|  |  |
| --- | --- |
| **附件一** | 國立臺灣科技大學  資訊管理研究所  研究方法期末報告 |
|  | |

|  |
| --- |
| **中文題目(標楷體，22級字)**  **English(Times New Roman, 22pt)** |
| 研究生：黃雅婄  學號：M11209202 |
| 指導教授：XXX博士  中華民國 年 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**。**

本研究旨在透過文字探勘（Text Mining）和自然語言處理（NLP）等技術，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投資者對於社群媒體內容所產生的情緒影響程度和投資人行為，進而調整模型中情緒字詞的權重、以提升分析結果的準確度，針對社群中正面、中立和負面情緒詞彙，製作出金融市場之社群情緒詞典。

透過社群情緒分數，分析台積電與台灣加權指數股價與社群情緒之關聯性，最後做出準確的股市預測，以幫助投資人即時獲取最新的股票資訊，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、內容分析法、文字探勘（Text Mining）、情緒分析（Sentiment Analysis）、VADER Sentiment

**目錄**

[**摘要** II](#_Toc166973357)

[**第一章 緒論** 1](#_Toc166973358)

[1.1 研究背景與動機 1](#_Toc166973359)

[現今的科技環境生成式AI幫助情緒分析與社群網路現在大家的使用現況 2](#_Toc166973360)

[社群用字的更新速度paper 2](#_Toc166973361)

[1.2 研究目的 2](#_Toc166973362)

[欲透過焦點小組結合情緒分析製作出金融情緒字典並判斷出，並結合生成式AI持續的更新情緒辭典 3](#_Toc166973363)

[1.3 研究架構 3](#_Toc166973364)

[第二章 文獻探討 4](#_Toc166973365)

[2.1 台灣股市之台積電與台股加權指數 4](#_Toc166973366)

[台灣研究台積電與台股加權指數的很少 4](#_Toc166973367)

[這中間的關聯性 4](#_Toc166973368)

[### 2.1.1 台灣研究台積電與台股加權指數的文獻回顧 4](#_Toc166973369)

[2.2 社群平台之情感分析 5](#_Toc166973370)

[社群情緒分析的相關論文發現與影響 5](#_Toc166973371)

[社群更新與網路流行術語的變化 5](#_Toc166973372)

[網路社群之影響與投資人的關聯 5](#_Toc166973373)

[應用於新聞中情緒分析的很多因為文字比較固定 5](#_Toc166973374)

[2.3 文字探勘與生成式AI的幫助 5](#_Toc166973375)

[技術的應用、英文與中文差距、中文斷詞的困境與新興技術 5](#_Toc166973376)

[第三章 研究方法 5](#_Toc166973377)

[3.1 實驗設計與架構 5](#_Toc166973378)

[專家法 焦點小組 建立平台或開啟 5](#_Toc166973379)

[3.2 研究假設限制 5](#_Toc166973380)

[3.3 內容分析法 5](#_Toc166973381)

[3.4 金融社群情緒字典與檢核 5](#_Toc166973382)

[3.5 情緒分析模型建置 5](#_Toc166973383)

[筆記 6](#_Toc166973384)

[參考文獻 7](#_Toc166973385)

**圖目錄**

**附件六**

[圖1.1. XXXX循環示意圖 1](#_Toc520412068)

[圖2.1. XXXXX結構示意圖 10](#_Toc520412072)

[圖2.2. XXXXX對準誤差的關係 11](#_Toc520412073)

**表目錄**

**附件七**

[表1.1. International XXX 2](#_Toc520412069)

表1.2. XXX比較 5

表1.3. XXX技術比較 6

表3.1. XXX材料參數 27

# **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 evelopment (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，使用者在在社群平台中討論的內容也包含多種主題，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。而社群平台中使用者的討論內容也作為一種資訊來源，加以分析後在諸多領域實際應用，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。

