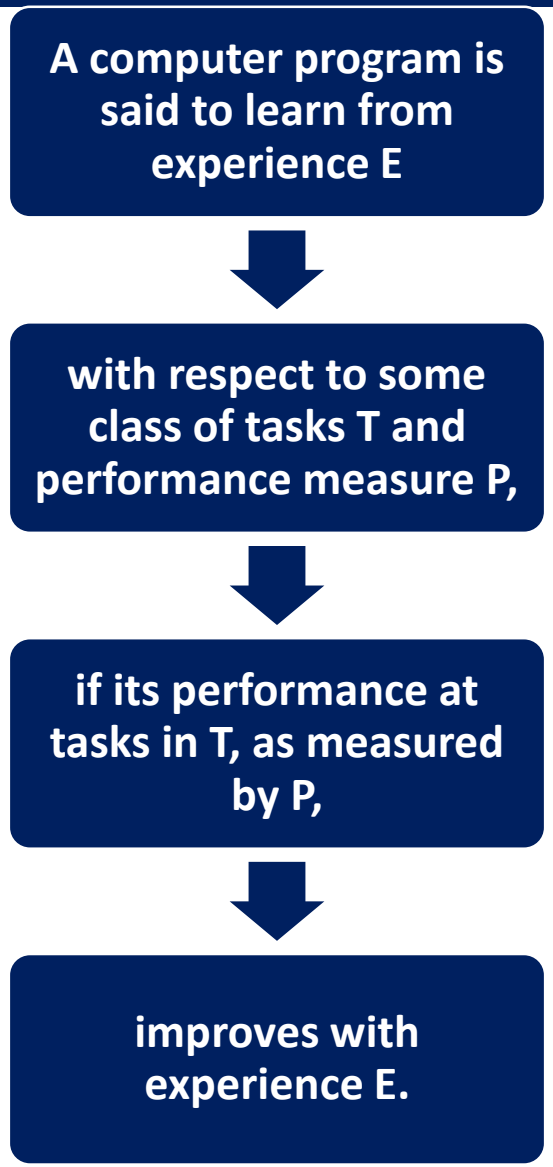


Introduction



Machine Learning v/s Human Learning

Machine Learning v/s Human Learning



Machine Learning v/s Human Learning

A computer program is said to learn from experience E

with respect to some class of tasks T and performance measure P,

if its performance at tasks in T, as measured by P,

improves with experience E.

How do machines learn?

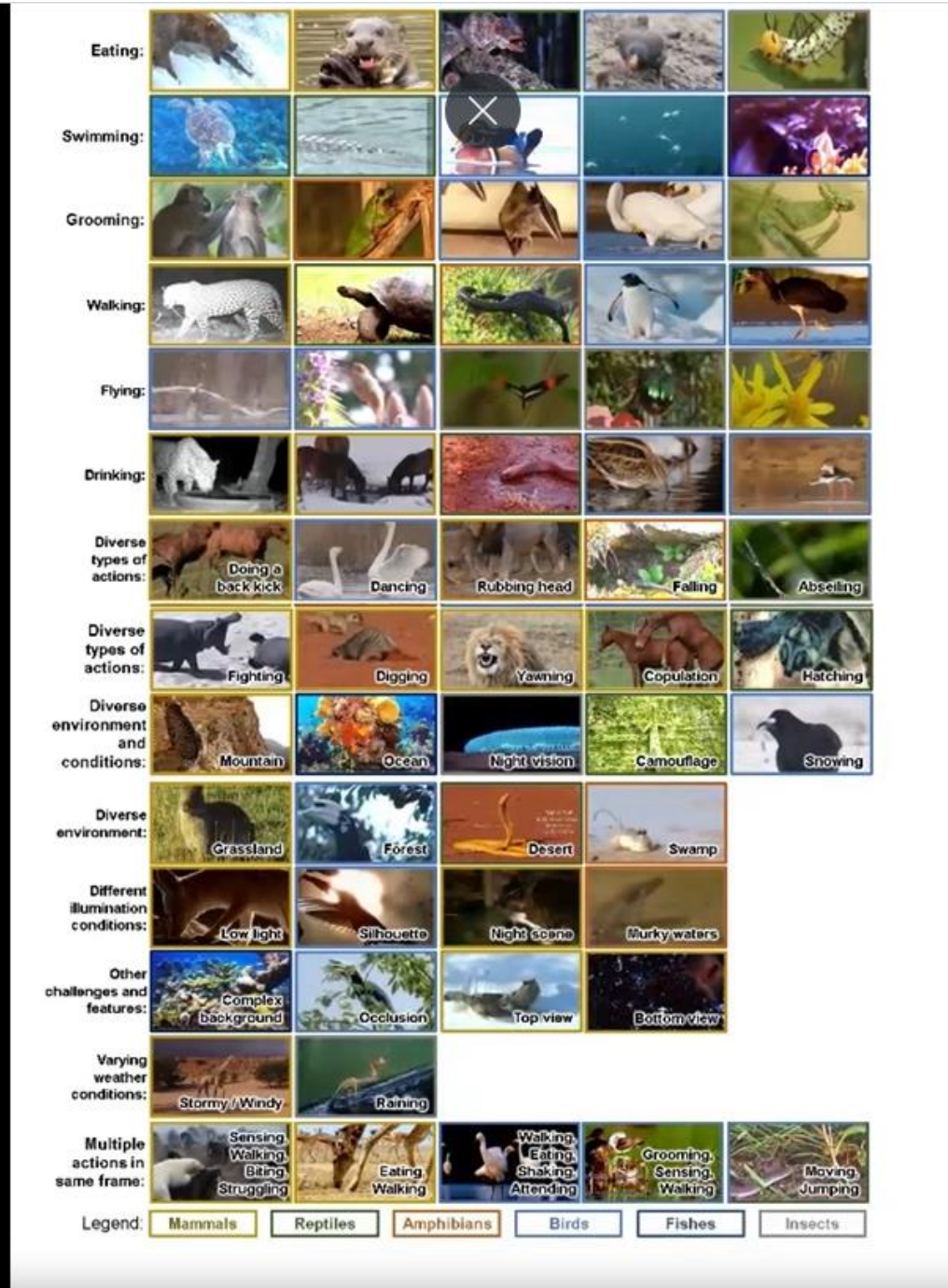
A broad pool of knowledge may consist of all living animals

their characteristics such as whether they

- **live in land or**
- **water, whether they lay eggs,**
- **whether they have scales or fur or none, etc.**

