

社群媒體分析 第一次讀書會專案

指導教授：黃三益 教授

助教：蔡易行、張家銘、呂育真、謝嘉霖

針對金秀賢事件－ 對 PTT 八卦版進行情緒分析

第十一組成員：

N124320004 莊筑雅

N124320006 陳勁廷

N124320009 吳亭緯

N124320027 楊毅婷

N124320030 傅才容

N124320023 蕭瑜堯

N124320003 馬永恩

N124320016 李明容

Contents

確定研究問題與目標.....2

資料收集.....3

資料前處理.....4

情緒分析.....5

關鍵詞分析.....7

NER（命名實體識別）分析.....8

POS（詞性標註）分析.....9

情緒波動時間段分析.....10

結論與建議.....11

附錄.....12

確定研究問題與目標

主題概述：

- 本研究聚焦於 PTT 八卦版對「金秀賢事件」的討論情緒分析，目的在探討該事件在社群媒體上的輿論走向與網友情緒反應。「金秀賢事件」指韓國演員金秀賢被指控與已故女演員金賽綸在其未成年時交往，並與後續發展產生關聯。該事件引發廣泛關注，輿論呈現多元聲音，部分媒體選擇低調處理，而社群平台則成為主要的意見交流場域。本研究透過文本情緒分析，解析該事件在 PTT 八卦版的討論內容，辨識主要情緒類別及影響因素，藉此了解社群輿論對事件的影響力與變化趨勢。

研究目標：

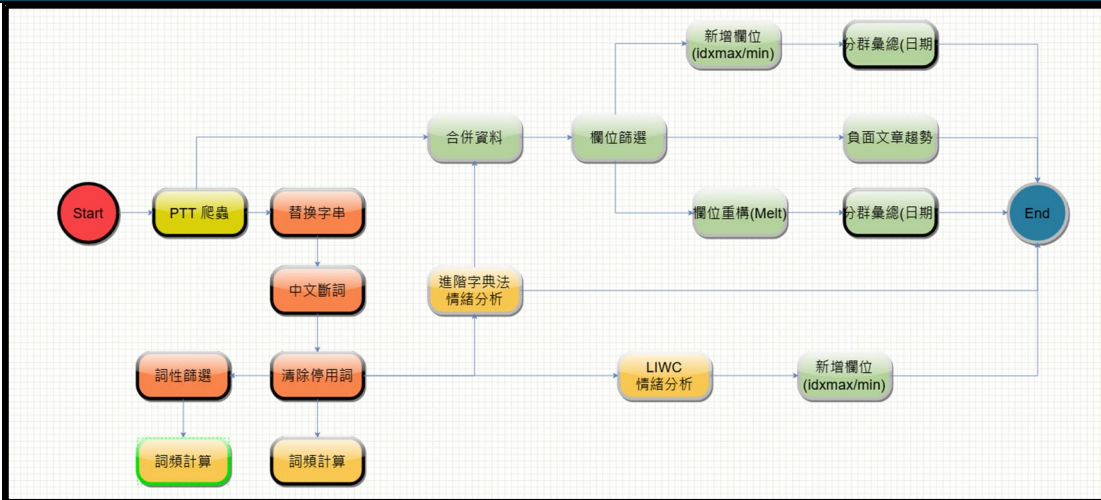
- 通過情緒分析，了解金秀賢事件在 PTT 八卦版中的影響，並探討網友對事件的態度與情感。
- 分析網友對事件的不同觀點，辨識主要情緒類別（如憤怒、支持、嘲諷等）及其影響因素。
- 透過關鍵詞與文本分析，了解討論內容的主要關注點，並比較本事件與其他類似事件的情緒反應異同。

資料收集

資料來源：

- 選定 PTT 八卦版的文章作為資料來源，時間範圍 2025 年 1 月~2025 年 3 月，收集相關討論的文章。

系統流程：



關鍵字設置：

- 針對事件的討論，設置合適的關鍵字，例如「金秀賢」、「八卦」、「事件」、「爭議」、「討論」等。

PTT 爬蟲 (6)

參數設定

任務結果

選擇看板 *

Finance(金融業)

Food(美食)

ForeignEX(外匯)

forsale(二手買賣)

gay(甲)

GetMarry(結婚)

Gossiping(八卦)

搜尋關鍵字 ⓘ

金秀賢

金賽綸

金賽倫

金賽羅

排除關鍵字 ⓘ

以換行區隔，e.g.
壽山動物園
猴子
...

