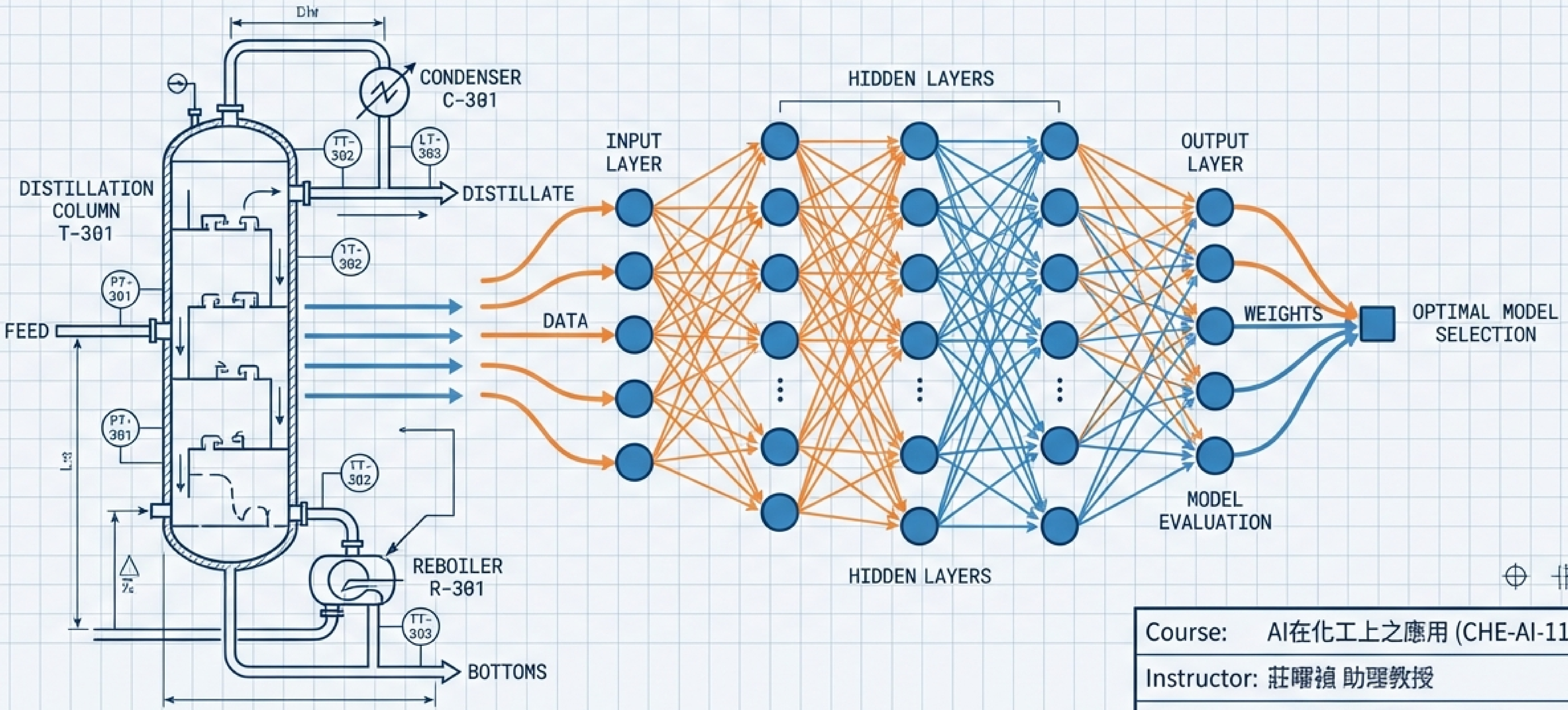
Unit 14 模型評估與選擇

工程設計驗證框架



Course:	AI在化工上之應用 (CHE-AI-114)	
Instructor:	莊曜禎 助理教授	
Concept:	從數學計算到工程品質保證	16:9

為什麼評估至關重要？

code

"In God we trust, all others bring data." — W. Edwards Deming

Specification Sheet Comparison

Academic ML (學術界)	Industrial AI (工業界)
目標：追求高準確度	目標：安全性、穩健性、經濟效益
1.0 Metric: $R^2 \approx 1.0$	Metric: ROI & Safety

Risk Factors

Techninical subtitilles

- 安全風險：故障預測失準 → 設備事故
- 經濟損失：產量預測偏差 → 原料浪費
- 法規問題：排放預測錯誤 → 環保罰款

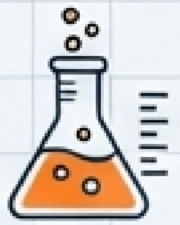
INDUSTRIAL BLUEPRINT v1

PROJECT: AI EVALUATION FRAMEWORK

回歸模型儀表板：連續數值的感測器

MAE (平均絕對誤差)

特性：直觀、對異常值不敏感
(Robust)



應用：預測反應物濃度

RMSE (均方根誤差)

特性：對大誤差敏感
(Penalizes Spikes)

