



資料前處理





資料(Data)



■ 什麼是資料？

- 透過觀測取得的紀錄
- 描述一個或多個人事物的特性

情況	資料範例
公司開會	會議記錄
交通管制	車流量、車速
問卷調查	填答紀錄
醫院	病歷紀錄
討論版	發文、留言記錄
股市	股票委託、交易紀錄

什麼是資料？

原始
資料

- 以觀察、蒐集等方式獲取未經加工過的數據

資訊

- 將原始資料經過整理，如統計、大數據分析後，可以提供使用者作出決策或參考之數據

智慧

- 根據彙整後的資訊，結合領域知識(Domain Know-How)，衍生出各類型不同的商業智慧(Business Intelligence, BI)

資料的重要性

- 世界經濟論壇在2012年發表的「Big Data, Big Impact: New Possibilities for International Development」報告指出，「資料」的地位等同於黃金和貨幣，將會是未來的新興的無形資產

資料的重要性

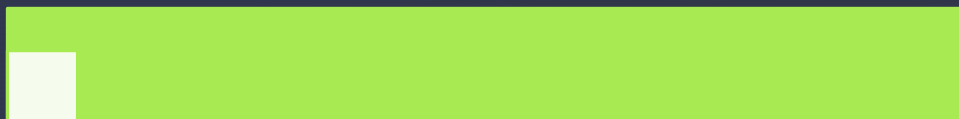
決策

數據分析

資料取得

資料的重要性

內部



☐ 發掘企業內部的優缺

☐ 客戶精準行銷

☐ 客製化服務

外部



☐ 探索市場趨勢

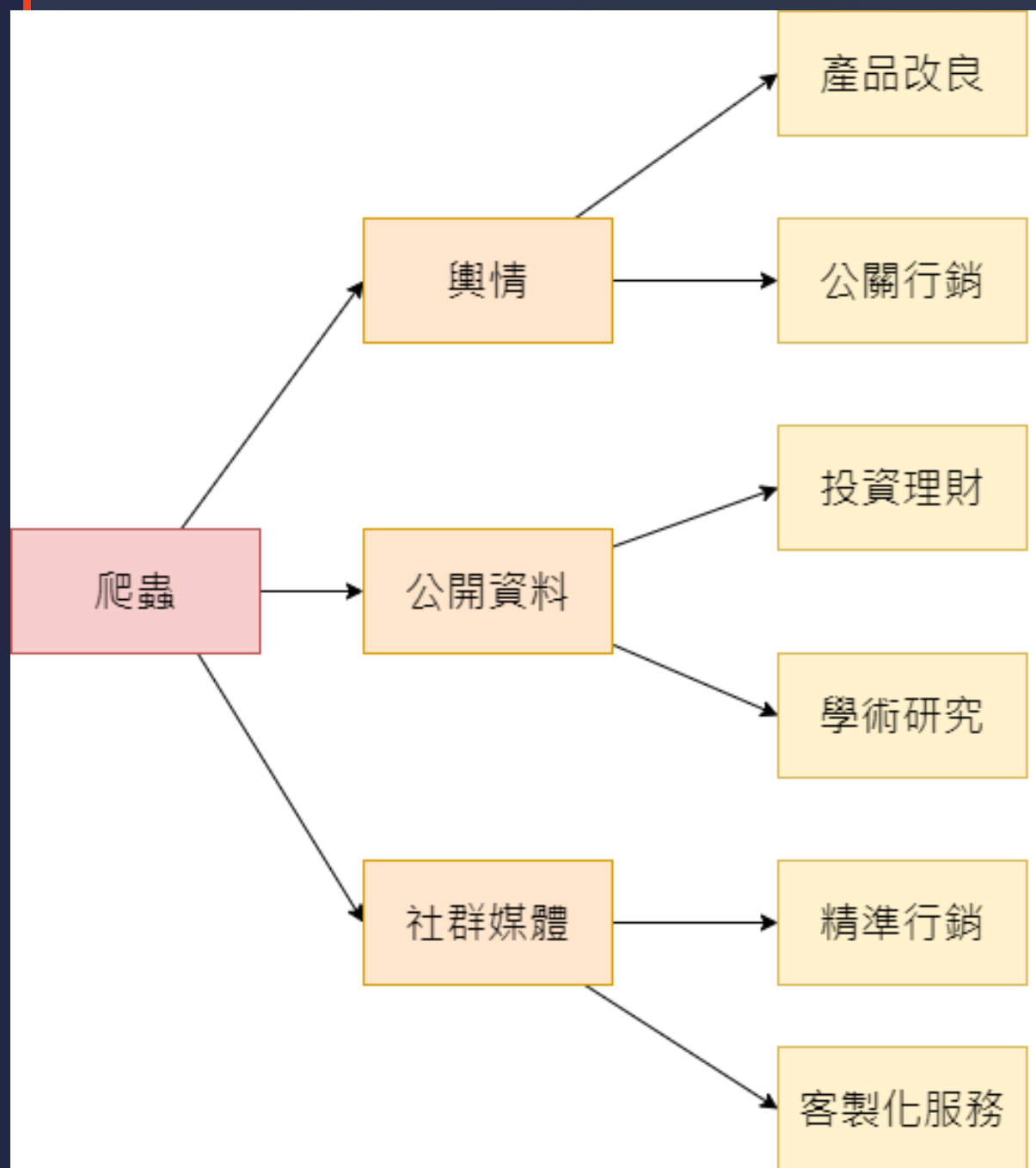
☐ 輿情分析


☐ 競爭者分析

數據分析實例

資料	用途
台灣證券交易所	程式交易
公開資訊觀測站	財報分析、審計查核
PTT、新聞	輿情分析
FB、Twitter	推薦行銷、客製化服務
氣候、公共衛生資料	登革熱防治

資料應用範例



- 
1. 把龐雜的資料進行整理
 - 將原始資料根據需求進行彙整
 2. 建構資訊系統
 - 將例行性資料建構成系統，自動化整理
 3. 預測分析
 - 利用歷史資料進行預測分析



