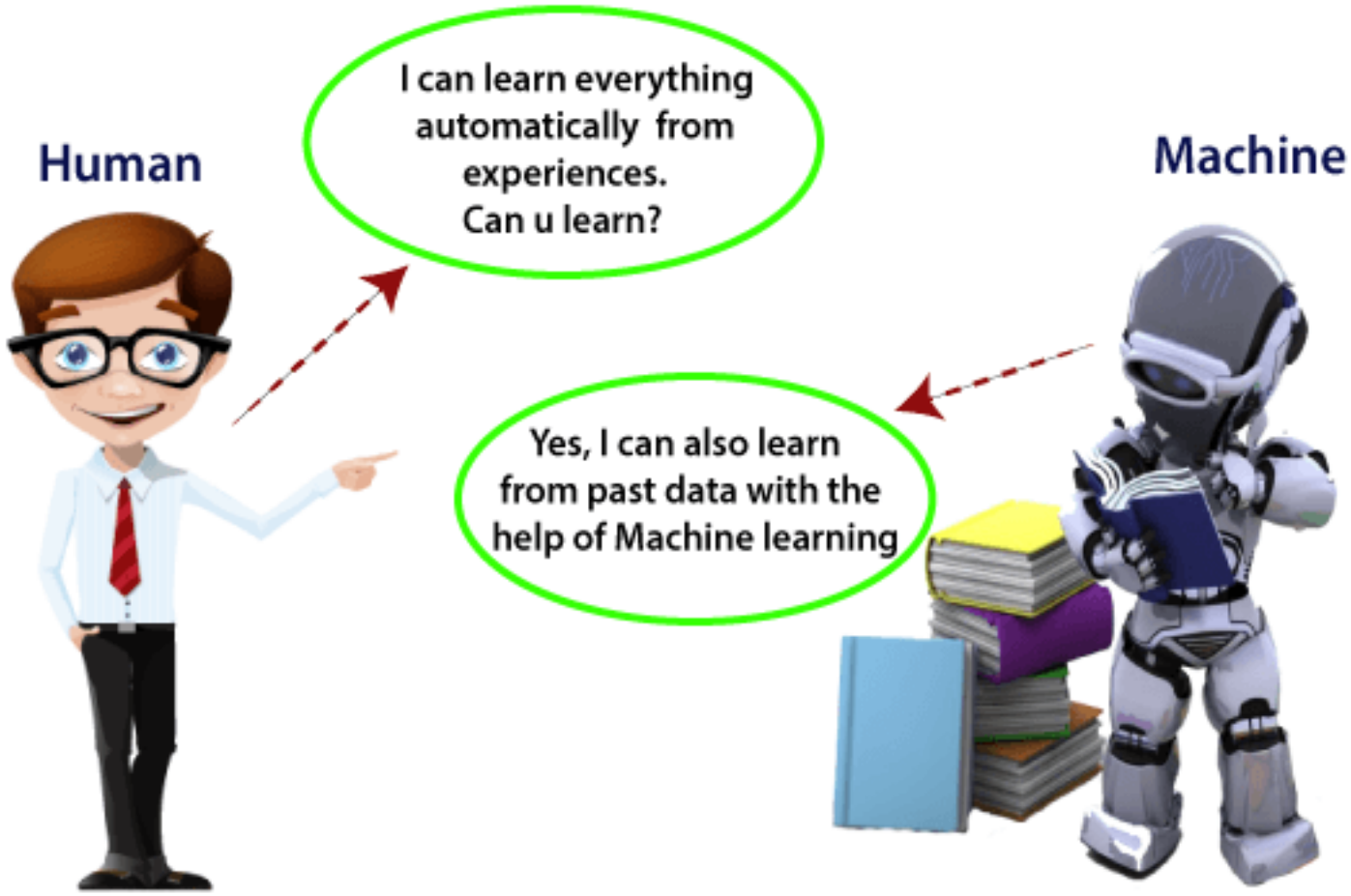
Introduction to Machine Learning

By
Sapna Yadav

What is Machine Learning?



What is Machine Learning?

- The term Machine Learning was first coined by **Arthur Samuel** in the year 1959.
- Humans can learn from their experiences with their learning capability,
- But can a machine also learn from experiences or past data like a human does?
- It is the branch of AI mainly concerned with the development of algorithms which allow a computer to learn from the **data** and **past experiences** on their own
- Machine Learning is a system that can learn from **example** through **self-improvement** and without being explicitly coded by programmer.

What is Machine Learning?

