

# Python 機器學習



資展國際

楊碩蓉 博士  
[shoujungyang@gmail.com](mailto:shoujungyang@gmail.com)



# Abstract

1. 數據團隊角色
2. 使用工具
3. 數據分析流程
4. 機器學習簡介
5. 挑戰與運用



# 1. 數據團隊角色



## IT 部門

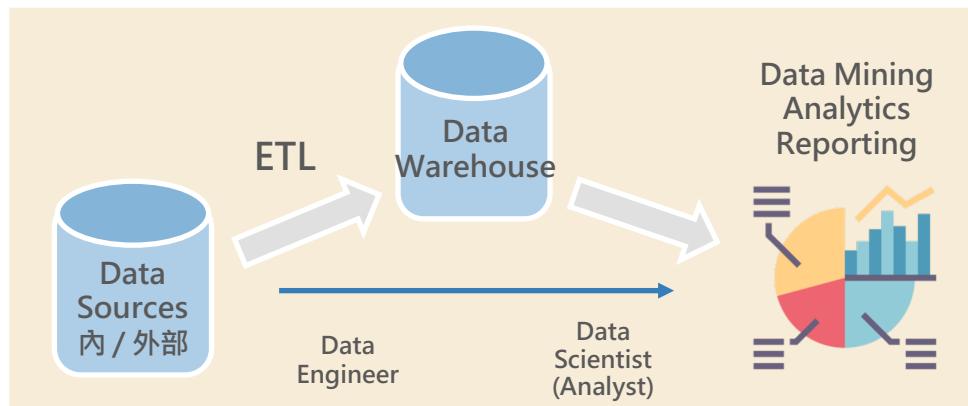
老闆  
Team Manager

數據工程師  
Data Engineer

資料科學家  
Data Scientist

Research track

Product Analytics  
track



數據分析師  
Data Analyst

領域專家  
Domain Expert

## 商業 部門

## 2. 使用工具

.....

## 數據工程師 Data Engineer

### Database systems

SQL & NoSQL

### Coding ability

Python

### ETL solutions

Airflow, Informatica

### Data warehouse software

Amazon Redshift, BigQuery

### Big Data Tools

Apache Hadoop, Spark

## 資料科學家 Data Scientist

### SQL Query

R / Python

Hadoop

Data visualization tools

Algebra, Statistics, ML, DL

Business acumen

Communication skills

## 數據分析師 Data Analyst

### SQL Query

Microsoft Excel

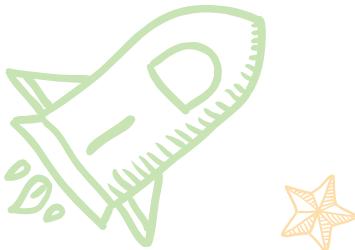
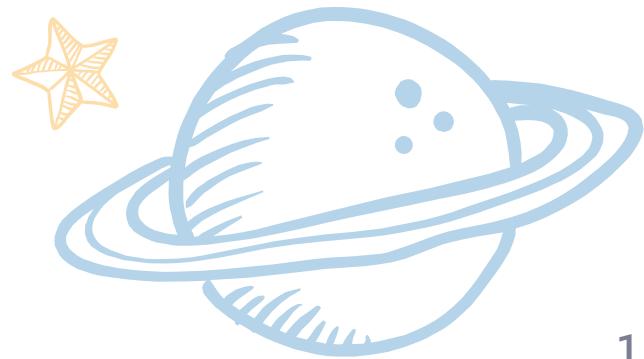
Data visualization tools

Statistics

Business acumen

Communication skills

Reference :MARYNA DEMCHENKO (2020). Data Engineer vs Data Scientist vs Data Analyst: What is the Difference?

