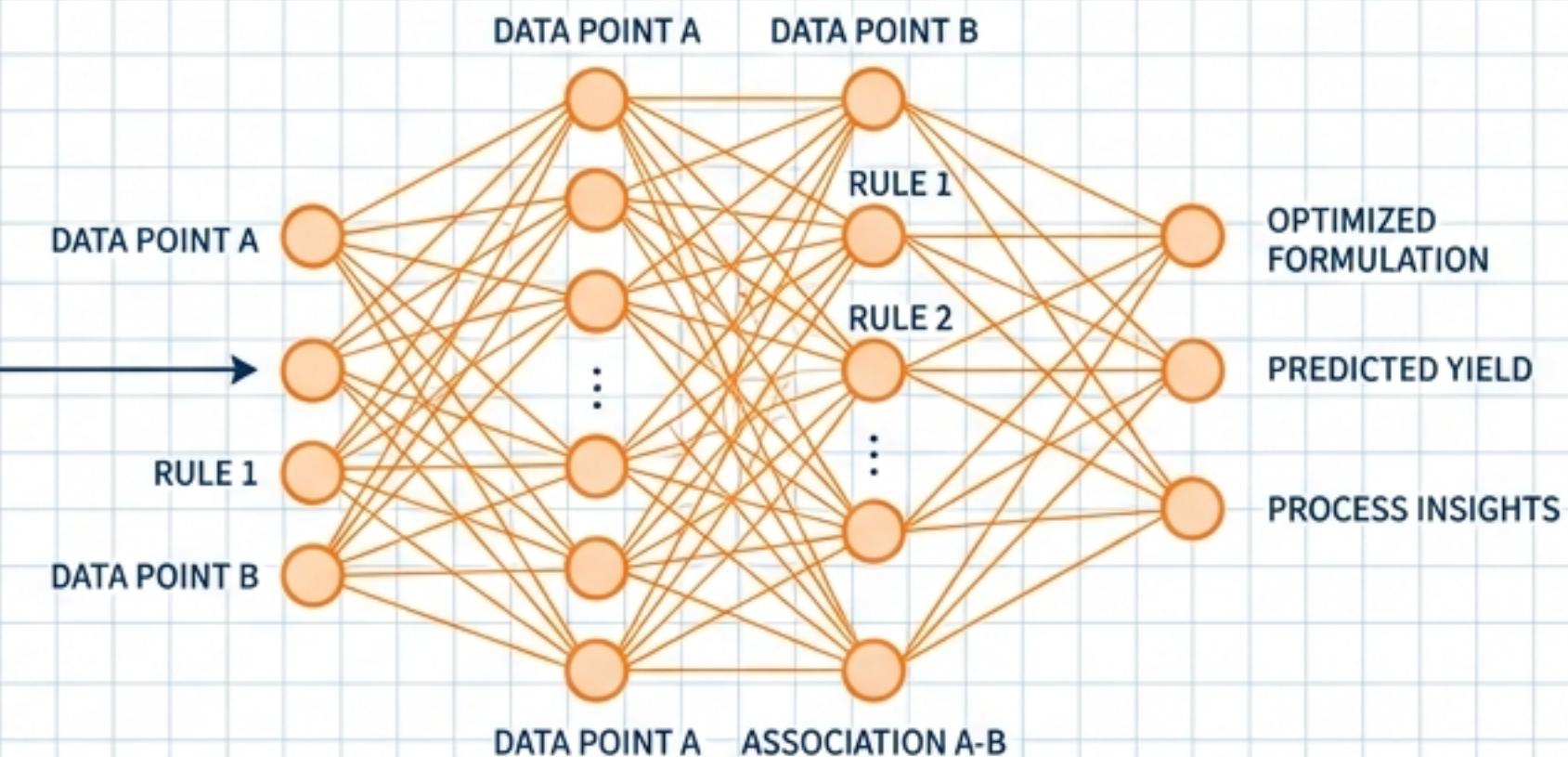
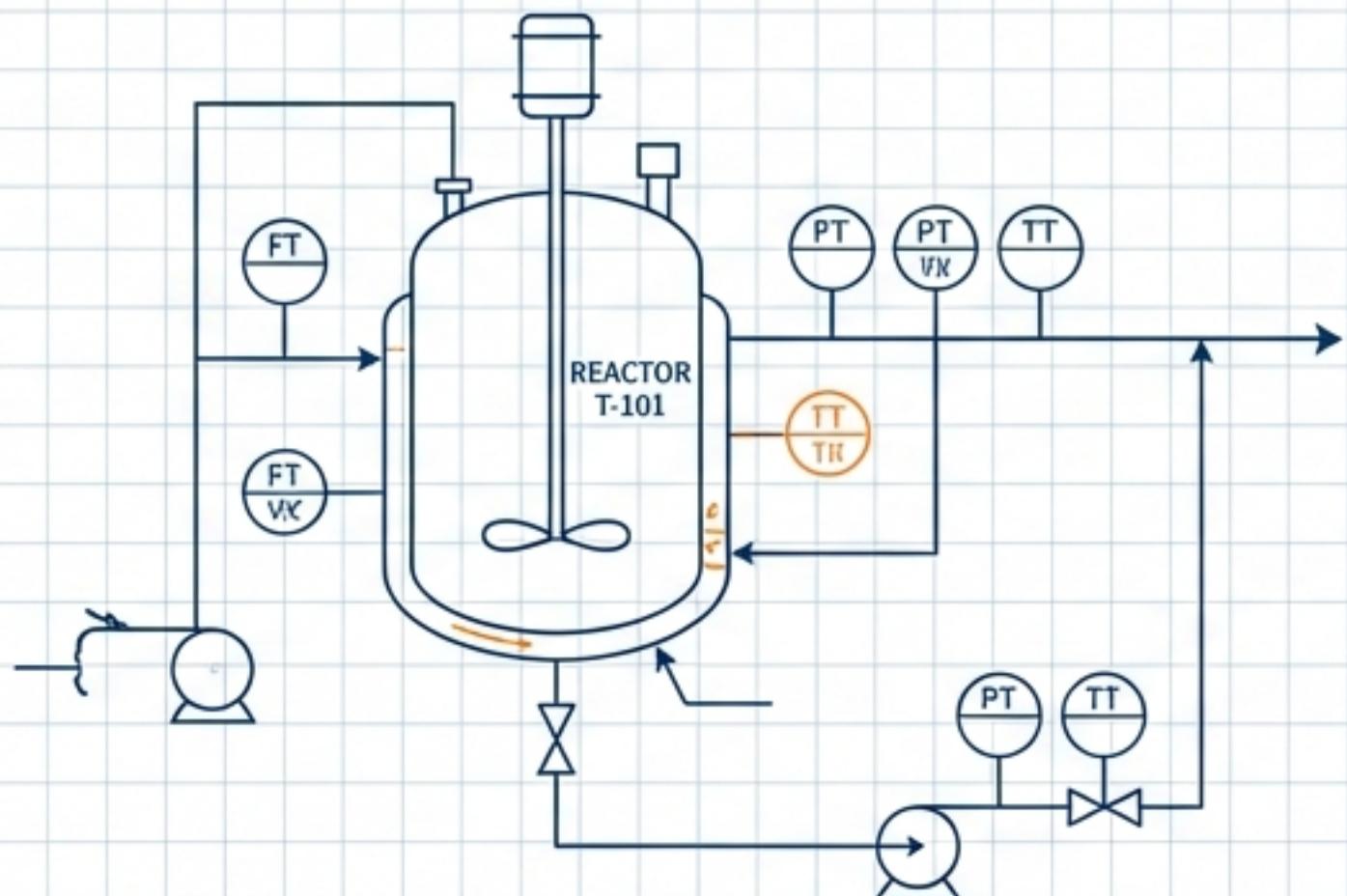
