



● 社群媒體分析期中報告 從批踢踢看2024總統大選

- 指導教授：黃三益教授
第11組成員
N104220015 陳韻元
N104220011 呂伉弘
N104220029 陳聖宗
N104020002 廖英捷

CONTENTS

01

動機和分析目的

02

資料集的描述

03

資料的分析過程

04

視覺化的分析結果與解釋

05

結論

01

動機和分析目的

動機與分析目的

有討論熱度

Gossiping(八卦)、
HatePolitics(政黑)版對政
治上的討論量都持續成長，
可藉此瞭解民眾關注的焦
點和選擇傾向。

可做評價分析

挖掘網友對各政治人物和政黨
的正面與負面評價，並分析其
背後的原因。

pt
t

預測選舉趨勢

根據分析結果，預測2024總
統大選可能的選舉趨勢，為
未來政治研究和選舉策略提
供參考依據。

可蒐集事件反應

透過時間序列分析，探討政治事件
對批踢踢上討論熱度的影響，以及
網友對事件的反應。

02

資料集
的描述

資料集的描述

起訖時間：
2022/1/1
至
2023/4/17
(共計16個月)

本組透過文字探
勘分析平台之
PTT爬蟲功能，
蒐集2024年大選
相關貼文

文字探勘工
作流程平台

- (PTT)爬蟲
Gossiping(八卦)
HatePolitics(政黑)

