

智慧社群洞察與熱門話題分析 平台

第10組
林君瑋、羅振豪



-



Introduction

1. 背景與動機

- 社群媒體使用量日益增長，各種議題發酵速度加快。
- 企業與研究單位需要即時了解社群趨勢與熱門話題。

2. 專案目標

- 開發「智慧社群洞察與熱門話題分析平台」，即時掌握網路聲量與趨勢。
- 協助決策者洞察消費者觀點、關鍵議題，便於更快做出決策。

3. 重要性與應用

- 有助於行銷策略制定、危機應對、產業動態追蹤…

related work

現有社群分析工具概述（精選 3 種）



功能與定位

主要提供多社群平台整合、排程發布、基礎數據分析

優勢（依官方與第三方評測歸納）

多帳號集中管理、操作介面相對友善

不足

深度語意分析或情緒偵測較為薄弱，偏向社群「管理」與「排程」

related work

現有社群分析工具概述（精選 3 種）



功能與定位

聚焦「品牌監測」與「社群聆聽」，可支援多語言大規模數據處理

優勢

多平台整合度高，圖表與儀表板呈現豐富

不足

系統設定複雜、付費成本較高，對中文（尤其繁體）的情感分析精確度需進一步測試

related work



現有社群分析工具概述（精選 3 種）

功能與定位

強調「內容行銷洞察」，分析關鍵字或主題的熱門程度與分享量

優勢

可快速找出在各社群上最受歡迎的內容，對行銷人員規劃內容策略相當有利

不足

偏重「社群分享量」與「話題熱度」數據，不具備深度語意或情緒偵測功能

related work

工具比較表

工具 / 指標	多平台管理	中文情感分析	即時更新	自動產出洞察	費用 / 複雜度
Hootsuite	有(主流社群)	弱(僅關鍵字層級)	中(排程可行)	無(需人工分析)	中等 / 操作簡單
Brandwatch	有(多來源)	有(但中文精度待測)	高(近即時)	有(可視化報告)	偏高 / 設定較複雜
BuzzSumo	部分(重內容分享)	無(不做情感判斷)	中(非即時)	無(以列表為主)	中等 / 上手容易

資料來源：官網功能說明、第三方評測平台（G2、Capterra）以及使用者分享。

註：圖片由專案團隊製作

related work

主要痛點歸納

1. 中文語意與深度情感分析有限：
 - 大多數工具以英文語料為主，對繁體中文網路用語的準確度不高。
2. 即時多平台整合不足：
 - 少數工具能即時更新，但未必能整合論壇、新聞等多元來源。
3. 自動產出洞察報告較少：
 - 大多提供數據可視化，深度議題脈絡剖析仍需仰賴人工。

Proposed scheme

方案概念

- **目的：**提供一個基於 PTT 即時關鍵字分析的平台，幫助使用者快速掌握熱門話題的情感趨勢、討論熱度及用戶行為模式。
- **原理：**透過即時爬蟲與分析模組，將關鍵字相關文章進行處理、分析並以視覺化方式呈現結果。

