
# Machine Learning

DATA AND
ARTIFICIAL INTELLIGENCE

# Supervised Learning: Regression and Classification

# Learning Objectives

By the end of this lesson, you will be able to:

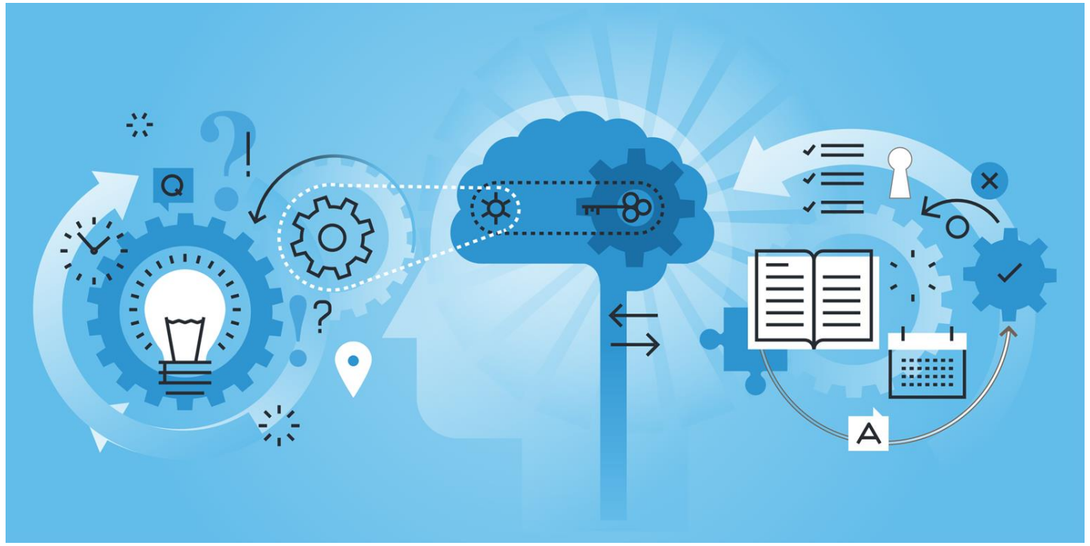
- Describe supervised learning and its types
- Summarize the flow of supervised learning
- Analyze the supervised learning use cases
- List the various types of classification algorithms

DATA AND
ARTIFICIAL INTELLIGENCE

# Supervised Learning: Overview

# Supervised Learning

Supervised learning is used to train models using labeled training data. It provides the ability to predict the output of future or unseen data.

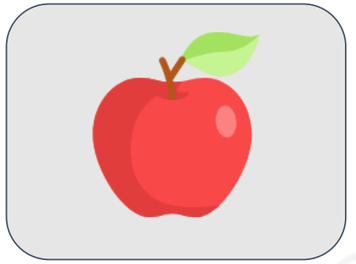


The goal of a supervised learning algorithm is to discover a mapping function that will translate the input variable (x) to the output variable (y).

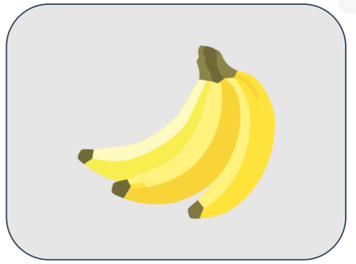
# Supervised Learning

Supervised learning algorithm

If the shape of the object is rounded having a **red** color, then it will be labeled as an Apple



If the shape of the object is a curving cylinder having the color **green-yellow** then it will be labeled a banana