資料來源

關鍵字

- 2024大選
- 2024總統大選
- 台灣總統大選

資料筆數

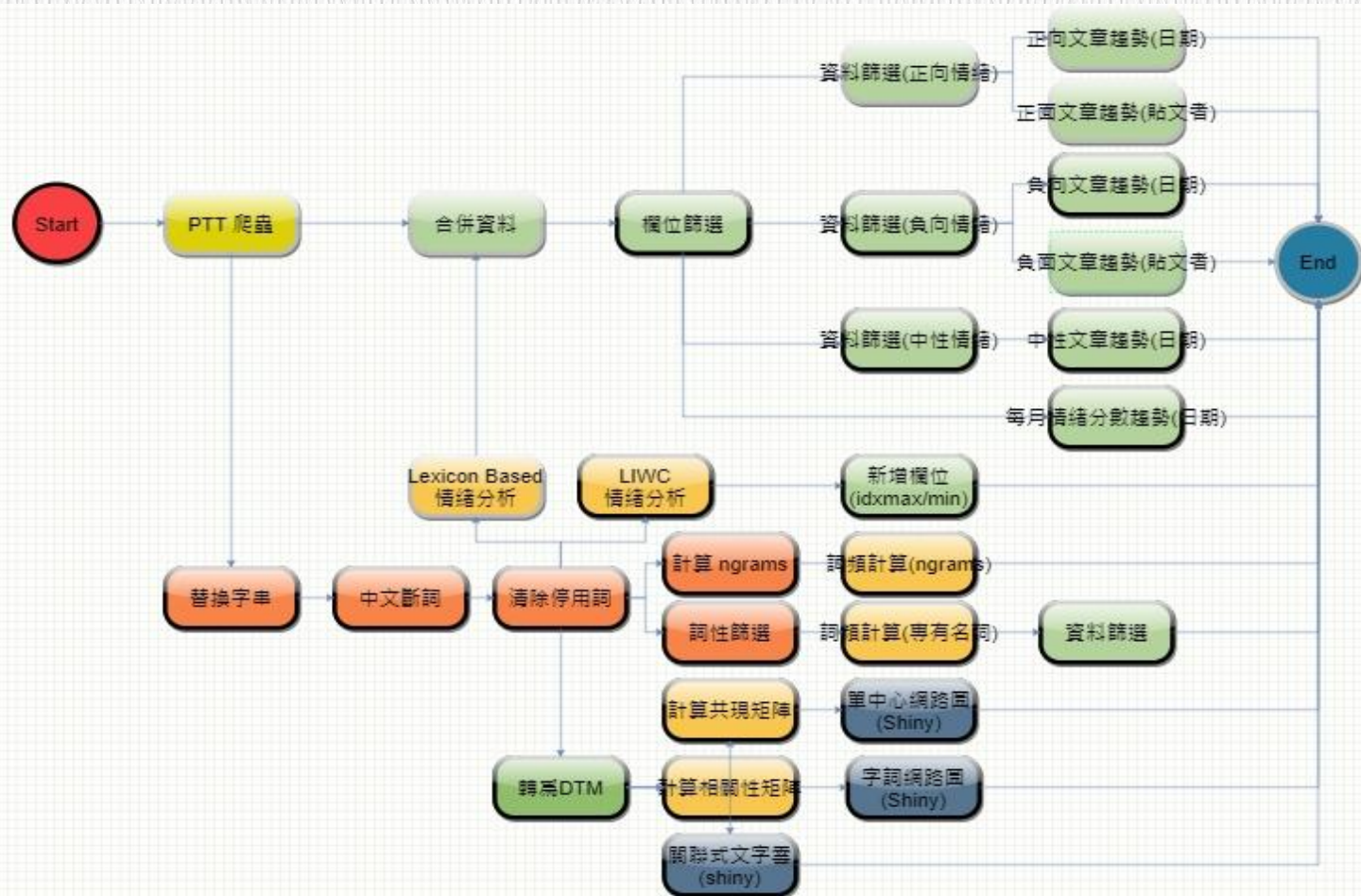
共1,543筆

03

資料的分析過程

- 分析模型
- 資料前處理

分析模型



資料前處理

1. 替換字串：

替換「\n」及「\n\n」，再替換「日期」、「時間」、「標號」及「網址」等，再替換藉由計算 ngrams 及詞頻計算功能找出五大潛在候選人相關字詞，如「郭台銘」類似字詞有「郭董」；「柯文哲」類似字詞有「阿北」、「柯P」、「柯p」。

☰ 替換字串 (7)


參數設定

Input - 4


任務結果

選擇處理欄位 *

artContent

選擇替換規則檔案 

-----請選擇-----

替換字串設定 

郭董>>郭台銘

阿北>>柯文哲

柯P>>柯文哲

柯p>>柯文哲

\n\n>>°

儲存更改

資料前處理

2.中文斷詞:

將常見於新聞字詞如「媒體來源」、「記者署名」、「發佈時間」、「最後更新時間」、「完整新聞標題」、「完整新聞內文」、「完整新聞連結」、「短網址」、「新頭殼」, 以及容易遭誤判的「韓國瑜」、「吳怡農」等設定為斷詞詞彙。後續將把新聞字詞進行停用字篩除。

☰ 中文斷詞 (8)


參數設定

Input - 7


任務結果

選擇處理欄位 *

result

定義詞彙 

媒體來源 100
記者署名 100
發佈時間 100
最後更新時間 100
完整新聞標題 100

選取字典 

-----請選擇-----

儲存更改

資料前處理

3.清除停用詞：

將「媒體來源」、「記者署名」、「發佈時間」、「最後更新時間」、「完整新聞標題」、「完整新聞內文」、「完整新聞連結」、「快訊」、「編輯」、「報導」、「短網址」、「備註」、「新頭殼」、「newtalk」、「ETtoday」、「網址」、「議題」、「網路」、「年底」、「現在」、「過去」、「未來」、「日前」、「今年」、「去年」、「今天」、「內容」、「明年」設定為自定義停用詞。

清除停用詞 (13)

參數設定	Input - 8	任務結果
清除英文字母 *	清除數字 *	
<div>是</div>	<div>否</div>	
清除換行符號 *	清除特殊標點符號 *	
<div>是</div>	<div>是</div>	
清除html tag *	自定義停止詞	
<div>是</div>	媒體來源 記者署名 發佈時間 最後更新時間	

04

視覺化的分析

- 五大潛在候選人熱度分析
- 文章情緒分析
- 字詞相關性

結果與解釋

五大潛在候選人熱度分析

1.詞性篩選:

本組為瞭解潛在的候選人的網路聲量熱度, 特地篩選專有名詞(Proper Noun)。

詞性篩選 (26)