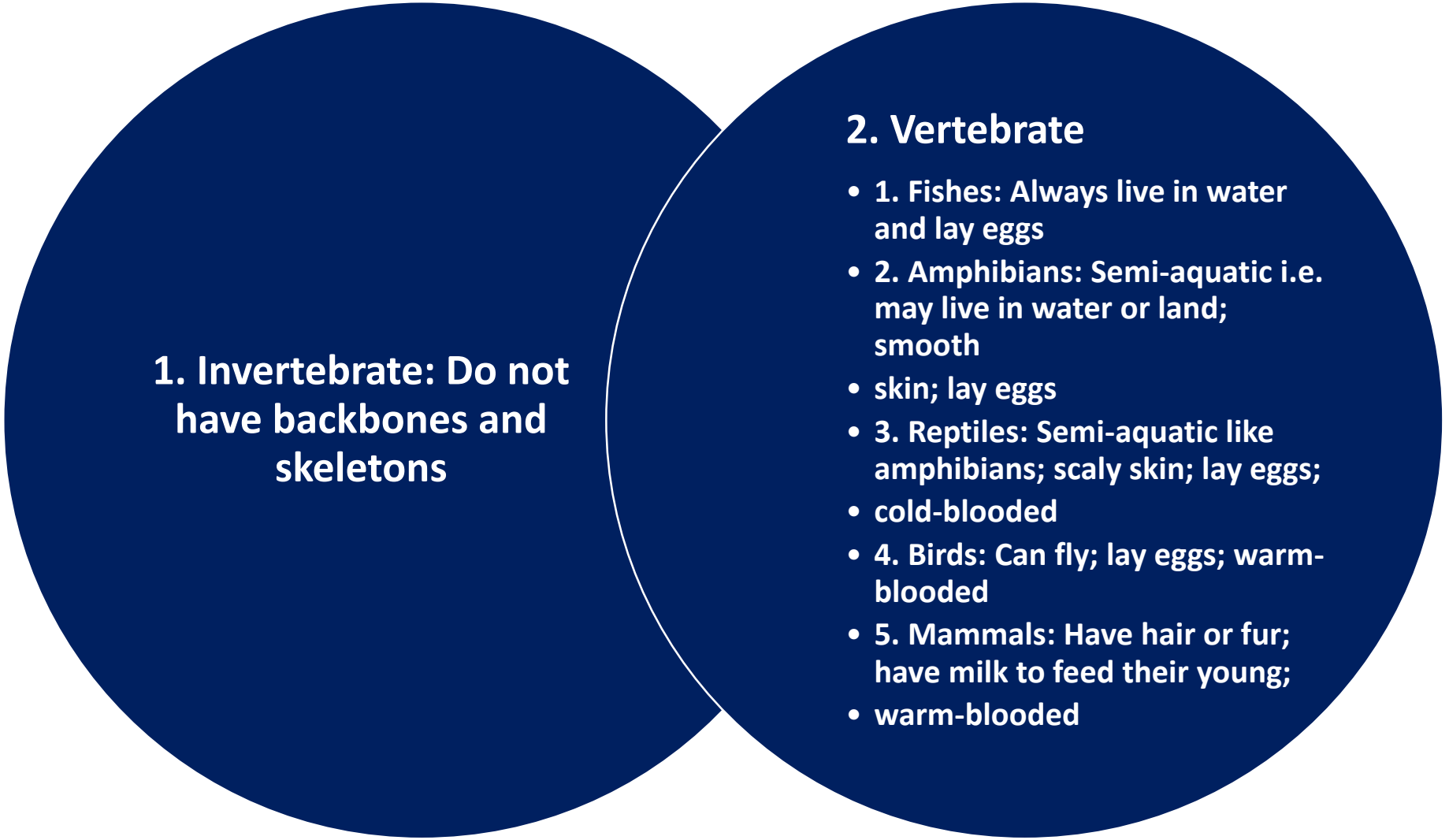
It is a difficult task for any student to memorize the characteristics of all living animals – no matter how much photographic memory he/she may possess.


Doing a back kick
Dancing
Rubbing head
Falling
Abseiling
Fighting
Digging
Yawning
Copulation
Hatching
Mourning
Ocean
Night vision
Camouflage
Shivering
Grassland
Forest
Desert
Swamp
Low light
Silhouette
Night scene
Murky waters
Complex background
Occlusion
Top view
Bottom view
Stormy / Windy
Raining
Sensing
Walking
Biting
Struggling
Eating
Walking
Walking
Eating
Shaking
Attending
Grooming
Sensing
Walking
Moving
Jumping



Activate Windows
Go to Settings to activate Windows.

How do machines learn?



How do machines learn?