搜尋起始日期

2025/01/01

搜尋結束日期

2025/03/16

資料量設定：

- 決定抓取文章的數量，範圍內的文章數量可以在 1,000 到 3,000 篇之間。

資料前處理

資料清理：

- 進行文本的清理，移除 URL、空值、特殊字符（如換行符號）等不必要的內容，並確保文本的乾淨與規整。

≡ 替換字串 (9)

參數設定	Input - 6	任務結果
選擇處理欄位 *	替換字串設定 ⓘ	
artContent	<pre>\n\n>> * \n>> >></pre>	
選擇替換規則檔案 ⓘ		
-----請選擇-----		

斷句與斷詞：

- 進行語句拆分與詞彙切分，並生成單詞列表。

≡ 中文斷詞 (13)

參數設定	Input - 9	任務結果
選擇處理欄位 *	定義詞彙 ⓘ	
result	<pre>金秀賢 500 金賽綸 500</pre>	
選取字典 ⓘ		
-----請選擇-----		

去除停用詞：

- 去除無實際意義的停用詞（如「的」、「是」、「在」等），提高後續分析的準確性。

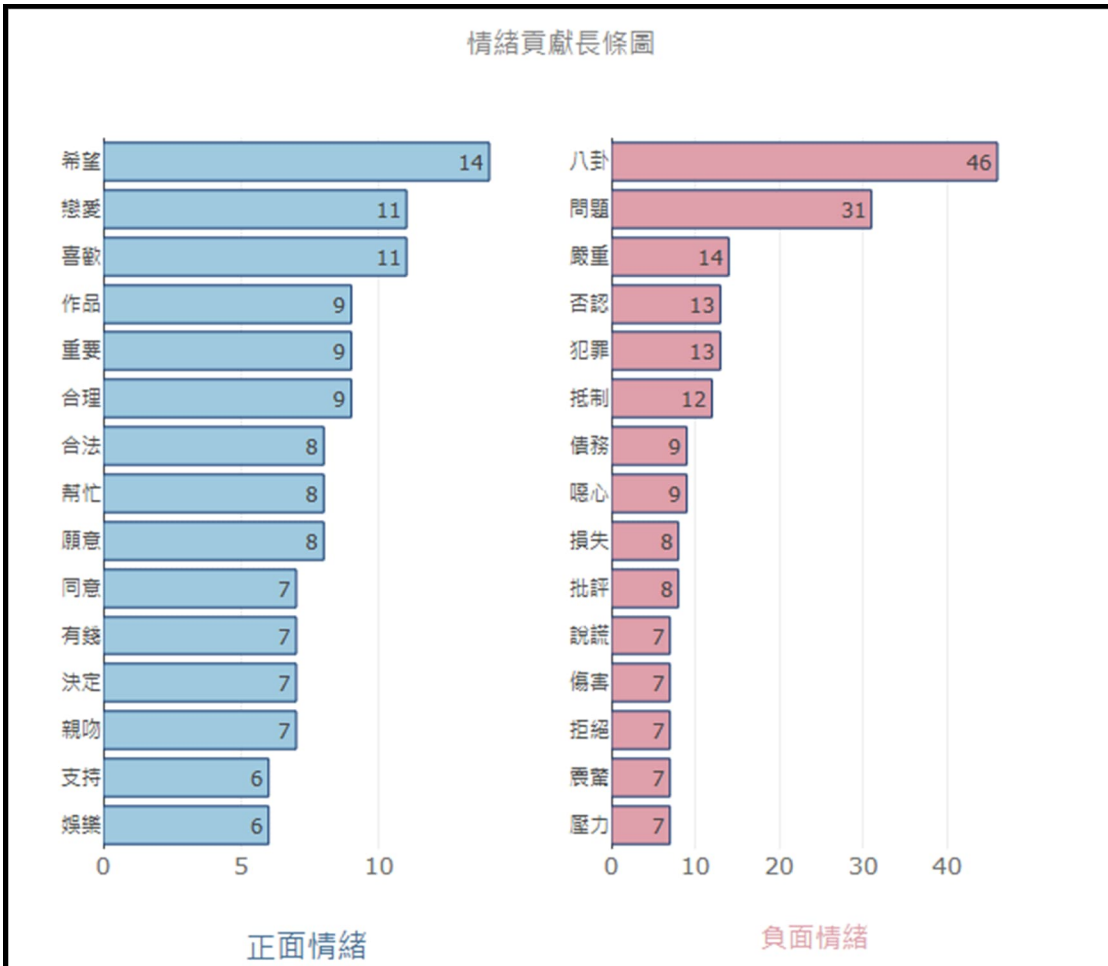
≡ 清除停用詞 (14)

