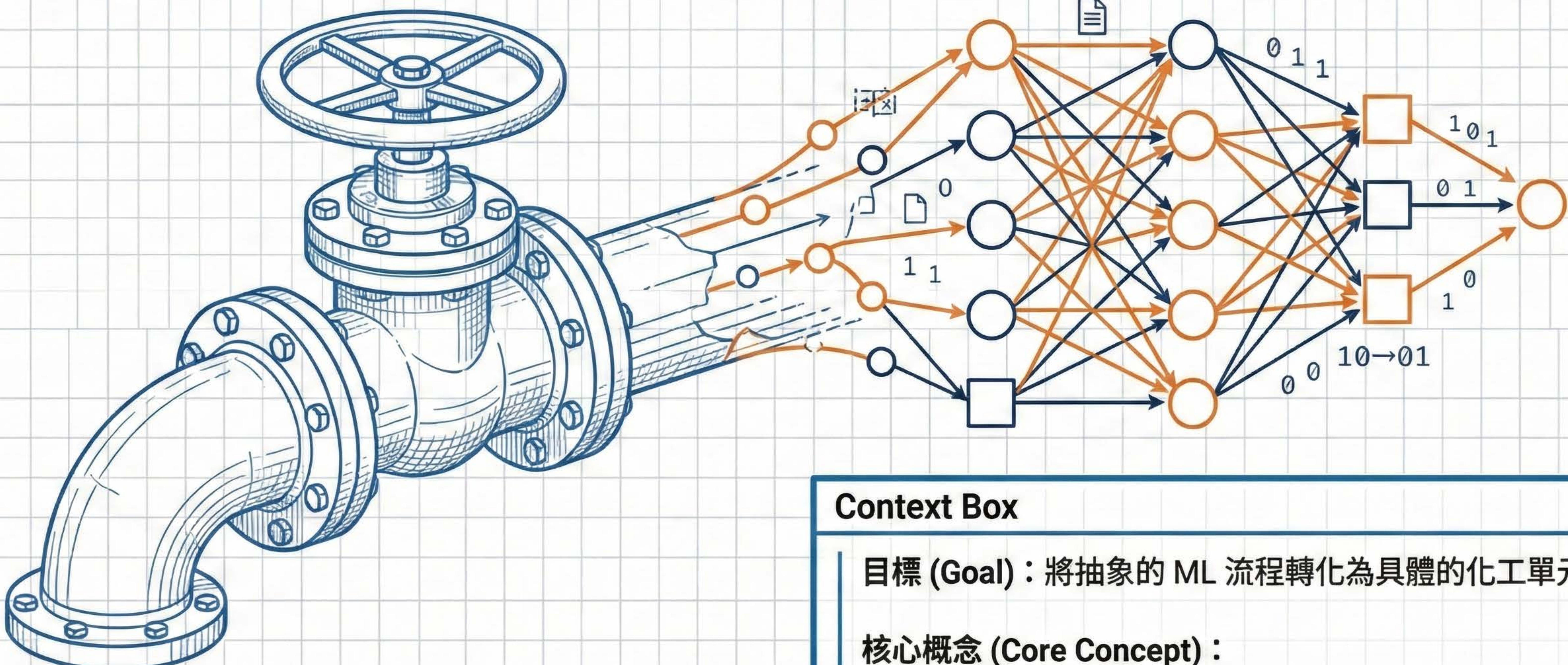


機器學習專案實戰：從數據到決策的完整工作流

Unit 01 - Section 05 | 建構化工 AI 模型的十個標準步驟



數位化工廠的製程圖 (The Process Flow Diagram of the Digital Plant)



1. 問題定義
2. 資料收集

3. 探索性資料分析 (EDA)
4. 資料前處理
5. 特徵工程 (Feature Engineering)

6. 模型選擇
7. 模型訓練
8. 模型評估
9. 模型優化
10. 部署與監控

機器學習不是魔法，它是系統化的工程程序。

Step 1: 問題定義 (Defining the Specs)

避免 'X-Y Problem'：精確定義輸入與輸出



Step 2: 資料收集 (Sourcing Raw Materials)

數據的來源、類型與品質檢驗

1. 歷史製程資料 (Structured)

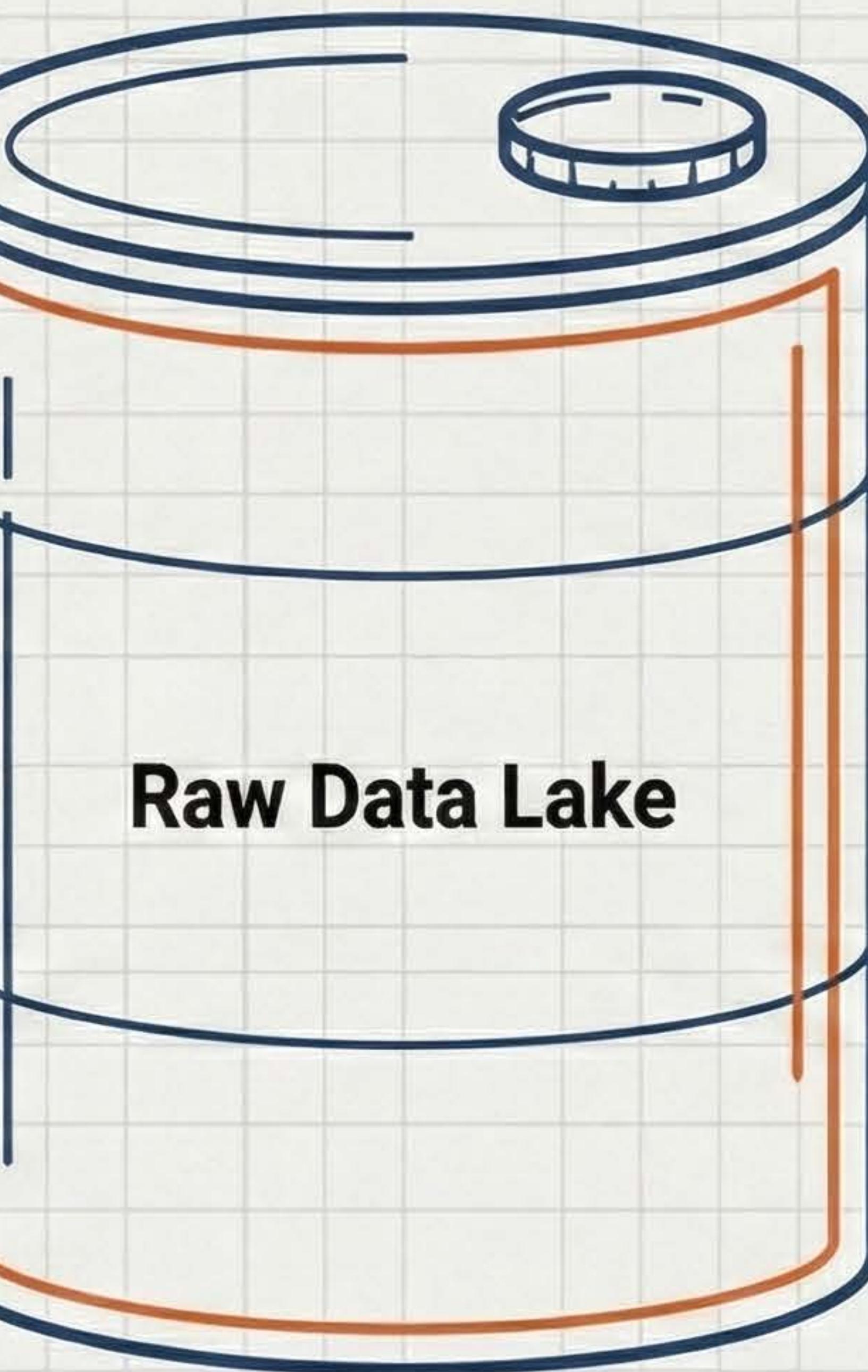
SCADA, DCS (T, P, Flow)

