CADT
Cambodia Academy of Digital Technology

# Introduction to Machine Learning

Lecturer: Soklay HENG

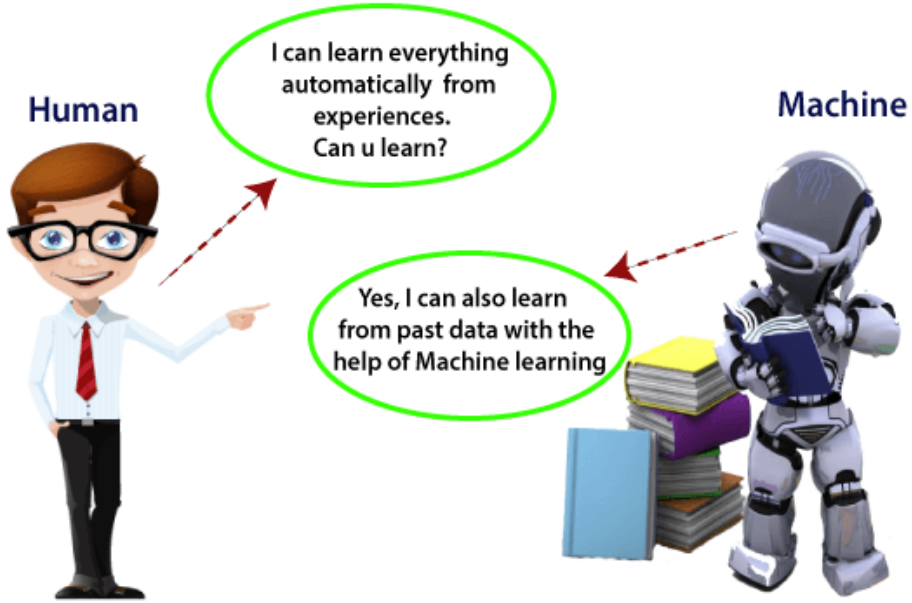
Academic Year: 2026-2027

# Contents

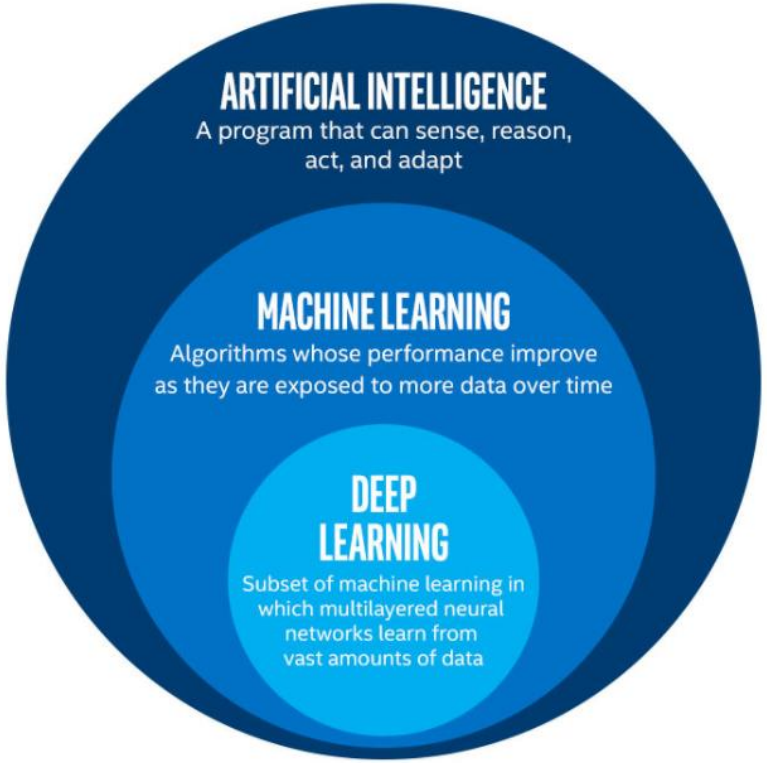
- What is Machine Learning?
- Why Machine Learning?
- Types of Machine Learning
- Machine Learning Algorithm
- Machine Learning Life Cycle

# Can a machine learn?

- In the real world, we are surrounded by humans who can learn everything from their experiences using their learning capability, and we have machines which work on our instructions. But can a machine learns from experiences like a human does? So here comes the role of **Machine Learning**.



# 1.What is Machine Learning?




- **Machine Learning** is a subset of AI that is mainly concerned with development of algorithms which provide computers the ability to solve problem by **learning from past experiences on their own** in the form of observations and real-world interactions.
- The term "Machine Learning" was first introduced by **Arthur Samuel** in 1959**.**
- Machine learning uses various algorithms from **Computer Science**, **Statistics** and other fields to build **mathematical predictive models**.

# 1. What is Machine Learning?

- **Arthur Samuel (1959)** Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
- **Tom Mitchell (1998)** Well-posed Learning Problem: A computer program is said to learn from experience $E$ with respect to some task $T$ and some performance measure $P$, if its performance on $T$, as measured by $P$, improves with experience $E$.

1. Classifying emails as spam or not spam
2. Watching you label emails as spam or not spam
3. The number (or fraction) of emails correctly classified as spam/not spam



# 2. Why Machine Learning?

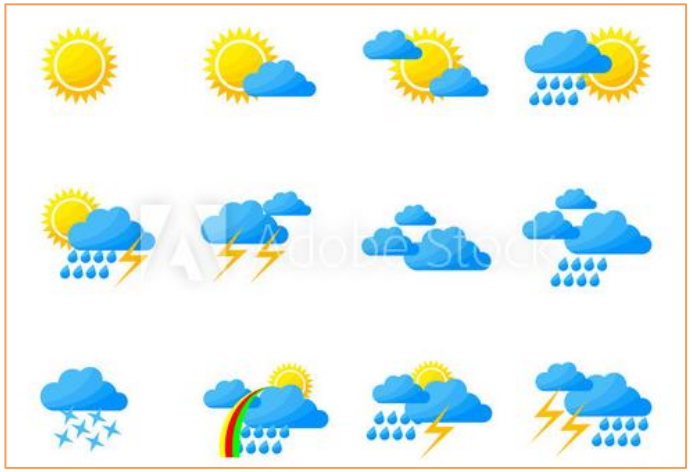
Calculate Rectangle Area vs. Weather Forecast



Weight

Length

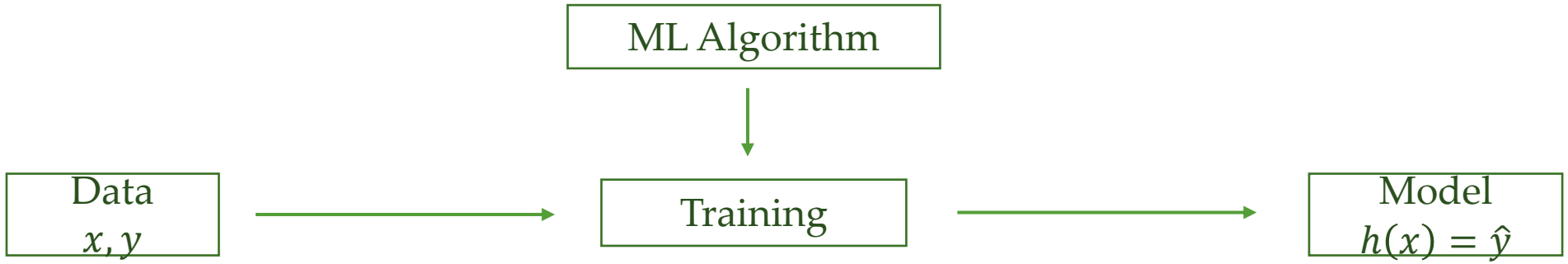
$h(x) = Weight \times Length$



How to define $h(x)$ to predict the weather?