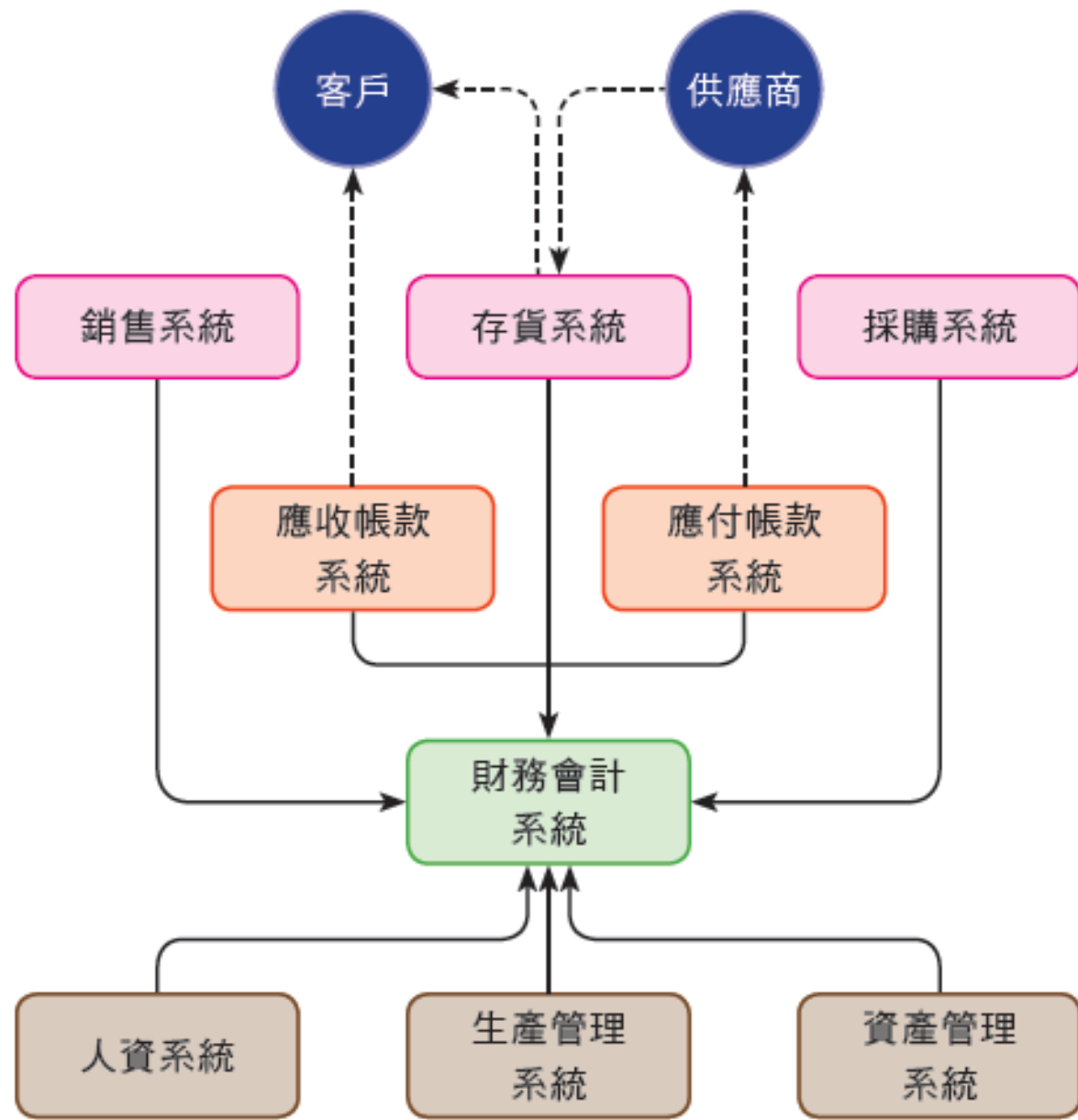
1. 龐雜的資料進行整理

運用領域：

1. 統計學
2. 程式設計
3. 資料庫