E * T = P

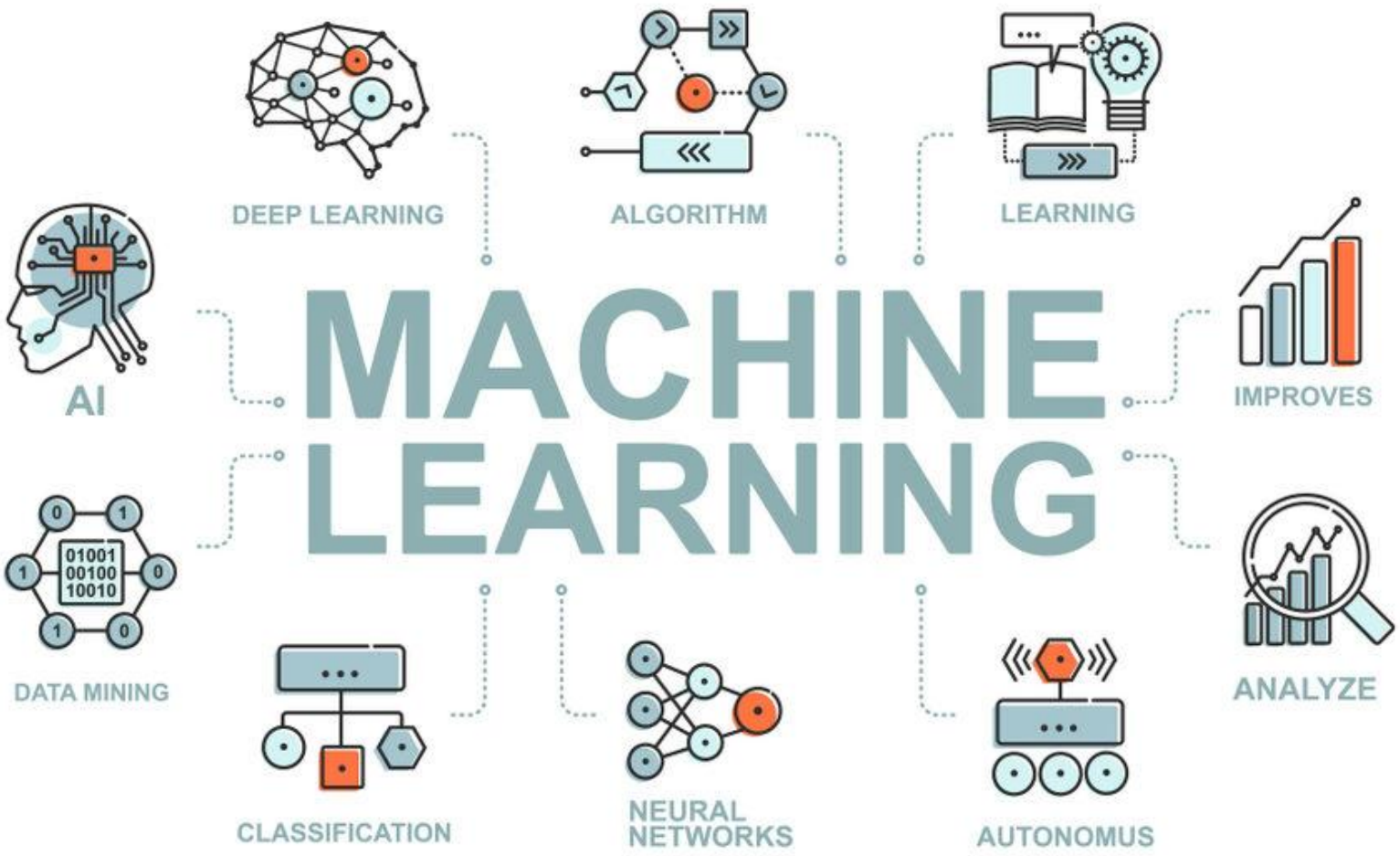
Experience * **Task** = **Performance**

Input Data:
- Housing prices
- Customer transactions
- Clickstream data
- Images

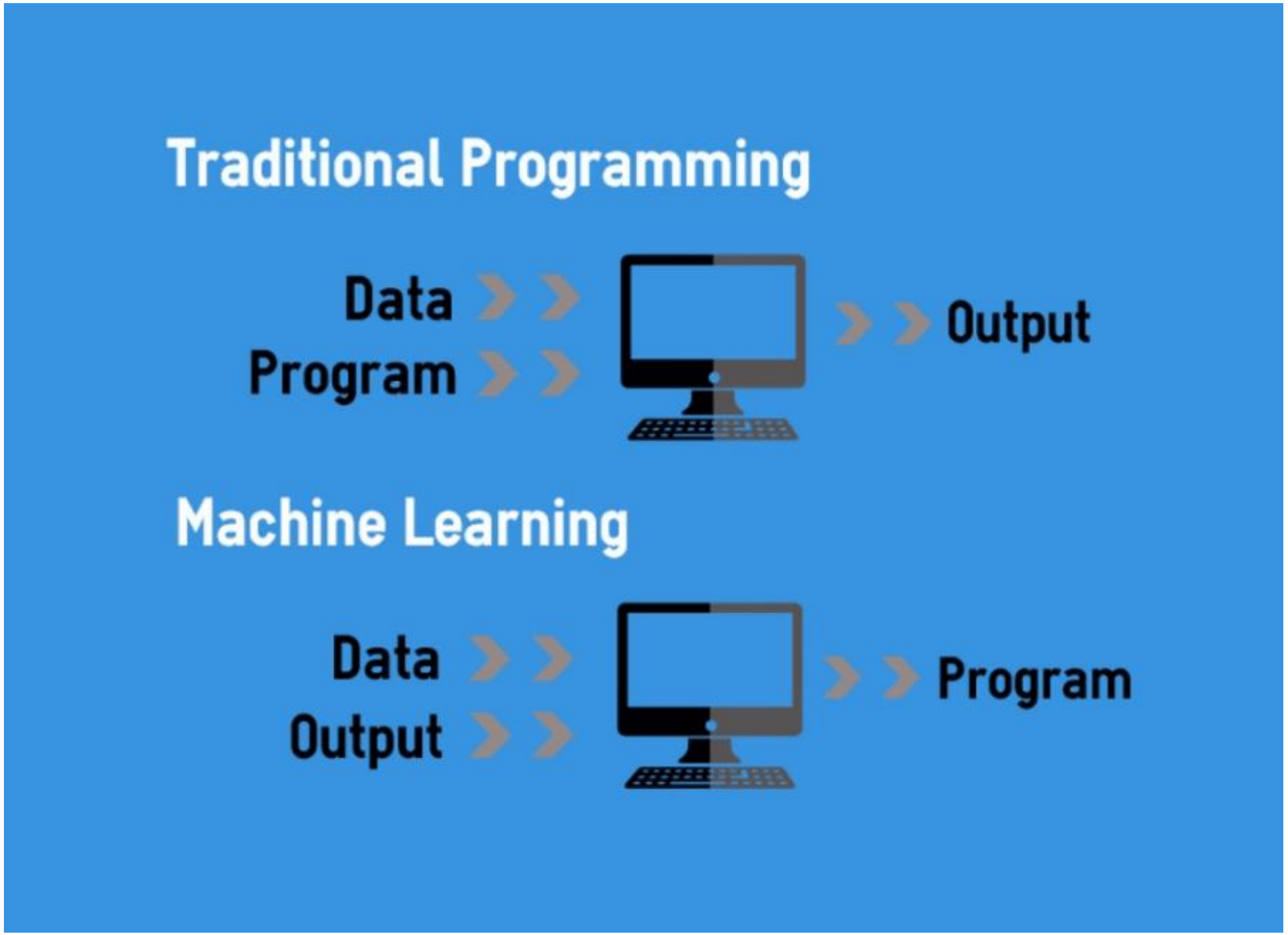
Task:
- Predict prices
- Segment customers
- Optimize user flows
- Categorize images

Performance:
- Accurate prices
- Coherent groupings
- KPI lifts
- Correctly sorted images

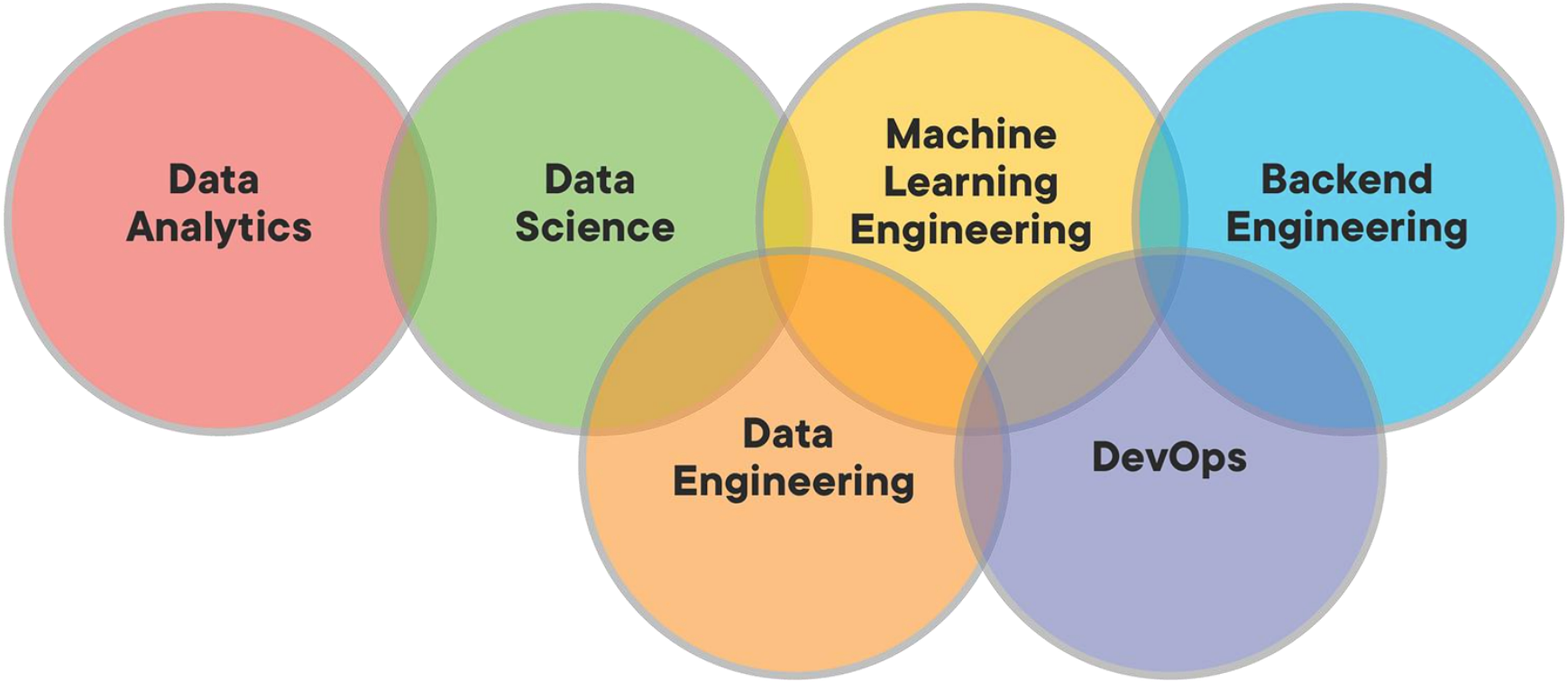
What is Machine Learning?



