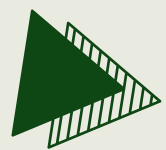


實作研討(3)

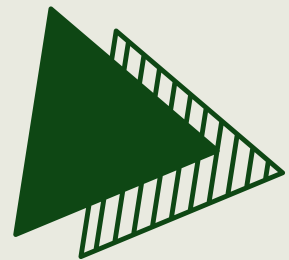


組長:陳鈺昕

組員:賴兆信, 黃子晏, 宋苡瑄, 黃天芸

目錄

- 1 HW_1 (包含交叉驗證，確保模型的穩定)
- 2 HW_2 模型評估 MAE, MSE, RMSE, R2
- 3 HW_3 模型評估 F1 Score, AUC, 混淆矩陣
Precision/ Recall
- 4 HW_4 decoder-only 改為支援繁體中文
- 5 HW_5 中文情感分析



HW_1

- 資料與目標

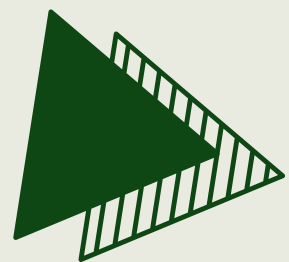
資料來源：UCI Red Wine Quality Dataset

樣本數量：1599 筆

特徵數量：11 項化學特徵（如酒精濃度、酸度等）

目標變數：quality（紅酒品質分數，整數型）

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68	9.8	5
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65	9.8	5
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58	9.8	6
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5
...
1594	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	10.5	5
1595	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	11.2	6
1596	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0	6
1597	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2	5
1598	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66	11.0	6



HW_1

- **模型與參數設定**

使用 決策樹回歸 (DecisionTreeRegressor) 預測紅酒品質。
為避免過擬合，使用以下策略：

- 📌 **參數搜尋範圍：**

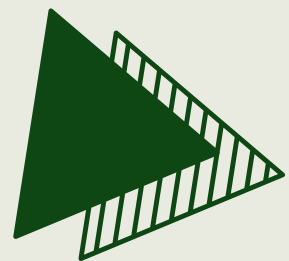
參數 搜尋值

max_depth [2, 3, 4, 5, 6]

min_samples_leaf [1, 3, 5, 10]

- 📌 **使用 GridSearchCV 配合 5 折交叉驗證 (KFold)，找出最佳參數組合。**

```
最佳參數: {'max_depth': 6, 'min_samples_leaf': 10}  
最佳平均 R2 分數: 0.295
```



HW_1

- 評估與結論

平均 R^2 分數 0.30：表示模型解釋約 30% 的品質變異。

R^2 標準差 0.021：模型在各折表現一致，穩定性佳。

模型優點：

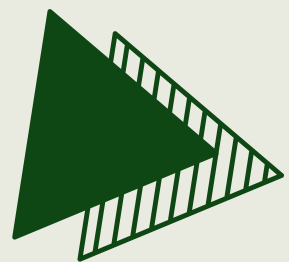
易於解釋，適合小型資料集

有效避免過擬合（透過 `max_depth` 與 `min_samples_leaf` 調整）

```
每折  $R^2$  分數: [0.3, 0.26, 0.29, 0.3, 0.32]
```

```
平均  $R^2$  分數: 0.3
```

```
 $R^2$  分數標準差: 0.021
```

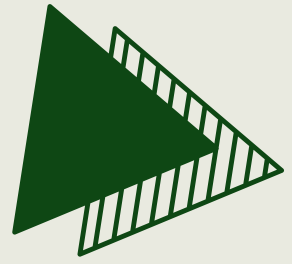


HW_2 資料集介紹

- 使用上課用到的紅酒資料集，1599筆資料、12項欄位
- 資料先經過特徵標準化
- 採用"quality"作為特徵與目標變數
- 切分資料集（80% 訓練，20% 測試）
- 使用訓練模型--決策樹

