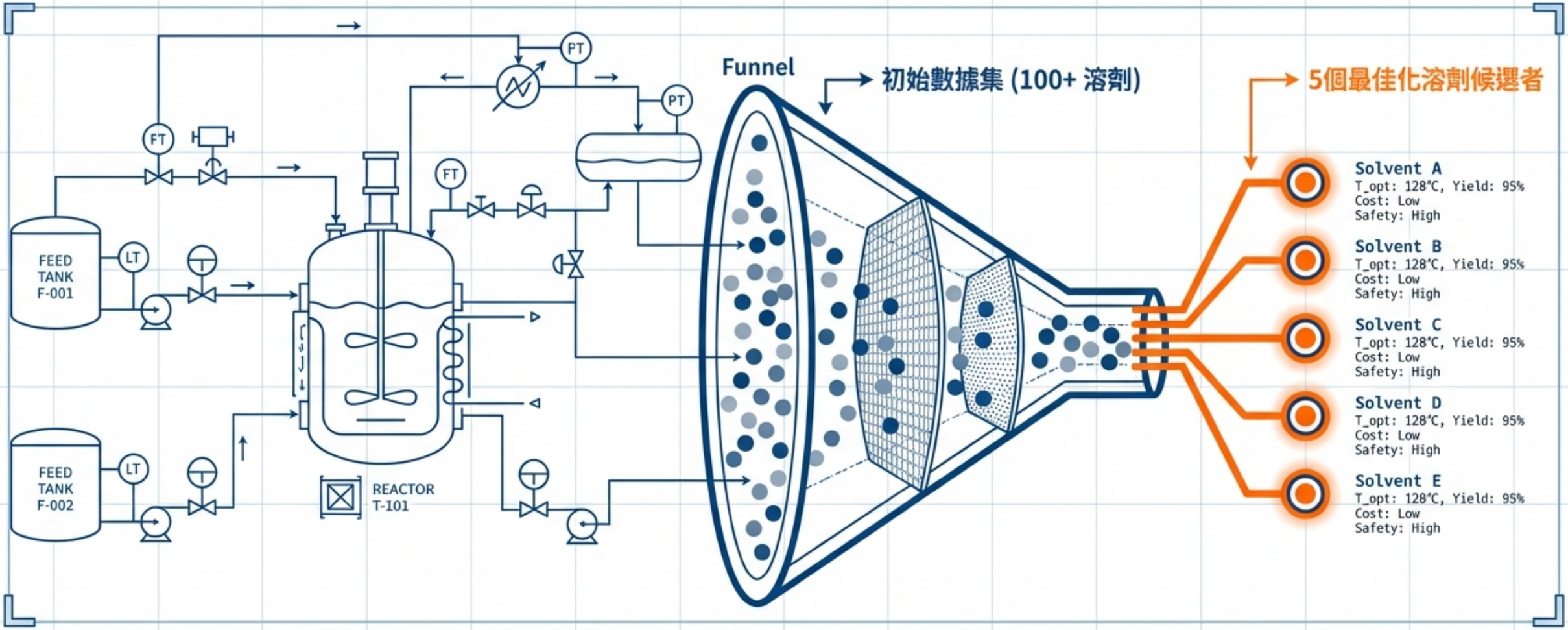
AI 在化工上的應用：溶劑篩選案例研究

Unit 09 - 從 100 到 5：非監督式學習的綜合應用



溶劑篩選的挑戰：多目標衝突與高昂成本

多重物理限制 (Conflicts)

- 高溶解度 (High Solubility) vs 低毒性 (Low Toxicity)
- 低沸點 (Low Boiling Point) vs 低揮發損耗 (Low Loss)
- 高性能 (High Performance) vs 低成本 (Low Cost)

方法比較 (Paradigm Shift)

傳統方法 (Traditional)

試誤法 (Trial & Error) - 高成本、依賴直覺、易受主觀影響。

AI 方法 (AI Approach)

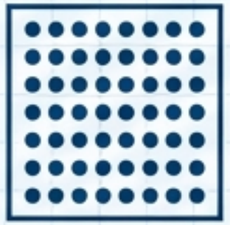
數據驅動、多維度評估、**快速篩選 (Fast Screening)**。

核心目標：如何在不進行 100 次實驗的情況下，找出那 5 個最佳候選？

AI 篩選漏斗：三種非監督式學習的協同效應

Input

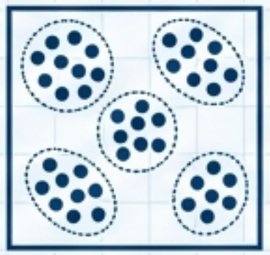
100
Solvents
Roboto Mono



Filter 1:
Clustering (分群)

