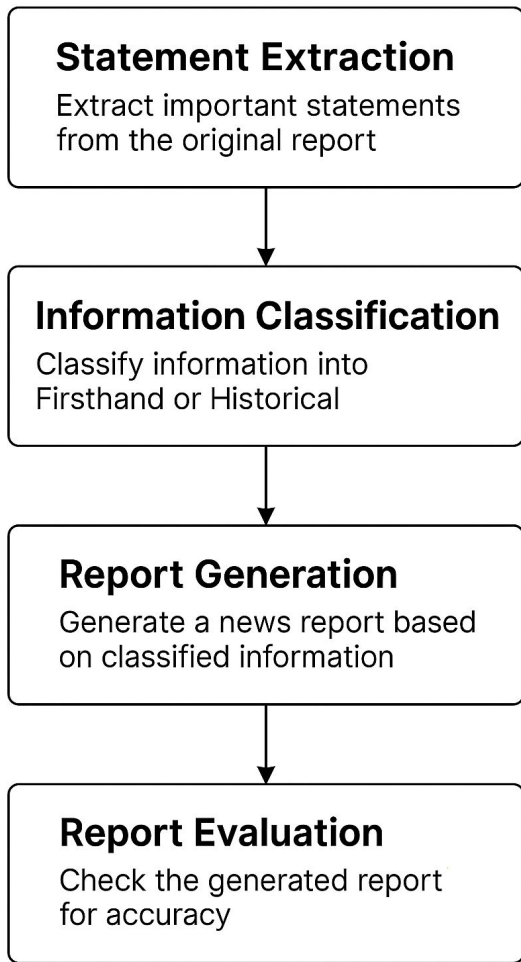


羽球專題LLM組

陳柏峯

目錄

1. (準備工作) 羽球新聞報導提取與重現	P3
2. (正式主題) 大型語言模型自動生成新聞報導概述	P4
3. 初步準備與規劃	P5-6
4. 操作篩選與執行	P7-10
5. 報告樹建構	P11
6. 生成報告	P12-13
7. 下學期規劃	P14



Flowchart

羽球新聞報導提取與重現

本測試與訓練在利用大型語言模型 (LLM) 提取新聞的簡短重點資訊，並把重要的第一手資訊與歷史資訊，重新給 LLM生成新聞報導。

黃色：抓到的重要訊息

紅色：可能的錯誤，因用詞較籠統(像似統整內容)非重點

****Croatian Badminton Makes History: Dinata Secures First European Championships Medal****^{1,2}

Zagreb, Croatia—Aria Dinata etched his name into Croatian sporting history today, **securing the nation's first-ever medal at the European Badminton Championships. His hard-fought 13-21, 21-18, 21-19 victory over Finland's Joakim Oldorf propels him into the semifinals.** This win mirrors Vivien Sándorházi's achievement for Hungary earlier today, marking a momentous day for both nations. Sándorházi also secured her country's inaugural European Championships medal.^{1,2}

Dinata, who relocated to Croatia from Indonesia three years ago, expressed his elation post-match: **"It was amazing because I have never played in such big halls, usually I play in the Future Series and International Series tournaments... I think it is a really great atmosphere, it has been really nice."** The scale of the competition was clearly a factor, a stark contrast to his previous experience in smaller tournaments. The visible emotion following the final point underscored the significance of this career milestone.^{1,2}

The match itself was a rollercoaster. While Oldorf took the first set decisively, Dinata demonstrated remarkable resilience, clawing back to win the final two sets in closely contested rallies. This comeback showcased not only his skill, but also his mental fortitude. This win sets up a semifinal clash against Toma Junior Popov, a formidable opponent. Undeterred, Dinata stated his resolve: **"I can say, let's bring it. I am always ready for a challenge, and I am not afraid of it."** His confident demeanor suggests a strong showing in the upcoming match, regardless of the underdog status.^{1,2}

The impact of Dinata's win extends beyond individual achievement. It marks a watershed moment for Croatian badminton, potentially inspiring a new generation of athletes and elevating the sport's profile within the nation. The implications for Croatia's seeding in future competitions are also significant, offering enhanced opportunities going forward.^{1,2}

