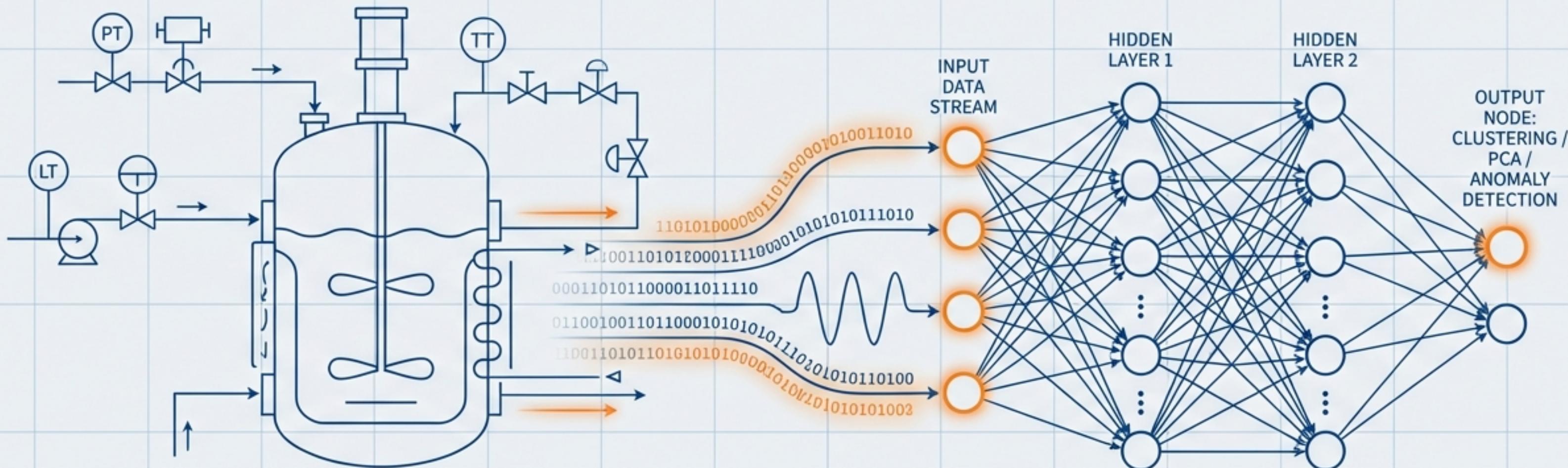


AI 在化工上的應用：綜合案例研究 (Unit 09)

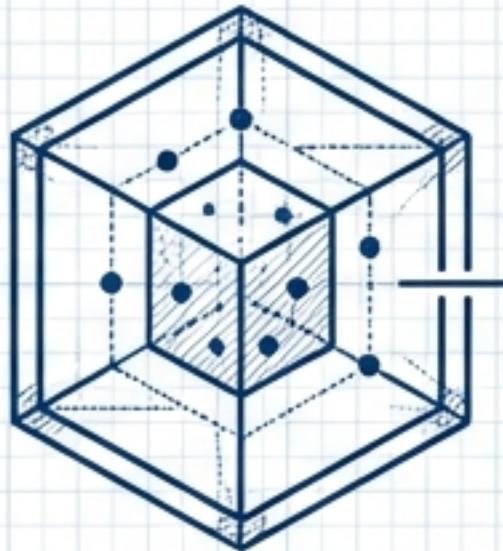


從單點技術到系統整合：田納西-伊士曼製程 (TEP) 實戰

Building the Pipeline: Integrating Clustering, PCA, and Anomaly Detection

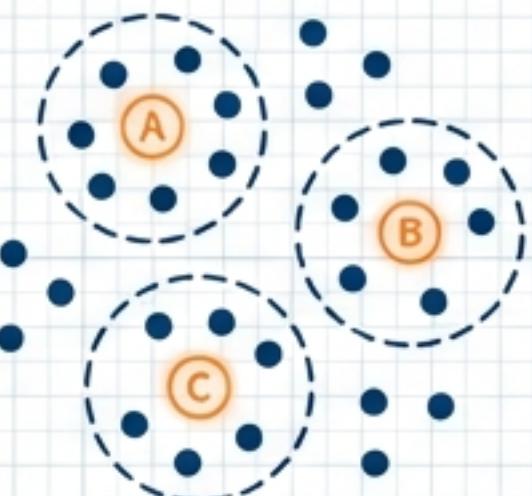
授課教師：莊曜禎 助理教授 | 2026 | 逢甲大學 智慧程序系統工程實驗室

為什麼需要整合分析？



降維
(Dimensionality Reduction)

`D_high` → `D_low` : Simplify
Complex Data Structures



分群
(Clustering)

`Group A`, `Group B`, `Group C`:
Identify Patterns & Segments



異常檢測
(Anomaly Detection)

`Outlier` : `X=12.5`, `Y=8.1` :
Detect Deviations & Risks



完整洞察
(Complete Insight)

`DISCOVERY` : [HIDDEN PATTERN],
[CRITICAL INSIGHT], [ROOT CAUSE]