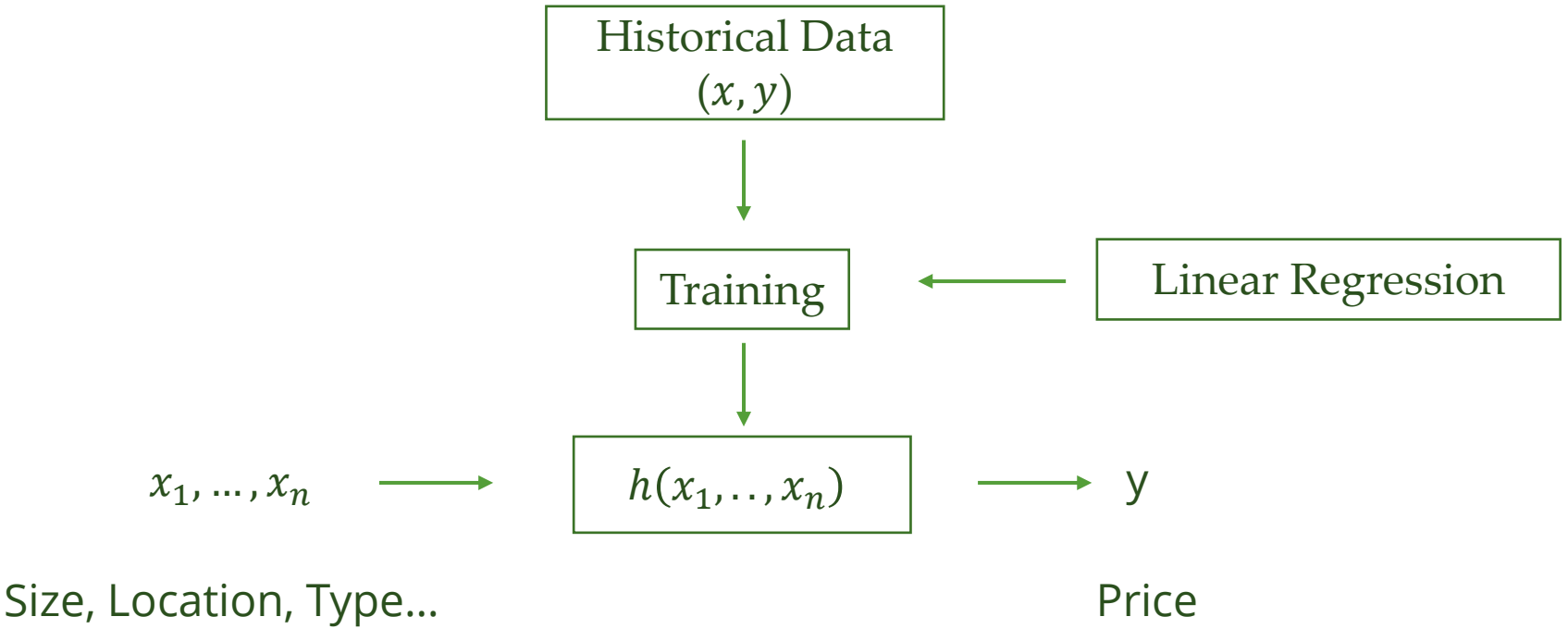
# 2. Why Machine Learning?

- Some problems cannot be modeled or interpreted easily (i.e. weather forecast, sale prediction…)
- Instead, they can be solved by learning a predictive function $\boldsymbol{h}(\boldsymbol{x})$ through experiences/data describing the relationship between input variables and output variable.
- Machine Learning is more feasible and scalable in solving problem because it is data-driven based solution.

