Introduction to Machine Learning

Dr. Singara Singh Kasana

Associate Professor

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology

Patiala, Punjab

Machine Learning

Machine learning is a field of artificial intelligence which enables a machine to learn automatically from the data and improve its performance with experience.
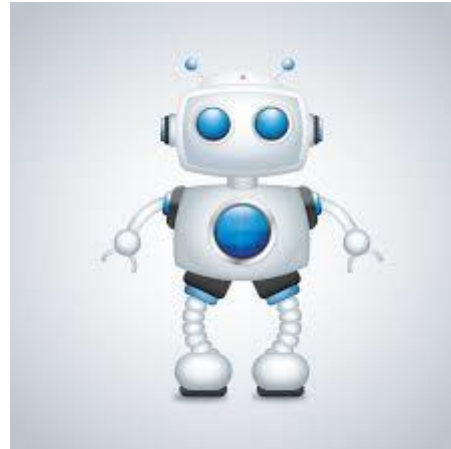
Objective of Machine Learning

To build a model using some statistical analysis of the data to predict an output which is capable to improve its performance as more data becomes available.
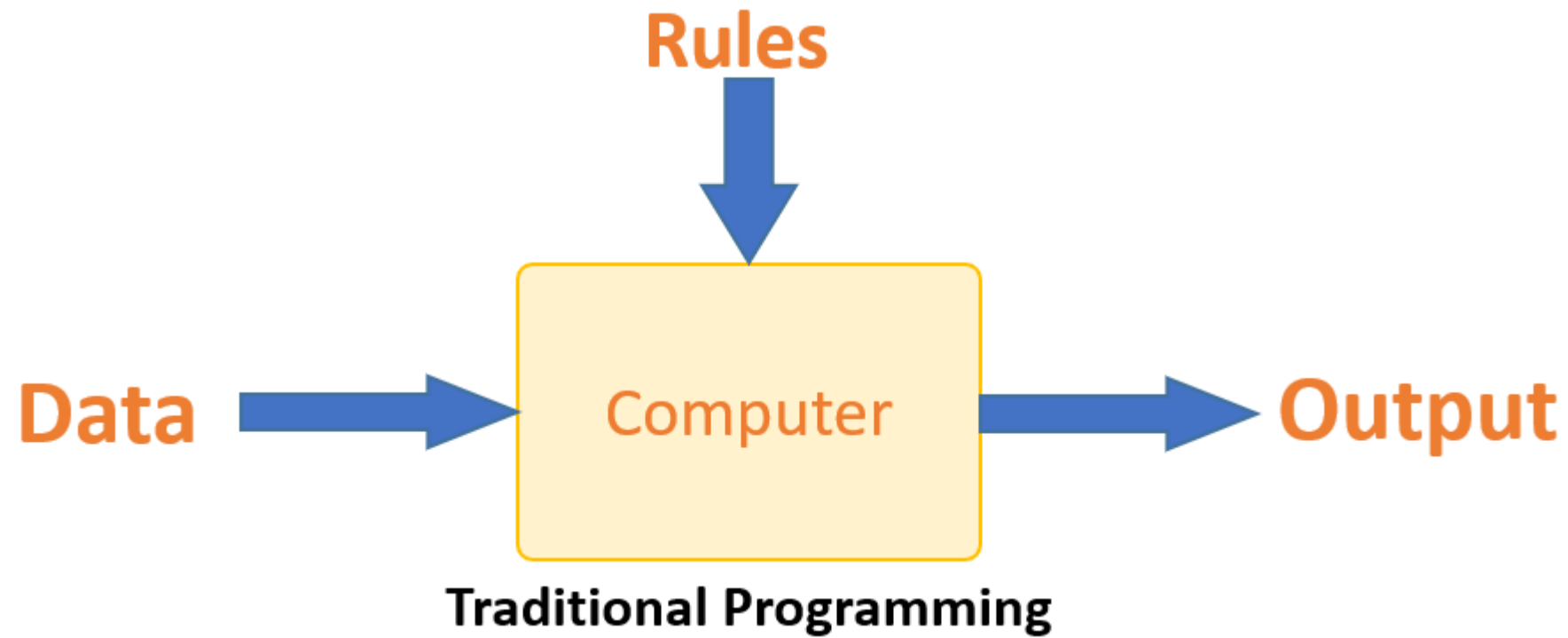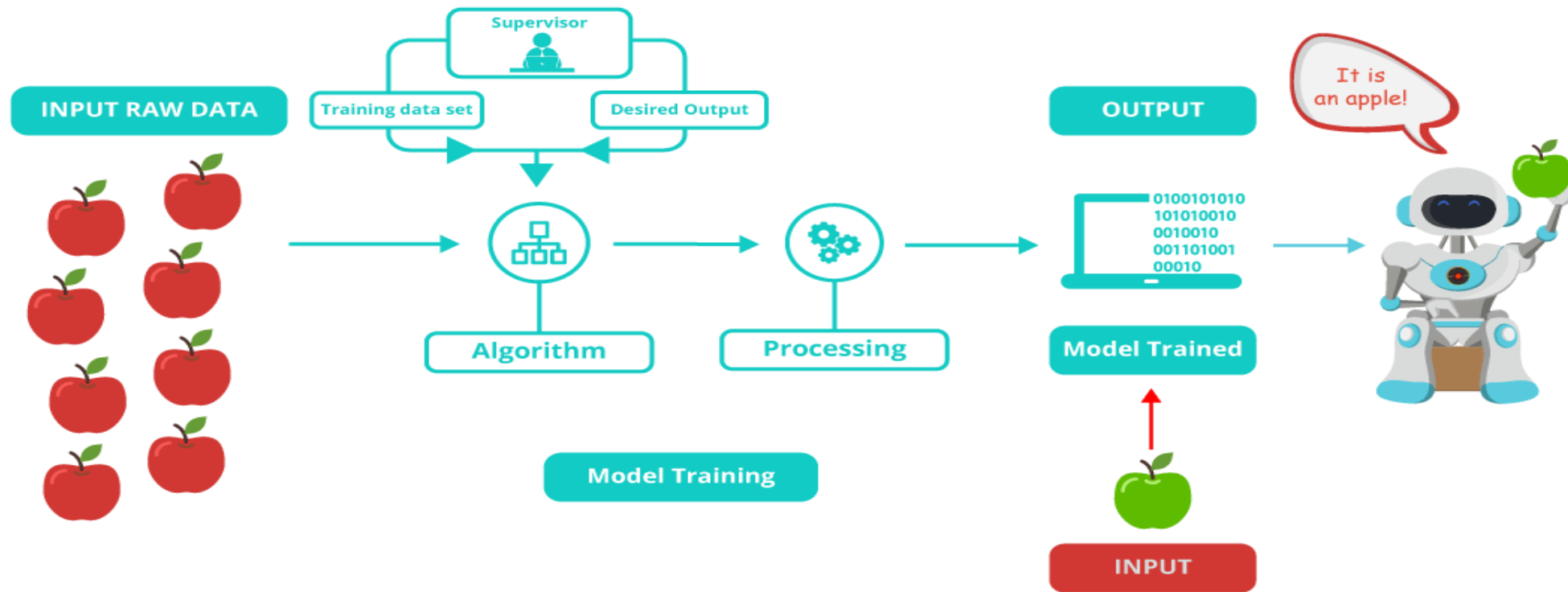
Learn from Experience

