

# 社群媒體分析 讀書會3

## Group3

B104020008 林冠儒 B104020044 黎懿綺  
B114020046 蔡佳倚 M134020017 李恩榮  
M134020024 盧荻原 M134020039 顏淳如

# 目錄

1. Week\_11 text embedding
2. Week\_12 bert

# 1. Week\_11 text embedding

資料集:

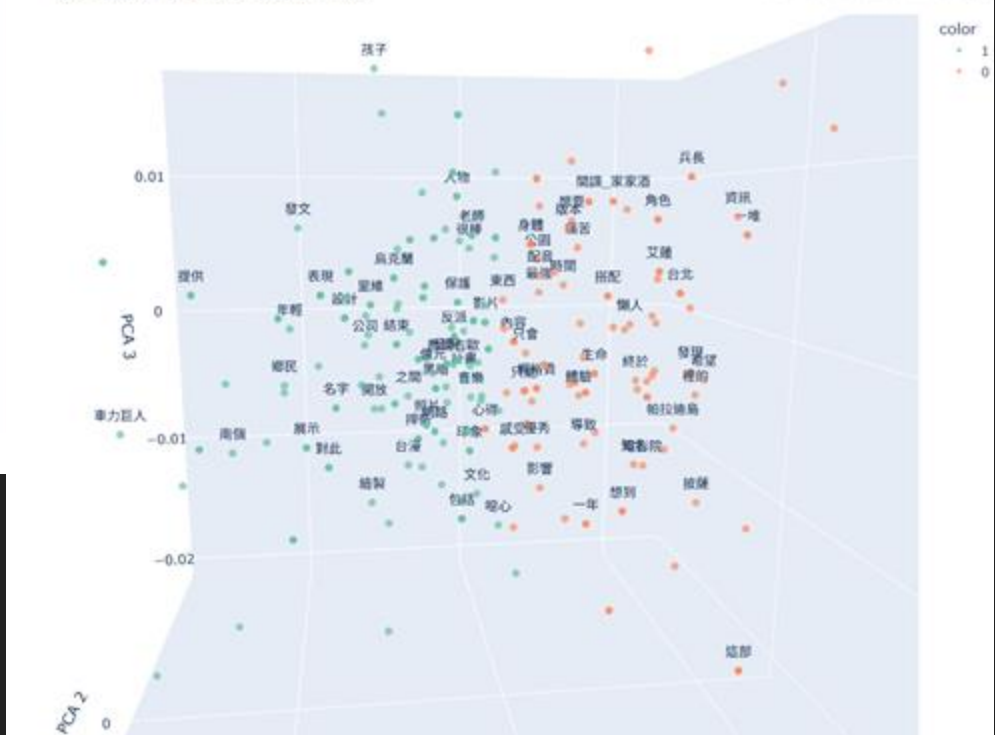
PTT八卦版、電影版

關鍵字:進擊的巨人

排除關鍵字:工作細胞、鬼滅之刃



◎ 4 · C · 1 8 8



=====

Query: [問卦]急！艾連要發動地鳴了

資料集中前五相似的文章：

[問卦]急！艾連要發動地鳴了 (Score: 1.0000)

[問卦]有沒有艾連發動地鳴的時候在想什麼 (Score: 0.9779)

[問卦]敬告所有尤米爾的子民== (Score: 0.9735)

Re:[問卦]進擊的巨人(有雷) (Score: 0.9548)

[問卦]地鳴有比瑪莉亞之牆被打破來的絕望嗎？ (Score: 0.9543)

=====

**embeddings : bert-base-chinese**

**資料集 : Dcard追星、寵物、時事、Youtuber版**

**分類器 : 邏輯回歸模型**

	precision	recall	f1-score	support
entertainer	0.85	0.94	0.89	127
pet	0.99	0.95	0.97	173
trending	0.93	0.92	0.92	253
youtuber	0.83	0.82	0.82	184
accuracy			0.90	737
macro avg	0.90	0.91	0.90	737
weighted avg	0.91	0.90	0.90	737

