程式設計與資料科學導論
期末專題報告

探討網路民眾對於
2024總統大選中民眾黨的定位

政治所 碩一 **R12322010** 李安妮

生醫電資所 碩二 **R11945020** 黃宣融

生醫電資所 碩二 **R11945022** 葉冠均

目錄

一、研究目的與動機

本研究旨在透過分析網路民眾在討論小黨政治時所使用的關鍵字，深入探討2024總統大選之中民眾黨在選民內心的定位。特別關注近年來民眾黨在網路上的高度討論度，以及其聲勢不斷壯大的現象。這種趨勢對於小黨的定位和實際選民態度可能產生深遠的影響，因此成為本研究的動機。

首先，本研究將透過情感分析技術，探討網路民眾在討論民眾黨時所表達的情緒傾向，包括正面、負面或中立的評價。這有助於我們了解民眾對於民眾黨的感受，以及他們在網路討論中所關注的焦點。進而推測在選民心中，民眾黨的形象與政治立場。

其次，本研究將探討網路生態中民眾黨的優勢是否能有效轉化為實際選票。藉由對比網路輿情分析結果與實際民調數據，我們將評估網路輿情在預測選情方面的準確度，進一步瞭解民眾黨在選民中的真實影響力。此外，我們也將考察民眾黨是否能夠有效地吸引中立選民，或者在網路聲勢上的增長是否可能導致中立選民的流失。

最後，本研究的目的是提供政治分析學者和政治策略制定者一個全面的視角，以更深入地了解小黨在網路生態中的角色和影響。我們希望透過研究結果，為政治行動者提供有價值的建議，協助他們更精準地應對民眾黨在網路空間和實際選民中的定位，並在政治競爭中取得競爭優勢。

二、研究方法

對「PTT政黑板」的關鍵字留言進行情緒分析後，比對當月各黨TVBS民調結果，藉由觀察最終呈現圖形是否有相關性，並探討其原因。
會選擇「PTT政黑板」而不是「八卦版」等討論人數更多的版的原因是因為此版兼具對於政治內容的一致性以及流量，更符合我們研究的目的。

三、研究流程

- ➢ 爬蟲：每個月民調採樣期間隨機挑選100頁進行爬蟲分析。
  a. 7/27~8/24
  b. 8/25~9/26
  c. 9/27~10/24
  d. 10/25~11/26
  e. 11/27~12/12
  f. 近一個月的所有留言資料(共1000頁)
- ➢ 篩選關鍵字：利用Python語法篩選出內容含有我們給定的特定關鍵字的所有留言。
- ➢ 情緒分析：利用snowNLP套件分析我們所篩選出的目標留言，給予評分用來區分正負面情緒以及其程度。
- ➢ 繪圖：以Python繪成長條圖的方式作呈現，用以比對民調的折線圖，觀察兩者之間是否有相關性。

四、Coding

爬蟲

以下程式碼是一個用於爬取 PTT（批踢踢實業坊）討論版的評論的爬蟲程式，使用了 Python 中的 requests 庫和 BeautifulSoup 模組。以下是程式碼的解釋：

1. 引入模組：
   這裡引入了用於網頁解析的 BeautifulSoup 模組，以及用於漂亮輸出的 pprint 模組，urllib.parse 用於處理 URL，time 用於時間相關操作，requests 用於發送 HTTP 請求。

```
from bs4 import BeautifulSoup
import pprint as pprint
import urllib.parse
import time
import requests
```

2. 使用者輸入：

```
index = str(input('想抓取哪個ptt看板？(ex: Movie版請輸入 https://www.ptt.cc/bbs/movie/index.html)：\n'))
pages = eval(input('想抓取幾頁呢？ex: 5：'))

not_exist = BeautifulSoup('<a>(本文已被刪除)</a>', 'lxml').a
```

程式會詢問使用者要抓取哪個 PTT 看板，以及要抓取幾頁的文章評論。在 PTT 看板中，被刪除的文章會有 "(本文已被刪除)" 的提示，這裡使用 BeautifulSoup 創建一個虛擬的被刪除的標籤，以便後續判斷。

3. 定義 get_comments_on_article 函式：

```
index = str(input('想抓取哪個ptt看板？(ex: Movie版請輸入 https://www.ptt.cc/bbs/movie/index.html)：\n'))
pages = eval(input('想抓取幾頁呢？ex: 5：'))

not_exist = BeautifulSoup('<a>(本文已被刪除)</a>', 'lxml').a
```