參數設定

Input - 13

任務結果

語言 *

Chinese

選擇保留詞性 *

Noun

Proper Noun

Verb

Adjective

Adverb

儲存更改

五大潛在候選人熱度分析

2.繪製文字雲：

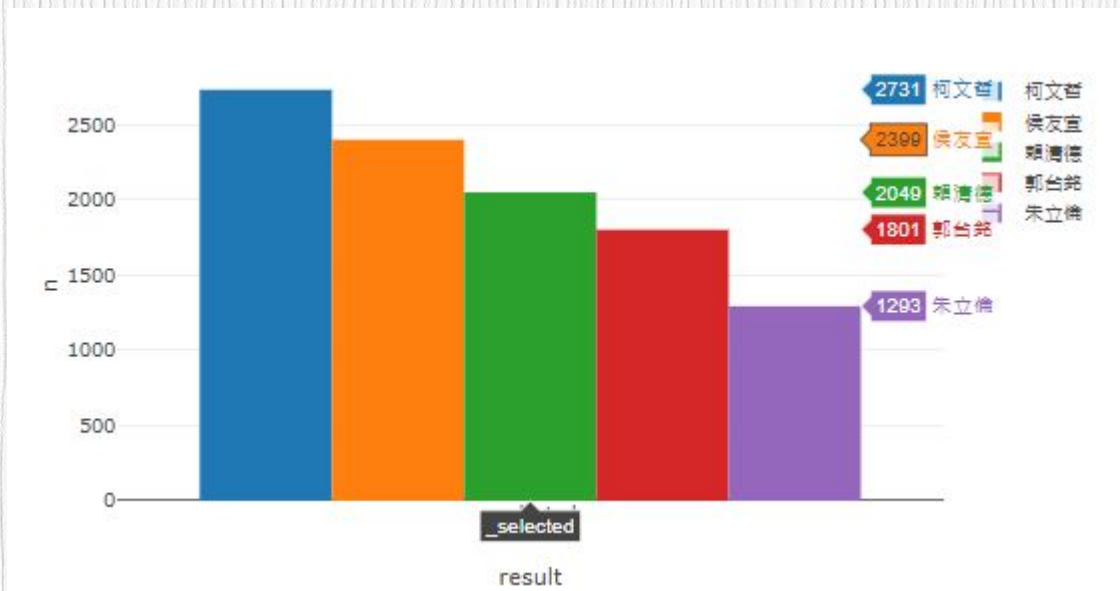
再透過詞頻計算，繪出文字雲，前十大字詞有「國民黨」、「民進黨」、「台灣」、「柯文哲」、「侯友宜」、「賴清德」、「郭台銘」、「朱立倫」、「美國」、「中國」。