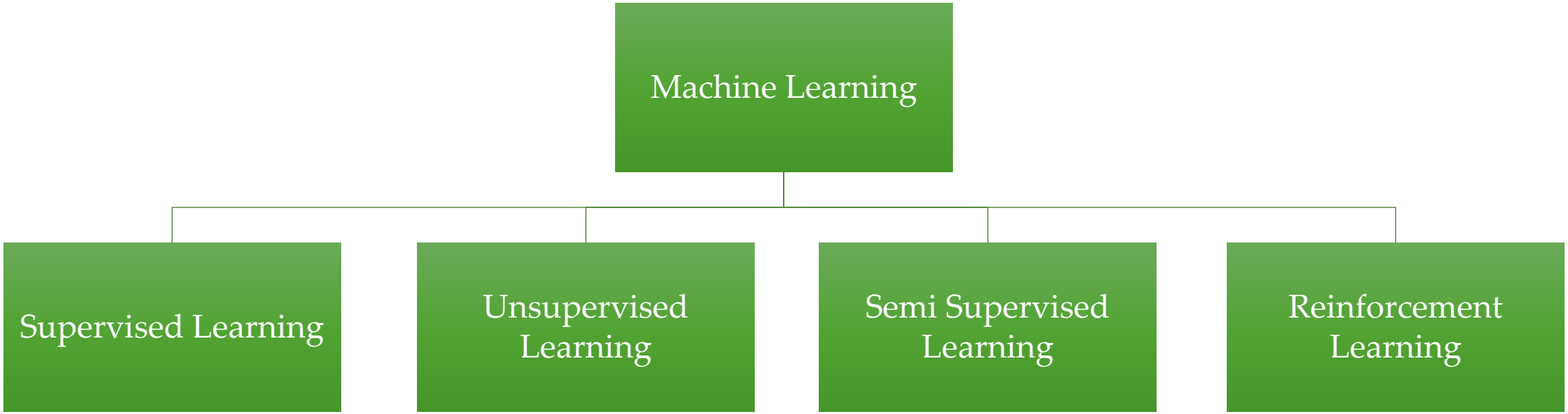


# 2. Why Machine Learning?

Example: House Price Prediction

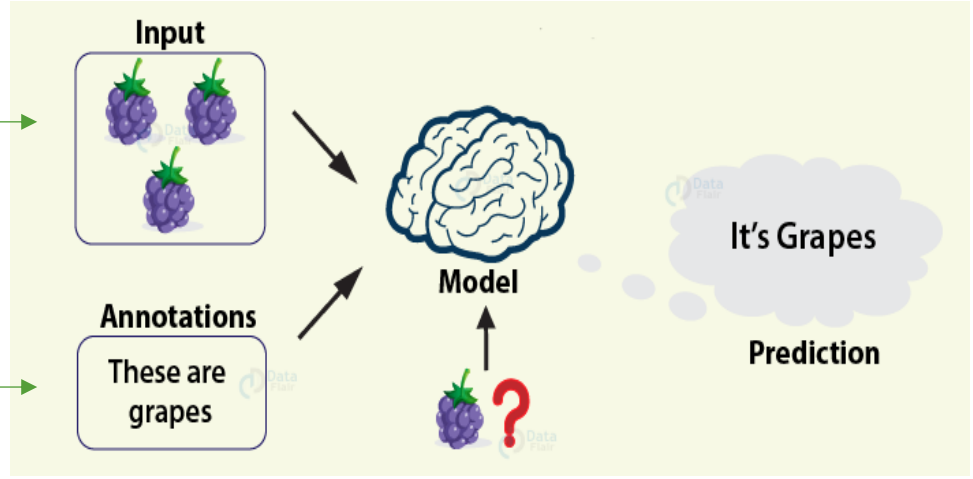


# 3. Types of Machine Learning



# 3.1 Supervised Learning

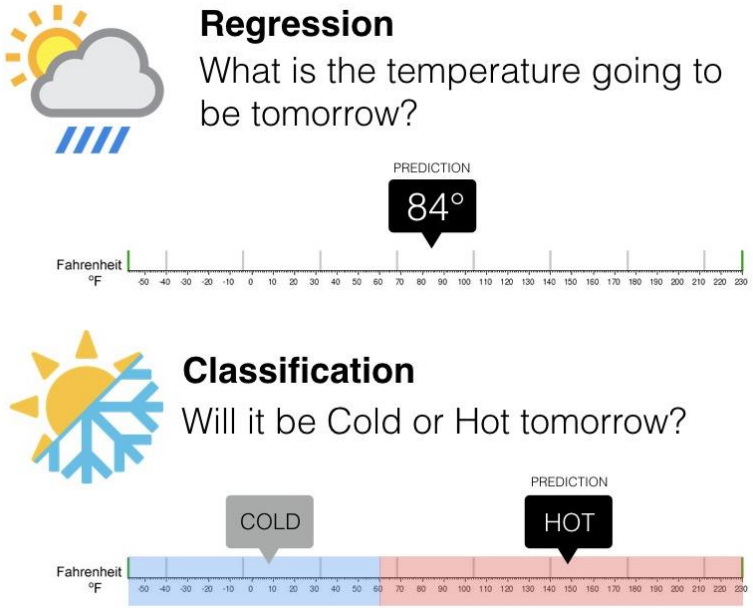
- **Supervised Learning** is a learning method which aims to learn an approximate function that maps an input to an output based on example input-output pairs.
- This learning is based on **supervision** which is the same as when a student learns things in the supervision of the teacher.




**Supervised Learning ⇒ Label is given**

# 3.1 Supervised Learning

- **Regression** algorithms attempt to estimate the mapping function from the input variables to numerical or **continuous output variables**.
- **Classification** algorithms attempt to estimate the mapping function from the input variables to **discrete or categorical output variables**.

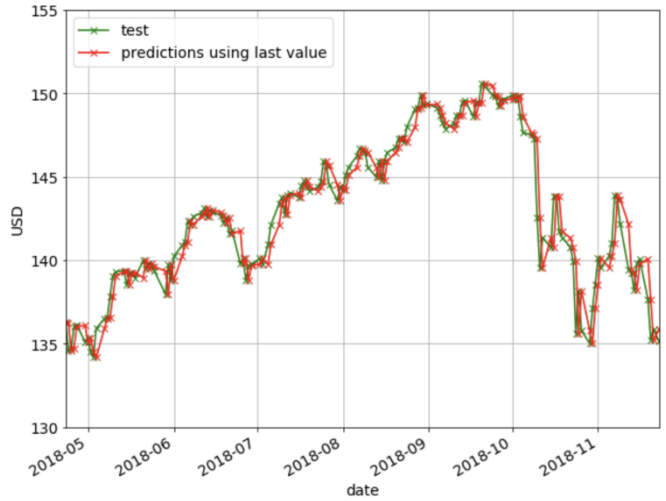



$x \rightarrow \boxed{h(x)} \rightarrow y$ $\in \mathbb{R}$ - Regression

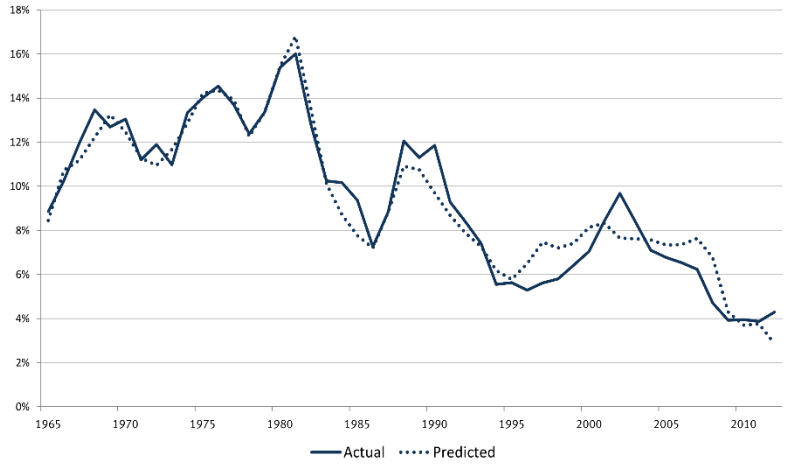
$y \in \mathbb{N}$ - Classification

# 3.1 Supervised Learning - Tasks

Examples: Regression




Stock Price Prediction

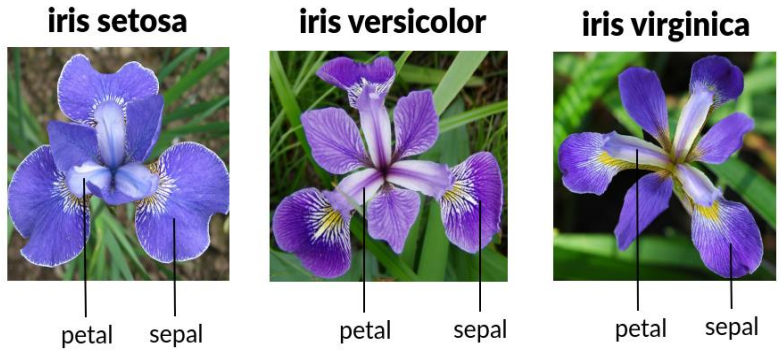



Health Spending Prediction

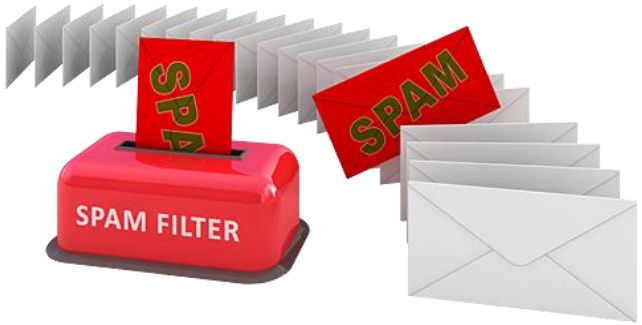
**Regression:** predict continuous valued output

