

股市社群之情緒分析 以台積電為例

專題組員：林姍如、黃雅培、廖劭其、張婕菱

指導教授：劉亮志 教授

技術指導：鄭麗珍 教授

CMoney

專題報告



🔍 前情提要

🔍 資料蒐集

🔍 資料分析

🔍 建議及總結

🔍 參考資料

研究動機



社群媒體不受
距離及時間限制

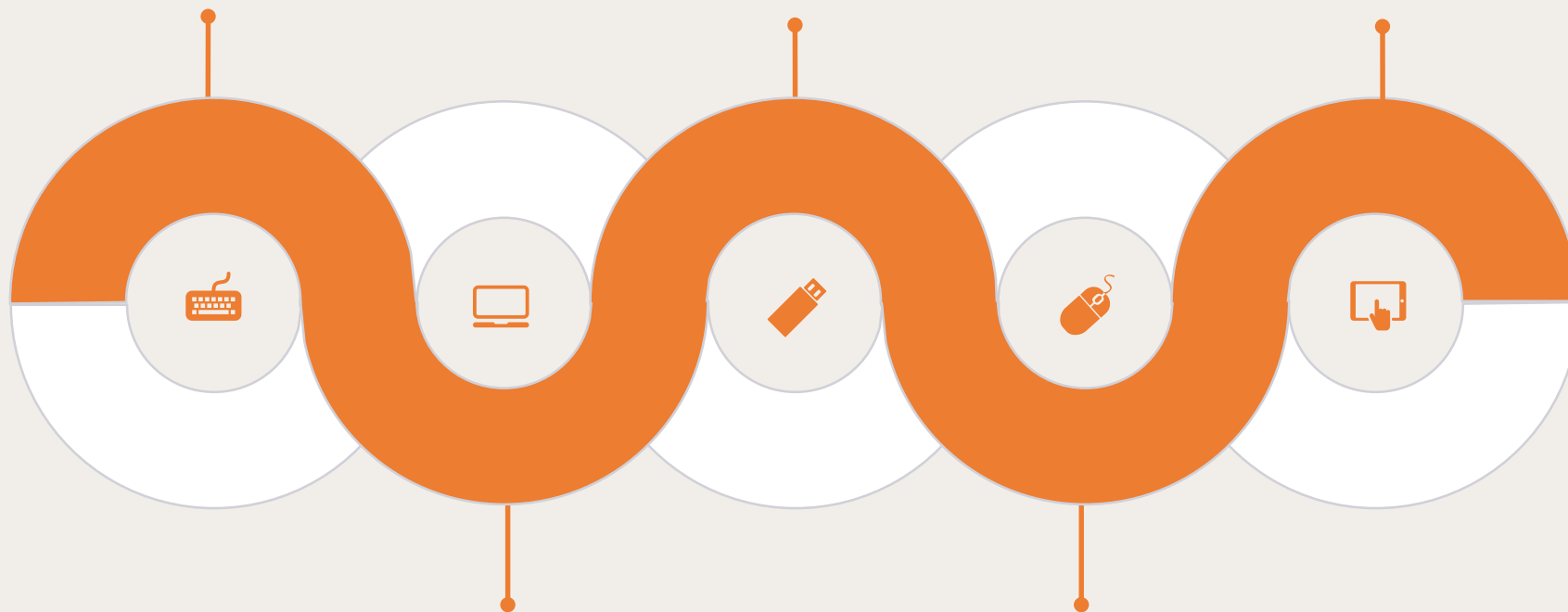
台灣股市
交易活動逐年熱絡

「護國神山」
台積電

一、爬蟲

三、正負向標記

五、系統情緒



二、斷詞

四、撿詞

使用自動化爬蟲套件進行資料蒐集

1.

Selenium

2.

PhantomJS

台積電一年的社群文章
(2021/3/27-2022/3/27) ►

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	發文時間	股票名稱與代碼	發文者	發文者等級	文章讚數	文章留言數	文章ID	文章內容					
2	2022/3/28 01:12	台積電2330	紅綠一瞬間	31	3	0	143843611	我覺得現在疫情大家都麻痺了；政府再管制一下又會降低疫情					
3	2022/3/28 00:49	台積電2330	3mr2L227L	57	5	1	143843516	可以進場了！					
4	2022/3/28 00:46	台積電2330	B布魯斯先生W	35	53	5	143843511	各位晚上好 Mr.布魯斯 OK 參照上週的規劃來看 可以看到盤後					
5	2022/3/28 00:28	台積電2330	IG搜尋🔍：趨勢	18	1	0	143843414	純分享市場資訊，想看更多請點IG連結 <a class="hplk" href="htt					
6	2022/3/27 23:44	台積電2330	莫忘世上韭菜多	14	13	0	143843041	明天開低走高收最高，嘎空嘎起來，600元見🤖🤖🤖🤖🤖					
7	2022/3/27 23:37	台積電2330	1fNiAulok	3	2	0	143843005	市場資金不足 類股輪動自然會快 短短一周時間 輪漲了至少5個					
8	2022/3/27 23:32	台積電2330	8qshA6jf	1	0	0	143842949	市場資金不足 類股輪動自然會快 短短一周時間 輪漲了至少5個					
9	2022/3/27 23:26	台積電2330	1fpVd0iB4	1	1	0	143842882	市場資金不足 類股輪動自然會快 短短一周時間 輪漲了至少5個					
10	2022/3/27 23:02	台積電2330	新竹工程師	27	3	1	143842696	OMG 上海封城 🤖🤖🤖					
11	2022/3/27 23:00	台積電2330	大金豬	24	26	7	143842664	上海突宣布半封城!先封浦東再封浦西 上海市官方27日晚間宣					
12	2022/3/27 22:53	台積電2330	帥哥哥	45	4	0	143842598	明天頂多就是回測589~590元而已~趨勢沒有變，參考					
13	2022/3/27 21:57	台積電2330	lkkoacsi	37	3	1	143842084	<a class="hplk" href="https://www.youtube.com/watch?v=WQovjtWVf					
14	2022/3/27 21:45	台積電2330	蚊子	19	28	2	143841987	近期 就 小心一點囉~~~~ 疫情有點小影響 大陸 香港 本地 都有					
15	2022/3/27 21:31	台積電2330	399	67	12	12	143841856	美股大漲，明天個股漲🚀🤖					
16	2022/3/27 21:16	台積電2330	易金看未來	30	62	4	143841728	E金看未來，坐三望四從美科技股看趨勢 NASDAQ100三月得					
17	2022/3/27 21:02	台積電2330	James	1	7	0	143841611	市場資金不足 類股輪動自然會快 短短一周時間 輪漲了至少5個					
18	2022/3/27 20:33	台積電2330	張書記	19	7	0	143841353	『發財三要素：想法、策略、執行』 3月行情你賺到錢了嗎?					
19	2022/3/27 20:32	台積電2330	秒盤專家	34	3	6	143841347	有人準備要放空囉 藉由疫情要來放空囉 禮拜一準備要放空囉					
20	2022/3/27 20:27	台積電2330	韭菜畢業班-叔叔	46	48	0	143841306	🔥美股目前概況🔥 請先按讚追蹤🔍韭菜畢業班-叔叔! 請					
21	2022/3/27 20:26	台積電2330	韭菜畢業班-叔叔	46	34	0	143841272	🔥03/27 台股彙整🔥 看之前記得先按下讚! 預祝各位明天掛					
22	2022/3/27 20:23	台積電2330	韭菜畢業班-叔叔	46	29	0	143841251	🔥台股盤後彙整 今天的盤後影音上架了~ 各位同學開之前記					
23	2022/3/27 20:10	台積電2330	怪醫股博士	57	90	12	143841164	大家可以看看國巨陳泰銘的面相。我不會說他老奸巨猾，但會					

文章範例

沒意外會再倒一波，636賣單加堆，635買單沒人掛，連假單都不掛，就是要倒了的意思。期指17747--17824高要過才會繼續攻，2330不夠強，637要過，今天期指17702不破也算還好，關鍵需要5-8天時間整理，上面還有缺口要回補也不要看太壞

斷詞範例

['意外', '倒', '賣單', '加堆', '買單', '沒', '人', '掛', '假單', '掛', '倒', '意思', '期指', '高', '過', '才', '繼續', '攻', '夠', '強', '過', '今天', '期指', '破', '還好', '關鍵', '需要', '時間', '整理', '上面', '缺口', '回補', '看', '太', '壞']

斷詞範例

'烏' , '俄' , '戰爭' , '貨櫃' ,
'三' , '雄' , '新' , '冠' , '肺炎'