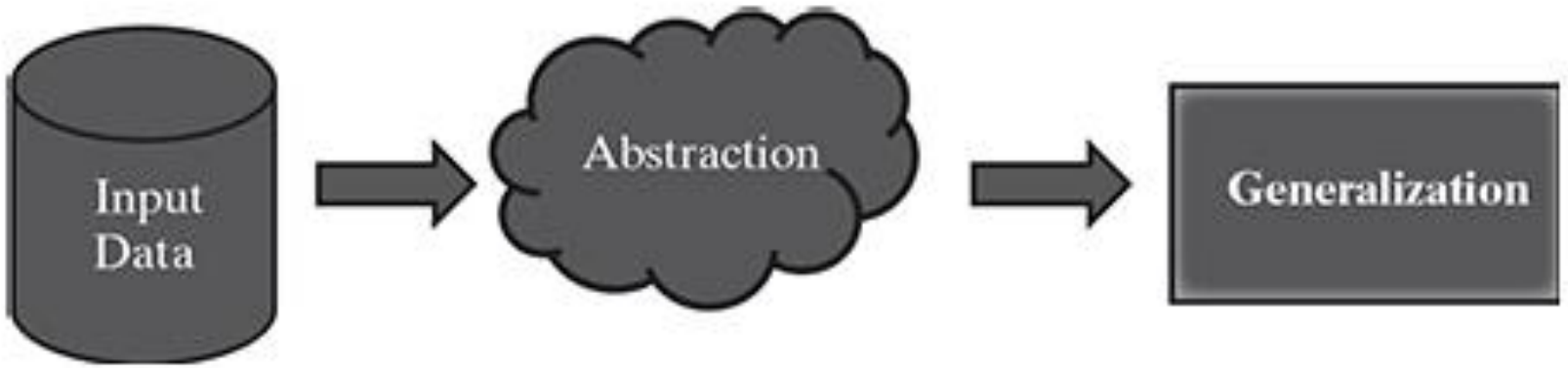


FIG. 1.2 Process of machine learning

How do machines learn?

- **Data cannot be used in the original shape and form**

- **Rather than using it in entirety, a concept map, much in line with the animal group to characteristic mapping is s drawn from the input data Vertebrate**

ABSTRACTION

How do machines learn?

Abstract the knowledge which comes as input data in the form of a model.

However, this abstraction process, or more popularly training the model, is just one part of machine learning.

The other key part is to tune up the abstracted knowledge to a form which can be used to take future decisions.

How do machines learn?

This is achieved as a part of generalization.

This part is quite difficult to achieve.

This is because the model is trained based on a finite set of data, which may possess a limited set of characteristics.

How do machines learn?

But when we want to apply the model to take decision on a set of unknown data, usually termed as test data, we may encounter two problems :

- **The trained model is aligned with the training data too much, hence may not portray the actual trend.**
- **The test data possess certain characteristics apparently unknown to the training data.**

How do machines learn?

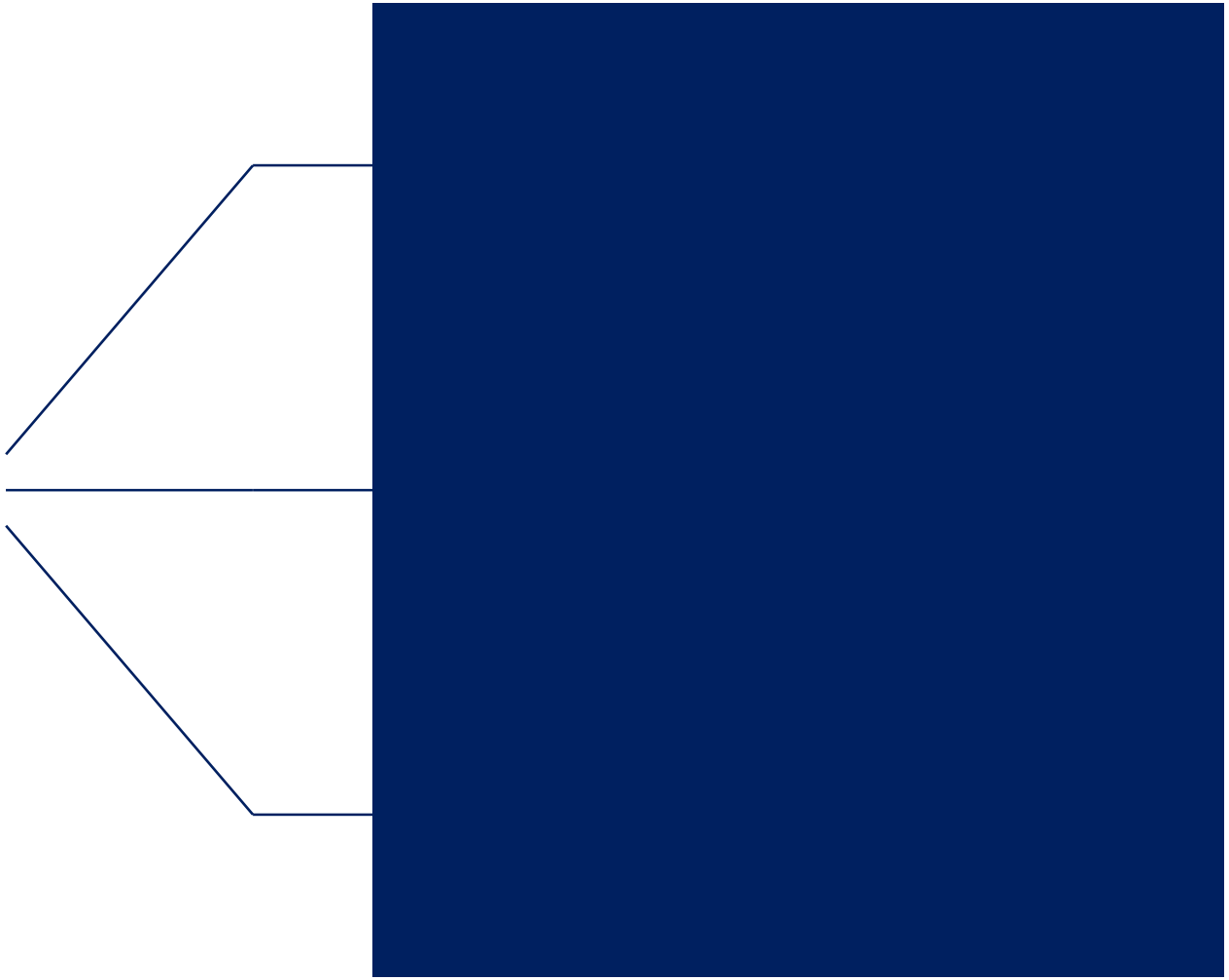
Hence, a precise approach of decision-making will not work.

An approximate or heuristic approach, much like gut feeling-based decision-making in human beings, has to be adopted.