<https://www.sciencedirect.com/science/article/pii/S1057521916301569>

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對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23%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，其中「消息面分析」帶來的影響也在通訊網路普及的時代背景下產生劇烈的影響，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，而這些消息面所帶來的影響和投資人總體的分析結果最終將決策反映到股市上。

生城市AI的發展

## 現今的科技環境生成式AI幫助情緒分析與社群網路現在大家的使用現況 大型語言模型

好的，让我更具体地解释一下在技术层面上如何利用大型语言模型和生成式AI来进行社群网络中股市讨论的情绪分析：

1. \*\*数据预处理\*\*：

- 使用自然语言处理（NLP）技术对收集到的文本数据进行预处理。这包括分词、去除停用词、标点符号和特殊字符，以及进行词干化或词形还原等操作。

- 这些预处理步骤有助于减少噪音，使文本更具可分析性。

2. \*\*情感分析模型训练\*\*：

- 利用监督学习方法，建立情感分析模型。可以使用已标记的情感数据集进行训练，其中文本标记为正面、负面或中立情绪。

- 可以选择预训练的大型语言模型作为基础，并在其之上进行微调，以适应股市讨论的情感分析任务。

3. \*\*模型选择\*\*：

- 可以选择适合情感分析任务的模型架构，如卷积神经网络（CNN）、循环神经网络（RNN）、长短期记忆网络（LSTM）、或者是Transformer模型等。

- 对于大规模数据和复杂语义的情感分析任务，Transformer模型可能更适合，因为它在处理长文本和捕获文本语义方面表现出色。

4. \*\*模型评估\*\*：

- 在训练好的情感分析模型上进行评估和验证，使用验证集或交叉验证方法来评估模型的性能。

- 使用指标如准确率、精确率、召回率和F1分数来评估模型的表现。

5. \*\*生成式AI辅助分析\*\*：

- 利用生成式AI模型，如GPT，生成与股市讨论相关的文本。

- 可以使用生成式AI模型生成虚拟的股市评论或预测，以扩展情感分析模型的训练数据集，提高模型的泛化能力。

6. \*\*模型集成\*\*：

- 将训练好的情感分析模型与生成式AI模型进行集成，以获得更全面的情感分析结果。

- 可以结合两种模型的预测结果，或者使用生成式AI模型生成的文本作为情感分析模型的输入之一。

7. \*\*结果可视化与解释\*\*：

- 将情感分析的结果进行可视化，并进行解释和解读。

- 可以使用图表、词云等方式呈现情感分析的结果，并结合股市的实际走势进行分析和解释，以帮助用户更好地理解市场情绪和趋势。

通过以上技术层面的方法，可以有效利用大型语言模型和生成式AI来进行社群网络中股市讨论的情绪分析，并为投资者提供有价值的信息和见解。

社群網路為何重要?有甚麼討論股市的平台?有多少人用 找報告書說明

平台可以做甚麼 所以社群網路帶來的影響也很大 找一篇paper中有說明社群網路對金融理財股市的影響的

是的，对大型语言模型和情绪分析来说，将焦点小组的研究融入其中是非常有帮助的。

1. \*\*增强情感分析模型的准确性\*\*：

- 通过焦点小组的研究，可以更深入地了解投资者对于不同情绪字眼的理解和反应。

- 这些理解可以用于优化情感分析模型，使其更准确地识别和理解文本中的情感倾向，从而提高情感分析的准确性。

2. \*\*改善生成式AI模型的生成效果\*\*：

- 通过焦点小组的研究，可以收集到投资者对不同情绪字眼的反应和偏好。

