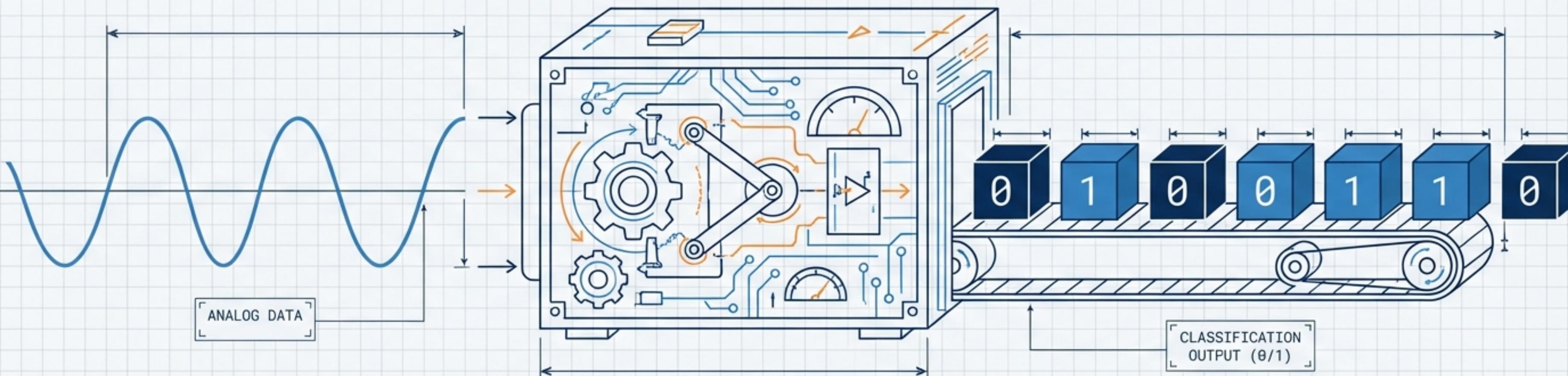


Unit 12 分類模型總覽

Noto Sans TC

從數值預測到數位決策



Noto Sans TC

Course: AI在化工上之應用

逢甲大學 智慧程序系統工程實驗室 | 莊曜楨 助理教授



分類問題的本質 (The Nature of Classification)

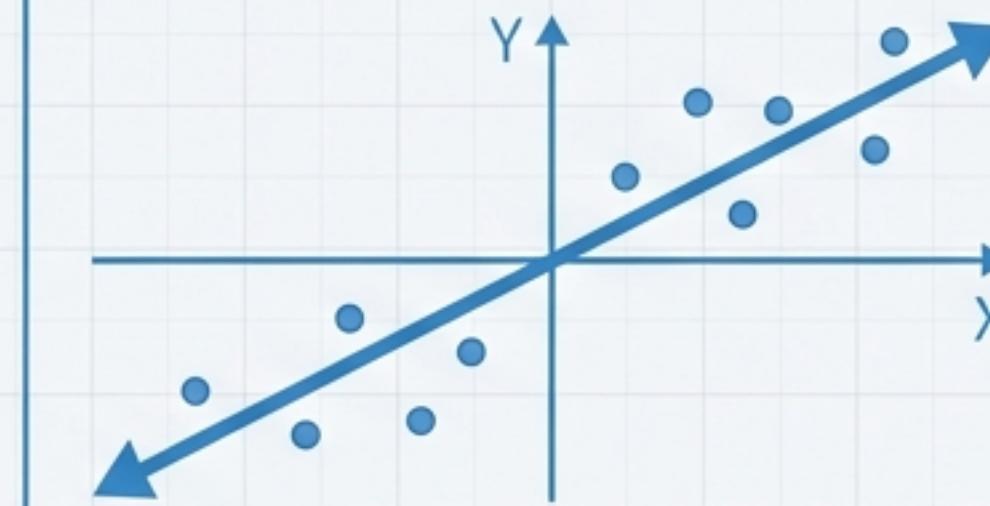
預測離散類別 (Predict Discrete Labels)

- 二元分類 (Binary): 安全 / 危險 (0/1)
 
- 多元分類 (Multi-class): 等級 A / B / C

- 多標籤分類 (Multi-label): 可燃 + 毒性

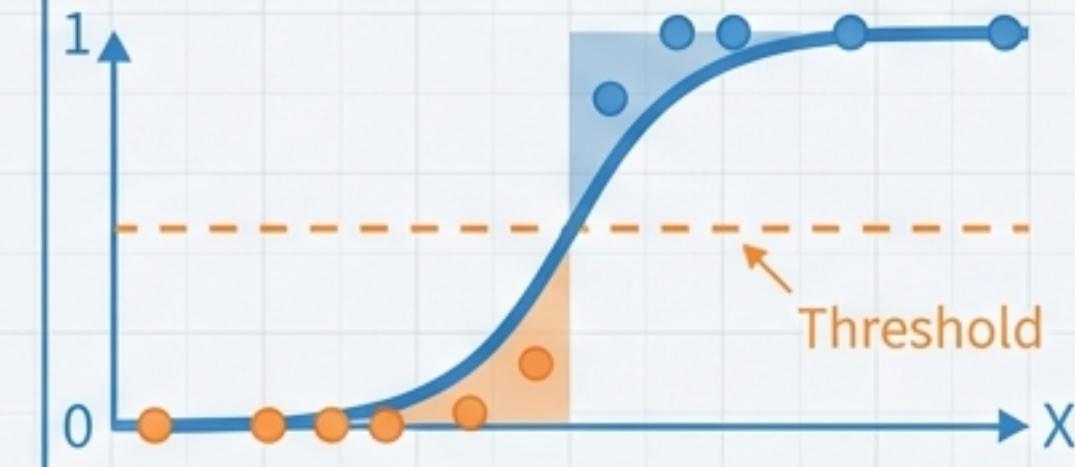


Regression (Continuous)



