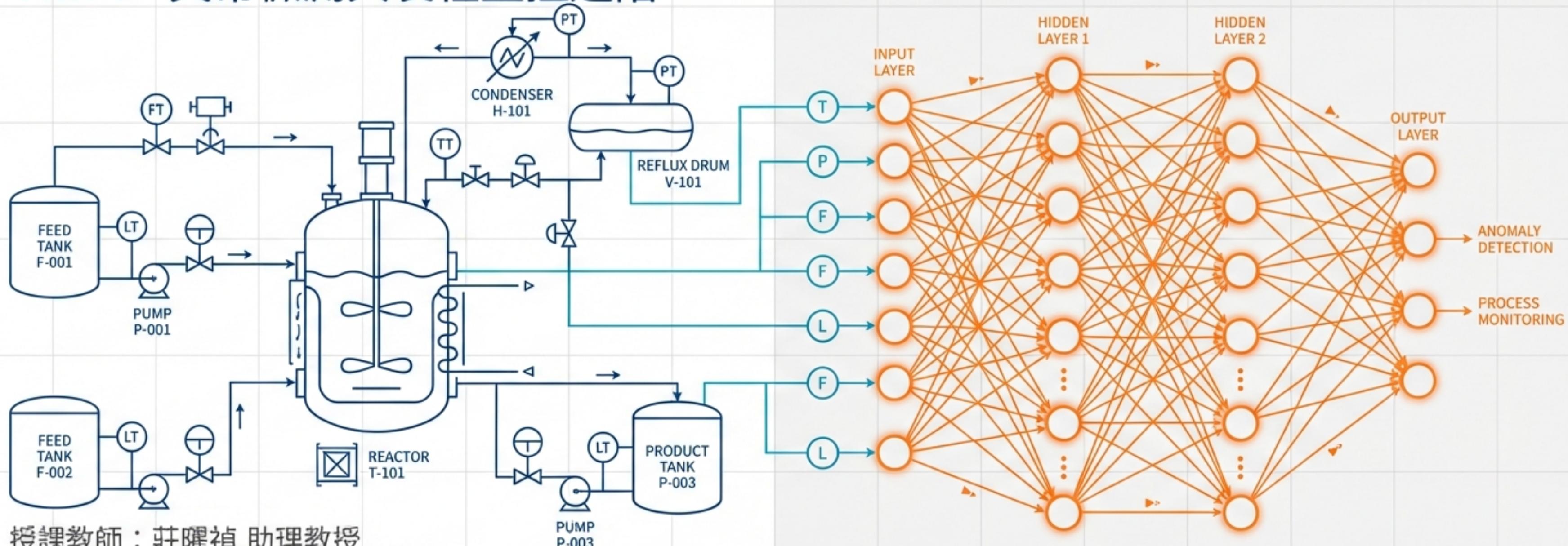


# 區域性離群因子：從原理到化工實踐

Local Outlier Factor: From Principles to ChemE Practice

Unit 07 異常檢測與製程監控進階



授課教師：莊曜禎 助理教授

實驗室：智慧程序系統工程實驗室

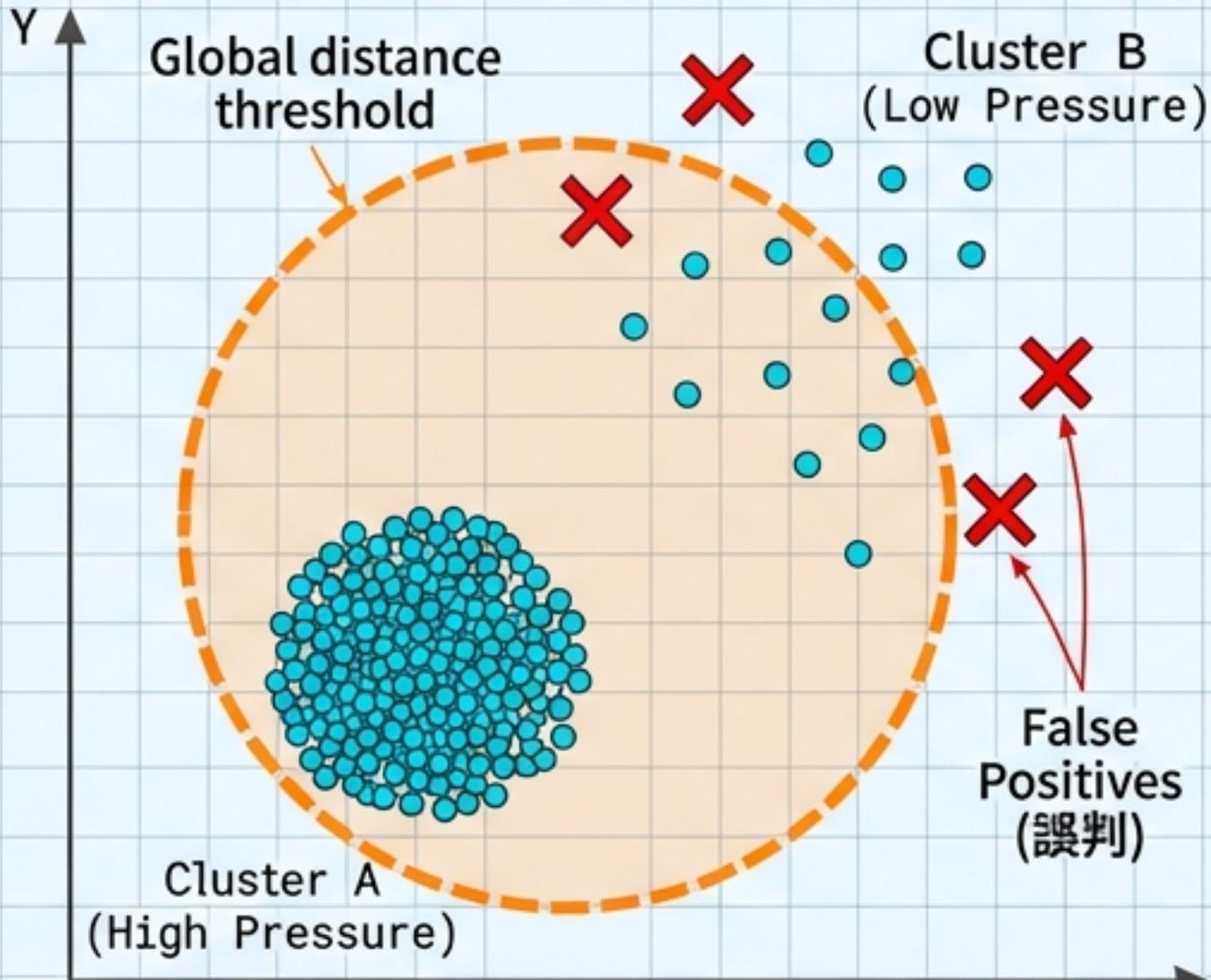
學期：114學年度第2學期

更新日期：2026-01-28

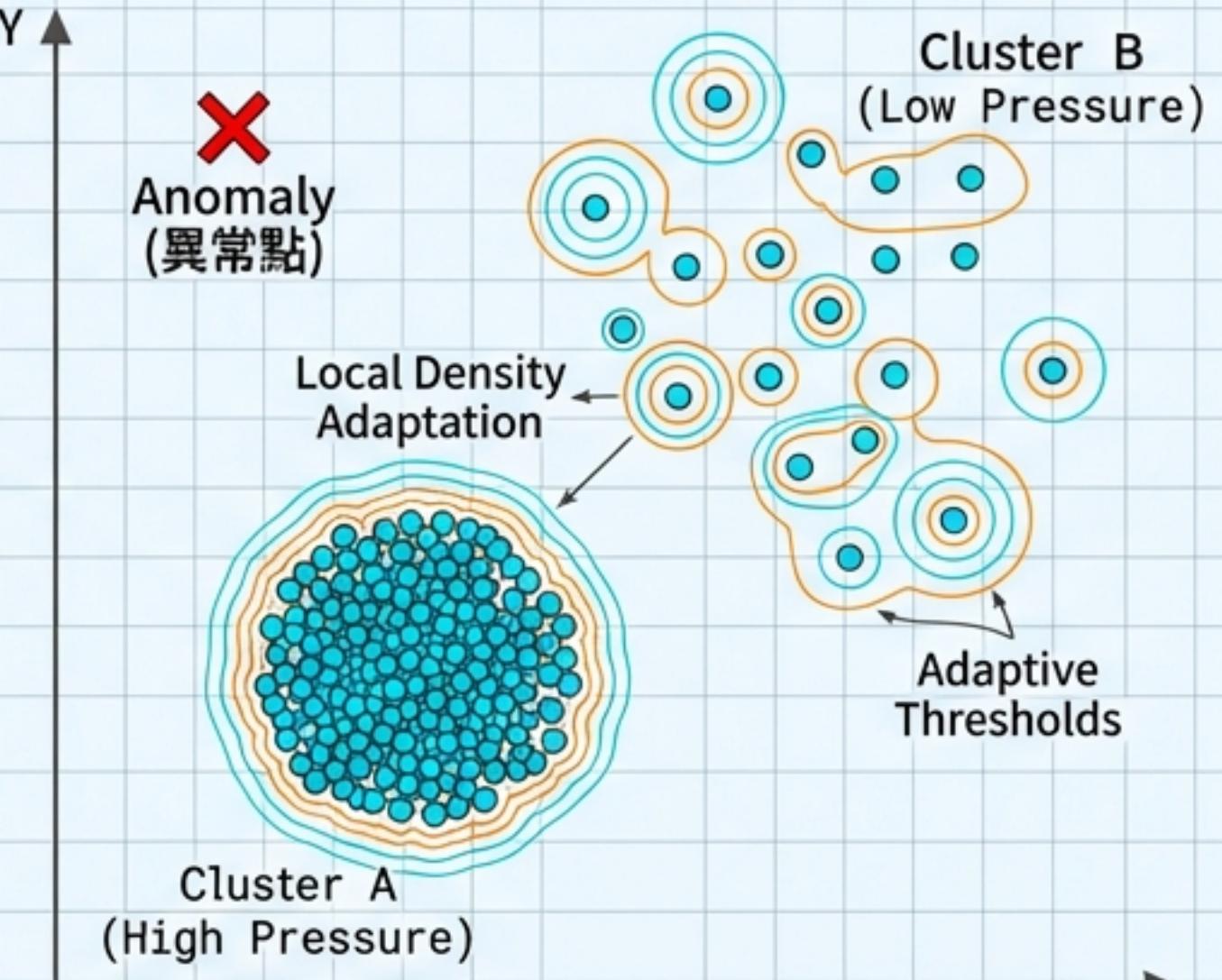
