國立政治大學商學院金融學系碩士班
碩士學位論文
Department of Money and Banking
College of Commerce
National Chengchi University (NCCU)
Master Thesis

基於自然語言分析建構預測企業信用評等變動之模型
Construction of Corporate Credit Rating Prediction Model Based on Natural Language Analysis

指導教授：江彌修 博士, 趙世偉 博士
Advisor: Mi-Hsiu Chiang, Ph.D., Shih-Wei Chao, Ph.D.
研究生: 陳明勝
Ming-Sheng Chen

中華民國 111 年 6 月
June, 2022

DOI:10.6814/NCCU202200901

誌謝

本研究承蒙江彌修老師與趙世偉老師的幫助下完成。在我大學生活乃至於研究所兩年的求學期間，江彌修老師一直給予我極大的鼓勵，不僅關心我的家庭經濟情況，在學業上更是大力提攜，也因為在老師的建議與幫助下，我在大二時就開始在校外修習資料分析相關課程，我也在那時打下學術研究能力的基礎，而後加入老師的研究團隊後更是獲益匪淺，在過程中有許多挑戰與磨練，更是要感謝老師和學長姐們的包容海涵。趙世偉老師非常幽默風趣且飽讀詩書，回顧過去修習趙老師的課程，無一不滿載而歸，或許當下會覺得老師上課的內容困難且不易理解，但在求學以及研究的過程中，老師教過的內容與概念總會一再出現，讓人覺得若求學時沒有上過趙老師的課程一定會後悔。非常感謝兩位老師對我的指點與提攜。

研究所期間，課業、研究計畫以及求職的壓力都讓人喘不過氣，很感謝金融所同學們和學長姐們在課堂報告和研究計畫上的幫助與切磋，這些都是人生中非常寶貴的經驗，我也在研究所的課程中更確定了自己未來的方向，期望未來能夠無愧我心、回饋社會。

陳明勝

DOI:10.6814/NCCU202200901

摘要

為改進過去語言分析模型無法辨認語言一字多義以及訓練域與預測域不一致之問題，本研究嘗試以 BERT(Bidirectional Encoder Representations from Transformers) 模型針對金融領域文本進行領域遷移 (Domain Adaption)，比較有無經過遷移對模型效能之改進，接著以遷移過之模型分析 RavenPack 資料庫內所含的美國企業相關新聞，並以此建構信用評等變動預警模型。

本研究實證結果顯示，經過遷移之模型預測財金文本情緒的預測準確率比未經遷移之模型高出 30.47%，且領域遷移後辨認的新聞情緒提升對未來企業信用評等變動的預測。另外，本研究建構四個隨機森林模型，用以證明企業金融財務面的媒體情緒隱含對企業未來評級可能變動的有效資訊。

關鍵詞: 自然語言分析、神經網路、領域遷移、企業信用預警。

DOI:10.6814/NCCU202200901

Abstract

To improve the inability of the language analysis model to recognize the polysemy of the language and the inconsistency between the training domain and the prediction domain, this study uses the BERT (Bidirectional Encoder Representations from Transformers) model to perform Domain Adaption for the financial corpus. The adaption improves the performance of the model, and we further use the adapted model to analyze the news related to US companies contained in the RavenPack database and construct an early warning model for credit rating changes.

The empirical results show that the prediction accuracy of the adapted model in predicting the sentiment of financial texts is 30.47% higher than that of the non-adapted one, which shows that adaption learning indeed improves the prediction of the corporate credit rating changes. Also, we developed four different random forest models to prove that the media sentiment on the company's financial news contains effective information on the possible changes in the company's future rating.

Keywords: Natural Language Analysis, Neural Network, Domain Adaption, Corporate Credit Prediction.

DOI:10.6814/NCCU202200901

目錄

DOI:10.6814/NCCU202200901

DOI:10.6814/NCCU202200901

圖目錄

DOI:10.6814/NCCU202200901

DOI:10.6814/NCCU202200901

表目錄

DOI:10.6814/NCCU202200901

第一章　緒論

1.1 研究動機與背景

綜觀過去幾次金融危機，包括 2001 年網際網路泡沫、2007-2008 年金融海嘯、2010 歐債風暴，甚至於 2020 年 Covid-19 疫情造成全球景氣蕭條、各國失業率暴增，無一不對市場產生巨大且深遠的影響。2020 年 1 月疫情剛開始蔓延時，沒有人預料到會對全球金融環境產生如此巨大的黑天鵝效應，美國政府直至 2020 年 3 月才宣布國家進入緊急狀態，但此時疫情已經一發不可收拾，標普 500 指數在十天內熔斷四次，而在此前美國股市只在 1997 年發生過一次熔斷。為了因應新型冠狀病毒的疫情所造成的經濟活動急速冷卻，美國聯準會不得不連續降息 6 碼至接近零水準，並同時啟動量化寬鬆，首次宣布購買稱為墮落天使債券的高收益債券提振市場信心。景氣衰退的同時，企業的無預警倒閉將造成投資人大量損失，2020 年 4 月 1 日懷汀石油公司 (Whiting Petroleum Corporation) 便因為疫情造成的需求衰退以及石油價格戰所引發的石油價格暴跌導致營業額銳減，向法院聲請破產。

幾次金融危機的出現來得快且猛烈，但並非毫無跡象可尋，網際網路普及以來，每年網路資訊量呈指數性爆炸增長，隨著網路傳遞資訊的速度越來越快，且資訊能見度越來越高，投資大眾可以輕而易舉的獲取各類訊息，若能以公開資訊提前捕捉到可能影響企業信用的風險因子並提前避險，便能減少投資損失，故如何提前預警企業潛在的信用風險是政府、學術界以及投資人最重要的課題。

DOI:10.6814/NCCU202200901

1.2 研究目的

過去的違約預警模型主要以Merton (1974) 所提出的結構模型和Jarrow and Turnbull (1995) 提出的縮減式模型等等為基礎開枝散葉，使用的風險因子不外乎公司槓桿比率、無風險利率以及公司的股票波動度等，屬於Liberti and Petersen (2019) 所指出的硬資訊範疇。Liberti and Petersen (2019) 將資訊分為硬資訊與軟資訊。其中硬資訊屬於能夠被量化且以數字型態表達的資訊，如同公司財報中的淨收益、淨利率等數值，而軟資訊則是難以用數字為載體表述，主要為文字、圖像、音頻等非結構化的資訊。Fama (1960) 效率市場假說認為所有市場資訊皆已充分反映在價格上，投資人無法藉由軟資訊對市場進行套利，其後學術界對相關領域研究的缺乏導致其中內蘊的資訊未被市場察覺，然而，近期的文獻指出文本內容包含了尚未反映於股價的市場資訊 (Tetlock et al., 2008)。

軟資訊資料量在網際網路普及後大幅增長，文字分析成為各領域的廣泛課題，過去財經領域主要使用的文字分析方式主要使用字典模型、文本易讀性分析和主題學習模型等等 (Loughran and McDonald, 2016)，藉由這些分析方法辨認文章的情緒、複雜度和文本類別。本研究嘗試以文字分析方法辨認媒體情緒，並以此分析企業信用風險，不過，若以前述方法進行分析，可能會產生偏誤，其最主要的缺陷是前述模型時常無法辨認一字多義的情況，也就是同一個字詞在不同語意環境下有截然不同的意義，以「負債」為例子來說，若此字詞出現在一般的文本中通常具有相當負面的意涵，而在財務領域文本中，「負債」一詞為相當普遍的財報用語，並無顯著的負面情緒，故無法針對文本背景產生相對語意情緒的模型在評估文本情緒時會有諸多偏誤。

綜上所述，本研究嘗試使用 BERT (Bidirectional Encoder Representations from Transformers) 模型以遷移學習方式改進前述模型之缺點，並用以分析企業新聞情緒建構違約預警模型。

DOI:10.6814/NCCU202200901

第二章　文獻回顧

2.1　衡量企業信用風險

學術界已有相當多學者建構企業信用預警之模型，利用會計資料、市場資料等方式預警可能的違約事件，期望能減少投資風險。本節將簡要回顧過去各種企業信用預警的相關設計及實證結果。

2.1.1　風險預警模型

風險預警模型根據其參考的風險因子可大致分為會計資料模型與市場資料模型兩種。Altman (1968) 使用多變量區別分析模型預測企業違約，其使用五項財務指標 (營運資金對總資產比率、保留盈餘對總資產比率、息前稅前盈餘對總資產比率、股本市值對總負債帳面價值比率、銷貨淨額對總資產比率) 作為風險因子，因此屬於會計資料模型。會計資料模型僅使用過去公司財務狀況來預測違約風險，而使用過去財報並不一定能符合當今情況，且忽略了市場風險對公司違約的影響。

而市場資料模型則有結構模型與縮減式模型之分，Merton (1974) 在Merton (1973) 提出選擇權評價模型之後，把公司的權益價值比做歐式買權，將歐式買權不履約的機率類比公司的違約風險，進而以股票市值、股票報酬率波動度和負債價值計算出違約距離，可視為結構模型之濫觴。結構式模型的缺點為其雖然考慮到公司負債價值，但事實上負債較少的公司仍可能因流動性而產生信用風險。縮減模型也屬於市場模型，但其假設企業違約是一種外生的隨機變量，Jarrow and Turnbull (1995) 提出第一個縮減式信用違約模型，其假設違約事件的發生為卜瓦松過程，違約事件以平均 λ 次的頻率發生，其中 λ 即為所謂之違約強度，故縮減式模型也被稱為違約強度模型。縮減式模型因假設企業違約為外生變數，其現實

DOI:10.6814/NCCU202200901

意義較差。

近年來機器學習領域因為資訊設備的改進而蓬勃發展，亦有學者利用各種機器學習方法來預測企業違約。Lee (2007) 使用 MDA 逐步迴歸方式挑選從 297 個財務變數中挑選 10 個作為模型輸入特徵，比較 SVM(Support Vector Machine)、BPN(Back-Propagation Neural Network)、MDA(Multiple Discriminant Analysis) 和 CBR(Case-Based Reasoning) 模型預測韓國 3017 家公司信用評等的優劣，實證結果 SVM 在企業違約預警方面為最佳機器學習模型。雖然機器學習模型可以胃納大量變數，但多數研究皆採取降維方法來降低模型訓練成本，Orsenigo and Vercellis (2013) 使用 PCA(Principal Components Analysis) 和 dbt(Double-Bounded Tree-connected)-isomap 方法將 23 個財務變數與比率降維，並利用 SVM、KNN(K Nearest Neighbor) 和 Naïve Bayes 模型，預測歐美和亞洲的銀行信用評等，實證 dbt-isomap 的降維方式更佳。Hajek and Michalak (2013) 亦在神經網絡模型、SVM、Naïve Bayes、隨機森林、線性判別分類器和最近均值分類器等模型中，使用 81 個財務變數，預測美國 852 家公司信用評等 (43 個變數預測歐洲 244 家公司)，比較各種不同降維方法的有效性。機器學習方法大部分不需要變數假設，相較以往模型可以納入更多資料特徵，但也因此模型可能會遇到維度災難，雖然可以利用降維方法解決，但也因此模型的結果也較難解釋。

2.1.2 信用違約交換

信用違約交換 (Credit Default Swap, CDS) 是 1990 年代出現的一種金融衍生性商品，其主要用途原是提供投資人規避企業違約風險所產生的金融商品，債券的風險承受方可和信用違約交換的承作機構進行交易，在契約期間支付一筆類似保險金的固定費用，若持有債券在契約期間違約，風險承受方能將債券以面額賣給承作機構以規避違約風險。而此固定費用稱為信用違約交換之利差 (CDS Spread)，若此利差越大，表示該債券的風險越高，所以承作機構須收取更多費用作為代價，反之利差越小，則表示該債券債務人之財務信用較佳。

信用違約交換出現後，學者開始採取信用違約交換之利差作為企業信用風險的代理變數，並尋找影響企業信用風險的主要影響因子。Pedrosa (1998) 觀察信用違約交換利差之變動在不同產業、品質、到期天數的債券中仍會根據長債利率、市場循環、市場波動度等總體因素有共同變化，初步指出影響企業風險的總體

DOI:10.6814/NCCU202200901

因子。信用違約交換利差的影響因子與上一節所述的違約預警模型有極大關連，Ericsson et al. (2009) 歸納傳統文獻中會影響信用違約交換利差的因子，包括無風險利率、公司股價波動度及公司槓桿。而Collin-Dufresn et al. (2001) 在控制上述因子後仍發現殘差相關，指出除傳統信用風險因子及流動性因子外，信用違約交換之利差變動量可能還有由其他重要因子所驅動。

Norden and Weber (2004) 發現股市和信用違約交換市場都能提前預期負面信用事件，表示在市場效率下信用違約交換利差的價格充分反映了未來可能發生的信用事件。Hull et al. (2004) 則研究信用事件與信用違約交換利差之間的關係，指出企業受到信評機構調升評等或是調降評等之前，市場已廣泛認知該企業的信用風險狀況已有變動，故其在信用平等發生變化時，對信用違約交換的利差在統計上並無顯著影響，然而，降級審查 (Review for downgrade) 則因市場普遍皆對該事件未有預期，對信用違約交換利差有顯著影響，與Galil and Soffer (2011) 的結果一致，另外此研究也指出信用違約交換利差及其變動量皆對負面信用事件提供有用訊息。Norden (2017) 則試圖探討新聞媒體等資訊是否揭露一間公司的信用風險，進而影響信用違約交換的利差，其使用事件研究法發現信用違約交換利差在較多媒體報導之公司的改變會更加強烈，而新聞的正負面將會影響信用違約交換利差的動向。

2.2 文字分析模型

Loughran and McDonald (2016) 將過去財經領域主要使用的文字分析方式主要使用字典模型、文本易讀性分析和主題學習模型等等，以下將分別簡述上述各類模型與本研究所使用的 BERT 模型之差異。

2.2.1 字典模型

字典模型一般依賴已分類過領域、正負面及詞性的字典剖析文字資訊，雖其模型建構方式簡易，但非相關領域字典所造成的偏誤及模型中不考慮前後文語意關係則令人詬病。Tetlock (2007) 利用 General Inquirer 字典分析《華爾街日報》內的 Abreast of the Market 專欄對於道瓊指數的影響，發現若專欄內負面字詞量上升，可以預測未來股票交易量將會增加，且股票價格將有下行之壓力。而Li et al.