Group by
Similarity

100 → 50
Candidates

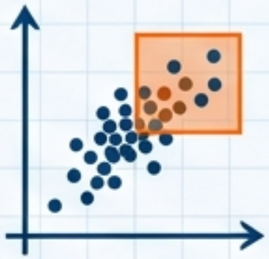


負責「廣度分類」，
排除表現差的群組

Filter 2:
PCA (降維)

Visual Navigation

50 → 34
Candidates

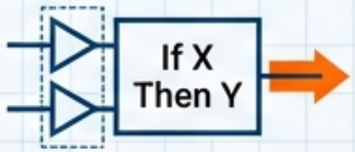


負責「深度視覺化」，
在二維平面鎖定高性能區

Filter 3:
Rules (規則)

Logic Discovery

34 → 5
Candidates



負責「挖掘規則」，
驗證篩選邏輯

Output

Top 5
Candidates
Roboto Mono

數據基礎與前處理：工欲善其事，必先利其器

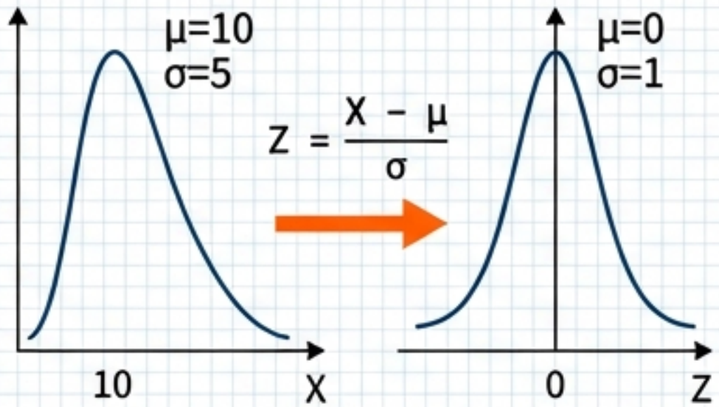
原始特徵 (Raw Features)

Variable	Example Value
Solubility	0.5
BoilingPoint	100
Viscosity	0.5
Density	19.75
Toxicity	0.38
Cost	200
Polarity	0.5
DielectricConstant	0.194
RefractiveIndex	0.001
FlashPoint	19.5

Scale Difference!
(e.g., BP ~100 vs Viscosity ~0.5)

Pre-processing

標準化 (Standardization)