單一方法的局限

單一模型可能無法全面揭示複雜數據中的模式。

互補性

降維簡化結構，分群識別模式，異常檢測發現偏離。

驗證性

不同方法的結果相互印證（例如：分群邊界應與異常檢測結果一致）。



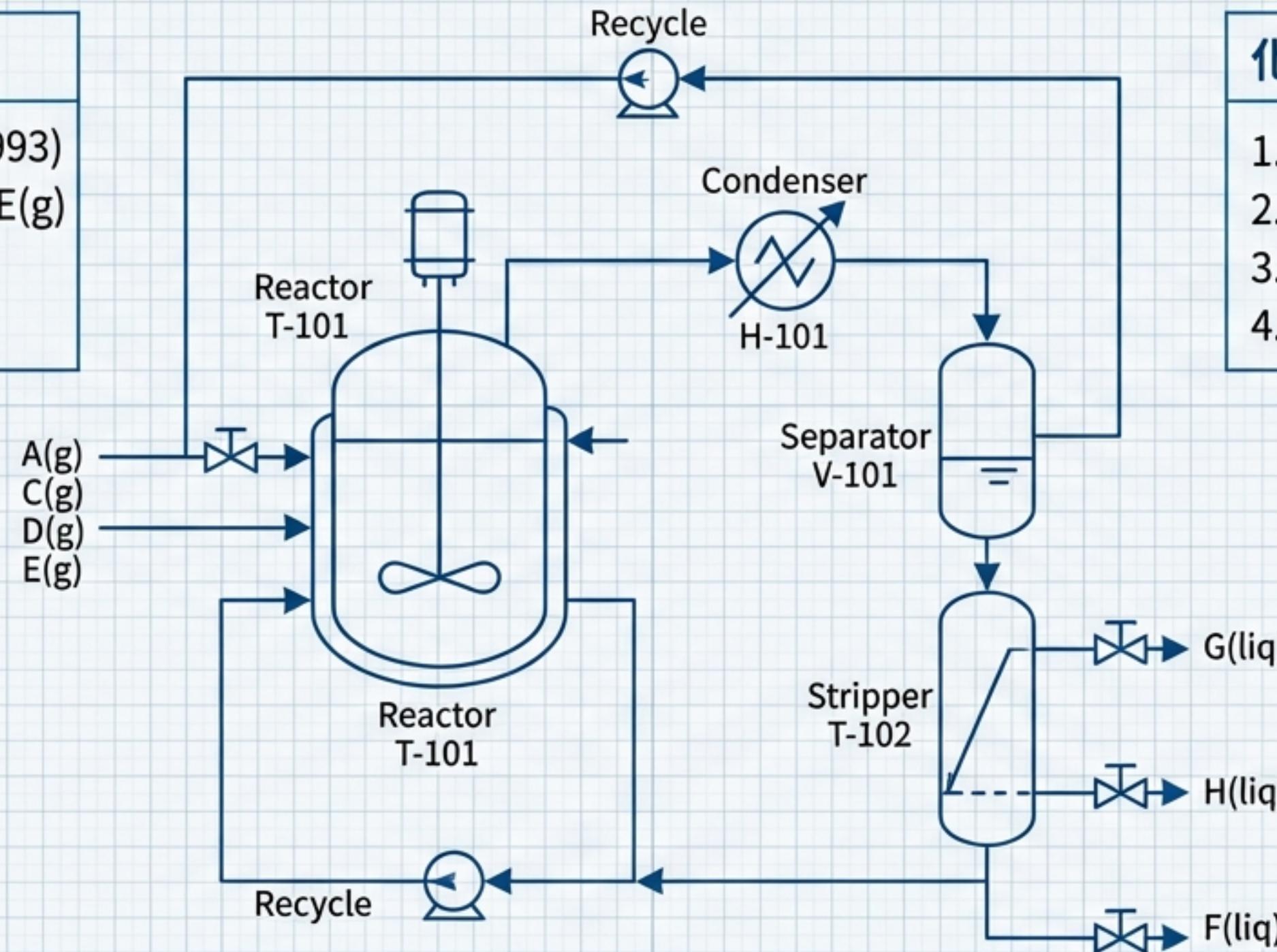
就像建造化工廠需要多個單元操作配合，數據分析也需要多種演算法協同工作。



案例背景：田納西-伊士曼製程 (TEP Benchmark)

製程概況

- 來源: Downs & Vogel (1993)
- 反應物: A(g), C(g), D(g), E(g)
- 產物: G(liq), H(liq)
- 副產物: F(liq)



化學反應 (Reactions)

- $A + C + D \rightarrow G$ (Product 1)
- $A + C + E \rightarrow H$ (Product 2)
- $A + E \rightarrow F$ (By-product)
- $3D \rightarrow 2F$ (By-product)

數據規格說明 (Data Specification)

41 測量變數 (Measured Variables - XMEAS)