Proposed scheme

主要特色

1. 即時響應

- 使用者只需輸入關鍵字，平台即可即時爬取 PTT 相關內容，無需等待長時間計算。

2. 多維度分析

- 涵蓋情感分析、趨勢預測及用戶分群，協助掌握話題整體熱度與細節。

3. 可視化結果

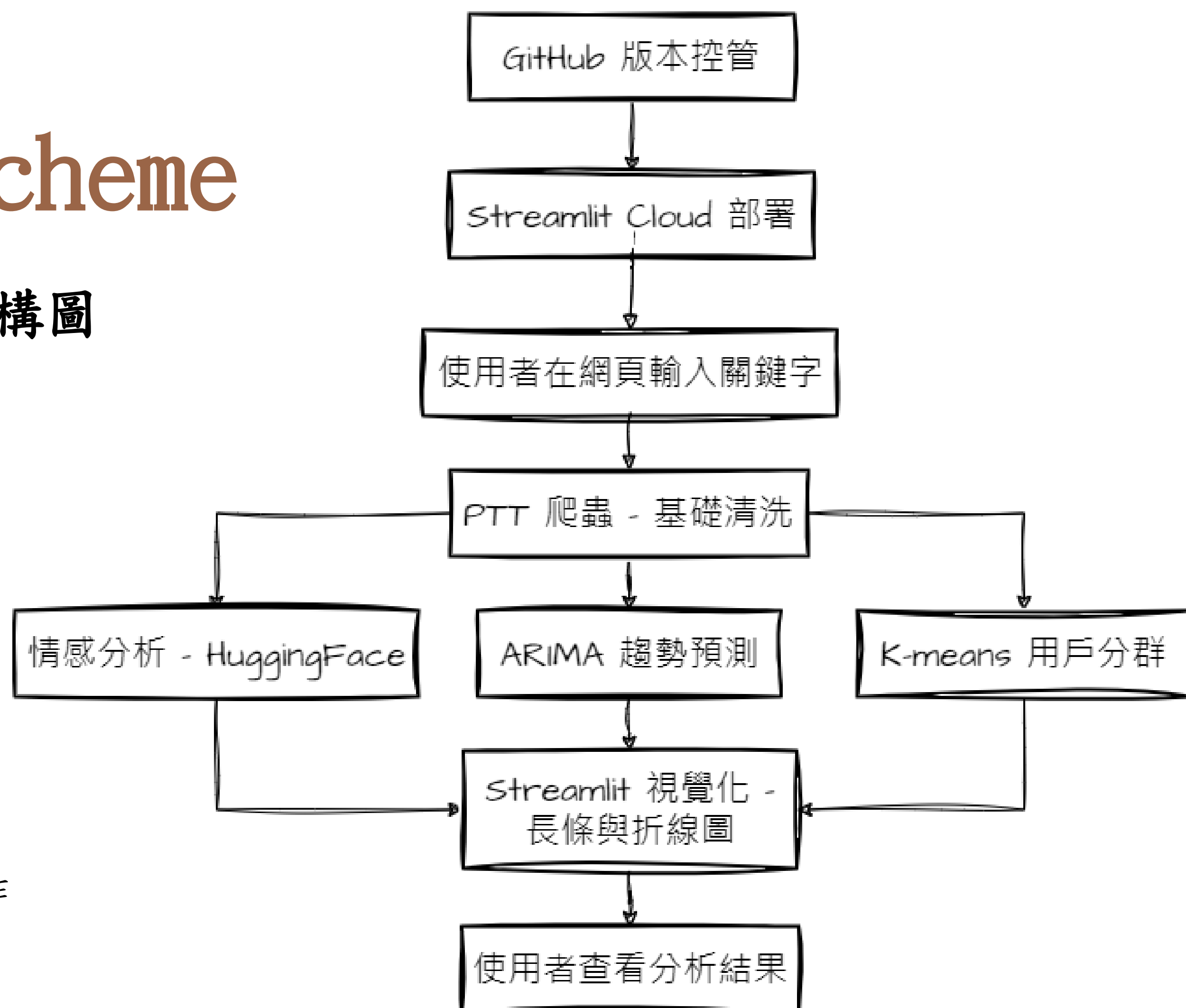
- 結果以長條圖、折線圖等直觀呈現，方便使用者快速解讀。

4. 雲端部署

- 前端部署於 Streamlit Cloud，方便多用戶隨時使用；程式碼版本控管集中於 GitHub。

Proposed scheme

方案流程 / 系統架構圖



註：圖片由專案團隊製作

Proposed scheme

使用技術

1. 爬蟲技術

- 透過 requests 與 BeautifulSoup，爬取 PTT 文章標題、內文與相關元數據。

2. 文本處理與清洗

- 使用正則表達式移除雜訊與無效元數據，保留核心內容進行分析。

3. 情感分析模組

- 使用 Hugging Face 的 roberta-base-finetuned-jd-binary-chinese 模型，對文章進行情感分類。

4. 趨勢預測模組

- 使用 ARIMA 時間序列模型，分析歷史數據並預測未來幾期的熱度走勢。

5. 用戶行為分群

- 利用 K-means 分群演算法，結合 PCA 降維，分析用戶發文與互動特徵。

6. 前端部署與可視化

- 前端採用 Streamlit，以長條圖、折線圖等形式即時呈現分析結果。

實作功能說明

實作環境與工具

Python 3.12

優點：

1. 廣泛的庫支持：擁有豐富的函式庫（如 requests, BeautifulSoup, transformers, statsmodels, sklearn），滿足爬蟲、NLP、數據分析等多種需求。
2. 開發效率高：語法簡潔，適合快速開發與測試。
3. 社群支持：Python 擁有活躍的開發者社群，遇到問題可以快速找到解決方案。