This approach has the risk of not making a correct decision – quite obviously because certain assumptions that are made may not be true in reality

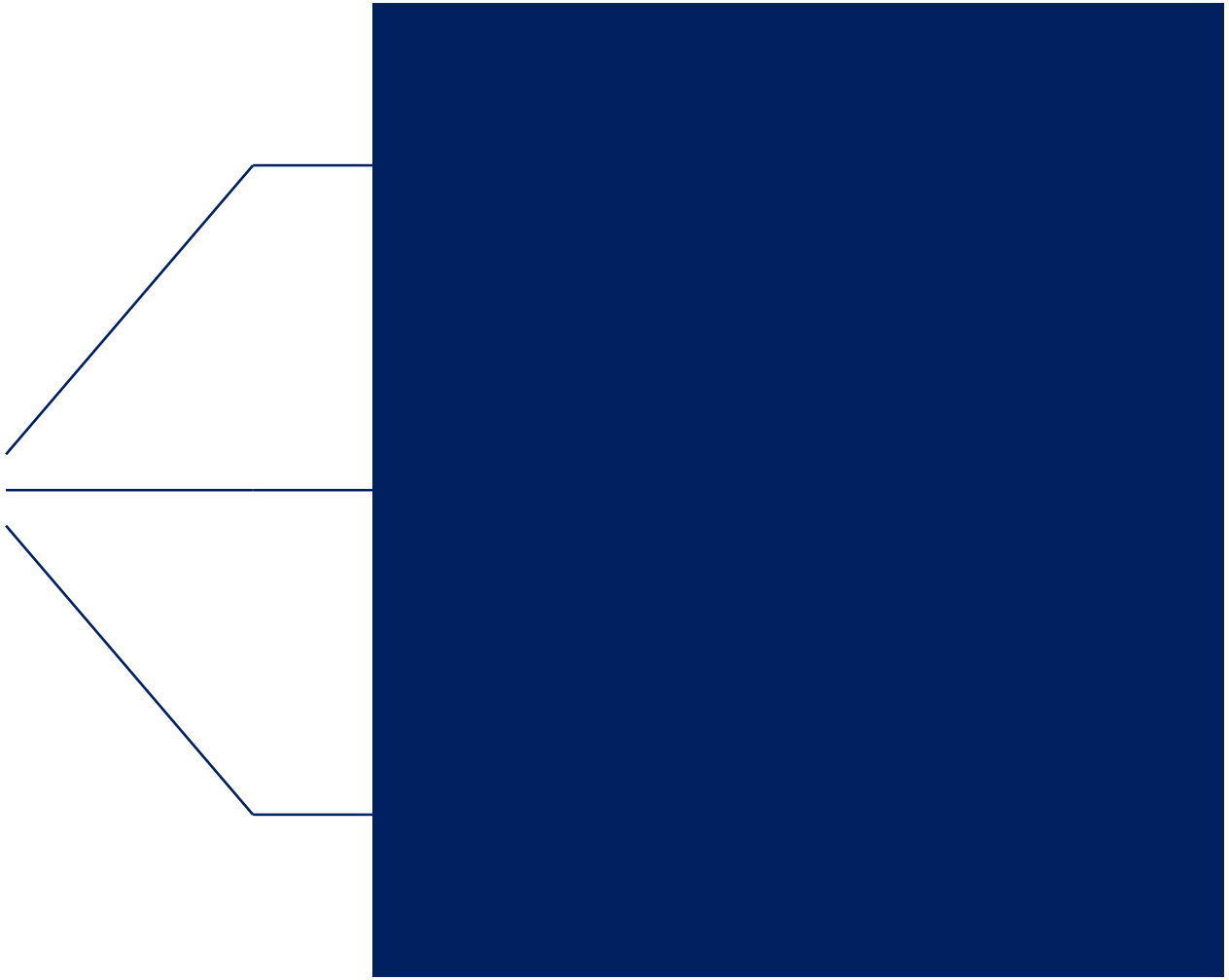
Well-posed learning problem

This framework also helps in deciding whether the problem is a right candidate to be solved using machine learning :



Well-posed learning problem

This framework also helps in deciding whether the problem is a right candidate to be solved using machine learning :