# Supervised Learning: Examples

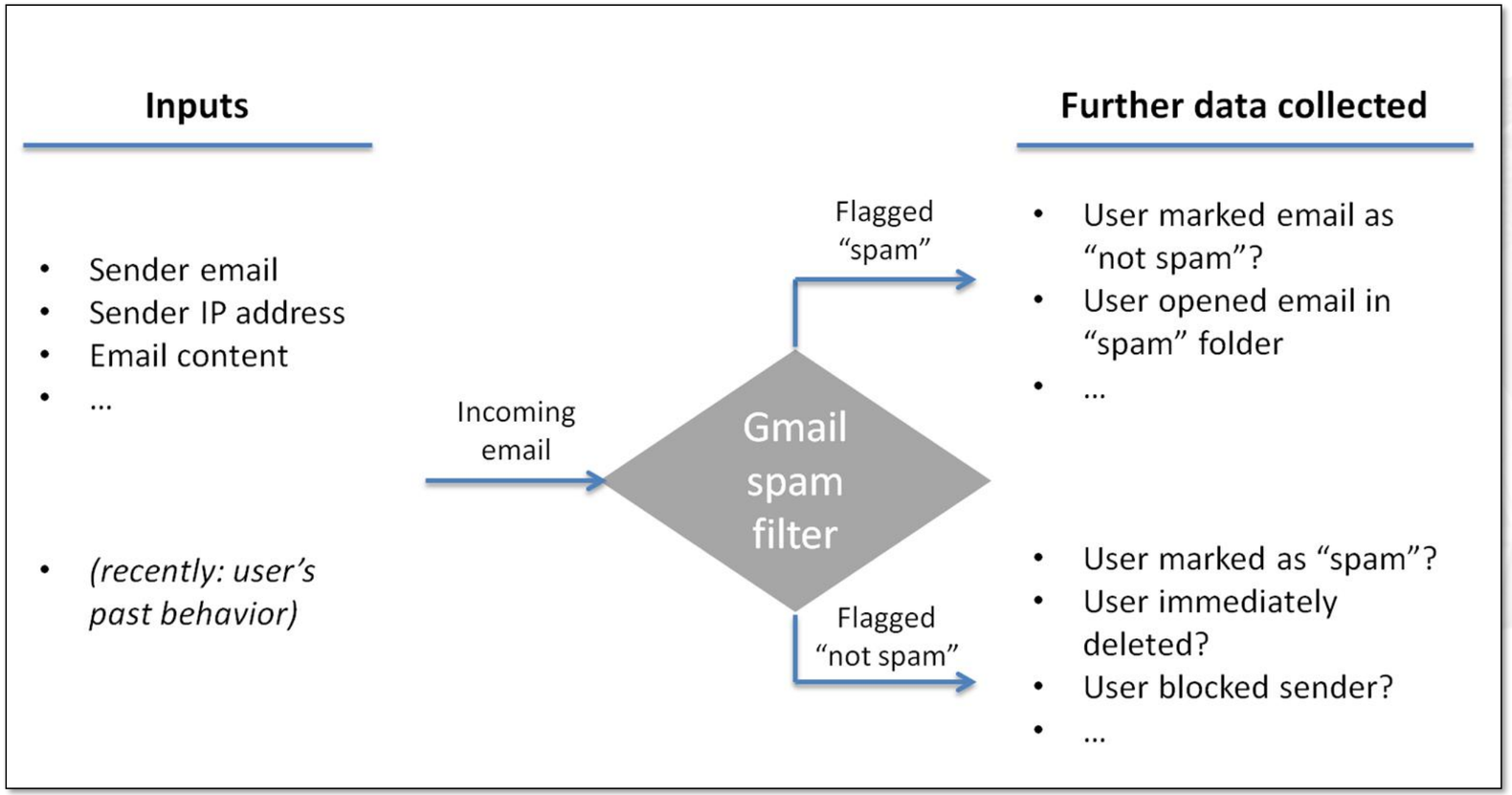
## Example 1: Weather Apps

The predictions made by weather apps at a given time are based on prior knowledge and analysis of weather over a period for a particular place.

# Supervised Learning: Examples

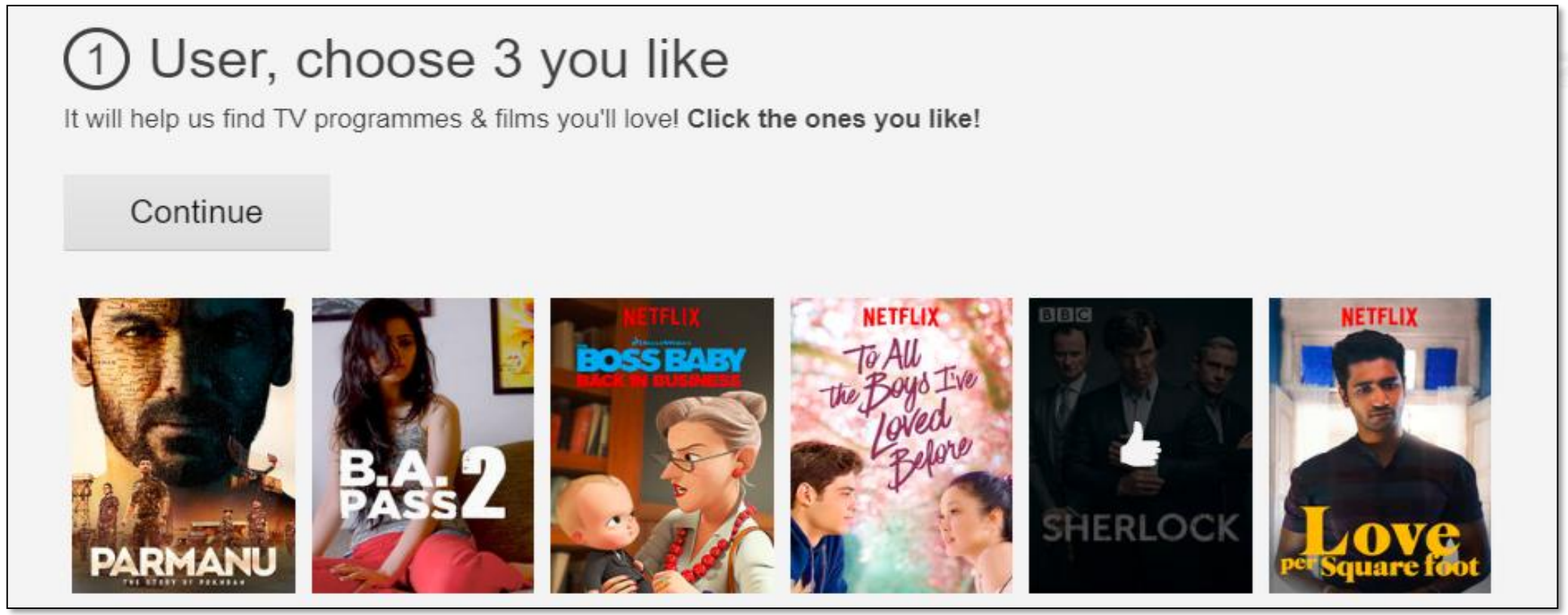
**Example 2: Gmail filters**

It helps to filter a new email into an inbox (normal) or junk folder (Spam) based on past information of spam.



