



國立臺灣科技大學
資訊管理研究所
研究方法期末報告

台積電與台灣加權指數金融社群情緒對投
資人行為影響之分析

**Analysis of the impact of financial
community sentiment on investor behavior
in TSMC and Taiwan Weighted Index**

研究生：黃雅婷
學號：M11209202

中華民國113年6月

摘要

數位科技的飛速進步及通訊設備的便利性，使得全球人們能透過網際網路在社群媒體上進行無地域、無時差限制的即時交流，社群媒體不僅是討論人文、社會和經濟發展議題的重要傳播平台，同時也是金融理財討論的重要場域。

過去人們只能透過新聞、報章雜誌或公開資訊觀測站提供的重大訊息來了解股市現況，如今投資者們能夠自由在金融理財社群中分享股票操作經驗並討論市場趨勢，因此除基本分析和技術分析等方式外，消息面對經濟、股價帶來的影響也在科技發展迅速的背景下日益劇增，金融訊息傳遞與接收管道的改變也影響了大眾投資行為，從新聞、網路社群各方得知的消息都會影響投資人對股票的預期心理和操作態度。

本研究旨在利用機器學習技術對社群媒體中台積電和台灣加權指數的評論和貼文進行情緒分析（Sentiment Analysis），並結合焦點小組來深度分析社群媒體所產生的社群情緒對投資人交易操作實際的影響程度和預期心理的影響，透過這些結果調整模型中情緒字詞的權重、以提升分析股價漲跌的準確度。

透過社群情緒分數與焦點小組的討論結果，進行質化和量化方法分析台積電與台灣加權指數股價之社群情緒關聯性，最後利用模型做出股市預測，以幫助投資人即獲取準確的股票資訊，提升其投資決策的準確性和效益。

關鍵字：台積電、台灣加權指數、社群媒體、情緒分析（Sentiment Analysis）、焦點小組（focus group）。

目錄

| | |
|------------------------|----|
| 摘要 | 2 |
| 目錄 | 3 |
| 圖目錄 | 4 |
| 表目錄 | 4 |
| 第一章 緒論 | 5 |
| 1.1 研究背景與動機 | 5 |
| 1.2 研究目的 | 7 |
| 1.3 研究架構 | 8 |
| 第二章 文獻探討 | 9 |
| 2.1 台積電與台股加權股價指數 | 9 |
| 2.2 社群情緒的影響 | 10 |
| 2.3 股市情緒分析 | 11 |
| 2.3.1 傳統媒體情緒分析 | 11 |
| 2.3.2 社群媒體情緒分析 | 12 |
| 第三章 研究方法 | 13 |
| 3.1 文本資料蒐集 | 13 |
| 3.2 資料前處理 | 13 |
| 3.3 社群文章詞典 | 14 |
| 3.4 專業文章詞典 | 14 |
| 3.5 情緒詞典分類 | 15 |
| 3.6 焦點小組 | 16 |
| 3.6.1 焦點小組目的 | 16 |
| 3.6.2 焦點小組的建立 | 17 |
| 3.6.3 焦點小組的對象選擇 | 17 |
| 3.6.4 分組人數與進行方式 | 17 |
| 3.6.5 訪談結構設計 | 18 |
| 3.6.6 焦點小組討論主題 | 18 |

| | |
|--------------------------------|----|
| 3.7 焦點小組結果應用於情緒分析模型之有效性驗證..... | 20 |
| 第四章 預期結果 | 21 |
| 參考文獻 | 22 |

圖目錄

| | |
|-------------------------|----|
| 圖 1.1. 研究架構流程圖 | 8 |
| 圖 3.1. 社群詞彙產生的方法 | 17 |
| 圖 3.2. 模型準確度優化與比較 | 17 |

表目錄

| | |
|----------------------------------|----|
| 表 2.1. 台灣加權指數市值佔比排行前 10 股票 | 9 |
| 表 3.1. 資料前處理範例 | 13 |
| 表 3.2. 社群文章詞典範例 | 14 |
| 表 3.3. 專業文章詞典範例 | 15 |
| 表 3.4. 正負向情緒詞典範例 | 15 |
| 表 3.5. 中立情緒詞典範例 | 15 |
| 表 3.6. 焦點小組主題列表 | 18 |
| 表 3.7. 參數調整項目 | 20 |

第一章 緒論

1.1 研究背景與動機

科技推動技術創新，使人類生活變得更智慧便利，各項技術發展在科技的幫助下進步飛速，進而促使世界連結更緊密。其中半導體產業為奠定科技進步的基礎，半導體製程產出的晶片元件應用範圍如雲端運算、物聯網、智慧家電等，近幾年熱度及技術越發穩定的自動駕駛汽車、人工智慧、生成式AI等多個重要領域都屬於其應用範圍，由此可見半導體技術可謂數位化世界不可或缺的重要先驅。

2022年，美國政府通過《晶片與科學法案》（CHIPS and Science Act），2024年2月，美國白宮宣布撥款鉅額執行《晶片研發計畫》（HIPS Research and development（R&D） programs），透過對半導體產業投資以激勵發展，進而加強美國整體經濟和供應鏈彈性，此法案間接促使全球半導體企業開始湧入大量的投資基金，也給產業帶來發展潛力可觀的前景。

科技發展也一併帶動了各國的經濟發展，近年來台灣國內商業、經濟也熬過了疫情的衝擊逐漸趨於穩定，特別是台灣股市交易活動也逐年熱絡起來，根據臺灣證券交易所統計之歷年股票市場概況表顯示，國內股市成交總金額從2019年的新台幣26兆元逐年上升至2023年的新台幣63兆元，股市的活躍程度也反映出股票這項投資工具越發受到投資人青睞。說到台股，不得不提及台灣股市中有「護國神山」美稱之台積電股票，作為台灣最重要的半導體代工企業，台積電不僅在技術創新方面領先，對國內金融經濟帶來的影響也十分顯著。

