

逢 甲 大 學
財 務 金 融 學 系
碩 士 論 文

金融重大訊息情緒分析之研究

A Study of Sentiment Analysis on Financial
Material Information

指導教授：陳麗君 博士
研 究 生：陳芊妤

中 華 民 國 一 百 一 十 一 年 八 月

逢 甲 大 學
財務金融學系碩士學位論文

金融重大訊息情緒分析之研究

A Study of Sentiment Analysis on Financial
Material Information

☒學術論文

☐作品(藝術類)

☐書面報告(藝術類)

☐成就證明(應用科技類)

☐技術報告(應用科技類)

☐專業實務報告(專業實務類)

研究生：陳芊妤

經碩士學位考試合格特此證明

評審委員

陳麗君

陳幸佑

陳宗圀

指導教授

陳麗君

系主任

劉昭傳

考試日期：中華民國 111 年 7 月 14 日

逢甲大學研究生無違反學術倫理聲明書 FCU Graduate Student Academic Ethics Statement

本人已完全瞭解學術倫理之定義與行為規範，謹此嚴正聲明，本人所呈繳之學位論文(包含作品、成就證明連同書面報告、技術報告或專業實務報告)：

(題目) 金融重大訊息情緒分析之研究，

如有抄襲、舞弊或違反著作權法等違反學術誠信與倫理之行為時，願自行承擔所有法律責任，以及概括承受一切後果，並無條件同意註銷本人之碩(博)士學位，絕無異議。

I am fully aware of and understand the University's regulations on plagiarism. And I declare herewith, that the thesis/ written report/ technical report/ professional practice report entitled "(Thesis title) A study of Sentiment Analysis on Financial Material" Information is a presentation of my original work. In addition, I understand that any false claim or plagiarism in respect of this work will result in disciplinary action in accordance with University's regulations. I assume legal liability for this and completely agree to the withdrawal of the Master's/ Doctor's degree if any violation of academic ethics in the thesis is confirmed to be true.

聲明人 Student Signature: 陳萃婷
(請親筆書寫簽名 Signature)

學號/ID.NO: M0901995

系所/Department: 財務金融學系碩士班

日期 Date: 111 / 7 / 14 (YYYY/MM/DD)

本人為 陳萃婷 之指導教授，經檢視其學位論文內容，確實無抄襲或剽竊之行為。
To the best of my knowledge, this thesis has no plagiarism or violations of academic ethics.

指導教授(Academic Advisor): 陳麗君 (親筆簽名 Signature)

聲明日期(Date): 111 年(Y) 7 月(M) 14 日(D)

*本聲明書正本請與學位考試申請文件一併附上，影本請裝訂於紙本學位論文內。

The original statement should be included in your thesis defense application; the copy should be included in your thesis.

誌 謝

本研究論文可以順利完成，非常感謝我的指導教授陳麗君老師的悉心指導，從最初論文的選題到論文最終完成，都在我出現難題時幫助我許多。在老師嚴謹地督促與細心地指導之下，我的論文才得以完成。

還要特別感謝李宗儒老師，每週不辭辛勞地撥出時間一起開會討論，將我當作他的指導學生一般給予極大的支持與幫助。在每週的開會討論中慢慢地讓論文越來越完整，雖然中間論文進行過程中面臨到瓶頸，老師們還是很有耐心地教導我論文相關專業的知識，讓我受益良多。也特別謝謝冠瑋同學給予程式專業方面的幫助，更耐心地教導我步驟的操作，除了要完成自己的論文之外，還要抽出空閒的時間協助我。

再者，在口試期間承蒙口試委員 陽明交通大學陳宗岡老師與逢甲大學陳宗佑老師細心指正本論文，讓本論文更加完整。

最後要謝謝我的家人朋友，在我完成論文的過程中，體諒我承受很大的壓力，對我有很大包容同時也鼓勵著我，給我很大的信心與支持。本論文能夠順利完成，要再次感謝所有幫助過我的人。

陳芊妤 謹誌於

逢甲大學金融學院

中華民國一百一十一年八月

摘要

本研究採用 2010 年 1 月 1 日至 2021 年 12 月 31 日所有金融股之重大訊息進行情緒分析，試圖使用字典法與 BERT 方法提取重大訊息之情緒，並檢視其預測結果是否與實際市場反應相符。透過三個詞典關鍵詞使用程度及文本中出現情緒詞彙個數與次數，檢驗我們選用之財經領域情緒詞典對於重大訊息文本之可用性。從精確度、準確度、召回率及 F1 值四個指標評估詞典在重大訊息有用性之成效。

研究發現，字典法與 BERT 方法針對情緒較中性之重大訊息文本成效並不好。然而，針對傳統上預測情緒的定義方式做改善，結果顯示有用性指標皆有明顯的提升。從詞典前 30 名出現頻率詞彙觀察到詞典裡的好壞關鍵詞會影響到重大訊息情緒的判讀。而由詞典中出現頻率為零的詞彙，了解到重大訊息與財經新聞文本上有很大差異。

關鍵詞：情緒分析、重大訊息、中文金融情緒詞典、BERT、金融股

Abstract

This study uses sentiment analysis on material information of financial stocks from January 1, 2010 to December 31, 2021. To test whether the predicted results are consistent with market reaction, we extract the sentiment from material information by Dictionary-based approach and BERT method. Through the use of three dictionary keywords and frequency of sentiment words in the text, we examine the usability of the financial sentiment dictionary for material information texts. The effectiveness of the dictionary is evaluated from the precision, accuracy, recall and F1-score.

The results show that the Dictionary-based approach and BERT method did not work well for neutral material information texts. However, we have improved the traditional definition of predicting sentiment, and found that the usefulness of the results has been significantly improved. From the top 30 frequency words in the dictionary, it is observed that the sentiment keywords will affect the interpretation of material information's sentiment. From the words with zero frequency in the dictionary, the material information is very different from the financial news text.

Keywords: Sentiment Analysis 、 Material Information 、 Chinese Financial Sentiment Dictionary 、 BERT 、 Financial Stock

目 錄

第一章 緒論	1
第一節 研究動機及背景	1
第二節 研究目的	3
第三節 研究架構	4
第二章 文獻探討	5
第一節 文本分析資料來源	5
第二節 情緒分析方法	8
第三章 研究設計與實施	13
第一節 研究流程	13
第二節 資料蒐集	14
第三節 前處理	15
第四節 字典法預測情緒	17
第五節 情緒分析方法成效指標	18
第四章 實證結果與分析	22
第一節 樣本基本描述	22
第二節 字典法成效分析	24
第三節 詞典詞彙出現頻率	31
第四節 BERT 方法成效	39
第五節 重大訊息案例探討	40
第五章 結論與建議	46
第一節 研究結論	46
第二節 未來展望	47
參考文獻	48

圖目錄

圖 3-1 研究流程.....	13
-----------------	----



表目錄

表2-1 文本來源文獻整理	7
表2-2 林宜萱 (2013) 財經領域情緒詞典 (FESD).....	9
表2-3 張津挺 (2015) 中文財務情緒詞典 (CFSL).....	10
表2-4 姜富偉、孟令超、唐國豪 (2021) 中文金融情感詞典 (CFSD)	11
表2-5 詞典整理	11
表3-1 公司清單	14
表3-2 門檻值定義	15
表3-3 好壞消息定義	15
表3-4 傳統上預測情緒定義	16
表3-5 本研究預測情緒定義	16
表3-6 指標定義	19
表4-1 樣本敘述統計	22
表4-2 重大訊息筆數占比	23
表4-3 關鍵詞使用比例	24
表4-4 關鍵詞出現個數	25
表4-5 關鍵詞出現個數小於 6 比例	26
表4-6 關鍵詞累計出現次數	26
表4-7 詞典指標有用性 傳統上定義	28
表4-8 詞典指標有用性 本研究定義	28
表4-9 正確預測各消息占比	29
表4-10 FESD 詞彙頻率前 30 名	31
表4-11 CFSL 詞彙頻率前 30 名	32
表4-12 CFSD 詞彙頻率前 30 名	33
表4-13 出現頻率零 FESD 好關鍵詞	34
表4-14 出現頻率零 FESD 壞關鍵詞	35
表4-15 出現頻率零 CFSL 好關鍵詞	36
表4-16 出現頻率零 CFSL 壞關鍵詞	36
表4-17 出現頻率零 CFSD 好關鍵詞	37

表4-18 出現頻率零 CFSD 壞關鍵詞.....	38
表4-19 BERT 資料集重大訊息筆數	39
表4-20 BERT 準確率.....	39
表5-1 重大訊息範例 1 元大金	40
表5-2 重大訊息範例 1 關鍵字表	42
表5-3 重大訊息範例 2 國泰金	42
表5-4 重大訊息範例 2 關鍵字表	45



第一章 緒論

第一節 研究動機及背景

台灣為淺碟型市場，易受消息面的影響，投資人從各種平台取得資訊，進而對股市進行趨勢預測，如何正確判讀消息的情緒是投資人重要的課題。

在過去網路與科技較不發達的時期，資訊的取得可從電視、報紙及廣播等傳統方式。而隨著時代的發展，網路與媒體的發展愈來愈進步。投資人透過新聞媒體、社交媒體與公司公開揭露文件等方式取得股市資訊。不論是打開電視聽著財經新聞、走進超商購買財經版面報紙，還是加入社群媒體的投資社團，亦或是觀看影音平台投資節目，無時無刻不在接收資訊。然而，新聞媒體與社交媒體的消息來源可能存在著假消息進而造成誤判的問題，反觀公司公開揭露文件為公司內部所發布，可信度較高，故本文選擇公開資訊觀測站重大訊息為研究對象。

隨著獲得資訊途徑與類型愈來愈多，也代表我們要開始面臨如何將這些豐富的資訊轉化成更能有效運用的知識，情緒分析與文字探勘成為相當常見的論文題材，應用面相當廣泛。情緒分析方法早期文獻可能會選擇人工判讀的方式判讀情緒，缺點是耗時費力且短時間無法處理大量文本。字典法的出現讓許多文本能夠藉由計算正負向詞彙數量，判讀文本的情緒，也有文獻指出通用詞典無法使用在財經領域之文本，因此出現針對財經領域所使用的情緒詞典。早期英文的情緒詞典較為發達，為了因應不同語言的語法差異，開始有文獻進一步建置中文情緒詞典，來針對中文的文本作情緒分析。

隨著科技的進步，機器學習的出現幫助處理大量的文本資料，以機器解決人工判讀所產生的缺點，利用自然語言處理技術從文本中理解和提取重要信息。在眾多情緒分析方法中，哪一個方法能夠有效判讀公開觀測站之重大訊息情緒？我們想深入探討情緒詞彙較少的重大訊息文本，使用字典法是否能夠有效判讀其情緒。

