

Machine Learning for Finance

Prepared by Dr. Diaa Salama

FNT633
Orientation Week

Lecture 1





مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY

Machine Learning for Finance

Lecture 1

FNT633
Intro

Prepared by
Dr. Diao Salama

www.midocean.ae

- **Associate Professor, Data Science Department**, Faculty of Computers Science , Misr International University, **On Leave from Faculty of Computers & Artificial Intelligence, Banha University, Banha, Egypt.**
- **Member**, National Committee of Communications and Information Technology, Academy of Scientific Research & Technology, CAIRO, Egypt.
- **Member**, International Telecommunication Union (ITU), United Nations (UN)
- **Member**, International Federation for Information Processing - IFIP
- **Winner of** the Arab Republic of **Egypt Encouragement Award** for Advanced Science and Technology in the Field of Engineering Sciences **2021**
- **Winner of** the **Benha University Encouragement Award** for Advanced Science and Technology in the Field of Engineering Sciences **2019**
- **Winner** of the DELL EMC AWARD for the academic year 2015/2016, 2016/2017, and 2017/2018



Dr.Diao Salama Ph.D

One of the top 2% Scientists worldwide 2023 by Stanford university ranking.

Email: diaasalama@midocean.edu.km

Week

02

Machine Learning

Today

- What is machine learning?
- Supervised learning
- Unsupervised learning

Week

02

Machine Learning

What is machine learning?

Machine learning!

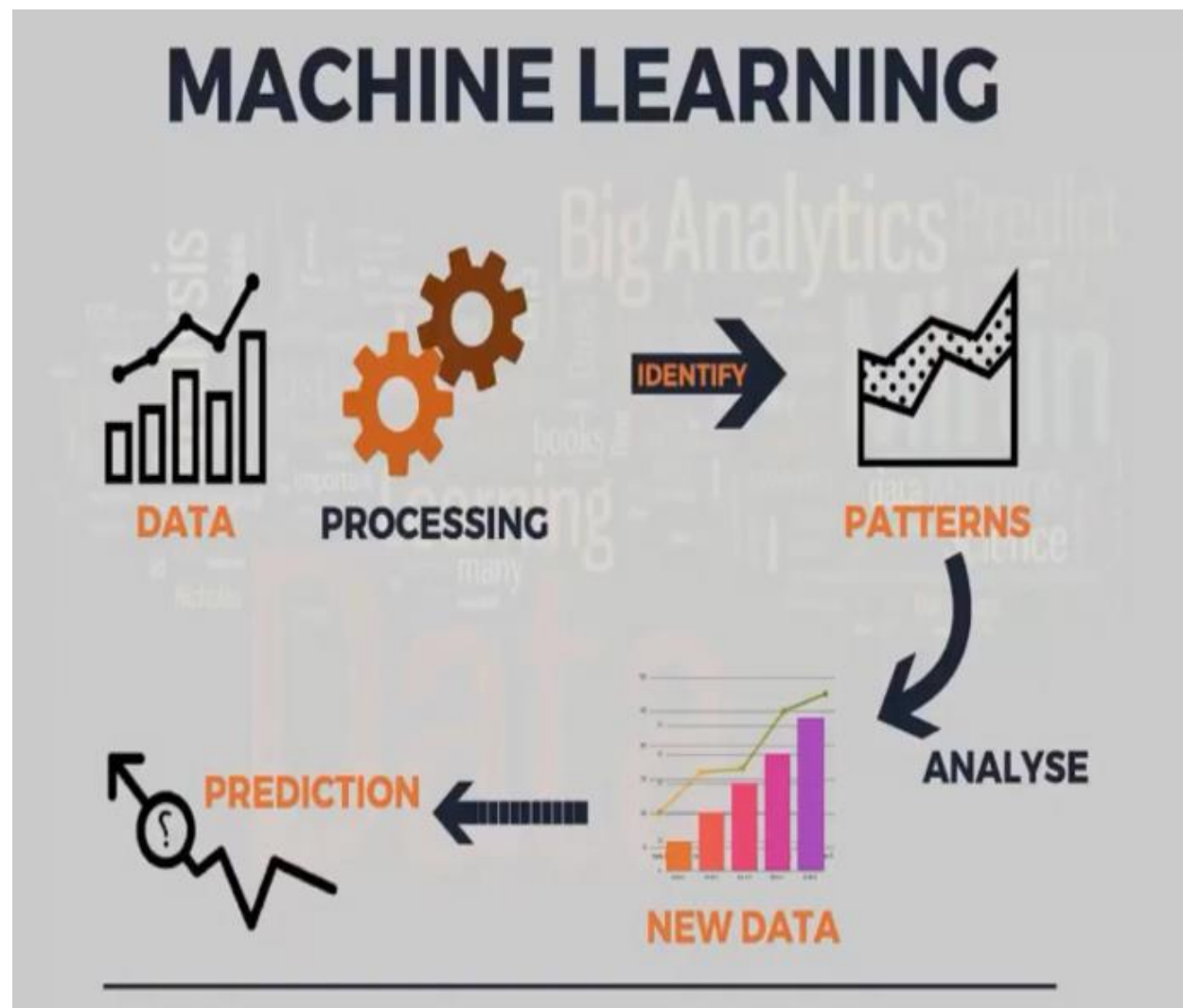
- Give the machine the ability how to solve some problems (e.g. statistical phenomenon) based on **incomplete information**
- Here, learning is not only a question of remembering but also of generalization to unseen cases
- Learn by example

Week

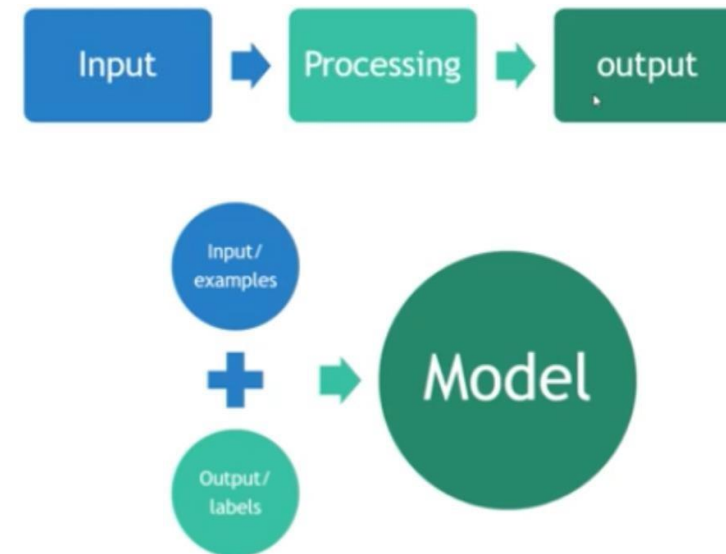
02

Machine Learning Definition:

- Arthur Samuel (1959).
 - Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed



- A branch of artificial intelligence, concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.
- As intelligence requires knowledge, it is necessary for the computers to acquire knowledge.
- Application
 - Finance
 - Image processing
 - Computer vision (Google Car)
 - NLP (monitor Twitters)
 - Speech Recognition
 - Bioinformatics