2. 實驗室數據 (Structured)

LIMS (Quality Analysis)

3. 非結構化資料 (Unstructured)

影像, 光譜, 日誌



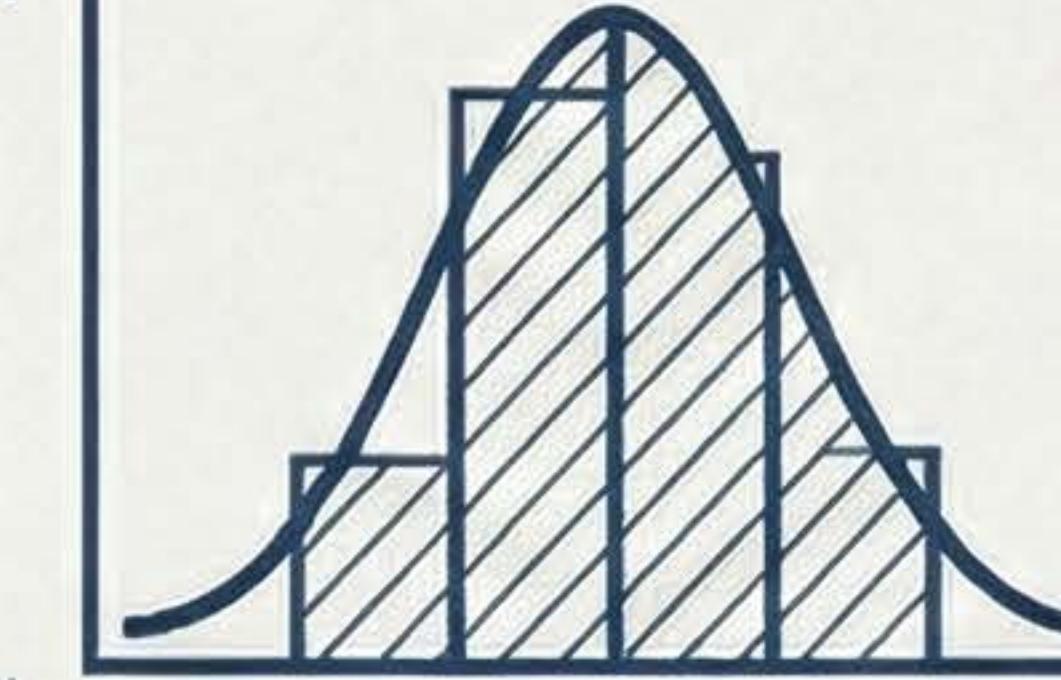
QA Criteria

- 代表性 (Representative)
- 完整性 (Completeness)
- 準確性 (Accuracy)

Step 3: 探索性資料分析 (Material Characterization)

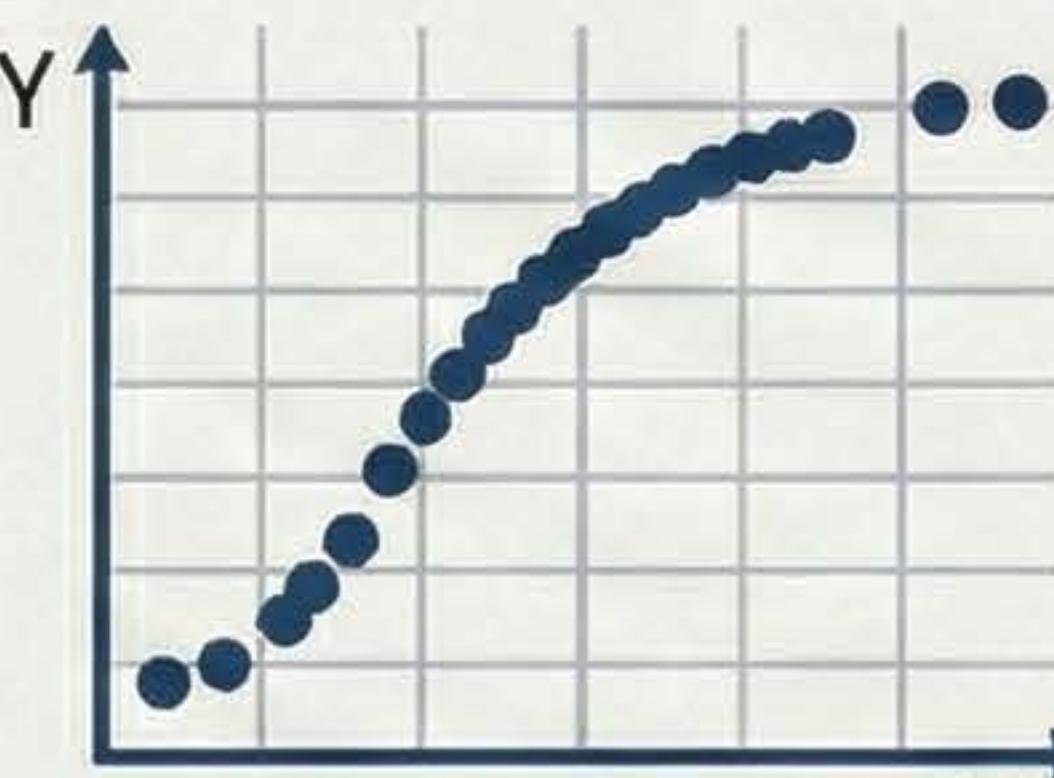
在進入模型之前，先理解你的數據

直方圖 (分布)



檢查常態分佈

散佈圖 (關聯)



發現非線性關係
(溫度 vs 產率)

描述性統計

Mean,
Std Dev,
Min/Max

箱型圖 (離群值)