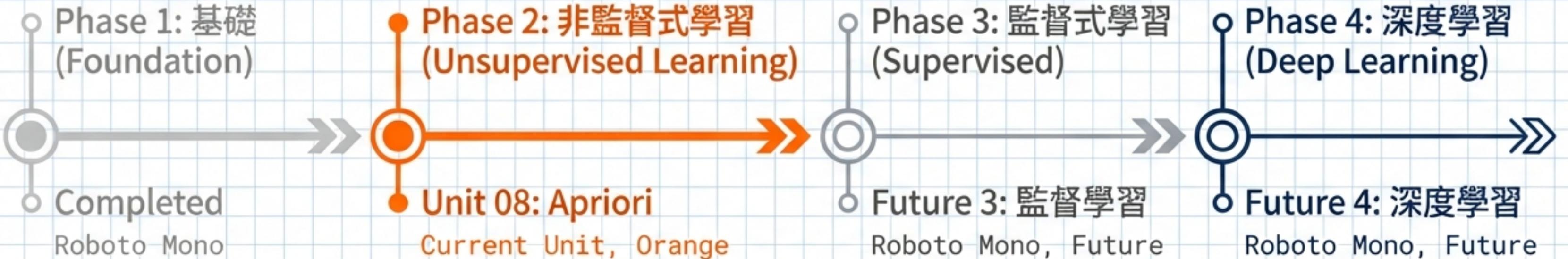