將資料放入統一格式，去除資料中的雜質

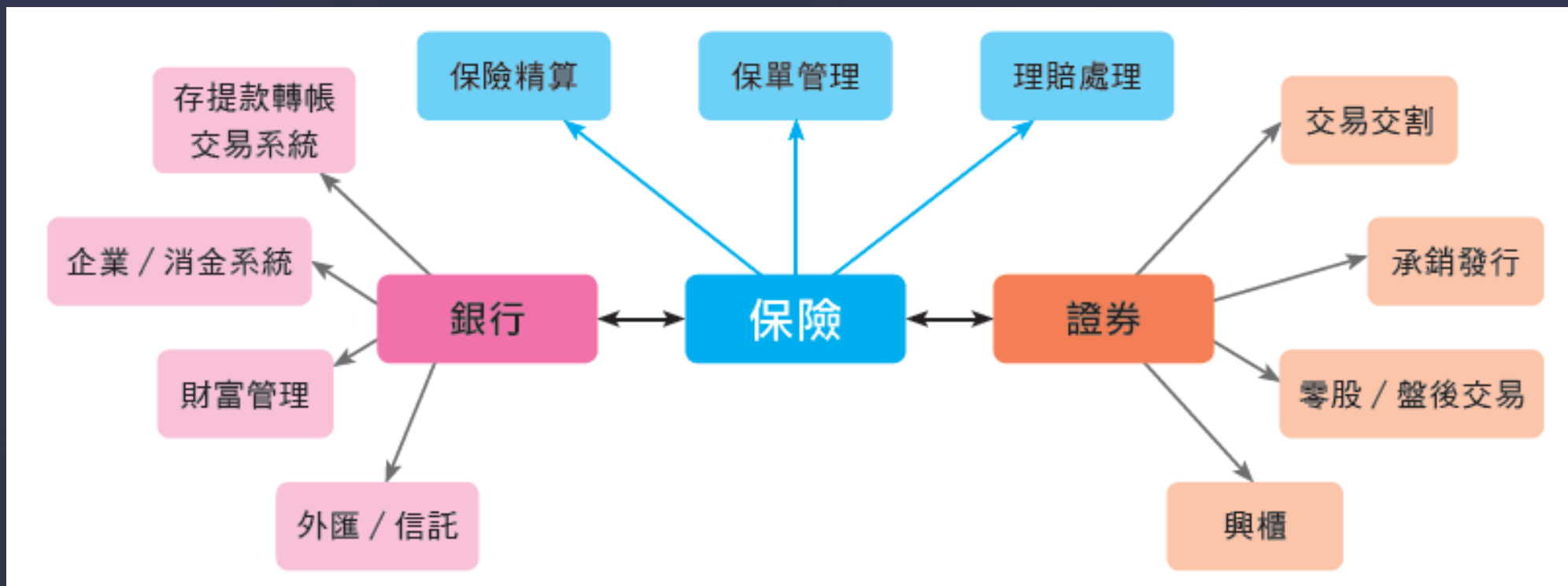




2. 建構資訊系統

製造業 ERP

2. 建構資訊系統