實作功能說明

實作環境與工具

Streamlit Cloud

優點：

1. **快速部署**：簡單一行指令即可將應用部署到雲端，無需額外伺服器配置。
2. **即時互動**：提供即時交互式的用戶界面，適合進行分析結果展示和使用者操作。
3. **免費方案**：適合中小型專案測試和初期開發。
4. **內建支持可視化**：輕鬆生成圖表與數據展示，減少額外設計工作量。



Streamlit

實作功能說明

實作環境與工具

GitHub

優點：

1. 版本控管：每次程式碼更新都會被記錄，方便回溯與管理歷史版本。
2. 多人協作：團隊成員可同時參與開發，減少衝突，提升效率。
3. 開源共享：方便公開分享專案進度，並能獲取社群建議或參與。



實作功能說明

情感分析

功能概述

- 對 PTT 爬取的文章及回文進行正負向情感分類，幫助掌握討論內容的整體情緒傾向。

技術細節

- 使用模型：uer/roberta-base-finetuned-jd-binary-chinese
 - 由 Hugging Face 提供，針對中文文本進行 fine-tune 的情感分析模型。
- 函式庫：transformers
 - 使用 Hugging Face 的 pipeline 方法，快速調用預訓練模型進行推論。
- 輸出結果：每篇文章的 label (POSITIVE/NEGATIVE) 和 score (情感分數)。

實作功能說明

情感分析

執行流程

1. 輸入文本：爬取到的 PTT 文章內容。
2. 文本截斷：使用 AutoTokenizer 將過長文本截斷為 512 字元以內。
3. 情感分類：對每篇文章進行情感分析，輸出正負向標籤與信心分數。