這個函式使用給定的文章 URL，發送 HTTP 請求並使用 BeautifulSoup 解析 HTML，獲取文章的評論信息，並返回一個評論列表。

4. 定義 get_articles_and_comments 函式：

```
def get_articles_and_comments(url):
    response = requests.get(url)
    soup = BeautifulSoup(response.text, 'lxml')
    articles = []

    for i in soup.find_all('div', 'r-ent'):
        meta = i.find('div', 'title').find('a') or not_exist
        article_url = urllib.parse.urljoin(url, meta.get('href'))
        articles.append({
            'title': meta.getText().strip(),
            'url': article_url,
        })

    return articles
```

這個函式用於爬取指定頁面的文章列表及其相關的評論，返回一個包含文章信息的列表。

5. 定義 get_pages 函式：

```
def get_pages(num):
    page_url = index
    all_articles = []

    for j in range(num):
        response = requests.get(page_url)
        articles = get_articles_and_comments(page_url)
        for article in articles:
            article_comments = get_comments_on_article(article['url'])
            article['comments'] = article_comments
            all_articles.append(article)
        page_url = urllib.parse.urljoin(index, BeautifulSoup(response.text, 'lxml').find('div', 'btn-group-paging').find_all('a', 'btn')[1].get('href'))

    return all_articles
```

這個函式用於爬取指定數量的頁面上的所有文章及其評論，返回一個包含所有文章信息的列表。

6. 執行爬蟲並儲存資料:

```
data = get_pages(pages)

for article in data:
    pprint.pprint(article)

csv_or_not = input('輸入 y 以匯出成csv檔，輸入其他結束程式: ')

if csv_or_not == 'y':
    board = index.split('/')[-2]
    csv = open('./ptt_%s版_前%d頁_評論.csv' % (board, pages), 'a+', encoding='utf-8')
    csv.write('標題,評論內容,評論時間,評論者,\n')
    for article in data:
        for comment in article['comments']:
            csv.write(article['title'] + ',' + comment['push_content'] + ',' + comment['push_time'] + ',' + comment['push_user'] + ',\n')
    csv.close()
    print('CSV檔案已儲存在您的資料夾中。')
else:
    quit()
```

爬蟲運行，並將爬取到的文章資訊顯示給使用者，然後詢問是否要將爬取到的資料儲存為 CSV 檔案。

總體而言，以上程式碼進行了一個簡單的 **PTT** 討論版爬蟲，使用者可以指定要爬取的看板和頁數，程式會回傳每篇文章的標題、**URL**，以及相關的評論內容。

情緒分析**snowNLP**

1. 使用**python snowNLP**情緒分析套件

   將爬蟲存取的csv檔匯入成dataframe，並且透過df.iloc函式抓取標題與評論內容。

```
! pip install snownlp
from snownlp import SnowNLP
```

```
#讀取資料
import pandas as pd
df = pd.read_csv('ptt_HatePolitics版_前100頁_12月.csv')
title = df.iloc[:, [0, 1]]
title.rename(columns={'標題': 'target', '評論內容': 'comment'}, inplace=True)
print(title)
```

```
                                target                comment
0                 [轉錄] 且看綠官們如何面對禮義廉恥             : 中國的恥在於台灣民主
1                 [轉錄] 且看綠官們如何面對禮義廉恥           : 壞人的恥 在於被警察抓
2                 [轉錄] 且看綠官們如何面對禮義廉恥       : 無恥之人 慣於以立場定恥與是非
3                 [轉錄] 且看綠官們如何面對禮義廉恥    : 讀了廉恥就有廉恥?不讀廉恥就不知廉恥?
4                 [轉錄] 且看綠官們如何面對禮義廉恥              : 超買不是走私!
...                                ...                   ...
85904  Re: [新聞] 高虹安告周玉蔻影射「與有婦之夫交往」不   : 看新聞報導的論述，好像是有重疊到是之後
85905  Re: [新聞] 高虹安告周玉蔻影射「與有婦之夫交往」不                 : 才離婚
85906  Re: [新聞] 高虹安告周玉蔻影射「與有婦之夫交往」不          : 沒離婚前 就是小三阿
85907  Re: [新聞] 高虹安告周玉蔻影射「與有婦之夫交往」不  : 李 就不敢講何時離婚啊 公布就一刀斃命w
85908              【討論】 阿北是否在用身體實踐科學                : 請正名柯學

[85909 rows x 2 columns]
```