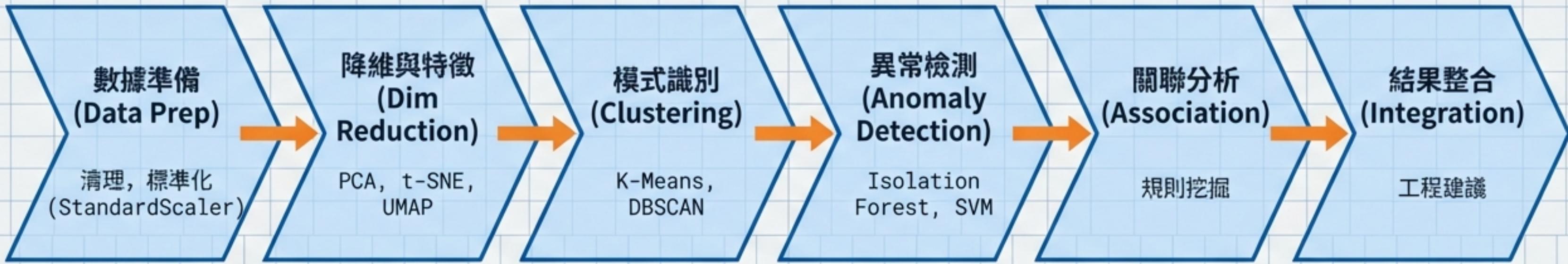
- 連續變數 (Continuous Variables - 1-22) Roboto Mono
 - 溫度 (T - XMEAS 1-22)
 - 壓力 (P - XMEAS)
 - 流量 (F - XMEAS)
 - 液位 (L - XMEAS)
- 取樣分析 (Sampled Analysis - 23-41) Roboto Mono
 - 組成分析 (Composition - XMEAS 23-41)
- 採樣頻率 (Sampling Frequency) Every 6-15 min
 - 每 6-15 分鐘

11 操作變數 (Manipulated Variables - XMV)

- 控制閥 (Control Valves - XMV) Roboto Mono
 - 進料閥 (Feed Valve)
 - 蒸汽閥 (Steam Valve)
 - 冷卻水閥 (Coolant Valve)
- 設備設定 (Equipment Settings - XMV) Roboto Mono
 - 攪拌器速度 (Agitator Speed)

總維度 (Total Dimensions): 52 變數 | 故障場景: 21 種 (IDV 1-21) + 正常操作 (IDV 0)

綜合分析工作流程 (The Analysis Pipeline)



階段 1 & 2：數據預處理與降維 (Preprocessing & PCA)

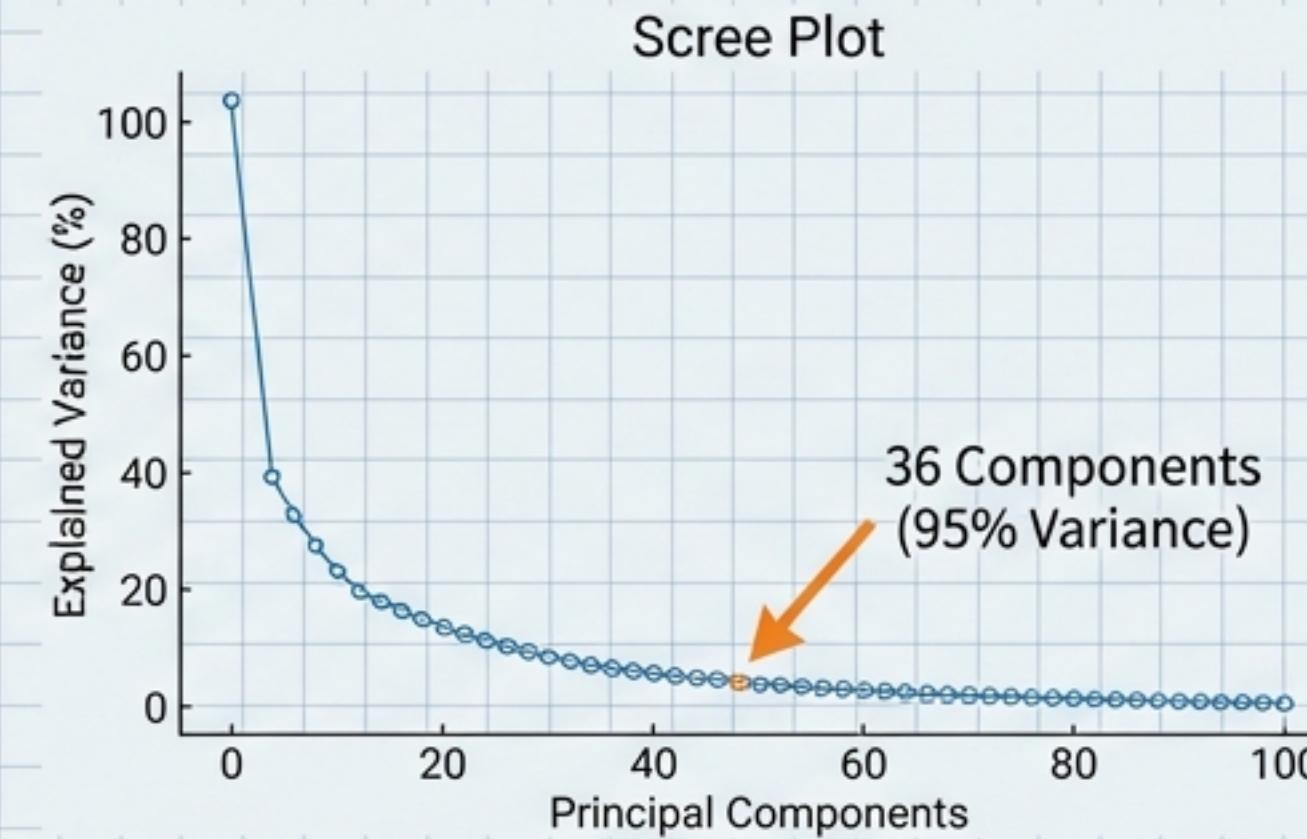
數據準備 (Data Preparation)