參數設定	Input - 13	任務結果
語言 *	使用預設停止詞	
Chinese	是	
是否清除單字元 ⓘ	是否轉為小寫英文	
是	是	
清除英文字母 *	清除數字 *	
是	是	
清除換行符號 *	清除特殊標點符號 *	
是	是	
清除html tag *	自定義停止詞	
是	<pre>台灣 小姐 先生 現在 備註 知道 這種 有沒有 研究所</pre>	

情緒分析

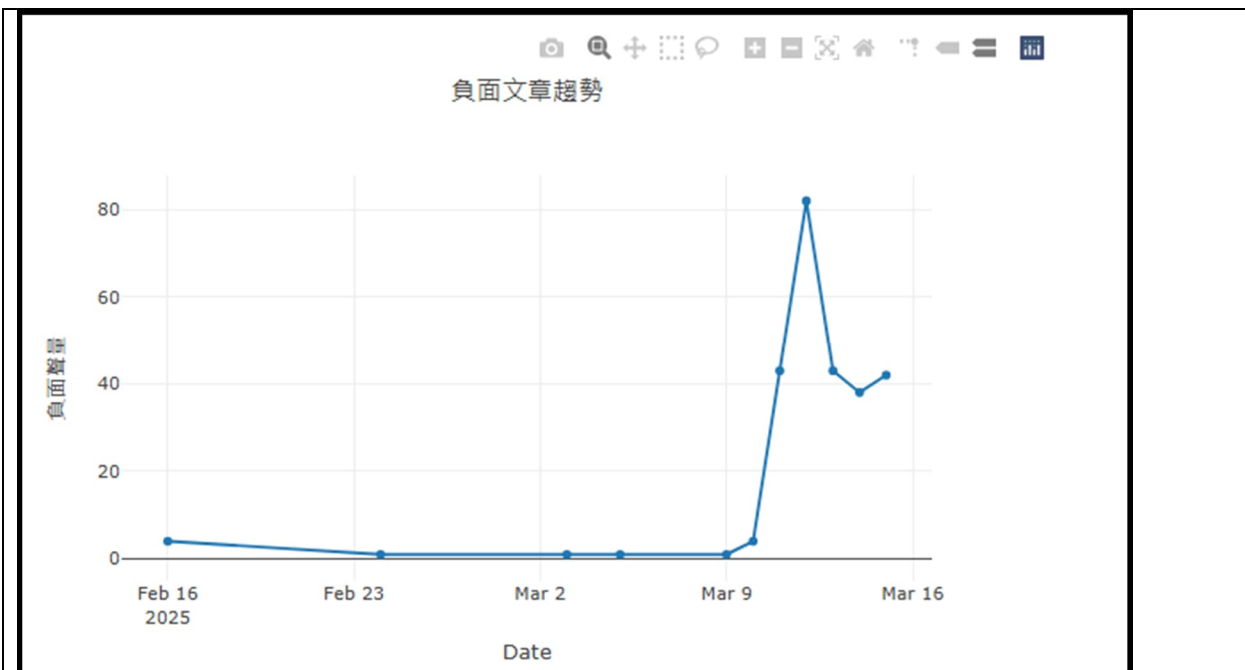
情緒分類：

- 使用情緒分析工具對每篇文章進行情緒分類，將文章分為正面、負面。這可以基於關鍵情緒詞彙、語氣等進行。



情緒波動：

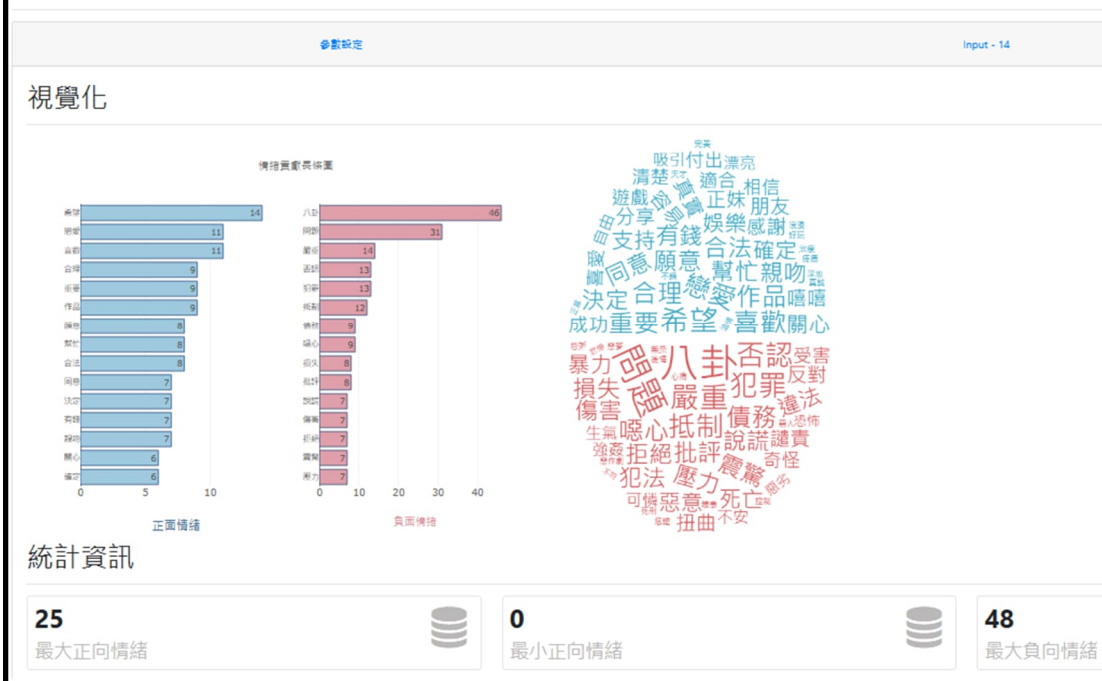
- 根據不同日期，繪製情緒變化的折線圖，觀察事件發酵前後網友情緒的變化，分析是否有情緒高潮（如熱烈的正面或負面討論）。