五大潛在候選人熱度分析

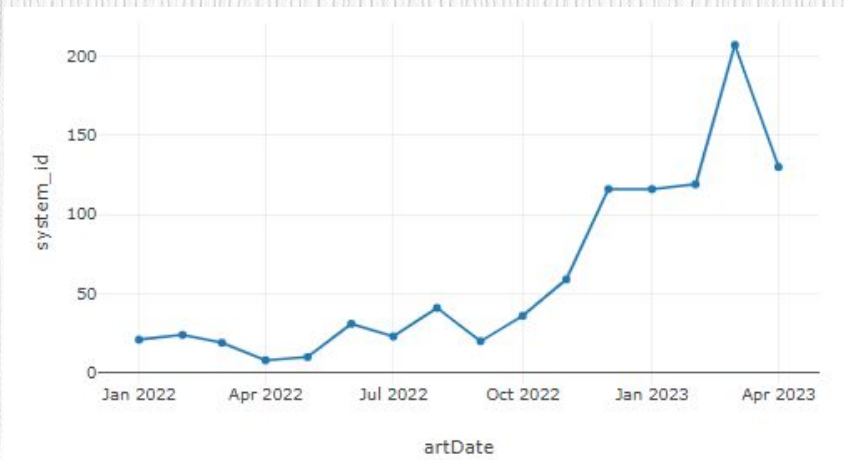
3.候選人熱度分析：

利用資料篩選，透過公式整理出前五大潛在候選人依序為「柯文哲」、「侯友宜」、「賴清德」、「郭台銘」、「朱立倫」。

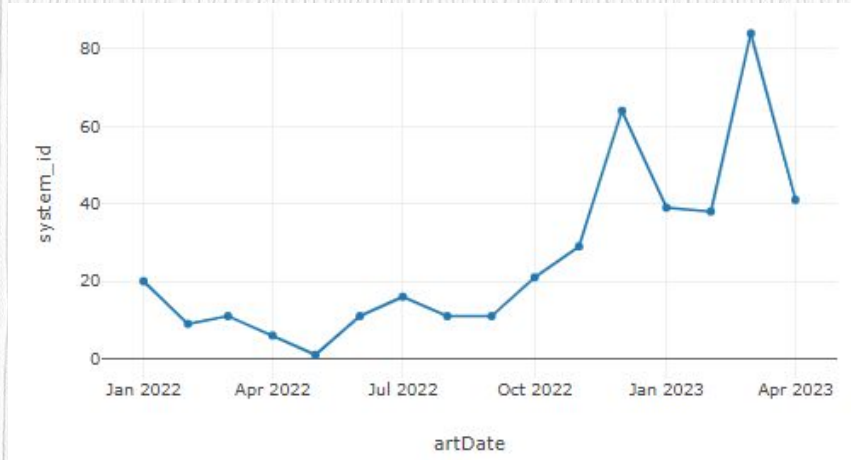


文章情緒分析

1.正向文章趨勢及負向文章趨勢



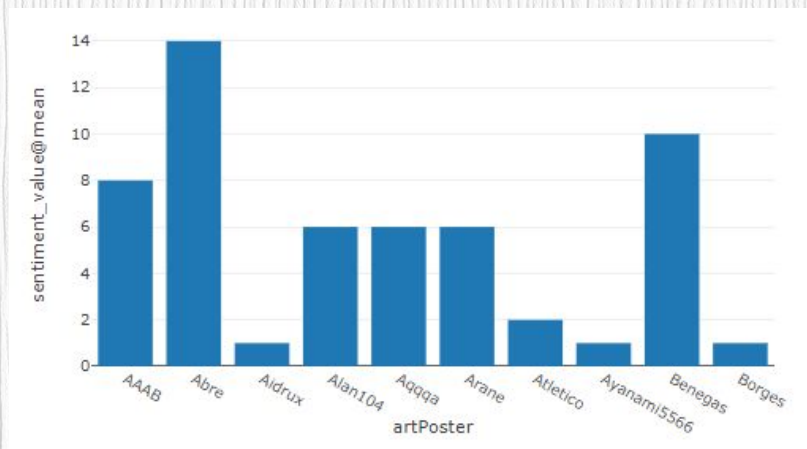
正向文章趨勢於111年12月逐漸增加, 平均有100則文章數以上。



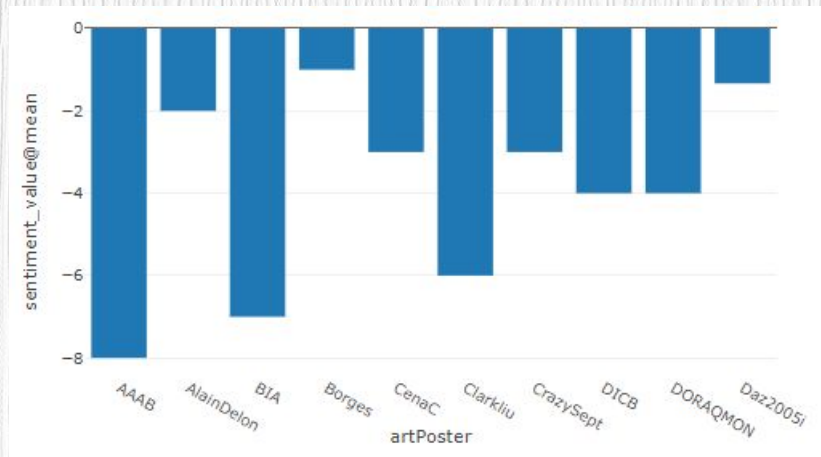
負向文章則於111年12月起逐漸增加, 可知與當年底的縣市長選情有關, 文章數量約30則以上90則以下。