Z-score Normalization.
使所有變數均值為 0，標準差為 1。

Feature Engineering

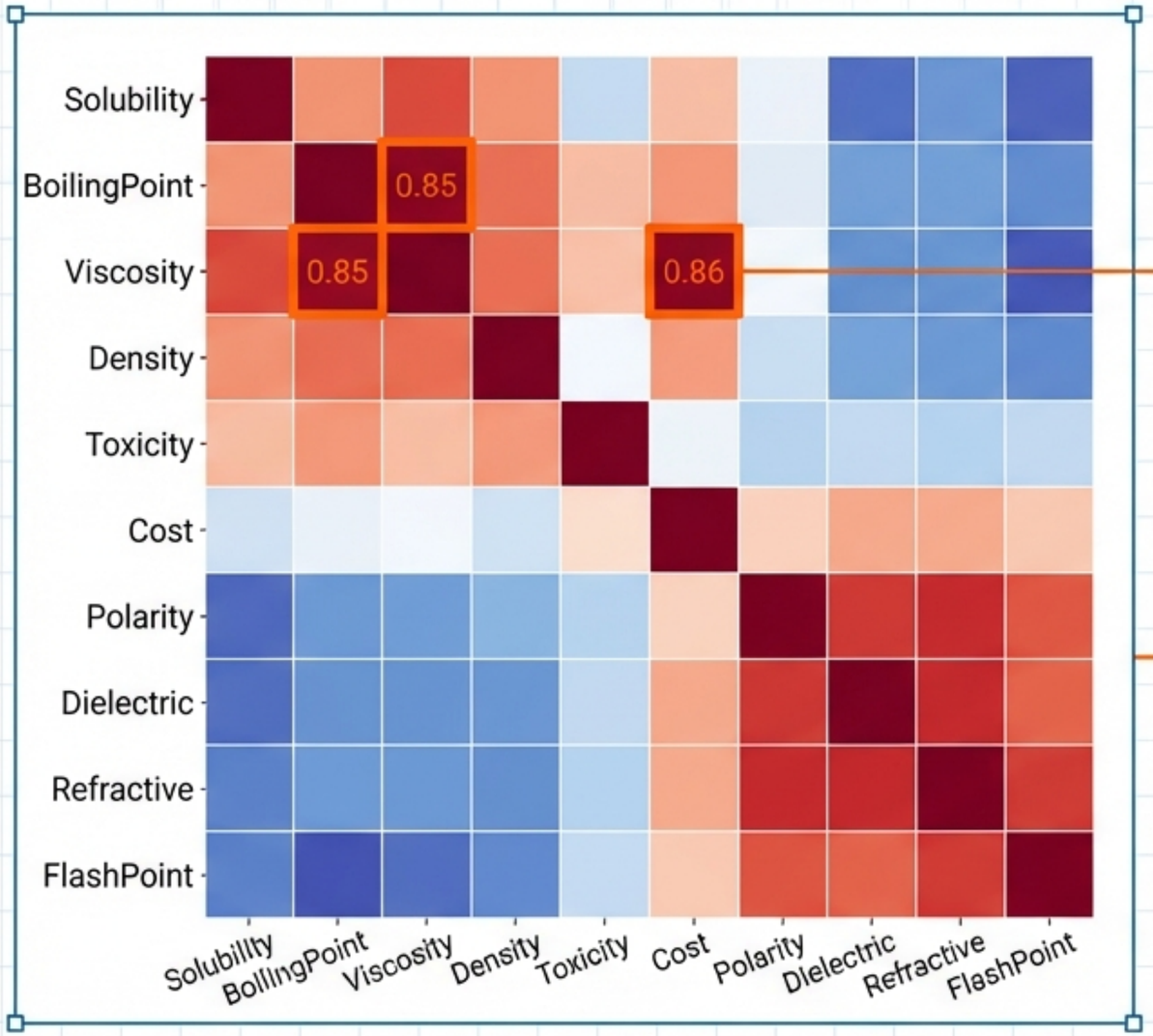
衍生指標 (Derived Indices)

PI (Performance) = Solubility / Viscosity × Toxicity

EI (Economic) = Solubility / Cost

SI (Safety) = FlashPoint / Toxicity

數據會說話：變數間的相關性分析



成本驅動因素 (Cost Drivers)

Viscosity ↔ Cost (r=0.86)
重質溶劑通常更昂貴。

物理性質聯動 (Physical Link)

BoilingPoint ↔ Viscosity (r=0.85)。
高沸點通常伴隨高黏度。

多重共線性 (Multicollinearity)

變數間高度相關，這正是降維分析 (PCA) 的最佳應用場景。

階段 1：物以類聚 (Clustering Analysis)

Method: K-Means (K=4)

Cluster 1: High Risk (Eliminate ❌)

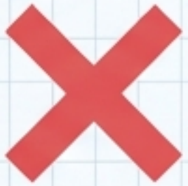
Boiling Point: 179°C | Toxicity: High

回收困難，毒性過高。

Cluster 2: High Polarity (Eliminate ❌)

Polarity: 8.1 | Dielectric: 25

與水互溶性過高，不利萃取。



Cluster 0: Economic (Keep ✅)