# Paradigm Shift

Roboto Mono

## 全域密度假設失效 (Global Assumption Failure)



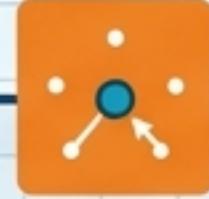
## 區域密度比較 (Local Density Approach)



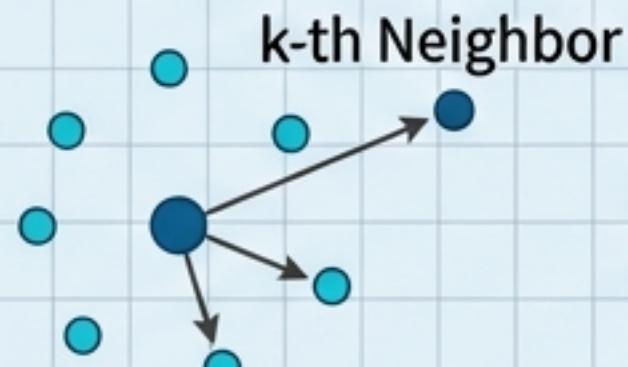
痛點 (Pain Point): 化工製程常有多種操作模式（如：不同配方 A/B/C），導致數據密度不均。全域方法容易誤判稀疏區域的正常點。