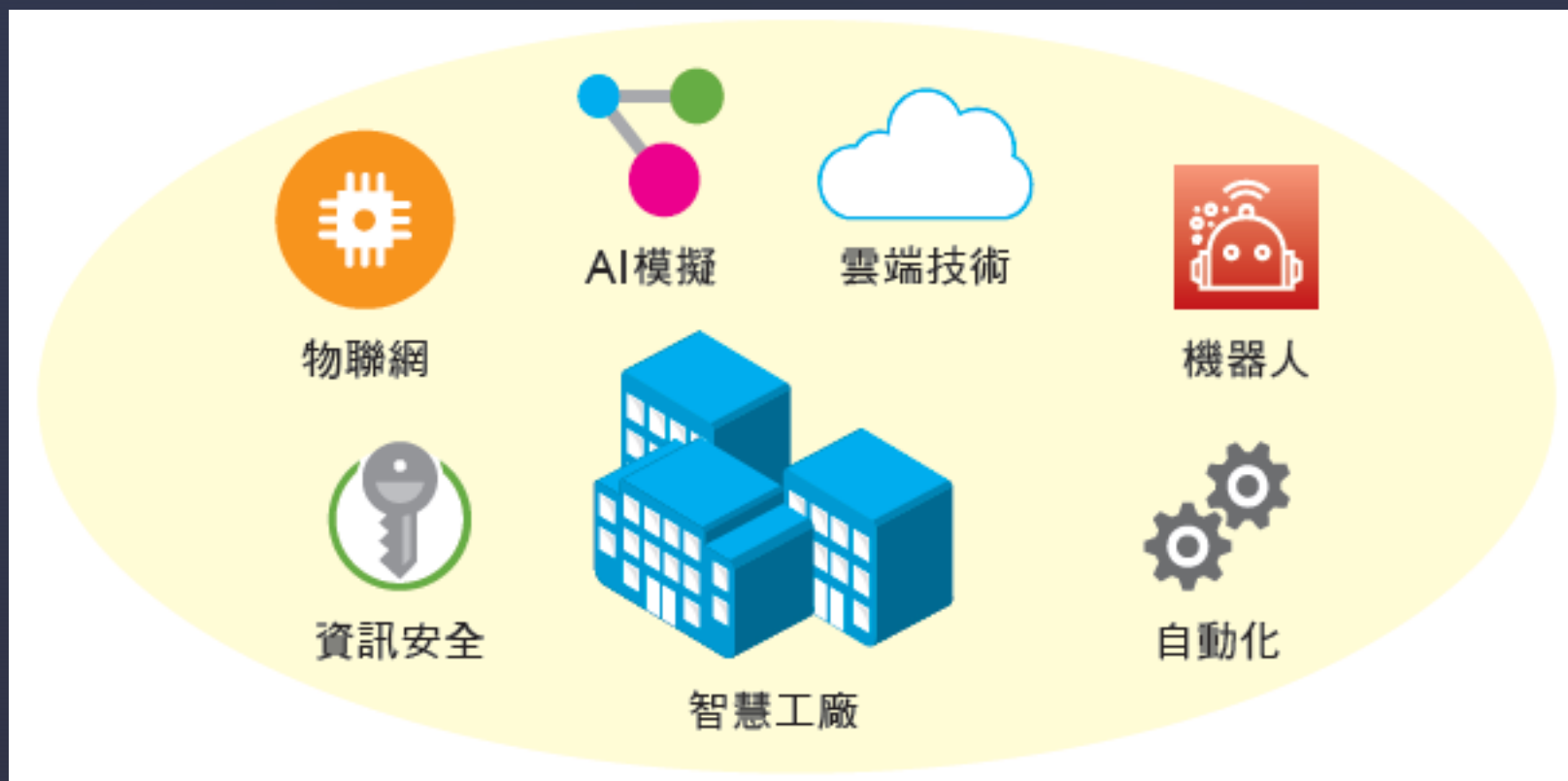
金融業資訊系統

2. 建構資訊系統