# Unit 08: Apriori 演算法 (關聯規則挖掘實務應用)

從試誤法到數據驅動：優化化工配方的關鍵技術



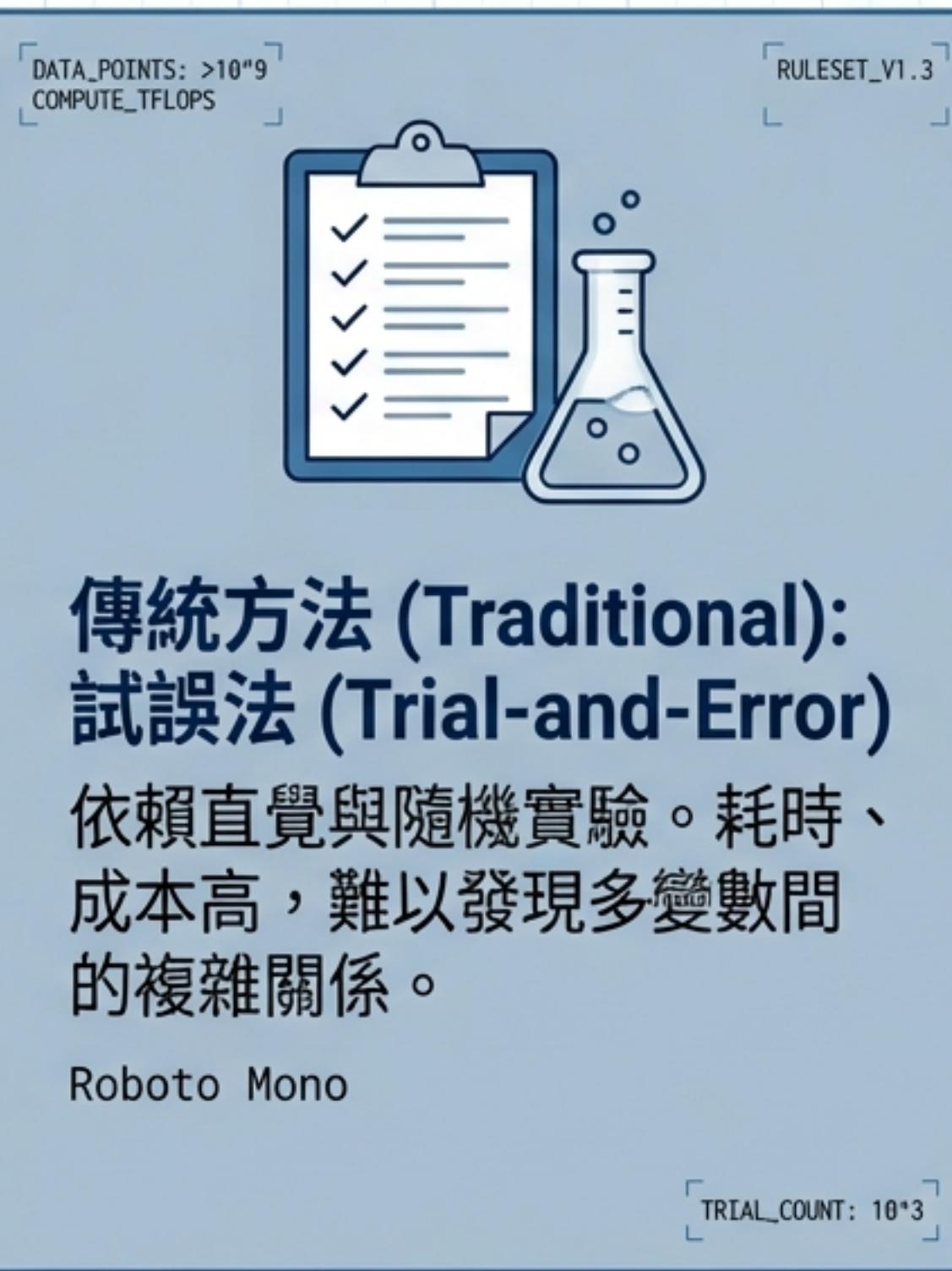
授課教師：莊曜楨 助理教授 | 過甲大學 化工系 智慧程序系統工程實驗室  
學期：114學年度第2學期