$$\frac{x - \mu}{\sigma}$$

- 原始維度: 52 變數
- 目標: 去除雜訊，保留訊號

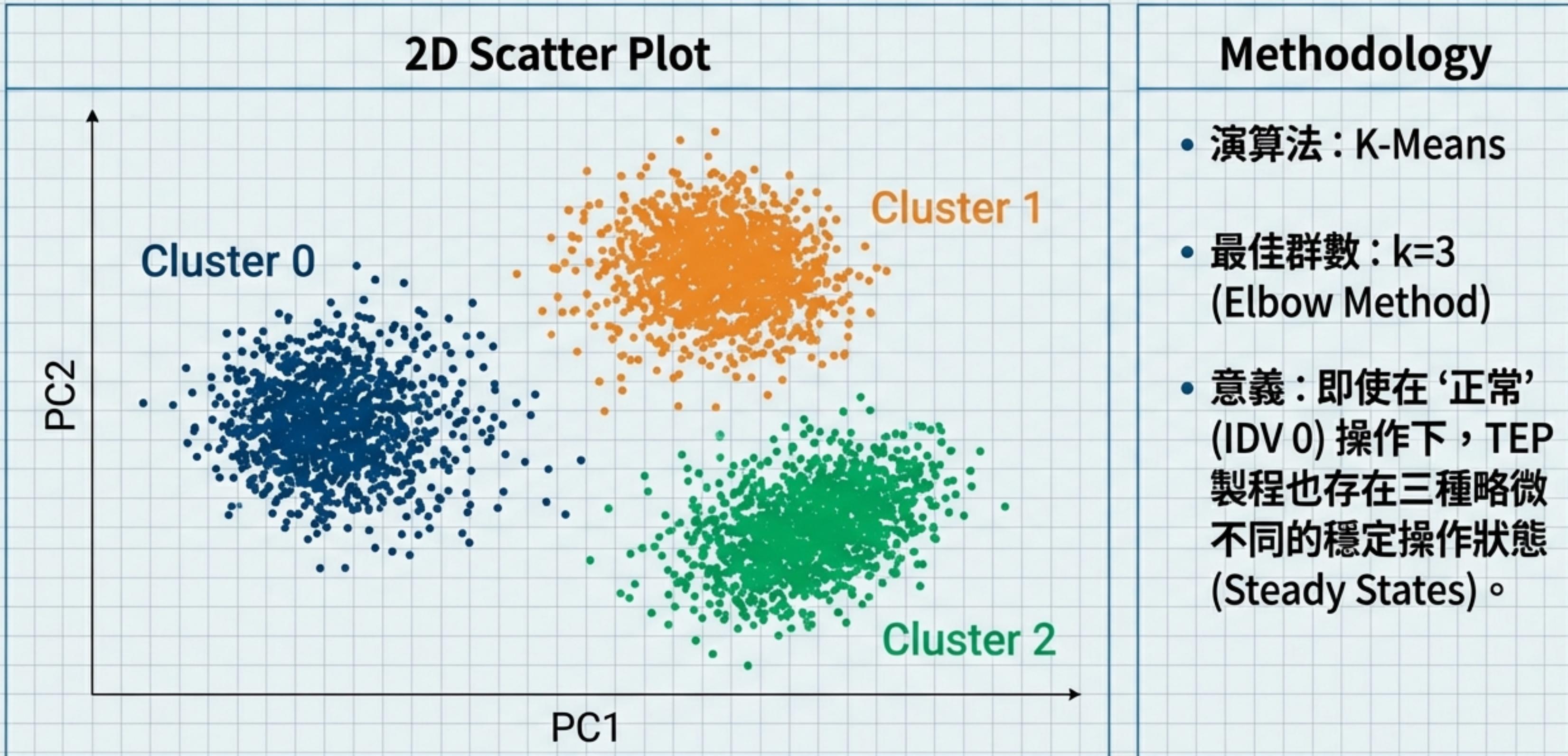
PCA 分析結果 (PCA Analysis Results)



PC1-PC3 貢獻最大 (~48%)，捕捉主要的
製程動態。

雖然有 52 個傳感器，但實際的獨立訊息維度約為 36。

階段 3：操作模式識別 (Operational Mode Clustering)



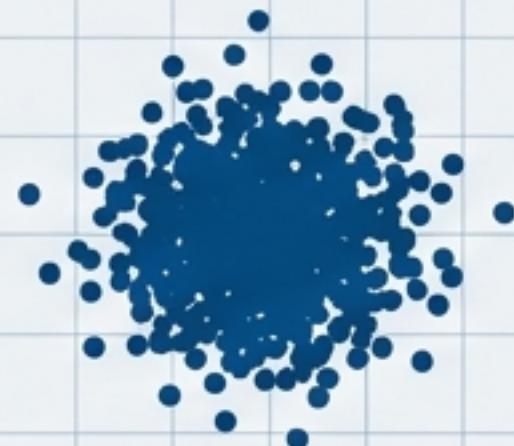
階段 4：異常檢測與監控 (Anomaly Detection)