2. 關鍵字篩選

分別設定三種政黨的特定關鍵字進行評論篩選，其中又對民眾黨進行雙層篩選，第一層使用民眾黨關鍵字對標題篩選，第二層使用小黨關鍵字對評論進行篩選，進而分析民眾黨在選民心中的定位。

- 國民黨**(KMT)**關鍵字：
  blue_key =['侯友宜','國民黨','侯侯做代誌', '侯', '警察','侯侯','藍','在野','侯粉','親中','馬英九','朱立倫','新北','趙少康','侯康','藍營','KMT']
- 民進黨**(DPP)**關鍵字：
  green_key = ['賴清德','民進黨','美德配', 'ㄉㄨㄚˋ', '賴','賴蕭','蕭美琴','執政','綠','反中','蔡英文','ㄘㄨㄚˋ','蔡','綠營','DPP']
- 民眾黨**(TPP)**關鍵字：
  keywords = ['柯文哲','民眾黨','白色力量', 'kp','柯P','阿北','柯','白','柯粉','KP','黃國昌','國蔥','蔥','TPP']
- 小黨關鍵字：
  keywords1 = ['小黨','席位','席次','國會','15%','第三政黨','立委','少數','關鍵','15趴','10趴','11趴','政黨票','第三','三大']

```
#民眾黨
#篩選標題中的小黨關鍵字
#篩target
keywords = ['柯文哲','民眾黨','白色力量', 'kp', '柯P','阿北','柯','白','柯粉','KP','黃國昌','國蔥','蔥']
filtered_df = title[title['target'].str.contains('|'.join(keywords), case=False)]
print(filtered_df)
```

```
                                target          comment
24          [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨    : 區域別做夢啦 整天打雞血
25          [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨       : 蔡璧如…....
26          [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨       : 新竹民眾黨又沒提名
27          [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨        : 柯妹哪有機會啦?
28          [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨     : 國民黨都不敢選了才給你
...                                  ...               ...
85878   Re: [討論] 快訊! 柯粉男神黃國昌下午要打賴老家!      : 黨當可以監督啊!!!
85879   Re: [討論] 快訊! 柯粉男神黃國昌下午要打賴老家!  : 日本民眾也會監督在野黨啊。
85880   Re: [討論] 快訊! 柯粉男神黃國昌下午要打賴老家!       : 南投國民黨是執政黨
85881   Re: [討論] 快訊! 柯粉男神黃國昌下午要打賴老家!           : 綠共急了
85908              【討論】 阿北是否在用身體實踐科學          : 請正名柯學

[33165 rows x 2 columns]
```

```
#篩comment
keywords1 = ['小黨','席位','席次','國會','15%','第三政黨','立委','少數','關鍵','15趴','10趴','11趴','政黨票','第
filtered_df1 = filtered_df[filtered_df['comment'].str.contains('|'.join(keywords1), case=False)]
print(filtered_df1)
```

```
                          target                    comment
39      [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨        : 笑死  說自己墨綠 然後立委反綠
65      [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨     : 其實有稍微跟過那些立委候選人發言的就知
73      [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨   : 統 政黨票雙崩耶 同時也代表TMD內參完全
97      [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨  : 不分區也就34席 你拿15席粗估政黨票也要拿
107     [新聞] 柯文哲民調雪崩「不分區難保5席」？爆黨      : 柯黨區域立委候選人魅力不足，名氣不夠,
...                                ...                   ...
85327          [討論] 柯現在是不是要祈禱兩件事       : 民眾黨關不關鍵，還是沒有他的票數重要
85512   [討論] 快訊! 柯粉男神黃國昌下午要打賴老家!        : 萬里也是黃國昌當立委時的選區，上
85547   [討論] 快訊! 柯粉男神黃國昌下午要打賴老家!        : 不分區立委不好好提自己政見是在?
85763   [新聞] 柯文哲如何說服入黨? 黃國昌秀2人會議紀     : 第三點就是服貿 直說就好 不用堆砌文字
85839    [討論] 李登輝: 柯文哲實在說起來，思想有問題        : 萬里也是黃國昌當立委時的選區，上

[414 rows x 2 columns]
```