情緒詞彙分析：

- 統計正向和負向情緒詞彙的出現頻率，並生成文字雲，觀察哪些詞彙與正負情緒相關（例如：「支持」、「憤怒」、「爭議」、「失望」等）。

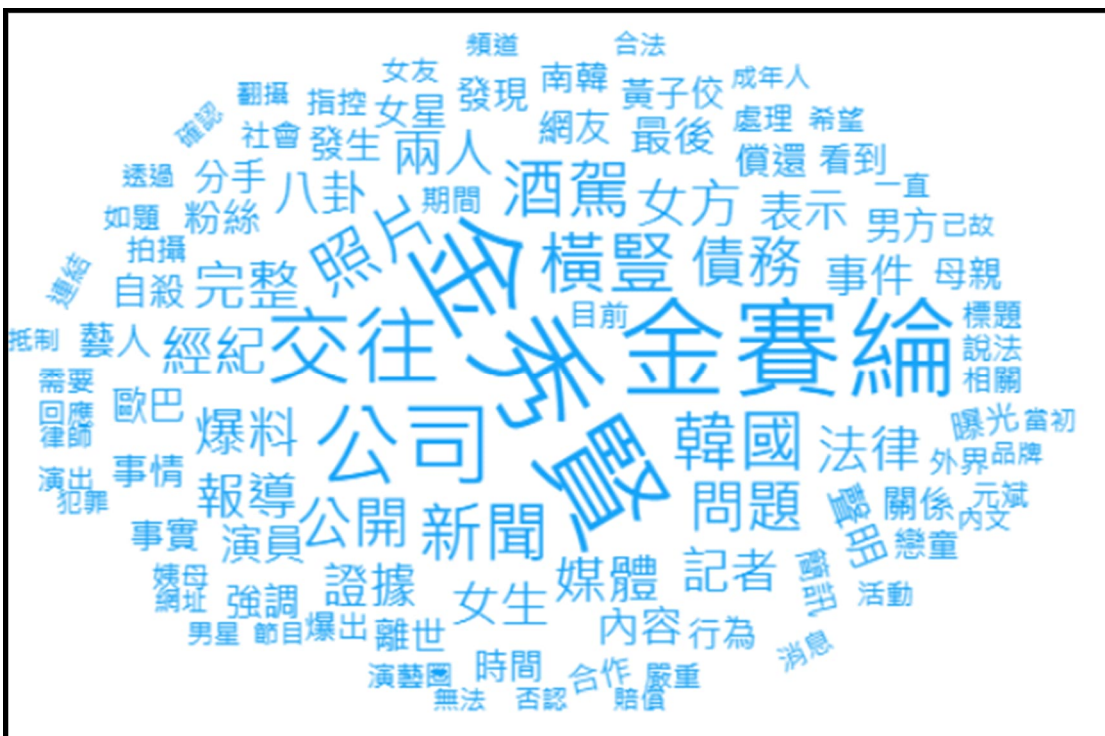
進階字典法 情緒分析 (21)