本研究針對金融股之重大訊息，深入探討字典法與機器學習方法是否能夠正確判讀重大訊息文本當中的情緒，並比較不同的財經領域詞典，評估其成效。



第二節 研究目的

過去對於情緒文本分析相關文獻，多著重於社群媒體及財經新聞文章，鮮少針對公司對外宣告較正式之內容形式作分析。使用字典法作為情緒分析方法的文獻中，也較多都是針對財經新聞作分析。公開資訊觀測站之重大訊息相比新聞或社群媒體所發布的財經資訊較為正式，對於發布的時間與格式有制式的規範，較無情緒性的詞彙，能否透過字典法與機器學習方法提取重大訊息文本中的情緒。

本文以台灣股票市場金融股為研究對象，以公開資訊觀測站之金融重大訊息為資料來源，蒐集公開觀測站之重大訊息發布，測試多種情緒分析方法判讀重大訊息情緒，期望對財經領域的情緒分析做出貢獻。研究目的如下：

- 一、 探討字典法與 BERT 是否能夠正確判讀重大訊息之情緒。
- 二、 透過效果指標分析字典法之情緒分析效果。
- 三、 探討財經情緒詞典對於重大訊息之可用性及有用性。

第三節 研究架構

本文主要分為五大章節：

第一章 緒論：說明本文研究背景與動機、研究目的及研究架構。

第二章 文獻探討：回顧相關領域之文獻。

第三章 研究設計與實施：說明本研究使用的方法。

第四章 實證結果與分析：陳述實證之結果，呈現研究方法之成效。

第五章 結論及未來展望：總結本文研究結果並建議後續研究之方向。



第二章 文獻探討

第一節 文本分析資料來源

接收資訊的途徑逐漸多樣化，而資訊對投資人的影響，可以透過文本分析做進一步的探討。金融文本情感分析的資訊來源主要以下列三個為主：媒體文章、公司公開揭露文件及社交媒體。

一、以社群媒體為資料來源之文本分析

在社群媒體發文成為現代人的日常，不過社群媒體平台上任何人都可以分享自己的意見，涉及主觀性且資訊較吵雜。許多文獻針對社群媒體文本當中的情緒作分析 (Gu & Kurov, 2020 ; Kouloumpis et al., 2011 ; Ranco et al., 2015 ; Giachanou & Crestani, 2016 ; Siganos et al., 2014)。

Das & Chen (2007) 闡述網路情緒包括潛在的有價值的見解、市場情緒、操縱行為以及對其他新聞來源的反應，可能會對金融市場造成重大影響。這些信息中可能包含很大一部分的噪音或非理性情緒。

二、以媒體表達之內容為資料來源之文本分析

過去許多文獻針對財經新聞之文本，分析其在金融市場的角色或對股票市場的影響 (Tetlock, 2007 ; Schumaker et al., 2012 ; Heston & Sinha, 2017 ; Kalyani et al., 2016 ; Mohan et al., 2019 ; 謝委霖，2015 ; 王聿庭，2019)。

媒體新聞之文章主要是二手甚至三手的消息，主要反映的是後見之明，也必須考慮當中是否有假新聞的問題，加上可能有些公司會利用新聞報導帶風向，可信度較低。

三、以公司公開揭露文件為資料來源之文本分析

公開資訊觀測站之重大訊息為投資人取得訊息管道之一，是來自內部人員的官方發布，與外部人員相比，他們更了解公司內部運作。這些文本中的用字遣詞和語氣可能會傳達公司未來預期業績的相關有用信息，而不只是財務報表中的數字。

鍾雨潼 (2003) 利用事件研究法以平均異常報酬與累積異常報酬來分析公開資訊觀測站。蔡侑達 (2012) 將公開資訊觀測站之重大訊息按照訊息種類手動分成 21 個種類，探討不同種類之重大訊息於不同時間段對股價波動度與股票報酬率之影響效果，研究發現當中盈餘宣告及高階主管異動這兩個類別在任何時段皆呈正向顯著，表示報酬率顯著增加。

黃娟娟 (2012) 顯示公司所發布之年報，其內容資訊公開完整，但使用頻率不高，探討年報在使用頻率不高的情況下，是否對投資人決策存在相關性，以及在問題公司當中，能否從其年報當中發現公司可能發生問題之跡象，研究結果發現公司年報敘述性揭露對問題公司發生企業失敗具有解釋能力。

陳予得 (2021) 透過人工判讀的方式針對致股東報告書前十行，探討其語調對未來公司績效之影響。我們與其不同的地方是重大訊息之資料量比較大，可能無法用人工判讀的方式做情緒分析，其主要探討致股東報告書之內容是否會影響到隔年的盈餘。

新聞媒體及社交媒體之文本情緒需要注意假消息與可信度的問題，而公司公開文件為公司公開之準確的消息，可信度較高，加上社交媒體的情緒較吵雜，因此我們想針對重大訊息做分析。

表2-1 文本來源文獻整理

文本來源	文獻	特點
社群媒體	Gu & Kurov (2020)、Kouloumpis et al. (2011)、Ranco et al. (2015)、Giachanou & Crestani (2016)、Siganos et al. (2014)、Das & Chen (2007)	資訊吵雜 信息包含噪音或 非理性情緒 可信度低
新聞媒體	Tetlock (2007)、Schumaker et al. (2012)、Heston & Sinha (2017)、Kalyani et al. (2016)、Mohan et al. (2019)、謝委霖 (2015)、王聿庭 (2019)	二手消息 假消息 可信度低
公司公開揭露文件	黃娟娟 (2012)、陳予得 (2021)、鍾雨潼 (2003)、蔡侑達 (2012)	由公司內部發布 可信度高



第二節 情緒分析方法

文本情感文獻中最常見的情緒分析方法是基於字典的方法 (Dictionary-Based Approach) 和機器學習 (Machine Learning)。早期文獻在字典法與機器學習還未發達之前，也會使用人工判讀的方式以判斷文章之情緒。

一、人工判讀相關文獻

Bryan (1997) 透過人工閱讀 250 篇 MD&A。Cole & Jones (2004) 閱讀 568 篇 MD&A。陳予得 (2021) 透過人工閱讀致股東報告書前十行，探討其語調對未來公司績效之影響。人工判讀的方式較耗時費力，較不適合處理大量資料。

二、字典法相關文獻

1. 英文情緒詞典建置

Henry (2008) 衡量收益新聞稿的語調。哈佛 General Inquirer (Harvard's GI) 被認為是最早的情感詞庫，Tetlock (2007) 使用 GI 來獲取文本中的情緒。GI 和 Diction 詞表已被研究人員廣泛用於量化金融和會計中文本信息的情緒。然而，它們已被證明不適合測量商業文本的情緒，因為它們並不是專門為商業用途而設計的，因此將金融文本中的常用詞分類錯誤 (Li, 2010; Loughran & McDonald, 2011)。Loughran & McDonald (2011) 發現在哈佛 GI 列表中幾乎四分之三 (73.8%) 的負面字字數在金融環境中通常不是負面的單詞。Loughran & McDonald (2011) 為財務和會計中最常用的詞表。Loughran & McDonald (2011) 表示衡量商業通信中的情緒的研究人員不應使用源自商業使用領域之外的分類方案。相反地，應使用專為商務交流設計的詞表來衡量商務文本的情緒 (Loughran & McDonald, 2016)。

2. 中文情緒詞典建置

文本分析越來越蓬勃發展，但大部分的研究依賴的以英文的文本居多，因中文與英文語法與思維模式的差異，特點也完全不同，常用的中文情感分類詞典有 3 種：HOWNET、DLUTSD 和 NTUSD。與哈佛 GI 和 Diction 類似，這三本詞典都是通用的，而不是商業用途，無法有效地分析財經文本。

表2-2 為林宜萱 (2013) 建置情緒詞典之來源整理。以 Loughran & McDonald 詞典為基礎，僅使用正、負極性詞彙，透過遠東線上字典將 Loughran & McDonald 詞典以半自動化的方式翻譯成中文，並以電腦比對中文新聞進行篩選，再進一步進行人工擴編與刪減，建置出財經領域情緒詞典。

編製成功之後，選取兩百篇新聞衡量之情緒分數，並與股票表現進行迴歸分析，以驗證其有效性，驗證結果發現確實能有效衡量新聞之情緒。

表2-2 林宜萱 (2013) 財經領域情緒詞典 (FESD)

來源		詞彙數量
財經詞典	LM 詞典中文翻譯	正 369
金融英漢詞典	遠東線上字典	負 522

表2-3 為張津挺 (2015) 建置詞典之來源整理。主要以 Loughran & McDonald 詞典為基底，透過四部英漢金融專業詞典及谷歌線上翻譯，建立一部中文財務情緒詞典，將財務情緒詞彙分為六大類：正面 (positive)、負面 (negative)、不明確性 (uncertainty)、爭議性 (litigious)、強語氣 (modal words strong) 和弱語氣 (modal words weak)。編製成功

之後，選取一萬九千篇雅虎新聞，擷取近 30 天之財務新聞進行分析比對，以確認中文情緒詞典之詞彙是否出現於中文財務新聞中，驗證其有效性，並與林宜萱 (2013) 所建置之財經領域情緒詞典進行比較。

表2-3 張津挺 (2015) 中文財務情緒詞典 (CFSL)

來源		詞彙數量
財經詞典	LM 詞典中文翻譯	
金融英漢詞典	高盛英漢字典	
	摩根史丹利英漢字典	正 207
	HK Rich-power	負 959
	MBAlib	
翻譯	谷歌翻譯	

表2-4 為姜富偉等 (2021) 建置詞典來源整理。使用三個方式建置詞典：(1) Loughran & MacDonald (2011) 詞典為基底，轉化為對應的中文版本。(2) 從中文通用情感詞典 (清華詞典、知網詞典和 NTUSD 詞典) 中篩選出在金融語境下仍然適用的情感詞彙。(3) 為了避免金融情感詞語的遺漏，以 word2vec 算法擴充詞典。將三個方式合併並去除重複詞，最後建置中文金融情感詞典。

表2-4 姜富偉、孟令超、唐國豪 (2021) 中文金融情感詞典 (CFSD)

負面詞彙 (5890)		
來源	詞彙數量	
財經詞典	LM 詞典中文翻譯	1562
通用詞典	清華大學詞典	1945
	知網詞典	534
	NTUSD 詞典	1243
Word2vec 詞典擴充		606
正面詞彙 (3338)		
來源	詞彙數量	
財經詞典	LM 詞典中文翻譯	458
通用詞典	清華大學詞典	1928
	知網詞典	304
	NTUSD 詞典	255
Word2vec 詞典擴充		393

以上三個詞典之比較，整理於表2-5。

表2-5 詞典整理

詞典	基礎	添加詞典	其他	有效性	詞彙數量
林宜萱 (2013) FESD	L&M 詞典 (正負類)	X	遠東線上字典	200 篇新聞	G:369 B:522
張津挺 (2015) CFSL	L&M 詞典 (六大類)	4 部金融英漢詞 典	Google translate	19000 篇 yahoo 新聞	G:207 B:959
姜富偉 (2021) CFSD	L&M 詞典	3 部通用詞典	Word2vec 詞典擴 充		G:3338 B:5890