核心假設：異常點的局部密度明顯低於其鄰近點 (Local Density of outlier << Neighbors)。

# 演算法解構：LOF 的四大數學支柱 (The 4 Pillars of LOF)



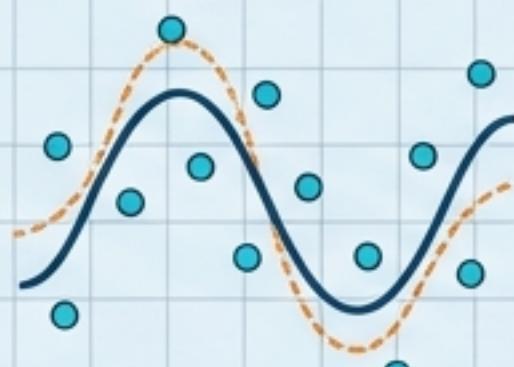
## 1. k-Distance (k-距離)



定義局部範圍。  
衡量  $x_i$  到第  $k$  個鄰居的距離。



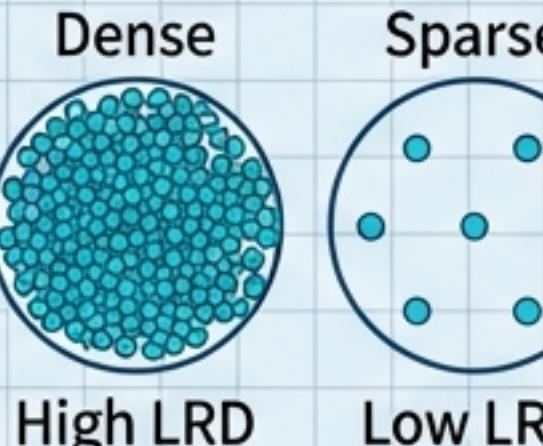
## 2. Reachability Distance (可達距離)



平滑化噪音，  
增加穩定性。  
Formula:  
 $\max\{k\text{-dist}, \text{dist}\}$



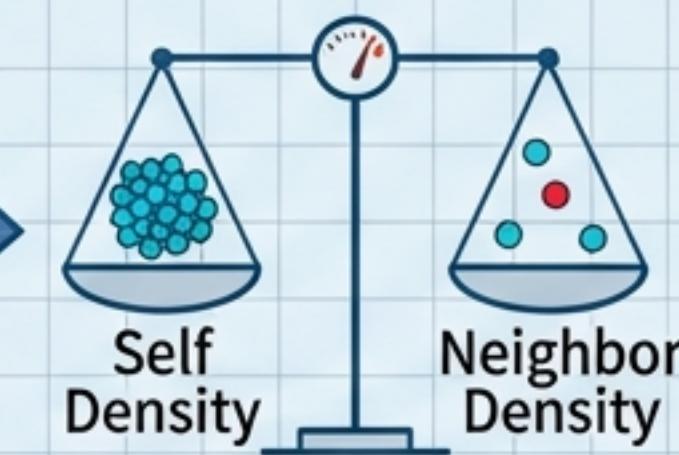
## 3. LRD (局部可達密度)



密度的倒數。  
數值越高，代表周圍越「擠」。



## 4. LOF Score (局部離群因子)



本身密度 vs.  
鄰居密度的比值。

核心概念：LOF 是一個「相對密度」的比值，而非絕對距離。

# 數據解讀：如何看懂 LOF 分數？(Interpreting the Score)



密度高於鄰居  
(罕見)