關鍵詞分析

頻率分析：

- 列出 PTT 文章中出現頻率較高的關鍵詞，生成文字雲，並根據頻率分析其對情緒的影響。



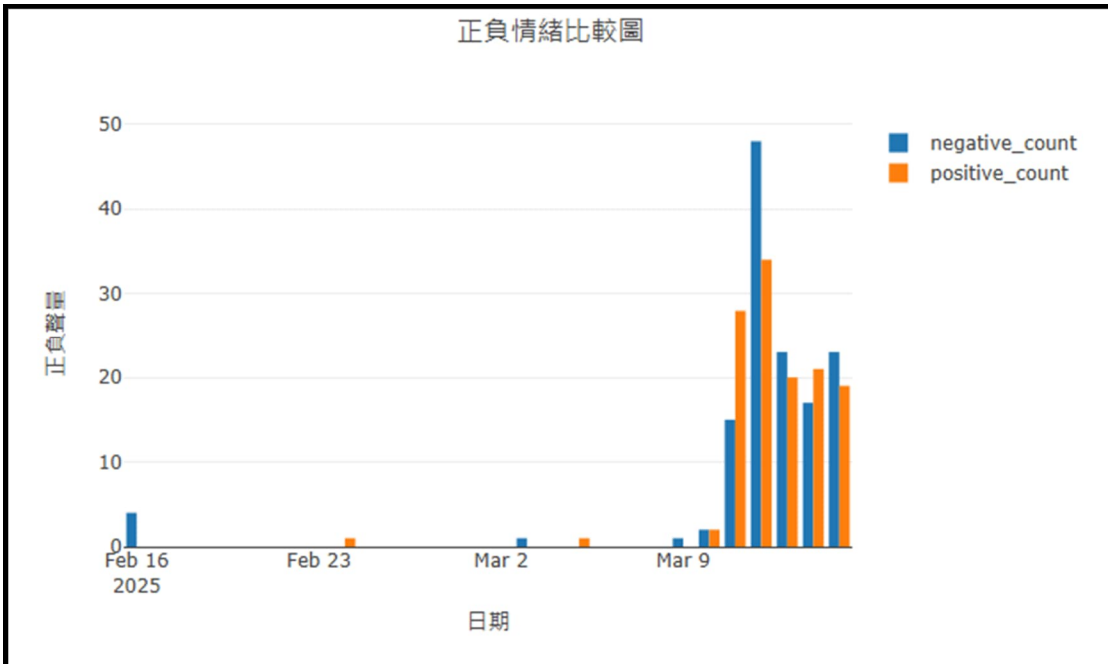
詞彙關聯：

- 進行詞彙的共現分析，觀察與「金秀賢」、「事件」等相關的常見詞彙，並分析其情緒走向。

NER（命名實體識別）分析

實體識別：