在全球市場對於半導體產業極度重視的背景下，推動產業發展之力量龐大，台積電在全球半導體供應鏈中的重要性也將進一步提升，而台積電股票對台灣股市帶來的影響也隨著產業發展產生更顯著的變化。

現代人們的日常交流逐漸社群化，使用者在社群媒體中討論的內容也包含多種面向，除了日常生活基本的食、衣、住、行、育、樂外，也在金融經濟、政治外交、

科技人文等領域有諸多的討論。將社群媒體中的討論內容加以分析後，可以在不同領域實際應用，例如透過產品點擊率與搜尋內容讓企業更了解使用者偏好，對特定顧客進行精準的廣告投放，達到更好的行銷效果；在政治方面，也有利用網路投票預估選情，判斷不同地區選民意向的案例。

作為人們創作、分享、交流意見和觀點及經驗之平台，社群媒體能快速反應人們對事物的看法，其中討論台灣股市的社群平台如PTT、CMoney、Histock、Dcard股市版、鉅亨網.....，使用者常在理財相關的社群平台分享股市預測或交易結果。根據美網MagnifyMoney，在2021年對1,536名18至40歲受訪者的調查結果顯示，40歲以下的投資者中，有六成的人是金融理財論壇的會員，說明投資人會在理財社群平台活動、參考平台中的投資建議或大眾評論。且有兩成的投資人會同時在多個社群平台瀏覽貼文、留言作為個人投資參考依據。可見除新聞媒體及報章雜誌等傳統媒體外，現代投資人也在金融網站或理財社群平台獲取股市新資訊。

影響投資人選股及評價的股票分析方法有基本面、技術面、籌碼面和消息面等面向分析，由於社群網路已成為現代社會中重要的訊息交流平台，「消息面分析」所帶來的影響也在通訊網路普及的時代背景下對股市的影響更深遠，因此，分析社群網路中的情緒對於理解市場趨勢和做出投資決策具有重要意義，投資人會在平台上瀏覽產業資訊和社群輿論，並結合自身的金融知識進行股票交易，這些投資決策結果最終將反映到股市上。

情緒分析是自然語言處理（NLP）中的一個重要應用，傳統上使用基於詞典的方法或簡單的機器學習技術來分析文本情緒。隨著生成式AI（Generative AI）和大型語言模型（Large Language Model）的出現，情緒分析的準確性和適用性得到了顯著提升。這些新技術利用深度學習和大量預訓練資料，能夠更準確地捕捉文本中情感語境的細微差別，為研究提供更可靠的情緒分析結果。

本研究將結合焦點小組和機器學習應用在情緒分析的技術，對台灣股市的台積電及台股加權指數的討論內容進行分析，利用焦點小組的方式對不同主題進行深入討論，分析社群情緒對投資人和整體市場的影響，最後將焦點小組的研究成果作為

後續模型調整的參考，以調整模型中的情緒字詞、各項參數的權重並考慮各項因素之相關性，以提升情緒分數計算的準確性，並探索社群情緒與整體股價走勢之間的關聯。

新技術的應用不僅能提升情緒分析的準確性，還能為投資者提供更有價值的見解，有助於投資人綜觀全局、快速分析社群媒體中所含的情緒狀態，以做出更明智的投資決策。通過研究社群媒體中的社群情緒，為投資分析提供新的工具和方法，並推動相關領域的研究和應用，探究網路輿論對股市的真實影響，協助投資人進行投資決策。

1.2 研究目的

本研究旨在分析社群媒體中有關台積電和台股加權指數討論版上使用者活動內容製成文本，對文本進行情緒分析產生社群情緒分數，並分析社群情緒對股市波動和投資人行為產生的影響。

透過在社群平台中建立獨立的討論區，結合焦點小組（focus group）的研究方式，蒐集投資人對社群媒體中大眾討論股票消息的影響，透過不同的討論主題，實際蒐集出對投資人及大眾而言較容易受影響的金融社群情緒字詞，為情緒分析模型建立字詞權重的部分提供有效的基礎，加速模型產生的速度和提升情緒分數的準確度，由於社群網路中網路用語的更新相當快速，期望透過焦點小組和生成式AI的方式掌握最新的社群情緒字和股市新網路用語，更完整優化模型及掌握股市資訊。

本研究的主要目的：

一、社群情緒對投資行為的影響

透過焦點小組的討論，深度分析社群情緒對投資行為的潛在影響，以便對後續預測股價漲跌的模型給與適當的情緒權重、提升模型優化的準確度，並且透過定期的焦

點小組討論，持續更新新的金融社群情緒用語，完整模型的情緒詞彙庫。

二、調整金融社群情緒詞典與情緒詞權重

延續目的一的結果，給予金融社群情緒詞彙有用的權重，包含常見的社群情緒詞彙和網路短語等，透過了解各項影響因素的相關性來更好的調整各項參數的權重，幫助提高情緒分析的準確性。

