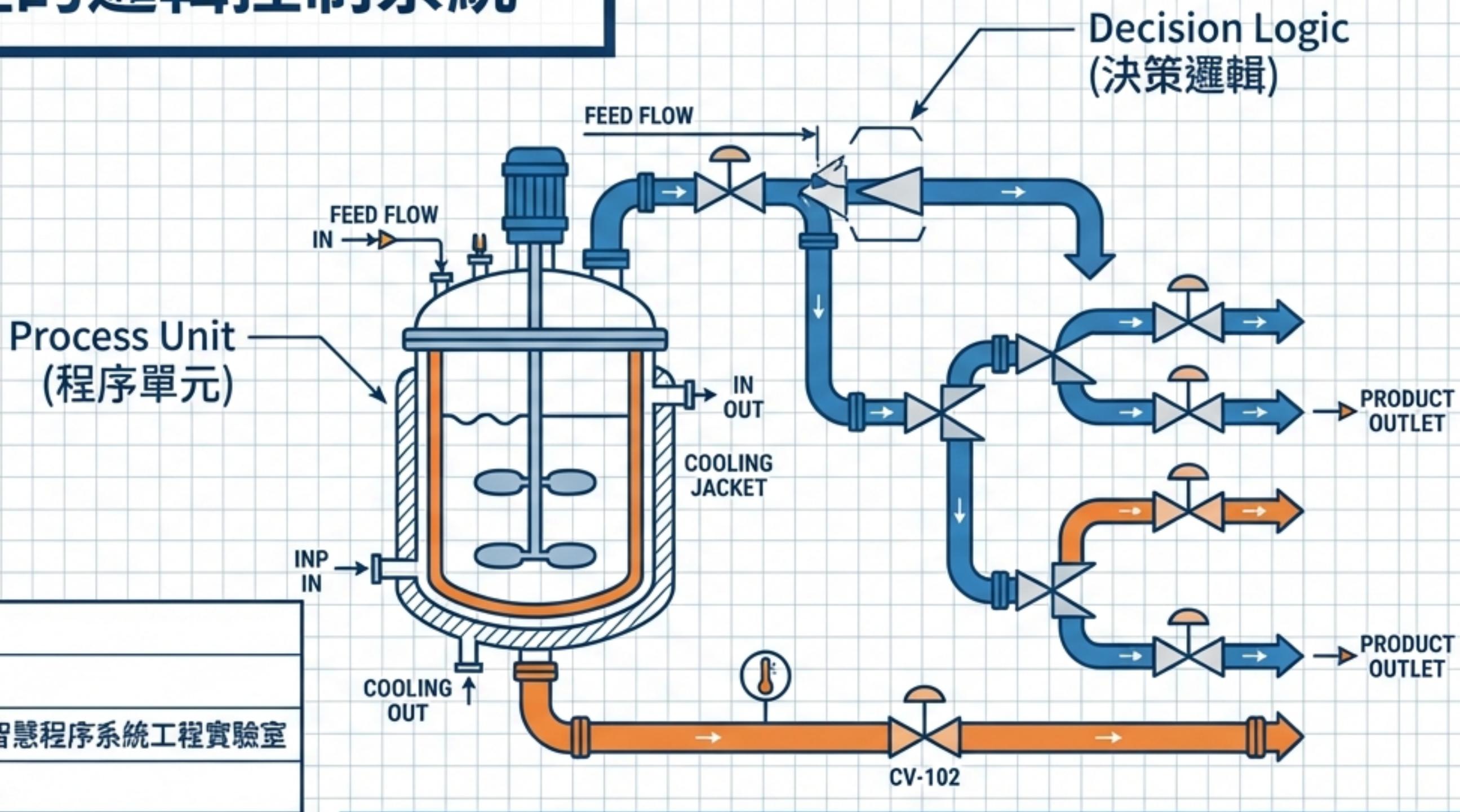


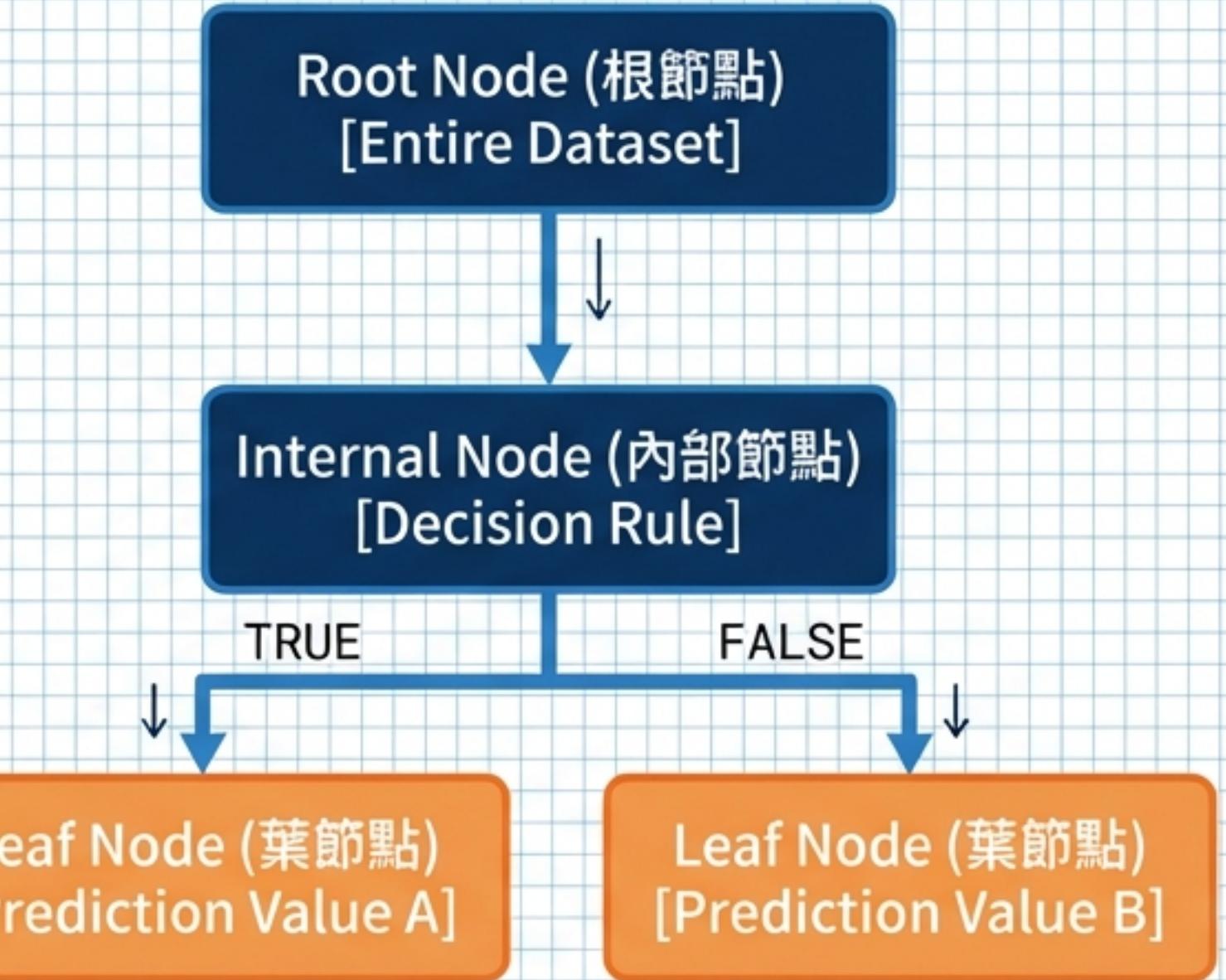
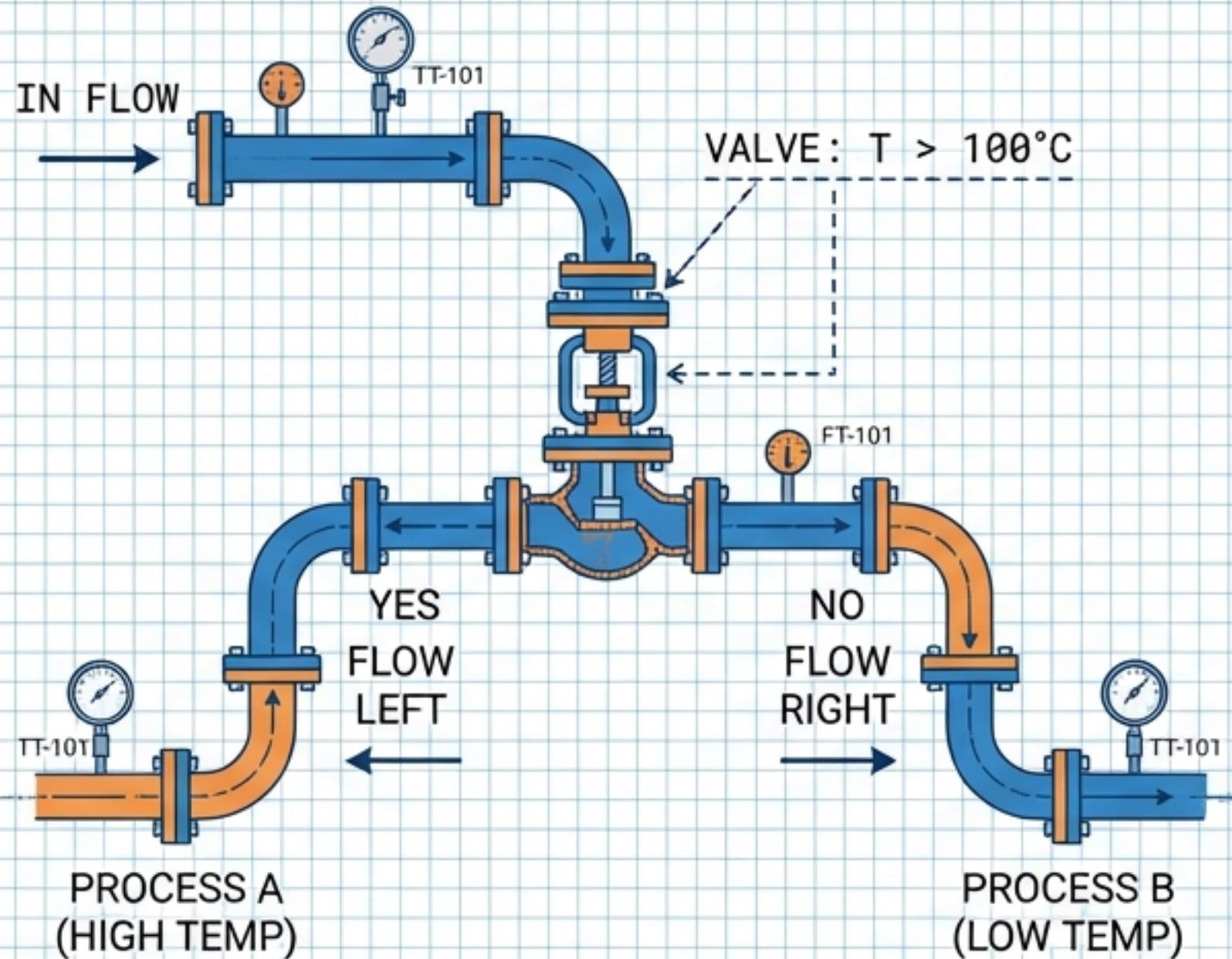
# Unit 11: 決策樹回歸

## 建構化學製程的邏輯控制系統



COURSE CODE:	CHE-AI-114
INSTRUCTOR:	莊曜楨 助理教授
DEPT:	逢甲大學 化工系 智慧程序系統工程實驗室
TARGET:	化學工程學系學生