DOI:10.6814/NCCU202200901

(2014) 則使用不同的字典，如Loughran and McDonald (2011) 所建構之商業字詞字典、Harvard IV-4 辭典等來建構文本的情感語意空間，然後將新聞文本映射到此空間上，將其用以預測香港恆生指數成分股上的漲跌趨勢，研究結果發現此情感分析有助於預測個股價格與股價指數趨勢，且其表現較詞袋模型更佳。此外，研究結果也發現使用不同的辭典所分析出來的情感分類也將有差異，故在文字分析上應嘗試使用不同辭典來獲得較佳的適配模型。

在財經領域上的實證研究上，因其在財務文本上的有效性，更多研究者傾向使用Loughran and McDonald (2011) 所發表的商業字詞字典 (Li et al., 2014, Mayew and Venkatachalam, 2012) 。字典模型之建構相對簡易且泛用性強，容易在此架構上繼續構築模型，如Da et al. (2015) 使用 General Inquirer 字典和 Lasswell Value 字典取出與經濟相關之正面及負面字詞，再經由 Google Trend 這些字詞的衍生詞綴之 SVI 指數建構情緒指數 FEARS (financial and economic attitudes revealed by search) 指標。結果發現，FEARS 指標能夠有效預測短期內股價的下跌壓力及波動，且能預測未來股票市場與債券市場的資金流動。

然而，在Loughran and McDonald (2011) 發表商業情感辭典之前，研究者皆使用其他與財經領域較不相關的預定義字典，而這些辭典將對財經文本的語意分析產生偏誤，如 *debt* 在其他領域將被分類為負面字詞，因其代表著負債以及財務壓力，但在財經領域文章，*debt* 被廣泛使用在描述公司的資產結構上，並非相當負面的字眼，故在對應領域上使用相對應的字典相當重要。

另外，同一文字在不同前後文下可能有不同語意，字典模型無法納入前後文的有效訊息，故Hutto and Gilbert (2014) 提出了 VADER 情感分析模型，此模型認為句子是情感構建之基礎，利用反向詞句、標點符號及大小寫等面向分析句子情緒，並實證此模型在分析社群媒體、電影評論及新聞文章等領域都較字典模型精準。Shapiro et al. (2020) 使用 VADER 結合 LM、GI 等辭典進行模型改進，發現經過改良的模型與人為分類的結果更為相似，此類改進方法亦可視為一種遷移學習之方式。

2.2.2 文本易讀性模型

相對於字典模型，測量文章的易讀性更為簡單，但也可能造成更多偏誤。過去文獻中測量文本易讀性的方法大多使用理解課本所需教育程度的計算方式，

DOI:10.6814/NCCU202200901

如Li (2008) 使用 Fog Index 來衡量公司財報的可讀性與其財報表現的關係，Fog Index 為句子長度以及超過雙音節之單字數量所建構，隱含的意義為能夠閱讀並理解該段文字所需要的教育時間長度；舉例來說，若該段文字所計算出的 Fog Index 為 17，則了解該段文字的所需教育程度為 17 年。Li (2008) 指出當一家公司獲利越少，則該公司的財報所計算出的 Fog Index 則會越高，此外，當一家公司的財報較容易閱讀時，未來將會有較持久的持續獲利。Miller (2010) 觀察到在企業財報發布前後，若該財報 Fog Index 和財報字數較高，投資者將會交易較少該標的股票，此結果表明，財報的可讀性將會影響投資者的交易意願，尤其是非專業的散戶投資者。Lawrence (2013) 也觀察到類似結果，其研究解果顯示，散戶投資者傾向於將資產配置在財報可讀性較高且字數較少的公司中。

然而，Loughran and McDonald (2014) 則指出在財務文件上使用 Fog Index 衡量文本易讀性不甚恰當，其爭論點在於，財務會計方面的文本易讀性應定義為「文本的寫作方式讓讀者能夠理解的程度」，而非「某一教育程度的人們認為一個文本本容易理解」，且使用音節來區分財務方面的艱難字彙也有失公允，如 Financial、Company、Employees 等多音節的字彙皆為投資者所熟知，將這些字詞計算入 Fog Index 將使分數產生偏誤。Loughran and McDonald (2014) 提出將直接將財報字數取自然對數作為文本易讀性的代理變數，實證結果顯示，財報字數越多的公司將會有較大的股票波動度和分析分歧，也指出此計算方式某種程度可以作為文本易讀性的代理變數，但其中也有可能是該公司的業務本就相當複雜。綜上，文本易讀性模型雖其建構相對簡易，但也因為相對簡易常有其他因素干擾計算結果，產生錯誤的分析推論。

2.2.3 主題模型

主題模型則較前述方法複雜許多，其原先的目的是解決詞袋模型過大的短語－文檔矩陣，如 Latent Semantic Analysis (LSA)，使用奇異質分解來縮減短語－文檔矩陣。而Blei et al. (2003) 提出的 Latent Dirichlet allocation (LDA) 模型為非監督式主題模型則被廣泛使用，LDA 為基於 Probabilistic LSA (pLSA) 所延伸的方法，估計方法採用貝氏統計，使用 Dirichlet 分佈作為詞語的先驗分佈，因其共軛分佈仍為 Dirichlet 分佈的特性使得模型演算得以簡化，採樣文本中的短語分佈及文本中的主題分佈，能為文章產生主題的權重。然而，簡化過的模型仍須非常大量的

DOI:10.6814/NCCU202200901

迭代計算，過慢的收斂速度為 LDA 主題模型之缺點。

作為 LDA 方法的首批應用之一，Huang et al. (2018) 使用 LDA 方法比較公司電話會議與會後大量分析師的報告之主題內容，發現分析師將會披露電話會議外的主題內容，若管理階層有更大的動機隱瞞公司與價值相關的訊息，投資者將會更看重分析師的報告。Dyer et al. (2017) 則使用 LDA 方法分析 1996-2013 的美國 10K 報表的文字資訊，發現 10K 報表的主題變化主要來自於財務會計標準委員會（Financial Accounting Standards Board, FASB）與美國證券交易委員會（United States Securities and Exchange Commission, SEC）的約束。

然而，pLSA、LDA 等主題模型雖使用大量文本進行非監督式訓練，其訓練結果仍需人為觀察其短語分佈命名主題，且若是文本本中有普遍存在的字詞未被視為斷字 (Stop Word) 去除，則短語分佈將存在雜訊。相較遷移學習模型，主題模型須研究者更頻繁的以其領域知識協助修正、調整模型參數。

2.2.4 遷移學習模型

遷移學習則是目前自然語言處理領域的發展趨勢，使用已經經過大量訓練的類神經網路模型針對目標任務進行微調，如本研究將使用的 BERT 模型就是當前類神經網路遷移學習的經典代表。Devlin et al. (2018) 提出的 BERT(Bidirectional Encoder Representations from Transformers) 模型源自 Transformer 模型中的 Encoder，為 Google 公司使用大量維基百科文本與 BooksCorpus 中的語料進行非監督訓練的預訓練模型，其訓練任務有兩個，其一是進行填空任務，其二是判斷句子是否為連接句子，透過這兩種方式，此神經網路模型也可以視為在學習字詞的向量編碼，相同語意的字詞將被投射在鄰近的向量空間中，而透過 Transformer 的自注意力機制（Self-attention mechanism），此模型在進行訓練時會考慮文字之上下文語意，使其能夠辨認同一字詞在不同背景所隱含的意思，讓自然語言模型做到指代消解的能力。透過預訓練模型所產生的向量編碼，BERT 可以輕易做到Devlin et al. (2018) 中提到的四個下游任務模型，包括單一句子分類任務、成對句子分類任務、單一句子標示任務及問答任務。

BERT 模型推出後，在各大自然語言分析任務達到 SOTA (State of the Arts)，許多改進模型應運而生。Lan et al. (2019) 提出 ALBERT (A Lite Bidirectional Encoder Representations from Transformers) 模型，針對 BERT 建構方式進行修正，包括跨

DOI:10.6814/NCCU202200901

層參數共享、修改訓練目標等，故其訓練速度較 BERT 更快。BERT 模型在做預測時雖有考慮上下文，卻未考慮被預測字詞的前後關係，而Yang et al. (2019) 所提出的 XLNet 模型結合 AutoRegressive (AR) 模型與 AutoEncoding (AE) 模型則改進了此缺點，此模型在自然語言分析任務上取得超越 BERT 的成績。Raffel et al. (2020) 所提出 T5 (Text-To-Text Transfer Transformer) 將所有自然語言任務轉換為文本至文本的任務，不同的任務只需在模型輸入的文字前加上欲達成的目標，就可以讓模型理解其任務目標，提供了所有自然語言模型一個通用框架。此外，Raffel et al. (2020) 在不同的模型建構階段皆嘗試前人所使用過的建構方法，最後選取各階段表現最佳的作為 T5 模型的建構方式。因 BERT 預訓練模型並非使用財務領域文本，Araci (2019) 將預訓練資料集增加 Financial PhraseBank、TRC2-financial 等財金相關文字資料，建構 fin-BEET 模型，實證此模型在財務領域上擁有較 BERT 更為精準的分析能力。

雖然此類模型具億級以上參數，訓練的時間與金錢成本相當巨大，但藉由遷移學習，研究者能輕易使用下游神經網路與目標任務對預訓練模型進行微調，透過遷移學習也可解決目標域訓練文本不足的問題。Transformer 納入上下文進行語義分析的特性也改進了字典模型及詞袋模型弱於辨認一字多義的情形。此外，下游任務的多元性可以匹配更多文字分析任務，具有相當大的彈性，包括文本的主題辨認以及情緒評價任務等，可用於本研究試圖辨認媒體內蘊情緒分析。綜上所述，使用類神經網路所建構之新興的自然語言處理模型與前述字典模型、文本易讀性和主題模型等具有相對優勢。

DOI:10.6814/NCCU202200901

第三章　研究方法

3.1 BERT 模型

BERT 模型之基本構造為多個 Transformer Encoders 堆疊而成，Google 最初發布時提供兩種預訓練模型，分別為 BERT-BASE 和 BERT-LARGE，其中 BERT-BASE 為具有 12 層 Transformer Encoder、每層 768 個神經元及 12 個自注意力接頭的神經網路，約有 1.1 億參數，而 BERT-LARGE 則有 24 層 Transformer Encoder、每層 1024 個神經元及 16 個自注意力接頭，約有 3.4 億參數。根據應用資料的不同，可以選擇適合的預訓練模型。BERT 預訓練模型為非監督式學習，學習文本為英文維基百科及 BooksCorpus，在預訓練階段時，模型被要求完成下列兩個任務:

(1) Masked Language Model

BERT 在預訓練階段時，會隨機把 15% 的字詞以 [mask] 遮蓋，模型則以被遮蓋的上下文來預測這些被遮蓋的文字，來訓練模型的參數。BERT 改進了過去模型只用單向字詞預測被遮蓋文字的缺點，並使用自注意力機制，使得每個字詞對於被遮蓋字詞的預測皆有影響。

(2) Next Sentence Prediction

此外，BERT 模型從訓練資料中隨機提取兩個句子，並讓模型試著預測這兩個句子是不是上下連續的，使模型能夠更了解自然語言語意，以此達成非監督式的訓練參數。

BERT 模型使用 WordPiece Tokenization 進行詞嵌入 (Token Embedding)，把原先的單字拆分成更小的詞條 (wordpieces)，並轉換為模型可讀的數字，拆分為更小詞條的用意為能夠有效處理不在字典內的字詞，避免字串無法轉換。

除此之外，BERT 模型另有 5 個特殊 tokens:

DOI:10.6814/NCCU202200901

(1) [CLS]: 分類任務時整個句子的代表 token

(2) [SEP]: 插入連續的兩個句子中，用以區隔兩個句子的 token

(3) [UNK]: 用此 token 代表未出現在 BERT 字典的字詞

(4) [PAD]:zero padding 遮罩，將長度不一的輸入序列補齊方便運算

(5) [MASK]: 未知遮罩，僅在預訓練階段會用到

除了 Token Embedding 外，BERT 模型的數據輸入還需要 Segment Embedding 和 Positional Embedding，其中 Segment Embedding 代表不同句子的位置，Positional Embedding 代表位置編碼。以輸入句子「My dog is cute,and he likes playing」為例，文字在輸入前會依據 WordPiece Tokenization 的規則進行拆分，句子中的字詞將先被轉換為小寫的詞條，而標點符號將轉換為 [SEP]，句首則加入代表句子的 [CLS]。轉換過的輸入特徵變為「[CLS]，my，dog，is，cute，[SEP]，he，likes，play，##ing 」接著，再將這些處理過的文字分別轉換為機器可讀的 Token Embeddings，也就是數字編碼; 以及區分句子的 Segment Embeddings，在此例子中，第一個 [SEP] 即其以前的字詞將被標示為 A 句，第二個 [SEP] 及其之前到第一個 [SEP] 之後的句子將被標示為 B 句。最後，Position Embeddings 再分別給予每個輸入字詞一個位置編碼，詳細編碼如圖3.1[1]。

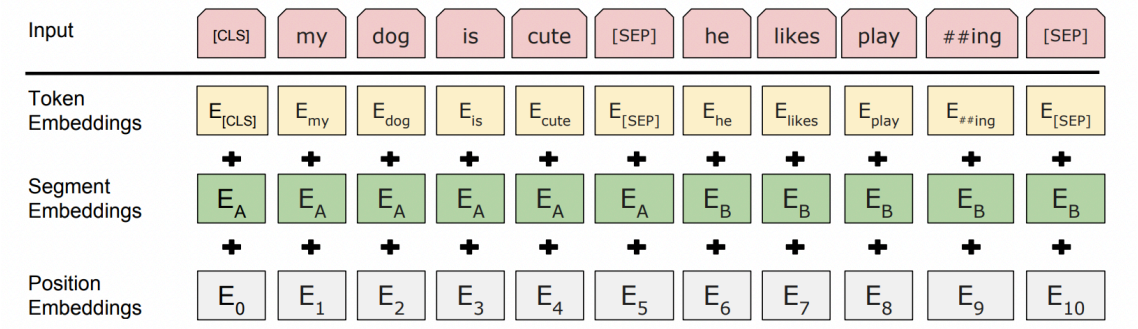

圖 3.1: BERT 編碼示意圖

BERT 使用源自於 Transformer 的自注意力機制，更擅長捕捉數據內部的相關性，自注意力方程可以簡單形容為將 q(Query) 和一組 k(Key) 和 v(Value) 映射至輸出 (Attention Value)，其中 q、k、v 和輸出皆為矩陣。

注意力方程可寫為式3.1：

[1] 資料來源：Devlin et al., 2018, p.5

DOI:10.6814/NCCU202200901

$$Attention(q,k,v) = softmax(\frac{qk^T}{\sqrt{d_k}})v \tag{3.1}$$

其中，

$$q = \beta_q X, k = \beta_k X, v = \beta_v X \tag{3.2}$$

d_k 為 k 矩陣之維度，β_q、β_k 及 β_v 皆為學習而來的映射矩陣，X 為輸入句子的每個元素向量所形成之矩陣。

Attention Value 本質上是給予序列中每個元素一個權重因子。這種通過 Q Query 和 Key 的相似性程度來確定 Value 的權重分佈的方法被稱為 Scaled Dot-Product Attention。為了讓模型在不同子空間內學習相關訊息，BERT 使用多頭注意力 (Multi-Head Attention) 同時計算句子中所有單字的注意力，然後將值連接起來，可寫為式3.3：

$$MultiHead(q,k,v) = Concat(head_1, ..., head_h)\beta^O \tag{3.3}$$

$$where\ head_i = Attention(q\beta_i^q, k\beta_i^k, v\beta_i^v)$$

其中 $\beta_i^q \in \mathbf{R}^{d_{model} \times d_k}$，$\beta_i^k \in \mathbf{R}^{d_{model} \times d_k}$，$\beta_i^v \in \mathbf{R}^{d_{model} \times d_v}$，$\beta^O \in \mathbf{R}^{hd_v \times d_{model}}$，圖3.3[2]為整體架構示意圖。