應用價值

- 輿情監測：快速定位正面或負面的高影響力討論，輔助危機管理或品牌評估。
- 數據可視化：將結果以長條圖呈現，清晰展現正負情緒的分布比例。

功能選單

選擇功能

- ☐ 首頁
- ☒ 情感分析
- ☐ 趨勢預測
- ☐ 用戶行為分析
- ☐ 個性化推薦
- ☐ 文本摘要

情感分析模組

此模組將分析 PTT 文章的情感傾向（正面、中立、負面）。

請輸入關鍵字：

AI

搜尋期間（月）

3

開始分析

正在抓取文章內容...（最多抓取 100 篇文章）

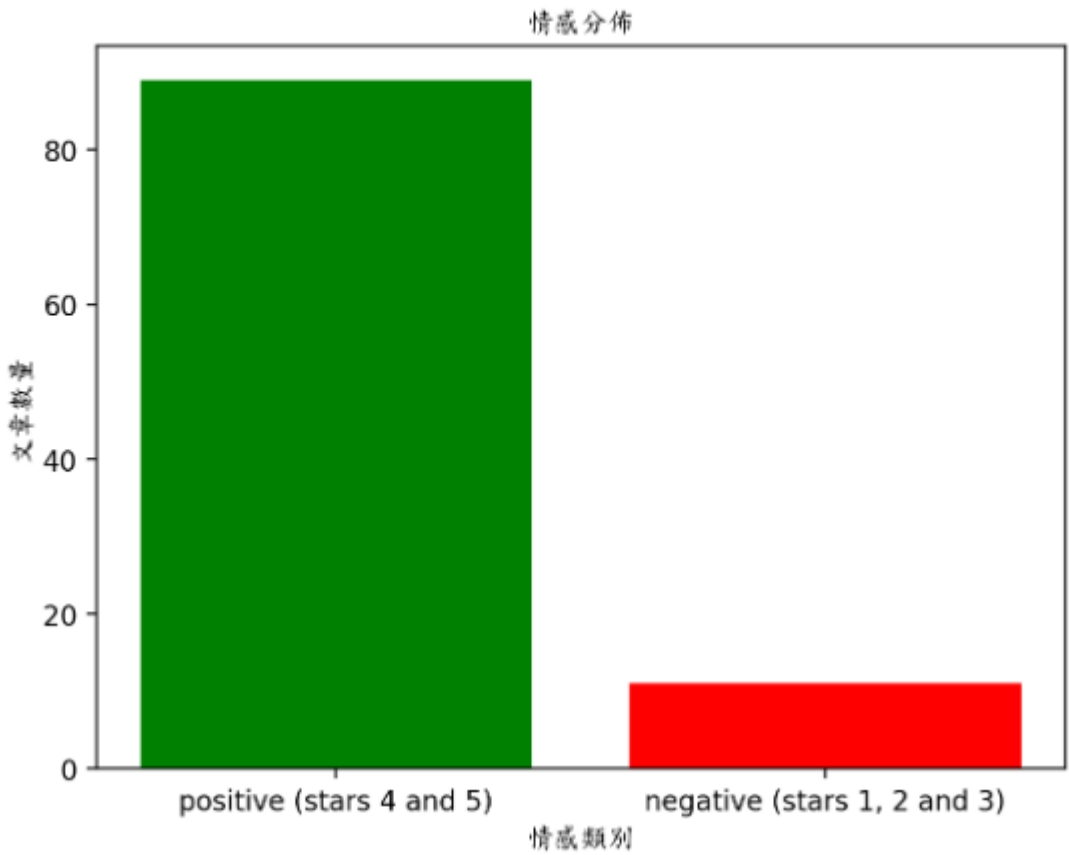
總共抓取到 100 篇文章

正在分析情感...

情感分佈統計：

positive (stars 4 and 5): 89 篇

negative (stars 1, 2 and 3): 11 篇



展示各情感類別的文章：

positive (stars 4 and 5) 類文章（展示三篇）：

文章 1：作者lovecomics 獅子頭三叔看板Gossiping標題問卦 AI機器人 真的能當老婆時間Tue Jan 7 140520 2025 鄉民老愛說AI機器人當老婆的時代就快到了 可是AI這麼聰明 ...

```
{
  "label": "positive (stars 4 and 5)"
  "score": 0.7927428483963013
}
```

[查看原文](#)

註：圖片由專案團隊製作

實作功能說明

趨勢預測

功能概述

- 以時間序列模型分析歷史文章數據，預測未來幾個月的討論熱度，判斷話題是否可能持續發酵。

技術細節

- **模型選擇**：ARIMA (Auto-Regressive Integrated Moving Average)
 - 適合分析時間序列數據的趨勢與週期性特徵。
- **函式庫**：statsmodels.tsa.arima.model
 - 提供專業的統計與建模工具，用於擬合與預測時間序列。
- **輸出結果**：
 - 歷史文章數量的折線圖。
 - 未來數月的熱度預測（每月文章數）。

實作功能說明

趨勢預測

執行流程

1. 數據準備：按月統計文章數量，生成時間序列數據。
1. 模型訓練：基於歷史數據擬合 ARIMA 模型，確定最佳參數（ p , d , q ）。
2. 趨勢預測：使用訓練好的模型進行 6 個月的熱度預測。
3. 結果可視化：以折線圖形式顯示歷史與預測數據。

應用價值

- 話題監測：提前發現可能爆發的熱門話題，為行銷或公關活動提供依據。
- 預測未來走勢：幫助使用者制定資源分配或策略調整計劃。

功能選單

選擇功能

- ☐ 首頁
- ☐ 情感分析
- ☒ 趨勢預測
- ☐ 用戶行為分析
- ☐ 個性化推薦
- ☐ 文本摘要

趨勢預測模組

此模組將分析熱門關鍵字的趨勢並進行未來預測。

請輸入關鍵字進行趨勢分析：

AI

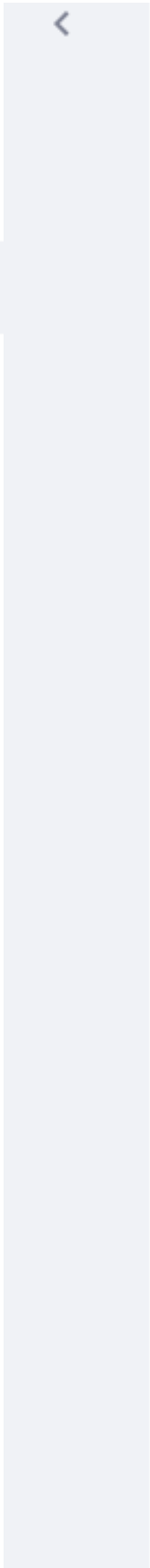
開始趨勢預測

正在爬取數據...