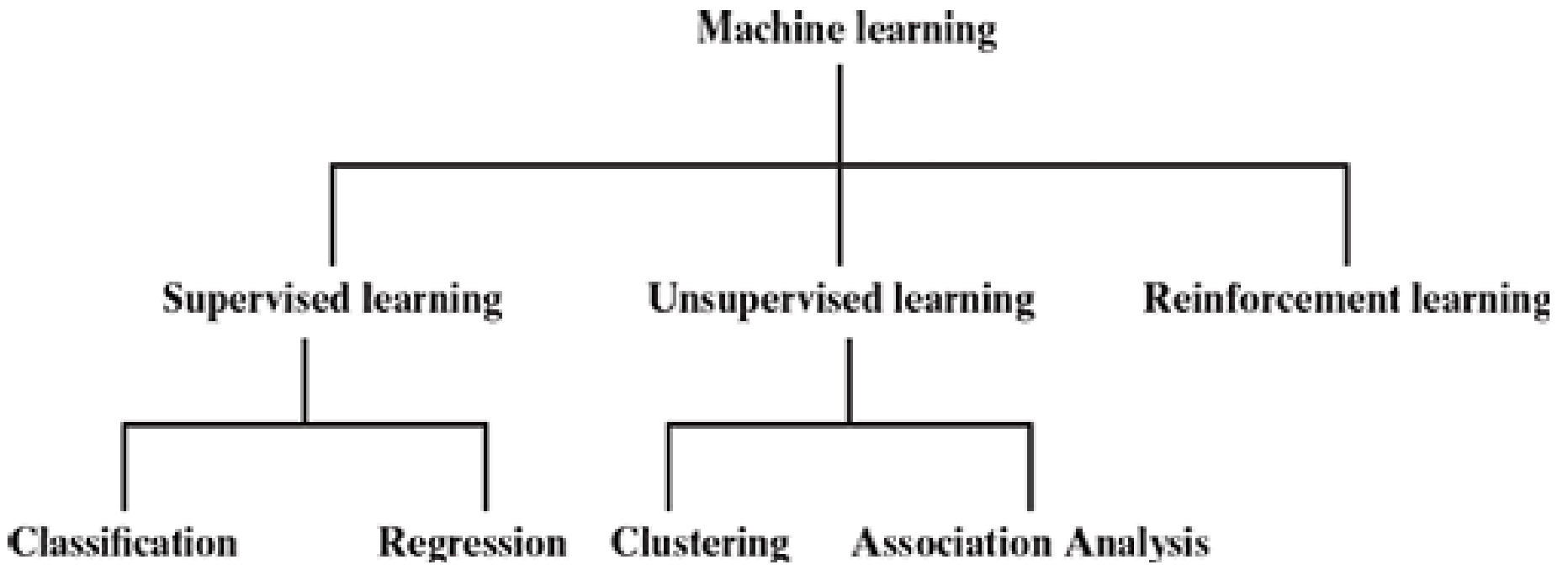


TYPES OF MACHINE LEARNING

RD_Spend	Administration	Marketing_Spend	State	Profit
16534920	13689780	47178410	New Delhi	19226183
16259770	15137759	44389853	Bangalore	19179206
15344151	10114555	40793454	Mumbai	19105039
14437241	11867185	38319962	New Delhi	18290199
14210734	9139177	36616842	Mumbai	16618794
13187690	9981471	36286136	New Delhi	15699112
13461546	14719887	12771682	Bangalore	15612251