文章情緒分析

2. 正向文章貼文者及負向文章貼文者



正向文章貼文者，以 Abre 最多(情緒分數 14)，Benegas 次之(情緒分數 10)。



負向文章貼文者，以 AAAB 最多(情緒分數 -8)，BIA 次之(情緒分數 -7)。

文章情緒分析

3.每月情緒分數趨勢



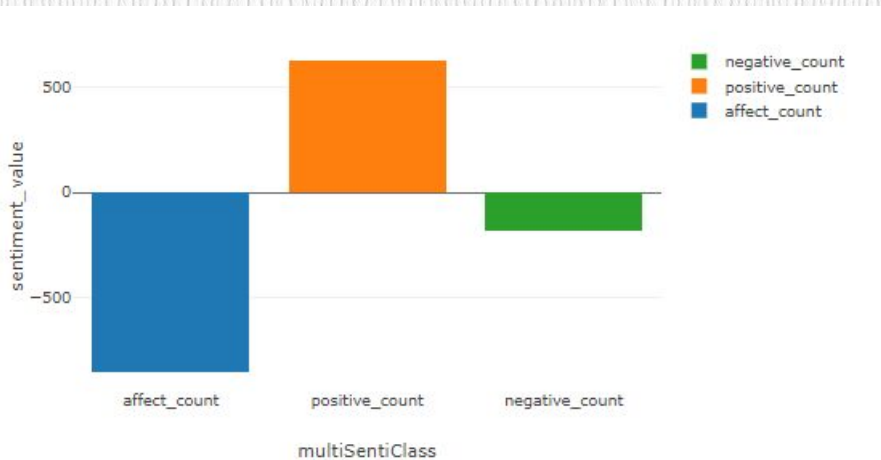
最低情緒出現於111年5月，而當年度7月亦偏低；之後隨11月縣市長選舉逐漸上升。