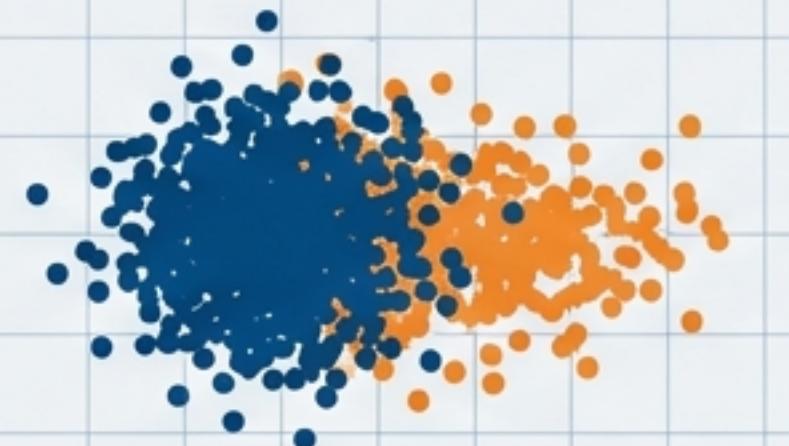
- 演算法: Isolation Forest (孤立森林)
- 參數: Contamination = 0.05 (預設 5% 異常率)
- 機制: 比較 Isolation Forest, One-Class SVM, LOF 以建立共識。
- 洞察: 異常點通常是隨機波動，但在故障發生時會形成連續的報警訊號。

實戰演練：故障診斷 (Fault Diagnosis - IDV 1)

正常操作 (Normal Operation)

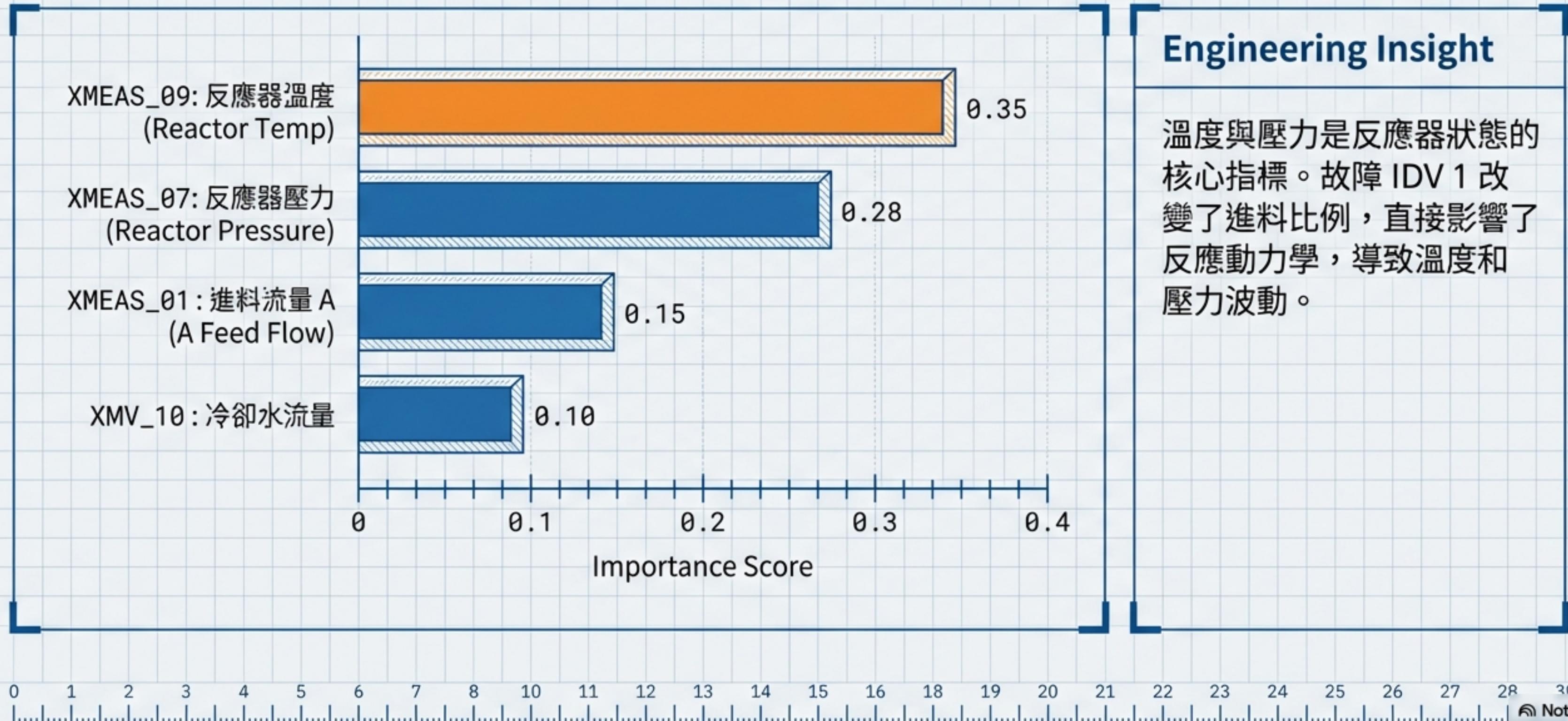


故障發生 (Fault IDV 1)