Cost: Low | Polarity: Low

經濟型選擇，適合成本敏感應用。

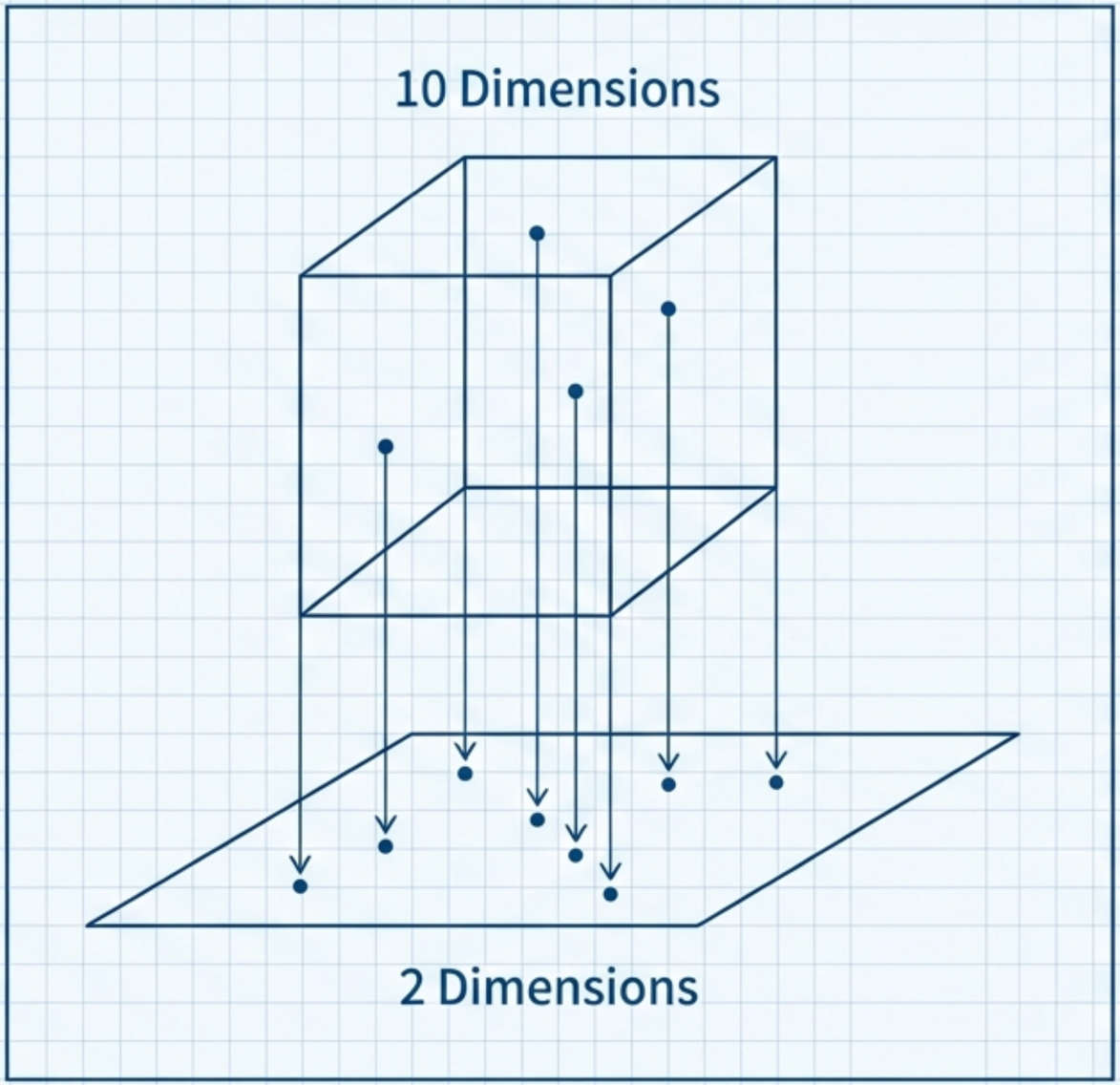
Cluster 3: Balanced (Keep ✅)

Viscosity: 0.54 | Toxicity: Low

綜合性能最佳，優質候選。

Result: 100 Candidates → 50 Candidates (Cluster 0 & 3)

階段 2：降維原理與物理意義 (PCA Principles)