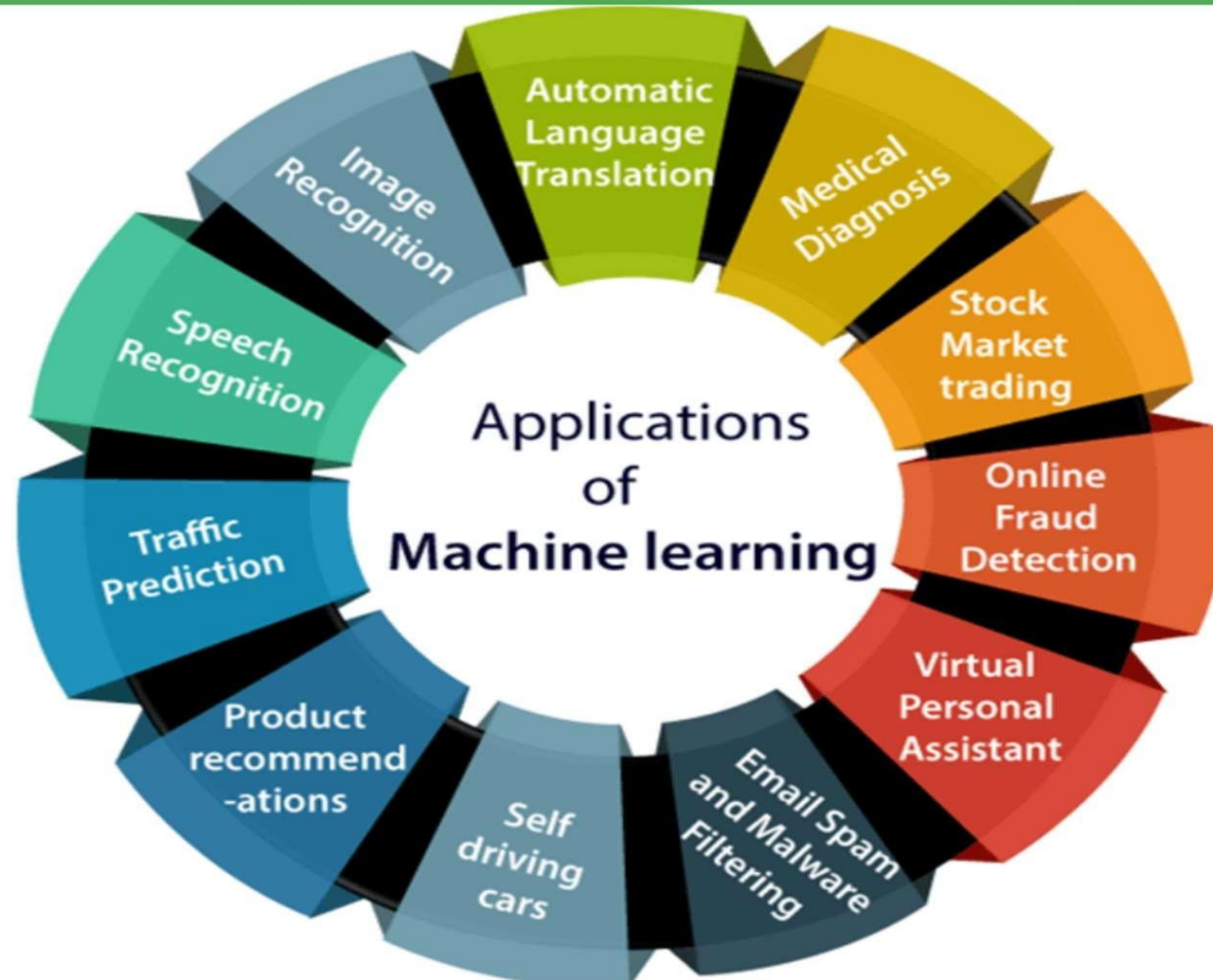
■ A Traditional Programming:



■ Machine Learning:



Applications of Machine learning



Applications of Machine learning



how do i m



how do i **make**

how do i **make sushi rolls**

how do i **make sushi rice**

how do i **get home**

how do i **get a passport**

Week

02

Applications of Machine learning

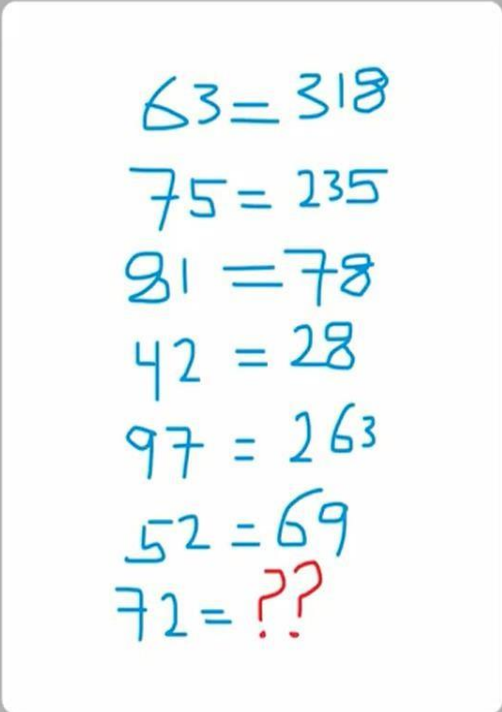


Week

02

Applications of Machine learning



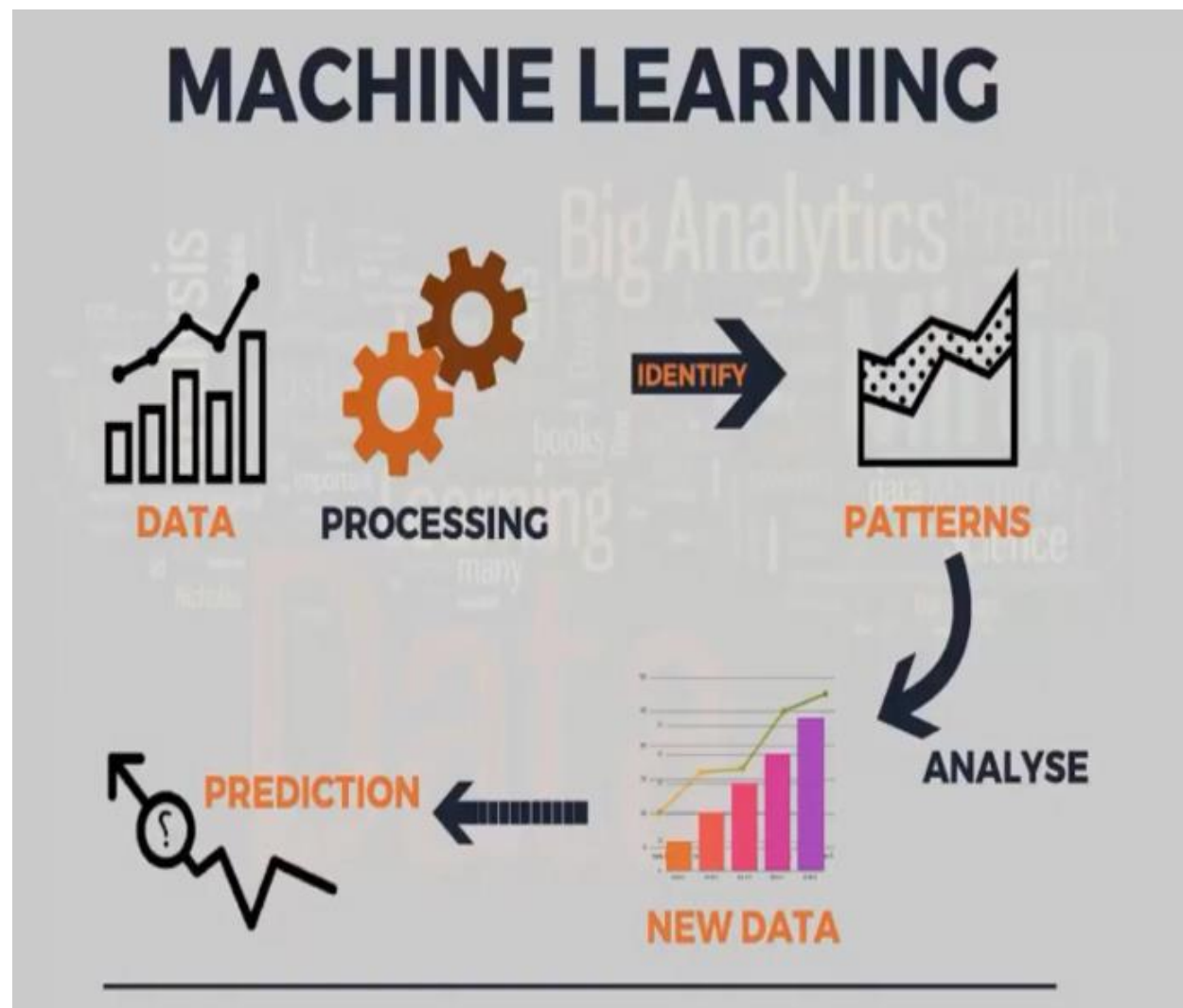


Handwritten numbers on a whiteboard:

$$\begin{array}{l} 63 = 318 \\ 75 = 235 \\ 81 = 78 \\ 42 = 28 \\ 97 = 263 \\ 52 = 69 \\ 72 = ?? \end{array}$$

Big Picture

- ≡ Dataset
- ≡ Find pattern that best match
- ≡ Problem in pattern
- ≡ predict



LEARNING

Acquisition of knowledge or skills through EXPERIENCE ...

