

# » 大數據與商業分析

## Big Data and Business Analytics

楊立偉教授

2026

- 
- ◆ 課程介紹
  - ◆ 選課介紹

# 授課老師 楊立偉教授

- ◆ 研究領域為資料庫、搜尋引擎、語意分析、大數據及商業分析

現任 台大工管系暨商研所兼任助理教授 2006~

台大資管系兼任助理教授 2017~

資訊及通信國家標準技術委員

台灣 AI 大聯盟大語言模型工作組召集人

意藍資訊 (股票代號6925) 創辦人及總經理

國內規模最大的 AI 智能數據及雲端平台服務商

經歷 獲選100 MVP最有價值經理人，擁有超過20項語意分析專利

獲國家雲端創新獎、數位時代「創業之星」首獎、台大教學傑出獎

獲數位時代雜誌選為代表台灣軟體的封面人物之一

# 課程介紹 (1)

- ◆ 為管理學院特色課程，並為商業數據分析等領域學程認列之選修課程
- ◆ 人工智慧 (AI) 和大數據 (Big Data) 近年蓬勃發展，相關書籍與論述眾多，各行各業對此充滿期待。現實商業情境中，Big Data 是各種想像的基礎條件。著眼於培養企業需求日殷的相關人才，此一跨域課程針對管理學院同學而設，透過講演與專案，概念與實作並重，以主題式地介紹數據分析的商管相關應用，為有志成為商業應用領域數據專家的修課同學，奠定未來發展的基礎。

## 課程介紹 (2)

- 本課程內容以金融及零售數據應用為主，涵蓋統計分析、文字探勘、機器學習入門等技巧，分為金融市場及電子商務二大專題實作。主題如下：

### Part I 文字及社群數據分析

期中專題：財經數據分析實作 (股市/股價預測)

### Part II 行銷及零售數據分析、個人化推薦

期末專題：電商數據分析實作 (客群分級及銷售預測)

# 課程介紹 (3)

- ◆ 兼具講課及實作，選修同學需可能需要額外投入不少時間
  - 、心力完成各項要求。
    - 限大三以上至研究所修習
    - 分組成績為主 (含組內互評)，每組5~6人，共同討論、學習觀摩
    - 同組內不可全為同系所
- ◆ 修課者最好具備管理學及統計基礎知識；有程式設計基礎或隨本課程自學 AI Coding 相關工具。
  - 程式語言建議採用Python、R、Java任一
  - 選配其他工具 (例如 Claude Code、統計軟體、資料庫等)
- ◆ 其他請參考課程說明網頁或學習平台
  - <http://homepage.ntu.edu.tw/~wyang>

# 主題介紹

L1: Text mining 文字分析

L2: Web mining 關聯分析

L3: Classification 監督式學習

L4: Clustering 非監督學習

期中專題：財經數據分析實作 (股票市場走勢分析)

L5: Sequence Tagging 序列分析

L6: Language Processing 語言處理

L7: Gen AI & Applications 生成式AI發展及應用

期末專題：行銷及零售數據分析 (客戶分群/商品推薦/精準行銷)

Special Issue: Social Data Analytics 社群數據分析

# 使用 AI Coding 作為輔助

- 不拘語法、釋放想像力
- 不同種類 及 不同層次的輔助工具
  - 通用性工具，提供技術問答，產生程式片段，如 ChatGPT、Claude、Gemini
  - 開發環境整合工具，與 IDE 整合，提供建議或用口語導引開發及除錯，如 Cursor、Windsurf、Google Antigravity、Google AI Studio
  - 命令列工具，賦予本地端權限，具備規劃及調用工具能力，如 Claude Code、OpenAI Codex CLI、Google Gemini CLI
  - 其他如 Lovable、Replit 等