三、社群情緒分析模型

建立一個高效的情緒分析模型。此模型將被訓練來識別和分類社群媒體中有關台積電與台灣加權指數討論的情緒分數，包括正面、負面和中立情緒。

四、社群情緒與股市活動之關聯

將情緒分析結果與實際股市進行比對，探討社群情緒變化與股市之間的關聯性，分析情緒波動是否能預測股市的短期或長期走勢、交易量的波動變化等股市交易活動。

1.3 研究架構

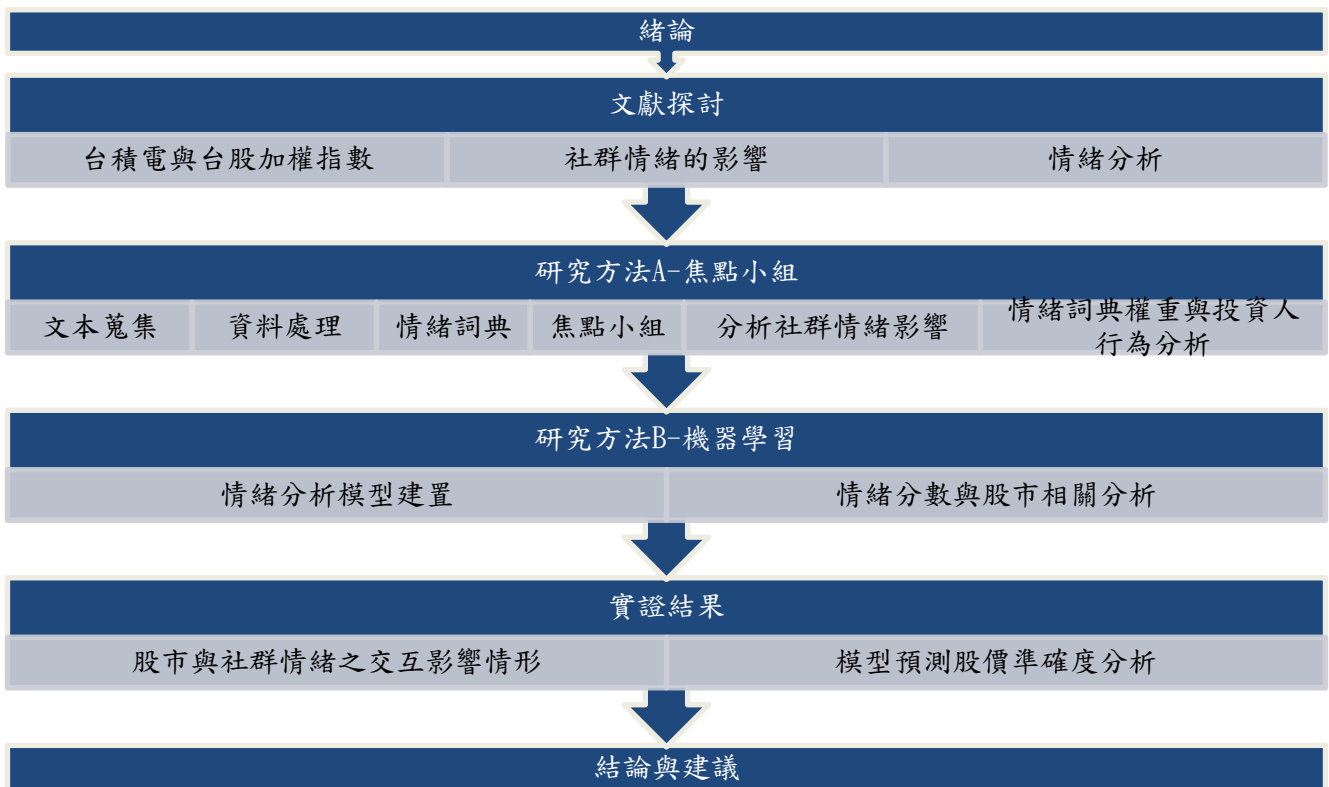


圖 1.1. 研究架構流程圖

第二章 文獻探討

2.1 台積電與台股加權股價指數

股票指數（stock index），又稱股價指數、股市指數，用於衡量整體股市表現的指標，股票指數的計算方式主要分為價格加權或市值加權，而各國股市皆有屬於自己的股價指數，如美國的道瓊工業指數、NASDAQ指數、S&P500等股價指數，還有歐洲的DAX指數、CAC-40和亞洲的富時中國A50指數、日經225指數等。

台灣股市也有台灣加權股價指數（Taiwan Capitalization Weighted Stock Index, TAIEX），常被國人簡稱為「大盤」，是台灣證券交易所（TWSE）編製的綜合性股票市場指數，用來衡量台灣股市整體的表現。

台灣加權指數是由台灣股市中具有代表性的上市公司股票所組成，每家公司的影響力取決於其市值，市值較大的公司在指數中的所佔權重就會較大，在台灣加權指數中市值佔比最重為台積電（2330）、第二為鴻海，緊接著是排名第三的聯發科，而大盤走勢反映了組成股票的整體波動情況，也代表著台灣股市整體漲跌的縮影。

表 2.1. 台灣加權指數市值佔比排行前10股票

| 排行 | 證券代碼 | 證券名稱 | 市值占大盤比重 |
|----|------|------|----------|
| 1 | 2330 | 台積電 | 31.7881% |
| 2 | 2317 | 鴻海 | 3.3553% |
| 3 | 2454 | 聯發科 | 2.462% |
| 4 | 2382 | 廣達 | 1.5581% |
| 5 | 2412 | 中華電 | 1.4924% |
| 6 | 2881 | 富邦金 | 1.3953% |
| 7 | 2308 | 台達電 | 1.2916% |
| 8 | 2882 | 國泰金 | 1.1493% |
| 9 | 6505 | 台塑化 | 1.0671% |
| 10 | 2891 | 中信金 | 1.0382% |

台灣加權股價指數反映了整體股市表現，是投資人判斷股市情況的重要指標，因此大盤漲跌對投資人的心理預期會產生一定的影響，可能在大盤指數上漲時，投資人會對股市呈現樂觀的態度，就更願意進行金融投資行為或對台灣股市抱有信心；反之當大盤指數下跌時，投資者心態就容易變得保守，對有興趣的股票採取觀望或出售的投資策略。

我們也閱讀了針對加權指數進行分析的研究文獻，如劉照群（2008）對美國與台灣股市關聯性進行分析，在研究發現費城半導體指數、台灣加權指數與台積電股價三者之間存在長期均衡關係。