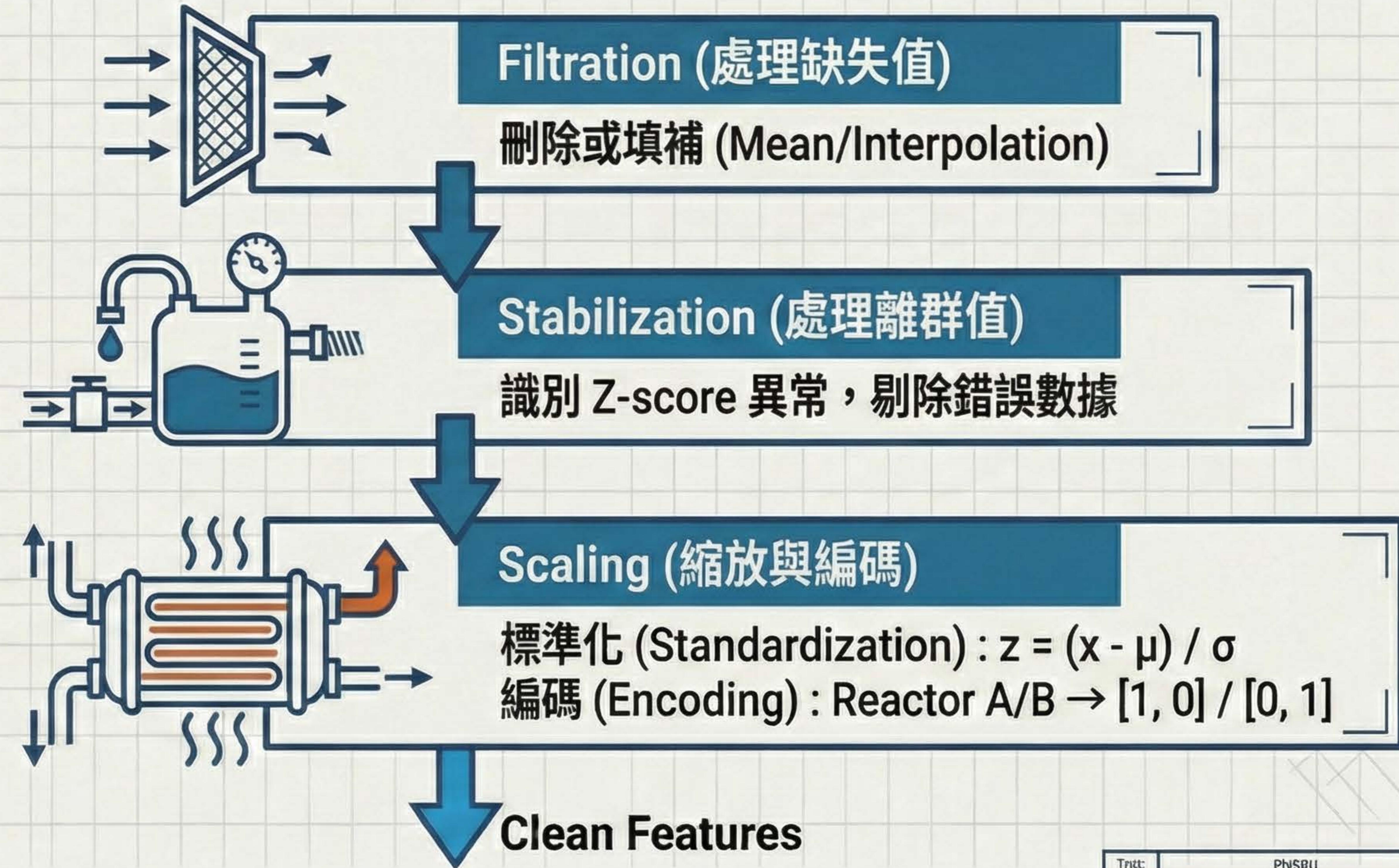


識別異常操作

ChemE Insight: 數據顯示週末的操作模式與平日不同，需分開處理。

Step 4: 資料前處理 (The Purification Unit)

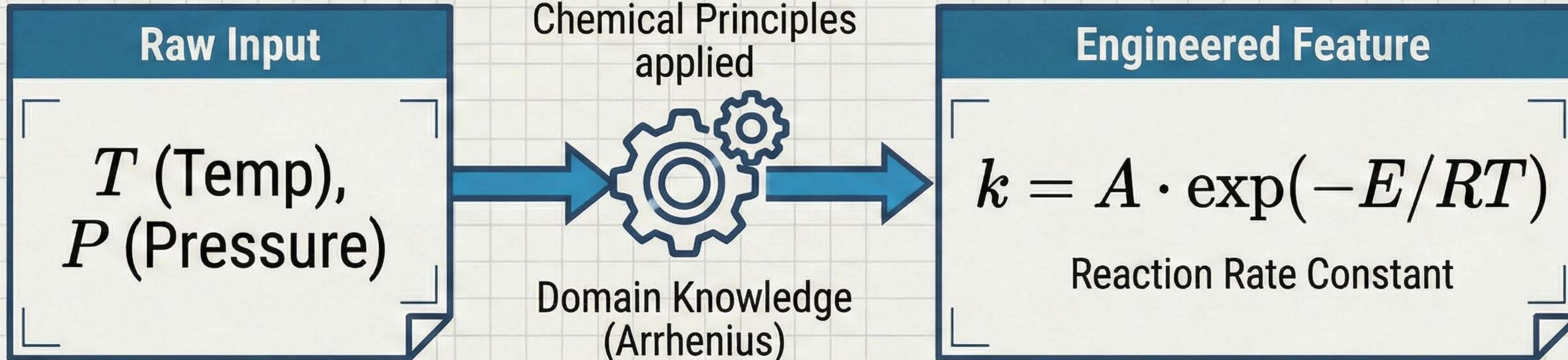
將原始數據轉換為模型可接受的格式



Tritt:	PhSRU	Res:
Calda	Roboje	19

Step 5: 特徵工程 (Reaction Kinetics)