應用：反應器溫度控制

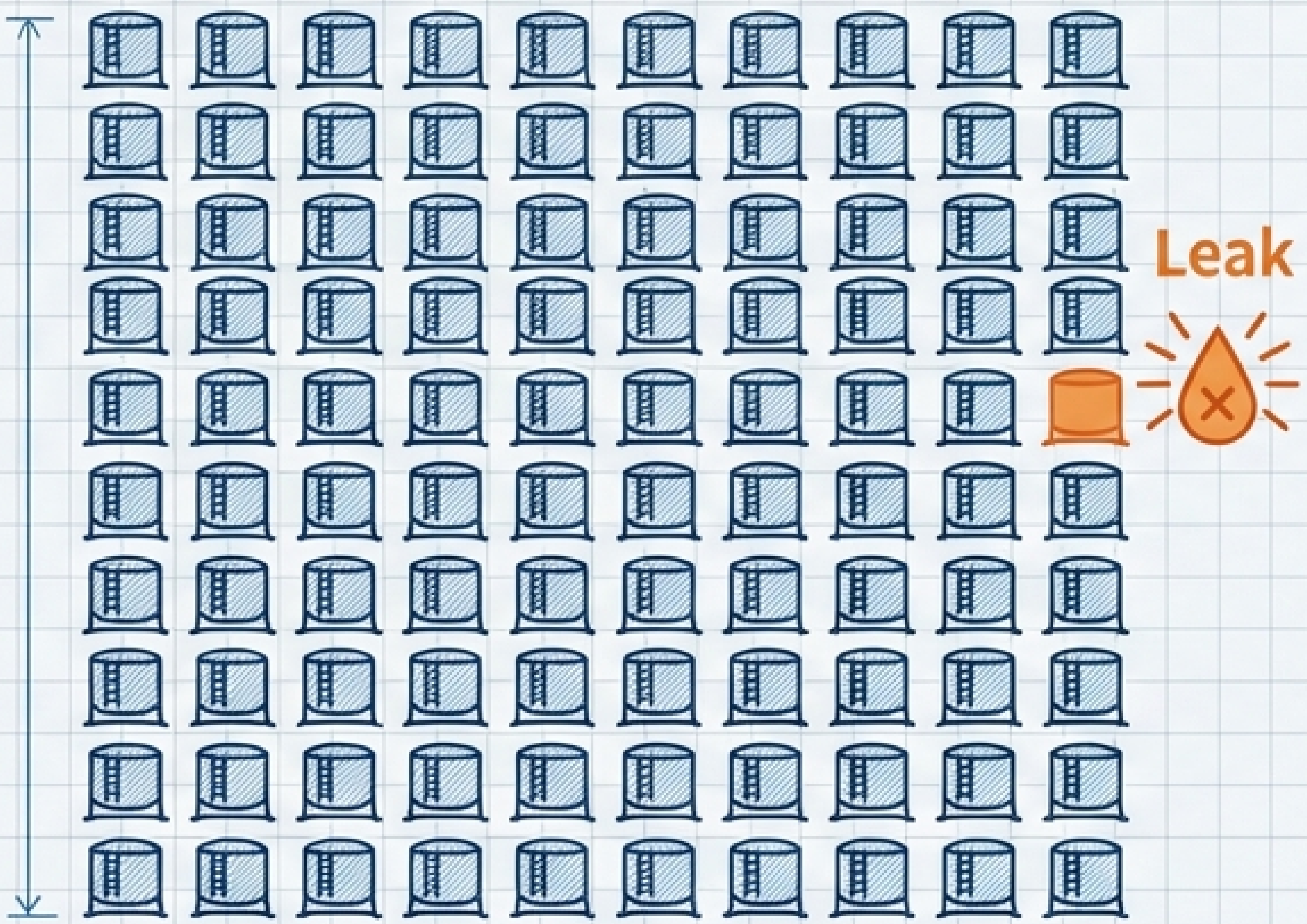
MAPE (平均絕對百分比誤差)

特性：相對誤差
(Relative Error %)

應用：跨尺度比較或商業溝通

分類指標陷阱：準確率的假象

Tank Farm Schematic



真實數據：99 Normal / 1 Fault

FAIL

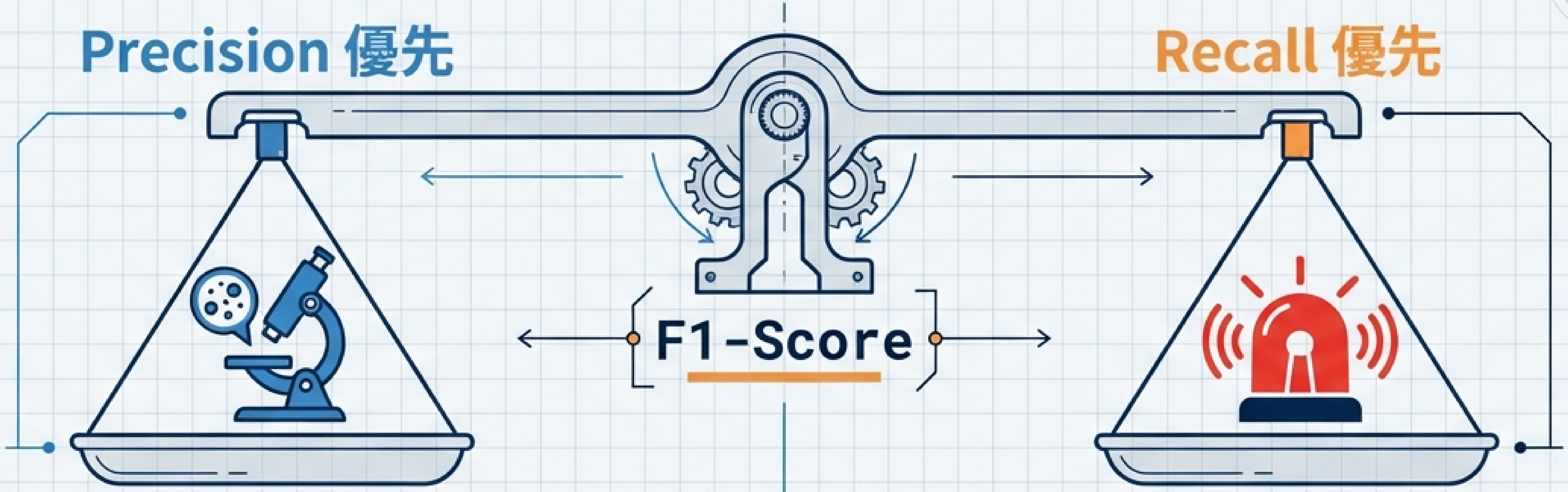
Model A (預測全部正常)