# 3.1 Supervised Learning - Tasks

- Examples: Classification



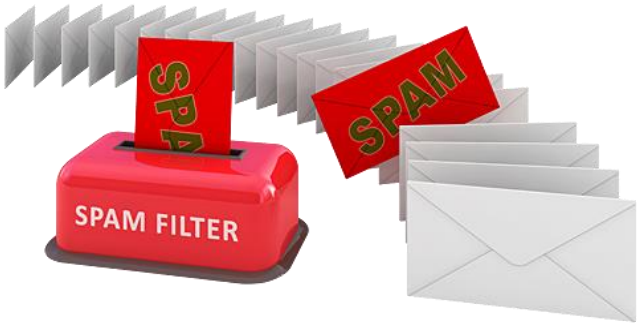

Iris Flower Prediction



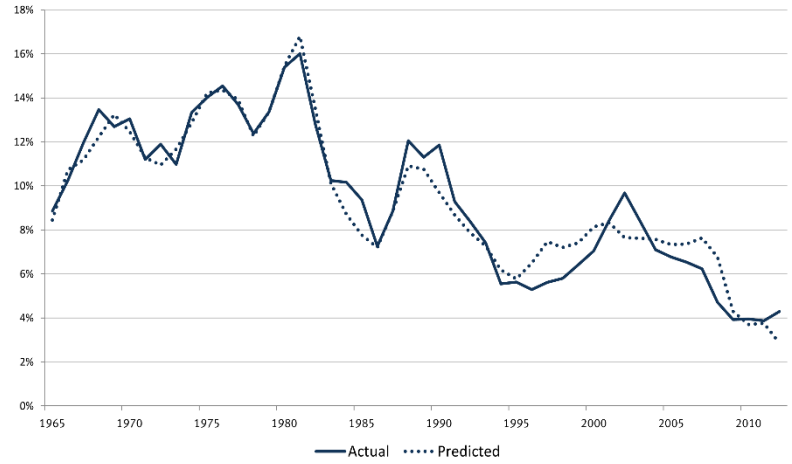

Spam Filter

**Classification:** predict discrete valued output

# 3.1 Supervised Learning - Tasks



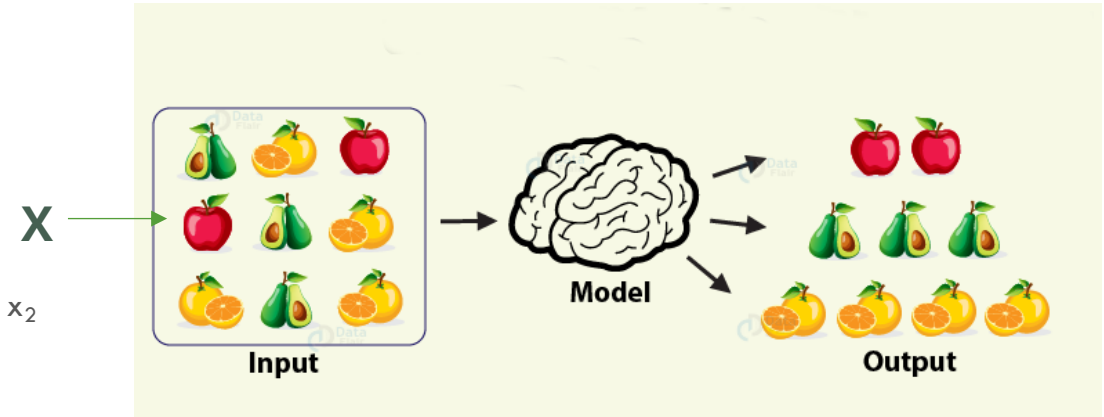

Spam Filter




- List down 5 classification & regression problems

# 3.2 Unsupervised Learning

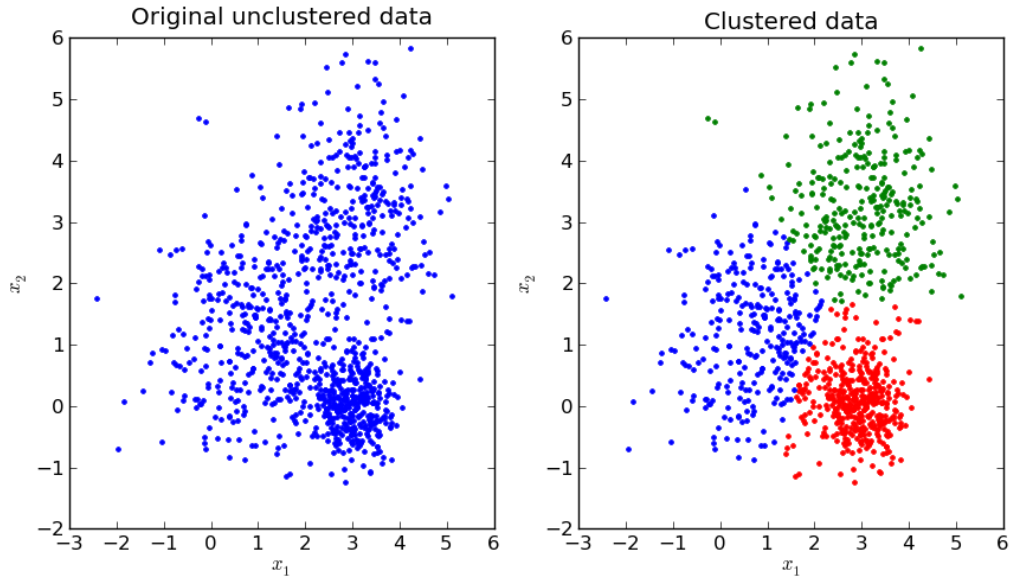
- **Unsupervised Learning** is a learning approach which is done without any supervision.
- The goal is to restructure the input data into new features or a group of objects with similar patterns. We don't have a predetermined result. The machine tries to find useful insights from the huge amount of data.




**Unsupervised Learning ⇒ no label is given**

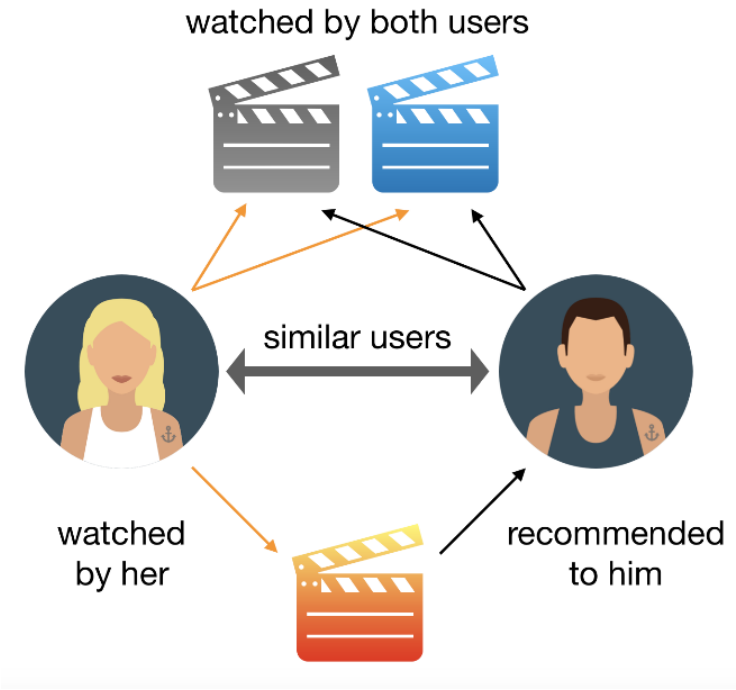
# 3.2 Unsupervised Learning - Tasks

- **Cluster analysis** or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups.