Experience is EXAMPLES

Find Missing Number

$$\text{Red Apple} + \text{Red Apple} + \text{Red Apple} = 90$$

$$\text{Red Apple} + \text{Two Green Apples} + \text{Two Green Apples} = 230$$

$$\text{Yellow Apple} + \text{Two Green Apples} + \text{Red Apple} = 210$$

$$\text{Green Apple} - \text{Red Apple} + \text{Yellow Apple} = ??$$

Google

LEARNING

$$3 \text{ (red apple)} = 90 \Rightarrow 1 \text{ (red apple)} = 30$$

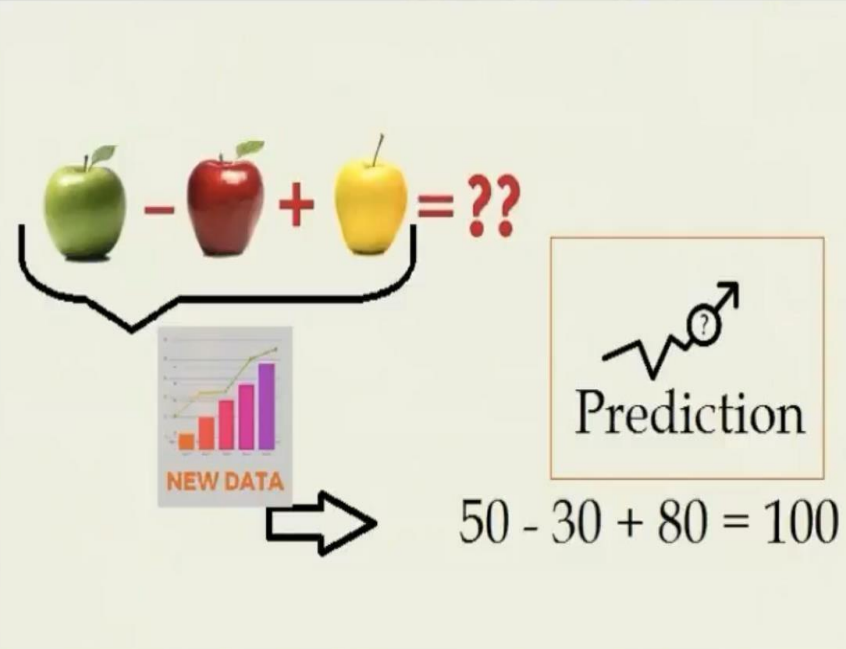
$$1 \text{ (red apple)} + 2 \text{ (green apple)} = 230 \Rightarrow 1 \text{ (green apple)} = 50$$

$$1 \text{ (yellow apple)} + 1 \text{ (green apple)} + 1 \text{ (red apple)} = 210 \Rightarrow 1 \text{ (yellow apple)} = 80$$

$$1 \text{ (green apple)} - 1 \text{ (red apple)} + 1 \text{ (yellow apple)} = ?? \Rightarrow 50 - 30 + 80 = 100$$