3. 使用字典法判讀財經文本

吳昱萱 (2018) 使用財經情緒詞典衡量公司財務預警程度，研究結果顯示公司當日、昨天和前天財務預警相關財經新聞情緒和新聞篇數與股價波動顯著相關。

三、機器學習相關文獻

隨著時代的發展，機器學習愈來愈進步。BERT 模型 2018 年由 Google 提出，是用於自然語言處理 (NLP) 的預訓練技術。夏鶴芸 (2020) 使用 BERT 模型進行新聞分類，發現在新聞情緒分類方面，BERT 模型搭配利用股價漲跌標籤的新聞資料集，比起利用情緒詞典正負面標註新聞的資料集，有不錯的成效。陳予得 (2021) 發現 BERT 模型分類準確率高達 86%，透過語境測試發現 BERT 有學習到語言結構，並比較字典法、TF-IDF 與 BERT 對致股東報告書情緒分析之準確率，發現 BERT 能夠理解語意結構，判斷否定詞與詞彙的關係，更能準確分辨語句的情緒。

相較於其它文本，重大訊息文本的情緒詞彙比較少，使用在重大訊息文本的話，效果可能沒那麼好，不過我們認為還是可以與機器學習的方法做比較，可以幫助了解兩個方法的差異與優缺點。

第三章 研究設計與實施

第一節 研究流程

本研究主要探討使用不同情緒分析方法對判讀重大訊息情緒的成效分析。研究流程如圖 3-1。



圖 3-1 研究流程

第二節 資料蒐集

本研究選取金融股產業之公開觀測站重大訊息，以爬蟲的方式擷取所有樣本。樣本期間為 2011 年 1 月 1 日至 2021 年 12 月 31 日。公司家數總共為 31 家，公司清單如表 3-1。

表 3-1 公司清單

公司名	代碼	公司名	代碼	公司名	代碼	公司名	代碼
彰銀	2801	聯邦銀	2838	富邦金	2881	國票金	2889
京城銀	2809	遠東銀	2845	國泰金	2882	永豐金	2890
台中銀	2812	安泰銀	2849	開發金	2883	中信金	2891
旺旺保	2816	新產	2850	玉山金	2884	第一金	2892
華票	2820	中再保	2851	元大金	2885	王道銀行	2897
台產	2832	第一保	2852	兆豐金	2886	上海商銀	5876
臺企銀	2834	三商壽	2867	台新金	2887	合庫金	5880
高雄銀	2836	華南金	2880	新光金	2888		

第三節 前處理

一、詞典前處理

將林宜萱財經領域正負向情緒詞典 (FESD) 之關鍵詞於 excel 中備置。張津挺中文財務情緒詞典 (CFSL) 只擷取正負類詞彙，於 excel 中備置。姜富偉中文金融情感詞典 (CFSD) 為簡體字，將其使用 excel 翻譯成繁體字，以符合重大訊息文本使用。

二、好壞消息定義

表3-2 為門檻值定義整理。本研究透過隔日股票漲跌幅來定義好壞消息，考慮到金融股對於漲跌停的情況不是很常發生，從 31 家公司報酬分配 1% 與 99% 的平均可以看到數值是落在 3% 左右，所以我們定 3% 為我們的門檻值。

表3-2 門檻值定義

常態分配	1%	25%	50%	75%	99%
平均	-2.9293	-0.4304	0.0000	0.5217	3.1179

因此將隔日漲幅 $> 3\%$ 定義為好消息 (G)，隔日跌幅 $< -3\%$ 定義為壞消息 (B)，其餘為無情緒消息 (N)。好壞消息定義如表3-3 所示。

表3-3 好壞消息定義

好消息 (G)	無情緒消息 (N)	壞消息 (B)
隔日報酬率 $> 3\%$	$-3\% < \text{報酬率} < 3\%$	隔日報酬率 $< -3\%$

三、預測情緒定義

我們計算重大訊息出現好壞詞之次數，來預測重大訊息之情緒，將預測消息以三分類分為正面 (G)、負面 (B) 及無情緒消息 (N)。定義以下兩種預測情緒標準，以測試兩種定義預測情緒標準之成效。

1. 傳統上定義

如果累計好詞次數大於累計壞詞次數，則預測為正面消息 (G)；如果累計好詞次數小於累計壞詞次數，則重大訊息預測為負面消息 (B)；如果累計好詞次數等於累計壞詞次數，則重大消息預測為無情緒消息 (N)。傳統上預測情緒定義如表3-4 所示。

預測情緒 = IF (累計 G > 累計 B, "G", IF(累計 G < 累計 B, "B", "N"))

表3-4 傳統上預測情緒定義

好消息 (G)	無情緒消息 (N)	壞消息 (B)
累計 G > 累計 B	累計 G = 累計 B	累計 G < 累計 B

2. 本研究定義

考量到重大訊息有大量的無情緒消息，本研究將定義調整為：如果累計好詞次數減累計壞詞次數大於 5，則預測為正面 (G)；如果累計好詞次數減累計壞詞次數小於 -5，則重大訊息預測為負面 (B)；如果都不符合，則重大消息預測為無情緒消息 (N)。本研究預測情緒定義如表3-5 所示。

預測情緒 = IF (累計 G - 累計 B > 5, "G", IF (累計 G - 累計 B < -5, "B", "N"))

表3-5 本研究預測情緒定義

好消息 (G)	無情緒消息 (N)	壞消息 (B)
累計 G - 累計 B > 5	-5 < 累計 G - 累計 B < 5	累計 G - 累計 B < -5

第四節 字典法預測情緒

在進行情緒分析作業之前，首先我們需要能夠測量文本語氣與情感的詞典，從中識別正面或負面的單詞。在以往的研究文獻當中，最常使用的四個英文詞典是 Henry (2008)、哈佛 General Inquirer (Harvard's GI)、Diction 詞表和 Loughran & McDonald (L&M)。

我們選擇的樣本為公開觀測站重大訊息，在語言差異之下選擇使用中文財經領域情緒的詞典，作為判斷情緒的工具，以研究情緒詞典是否能夠成功判讀重大訊息的情緒。使用之情緒詞典如下：

- (1) 林宜萱 (2013) 財經領域正負向情緒詞典 (FESD)。
- (2) 張津挺 (2015) 中文財務情緒詞典 (CFSL)。
- (3) 姜富偉等 (2021) 中文金融情感詞典 (CFSD)。

若重大訊息實際之情緒與字典法預測之情緒相符，代表字典法成功判讀出重大訊息之情緒，正確判斷記 1，不正確判斷記 0。

$$\text{正確情緒判斷} = \text{IF (預測情緒} = \text{實際情緒, 1, 0)}$$

第五節 情緒分析方法成效指標

一、有用性指標

本研究測試多種情緒分析的方法，以了解方法的成效。透過下列四種指標對情緒分析方法的成效做評價。

1. 精確度 (Precision)：預測為正/負/無情緒消息中，實際為正/負/無情緒消息之比率

$$\text{Precision} = \frac{\text{實際正/負/無情緒}}{\text{預測正/負/無情緒}}$$

即針對我們預測結果當中，預測為正/負/無情緒的樣本中有多少是正/負/無情緒的樣本。

2. 準確度 (Accuracy)：預測所有消息中，判斷正確之比率

$$\text{Accuracy} = \frac{\text{正確預測}}{\text{所有預測}}$$

即所有樣本當中，正確預測之樣本數有多少，準確度越高，效果越好。

3. 召回率 (Recall)：實際為正/負/無情緒消息，能夠預測出為正/負/無情緒消息的比率

$$\text{Recall} = \frac{\text{預測出正/負/無情緒}}{\text{實際為正/負/無情緒}}$$

即代表實際為正/負/無情緒的樣本，有多少樣本被預測正確。

4. F 值 (F1-score)：權衡綜合 Precision 與 Recall 兩個指標的結果

$$F1 - score = \frac{2 * Recall * Precision}{Recall + Precision}$$

即召回率和精確度之加權調和平均值，F 值愈高，說明方法愈有效。

表3-6 指標定義

精確度 Precision	預測為正面消息中，實際為正面消息之比率
準確度 Accuracy	預測所有消息中，判斷正確之比率
召回率 Recall	實際為正面消息，能夠預測出正面消息之比率
F 值 F1-score	召回率和精確度之加權調和平均值，綜合兩個指標的結果

二、可用性評估

本研究計算三個詞典中詞的使用程度、每篇重大訊息所出現詞的個數與累計次數之平均及標準差，針對所使用的詞典做可用性評估。

1. 關鍵詞使用比例

(1) 好/壞關鍵詞使用比例

$$\text{好/壞詞使用比例} = \frac{\text{使用到的好/壞詞 (個)}}{\text{詞典所有好/壞詞(個)}}$$

(2) 關鍵詞總使用比例

$$\text{總使用比例} = \frac{\text{使用到的好詞 + 壞詞 (個)}}{\text{詞典所有好 + 壞詞(個)}}$$

透過計算好/壞關鍵詞的使用比例可以了解詞典中有多少好/壞關鍵字被使用，也可以掌握到在重大訊息中會出現哪些詞彙。總使用比例代表是好詞與壞詞總使用率，幫助了解詞典之可用性。

2. 關鍵詞出現個數

(1) 好/壞關鍵詞每篇平均出現數(個)


$$\text{好/壞詞平均出現個數} = \frac{\text{好/壞詞總出現個數}}{\text{重大訊息(則)}}$$

好/壞關鍵詞平均出現個數，代表在一則重大訊息中平均出現幾個好/壞關鍵詞。

(2) 標準差

標準差數值愈大表示數值與平均值的差距愈大，反映這些重大訊息裡面出現關鍵字個數之離散程度。

(3) 好壞詞出現率


$$\text{好/壞詞比例} = \frac{\text{好/壞詞總出現個數}}{\text{好 + 壞詞總出現個數}}$$

即好/壞關鍵詞之出現率，可以幫助了解在重大訊息中好壞詞出現的比例。

3. 關鍵詞累計出現數(次)

(1) 好/壞關鍵詞每篇平均累計出現數(次)

$$\text{好/壞詞平均累計出現次數} = \frac{\text{好/壞詞總累計出現次數}}{\text{重大訊息(則)}}$$

好/壞關鍵詞平均累計出現次數，代表在一則重大訊息中平均出現幾次好/壞關鍵詞。

(2) 標準差

標準差數值愈大表示數值與平均值的差距愈大，反映這些重大訊息裡面累計出現關鍵字個數之離散程度。

(3) 好壞詞出現率

$$\text{好/壞詞比例} = \frac{\text{好/壞詞總累計出現次數}}{\text{好} + \text{壞詞總累計出現次數}}$$

即好/壞關鍵詞之累計出現率，可以幫助了解在重大訊息中好壞詞累計出現次數的比例。



第四章 實證結果與分析

第一節 樣本基本描述

本研究欲使用字典法與 BERT 方法提取重大訊息之情緒，檢視其預測結果是否與實際市場反應相符。研究期間為 2010 年 1 月 1 日至 2021 年 12 月 31 日。因非上市上櫃公司無受到公開市場行情之規定，無法取得及進一步觀察其價格變化，故樣本採用公開資訊觀測站上市上櫃金融類股公司之重大訊息，共有 31 家公司及 39301 則重大訊息。而股票資料取自台灣經濟新報資料庫。