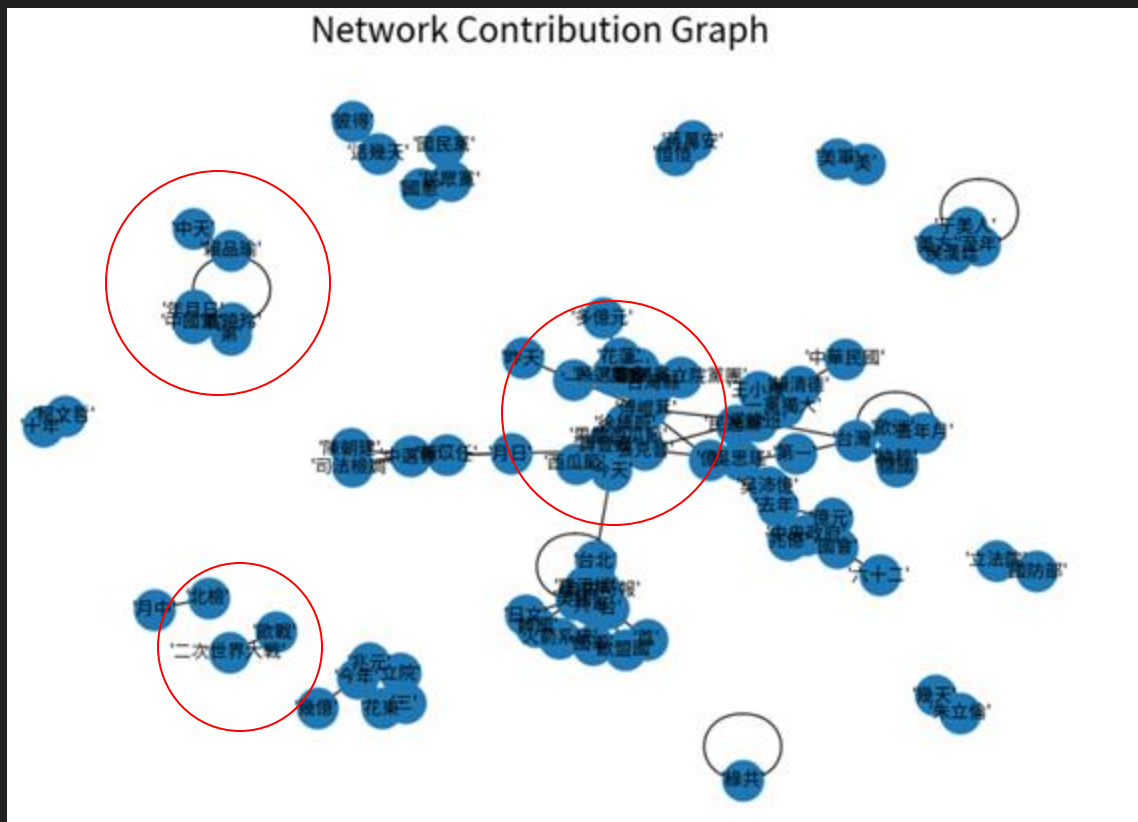


## **2. Week\_12 bert**

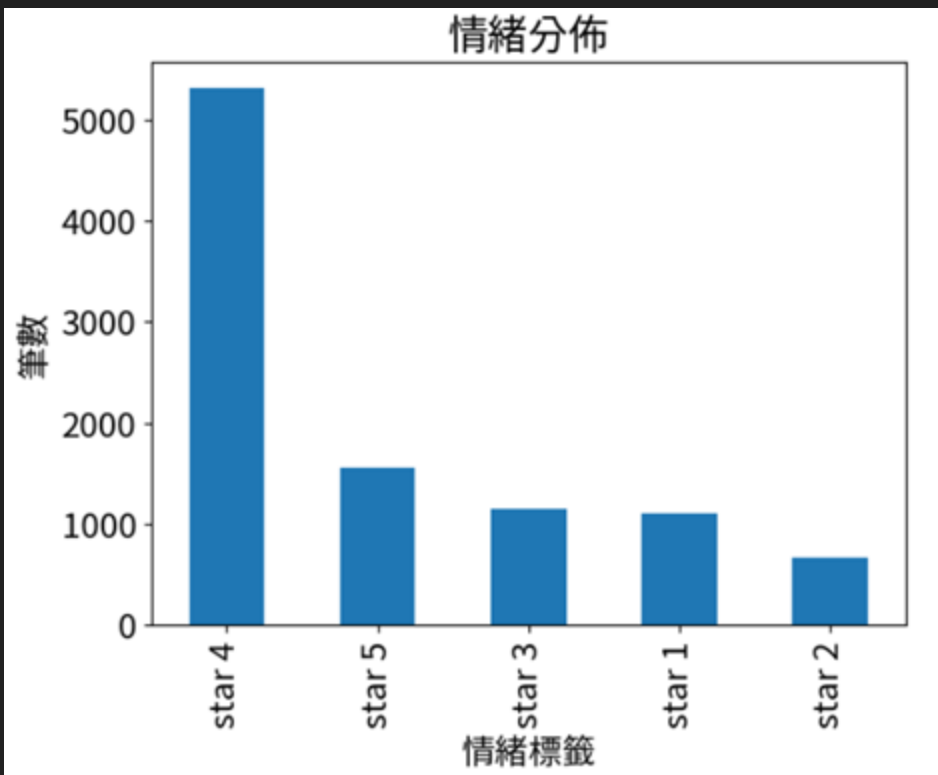
**資料集1:**

**PTT政黑板2025年5月資料集， 並隨機抽樣500筆**

# Token classification (NER)



# Sequence classification (Sentiment Classification)



	sentence	label	score
9049	第九十八條之二利用他人個人資料	star 4	0.988665
9605	證券交易法第條第項第款	star 4	0.988232
5770	公職選罷法第條第項規定	star 4	0.987223
9364	四刑事訴訟法第條第條之立法理由	star 4	0.986962
5218	寫在選罷法第七十六條第一項	star 4	0.983777
9389	該審判外之陳述	star 4	0.983189
9452	本院年台上字第號判例	star 4	0.982085
9491	刑事訴訟法第條第項定有明文	star 4	0.968635
9502	其中第點第款明定檢察機關為偵查犯證券交易法第條之罪之案件	star 4	0.967911
9289	核原判決所為論斷說明	star 4	0.965125