Google

LEARNING

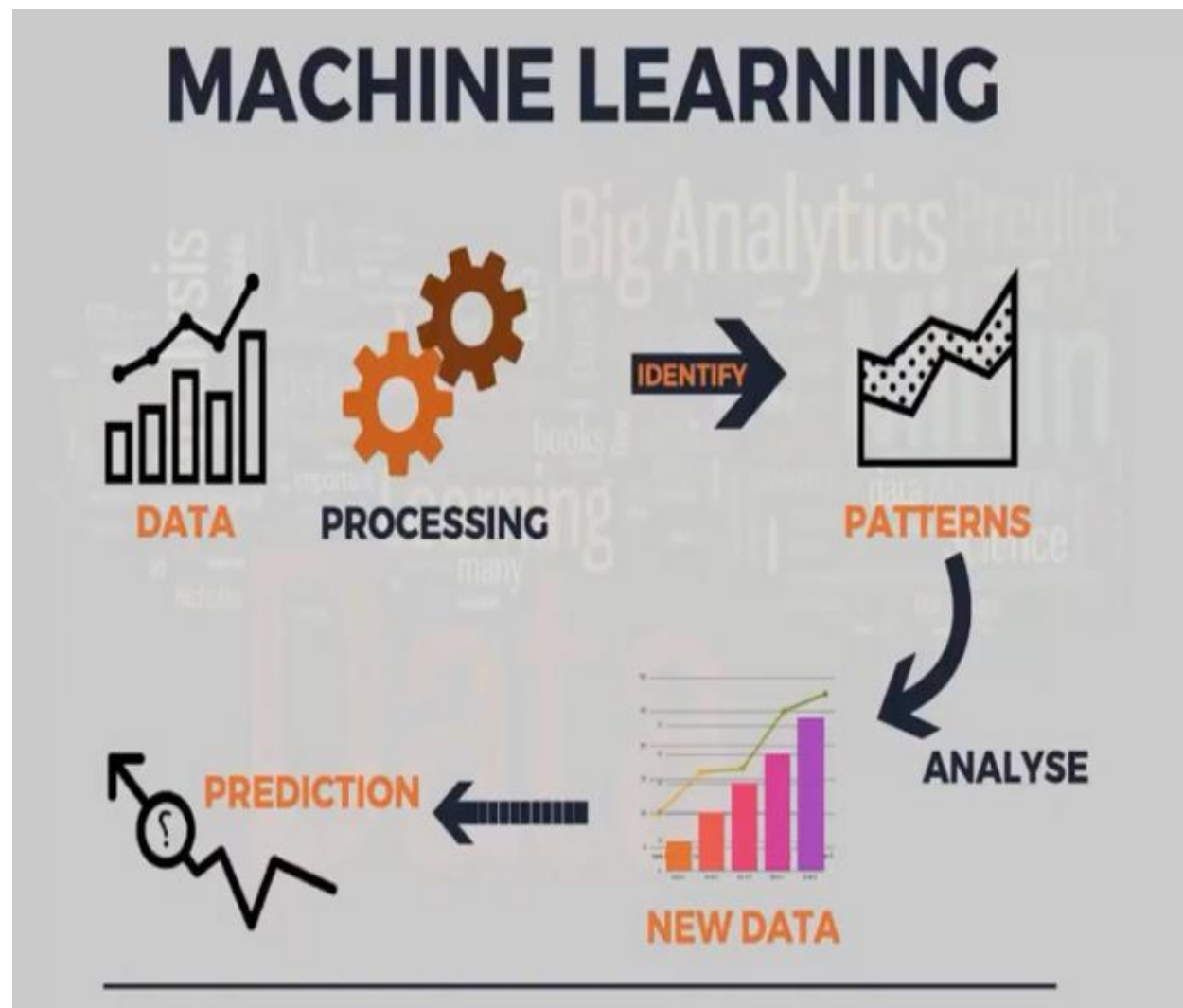


$\text{Green Apple} - \text{Red Apple} + \text{Yellow Apple} = ??$

NEW DATA

Prediction

$50 - 30 + 80 = 100$

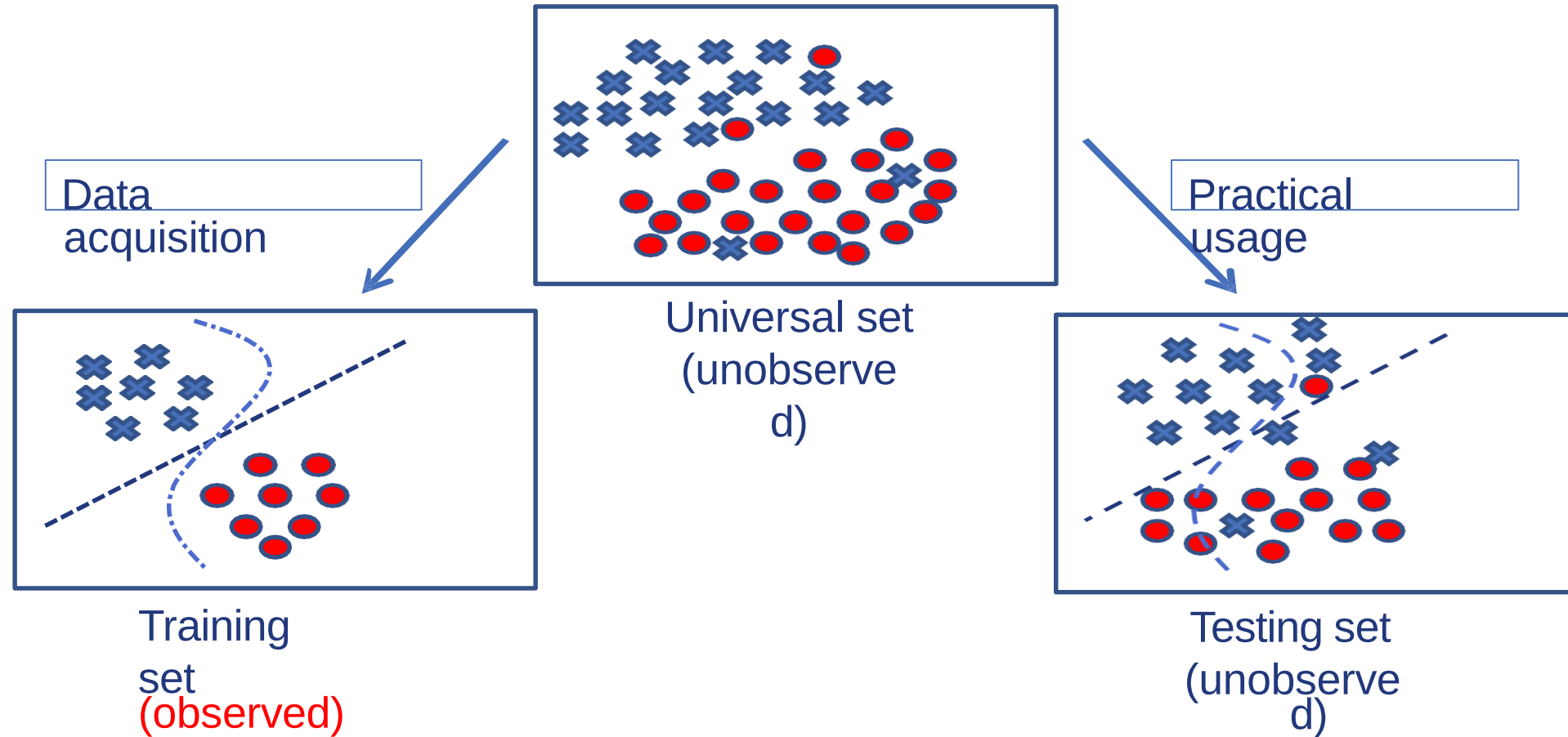


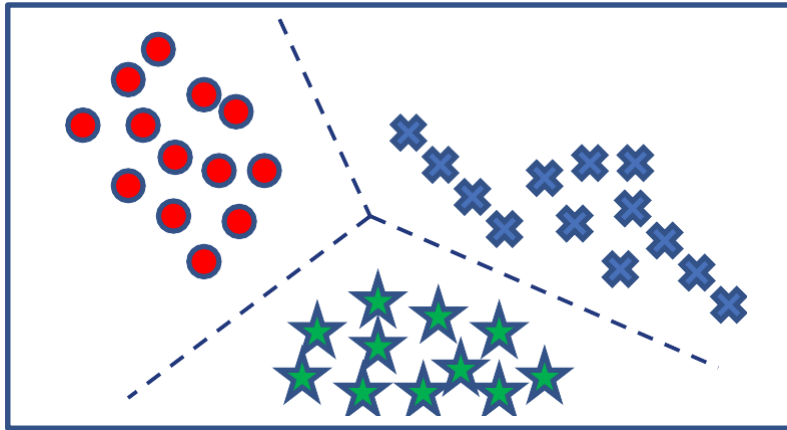
Week

02

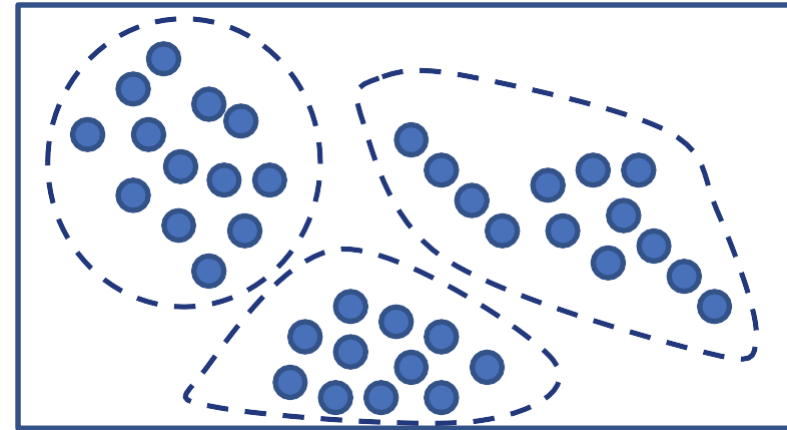
Master of Financial Technology

- After completion of the program, the graduate becomes eligible to
- The Master of Financial Technology focuses on developing FinTech services.
 - The program emphasizes the systematic understanding of new digital business models in financial services,
 - how innovation through technology is changing business practices,
 - quantitative finance skills, and the application of emerging technologies to financial services.





Supervised learning



Unsupervised learning

Week

02

Machine learning Algorithms Types

- Supervised learning
- Unsupervised learning
- Others: Reinforcement learning, recommender systems.

Machine Learning Types

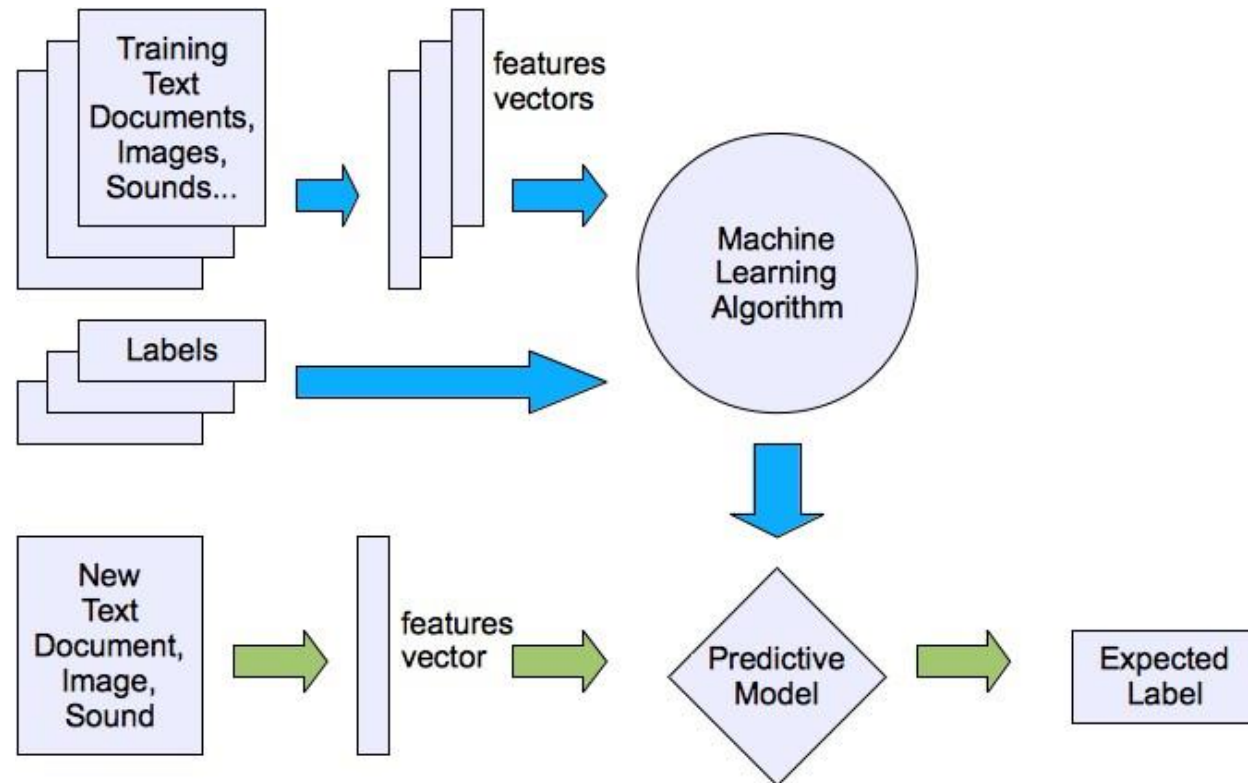
- Suppose you are building a Face recognition program
- Supervised learning (Classification):
 - You have a dataset of faces image and other images.
 - The dataset is labeled (classified) into Faces images, and other images.
 - Program task know is to classify new images based on the available dataset.
- Unsupervised learning (Clustering):
 - You don't have a dataset, and program role know is to cluster similar images for a new given dataset into different group, e.g. it can distinguish that faces are very different from landscapes, which are very different from horses.