code

- 預測結果：100 Normal / 0 Fault
- Accuracy : 99% (看似完美 ✨)
- 故障檢出率 : 0% (工廠爆炸 ❌)

關鍵結論：在不平衡數據 (Imbalanced Data) 下，Accuracy 毫無意義。必須使用 Precision 與 Recall。

權衡的藝術：Precision vs. Recall

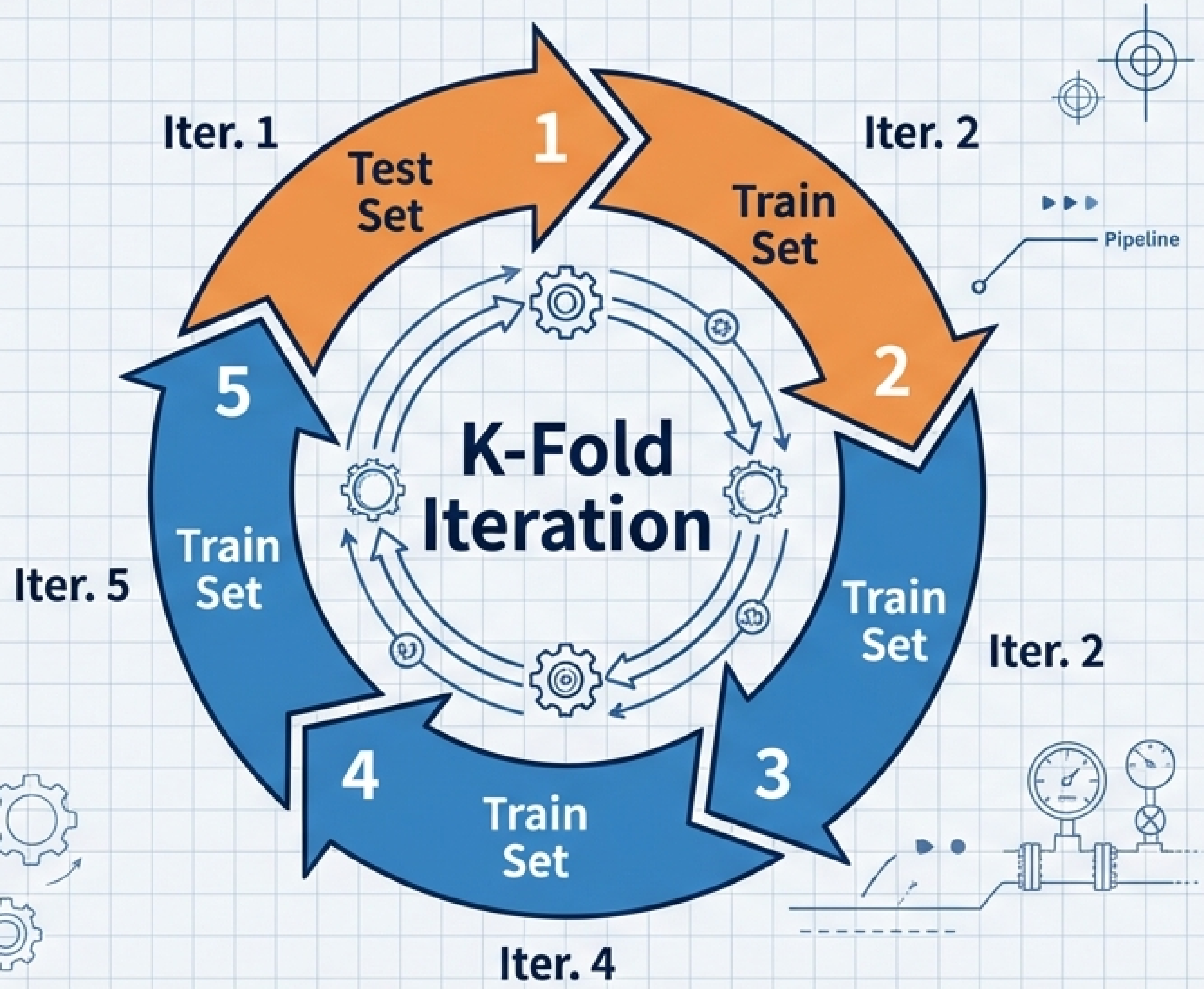


場景：新材料篩選
Goal: 避免誤判 (FP 代價高)
Formula: Precision = TP / (TP + FP)
寧缺勿濫

場景：設備故障預警
Goal: 不可漏報 (FN 代價極高)
Formula: Recall = TP / (TP + FN)
寧可誤報，不可漏報

壓力測試：交叉驗證 (Cross-Validation)

K-Fold Iteration



標準程序 (Standard K-Fold)

- 輪流將每一份當作測試集，評估更穩健。

進階技術 (Stratified K-Fold)

