|  |  |
| --- | --- |
| **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](#_Toc166877779)

[**第一章 緒論** 1](#_Toc166877780)

[1.1 研究背景與動機 1](#_Toc166877781)

[1.2 研究目的 1](#_Toc166877782)

[1.3 研究架構 2](#_Toc166877783)

[第二章 文獻探討 4](#_Toc166877784)

[2.1 台灣股市之台積電與台股加權指數 4](#_Toc166877785)

[台灣研究台積電與台股加權指數的很少 4](#_Toc166877786)

[這中間的關聯性 4](#_Toc166877787)

[2.2 社群平台之情感分析 4](#_Toc166877788)

[社群情緒分析的相關論文發現與影響 4](#_Toc166877789)

[社群更新與網路流行術語的變化 4](#_Toc166877790)

[網路社群之影響與投資人的關聯 4](#_Toc166877791)

[應用於新聞中情緒分析的很多因為文字比較固定 4](#_Toc166877792)

[2.3 文字探勘與生成式AI的幫助 4](#_Toc166877793)

[技術的應用、英文與中文差距、中文斷詞的困境與新興技術 4](#_Toc166877794)

[第三章 研究方法 4](#_Toc166877795)

[3.1 實驗設計與架構 4](#_Toc166877796)

[專家法 焦點小組 建立平台或開啟 4](#_Toc166877797)

[3.2 研究假設限制 4](#_Toc166877798)

[3.3 內容分析法 5](#_Toc166877799)

[3.4 金融社群情緒字典與檢核 5](#_Toc166877800)

[3.5 情緒分析模型建置 5](#_Toc166877801)

[參考文獻 5](#_Toc166877802)

**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

近年來台灣股市交易活動逐年熱絡，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2016年的新台幣16兆元逐年上升至2020年的新台

幣45兆元，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，台積電隸屬晶圓代工

產業，2021年台積電代工生産份額更是佔全球的56%，是全球半導體產業中重要的生

產公司，其產品隨著技術多元化而精進。台積電影響著各行業的經濟命脈，可謂股市

中不容忽視的一股力量。[1]

隨著科技日新月異，網路的普及讓人們隨時可以傳遞消息，不同類型的社群媒體

也相繼推出。在TWNIC台灣網路報告中提及，累積至2020年，台灣12至24歲的Z

世代網路使用率達到100%、25至55歲的X、Y世代的網路使用率也高達95.3%，且

網路服務使用項目中，12至39歲的用戶使用社群論壇的比例高達95.6%，40至55歲

之使用比率也達到79.5%，可見各年齡層對網路的依賴程度與日俱增。現代人的日常

交流逐漸社群化，使用者在社群媒體上的活動產生了大量資料流量，而這些大數據資

料目前也應用在經濟、行銷、政治、人文等領域。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。[2]

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，已成為現代人日常生活密不可分的一部份。常見的社群媒體例如

Facebook、Podcast、Instagram、Twitter......，都是近年來火紅的社群平台。其中也有討

論股市的社群平台如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」卻常被視為輔助角色而忽視了其對整體經濟和個股評價的價值，鑒於多種調

查及相關論文實證現代人交流趨向社群化，不只有社群活動的大數據資訊產生的附加

價值，人們也開始關注社群對現實生活所帶來的影響。[3][4]

社群上股票評論已成為投資決策中重要的影響因素，投資人在平台上瀏覽產業資

訊和社群輿論，結合自身的金融知識並進行股票交易，最後將決策反映到股市上。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。

為了持XXXXX的製程節點，預計2021年後更將進入5nm的製程節點[2]，如何在持續縮小電晶體尺寸的同時保持製程品質，是一個相當嚴峻的挑戰。

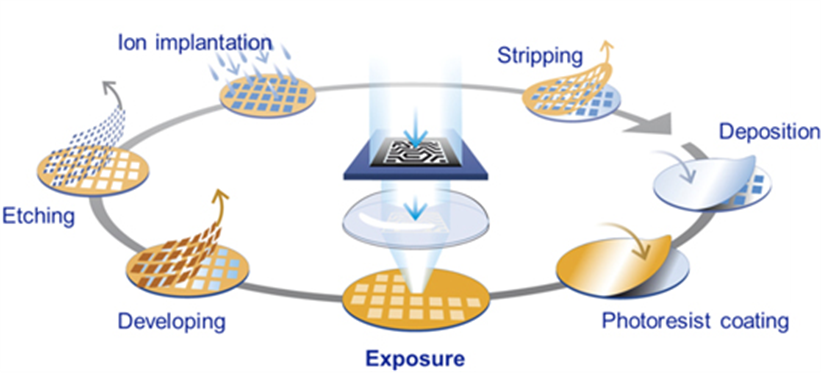


圖 1.1. 積體電路製程循環示意圖

積體電XXXXX，如圖1.1. 所示，一次基本的製程循環包含光阻塗佈(Photoresist Coating)、微影曝光(Exposure)、顯影(Developing)以及蝕刻(Etching)等步驟[3]。XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX。

## 1.2 研究目的

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX．

|  |  |
| --- | --- |
|  | (1.1) |
|  |  |

## 1.3 研究架構

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX，持續成長中。

|  |  |  |  |
| --- | --- | --- | --- |
| 表 1.1. 本文中之目標產業線寬與產值 | | | |
| 目標產業 | PCB | 封裝 | TFT-LCD |
| 線寬 | 25~30 [um] (5G) | 1~2 [um] (先進封裝) | 2 [um] |
| 全球產值 | 588億(USD) | 3456億(台幣) | 1兆(台幣) |

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX，如表1.1.。

# 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.