Week

02

Machine learning structure

■ Supervised learning



Week

02

Supervised learning:

- After Input (x) and correct output (y) are given.
- Used for Regression and classification problems.

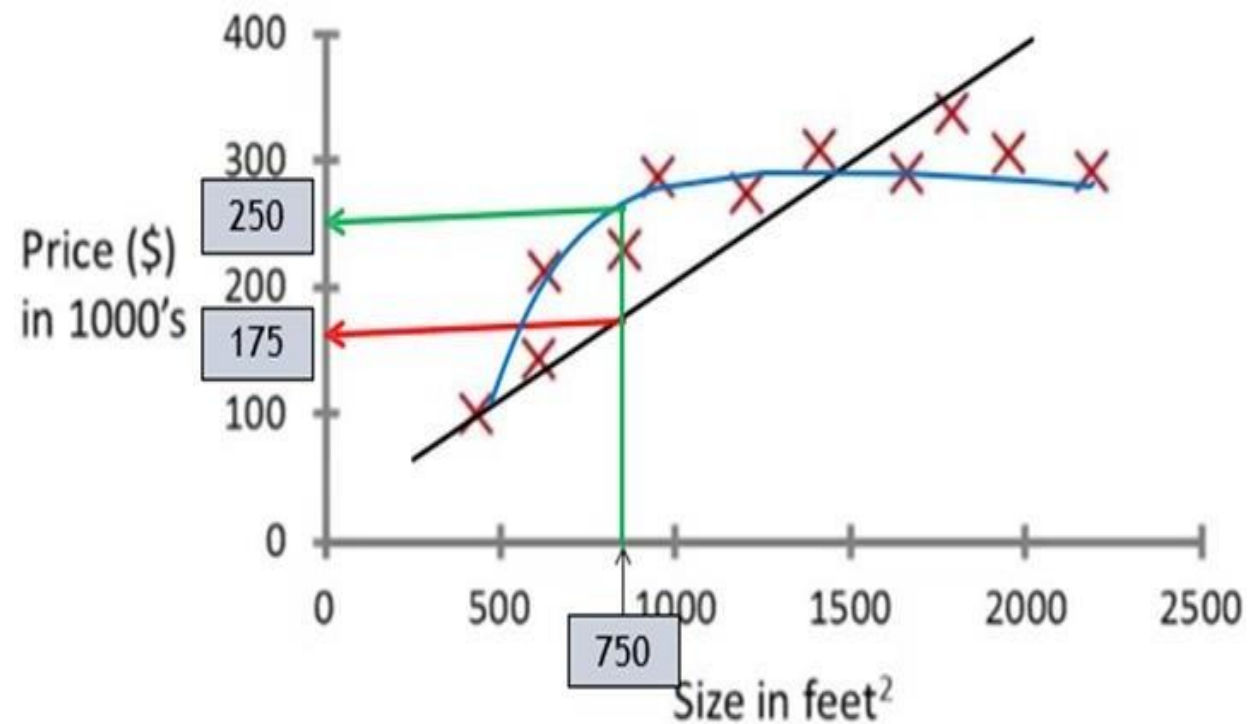
Input (x)	Output (Y)	Application
email	spam? (0/1)	spam filtering
audio	text transcripts	speech recognition
English	Spanish	machine translation

Week

02

Regression:

- Predict continuous valued output.
- Many possible outputs.
- Example: Housing price prediction

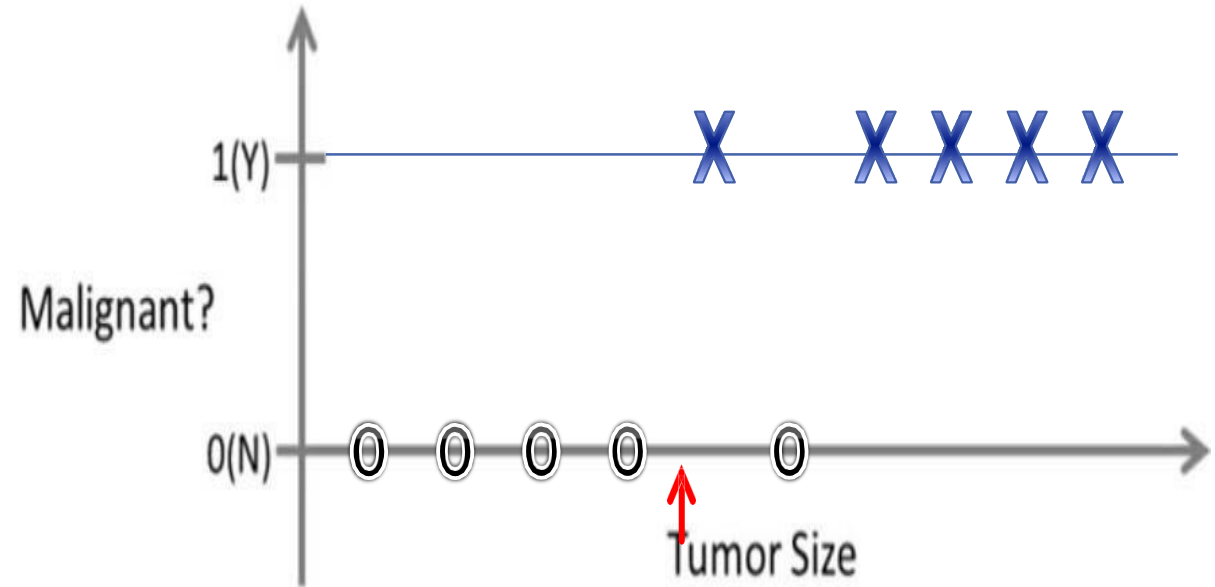


Week

02

Classification:

- Discrete valued output (0 or 1)
- Classification problem: small number of possible outputs.
- Example: Breast cancer (malignant, benign)
- only one feature or one attribute, mainly, the tumor size is used in order to predict whether the tumor is malignant or benign