- 確保每個 Fold 的類別比例與原始數據相同。

⚠ 適用於不平衡數據 (如故障檢測)

時間序列的特殊考量

Time Series Split (Expanding Window)

問題：隨機劃分 (Random Split) 會導致“用未來預測過去”(Look-ahead Bias)。
解決方案：訓練集始終在測試集之前。
案例：反應器溫度預測。

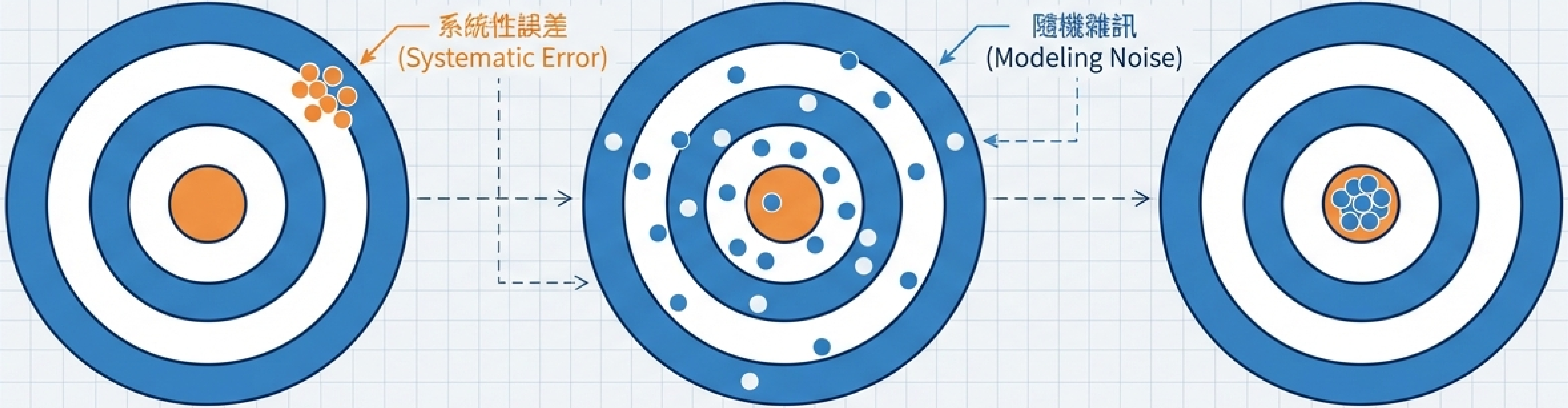
模型診斷：偏差與方差 (Bias-Variance Tradeoff)

Noto Sans TC | Roboto Mono

Underfitting / High Bias

Overfitting / High Variance

Ideal / Low Bias & Low Variance



模型太簡單
(e.g., Linear)

模型太複雜
(e.g., Poly-20)

理想模型
(Goldilocks Zone)