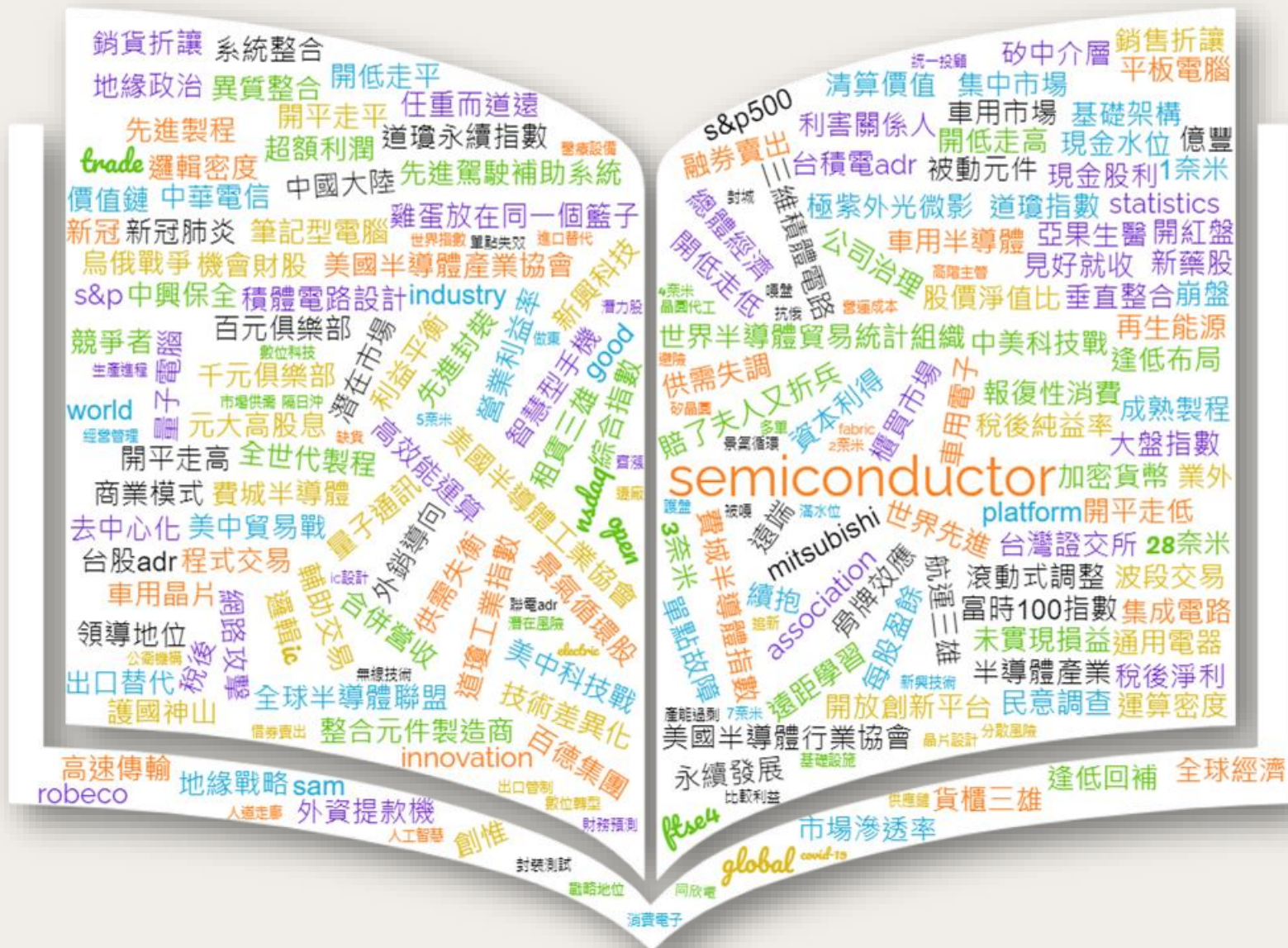
撿詞範例

- 烏俄戰爭
- 貨櫃三雄
- 新冠肺炎

四

五

8





斷詞範例

['意外', '倒', '賣單', '加堆', '買單', '沒',
'人', '掛', '假單', '掛', '倒', '意思', '期
指', '高', '過', '才', '繼續', '攻', '夠', '強',
, '過', '今天', '期指', '破', '還好', '關鍵',
, '需要', '時間', '整理', '上面', '缺口',
'回補', '看', '太', '壞']

正負向標記

['意外', '倒', '賣單', '加堆', '買單', '沒',
, '人', '掛', '假單', '掛', '倒', '意思', '
期指', '高', '過', '才', '繼續', '攻', '夠',
'強', '過', '今天', '期指', '破', '還好', '
關鍵', '需要', '時間', '整理', '上面', '
缺口', '回補', '看', '太', '壞']