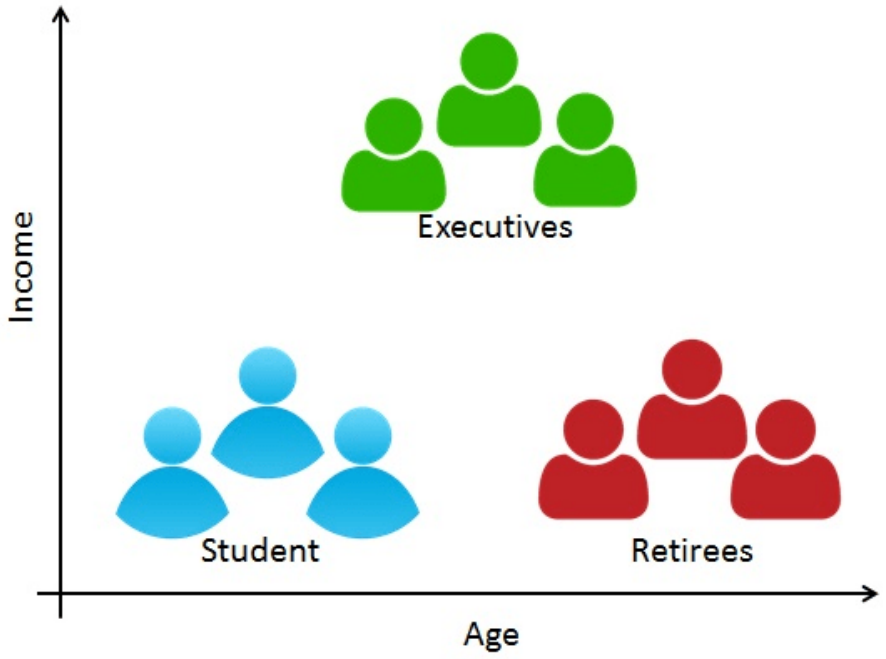


# 3.2 Unsupervised Learning - Tasks

- Examples: Clustering



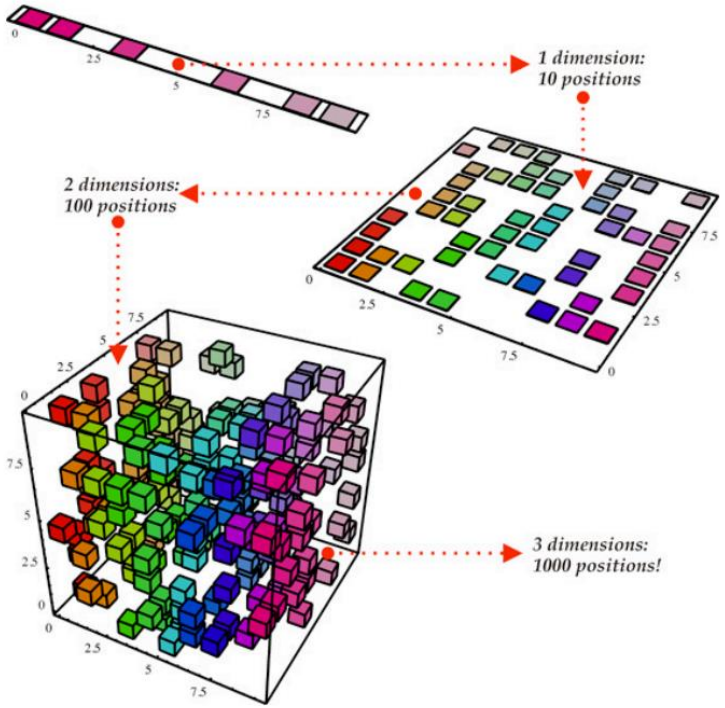

Recommender System



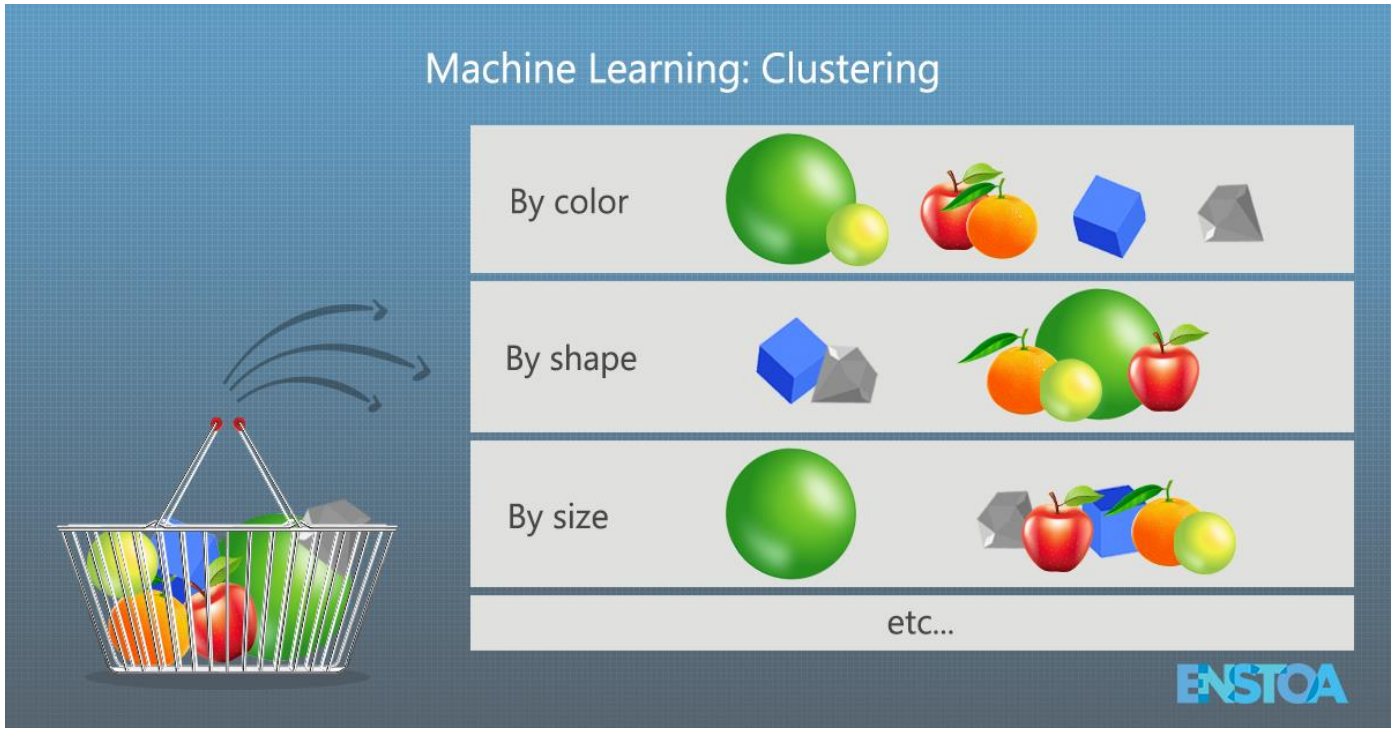

Customer Segmentation
Health Insurance

# 3.2 Unsupervised Learning - Tasks

- **Dimensionality reduction** or **dimension reduction** is the process of reducing the number of random variables under consideration, and can be divided into feature selection and feature extraction.