Learn from Data

Computer Follows Instructions

Traditional Programming

ML gives the computer that which makes it more similar to humans: *The ability to learn*.

Arthur Samuel Definition of ML

"Field of study that gives computers the ability to learn without being

explicitly programmed".

Tom Mitchell Definition of ML

"A computer program is said to learn from experience E with respect to some task *T* and performance measure P, if its performance on task T, as measured by P, improves with experience E".

Example

Suppose in first attempt, the child performance measure (P) is 1/4, which means that the child found 1 out of 4 correct shape holes.

After repeating this task (T) 100 times or more, the baby may be able to figured out which shape goes into which hole.

So his experience (E) increases, as the number of attempts at this toy increases. The performance (P) also increases, which results in higher accuracy

Task(T) is to find the appropriate hole for a shape.
Experience(E) is Number of attempts
Performance (P) is the **Accuracy**

A handwritten digits recognition System

Task T: To recognize the handwritten digits

Performance measure P: Prediction Accuracy

Training experience E: Data set of sufficient size containing images of handwritten digits along

with their labels.

In general, to have a well-defined learning problem, we must identity these three features:

➢ class of tasks

➢ measure of performance to be improved

➢ source of experience.

A checkers learning problem:

Task T: playing checkers

Performance measure P: percent of games won against opponents
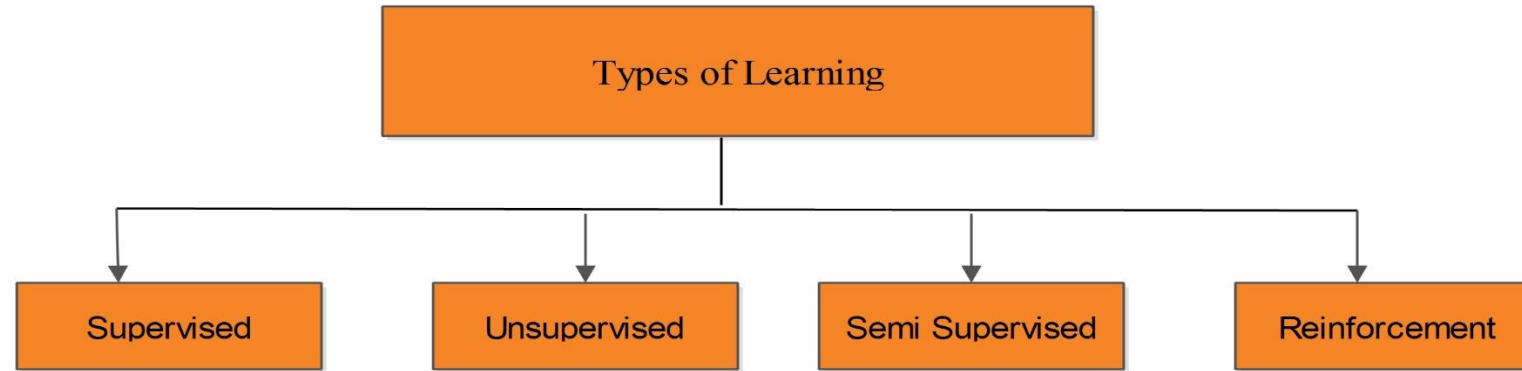
Training experience E: playing practice games against itself

A handwriting recognition learning problem:

Task T: ??

Performance measure P: ??

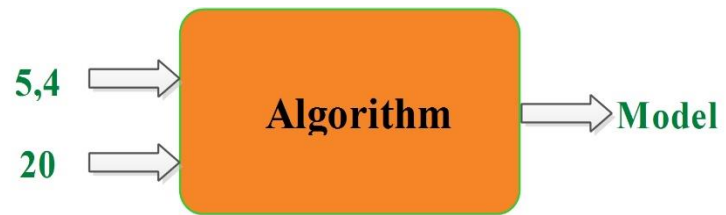**Training experience E**: ??

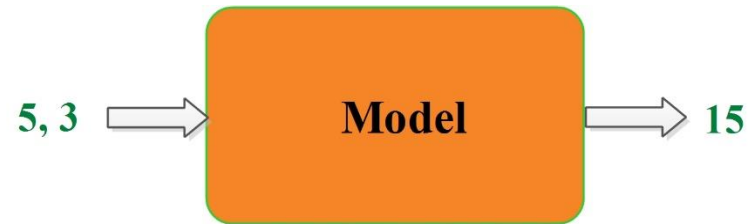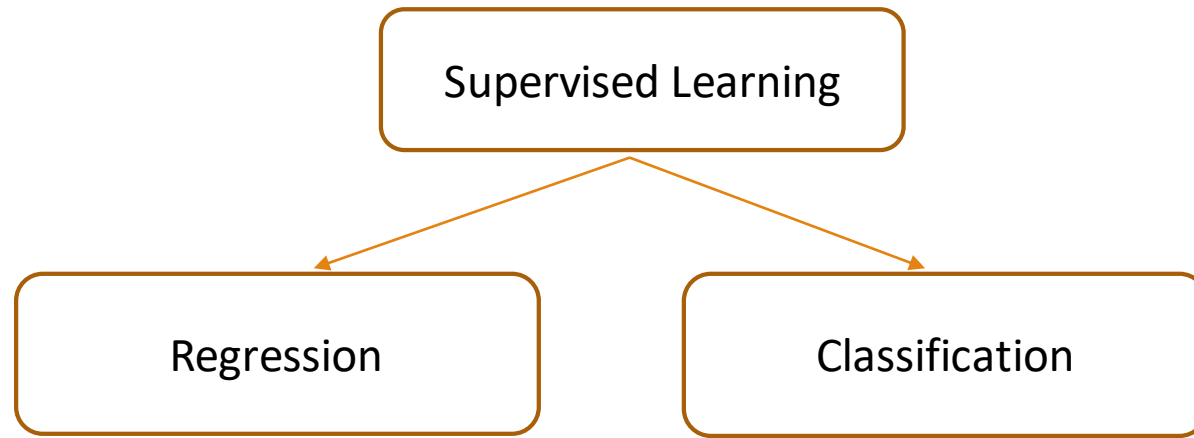# Supervised Learning

Figure 1 (a)

Figure 1 (b)