# 系統開發流程與問題解決步驟



研究整理：楊立偉 · 意藍資訊

# 課程參考資料

## ◆ 課程講義 / 實習講義

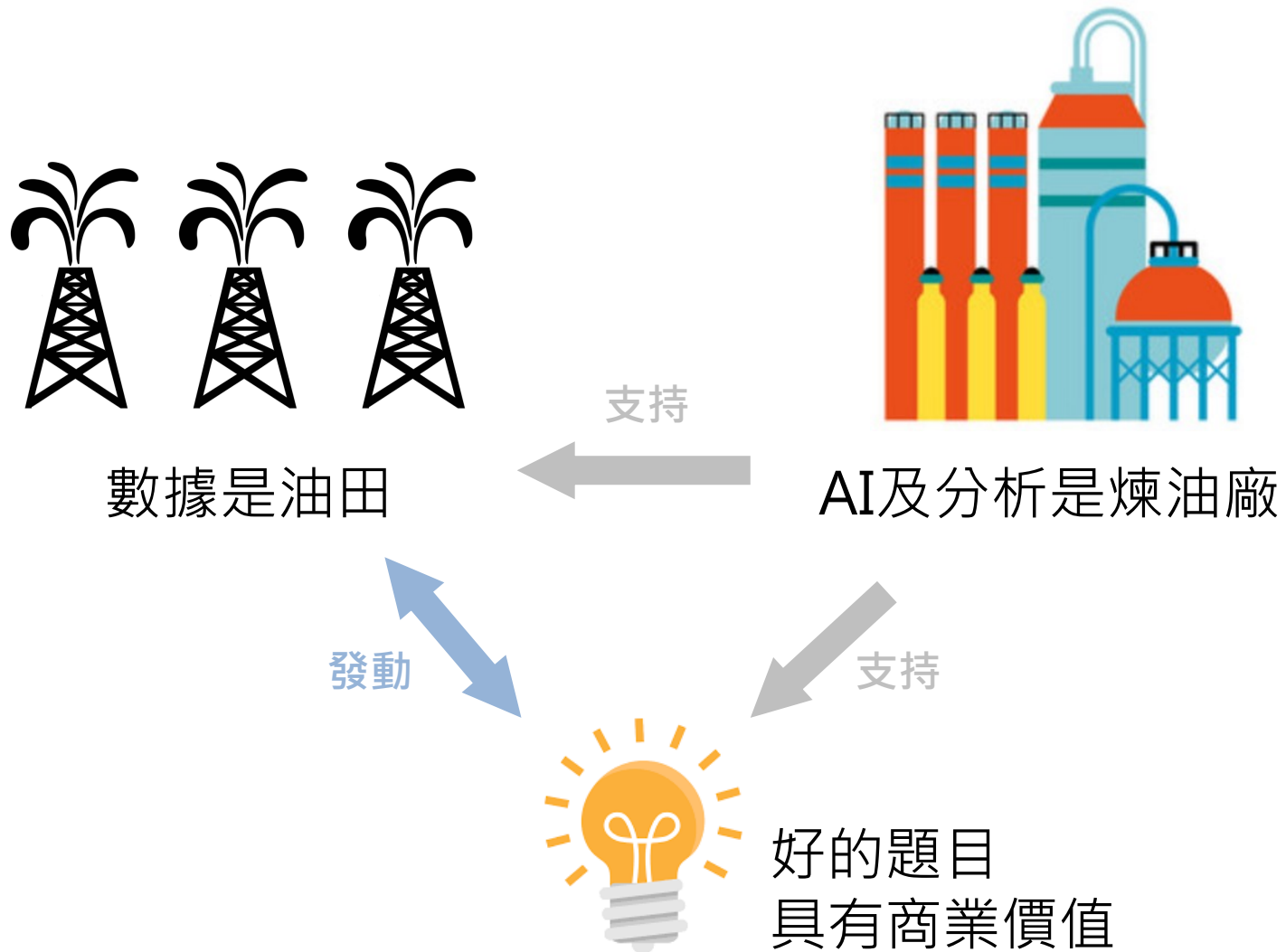
## ◆ Stanford AI Labs, NLP Group

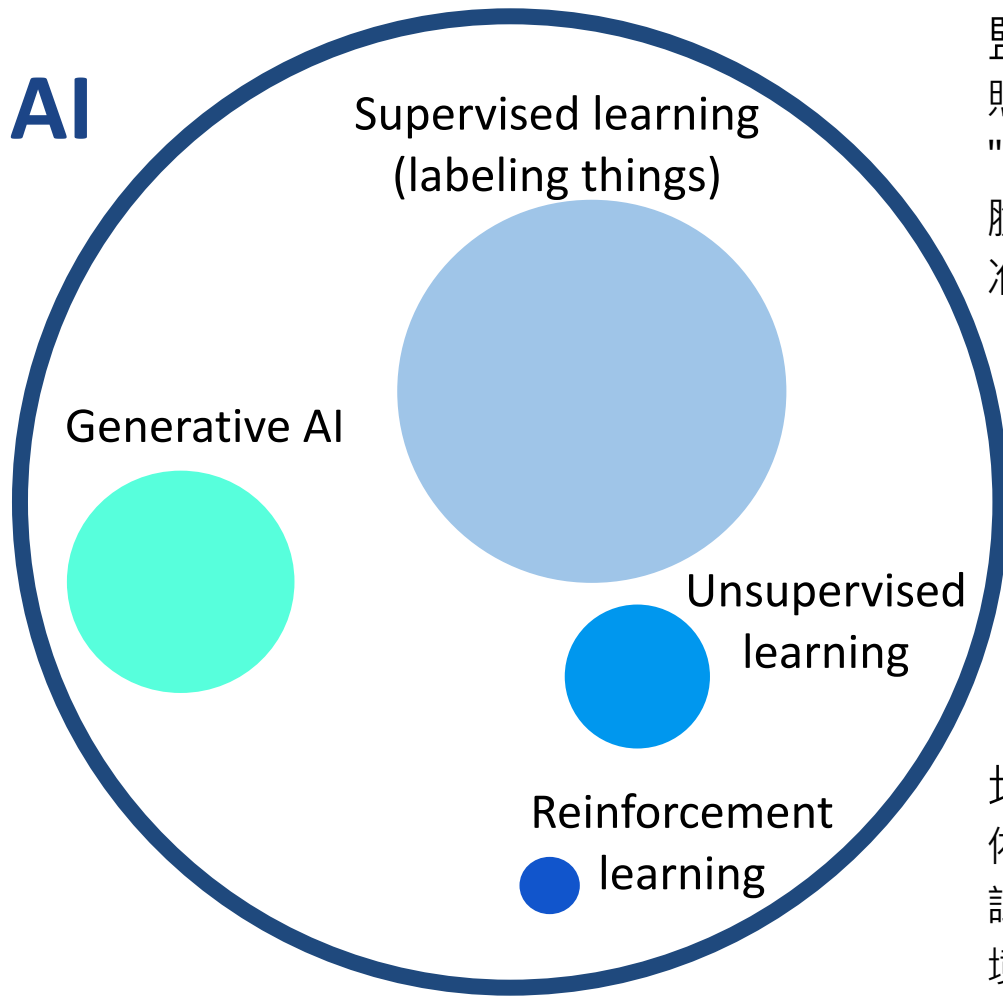
- Class CS276
- Class CS224N
- Introduction to Information Retrieval by C. Manning, MIT 2008

## ◆ 中文參考書

- 理論到實作都一清二楚：機器學習原理深究，深智 2023
- 全中文自然語言處理：Pre-Trained Model方法最新實戰，深智 2022
- PyTorch深度學習與自然語言中文處理，博碩 2018

# AI & Big Data





生成式AI  
學習資料分布，  
生成更多資料。  
如內容生成、對  
話問答等

Supervised learning  
(labeling things)

