Review of Securities and Futures Markets, 36:2, 159-206 (2024)
DOI:10.6529/RSFM.202406_36(2).0004

年報文字基礎溝通價值與公司信用評等預測：機器學習模型之應用

陳宗岡　廖咸興　官顥　林育均　郝昀*

在以往信用評等預測相關文獻中，大多以統計模型來進行信用評等的分類預測，並財務特徵變數作為主要的解釋變異來源。然而，根據標準普爾的信用評等準則可知，企業信用評等乃先以企業經營風險及財務風險決定一初值後，再經由多面向的非量化因素調整而得，是故財務變數並無法捕抓信用評等資訊的全貌。不同於以往文獻，本研究以機器學習模型 (隨機森林、支持向量機、極限梯度提升、羅吉斯回歸) 為基礎，並在既有財務會計特徵變數為標竿設定下，額外引入年報文字基礎溝通價值變數 (e.g.可讀性與語意) 來捕抓信評公司所考量的非量化調整因素，特別是不完全會計資訊的部份。本研究利用美國市場 1994 年至 2017 年的公司信用評等資料來進行分析，實證結果顯示：在額外投入年報文字基礎溝通價值變數後，模型預測效力 (e.g. F1 分數) 均有一定程度的提升，其中又以隨機森林及極限梯度提升模型總體表現最好 (F1 分數可達 0.76～0.77，增額提升約 6%)。此顯示年報文字基礎溝通價值資訊對信用評等

「政策與管理意涵」

本研究發現年報文字基礎溝通價值變數與傳統財務變數兩者對信用評等的分類預測效力具有互補性。如對信用品質較差的公司而言，年報文字基礎溝通價值變數所能額外補抓的信用風險資訊效果較財務變數為多。對主管當局來說，應更留意信用品質較差公司的年報文字基礎溝通價值。此亦可作為未來年報文字表達規範建議之政策參考依據。

* 通訊作者：郝昀，國立陽明交通大學管理科學系博士生，通訊地址：新竹市東區大學路 1001 號管理一館 212 室，電話：(03)5712121 ext 57124，E-mail: peterhao.c@nycu.edu.tw；陳宗岡為國立陽明交通大學管理科學系教授暨國立臺灣大學計量理論與應用研究中心特約研究員，E-mail: vocterchen@nycu.edu.tw；廖咸興為國立臺灣大學財務金融學系教授，E-mail: hliao@ntu.edu.tw；官顥為國立陽明交通大學管理科學系碩士，E-mail: kuanhao.861029@gmail.com；林育均為國立陽明交通大學管理科學系學生，E-mail: magic.mg09@nycu.edu.tw。

作者感謝期刊主編鍾惠民教授、兩位匿名審查人、以及 2023 年臺灣財務金融學會年會的評論人葉宗穎教授及其他與會者提供的寶貴建議。作者感謝教育部高等教育深耕計畫補助國立臺灣大學計量理論與應用研究中心之研究經費 (編號 112L900201)。

有不同於傳統財務變數的額外解釋能力。此外，本研究亦發現年報文字基礎溝通價值資訊能進一步降低非投資級公司被誤判為投資級公司的比率，亦即對非投資級公司的評等分類有更高的預測效力。因此，本研究驗證年報文字基礎溝通價值變數可捕抓到信評公司的非量化因素調整之資訊內涵。