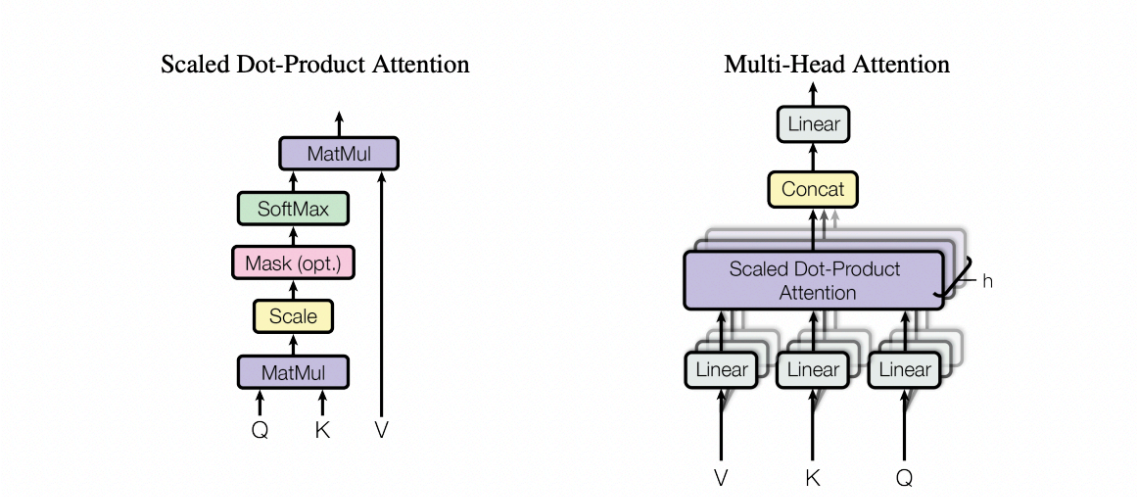

圖 3.2: 多頭注意力機制

獲取已經預訓練好的模型後，研究者可依自己的需求架設下游神經網路，使

[2]資料來源：Vaswani et al., 2017, p.4

DOI:10.6814/NCCU202200901

用具有標註的樣本進行監督式訓練。BERT 的下游任務可以分為四大類，分別為單一句子分類任務、句子內單詞的標注任務、成對句子的分類任務以及問答任務。

(1) 單一句子的分類任務

常見的單句分類任務有情緒分類任務和主題分類任務等，將單一句子輸入模型後，模型將依據句子中的用字產生分類標籤。

(2) 句子內單詞的標注任務

此處的標註任務實際上與單句分類任務相同，只是預測主體改變為句中的單詞，將單一句子輸入模型後，模型將對句子中的字詞產生分類標籤。

(3) 成對句子的分類任務

成對句子分類任務的常見應用為兩個句子是否具有一致推論，如上文的意思是否與下文一致，將兩個句子輸入模型後，模型將對此對句子產生分類標籤。

(4) 問答任務

問答任務為給予模型一個問題及一個與答案有關的文本，訓練模型能針對問題給出相應的答案，模型的運作方式為，將問題與具有的答案文章輸入模型後，模型產生對應的答案文字。

本研究嘗試以 BERT 預訓練模型為基礎，架設下游神經網路針財務新聞文本進行文本遷移，下游神經網路將以單一句子的情續分類任務為主，並期望透過遷移學習克服文本異質性所造成的預測困難。

首先，本研究將使用 The Stanford Sentiment Treebank (SST)、Financial Phrase-bank 等語料庫進行監督式訓練，這些資料具有每個文字的的正負面標籤。值得注意的是，不同語料庫對模型將有不同影響，如上述 SST 語料庫的文字資料主要來自於電影評論，而 Financial Phrasebank 資料庫內的文字主要來自於財務新聞，而本研究假設，同一字詞在不同領域的語意將有不同情緒，故本研究預期，使用財務新聞語料訓練的神經網路模型將有較佳表現。經過訓練後，模型將產生對輸入的句子產生分類輸出值，也就是情緒分數。本處神經網路一樣將使用交叉驗證法進行超參數配置，啟動函數將使用整流線性單位函式。本處神經網路使用交叉驗證法 (Cross Validation) 進行超參數配置，而啟動函數 (Activation Function) 將使用整流線性單位函式 (Rectified Linear Unit，ReLU)，此函數相較於邏輯函式 (Logistic Sigmoid) 和 tanh 等雙曲函式能更有效率的進行梯度下降，解決梯度消失

DOI:10.6814/NCCU202200901

及梯度爆炸的問題，此外，其運算效率也較指數函數高，使神經網路的計算成本下降。輸出值將再根據模型運算的閾值標準給出分類。

$$ReLU(x) = max(0, x) \tag{3.4}$$

隨後，本研究將根據此生成的特徵，為每家公司在特定時間點生成基於不同主題的情緒特徵，並用以建構違約預警模型。

3.2 隨機森林

隨機森林是基於決策樹發展而來的類集成機器學習算法，透過 Bootstrap 算法產生多個決策樹分類結果後，最終產生一個強分類器，故會有較單個決策樹模型有更好的迴歸及分類表現。

本研究所使用之決策樹演算法為 CART(Classification And Regression Tree) 演算法，其演算基礎為 CLS(Concept Learning System)，其中決策樹設計有三種節點：根節點、葉結點和內部節點。空決策樹任意選擇的第一個節點為根節點；若我們按照某種條件進行劃分，劃分到某個子集為空或該子集內所有樣本皆屬於同一類別，則該節點為葉節點，否則這些子集就對應於決策樹內部節點，需要繼續進行劃分，直到所有的子集皆為葉節點，即為空或屬於同一種類別。

本研究所建立之決策樹以吉尼不純度 (Gini Impurity) 為基礎，吉尼不純度介於 0 至 1 之間，0 完全相等，1 完全不相等，可用來度量類別分佈不均勻之程度，當子集內包含的類別越雜亂，吉尼不純度越高。針對有 C 種類別的資料集 S，吉尼不純度之定義為：

$$Gini(S) = 1 - \sum_{i \in C} p_i^2, 0 \leq p_i \leq 1 \tag{3.5}$$

其中 p_i 為資料集 S 中屬於類別 i 之機率，若將資料集 S 以特徵 A 進行分割，則此分割條件下，分割後之吉尼不純度定義為：

$$Gini(S)_A = \sum_{j \in S'} \frac{|S_j|}{|S|} Gini(S_j) \tag{3.6}$$

其中 S_j 為分割後的第 j 個子集，因此式3.6也可看作所有子集的吉尼不純度

DOI:10.6814/NCCU202200901

加權平均，此時不純度降低量可表示為：

$$\Delta Gini(S)_A = Gini(S) - Gini(S)_A \tag{3.7}$$

決策樹會選擇降低最大不純度之特徵作為分割屬性。

隨機森林使用 Bootstrap 方法建構，決策樹皆以下列步驟形成：

1. 隨機抽取 n 組樣本作為訓練集，樣本可被重複抽取

2. 每組樣本隨機抽取 m 個特徵作為分割方式訓練決策樹

如圖3.3所示，最後模型會整合每個若分類器的輸出結果，若為分類資料，則採簡單多數投票法；若為迴歸資料，則採簡單平均法。

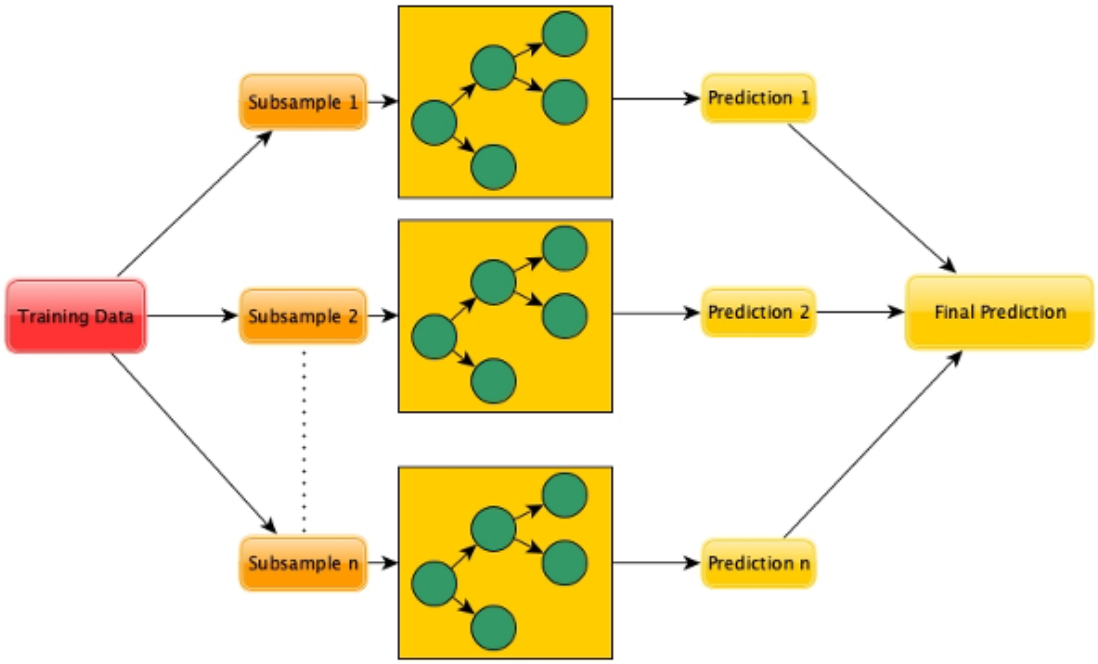

圖 3.3: 隨機森林架構

因本研究所生成之主題情緒特徵較多，若使用全部特徵作為模型輸入，將降低模型訓練效能，以及產生過擬合問題，故本研究將對特徵進行篩選，降低特徵維度。

特徵篩選可分為三種，分別為過濾式方法 (Filter Method)、封裝式方法 (Wrapper Method) 和嵌入式方法 (Embedded Method)。過濾式方法為對特徵設定一個門檻，如相關係數、資訊增益等，刪除達不到門檻的變數來達到減少變數的目的，此方法之優點為速度快，不需要大量的計算，而缺點是有可能刪除某些重要

DOI:10.6814/NCCU202200901

特徵。封裝式方法則嘗試使用不同的特徵子集訓練模型，最後找出擁有最佳績效的子集作為模型特徵，雖然此方法可能將擁有比過濾式方法有更佳的模型績效，但其計算量較為龐大。嵌入式方法則是在模型訓練時，同時篩選重要的特徵，相關的模型有 LASSO(Least Absolute Shrinkage and Selection Operator)、脊回歸 (Ridge Regression) 等，隨機森林模型也是嵌入式方法的例子之一，因其模型特性，隨機森林可以計算每一個特徵的特徵重要度，進而篩選出較重要的特徵。

本研究採用嵌入式特徵篩選法，將特徵為模型所降低之吉尼不純度作為其重要度，假設隨機森林模型總共有 T 棵樹、F 個特徵，而其中樹 t 有 Q_t 個節點，則第 f 個特徵在樹 t 的第 q 個節點之特徵重要度 $FI^t_{f,q}$ 計算如下：

$$FI^f_{t,q} = \Delta Gini(S)_f \tag{3.8}$$

若特徵 f 在樹 t 中出現的次數為 $Q_{t,f}$ 次，則該特徵在樹 t 之特徵重要度為 FI^t_f 為：

$$FI^f_t = \sum_{q \in Q_{t,f}} FI^f_{t,q} \tag{3.9}$$

因此特徵 f 在整個模型之重要度為：

$$FI^f = \sum_{t \in T} FI^f_t \tag{3.10}$$

最後再進行標準化處理：

$$FI^f = \frac{FI^f}{\sum_{f' \in F} FI^{f'}} \tag{3.11}$$

得到每個特徵的重要度之後，我們便可以設定一個門檻值來捨去不重要的特徵，本研究門檻值設定為所有特徵之重要度之平均。

然而，當某特徵之重要度很高，我們仍然無法得知該特徵對於模型分類的正負向影響為何，故本研究使用Lundberg and Lee (2017) 提出之 SHAP(Shapley Additive exPlanations) 解釋在模型中每個變數對於模型預測值的貢獻。其概念為延伸合作博弈理論中 Shapley Value 的概念，透過排列所有可能的變數組合，求出每個變數對於模型預測之貢獻，其公式如式:

DOI:10.6814/NCCU202200901

$$\phi_j(val) = \sum_{S \subseteq \{x_1,\ldots,x_F\}/\{x_j\}} \frac{|S|!(p-|S|-1)!}{p!}(val(S \cup \{x_j\}) - val(S)) \tag{3.12}$$

其中，$\phi_j(val)$ 為 x_j 之 SHAP 值，$x_1, \ldots, x_F$ 為資料中的 F 個變數，S 為所有變數排除 x_j 之集合，$val(S)$ 為集合 S 之模型預測值減去模型期望值之函數，其公式如下:

$$val(S) = \int \hat{f}(x_1, \ldots, x_F) d\mathbb{P}_{x \notin S} - E(\hat{f}(X)) \tag{3.13}$$

其中，$\hat{f}(\cdot)$ 為研究使用之模型。透過計算每個樣本之 SHAP 值，不僅可以得知各個變數對於模型預測之貢獻，也可以知道該變數對於模型預測之方向 (亦即正向影響或負向影響)。

3.3 模型績效衡量指標

3.3.1 混淆矩陣

本研究使用混淆矩陣作為模型衡量指標之一，混淆矩陣中，橫列代表實際樣本類別，直欄則為預測類別，本研究採用三元分類之混淆矩陣，如表3.1所示。以預測 A 類別為例，當真實不為 A 類別，模型預測卻為 A 類別的樣本，以偽陽性 (False Positive) 表示，也就是型一錯誤 (Type I Error)；而當真實為 A 類，模型預測卻非 A 類別時，則以偽陰性 (False Negative) 表示，型二錯誤 (Type II Error)，兩者皆為未被正確預測的類別，故為一個相當簡易的模型衡量方式。本研究模型在情緒分類所分類之類別則分為正面情緒、中立情緒及負面情緒，而在信用評級變度預測模型則分為上調評級、評級不變及下調評級。

表 3.1: 三元分類混淆矩陣

	預測 A 類	預測 B 類	預測 C 類
實際 A 類	真陽性 (True Positive)	偽陰性 (False Negative, B)	偽陰性 (False Negative, C)
實際 B 類	偽陽性 (False Positive, B)		
實際 C 類	偽陽性 (False Positive, C)		

DOI:10.6814/NCCU202200901

3.3.2 模型準確率

準確率亦為相當常見之衡量指標，代表模型預測正確之樣本比例，其可由混淆矩陣直接求得：

$$Accuracy = \frac{\sum_{c \in C} TP_c}{n} \tag{3.14}$$

其中 C 為全部類別之集合，TP_c 為預測 c 類別中，預測結果為真陽性之樣本數，n 為樣本總數。

然而，在資料不平衡的狀況下，準確率未必能衡量機器學習模型的好壞。例如當陽性樣本佔總體 99% 以上時，若分類器將所有樣本皆預測為陽性，此分類器之分類結果完全無意義，但卻有相當高的預測準確率，故使用使指標作為衡量標準時必須搭配其他指標，以確保預測品質。

3.3.3 精確度、召回率及 F1-Score

精確度指標又稱陽性預測值 (Positive Predictive Value，PPV)，代表預測為為陽性之樣本中多少比例為真實陽性，類別 c 的精確度可寫為：

$$Precision_c = \frac{TP_c}{TP_c + FP_c} \tag{3.15}$$

其中 FP_c 為預測 c 類別中，預測結果為偽陽性的樣本數。

根據 Micro-Averaging 和 Macro-Averging 兩種方式，整個模型的精確度可寫為：

$$Precision_{Micro} = \frac{\sum_{c \in C} TP_c}{\sum_{c \in C} TP_c + \sum_{c \in C} FP_c}$$

$$Precision_{Macro} = \frac{\sum_{c \in C} Precision_c}{|C|} \tag{3.16}$$

而召回率指標又稱真陽性率 (True Positive Rate) 或敏感度 (Sensitivity)，代表所有真實為陽性的樣本多少比例被正確預測，故類別 c 的召回率可寫為：