Traditional Programming VS Machine Learning?



Related fields



Applications

Steps- Machine Learning Process

- The Machine Learning process involves building a Predictive model that can be used to find a solution for a Problem Statement.
- To understand the Machine Learning process let's assume that you have been given a problem that needs to be solved by using Machine Learning.
- ***The problem is to predict the occurrence of rain in your local area by using Machine Learning.***

Steps- Machine Learning Process

Step 1: Define the objective of the Problem Statement

- the objective is to predict the possibility of rain by studying weather conditions. Decide what kind of **data** can be used to solve this problem or the type of **approach** (algorithm) you must follow to get to the solution.

Steps- Machine Learning Process

Step 2: Data Gathering

At this stage, you must be asking questions such as,

- What kind of data is needed to solve this problem?
- Is the data available?
- How can I get the data?

- Data collection can be done **manually** or by **web scraping**. There are 1000s of data resources on the web, you can just download the data set and get going.
- the data needed for weather forecasting includes measures such as ***humidity level, temperature, pressure, locality, whether or not you live in a hill station***, etc. Such data must be collected and stored for analysis.

Steps- Machine Learning Process

Step 3: Data Preparation

- The data you collected is almost never in the right format. You will encounter a lot of inconsistencies in the data set such as missing values, redundant variables, duplicate values, etc. Removing such inconsistencies is very essential because they might lead to wrongful computations and predictions. Therefore, at this stage, you scan the data set for any inconsistencies and you fix them then and there. It may involve steps like **Cleaning, Formatting, Sampling, Decomposition, Scaling**

Steps- Machine Learning Process

Step 4: Exploratory Data Analysis

- EDA or Exploratory Data Analysis is the **brainstorming stage** of Machine Learning. Data Exploration involves **understanding the patterns** and **trends** in the data. At this stage, all the useful **insights are drawn** and **correlations between the variables** are understood.
- For example, in the case of predicting rainfall, we know that there is a strong possibility of rain if the temperature has fallen low. Such correlations must be understood and mapped at this stage.

Steps- Machine Learning Process

Step 5: Building a Machine Learning Model

- All the insights and patterns derived during Data Exploration are used to build the Machine Learning Model. This stage always begins by **splitting the data** set into two parts, **training data, and testing data**.
- The training data will be used to build and analyze the model. The logic of the model is based on the Machine Learning Algorithm that is being implemented.
- In the case of predicting rainfall, since the output will be in the form of True (if it will rain tomorrow) or False (no rain tomorrow), we can use a Classification Algorithm such as Logistic Regression.

Steps- Machine Learning Process

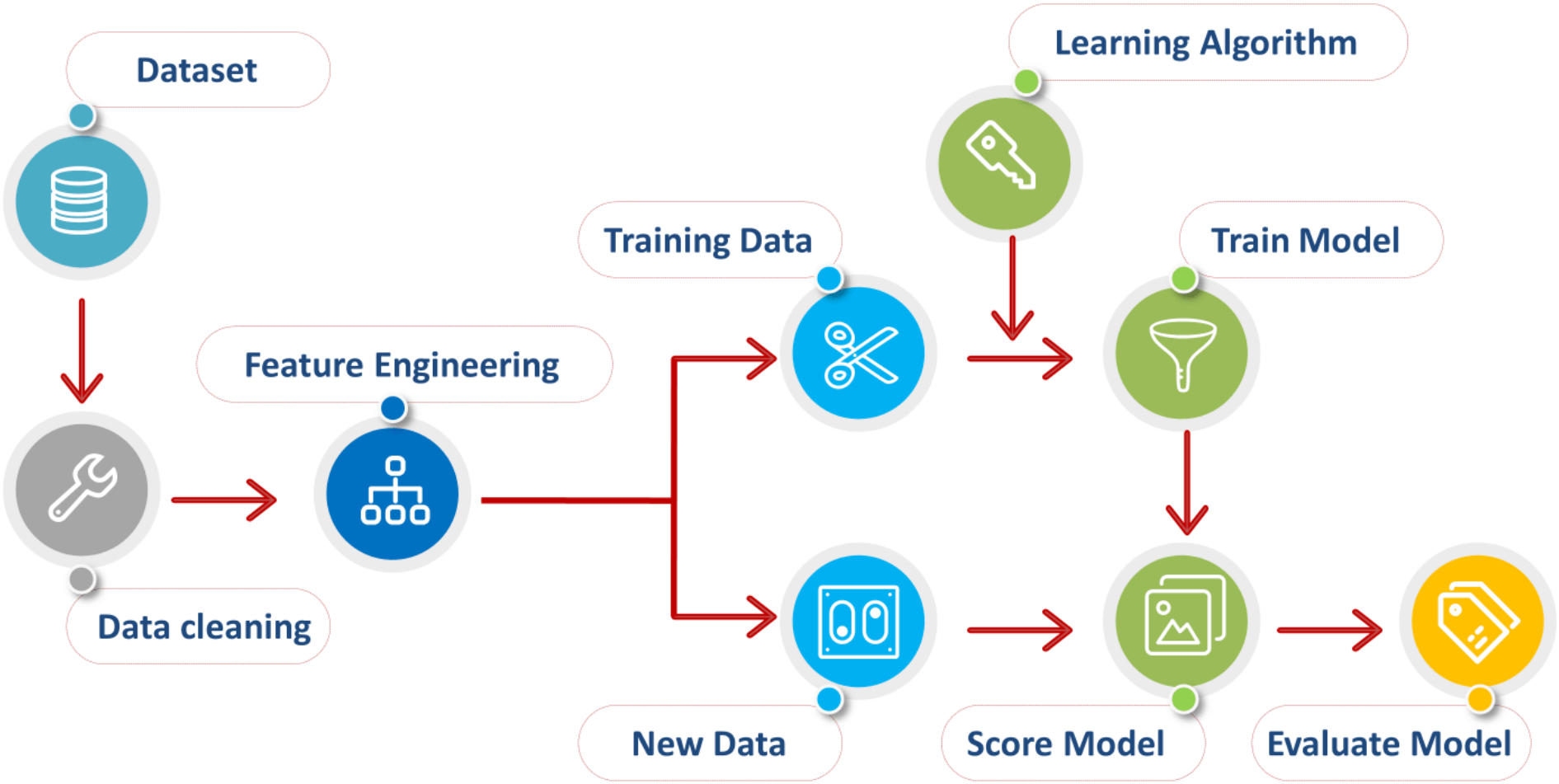
Step 6: Model Evaluation & Optimization

- After building a model by using the training data set, it is finally time to put the model to a test. The testing data set is used to check the **efficiency of the model** and **how accurately** it can predict the outcome. Once the accuracy is calculated, any further improvements in the model can be implemented at this stage. Methods like **parameter tuning** and **cross-validation** can be used to improve the performance of the model.