# 決策樹的運作邏輯：數位化的管路系統

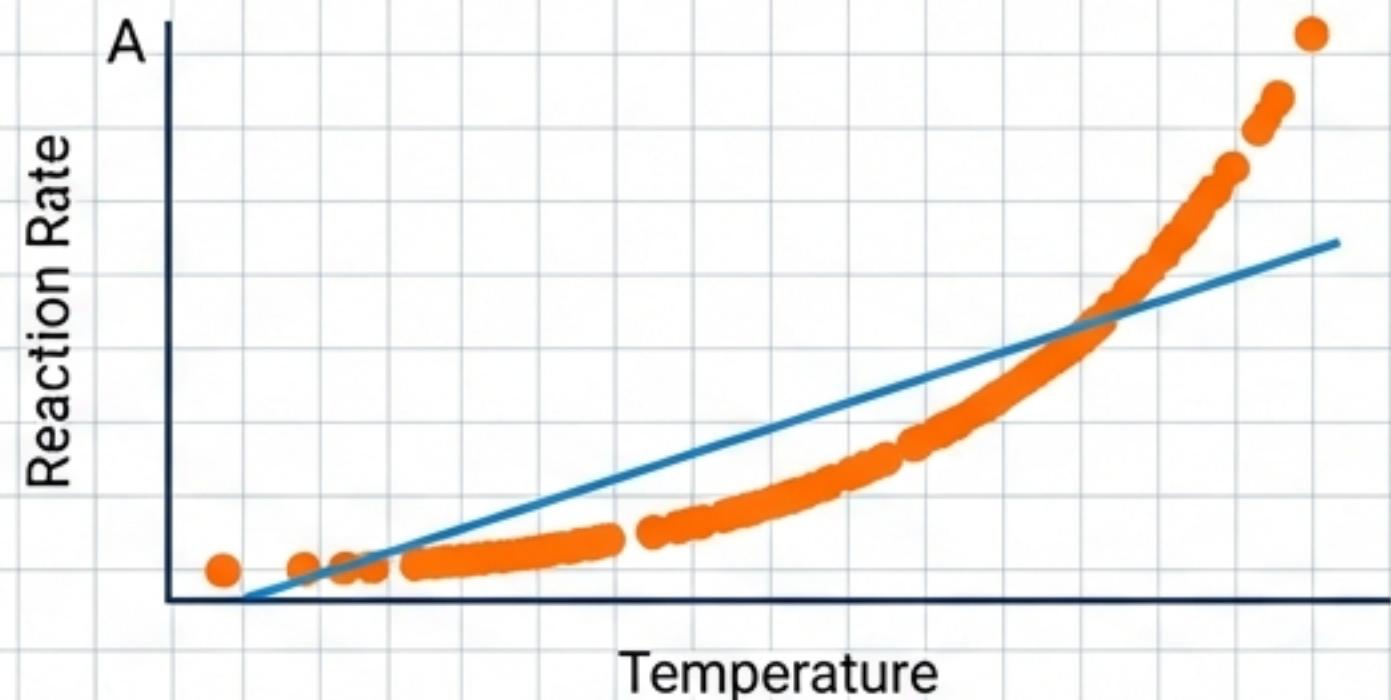


定義：一種基於樹狀結構的非參數監督式學習方法。通過一系列二元決策規則 (Binary Decision Rules) 將特徵空間遞迴分割。

預測原理：區域內樣本的平均值 (Mean Value)。

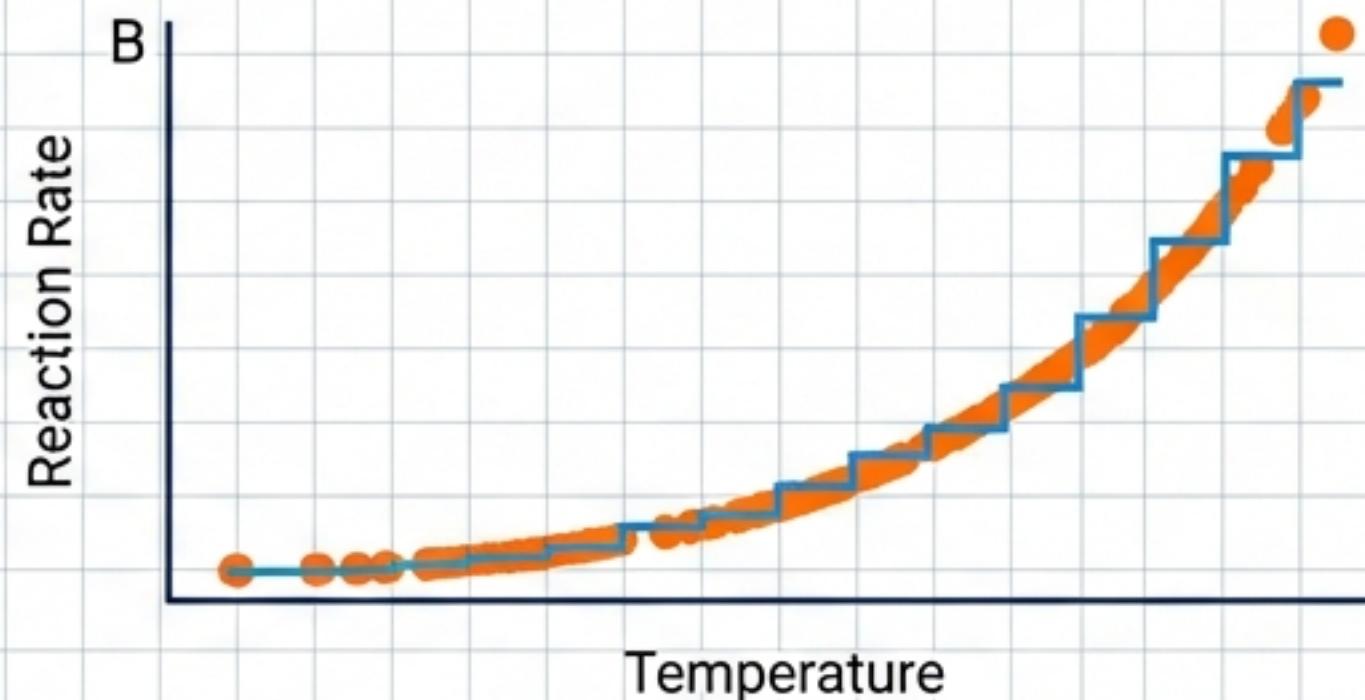
# 為什麼線性模型不足以應對化工製程？

線性回歸：無法捕捉動力學曲線



Underfitting (低度擬合)

決策樹：分段常數擬合



Piecewise Constant (分段常數)

	線性模型	決策樹
特徵關係 (Feature Relationship)	線性模型：加性關係 (Additivity)	決策樹：交互作用 (Interaction)
邊界形狀 (Boundary)	線性模型：平滑直線	決策樹：階梯狀 (Staircase)

關鍵洞察：化工現象（如阿倫尼烏斯定律）通常是非線性的，決策樹能自動捕捉特徵間的交互作用。

# 分裂的數學原理：尋找純度的極大值

## 目標函數 (The Goal)

最小化均方誤差 (Minimize MSE)

$$MSE(t) = \frac{1}{N_t} \sum_{y_i} (y_i - \bar{y}_t)^2$$

目標是讓分裂後的子節點誤差最小化。

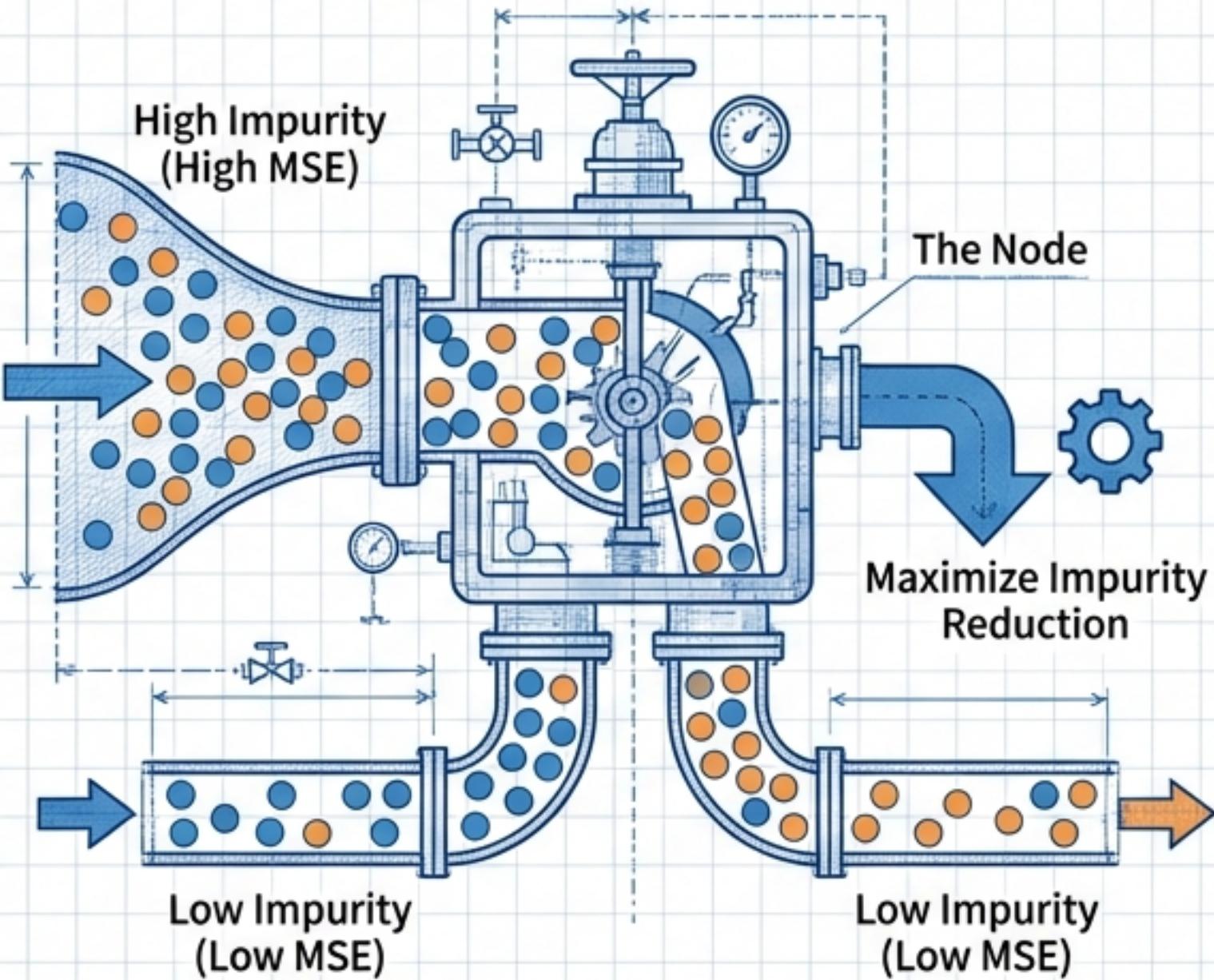
## 分裂決策 (The Decision)