- 这些反馈可以用于指导生成式AI模型的训练，使其生成的文本更符合投资者的期望和需求，从而提高生成效果。

3. \*\*丰富模型训练数据\*\*：

- 焦点小组的研究可以提供丰富的文本数据，包括投资者在讨论股市时使用的情绪字眼。

- 这些数据可以用于模型的训练和优化，帮助模型更好地理解和模拟投资者的语言使用习惯，提高模型的泛化能力和适应性。

4. \*\*指导平台功能和服务的改进\*\*：

- 焦点小组的研究结果可以为平台的功能和服务改进提供重要参考。

- 通过了解投资者对情绪字眼的反应和偏好，平台可以调整其使用方式，并设计更符合投资者需求的功能和服务，提升用户体验。

因此，将焦点小组的研究融入大型语言模型和情感分析中可以帮助优化模型的性能，提高模型的准确性和生成效果，并指导平台的改进和优化，从而更好地满足用户的需求和期望。

## 社群用字的更新速度paper

用paper證明社群更新很快

找生成式AI的應用發展與應用於情緒分析的部分

本研究認為，若將社群平台中關於股市的貼文、留言、論壇內容透過自然語言處理1 （NLP）之技術初步處理後，再對其進行語意分析（Semantic Analysis），得到之資料價值將有助於投資人綜觀全局、有效分析社群評論對股市的情緒狀態。

本研究以台積電為例，蒐集為期一年有關台積電股票之網路留言或貼文資料。不

同於財經節目或新聞媒體制式化的報導內容，社群網路的評論內容較亂無章法，因此本研究須先以Jieba（結巴）、CkipTagger斷詞等相關技術，對文字進行斷詞、斷句等初步處理，後將文字資料轉換成規則的結構化資料，再透過自然語言處理（NLP）之技術進行情緒分析（Sentiment Analysis），將資料分類成正面和負面等情緒標籤，並統整成相關的情緒字典，驗證股價漲跌走勢與社群評論之關係，探究社群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

## 欲透過焦點小組結合情緒分析製作出金融情緒字典並判斷出，並結合生成式AI持續的更新情緒辭典

判斷大盤與台積電中論壇等級、是否有影響 提出研究假設?

如何透過研究設計找出、社群情緒與股價漲跌、交易量、等等各項因素是否有影響?

|  |  |
| --- | --- |
|  |  |

## 1.3 研究架構

# 第二章 文獻探討

## 2.1 台灣股市之台積電與台股加權指數

## 台灣研究台積電與台股加權指數的很少

## 這中間的關聯性

## ### 2.1.1 台灣研究台積電與台股加權指數的文獻回顧

過去有關美國股市與台灣股市關聯性的研究顯示，道瓊工業指數對台灣股市的影響最為顯著，其次是NASDAQ指數或費城半導體指數。隨著產業結構的變遷，特別是台灣成為世界電子科技業的代工重鎮，電子業已成為台灣的主流產業。尤其是半導體業者，台積電在晶圓代工領域的主導地位，使其成為投資者矚目的焦點。

近年來，費城半導體指數的重要性日益增加，因其包含了19家重要的半導體公司，其中唯一的亞洲公司便是台積電。研究顯示，費城半導體指數對台積電股價的影響程度比台灣加權股價指數更強烈。在Granger因果關係檢定、共整合檢定和誤差修正模型中，費城半導體指數對台積電股價的影響均顯著。

在共整合檢定中，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。費城半導體指數與台積電股價、台灣加權指數與台積電股價之間也存在長期均衡關係，這意味著投資者可以利用某一數列資料來預測另一數列資料的變動。費城半導體指數與台灣加權指數之間沒有共整合現象，可能是由於半導體產業的摩爾定律現象，使得費城半導體指數波動更為顯著。

在誤差修正模型中，研究發現費城半導體指數與台積電股價均受到雙方前一期的影響，表明這兩者之間存在短期及長期的相互影響關係。然而，台灣加權指數與台積電股價之間的關係較為薄弱，僅台積電股價前四期對台灣加權指數有顯著影響，其餘前1至4期的交叉變數對當期變數的解釋均不顯著。