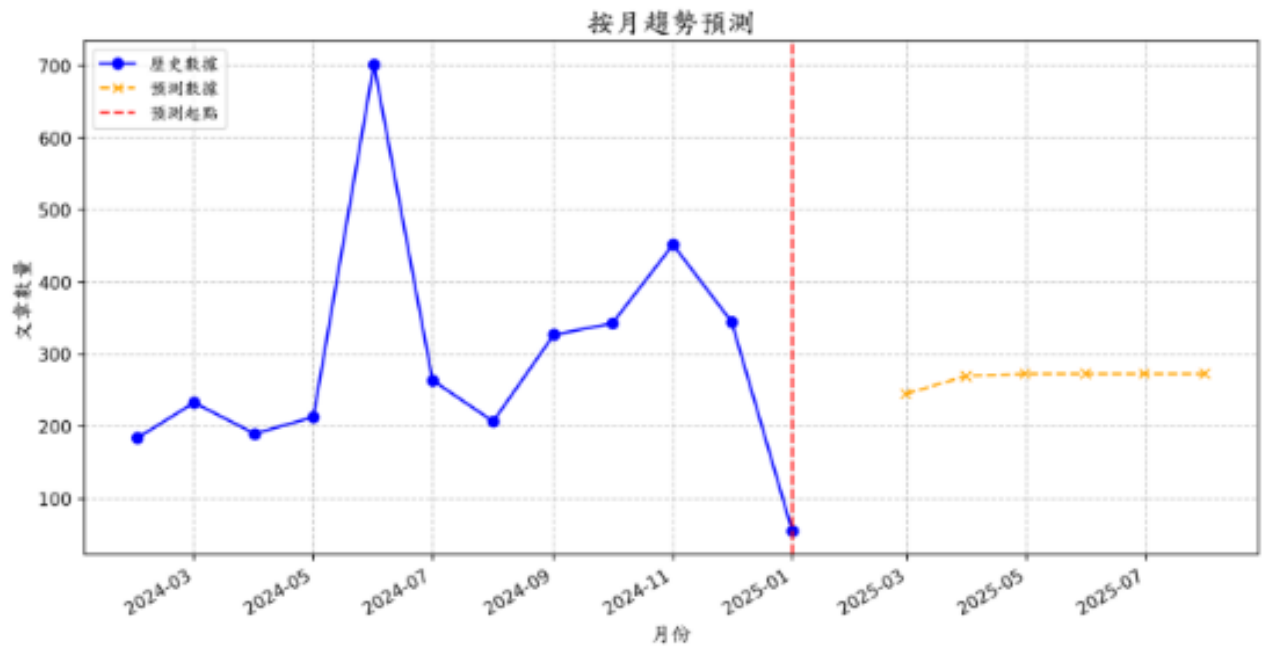
爬取完成！數據範圍：2024-01-01 00:00:00 到 2025-01-01 00:00:00

數據抓取成功！

	month	value
12	2024-01-01 00:00:00	106
11	2024-02-01 00:00:00	184
10	2024-03-01 00:00:00	233
9	2024-04-01 00:00:00	190
8	2024-05-01 00:00:00	213
7	2024-06-01 00:00:00	702
6	2024-07-01 00:00:00	264
5	2024-08-01 00:00:00	207
4	2024-09-01 00:00:00	327
3	2024-10-01 00:00:00	343



未來 6 個月的趨勢預測：



註：圖片由專案團隊製作

實作功能說明

用戶行為分群

功能概述

- 基於 PTT 用戶的發文與回文行為數據，將用戶劃分為不同群體，分析其活躍度與行為模式。

技術細節

- 演算法：K-means 分群
 - 利用聚類技術將用戶分為不同行為特徵的群體。
- 數據降維：PCA (Principal Component Analysis)
 - 將高維數據壓縮到 2 維，便於視覺化展示。
- 函式庫：sklearn
 - 提供高效的 K-means 與 PCA 模組。
- 輸出結果：每個用戶的分群標籤 (Cluster ID)、行為特徵分布圖。

實作功能說明

用戶行為分群

執行流程

1. 數據準備：爬取 PTT 用戶的發文數與回文數，計算特徵統計（如總發文量、總回文量）。
2. 數據標準化：對數據進行標準化處理，消除特徵單位差異的影響。
3. 聚類分析：使用 K-means 將用戶分為 3~5 群，根據用戶特徵劃分群體。
4. 降維與視覺化：通過 PCA 將聚類結果壓縮到 2 維，並以散佈圖形式展示。

應用價值

- 高效用戶洞察：快速識別高活躍用戶（核心參與者）與低互動用戶（潛在觀察者）。
- 行銷策略優化：針對不同用戶群制定差異化的溝通與互動策略。
- 輿情分析支援：發現對事件推波助瀾的關鍵用戶，輔助危機處理。

功能選單

- 選擇功能
- 首頁
 - 情感分析
 - 趨勢預測
 - 用戶行為分析
 - 個性化推薦
 - 文本摘要

用戶行為分析模組

此模組將分析 PTT 用戶發文行為模式與特性。

請輸入關鍵字：

AI

搜尋期間 (月)

3 - +

開始分析

已爬取 551 篇文章，目前爬取到的文章時間：2024-10-21，搜尋範圍最早日期：2024-10-07

停止爬取

正在抓取 PTT 數據...