- 針對文章中提到的實體（如「金秀賢」、「相關人員」、「事件地點」等）進行命名實體識別（NER）分析，並分析其討論情緒。



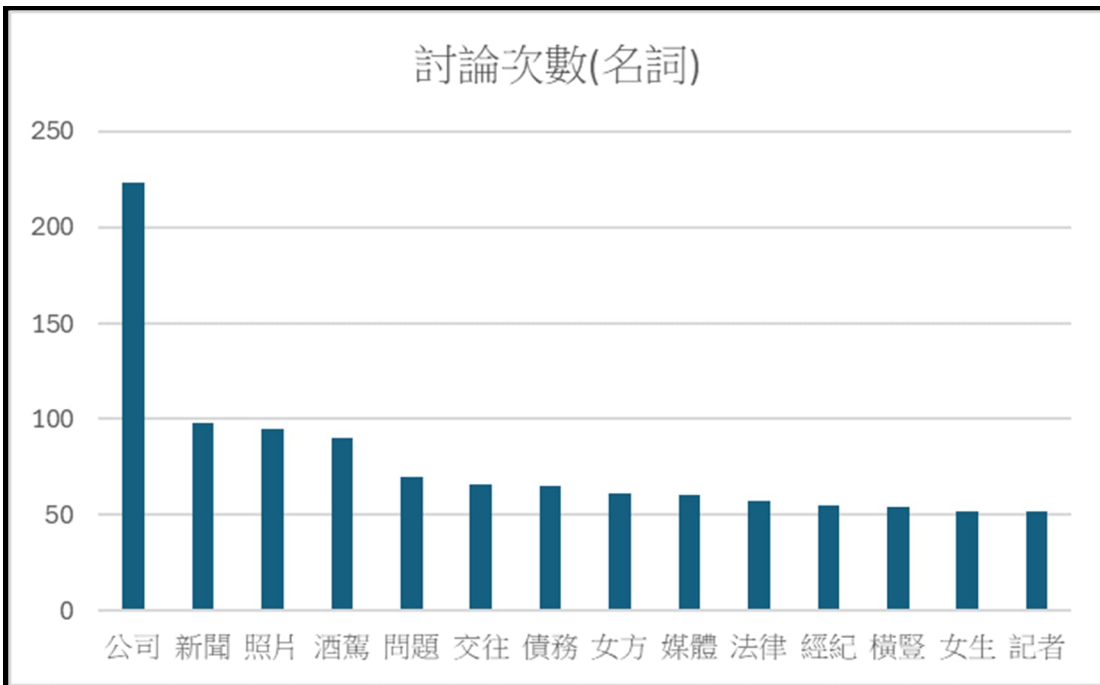
影響人物分析：

- 如果有提到其他人物或公眾人物（例如演藝圈其他名人），可進行其情緒分析與輿論走向分析。

POS（詞性標註）分析