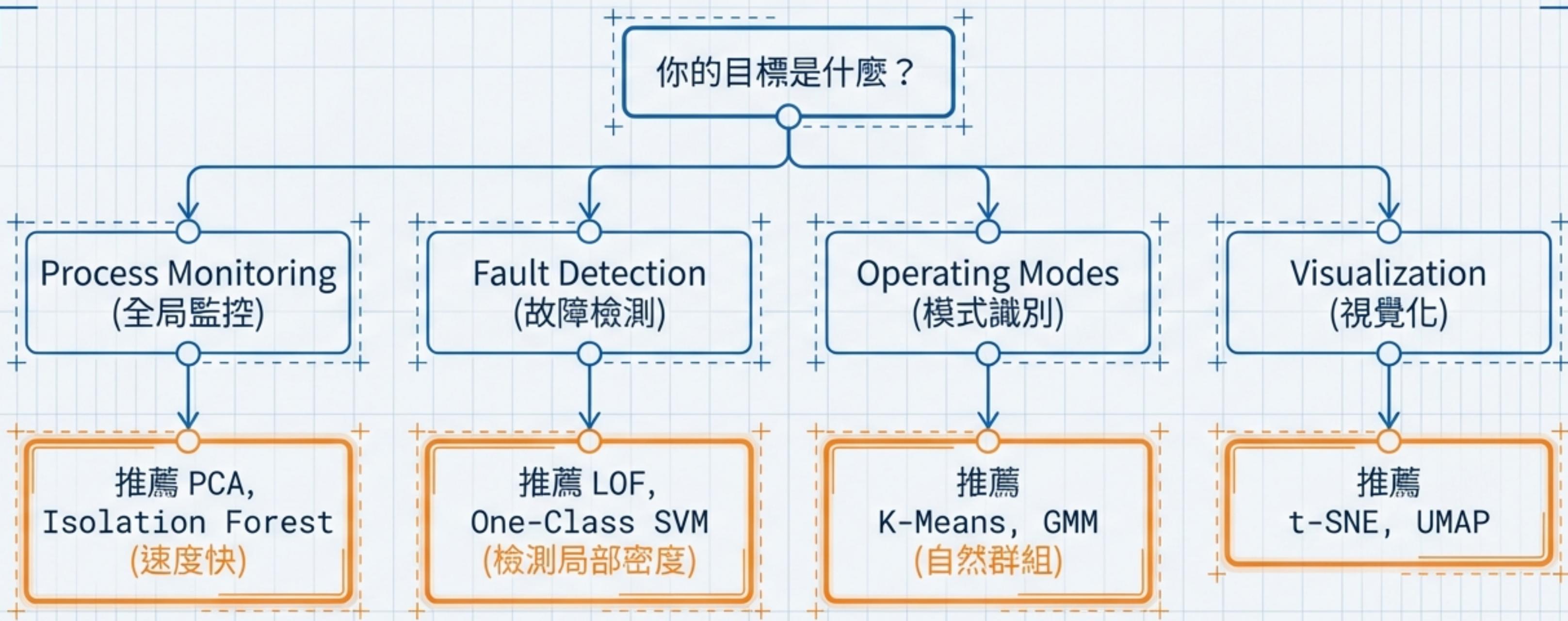


- 場景: IDV 1 (A/C Feed Ratio Change)
- 準確率 (Accuracy): ~85% (Elliptic Envelope)
- 召回率 (Recall): ~88% (Conservative Strategy - 寧可誤報，不可漏報)
- 結論: 模型成功區分了正常操作與故障偏移。

關鍵特徵與變數重要性 (Feature Importance)



工程師指南：如何選擇正確的方法？(Method Selection Guide)



不同的化工問題需要對應不同的演算法類型。

現實世界的挑戰 (Real-World Challenges)

化工廠不是 Kaggle 競賽，物理世界有其限制。

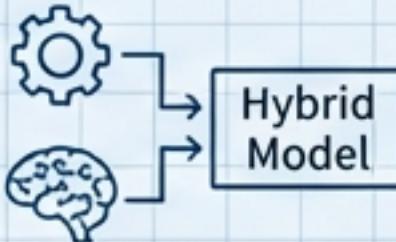
物理限制 (Physics Constraints)



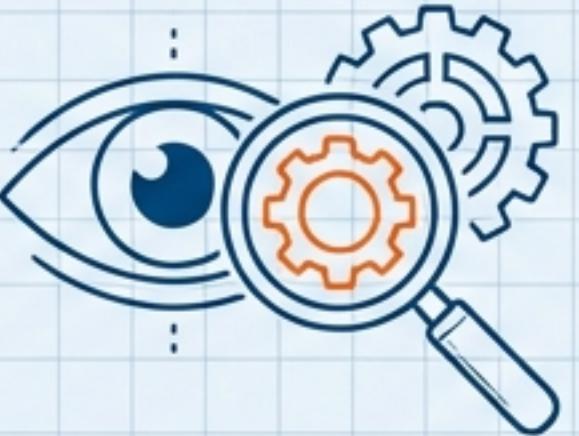
純數據模型可能會預測出違反質量守恆 (Mass Balance) 的結果。

→ 解法 (Solution):