$$\text{Total Error} = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$

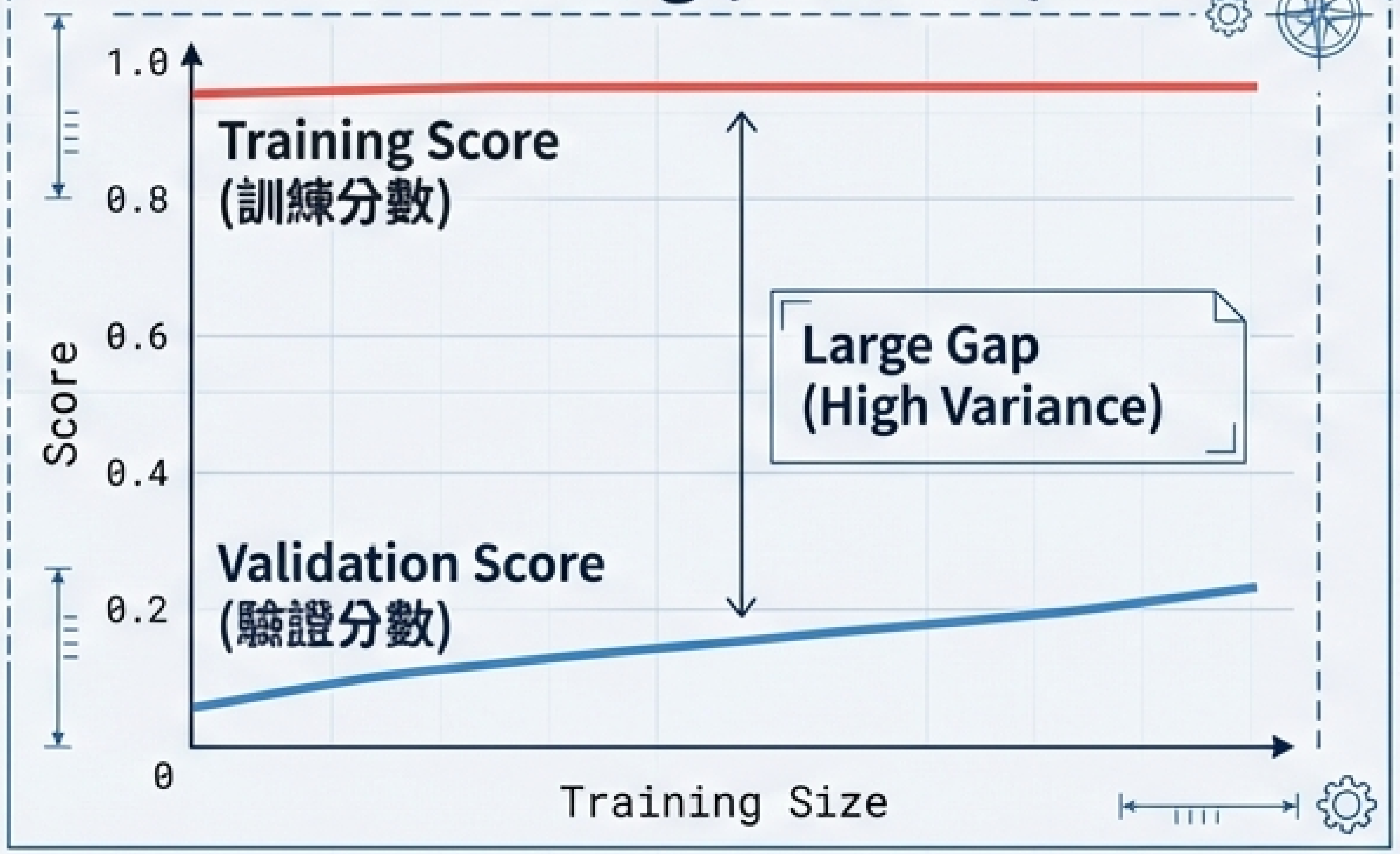
INDUSTRIAL BLUEPRINT v1

PROJECT: AI EVALUATION FRAMEWORK | 模型診斷

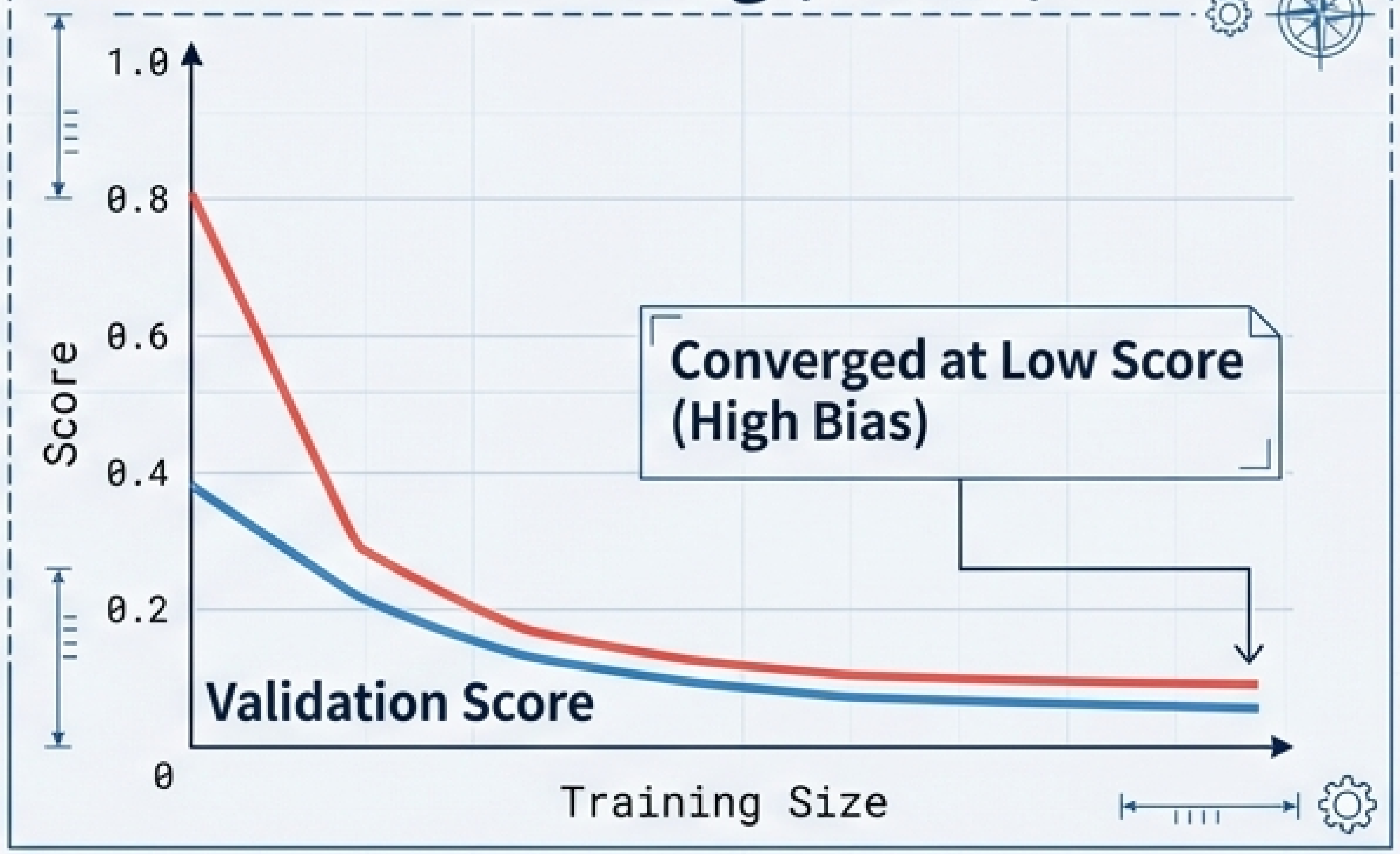
讀懂性能曲線 (Learning Curves)

Roboto Mono

Overfitting (過度擬合)