化工人的核心優勢領域 — 注入領域知識



- 無因次數 (Dimensionless Numbers): Re, Sc, Nu



- 交互作用 (Interactions): $P \times V$

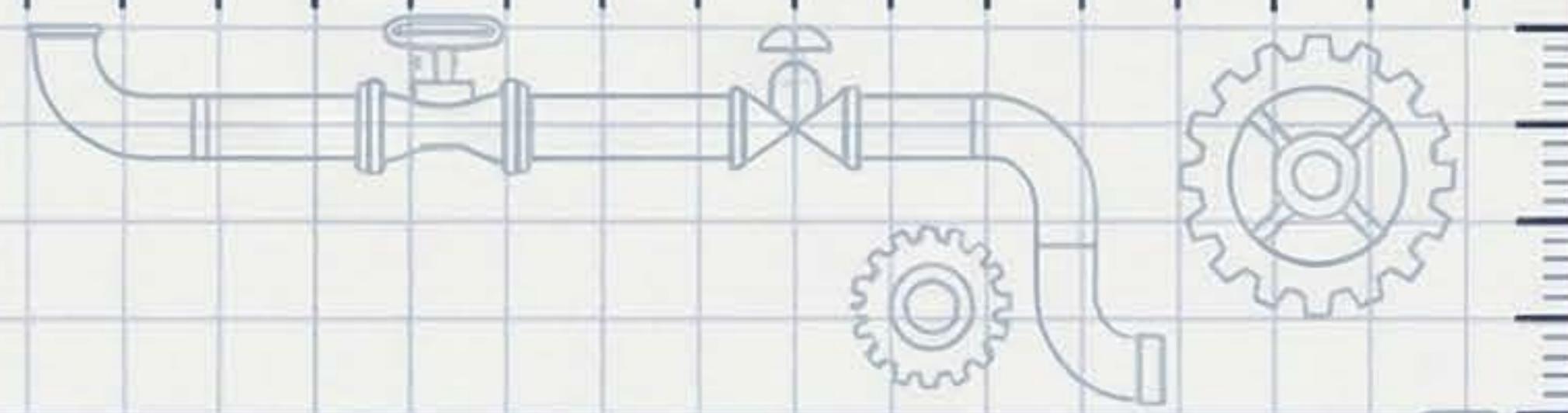


- 時間特徵 (Time-Series): 變化率 dT/dt

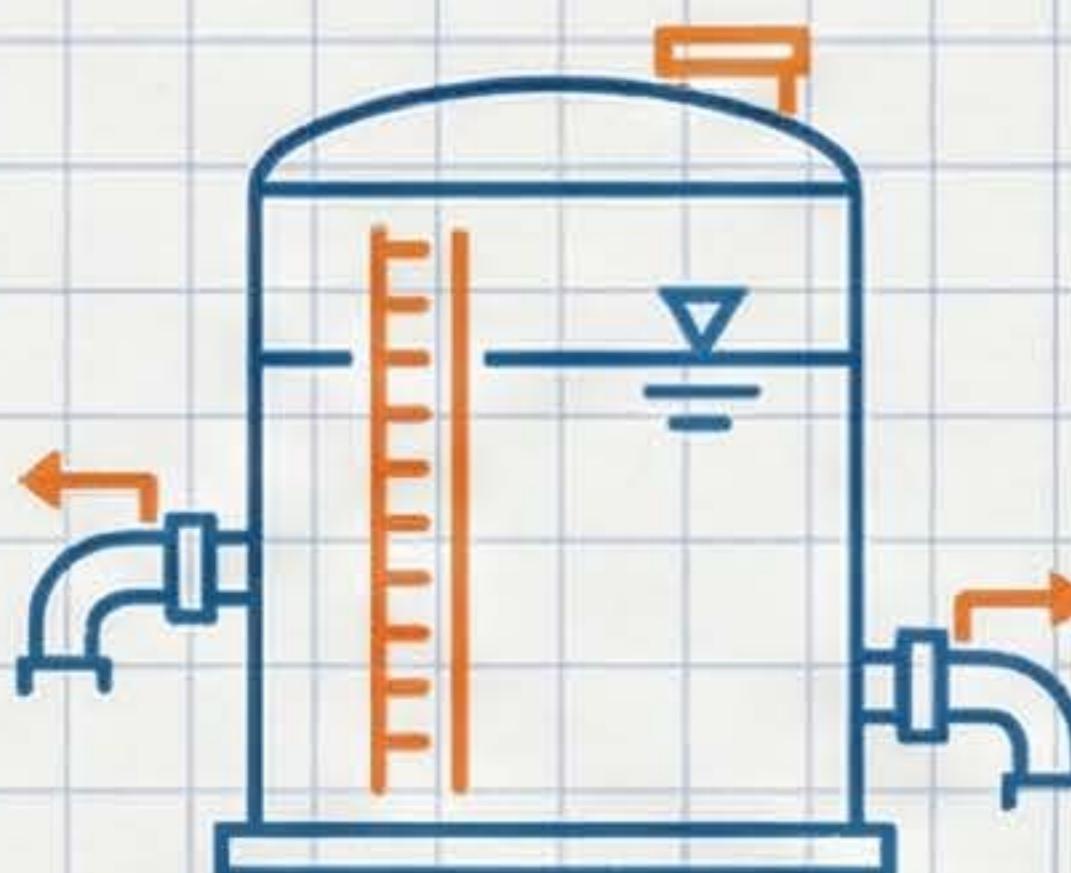
不要只是餵入原始數據，要轉換它們！

Step 6: 模型選擇 (Choosing the Reactor)

根據數據特性與目標選擇合適的演算法



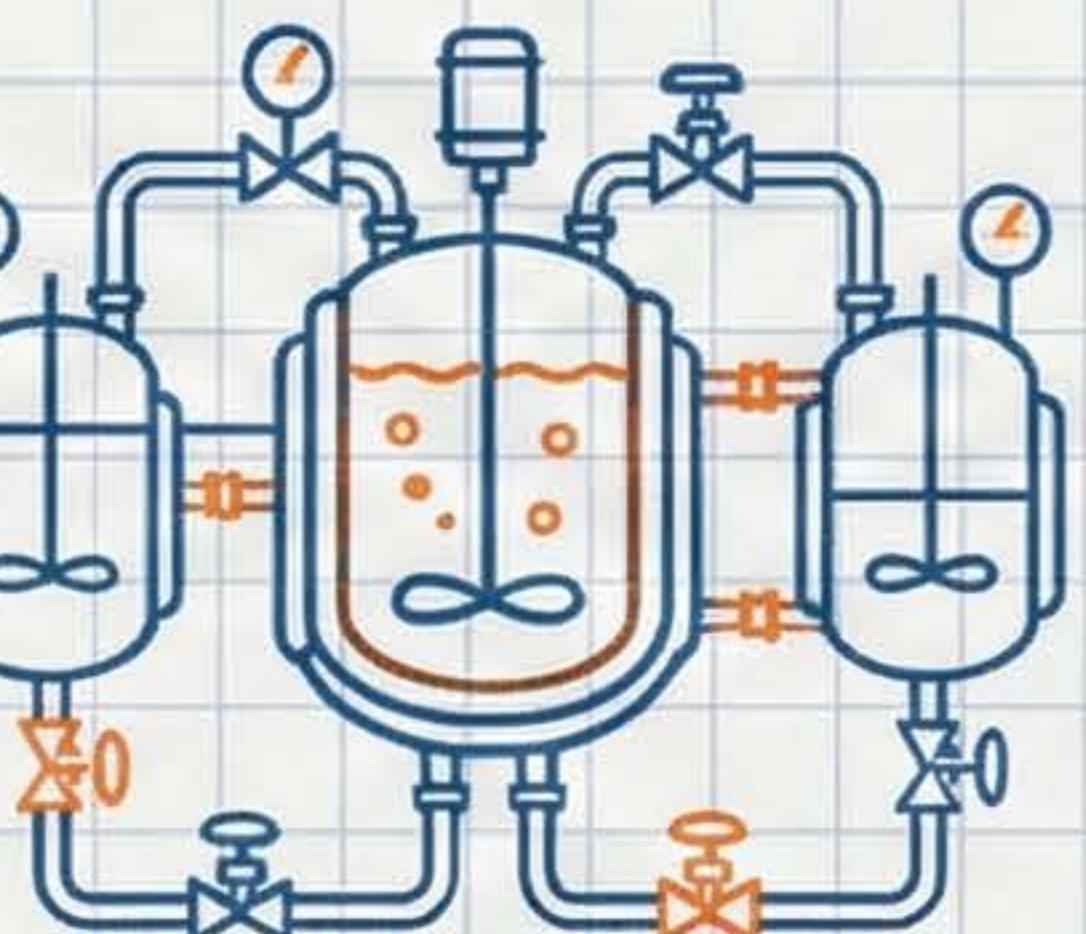
小資料量 / 高可解釋性



線性回歸
(Linear Regression)
決策樹 (Decision Tree)