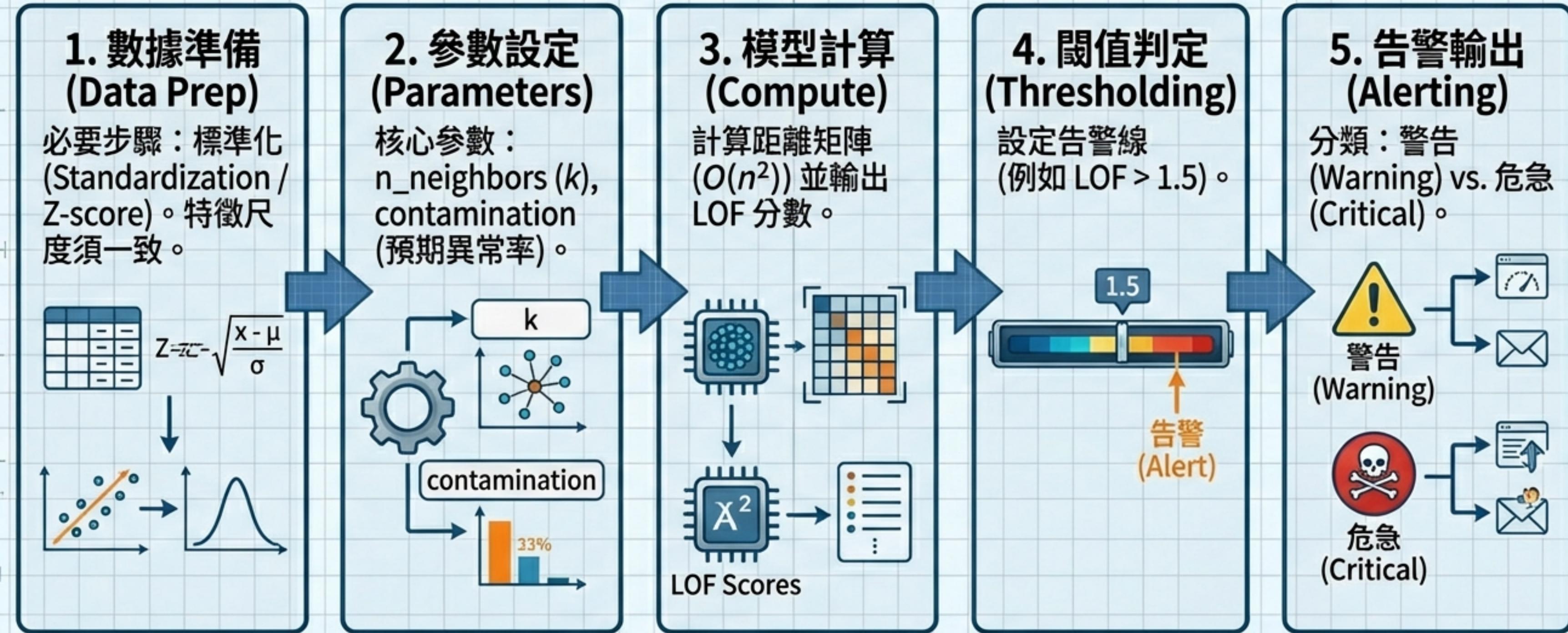
密度與鄰居  
一致  $\approx 1$

密度略低於鄰居  
(感測器漂移)

密度顯著低於鄰居  
(設備故障)