四、正負向

情緒字典範例

正向詞

買單、攻、強、整理、回補

負向詞

倒、賣單、缺口、壞

負向字詞文字雲



下跌、跌、賣、
賣超、升息

在全年的所有文章
中各出現超過
5000次

五、系統情緒

資料樣本區間 ▼

2021年3月27日至2022年3月27日

股市交易日244天

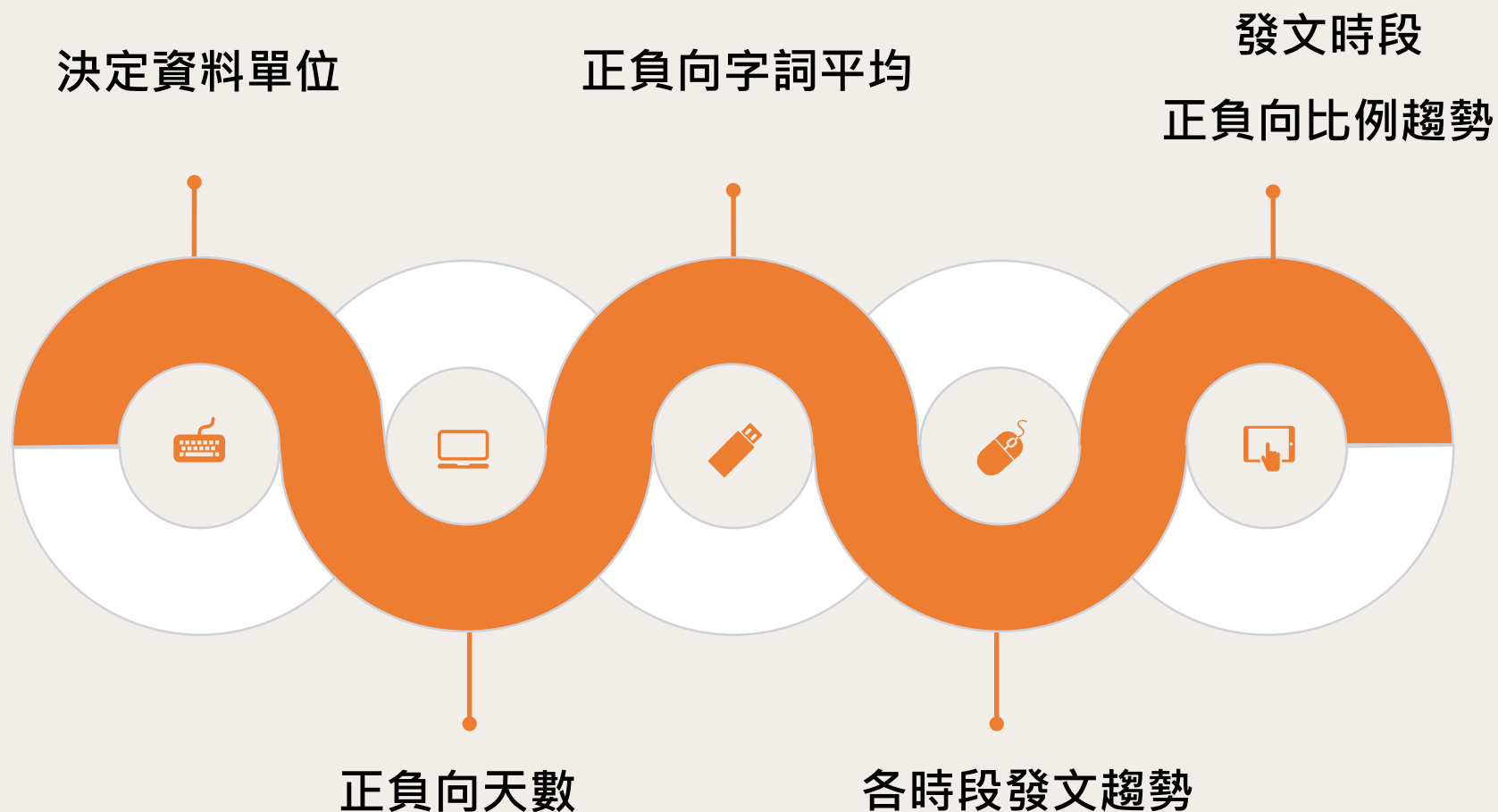
總文章篇數57798篇

① 發文內容分析

② 系統準確度分析



① 發文內容分析





① 發文內容分析

✓ 決定資料單位

系統情緒 (各篇加總) ▼

以一篇社群文章為一個情緒單位

系統情緒 (全天加總) ▼



以一天為一個情緒單位

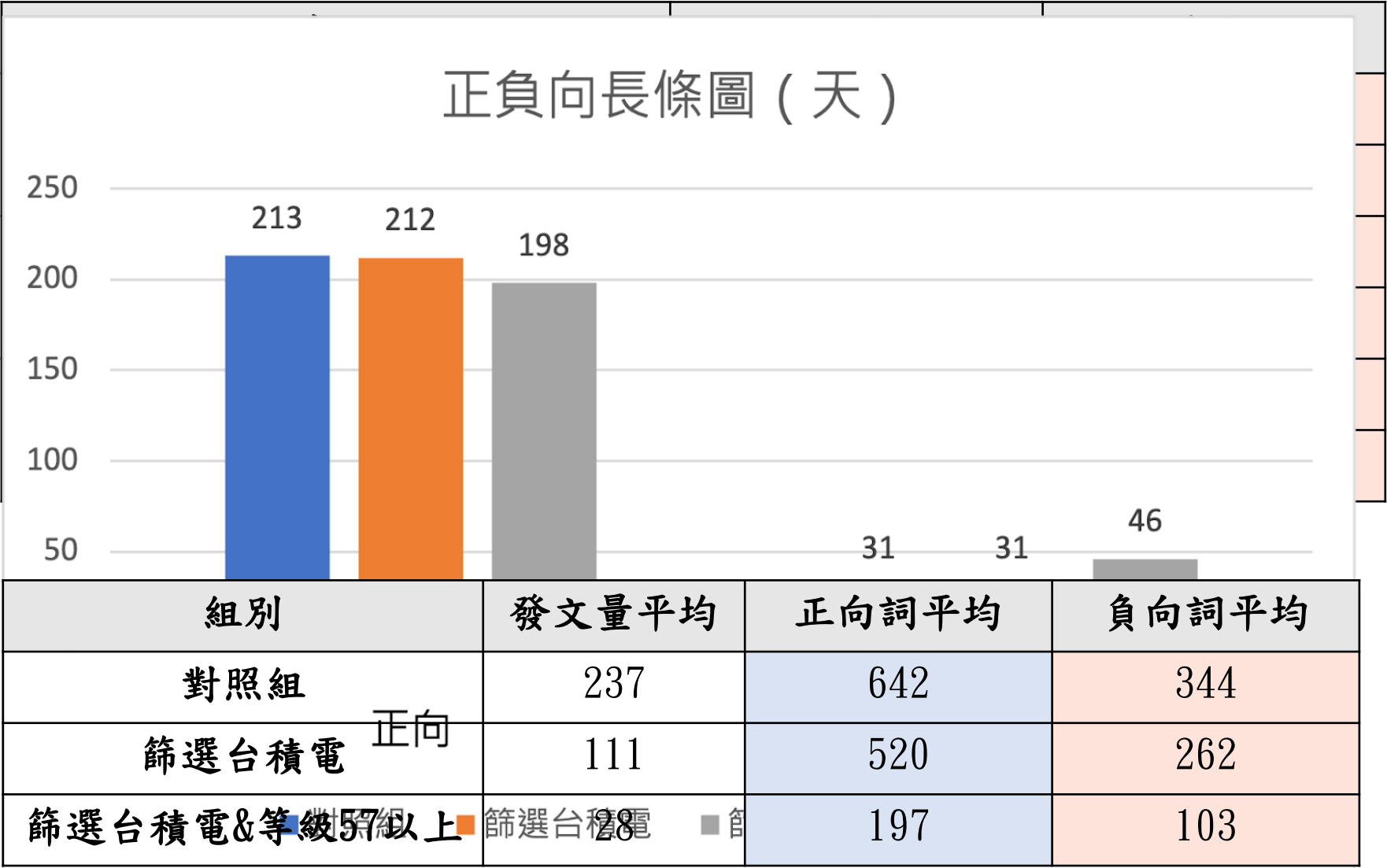
	對照組			實驗組 (篩選台積電)			實驗組 (篩選台積電&等級57以上)		
		當日	隔日		當日	隔日		當日	隔日
系統情緒 篇加總	符合	125	105	符合	118	104	符合	122	109
	不符合	119	139	不符合	126	140	不符合	122	135
	總數	244	244	總數	244	244	總數	244	244
	符合率	51.23%	43.03%	符合率	48.36%	42.62%	符合率	50.00%	44.67%
系統情緒 天加總		當日	隔日		當日	隔日		當日	隔日
	符合	129	115	符合	130	114	符合	140	113
	不符合	115	129	不符合	114	130	不符合	104	131
	總數	244	244	總數	244	244	總數	244	244
	符合率	52.87%	47.13%	符合率	53.28%	46.72%	符合率	57.38%	46.31%



① 發文內容分析

✓ 正負向天數

✓ 正負向字詞平均



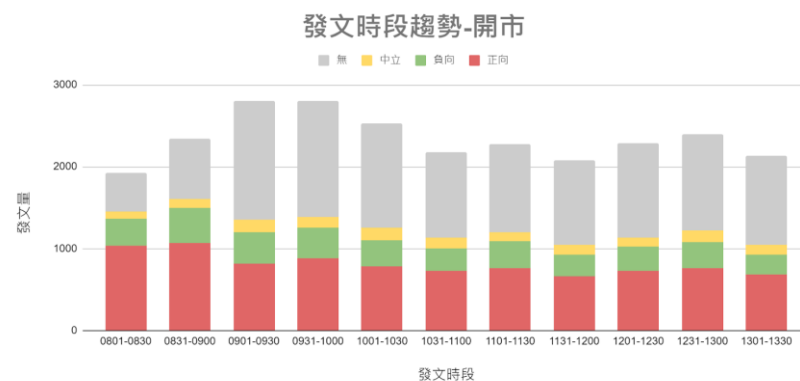
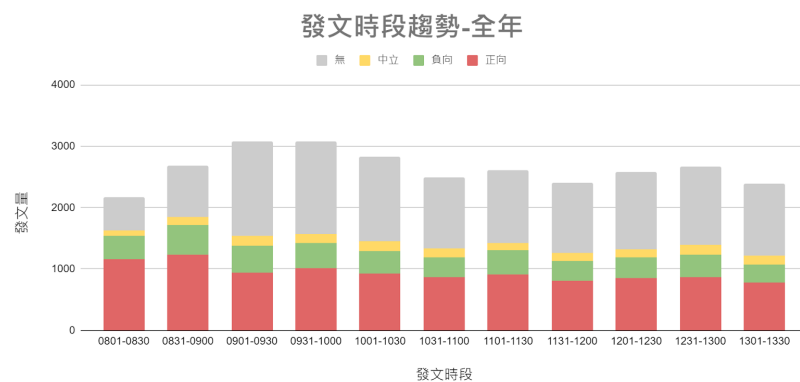


① 發文內容分析

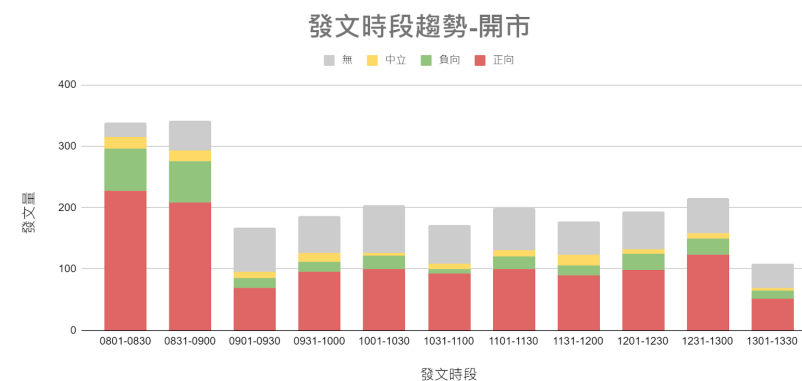
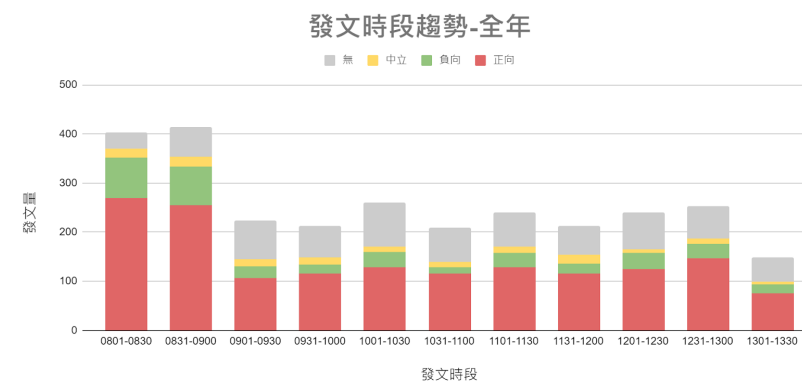
✓ 各時段發文趨勢

- 上午08:00
- 下午13:30
- 每30分鐘為一區間
- 全年資料(366天)
- 交易日資料(244天)

對照組



實驗組 (篩選台積電&等級57以上)



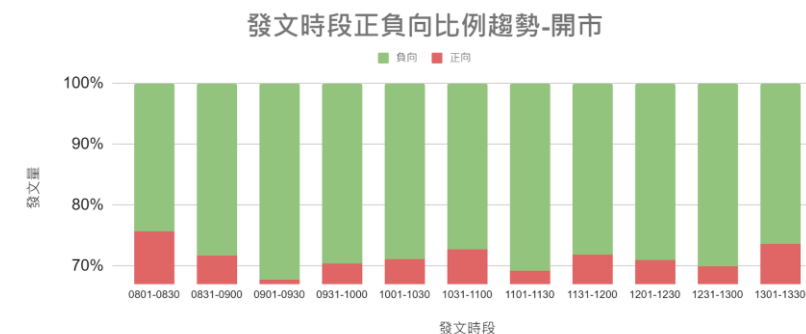
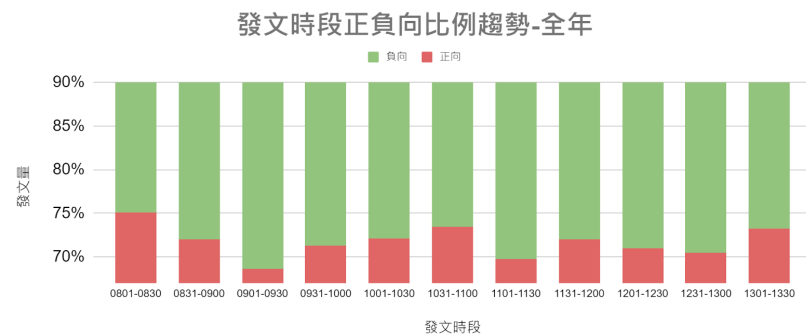