3. 篩選評論欄
   將經由特定關鍵字篩選後的評論列印出來:

```
#篩選評論欄
key_comment = filtered_df1.iloc[:, 1]
print(key_comment)
```

```
39              : 笑死   說自己墨綠 然後立委反綠
65           : 其實有稍微跟過那些立委候選人發言的就知
73         : 統 政黨票雙崩耶 同時也代表TMD內參完全
97       : 不分區也就34席 你拿15席粗估政黨票也要拿
107          : 柯黨區域立委候選人魅力不足，名氣不夠，
                    ...
85327         : 民眾黨關不關鍵，還是沒有他的票數重要
85512           : 萬里也是黃國昌當立委時的選區，上
85547           : 不分區立委不好好提自己政見是在?
85763        : 第三點就是服貿 直說就好 不用堆砌文字
85839           : 萬里也是黃國昌當立委時的選區，上
Name: comment, Length: 414, dtype: object
```

4. 使用**snowNLP**進行情緒分析

```
comment = str(key_comment)
s = SnowNLP(comment)
#所有句子的情緒分析
print(s.sentiments)
```

```
1.3620768213318257e-07
```

snowNLP套件進行情緒分析後得到的數值介於**0～1**，越接近**1**表示正面情緒，越接近**0**表示負面情緒。

圖表呈現

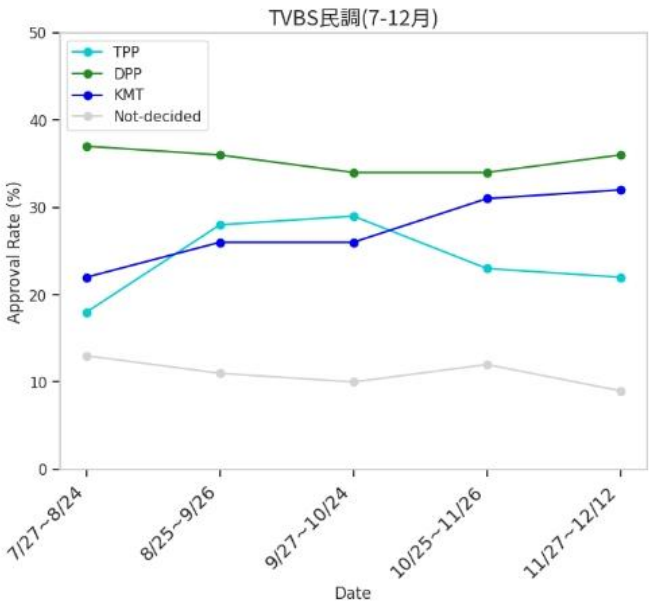
1. 長條圖

```
#長條圖
#未篩選鄉民對於選舉的情緒分析
sns.set(style="whitegrid")
plt.figure(figsize=(8, 6))
sns.barplot(x='Date', y='Score of sentiment', data= score_df,hatch='//',color='peru')
plt.title('整體鄉民對於選舉之情緒分析', fontproperties='Taipei Sans TC Beta', fontsize=16)
plt.xlabel('Date', fontsize=16)
plt.ylabel('Score of sentiment (log scale)', fontsize=16)
plt.xticks(rotation=45, ha='right', fontsize=14)
plt.yticks(fontsize=14)
bar_width = 0.2
plt.ylim(0, 10)
plt.show()
```

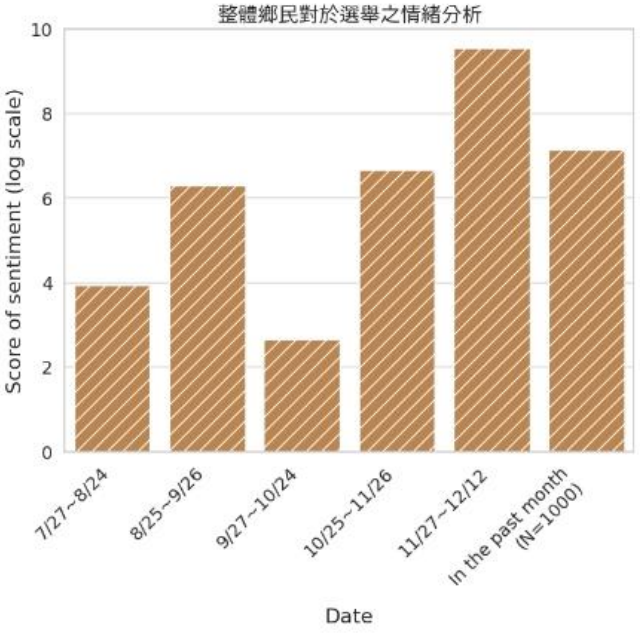
X軸為民調採樣的月份區間
Y軸為所篩選留言的正負情緒評分(以log scale的方式作呈現，因此長條圖的高度越高所代表的意義就是留言內容的情緒越負面)