May tomorrow's semifinal bring forth a display of skill, sportsmanship, and mutual respect between all competitors.^{1,2}

LLM生成成果

Made with GAMMA



大型語言模型自動生成新聞報導

本專案旨在利用大型語言模型 (LLM) 根據使用者提供的大綱和資料，自動生成完整的新聞報導。此系統可應用於羽球賽事報導，並能擴展至其他結構化表格資料的報導生成。

整題流程：

[讀取大綱] → [欄位篩選] → [分析規劃] → [列出操作] → [選定操作] → [執行操作] → [建構報告樹] → [生成報告]

原始資料與初步準備 [讀取大綱] → [欄位篩選]

輸入資料

- 新聞報導大綱檔案 (main.txt)
- 表格欄位描述檔案 (data_description.txt)
- 原始 CSV 資料檔案 (set1.csv)

資料篩選

利用 Gemini API 判斷欄位相關性，篩選出新聞報導可能需要的欄位，例如 'rally', 'time', 'player' 等，生成新的 CSV 檔案。

Gemini 回應 (1): ```python

```
['roundscore_A', 'roundscore_B', 'type', 'lose_reason', 'rally', 'player']
```

```

Gemini 回應 (2): ```python

```
['roundscore_A', 'roundscore_B', 'type', 'rally', 'lose_reason', 'getpoint_player', 'backhand', 'aroundhead']
```

```

Gemini 回應 (3): ```python

```
['roundscore_A', 'roundscore_B', 'type', 'rally', 'lose_reason', 'getpoint_player', 'player_location_area', 'opponent_location_area']
```

```

✓ 經過三次篩選後，出現兩次以上的欄位: ['roundscore\_A', 'roundscore\_B', 'type', 'lose\_reason', 'rally', 'getpoint\_player']

最終欄位清單:

```
['roundscore_A', 'roundscore_B', 'type', 'lose_reason', 'rally', 'getpoint_player']
```

# 分析規劃

```
Understanding the Data and Objectives:

* **Goal:** Analyze badminton rally data to create content for a badminton news outlet.
* **Perspective:** A senior badminton news reporter with deep badminton knowledge.
* **Data:** Rally-by-rally data, including time of shots, scores, players, shot types, reasons for losing the rally, and who won the rally.
* **Restriction 1:** Avoid common sense badminton knowledge, focusing on data-driven insights that might not be immediately obvious to the average fan.
* **Restriction 2:** All ideas and questions should be meaningful for the audience. Avoid trivial observations.

Chain-of-Thought Reasoning & Idea Generation:

1. **Outcome & Score (Idea 1 - Observed):** While observing the final score is a starting point, to make it interesting we need to dig deeper. Instead of
```

## :Chain-of-Thought [分析規劃]



LLM 作為規劃助手

使用 Gemini API, 輸入生成報導的注意事項和篩選後的欄位描述。



Chain-of-Thought 推理

生成相關且有意義的問題或規劃策略, 輸出為英文分析思路。



分析結果

生成 analyze\_response.txt, 包含詳細的分析思路。

# 操作識別與篩選 (第一階段) [列出操作]

## 1 識別操作

由扮演新聞記者的  
LLM 閱讀分析結果，  
識別對資料分析有用  
的表格操作。

## 2 操作列表

輸出為編號列表，包  
含 write,  
select\_row,  
select\_column 等，

## 3 初步結果

生成 operations\_info.json, 包含 add\_column  
value\_counts, crosstab 等操作。

✓ 操作清單與描述已儲存至 operations\_info.json

✓ 操作名稱陣列:

```
['**write**', '**select_row**', '**select_column**', '**groupby**', '**pivot_table**', '**sort**', '**merge**', '**add_column**', '**value_counts**', '**crosstab**']
```

# 操作篩選 (第二階段) 與重要性排序 [選定操作]

|    |                                     |
|----|-------------------------------------|
| 方法 | 將資料概要、完整資料、新聞大綱和操作清單提供給 Gemini API。 |
| 標準 | 評估操作是否能提供更多資訊、是否能運用在多個欄位、是否能提取大綱重點。 |
| 結果 | 獲得一個操作編號的重要性排序列表。                   |