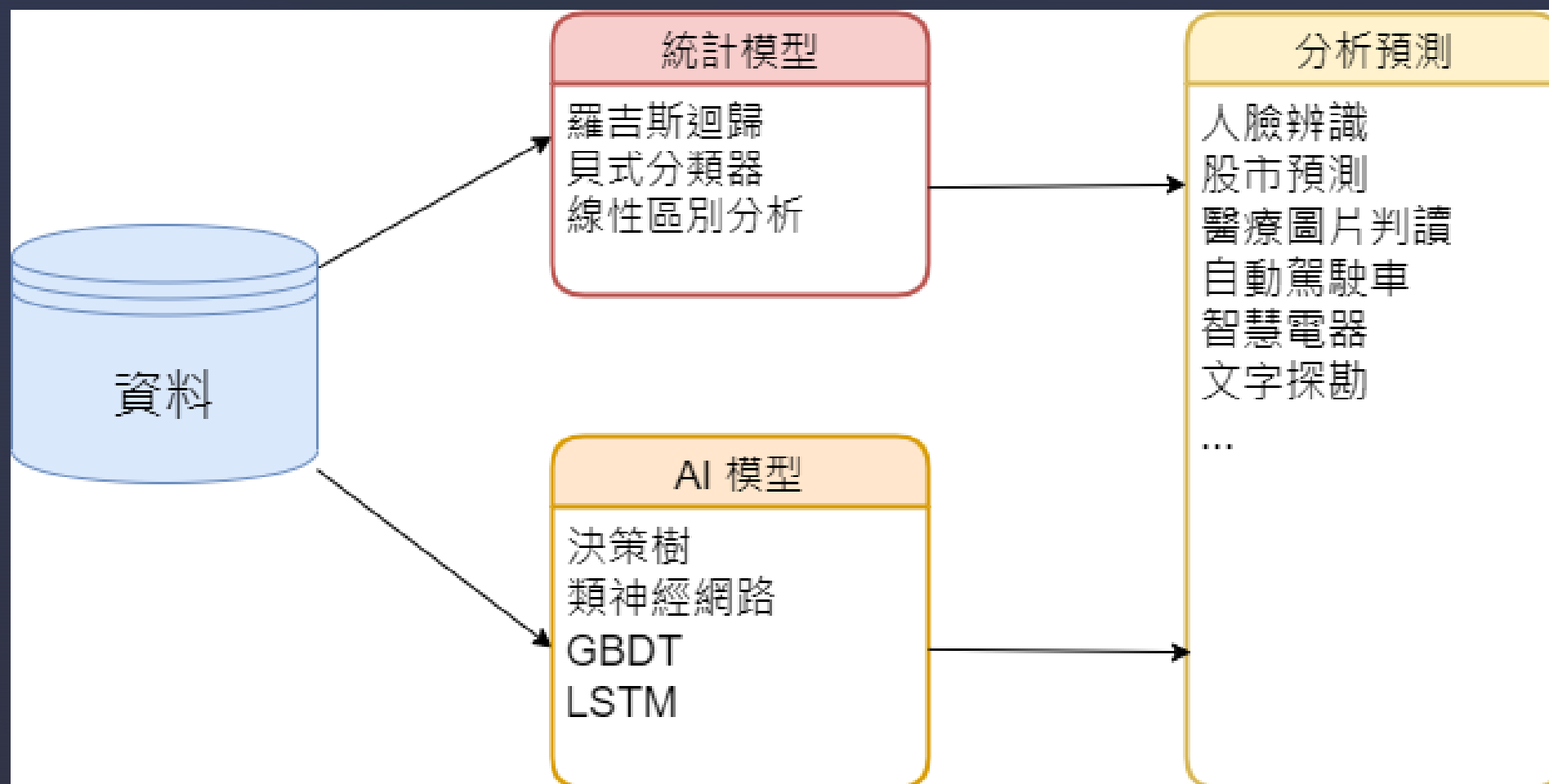
醫療資訊系統

2. 建構資訊系統



工業 4.0

3. 預測分析





測量尺度 (scale of measure)