① 發文內容分析

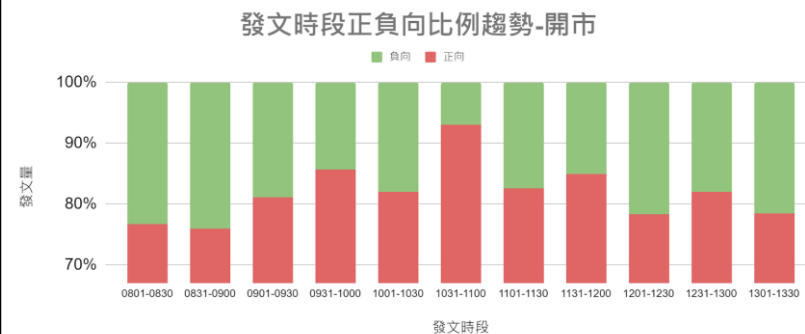
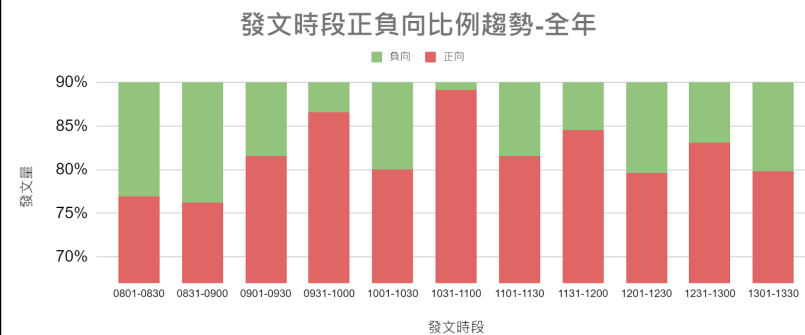
✓ 發文時段正負向 比例趨勢

- 上午08:00
- 下午13:30
- 每30分鐘為一區間
- 全年資料(366天)
- 交易日資料(244天)

對照組



實驗組 (篩選台積電&等級57以上)



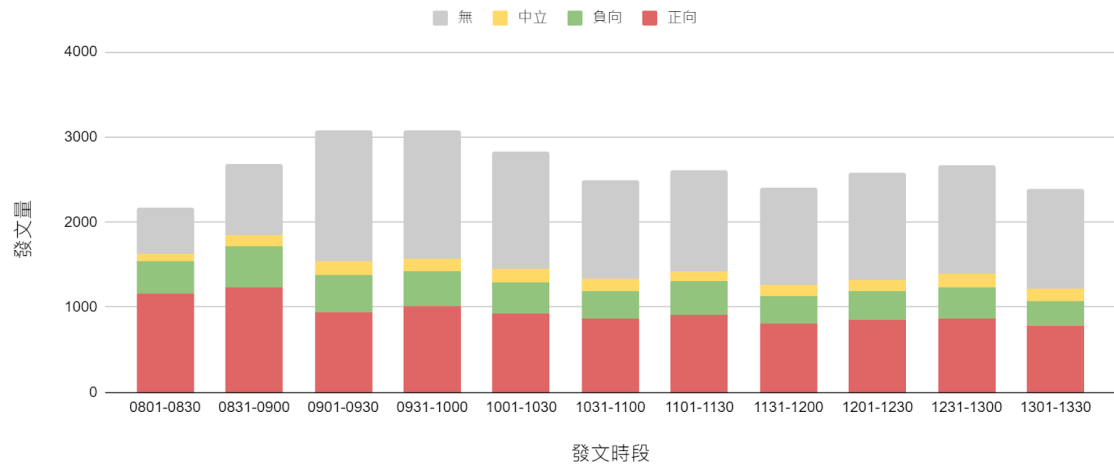


① 發文內容分析

篩選條件增加，無情緒的文章數量越少。

對照組

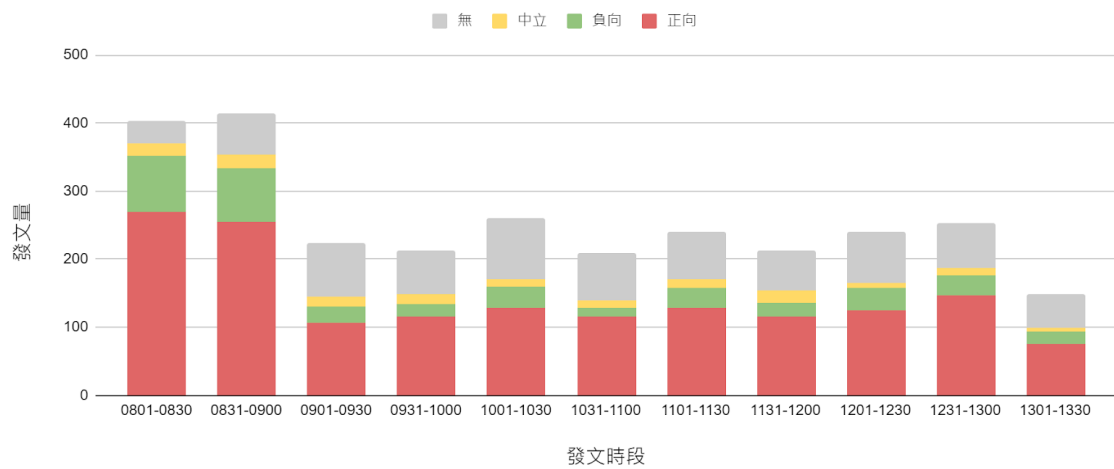
發文時段趨勢-全年



實驗組

(篩選台積電&等級57以上)

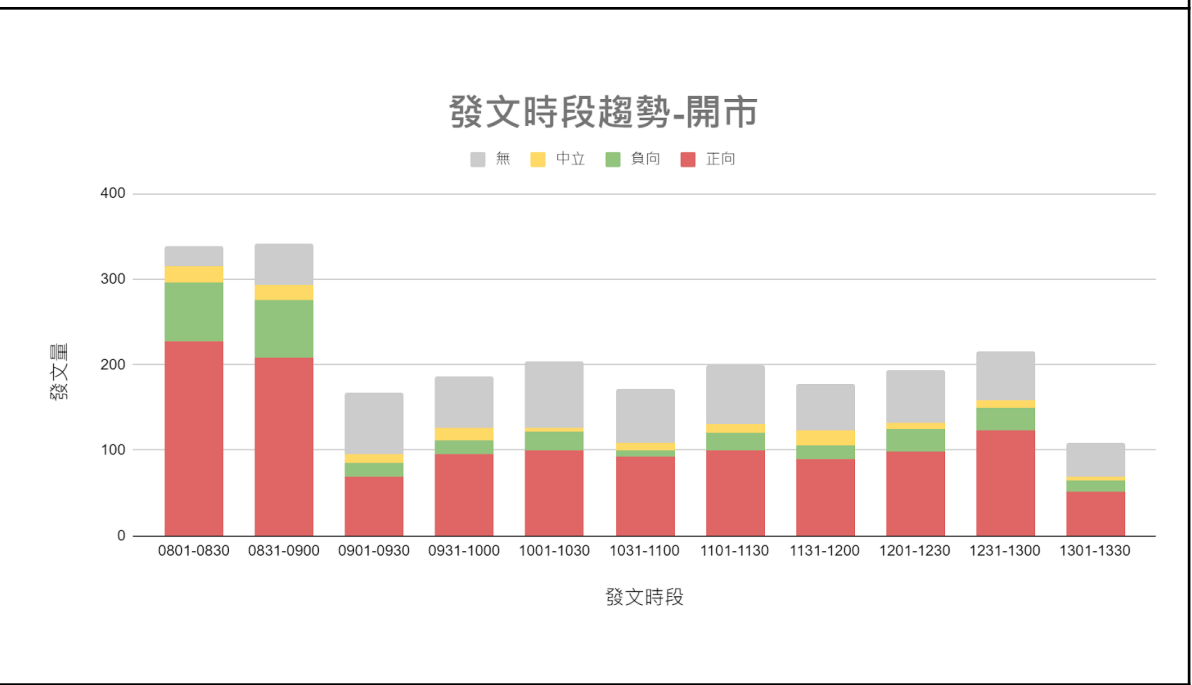
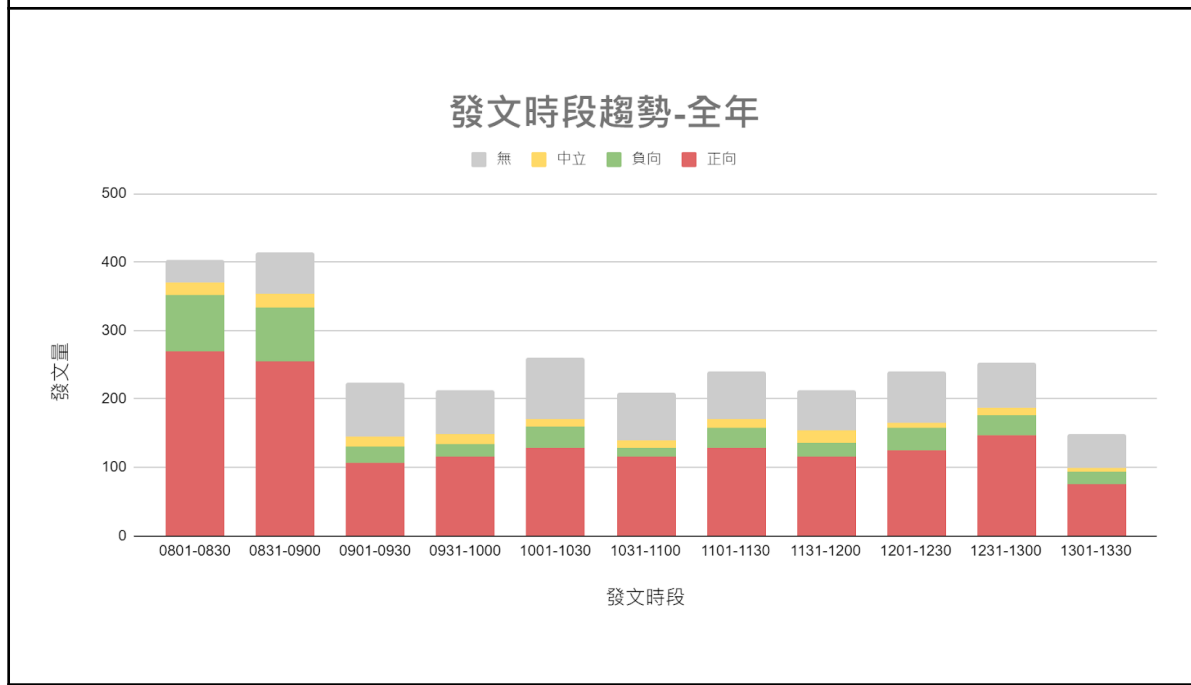
發文時段趨勢-全年