表4-1 顯示重大訊息之敘述統計。可觀察到標準差在 50 以上總共有 6 間公司，分別是「三商壽」、「富邦金」、「國泰金」、「開發金」、「中信金」及「合庫金」。這十二年以來發過最多重大訊息之前三名為「富邦金」、「國泰金」及「元大金」。

表4-1 樣本敘述統計

公司名稱	代碼	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	總計	平均	標準差
彰銀	2801	41	54	38	47	44	43	37	44	34	37	53	62	534	44.5	8.28
京城銀	2809	31	45	36	37	45	41	37	33	45	39	39	36	464	38.67	4.64
台中銀	2812	45	53	38	36	43	36	28	45	35	28	38	38	463	38.58	7.14
旺旺保	2816	27	35	18	19	21	14	20	10	15	15	11	12	217	18.08	7.18
華票	2820	14	15	11	12	12	14	8	10	15	22	18	24	175	14.58	4.74
台產	2832	20	39	14	17	18	16	13	23	10	11	24	20	225	18.75	7.76
臺企銀	2834	47	41	46	42	51	50	41	42	44	48	44	49	545	45.42	3.58
高雄銀	2836	33	47	34	29	30	25	27	46	35	44	53	42	445	37.08	9.03
聯邦銀	2838	21	13	24	25	23	26	24	30	29	29	32	51	327	27.25	8.99
遠東銀	2845	26	19	24	20	29	24	35	27	24	15	15	31	289	24.08	6.11
安泰銀	2849	41	30	32	17	14	20	23	14	13	19	19	29	271	22.58	8.67
新產	2850	19	24	26	25	26	24	26	26	24	16	22	25	283	23.58	3.15
中再保	2851	16	13	17	10	17	13	12	12	10	7	21	19	167	13.92	4.12
第一保	2852	25	18	11	21	18	15	18	17	17	23	15	22	220	18.33	3.89
三商壽	2867	3	8	181	133	65	43	62	52	45	29	32	46	699	58.25	50.84
華南金	2880	135	133	128	122	95	141	122	122	105	117	181	131	1532	127.67	21.12
富邦金	2881	248	368	421	450	421	389	367	309	317	392	369	443	4494	374.50	59.51

公司名稱	代碼	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	總計	平均	標準差
國泰金	2882	320	395	401	439	337	311	348	347	341	389	263	285	4176	348.00	50.90
開發金	2883	198	172	201	352	246	291	191	183	197	214	222	215	2682	223.50	51.30
玉山金	2884	101	115	106	86	112	113	107	106	85	88	108	102	1229	102.42	10.54
元大金	2885	222	293	339	287	338	310	406	335	317	372	270	291	3780	315.00	48.22
兆豐金	2886	149	162	222	170	170	176	215	138	148	154	133	148	1985	165.42	28.02
台新金	2887	159	140	115	108	144	154	136	120	147	124	116	137	1600	133.33	16.50
新光金	2888	145	94	99	160	158	136	103	99	147	145	201	197	1684	140.33	36.50
國票金	2889	78	102	53	68	91	61	71	95	66	91	119	111	1006	83.83	20.83
永豐金	2890	203	222	192	195	202	172	182	221	144	166	107	148	2154	179.50	34.03
中信金	2891	185	252	299	281	253	389	329	319	349	356	294	328	3634	302.83	55.17
第一金	2892	172	164	148	165	142	154	152	146	135	121	102	149	1750	145.83	19.47
王道銀行	2897	30	23	23	57	44	41	60	65	50	49	65	62	569	47.42	15.49
上海商銀	5876	0	0	0	11	13	28	20	31	51	31	29	35	249	20.75	16.23
合庫金	5880	0	30	176	177	218	173	126	145	154	159	143	148	1649	137.42	62.02
		2754	3119	3473	3618	3440	3443	3346	3212	3148	3350	3158	3436	39301		

表4-2 為重大訊息筆數占比。重大訊息中包含高達 94.5% 之無情緒消息，好消息與壞消息只占約 3% 與 2%。

表4-2 重大訊息筆數占比

情緒	筆數	占比
G	1227	3.122
B	918	2.336
N	37156	94.54
	39301	100

第二節 字典法成效分析

一、字典法之可用性

1. 關鍵詞使用比例

計算好/壞關鍵詞的使用比例可以了解詞典中有多少好/壞關鍵字被使用，結果呈現於表4-3。FESD 好詞有 369 個，其中使用 216 個，使用率為 58.70%，壞詞有 522 個，其中使用 304 個，壞詞使用率為 58.46%；CFSL 好詞有 207 個，其中使用 139 個，使用率為 67.15%，壞詞有 959 個，其中使用 504 個，壞詞使用率為 52.56%；CFSD 好詞有 3338 個，其中有 719 個被使用，使用率為 21.54%，壞詞有 5890 個，其中使用 664 個，壞詞使用率為 11.27%。

三個詞典之總使用率由大到小為 FESD > CFSL > CFSD。FESD 總使用率為 58.56%，CFSL 總使用率為 55.15%，CFSD 總使用率為 14.99%。CFSD 總使用率較低可能是因為詞典詞彙數量較多，加上本身建置過程有添加三部通用詞典，並不是添加財經領域詞典。CFSL 與 FESD 總使用率皆超過 55%，原因可能是總詞彙數量不多，加上編製過程所添加的是財經領域詞典。

表4-3 關鍵詞使用比例

	G	B	總使用率
FESD	58.70%	58.46%	58.56%
CFSL	67.15%	52.56%	55.15%
CFSD	21.54%	11.27%	14.99%

2. 關鍵詞出現個數

藉由關鍵詞出現個數可幫助了解詞典在重大訊息使用之情況。表4-4 顯示關鍵詞出現個數，每則重大訊息平均會出現 FESD 3 個好

關鍵詞及 5 個壞關鍵詞；CFSL 3 個好關鍵詞及 9 個壞關鍵詞；CFSD 4 個好及 4 個壞關鍵詞。

出現個數之標準差可觀察每則重大訊息出現個數之離散程度，其中 CFSL 之好壞詞標準差較懸殊，顯示好壞詞出現個數較不平均。

好壞詞出現率代表好/壞詞占總出現個數之比率。FESD 之好壞詞出現率為 30% 及 70%，壞詞出現比例較多。CFSL 之好壞詞出現率為 16% 及 84%，壞詞出現比例較多。CFSD 之好壞詞出現率各占一半左右，較為平均。綜合來說，三個詞典之壞詞出現個數都比較多，可以推測詞典對重大訊息的判讀結果會傾向壞消息比較多。

表4-4 關鍵詞出現個數

		每則平均	標準差	好壞詞出現率	總計	
FESD	G	2.6138	2.1661	29.61%	73953	249738
	B	4.9793	4.2645	70.39%	175785	
	Total	6.3545	6.0682			
CFSL	G	2.5526	1.8477	15.74%	66005	419320
	B	9.1011	6.4158	84.26%	353315	
	Total	10.6695	8.1533			
CFSD	G	3.7613	3.2936	48.61%	120177	247235
	B	3.9673	3.0161	51.39%	127058	
	Total	6.2908	5.9993			

三個詞典在每則重大訊息中平均出現 6 個以上的好壞關鍵詞，表4-5 為出現個數 6 個以下之重大訊息筆數及比例。FESD 出現 6 個以下好壞關鍵詞有 62%，其中出現個數為 0 的約 7%；CFSL 出現 6 個以下好壞關鍵詞有 31%，其中出現個數為 0 的約 4%；CFSD 出現 6 個以下好壞關鍵詞有 58%，其中出現個數為 0 的約 20%。此比例可連結到詞典之有用性，當重大訊息普遍出現關鍵詞個數不多時，可能會影響到重大訊息情緒判讀的準確度。

表4-5 關鍵詞出現個數小於 6 比例

	出現個數 < 6	占全部比例	出現個數 = 0	占 <6 比例
FESD	24301	62%	1796	7%
CFSL	12145	31%	480	4%
CFSD	22730	58%	4510	20%

3. 關鍵詞累計出現次數

藉由關鍵詞出現次數可以幫助了解詞典在重大訊息累計使用之頻率，如表4-6。每則重大訊息累計平均會出現 FESD 5 次好關鍵詞及 12 次壞關鍵詞，CFSL 6 次好關鍵詞及 25 次壞關鍵詞，CFSD 8 次好/壞關鍵詞左右。其中以 CFSL 累計平均出現 25 次壞關鍵詞為最多，顯示出其詞典之壞詞累計出現次數最多。

出現次數之標準差可以觀察每則重大訊息出現次數之離散程度，其中 CFSL 之好壞詞出現次數標準差較懸殊，顯示好壞詞出現次數較不平均。

好壞詞累計出現率代表好/壞詞出現次數占總出現次數之比率。FESD 之好壞詞出現率為 23% 及 77% 左右，壞詞累計出現次數比例較多；CFSL 之好壞詞累計出現率為 13% 及 87% 左右，壞詞累計出現次數較多；CFSD 之好壞詞出現率各占一半左右，較為平均。

表4-6 關鍵詞累計出現次數

		每則平均	標準差	好壞詞出現率	總計
FESD	G	4.4688	4.4556	23.26%	126436
	B	11.8173	14.0622	76.74%	417186
	Total	13.8323	16.6977		543622
CFSL	G	5.6123	5.3769	12.83%	145123
	B	25.4092	25.8471	87.17%	986412
	Total	28.7915	30.1769		1131535
CFSD	G	7.5855	8.8358	45.78%	242363
	B	8.9646	9.4393	54.22%	287100
	Total	13.4720	16.8193		529463

二、字典法之有用性

本研究從精確度、準確度、召回率及 F 值四個指標觀察詞典其有用性。

1. 傳統上定義

表4-7 為傳統上定義所計算之詞典指標。精確度代表在預測結果當中，預測為 G/B/N 消息的樣本中有多少是 G/B/N 消息的樣本，我們分別計算 G/B/N 消息使用字典法之精確度。以好消息 (G) 來看，三個詞典之精確度由大到小為：CFSD > FESD > CFSL。FESD 精確度為 2.93%，CFSL 精確度為 1.67%，CFSD 精確度為 3.38%。顯示出三個詞典預測為好消息的樣本，實際為好消息的樣本的比例都不超過 5%，推測可能是跟詞典可用性之好關鍵詞出現率較低有關，代表詞典之好關鍵詞識別出重大訊息之好消息能力較弱。以壞消息 (B) 來看，三個詞典之精確度由大到小為：CFSL > CFSD > FESD。FESD 精確度為 2.29%，CFSL 精確度為 2.36%，CFSD 精確度為 2.31%，顯示出三個詞典預測為壞消息的樣本，實際為壞消息的樣本的比例都不超過 3%。以無情緒消息 (N) 來看，三個詞典之精確度由大到小為：CFSD > FESD > CFSL。FESD 精確度為 81.32%，CFSL 精確度為 21.45%，CFSD 精確度為 88.89%。從精確度來看三個詞典預測 G、B 消息之效果比較差。