最大化雜質減少量 (Impurity Reduction)

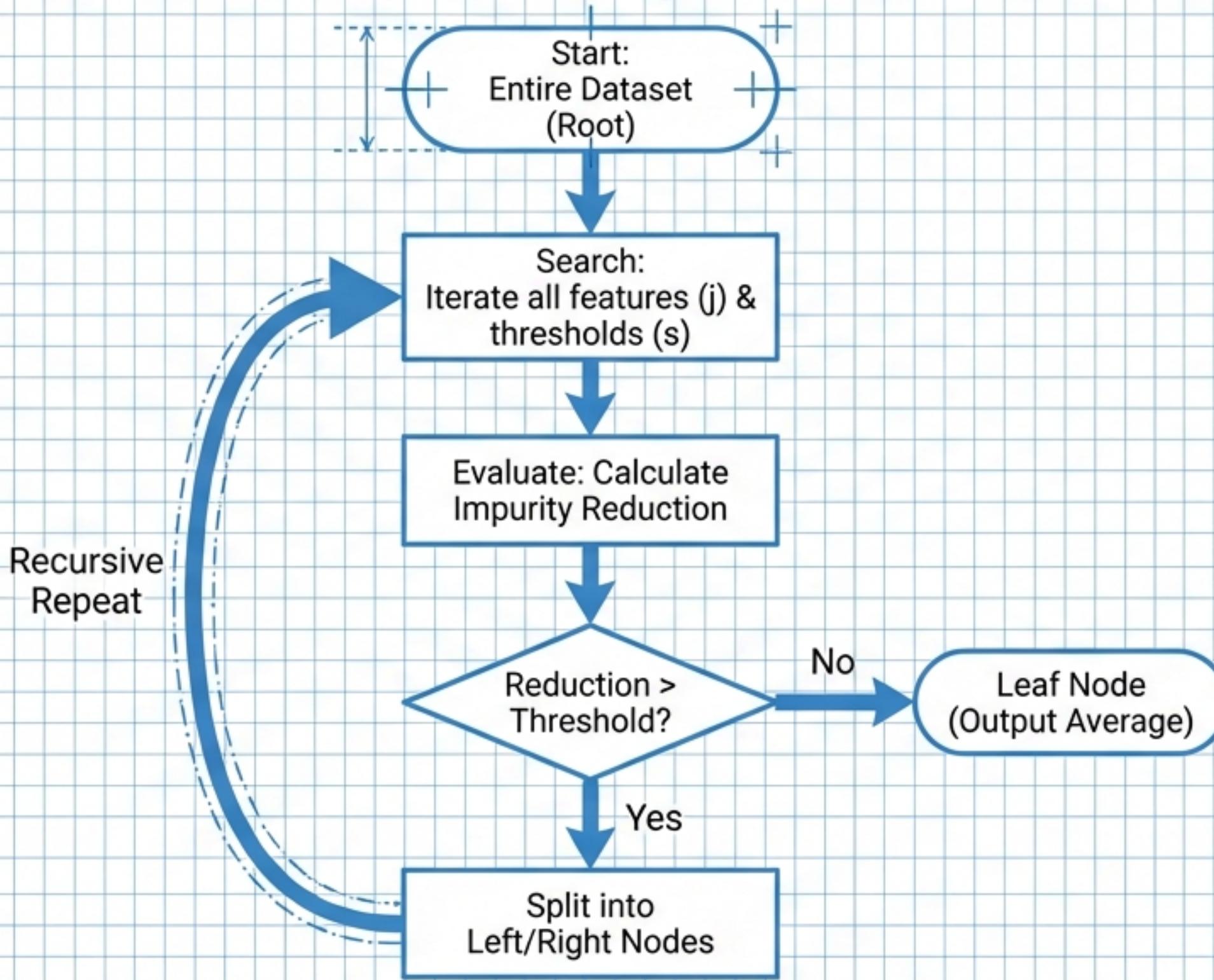
$$(j^*, s^*) = \arg \min_{j,s} [N_{left} \cdot MSE_{left} + N_{right} \cdot MSE_{right}]$$