① 發文內容分析

發文等級較高的使用者發文時段較集中在開盤前一小時，也就是上午08:00至09:00之間。

實驗組
(篩選台積電&等級57以上)

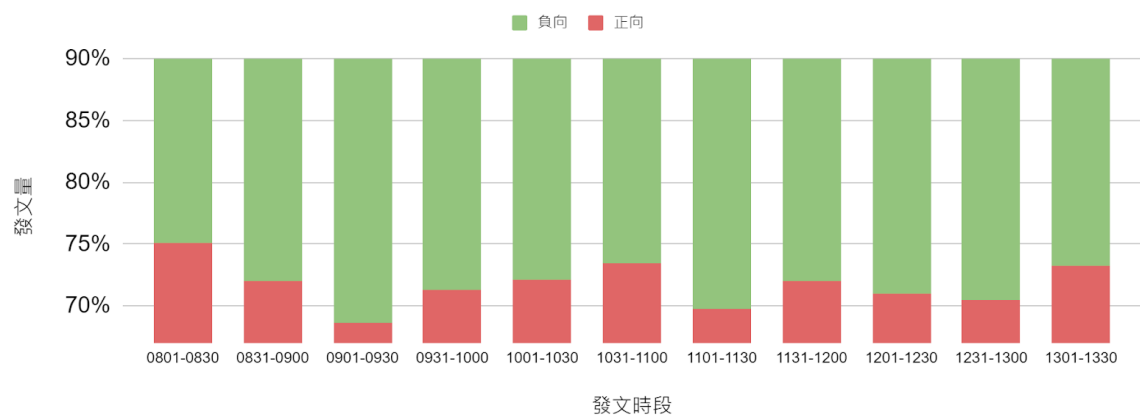


① 發文內容分析

與無篩選相比，高等級使用者的發文內容中，整體正向文章的比例高出不少，特別是全年區間的實驗組資料中，上午10:30至11:00這半小時的正向文章比例更接近90%

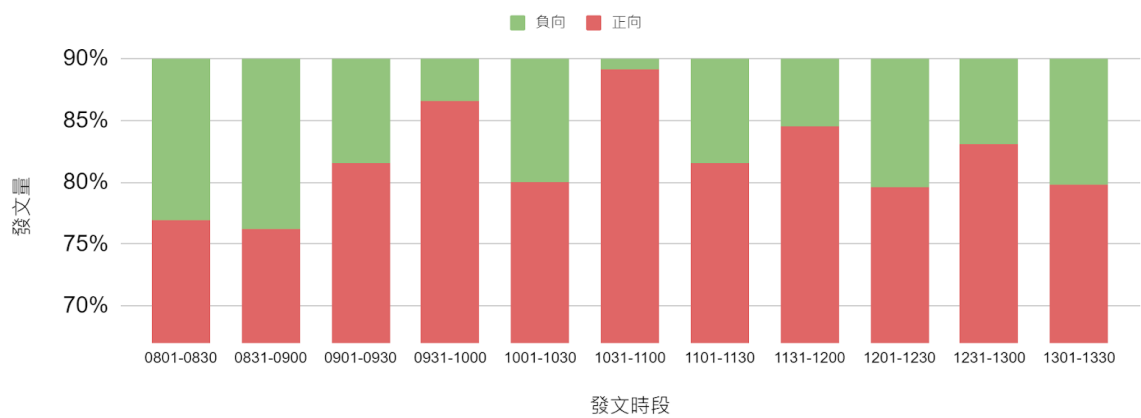
對照組

發文時段正負向比例趨勢-全年



實驗組
(篩選台積電&等級57以上)

發文時段正負向比例趨勢-全年





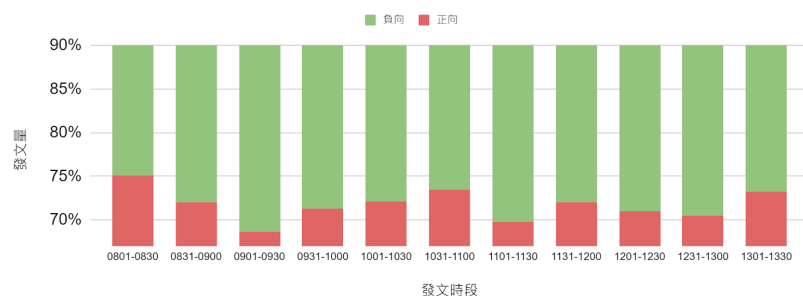
① 發文內容分析

✓ 發文時段正負向 比例趨勢

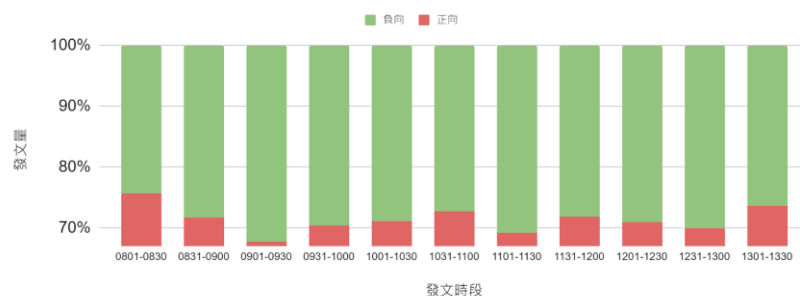
發現開市與全年的發文型態無論在發文時段趨勢或是正負向文章比例都極相似。

對照組

發文時段正負向比例趨勢-全年



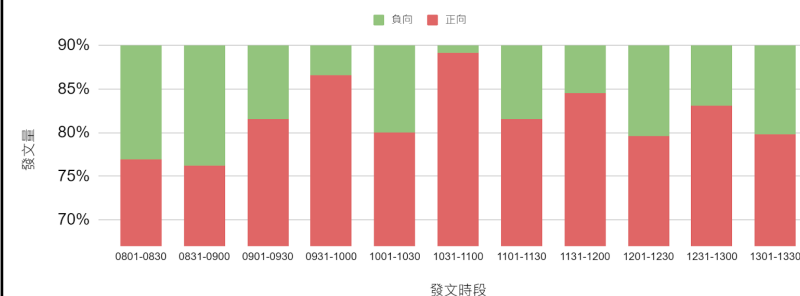
發文時段正負向比例趨勢-開市



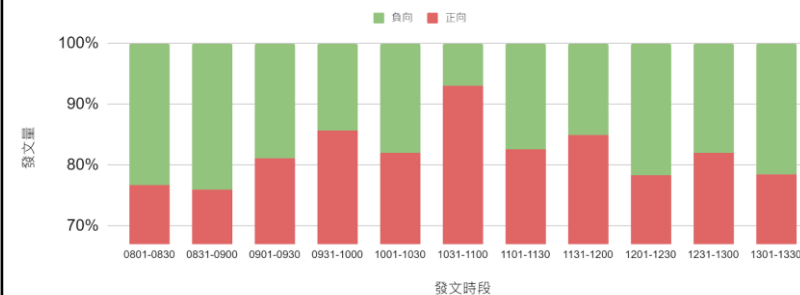
實驗組

(篩選台積電&等級57以上)