三個詞典在傳統上定義時候的準確度都很低，FESD 準確度不到 10%，CFSL 準確度不到 3%，CFSD 準確度有 21%。可以發現到好情緒消息 (G) 的精確度都很低，或許與詞典可用性之好關鍵詞出現率較低有關，詞典之好關鍵詞識別出重大訊息之好消息能力較弱。

表4-7 傳統上定義詞典指標有用性

	實際	預測	精確度	準確度	預測	實際	召回率	F1 值
G	194	6613	0.0293		194	1227	0.1581	0.1212
FESD B	665	28999	0.0229	9.82%	665	918	0.7244	0.1729
N	3000	3689	0.8132		3000	36676	0.0818	0.0893
G	4	240	0.0167		4	1227	0.0033	0.0058
CFSL B	908	38441	0.0236	2.66%	908	918	0.9891	0.0518
N	133	620	0.2145		133	36677	0.0037	0.0064
G	361	10668	0.0338		361	1227	0.2942	0.2473
CFSD B	465	20135	0.0231	21.32%	465	918	0.5065	0.3001
N	7554	8498	0.8889		7554	36676	0.2060	0.2095

2. 本研究定義

如果以傳統上累計好詞大於累計壞詞就為好消息之定義，可以發現三個詞典準確度都很低。表4-8 為根據本研究定義所計算之指標，可以看到準確度都有大幅提升。FESD 從原本 9.82% 提升到 59.02%，CFSL 從 2.66% 提升到 16.16%，CFSD 從 21.32% 提升到 74.07%。將有改善之精確度、準確度及 F1 值以粗體字顯示，可發現數值皆有明顯提升，推測是因為我們重大訊息文本中有大量的無情緒消息，如果使用傳統上的定義，會出現較多 G、B 消息。如果使用本研究所定義之標準，能夠考慮到文本的 N 消息，準確度自然也就提高了。

表4-8 本研究定義詞典指標有用性

	實際	預測	精確度	準確度	預測	實際	召回率	F1 值
G	36	1045	0.0344		36	1227	0.0293	0.0559
FESD B	296	13609	0.0218	59.02%	296	918	0.3224	0.4170
N	22864	24647	0.9277		22864	36676	0.6234	0.6064
G	0	58	0		0	1227	0	0
CFSL B	783	32900	0.0238	16.16%	783	918	0.8529	0.2717
N	5566	6343	0.8775		5566	36676	0.1518	0.1565
G	68	1976	0.0344		68	1227	0.0554	0.1031
CFSD B	147	6263	0.0235	74.07%	147	918	0.1601	0.2633
N	28894	31062	0.9302		28894	36676	0.7878	0.7635

準確度代表詞典中預測正確之比例，詞典準確度由大到小為：CFSD > FESD > CFSL。FESD 準確度為 59.02%，CFSL 準確度為 16.15%，CFSD 準確度為 74.07%，顯示 CFSD 比起其他兩個詞典，較能夠正確判讀重大訊息之情緒。從表4-9 正確預測各消息的比例可以發現，FESD 正確預測 N 消息占了 98.57%，CFSL 正確預測 N 消息占了 87.67%，CFSD 正確預測 N 消息占了 99.26%，表示 N 消息正確預測比例占大多數，也就是三個字典法預測 N 消息的效果比較好。三個詞典正確 G 消息之筆數很低，推測可能是跟詞典可用性之好關鍵詞出現率較低有關，代表詞典之好關鍵詞識別出重大訊息之好消息能力較弱。

表4-9 正確預測各消息占比

FESD			CFSL			CFSD		
消息	正確筆數	比例	消息	正確筆數	比例	消息	正確筆數	比例
G	36	0.15%	G	0	0%	G	68	0.23%
B	296	1.28%	B	783	12.33%	B	147	0.51%
N	22864	98.57%	N	5566	87.67%	N	28894	99.26%
總計	23196		總計	6349		總計	29109	

召回率代表實際為正面消息，能夠預測出正面消息之比率。我們分別計算好消息與壞消息之召回率。以好消息來看，三個詞典之召回率由大到小為：CFSD > FESD > CFSL。FESD 召回率為 2.93%，CFSL 召回率為 0%，CFSD 召回率為 5.54%，表示三個詞典實際為好消息的樣本，能夠預測為好消息的樣本的比例。以壞消息來看，三個詞典之召回率由大到小為：CFSL > CFSD > FESD。FESD 召回率為 41.70%，CFSL 召回率為 85.29%，CFSD 召回率為 16.01%，顯示出三個詞典實際為壞消息的樣本，能夠預測為壞消息的樣本的比例。以無消息 (N) 來看，三個詞典之召回率由大到小為：CFSD > FESD > CFSL。FESD 召回率為 62.34%，CFSL 召回率為 15.18%，

CFSD 召回率為 78.78%，顯示出三個詞典實際為無消息的樣本，能夠預測為無消息的樣本的比例。

F1 值為召回率和精確度之加權調和平均值，綜合兩個指標的結果。以好消息 (G) 來看，三個詞典之 F1 值由大到小為：FESD > CFSD > CFSL。FESD F1 值為 5.59%，CFSL F1 值為 0%，CFSD F1 值為 10.31%。以壞消息 (B) 來看，三個詞典之 F1 值由大到小為：FESD > CFSL > CFSD。FESD F1 值為 41.70%，CFSL F1 值為 27.17%，CFSD F1 值為 26.33%。以無消息 (N) 來看，三個詞典之 F1 值由大到小為：CFSD > FESD > CFSL。FESD F1 值為 60.64%，CFSL F1 值為 15.65%，CFSD F1 值為 76.35%。F1 值愈高，表示方法愈有效。綜合來看，FESD 預測負面及無消息效果較好，CFSL 預測負面消息效果較好，CFSD 預測無消息效果較好。



第三節 詞典詞彙出現頻率

一、出現頻率前 30 名

1. 林宜萱 (2013) 財經領域正負向情緒詞典

表4-10 為 FESD 詞彙頻率前 30 名。可發現意義為增的好字，像是增、高、強、增加、提高，都有在前 30 名以內。意義為減的壞字，像是失、損失、差、減、低也都有在前 30 名的出現頻率。壞關鍵詞「關」出現頻率最高，不過單看這個字無法判斷其情緒，需要考慮到使用的情境與用途。壞關鍵詞「有限」出現頻率有 14706 次，不過在重大訊息文本中，當公司名字出現時，可能會包括「有限」這個詞，顯示這個詞在重大訊息文本不是壞詞。

表4-10 FESD 詞彙頻率前 30 名

好關鍵詞	次數	好關鍵詞	次數	壞關鍵詞	次數	壞關鍵詞	次數
重要	12566	許可	1226	關	20971	停止	2946
利益	7285	發展	1198	有限	14706	差	2727
擔保	6678	保證	1159	處分	8587	重大	2589
特別	6477	獲利	987	限制	8532	死	2422
增	5177	提升	946	意見	8229	死亡	2420
第一	3137	完成	889	失	7643	減	2167
重大	2589	預定	831	目的	7536	訴訟	2121
符合	2033	改善	771	損失	7411	低	1831
高	1968	增加	636	相對	7351	轉換	1699
強	1819	報酬	510	費用	6973	不足	1654
最後	1774	效益	487	辭職	5766	修正	1645
合作	1728	提高	429	調整	5147	監督	1532
進行	1414	業績	425	開發	4929	轉讓	1420
超過	1382	慈善	346	執行	3861	禁止	1393
強化	1346	佳	315	召開	3165	超過	1382

2. 張津挺 (2015) 中文財務情緒字典

表4-11 為 CFSL 詞彙頻率前 30 名。可發現到 (1) 壞關鍵詞前幾名累計出現次數很多可能是因為是單詞比較多 (7/10)，這些單詞出現頻率很高，但不確定是否能夠把它當作判讀情緒的基準，因為單看像是「說」、「對」、「稱」等單詞 (紅字處)，可能無法判讀其情緒，如果出現頻率高卻是無法判讀情緒的字詞，可能會影響到重大訊息文本判讀的結果，造成誤判的狀況。(2) 前 30 名詞彙總計 10 個好關鍵詞及 9 個壞關鍵詞與林宜萱財經領域正負向情緒詞典重複到 (粗體處)，可能是因為兩者詞典的基礎皆為 Loughran & McDonald 詞典，加上此詞典之建置過程有納入林宜萱財經領域正負向情緒詞典的詞彙。

表4-11 CFSL 詞彙頻率前 30 名

好關鍵詞	次數	好關鍵詞	次數	壞關鍵詞	次數	壞關鍵詞	次數
大	9281	完成	889	公司	35103	擔保	6678
取得	8792	放款	838	會	30123	保險	6445
需	7490	最終	719	有	22532	合理	5972
利益	7285	收益	664	稱	21511	辭職	5766
影響	3351	贖回	551	不適	18573	調整	5147
資訊	3201	報酬	510	為	16907	開發	4929
第一	3137	提高	429	對	13354	將	3922
使	2369	慈善	346	說	11589	分配	3509
授權	2334	有效	247	損	9906	結果	3198
特殊	2329	償還	237	限制	8532	停止	2946
強	1819	有利	220	承認	8294	解除	2742
合作	1728	穩定	218	異議	7714	差	2727
超過	1382	最高	196	公開	7597	死亡	2420
保證	1159	加強	195	目的	7536	使	2369
獲利	987	誠信	177	損失	7411	符	2198

3. 姜富偉、孟令超、唐國豪 (2021) 中文金融情感詞典

表4-12 為 CFSD 詞彙頻率前 30 名。可以觀察到以出現頻率前 30 名來看，總共有 7 個好關鍵詞與 11 個壞關鍵詞分別與前兩個詞典之詞彙重複到 (粗體處)，可能是因為三個詞典皆以 Loughran & McDonald 詞典為基礎。

表4-12 CFSD 詞彙頻率前 30 名

好關鍵詞	次數	好關鍵詞	次數	壞關鍵詞	次數	壞關鍵詞	次數
適用	18678	完整	1283	不適	18573	差異	1662
生效	9047	發展	1198	處分	8587	不足	1654
公開	7597	保證	1159	限制	8532	禁止	1393
權益	7341	獲利	987	生變	8408	事變	1304
具體	7270	流通	983	異動	8306	不確	1264
有權	6783	提升	946	異議	7714	逾期	1240
合理	5972	上市	829	損失	7411	資金不足	1171
理性	5797	發現	810	揭露	6821	反對	1161
同意	5355	捐贈	794	辭職	5766	負數	1158
開發	4929	充實	792	受限	5734	控制	895
獨立	3107	改善	771	停止	2946	減少	780
解除	2742	澄清	656	差	2727	呆帳	747
聯合	2189	動人	563	死亡	2420	審查	688
合作	1728	提高	429	訴訟	2121	涉及	588
強化	1346	創業	420	尚未	1778	案件	559