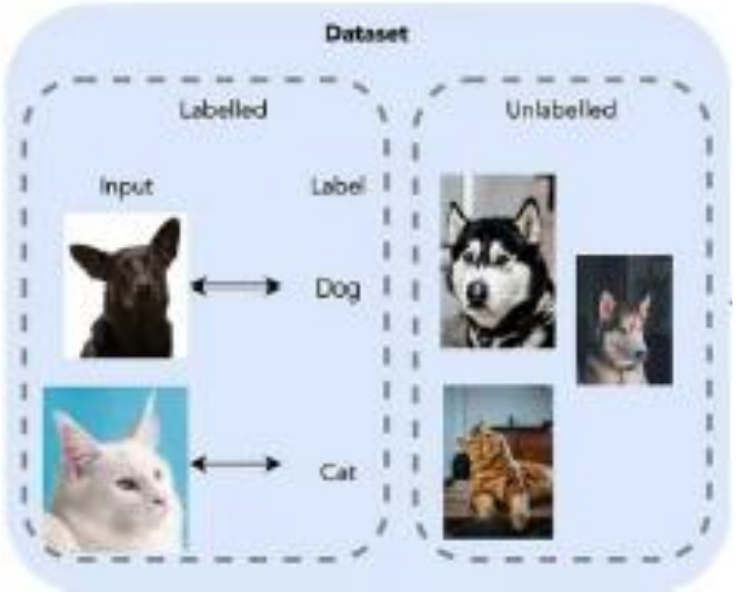
# 3.2 Unsupervised Learning - Tasks




- List down 5 clustering problems

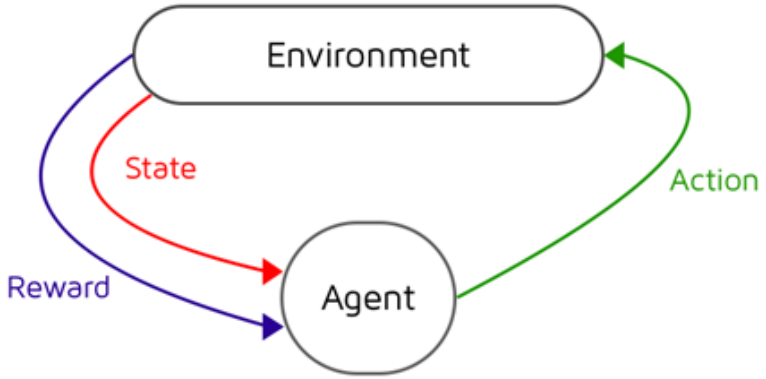
# 3.3 Semi-supervised Learning

- Semi-supervised Learning combines elements of both supervised and unsupervised learning. It uses a small amount of labeled data along with a larger amount of unlabeled data to improve the training model's performance.
- It is often used when obtaining labeled data is expensive or time-consuming.



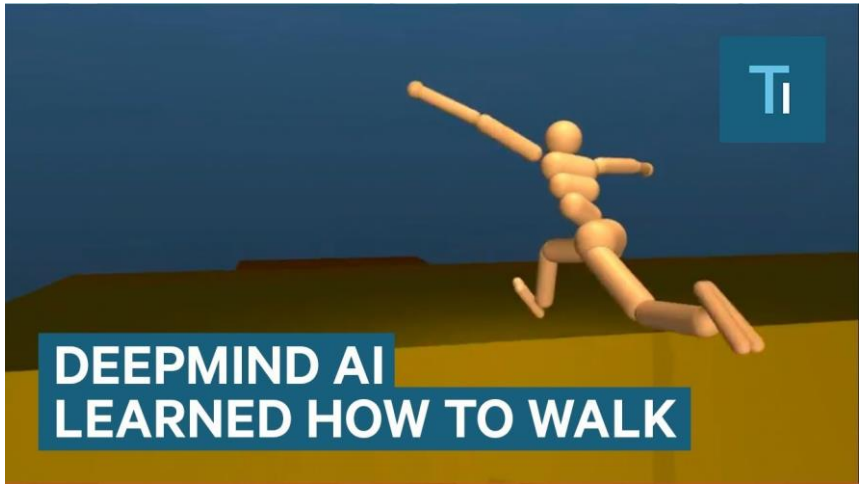
# 3.4 Reinforcement Learning

- **Reinforcement Learning** is a feedback-based learning method, in which a learning agent gets a reward for each right action and gets a penalty for each wrong action.
- The agent learns automatically via feedback from the actions during interacting and exploring the environment. The goal of the agent is to get the most reward points, and hence, it improves its performance.