```
Chain of Thought 分析:

1. **了解新聞大綱目標:** 新聞報導需要從數據中提煉出有意義的故事，並以清晰易懂的方式呈現。因此，分析操作的目標是找到能夠揭示比賽關鍵信息、模式和洞見的方法。

2. **評估各操作的功能:**

 * **write (1):** 直接生成文本，看起來很有用，但如果沒有好的分析結果作為輸入，LLM 可能會產生無意義或不準確的內容。因此，`write` 依賴於其他分析操作的結果。
 * **groupby (4):** 能夠將數據按特定欄位分組，例如按球員、球的類型、失分原因等分組。這有助於找出不同球員的優勢和劣勢，不同球種的使用頻率，以及常見的失分模式。這對於深入理解比賽至關重要。
 * **value_counts (9):** 快速統計某個欄位中各個值的出現次數。例如，統計每種球的類型被使用的次數，或者每種失分原因的次數。這是一個簡單但有效的初步分析工具。
 * **crosstab (10):** 用於分析兩個或多個因素之間的關係，例如球員和球的類型，或者球的類型和失分原因。這可以幫助我們發現哪些球員更擅長使用哪種球，或者哪些球的類型更容易導致失分。
 * **sort (6):** 排序操作本身不能直接提供分析洞見，但它可以幫助我們更容易地識別數據中的模式。例如，按 rally 的時間排序可以幫助我們分析比賽的節奏。
 * **select_row (2) & select_column (3):** 這兩個操作用於選擇數據的子集。在分析特定方面時非常有用，例如只關注某個球員的數據，或者只關注特定的欄位。
 * **pivot_table (5):** 產生一個樞紐分析表，可以讓我們以不同的方式總結和查看數據。它在探索多維度數據時非常有用。
 * **add_column (8):** 添加新欄位，例如計算球員的得分率或失誤率。這些新欄位可以提供更有意義的分析。
 * **merge (7):** 合併表格在單一比賽數據分析中可能用處不大，除非要合併其他相關數據（例如球員信息）。

3. **排序依據:** 根據「重要性描述」的標準，著重考慮哪些操作能直接提煉信息、適用範圍廣、並能提取大綱重點。

排序結果:
...

排序操作重要性編號: [4, 9, 10, 8, 5, 2, 3, 6, 1, 7]
```



# 確定最終操作集 [選定操作]

## 篩選策略

從排序結果中保留最重要的操作。

## 強制保留

write, select\_row,  
select\_column 這三個基本操作  
必須保留。

## 數量控制

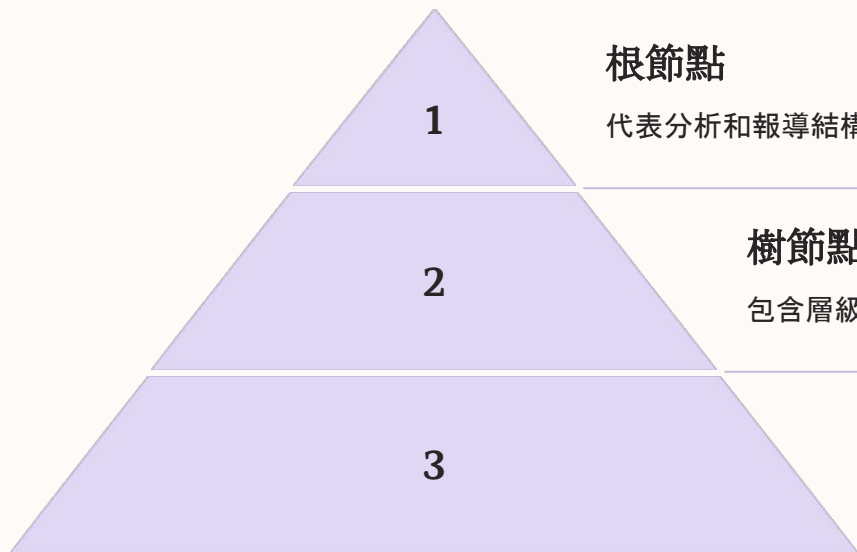
保留前約 70% 的操作, 最終生成 selected\_operations.json。