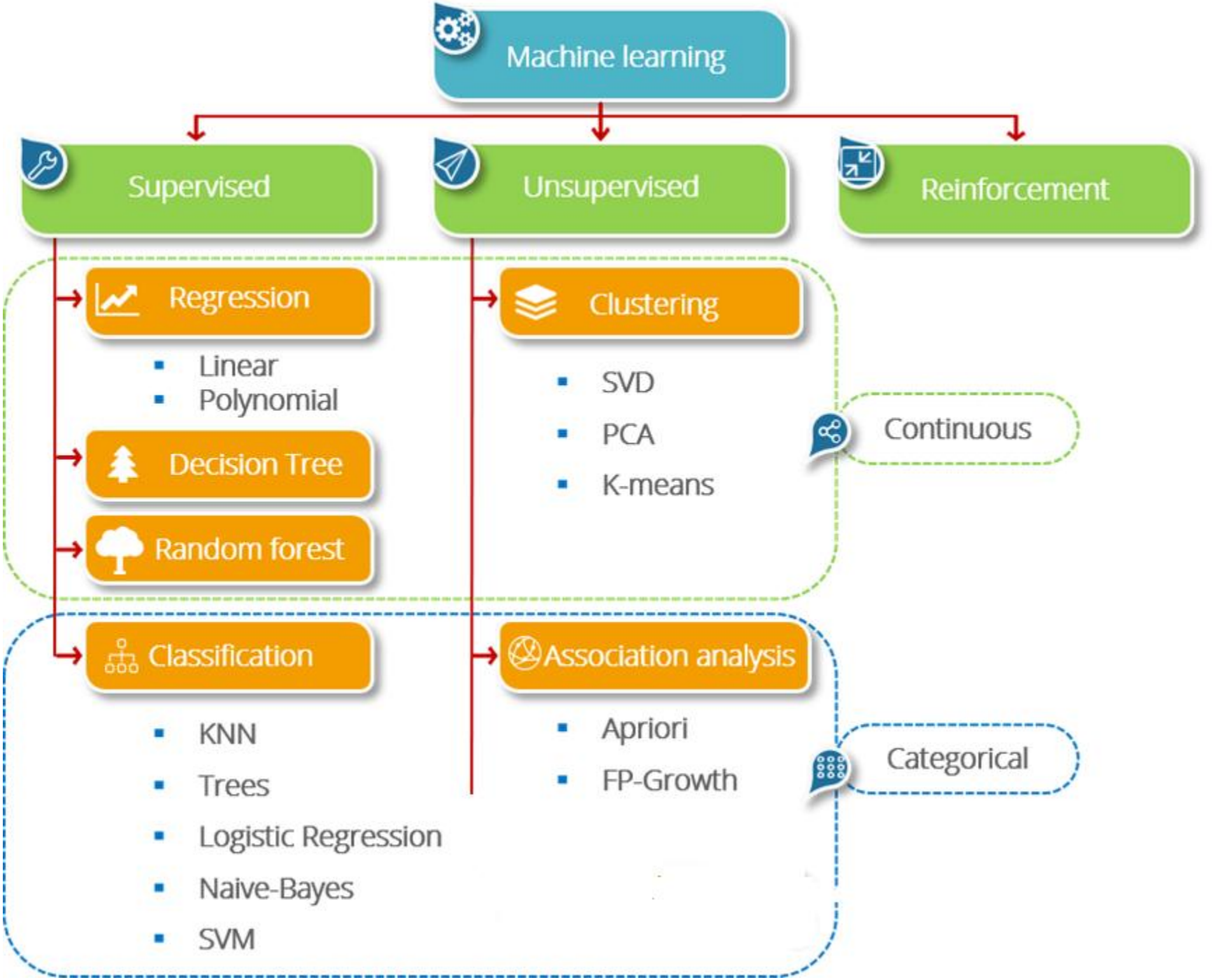
Step 7: Predictions

- Once the model is evaluated and improved, it is finally used to **make predictions**. The final output can be a **Categorical variable** (eg. True or False) or it can be a **Continuous Quantity** (eg. the predicted value of a stock).
- In our case, for predicting the occurrence of rainfall, the output will be a categorical variable.

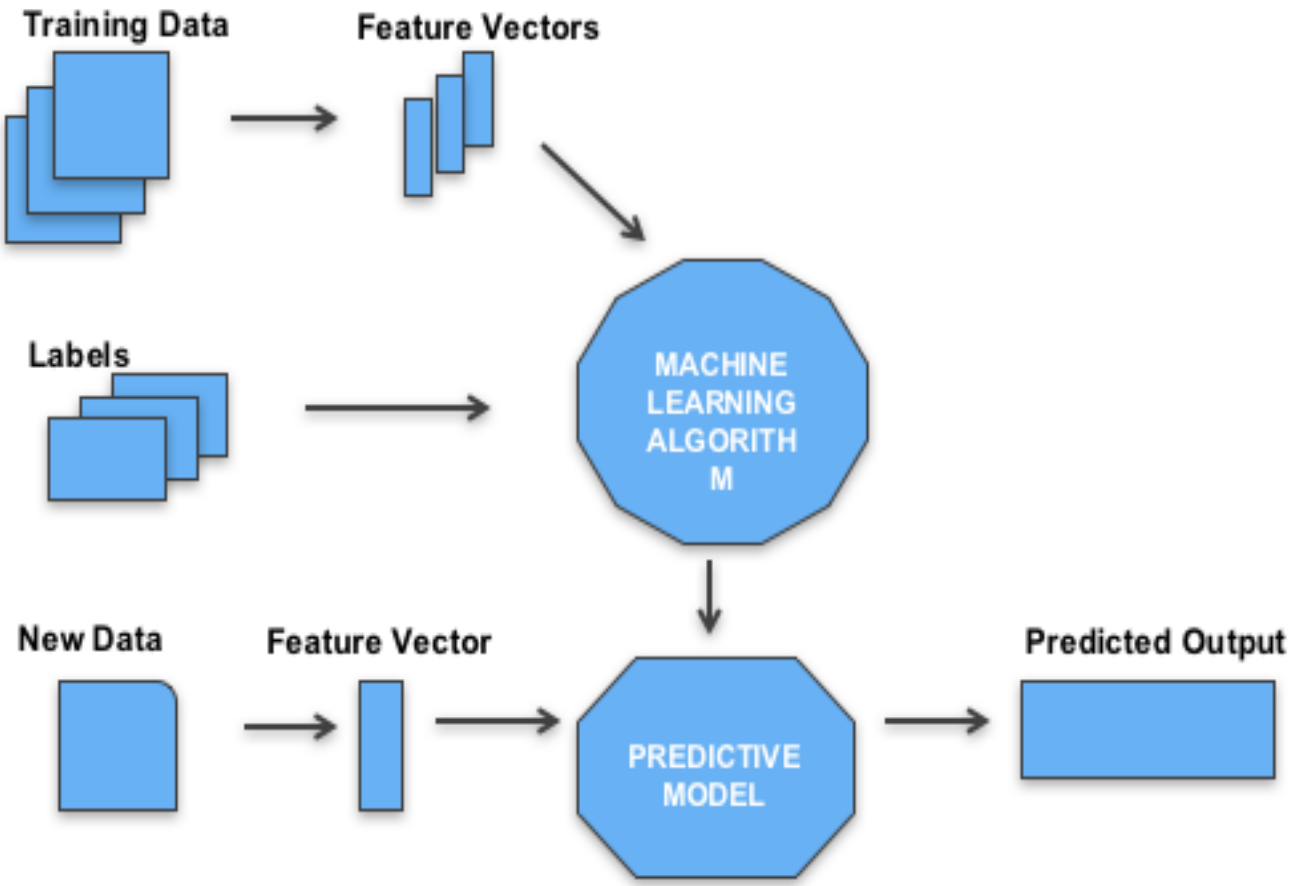
Life Cycle


Dataset
Data cleaning
Feature Engineering
Training Data
New Data
Learning Algorithm
Train Model
Score Model
Evaluate Model