已完成爬取 551 篇文章，正在進行分析...

正在進行數據預處理 (包括極端值處理) ...

處理前的回文數摘要：

mean	32.4175
std	50.4547
min	0
25%	7
50%	15
75%	32
max	266.2323

數據預處理完成，開始用戶分群分析...

分群數據摘要

	cluster	avg_post_count	avg_reply_count	user_count
0	0	1.146	25.2066	363
1	1	3.72	75.88	25
2	2	1.4	227.1673	30

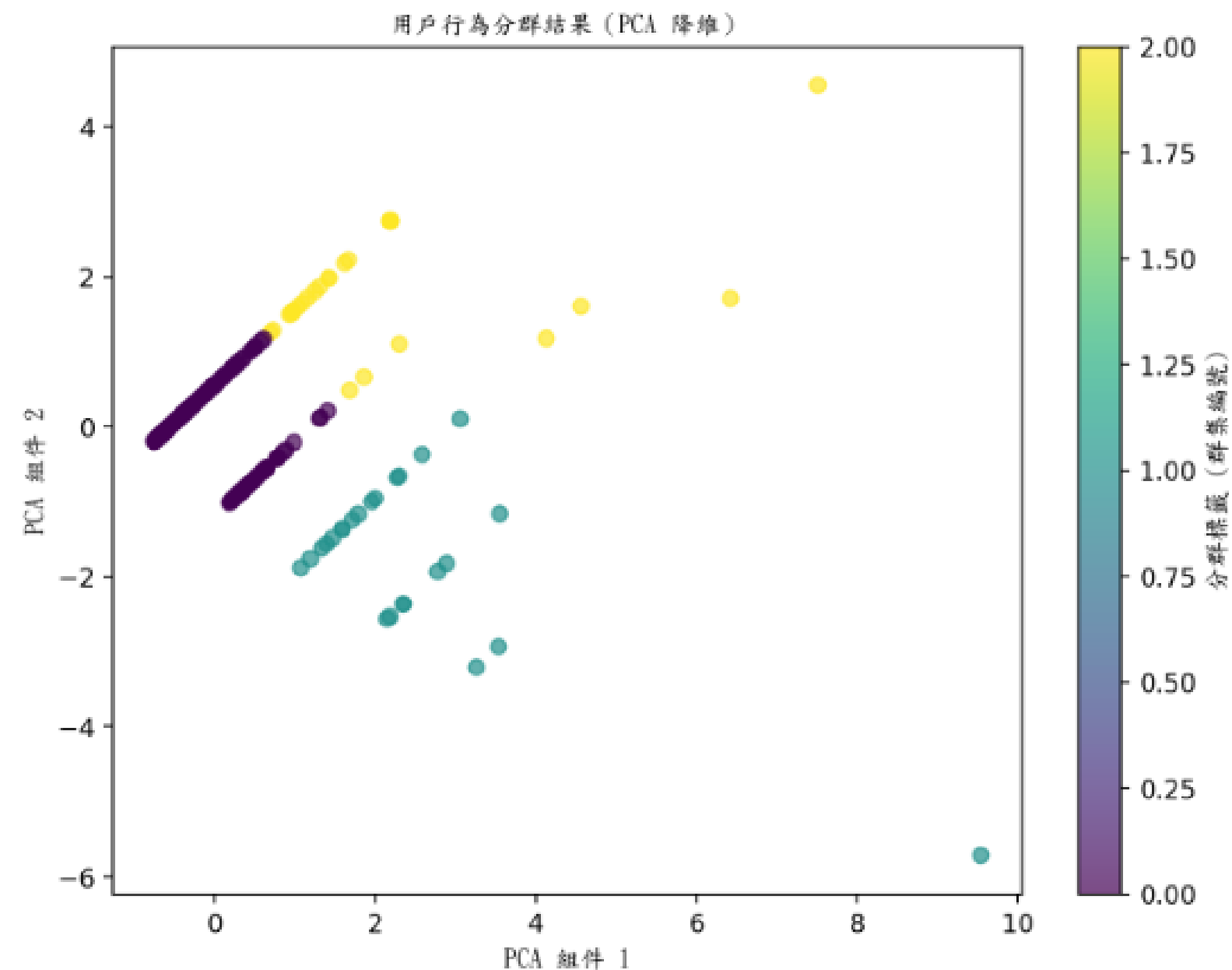
分群數據的應用價值