## 本單元學習目標 (Learning Objectives)

- 1. 理解 Apriori 演算法核心原理 (向下封閉性)
- 2. 掌握 mlxtend Python 套件實作
- 3. 應用於聚合物配方優化 (Polymer Formulation Optimization)

# 化工研發的典範轉移：為什麼需要關聯規則？



典範轉移  
(Paradigm Shift)

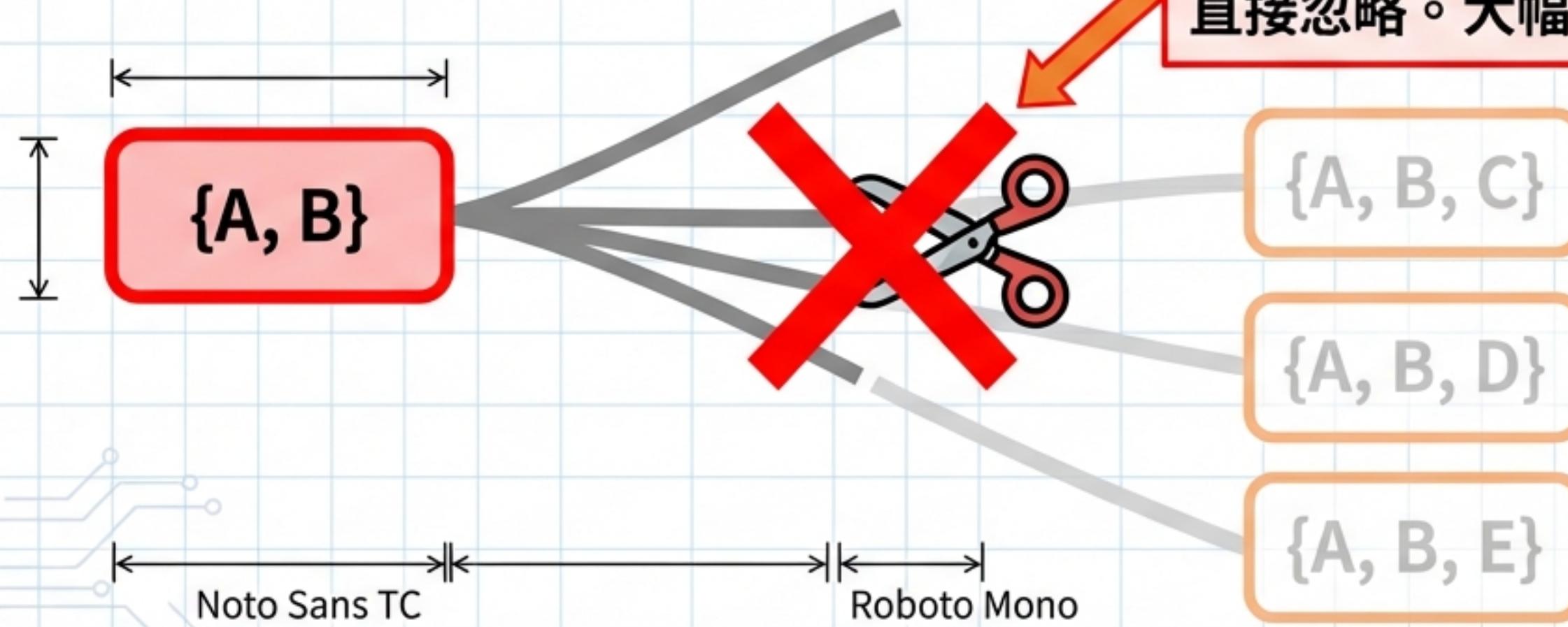
If {Catalyst A + Solvent X}, Then {High Yield}?