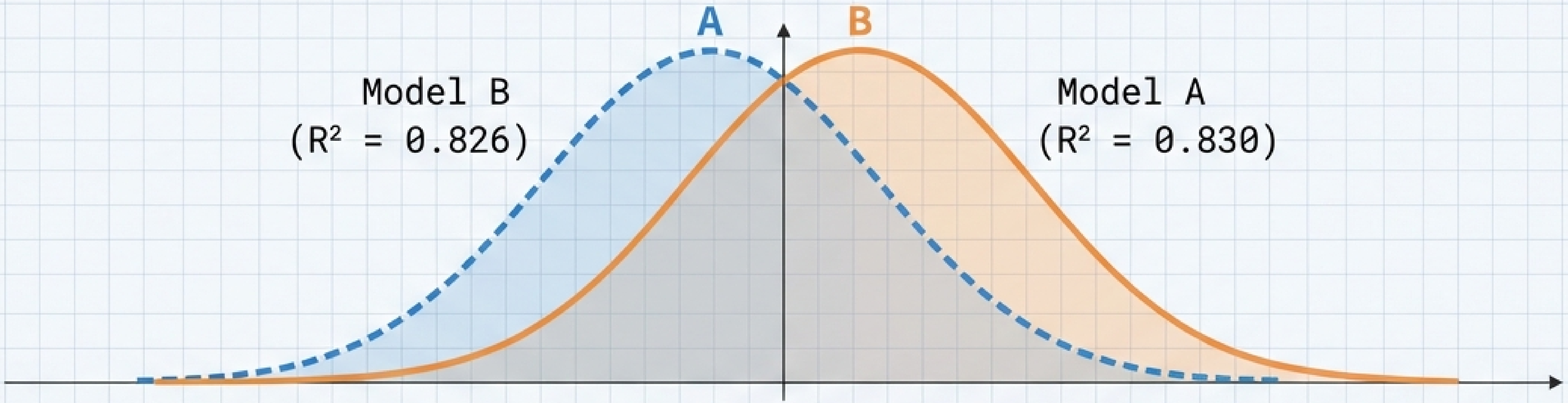
Underfitting (欠擬合)



策略: 增加數據 或 正則化

策略: 增加特徵 或 提高複雜度

統計檢定：確認改進的顯著性



差異僅 0.004。是真實改進還是隨機波動？

Paired t-test
(配對 t 檢定)

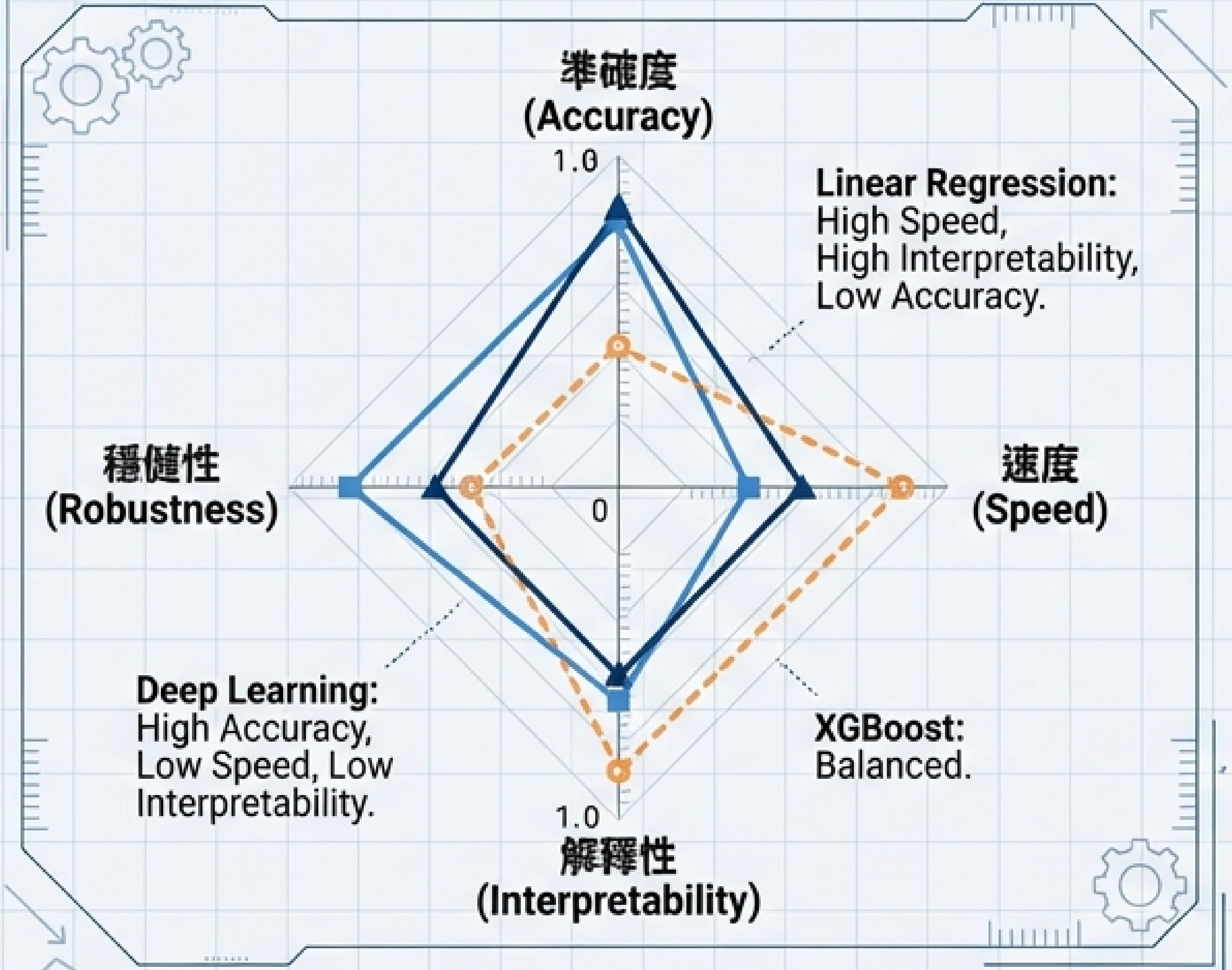
$p < 0.05$ (顯著 Significant)
選擇 Model A

$p > 0.05$ (不顯著 Not Significant)
選擇較簡單的模型 (Occam’s Razor)