關鍵詞：年報文字基礎溝通價值、信用評等、機器學習、不完全資訊。

壹、緒論

近年來，影響金融市場劇烈波動的事件層出不窮，如 2008 年發生的美國次貸金融風暴及 2020 年間的新冠肺炎疫情皆對企業的信用風險帶來嚴重的影響。近期甚有矽谷銀行違約及瑞士信貸債務危機等事件。重大金融事件導致了股市的劇烈跌幅、原物料價格的衝擊、供應鏈的長鞭效應，而因重大金融事件受到衝擊之企業和產業亦可能面臨資金短缺的困境。根據惠譽 (Fitch Ratings) 報告指出，全球金融機構和非金融機構產業均在重大金融事件後面臨巨大的信用風險危機，如信用評等降級或債務違約等。而大部分受重大金融事件所影響之企業股價也會受到顯著的負面影響，如台灣的國泰人壽、富邦人壽皆在新冠疫情爆發後，被惠譽 (Fitch Ratings)、標準普爾 (S&P Global Ratings) 評等展望調為負向，直接導致了其股價的跌幅；然除了信用評等調降外，嚴重受影響之企業亦可能面臨前所未有的企業破產或違約之嚴重事件。因此，企業之信用危機，如企業破產、債務違約、信評降級、股價劇烈跌幅等，皆會為投資人、債權人、發行人、政府等外部利害關係人帶來甚大的影響。其中，如何審慎地衡量其企業的信用風險，即為外部利害關係人非常注重一項課題。

實務上，目前信用評等已經廣泛的被外部利害關係人視為評估企業信用風險的重要衡量指標。投資人會依據信用評等之等級高低作為主要的投資參考依據，而信評的高低也代表公司的風險承受能力。此外，當公司信用評等被降級時，亦可能會面臨需要履行特別的契約責任，並承擔違約的風險。而政府亦可能會視企業的信用評等來決定其投資額度或作為財政政策參考之依據。綜上所述，信用評等是風險管理中非常重要的一項議題。其中，信用評等之預測或是破產 (違約) 機率之估計將有助於外部利害關係人可以更精確的做部位的管理與風險的胃納之評估。然而，標準普爾 (S&P Global Ratings)、惠譽 (Fitch Ratings)、穆迪 (Moody's) 等公司之信評工作成本相當高，其過程需要投入大量時間和人力成本，並需要依據各信評公司不一樣的評等方法論，且多方面評估企業財務體質、產業前景、國家政治風險甚至是外部利害關係人所不知道的

調整機制。除了需要投入至少數個月的時間，且無法提供外部利害關係人非常即時的企業信用風險監管資訊，故信評機構之評等資訊基本上屬於「落後指標」。同時，並非每家企業都有能力負擔其信用評等之高額評估成本，亦即並非每家企業皆有信評資訊。雖然信用評等已經成為全球主權國家、企業、債券的重要風險衡量指標，但是其評估流程之的客觀性亦經常被外部利害關係人所詬病。如經濟合作暨發展組織 (OECD) 於 2020 年出版的「全球債券市場趨勢、新興風險與貨幣政策」報告指出，信評機構之信評在金融風暴之後有「膨風 (Rating Inflation) 」之情形。為更精確瞭解信評公司的評估考量，本研究嘗試利用機器學習模型來模擬標準普爾之信評方法論，並結合引入「年報文字基礎溝通價值」(Communicative value; Seebeck and Kaya, 2023) 變數來進行探究，期能提供給外部利害關係人更準確且即時的評等資訊。

在過往信用風險預測相關研究中，多數學者乃著重在公司破產預測相關研究，且多以公司財務會計比率及股價相關資料作為模型投入變數，如 Altman's (1968) Z-score、及 Ohlson's (1980) O-score。上述研究皆發現財務會計比率及股價相關資料均可作為企業信用風險衡量的重要指標。此外，根據標準普爾的評等方法論可知，信用評等的決定乃初步先以企業的業務風險及財務風險來形成定錨評級後再予以非量化因素的調整。由此可知財務會計比率和股價相關資料為分析企業信用風險的一大基礎，故本研究以此作為基礎模型的投入變數。

過往信用風險研究乃多著重於定量或結構化之數值資料，惟近來研究亦開始重視定性或非結構化的數據，如企業年報文本內容中所隱含的前瞻性資訊或潛在訊息。而在信評機構之評等方法論中，「不完全資訊」亦在信用評等調整的非量化因素中扮演一重要的角色。Mayew et al. (2015) 發現公司年度報告 (10-K) 中之 MD&A (Management Discussions and Analyses) 包含了評估企業信用風險相關的前瞻性資訊；而 Campbell et al. (2014) 和 Loughran & Mcdonald (2011) 的研究亦表明，企業之財務文本包含信用風險相關的前瞻性資訊。是故，本研究除了以企業的財務會計資訊和股票市場資訊當作為模型之基礎投入變數之外，亦額外以公司「年報文字基礎溝通價值」變數 (e.g.可讀性及語意) 作為模型的額外投入變數。其中，「年報文字基礎溝通價值」(Seebeck and