Liao, Kao和Yuan的研究中利用深度學習技術中的卷積神經網路（CNN），來對台灣加權股價指數（TAIEX）進行股價預測。透過將台灣加權股價指數（TAIEX）的日金融數據轉換為圖像數據，並使用Xception模型進行訓練，結合Gramian技術將時間序列數據轉化為極座標表示，提升了模型對股市趨勢的預測準確性。

對美國的S&P500加權股價指數也有相關研究，CHEN（2022）利用Glassdoor的員工評論作為文本，並結合Yahoo Finance和Macro Trends的數據，使用（Random Forest）、極限梯度提升（EGB）等多個模型技術來預測股價走勢，並試圖找影響股價的因子。Bollen等人分析社群網站Twitter中的正負面情緒，最後透過回歸的模糊神經網路預測道瓊指數的收盤價。

2.2 社群情緒的影響

由於人們接受和傳遞訊息的管道改變，網路社群對人們生活息息相關，而社群情緒對現實世界的影響也越來越大，而社群中正面和負面情緒對人們的影響也不盡相同。

Ferrara（2015）的研究發現，社群情緒會影響人們對事件或現象的看法，進而影響人們決策，容易共情、受情緒影響的人更傾向採納正面情緒，且正向情緒的蔓

延機率會大於負面情緒。但負面情緒如憤怒、恐懼等，若在社群中快速蔓延，可能導致人們的偏激行為（Ruzzene, 2022）。正面情緒例如喜悅、感激等，則有助於增進人際關係和社會凝聚力（Wang et al., 2021）。

此外，Stieglitz和Dang-Xuan（2013）的研究發現，社群情緒也會影響現實政治走向，如同大量的社群負面情緒可能引發社會運動，而正面情緒則更能穩定社會體制。在經濟層面，Baker等人的論文發現（2020），社群中若是充滿著負面情緒，可能會抑制現實世界的經濟投資和整體消費行為，可見社群中的大眾負面情緒對經濟發展、成長都會帶來不小的阻力。

網路社群平台內容對現實世界產生的影響包括社會、政治、經濟等各個層面，因此瞭解和分析社群情緒的重要性就顯而易見。

2.3 股市情緒分析

2.3.1 傳統媒體情緒分析

社群媒體是自由、碎片化、具時效性的討論平台，與傳統媒體如新聞、報章雜誌等不同，社群媒體的交流特點之一是用字較不嚴謹、充斥著各種流行用語，使用者能隨時在社群平台上表達當下的想法及最新動態。

過去的文獻也有對傳統媒體做情緒分析的研究，CHEN, CHIAN-YU（2022）的研究中以公開資訊觀測站的重大訊息為文本，利用三種詞典：財經領域情緒詞典（FESD）、中文財務情緒詞典（CFSL）、中文金融情感詞典（CFSD）等對文本分析，但在研究中單純是利用累積關鍵詞的字數去判斷情緒值，不過公告消息中情緒詞的比例較少，所以在判斷情緒的結果較不準確，且在研究中發現相較於PTT、facebook、CMoney三平台中，CMoney是投資者情緒指數最大的。

Juan,Zong-Yong（2022）的研究中利用Bert對中大型股票的新聞資料進行分析，

發現了對情緒詞做加權的處理方式能將準確率提高，但也要考慮文本的內容特點，發生重大事件、產業環境改變、新聞寫作風格等都會影響分析結果，另外發現「詞頻低影響大的詞」對情緒分數的影響容易被「詞頻高但較無意義的詞」所稀釋，故文本、情緒加權、詞頻調整等對於預測結果有顯著影響。

2.3.2 社群媒體情緒分析

隨著社群媒體的迅速發展和網路流行術語的頻繁變化，社群更新對投資者的情緒產生了直接影響，下列文獻分析了社群媒體中的情緒與股市波動的關聯，且隨著科技進步，分析情緒的工具與方法也有了顯著的變化。

過去沒有機器學習技術，人們只能透過人工判讀的方式進行情緒分析，Bryan (1997) 透過人工的方式分析MD&A文章。Cole & Jones (2004) 同樣閱讀MD&A文章，但人工閱讀分析非常耗時也無法處理大量資料。

近年來社群情緒分析的相關研究不斷增加，Fan, Y., Zhou, F., An, Y., Yang, J. (2021) 等人的研究中發現情緒對股市有影響，甚至有可能帶來股市崩盤的極端可能性。另外，Shangkun Deng 等人 (2023) 利用整合機器學習對股市進行預測，主要整合Boosting、Bagging、和 NSGA-II 方法，而結果也證明了此方法對於股價指數和模擬交易的預測準確度高於基準方法。Nandini Jayaram等人 (2021) 研究對Twitter上投資相關情緒與用戶活動的關聯性，發現社群情緒與股價波動之間存在正相關。

第三章 研究方法

3.1 文本資料蒐集

根據前述提及，「金融社群網站CMoney」是前研究中實證結果情緒指數最大平台，故我們選定此CMoney作為資料蒐集的平台，將「台積電個版」和「台股大盤」中有關台積電（2330）和台灣加權指數（TWA00）的發文內容、留言等資料透過爬蟲技術進行蒐集。

為了讓模型建立足夠的訓練資料集，且考慮到疫情期間對股市衝擊帶來的特殊影響，我們將爬蟲資料樣本區間涵蓋疫情開始以及疫情影響趨緩後，資料蒐集區間設定為2018年4月30日至2024年4月30日，由於股市未開市時並無股價漲跌數據，在後續研究中的部分分析將排除股市未開市的天數。

3.2 資料前處理

本研究使用中央研究院中文詞知識庫小組的套件CkipTagger（又稱Ckip）來進行斷詞、停用詞、標點符號處理、詞性標註等前置處理。制定分類邏輯確保判斷一致性，再製作成社群情緒詞典，作為模型辨別文章情緒的依據。