尋找最佳特徵  $j$  與切分點  $s$ 。

## Visual Metaphor



# 建構演算法 (CART) : 遞迴分割的 SOP

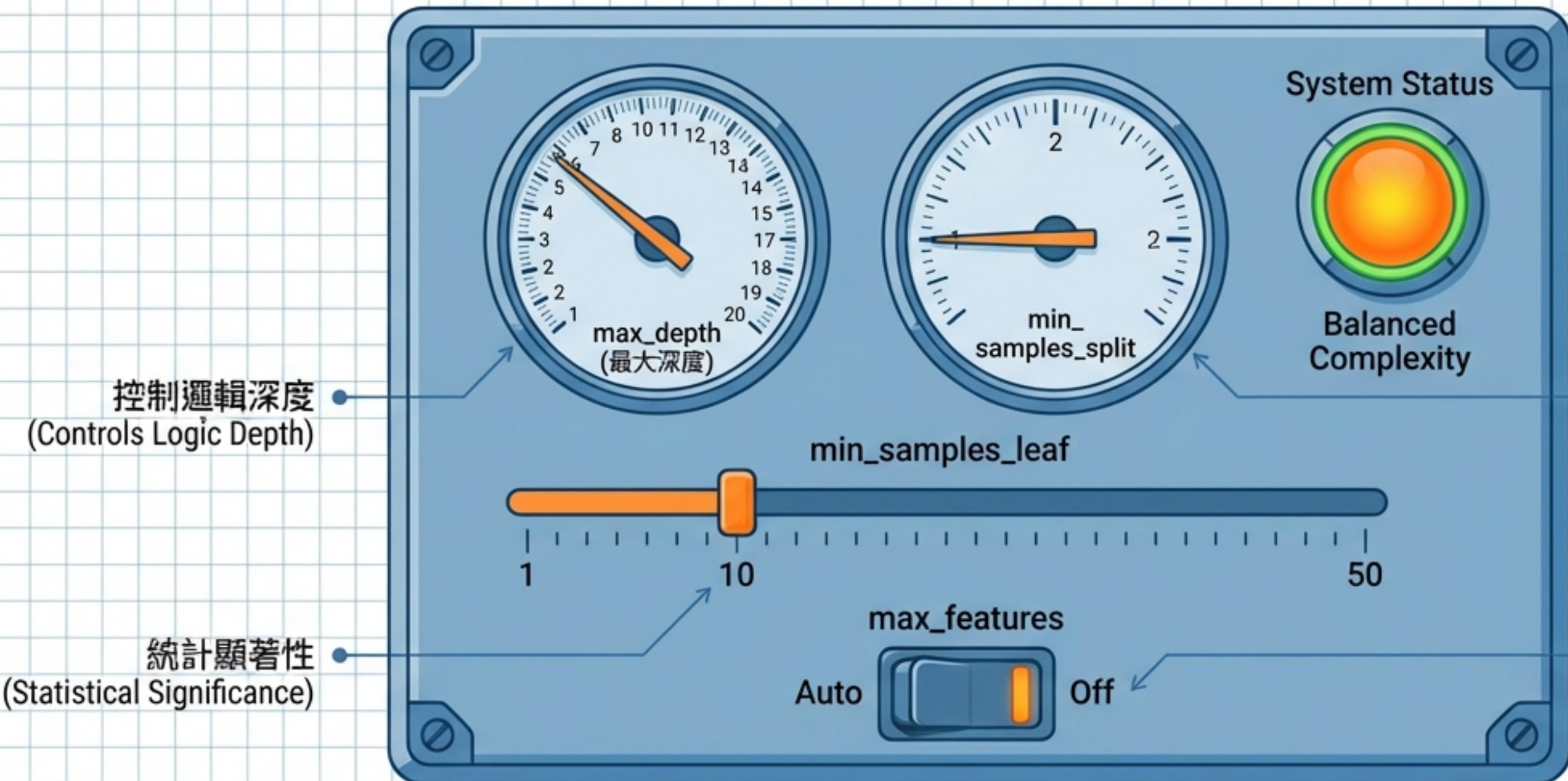


## Technical Note

### SPECIFICATIONS

- **Method:** Greedy Algorithm (貪婪演算法)
- **Complexity (Train):**  $O(N p \log N)$
- **Complexity (Predict):**  $O(\log N)$

# 系統控制面板：決定模型的複雜度

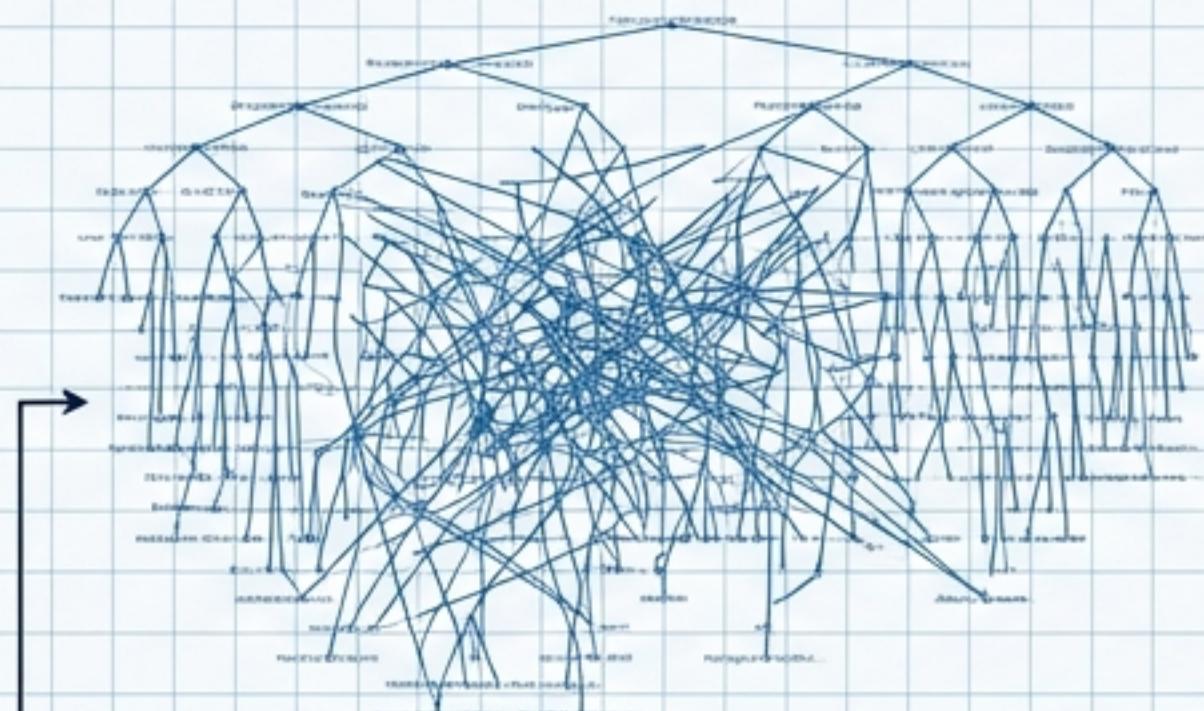


# 安全連鎖系統：防止過度擬合

Safety Interlock System: Preventing Overfitting



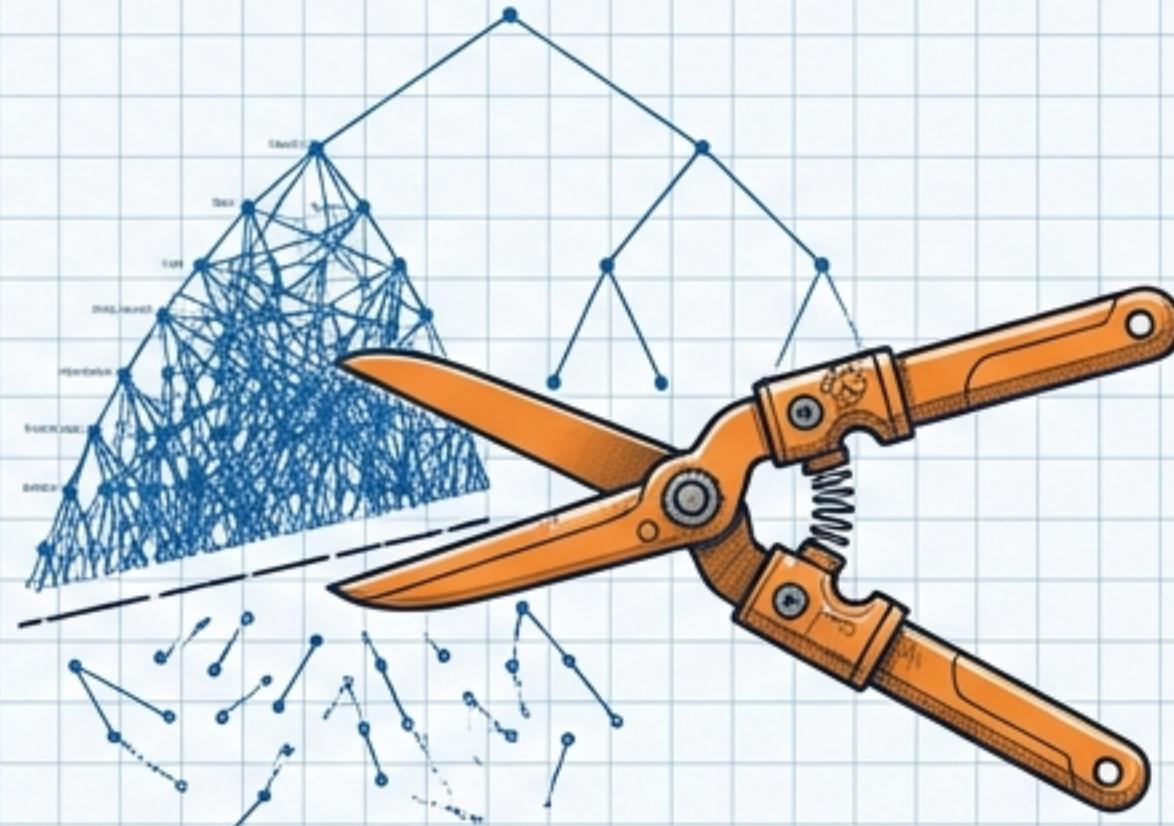
## Danger: Overfitting (過擬合)