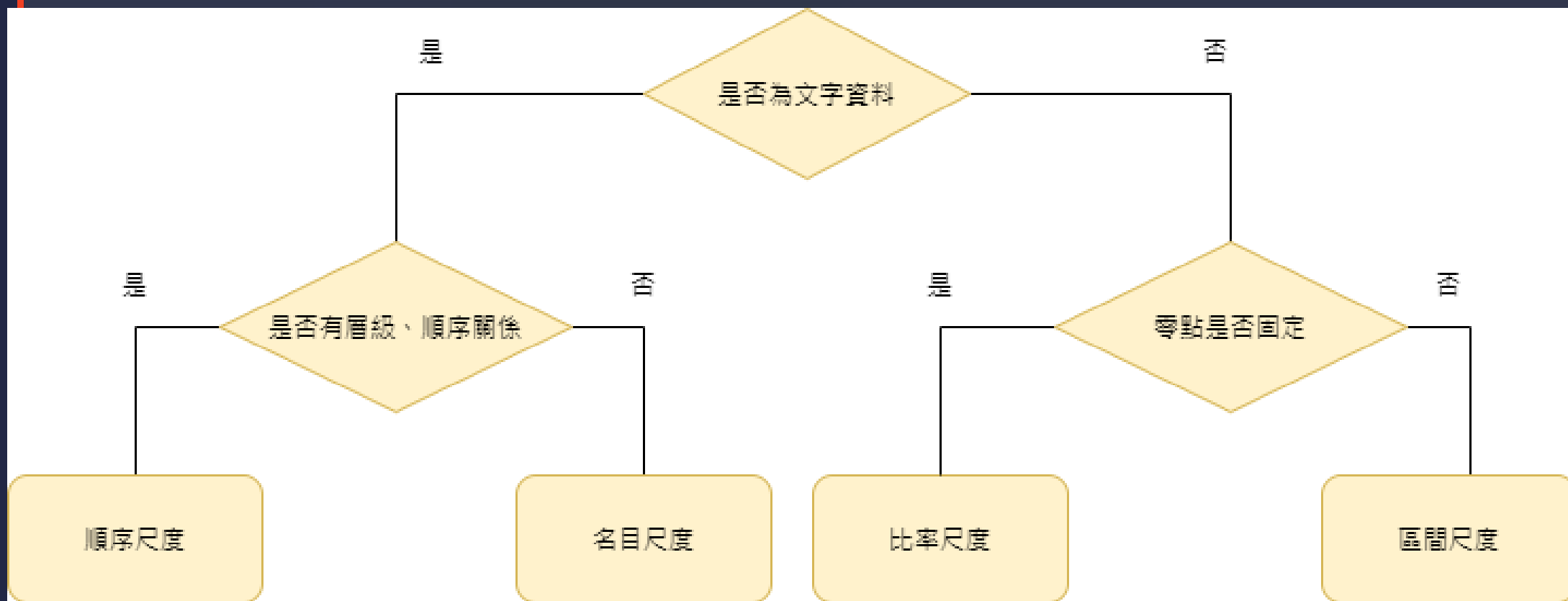
統計學中的資料類別

資料類別	又稱	四則運算	零點	比較大小	結構性	範例
名目尺度 (Nominal)	名義、類別	X	X	X	最弱	性別、職業、血型
順序尺度 (Ordinal)	等級、序列	X	X	V	第三	排名、米其林等級、PR值
區間尺度 (Interval)	間隔、間距、等距	通常 僅加減	不固定	V	第二	攝氏溫度、年份
比率尺度 (Ratio)	等比、比例	可	絕對零點 (true zero)	V	最強	金錢、長度、重量

各種尺度的統計方法

資料類別	範例	眾數	中位數	差值、極值	平均值	標準差
名目尺度 (Nominal)	性別、職業、 血型	V				
順序尺度 (Ordinal)	排名、職等、 米其林等級、 PR值	V	V	V		
區間尺度 (Interval)	攝氏溫度、 年份	V	V	V	V	V
比率尺度 (Ratio)	金錢、長度、 重量	有時可以	V	V	V	V

資料分類參考





謝謝聆聽！