別被微小的數字差異誤導。
(Don't be misled by minor differences).

多目標決策矩陣 (Multi-Objective Decision Making)

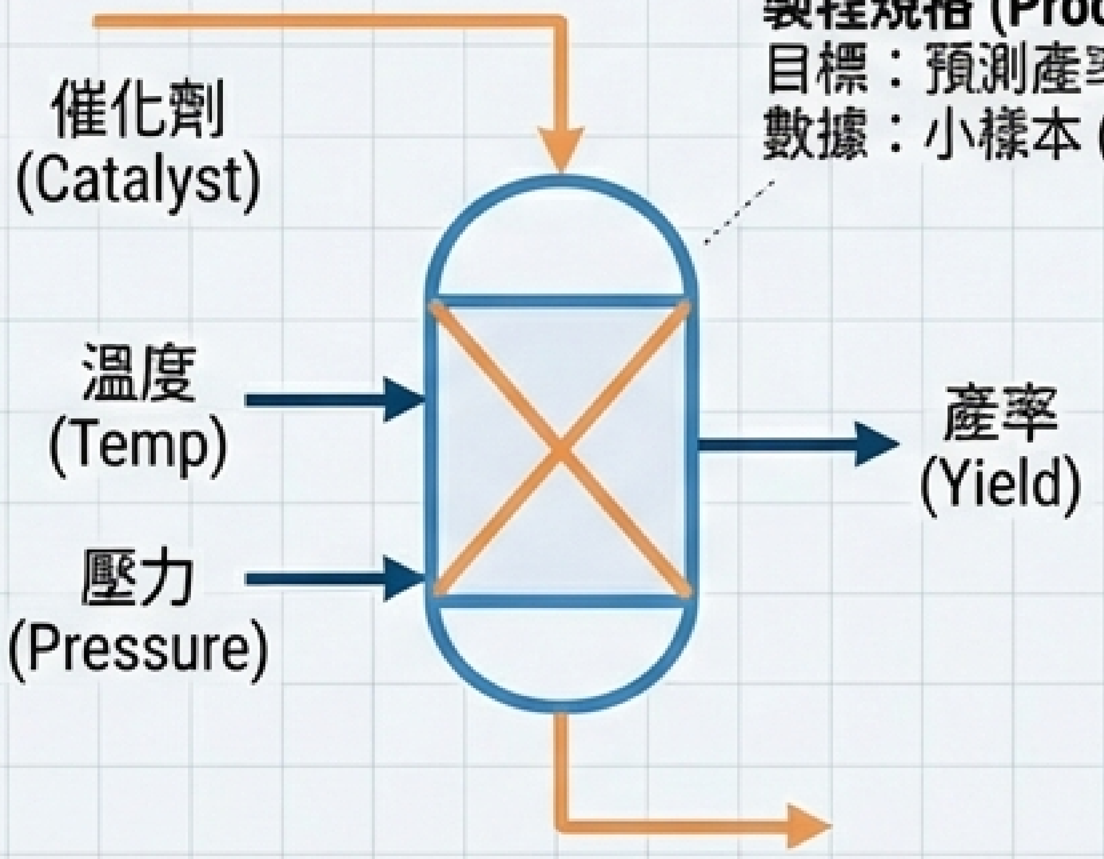
Roboto Mono : Primary Blue



Model	準確度 (Accuracy)	解釋性 (Interpretability)	速度 (Speed)
Linear Regression	準確度：低	解釋性：極高 (Safe)	速度：<0.1ms
XGBoost	準確度：高	解釋性：中等	速度：中等
Deep Learning	準確度：最高	解釋性：黑箱 (Black Box)	速度：慢

決策邏輯：即時控制 (Real-time) **需犧牲準確度換取速度；法規合規需犧牲準確度換取解釋性。**

案例分析 I：催化劑產率預測 (Regression)



製程規格 (Process Specs):
目標：預測產率 (60-95%)
數據：小樣本 (100 runs)