症狀：Perfect Memory of Noise

Train  $R^2 \approx 1.0$  | Test  $R^2$  Low

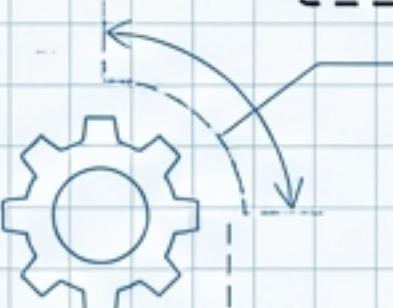
## Solution: Pruning (剪枝)



策略：像安全閥一樣限制系統壓力

### STRATEGIES (策略)

1. Pre-Pruning (預剪枝): 提前停止 (e.g., max\_depth=5)
2. Post-Pruning (後剪枝): 成本複雜度剪枝 (CCP Alpha)



# 實作藍圖：Python 程式碼規格

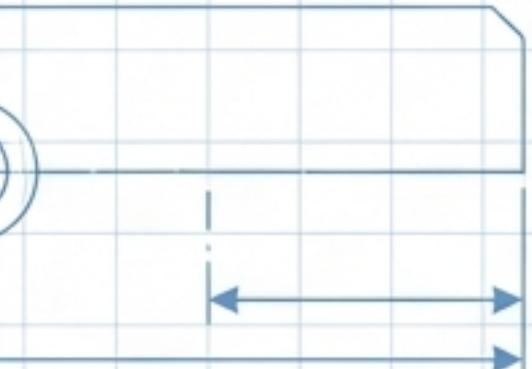
Good Code

```
1 from sklearn.tree import DecisionTreeRegressor  
2  
3 # Initialize the "Control System"  
4 model = DecisionTreeRegressor(  
5     max_depth=5,          # Limit logic depth  
6     min_samples_leaf=10,   # Ensure statistical mass  
7     random_state=42       # Reproducibility  
8 )  
9  
10 # Commissioning (Training)  
11 model.fit(X_train, y_train)  
12  
13 # Operation (Prediction)  
14 y_pred = model.predict(X_test)
```

Core Engine  
(核心引擎)

Safety Constraint  
(安全限制)

Data Loading  
(載入數據)



# 專案案例：催化反應器產率優化

**Temperature (T)**

150 - 250 °C



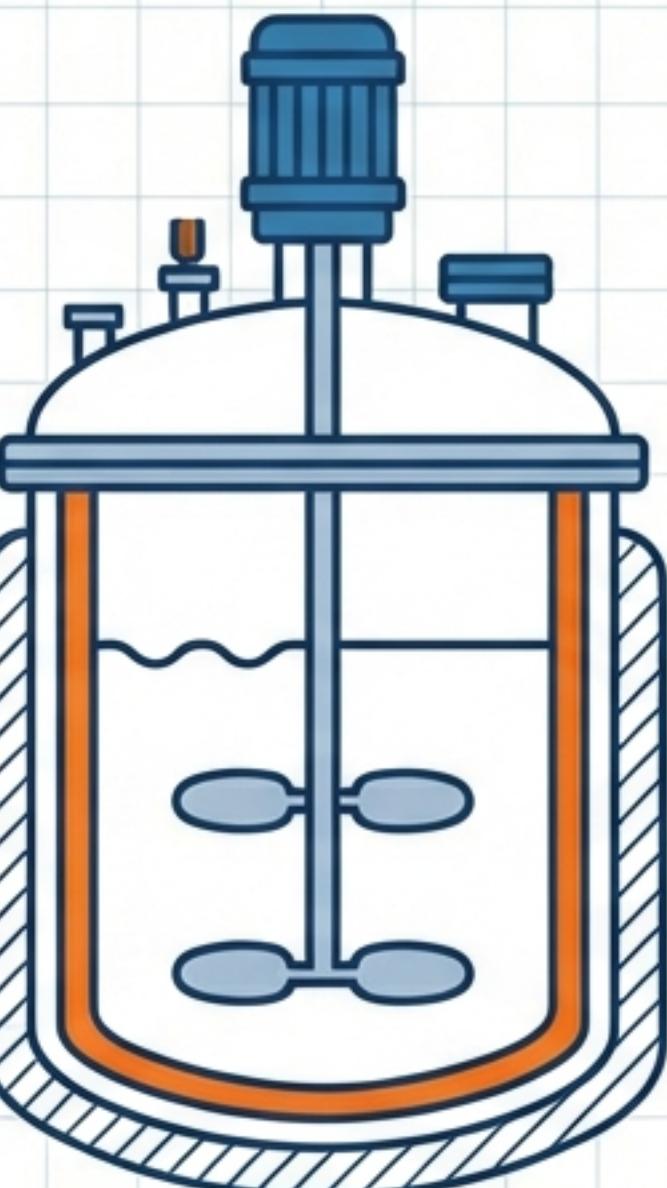
**Pressure (P)**

1 - 5 bar



**Catalyst (Cat)**

0.5 - 2.0 g



**Yield (Y)**

Target: 70 - 95%

## Data Spec

**Dataset:** 100 Simulated Runs

**Split:** 70% Training / 30% Testing

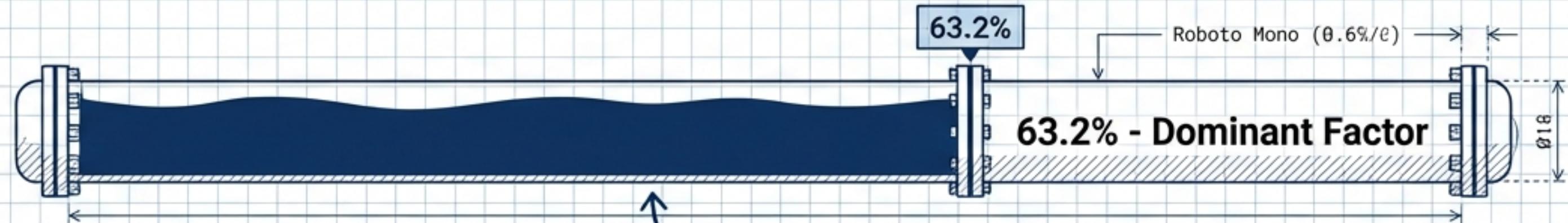
# 模型效能測試：無限制 vs. 受控系統

Parameter	Model A: Unrestricted (無限制)	Model B: Controlled (受控)	Optimized (Grid Search)
Depth (深度)	11 (Complex)	4 (Simple)	7
Train R <sup>2</sup>	1.0000 !	0.9019	0.9736
Test R <sup>2</sup>	0.9013	0.7977	0.8759 ✓
RMSE	1.92	2.75	2.15
Status	Overfitting (過擬合)	Stable (穩定)	Best Fit (最佳擬合)