112年後，隨著選情加溫，目前最高情緒出現於112年4月。

文章情緒分析

4.正負向文章數量



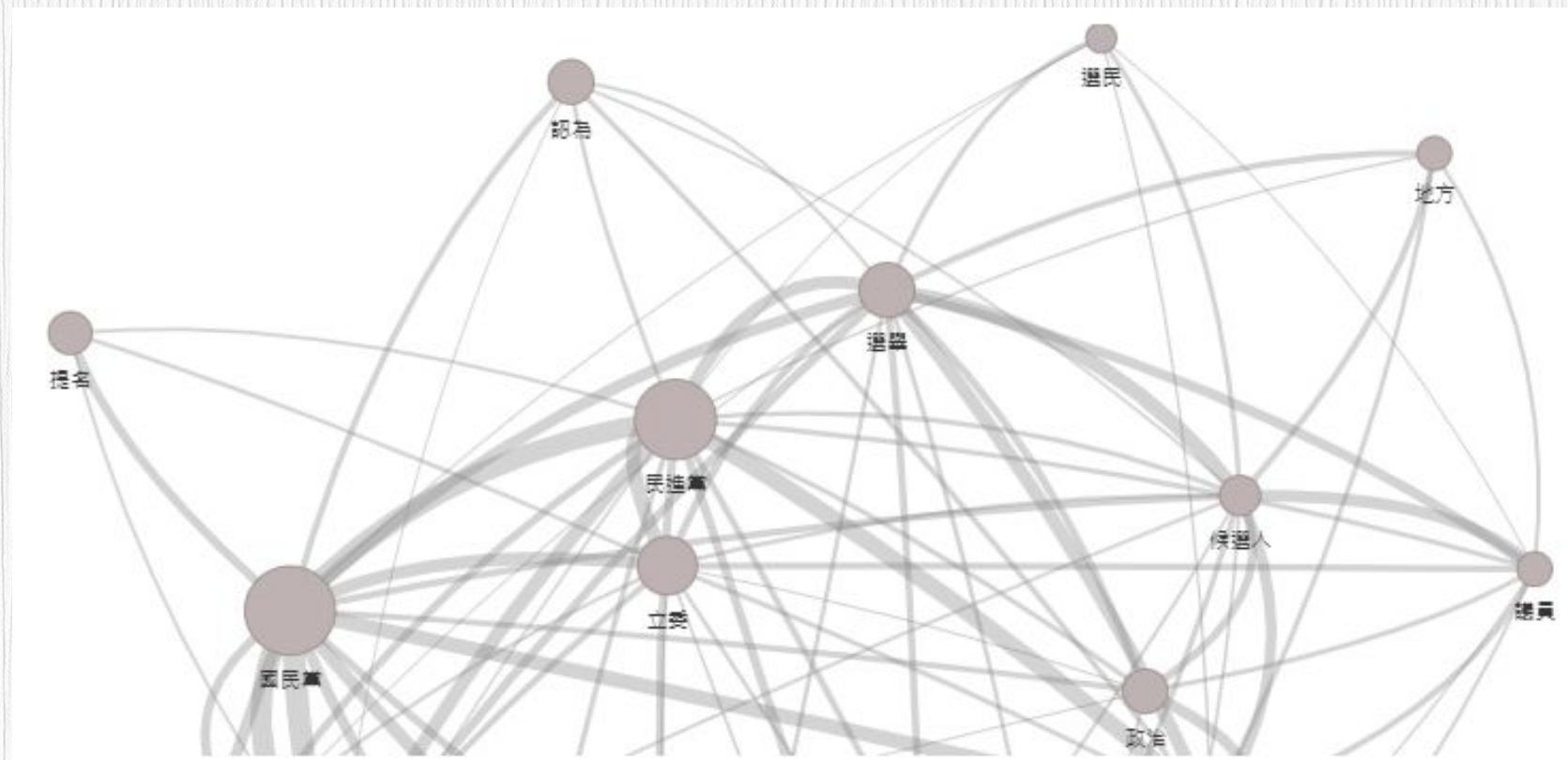
本組發現如分析PTT文章，正向情緒數會大於負向情緒數。



透過盒狀圖(Boxplot)，發現正向及負向情緒數均蠻集中且數值接近。

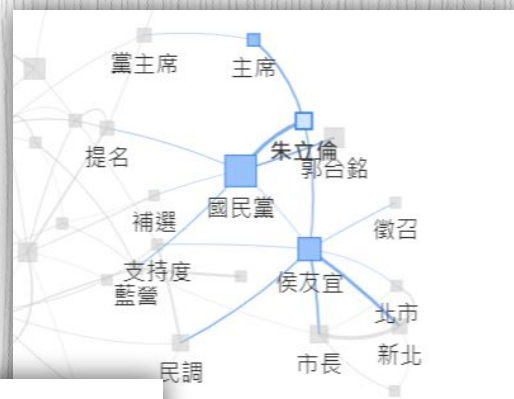
字詞相關性

1.單中心網路圖



字詞相關性

2.字詞網路圖



字詞相關性

3.關聯式文字雲

05

結論

結論與小組討論

● 總統候選人

民進黨與民眾黨候選人已經很明確，國民黨則尚未明朗。

● 總體聲量

總體討論聲量趨勢，自今年 1 月起逐步往上，顯示各政黨已開始佈局且受關注。

● 候選人聲量

柯文哲>侯友宜>賴清德>郭台銘>朱立倫

● KOL情緒分析

AAAB、BIA為負向情緒發文者、Abre、Benegas為正向情緒發文者

● 整體情緒分析

整體情緒分數今年 4 月來到高點，與各政黨積極佈局，營造參選人形象，尚未互相攻訐有關。

● 字詞相關性

透過字詞相關性可了解目前為各政黨黨內提名階段，因此多為尋求黨內支持，並進行民調。





THANK YOU!