五、結果

1. TVBS民調 (本研究結果所比對的實際民調)

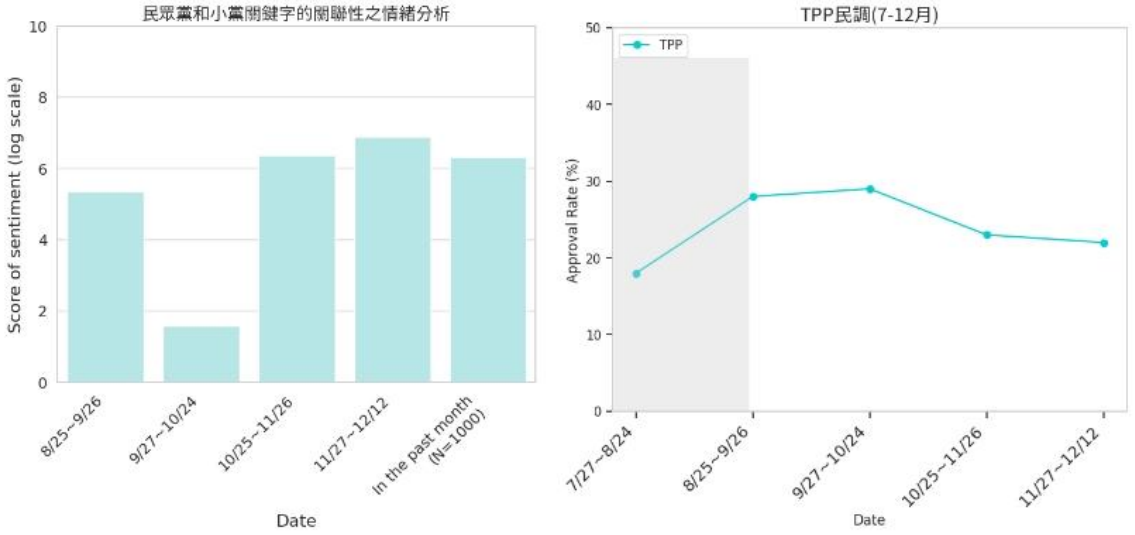


2. 整體政黑板鄉民對於2024總統大選之情緒分析(無關鍵字篩選)