綜合以上，發現到詞典的總使用率雖然不高，但都集中在某些詞彙出現。由此可以對重大訊息文本會使用到哪些好/壞關鍵詞有基本的掌握。對於詞典內有許多頻率出現很高，卻看不出為何詞典將他分類在好詞或壞詞的狀況，可能會影響到重大訊息文本判讀的結果，造成誤判的狀況。

二、出現頻率為零

1. FESD

(1) 好關鍵詞

表4-13 為 FESD 出現頻率零次的好關鍵詞，可以觀察到很多都為三個字以上的詞彙，在重大訊息中未出現過，可能是因為有些詞彙較白話，而重大訊息的語法形式較精簡。

表4-13 出現頻率零 FESD 好關鍵詞

「twA-1」、一枝獨秀、上攻、上修、大放異彩、大幅挺進、大買小賣、大豐收、不錯、不錯的、互相扶持、止跌回升、付諸實施、令人滿意、功能性強、可獲利、平易、正面向上、用心、由黑翻紅、由賣轉買、企業社會責任獎、先例、全面獲利、多頭、多頭格局、好成績、好很多、好時機、好消息、好評、好運、如日中天、成長力道強勁、成長性極高、成長茁壯、有把握、有效帶動、有能力、佔領、免費、利多、助長、努力深耕、技能、更好的、狂飆、走高、佳績、供不應求、刷新、刺激、卓越的、受惠良多、明白、附加價值高、亮眼、亮眼表現、信譽品牌、前茅、威信、看漲、突出的、紅盤、重返榮耀、革新、首獎、殊榮、特別的、站台、貢獻獲利、送暖、逆勢成長、逆勢走高、高品質、高度肯定、高度競爭力、高點、強壯、第一名、連日走高、頂級、備受看好、喜悅、富的、景氣回升、最夯、最有價值、最有價值國際品牌、最初的、發揚光大、絢爛、著名、買氣大升、買氣狂飆、買超、買盤、越走越順、進步、開高、開高走高、新高點、新進展、業界之冠、跳升、零庫存、預期的、奪得、暢旺、漲幅「第一名」、稱王、稱讚、需求旺盛、需求量龐大、領先、領導能力、領頭、價量齊揚、寬大、樂觀的、標竿、潛力可期、熱忱、熱烈反應、熱銷、銷量一夕暴增、儘管、樹立典範、獨霸、積極的、優秀的、獲利大躍進、獲利成長佳、獲利能力增強、獲利最好、獲利稱王、獲勝、聲望、賺錢的、蟬聯信譽品牌、蟬聯信譽品牌金獎、豐富的、贏得、續創新高、護航、飆漲、歡呼、驕傲、驕傲的、驚人、驚人的、驚豔

(2) 壞關鍵詞

表4-14 為 FESD 出現頻率零次的壞關鍵詞，可以觀察到有些詞彙是比較口語話的說法，有些是用來形容股市的詞彙，比較不會使用在重大訊息文本裡。

表4-14 出現頻率零 FESD 壞關鍵詞

人工薪資不斷上漲、人心惶惶、下降的、下探、下跌的、口令、大砍、大降不合適、歹戲拖棚、包圍、失去、市場緊縮、必然的、打破、生意蒸發、由紅翻黑、曲折、利潤下滑、告誡、批評、抑制、沒有成長、沒指望、系統故障供過於求、刺激、受到衝擊、命運、坦承、延緩、爭議事件太多、表現不佳、表現較差、阻止、削減、勉強、後果、故障、看淡、突然發生迫切、迫切的、面臨挑戰、消極、疲弱、疲弱不振、耗損退位、逆境、逆轉、偏弱、停工、瓶頸、脫去、雪上加霜、插曲、摒棄、最低價、最差、最壞、最壞的、減產、減碼、無一倖免、無力、無效率、痛失股王寶座、買氣急凍、費用提高、跌破、跌停、跌到一半、跌勢、跌勢仍未止穩、跌逾、跌逾二成、開低走低、閒置的、新低點、經濟放緩、經營不善、落空、跳空、跳空跌停、過大的、過去的、道德勸說不成、電費上漲、圖示不符、慘跌、慢、攢壓、滲透、蓄意、駁斥、廢棄物、廣告不實、緩慢、罷工、賣超、賣壓、賣壓則大幅湧現、遭殃、震驚、獨占、獲利不理想、獲利衰退、艱苦、避開、隱藏、擺脫、轉向、嚴重打擊、嚴峻的、懸浮、競爭力下滑、議論、躊躇、露出、彎曲的、聽從的、襲擊、驚人的.....

2. CFSL

(1) 好關鍵詞

表4-15 為 CFSL 出現頻率零次的好關鍵詞，可觀察到多數詞彙並不是重大訊息會出現的用字。

表4-15 出現頻率零 CFSL 好關鍵詞

偉大、價格上漲、儘管、光榮、免費、創造力、創造性、勤奮、印象、取悅、喜悅、圓滿、壯觀、多才多藝、夢想、實益、巧奪天工、平滑、建設性、得天獨厚、恭維、愉快、排他性、推進、擅長、收藏、放心、有利可圖、活力、無與倫比、熟練、熱情、特殊性、發明家、確鑿、稱讚、穩健性、立功、精美、精通、素養、繁榮、聰明才智、聲望、自信、致意、興奮、著名、蓬勃發展、註定、讚譽、賦予、贏利、贏得、赫然、輝煌、進步、銘記、難以置信、靈感、青睞、革命、革命性、革新、首映、驚喜、鼓舞人心

(2) 壞關鍵詞

表4-16 為 CFSL 出現頻率零次的壞關鍵詞，同樣可觀察到很多詞彙並不像是重大訊息文本會出現的情境。

表4-16 出現頻率零 CFSL 壞關鍵詞

一文不值、下落不明、不公平、不切實際、不可逆轉、不可避免、不合情理、不得人心、不景氣、不由自主、不知不覺、不道德、中傷、串謀、令人不安、位移、佔領、作證、併發症、供過於求、侵蝕、倒閉、偏見、停工、停滯不前、停用、停頓、偽證、傳喚、僵局、克服、免責、公害、冒犯、冤枉、分歧、削減、剝奪、創傷、反壟斷、反托拉斯、古怪、召回、合理化、同謀、否定、嚴重性、回避、堅持、失寵、失職、失衡、失調、好色、威懾、完好無損、定罪、容忍、寬恕、專利、對簿公堂、崩潰、弱點、心疼、忽視、性騷擾、恐嚇、恐懼、恥辱、悲慘、情有可原、惡化、惹惱、意想不到、慘澹經營、懸浮、打擾、扣除額、抑制、投訴、抱怨、抵押、拒付、拒收、拖延、拘留、拯救、指責、挑釁、推遲、損失賠償、擊敗、收尾、故障、敏感性、教唆、敲詐、斷言、昂貴、暴跌、暴露、曲折、校正、欺詐、歪曲、死刑、死胡同、氣餒、沮喪、波幅、消耗、清倉、滋擾、滯銷、災難、無能。。

3. CFSD

(1) 好關鍵詞

表4-17 為 CFSD 出現頻率零次部份的好關鍵詞，可觀察到有一些詞彙為中國日常用語，像是靠譜。有些詞彙字數較多較白話，也有許多為成語和形容詞，比較不常於重大訊息中出現。

表4-17 出現頻率零 CFSD 好關鍵詞

榜樣、飽滿、筆路藍縷、變得更好、標杆、不可思議的、破格錄用、才思、喜兆、彩頭、昌盛、超群的、成功者、充裕的、寵愛、從容不迫、達到目的、打動人、大獲成功、得天獨厚的、巔峰時刻、鬥志昂揚、獨家報導、獨家經營、獨特魅力、多才多藝的、奮勇爭先、風度翩翩、高瞻遠矚、給人深刻印象、功成名就、關鍵時刻、國泰民安、合算的、恒心、歡呼雀躍、活力四射、獲得授權的、積極的、急速發展、建設性的、金點子、錦囊妙計、精巧、精通、犒勞、靠譜、可預測、空前未有的、寬恕的、禮貌的、理想的、廉正、領導能力、領導者、令人高興的事、令人滿意地、令人信服地、滿腔熱情地、夢想成真、密切地、妙計、名列前茅、明確地、能手、弄潮兒、普及的、奇特、謙遜的、享有聲望的、情況好轉、如願以償、潤色、善行、令人滿足、令人滿足的、石破天驚、受尊敬的、舒適的、討人喜歡的、套期保值、特許經營權、同舟共濟、頭籌、突破性進展。。

(2) 壞關鍵詞

表4-18 為 CFSD 出現頻率零次部份的壞關鍵詞，可觀察到有些是形容詞的詞彙，比較不會出現在重大訊息。

總結以上，三個詞典中出現頻率為零的詞彙，在重大訊息中未出現過的原因可能是詞彙較為白話，以及比較常在日常生活中的用語使用，有些是建置環境上的差異。加上重大訊息大多數時候是在敘述事實，比較不會用到情緒性的詞彙，不會誇大說明，形容詞類的詞彙也比較少會出現在重大訊息文本。由此可知，重大訊息文本

與財經新聞有文本上的差異，用字遣詞差異大，也可以連結到對於重大訊息文本詞典之可用性為何不好。

表4-18 出現頻率零 CFSD 壞關鍵詞

敗壞名聲、被沒收的、變節、不適當、妨礙、焦頭爛額、困惑、落後於、密謀策劃、欺詐地、刑事的、受傷害的、危害、有害的、暗地裡破壞、暗中破壞、揠苗助長、敗壞、薄弱環節、暴跌、卑劣地、悲劇、背叛、被懲罰的、被毀、被禁止的、被拋棄的、被誤解的、崩潰、崩盤、編造的、貶黜、貶低、皮相、剝奪財產、剝離、捕風捉影、裁減人數、蠶食、慘案、慘敗、操之過急、操縱的、查抄、查封的、差於預期、拆臺、攙假、攙雜的、產權負擔、遲鈍、恥辱、醜聞、出故障的、處理不當、處於不利地位的、處置失當、次品、刺痛、粗心、篡奪、篡權、挫敗、挫敗的、挫折、打碎、大禍、逮捕的、帶來痛苦、怠工、倒閉、低價處理、低劣、低市值、敵對、抵觸、底穀、定價過高、定罪、丟臉、動盪的、動盪的局面、動亂、瀆職、斷絕、多付的款項、奪權、惡化、惡習、遏制的、發愁、發生故障、罰錢、反駁、反對的、反壟斷的、犯錯誤、坊間傳聞、仿冒品、放縱、非法佔有.....