Loadings Analysis

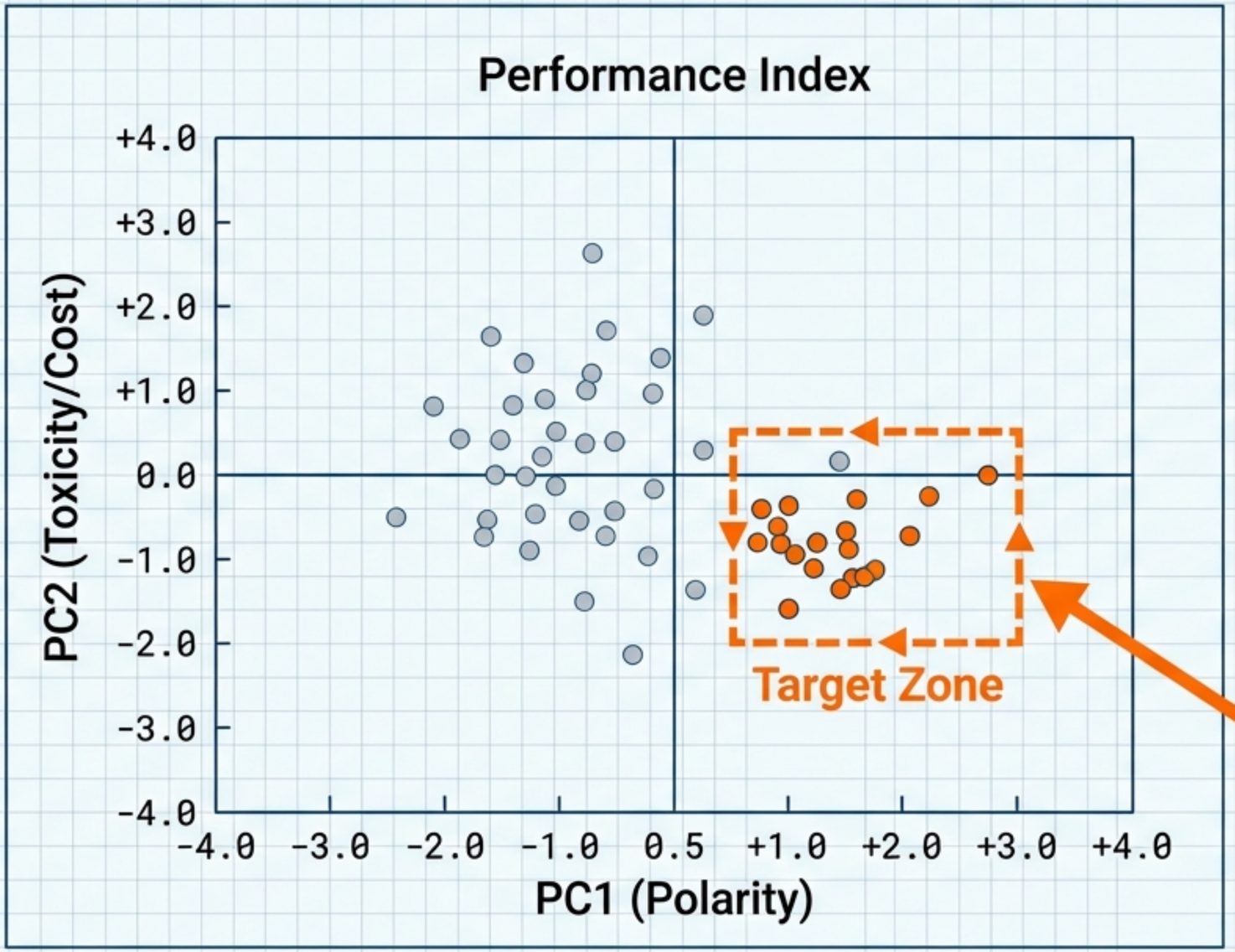
PC1 - The Polarity Axis (42.2% Variance)

PC2 - The Cost/Risk Axis (11.9% Variance)

Key Insight Box

PCA 就像是一個指南針，PC1 指示極性方向，
PC2 指示風險方向。

階段 2：在 PC1-PC2 平面上尋找黃金區域



篩選邏輯 (Screening Logic)

目標：保留高性能且低風險的溶劑。

Result: 50 Candidates
Candidates → 34 Candidates

1. High Performance Index
2. Low PC2 (Low Toxicity)

階段 3：挖掘隱藏邏輯 (Association Rules)

Method: Apriori Algorithm

數據離散化 (Discretization) : High / Medium / Low

Rule Card

```
IF   { BoilingPoint = Medium }
THEN { Toxicity = Medium }
```

`Lift = 1.288 | Confidence = 90.9% | Support = 0.294`

工程意義：不需要為了高性能去冒險。
'Medium' (80-120°C) 是平衡回收成本與安全性的甜蜜點。

Result: Validates strategy to target **balanced solvents**. Candidates refined to **Top 5**.

決策時刻：建立綜合評分標準 (Final Scoring Model)

$$\text{Composite Score} = 0.4 \times \text{PI} + 0.3 \times \text{EI} + 0.3 \times \text{SI}$$

40% Performance (PI)

效率至上 (Efficiency First).
Solubility / Viscosity.

30% Economy (EI)

經濟可行 (Viability).
Solubility / Cost.