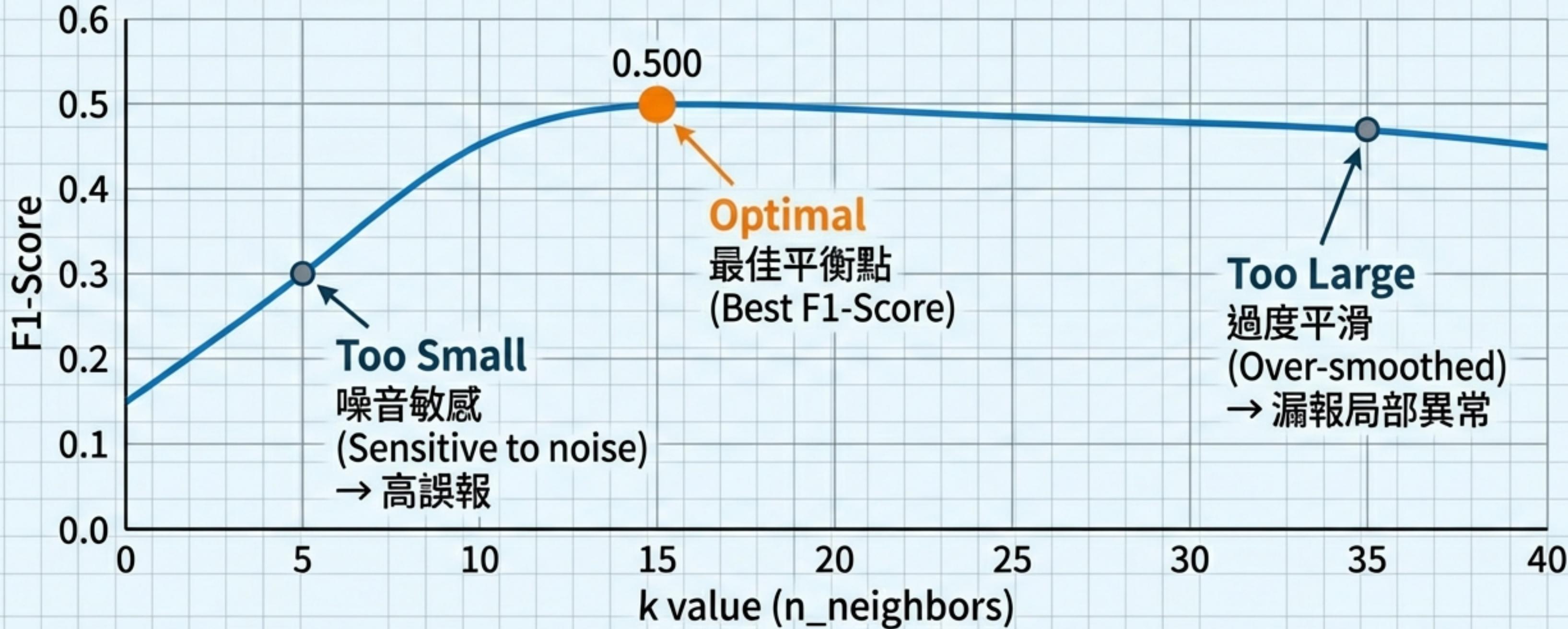
⚠ 注意：LOF 分數是相對值，提供異常程度的量化指標，不僅是二元分類。

# 實作流程：從數據到告警 (Implementation Workflow)



# 關鍵參數優化：尋找最佳 $k$ 值 (Optimizing n\_neighbors)

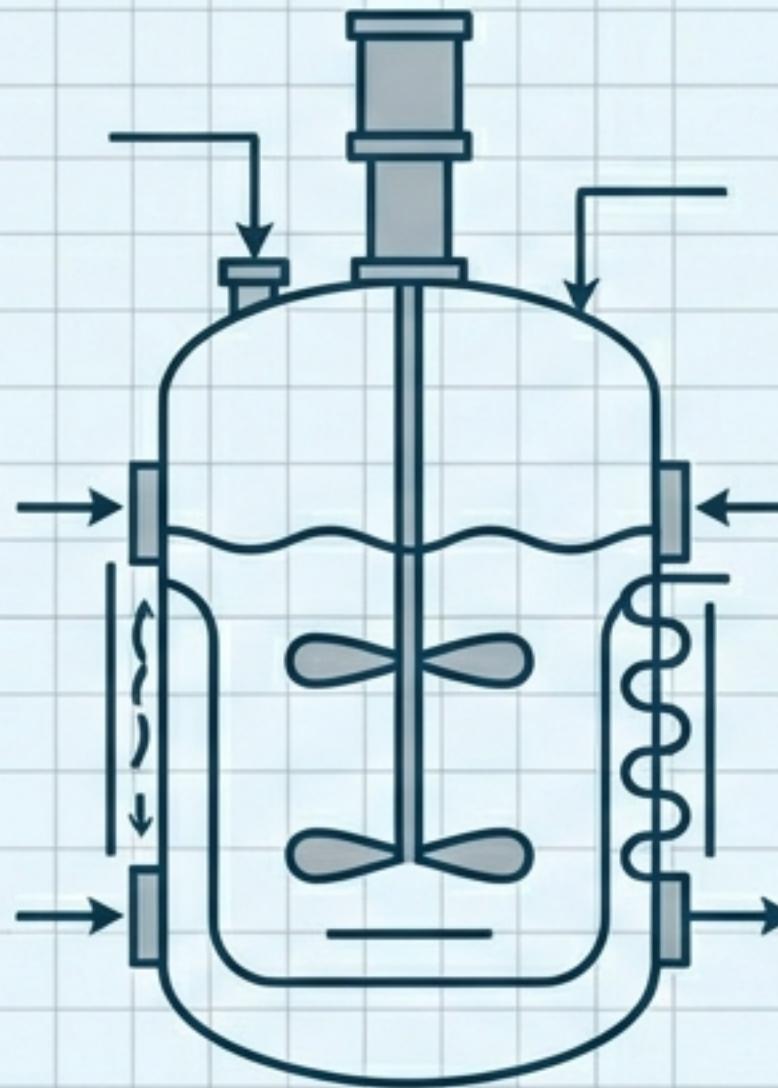
NotebookLM



建議：中型數據集 (300-500 筆) 建議  $k$  範圍為 10-20。