DOI:10.6814/NCCU202200901

$$Recall_c = \frac{TP + c}{TP_c + FN_c} \tag{3.17}$$

其中 FN_c 為預測 c 類別中，預測結果為偽陰性的樣本數。

根據 Micro-Averaging 和 Macro-Averging 兩種方式，整個模型的精確度可寫為：

$$Recall_{Micro} = \frac{\sum_{c \in C} TP_c}{\sum_{c \in C} TP_c + \sum_{c \in C} FN_c}$$

$$Recall_{Macro} = \frac{\sum_{c \in C} Recallc}{|C|} \tag{3.18}$$

在信用預警的模型當中，模型績效通常較重視召回率之表現，例如銀行放貸時，若信用模型之召回率較低，則表示將有大量違約的客戶被模型誤分類為不會違約，則放貸方則會蒙受損失，相反的，若模型之精確度較低，則表示大量不會違約的用戶被模型標註為有違約風險，銀行不選擇放貸，則放貸方則是損失了獲利來源，故相比之下，召回率較低的損失將比精確度較低的損失來得大。

召回率與精確度之值應越高越好，而當模型判斷之臨界機率變化時，型一誤差與型二誤差成反向變動，兩者不能同時降低，表示精確度與召回率無法同時改善，故其為取捨問題，F1-Score 為兩者之綜合評量標準。F1-Score 為精確度 (Precision) 和召回率 (Recall) 之調和平均，用來衡量預測能力之綜合指標。

$$F1 - Score_c = 2 \times \frac{1}{\frac{1}{Precision_c} + \frac{1}{Recall_c}} = 2 \times \frac{Precion_c \times Recall_c}{Precion_c + Recall_c} \tag{3.19}$$

最後，本研究使用 Macro-Averaging F1-Score 來衡量模型績效：

DOI:10.6814/NCCU202200901

$$Macro - Averaging\ F1 - Score$$

$$= 2 \times \frac{1}{\frac{1}{Precision_{Macro}} + \frac{1}{Recall_{Macro}}} \tag{3.20}$$

$$= 2 \times \frac{Precion_{Macro} \times Recall_{Macro}}{Precion_{Macro} + Recall_{Macro}}$$

3.3.4 ROC 曲線

ROC 曲線 (Receiver Operating Characteristic Curve) 為呈現召回率與 1-特異度 (Specificity) 之圖形，其中橫軸為 1-特異度，縱軸為召回率，詳見圖3.4。特異度指標又稱真陰性率 (True Negative Rate)，如式3.21所示，在二元分類中本代表所有真實為陰性的樣本中有多少比例被預測正確，但因本研究皆為三元分類模型，以預測 A 類別為例，此處之真陰性樣本則代表所有真實不為 A 類別的樣本中，有多少比例被分類為 B、C 類，故可將表3.1改寫為表3.2，如此以來，我們便能以計算二元分類之分類特異度的方式來計算三元分類中每個分類的分類特異度，並以此繪製 ROC 曲線。而整體模型的 ROC 曲線，一樣可以使用 Micro-Averaging 和 Macro-Averaging 的方式繪製，整體模型之特異度可以以式3.22計算，再搭配式3.18所計算之整體模型的召回率，最終繪製出 Micro-Averaging 和 Macro-Averaging 之 ROC 曲線。

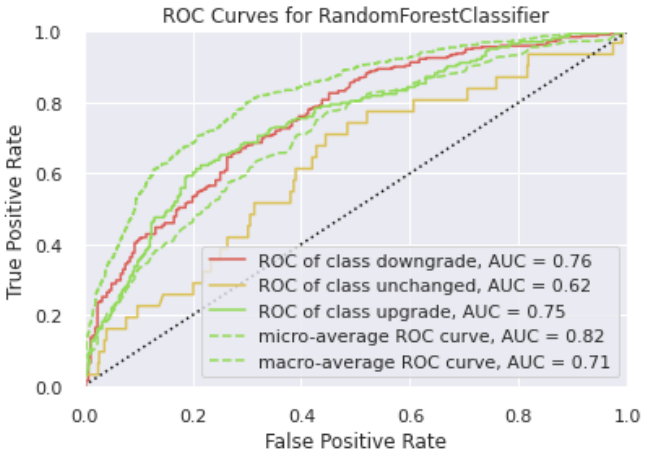

圖 3.4: ROC 曲線示意圖

DOI:10.6814/NCCU202200901

$$Specificity_c = \frac{TN_c}{FP_c + TN_c} \tag{3.21}$$

其中 TN_c 為預測 c 類別中，預測結果為真陰性的樣本數。

$$Specificity_{Micro} = \frac{\sum_{c \in C} TN_c}{\sum_{c \in C} FP_c + \sum_{c \in C} TN_c}$$

$$Specificity_{Macro} = \frac{\sum_{c \in C} Specificity_c}{|C|} \tag{3.22}$$

而 1-特異度之值代表真實為陰性的樣本被預測為陽性樣本之比例，即偽陽性率，此值應越低越好。若欲召回率增加，應減少型一誤差，此舉造成型二誤差增加，而型二誤差增加卻導致偽陽性率增加，故兩項指標亦為取捨問題，不能同時改善。故 ROC 曲線提供我們一種方式檢驗模型的分類門檻值對型一錯誤與型二錯誤的影響，並讓研究者可以簡單觀察模型的適配程度。

表 3.2: 三元分類真陰性樣本示意

	預測 A 類	預測 B 類	預測 C 類
實際 A 類	真陽性 (True Positive)	偽陰性 (False Negative)	
實際 B 類	偽陽性 (False Positive)	真陰性 (True Negative)	
實際 C 類			

從 ROC 曲線我們並不能很快地看出各分類模型的好壞，因此需採用其曲線下的面積值 (Area Under Curve, AUC) 來進行分類模型的評估與比較，其定義為 ROC 曲線與橫軸之間的面積，一般範圍介於 0.5 至 1 之間，此值越大代表模型越靠近左上方，代表模型的區分能力與配適度越好。

DOI:10.6814/NCCU202200901

第四章　實證分析

4.1 資料處理

4.1.1 財務變數資料

為對比媒體情緒是否增進傳統財務變數模型的分類預測效果，故本研究也將納入財務變數比較，財務變數資料參考Lu et al. (2012) 建構，公司財務資料來源使用 Compustat 資料庫，違約距離變數使用 CRI 資料庫，資料期間為 2000 年至 2017 年，Lu et al. (2012) 將能夠捕捉公司信用風險的財務變數分為五大類，包含公司規模、槓桿程度、獲利能力、清償能力及流動性，並使用此五大類變數作為模型輸入Lu et al. (2012) 亦使用違約距離作為輸入變數，違約距離是依據結構模型，以股票市值、股票報酬率波動度與負債價值所計算出公司資產價值與資產報酬標準差，進而推估公司資產市值與違約距離。表4.1為本研究所使用之財務變數，共使用 10 個財務比率及 1 個違約距離變數。

表 4.1: 財務資料變數

變數名稱	說明
Total_Asset	總資產
BM_Ratio	帳市比
LTD/Total_Invested_Capital	長期債務與投入資本總額比率
Debt/Equity	債務與股本比率
Operating_Income	營業收入
Net_Income_Before_Tax	稅前淨利
Gross_Profit_Margin	毛利率
Earnings_Per_Share	每股盈餘
EBIT_Interest_Coverage	利息保障倍數
Quick_Ratio	速動比率，速動資產與流動負債比率
DTD	違約距離

DOI:10.6814/NCCU202200901

4.1.2 新聞文本資料

本研究使用 Raven Pack 資料庫中 2000 年至 2017 年之美國新聞資料，Raven-Pack 資料庫是成立於 2003 年的新聞分析數據提供者，不同於以往的金融數據提供商提供關於金融市場交易數據或資產結構等量化數據，RavenPack 主要提供來自主流新聞媒體以及社交媒體等文字的質化訊息，其資料的提供來源分別為 Premium Sources 和 Web Sources，包含超過 22000 個資料來源。其中，Premium Sources 的資料為 Dow Jones、Wall Street Journal、Barrons、MT Newswires、Ben-zinga 等專業金融媒體所提供，Web Sources 的資料則取自區域性或地方報紙以及知名金融資訊網站。RavenPack 提供的的資料除了文本本身、文本發布時間等一手資料數據以外，也提供了該文本主題、該文本所提及的資產實體、文本與其資產實體的相關性和情緒分數等分析資料。RavenPack 提供了大量有精確時間及主題標註的文本，除此之外，文本內容所提及用於語意訓練的實體標註更是不可或缺，RavenPack 的 NER (Named Entity Recognition) 演算法能夠辨別文章中資產實體為何，對於本研究辨認不同公司新聞的情緒計算為一大助益。Raven Pack 資料集中共含 52 個變數，本研究建構樣本資料之情緒只採用其中 11 個變數，表4.2為其詳細內容。

因資料庫完善，資料處理相對容易。本研究所研究之新聞內容將以公司為主體，故以 *ENTITY_TYPE* 之分類刪除新聞描述主體非公司之新聞，接著再將 *HEADLINE* 和 *EVENT_TEXT* 合併為 *CONTENT* 作為後續情緒分析之用，最後共有近 1900 萬筆新聞資料。

4.1.3 評級變動資料

本研究以 Raven Pack 資料庫內評級機構的新聞發布作為評級變動之標註來源，首先以 *TYPE* 之分類留下所有屬於"credit-rating-change" 之新聞，此類別之新聞只包括信評機構對企業的評級變動，信評機構包括標準普爾、穆迪及惠譽，再以 *EVENT_SIMILARITY_KEY* 辨認並刪除重複報導，只留下首篇報導，再依據資料庫中的 *SUB_TYPE* 變數進行上升、不變、下調的分類，最後產生近 35000 多筆信評變動資料，包含 3588 家公司，詳細如表4.3所示。然而，因信評機構未必會對信評不變的公司發布新聞稿，故本研究在採樣樣本時，若該樣本未被信評機構

DOI:10.6814/NCCU202200901

表 4.2: Raven Pack 資料變數

變數名稱	說明
TIMESTAMP_UTC	時間戳記
RP_ENTITY_ID	Raven Pack 資料庫對新聞主體之標示，用於辨認新聞所描述的對象
ENTITY_TYPE	主體類別，用於辨認新聞描述對象為企業或是國家
ENTITY_NAME	主體名稱
EVENT_SIMILARITY_KEY	相似金鑰，用以辨認及排除相同事件的報導
TOPIC	新聞主題，共有商業、社會、政治、環境、經濟五大主題
GROUP	新聞群組，屬於新聞主題的子類別，共有 53 個分類
TYPE	新聞類別，屬於新聞群組的子類別，共有 359 個分類
SUB_TYPE	新聞副類別，屬於新聞類別的子類別，共有 119 個分類
EVENT_TEXT	新聞內文
HEADLINE	新聞標題

標註其為上調或下調評級，本研究則將其標註為評級不變，故實際評級不變的樣本數量將較表4.3多。

4.2 特徵生成

4.2.1 訓練 BERT 模型

本研究使用 SST5(The Stanford Sentiment Treebank - 5) 資料集訓練未經遷移的 BERT 模型，遷移 BERT 模型則使用 Financial Phrasebank 資料集。SST5 資料集是由史丹佛大學發布的情感資料集，由三名評委對 10754 個單句標註情感，情感分類為分為負面、有些負面、中立、有些正面及正面五類，因其文字之語意背景貼近一般文本的使用環境，本研究使用其作為未經遷移的 BERT 模型之訓練文本。Financial Phrasebank 資料集由 4846 個金融新聞語句組成，由 5 至 8 名評委對其情感進行分類，情緒類別有負面、中立及正面三類，本研究採用其作為遷移 BERT 模型之訓練文本以及測試文本，期望模型的情感認知能從一般文字背景遷移至財金相關文本，將此資料集作為測試文本也能比較兩個模型在財金文本上的情感分類差異。因兩資料集的情感分類有些許不同，本研究將 SST5 資料集中「負面、

DOI:10.6814/NCCU202200901

表 4.3: 每年度信評變動數量

年/信評變動	降評	不變	升評	總和
2000	255	25	31	311
2001	806	161	132	1099
2002	895	139	145	1179
2003	737	335	266	1338
2004	749	1228	557	2534
2005	542	444	375	1361
2006	663	874	526	2063
2007	817	733	574	2124
2008	1242	529	377	2148
2009	2047	750	512	3309
2010	763	919	1020	2702
2011	1184	1697	1223	4104
2012	1069	1132	816	3017
2013	578	445	540	1563
2014	434	379	482	1295
2015	658	691	461	1810
2016	772	436	401	1609
2017	726	638	394	1758
合計	14937	11555	8832	35324

表 4.4: 訓練及測試文本資料分布

情緒/資料集	SST5 訓練集	Financial Phrasebank 訓練集	Financial Phrasebank 測試集
負面	4222	427	885
中立	2013	1994	392
正面	4519	971	177
總和	10754	3392	1454

有些負面」的句子類別改為「負面」、「正面、有些正面」的句子類別改為「正面」，方便訓練後的 BERT 模型進行比較。

表4.4為訓練及測試集樣本分佈，未經遷移之 BERT 模型使用 SST5 所有的單句 (10754 句) 進行訓練，而遷移之 BERT 型使用其 70% 之句子 (3392 句) 進行訓練，剩下的 30%(1454 句) 則作為測試集供兩模型驗證及比較。

未經遷移 BERT 模型以及遷移 BERT 模型皆使用 Google 提供之預訓練 BERT 模型為基底，在額外新增一層丟失率為 0.1 的丟棄層及簡單線性分類器，丟棄層的用途為防止過擬合，簡單線性分類層則可把模型隱藏層之輸出結果以線性分類，故 BERT 模型之訓練主要訓練最後的分類器，只對預訓練模型基底進行微調。兩模型使用 Adam(Adaptive Moment Estimation) 演算法進行參數更新，批量大小

DOI:10.6814/NCCU202200901

(batch size) 皆為 128，而訓練輪數 (epoch) 部分以模型訓練集準確率至 90% 為止，未經遷移之 BERT 模型為 25 次，遷移之 BERT 模型為 10 次。

模型訓練完後皆以 Financial Phrasebank 之測試集進行驗證，圖4.1及圖4.2為其混淆矩陣，經過遷移之 BERT 模型辨認情緒之準確率達到 83.43%，而未經遷移之模型準確率只有 52.96%，顯示經過遷移之 BERT 模型的確較能分辨財金新聞的情緒。通過混淆矩陣可以發現未經遷移之模型在分辨中立語句發生錯誤的比率較高。

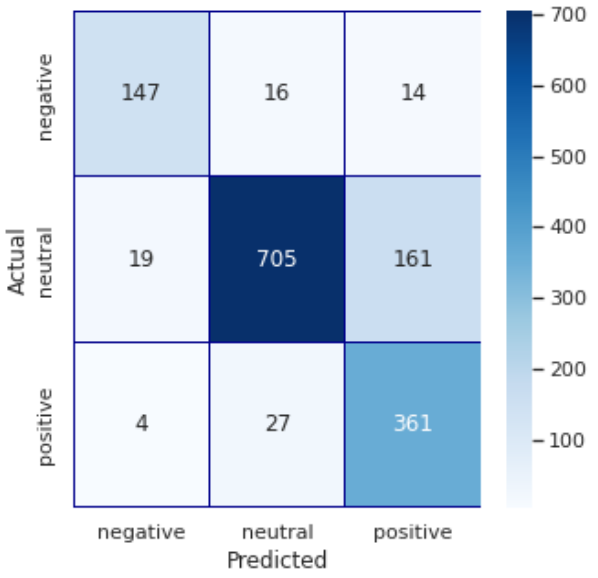

圖 4.1: 遷移之 BERT 模型情緒分類混淆矩陣

4.2.2 生成情緒特徵

模型訓練完成後，兩個模型分別對 Raven Pack 資料庫內近 1900 萬筆新聞文本進行情緒分析，將文章情緒分成正面、中立及負面三類，情緒分析結果如圖4.3與圖4.4所示，縱軸為該情緒之文章數量，已遷移之 BERT 模型分類為正面情緒的新聞數量明顯較未遷移之模型多，而未經遷移之模型則有較多分類為負面情緒的文章數量，顯示兩模型的語意環境確實具有差異。