# 關鍵製程參數分析 (Feature Importance)

Temperature



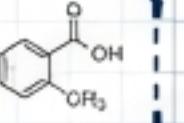
Pressure



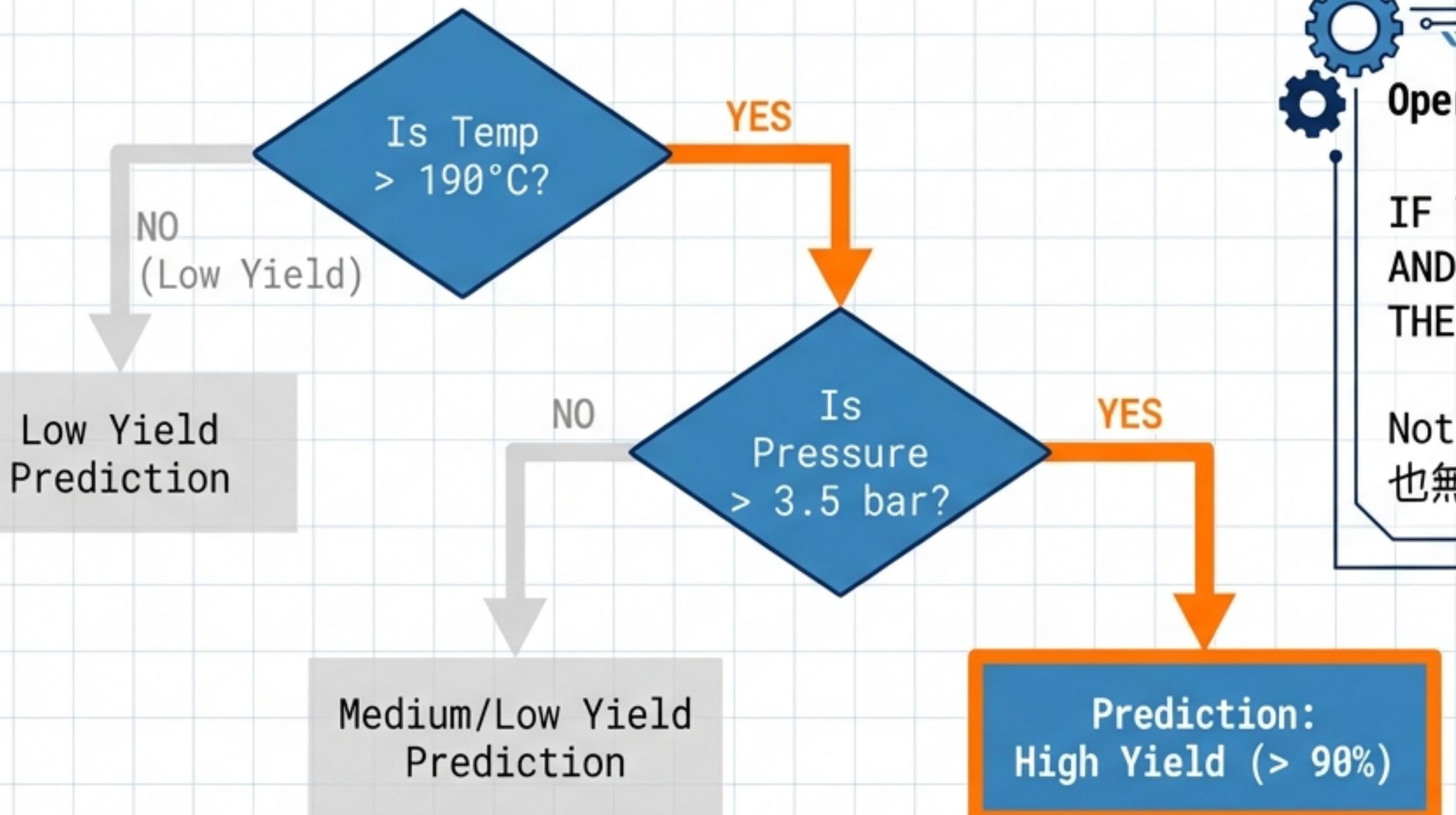
Catalyst



Chemical Insight: 符合動力學理論 (Arrhenius),  
Insight: 溫度對反應速率呈指數影響。



# 解讀決策規則：操作指導方針



**Operational Rule (操作規則):**

IF Temp > 190°C  
AND Pressure > 3.5 bar  
THEN Expect High Yield.

Note: 低溫 (<190°C) 即使加壓  
也無效。



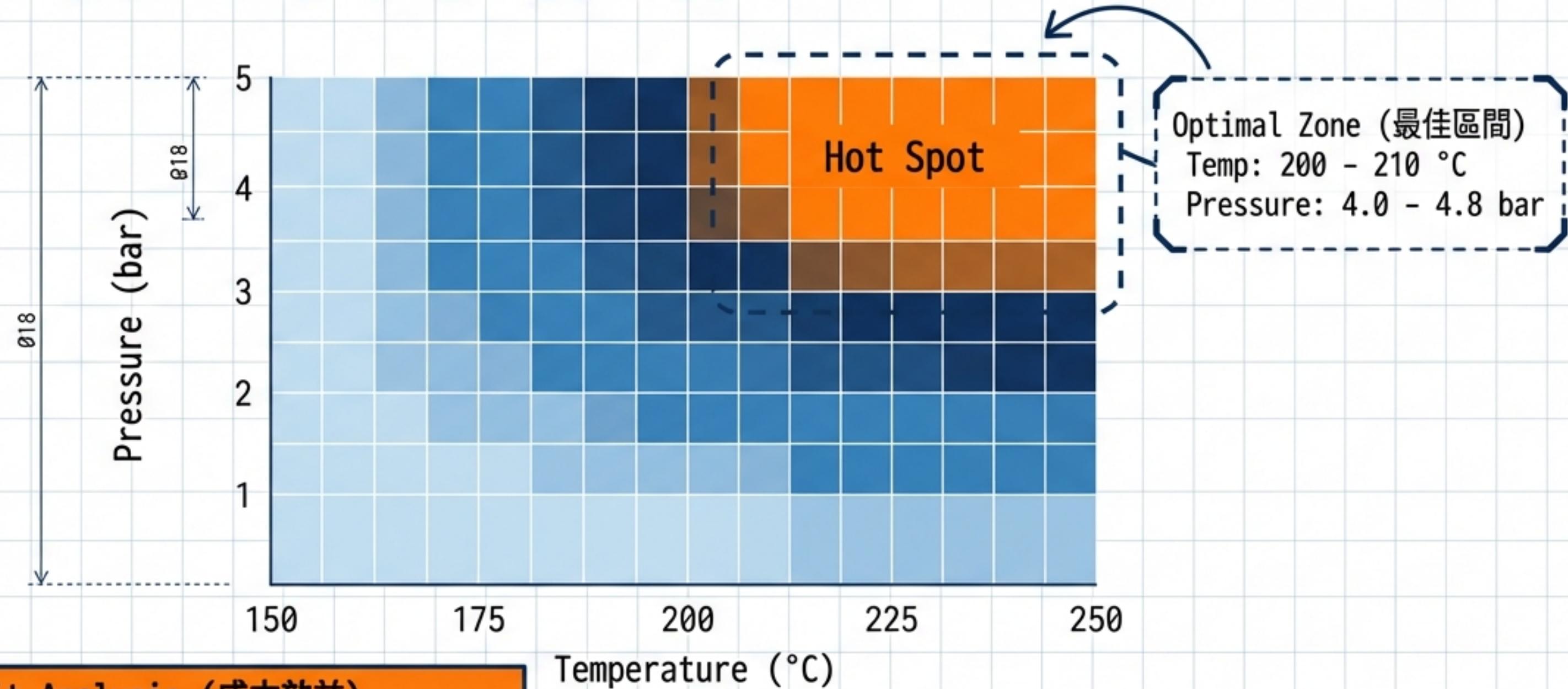
Industrial Blueprint v1

Project

Date



# 優化結果：最佳操作視窗



## Cost-Benefit Analysis (成本效益)

Catalyst Reduction: 1.5g  $\rightarrow$  0.78g (Save 48%)