Supervised Learning

Learning using Training Data

Prediction with Testing Data

```
                    ┌─────────────────────────┐
                    │   Supervised Learning   │
                    └─────────────────────────┘
                          ╱           ╲
                        ╱               ╲
          ┌─────────────────┐   ┌─────────────────┐
          │   Regression    │   │  Classification │
          └─────────────────┘   └─────────────────┘
```
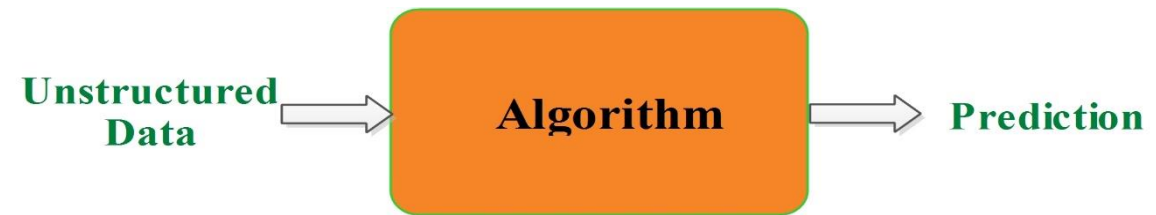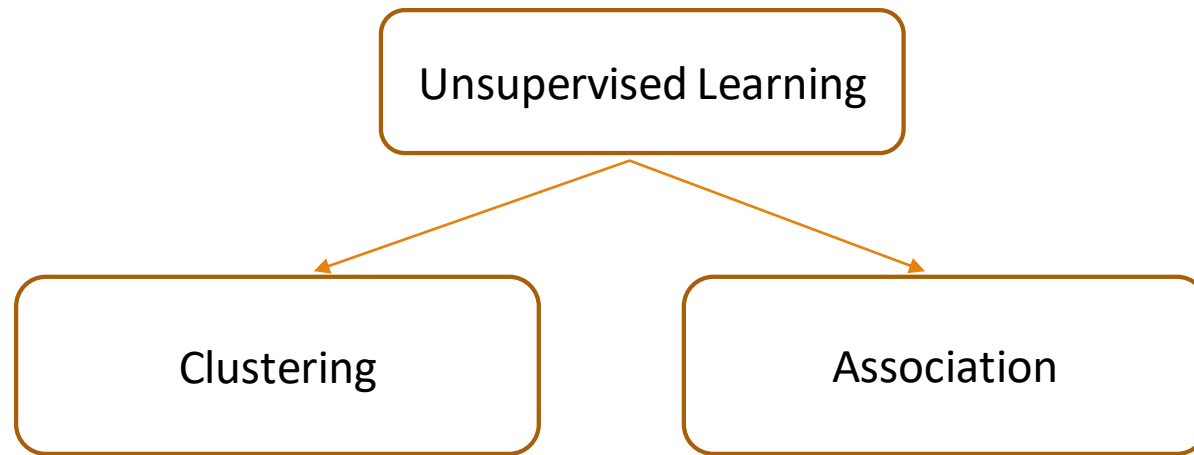
Regression: This is a type of learning in which need to predict
the continuous output feature.

Classification: This is a type of learning using which we can
predict the categorical output feature.

Unsupervised Learning

Unstructured Data → **Algorithm** → Prediction

```
                    ┌─────────────────────────┐
                    │                         │
                    │  Unsupervised Learning  │
                    │                         │
                    └─────────────────────────┘
                          ╱           ╲
                         ╱             ╲
              ┌──────────────┐    ┌──────────────┐
              │              │    │              │
              │  Clustering  │    │  Association │
              │              │    │              │
              └──────────────┘    └──────────────┘
```

Clustering: It is applied to group the data based on different patterns, found by the machine learning model. K Means is one example of clustering.

Association: This technique is a rule based technique which finds out some very useful relations between parameters of a large data set. For example, shopping stores use algorithms based on this technique to find out relationship between sale of one product w.r.t to sale of others product. Once trained well, such models can be used to increase their sales by planning different offers.

Semi Supervised Learning

As the name suggests, its working lies between Supervised and Unsupervised techniques. We use this technique, when we are dealing with a training data which is a little bit labelled and rest large portion of it is unlabeled. We can use unsupervised technique to predict labels and then feed these labels to supervised techniques. This technique is mostly applicable in case of image data-sets where usually all images are not labelled.
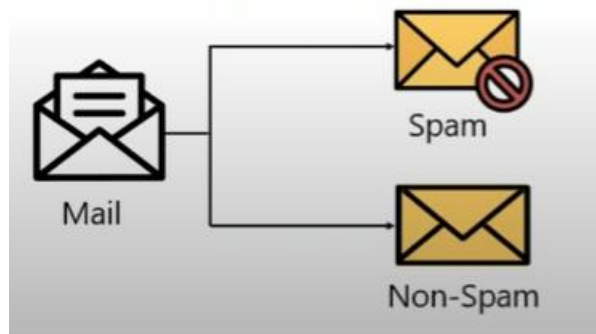
Reinforcement Learning