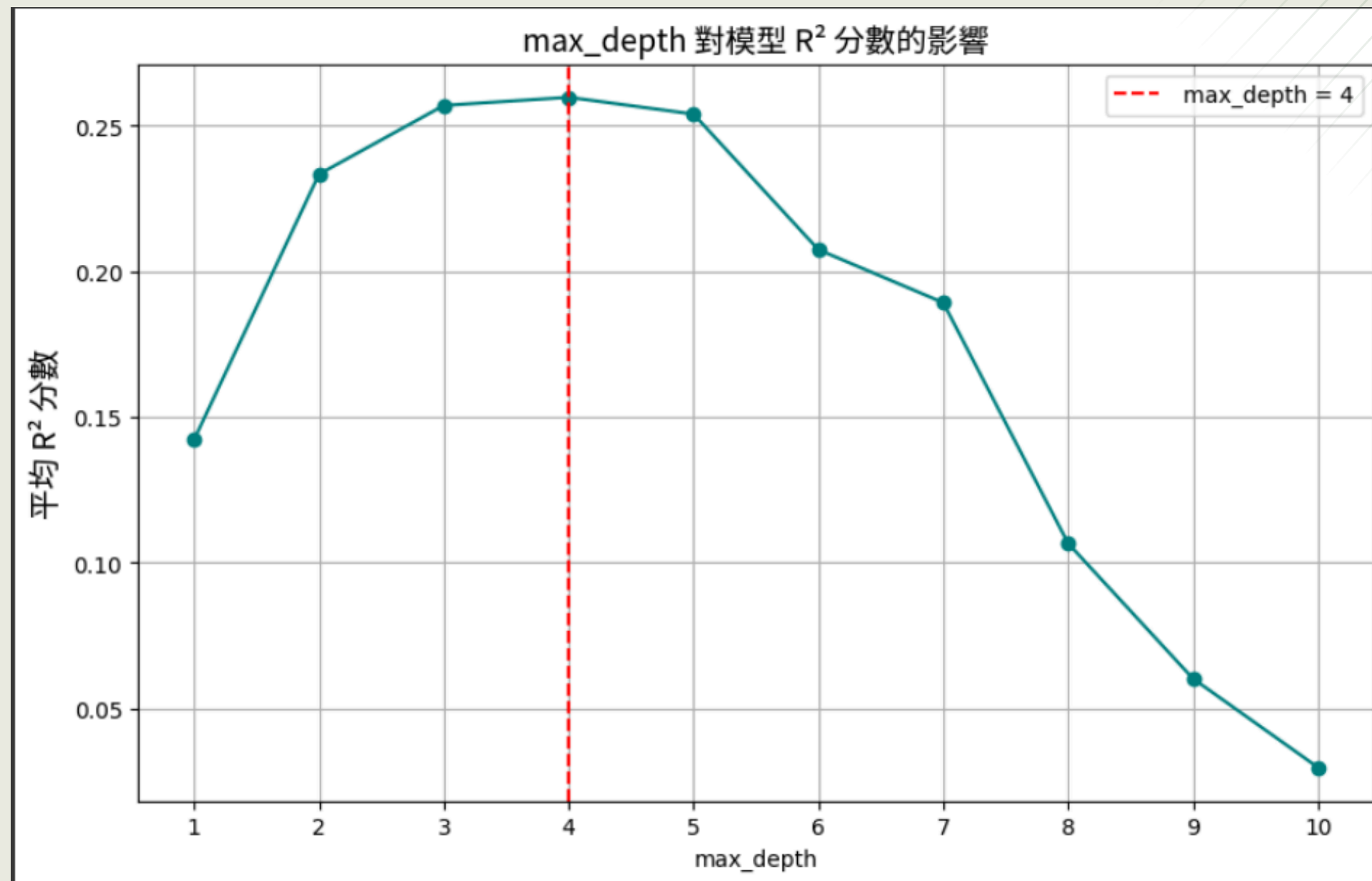
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68	9.8	5
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65	9.8	5
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58	9.8	6
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5
...
1594	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	10.5	5
1595	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	11.2	6
1596	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0	6
1597	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2	5
1598	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66	11.0	6