13029813	14553006	32387668	Mumbai	15575260

Recency (months)	Frequency (times)	Monetary (c.c. blood)	Time (months)	whether he/she donated blood in March 2007
2	50	12500	98	1
0	13	3250	28	1
1	16	4000	35	1
2	20	5000	45	1
1	24	6000	77	0
4	4	1000	4	0
2	7	1750	14	1

TYPES OF MACHINE LEARNING



TYPES OF MACHINE LEARNING

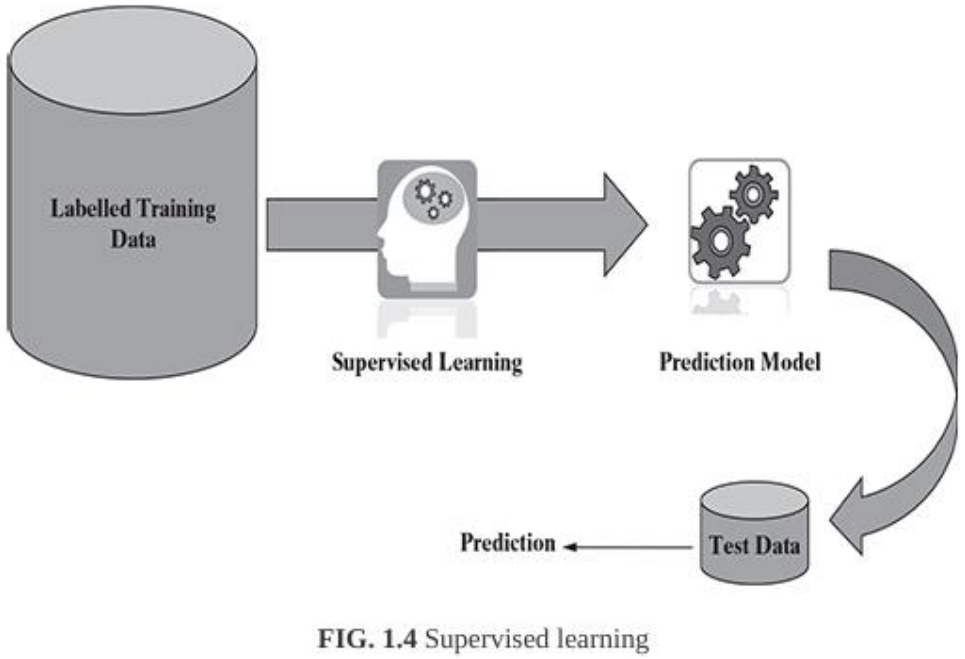
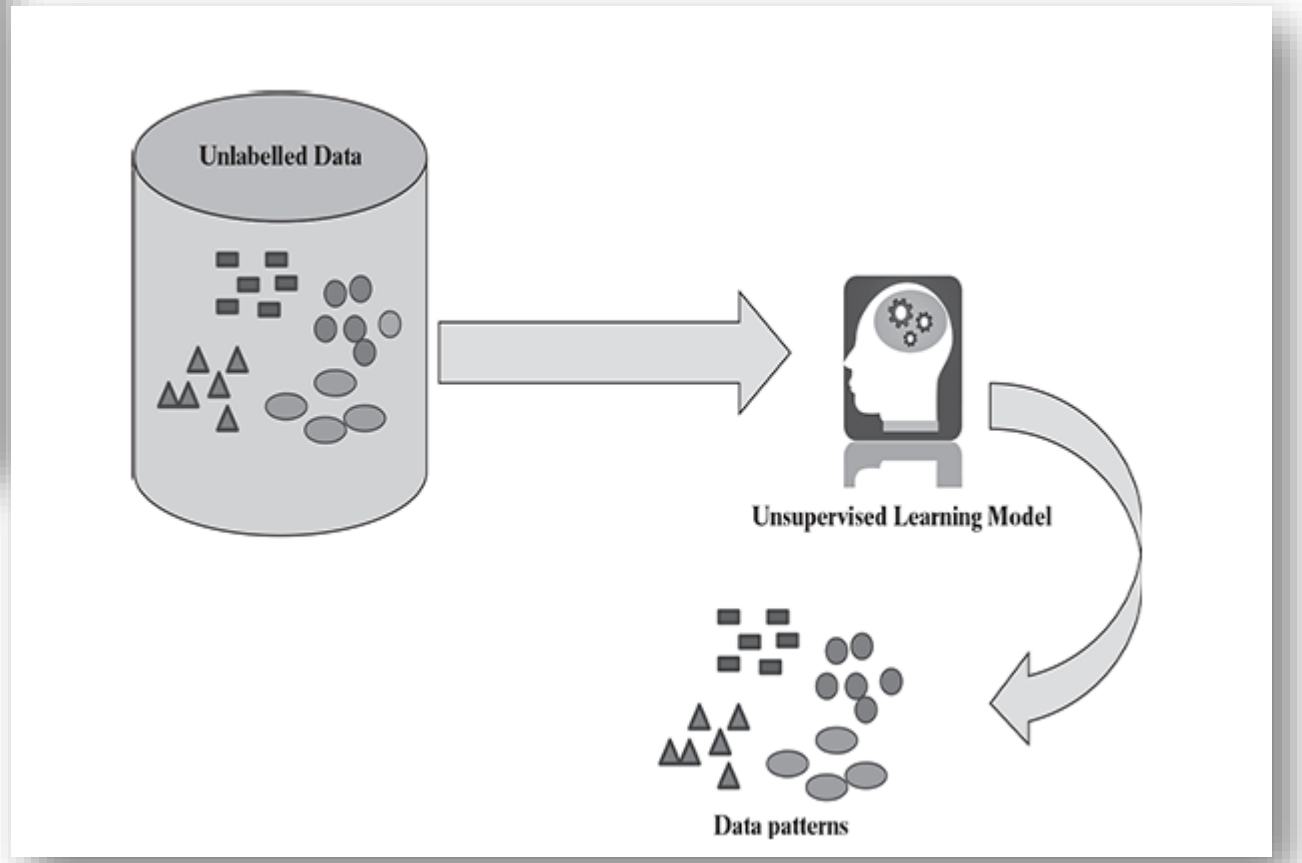
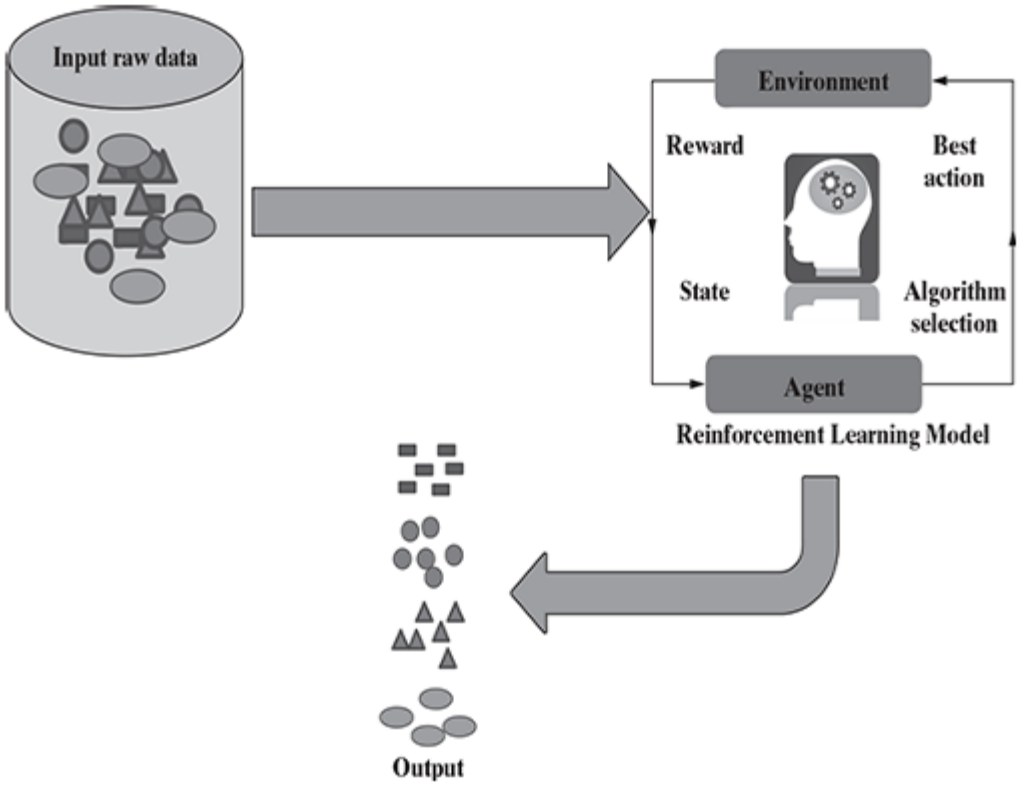