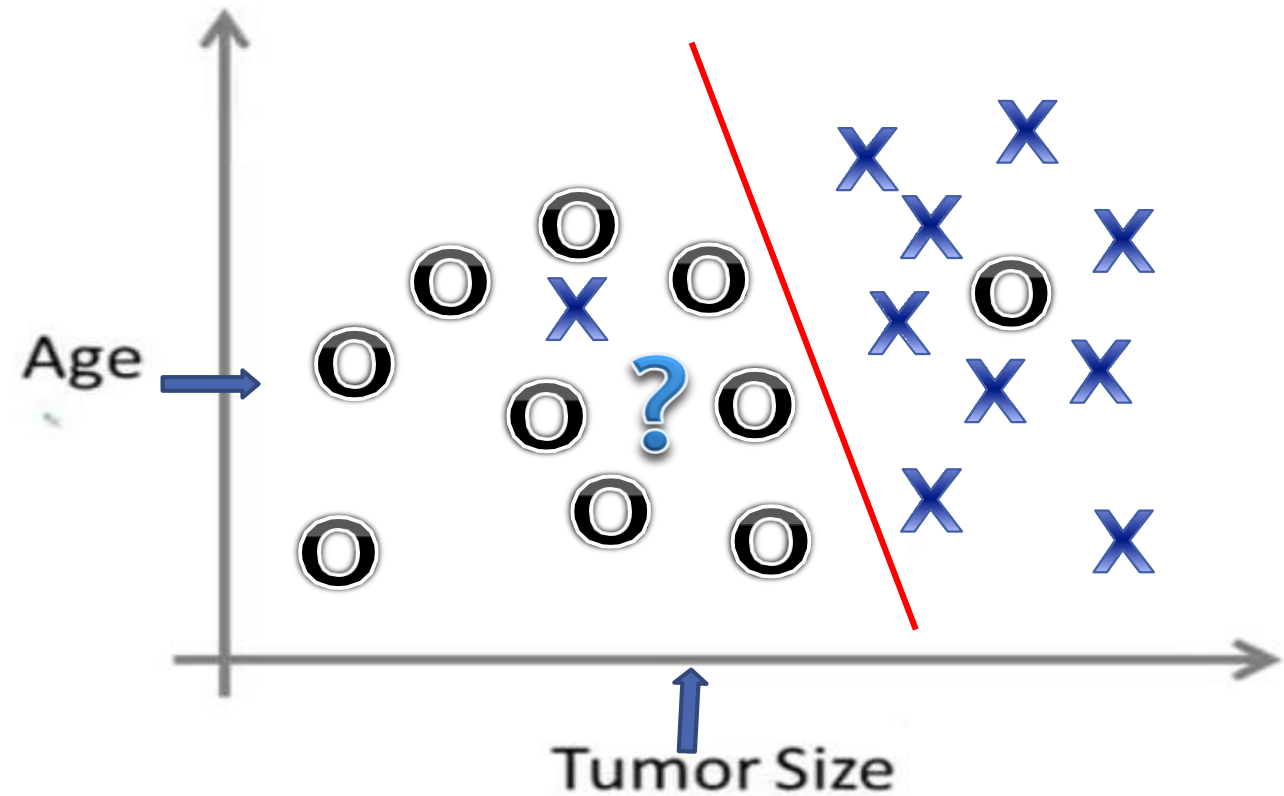


Week

02

Classification:

- Two or more inputs and one output (Data set)



³₁Example:

- You're running a company, and you want to develop learning algorithms to address each of two problems.
- Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
- Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.
- Should you treat these as classification or as regression problems?
 - Treat both as classification problems.
 - Treat problem 1 as a classification problem, problem 2 as a regression problem.
 - Treat problem 1 as a regression problem, problem 2 as a classification problem.
 - Treat both as regression problems.

Week

02

Example

- After You're running a company, and you want to develop learning algorithms to address each of two problems.
- Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
- Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

Week

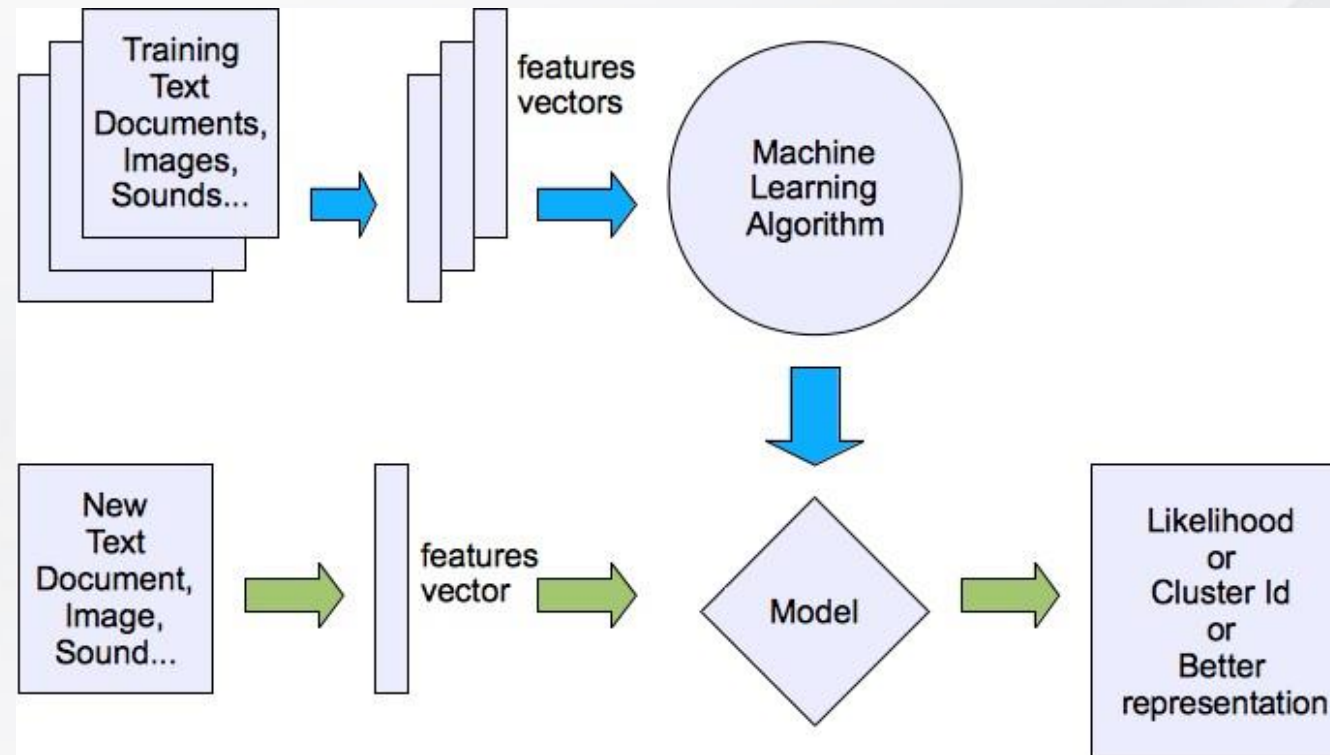
02

Example

- Should you treat these as classification or as regression problems?
 - Treat both as classification problems.
 - Treat problem 1 as a classification problem, problem 2 as a regression problem.
 - Treat problem 1 as a regression problem, problem 2 as a classification problem.
 - Treat both as regression problems.
 - how innovation through technology is changing business practices,
 - quantitative finance skills, and the application of emerging technologies to financial services.