Types of Learning

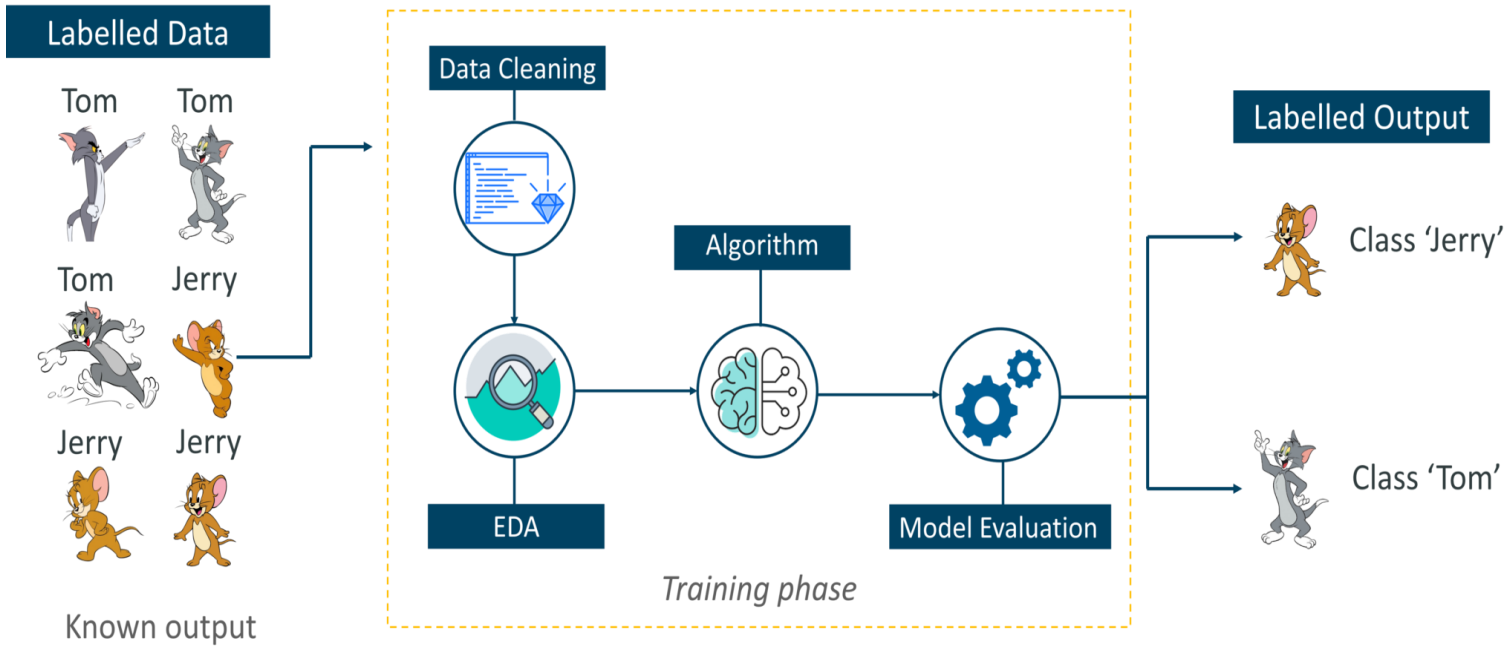

Machine learning
Supervised
Unsupervised
Reinforcement
Regression
Linear
Polynomial
Decision Tree
Random forest
Clustering
SVD
PCA
K-means
Continuous
Classification
KNN
Trees
Logistic Regression
Naive-Bayes
SVM
Association analysis
Apriori
FP-Growth
Categorical

Supervised Learning



Supervised Learning

- Supervised learning is a technique in which we teach or train the machine using data which is well labeled.
- *Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors are known as supervised learning problems.*



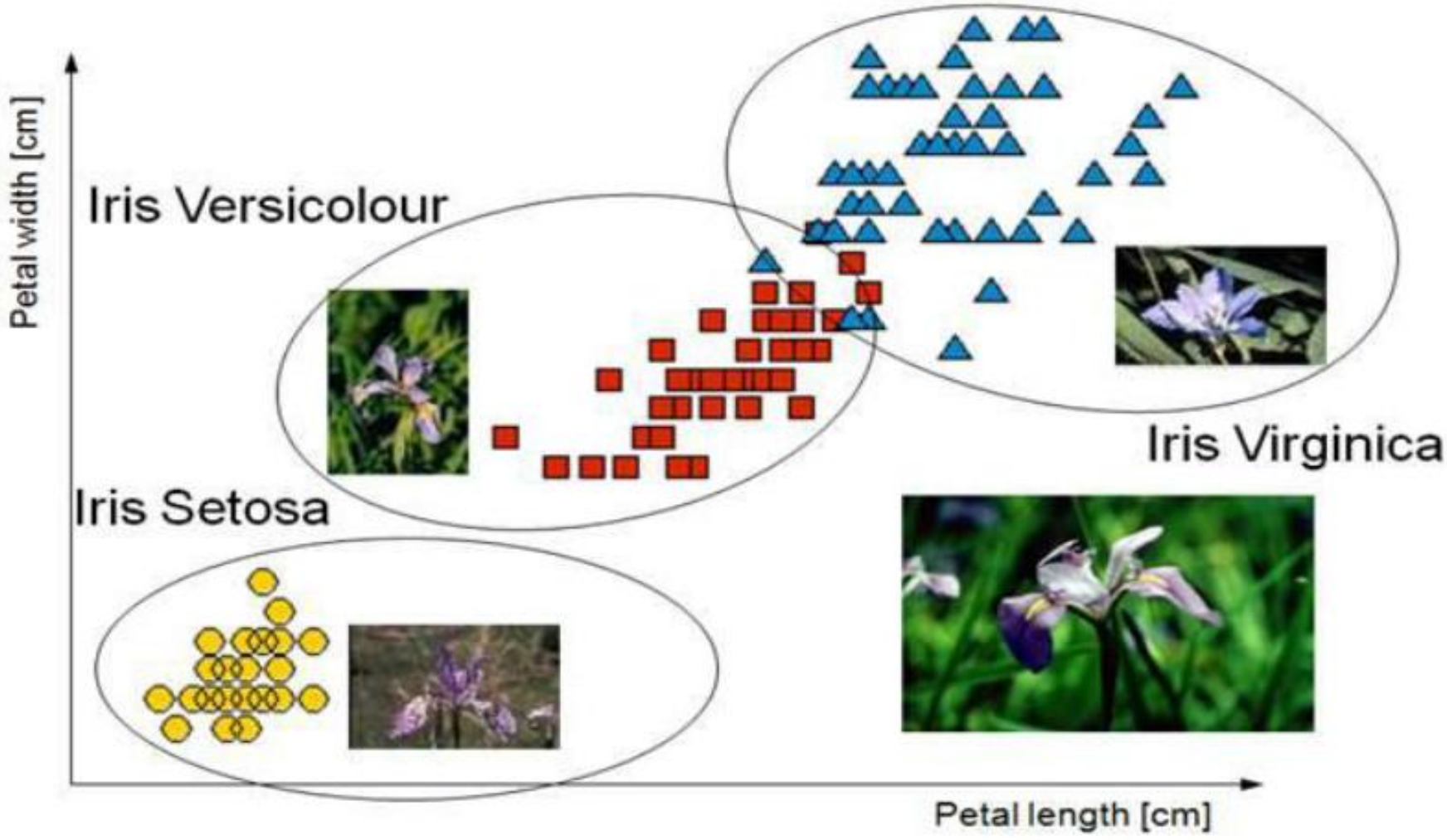
Classification VS Regression

- There are two main types of supervised learning problems:
- **Classification**: Supervised learning problem that involves predicting a class label.
- **Regression**: Supervised learning problem that involves predicting a numerical label.