# Netflix: Real-Life Scenario

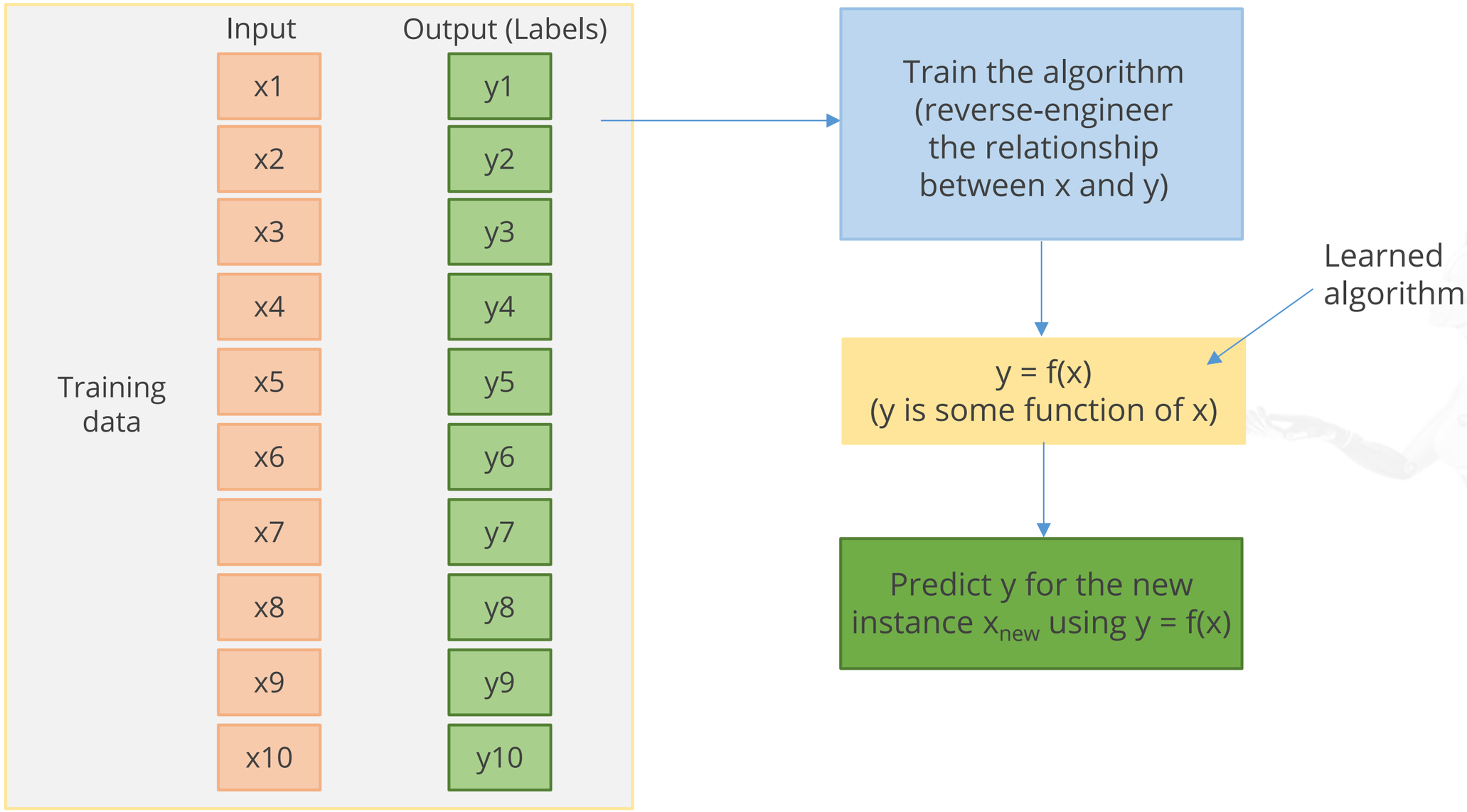
Ever wondered how Netflix makes recommendations?

# Netflix: Real-Life Scenario

Netflix uses **supervised learning** algorithms to recommend shows for the users based on the viewing history and ratings by similar classes of users.