適合快速原型驗證與
解釋性要求高的場景

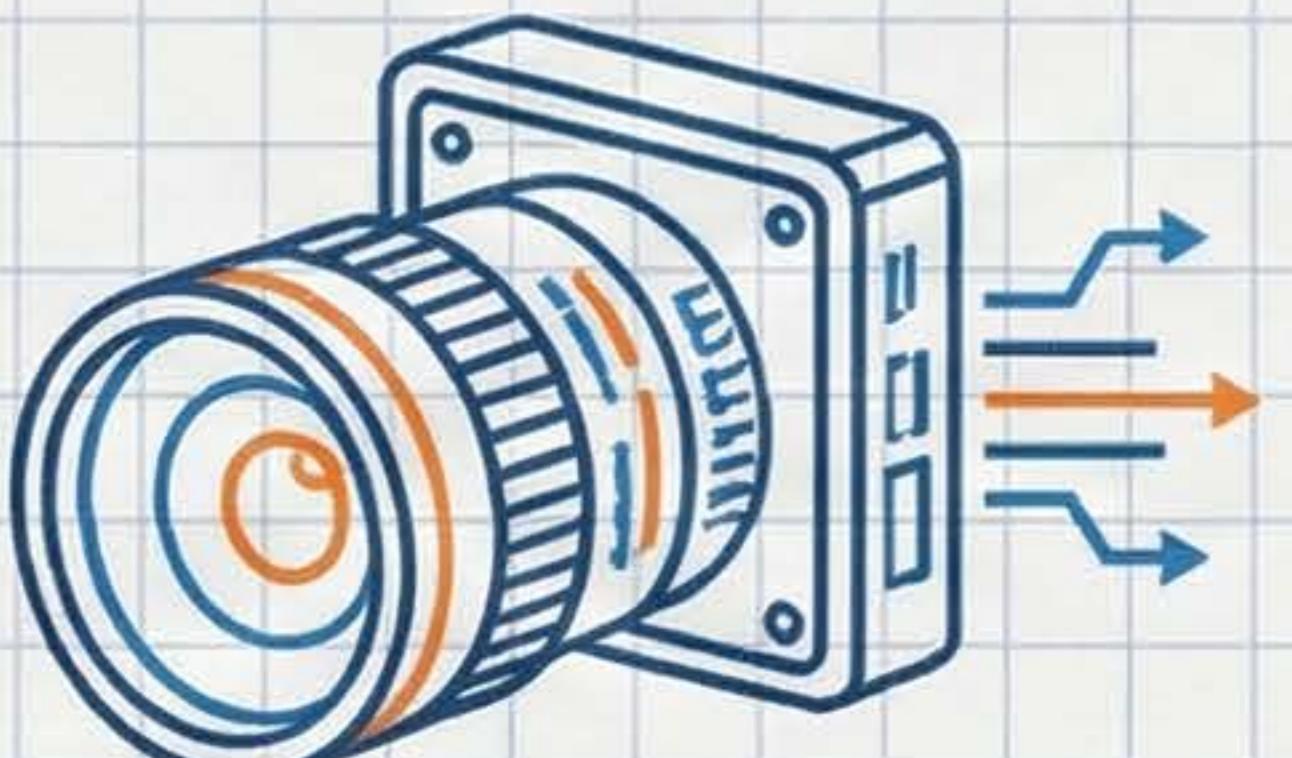
複雜非線性 / 中等資料量



隨機森林
(Random Forest)
XGBoost

化工品質預測首選

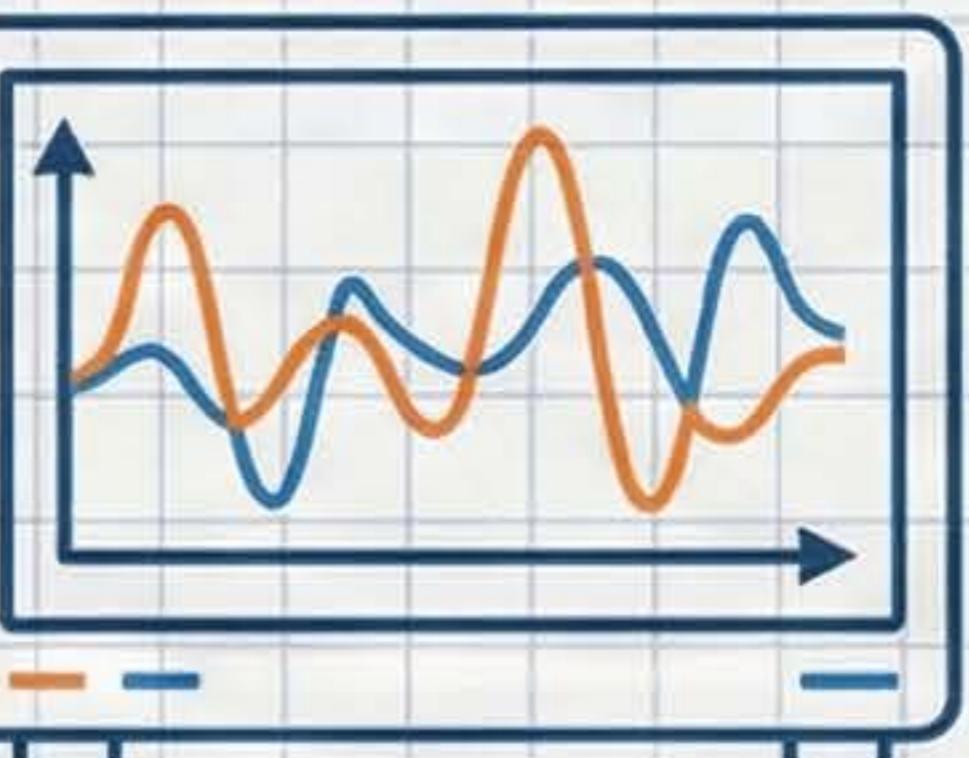
影像資料 (Image)



CNN
(卷積神經網路)