表3.1. 資料前處理範例

| 文章範例 | |
|------|---|
| 範例一 | 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 |
| 範例二 | 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 | |
| 範例一 | [['意外', '倒', '賣單', '加堆', '買單', '人', '掛', '假單', '掛', '倒', '了', '意思'], |
| 範例二 | ['期指', '高', '才', '繼續', '攻', '夠', '強', '今天', '期指', '破', '還好', '關鍵', '需要', '時間', '整理', '上面', '缺口', '回補', '看', '太', '壞'] |

3.3 社群文章詞典

本研究利用文本資料製作社群文章詞典，使用者在社群中的發文內容不似較專業的報章雜誌，社群中的用詞大多都較口語化，同時充斥著大量的網路用語、流行用語，故先結合大型語言模型(LLM)、生成人工智慧(GAI)等方式將社群字詞詞典蒐集、更新，最後將被斷開的字詞重新檢回，透過這個動作，斷詞模型將更新斷詞的方式，將斷詞的流程進行的更精確、降低字詞被斷開的可能，其中檢詞也考慮了上下文影響，如拉/不拉、漲/不會漲、站穩/不會站穩等類似的上下文字詞，若被斷開則有不同含意。

表3.2. 社群文章詞典範例

| 社群文章詞典範例 | | | | | | | |
|----------|-----|------|------|------|------|-------|--------|
| 不拉 | 業外 | 漲跌- | 台積電+ | 百德集團 | 開低走低 | 資本利得 | 任重而道遠 |
| 小跌 | 護盤 | 滿水位 | 台積電- | 借券賣出 | 開低走平 | 機會財股 | 股價淨值比 |
| 多單 | 續抱 | 隔日沖 | 加密貨幣 | 烏俄戰爭 | 開平走高 | 融券賣出 | 百元俱樂部 |
| 走跌 | 避險 | 會不會 | 月線往上 | 租賃三雄 | 開平走低 | TSM+ | 景氣循環股 |
| 抗俄 | 不看漲 | 噴起來 | 月線向上 | 航運三雄 | 開平走平 | TSM- | 殖利率倒掛 |
| 站回 | 不看跌 | 潛力股 | 中興保全 | 逢低布局 | 跌不下去 | IC設計 | 費城半導體 |
| 被嘎 | 不看好 | 競爭者 | 中華電信 | 通貨膨脹 | 超額利潤 | 護國神山 | TSM + |
| 崩盤 | 不是賣 | 不要買進 | 再度站上 | 通用電器 | 買賣超- | 櫃買市場 | TSM - |
| 做東 | 3奈米 | 不要再追 | 回補空單 | 貨櫃三雄 | 統一投顧 | 獲利了結 | 聯電ADR |
| 追新 | 不會漲 | 大盤指數 | 波段交易 | 貨幣緊縮 | 程式交易 | 千元俱樂部 | 台積電ADR |

3.4 專業文章詞典

除網路用語外，社群討論內容也會充斥著各種專業術語。本研究參考許多機構公開的產業報告書、半導體產業調查報告及多篇經濟期刊等專業文章，將台灣半導體產業中常見的專業字詞製成專業文章詞典。

表3.3. 專業文章詞典範例

| 專業文章詞典範例 | | | | | | | |
|----------|------|------|------|------|-------|--------|---------------------|
| 封城 | 醫療設備 | 每股盈餘 | 數位科技 | 世界先進 | 邏輯IC | 美中貿易戰 | 先進駕駛補助系統 |
| 新冠 | 新興技術 | 現金股利 | 總體經濟 | 戰略地位 | 車用晶片 | 美中科技戰 | 3D Fabric |
| 遷廠 | 財務預測 | 領導地位 | 集成電路 | 供需失衡 | 量子電腦 | 中美科技戰 | 美國半導體工業協會 |
| 5G | 高階主管 | 邏輯密度 | 地緣戰略 | 景氣循環 | 量子通訊 | 智慧型手機 | 美國半導體行業協會 |
| 缺貨 | 全球經濟 | 異質整合 | 新冠肺炎 | 晶圓代工 | 車用市場 | 筆記型電腦 | 美國半導體產業協會 |
| 遠端 | 遠距學習 | 矽中介層 | 先進封裝 | 成熟製程 | 垂直整合 | 報復性消費 | 雞蛋放在同一個籃子 |
| 2奈米 | 地緣政治 | 利益平衡 | 去中心化 | 銷售折讓 | 商業模式 | 市場滲透率 | 世界半導體貿易統計組織 |
| 供應鏈 | 基礎架構 | 中國大陸 | 分散風險 | 銷貨折讓 | 半導體產業 | 車用半導體 | FTSE4 Good |
| 7奈米 | 潛在市場 | 公衛機構 | 外銷導向 | 平板電腦 | 技術差異化 | 三維積體電路 | Mitsubishi Electric |