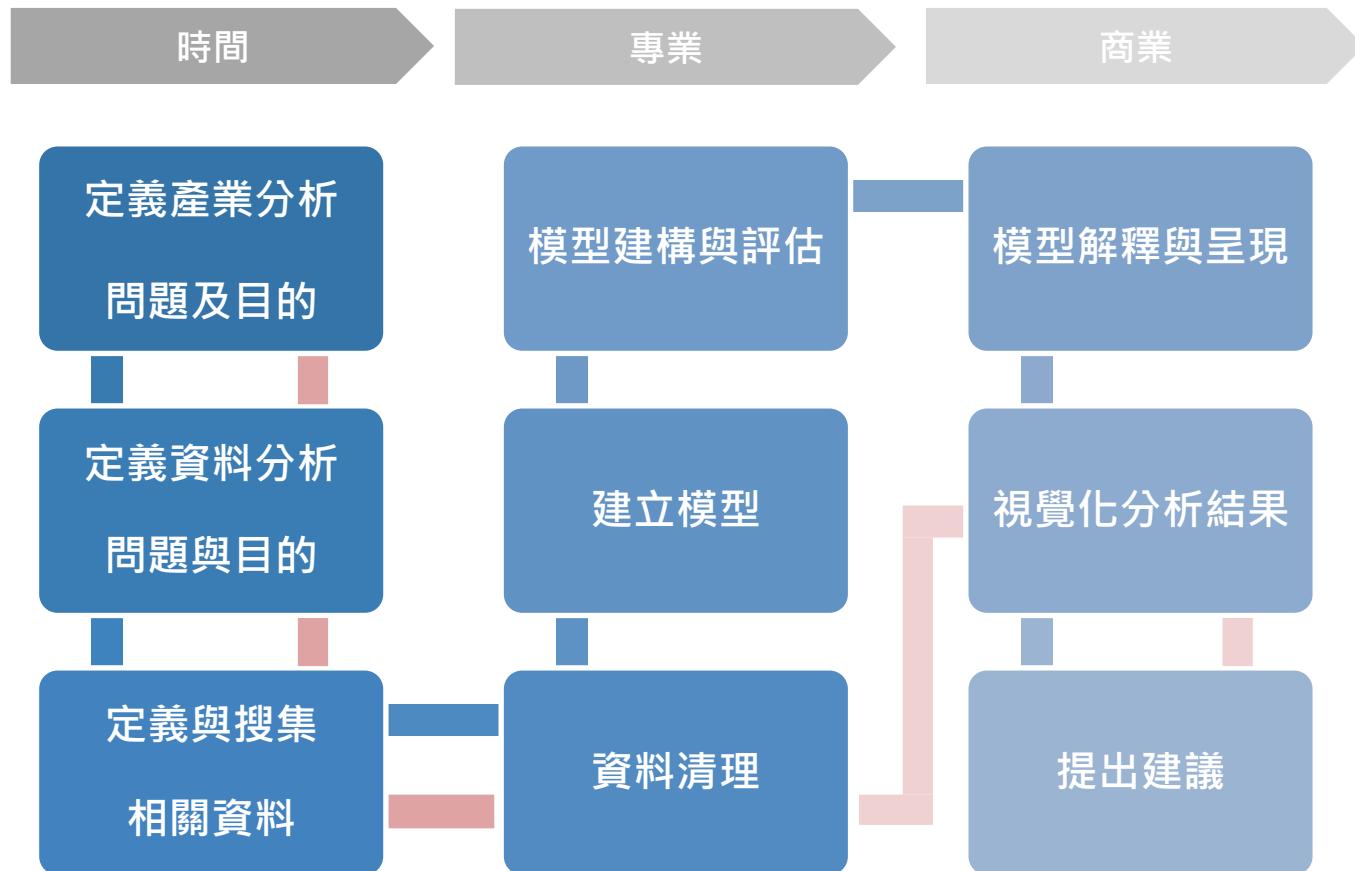


## Choose tools

1. 你是什麼角色？要解決什麼樣的問題？
2. 學習一門程式語言的成本？所花費時間？熟悉的軟體能不能解決你要解決的問題？
3. 在你的領域最常用的工具是什麼？有什麼可用的工具，它們和那些常用工具有什麼關係？