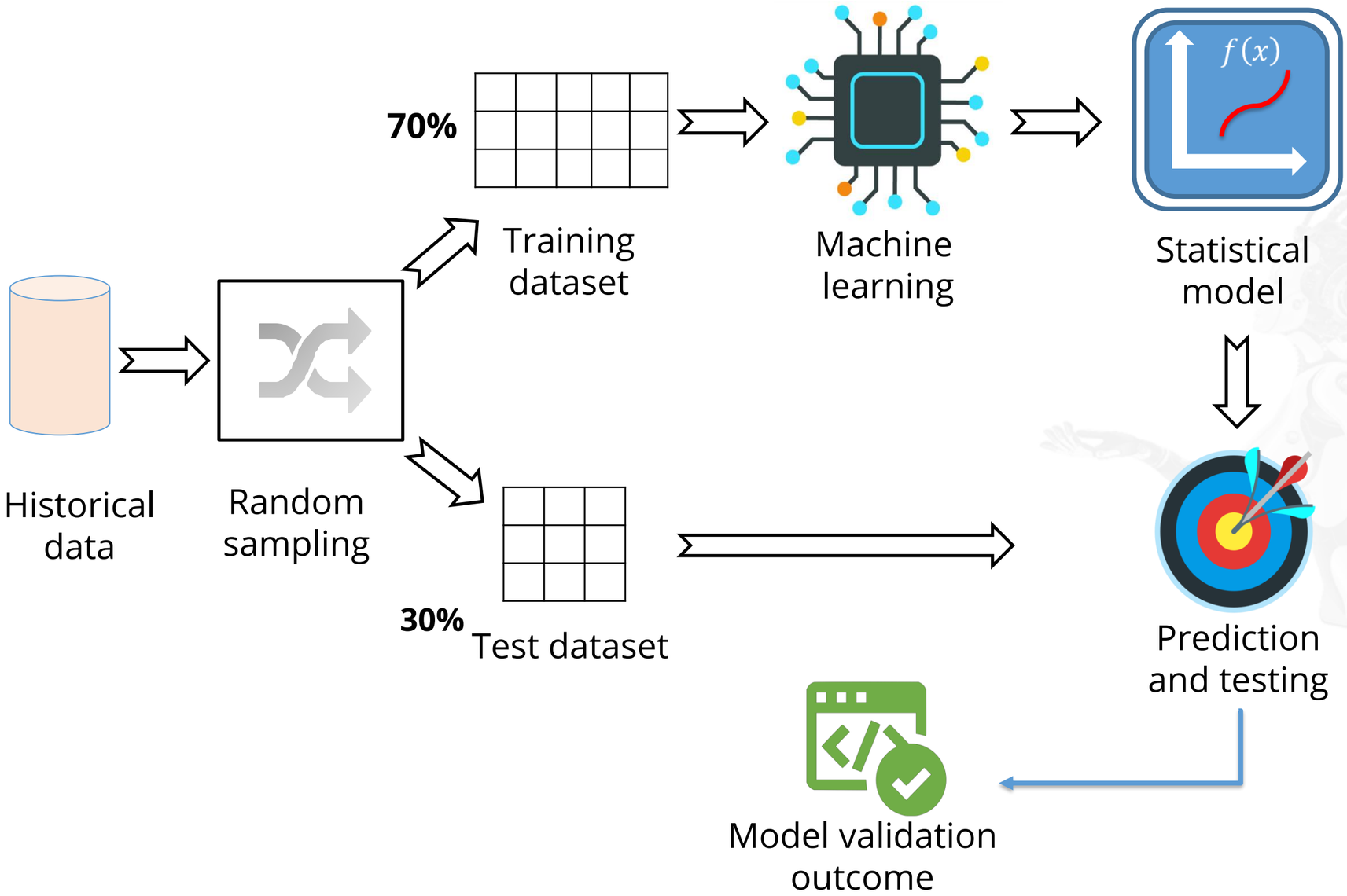
# Understanding the Algorithm



# Supervised Learning: Process Flow

**Training and Testing**

**Prediction**

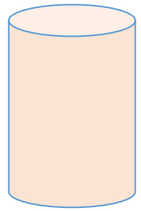


# Supervised Learning: Process Flow

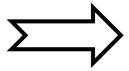
Training and testing

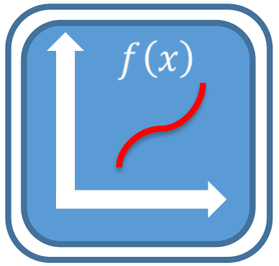
**Prediction**

Use the learned algorithm y = f(x) to predict production data

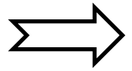


New data






Model

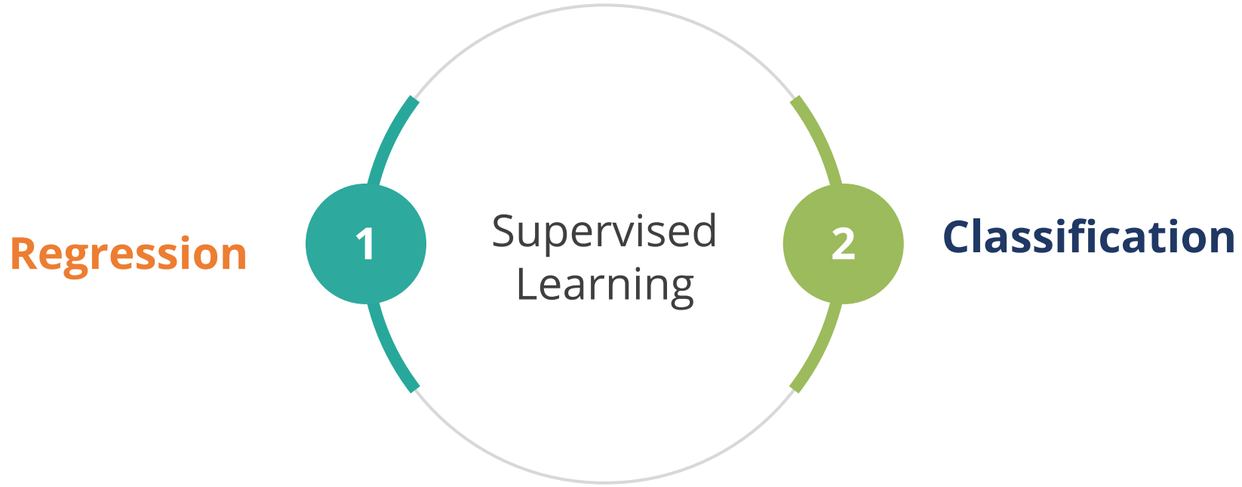


Prediction outcome

Algorithm prediction can be improved by more training data, capacity, or algorithm redesign.