Activation Function
(激活函數)
Roboto Mono

Classification (Discrete)



分類是將連續的機率映射為離散的決策。

基礎工具組：線性與機率模型 (Linear & Probabilistic Models)

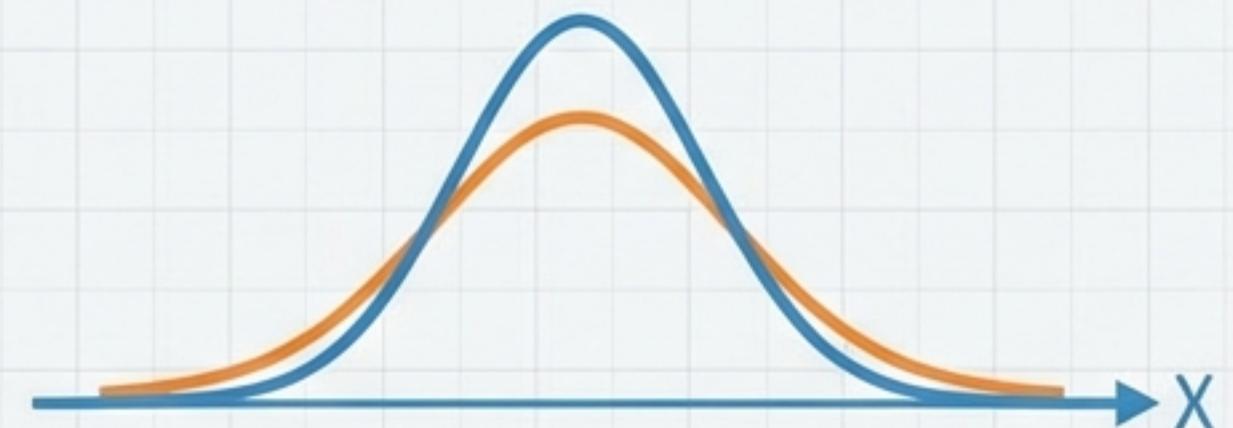
1. 邏輯迴歸 (Logistic Regression)



- 名為「回歸」，實為分類器
- 適用：線性可分問題、高可解釋性
- 核心公式 (Sigmoid)：

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

2. 貝氏分類器 (Gaussian Naive Bayes)