```
1 write
2 select_row
3 select_column
4 groupby
5 pivot_table
6 sort
7 add_column
8 value_counts
9 crosstab
```

# 數據操作執行與程式碼生成 [執行操作]



# 報告樹 (Tree-of-Report) 建構 [建構報告樹]



## 根節點

代表分析和報導結構的起點。

## 樹節點

包含層級、文本、DataFrame、操作等資訊。

## 建構邏輯

非葉節點根據規劃器生成子節點，避免近期冗餘操作。

```
2025-06-07 20:58:40,960 - INFO - 開始建構報告樹...
2025-06-07 20:58:40,975 - INFO - 處理節點 - Level: 0, Operation: root(None)
2025-06-07 20:58:40,989 - INFO - 嘗試第 1 次向 Gemini 發送請求...
2025-06-07 20:58:42,102 - INFO - 成功獲得 Gemini 回應
2025-06-07 20:58:42,102 - INFO - 生成操作: ['select_column(rally, player, type, lose_reason, getpoint_player)', 'value_counts(type)', 'value_counts(lose_reason)']
2025-06-07 20:58:45,622 - INFO - 操作成功，結果形狀: (315, 5)
2025-06-07 20:58:45,623 - INFO - 創建數據操作節點: select_column(rally, player, type, lose_reason, getpoint_player)，結果形狀: (315, 5)
2025-06-07 20:58:45,623 - INFO - 添加子節點: efa7c01a to c56cc883
```

# 文本生成與合併 [生成報告]

## 文本生成器

負責為報告樹中的節點生成文本。

## write() 節點文本

根據 DataFrame 內容生成簡短、符合語氣的英文文本，提取亮點。

## 父節點文本

合併並改寫子節點文本，形成結構化和連貫的文本。

```
node.table: lose_reason count
0 對手落地致勝 12
1 出界 12
2 掛網 10
3 未過網 2
```

節點文本：本場比賽失分多集中在對手落地致勝和自身出界，各有12次，可見防守端需加強，控制自身失誤也至關重要。

# 最終報告生成與輸出

## [生成報告]



### 報告生成

從報告樹的根  
節點開始，遞歸  
生成最終報導  
文本。



### 最終報導 格式

LLM 生成全面  
賽事新聞報導，  
使用英文、生動  
文句。



### 輸出文件

生成  
tree\_of\_report.txt,  
tree\_structure.json,  
execution\_report.md  
等。



### 執行監控

記錄節點處理  
日誌，呈現統計  
信息和錯誤摘  
要。

## table-to-report 一局比賽新聞稿

## Player A Dominates with Strategic Play, Secures Victory<sup>4,5</sup>

In a thrilling badminton match characterized by a dynamic shift between tempos and calculated maneuvers, Player A emerged victorious against Player B with a score of 21-15.<sup>4,5</sup>

The match unfolded as a chess game on the court, with both players exhibiting a wide array of shots. Long, high clears were frequently employed to dictate the pace, punctuated by powerful smashes and delicate drop shots aimed at disrupting the opponent's rhythm. Lifts and net play, including drops and pushes, added further complexity to the tactical battle.<sup>4,5</sup>

While both players displayed skillful techniques, critical errors proved costly. Too often, shots sailed beyond the boundary line, gifting precious points to their rival. Net errors also became a significant factor, particularly benefiting Player B, who capitalized on these miscues.<sup>4,5</sup>

Player A demonstrated an advantage, successfully leveraging a particular type of shot to secure decisive points. However, the rallies often turned into dramatic exchanges, with both athletes showcasing their ability to execute slicing shots followed by delicate net play.<sup>4,5</sup>

Despite the intense back-and-forth, Player A's scoring proved too much for Player B to overcome. A considerable number of points remain a mystery, lacking detailed information, but the overall picture painted a clear narrative of Player A's strategic dominance and ultimate triumph. The match served as a testament to the importance of both offensive prowess and minimizing errors in the pursuit of victory.<sup>4,5</sup>

使用的表格

ShuttleSet:An\_Se\_Young\_Pornpawee\_Chochuwong\_TOYOTA\_THAILA  
ND\_OPEN\_2021\_QuarterFinals/set1.csv