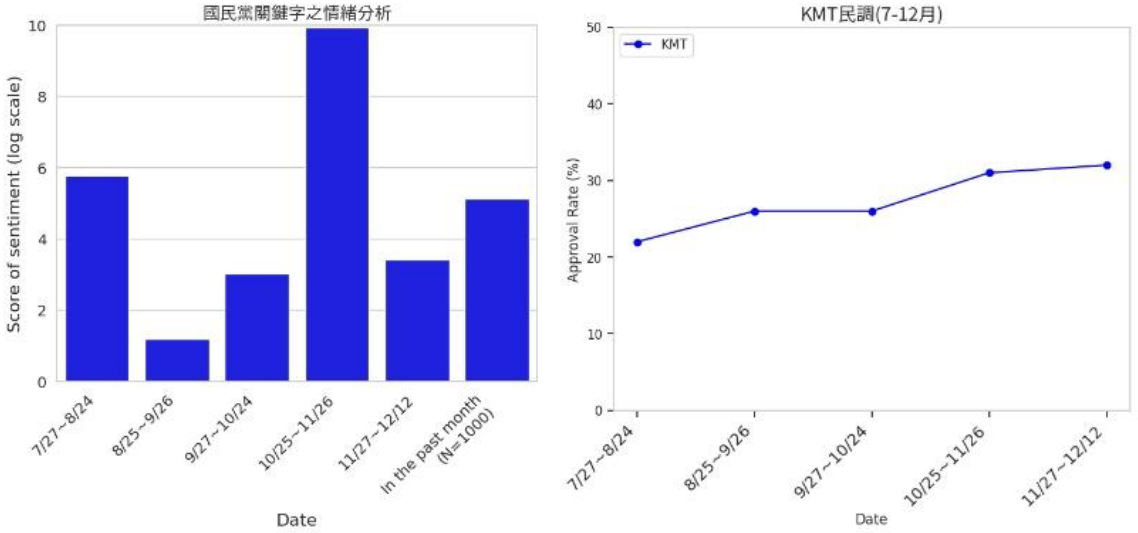
可以觀察出近三個月鄉民留言對於總統大選的情緒越來越高漲也越負面

3. 民眾黨和小黨關鍵字的關聯性之情緒分析(雙層關鍵字篩選)



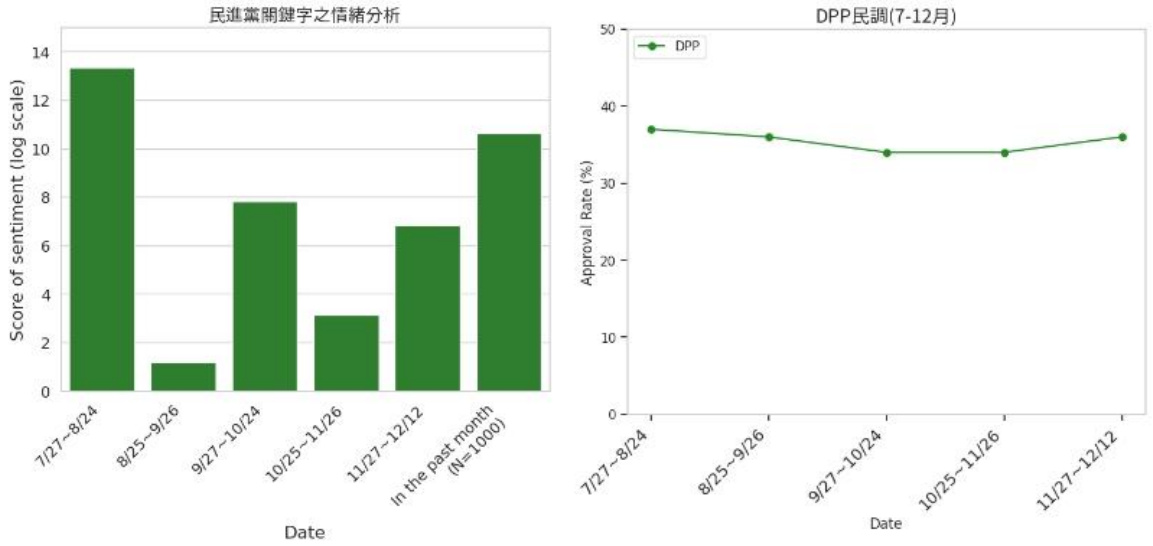
由於我們所設定的民眾黨以及小黨兩層關鍵字的留言篩選在7~8月並沒有交集，也因此那部分不予以比較。自8月起至12月的民調漲跌趨勢(右圖)與左圖情緒分析具有一定的相關性(情緒愈加負面對應民調逐漸下滑)。

4. 國民黨關鍵字的關聯性之情緒分析(單層關鍵字篩選)



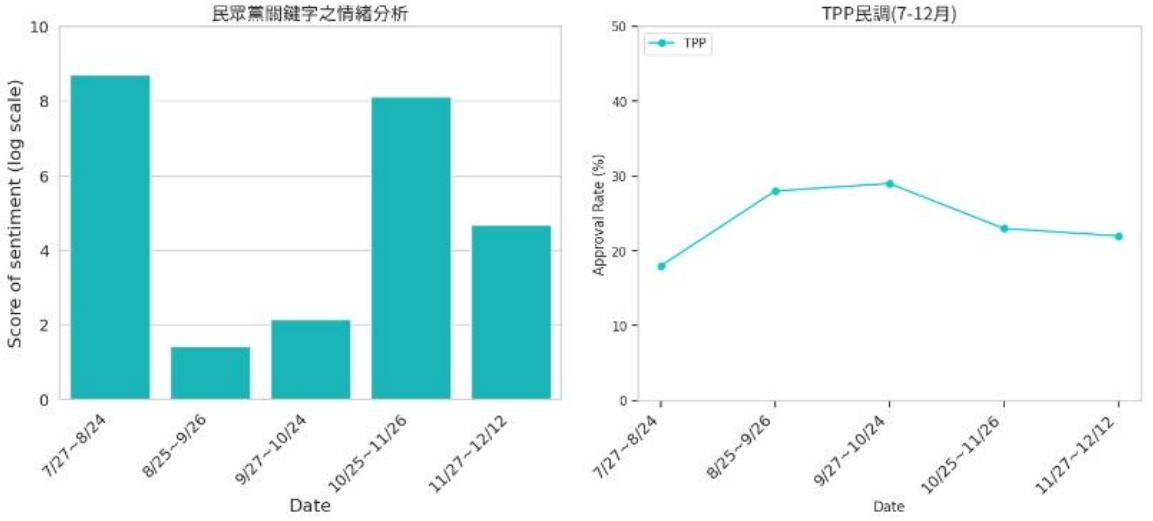
分析結果與國民黨民調(右圖)結果比對後並沒有明顯的關聯性(情緒分數的變動並沒有影響民調)