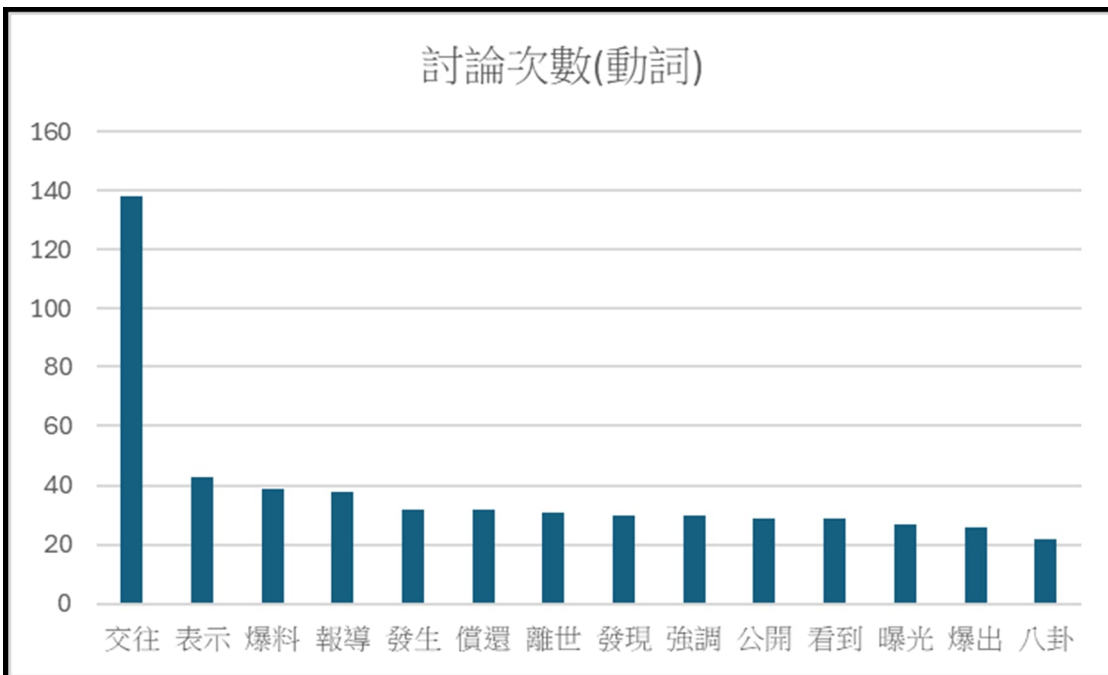
名詞分析：

- 統計與金秀賢事件相關的名詞，找出最常出現的人物、事件、場景等。



動詞分析：

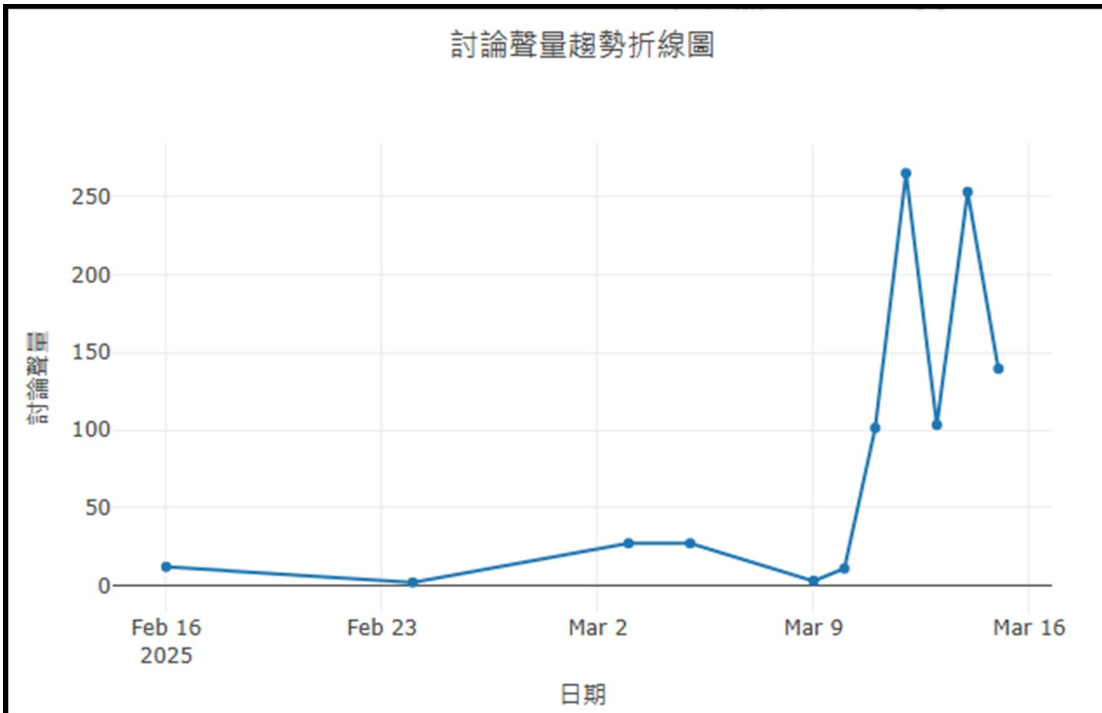
- 分析文章中動詞的頻率，看看「支持」、「反駁」、「討論」等行為是否有高頻次出現，這些動作能顯示出網友的情緒動向。