- 群 0：以普通用戶為主，活躍度較低，可能是大部分用戶的行為模式。
- 群 1：回文數明顯較高，這些用戶可能是社群的高互動參與者，對回應討論感興趣。
- 群 2：活躍度最高，發文和回文數都非常高，可能是熱門話題的主要貢獻者或核心用戶。

註：圖片由專案團隊製作

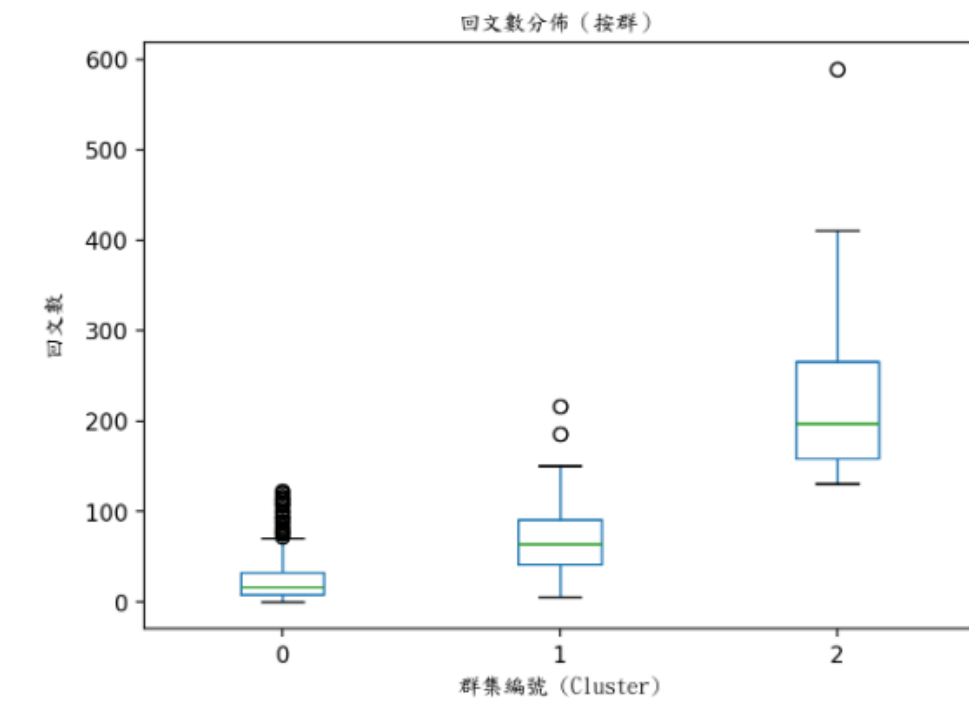
觀看方式：

- 群集 (顏色區分)：顏色對應用戶的分群結果。
- 點的位置 (分佈)：表示用戶行為的特徵相似性，距離越近的点表明用戶行為越相似。
- 可觀察：
 - 各群是否有明顯分界 (不同顏色的點是否聚集)。
 - 某些群是否存在異常行為 (例如遠離其他點的用户)。

