



大數據的處理 目標希望找到 有關聯性 的資料  
要找到這樣的資料 當輸入 然後 輸出 有關聯性的部分

但事實上 這個有用的 關聯性 十分難找  
所以會一次輸入 很多種類 很大的資料  
然後讓電腦幫我們找出這個 關聯性 」

我們人類也需要幫忙一些  
如 找出不合理值 幫忙篩選  
合併欄位 合併分散的資料  
或是 資料擴增  
或是 判斷是不是 高斯分佈 代表可以合理降維度

取特徵值

資料降維 是從已經有紀錄的資料 去查看  
哪些東西是可以

資料的標示 有四種 ：  
連續 離散 圖像 時空

資料處理的步驟：  
清洗資料  
去特徵值  
資料降維 t-SNE  
建模型  
分類  
報告論文

論文 需要做 量化 值化 的驗證  
甚至需要做一套SOP  
量化 就是用數據 去驗證

## ➤ 利用簡易案例加強分析流程的概念

### ● 案例2：利用大數據分析降低學生餐廳廚餘量

定義題目：廚餘？剩餘的 ✓  
 收集資料：沒吃完的 ✓  
 清洗資料：含水分量 ✓  
 取特徵值：重量 ✓  
 降維：天氣、節慶、行事曆、期中、期末考 ✓  
 分類：中餐、西餐、歷史資料 ✓  
 建模：✓  
 報告製作：✓

對目標的客制化分析  
 5G 階段 各種殘人數  
 科學人數 男生人數  
 女生人數  
 環境保護

最後步驟 要把這東西實現 並作出SOP 防呆 等等的

希望再用 分類 的技術 提升分析準確度

分類 這技術 超級重要

之後要把準確率提升 有一部分 就是要靠 分類

## ➤ 利用簡易案例加強分析流程的概念

### ● 案例2\*：利用大數據分析降低學生餐廳廚餘量 (如何使用分類技術來提升分析準確度?)

定義題目：分類：早、中、晚各建一個模型

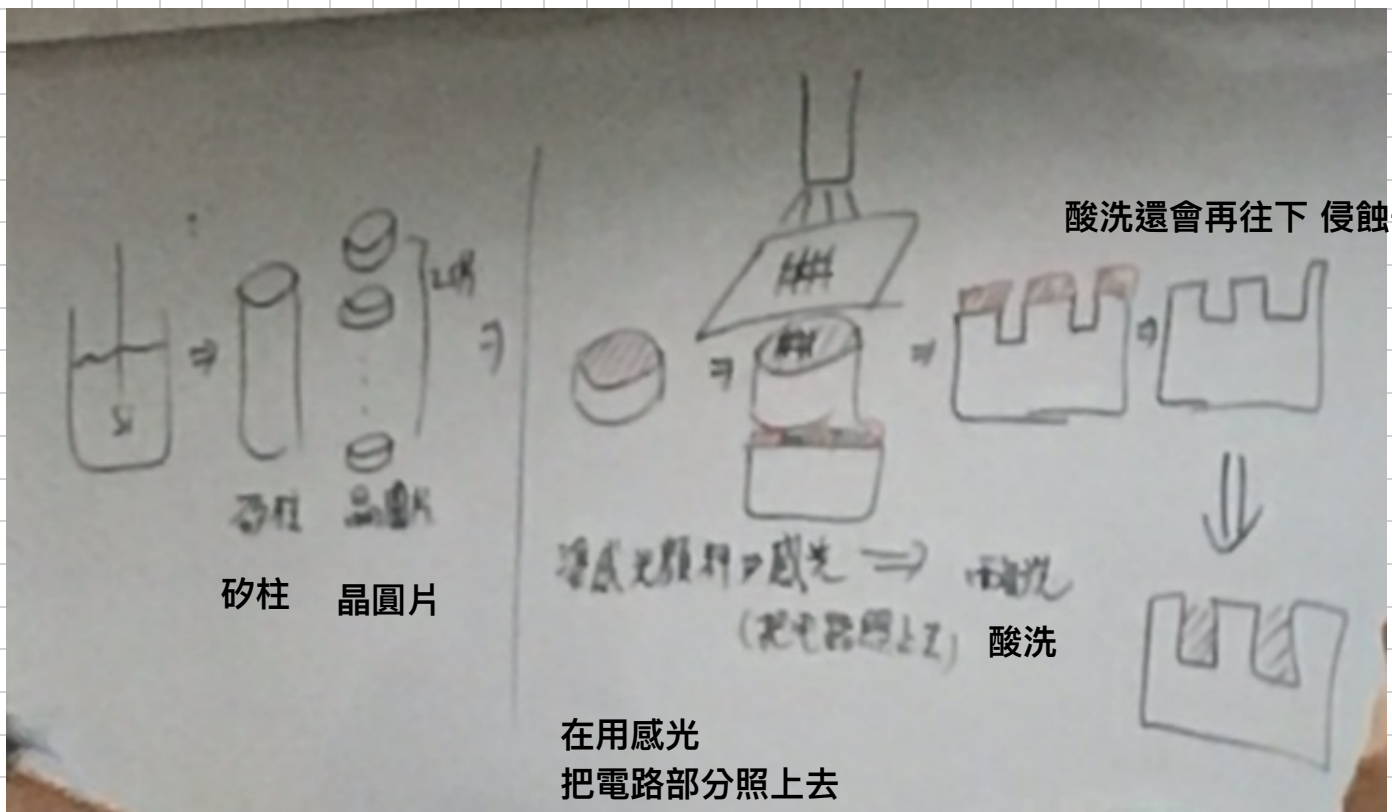
收集資料：

清洗資料：不同國家學生的飲食行為不同

取特徵值：把中式料理 & 西式料理分開處理

降維：

建模：不同時間不同模型



職安 講的 2.0Ph值 強酸  
就是在這裡

定義好題目 和 要做的事 超重要  
要知道做什麼

➤ 分析流程各部分執行關鍵 [有球人會需要]

● 定義題目

1. 沒有題目就不可能分析
2. 定義題目請從簡單到難
3. 與第一線人員而非主管對話

● 收集資料

1. 請確定資料是否可用，如視覺化，或了解資料收集方式
2. 請確定資料分佈跟題目所想是否一樣

● 清洗資料：

1. 大數據分析最關鍵地方
2. 技術不難，而while就會做
3. 但他是...最痛苦的地方

Handwritten notes and diagrams on the right side of the slide, including a scatter plot with axes labeled 'x' and 'y', and some text in Chinese characters.

把資料視覺化 超重要 不然模型 和 結果 做出來很可能是錯的

像這個要分 圈 叉 的狀況

明顯是要分成三個區域

但只有兩種資料很容易只分兩類

也就是只有畫一條線 但只畫一條線去分類 怎麼樣都會是錯的

有的時候資料處理 準確率 80多%就算很好了

Fuzzy Logic 模糊邏輯

問溫度

精確：25度

模糊：好熱 (29~35)

模糊的數學太難算

後面才會再把它改成模糊類神經