混合建模 (Hybrid Modeling)



可解釋性 (Interpretability)



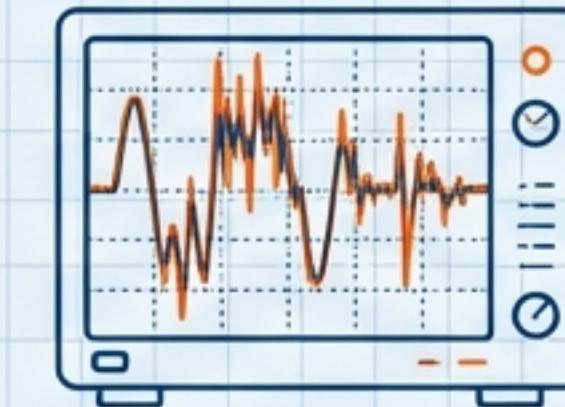
操作員需要知道 “為什麼” 報警。

→ 解法 (Solution):

SHAP Values, Contribution Plots



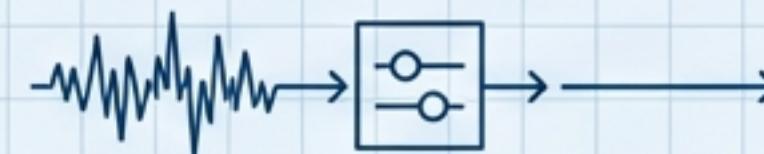
數據品質 (Data Quality)



現實數據充滿雜訊、缺失值和時間延遲。

→ 解法 (Solution):

穩健的預處理與濾波 (Robust Preprocessing & Filtering)



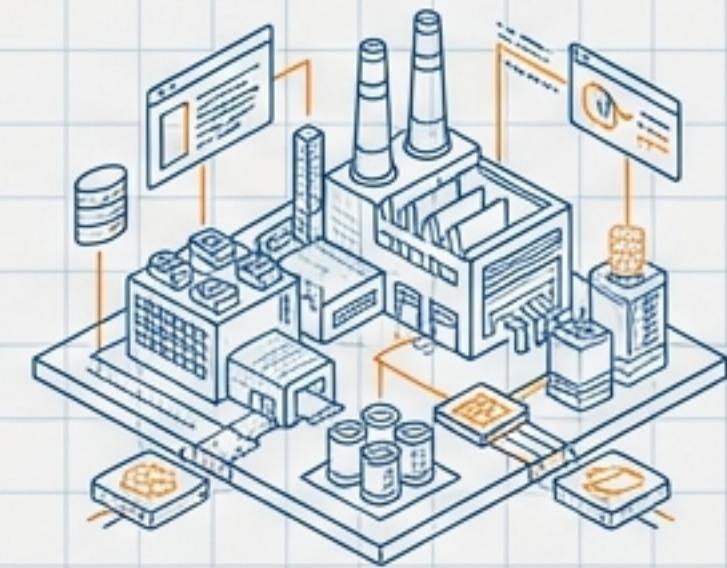
商業價值與工程效益 (Engineering Value)