3.5 情緒詞典分類

對詞典進行正向、負向分類，中立詞彙等。

表3.4. 正負向情緒詞典範例

| 正向詞 | | | | 負向詞 | | | |
|-----|----|----|------|-----|----|-----|-------|
| 攻 | 拉高 | 漲勢 | 買回來 | 弱 | 扼殺 | 跌幅 | 不要買進 |
| 抱 | 長多 | 漲價 | 漲上來 | 挫 | 走弱 | 落後 | 不會上漲 |
| 拉 | 看好 | 領先 | 漲停板 | 割 | 走跌 | 緊縮 | 台積電- |
| 強 | 看漲 | 領漲 | 漲跌+ | 跌 | 軋空 | 賠錢 | 回補空單 |
| 買 | 突破 | 增長 | 噴起來 | 賣 | 看跌 | 賣出 | 忐忑不安 |
| 漲 | 留倉 | 樂觀 | 不會升息 | 下挫 | 重挫 | 賣掉 | 貨幣緊縮 |
| 上去 | 站上 | 獨強 | 月線向上 | 下殺 | 降溫 | 賣超 | 通貨膨脹 |
| 上揚 | 站回 | 獲利 | 月線往上 | 下跌 | 倒掛 | 賣壓 | 無法站穩 |
| 上漲 | 站穩 | 賺到 | 台積電+ | 下彎 | 除息 | 壓回 | 買賣超- |
| 大單 | 做多 | 賺錢 | 再度站上 | 大跌 | 做空 | 縮表 | 漲跌 - |
| 大買 | 強勁 | 擴產 | 均線向上 | 小跌 | 停利 | 虧損 | 獲利了結 |
| 大漲 | 強勢 | 翻紅 | 均線往上 | 不利 | 停損 | 轉弱 | TSM- |
| 止跌 | 創高 | 轉強 | 供不應求 | 不拉 | 通膨 | 不看好 | 不會再創高 |

表3.5. 中立情緒詞典範例

| 中立詞 | | | | | | | |
|-----|----|----|----|----|----|----|----|
| 市場 | 利率 | 交易 | 風險 | 預算 | 平穩 | 配股 | 報告 |
| 投資 | 貿易 | 負債 | 現金 | 配息 | 股東 | 評估 | 資訊 |
| 股票 | 基金 | 資產 | 支出 | 市值 | 分析 | 估值 | 研判 |
| 貨幣 | 收益 | 水位 | 股價 | 併購 | 趨勢 | 盈虧 | 策略 |
| 資本 | 波動 | 財務 | 波動 | 可能 | 持股 | 預測 | 變動 |

3.6 焦點小組

3.6.1 焦點小組目的

本研究透過焦點小組的研究方式深入了解投資者對社群媒體內容的情緒反應及社群情緒對實際投資決策的影響，討論發文內容、發文者、文章熱度等各項變數是否會對投資行為產生不同的影響，並且由於過去文獻無法及時蒐集最新的社群用詞，本研究希望透過焦點小組和生成式AI的方式即時掌握最新的社群情緒字和股市新的網路用語。

透過此研究方法蒐集投資者認為重要的情緒詞和網路短語，來更精準的判別文章情緒，並將焦點小組的研究結果作為後續情緒分析模型調整的依據，利用焦點小組找出社群情緒對股市的影響，後續將結果加入模型中，驗證情緒對模型預測效果的影響程度與準確度。