Machine learning structure

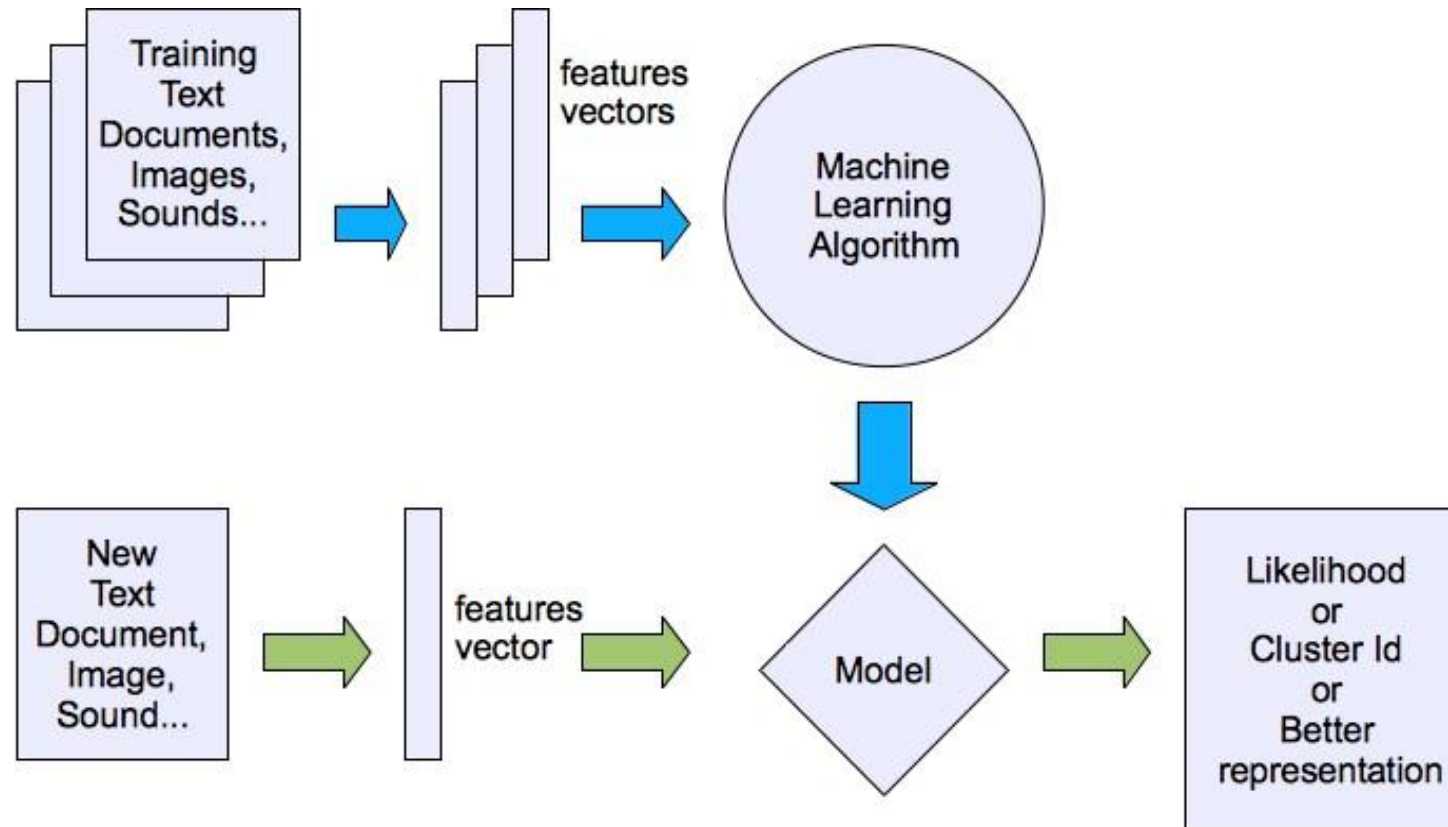
❖ Unsupervised learning



Week

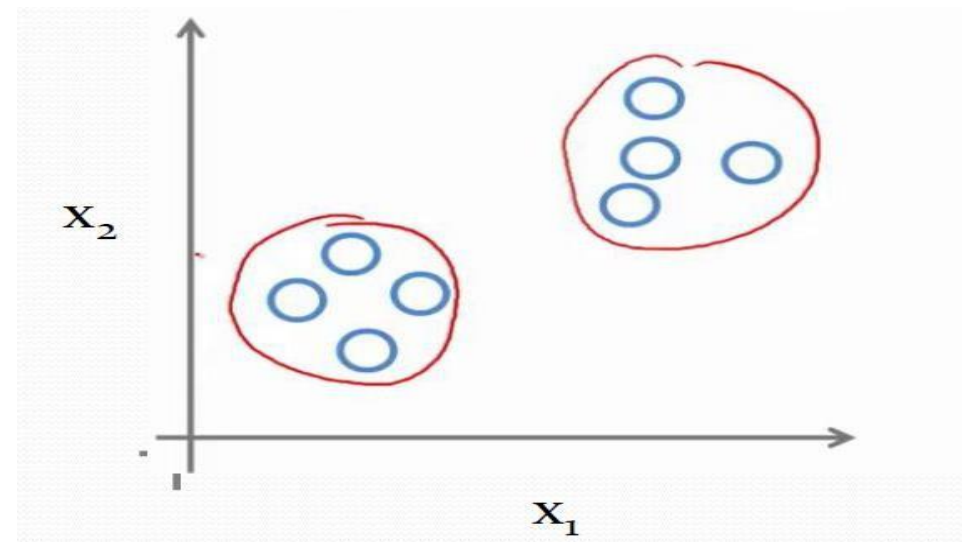
02

Unsupervised learning



Unsupervised Learning:

- Unsupervised learning is where you only have input data and no corresponding output variables.
- Data only comes with inputs x , but not output labels y .
- Algorithm has to find structure in the data.
- Used for Clustering.



Application of Unsupervised Learning: Example: Clustering Google news

Giant panda gives birth to rare twin cubs at Japan's oldest zoo

USA TODAY · 6 hours ago

- Giant panda gives birth to twin cubs at Japan's oldest zoo

CBS News · 7 hours ago

- Giant panda gives birth to twin cubs at Tokyo's Ueno Zoo

WHBL News · 16 hours ago

- A Joyful Surprise at Japan's Oldest Zoo: The Birth of Twin Pandas

The New York Times · 1 hour ago

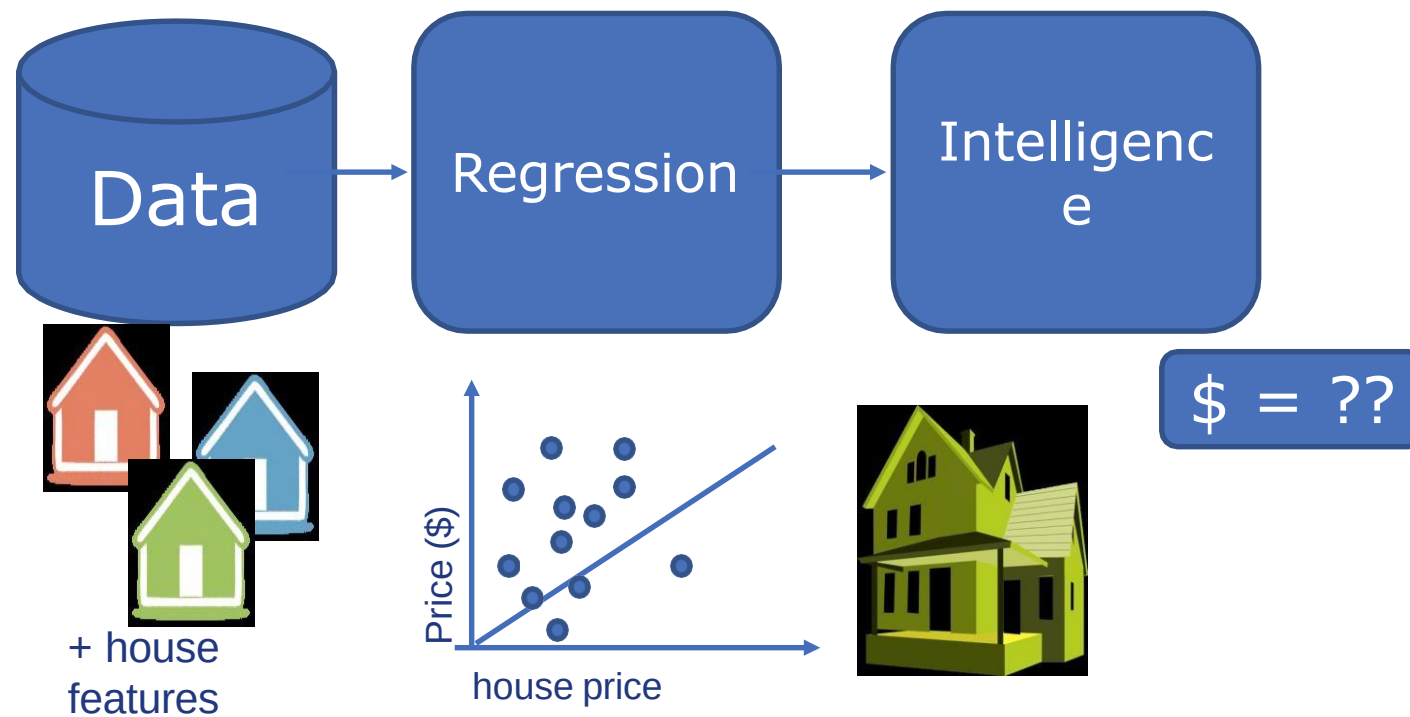
- Twin Panda Cubs Born at Tokyo's Ueno Zoo

PEOPLE · 6 hours ago

 [View Full Coverage](#)