預測性維護
(Predictive Maintenance)
提前識別異常
→ 減少非計畫停機



製程優化
(Process Optimization)
識別最佳操作模式
→ 節能降耗



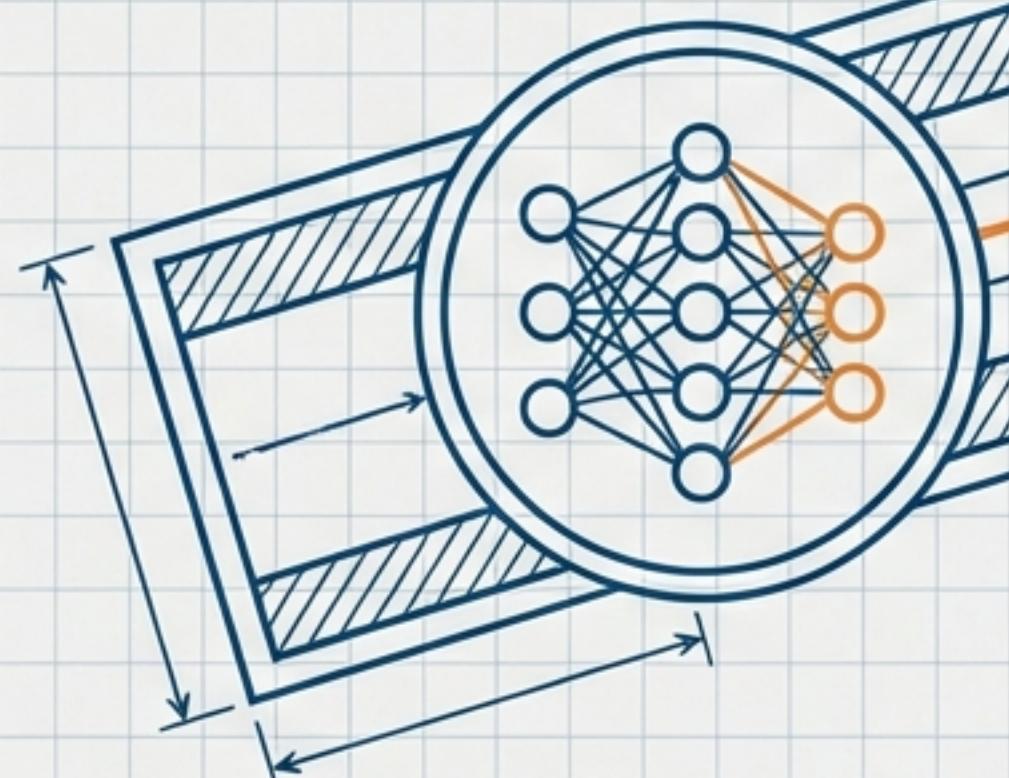
數位孿生
(Digital Twin)
數據模型是建立工廠
數位分身的基礎

未來展望：邁向智慧製造 (Future Outlook)

Deep Learning

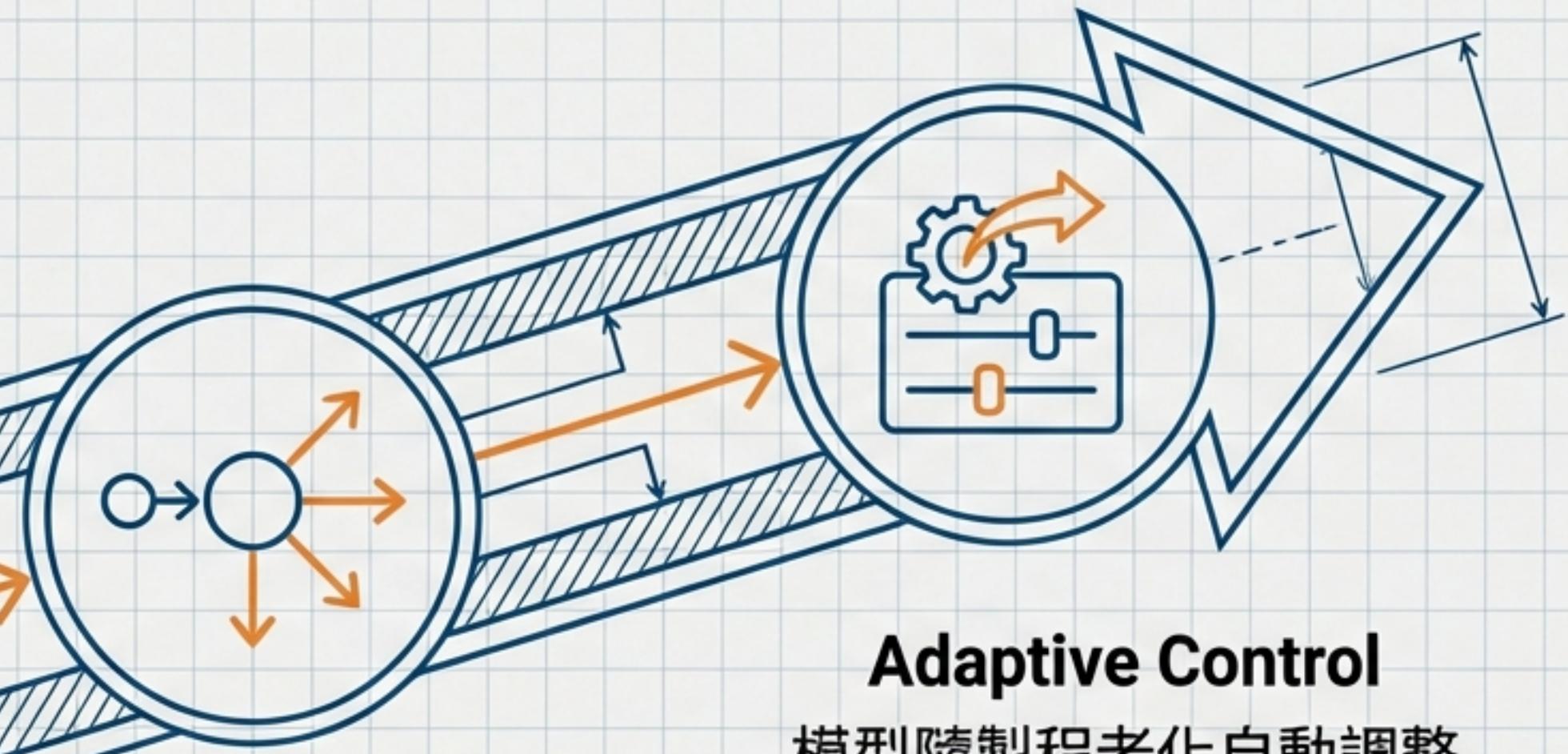
LSTM/Autoencoders

用於處理時間序列動態。



Causal Inference

從“相關性”邁向
“因果關係”。



Adaptive Control

模型隨製程老化自動調整
(Online Learning)。

數據分析是工業 4.0 的大腦。

結語：您是未來的定義者 (Conclusion)

“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。”

Next Steps (Actionable)



- 完成 Jupyter Notebook 實作 (Unit 09)
- 嘗試調整參數 (Contamination, k-clusters)
- 將方法應用於您自己的製程數據