### 3. 數據分析流程

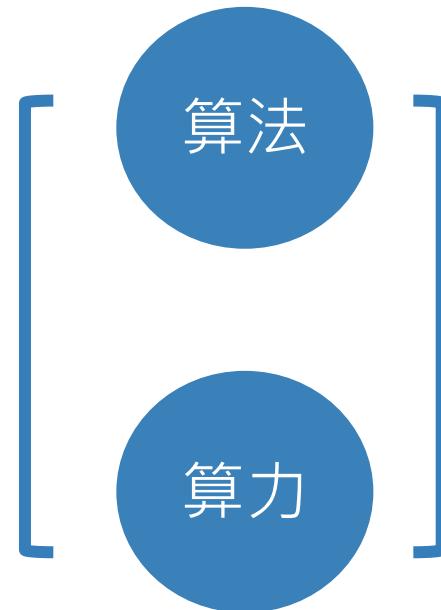




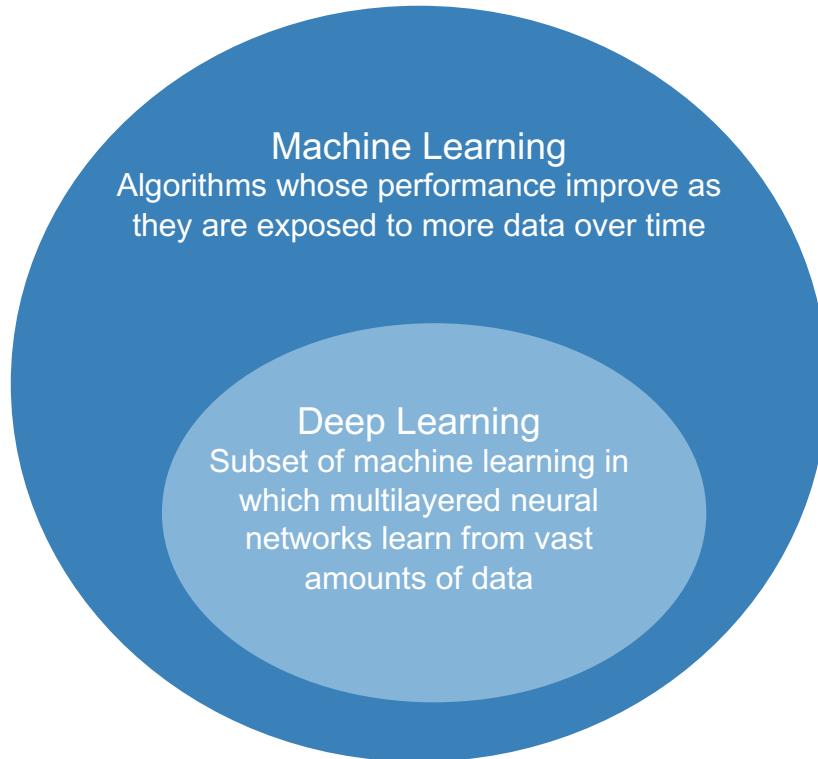
## 4. 機器學習簡介