Kaya, 2023) 能反映外部投資人接收公司年報文字信息的程度，故能反映企業的不完全資訊程度。故額外引入「年報文字基礎溝通價值」變數乃與 Duffie & Lando (2001) 及信評公司評等方法論之「不完全資訊」立論一致，故本研究使用年報文字基礎溝通價值變數來進行信用評等分類預測具有相當的理論基礎。而本研究擬透過量化公司財務文本非結構化資訊導入至機器學習模型中，提供外部利害關係人更精確之評等資訊。

目前在結合年報文本資訊及機器學習模型來進行公司信用風險預測的相關文獻中，主要以破產預測研究為主，如 Mai et al. (2019) 及 Chen et al. (2023)。其中，Mai et al. (2019) 及 Chen et al. (2023) 分別以年報中 MD&A 的詞頻與整份年報的可讀性及語意來作為破產預測模型之新增投入變數。然而，在信用評等預測之相關研究方面，仍多以公司財務比率變數結合機器學習或深度學習模型來進行信用評等預測評估 (e.g. Wang and Ku, 2021，註 1)，鮮少有以年報文字基礎溝通價值變數作為機器學習模型的新增投入變數來進行分析。然而，依前段討論可知，在預測企業信用評等上，傳統的財務會計變量 (結構化數據) 與年報文字基礎溝通價值變數 (非結構化數據) 可能具有互補的效果，故預期可提升機器學習模型的預測效力。是故，本研究依循 Mai et al. (2019) 及 Wang and Ku (2021) 所使用的財務及股價特徵變數作為基礎模型的財務會計投入變量，並進一步比較額外加入年報文字基礎溝通價值特徵變量 (可讀性和語意) 後，機器學習模型是否會有更佳的預測效力。此亦即探究公司非結構化且定性的數據是否可以額外補充傳統財務會計投入變量無法解釋到的企業信用評等資訊。

本研究利用 1994 年至 2017 年的美國公司信用評等資料，並進行資料預處理、特徵變數選取、及資料不平衡處理 (RandomOverSampling、SMOTE) 等程序後，並利用羅吉斯迴歸 (Logistic Regression)、隨機森林 (Random Forest)、極限梯度提升(XGBoost)、及支援向量機 (Support Vector Machine; SVM) 等四種機器學習模型來探討在額外引入年報文字基礎溝通價值變數後對模型的增額預測效力。此外，為強化分析結果的穩健性，本研究採用 Rolling Window 的方

註 1：Wang and Ku (2021) 使用 28 項公司財務特徵變數作為模型投入變數，重要性前五名之特徵變數依序為：負債比率、公司市值、息前淨利、股東權益、及營業收入。

法來訓練集與測試集資料之切分，亦即以 1994 年至 2011 年為訓練組，逐年滾動式地來預測未來一年的公司信用評等，最後將上述六項結果平均而得最終預測結果。實證結果顯示，在新增投入年報文字基礎溝通價值變數後，模型績效指標如準確率 (Accuracy)、精確率 (Precision)、召回率 (Recall)、F1 分數 (F1-Score) 皆有顯著的改善，特別是極限梯度提升及隨機森林模型表現相對優異 (如 F1 分數可達 0.76～0.77，增額提升約 6%)。此外，信用評等被模型錯誤高估的比例也有顯著的減少，這顯示年報文字基礎溝通價值變數資訊能額外提供傳統財務會計變數無法捕抓的信用風險相關之不完全資訊。再者，本研究亦發現年報文字基礎溝通價值變數能進一步降低非投資級公司被誤判為投資級公司的比率，亦即對非投資級公司的評等分類有更高的預測效力。因此，年報文字基礎溝通價值變數可降低信評公司的評級誤判風險及金融機構的授信風險。