發文時段正負向比例趨勢-全年

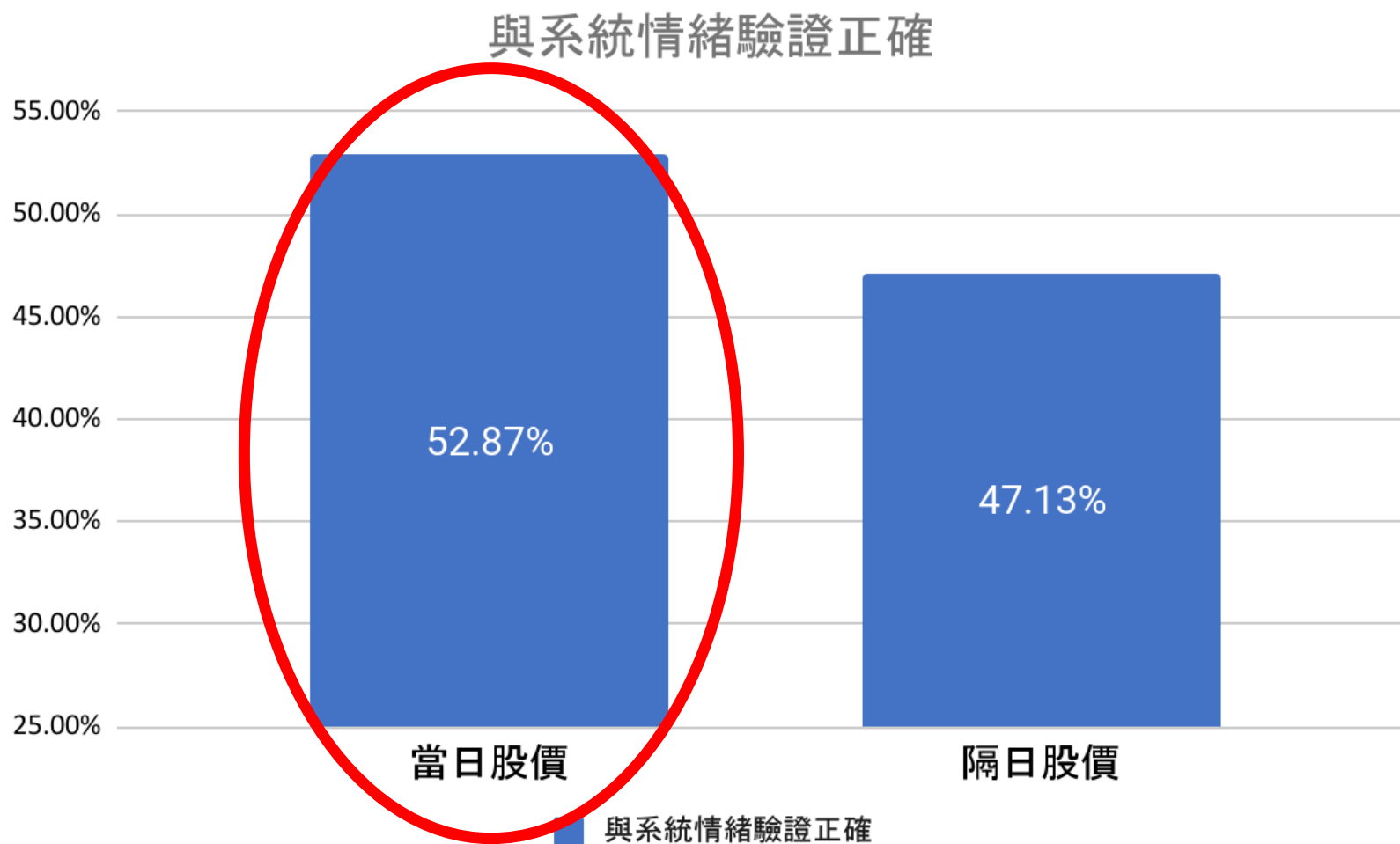


發文時段正負向比例趨勢-開市

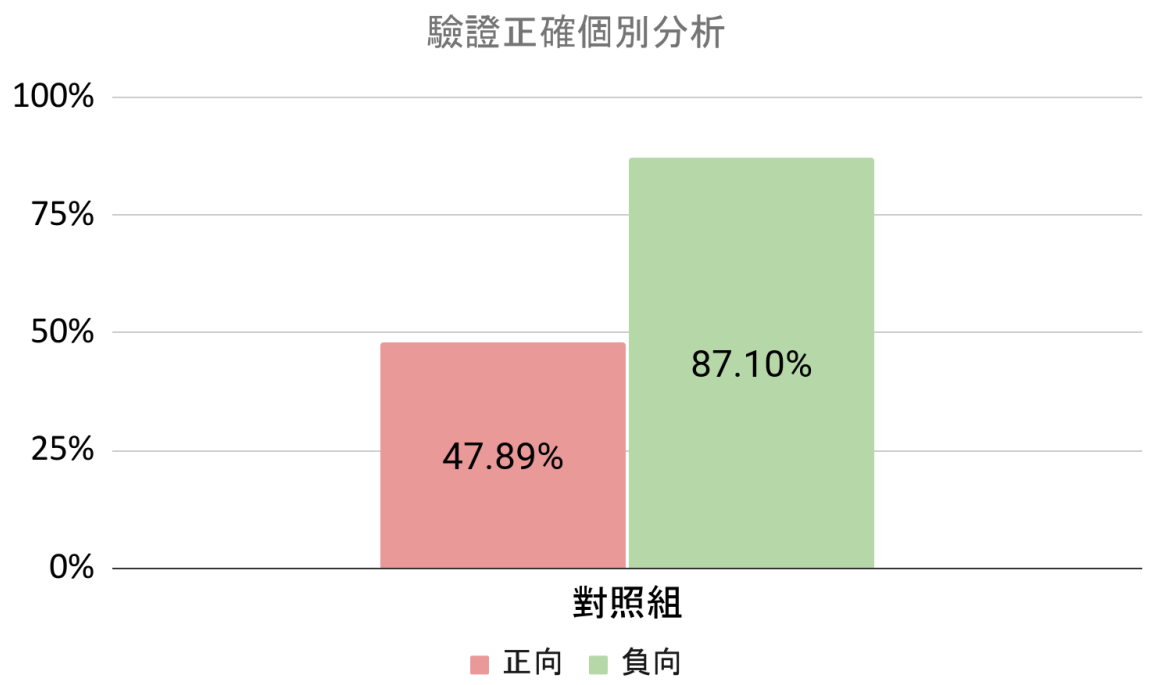
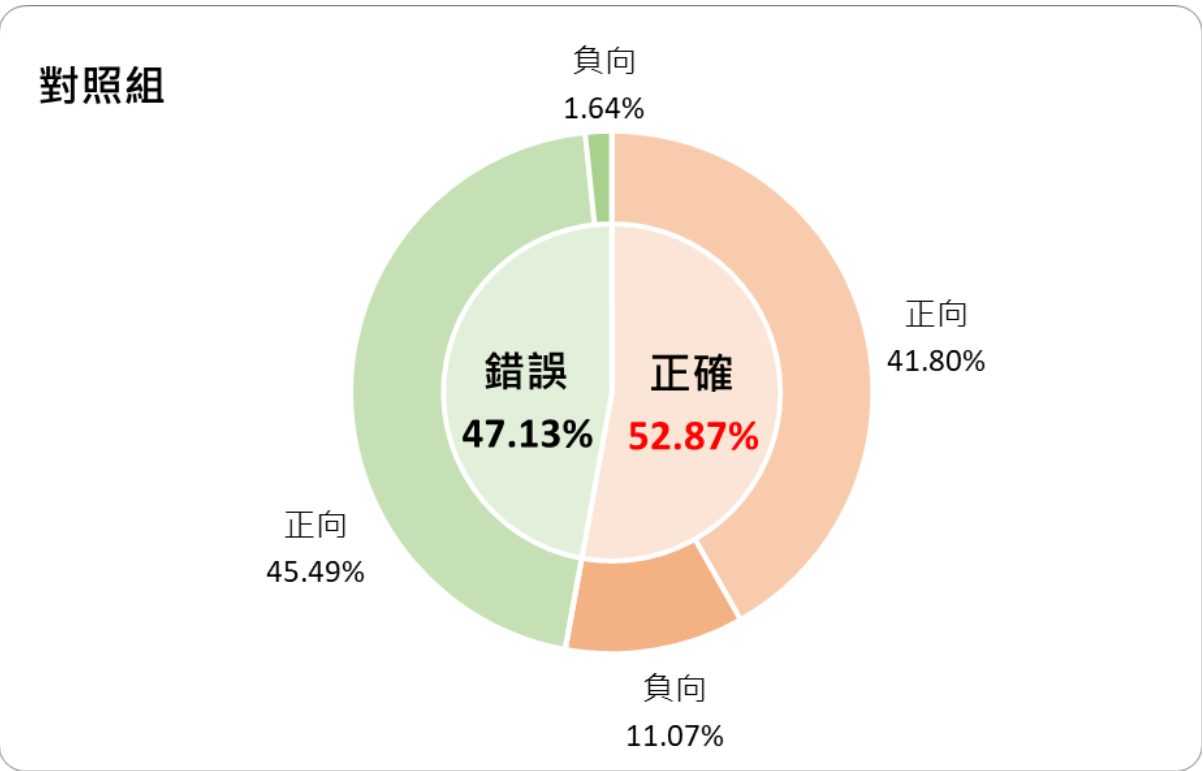