因本研究試圖探討媒體情緒對未來企業信用評等變化的影響，故本研究之情緒特徵採用樣本公司某一時點前半年內各種主題類別下，不同情緒的文章數量之比例來建構樣本特徵，並以此特徵來預測未來該樣本公司的信用評等變化。舉

DOI:10.6814/NCCU202200901

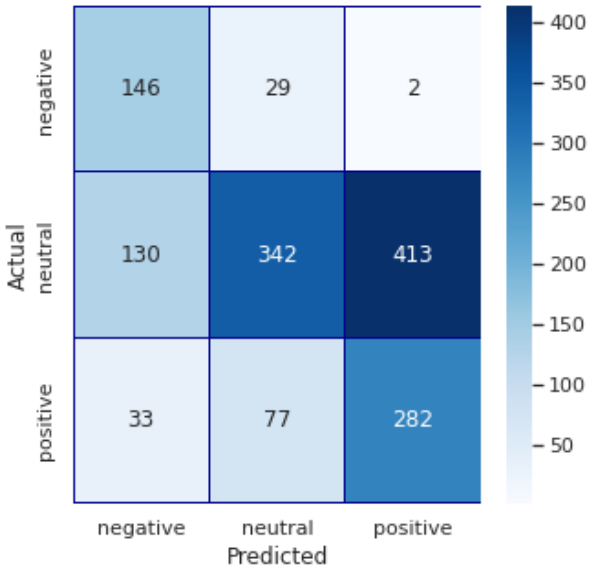

圖 4.2: 未經遷移之 BERT 模型情緒分類混淆矩陣

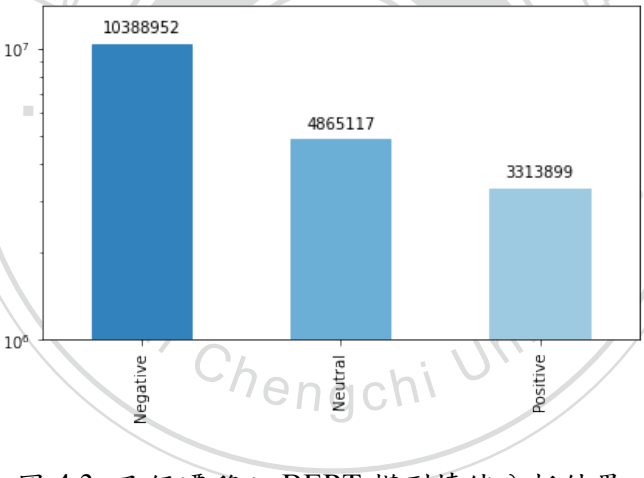

圖 4.3: 已經遷移之 BERT 模型情緒分析結果

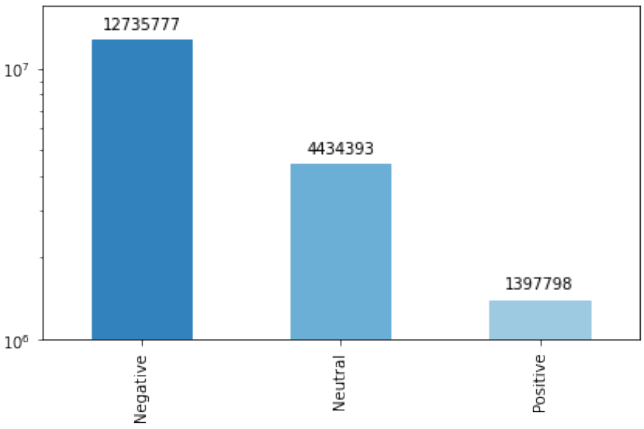

圖 4.4: 未經遷移之 BERT 模型情緒分析結果

DOI:10.6814/NCCU202200901

例來說，A 公司在 2010/10/1 前 180 天的的商業 (Business) 文章中，若有 20 篇為正面情緒文章，且該公司在該時點過去 180 天內的相關新聞總數為 100，則該公司在 2010/10/1 的特徵之一 $sentiment_{business,positive}$ 為 0.2，因此每個樣本點將有相當龐大的特徵數量，用以記錄每家公司在該段時間內所有不同主題下的情緒文章數量。本研究採用的主題變數如表4.2所述，使用 Raven Pack 資料庫中 *TOPIC*、*GROUP*、*TYPE* 內所有的主題類別，不採用 *SUB_TYPE* 中的分類原因除了避免特徵數量過多外，*SUB_TYPE* 中的分類多為形容詞，如正面及負面抑或上升或下降等，以此作為特徵會產生結果難以解讀的問題。另外，Norden (2017) 指出新聞數量對於衡量企業信用風險有顯著影響，故本研究也放入新聞總數作為媒體情緒特徵之一。最後，本研究產生 1169 個媒體主題情緒特徵用以形成企業信用評等預警模型。

4.3 建構信用評等預警模型

4.3.1 訓練集與測試集

本研究以前述變數作為樣本特徵建構隨機森林信用評等變動之預警模型，共建構四種信用評等預警模型，每個模型納入不同風險因子，用以比較 BERT 萃取出之媒體情緒是否增進模型預警能力，並且比較有無遷移對於風險預警之影響。第一個模型納入未遷移過的媒體情緒特徵作為模型輸入，第二個模型則使用已經過遷移的媒體情緒特徵作為模型輸入，兩個模型用以比較遷移與否對信用評等預警的影響。第三個模型納入已過遷移的媒體情緒特徵與財務比率變數作為模型輸入，第四個模型則僅納入財務比率變數作為模型輸入，兩個模型用以比較媒體情緒特徵是否在傳統財務比率外增加企業信用評等預警的能力。

本研究採用樣本時間後 90 天內的評級變動作為預測變數，各模型訓練區間為 2000/1/1 至 2012/12/31，預測區間為 2013/1/1 至 2017/12/31，預測類別則分為上調評級、下調評級與評級不變。然而，因為一般情況下，評級不變之樣本數會遠多於上調與下調評級的樣本數，以不平衡類別之資料訓練模型會產生模型過度預測多數類樣本之問題，故本研究採用欠採樣方法平衡三個類別的樣本數量，使三個類別的數量相同。訓練與預測樣本數以 9:1 進行建構，故 2000/1/1 至 2012/12/31

DOI:10.6814/NCCU202200901

表 4.5: 各模型特徵篩選前後之特徵數量

模型	模型納入特徵	特徵篩選前	特徵篩選後
模型一	未遷移之媒體情緒特徵	1169	247
模型二	已遷移之媒體情緒特徵	1169	245
模型三	已遷移之媒體情緒特徵與財務比率	1180	251
模型四	財務比率	11	11

的訓練樣本共 4500 筆，其中評級上調、評級不變與評級下降各 1500 筆，2013/1/1 至 2017/12/31 之預測樣本共 498 筆，評級上調、評級不變與評級下降各 166 筆。

因本研究特徵數量多達 1180 個，包含 1169 個媒體情緒特徵與 11 個財務變數特徵。而媒體情緒特徵中，常常含有大量空值，因爲不一定每間樣本公司皆有在某個主題下的情緒文章，為避免特徵數量過多影響模型效率以及過擬合問題，本研究在訓練隨機森林模型前皆使用嵌入式特徵篩選法篩選重要特徵，嵌入法之特徵篩選器亦使用隨機森林，而門檻值設定為特徵重要度的平均值，換句話說，在訓練隨機森林之過程中，若該特徵之特徵重要度小於所有特徵之特徵重要度的均值，則將該變數排除，因此，我們可以將較不具影響力的特徵刪除，減少特徵維度。而僅含有財務比率變數的模型則因特徵數量本就較為稀少，故不使用特徵篩選。各模型在特徵篩選前後的特徵個數如表4.5所示。值得注意的是，第三個模型在進行特徵篩選後，財務比率變數皆沒有被淘汰，顯示本研究所使用之財務比率變數在信用評等預警上有一定的重要程度。

4.3.2 模型訓練與超參數設置

本研究使用隨機森林建構企業信用評級變動預警模型，每棵樹的分類標準為降低最大的吉尼不純度，其中超參數包含 *n_estimators*(隨機森林模型之決策樹數量)、*max_depth*(每棵樹之最大深度) 及 *max_features*(每棵樹之最大採用特徵數)，除了只含有 11 個財務變數的模型四之外，其他各模型之 *max_features* 使用袋外錯誤率最低的配置進行設置，其結果如圖4.5、圖4.6及圖4.7所示，其中 *log2* 代表將總特徵數以 2 為底取對數作為每棵樹之最大採用特徵數，而 *sqrt* 代表將總特徵數開根號作為每棵樹之最大採用特徵數，*None* 則代表將總特徵數作為每棵樹之最大採用特徵數。

根據圖表的實驗結果，各模型 *max_features* 皆以使用總特徵數的開平方後的數量作為模型超參數有較低的袋外錯誤率，然而，使用 *sqrt* 和 *log2* 作為模型超參

DOI:10.6814/NCCU202200901

數也並未對模型有太嚴重的影響，袋外錯誤率雖然較高但並沒有上升太多，因此本研究模型一至模型三的 *max_features* 皆使用 *sqrt* 作為超參數。而各模型使用決策樹數量分別設定在 2000 棵樹、1600 棵和 2800 棵時有較低的袋外錯誤率，使用不同數量的決策樹對於模型袋外錯誤率也一樣無太嚴重的影響，故超參數之選擇與 *max_depth*(模型深度) 一起由網格調參法決定。模型四因只含有財務變數 11 個，每棵樹所使用的最大特徵數不作限制，而模型四使用的決策樹數量則也以下階段之網格調參決定。

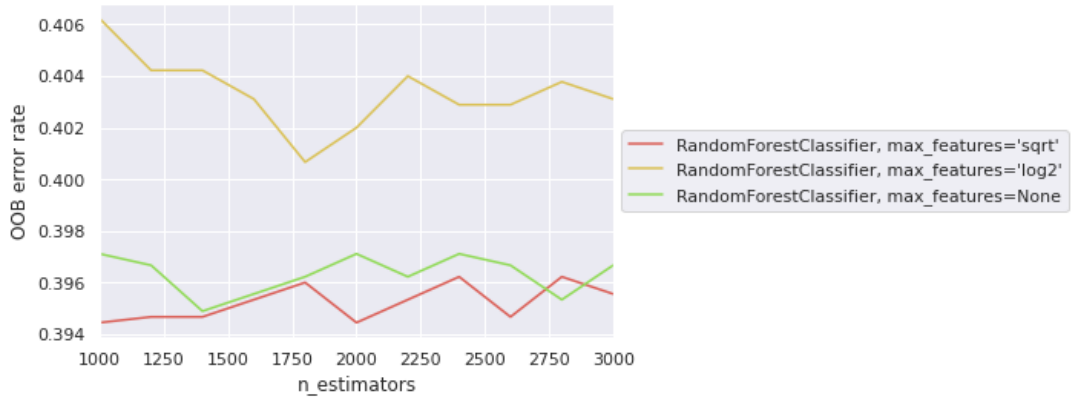

圖 4.5: 模型一之袋外錯誤率

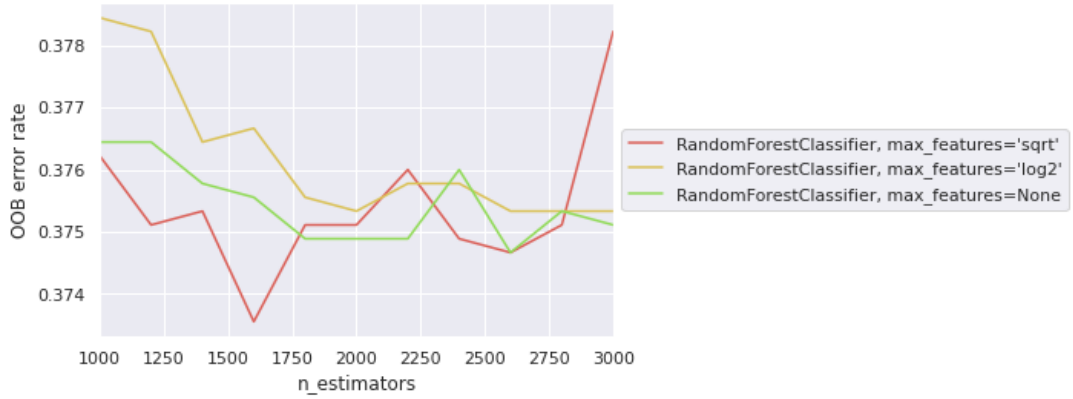

圖 4.6: 模型二之袋外錯誤率

每個模型的超參數 *n_estimators* 和 *max_depth*(每棵樹之最大深度) 由網格調參法決定，網格調參法的原理為窮舉各個超參數之組合放入模型中，而模型之結果以特定評價方法測量，擁有最高評價的組合則設定為模型最佳超參數。本研究網格調參法之評價方法使用加權召回率 (recall weighted)，各參數組合的模型有擁有最高加權召回率的則設定為模型超參數，而使用加權召回率的原因為信用評價模

DOI:10.6814/NCCU202200901

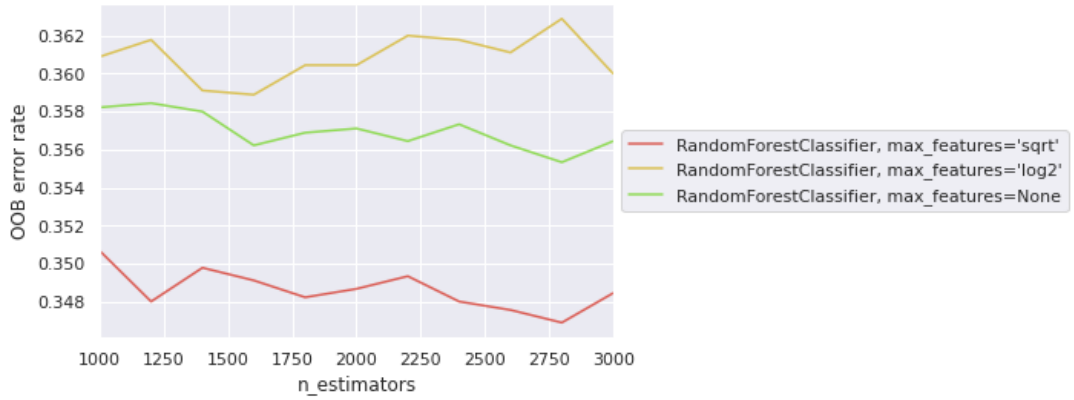

圖 4.7: 模型三之袋外錯誤率

型一般希望能夠捕捉所有潛在信用變動對象。各模型之網調參法設置與最終使用超參數如表所4.6示。

表 4.6: 模型超參數設置

	n_estimators	*max_depth*	最佳配置
模型一	1700, 1800, ..., 2300	3, 4, 5, 6, 7	*n_estimators* = 1700, *max_depth* = 6
模型二	1300, 1400, ..., 1900	3, 4, 5, 6, 7	*n_estimators* = 1800, *max_depth* = 6
模型三	2500, 2600, ..., 3100	3, 4, 5, 6, 7	*n_estimators* = 2800, *max_depth* = 6
模型四	200, 300, ..., 3000	3, 4, 5, 6, 7	*n_estimators* = 2400, *max_depth* = 6