整體而言，本研究發現年報文字基礎溝通價值變數與傳統財務變數兩者對信用評等的分類預測效力具有互補性，如對信用品質較差的公司而言，年報文字基礎溝通價值變數所能額外補抓的信用風險資訊效果較多；而對信用品質較好的公司而言，財務變數所能補抓的信用風險資訊效果則較多。這是因為信用品質較差隱含資訊不完全程度較高，故財務變數的可信度較低，故年報文字基礎溝通價值變數所能額外補抓到的資訊效果相對較高，此與 Duffie and Lando (2001) 的不完全資訊觀點一致。最後，透過本研究模型模擬標準普爾之評等方法論，也可以進一步延伸應用於無信評之公司，提供外部利害關係人該企業信用風險的一項參考依據。

需注意的是，Wang and Ku (2021) 雖有利用機器學習模型及深度學習模型來探討公司信用評等之預測，惟其所使用的投入變數為財務會計及股價相關變數，而非年報文字基礎溝通價值變數 (可讀性及語意，Seebeck and Kaya, 2023; Chen et al., 2023)。故本研究的增額貢獻包含：(1)引入 Seebeck and Kaya (2023) 及 Chen et al. (2023)所提之文字基礎溝通價值變數(可讀性及語意)於公司信用評等預測模型中；(2)提供年報文字基礎溝通價值變數能顯著改善將非投資級公司被誤判為投資級公司的證據，並能有效改善信評公司的評級誤判風險及銀行的授信風險；(3)提供年報文字基礎溝通價值變數與傳統財務變數兩者對信

用評等的分類預測效力具有互補性的相關證據；(4)本研究模型可在基於模擬標準普爾之評等方法論下，延伸應用於無信用評等的公司並得出初步的信用評級。綜上所述，本研究結果不僅可補足信用評等預測文獻的學術缺口，亦能進一步提供信用評等及授信實務的改善建議。

貳、文獻回顧

本研究之文獻回顧會分成三部分做介紹。首先為信用風險基礎理論與相關研究，其次為公司年報文字基礎溝通價值之討論，最後為機器學習應用於信用風險事件預測效力之研究。其中，本研究所使用機器學習模型之建立與投入變數之選擇，包含結構化數據 (會計和股票市場資訊) 和非結構化數據 (公司財務文本之文字特徵)，均為依循過往相關文獻所提及之機器學習模型與投入變數作為本文之基礎參考依據，並以此延伸作更深入的研究與分析。

一、信用風險基礎理論與相關研究

（一）信用風險基礎理論

信用風險是公司財務相關研究中非常重要的問題之一，尤其在金融危機後，各企業之信用評級成為投資人、債權人、政府所等等外部利害關係人關注的重要指標，金融機構亦為了降低企業破產 (違約) 風險而使用信用評級作為衡量企業信用風險之標準 (Baesens et al. , 2003；Kim & Ahn, 2012；Tsai & Chen, 2010；Yu et al., 2008)。因此，企業信用評級是信用風險中的重要議題之一；而信用風險相關事件亦為外部利害關係人相當關注的一項資訊，如企業破產 (違約)、信評升降級、股價顯著的負報酬等是 (Donovan, Jennings, Koharki & Lee, 2021)。

根據結構型信用風險模型 (Merton, 1974; Duffie & Lando, 2001) 可知，資產價值分配 (含資產價值水準及資產價值波動)、負債水準、及不完全會計資訊乃公司信用風險的四項主要決定因子。其中，在 Duffie & Lando (2001) 所提及

