

## 一、研究目的

利用中央氣象署所提供的「小時溫度觀測分析格點資料 (O-A0038-003)」，進行資料前處理與機器學習模型建構。透過轉換原始格點溫度資料，分別建立 **分類資料集** 與 **回歸資料集**，並利用此兩類資料集訓練對應的機器學習模型，以驗證分類與回歸在氣象數據上的應用可能性。

---

## 二、資料處理流程

### 1. 資料來源

- XML 檔案包含 **67 × 120** 的溫度格點矩陣，每個格點對應到一個經緯度位置。
- 左下角座標為 (120.00, 21.88)，經緯度間距皆為 **0.03** 度。
- 格點值若為 -999 表示無效，否則為有效溫度值 (單位：攝氏)。

### 2. 分類資料集 (classification\_data.csv)

- 欄位格式：(Longitude, Latitude, Label)
- 若原始值 = -999，則 Label=0；否則 Label=1。
- 資料完整保留所有格點，共約 **8040** 筆。

### 3. 回歸資料集 (regression\_data.csv)

- 欄位格式：(Longitude, Latitude, Temperature)
  - 僅保留有效溫度值，無效格點全部剔除。
  - 資料數量比分類資料少，僅含真實觀測溫度。
- 

## 三、模型建構

### 1. 分類模型 (Decision Tree Classifier)

- 輸入特徵：經度、緯度
- 輸出標籤：是否為有效格點 (0/1)

- 訓練方式：8:2 分割訓練/測試集

## 2. 回歸模型 (Random Forest Regressor)

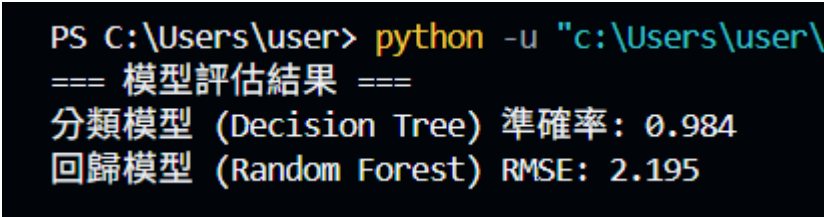
- 輸入特徵：經度、緯度
- 輸出數值：對應格點的溫度 (°C)
- 訓練方式：8:2 分割訓練/測試集，隨機森林樹數量設定為 100。

---

## 四、實驗結果

根據 train\_results.txt，我們得到以下指標：

- 分類模型準確率 (Accuracy)：0.984
- 回歸模型誤差 (RMSE)：2.195 °C



```
PS C:\Users\user> python -u "c:\Users\user\  
=== 模型評估結果 ===  
分類模型 (Decision Tree) 準確率: 0.984  
回歸模型 (Random Forest) RMSE: 2.195
```

---

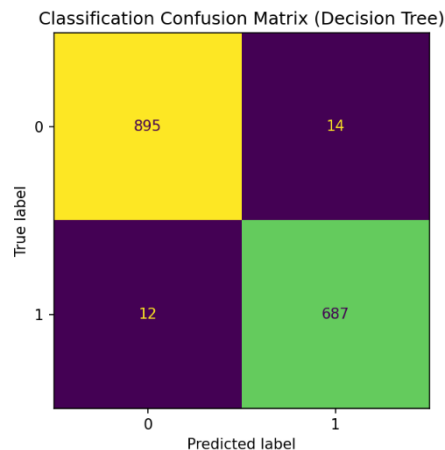
## 五、結果分析與圖表說明

### (1) 分類模型：混淆矩陣

混淆矩陣結果顯示，絕大部分無效格點與有效格點皆被正確分類。

- True Negative：895
- True Positive：687
- False Positive：14
- False Negative：12

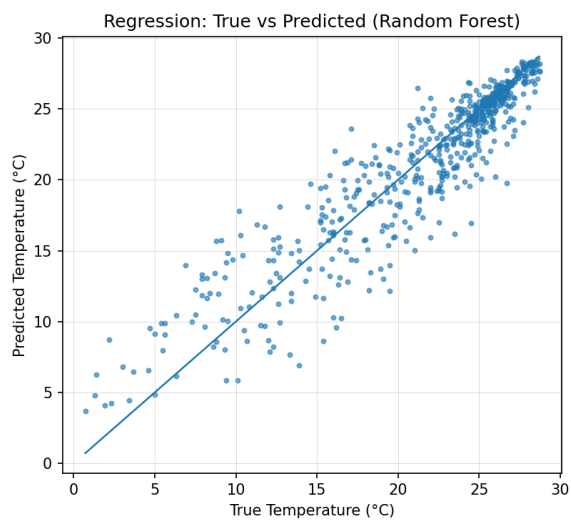
表示模型能夠精準判斷格點是否含有效資料，錯誤率非常低。



## (2) 回歸模型：真實值 vs 預測值

可以從圖中看到，大部分點落在對角線附近，表示模型預測接近真實資料。然而在部分低溫與高溫範圍，仍存在少量點偏離對角線，代表在極端值的預測能力較弱。

但總得來說，RMSE 僅約 2.195 °C，已能提供合理的溫度預測。

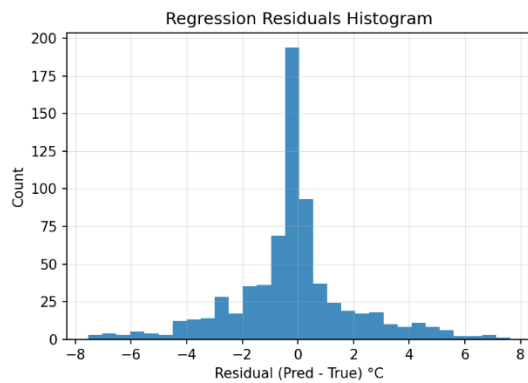


## (3) 回歸模型：殘差直方圖

殘差 (Pred - True) 分布以 0 為中心，近似對稱。

- 大部分誤差集中於  $\pm 2$  °C 之內
- 少數誤差超過  $\pm 5$  °C，屬於異常情況

代表模型預測未出現系統性偏差，僅在局部地區有較大誤差。



---

## 六、結論與建議

### 1. 分類任務

- **Decision Tree** 模型能有效判斷格點是否含有效溫度，準確率高達 98%。
- 結果顯示分類任務藉由「空間位置」判斷有效性，具有一定規律性。

### 2. 回歸任務

- **Random Forest** 模型在經緯度到溫度的非線性映射中平均誤差小於 2.5 °C。
- 在部分極端值仍有偏差。

### 3. 結論

- 成功從原始氣象 XML 格點資料轉換為機器學習可用的監督式資料集，並建立分類與回歸模型。