# 演算法核心：Apriori 性質與剪枝策略

**Apriori 性質 (Downward Closure):** 如果一個項目集是不頻繁的，那麼它的所有超集也必定是不頻繁的。

**剪枝 (Pruning) :** 不需計算，直接忽略。大幅提升運算效率。



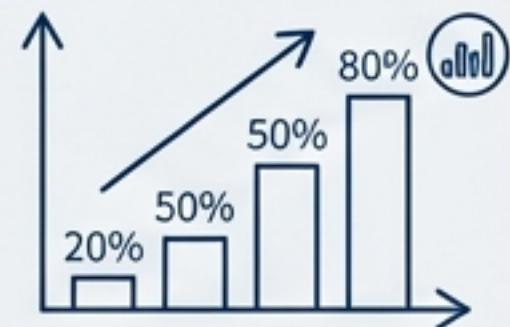
# 評估指標：解讀規則的三把尺

## 1. Support (支持度)

### 覆蓋率 / 頻率

規則在所有數據中出現的頻率。

⚙️ Engineering Insight  
這組配方多常被使用？



## 2. Confidence (置信度)

### 可靠性 / 機率

$$P(Y|X)$$

🔧 Engineering Insight  
使用了催化劑 A，有多大把握會高產率？

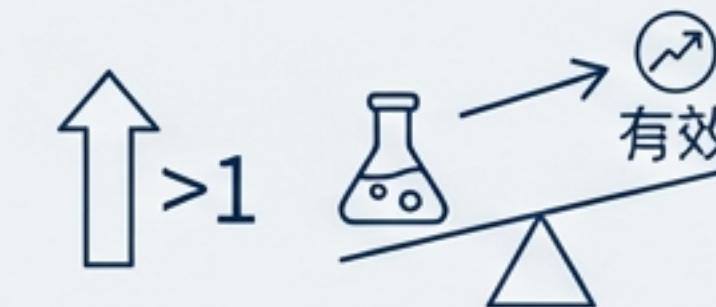


## 3. Lift (提升度)

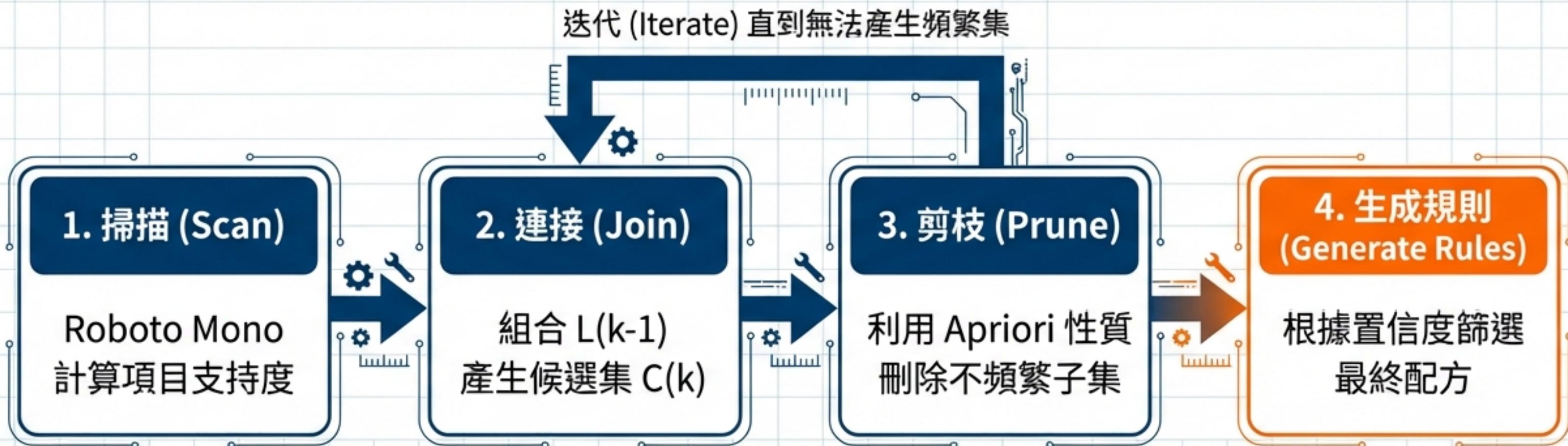
### 關聯強度

Lift > 1 表示正相關（有效）。

🧪 Engineering Insight  
這是真實的化學協同效應，還是隨機巧合？



# 演算法運作流程：逐層搜索



# Python 實作工具：mlxtend

## 核心函式庫

```
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import
association_rules

#
# 1. 找出頻繁集
frequent_itemsets = apriori(df,
min_support=0.1)

# 2. 產生規則
rules = association_rules(frequent_itemsets,
metric='lift')
```

## 數據格式要求：One-Hot Encoding

Catalyst A	Catalyst B	Solvent X
1	0	1

0 = Not Used, 1 = Used

# 演算法特性分析 (Pros & Cons)



## 優勢 (Pros)



可解釋性強  
(Explainable AI)：  
規則直觀 (If-Then)，  
易於向化工專家解釋。



稀疏數據友善：適合原  
料多但單一配方使用  
少的場景。

## 限制 (Cons)



I/O 開銷大：需多次  
掃描資料庫。



候選集爆炸：  
 $\text{Min\_support}$  設定過  
低時計算量呈指數增  
長。

## 適用場景 (Sweet Spot)



中小型數據集  
( $< 10,000$  筆)