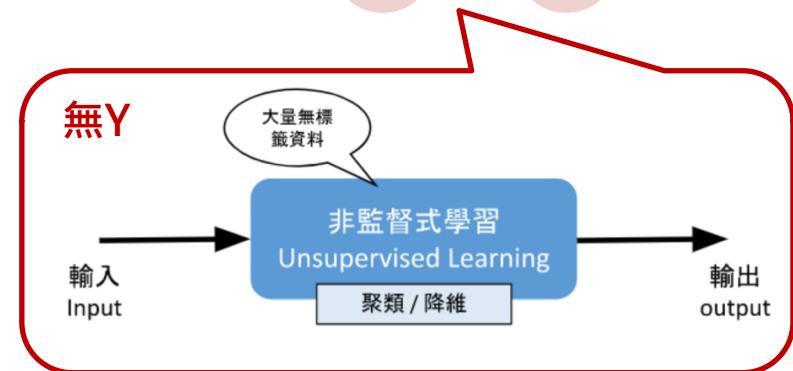
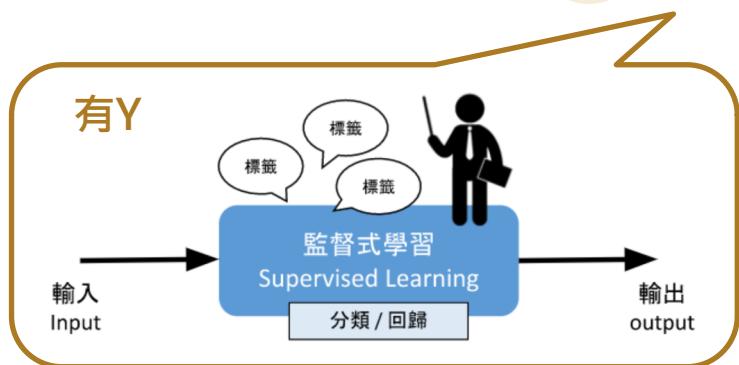
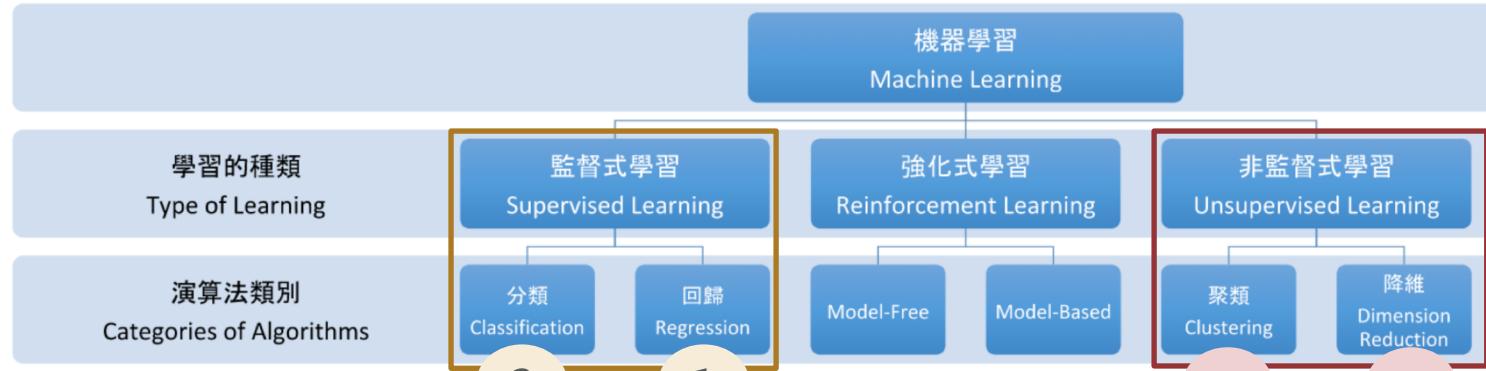
# AI 發展關鍵