5. 民進黨關鍵字的關聯性之情緒分析(單層關鍵字篩選)



分析結果與民進黨民調(右圖)結果比對後並沒有明顯的關聯性(情緒分數的變動並沒有影響民調)

6. 民眾黨關鍵字的關聯性之情緒分析(單層關鍵字篩選)



分析結果與民眾黨民調(右圖)結果比對後可以發現每個月分留言情緒分數的起伏與其民調結果有相當高程度的相關，情緒越負面(長條圖長度越高)，當月的民調則會相應的下降，反之亦然。

六、結論

本研究旨在探討網路輿論與實際民調之間的關聯性，以及其對民眾黨在選民心中的定位和選舉結果可能產生的影響。經過研究假設的檢驗，我們得出以下結論。

首先，對於研究假設中的第一點，我們發現網路輿論與民調結果之間的正相關性並不明顯。儘管在網路輿論中存在著對民眾黨的關注，但這並未直接轉化為與實際選民態度的一致性。這可能意味著民眾黨在網路上的形象相對模糊，未能在選民心中明確定位為小黨，使其在選舉中獲利的機會降低。

其次，我們觀察到藍綠兩大政黨在網路情緒聲量與實際民調之間存在脫鉤的現象。這顯示網路輿論對於藍綠陣營的情感表達並不完全反映實際選民的態度。這可能是因為網路平台上的言論更容易受到極端意見的影響，而實際選民的態度可能更加多元和複雜。

最後，我們發現在特別事件發生時，網路輿論的情緒聲量會有明顯的改變。例如：11/24發生藍白合君悅事件，藍白合議題於整個11月便不斷被討論，因此我們也觀察到十一月網路聲量有明顯變化。從這表明特殊事件對於網路輿論的形成和表達具有重要影響，可能導致民眾黨在特定時刻的形象變動，進而影響其在選民中的評價。

總體而言，本研究強調了網路輿論與實際民調之間複雜且不一致的關係。在未來的研究中，有必要進一步深入探討網路輿論對於小黨的定位和選情的實際影響，以提供更全面的理解，並為政治行動者提供更精準的策略建議。此外，我們也希望在未來有機會將收集到的數據進行更精確的統計分析，包括迴歸分析、顯著性等等。

七、參考資料

- **https://github.com/isnowfy/snownlp**
- **TVBS民意調查中心**
- **pttHatePolitics**

八、組員分工表

姓名	學號	分工內容
李安妮	**R12322010**	1. 研究目的與動機 2. Coding – 爬蟲 (資料蒐集) 3. 結論 4. 提案口頭報告 5. 期末口頭報告 6. 政治學顧問
黃宣融	**R11945020**	1. Coding – 情緒分析 2. 結果 – 國民黨關鍵字的關聯性之情緒分析 民進黨關鍵字的關聯性之情緒分析 民眾黨關鍵字的關聯性之情緒分析 3. PPT美編與排版 4. 期末口頭報告 5. debug python程式
葉冠均	**R11945022**	1. 研究方法 2. 研究流程 3. Coding – 圖表呈現 4. 結果 – TVBS民調 整體政黑板鄉民對於2024總統大選之情緒分析 民眾黨和小黨關鍵字的關聯性之情緒分析 5. PPT封面設計 6. 期末口頭報告