# 實驗案例：批次反應器監控 (Case Study: Batch Reactor)

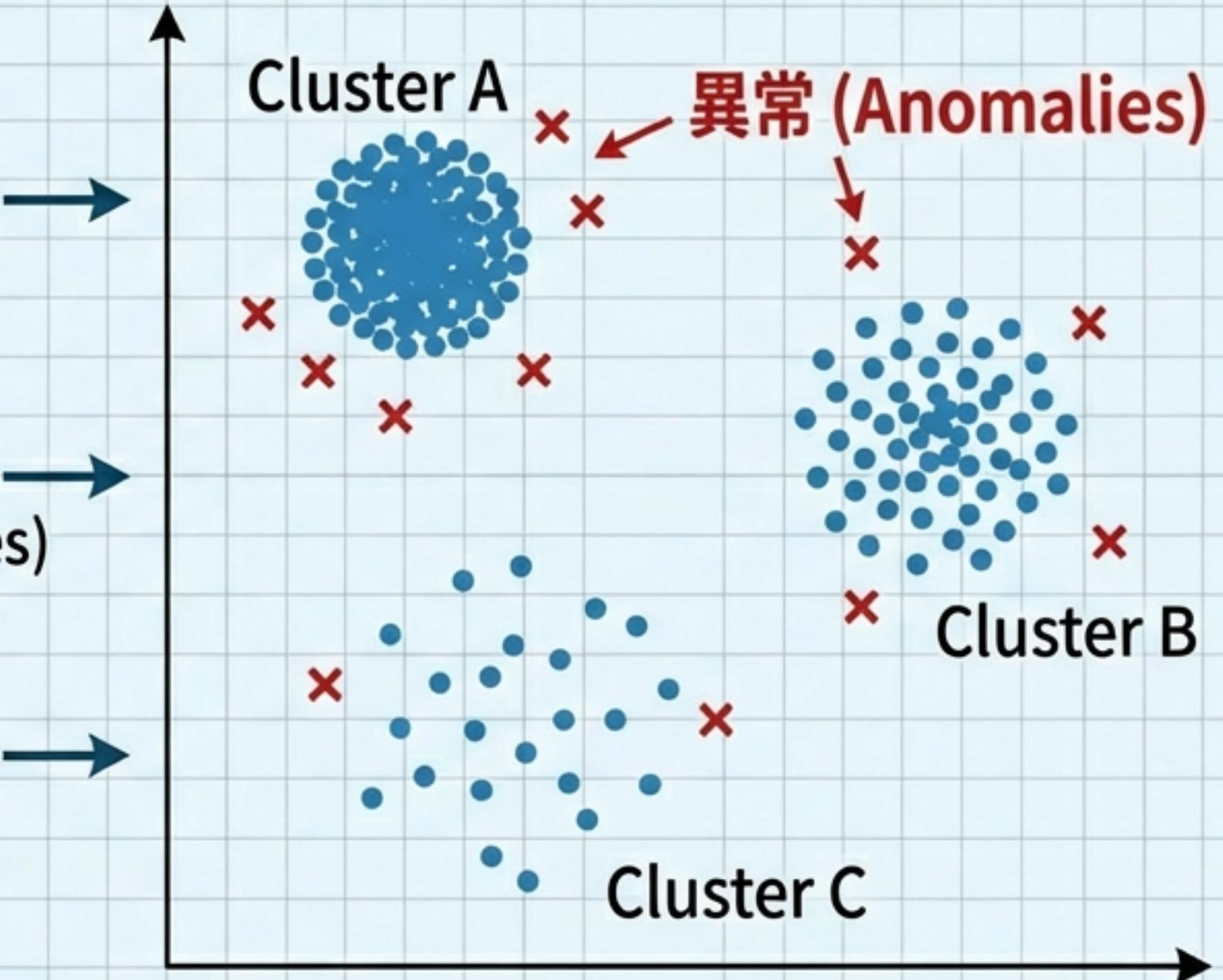
NotebookLM



**Recipe A: 高溫/高壓**  
(Dense Data, 150 samples)

**Recipe B: 中溫/中壓**  
(Medium Data, 150 samples)

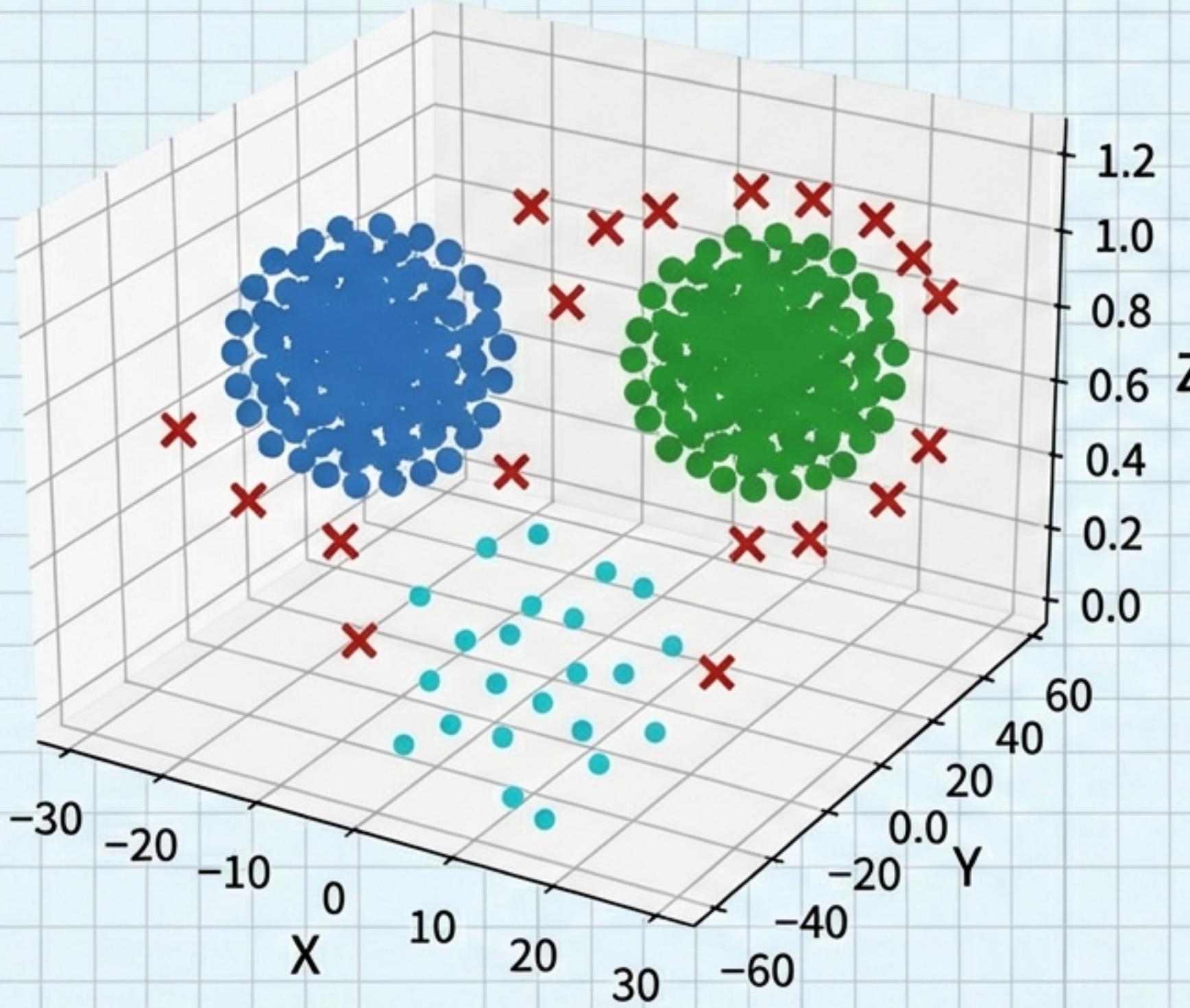
**Recipe C: 低溫/低壓**  
(Sparse Data, 60 samples)