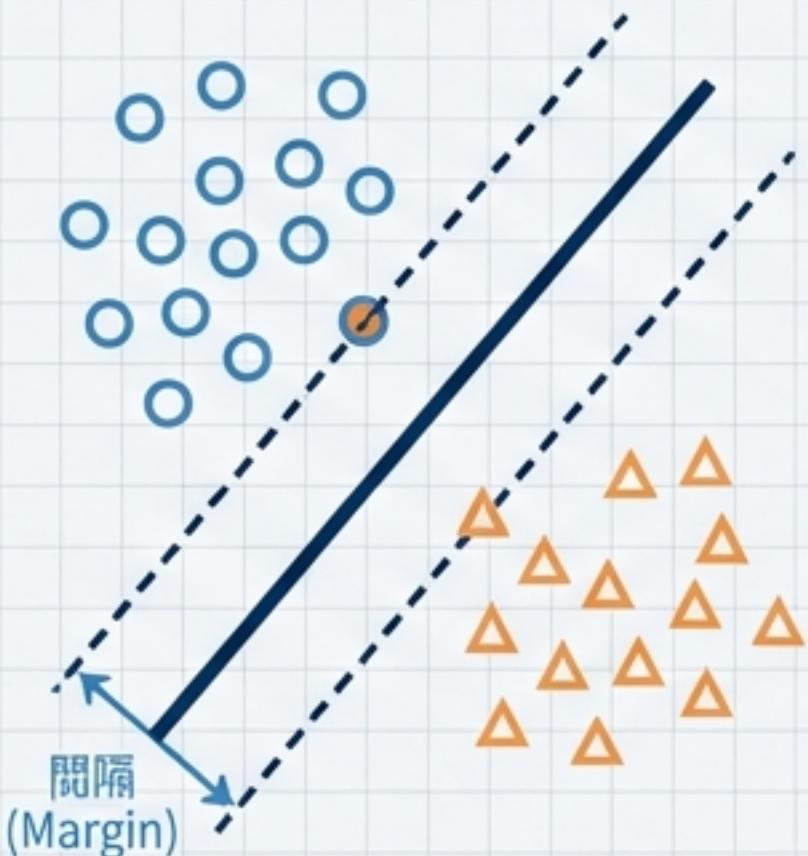


- 基於貝氏定理與特徵獨立假設 (Naive)
- 優勢：訓練極快，適合建立基準線 (Baseline)
- 核心公式：

$$P(y|x) \propto P(y) \prod_{x_i} P(x_i|y)$$

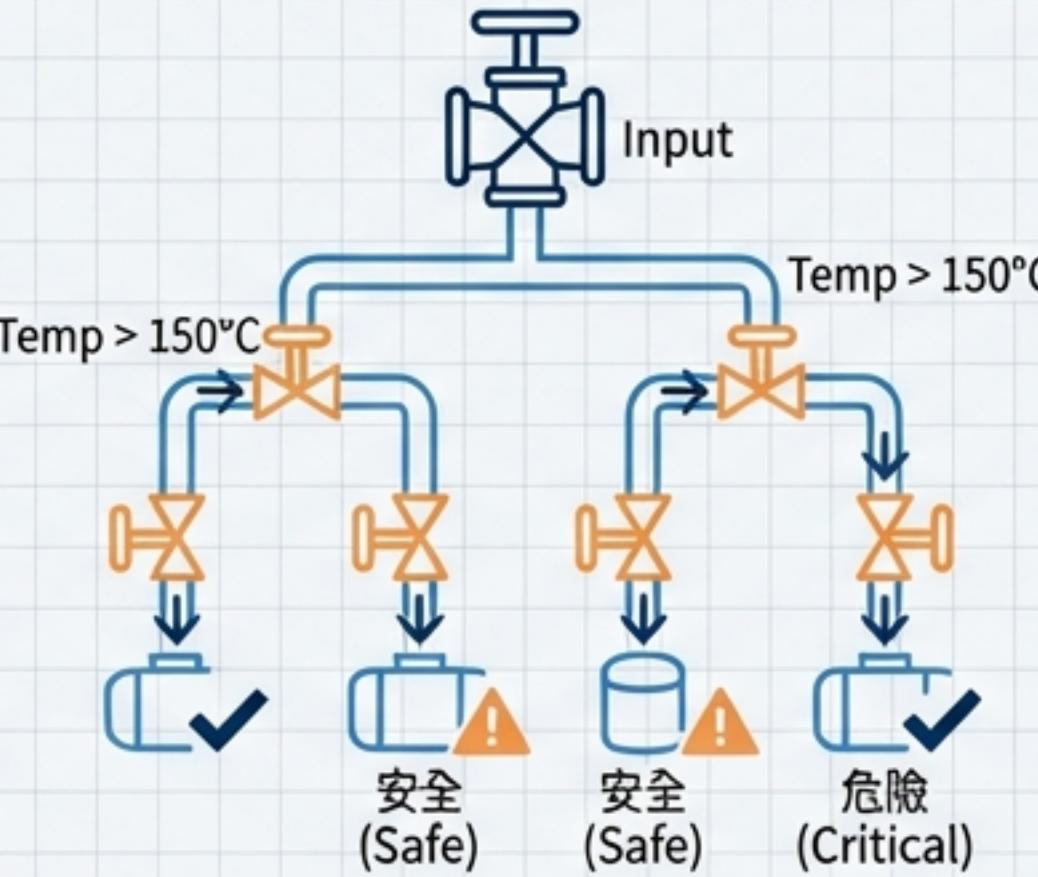
進階工具組：非線性與複雜模型

支持向量分類 (SVC)



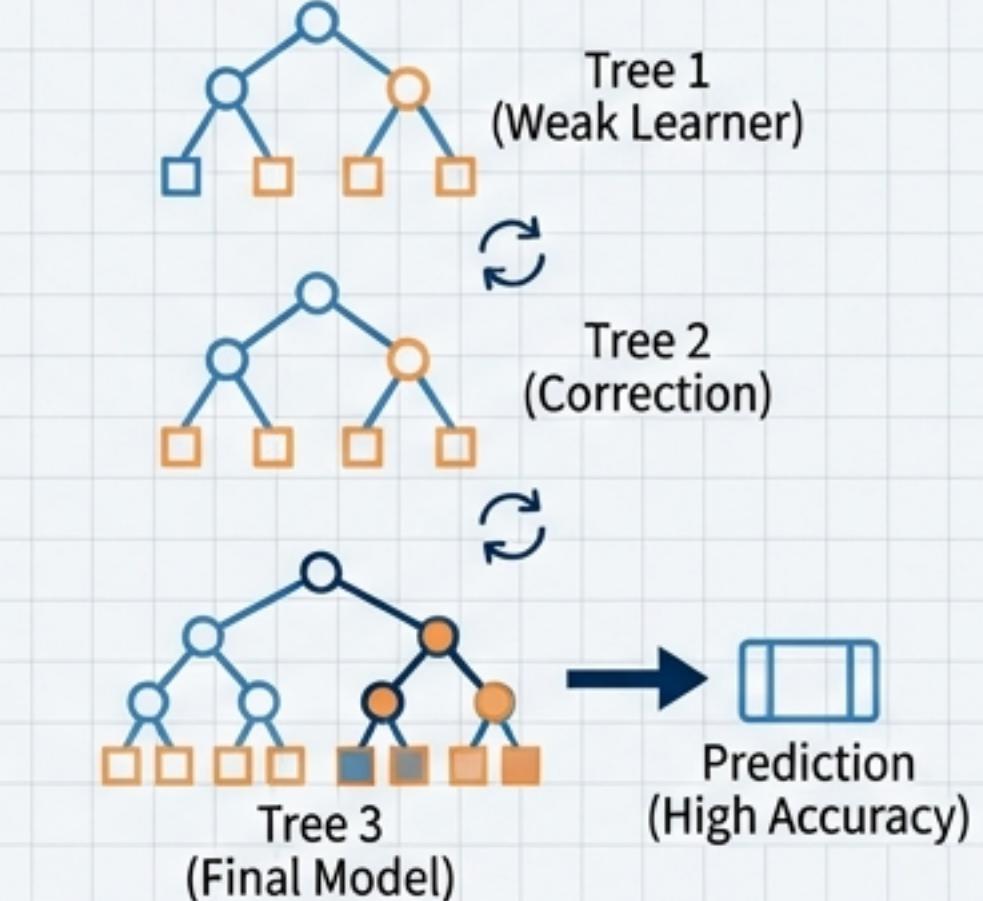
- 尋找最寬的間隔 (Margin)
- 核函數 (Kernel) 處理高維非線性