用於製程監控與視覺檢測

動態製程 / 時間序列

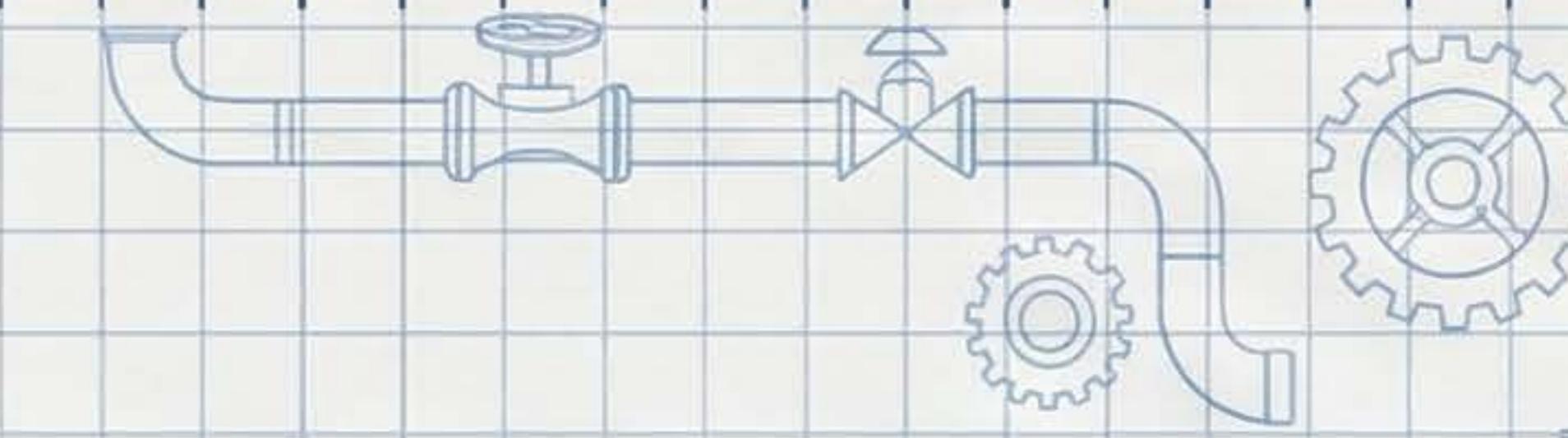


LSTM, GRU
(遞迴神經網路)

處理時間相關數據與
預測未來趨勢

Step 7: 模型訓練 (Start-up Process)

教導模型學習數據中的規律



警告：測試集絕不可用於訓練！
(Avoid Data Leakage)

Dataset



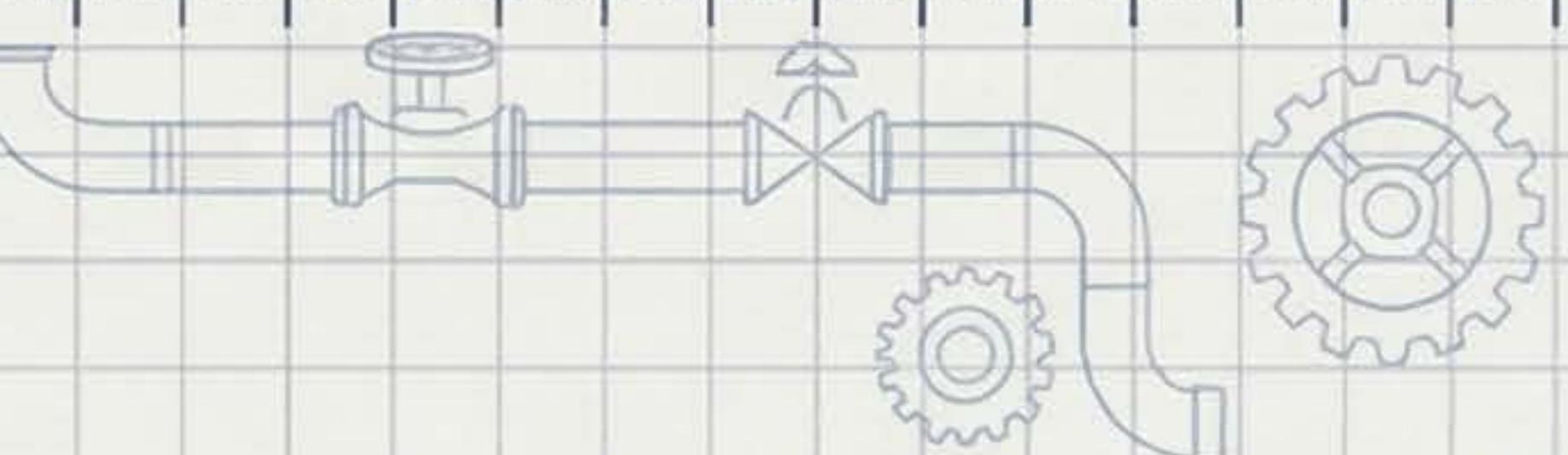
用於建立模型參數
(Processing)

用於調整超參數
(In-process Check)

用於最終評估
(Final QA)

Step 8: 模型評估 (Quality Control)

驗證模型的預測能力與可靠度



Model Performance Report

Regression Metrics (數值預測)

R^2 Score



RMSE

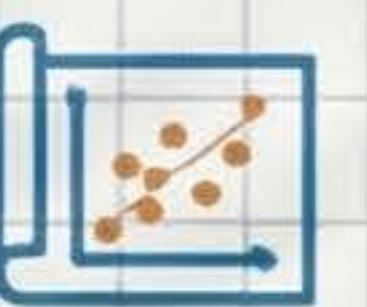


Classification Metrics (分類判斷)