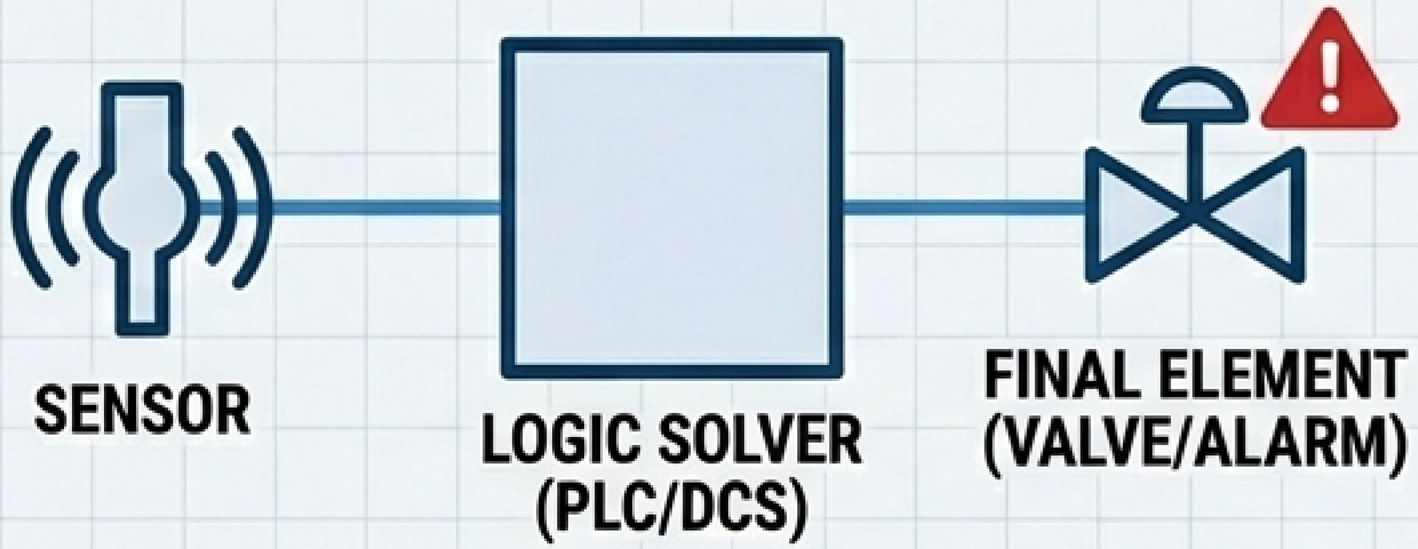
模型	MAE (%)	R^2	說明 (Note)
Linear Regression	2.3	0.85	低擬合 (Underfitting)
Random Forest	1.8	0.91	良好解釋性 (Good Interpretability)
XGBoost	1.5	0.93	最佳準確度 (Best Accuracy)

最終決策：選擇 XGBoost
理由：追求最高產率優化，配合 SHAP 分析補足解釋性。

案例分析 II：反應器故障檢測 (Classification)

Roboto Mono : Primary Blue

Safety Interlock System



編程說明：
背景：類別不平衡 (Normal 90%, Fault 10%)
關鍵要求：嚴重故障 (Class C) 不能漏報

	Model 1 Logistic Regression	Model 2 XGBoost (Optimized)
1	Accuracy: 92%	Accuracy: 95%
2	Class C Recall: 45% ❌ (Dangerous!)	Class C Recall: 85% (Better)

Optimization Graphic (Threshold Tuning)

犧牲部分 Precision，將 Class C Recall 提升至 94% ✅

安全系統必須優先優化 Recall。

最佳實踐檢查清單 (Best Practices Checklist)

Do's (正確做法 ✅)

- ✅ 使用 Cross-Validation 評估穩健性
- ✅ 針對業務目標選擇正確指標 (F1, RMSE...)
- ✅ 保留完全獨立的測試集 (Test Set)

WARNING: Data Leakage (數據洩漏)

嚴禁在特徵工程或調參時 “偷看” 測試集數據。

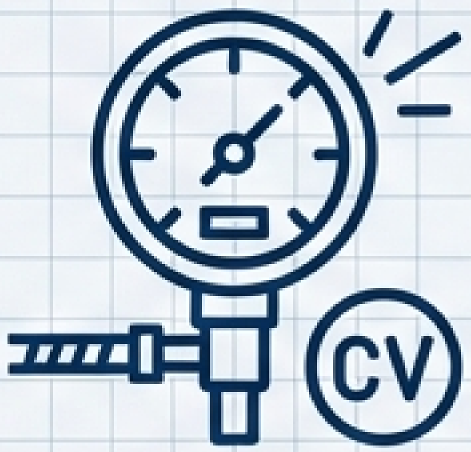
Don'ts (錯誤做法 ❌)

- ❌ 忽略時間序列的順序性
- ❌ 僅憑單一指標 (如 Accuracy) 做決策

總結與展望 (Conclusion & The Road Ahead)

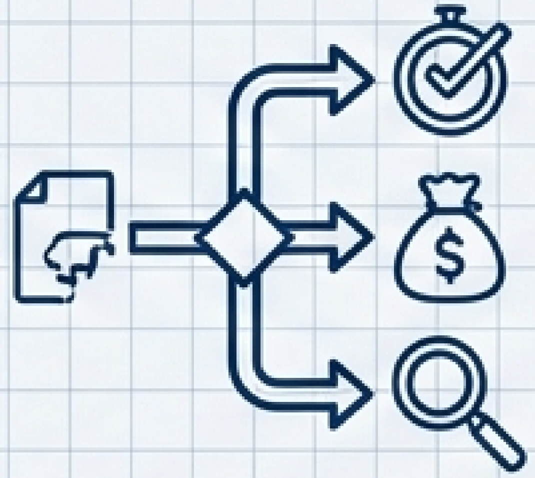
Metric (指標)

根據工程目標選擇儀表 (MAE vs Recall)。



Validate (驗證)

使用 CV 進行壓力測試與診斷 (Bias-Variance)。



Decide (決策)

綜合考量準確度、成本與解釋性。

→ Next Step: Unit 15 - 深度學習模型 (Deep Learning)

優秀的工程師不只會訓練模型，更懂得如何評估與信任模型。