Regression

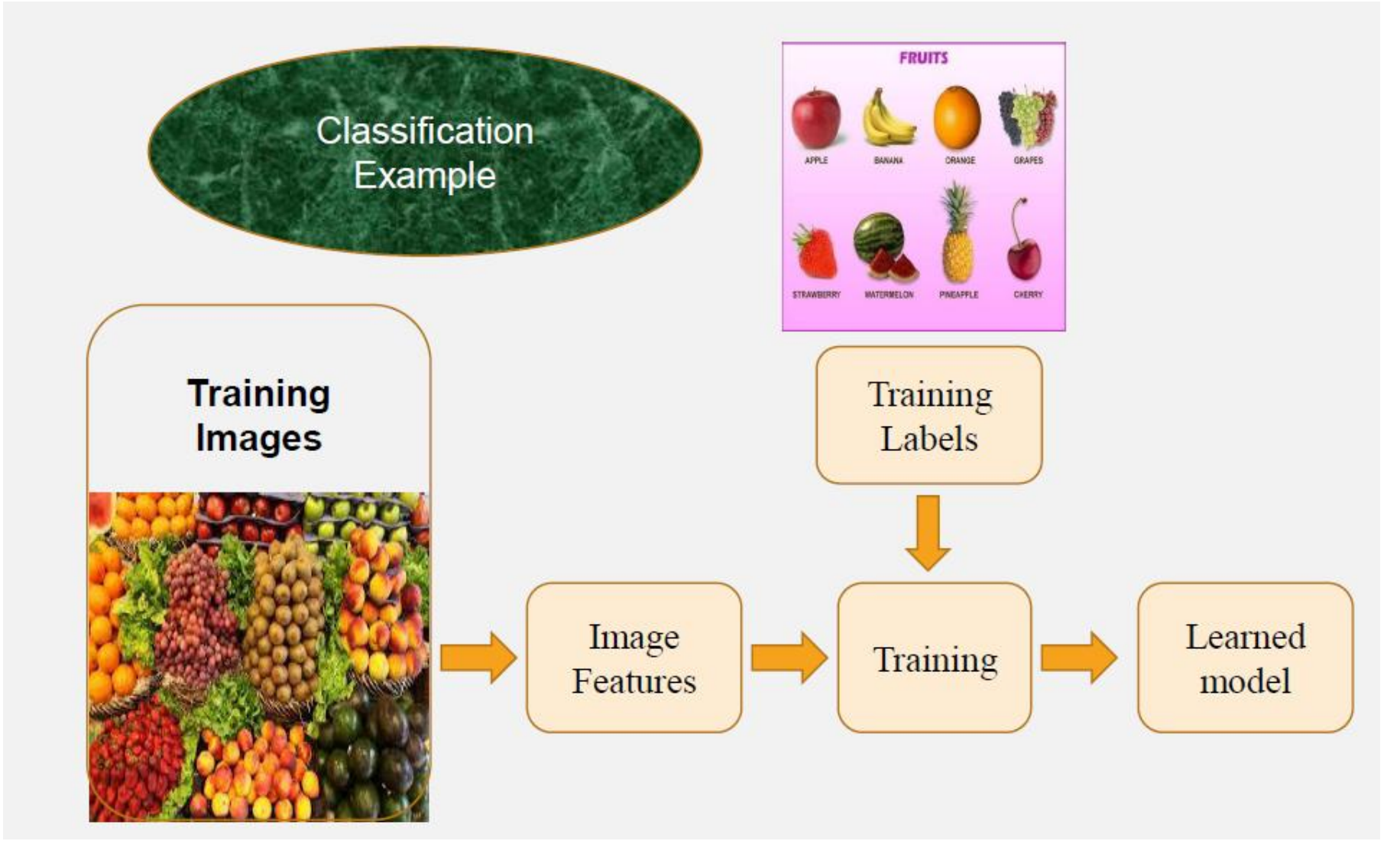
SUPERVISED LEARNING

Classification

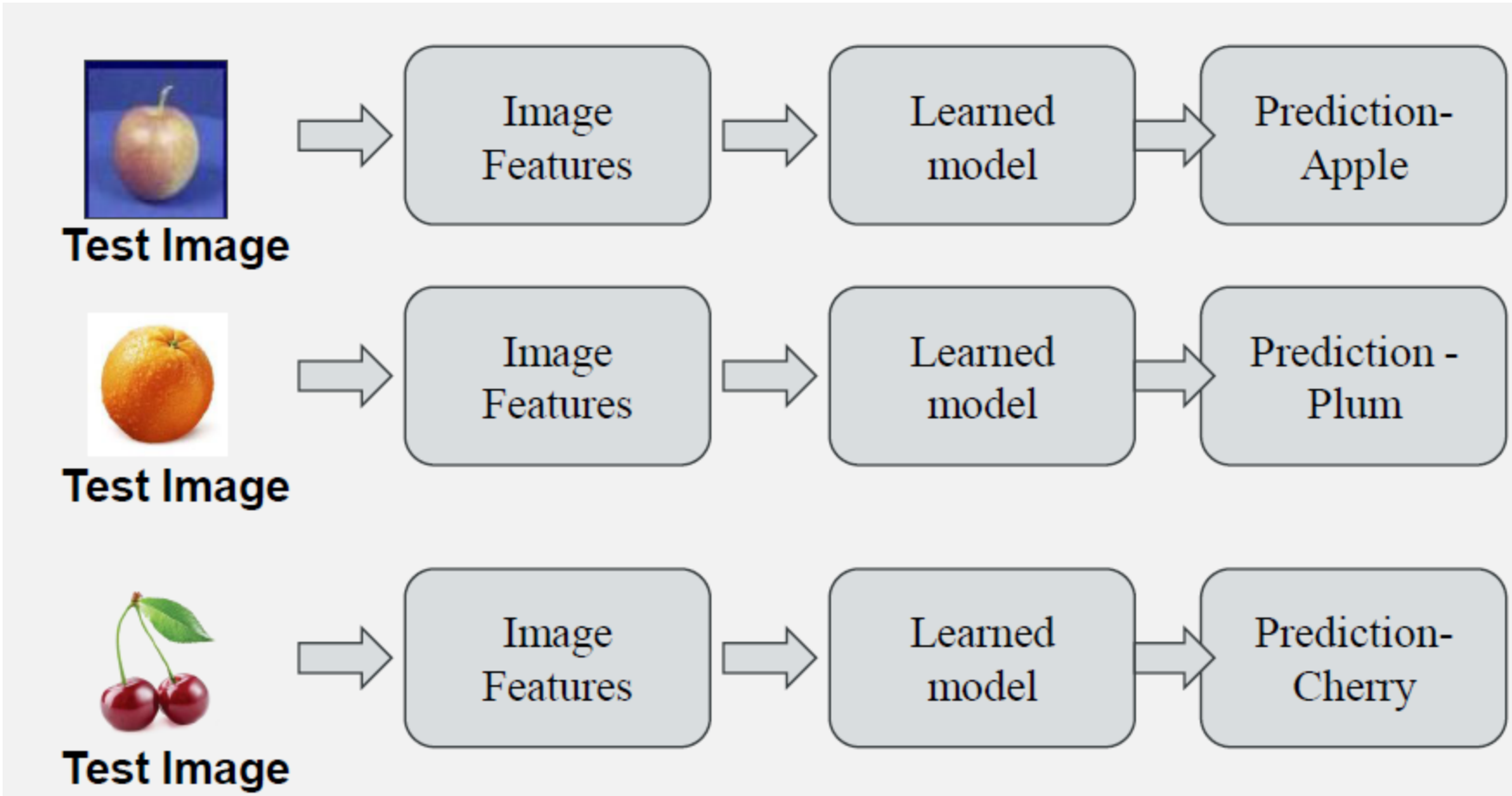


SUPERVISED LEARNING

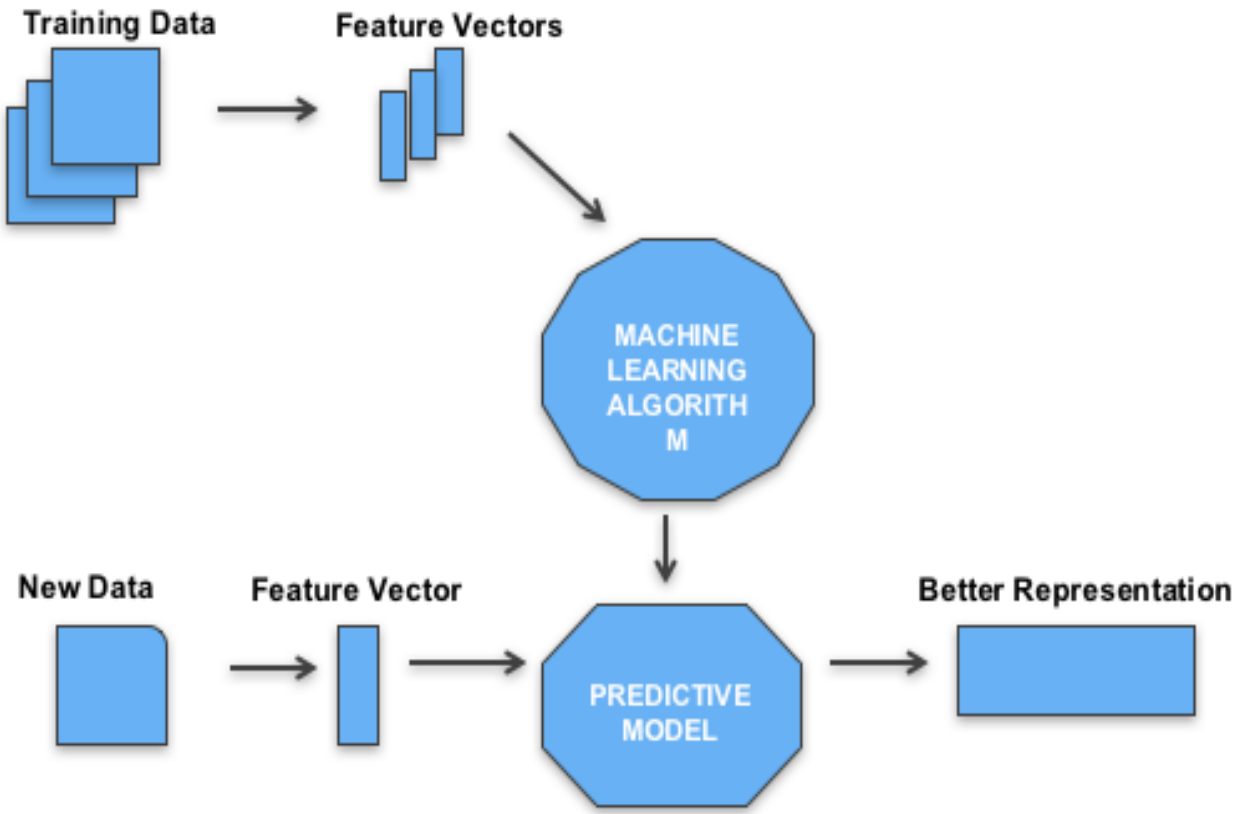
Training Framework



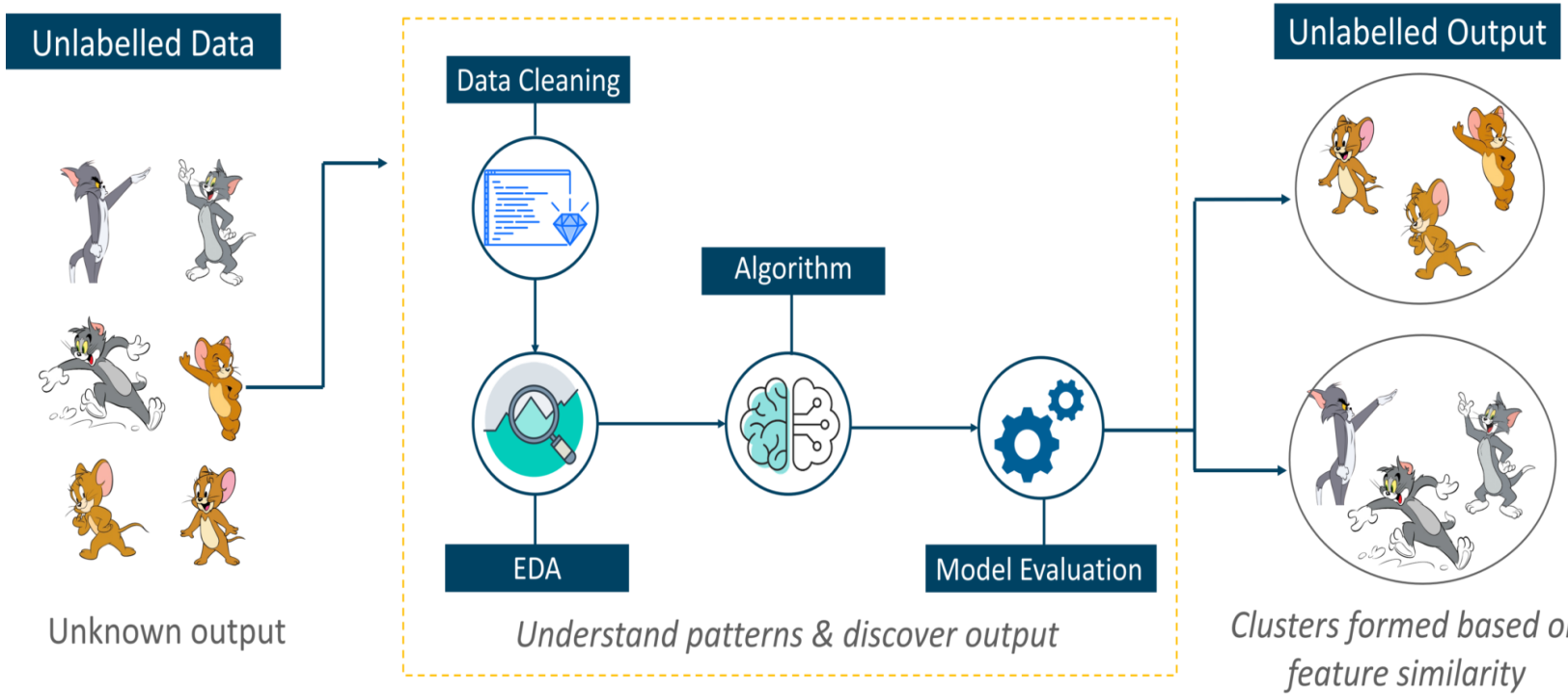
Testing Framework



Unsupervised Learning



Unsupervised Learning



For example, it identifies prominent features of Tom such as pointy ears, bigger size, etc, to understand that this image is of type 1. Similarly, it finds such features in Jerry and knows that this image is of type 2. Therefore, it classifies the images into two different classes without knowing who Tom is or Jerry is.

Clustering

sample

Cluster/group

It mainly deals with finding a structure or pattern in a collection of uncategorized data. Clustering algorithms will process your data and find natural clusters(groups) if they exist in the data.

Example of Unsupervised Learning

- Let's, take the case of a baby and her family dog.

She knows and identifies this dog. Few weeks later a family friend brings along a dog and tries to play with the baby.

Baby has not seen this dog earlier. But it recognizes many features (2 ears, eyes, walking on 4 legs) are like her pet dog. She identifies the new animal as a dog. This is unsupervised learning, where you are not taught but you learn from the data (in this case data about a dog.)

Had this been supervised learning, the family friend would have told the baby that it's a dog.