Predicted
Positive Negative

Actual	Positive	Negative
Positive	TP	FP
Negative	FN	TN

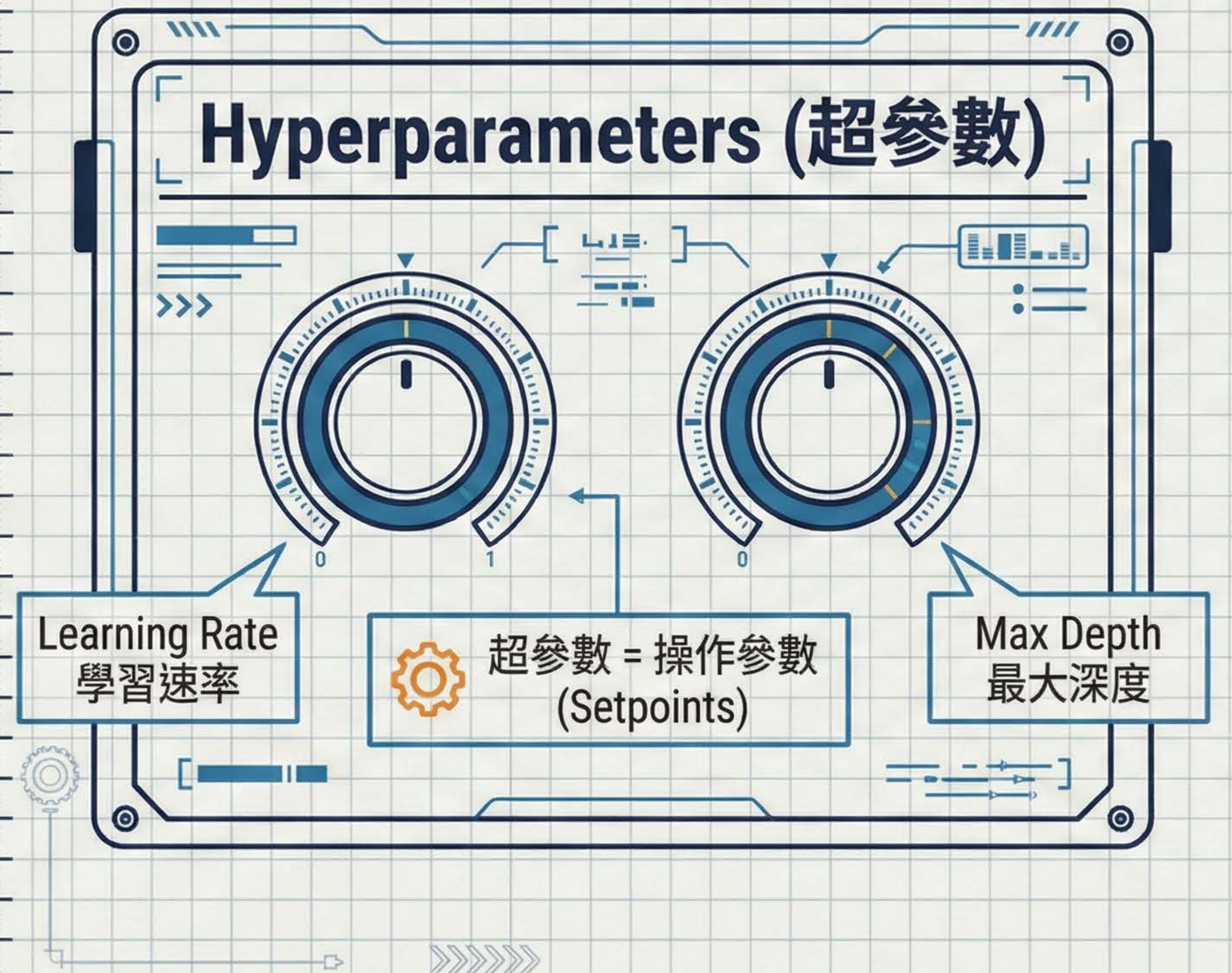
Accuracy, Precision, Recall



Diagnostics: Check Residuals for randomness

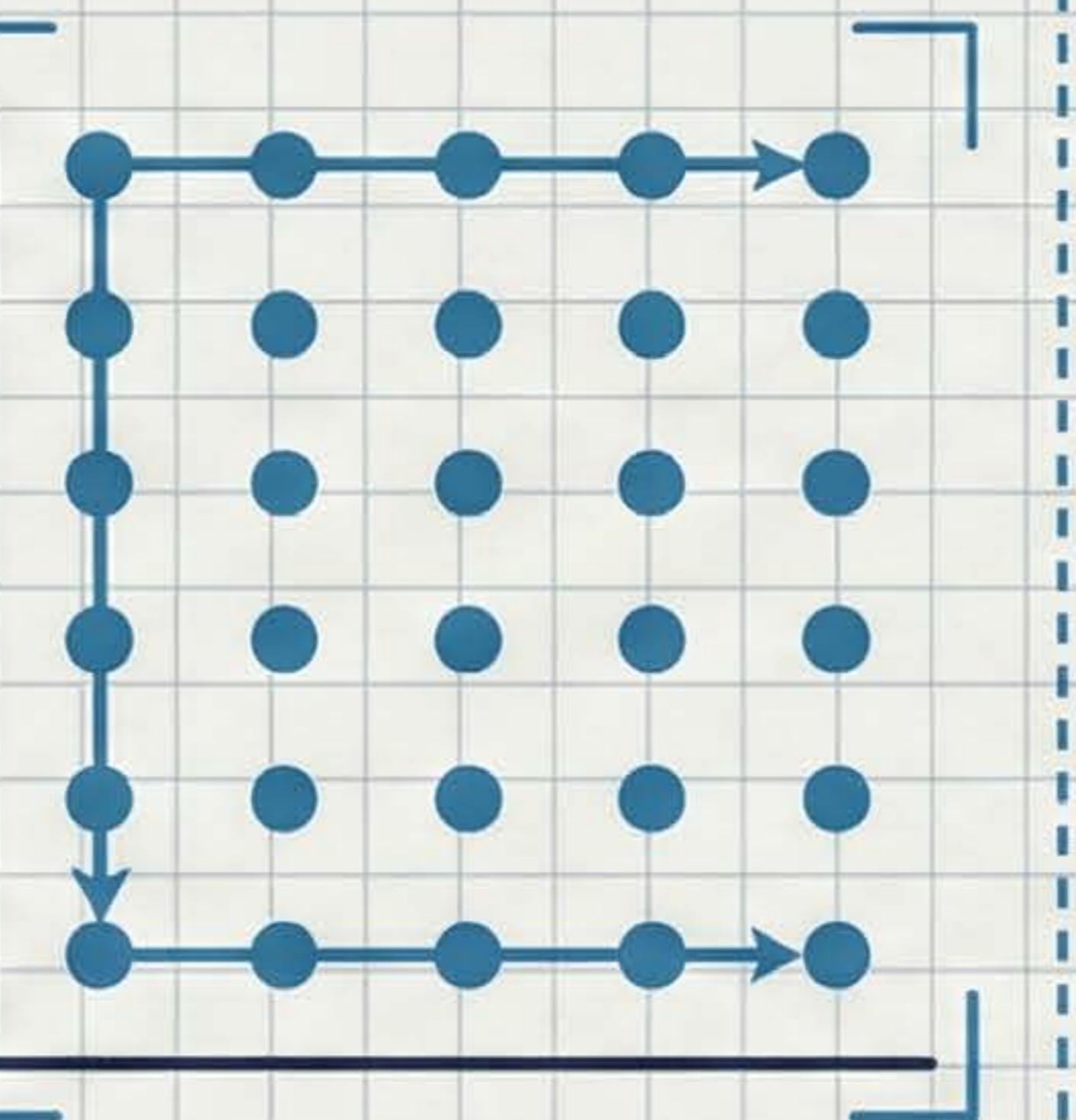
Step 9: 模型調整與優化 (Process Optimization)

微調操作參數以達到最佳效能



Grid Search

網格搜尋
(暴力測試)