的不完全會計資訊信用風險模型中，基於資訊不對稱假設，外部關係人僅能根據企業財務報表感知企業的資產價值分配，故會計資訊之不完全程度於外部投資人對企業真實資產價值分配之評估具有一定的影響。由於年報文字基礎溝通價值 (e.g. 可讀性及語意) 可代表公司資訊傳遞的速度及訊息表達之程度，故其可描繪企業的不完全資訊程度。本研究在傳統的基礎財務會計投入變數外，依循 Duffie & Lando (2001) 和信評公司評等方法論之「不完全資訊」之考量，透過量化公司年報文本之非結構化資料得出其文字基礎溝通價值資訊，並將其導入信用評等預測模型中，期能提供外部關係人更準確且更即時的評等資訊。

（二）信用風險相關研究

1.企業破產與違約機率相關研究

過往企業破產預測之投入變數中基本上可分為會計基礎 (Accounting-Based) 及股票市場基礎 (Market-Based) 兩大投入變數，因為這兩大類投入變數基本上就足夠反映了公司的財務體質是否健康、營運狀況是否良好。首先會計基礎變數是以公司歷史的財務報表之會計資訊為基礎，並結合企業實際破產 (違約) 事件，透過傳統統計方法進行事件之預測與分析，其中較著名之研究為 Altman (1968) 和 Ohlson (1980) 之企業破產預測。市場基礎變數則是以股票市場之交易資訊來衡量企業之信用風險，較著名之研究如 Merton (1974) 之選擇權評價模型所衍生的 KMV 違約機率風險模型。

近期企業破產預測相關研究多結合財務會計和股票市場兩類投入變數來進行企業破產事件之預測。Mai et al. (2019) 乃利用企業會計資料和股票市場資料並額外加入公司年報 MD&A 詞頻變數來進行企業破產預測；而 Donovan et al. (2021) 則以財務會計及股票市場資料來深入探討 Z-score、O-score、EDF、企業破產事件與信用利差 (Credit Spread) 之間的關聯性。上述研究皆表明企業之破產事件與會計和股票市場變數之間呈高度相關。因此，本研究乃以過去文獻納入考量財務會計特徵變數作為基礎的投入變數。

2.信用評等相關研究

信用評等乃為信用風險中重要的議題之一，係指企業或債券之信用狀況與

其還款能力之等級評分 (Huang et al., 2004；Lee, 2007)。而信用評等之等級亦能反應其企業或債券違約之可能性 (破產或違約機率) (Martens et al., 2010)。因此，企業或債券的信用評等可為外部利害關係人提供一個重要的投資參考指標。以企業信用評等為例，信評機構之評等方法論普遍以統計模型做為擬定評等之方法，並將所評估企業之各種風險予以量化。如標準普爾公司 (S&P) 或穆迪公司 (Moody's) 乃透過分析企業之財務體質狀況來得出信用評等 (Huang et al., 2004)；其中，標準普爾之評等方法論係透過將企業之業務風險 (Business Risk Profile) 和財務風險 (Financial Risk Profile) 結合形成定錨等級 (Anchored Rating)，並透過不同之調整機制形成公司獨立之評等 (Stand-Alone Credit Profile)，再將國家、政府或集團納入考量因素，最終形成企業之最終信用等級 (Issuer Credit Rating)。惟信用評等之評估過程是一個高成本且複雜過程，且通常需要數個月才能決定最終信用等級，故信評機構之評等基本上屬於「落後指標」。因此，過往的信評研究方向皆是以過往的會計或股票市場數據中建立一個更準確的信用風險評估模型，並將其模型應用於其他企業信用風險事件中。

綜上所述，過往的信用評等相關研究多以評等等級預測為主，其中以統計方法為基礎之研究如 Fisher (1959) 利用最小平方法 (OLS) 來預測債券之信用評等；Pinches & Mingo (1973) 則利用多重判別分析 (MDA) 產生線性判別函數將債券之評等等級數據做適當的分類並進行預測；除了債券評等外，Gogas et al. (2014) 和 Karminsky & Khromova (2016) 利用序列機率回歸模型 (Ordered Probit Regression Model) 來預測金融機構之信用評等；Hajek & Prochazka (2017) 則建立樸素貝葉斯 (Naive Bayes) 之模型來預測企業之信用評等；Irmatova (2016) 則使用主成分分析 (PCA) 來預測 30 個國家之信用評等。