4.4 各模型預警成效

4.4.1 混淆矩陣及相關衡量指標

下列各圖表為各模型之混淆矩陣之績效，每個模型皆以該樣本過去 180 天所形成之樣本的特徵預測該樣本時間點後 90 天內評級變動之情況。為比較 BERT 文字分析模型在經過文字領域遷移後，是否改善新聞情緒分析結果，進而提高情緒變數對企業評級變動的預警能力，本研究以模型一與模型二之預警績效作為實驗比較基準。

圖4.8與圖4.9為模型一與模型二之分類混淆矩陣，可從混淆矩陣上綜觀兩模型分類效果，兩個模型都過度將樣本分類為未來信用評等將上調，導致實際降評卻被錯誤分類為升評的樣本過多，模型一的預測準確率為 49.60%，而模型二為 52.00%。

表4.7與表4.8為模型一與模型二之分類績效，在分類召回率的部分，模型二上

DOI:10.6814/NCCU202200901

表 4.7: 模型一分類績效

	Downgrade(-1)	Unchanged(0)	Upgrade(1)	Macro - Averaging
召回率	0.3494	0.3614	0.7771	
精確度	0.6237	0.6818	0.4069	
F1 - Score	0.4479	0.4724	0.5342	0.4848

調評級與下調評級的分類召回率皆較模型一高，顯示經過遷移的 BERT 文字分析模型所產生的特徵，更能衡量財金新聞媒體的文字情緒，提高了模型辨認未來公司被上調評級與下調評級的能力，也就是在評級被上調與下降的樣本中，模型二較模型一更能正確的分類公司未來評級變動狀況。雖然在精確度的部分，模型一在評級下調及評級不變的分類精確度較模型二高，顯示模型二的型一錯誤較高，不過以 F1-Score 觀察召回率與精確率的調和平均，模型二在分類信用評等上升及調降的樣本在整體上比模型一好，且模型二之 Macro-Averaging F1-Score 較模型一高，表示以調和平均衡量各分類的召回率與精確度後，整體上，模型二較模型一有更好的分類能力。

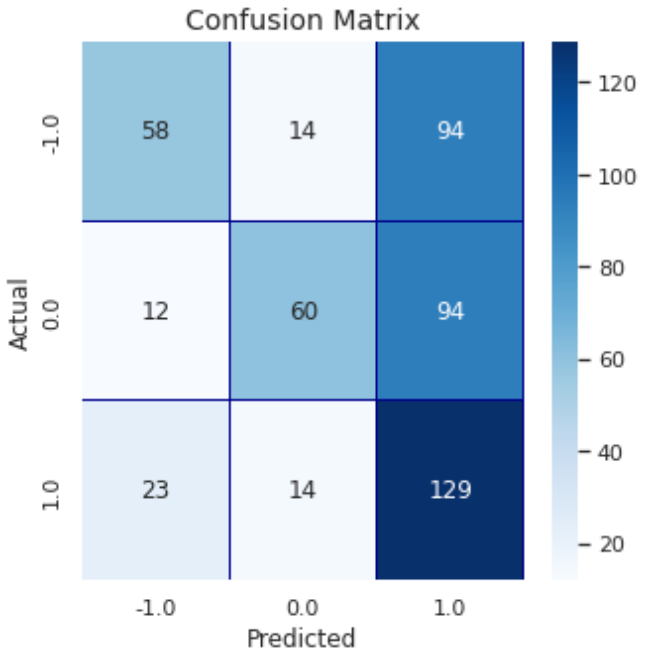

圖 4.8: 模型一分類混淆矩陣

為比較納入媒體情緒變數是否增進與改善僅納入財務變數的預警模型之預測效果，本研究以模型三與模型四之預警績效作為實驗比較基準。

僅觀察圖4.10與圖4.11，模型三在三個類別的分類效果都較只含有財務變數的模型四佳，模型三之預測準確率達到 58.43%，而模型四預測準確率為 45.98%，

DOI:10.6814/NCCU202200901

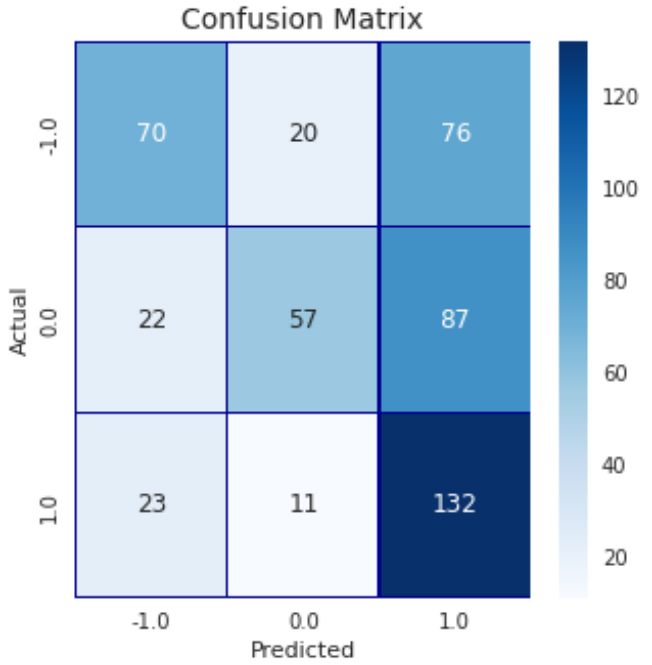

圖 4.9: 模型二分類混淆矩陣

表 4.8: 模型二分類績效

	Downgrade(-1)	Unchanged(0)	Upgrade(1)	Macro - Averaging
召回率	0.4217	0.3434	0.7952	
精確度	0.6087	0.6477	0.4475	
F1 - Score	0.4982	0.4488	0.5727	0.5065

DOI:10.6814/NCCU202200901

表 4.9: 模型三分類績效

	Downgrade(-1)	Unchanged(0)	Upgrade(1)	Macro - Averaging
召回率	0.5422	0.3795	0.8313	
精確度	0.6818	0.6702	0.5074	
F1 - Score	0.6040	0.4846	0.6301	0.5729

比較圖4.9與圖4.10則可以觀察出加入財務變數後，模型二分類過度分類樣本為評級上調的問題得到緩解，整體預測準確率進一步提升。

表4.9與表4.10為模型三與模型四之分類績效，顯而易見的，模型三在評級上調與評級下調的樣本召回率與精確度都較模型四佳，雖然評級不變的樣本召回率以模型四勝出，但在考量將確度後計算評級不變的 F1-Score 後，模型三還是高於模型四，且模型三之 Macro-Averaging F1-Score 為所有模型中最高，模型二居次，顯示使用財務變數與媒體情緒變數之模型擁有最好的信用評級變動預警能力，且若如模型二一樣只使用經過領域遷移過後的媒體情緒變數，模型仍有一定程度的預警效果。

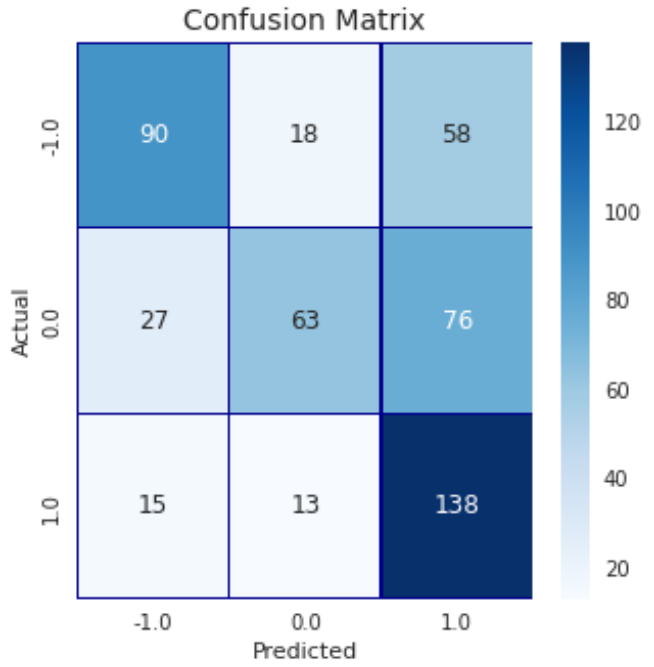

圖 4.10: 模型三分類混淆矩陣

DOI:10.6814/NCCU202200901

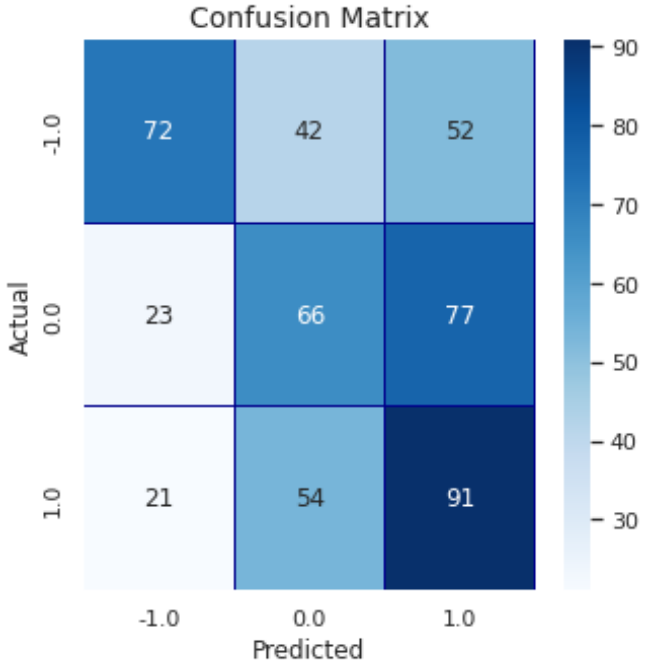

圖 4.11: 模型四分類混淆矩陣

表 4.10: 模型四分類績效

	Downgrade(-1)	Unchanged(0)	Upgrade(1)	Macro - Averaging
召回率	0.4337	0.3976	0.5482	
精確度	0.6207	0.4074	0.4136	
F1 - Score	0.5106	0.4024	0.4715	0.4615

DOI:10.6814/NCCU202200901

4.4.2 模型 ROC 曲線與 AUC

ROC 曲線與 AUC 用於衡量模型在不同閾值下的整體表現與分類效果，ROC 曲線愈往左上方靠攏表示該模型之分類效果越佳，下列各圖表為不同模型在分類評級上升、評級下調與評級不變的樣本時的 ROC 曲線圖。

圖4.12為模型一各分類之 ROC 曲線，各分類之 AUC 皆在 0.7 左右，顯見模型已有相當分類效果，且以分類評級下調的樣本有較佳之 AUC。圖4.12為模型二的各分類 ROC 曲線，其 Micro-Average ROC 與 Macro-Average ROC 皆較模型一高，表示使用經過遷移之媒體情緒變數，可以建構更好的信評變動預警模型，且信評不變與信評下調兩類為改進的較多的類別，也就是經過遷移的媒體情緒變數更能挖掘企業未來信用下調的風險。

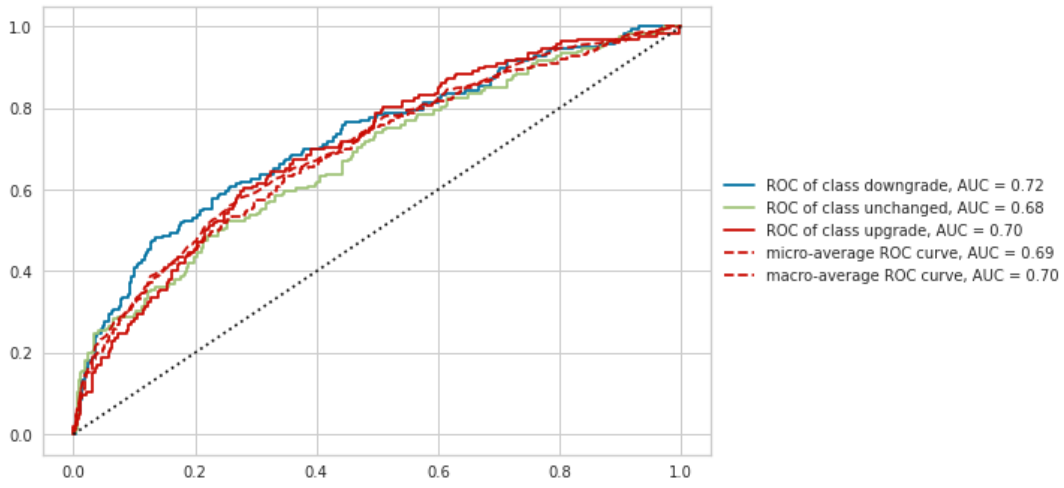

圖 4.12: 模型一之 ROC 曲線

圖4.14為模型三之各分類 ROC 曲線，雖然信用評等不變的類別 AUC 值相較模型二下降，但評級上調與與評級下調的 AUC 已提升到 0.8 左右，且 Micro-Average ROC 與 Macro-Average ROC 又為全部模型中最高值，表示納入領域遷移過後的財務媒體情緒與財務變數所建構之模型擁有最佳信用評等變動預警效果。相較之下，圖4.15中模型三個類別的 ROC 曲線，皆為所有模型中最低，顯見在信用評級變動預警模型中應納入不同於傳統結構型資料的重要性。