第四節 BERT 方法成效

將隔日報酬率漲超過 3% 之重大訊息標註成好消息 (G)；跌超過 -3% 標註成壞消息 (B)；其他為無情緒消息 (N)，總共 39301 筆重大訊息，包含 1227 的好消息、918 筆壞消息及 37156 筆無消息，因考慮到 BERT 字數限制，將超過 512 字的重大訊息資料過濾掉，過濾後剩下 836 筆好消息、687 筆壞消息與 26541 筆無消息，將每個屬性的資料各取 687 筆，其中每種各 654 筆拿去做訓練資料，其餘資料做為測試資料，訓練資料集占 85% 測試資料集為 15%。表 4-19 為資料集之重大訊息筆數。

表4-19 BERT 資料集重大訊息筆數

消息屬性	G	B	N	總數
筆數	1227	918	37156	39301
過濾	391	231	10615	
剩餘	836	687	26541	
資料集	687	687	687	2061
訓練資料集	654	654	654	1962
測試資料集(15%)	33	33	33	99

將 1962 筆資料放入 BERT 模型訓練後，用 99 筆測試資料集測試 BERT 是否能正確判讀文本情緒，發現 99 筆測試資料集皆預測為無消息，也就是說 BERT 模型將 33 筆好消息與 33 筆壞消息皆預測為無消息，準確率為 33.3%。表 4-20 為使用 BERT 方法之準確率。

表4-20 BERT 準確率

好消息	0/33 (預測成無消息 33 筆)
壞消息	0/33 (預測成無消息 33 筆)
無情緒消息	33/33
準確率	33.3%

第五節 重大訊息案例探討

1. 實際壞消息

表5-1 重大訊息範例 1 元大金

序號	1	發言日期	109/03/13	發言時間	14:54:48
發言人	翁健	發言人職稱	總經理	發言人電話	02-27811999
主旨	元大金控補充公告認購子公司元大期貨 108 年度現金增資事宜。				
符合條款	第 20 款	事實發生日	109/03/13		
說明	<p>1.標的物之名稱及性質（屬特別股者，並應標明特別股約定發行條件，如股息率等）： 元大期貨股份有限公司普通股股票。</p> <p>2.事實發生日:109/3/13~109/3/13</p> <p>3.交易數量、每單位價格及交易總金額: (1)交易數量：9,721,425 股(補充公告); (2)每單位價格：每股新臺幣 48 元(補充公告); (3)交易總金額：466,628,400 元(補充公告)。</p> <p>4.交易相對人及其與公司之關係（交易相對人如屬自然人，且非公司之關係人者，得免揭露其姓名）： 元大期貨股份有限公司； 該公司為本公司 68.65% 持有之子公司。</p> <p>5.交易相對人為關係人者，並應公告選定關係人為交易對象之原因及前次移轉之所有人、前次移轉之所有人與公司及交易相對人間相互之關係、前次移轉日期及移轉金額: 不適用。</p> <p>6.交易標的最近五年內所有權人曾為公司之關係人者，尚應公告關係人之取得及處分日期、價格及交易當時與公司之關係: 不適用。</p> <p>7.本次係處分債權之相關事項（含處分之債權附隨擔保品種類、處分債權如有屬對關係人債權者尚需公告關係人名稱及本次處分該關係人之債權帳面金額: 不適用。</p> <p>8.處分利益（或損失）（取得有價證券者不適用）（原遞延者應列表說明認列情形）： 不適用。</p> <p>9.交付或付款條件（含付款期間及金額）、契約限制條款及其他重要約定事項: (1)交付或付款條件：於繳款期間內繳付本次現金增資款(補充公告)。 (2)契約限制條款及其他重要約定事項：本案業經金管會民國 109 年 1 月 2 日金管銀控字第 1080225627 號函核准投資(補充公告)。</p> <p>10.本次交易之決定方式、價格決定之參考依據及決策單位: (1)交易之決定方式：係以現金增資方式認購。 (2)價格決定之參考依據：依元大期貨 109 年 2 月 6 日董事會決議之現增每股發行價格(補充公告)。 (3)決策單位：本公司審計委員會同意及董事會通過後辦理。</p> <p>11.取得或處分有價證券標的公司每股淨值: 37.09 元</p> <p>12.迄目前為止，累積持有本交易證券（含本次交易）之數量、金額、持股比例及權利受限情形（如質押情形）： 本公司持有元大期貨股份有限公司 (1)累積持有本交易證券數量：普通股 169,188,707 股(補充公告)。 (2)累積持有金額：3,887,693,035 元(補充公告)。 (3)持股比例：67.68%(補充公告)。</p>				

	<p>(4)權利受限情形：無</p> <p>13.迄目前為止，依「公開發行公司取得或處分資產處理準則」第三條所列之有價證券投資（含本次交易）占公司最近期財務報表中總資產及歸屬於母公司業主之權益之比例暨最近期財務報表中營運資金數額（註二）：</p> <p>(1)有價證券投資占公司最近期財務報表中總資產比例：96.9%(補充公告)。</p> <p>(2)占歸屬於母公司業主之權益之比例：111.38%(補充公告)。</p> <p>(3)最近期財務報表中營運資金數額：不適用。</p> <p>14.經紀人及經紀費用：不適用。</p> <p>15.取得或處分之具體目的或用途：為子公司元大期貨股份有限公司充實營運資金、強化財務結構、提高調整後淨資本額(ANC)及淨值。</p> <p>16.本次交易表示異議董事之意見：無。</p> <p>17.本次交易為關係人交易：是</p> <p>18.董事會通過日期：民國 108 年 11 月 27 日</p> <p>19.監察人承認或審計委員會同意日期：民國 108 年 11 月 26 日</p> <p>20.本次交易會計師出具非合理性意見：不適用</p> <p>21.會計師事務所名稱：不適用</p> <p>22.會計師姓名：不適用</p> <p>23.會計師開業證書字號：不適用</p> <p>24.是否涉及營運模式變更：否</p> <p>25.營運模式變更說明：不適用</p> <p>26.過去一年及預計未來一年內與交易相對人交易情形：過去一年與交易相對人無交易</p> <p>27.資金來源：自有資金</p> <p>28.其他敘明事項：無</p>
--	--

從表 5-2 可以發現 CFSL 與 CFSD，將「開發」、「合理」兩個詞彙分類在不同類別。重大訊息範例 1 為元大金控補充公告認購子公司元大期貨 108 年度現金增資事宜，市場反應實際為壞消息。三個詞典累計壞關鍵詞皆比好關鍵詞多，結論顯示透過字典法判讀與這則重大訊息之市場反應相符合。

表5-2 重大訊息範例 1 關鍵字表

詞典出現關鍵字		累 G	累 B	預測	T/F
林宜萱	G: 增,高,強,強化,提高,重要,利益,特別,擔保				
	B: 失,目的,有限,相對,限制,涉及,處分,費用,開發,意見,損失,調整,關	14	43	B	T
張津挺	G: 利益,取得,大,強,提高,需				
	B: 不適,充,充公,公司,公開,合理,對,承認,損,損失,擔保,會,有,為,異議,目的,稱,說,調整,開發,限制	16	120	B	T
姜富偉	G: 強化,提高,充實,有權,同意,開發,理性,適用,合理,權益,具體,公開,				
	B: 異議,揭露,涉及,損失,限制,處分,受限,充公,不適	25	41	B	T

2. 實際好消息

表5-3 重大訊息範例 2 國泰金

序號	1	發言日期	100/08/10	發言時間	08:48:11
發言人	李偉正	發言人職稱	副總經理	發言人電話	27087698#7760
主旨	本公司董事會通過買回公司股份				
符合條款	第 35 款	事實發生日	100/08/10		
說明	1.董事會決議日期:100/08/10 2.買回股份目的:股權轉換 3.買回股份種類:普通股 4.買回股份總金額上限(元):9000000000 5.預定買回之期間:100/08/10~100/10/09 6.預定買回之數量(股):200000000 7.買回區間價格(元):33.00~45.00 8.買回方式:自集中交易市場買回 9.預定買回股份占公司已發行股份總數之比率(%):1.93 10.申報時已持有本公司股份之累積股數(股):0 11.申報前三年內買回公司股份之情形:無 12.已申報買回但未執行完畢之情形:無 13.董事會決議買回股份之會議紀錄: 案由:擬實施 100 年第 1 次庫藏股買回本公司股份,提請審議。 決議:經全體出席董事無異議照案通過。 14.「上市上櫃公司買回本公司股份辦法」第十條規定之轉讓辦法: 不適用 15.「上市上櫃公司買回本公司股份辦法」第十一條規定之轉換或認股辦法: 國泰金融控股股份有限公司以買回本公司股份參與發行海外存託憑證暫訂發行辦法 一、發行公司 國泰金融控股股份有限公司(以下簡稱「國泰金控」或「發行公司」) 二、發行海外存託憑證之目的				

	<p>國泰金控為使公司籌集資金管道多樣化及便利化，擬以買回之已發行股份參與發行海外存託憑證。</p> <p>三、海外存託憑證發行條件</p> <p>(一) 預定發行日期</p> <p>預定於財政部證券暨期貨管理委員會（以下簡稱「證期會」）核准後三個月內發行，必要時得向證期會申請延期。</p> <p>(二) 預定發行總單位數</p> <p>暫定以 44,204,032 單位為上限。</p> <p>(三) 海外存託憑證所表彰之有價證券股數</p> <p>暫定每單位海外存託憑證表彰發行公司普通股 10 股。</p> <p>(四) 預計單位發行價格之訂定方式</p> <p>本次海外存託憑證發行價格擬參酌發行價格訂價日（以下簡稱「訂價日」）當日發行公司普通股於台灣證券交易所之收盤價或訂價日之前一段期間發行公司普通股於台灣證券交易所之平均收盤價，擇一作為海外存託憑證發行參考價格。</p> <p>(五) 預定發行總金額</p> <p>預計發行總金額上限為美金__億元，惟實際發行金額依海外存託憑證實際發行單位總數及以訂價日（定義如第三(四)段）之發行價格計算之。</p> <p>四、海外存託憑證持有人權利義務之行使</p> <p>海外存託憑證持有人之權利義務，應依中華民國相關法令規定及存託契約相關約定規範之。存託契約主要約定事項如下：</p> <p>(一) 表決權之行使</p> <p>海外存託憑證持有人不能直接行使其持有存託憑證所表彰普通股之表決權，但仍有權指示存託機構依存託契約行使表決權。</p> <p>(二) 海外存託憑證兌回方式</p> <p>海外存託憑證持有人得依中華民國相關法令及存託契約之規定，請求存託機構兌回及交付由海外存託憑證所表彰之國泰金控普通股股份；或存託機構依持有人請求兌回及出售由海外存託憑證所表彰之國泰金控普通股股份後，依中華民國相關法令及存託契約之規定，將出售所得之款項扣除相關費用、稅捐及政府規費，結售成美金，並交付予請求出售之海外存託憑證原持有人。</p> <p>(三) 股利分配</p> <p>除存託契約另有規定外，海外存託憑證持有人享有與發行公司普通股股東分派股利之同等權利。未來發放股票股利時，存託機構得依海外存託契約之規定，按海外存託憑證持有人原持有比例增發海外存託憑證予海外存託憑證持有人，或代海外存託憑證持有人出售該股票股利，並將出售所得之金額扣除相關費用及稅捐後，按比例分配與海外存託憑證之原持有人。</p> <p>(四) 新股優先認購權</p> <p>發行公司辦理現金增資或其他認購股權時，除存託契約另有規定外，海外存託憑證持有人享有與發行公司普通股股東同等之優先認股權利，存託機構應於法律許可之範圍內，按存託契約之規定，提供海外存託憑證持有人該等權利，或代海外存託憑證持有人出售此權利，並將出售所得之價金扣除相關費用及稅捐後，按比例分配與海外存託憑證之原持有人。</p> <p>五、海外存託憑證所表彰之有價證券之來源</p> <p>由國泰金控提供買回庫藏股股份於不超過 442,040,320 股之範圍內，作為參與發行海外存託憑證之原有價證券。</p> <p>六、發行方式及交易地點</p> <p>(一) 發行方式</p> <p>發行公司擬採全數公開發行方式，於中華民國境外依相關法令規定參與發行海外存託憑證。</p> <p>(二) 發行地點</p> <p>發行公司擬於歐洲、亞洲及美洲依相關法令募集銷售海外存託憑證，但不在中華民國境內銷售。</p> <p>(三) 交易地點</p> <p>海外存託憑證暫訂於盧森堡證券交易所上市。</p> <p>七、海外存託憑證發行額度之增加及於原發行額度內之再發行</p> <p>依中華民國現行法令規定，海外存託憑證發行完成後，除發行公司辦理現金增資發行新股、盈餘或資本公積配發新股外，未經主管機關核准，存託機構不得增加海外存託憑證之發行單位。惟如經存託契約及保管契約載明，於海外存託憑證兌回且所兌回之發行公司普通股經出售後，存託機構得依中華民國法令及存託契約之規定在原發行額度內再發行海外存託憑證。</p> <p>八、海外存託憑證發行及存續期間相關費用之分攤方式</p> <p>所有海外存託憑證發行之相關費用，均由參與股東負擔。海外存託憑證發行後存續期間之費用（包括交易所上市、資訊揭露及其他相關費用）原則由發行公司負擔。</p>
--	---