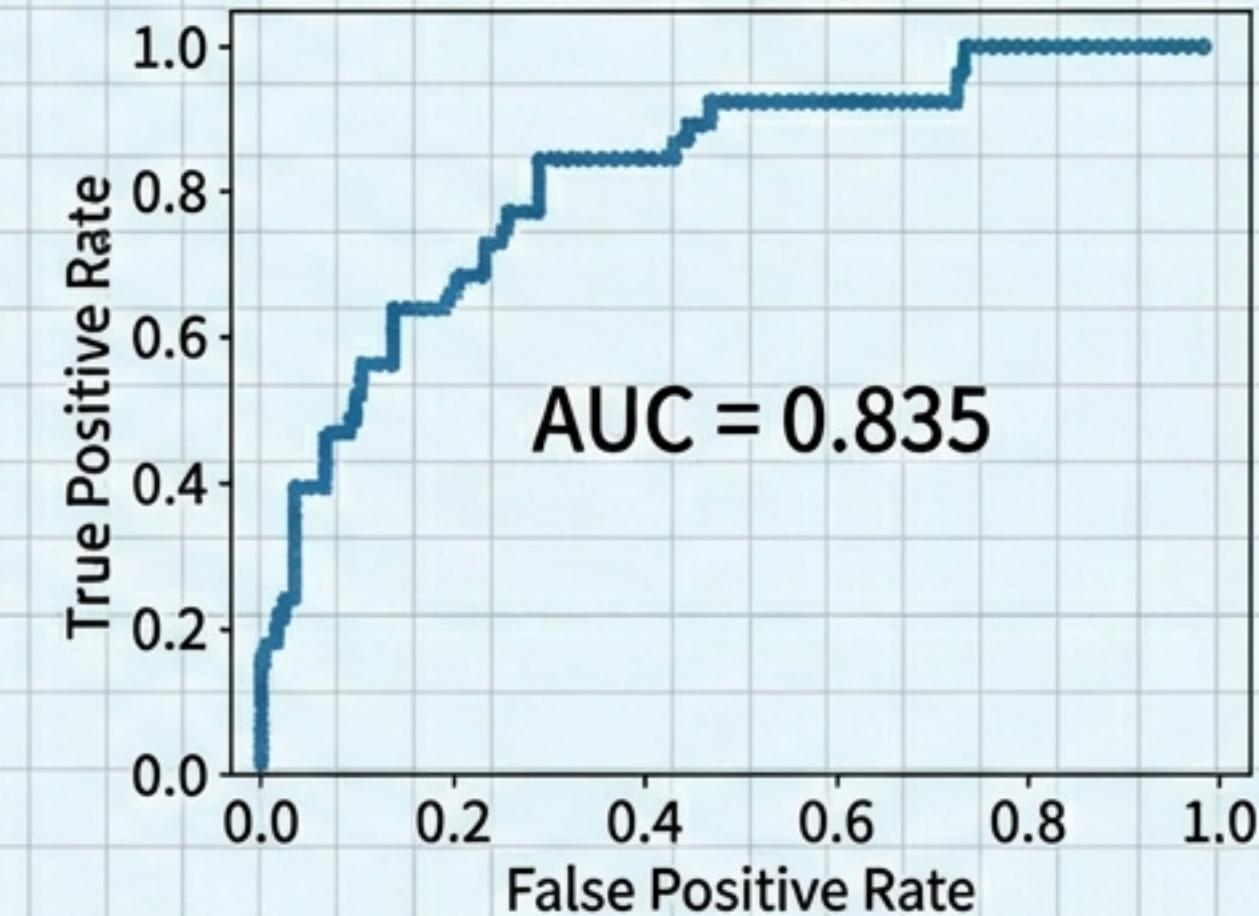
**目標：在多模式操作下，準確抓出異常，且不將稀疏的 Recipe C 誤判為異常。**

# 實驗結果可視化 (Visualization & Metrics)

3D Feature Space



ROC curve



Roboto Mono

True Positive (TP): 15  
False Positive (FP): 15  
Precision: 0.500  
Recall: 0.500

# 模式選擇：內部檢測 vs. 新穎性檢測

Outlier Detection vs. Novelty Detection

內部檢測 (Internal Detection)  
`novelty=False`



- 訓練數據包含異常 (Dirty Data)
- Use Case: 歷史數據事後分析 (Root Cause Analysis)

F1 Score = 0.500

新穎性檢測 (Novelty Detection)  
`novelty=True`



- 訓練數據僅含正常樣本 (Pure Data)
- Use Case: 線上即時監控 (Real-time Monitoring)

F1 Score = 1.000 (Perfect)

結論：工廠產線監控應優先採用 `novelty=True` 模式。

# 算法對決：LOF vs. Isolation Forest

## LOF

(Local Outlier Factor)