DOI:10.6814/NCCU202200901

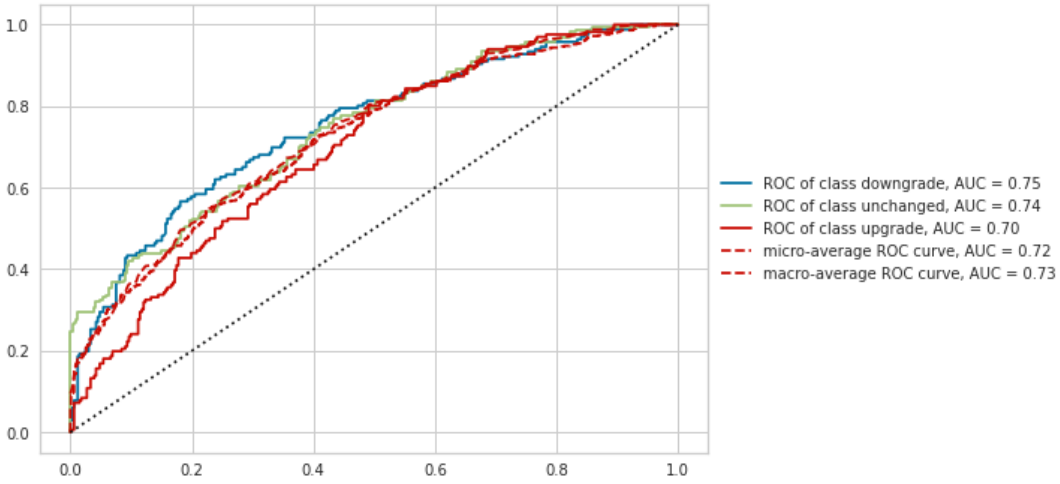

圖 4.13: 模型二之 ROC 曲線

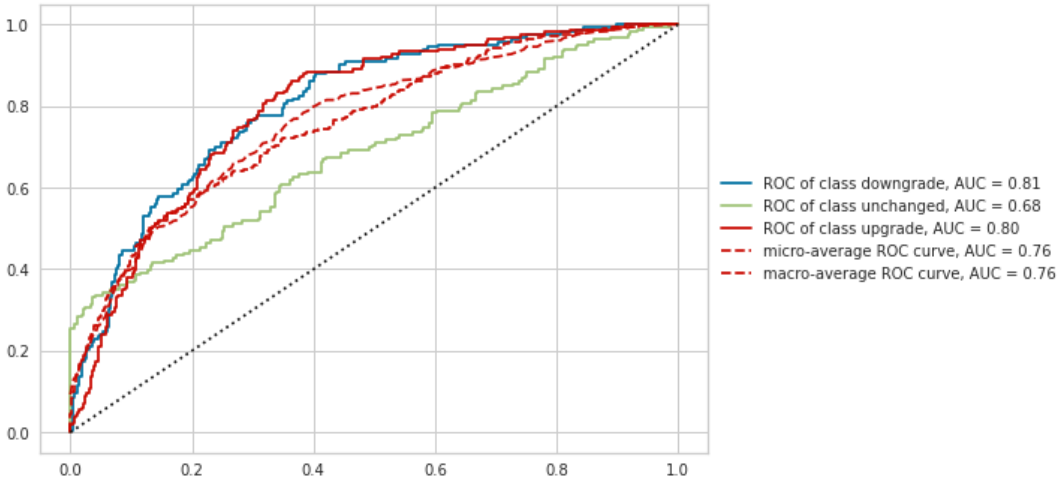

圖 4.14: 模型三之 ROC 曲線

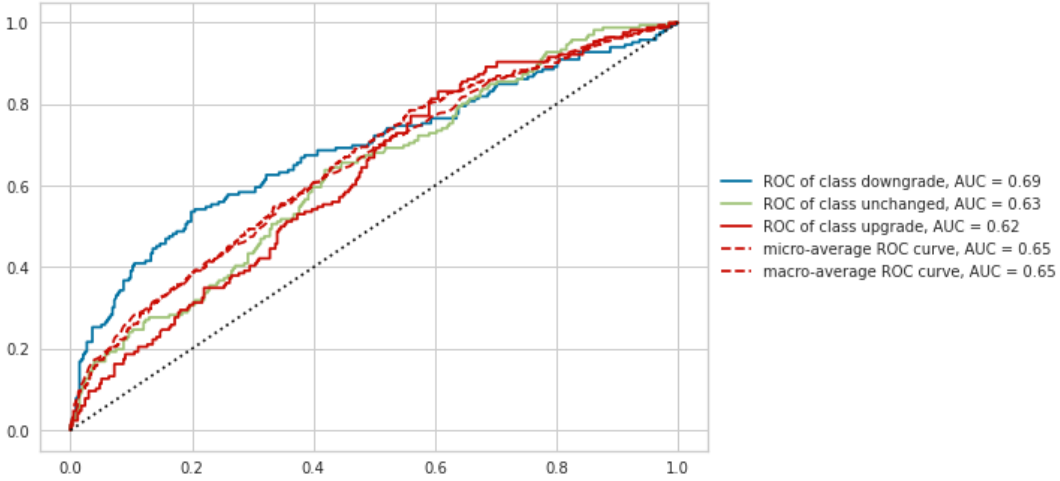

圖 4.15: 模型四之 ROC 曲線

DOI:10.6814/NCCU202200901

4.4.3 特徵重要性

依據隨機森林中使用某一特徵時平均降低的吉尼不純度，我們可以觀察出該特徵在整個模型的影響程度，進而排列出每個特徵的在模型中的重要性，下列各圖表為各模型建構過程中所排列出的特徵重要度。圖4.16與圖4.17為模型一與模型二的特徵重要度排序，由於這兩個模型只納入媒體情緒變數，因此特徵重要度的排序意味著何種情緒與何種主題主要影響模型判斷樣本未來的信用評級變化。從情緒分類的結果來看，本研究萃取之情緒當中，以負面情緒因子主要影響企業未來信評變化，模型一前 20 大重要影響因子中，有 9 個為負面情緒因子、0 個正面情緒因子，模型二則有 11 個負面情緒因子、2 個正面情緒因子，這與Tetlock (2007)、Smales (2016) 的研究結果一致，因此，媒體正負面情緒除了對市場價格外，也對企業未來信用評等狀況有不對稱的影響力。而從主題的角度來看，與信用評等相關的主題，如評等機構對企業的信評觀望、財經專欄對企業未來信評的評論等，為模型區分未來信用評等變化的主要因素。除此之外，公司的營收狀況及未來預期營收等相關新聞也是區分公司未來信評變化的重要因子。

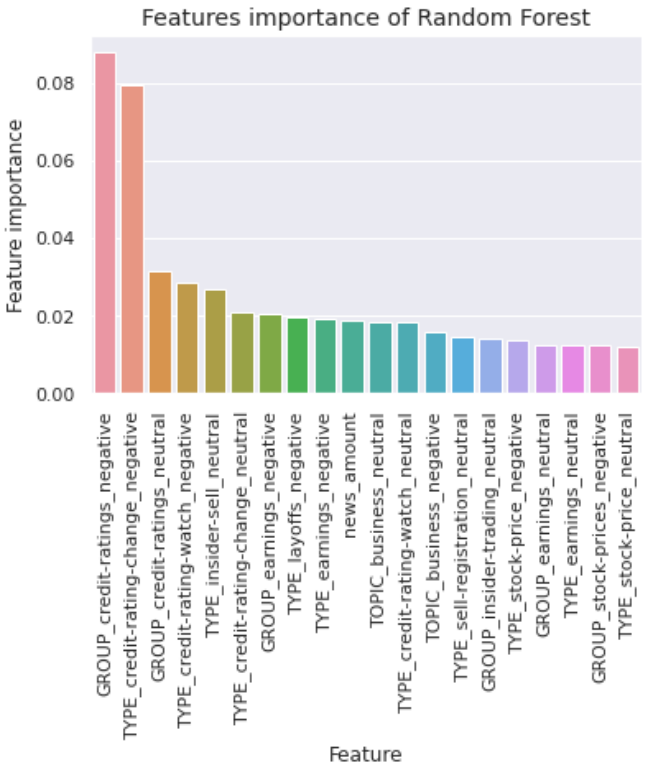

圖 4.16: 模型一特徵重要度

圖4.18為納入財務變數的模型三特徵重要度排序，模型三的前 20 大重要特徵

DOI:10.6814/NCCU202200901

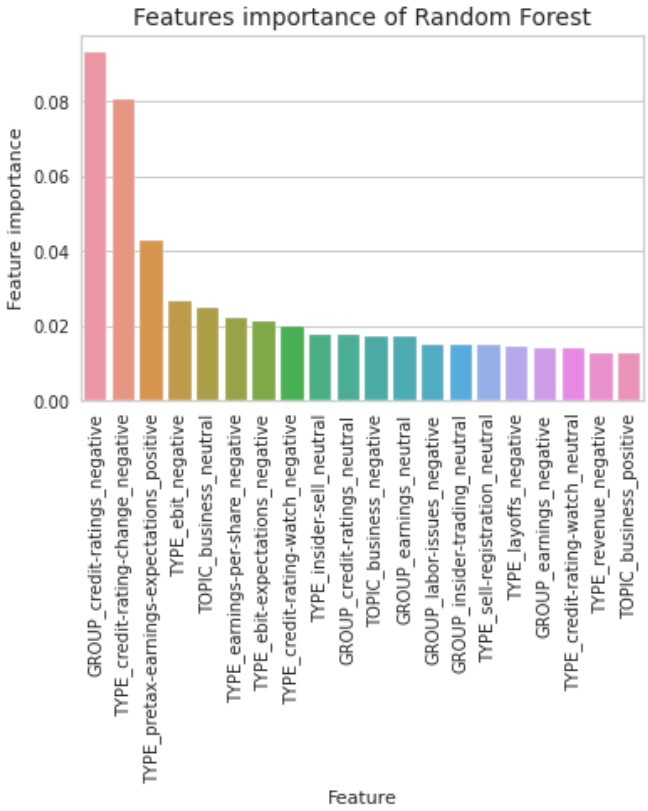

圖 4.17: 模型二特徵重要度

中有 6 個屬於財務變數，依序分別為稅前淨利、每股盈餘、利息保障倍數、違約距離、帳市比及營業收入，與只納入財務因子的模型四之特徵重要度 (圖4.19) 順序大致相同，財務因子仍為評價一間公司未來信評狀況的重要因素。而媒體情緒的排序與模型一和模型二差距不大，仍是以信評相關主題新聞與營收狀況為重要特徵。

除了信用評等、營收等主題的新聞直觀地影響企業未來信用評等外，本研究將模型三之重要特徵中除去財務變數、信評相關新聞與營收相關新聞的特徵，繪製圖4.20。可以觀察到新聞總數也是模型分類時的重要參考因子，相關新聞數量某種程度上屬於大眾對該企業的關注程度，而關注程度越高隱含著該公司財務狀況未來可能發生改變，此結果與Norden (2017) 一致。除此之外，負面的資遣、勞工權益與資產處置等新聞也是模型分類的重要特徵，這類新聞隱含著公司財務狀況不佳，必須節流所帶來的資產處理事件與勞資問題。

為了了解每個特徵對模型分類的正負向影響，本研究針對模型三使用 SHAP 方法解構每個特徵對模型分類之貢獻。圖4.21為模型三將樣本分類為上調評級時，每個特徵對於其模型輸出之貢獻。在這個圖表中，x 軸代表 SHAP 值，y 軸代表

DOI:10.6814/NCCU202200901

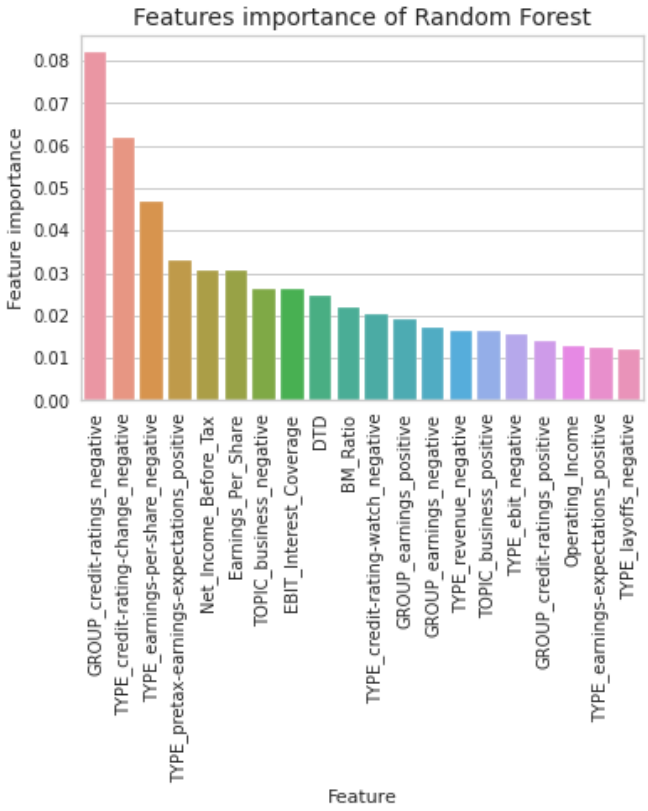

圖 4.18: 模型三特徵重要度

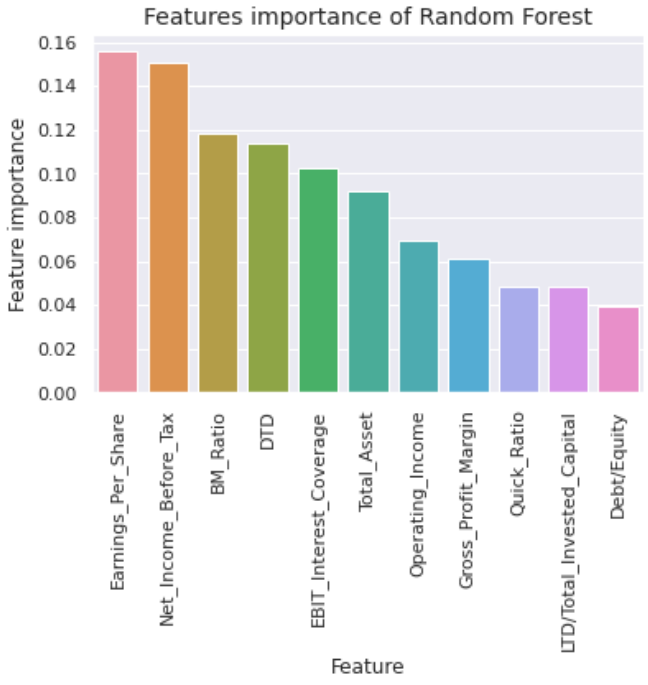

圖 4.19: 模型四特徵重要度

DOI:10.6814/NCCU202200901

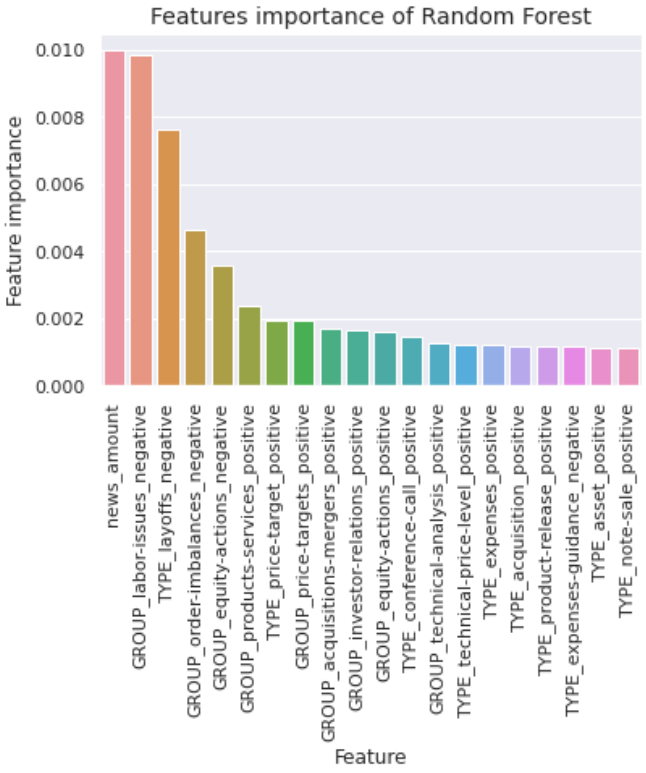

圖 4.20: 模型三其他重要特徵

所有特徵，其中的每個點都代表某個樣本在該特徵的 SHAP 值，而圖中的黑色實線為 SHAP 值為 0 之標線。而紅色表示該樣本之特徵數值較高，藍色則表示該特徵數值較低。我們可以根據紅點和藍點的分佈來大致了解特徵的方向性影響。以特徵 *GROUP_credit-ratings_negative*(負面信評主題新聞數量) 為例，其紅色的樣本點集中在左側，表示所有的樣本中，擁有較高的負面信評新聞數量的樣本，其 SHAP 值為負，也就是模型在納入該特徵時，會降低該樣本分類為上升評等之機率。簡單來說，若該特徵與模型分類結果為正相關，則該特徵的紅點會集中分布在右側，反之，若為負相關，則紅點集中在左側，且我們通常希望紅點與藍點的分佈能夠分離開來，表示該特徵的高低值對模型輸出影響差別較大。

圖4.21為模型分類樣本為上調評級時，前 20 個貢獻最大之特徵。圖表的分布狀況大致符合預期，也就是負面的媒體情緒特徵如負面的預期營收新聞、負面商業類別新聞等，將降低模型分類該樣本為上調評級的機率，而越高的獲利能力與預期營收的正面新聞數量將提升模型分類該樣本為上調評級的機率。