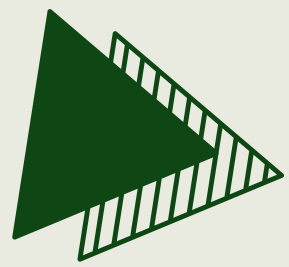
1599 rows × 12 columns



HW_2 判斷max_depth

- 使用GridSearchCV得知最合適的max_depth為4
- 交叉驗證cv=5
- 比較欄位為" R^2 "





HW_2

評估結論

- **MAE(平均誤差的絕對值)**

預測平均誤差約 ± 0.52 分，算是偏小的誤差，代表模型預測相對準確

- **MSE(平均平方誤差)**

表示整體誤差偏小。

因為平方的關係，越小代表越穩定

- **RMSE(均方根誤差)**

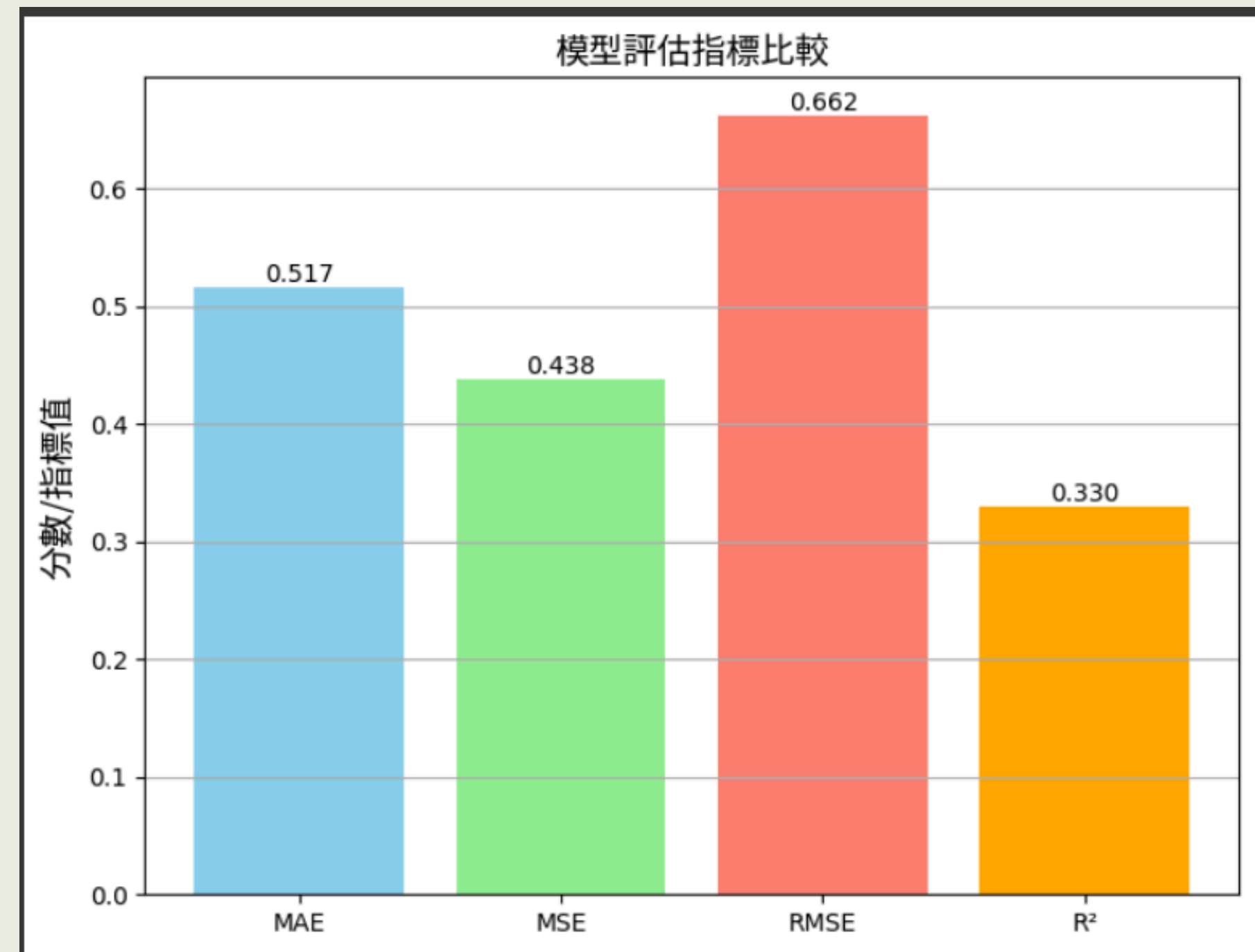
與 MAE 差距不大，

表示模型錯誤分布滿平均，沒太多離群值

- **R^2 (決定係數)**