機制：  
局部密度

(Local Density)  
Roboto Mono

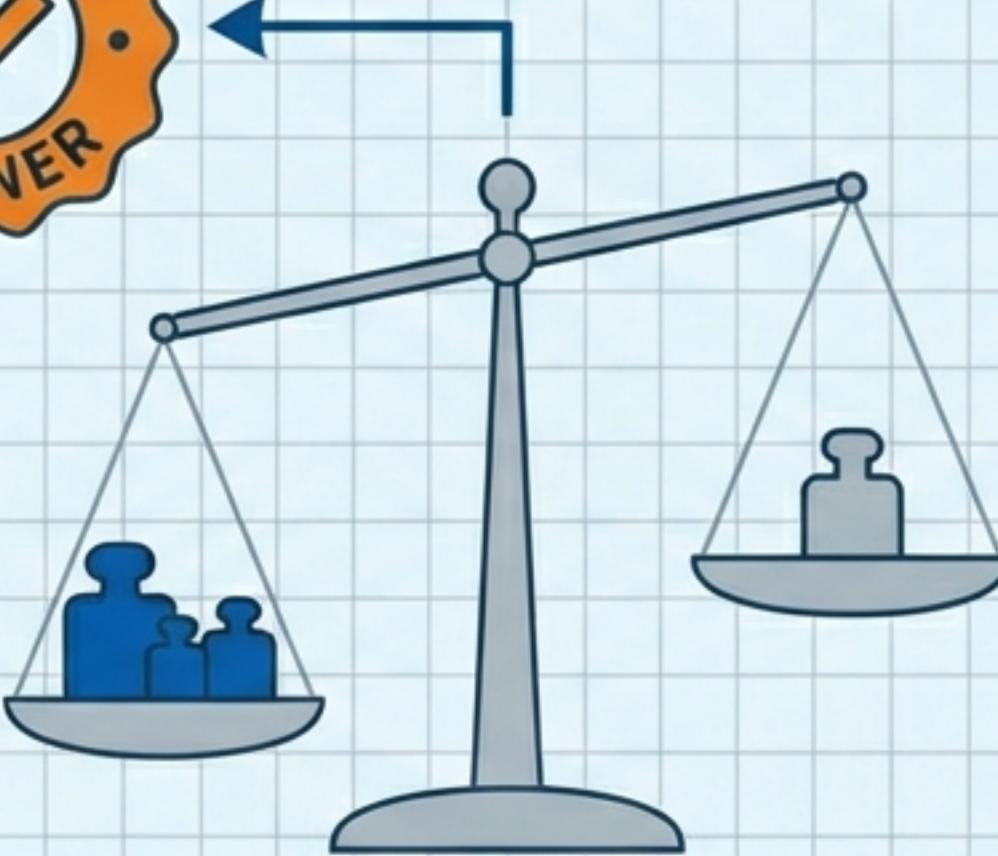


F1 Score

50%

0.500

完美適應 Recipe C 的稀疏  
特性，不誤判。



## Isolation Forest

機制：  
全域隔離

(Global Isolation)  
Roboto Mono



F1 Score

40%

0.400

 容易誤判稀疏區域  
(Recipe C) 為異常。



# 工程評估：優缺點分析 (Pros & Cons)

基於局部離群因子 (Local Outlier Factor, LOF) 的評估

## Pros (優勢)



**局部適應性:** 完美處理多配方/多密度數據。



**量化指標:** LOF 分數直接反映異常程度。



**無分布假設:** 不需要數據符合高斯分布。

## Cons (限制)



**計算成本高:** 時間複雜度  $O(n^2)$ 。  
不適合超大數據 ( $>10k$ )。



**維度災難:** 高維度數據需先做 PCA 降維。



**記憶體需求:** 需儲存距離矩陣。

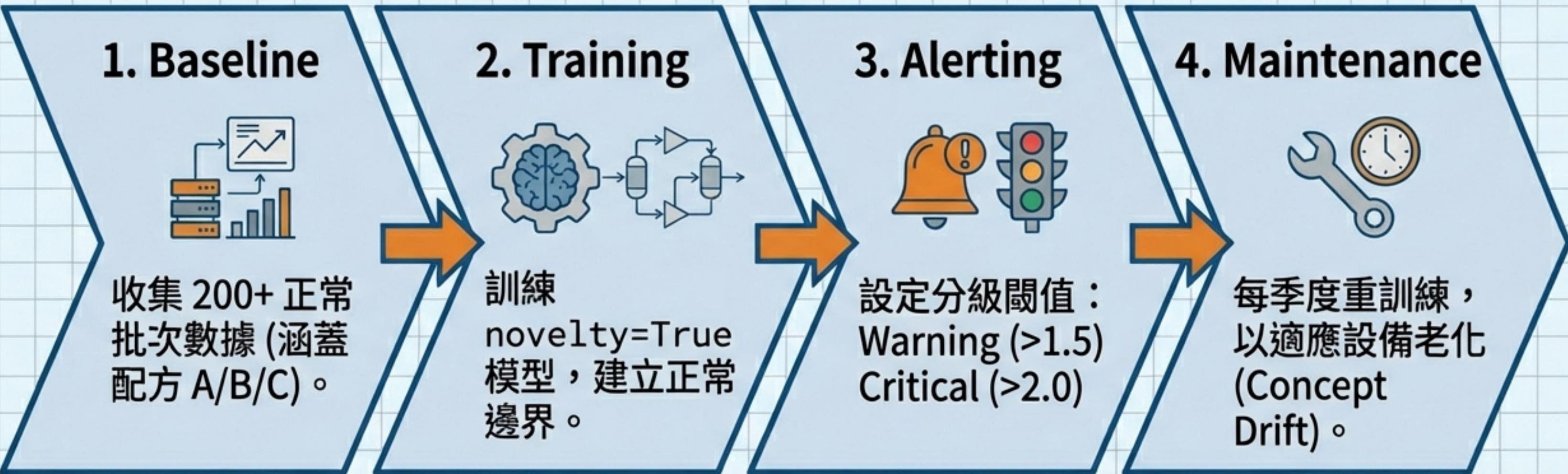
# Python 實作範例 (Implementation with Scikit-Learn)

```
from sklearn.neighbors import LocalOutlierFactor  
from sklearn.preprocessing import StandardScaler  
  
# 1. 關鍵步驟：標準化 (Standardization)  
X_scaled = StandardScaler().fit_transform(X_raw)  
  
# 2. 初始化模型 (開啟新穎性檢測模式)  
lof = LocalOutlierFactor(n_neighbors=20,  
                         contamination=0.05,  
                         novelty=True)  
  
# 3. 訓練與監控  
lof.fit(X_normal_batches)          # 僅用正常數據訓練  
y_pred = lof.predict(X_new)        # 檢測新批次  
scores = -lof.score_samples(X_new) # 取得異常分數
```

重要：必須使用標準化，  
避免距離計算失真。



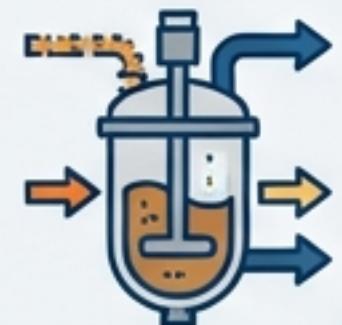
# 工業部署藍圖 (Deployment Roadmap)



# 課程總結 (Key Takeaways)

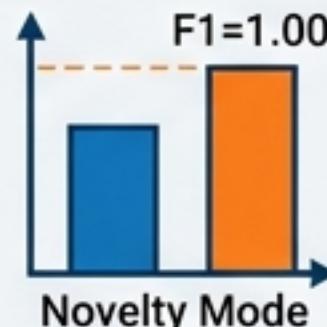
## 解決密度難題

LOF 是多配方、多密度製程的最佳異常檢測工具。



## 模式決定效能

實驗證明：在新穎性檢測 (Novelty) 模式下，可達  $F1=1.00$ 。



## 工具選擇指南

小數據/多模式 →

→ 選 LOF。

大數據/高維度 →

→ 選 Isolation Forest。

**Impact:** 精準定位局部異常，提升化工製程安全性。

# 下一步 (Next Steps)

Next Unit:

Unit 08 - 孤立森林 (Isolation Forest) 與自動編碼器



## Call to Action

- 下載 `Unit07\_LOF.ipynb` 進行實作練習。
- 嘗試調整 `n\_neighbors` 觀察 ROC 曲線變化。

AI 不會取代化工工程師，但懂得使用 LOF 的工程師將取代不懂的人。