情緒波動時間段分析

高峰期分析：

- 根據文章的情緒波動，集中分析在某一時間段內的情緒波動，分析事件發生後幾天內的網絡反應。



社會事件關聯：

- 如果有其他社會事件與金秀賢事件相關，可以分析這些事件對網友情緒的影響（例如其他名人的聲明、媒體報導等）。

結論與建議

總結情緒趨勢：

- 根據 PTT 八卦版的情緒分析，金秀賢事件的討論情緒呈現以下趨勢：
 1. 負面情緒為主：整體而言，網友對此事件的討論以負面情緒為主，特別是憤怒、失望與嘲諷等情緒較為突出。
 2. 情緒波動明顯：事件初期因媒體曝光引發情緒高峰，負面情緒占主導地位；隨後討論熱度下降，但在關鍵新聞發布或其他名人發聲時，情緒波動再次上升。
 3. 關鍵詞分析：與「金秀賢」、「未成年」、「爭議」、「經紀公司」等詞彙關聯的討論多數帶有批評意味，而「支持」、「粉絲」等詞彙則相對較少，顯示社群輿論較不利於金秀賢。
 4. 情緒分類結果：情緒分析顯示，PTT 八卦版的討論主要可分為「憤怒批評」、「嘲諷與戲謔」、「關注後續發展」三大類別，其中前兩者占比較大，顯示網友對此事件持高度負面評價。

未來研究建議：

- 比較不同社群平台的情緒趨勢
擴展分析範圍，對比 PTT 與 Dcard、Facebook、X(Twitter) 等不同平台的討論情緒，探討各社群對此事件的不同態度。
- 深入分析情緒演變過程
透過時間序列分析，觀察討論熱度與情緒變化，分析哪些事件節點對輿論影響最大，例如媒體報導、官方聲明或其他名人發聲等。
- 探索輿論影響因素
透過主題模型 (Topic Modeling) 分析，識別影響情緒變化的核心話題，例如道德批判、法律責任、娛樂圈影響等，以更深入理解輿論的形成機制。
- 應用情緒分析於其他社會事件
可將本研究方法應用於其他類似爭議性事件，分析大眾對名人醜聞的情緒模式，並探討社會輿論如何影響個人聲譽與媒體報導方向。

分析工具與方法：

- 情緒分析工具：使用 Tarflow 進行文本情緒分類，將 PTT 八卦版的討論內容標註為正面、負面或中立，並分析情緒詞彙的出現頻率與關聯性。
- 自然語言處理 (NLP) 技術：應用文本預處理技術，包括斷詞 (使用 Jieba)、去除停用詞、詞性標註 (POS tagging)，以提高分析精準度。
- 關鍵詞與語意分析：透過 TF-IDF 計算關鍵詞權重，並使用詞彙共現分析 (Co-occurrence Analysis) 與命名實體識別 (NER) 技術，識別討論核心議題及主要人物。
- 情緒變化與趨勢分析：透過時間序列分析，繪製情緒變化折線圖，追蹤事件不同階段的輿論波動，並結合外部新聞或社群影響因子進行交叉比對。

圖表展示：

- 文字雲 (Word Cloud)：視覺化呈現 PTT 八卦版討論中的高頻關鍵詞，突出與金秀賢事件相關的重要詞彙與情緒詞語。
- 情緒趨勢折線圖 (Sentiment Trend Line Chart)：透過時間序列分析，繪製不同情緒類別 (正面、負面) 隨時間變化的趨勢，觀察事件發酵期間的輿論波動。
- 情緒波動圖 (Sentiment Fluctuation Chart)：針對特定時間點 (如媒體報導、官方聲明發佈後) 進行情緒變化分析，評估關鍵事件對社群情緒的影響。
- 關鍵詞關聯圖 (Keyword Co-occurrence Network)：透過詞彙共現分析，繪製關鍵詞之間的關聯網絡，幫助理解不同議題間的連結與輿論方向。