# Types of Supervised Learning

# Types of Supervised Learning




In supervised learning, an algorithm is selected based on the target variable.

DATA AND
ARTIFICIAL INTELLIGENCE

# Regression Algorithms

# Regression

Regression is a statistical model used to predict the relationship between independent and dependent variables denoted by x and y respectively

Examine 2 factors

1

**How closely are x and y related?**

Linear regression gives a number between -1 and 1 indicating the strength of correlation between the two variables

**0**: no correlation
**1**: positively correlated
**-1**: negatively correlated

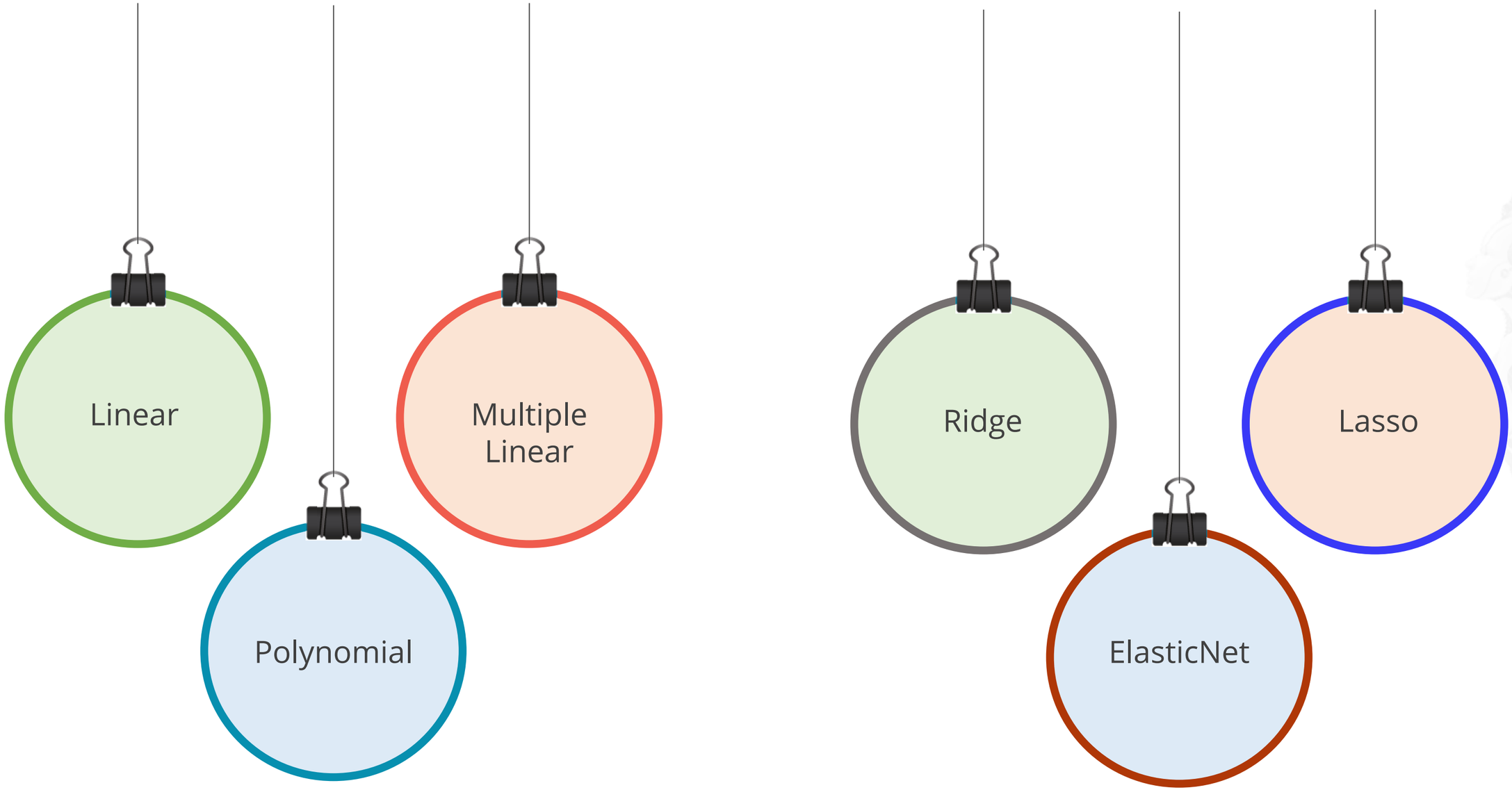
2

**Prediction**

When the relationship between x and y is known, use this to predict future values of y for a value of x. This is done by fitting a regression line and represented by a linear equation