Specific Setpoint: 203°C / 4.51 bar / 0.78g

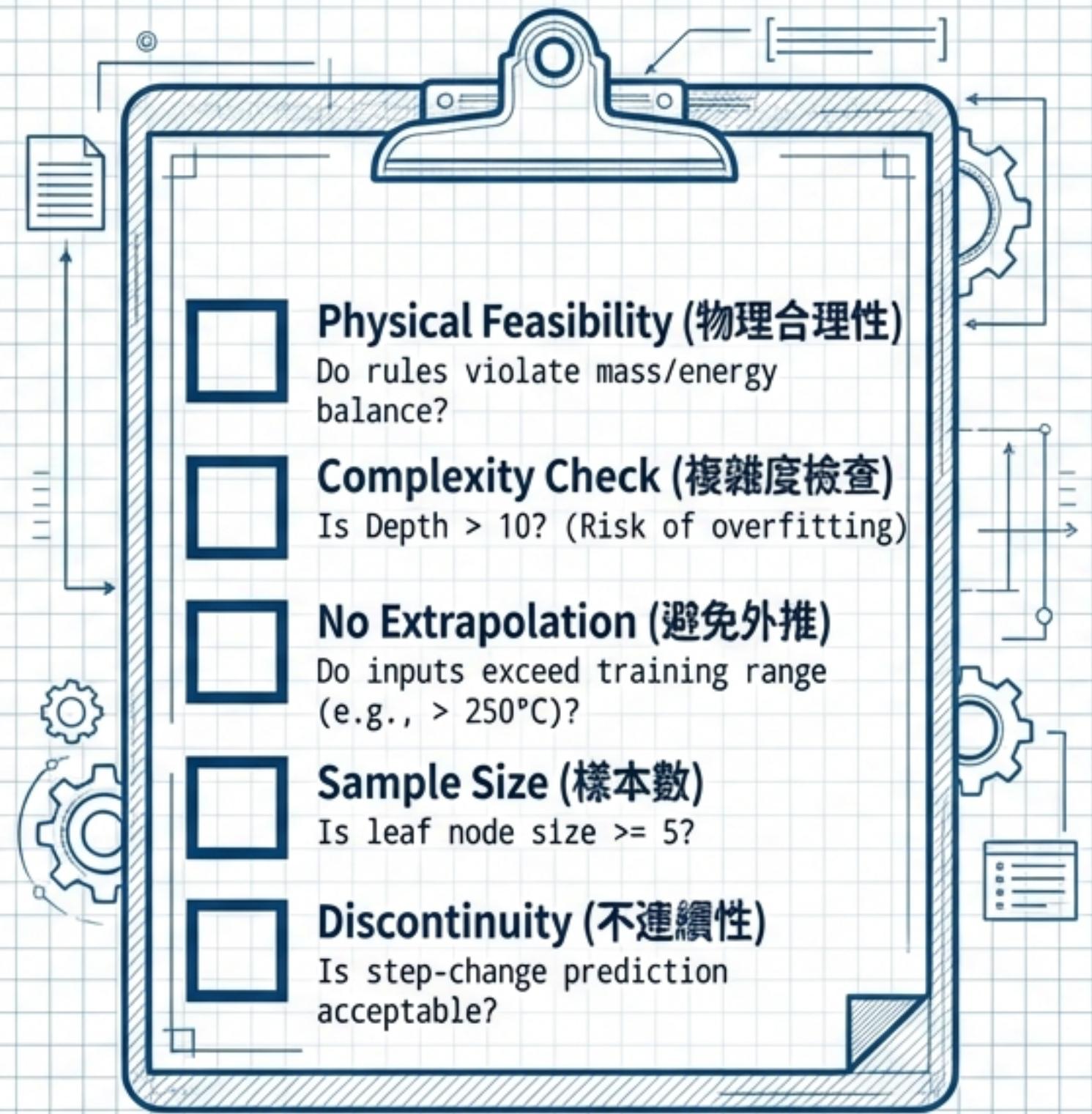
Industrial Blueprint v1

Project

Date

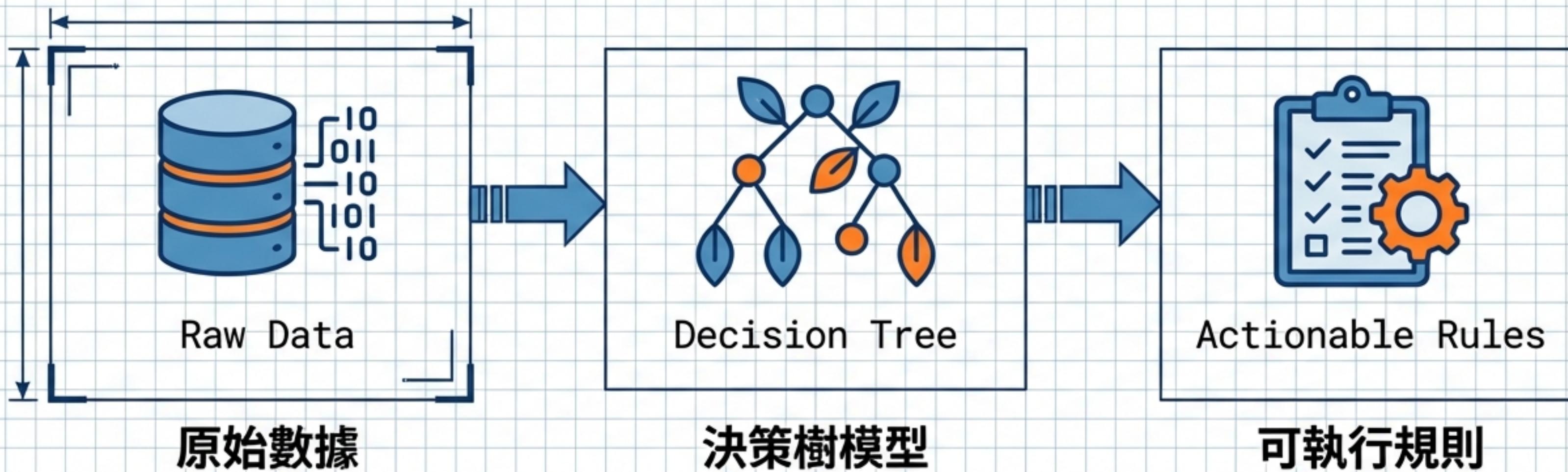


# 工程師檢查清單 (Engineer's Checklist)





# 總結：從數據到規則的橋樑



## Key Takeaways

1. Non-Linearity: 適合處理化工動力學與閾值效應。
2. Transparency: 提供清晰的 IF-THEN 規則，非黑箱作業。
3. Evolution: 下一單元將介紹隨機森林 (Random Forests) 以提升精度。