決策樹 (Decision Tree)



- 類似化工流程圖 (Flowchart)
- 白箱模型，邏輯可見

梯度提升 (Gradient Boosting)



- 迭代修正錯誤 (Boosting)
- 追求最高準確度的重型機具

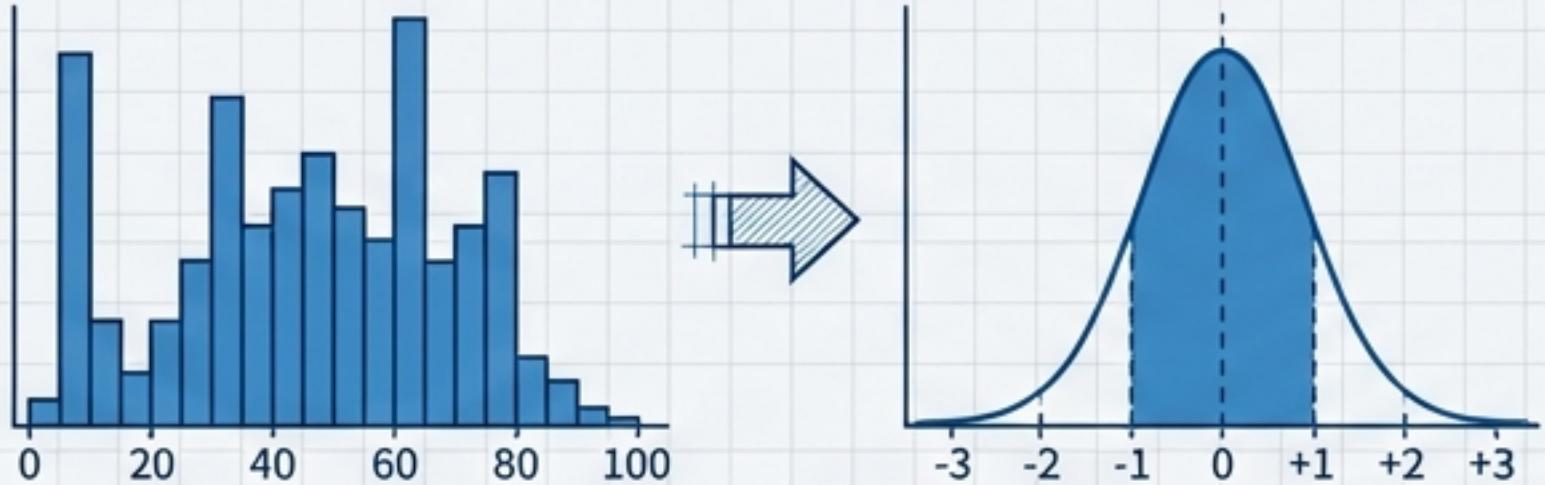
模型選擇矩陣 (Model Selection Matrix)

| 模型 | 線性/非線性 | 可解釋性 | 訓練速度 | 資料量需求 |
|---------------------|--------|---|--|--------------|
| Logistic Regression | 線性 | <input checked="" type="radio"/> High (高) | <input checked="" type="radio"/> Fast (快) | Medium-Large |
| Naive Bayes | 線性 | <input checked="" type="radio"/> High (高) | <input checked="" type="radio"/> Very Fast | Small-Medium |
| Decision Tree | 非線性 | <input checked="" type="radio"/> High (高) | <input checked="" type="radio"/> Fast (快) | Medium |
| SVC | 非線性 | <input type="radio"/> Low (低) | <input type="radio"/> Slow (慢) | Small-Medium |
| Gradient Boosting | 非線性 | <input type="radio"/> Medium (中) | <input type="radio"/> Slow (慢) | Medium-Large |

沒有萬能的模型，只有最適合場景的權衡 (Trade-off).

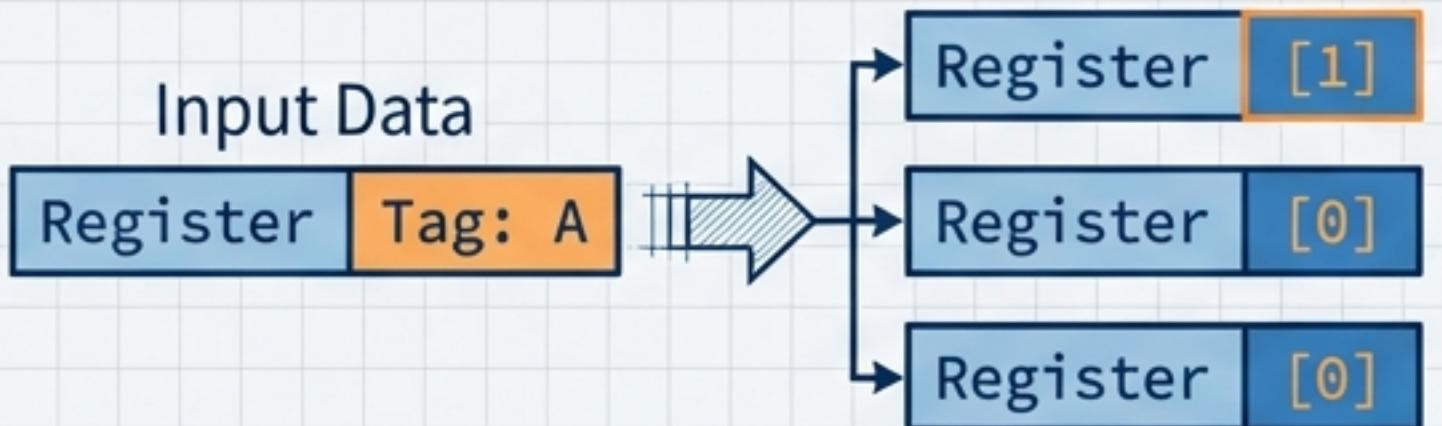
進料前處理：校準與設置 (Data Preprocessing)