$$y = a * x + b$$

# Types of Regression Algorithms



# Regression: Use Case

Predicting profit based on expenditures of the company

1 R&D Spend

2 Administration

3 Marketing Spend

4 State

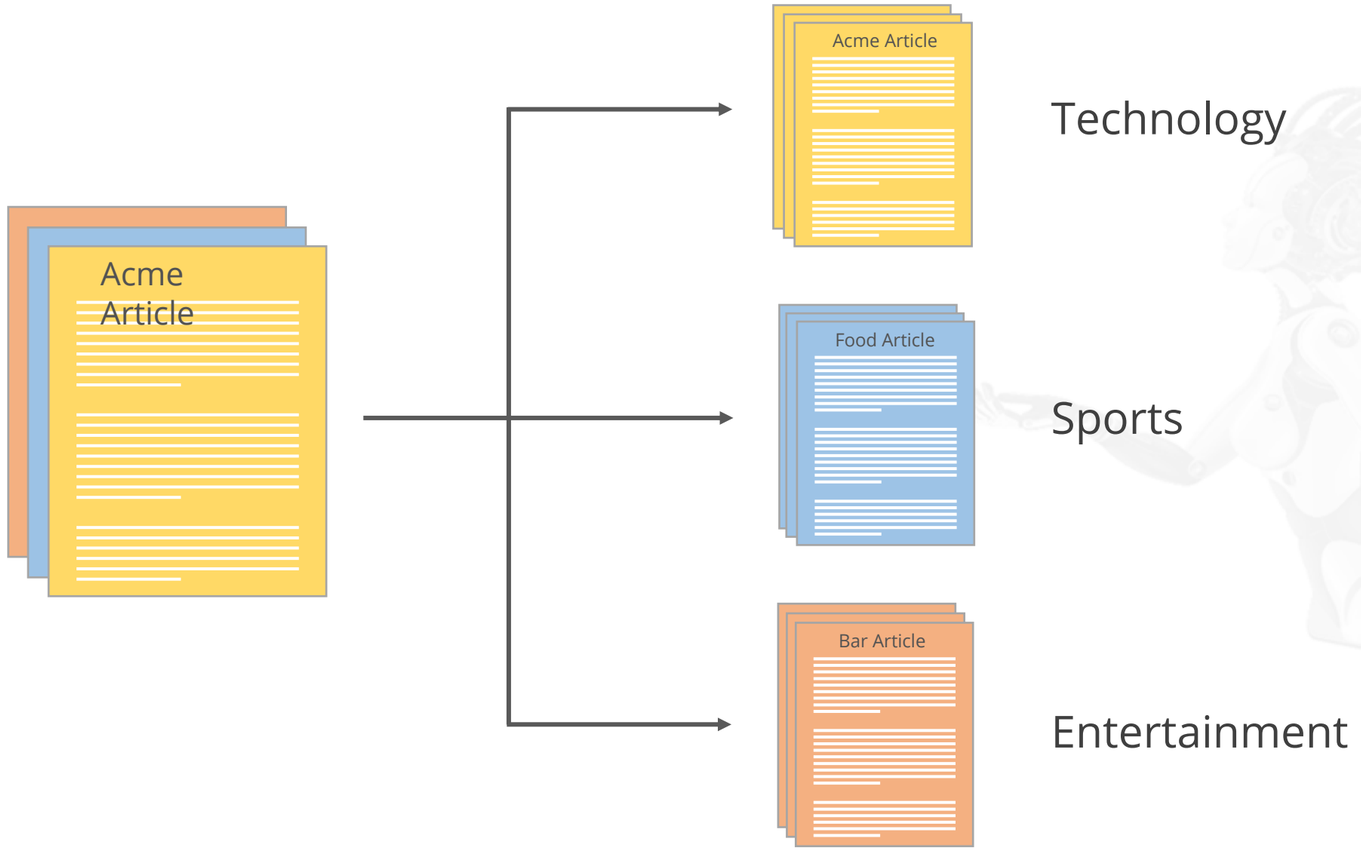
5 Profit

Profit Estimation

# Classification Algorithms

# Classification

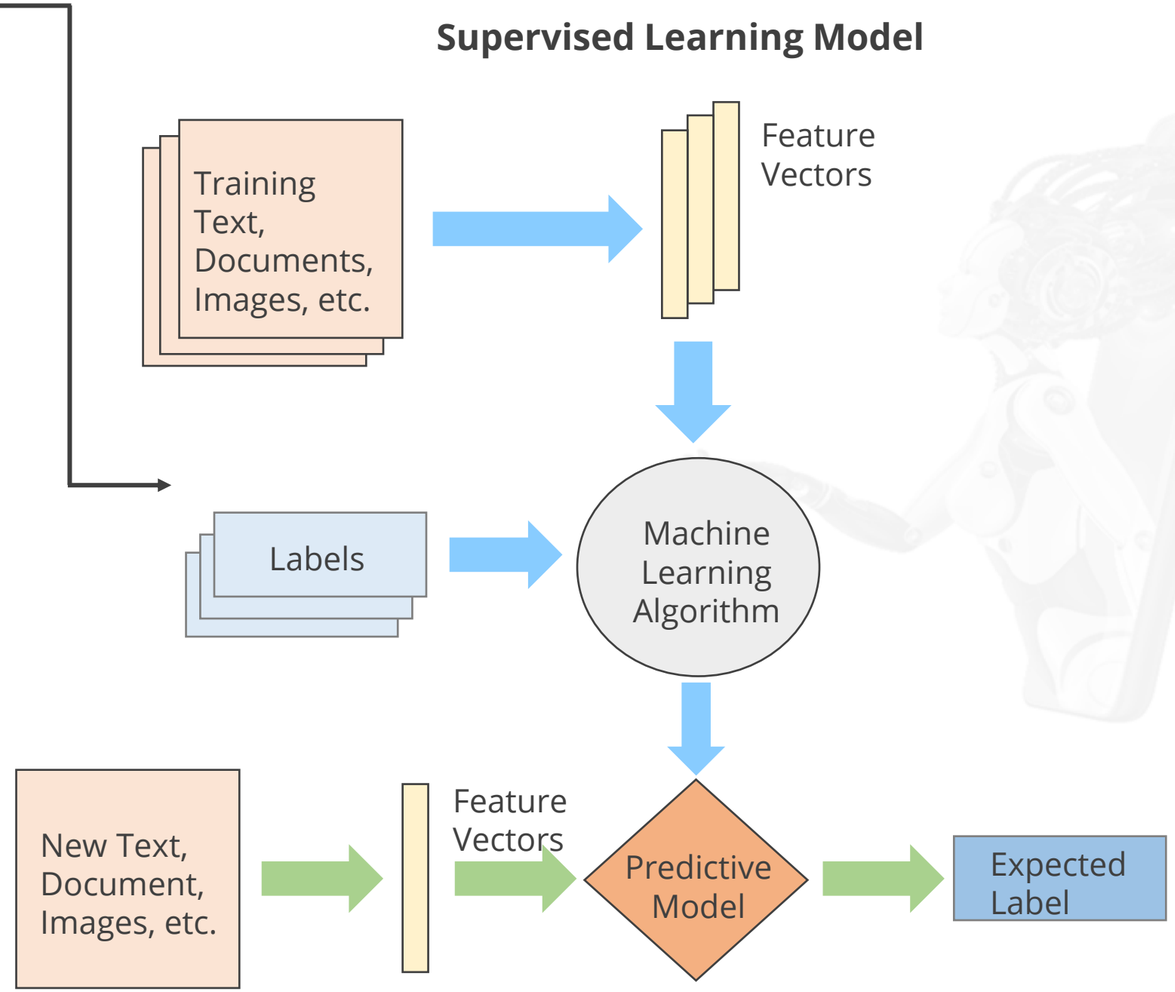