Generative AI

Unsupervised  
learning

Reinforcement  
learning

監督式學習  
照範例訓練到自動  
"依樣畫葫蘆"。如  
臉部辨識、貸款核  
准、消費預測等

非監督式學習  
從資料中找到洞見  
如客戶分級、商品分群

增強式學習  
依目標和獎勵規則，嘗  
試學習。如下圍棋、環  
境感測

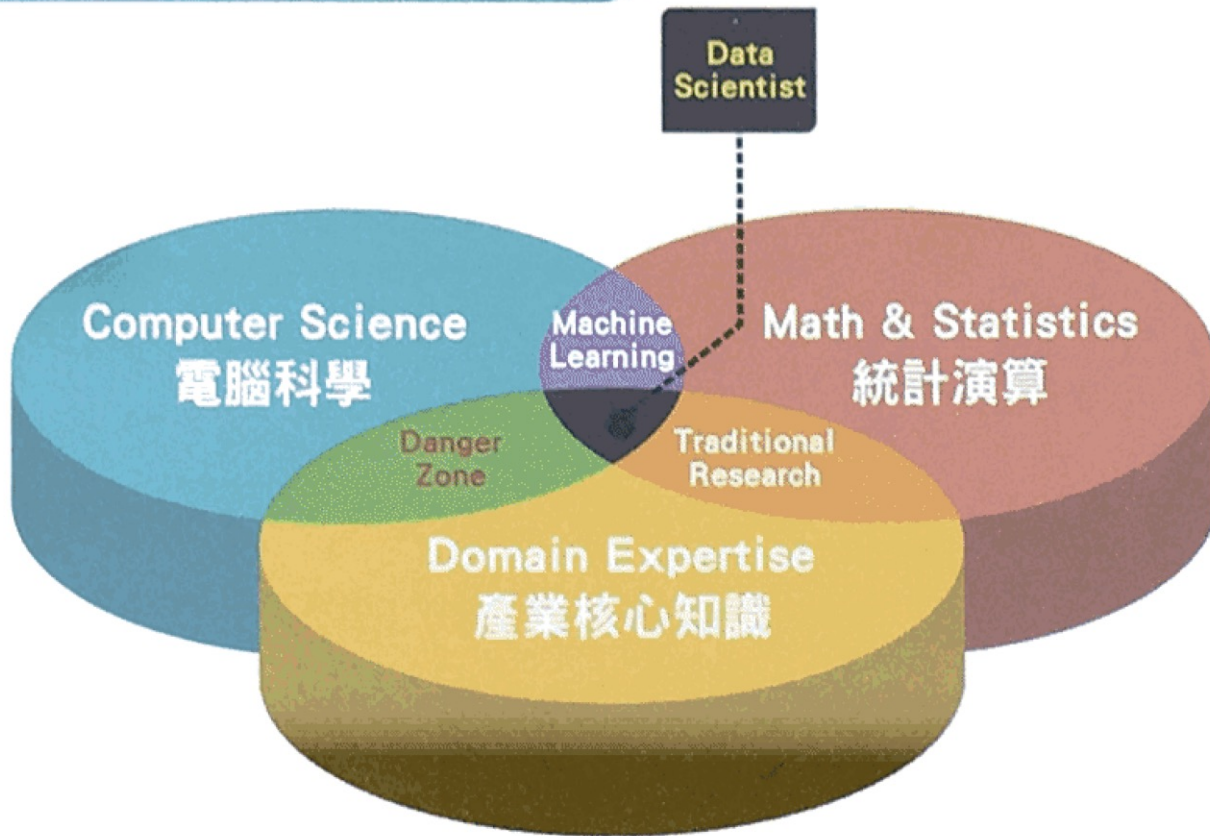
有數據有答案

只有數據

只有目標

合成數據

圖13：大數據需要跨界人才



數據科學家就像傳說中的獨角獸，得熟悉橫跨三大領域的知識，包含電腦科學、統計分析和產業核心知識。前兩者著重於技術層面，而產業核心知識即是大數據的哲學思考來源，愈理解產業核心，就愈不容易迷失在數據海。但要注意danger zone，擁有電腦科學和產業核心知識的人才，容易低估統計演算邏輯的重要性，而落入錯誤判斷的陷阱

Source: 大數據玩行銷 Big Data Marketing

# Introduction

## ◆ 什麼是大數據 Big Data

- Data that exist in very large volumes and many different varieties (data types) and that need to be processed at a very high velocity (speed).