30% Safety (SI)

安全無虞 (Safety).
FlashPoint / Toxicity.

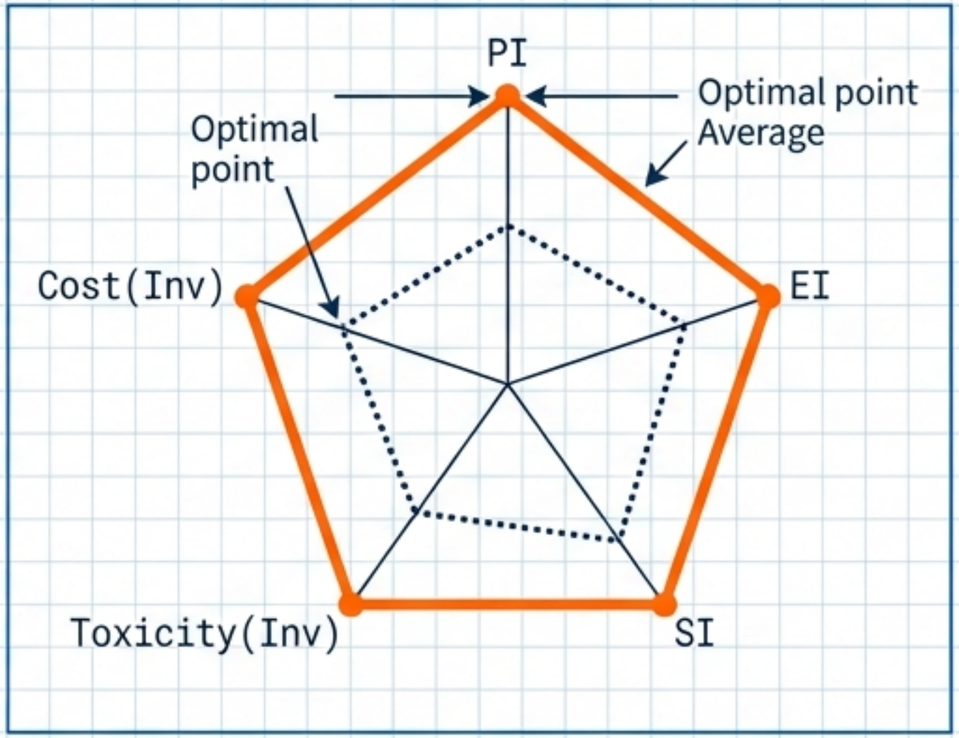
所有指標在計算前均經過歸一化處理 (Normalized).