實驗室配方分析



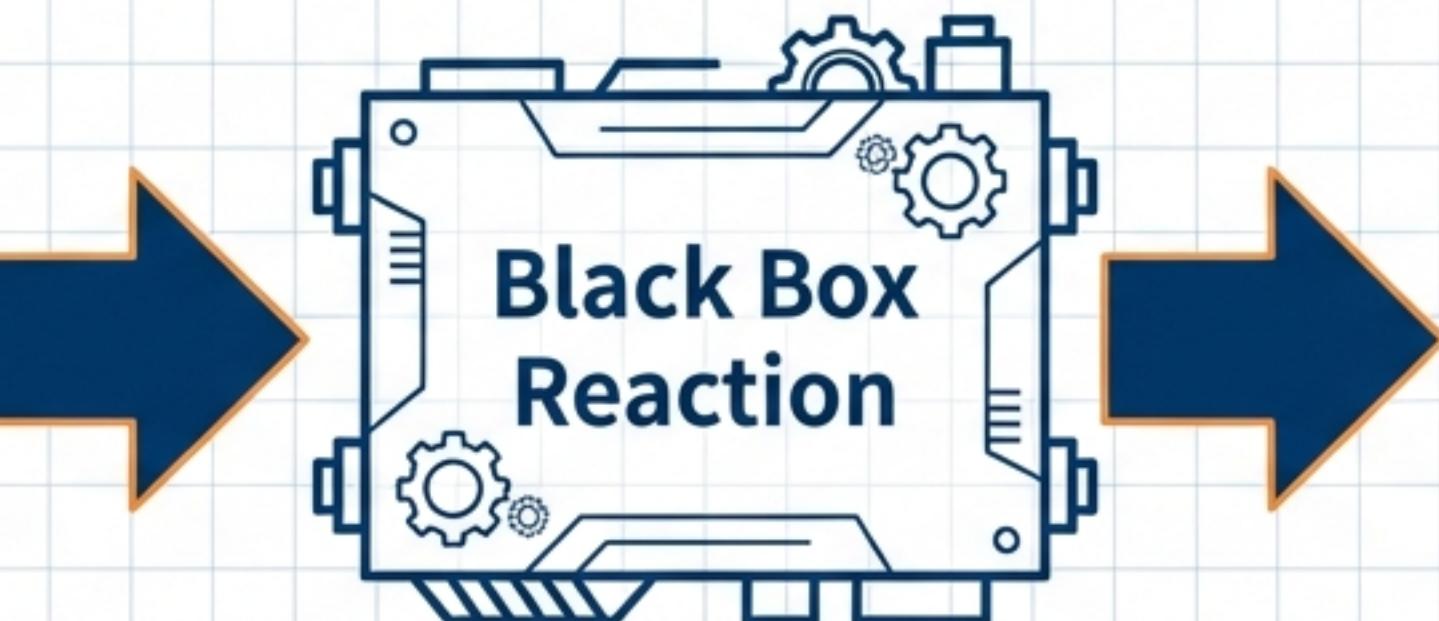
故障根因診斷

# 實戰案例：聚合物配方優化 (Polymer Optimization)

目標：從 200 筆歷史實驗數據中找出最佳配方

## Input Variables (X)

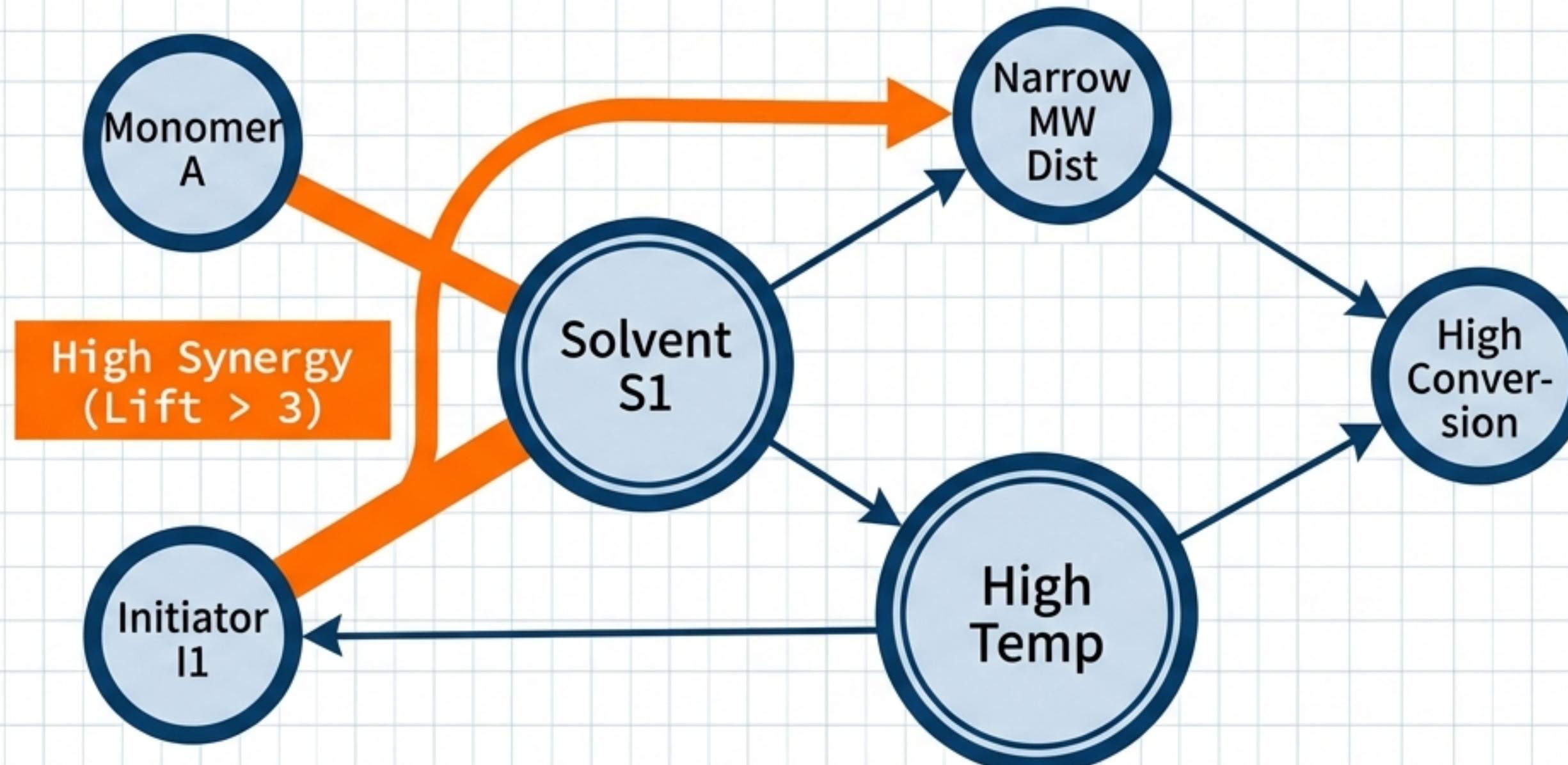
	單體 (Monomer A-E)
	引發劑 (Initiator I1-I3)
	溶劑 (Solvent S1-S3)
	操作條件 (Temp, Time)



## Quality Targets (Y)

	窄分子量分布 (Narrow MW Dist)
	高轉化率 (High Conversion)
	高純度 (High Purity)