模型解釋資料變異程度

約 33% 的變異可以被模型解釋，代表模型表現普通



HW3

資料蒐集

鐵達尼號資料集 (簡化版 891 ROWS × 12 COLUMNS)

資料前處理

- 補上 Age 缺失值 (使用平均數)
- 將 Sex 轉為數值 (male=0, female=1)

特徵選擇

使用 Pclass、Sex、Age、SibSp、Parch、Fare 等欄位

建立模型

使用 Decision Tree (決策樹) 分類模型

模型評估

指標包含 F1 Score、AUC、Precision、Recall、混淆矩陣

HW3

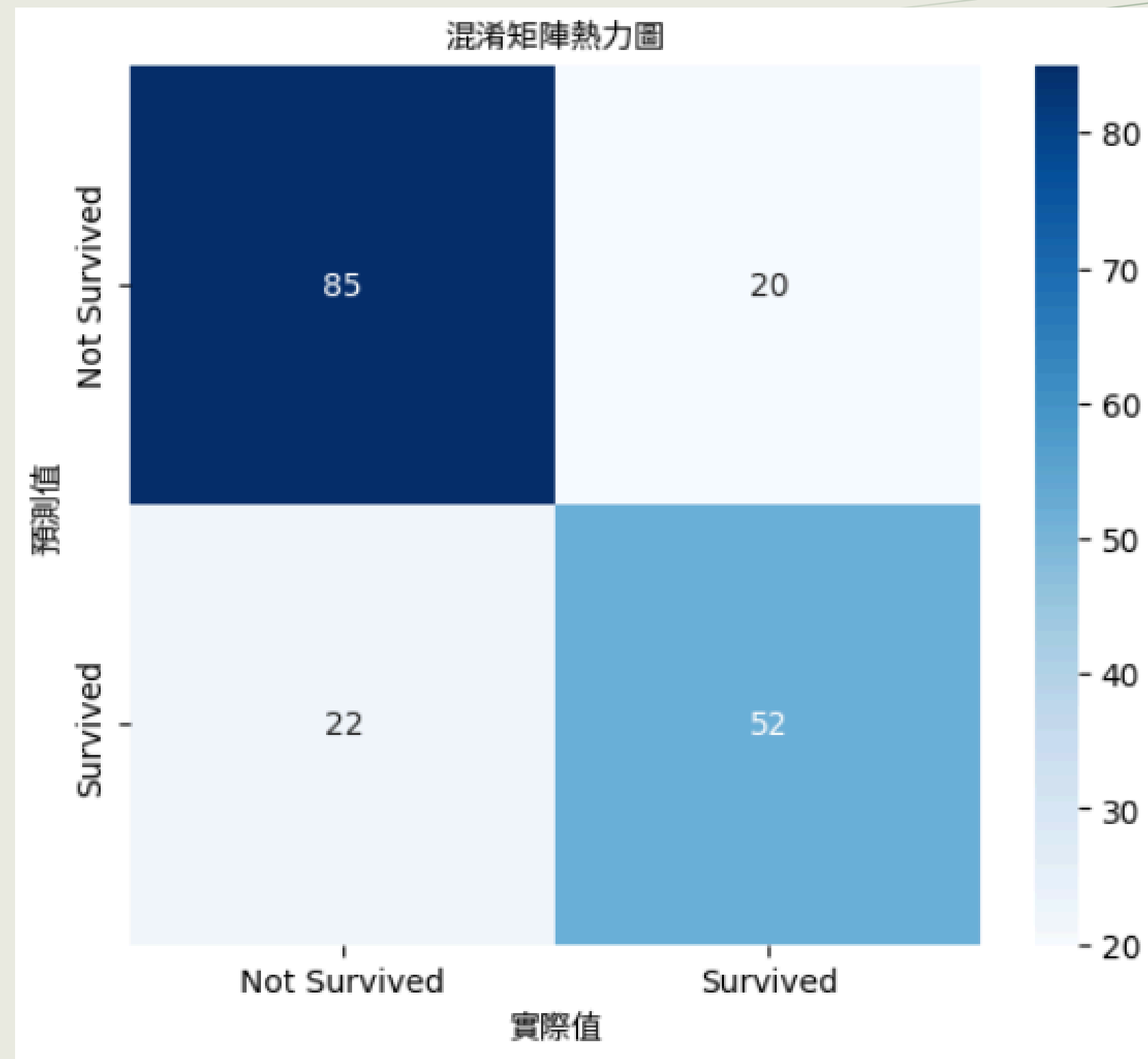
['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']