1. 特徵縮放 (Scaling)



Logistic / SVC 必須標準化 (Standardization)。決策樹不需要。

2. 編碼 (Encoding)



One-Hot Encoding: 將類別轉為二進位向量。

3. 類別不平衡 (Class Imbalance) - 關鍵風險



- 問題：模型容易忽略少數類
- 解法：類別權重 (Class Weights) 或 重採樣 (SMOTE)

評估儀表板：混淆矩陣 (The Confusion Matrix)

在化工安全中，漏報 (FN) 比誤報 (FP) 更致命

| | | Predicted Class (預測) |
|-------------------|-----------------------------------|----------------------|
| | | TP (真陽性) 成功預測危險 |
| Actual Class (實際) | FN (假陰性) 漏檢危險 (Critical Error) | |
| | FP (假陽性) 誤報警 (False Alarm) | TN (真陰性) 正確預測安全 |

關鍵性能指標 (Critical Metrics)

準確率 (Accuracy)

$$\frac{TP+TN}{Total}$$

注意：在不平衡資料中可能誤導



精確率 (Precision)

$$\frac{TP}{TP+FP}$$

預測為「正」的可信度



召回率 (Recall / Sensitivity)

$$\frac{TP}{TP+FN}$$

安全關鍵指標 — 能找出多少比例的異常？



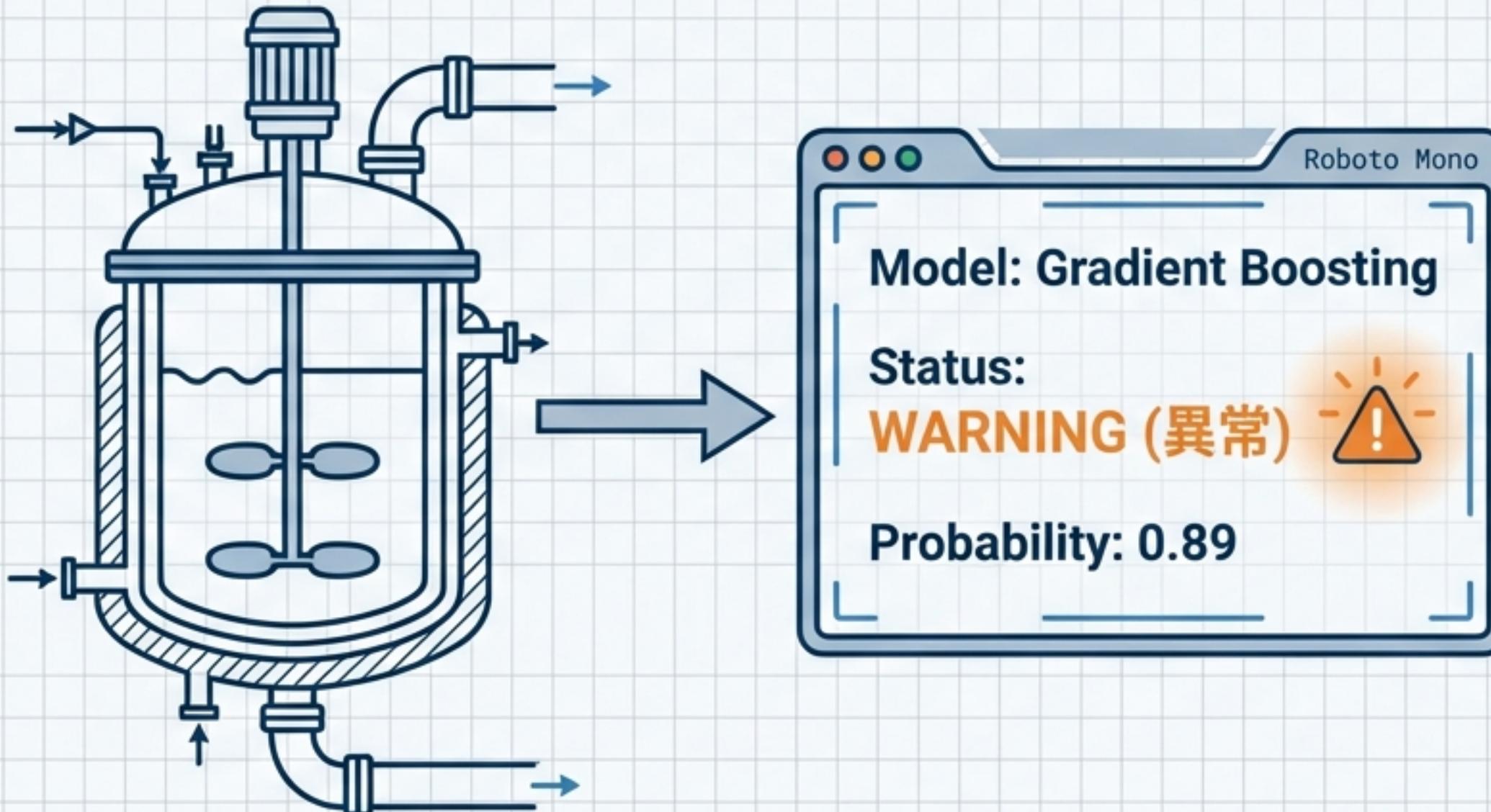
F1-Score

Precision 與 Recall 的平衡點



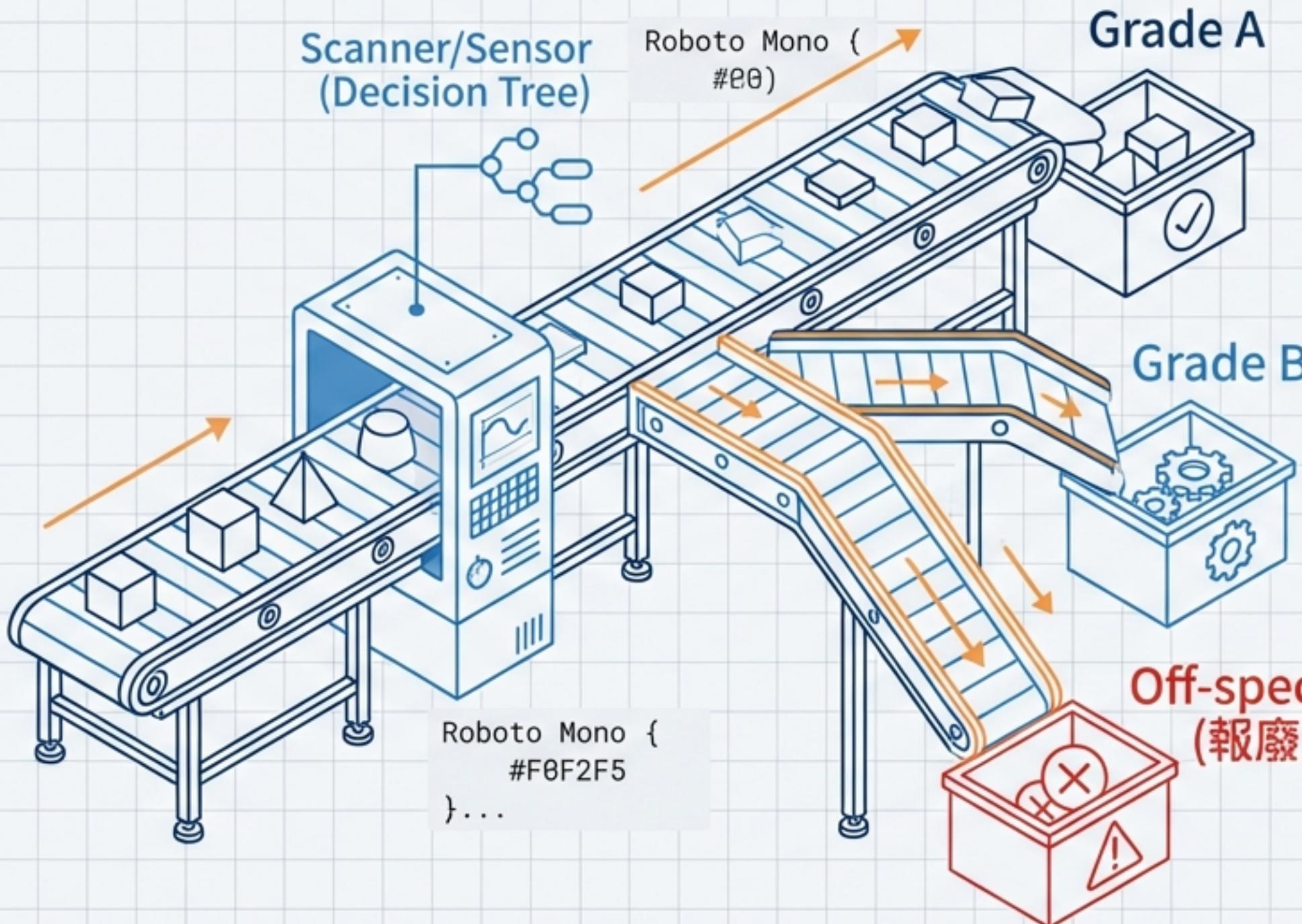
AUC-ROC：評估模型在各種閾值下的整體能力

應用場景 I：製程異常檢測 (Process Safety)