② 系統準確度分析

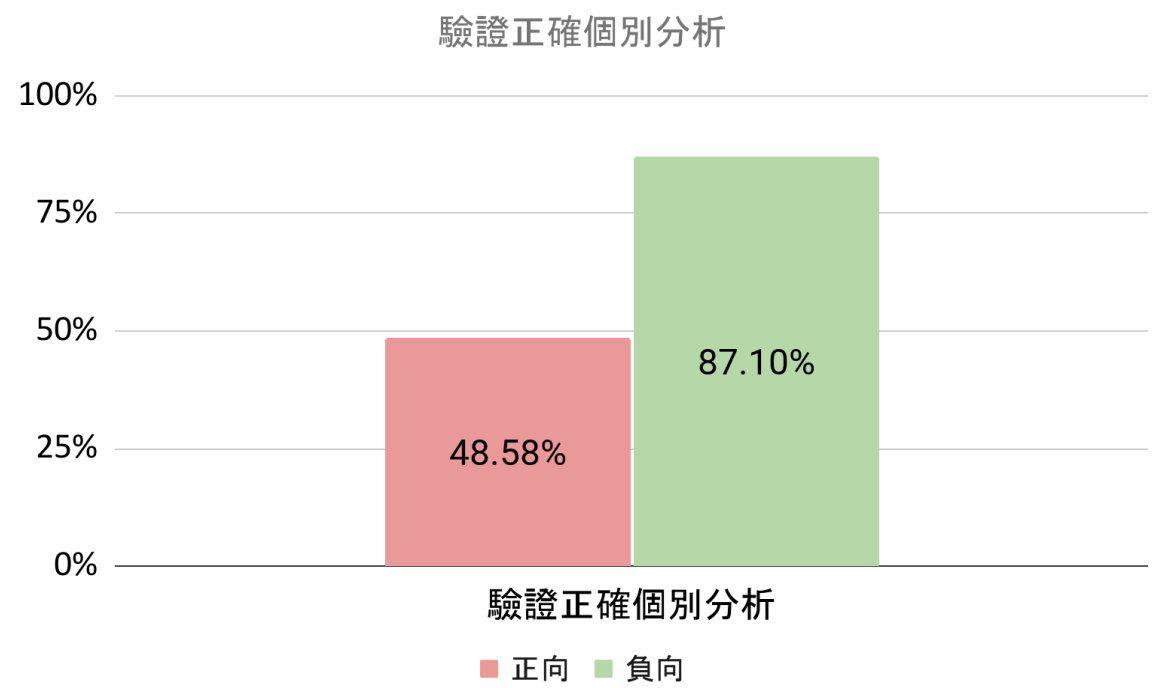
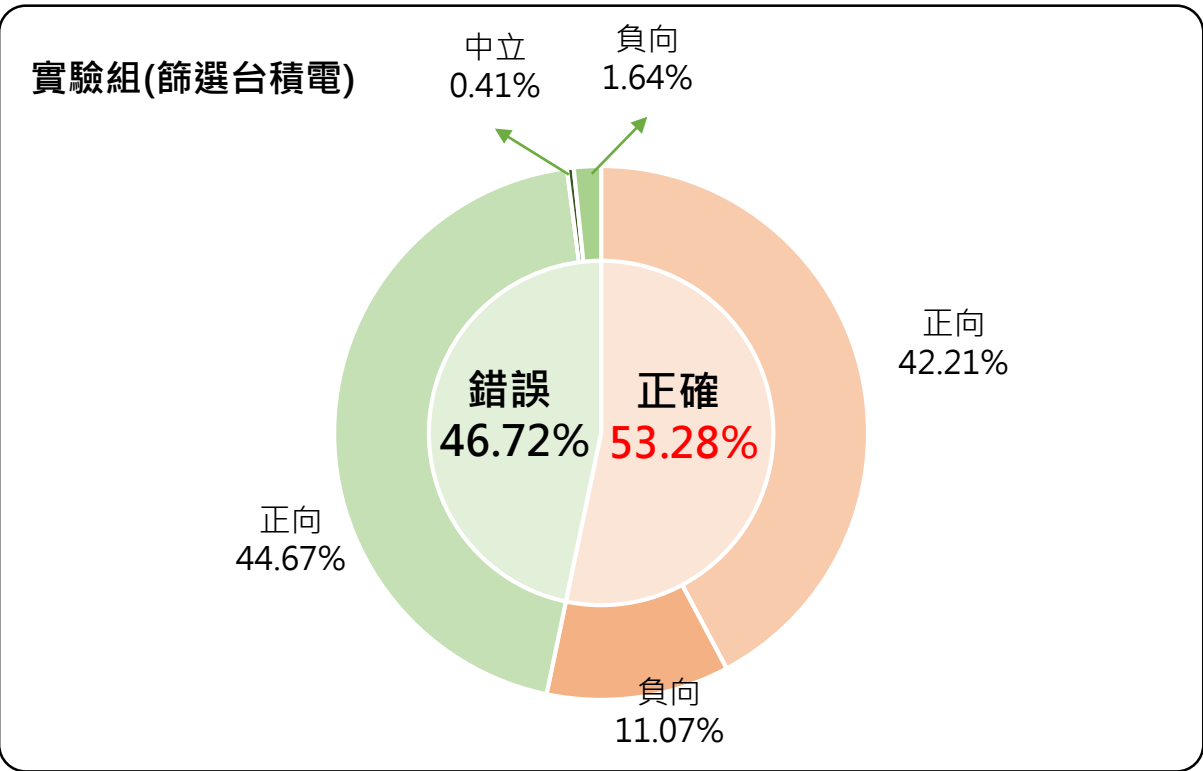
當日股價
精準度較高



② 系統準確度分析

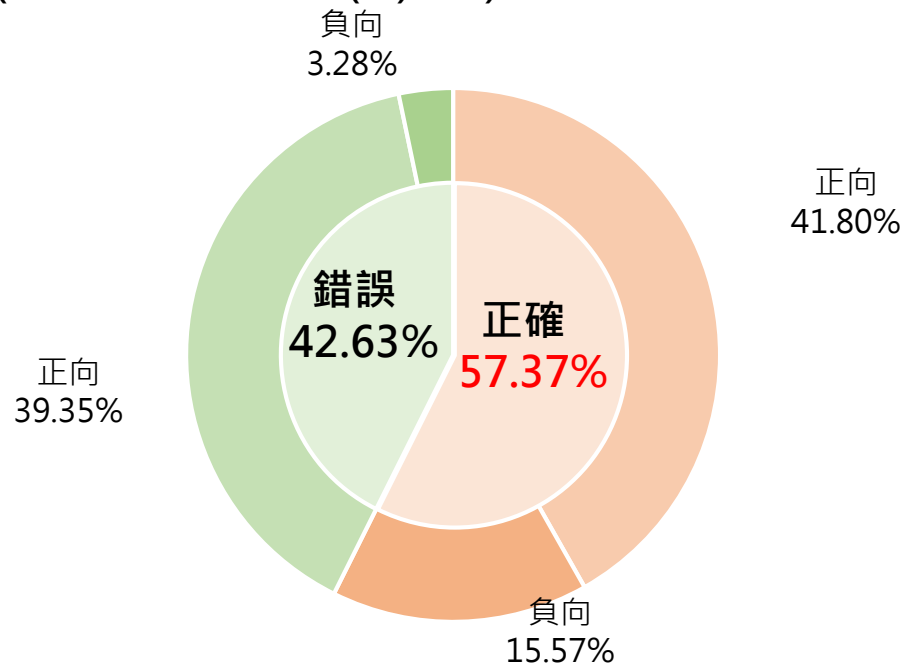


② 系統準確度分析

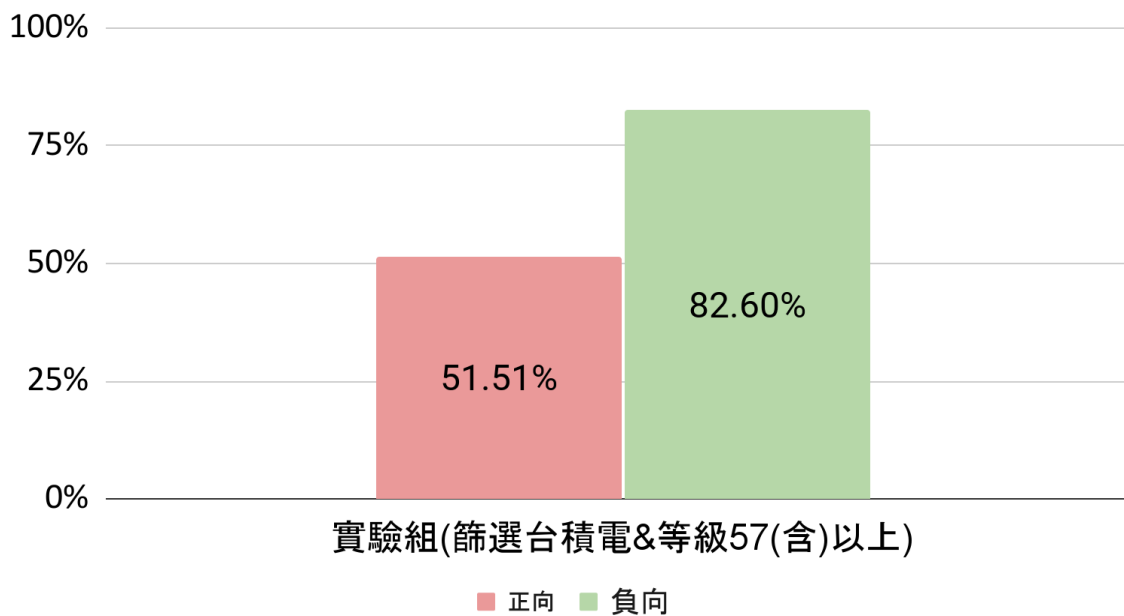


② 系統準確度分析

實驗組(篩選台積電&等級57(含)以上)