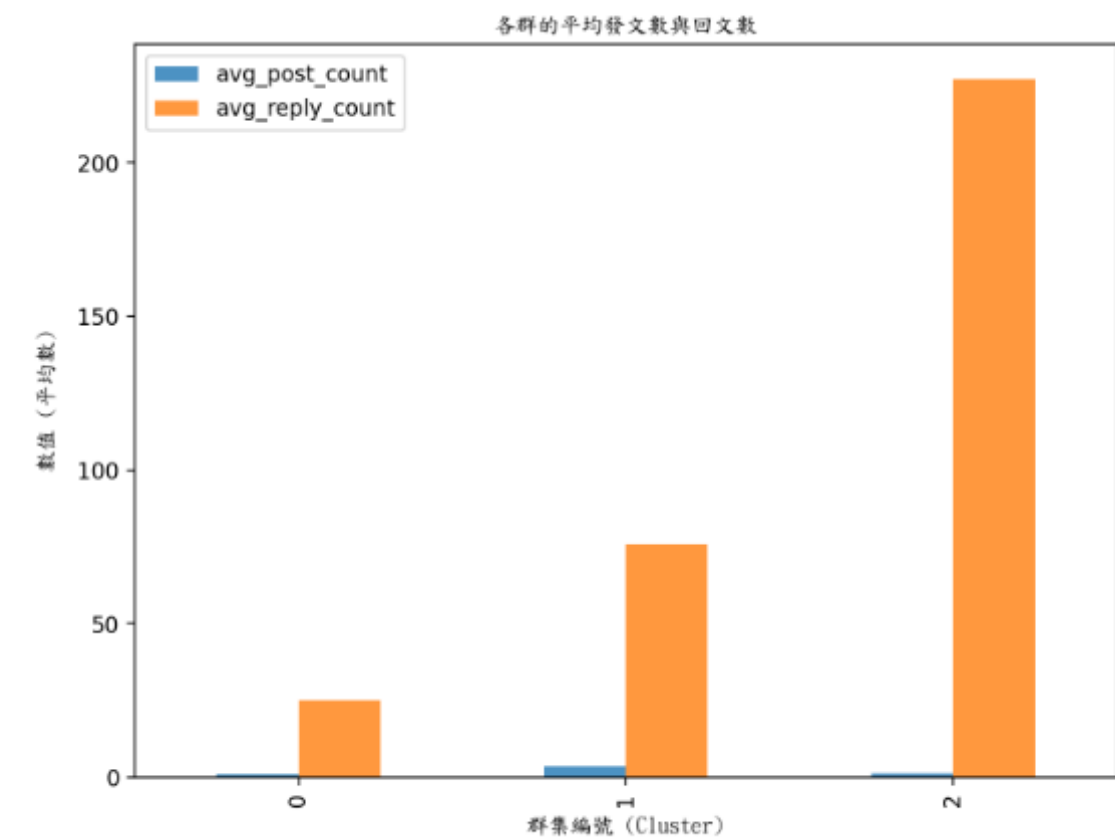


觀看方式：

- 群集編號 (X 軸) 表示不同用戶群。
- 回文數 (Y 軸) 表示該群用戶的回文行為數據。
- 可以觀察：
 - 群間的回文數分佈差異。
 - 是否存在某些群有極高或極低的異常值。



各群平均特徵比較



總結

「從數據中挖掘價值，用智慧洞察引領決策。」

本專案透過即時關鍵字分析與多維度輿情洞察，為使用者提供了一個高效、直觀的工具。未來，我們將持續優化技術，擴展應用場景，讓社群數據真正成為決策的力量。



參考文獻

技術與工具

1. Hugging Face

- Transformers 文檔: <https://huggingface.co/docs/transformers>
- 模型: uer/roberta-base-finetuned-jd-binary-chinese
 - 模型頁面: <https://huggingface.co/uer/roberta-base-finetuned-jd-binary-chinese>

2. ARIMA 時間序列模型

- Statsmodels 文檔: <https://www.statsmodels.org>
- 使用指南: "Introduction to ARIMA Models"

3. K-means 與 PCA

- scikit-learn 文檔: <https://scikit-learn.org/stable/>
- K-means 使用指南: "Clustering with K-Means"

參考文獻

參考文章與資源

4. Streamlit

- 官方文檔: <https://docs.streamlit.io>
- Streamlit Cloud 部署指南: <https://streamlit.io/cloud>

5. 爬蟲技術

- [Requests 文檔](#)
- [BeautifulSoup 文檔](#)

6. [ARIMA時間序列模型python應用-銅價格預測\(一\)](#)

7. [Python爬下PTT文章內容技巧\(含程式碼\)](#)

參考文獻

圖書與學術資源

9. "Python for Data Analysis, 3rd Edition"

- 作者: Wes McKinney
- 出版社: O'Reilly Media, 2022

10. "Applied Text Analysis with Python"

- 作者: Benjamin Bengfort, Rebecca Bilbro
- 出版社: O'Reilly Media, 2020

圖片來源

[Hottsuite](#)
[Brandwatech](#)

[buzzsumo](#)
[pyhton](#)

[streamlit](#)
[github](#)

- GitHub 專案連結：
「GitHub 專案程式碼」：https://github.com/jun-weilin/NCHU/tree/main/AIoT-DA/Final_Project
- Streamlit Cloud 前端連結：
「Streamlit Cloud 使用介面」：
<https://eefzbzjg62yh54cyzxez5q.streamlit.app/>