這 6 欄是比較乾淨、沒有太多缺失值，而且對預測生存有幫助的特徵欄位

欄位名稱	說明	資料型態
PassengerId	乘客編號（1 ~ 891）	整數
Survived	是否生存（0=死亡，1=生存）	二元分類
Pclass	艙等（1=頭等，2=二等，3=三等）	類別整數
Name	姓名（包含頭銜、婚姻等）	文字
Sex	性別（male / female）	類別文字
Age	年齡（可能有缺值）	浮點數
SibSp	同行兄弟姊妹 / 配偶人數	整數
Parch	同行父母 / 子女人數	整數
Ticket	船票號碼	文字
Fare	票價	浮點數
Cabin	客艙編號（大量缺值）	文字 / 缺值
Embarked	登船港口（S=南安普敦，C=瑟堡，Q=皇后鎮）	類別文字

HW3

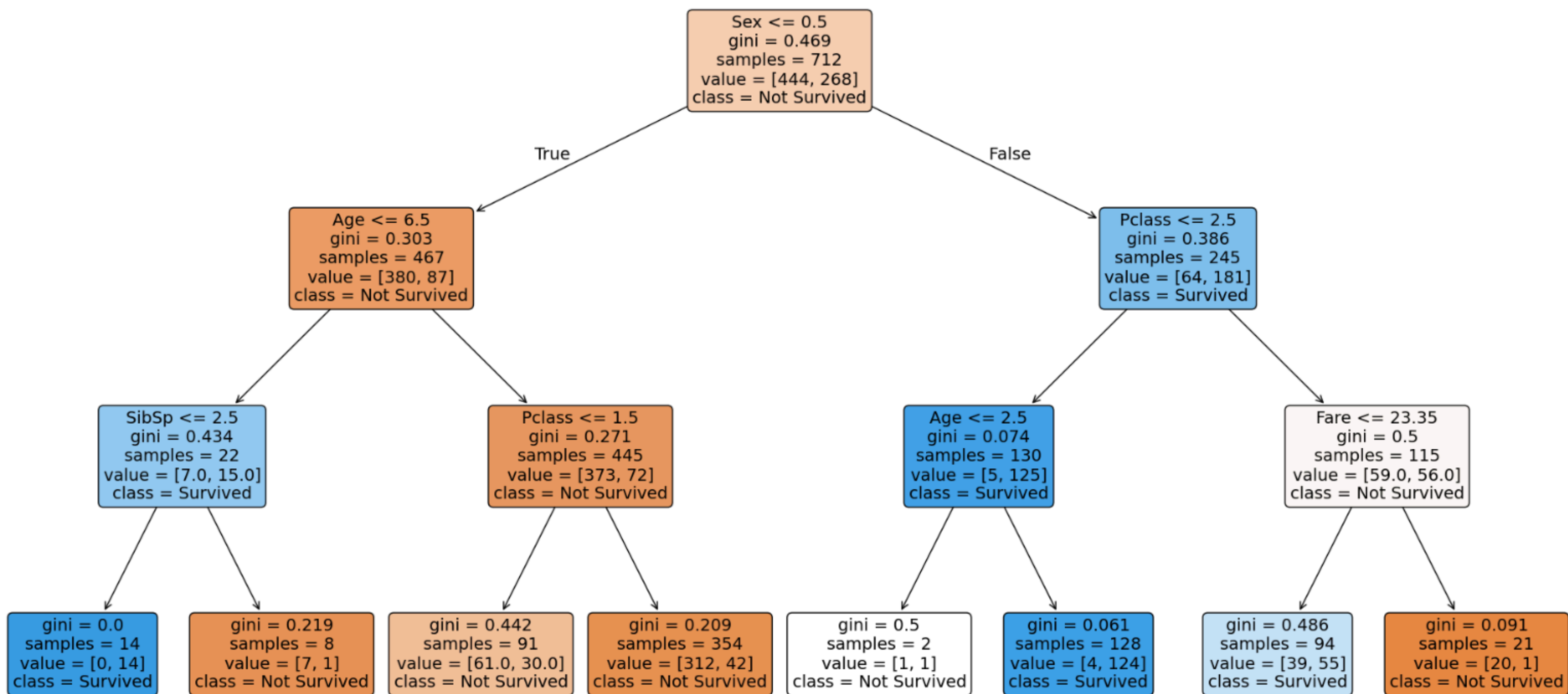
- ✓ F1 Score: 0.7542
- ✓ AUC Score: 0.8261
- ✓ Precision: 0.7714
- ✓ Recall: 0.7368



HW3

決策樹:「性別」與「艙等」對預測結果影響最大

Titanic 生存預測決策樹 (max_depth=3)



HW4

使用者輸入 prompt

加上風格標籤引導模型（如 [風格: 古文]）

幫助模型切換不同語氣與文體

max_new_tokens: 控制生成長度

temperature: 調整隨機性

top_k / top_p: 控制詞彙選擇的多樣性

do_sample=True: 啟用隨機取樣（更自然）

```
# 根據 style 參數取得對應的中文風格標籤，若未指定則預設為「一般」
style_tag = style_labels.get(style, "一般")

# 將風格標籤加入提示詞中，以引導模型產生指定風格的內容
styled_prompt = f"[風格: {style_tag}]\n{prompt}"

# 將 styled_prompt 編碼成模型可接受的 token ID，並轉成 tensor
input_ids = tokenizer.encode(styled_prompt, return_tensors="pt").to(device)

# 建立 attention_mask，避免模型注意到 padding 的位置（這裡預設沒有 padding，但為保險起見仍處理）
attention_mask = (input_ids != tokenizer.pad_token_id).long()

# 使用模型產生文字
output_ids = model.generate(
    input_ids,                                # 輸入的 token
    attention_mask=attention_mask,            # 注意力遮罩，讓模型知道哪些位置是有效輸入
    max_new_tokens=max_new_tokens,           # 生成的最大 token 數量
    temperature=temperature,                 # 溫度值控制隨機性，越高越多樣化
    top_k=top_k,                             # top-k 篩選（僅在前 k 個最可能的詞中選擇）
    top_p=top_p,                             # top-p (nucleus sampling)，保留機率總和為 p 的詞集中選擇
    do_sample=True,                          # 啟用隨機取樣（非貪婪解碼）
    pad_token_id=tokenizer.eos_token_id      # 設定 pad token 為結束符號，以避免警告
)

# 解碼模型輸出的 token ID，轉回可讀文字，並移除多餘空白
output_text = tokenizer.decode(output_ids[0], skip_special_tokens=True).replace(" ", "")
return output_text                          # 回傳最終生成的文字
```