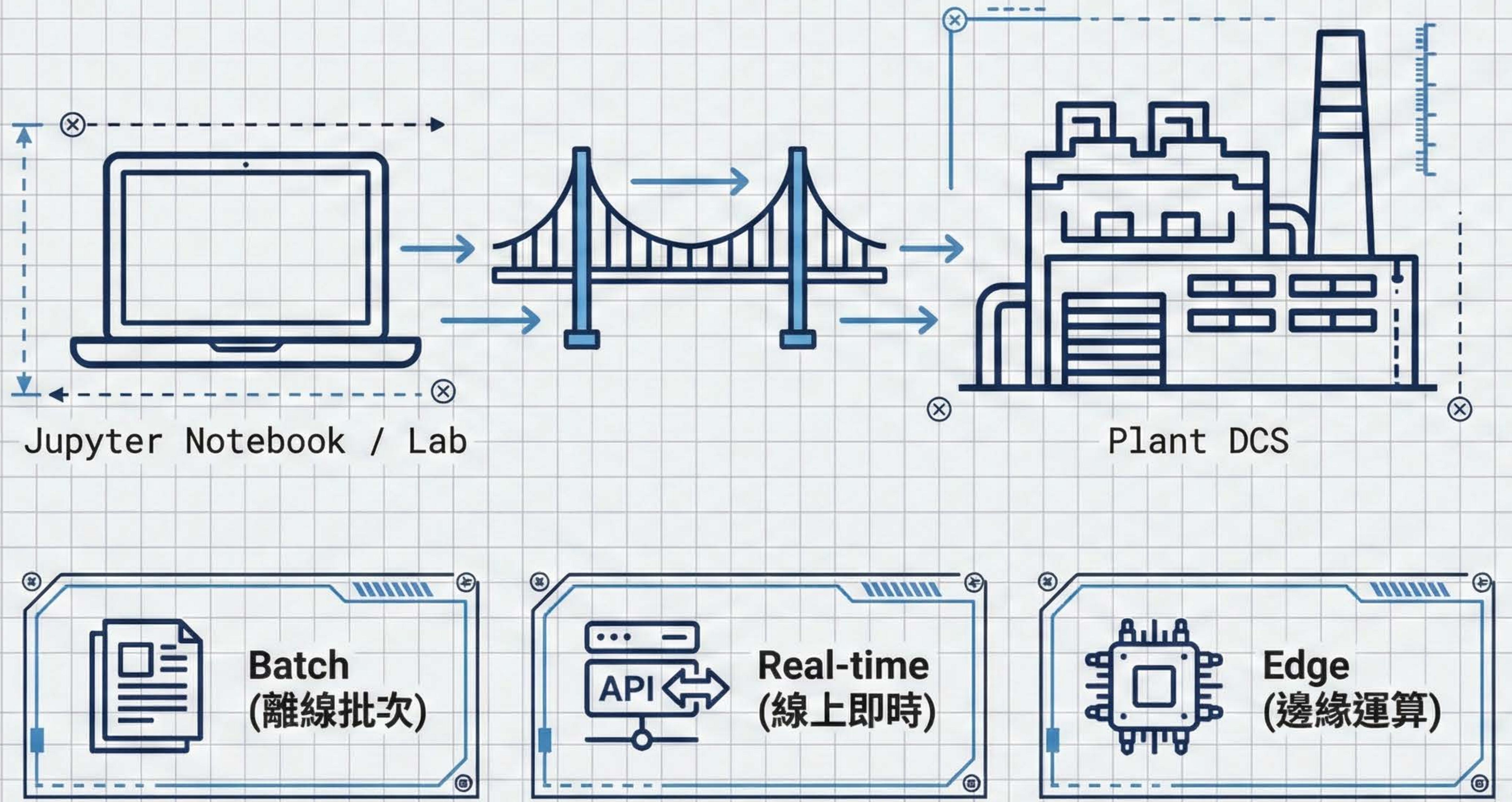
Bayesian Optimization

貝氏優化 (智慧搜尋)



Step 10: 部署與監控 (Plant Commissioning)

從實驗室模型走向工廠實際應用



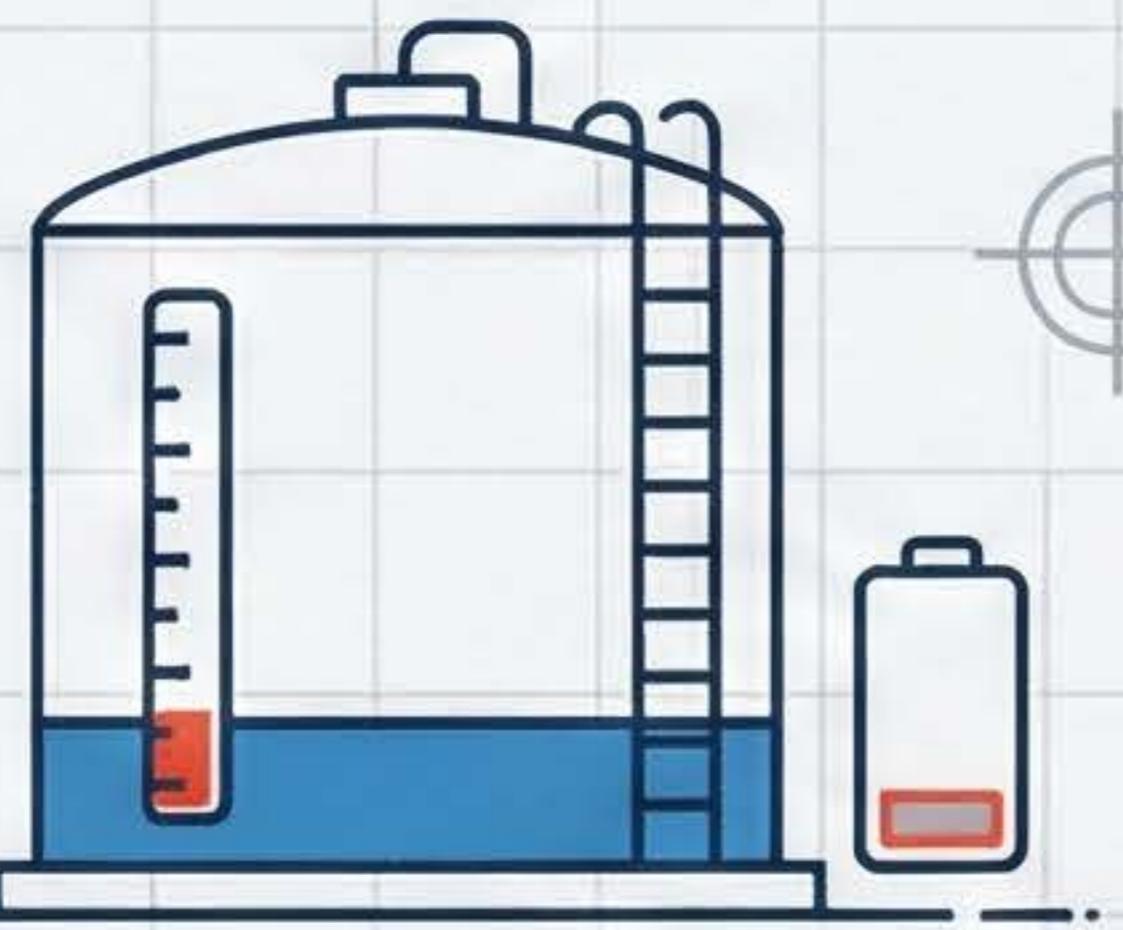
Model Health

Monitor for
Concept Drift
(觸媒老化效應)
Noto Sans TC

Retrain needed
if performance
drops.

常見挑戰與解決方案 (Troubleshooting)

克服化工領域的特殊限制



Challenge: 資料量不足 (Small Data)

💡 資料擴增
(Augmentation)，遷移學習 (Transfer Learning)



Challenge: 物理限制 (Physical Constraints)

💡 物理約束嵌入
(Physics-informed ML) – 確保守恆



Challenge: 黑箱問題 (Black Box)

💡 SHAP/LIME 分析 –
確認特徵重要性符合直覺

核心原則：品質決定成敗

Garbage In, Garbage Out

