

# 氣象格點溫度資料分析報告

## 摘要

本報告以資料集 O-A0038-003.xml (氣象資料觀測平台之"溫度分布-小時觀測分析格點資料")為範例，完成以下任務：

1. 將原始 XML 內容轉換為兩個監督式學習資料集：分類（是否為有效值）與回歸（有效格點溫度）。
2. 建立並訓練簡單機器學習模型（Random Forest）分別完成分類與回歸任務。
3. 評估模型性能並以圖表輔助說明。

---

## 1. 資料說明

- 原始資料：O-A0038-003.xml。
- 格點說明：
  - 經向解析度： $0.03^{\circ}$ ，共 67 個格點(東向遞增)。
  - 緯向解析度： $0.03^{\circ}$ ，共 120 個格點(北向遞增)。
  - 左下角座標：經度 $120.00^{\circ}$ ，緯度 $21.88^{\circ}$ 。
  - 網格大小： $67 \times 120$ (總共 8040 個格點)。
  - 無效值：-999 表示。
- 溫度單位：攝氏度( $^{\circ}\text{C}$ )。

## 2. 資料前處理

### 1. 讀取 XML 並抓住 `< Content >` 內容

- ◆ 使用正規表達式擷取科學記號格式的數值，轉換為浮點數。
- ◆ 重新整理成  $120 \times 67$  的二維陣列。

### 2. 生成經緯度座標

- ◆ 經度： $120.00 \rightarrow 122.00$
- ◆ 緯度： $21.88 \rightarrow 25.47$
- ◆ 建立 (lon, lat) 格點網格。

### 3. 生成兩個監督式資料集

- ◆ 分類資料集：格式 (lon, lat, label)，label = 0 (無效值 -999)，label = 1 (有效值)。
- ◆ 回歸資料集：格式 (lon, lat, value)，僅保留有效值。

## 3. 模型與訓練

- 分類模型：Random Forest Classifier( `n_estimators=100` )
  - 特徵：經度、緯度
  - 標籤：0 或 1 (無效/有效)
- 回歸模型：Random Forest Regressor( `n_estimators=100` )
  - 特徵：經度、緯度
  - 目標：有效溫度值(°C)
- 訓練測試比：80% / 20%

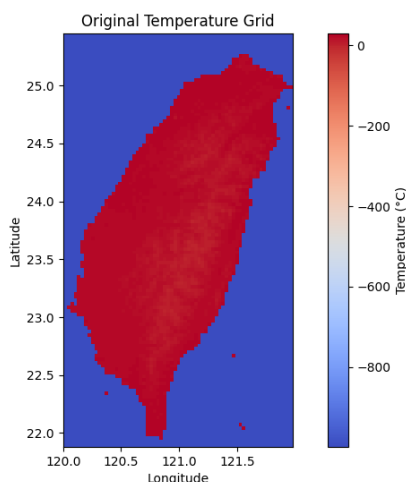
#### 4. 結果與圖表

##### 數值結果

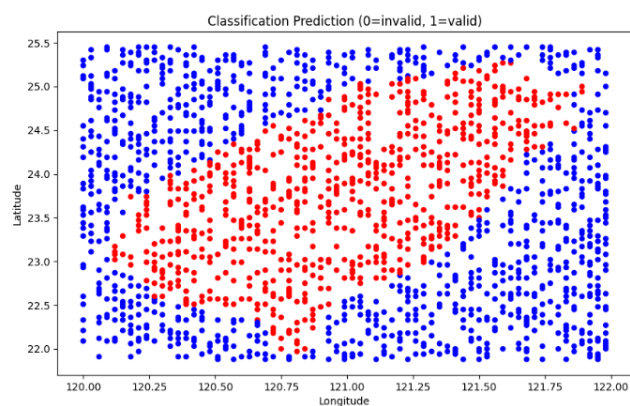
- 分類模型：
  - ◆ Accuracy = 0.9845
- 回歸模型：
  - ◆ RMSE = 2.20 °C

##### 圖表說明

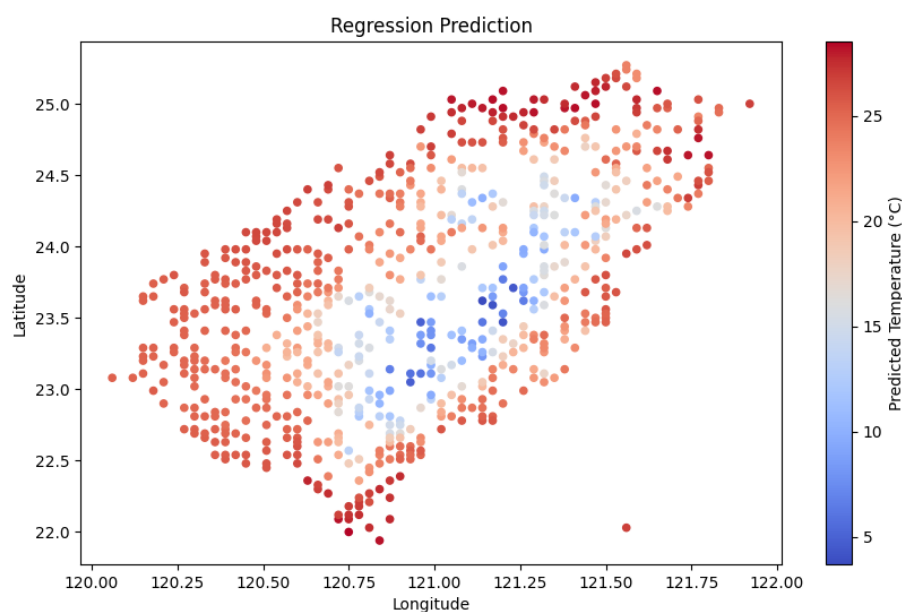
- 圖一 – 原始溫度熱圖( `temperature_heatmap.png` )
  - ◆ 顯示原始 67 × 120 網格的溫度分布。
  - ◆ 可以直觀觀察有效/無效值區域。
- 圖二 – 分類結果散點圖( `classification_cm.png` )
  - ◆ 測試資料集的預測結果（紅/藍表示無效/有效）。
  - ◆ 準確率高，幾乎完全正確辨識無效與有效格點。
- 圖三 – 回歸預測散點圖( `regression_scatter.png` )
  - ◆ 測試資料集的預測溫度（以顏色表示）。
  - ◆ 分布趨勢與原始溫度場接近，但在細部區域存在  $\pm 2^{\circ}\text{C}$  誤差。



圖一



圖二



圖三

## 5. 討論

### ● 分類模型

- ◆ 準確率高達 **98.45%**，代表僅憑地理位置即可有效辨識無效與有效格點。
- ◆ 無效值在空間上呈現一定規律性，因此 Random Forest 容易學習。

### ● 回歸模型

- ◆ RMSE 為 **2.20 °C**，顯示模型能夠大致捕捉溫度分布。
- ◆ 誤差可能來自於：
  1. 僅使用 (lon, lat) 當特徵，未包含鄰近格點溫度或地形因子。
  2. Random Forest 雖能處理非線性，但對於連續場的細微變化表現有限。

## 6. 結論

- 本研究展示了如何從原始 XML 格點氣象資料建立監督式學習資料集。
- Random Forest 分類模型能準確預測格點有效性，回歸模型能在  $\pm 2^{\circ}\text{C}$  誤差範圍內估計溫度分布
- 未來若能結合更多空間與物理特徵，預測表現可望進一步提升。