FIG. 1.4 Supervised learning



TYPES OF MACHINE LEARNING



Preparing to Model

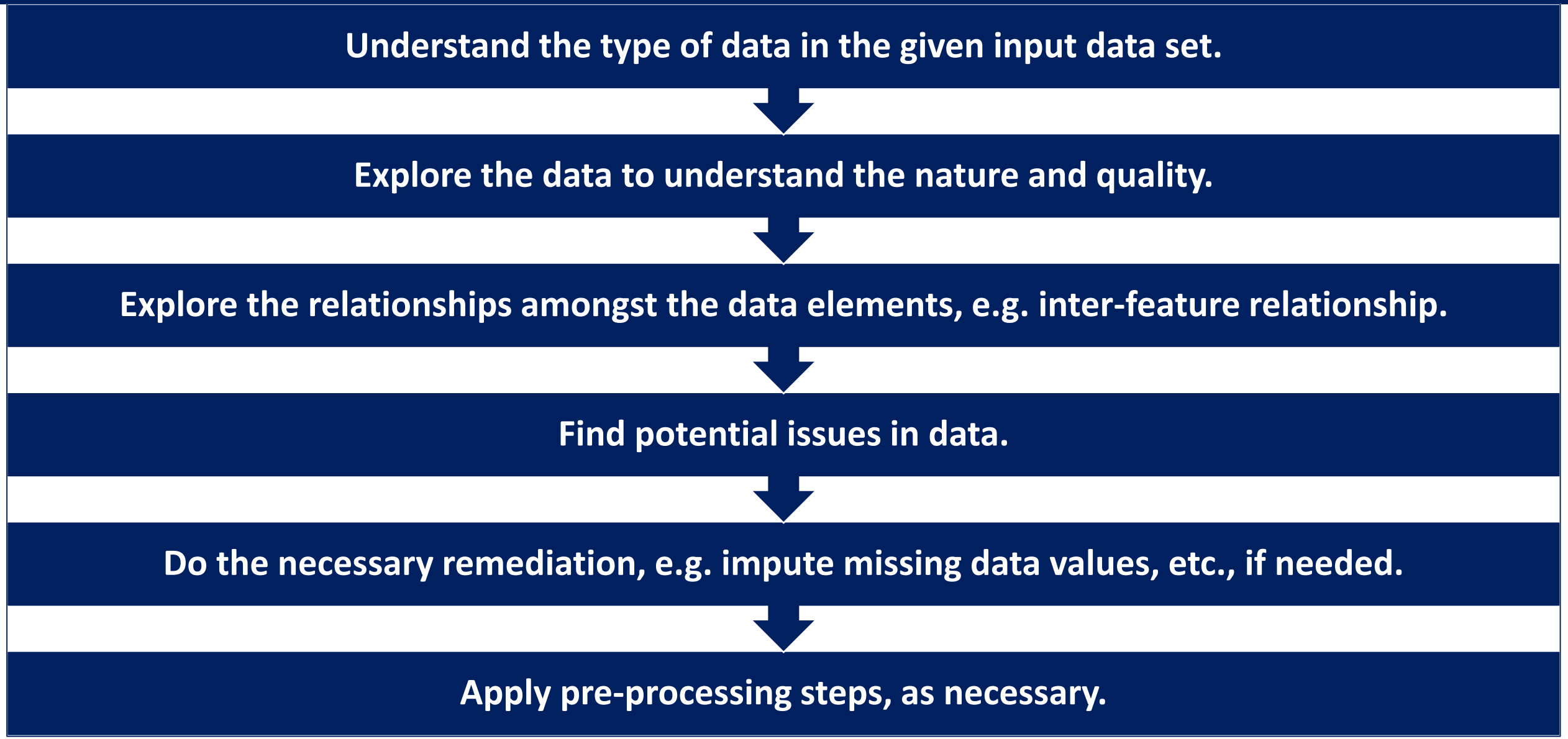
- **The first step in machine learning activity starts with data.**
- **In case of supervised learning, it is the labelled training data set followed by test data which is not labelled.**
- **In case of unsupervised learning, there is no question of labelled data but the task is to find patterns in the input data**

Preparing to Model

- **A thorough review and exploration of the data is needed to understand the type of the data, the quality of the data and relationship between the different data elements.**

- **Multiple pre-processing activities may need to be done on the input data before we can go ahead with core machine learning activities**

Preparing to Model



Preparing to Model

Once the data is prepared for modelling, then the learning tasks start off.

As a part of it, do the following activities:

- **The input data is first divided into parts – the training data and the test data (called holdout).**
- **This step is applicable for supervised learning only.**
- **Consider different models or learning algorithms for selection.**
- **Train the model based on the training data for supervised learning problem and apply to unknown data.**
- **Directly apply the chosen unsupervised model on the input data for unsupervised learning problem.**