圖4.22為模型分類樣本為下調評級時，前 20 個貢獻最大之特徵。圖表的分布狀況與圖4.21相反，表示這些特徵對模型辨認公司未來評級狀況具有一致性，負

DOI:10.6814/NCCU202200901

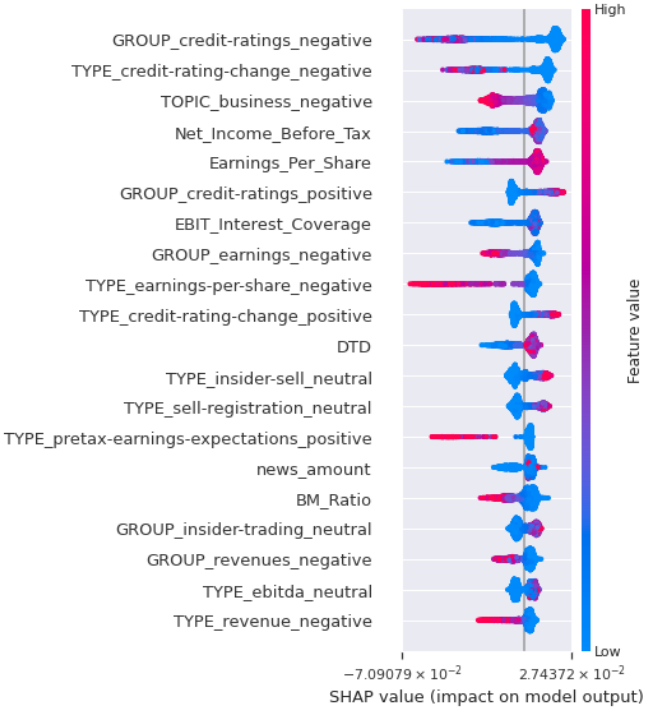

圖 4.21: 分類為上調評級之特徵貢獻

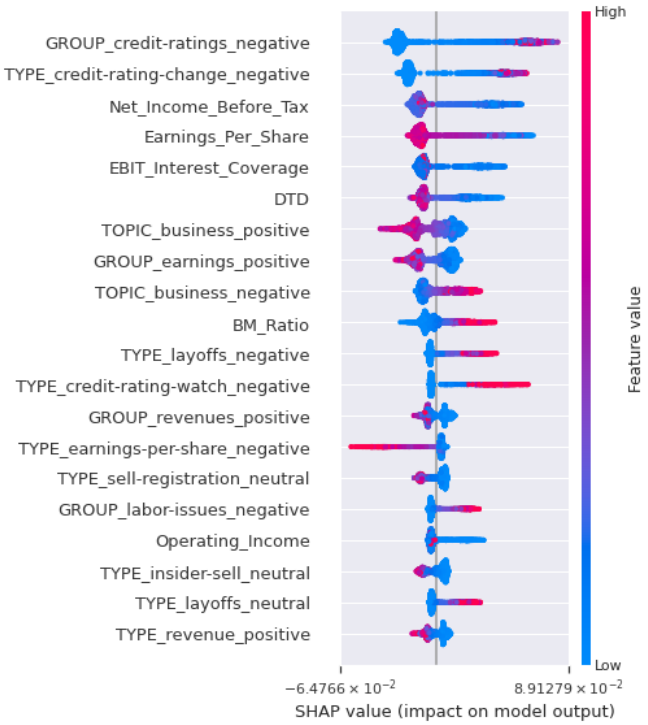

圖 4.22: 分類為下調評級之特徵貢獻

DOI:10.6814/NCCU202200901

面的媒體情緒特徵及越低的獲利能力將提高模型分類該樣本為下調評級之機率。

一樣地，我們排除財務變數、信評相關新聞與營收相關新聞的特徵，繪製圖4.23與圖4.24觀察其他特徵之貢獻，可以觀察到新聞總數在模型分類樣本為上調與下調時皆有影響力，但其影響力分佈較為模糊，藍點與紅點都相當集中，表示新聞總數增加並不一定代表該公司上升或下降評等的機率增加，而是加強其他特徵對於評級變動之影響，與一般認知相符。與上調評級正相關的特徵有正面的季度電話會議新聞、正面的市場份額新聞和產品新聞等。在下調評級的部分，我們也驗證了前述的推論，負面的資遣新聞、勞資糾紛新聞及資產處置與重組等新聞將提升模型將樣本分類為下調評級的機率。

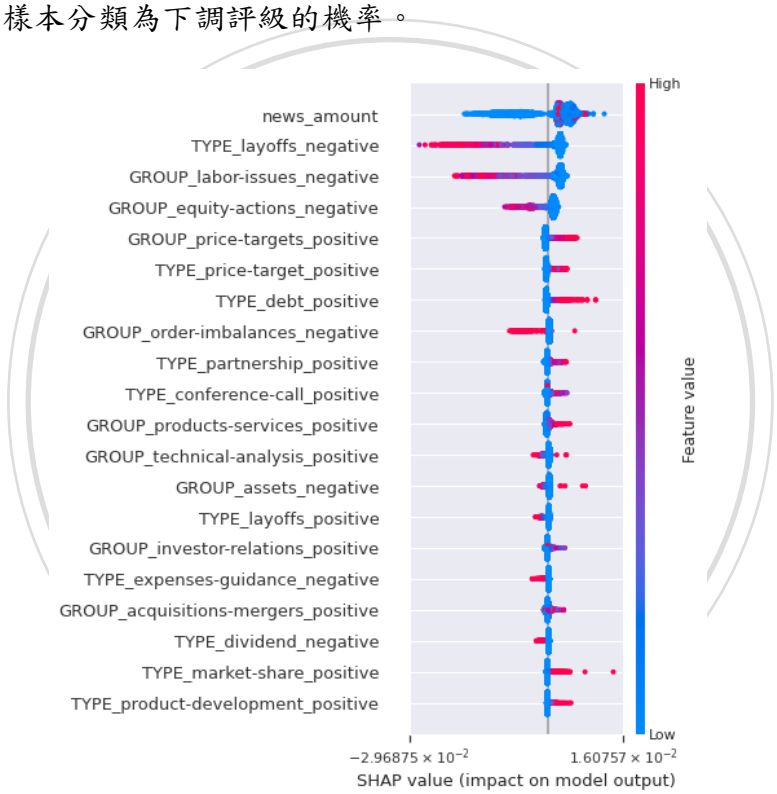

圖 4.23: 分類為上調評級之其他特徵貢獻

DOI:10.6814/NCCU202200901

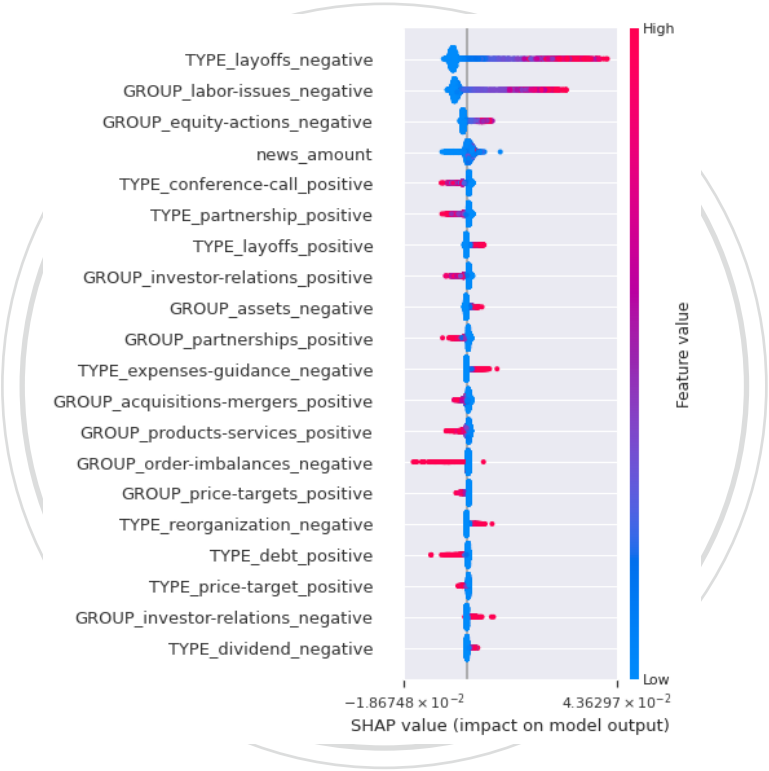

圖 4.24: 分類為下調評級之其他特徵貢獻

DOI:10.6814/NCCU202200901

第五章　結論與建議

本研究嘗試以遷移學習的方式，改進傳統字典模型辨認語句情緒可能造成的偏誤，進而建構更能萃取財金新聞媒體情緒的信用評等變動預警模型。以 Google 發布的 BERT 模型為基礎，本研究在此模型下游建置線性神經網路，並以 SST5 資料庫與 Financial Phrasebank 資料庫內經過標籤情緒之語句進行訓練，分別建置兩個能夠辨認文章情緒的神經網路模型。由此兩個模型生成新聞之情緒特徵，進而比較與探討文字處理模型是否經過遷移對特徵生成之影響。接著，本研究將 Raven Pack 資料庫內的財金新聞輸入上述兩個文字情緒分析模型，以兩個模型所生成之特徵建構企業信用評等變動預警模型，嘗試在各模型間探討將財金媒體情緒應用於對企業信用評等預警之成效。

本研究建構兩個文字語意分析的神經網路模型，其一為經過領域遷移至財金相關文本的 BERT 模型，另一個則為未遷移的 BERT 模型，並使用兩個模型在本研究所使用的 Financial Phrasebank 財金新聞語庫測試集內進行情緒分類，經過遷移的 BERT 模型在情緒分類的準確率達到 83.43%，而未經遷移之 BERT 模型的情緒分類準確率只有 52.96%。文字分析模型在其訓練語庫的背景不同下，對於情緒的分類也會有差異，本研究的研究結果指出在評估財金相關文字情緒時，若原模型訓練時所使用的文字不屬於財金文本，應將模型進行語意領域遷移，以避免模型錯估語意而造成分類錯誤。

同時本研究分別建構四個隨機森林模型，使用企業 180 天內的財金媒體主題情緒特徵與財務因子預測該企業未來 90 天內的信用評級變動狀況。實證結果指出，使用經過遷移的模型產生之特徵建構信用評級變動預警模型，其分類召回率在上調評級與下調評級的類別均較未遷移的特徵有所提升，且整題模型之 Macro-Averaging F1-Score 也較高，領域遷移改善了模型的評級變動預測表現。而在納入財務因子後，模型召回率及準確率進一步提升，且模型之 Macro-Averaging

DOI:10.6814/NCCU202200901

F1-Score 為四模型中最高，本研究實證媒體情緒提供企業評級變動預測除去財務因子以外的解釋能力。此外，就本研究所萃取之財金媒體主題情緒中，我們的發現與Tetlock (2007) 和Smales (2016) 一致，負面的媒體情緒對模型分類的影響較大，故媒體情緒對未來公司評級變動有不對稱之影響效果。而在排除信評展望、公司營收等媒體文章所產生之特徵所帶來的影響，本研究發現新聞總數也是模型分類時的重要參考因子，此結果與Norden (2017) 一致，另外，負面的資遣、勞工權益與資產處置等新聞也是模型分類的重要特徵。

本研究比較了有無領域遷移的 BERT 模型對財金文本的情緒分類績效，故兩模型皆為神經網路模型，本研究建議未來可納入字典模型、主題模型等其他文字分析模型來比較是否進行領域遷移對模型分析財務文本的影響，而在評級變動預測模型的部分，本研究因解釋性原因使用隨機森林作為基礎模型，未來研究則可以尋找預測效果更好亦能維持解釋力的模型進一步改善模型預測效能。另外，本研究使用媒體情緒因子與財務比率因子預測企業評級的變動狀況，提供了前述因子對企業評級變動的影響，然而，此影響是否如同市場效率假說已經被納入市場交易考量，亦或是市場價格尚未納入的潛在資訊，未來進一步研究可以以實際結構型商品價格的資訊進行深入挖掘。

DOI:10.6814/NCCU202200901

參考文獻

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609.

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Collin-Dufresn, P., Goldstein, R. S., and Martin, J. S. (2001). The determinants of credit spread changes. *The Journal of Finance*, 56(6):2177–2207.

Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all fears investor sentiment and asset prices. *The Review of Financial Studies*, 28(1):1–32.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dyer, T., Lang, M., and Stice-Lawrence, L. (2017). The evolution of 10-k textual disclosure: Evidence from latent dirichlet allocation. *Journal of Accounting and Economics*, 64(2-3):221–245.

Ericsson, J., Jacobs, K., and Oviedo, R. (2009). The determinants of credit default swap premia. *Journal of Financial and Quantitative Analysis*, 44(1):109–132.

Fama, E. F. (1960). Efficient market hypothesis. *Diss. PhD Thesis, Ph. D. dissertation*.

DOI:10.6814/NCCU202200901

Galil, K. and Soffer, G. (2011). Good news, bad news and rating announcements: An empirical investigation. *Journal of Banking & Finance*, 35(11):3101–3119.

Hajek, P. and Michalak, K. (2013). Feature selection in corporate credit rating prediction. *Knowledge-Based Systems*, 51:72–84.

Huang, A. H., Lehavy, R., Zang, A. Y., and Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, 64(6):2833–2855.

Hull, J., Predescu, M., and White, A. (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking & Finance*, 28(11):2789–2811.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 216–225.

Jarrow, R. A. and Turnbull, S. M. (1995). Pricing derivatives on financial securities subject to credit risk. *The Journal of Finance*, 50(1):53–85.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., and Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*, 56(1):130–147.

Lee, Y.-C. (2007). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications*, 33(1):67–74.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3):221–247.

Li, X., Xie, H., Chen, L., Wang, J., and Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69:14–23.

DOI:10.6814/NCCU202200901

Liberti, J. M. and Petersen, M. A. (2019). Information: Hard and soft. *Review of Corporate Finance Studies*, 8(1):1–41.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Loughran, T. and McDonald, B. (2014). Measuring readability in financial disclosures. *the Journal of Finance*, 69(4):1643–1671.

Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230.

Lu, H.-M., Tsai, F.-T., Chen, H., Hung, M.-W., and Li, S.-H. (2012). Credit rating change modeling using news and financial ratios. *ACM Transactions on Management Information Systems (TMIS)*, 3(3):1–30.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

Mayew, W. J. and Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, 67(1):1–43.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, pages 141–183.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, 85(6):2107–2143.

Norden, L. (2017). Information in cds spreads. *Journal of Banking & Finance*, 75:118–135.

Norden, L. and Weber, M. (2004). Informational efficiency of credit default swap and stock markets: The impact of credit rating announcements. *Journal of Banking & Finance*, 28(11):2813–2843.

DOI:10.6814/NCCU202200901

Orsenigo, C. and Vercellis, C. (2013). Linear versus nonlinear dimensionality reduction for banks ' credit rating prediction. *Knowledge-Based Systems*, 47:14–22.

Pedrosa, M. (1998). Systematic risk in corporate bond credit spreads. *Journal of Fixed Income*, 8(3):7–26.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P. J., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.

Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2020). Measuring news sentiment. *Journal of Econometrics*.

Smales, L. A. (2016). News sentiment and bank credit risk. *Journal of Empirical Finance*, 38:37–61.

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.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The journal of finance*, 63(3):1437–1467.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., and Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.

DOI:10.6814/NCCU202200901