- Classification is a technique used for determining which class the dependent variable belongs to based on one or more independent variables.
- It is mainly used for predicting discrete values.



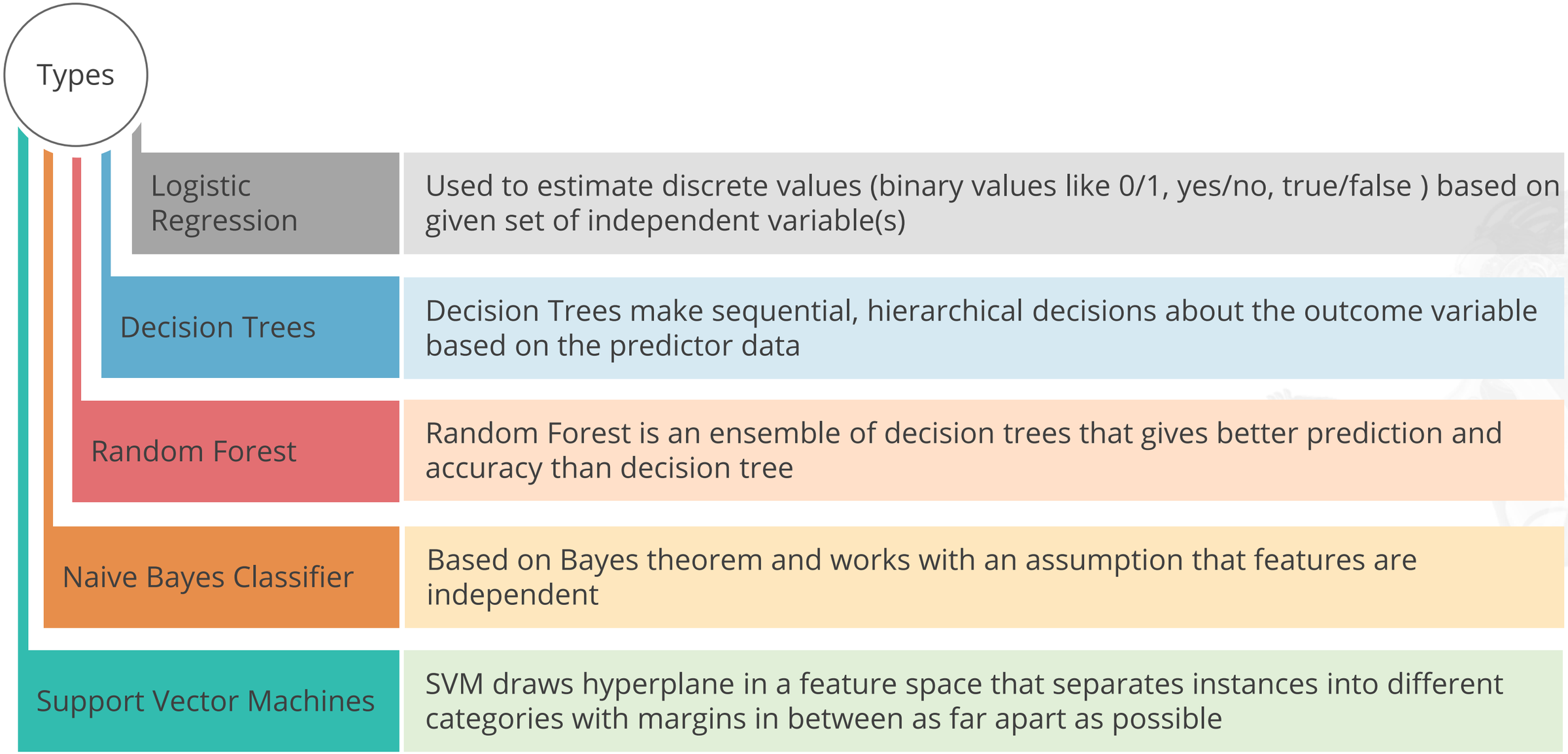
# Classification: A Supervised Learning Algorithm