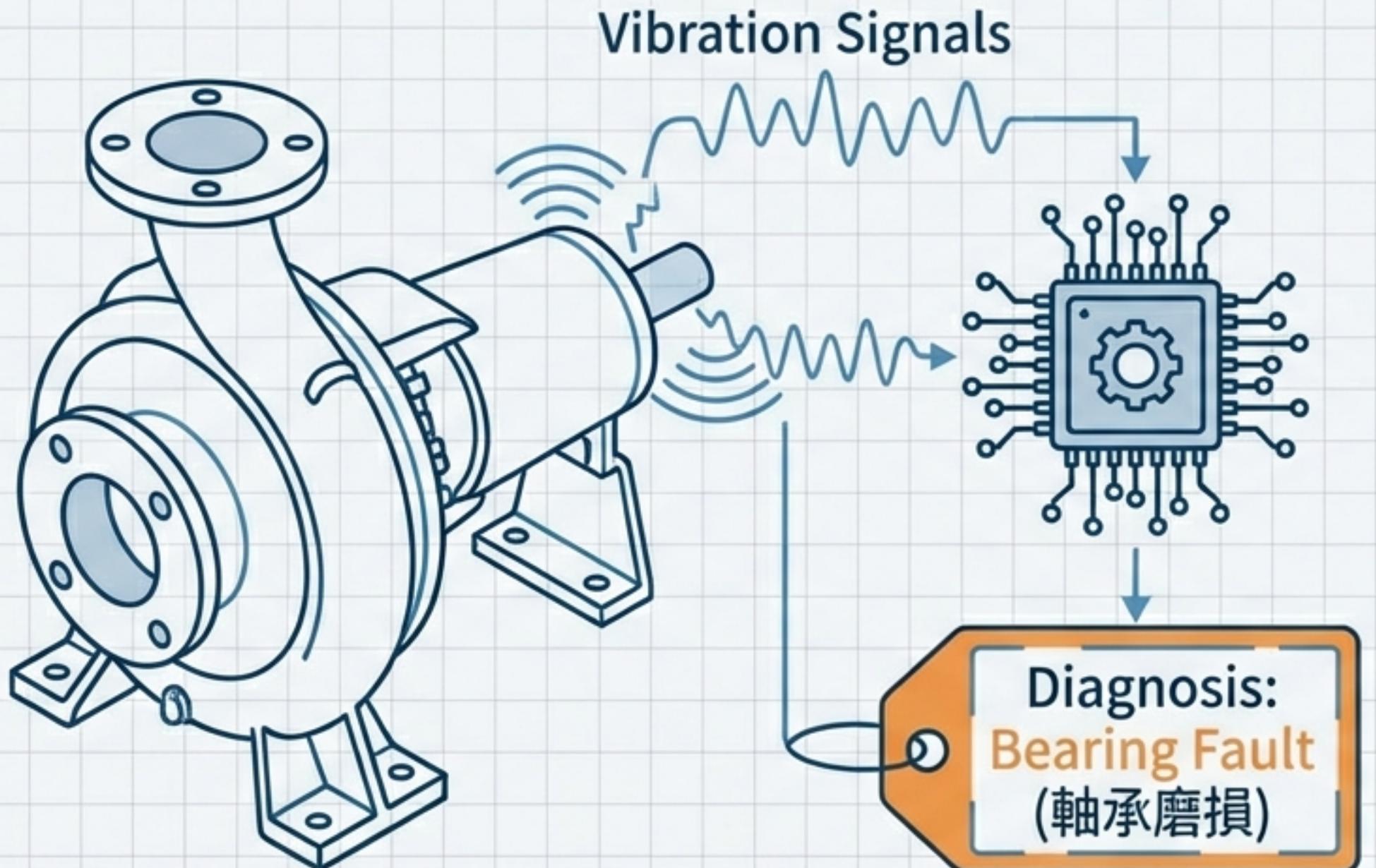
- 目標：區分「正常操作」與「異常狀態」
- 模型：Isolation Forest 或 Gradient Boosting
- 關鍵需求：High Recall (高召回率)
- 策略：寧可誤報 (False Alarm)，絕不能漏檢 (Missed Danger)

應用場景 II：產品品質分級 (Product Quality)



- 目標：多元分類 (Multi-class Classification)
- 模型：Decision Tree (決策樹)
- 優點：QC 人員可檢視分級規則 (可解釋性)
- 關鍵需求：**High Precision (高精確率)** — 避免將劣質品賣給客戶

應用場景 III：設備故障診斷 (Fault Diagnosis)

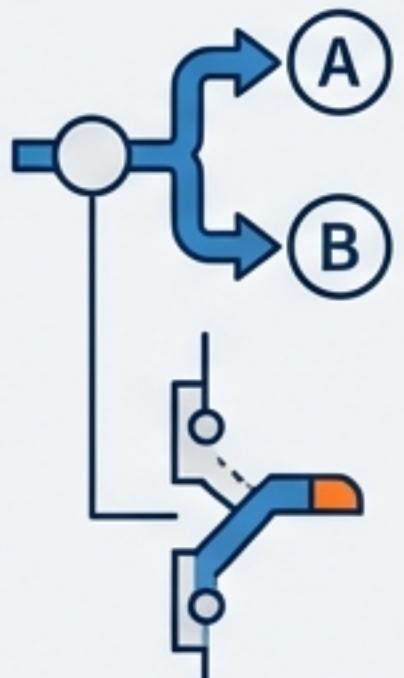


- 場景：泵浦振動頻譜分析
- 目標：識別具體故障類型（正常 / 軸承 / 葉輪 / 不平衡）
- 模型：SVC 或 Random Forest (處理高維特徵)
- 效益：**預測性維護 (Predictive Maintenance)**，減少非計畫停機

實務教戰守則：從實驗室到工廠 (Best Practices)

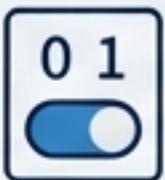


分類 vs. 回歸：決策指南 (The Decision Guide)



Classification (分類)

- Output: 標籤 (Label)
- Question: 是 A 還是 B?
(Which one?)
- Action: 決策 (Decision) –
停機/繼續，合格/報廢