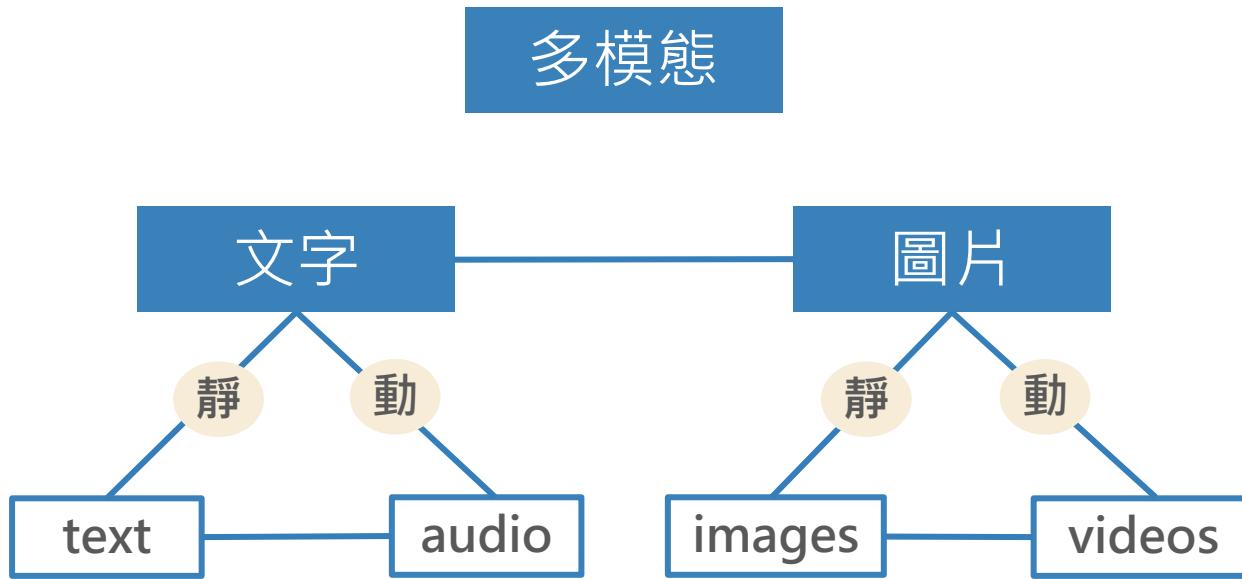
# 深度學習與機器學習的差別



# 機器學習課程框架



# 深度學習運用



# 常見深度學習框架及選用考量

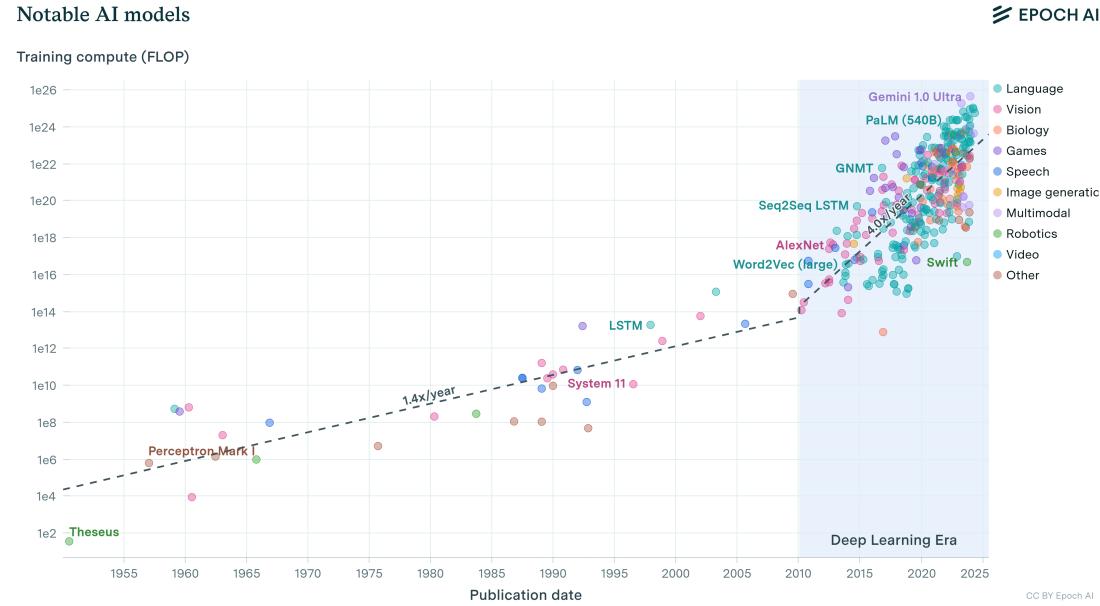


## 5. 挑戰與運用



# 挑戰

1. 定義問題：關注在企業的重要問題 ( \$\$ ) 。
  2. Data：高品質數據、隱私問題。
  3. Train model：model越來越複雜，數據量需求大，訓練成本高。
  4. 部署：即時/非即時，考慮計算成本。
  5. 監控模型：模型和數據會因時間而改變。數據/人為決策。



# 運用

商業

健康

製造業

推薦, chatbot, 銷售預測, 詐騙偵測

即時監測, 疾病檢測

機器維護, 產品品質控制, 自駕

# AI 發展

## 道德與倫理

THANKS