Supervised vs. Unsupervised Machine Learning

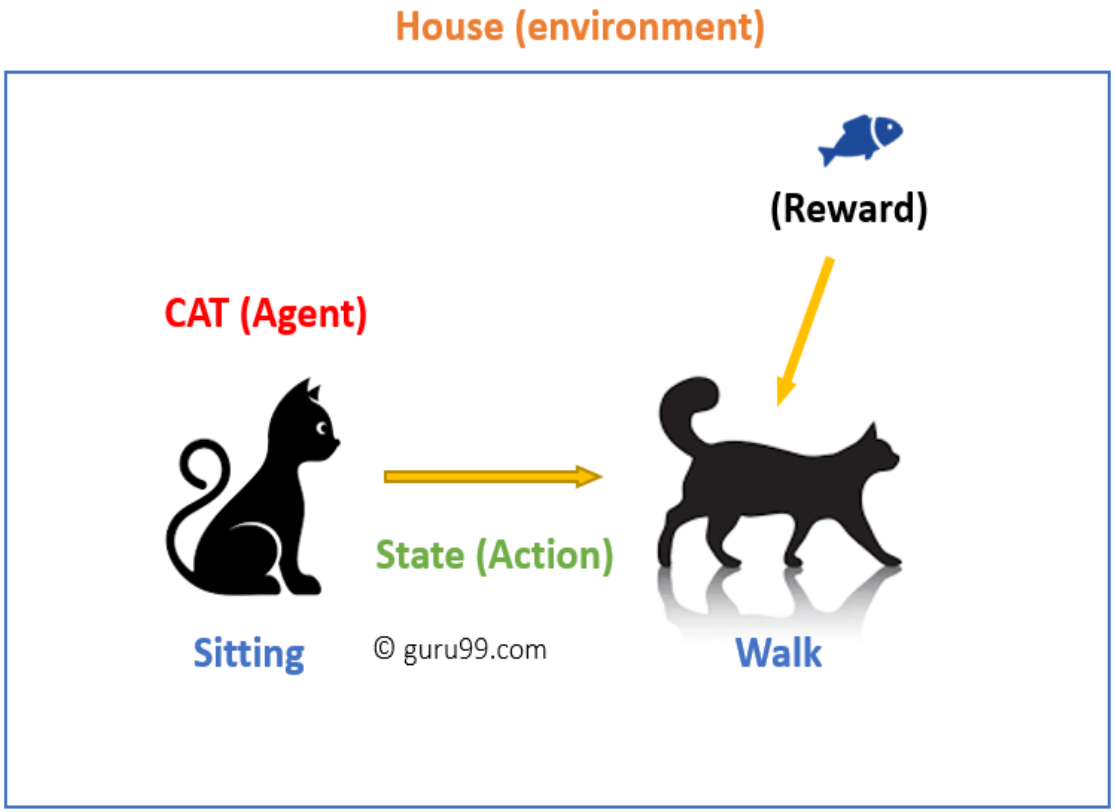
Parameters	Supervised machine learning technique	Unsupervised machine learning technique
Input Data	Algorithms are trained using labeled data.	Algorithms are used against data which is not labelled
Computational Complexity	Supervised learning is a simpler method.	Unsupervised learning is computationally complex
Accuracy	Highly accurate and trustworthy method.	Less accurate and trustworthy method.

Reinforcement Learning

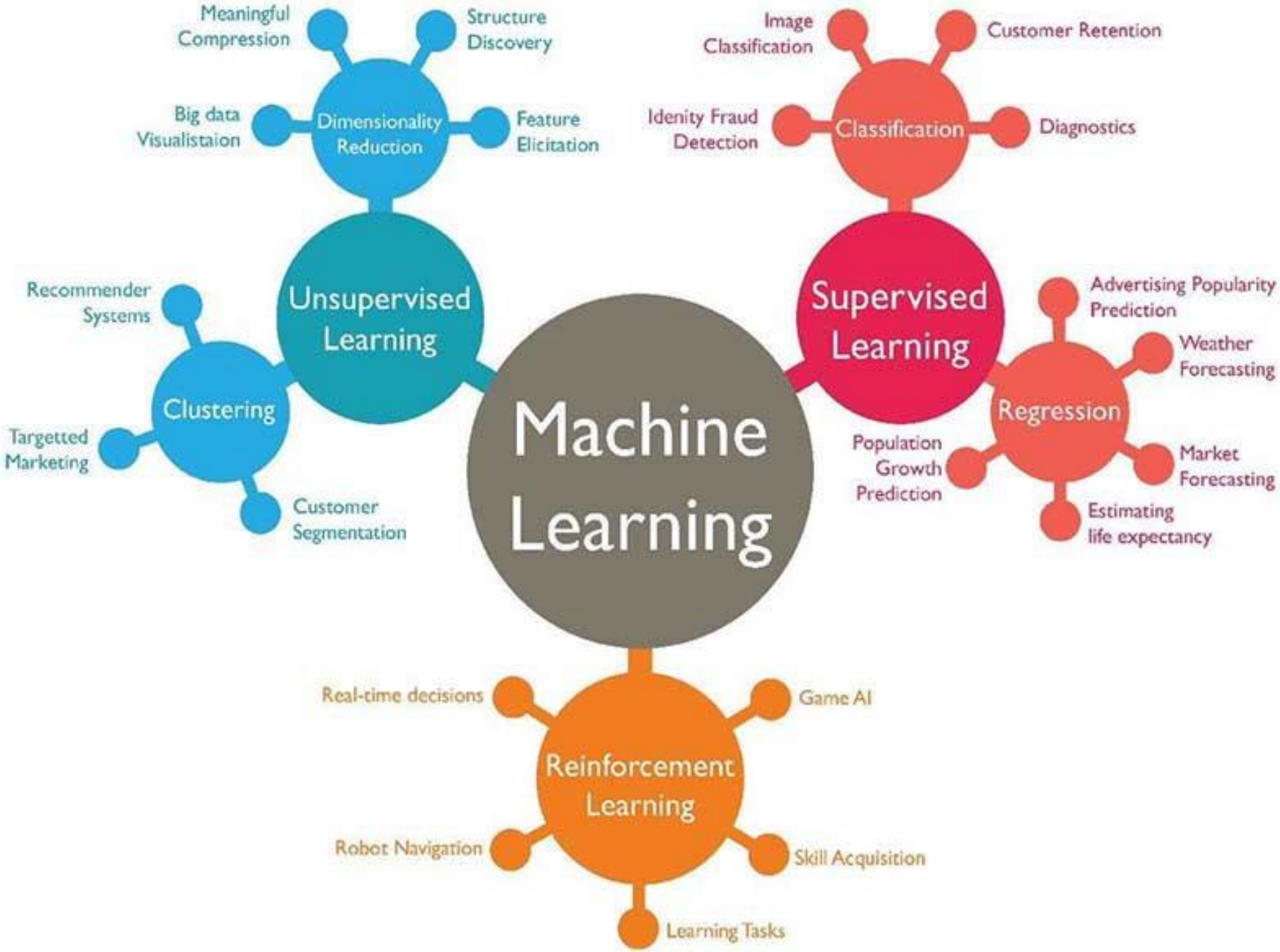
- Reinforcement Learning is a part of Machine learning where an agent is put in an environment and he learns to behave in this environment by performing certain actions and observing the rewards which it gets from those actions.
- Imagine that you were dropped off at an isolated island! What would you do?
- As time passes by, you will learn how to live on the island. You will explore the environment, understand the climate condition, the type of food that grows there, the dangers of the island, etc. This is exactly how Reinforcement Learning works,
- it involves an **Agent** (**you, stuck on the island**) that is put in an unknown environment (island), where he must learn by observing and performing actions that result in rewards.
- Reinforcement Learning is mainly used in **advanced Machine Learning areas** such as self-driving cars etc.

How Reinforcement Learning works?

- Consider the scenario of teaching new tricks to your cat
- As cat doesn't understand English or any other human language, we can't tell her directly what to do. Instead, we follow a different strategy.
- We emulate a situation, and the cat tries to respond in many different ways. If the cat's response is the desired way, we will give her fish.
- Now whenever the cat is exposed to the same situation, the cat executes a similar action with even more enthusiastically in expectation of getting more reward(food).
- That's like learning that cat gets from "what to do" from positive experiences.
- At the same time, the cat also learns what not do when faced with negative experiences.



- There is no supervisor, only a real number or reward signal
- Sequential decision making
- Time plays a crucial role in Reinforcement problems
- Feedback is always delayed, not instantaneous
- Agent's actions determine the subsequent data it receives


Machine Learning
Unsupervised Learning
Supervised Learning
Reinforcement Learning
Dimensionality Reduction
Clustering
Classification
Regression
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Recommender Systems
Targetted Marketing
Customer Segmentation
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Challenges in Machines Learning

- **Quality of data** − Having good-quality data for ML algorithms is one of the biggest challenges.
- **Time-Consuming task** − data acquisition, feature extraction and retrieval.
- **Issue of overfitting & underfitting** − If the model is overfitting or underfitting, it cannot be represented well for the problem.
- **Curse of dimensionality** − too many features of data points. This can be a real hindrance.
- **Difficulty in deployment** − Complexity of the ML model makes it quite difficult to be deployed in real life.

Thank You