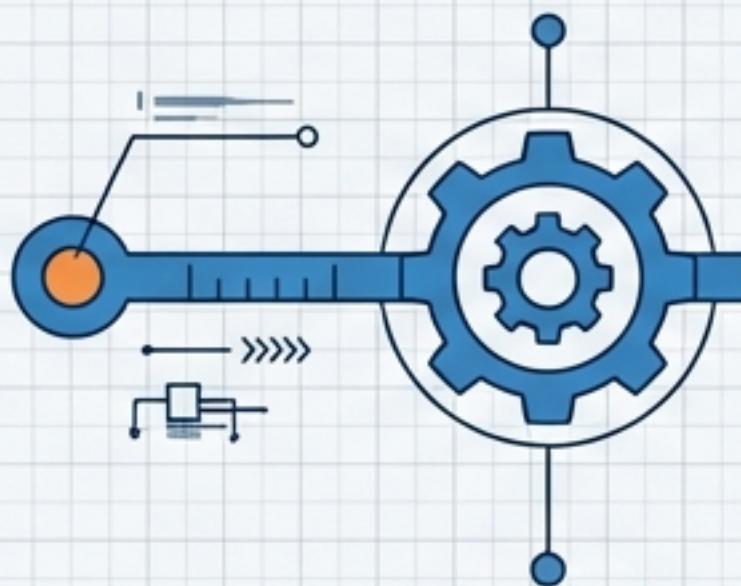
Regression (回歸)

- Output: 數值 (Number)
- Question: 是多少? (How much?)
- Action: 量化 (Quantification) –
產量預測，溫度設定

Rule of Thumb: 若需要立即採取行動 (Act)，通常是分類問題。

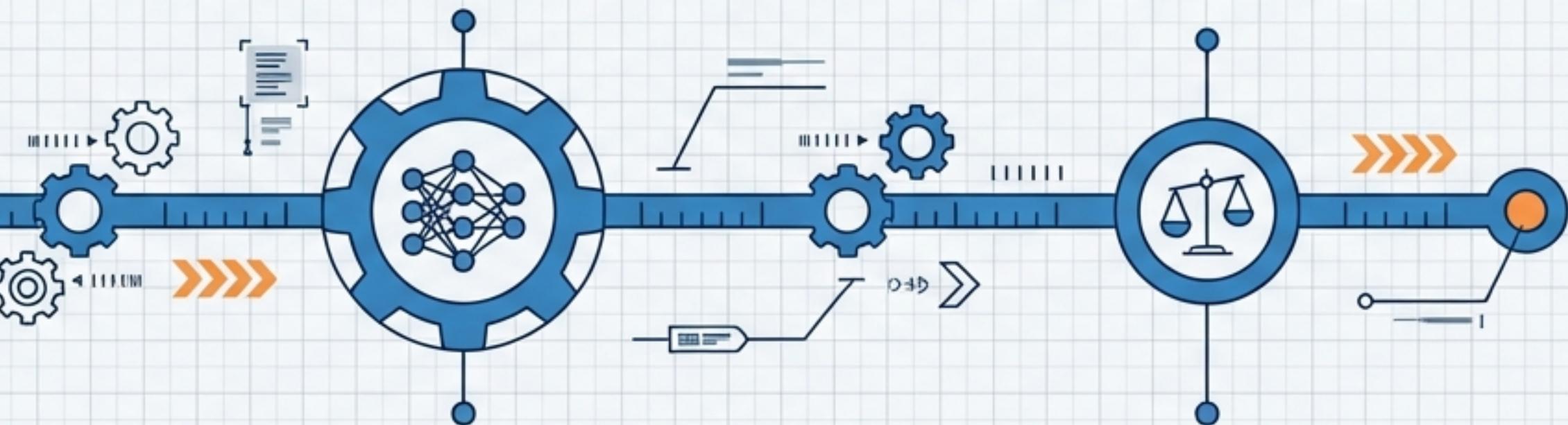
未來技術藍圖 (Future Roadmap)

集成學習進階



Stacking, Voting,
XGBoost

深度學習 (Deep Learning)

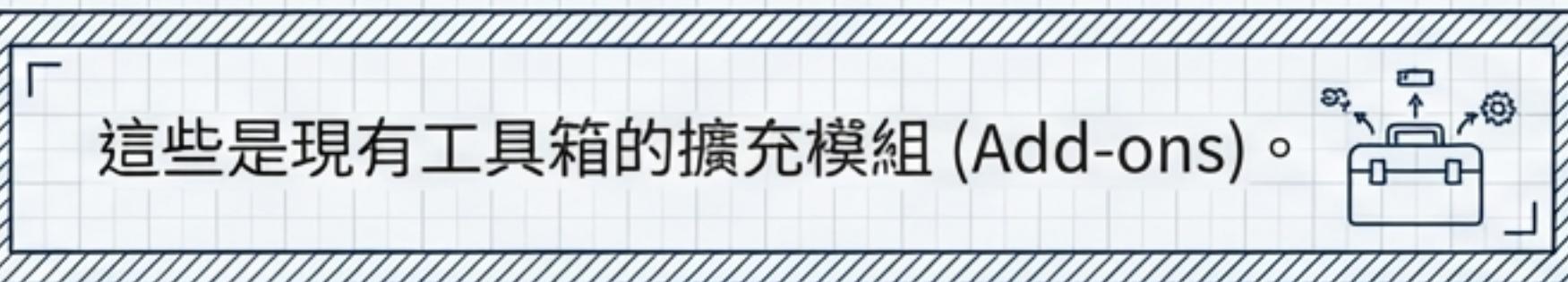


CNN (影像瑕疵檢測),
RNN (時序)

進階不平衡處理



ADASYN,
Borderline-SMOTE



結語：您的數位工具箱已備妥

- 模型將數據轉化為明確的 **決策 (Decisions)**。
- 評估指標必須對應 **業務目標** (安全優先 vs. 效率優先)。
- 從簡單模型開始，逐步優化。

**「Safety First: 寧可接受虛驚一場 (False Alarm)，
不可錯失一次危機 (Missed Danger)。」**



下一步 (Next Step)：
開啟 **Unit12_Logistic_Regression.ipynb**
現在就開始您的實作練習！