# 3.4 Reinforcement Learning

- Examples




Learning to Walk



Playing Chess

# 3.4 Reinforcement Learning

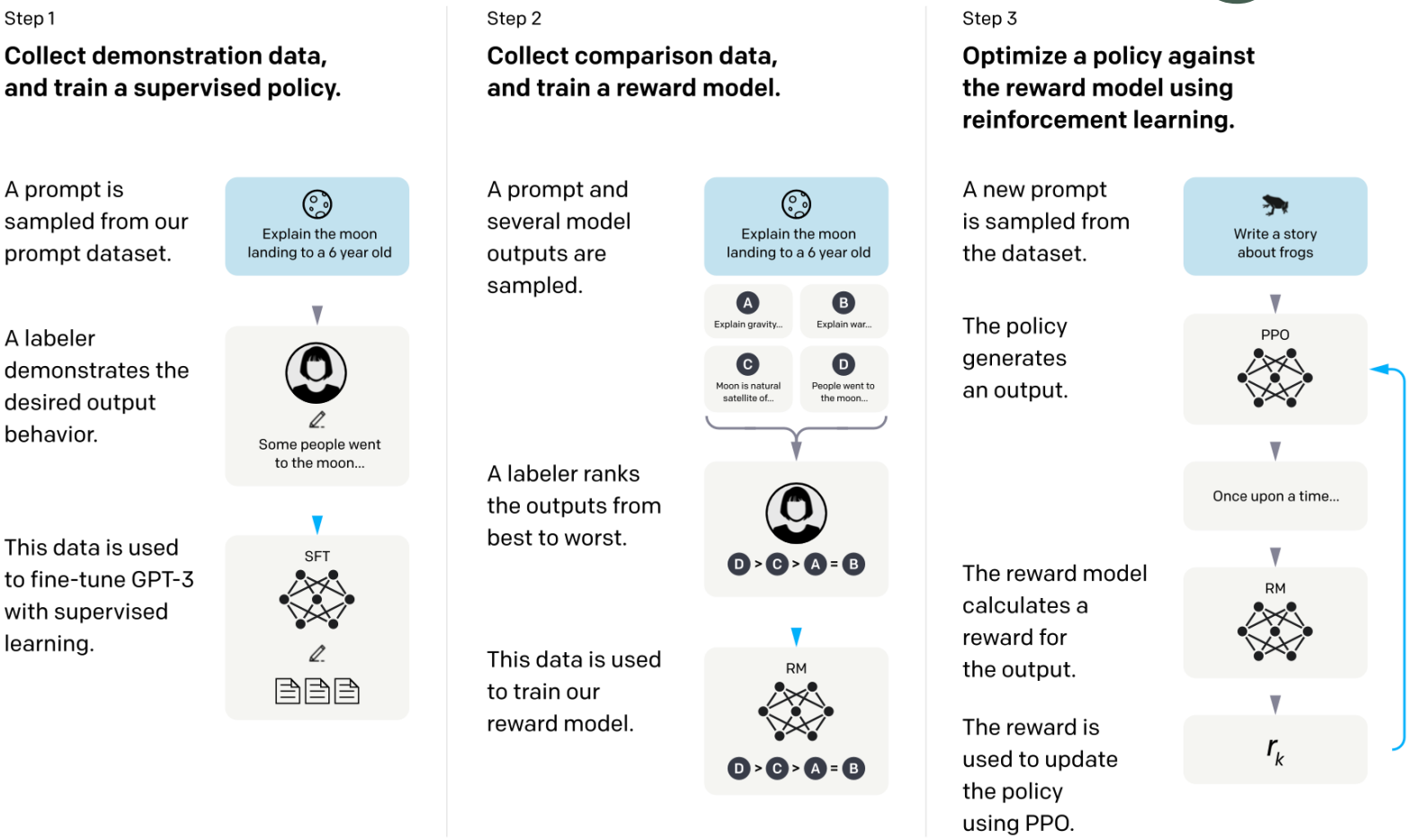
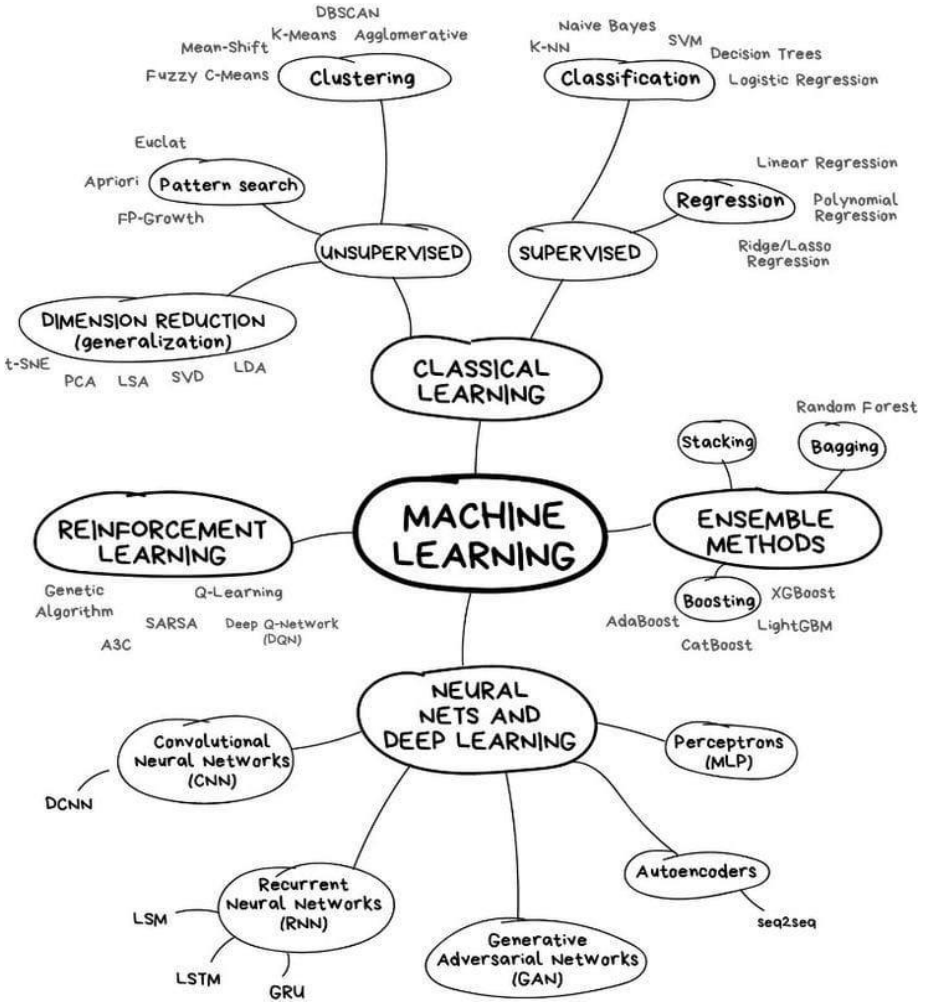
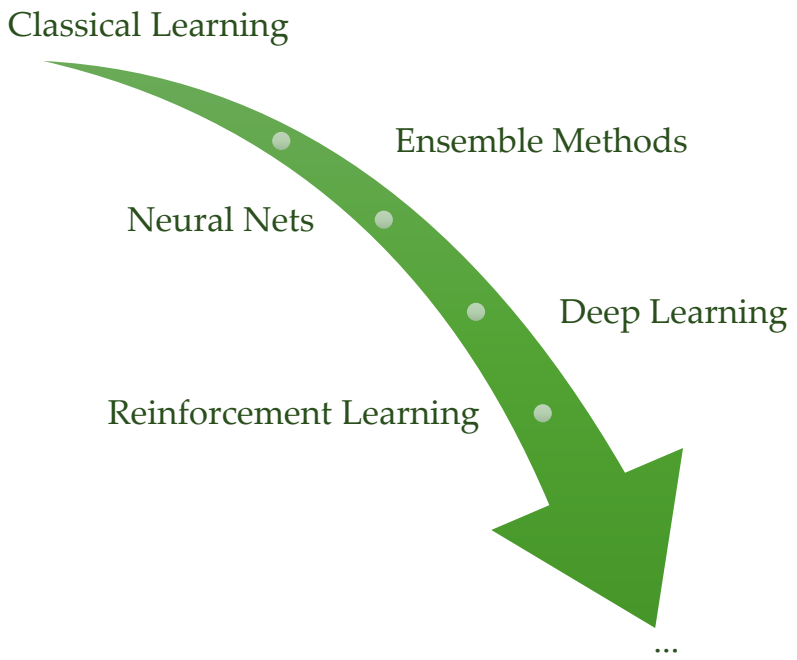