總結來說，過去的研究表明，費城半導體指數對台積電股價的影響力顯著超過台灣加權指數，強調了美國半導體市場對台積電及台灣整體電子產業的關鍵影響。這些研究結果為投資者提供了寶貴的參考意見，使他們能更好地理解美國股市與台灣股市之間的動態交互關係。

## 2.2 社群平台之情感分析

## 社群情緒分析的相關論文發現與影響

## 社群更新與網路流行術語的變化

## 網路社群之影響與投資人的關聯

## 應用於新聞中情緒分析的很多因為文字比較固定

## 2.3 文字探勘與生成式AI的幫助

## 技術的應用、英文與中文差距、中文斷詞的困境與新興技術

# 第三章 研究方法

## 3.1 實驗設計與架構

## 專家法 焦點小組 建立平台或開啟

## 3.2 研究假設限制

## 3.3 內容分析法

## 3.4 金融社群情緒字典與檢核

## 3.5 情緒分析模型建置

# 筆記

### 第二章 文獻探討

#### 2.1 台灣股市之台積電與台股加權指數

- \*\*2.1.1 台灣研究台積電與台股加權指數的文獻回顧\*\*

- 過去研究的概述

- 研究的不足之處

- \*\*2.1.2 台積電與台股加權指數之間的關聯性\*\*

- 影響因素分析

- 實證研究的發現

#### 2.2 社群平台之情感分析

- \*\*2.2.1 社群情緒分析的相關研究\*\*

- 情緒分析的基本概念

- 主要發現與影響

- \*\*2.2.2 社群更新與網路流行術語的變化\*\*

- 新興術語的出現及其影響

- \*\*2.2.3 網路社群之影響與投資人的關聯\*\*

- 社群媒體對投資決策的影響

- 投資人行為分析

- \*\*2.2.4 應用於新聞中的情緒分析\*\*

- 新聞情緒分析的應用

- 固定文本情緒分析的挑戰和解決方案

#### 2.3 文字探勘與生成式AI的應用

- \*\*2.3.1 技術應用概述\*\*

- 文字探勘技術的發展

- 生成式AI在情緒分析中的應用

- \*\*2.3.2 英文與中文文本處理的差異\*\*

- 英文文本處理技術

- 中文文本處理的特殊挑戰

- \*\*2.3.3 中文斷詞的困境與新興技術\*\*

- 中文斷詞技術的發展

- 相關新興技術的介紹

### 第三章 研究方法

#### 3.1 實驗設計與架構

- \*\*3.1.1 專家法\*\*

- 專家訪談的設計與實施

- \*\*3.1.2 焦點小組\*\*

- 焦點小組的組織與運作

- \*\*3.1.3 建立平台或開啟\*\*

- 平台建設的步驟與方法

#### 3.2 研究假設與限制

- \*\*3.2.1 研究假設\*\*

- 主要研究假設的提出

- \*\*3.2.2 研究限制\*\*

- 研究範圍與局限性

#### 3.3 內容分析法

- \*\*3.3.1 內容分析法的介紹\*\*

- 方法論的基礎

- \*\*3.3.2 應用於本研究的內容分析法\*\*

- 具體應用步驟

#### 3.4 金融社群情緒字典與檢核

- \*\*3.4.1 情緒字典的構建\*\*

- 字典構建的過程

- \*\*3.4.2 情緒字典的檢核\*\*

- 字典驗證的方法

#### 3.5 情緒分析模型建置

- \*\*3.5.1 模型建置的步驟\*\*

- 模型選擇與建置

- \*\*3.5.2 模型驗證與調整\*\*

- 模型的測試與優化

這樣的細化和結構化有助於讓讀者更清晰地了解每個部分的具體內容和研究重點，並更容易追蹤您的研究進程和結果。

# 參考文獻

[1] G. O. Young, “Synthetic structure of industrial plastics,” in *Plastics,* 2nd ed., vol. 3, J.Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.