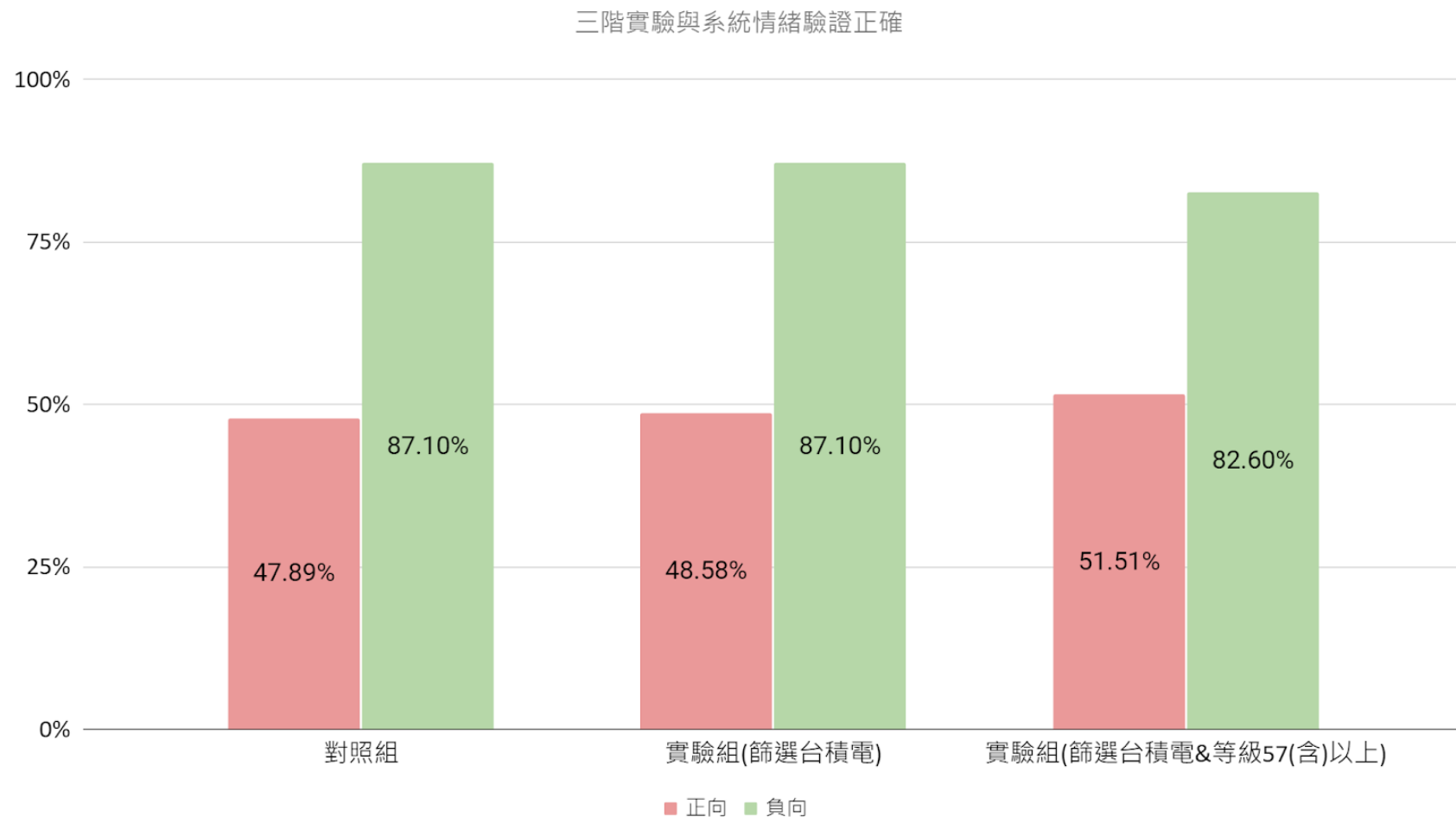
驗證正確個別分析





② 系統準確度分析

- 正向數量較多
- 負向正確率較高



總結

發文內容分析

1. 論壇使用者的發文情緒偏正面
2. 平台使用者在上午的活動較活躍
3. 高等級發文者文章情緒較明顯
4. 當天是否為交易日

對論壇使用者的發文內容及習慣都沒有太大影響



總結

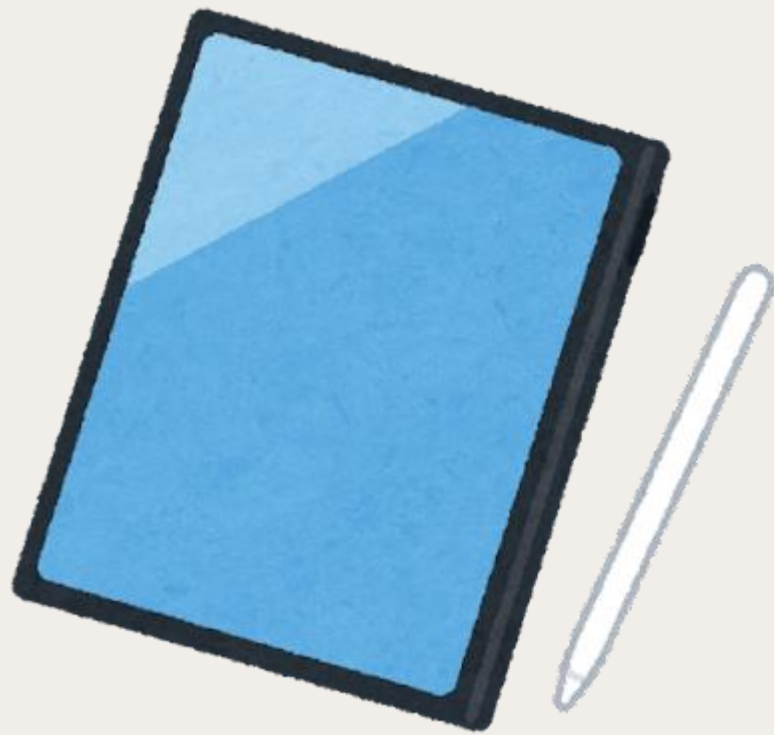
系統準確度分析

1. 社群情緒和當日股價較有關連
2. 社群情緒判斷為負向的正確率較高
3. 高等級發文者提及台積電的相關文章較有研究價值



建議及未來望

1. 斷詞斷句與製作字典
2. 給定正負向字詞動態權重
3. 結合機器學習
4. 尋找平均發文量較多的平台
5. 建議擴大CMoney資料樣本區間



- [1] [台灣證券交易所【歷年股票市場概況表】年報](#)
- [2] [台灣網路報告TWNIC](#)
- [3] [股票分析四大面向：基本面、技術面、籌碼面、消息面](#)
- [4] [Nearly 60% of Young Investors Are Collaborating Thanks to Technology, Often Turning to Social Media for Advice](#)
- [5] 朱夢琚，蔣洪汎，許偉. (2016). 基於金融微博情感與傳播效果的股票價格預測.
- [6] 文字探勘與機器學習於股票市場的應用與三大步驟
- [7] [信心理論](#)
- [8] 顏士杰. (2021). 透過傾向分析進行股價趨勢預測的實務經驗分享.
- [9] [內容分析法](#)
- [10] Yu-Teng Su. (2019). Research on the Analysis of Marketing Strategy of Fitness Club by Content Analysis.
- [11] Sheng-Cheng Chung. (2019). A Study on the Development of Hydropower Industry by Content Analysis.
- [12] [美國金融研究公司台積電全球地位文章](#)

[13] [爬蟲技術](#)

[14] 张晔 孙光光 徐洪云 庞婷 曲潇洋, (2020) . 国外科技网站反爬虫研究及数据获取对策研

[15] [從word2vec到情感分析](#)

[16] [word2vec簡介](#)

[17] Es, S. (2020) . Sentiment Analysis in Python: TextBlob vs. Vader Sentiment vs. Flair vs. Building It From Scratch.

[18] Shihab Elbagir and Jing Yang. (2019) . Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment

[19] [資料探勘與文字探勘之比較](#)

[20] Sung-Shun Weng. (2008) . Using Support Vector Machine and Text Mining For Stock Price Trends Prediction.

[21] Nielsen, A. (2020) . Practical Time Series Analysis: Prediction with Statistics & Machine Learning, O' Reilly

[22] Uhr, P., J. Zenkert, and M. Fathi. (2014) . Sentiment Analysis in Financial Markets,' ' 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC) : San Diego, CA, USA.

- [23] Bhati, R. G. (2020) . Sentiment Analysis: a Deep Survey on Methods and Approaches. Int' l Journal of Disaster Recovery and Business Continuity. Vol. 11, No. 1, pp. 503-51.
- [24] Bohmian. (2020) . Sentiment Analysis of Stocks from Financial News Using Python.
- [25] Briggs, J. (2020) . Sentiment Analysis for Stock Price Prediction: How we can predict stock price movement using Twitter.
- [26] Lin Yu. (2020) . Pricing Anomaly from the Text Sentiment in Social Community Forum.
- [27] 莊凱翔. (2018) . The prediction of trend toward stock price by text mining and sentiment analysis on social media: Using SVM and LDA Algorithm