Classification is a supervised learning algorithm, and the training data contains labels.

| Record ID | Age | Spectacle Prescription | Astigmatic | Tear production Rate | Class Label Lenses |
|---|---|---|---|---|---|
| 1 | Young | Myope | No | Reduced | Noncontact |
| 2 | Young | Myope | No | Normal | Soft contact |
| 3 | Young | Myope | Yes | Reduced | Noncontact |
| 4 | Young | Myope | Yes | Normal | Hard contact |
| 5 | Young | Hypermetrope | No | Reduced | Noncontact |
| 6 | Young | Hypermetrope | No | Normal | Soft contact |
| 7 | Young | Hypermetrope | Yes | Reduced | Noncontact |
| 8 | Young | Hypermetrope | Yes | Normal | Hard contact |
| 9 | Pre-presbyopic | Myope | No | Reduced | Noncontact |
| 10 | Pre-presbyopic | Myope | No | Normal | Soft contact |
| 11 | Pre-presbyopic | Myope | Yes | Reduced | Noncontact |
| 12 | Pre-presbyopic | Myope | Yes | Normal | Hard contact |
| 13 | Pre-presbyopic | Hypermetrope | No | Reduced | Noncontact |
| 14 | Pre-presbyopic | Hypermetrope | No | Normal | Soft contact |
| 15 | Pre-presbyopic | Hypermetrope | Yes | Reduced | Noncontact |
| 16 | Pre-presbyopic | Hypermetrope | Yes | Normal | Noncontact |
| 17 | Presbyopic | Myope | No | Reduced | Noncontact |
| 18 | Presbyopic | Myope | No | Normal | Noncontact |
| 19 | Presbyopic | Myope | Yes | Reduced | Noncontact |
| 20 | Presbyopic | Myope | Yes | Normal | Hard contact |



# Types of Classification Algorithms

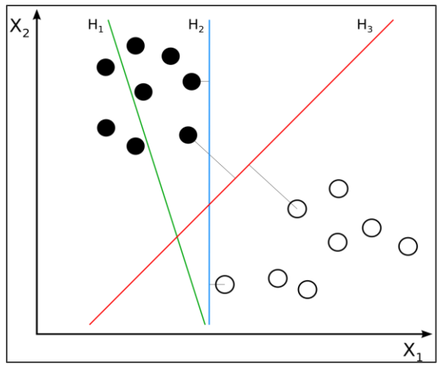


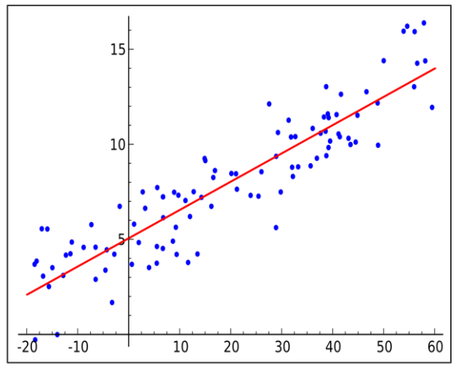
# Classification: Use Case

Sentiment analysis, fraud detection, spam filtering, and product categorization are examples of classification.

# Difference between Regression and Classification Outputs

# Classification and Regression: Graphs




## Classification

- It classifies discrete values such as male or female, true or false, and spam or not spam.
- It must identify the best fit line that predicts the output more accurately.

## Regression

- It predicts continuous values like price, wage, and age.
- It must identify the best decision boundary that divides the dataset into different classes.

# Key Takeaways

- In supervised learning, the program is given labeled input data and expected output results.
- Regression is preferred to predict a continuous quantity and classification is preferred to predict a discrete class label.
- Regression can be used for profit estimation in different departments such as R & D, administration, and marketing.
- In supervised learning, an algorithm is selected based on the target variable.