Image Source: OpenAI. (n.d.). Retrieved from https://openai.com/blog/chatgpt

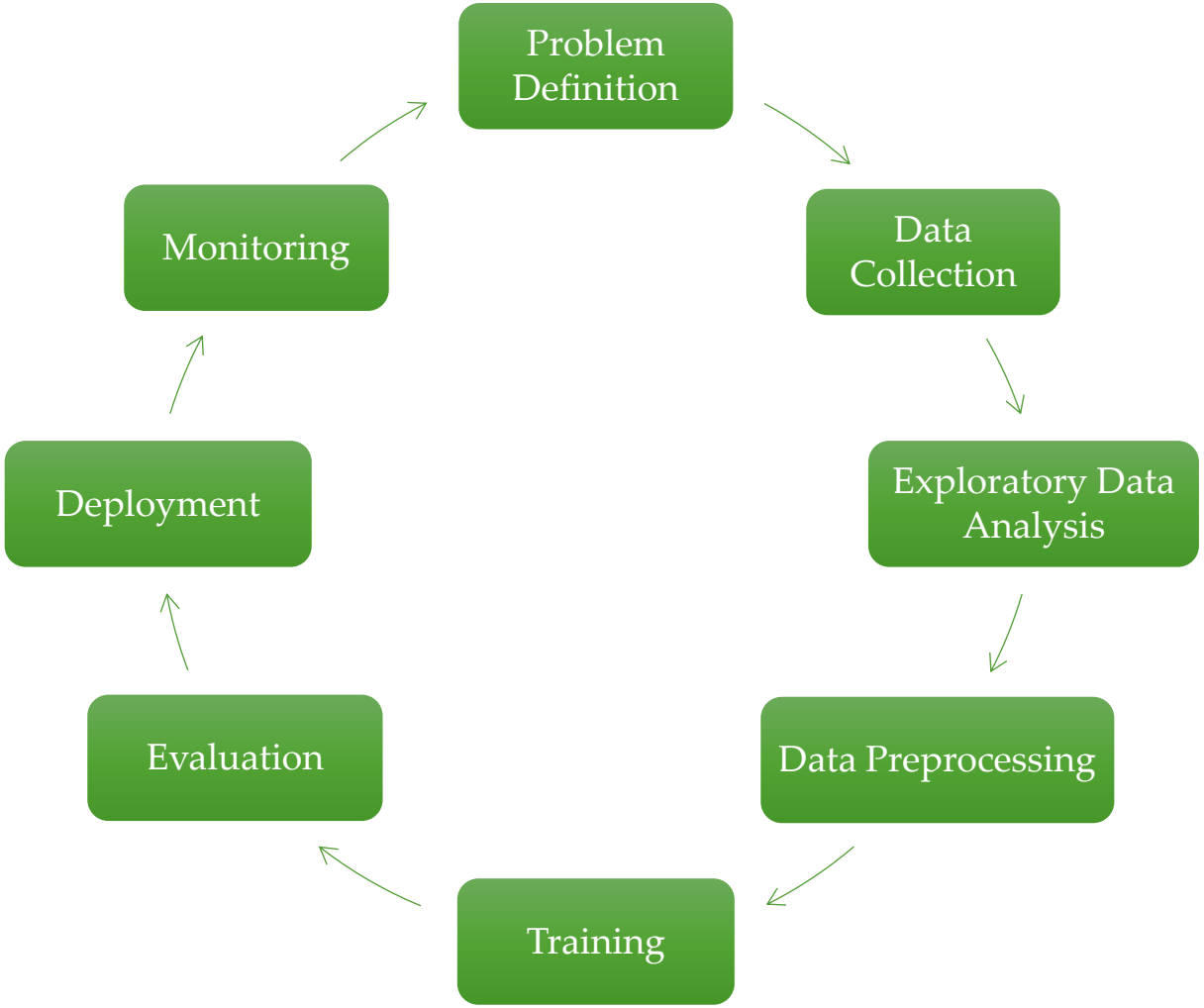
# 4. Machine Learning Algorithms




Recommended Learning Curve

# 5. Machine Learning Life Cycle



# 5.1 Problem Definition

- Before starting Machine Learning project, we first need to understand and analyze the business problem and then try to determine how it can be formulated into Machine Learning task.
- This step consists of:
  - Identifying what we have and what we want to predict.
  - Defining Machine Learning task with respect to our problem.
  - Defining what kind of the data we need (i.e. the input and output variables).
  - Identifying the assumptions and constraints that exist in our problem.
  - Etc.

# 5.2 Data Collection

- Data Collection is the process of gathering and measuring information from countless different sources. In order to use the data we collect to develop machine learning solutions, it must be collected and stored in a way that makes sense for the business problem at hand.
- This step includes:
  - Identifying various data sources (i.e. survey, web, file, database).
  - Collecting the data from the sources and annotating/labeling them.
  - Integrating the data obtained from different sources.
- After these tasks, we obtain a coherent data - **dataset** which will be used in further steps.

# 5.3 Exploratory Data Analysis

- Exploratory Data Analysis (EDA) refers to the process of performing initial investigations on data to discover patterns and characteristics, to spot anomalies, and to check assumptions with the help of summary statistics and graphical representations.
- Descriptive Statistics is one of the popular techniques of EDA to understand the data by using graphs (i.e. bar chart, histogram, scatter plot, box plot).

# 5.4 Data Preprocessing

- Data Preprocessing/Data Wrangling is the process of converting raw data into a useable format. It involves cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step.
- In real-world applications, collected data may have various issues, including:
  - Missing Values
  - Duplicated Data
  - Outliers/Noise
  - Imbalance Data
  - Invalid Data
  - ...

# 5.5 Training

- This step aims to build inference models of Machine Learning, which are obtained through training learning algorithms (i.e. Linear Regression, Logistic Regression…) with a given dataset with respect to the problem.
- This step involves:
  - Choosing learning algorithms and their hyperparameters based on the problem.
  - Preparing the dataset which are useful for training and testing.
  - Defining the objective function, optimizer…, which respect to the learning algorithms.
  - Training the learning algorithms to obtain the models.

# 5.6 Evaluation

- Once the model has been trained with a given learning algorithm, we can then evaluate the model with the new unseen data to determine how good the model is and whether it responds to the problem requirement.
- In order to evaluate the model, the evaluation measures or metrics are defined with respect to the problem (i.e. Classification, Regression, Clustering…).

# 5.7 Deployment

- Deployment is the process involving how our model is integrated and deployed into real-world system such as enterprise cloud services (i.e. AWS, Google App Engine) or simple web service.
- Before the deployment can be done, we need to make sure our model produces well enough accurate result in term of generalization of the problem as per requirement.

# 5.8 Monitoring

- In certain cases, monitoring might not be implemented. The goal of this step is to check the progress of the model during the production to verify if everything is acceptable with respect to what we defined during problem requirement.
- For future development, it is better to continue collecting and validating new data so that it can be used in the next phase for improving or tuning our baseline model.