HW5_中文情感分類模型

資料集

採用 tyqiangz/multilingual-sentiments 中的中文資料，僅保留標記為正面（0）與負面（1）的樣本

主要套件：

- transformers：用於載入預訓練模型和分詞器
- datasets：用於處理和載入資料集
- scikit-learn：用於資料分割和評估指標計算
- torch：用於模型訓練和推論

資料處理步驟

- 載入中文情感資料集
- 資料篩選：僅保留正面與負面樣本，排除中性資料
- 類別平衡：為避免模型偏倚，對兩類樣本數進行平衡處理。
- 資料分割：將資料集分為訓練集和驗證集，比例為 80% 與 20%

HW5_中文情感分類模型

模型建構與訓練

- 模型選擇：使用 bert-base-chinese 作為預訓練模型，進行微調以適應情感分類任務
- 分詞處理：利用 BERT 分詞器對文本進行分詞，並轉換為模型可接受的輸入格式
- 自定義 Dataset：建立 SentimentDataset 類別，將分詞後的資料與標籤封裝為可供模型訓練的格式
- 訓練參數

訓練參數與流程

- 使用 Huggingface Trainer 模組
- 訓練 3 個 epoch
- Batch Size: 16 (train) / 64 (eval)
- 使用 F1 與 Accuracy 評估模型表現
- 學習率調整、權重衰減等其他參數設定

HW5_中文情感分類模型

模型表現評估

- 使用驗證集進行評估
- Accuracy: 約 90%
- F1-score: 考慮 precision/recall 平衡

```
Evaluation result: {'eval_loss': 0.4645451009273529, 'eval_accuracy': 0.86125, 'eval_f1': 0.861244579866401, 'eval_runtime': 38.2065,
```

```
'eval_samples_per_second': 41.878, 'eval_steps_per_second': 0.654, 'epoch': 3
```

實際推論案例

```
[{'confidence': 0.9935, 'prediction': '負面', 'text': '這家店的服務態度真的非常差。'},  
{ 'confidence': 0.9911, 'prediction': '正面', 'text': '我非常喜歡這部電影，演員很棒！'},  
{ 'confidence': 0.9905, 'prediction': '負面', 'text': '東西還可以，但沒有我期待的那麼好。'},  
{ 'confidence': 0.9908, 'prediction': '負面', 'text': '太失望了，再也不會來了。'},  
{ 'confidence': 0.9917, 'prediction': '正面', 'text': '這是我吃過最好吃的牛肉麵！'},  
{ 'confidence': 0.9966, 'prediction': '正面', 'text': '店員很親切，讓人感覺很舒服。'},  
{ 'confidence': 0.9945, 'prediction': '負面', 'text': '產品質感普通，有待加強。'},  
{ 'confidence': 0.9614, 'prediction': '正面', 'text': '整體體驗還不錯，下次會再來。'},  
{ 'confidence': 0.9924, 'prediction': '負面', 'text': '不值得這個價錢。'},  
{ 'confidence': 0.9747, 'prediction': '正面', 'text': '客服反應迅速，問題很快就解決了。'},  
{ 'confidence': 0.9939, 'prediction': '負面', 'text': '這款手機的續航力真的很弱。'},  
{ 'confidence': 0.9921, 'prediction': '負面', 'text': '功能齊全，但外觀設計不夠吸引人。'},  
{ 'confidence': 0.9853, 'prediction': '正面', 'text': '雖然延遲了一下，但還是準時送達。'},  
{ 'confidence': 0.9879, 'prediction': '正面', 'text': '超出我的預期！真的太棒了。'},  
{ 'confidence': 0.9944, 'prediction': '負面', 'text': '包裝破損，內容物也有刮傷，失望。'},
```