# 視覺化關聯網絡 (Visualizing the Chemical Network)

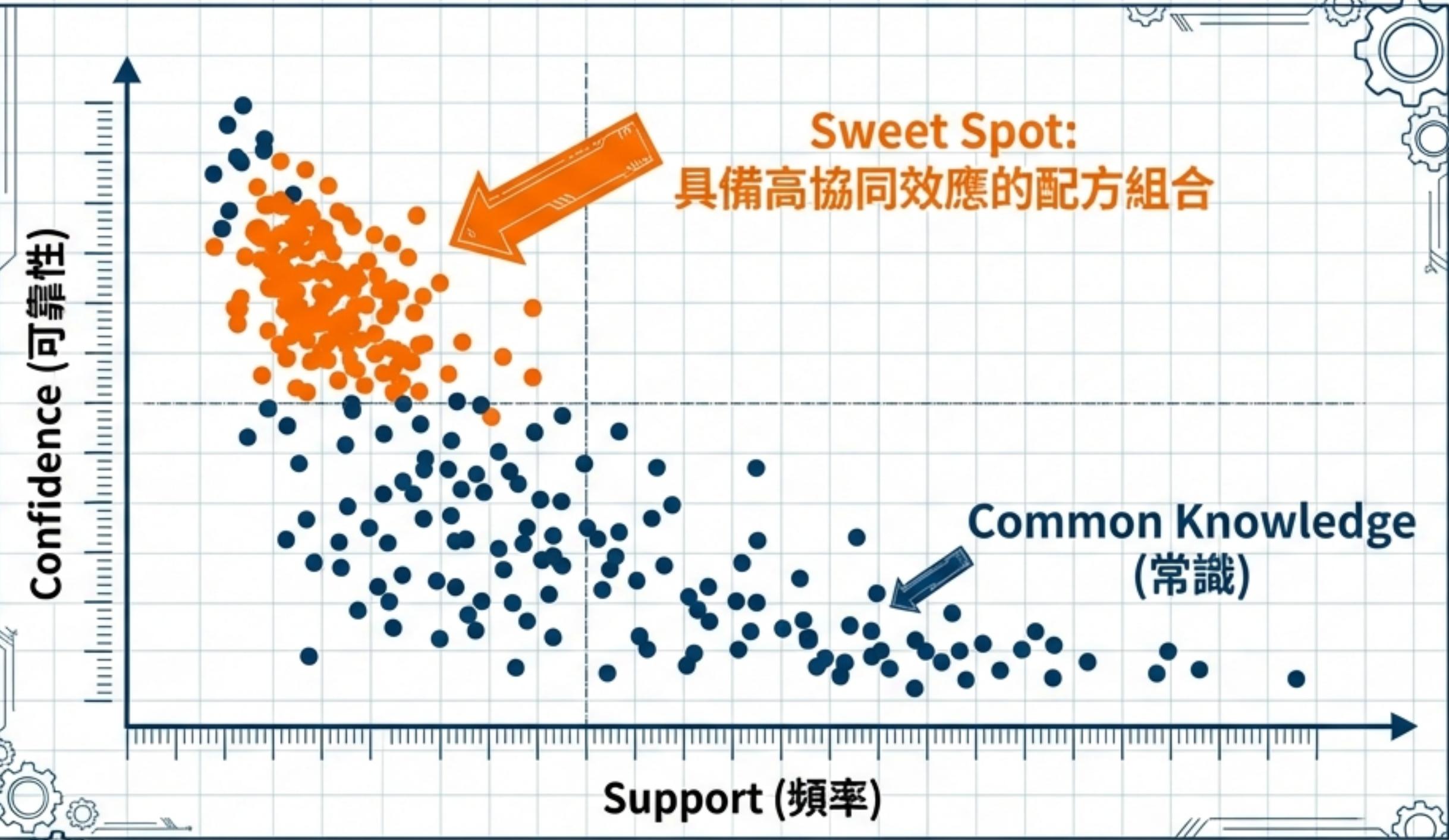


線條粗細代表 Lift 強度。中心節點代表配方中的關鍵驅動因子。

# 關鍵發現：高價值的配方規則 (Top Rules)

Rule (Recipe -> Result)	Lift (Strength)	Confidence (Reliability)	Insight
{Monomer_A + Initiator_I1} -> {Narrow_MW_Dist}	3.66	67.7%	關鍵協同效應
{Solvent_S1 + High_Temp} -> {High_Conversion}	3.76	82.6%	高溫下溶劑 S1 表現最佳
{Chain_Transfer_T2 + Medium_Temp} -> {High_Purity}	3.26	94.4%	中溫控制對純度至關重要

# 規則分佈與篩選策略 (Rule Distribution)



避免只看高頻率  
規則，要尋找高  
Lift 的隱藏洞察。

# 工程決策與行動建議 (Actionable Decisions)



# 實務建議：參數設定指南 (Tuning)

**Min\_Support**  
(頻率閾值)



**Min\_Support (頻率閾值)**

- Small Data: Set **0.1 - 0.3** (避免噪音)
- Big Data: Set **0.01 - 0.05** (挖掘稀有模式)

**Min\_Confidence**  
(信心閾值)



**Min\_Confidence (信心閾值)**

- Critical Safety: Set **> 0.9**
- R&D Exploration: Set **> 0.5**



**Warning:** 參數過低將導致計算量爆炸 (Computation Explosion)。

# 結語：從數據中提煉化學習智慧

**Apriori 演算法將工程師的『經驗法則』轉化為可量化的『數學規則』。**

揭示了肉眼看不見的協同效應 (Synergy)。

Next Step: Unit 08b - FP-Growth Algorithm (更高效的大數據挖掘)。

AI 不會取代化工工程師；  
但懂得使用 AI 的化工工程師將取代不懂的人。