	<p>九、中華民國相關稅務規定 根據現行中華民國稅法規定，對於以保留盈餘發放之現金及股票股利，持有海外存託憑證之外國人將被課徵 20% 之股利所得稅（現金以實際發放金額計，股票股利以票面金額計）。依現行中華民國法令，出售表彰海外存託憑證之國泰金控普通股應依出售價格繳納 0.3% 之證券交易稅，出售國泰金控普通股之所得目前不需繳納證券交易所稅。</p> <p>十、出售限制 海外存託憑證不得直接或間接於中華民國境內募集、銷售或交付。</p> <p>十一、發行條件之修改 發行公司與主辦承銷商及參與股東得於協商同意並經主管機關核准後修改本發行辦法所載之海外存託憑證發行條件。</p> <p>十二、準據法 出售海外存託憑證之承銷契約及發行海外存託憑證之存託契約將以美國紐約州法律為準據法。但海外存託憑證之核准參與發行則依中華民國法令辦理並受中華民國法令之限制。</p> <p>16. 董事會已考慮公司財務狀況，不影響公司資本維持之聲明： 國泰金融控股股份有限公司董事會聲明書 一、本公司經一〇〇年八月十日第四屆第八次臨時董事會三分之二以上董事之出席及出席董事超過二分之一之同意通過，自申報日起二個月內於集中交易市場（證券商營業處所）買回本公司股份二億股。 二、上述買回股份總數，僅占本公司已發行股份之百分之一．九三，且買回股份所需金額上限僅占本公司流動資產之百分之三五．四七，茲聲明本公司董事會已考慮公司財務狀況，上述股份之買回並不影響本公司資本之維持。 三、本聲明書業經本公司上述同次董事會議通過，出席董事均同意本聲明書之內容，併此聲明。</p> <p>17. 會計師或證券承銷商對買回股份價格之合理性評估意見： 此次買回股份對國泰金融控股股份有限公司之財務結構、每股淨值、每股盈餘及股東權益報酬率，尚無產生重大影響或變動。</p> <p>18. 其他證期局所規定之事項： 無</p>
--	---

從表 5-4 可以發現林宜萱好壞關鍵詞中皆有「重大」這個詞彙。CFSL 與 CFSD，將「開發」、「合理」兩個詞彙分類在不同類別。

重大訊息範例 2 為國泰金董事會通過買回公司股份之公告，市場反應實際為好消息，三個詞典當中只有 CFSD 之累計好關鍵詞比累計壞關鍵詞多，這則重大訊息透過字典法判讀與實際市場反應符合，其餘兩個詞典皆判讀失敗。

表5-4 重大訊息範例 2 關鍵字表

詞典出現關鍵字		累 G	累 B	預測	T/F
林宜萱	G: 完成,享有,超過,許可,報酬,便利,增,增加,預定,重大				
	B: 不在,不能,目的,扣除,有限,保留,負擔,重大,限制,執行,費用,超過,開發,意見,轉換,轉讓,證實,關	23	48	B	F
張津挺	G: 享有,使,報酬,大,完成,影響,資訊,超過				
	B: 不能,不適,使,公司,公開,分配,合理,需,圍,實施,審議,將,對,延期,必要,按,收盤,會,有,未經,為,異議,目的,稱,範圍,聲明,臨時,評估,請求,證實,負擔,開發,限制,完成	18	150	B	F
姜富偉	G: 優先,享有,適用,理性,有權,同意,多樣化,人權,上市,多樣,表彰,便利,權益,公開,合理,開發	29	9	G	T
	B: 異議,負擔,揭露,限制,延期,不適,不能				



第五章 結論與建議

第一節 研究結論

本研究針對 2010 年 1 月 1 日至 2021 年 12 月 31 日金融股之重大訊息，嘗試透過字典法與 BERT 方法對重大訊息文本作情緒分析。使用過去文獻所建置的三個財經領域情緒詞典進行情緒分析，觀察其預測結果是否與實際市場反應相符。透過三個詞典關鍵詞使用程度及文本中出現情緒詞彙個數與次數，觀察我們選用之財經領域情緒詞典對於重大訊息文本之可用性。從精確度、準確度、召回率、F1 值四個指標評估詞典在重大訊息有用性之成效。

研究發現，字典法與 BERT 方法針對情緒較中性之重大訊息文本成效並不好。不過我們針對傳統上預測情緒的定義方式做改善，本研究定義累計好詞與累計壞詞之差距需要超過五個以上才能定義為好壞消息，否則定義為無情緒消息。可以發現，有用性指標皆明顯改善，表示效果有提升，推測是透過本研究預測情緒定義，讓重大訊息大量無情緒消息被預測正確。從詞典前 30 名出現頻率詞彙觀察到詞典裡的好壞關鍵詞會影響到重大訊息情緒的判讀，例如：CFSL 中「說」、「對」、「稱」等單詞，可能無法判讀其情緒，如果出現頻率高卻是無法判讀情緒的字詞，可能會影響到重大訊息文本判讀的結果，造成誤判的狀況。而由詞典中出現頻率為零的詞彙，推測可能是關鍵詞彙較為白話，以及比較常在日常生活中的用語使用，加上重大訊息大多數時候是在公告事實，比較不會用到情緒性的詞彙，用詞也較精簡。由此可知，重大訊息文本與財經新聞有文本上的差異，用字遣詞差異大，也可連結到詞典之可用性對於重大訊息文本效果不好的原因。

第二節 未來展望

本文探討字典法與 BERT 方法對於金融股重大訊息情緒分析之成效，未來研究方向可針對其他不同產業之重大訊息作分析。本研究在使用三個詞典進行情緒分析時，發現到有些字詞分類不適合的地方，未來可以針對詞典建置方面做改善。我們在機器學習方法中選擇 BERT 方法來做測試，未來可添加不同特徵選取，也可測試其他機器學習的方法，或是將字典法結合機器學習方法來對重大訊息文本作分析。

公開觀測站重大訊息之文本內容較制式，情緒較中性，對於情緒分析方面或許效果並不好，若想針對情緒較多之公司公開揭露文件作研究，建議可以研究法說會。



參考文獻

- 王聿庭 (2019)。探討股市新聞與股票價格之關係。國立雲林科技大學應用外語系研究所碩士論文。
- 吳昱萱 (2018)。股價波動與財務預警-數據分析觀點。國立政治大學會計學研究所碩士論文。
- 林宜萱 (2013)。財經領域情緒詞典之建置與其有效性之驗證-以財經新聞為元件。國立臺灣大學會計學研究所碩士論文。
- 姜富伟、孟令超、唐国豪 (2021)。媒体文本情绪与股票回报预测。经济学 (季刊), 21 (4), 1323-1344。
- 夏鶴芸 (2020)。應用深度學習與自然語言處理新技術預測股票走勢－以台積電為例。國立臺北大學資訊管理研究所碩士論文。
- 陳予得 (2021)。應用遷移學習語文字探勘分析置股東報告書。國立臺灣大學會計學研究所碩士論文。
- 張津挺 (2015)。中文財務情緒字典建構與其在財務新聞分析之應用。臺北市立大學資訊科學系碩士論文。
- 黃娟娟 (2012)。公司年報文字探勘與財務預警資訊內涵。私立逢甲大學商學博士學位學程博士論文。
- 蔡侑達 (2011)。公開資訊觀測站訊息如何影響股票報酬？台灣股票市場日內及日間資料分析。國立暨南國際大學財務金融系碩士論文。
- 鍾雨潼 (2003)。臺灣 2002 年公開資訊觀測站重大訊息之資訊內涵分析。淡江大學管理科學學系碩士論文。
- 謝委霖 (2015)。從財金新聞預測公司財報之營收走勢。國立中山大學資訊管理學系研究所碩士論文。

- Black, F. (1986). Noise. *The journal of finance*, 41(3), 528-543.
- Bryan, S. H. (1997). Incremental information content of required disclosures contained in management discussion and analysis. *Accounting Review*, 285-301.
- Cole, C. J., & Jones, C. L. (2004). The usefulness of MD&A disclosures in the retail industry. *Journal of Accounting, Auditing & Finance*, 19(4), 361-388.
- Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management science*, 53(9), 1375-1388.
- Gu, C., & Kurov, A. (2020). Informational role of social media: Evidence from Twitter sentiment. *Journal of Banking & Finance*, 121, 105969.
- Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2), 1-41.
- Heston, S. L., & Sinha, N. R. (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, 73(3), 67-83.
- Henry, E. (2008). Are investors influenced by how earnings press releases are written?. *The Journal of Business Communication*, 45(4), 363-407.
- Kalyani, J., Bharathi, P., & Jyothi, P. (2016). Stock trend prediction using news sentiment analysis. arXiv preprint arXiv:1607.01958.
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg!. In Proceedings of the international AAAI conference on web and social media (Vol. 5, No. 1, pp. 538-541).
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of finance*, 66(1), 35-65.

- Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., & Anastasiu, D. C. (2019, April). Stock price prediction using news sentiment analysis. In 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService) (pp. 205-208). IEEE.
- Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PloS one*, 10(9), 1-21.
- Schumaker, R. P., Zhang, Y., Huang, C. N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3), 458-464.
- Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*, 107, 730-743.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3), 1139-1168.