- Case Study 1 :
- Predicting house prices



- Case Study 1 :
- Predicting house prices



- Case Study 1 :
- Predicting house prices

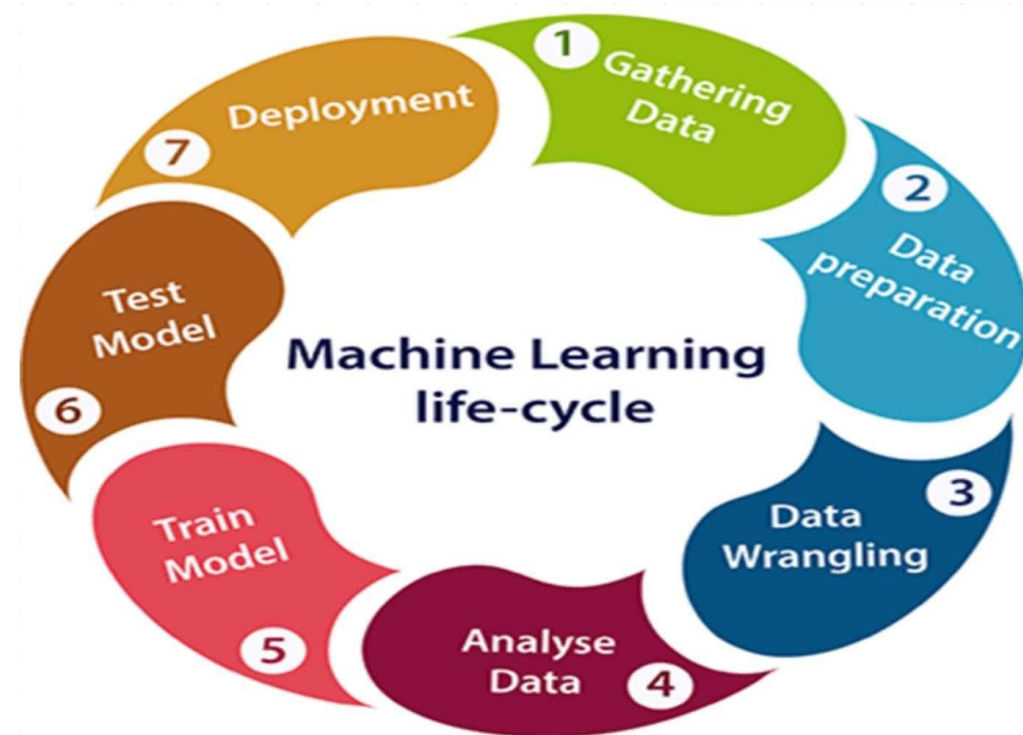


- Case Study 1 :
- Predicting house prices



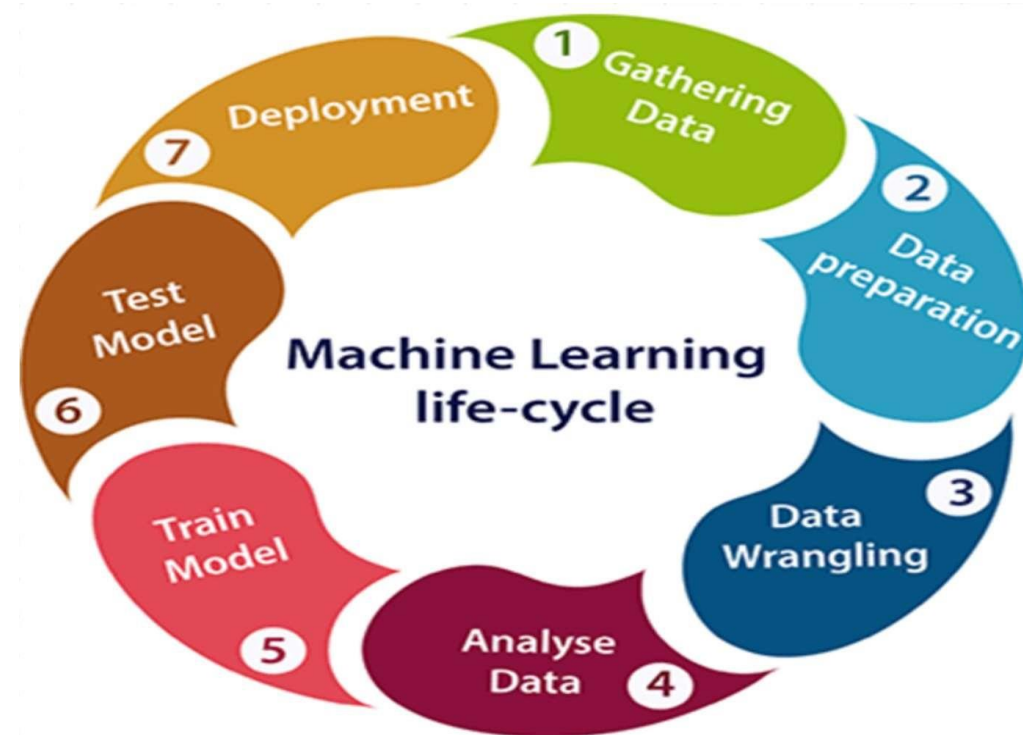
Machine learning Life cycle

- The most important thing in the complete process is to understand the problem and to know the purpose of the problem. Therefore, before starting the life cycle, we need to understand the problem because the good result depends on the better understanding of the problem.



- The Machine learning life cycle involves seven major steps, which are given below:

- Gathering Data
- Data preparation
- Data Wrangling
- Analyze Data
- Train the model
- Test the model
- Deployment



1. Gathering Data

- It is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.
- In this step, we need to identify the different data sources,
 - as data can be collected from various sources such as files, database, internet, or mobile devices.It is one of the most important steps of the life cycle.
- The quantity and quality of the collected data will determine the efficiency of the output. The more will be the data, the more accurate will be the prediction.
- By performing the above task, we get a coherent set of data, also called as a dataset. It will be used in further steps.

2. Data preparation

- The data is explored, putted into a suitable place and prepared to
- use in our machine learning training. In this step, we understand
- the nature of data that we must work with. We need to understand the characteristics, format, and quality of data.
- A better understanding of data leads to an effective outcome. In this, we find [Correlations](#), [general trends](#), and [outliers](#).
- [3. Data Wrangling \(Cleaning\)](#)
- Data wrangling is the process of cleaning and converting raw data into a useable format.
- In real-world applications, collected data may have various issues,

including:

- [Missing Values](#)
- [Duplicate data](#)
- [Invalid data](#)
- [Noise](#)

So, we use various filtering techniques to clean the data.

2. Data preparation

- The data is explored, putted into a suitable place and prepared to
- use in our machine learning training. In this step, we understand
- the nature of data that we must work with. We need to understand the characteristics, format, and quality of data.
- A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.
- Data wrangling is the process of cleaning and converting raw data into a useable format.
- In real-world applications, collected data may have various issues

Week

02

3. Data Cleaning

- including:
 - Missing Values
 - Duplicate data
 - Invalid data
 - Noise
 - So, we use various filtering techniques to clean the data.

Week

02

4. Data Analysis

- including:
 - We use the cleaned and prepared data, in this step which involves:
 - Data Visualization
 - Selection of the machine learning technique (a machine learning model) such as Classification, Regression, Cluster analysis, Association, etc.
 - Building and evaluating the model using the prepared data
 - Review the result

Week

02

5. Train Model

- including:
 - The model is trained to improve its performance for better outcome of the problem.
 - We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and features.

Week

02

6. Test Model

- We test the model and check for the accuracy of our model by providing a test dataset to it.
- Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

Week

02

7. Deployment

- The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system.

Week

01

References

1. *"Machine Learning for Asset Managers"* by Marcos López de Prado
2. "Advances in Financial Machine Learning" by Marcos López de Prado
3. "Machine Learning for Finance: Principles and practice for financial insiders" by Jannes Klaas
4. "Hands-On Machine Learning for Algorithmic Trading: Design and implement investment strategies based on smart algorithms that learn from data using Python" by Stefan Jansen
5. "Financial Machine Learning" by David Blazard
6. "Machine Learning and Data Sciences for Financial Markets: A Guide to Contemporary Practices" by Agostino Capponi and Charles-Albert Lehalle
7. "Python for Finance: Mastering Data-Driven Finance" by Yves Hilpisch



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You