A General Machine Learning System
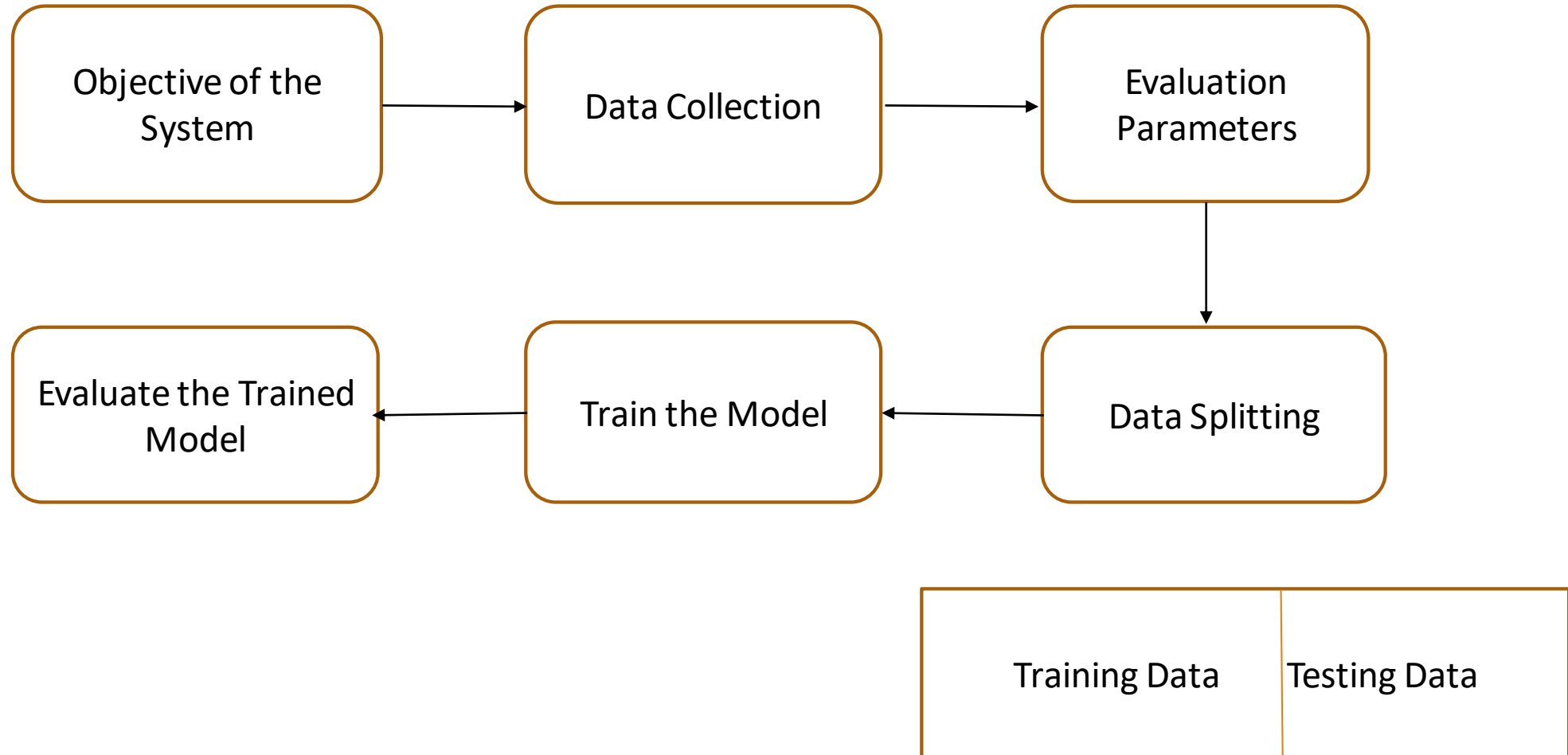
Types of Classification

➢ A classification problem with two classes is called binary classification while a classification problem with more than two classes is called multiclass classification.

IRIS Dataset

Sepal Length	Sepal Width	Petal Length	Petal Width	Species
5.1	3.5	1.4	0.2	0
4.7	3.2	1.3	0.2	0
6.6	3.0	4.4	1.4	1
5.7	3.0	4.2	1.2	1
5.9	3.0	5.1	1.8	2

Iris Setosa

Iris Versicolor

Iris Virginica

Thanks