最終結果：Top 5 最佳候選溶劑

Top 5 Candidates

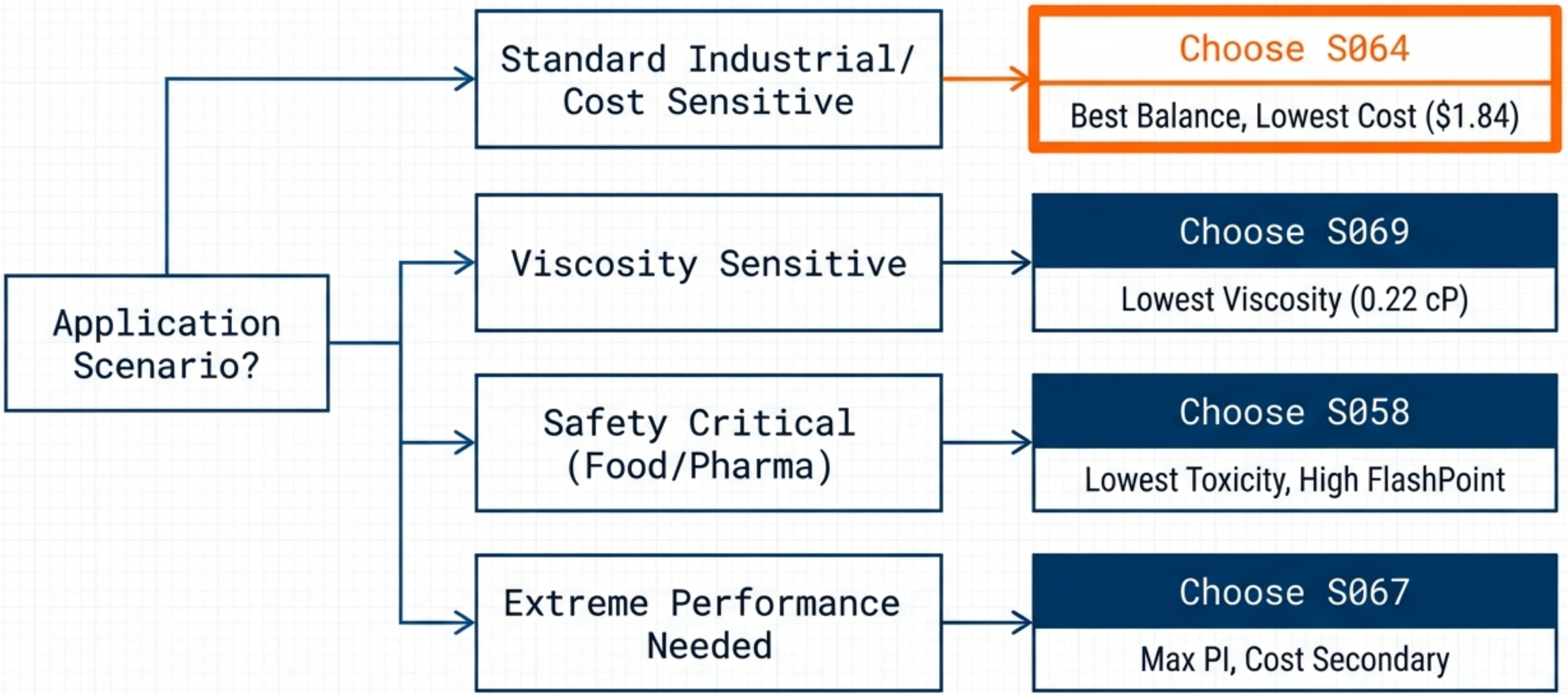
Rank	ID	Score	Toxicity	Cost	Note
1 |	S064	0.68	2 (Low)	$1.84	Best Balance
2 |	S069	0.57	4 (Med)	$1.65	Best PI
3 |	S058	0.57	2 (Low)	$2.86	Safest
4 |	S067	0.47	4 (Med)	$4.50	Max Perf.
5 |	S052	0.47	3 (Low)	$2.11	High Sol.

Champion: S064



Score: 0.68 (19% higher than 2nd place). No weaknesses.

情境分析：沒有最好的溶劑，只有最適合的溶劑



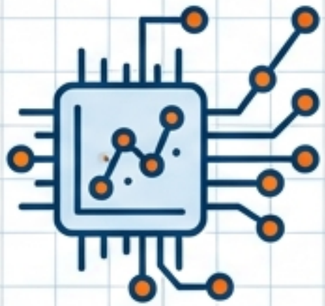
方法論協同效應：為何需要組合拳？

Method	Strength	Blind Spot	Role in Funnel
Clustering	Groups similar items	Doesn’t rank ‘better’ vs ‘worse’	Filter Noise (Stage 1)
PCA	Visual Map of space	Loses ~45% variance info	Navigation (Stage 2)
Association Rules	Discovers Logic	Hard to rank specific candidates	Validation (Stage 3)

Combined Strength: Filter → Map → Validate → Rank

領域知識的價值：AI 與化工工程師的協作

Domain Knowledge
(化工專業)



Data Science
(數據技能)

Value Creation
(價值創造)

工程師的角色 (The Engineer's Role)

- 定義指標 (Defining PI, EI, SI)
- 解釋物理意義 (Interpreting PC1 as Polarity)
- 權衡決策 (Trade-off: Cost vs Safety)

AI 的角色 (The AI's Role)

- 快速處理 (Processing 100 samples)
- 發現關聯 (Correlations: Viscosity ↔ Cost)
- 降維視覺化 (Mapping 10D to 2D)

“AI 不會取代化工工程師，但懂得使用 AI 的工程師將取代不懂的人。”

結論與下一步：從模擬走向實驗

Impact Summary

95% Reduction
Experimental Effort

Seconds vs Weeks
Screening Speed

Next Steps (Action Plan)

Lab Verification (Test S064)
Noto Sans TC & Roboto Mono, Deep Engineering Blue

Feedback Loop (Update Model)
Noto Sans TC & Roboto Mono, Deep Engineering Blue

Scale Up (Pilot Test)
Noto Sans TC & Roboto Mono, Deep Engineering Blue

Unit 09 End. Next: Unit 10 - Supervised Learning.
Roboto Mono, Deep Engineering Blue