有些中性句子被分類到star4

# Text Clustering (KMeans)

Intertopic Distance Map

D2



Topic 1、5、9  
跟市場情況有關



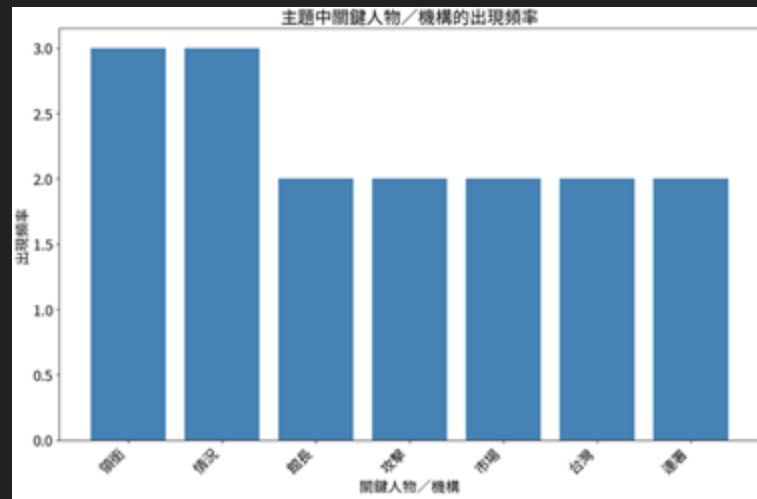
Topic 3、4、6、7、8  
跟政治、戰爭有關

D1



Topic 0、2  
跟政黨有關

	Topic	Top Words	Example Sentences
0	0	國民黨, 支持者, 柯文, 民進黨	[很大程度也是在年終替總統柯文哲, 此時戴帽子那個國民黨支持者借出拳要打白衣男]
1	1	, 題, 信, 並說, 第一	[如題, 如題]
2	2	, 西瓜, 說, 說, 其	[傅崐冀說, 傅崐冀今天與妻子縣長徐榛蔚連袂出席在壽豐鄉舉辦的果饅西瓜節活動]
3	3	提議, 避免, 中選會, 黨, 告發	[在提議連署第一階段後就進行告發, 吳思遠避免黨在第一階段共送出件避免提議書]
4	4	, 國家, 歷史, 二戰, 歌戰	[共同反侵略國家面追求真正和平的決心和行動, 我們與參加歌戰的許多民主國家]
5	5	, 合規, 說, 市場, 情況	[是市場視作壓力, 本是同根生相何太急]
6	6	優著, 偽造, , 檢調, 員工	[部長直插稿單針對拍影片自述稿底下員工遭到青島攻擊的事件, 我不認同偽造優著]
7	7	納粹, , 媒體, 扭曲, 事實	[但當事者德國最知道台灣現況根本就納粹八竿子打不著所以出來說話, 德國表達的簡單說就是納粹...
8	8	飛彈, 文匯, 系統, 枚, 套	[第批遠程精準火力打擊系統套及訓練模擬器套已於年月日接收, 採軍購方式向美購買海馬斯多管火箭...
9	9	剷除, 轉錄, 不計, 代填, 一半	[可直接剷除不計, 轉錄網址轉錄來源]



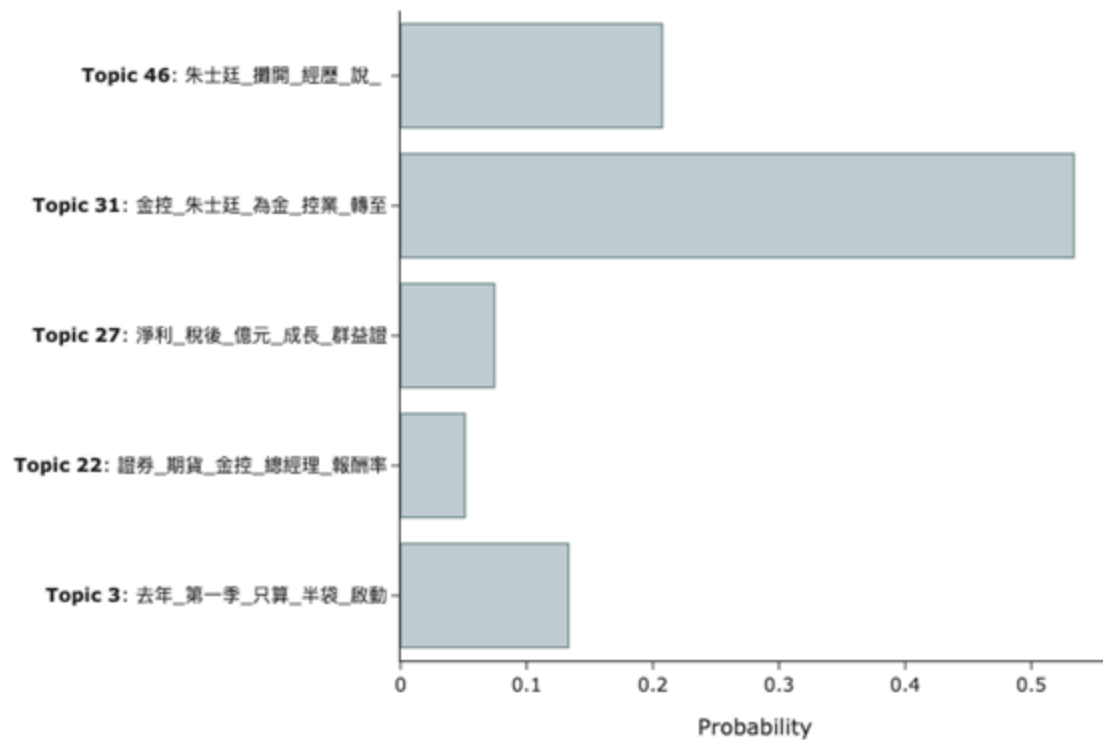
**資料集 2：**

**PTT財經版2024年到今年4月資料集， 共400多篇**

## Intertopic Distance Map



## Topic Probability Distribution





主題 2 詞彙機率前10名