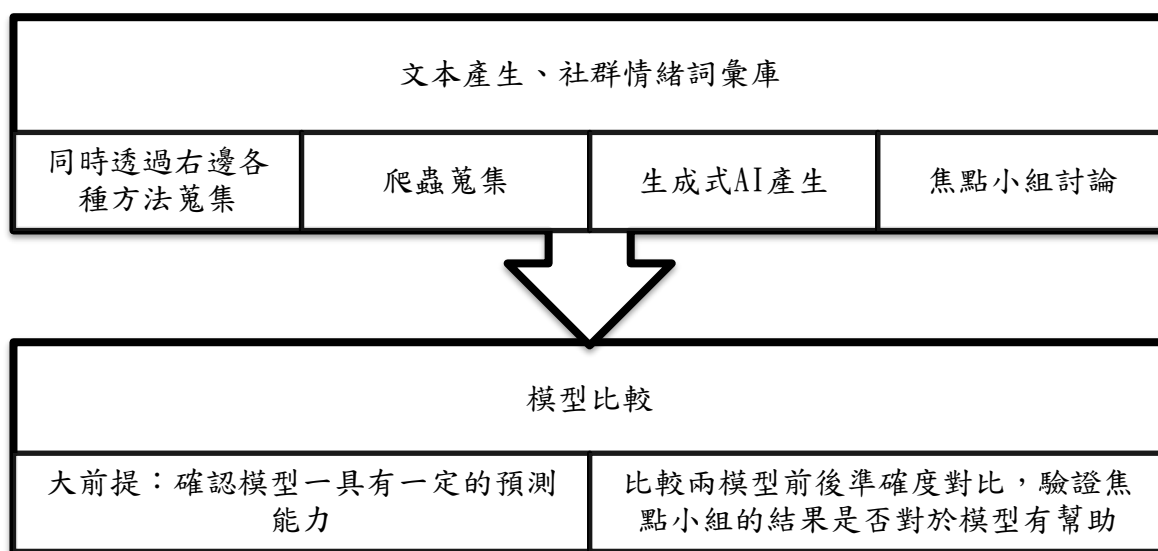


圖3.1. 社群詞彙產生的方法

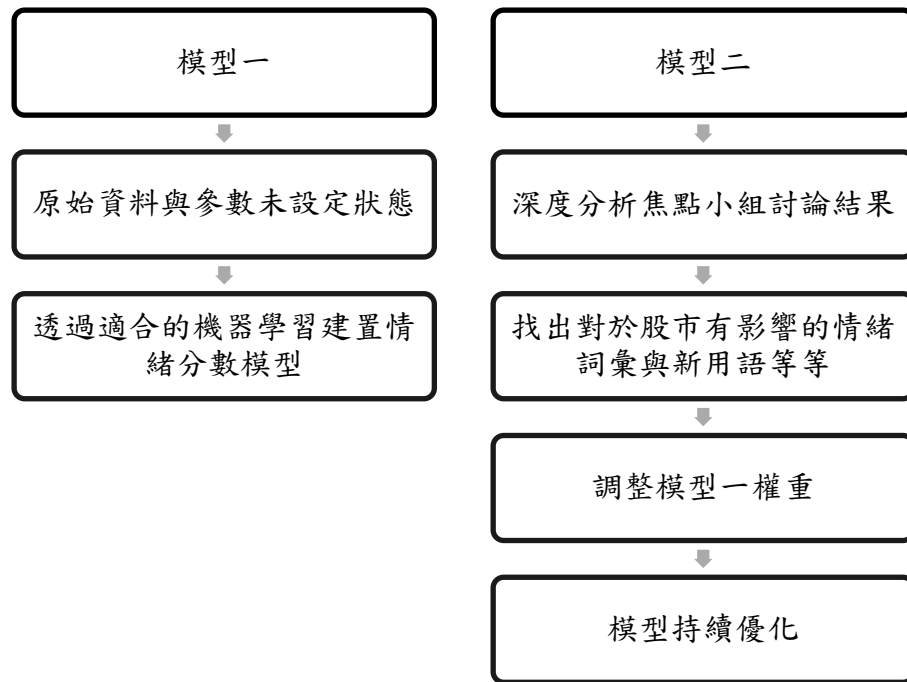


圖3.2. 模型準確度優化與比較

3.6.2 焦點小組的建立

本研究與CMoney平台合作，在平台上建立一個焦點小組討論區，透過線上論壇的方式進行焦點小組的資料蒐集及訪談。

3.6.3 焦點小組的對象選擇

透過CMoney提供的會員基本資料進行篩選，將會員的投資偏好、年齡、性別、職業、經濟情況加入為篩選依據，確保組成的焦點小組成員包括不同背景和經驗層次的投資者，確保研究結果代表多元的投資觀點。向這些會員發送焦點小組邀請，告知焦點小組的目的、進行方式等，確保焦點小組成員是自願且可以完全參與研究。

3.6.4 分組人數與進行方式

為了深入探索研究主題採用焦點小組的研究方式，預計將以每個主題為一場焦

點小組，每組由人數6-12人組成，每場的討論時間限制在60至90分鐘，活動在金融社群平台的討論區建立一焦點小組專區進行討論，過程中將全程透過軟體紀錄成文字檔案供後續分析，記錄如發言時間、發言者名稱（後續將發言者名稱經過匿名處理確保隱私不會外洩）、發言內容、發言者對應的個人資料（年齡、職業、投資偏好等），供後續深度分析結果。

3.6.5 訪談結構設計

為了避免參與人員提出的意見過於繁雜無法收斂，採取設計半結構式訪談，透過主持人主持焦點小組訪談，讓討論內容圍繞主題，但參與者的發言內容仍以開放式的方式進行，確保參與者能自由表達意見，並經由主持人引導彙整所有成員發表的意見與看法。

3.6.6 焦點小組討論主題

表3.6. 焦點小組主題列表

| | 主題 |
|----------|---|
| 1 | 投資人對正負向文章情緒的影響程度會相同嗎？ |
| 內容 | 若大眾情緒分數結果為中立，在判斷投資行為時會同樣中立觀望或是被正面、負面情緒的文章影響更多？ |
| 2 | 高詞頻情緒詞彙的影響程度及情緒標註。 |
| 內容 | 透過前述提及的情緒詞典結果展示給焦點小組成員，對平台中出現詞頻較高的字詞進行討論，並引導討論哪些詞彙最能影響他們的情緒反應、進而改變投資決策。 請成員對高詞頻詞彙進行情緒標註（正面、中立、負面）並對詞彙的情緒強度評分，作為後續模型中「高詞頻情緒詞」的權重調整基礎。 |
| 3 | 低頻率情緒詞彙的影響程度及情緒標註。 |
| 內容 | 對容易被忽略的中低頻情緒詞彙進行討論，將詞頻中低的詞彙展示，並引導討論是否有「詞頻低但影響大」的情緒詞，或是討論「低詞頻對投資人 |

| | |
|----------|---|
| | 情緒及投資行為影響也較低」的可能，作為後續模型中「低詞頻情緒詞」的權重調整基礎。 |
| 4 | 台灣加權指數與台股電社群情緒及股價互相影響的程度。 |
| 內容 | 分析是否會因素兩者之一股票漲跌而對另一股票的投資行為有影響、是否會交錯分析兩股票中社群情緒的變化，作為投資者預期股票漲跌的參考，及投資者對於兩股票之間影響程度的討論。 |
| 5 | 不同消息來源對投資者情緒變化的影響程度。 |
| 內容 | 分析社群媒體上哪些類型的消息（如公司財報、新聞事件、行業趨勢）最能影響投資者的情緒。社群媒體內容對投資決策的影響，分析投資者對不同來源的訊息影響程度。 |
| 6 | 社群媒體情緒與市場波動的關聯、反應時間長短。 |
| 內容 | 探討社群媒體上的情緒如何反映市場波動，並分析情緒變化與市場走勢之間的具體關聯，是否會因為平台情緒低迷或看好而馬上反映在市場上？或是因為社群情緒極好或極差而對投資策略長短的改變，分析平台情緒對投資者的實際投資行為的影響程度與影響時間長短，了解投資者如何看待社群情緒的變化，並討論這些變化對市場的潛在影響。 |
| 7 | 平台中不同發文者與貼文熱度對投資者投資行為或情緒的影響。 |
| 內容 | 討論不同發文者等級或貼文的熱門程度、按讚或留言情況是否會成為影響預期心理的主要因素？越多人關注越能代表大眾情緒？對後續模型調整發文者等級、留言、按讚等文章熱門度加入模型權重參考因素。 |
| 8 | 社群媒體與傳統媒體的影響力比較。 |
| 內容 | 討論社群媒體與傳統媒體（如報紙、電視）在影響投資者情緒和決策方面的異同，討論投資人對傳統媒體的觀看偏好、習慣及影響程度，將其與社群媒體的內容作比較。 |
| 9 | 金融社群新用語與用詞、網路用語蒐集與更新。 |
| 內容 | 討論社群中最新的網路用語與情緒詞彙，來更完善情緒分析預測股價的模型資料庫，且對於最新的情緒詞也討論情緒權重、影響程度，或是針對發想類似的情緒詞彙互相比較，對後續模型建置提供良好的基礎。 |