Preparing to Model

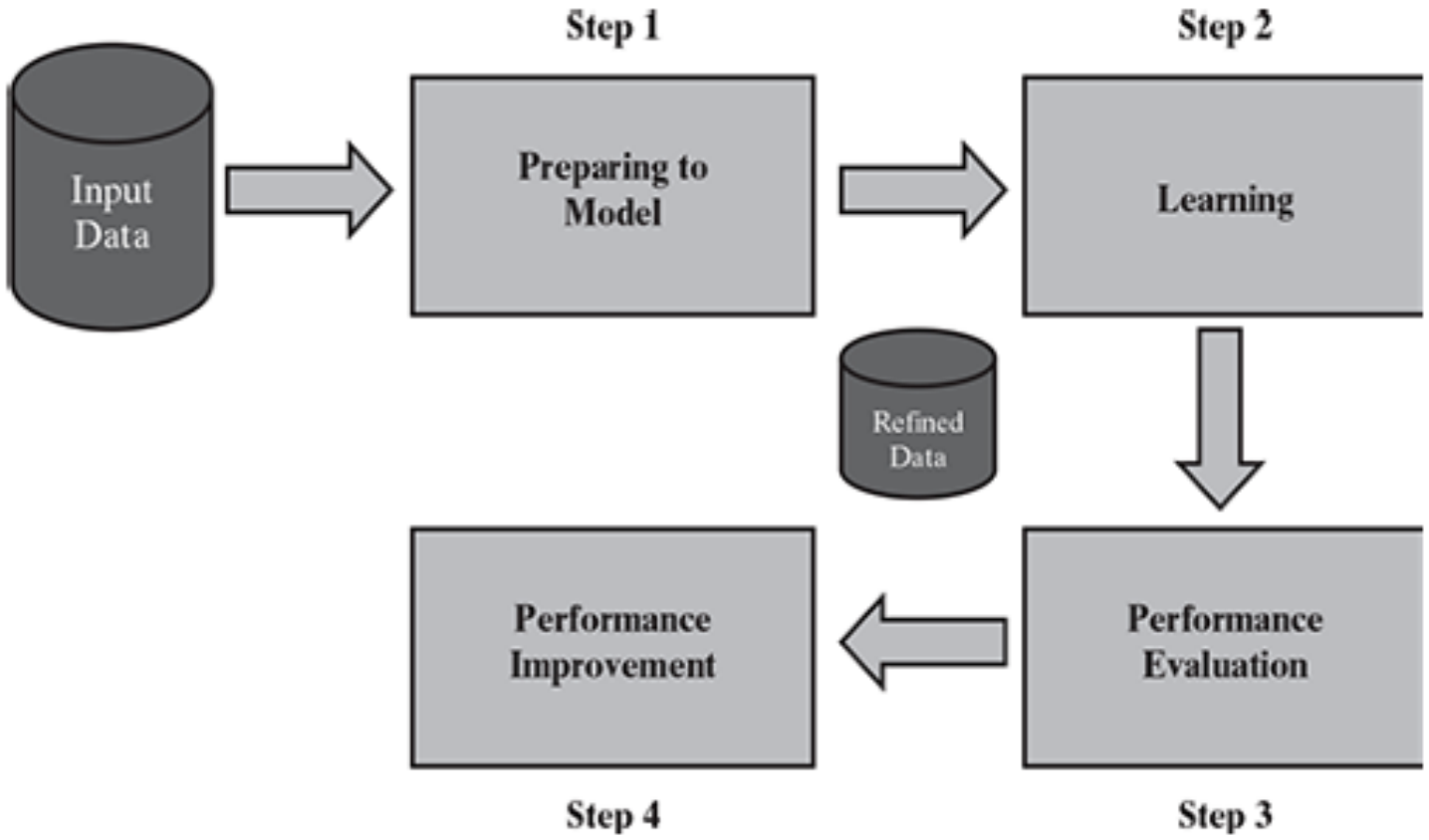


FIG. 2.1 Detailed process of machine learning

BASIC TYPES OF DATA IN MACHINE LEARNING

Student data set:

Roll Number	Name	Gender	Age
129/011	Mihir Karmarkar	M	14
129/012	Geeta Iyer	F	15
129/013	Chanda Bose	F	14
129/014	Sreenu Subramanian	M	14
129/015	Pallav Gupta	M	16
129/016	Gajanan Sharma	M	15

Student performance data set:

Roll Number	Maths	Science	Percentage
129/011	89	45	89.33%
129/012	89	47	90.67%
129/013	68	29	64.67%
129/014	83	38	80.67%
129/015	57	23	53.33%
129/016	78	35	75.33%

FIG. 2.2 Examples of data set

Roll Number	Name	Gender	Age
129/011	Mihir Karmarkar	M	14
129/012	Geeta Iyer	F	15

BASIC TYPES OF DATA IN MACHINE LEARNING

- **Qualitative data provides information about the quality of an object or information which cannot be measured.**
- **Qualitative data is also called categorical data.**
- **Qualitative data can be further subdivided into two types as follows:**
  - **1. Nominal data**
  - **2. Ordinal data**

BASIC TYPES OF DATA IN MACHINE LEARNING

- **1. Blood group: A, B, O, AB, etc.**
- **2. Nationality: Indian, American, British, etc.**
- **3. Gender: Male, Female, Other**

- **Nominal data is one which has no numeric value, but a named value.**
- **It is used for assigning named values to attributes.**
- **Nominal values cannot be quantified.**

BASIC TYPES OF DATA IN MACHINE LEARNING

- **1. Customer satisfaction: ‘Very Happy’, ‘Happy’, ‘Unhappy’, etc.**
- **2. Grades: A, B, C, etc.**
- **3. Hardness of Metal: ‘Very Hard’, ‘Hard’, ‘Soft’, etc**

- **Like nominal data, basic counting is possible for ordinal data.**
- **Hence, the mode can be identified.**
- **Since ordering is possible in case of ordinal data, median, and quartiles can be identified in addition.**
- **Mean can still not be calculated.**

BASIC TYPES OF DATA IN MACHINE LEARNING

- Quantitative data relates to information about the quantity of an object – hence it can be measured.
- Quantitative data is also termed as numeric data.
- There are two types of quantitative data:
  - 1. Interval data
  - 2. Ratio data
- Interval data is numeric data for which not only the order is known, but the exact difference between values is also known.
- An ideal example of interval data is Celsius temperature.

BASIC TYPES OF DATA IN MACHINE LEARNING

- **Ratio data represents numeric data for which exact value can be measured.**
- **Absolute zero is available for ratio data.**
- **Also, these variables can be added, subtracted, multiplied, or divided.**
- **The central tendency can be measured by mean, median, or mode and methods of dispersion such as standard deviation.**

BASIC TYPES OF DATA IN MACHINE LEARNING

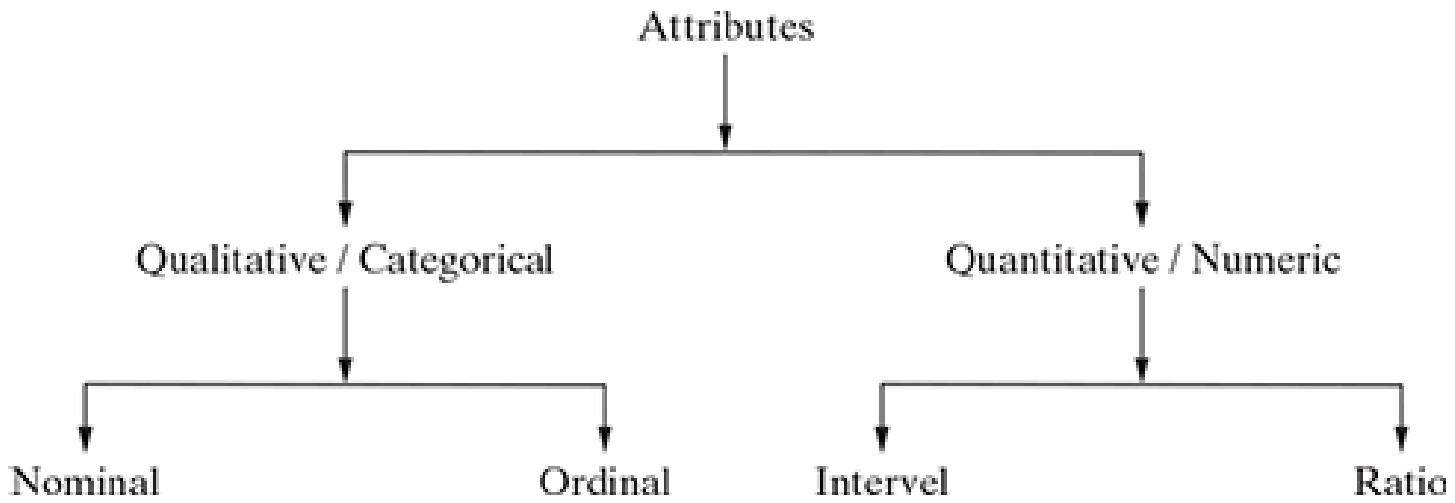


FIG. 2.4 Types of data