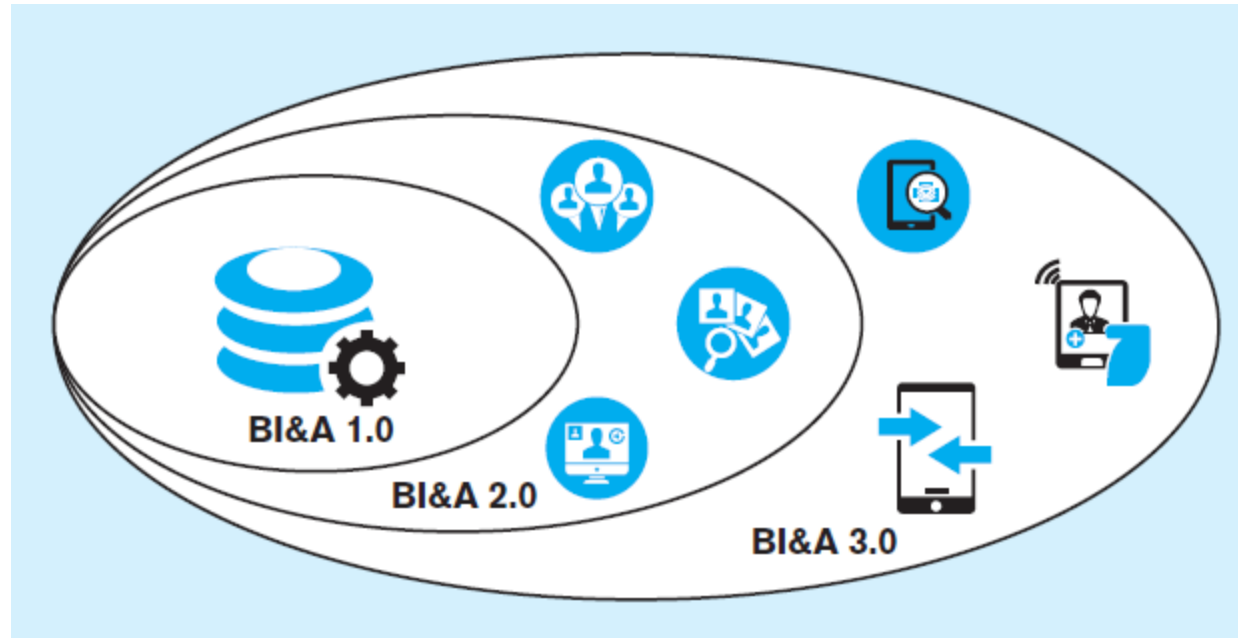
## ◆ 什麼是分析 Analytics

- Systematic analysis and interpretation of data—typically using mathematical, statistical, and computational tools—to improve our understanding of a real-world domain.

# Types of Analytics 分析的類型

- ◆ Descriptive analytics 敘述性分析
  - describes the past status of the domain of interest using a variety of tools through techniques such as reporting, data visualization, dashboards, and scorecards
- ◆ Predictive analytics 預測性分析
  - applies statistical and computational methods and models to data regarding past and current events to predict what might happen in the future
- ◆ Prescriptive analytics 指示性分析
  - uses results of predictive analytics along with optimization and simulation tools to recommend actions that will lead to a desired outcome

# Figure - Generations of Business Intelligence and Analytics



Adapted from Chen et al., 2012

## **BI&A 1.0**

Focus on structured quantitative data largely from relational databases

## **BI&A 2.0**

Include data from the Web (web interaction logs, customer reviews, social media)

## **BI&A 3.0**

Include data from mobile devices, (location, sensors, etc.) as well as Internet of Things

# from Reporting to Analysis

警告: 14號AP伺服器在過去5分鐘  
CPU使用率超過98%

原來是10:01時網友討論推薦本公司服務，建議應該如何...

Reporting	Analysis
Descriptive 敘述	Prescriptive 處置
What? 發生什麼	Why? 為什麼發生
Backward-looking 解釋過去	Forward-looking 預測未來
Raise questions 提出問題	Answer questions 找到答案
Data → information 產生資訊	Data + information → insights 產生看法
Reports, dashboards, alerts 報表, 儀錶板, 通報	Findings, recommendations, predictions 發現, 建議, 預測
No context 獨立進行	Context + storytelling 找到脈絡

# 選課介紹 (1)

- ◆ 本學期採混合式教學 (實體+線上) ，進行方式如下
  - 預錄講義及解說，於學習平台上自主學習觀看
  - 實體周次之課堂時間為重點解說、討論問答、展示報告
  - 同學分成小組，自行討論及實作專題報告 (建議5~6人，自行分組並推選組長，分組報告均含互評成績)；2周內向助教登記名單，超過期限則由助教隨機編組
  - 個人作業自行繳交；分組作業除繳交報告外，需錄製簡報影片，放上平台後彼此學習觀摩
    - 遲交處理原則: 超過期限24小時內8折計分; 超過24小時後不收

## 選課介紹 (2)

- ◆ 總人數上限150人，分為 1.工管系所35人、2.資管系所35人、3.管院其他三系所35人、4.學程名額25人(管院商業數據分析學程、會研所商業智慧與數據分析組、經濟系跨域專長選課)、5.其他不限系所20人。
- ◆ 線上登記加選或旁聽。截止後，助教將抽籤進行修課學生加選作業。請注意授權碼限本人使用，抽籤未中者不遞補，依意願登記線上旁聽 (無學分)

# » 問題討論