3.7 焦點小組結果應用於情緒分析模型之有效性驗證

為了確認焦點小組結果對模型準確度提升的程度，將對模型一和模型二進行準確度對比。

1. 原始模型資料對比

在焦點小組開始前，先對原始的情緒分析模型一（後面也將以「模型一」代稱原始模型）進行結果測試，記錄模型一對歷史數據的預測準確度，此模型一的預測結果在後續將用來比較參考焦點小組調整後的模型二預測準確度提升情況。

2. 資料蒐集與整合

透過焦點小組的討論結果，收集參與者提供最新的情緒詞彙和網路用語、情緒權重調整和各項可調整參數（發文者、按讚數等），並將這些參數整合更新到模型二中。具體調整參數包括：

表3.7. 參數調整項目

| 參數 | 內容 |
|-------|--|
| 情緒詞 | 將最新的情緒詞彙更新到模型資料庫，並根據焦點小組提供的正負向情緒詞影響程度，調整模型中各情緒詞彙的權重。 |
| 發文者等級 | 根據前章節分析結果，如發文者等級若越高則影響情緒程度越大，則調整有關發文者等級的參數初始權重。 |
| 文章按讚數 | 根據前章節分析結果，文章按讚數、留言數量越高則影響情緒程度越大，則調整按讚數、留言數量參數之初始權重。 |
| 留言數量 | |
| 文章內容 | 根據前章節分析結果，辨別社群媒體上哪些類型的消息（如公司財報、新聞事件、行業趨勢）最能影響投資者的情緒，對此參數的初始權重進行調整。 |

3. 模型再訓練與驗證

使用新模型（模型二）對歷史數據進行重新預測，並模型一結果進行對比，並

且持續的利用更多資料集對模型進行優化、調整，讓預測準確度提升。

準確度比較：比較更新後模型和初始模型在情緒分類和股市預測準確度上的差異。

雙盲測試：將兩組不同的模型分別應用於相同的數據集，評估其預測效果。

效果評估：使用統計方法檢驗模型預測準確度的差異是否顯著。

第四章 預期結果

預計利用研究方法一的焦點小組方法和其他研究方法同時進行蒐集文本的動作，並透過調整金融社群情緒詞典與情緒詞權重，結合質化與量化的研究法，了解各項影響因素的相關性來更好的調整各項參數的權重，幫助提高情緒分析的準確性。

透過建立高效、精準分析社群情緒的的情緒分析模型，找出社群情緒與股市活動之關聯，探討社群情緒變化與股市之間的關聯性，分析情緒波動是否能預測股市的短期或長期走勢、交易量的波動變化等股市交易活動。

參考文獻

- [1] Evans, Julie Ryan. (2021) . *Nearly 60% of Young Investors Are Collaborating Thanks to Technology, Often Turning to Social Media for Advice*. Magnifymoney. <https://www.magnifymoney.com/news/young-investors/>.
- [2] Liu, Chao-Chun. (2008) . An Empirical study of the Relationship Among Philadelphia Semiconductor Index in American, Taiwan Stock Index and TSMC Stock Price (Unpublished master's thesis) . National Taipei University , Taiwan, ROC.
- [3] Liao, C. H., Kao, T. L., & Yuan, S. M. (2020) . Forecasting Taiwan Capitalization Weighted Stock Index by Using Convolutional Neural Network. IEEE Xplore Logo. <https://ieeexplore-ieee-org.ntust.idm.oclc.org/document/9301956>
- [4] Chen, Kuan-Yu. (2020) . Predicting Stock Price Trends using Social Reviews: Evidence from the S&P 500 Firms, (Unpublished master's thesis) . National Chung Cheng University, Taiwan, ROC.
- [5] Ferrara , E., & Yang , Z. (2015) . Measuring Emotional Contagion In Social Media. *Plos One*.
- [6] Ruzzene, A. (2022) . Exploring the role of moral emotions in online hate speech. *Social Science Computer Review*, 40 (2) , 377-396.
- [7] Wang, H., Wang, L., Su, F., Yi, X., & Tsou, B. K. (2021) . How do emotion expressions transform interpersonal resonance? Exploring the emotional influence process in online social interaction using facial expression analysis. *International Journal of Information Management*, 57, 102293
- [8] Stieglitz, S., & Dang-Xuan, L. (2013) . Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of management information systems*, 29 (4) , 217-248.

- [9] Baker, S. R., Bloom, N., Davis, S. J., & Sammon, M. (2020) . What triggers stock market jumps?. Available at SSRN 3680913.
- [10] Chen, Chian-Yu. (2022) . A Study of Sentiment Analysis on Financial Material Information, (Unpublished master's thesis) . Feng Chia University , Taiwan, ROC.
- [11] Juan, Zong-Yong. (2022) . Use BERT to Mine Medium and Large-Cap Stock News Data to Explore The Rise and Fall of Its Stock Price - Taking TSMC as An Example, (Unpublished master's thesis) . National Yunlin University of Science and Technology , Taiwan, ROC.
- [12] Bryan, S. H. (1997) . Incremental information content of required disclosures contained in management discussion and analysis. *Accounting Review*
- [13] Nandini Jayaram (2021) The Rising Power of the Individual Investor: How Social Media Sentiments and User Activity Impact Stock Price Volatility and Trading Volume
- [14] Fan, Yunqia, Zhou, Fangzhao, An, Yunbib, An Y. & Yang, Junc (2021) Investor Sentiment and Stock Price Crash Risk: Evidence from China, *Global Economic Review*