以上過往信用評等相關研究都是以債券評等和企業評等為主，雖然債券評等與企業評等是相關且密不可分的問題，但由於債券有擔保品、償債先後順序等問題，加上債券有諸多不一樣的交易條件，這使得債券評等模型變得更難以泛化與分析，且債券的數量遠多於公司的數量，直接限制了其信評數據之可用性，係本文將研究重點放在企業之信用評等預測上，希望能藉由結構化數據 (企業財務會計資訊) 和非結構化數據 (年報文字基礎溝通價值)，建構出一個泛

化能力更佳的企業信評預測模型。

二、公司年報文字基礎溝通價值資訊

過往相關信用風險事件研究大多以會計及股票市場等結構化 (定量) 數據為投入變數來進行信用風險相關事件研究，鮮少有考慮公司財務文本之非結構化數據 (定性)。隨著資料集越來越多元化，財務文本披露作為一種非結構化且定性的數據，其亦代表了一種傳遞重要資訊給大眾的一種工具，如上市公司向美國監管機構 (SEC，證券交易委員會) 提交的年度文件，如 10-K (年報)、10-Q (季報)、8-K (臨時報告)、DEF 14A (致股東會報告書)、及 Conference Call (公司電話會議)中有很大一部分是文本之披露。近期許多研究也表明，企業之文本披露內容中包含了信用風險相關的資訊，如 Mayew et al. (2015) 之研究表明 10-K 中的 MD&A 包含了信用風險相關的前瞻性資訊。此外，Mayew et al. (2015) 同時建立了關於公司未來成長動能和經營能力相關的字典，來對企業未來的破產風險進行更深入的分析探討。而除了 10-K 中的 MD&A 之外，Campbell et al. (2014) 和 Loughran & Mcdonald (2011) 之研究亦表明企業所披露的其他財務文本亦包含了信用風險相關的前瞻性資訊; Ganguin & Bilardello (2005) 亦建議信評機構定期從電話會議 (Conference Call) 中蒐集和處理結構化 (定量) 和非結構化 (定性) 的資訊，以更精確的評估企業未來的信用風險。

過往文獻結合年報文字特徵資訊及機器學習模型的研究多著重於公司破產預測，如 Mai et al. (2019)及 Chen et al. (2023)。Mai et al. (2019) 利用年報中 MD&A 的詞頻作為新增投入變數並以深度學習模型進行企業破產預測之研究，其研究結果顯示 MD&A 的文字特徵可提升模型的破產預測準確率；而 Chen et al. (2023) 則利用整份年報的可讀性及語意來作為新增投入變數並以機器學習模型來進行企業破產預測，實證結果發現年報的可讀性及語意變數可有效提升企業破產預測效力。這都顯示年報文字資訊可以補充會計和股票市場資訊所捕捉不到的企業信用風險資訊。

此外，Donovan et al. (2021) 利用 10-K 中的 MD&A 及公司電話會議之文字特徵，先將前述之文本直接投入於深度學習模型並以信用利差 (Credit Spread)

當作目標變數進行訓練，再對未來信用風險事件如信用評等降級、企業破產、債券利率之關聯性進行分析和探討，而其代理變數乃為基於會計和股價基礎之相關變數，如 Altman Z-score，Ohlson O-score，EDF 等信用風險代理變數。研究結果顯示，電話會議和 MD&A 之文字特徵與信用評等降級、企業破產 (違約)、公司合約利差等信用風險事件有顯著相關。

由上述可知，公司財務文本特徵在預測信用風險上比傳統上僅考慮會計和股票市場變量在模型績效上有顯著的改進。因此，為了捕抓更多來自公司財務文本的資訊，本研究將 10-K 之全部文本內容之文字特徵納入投入變數。然而，在過往量化非結構化數據中，大多以詞頻 (TF-IDF)、可讀性 (Readability)、語意 (Tone/Sentiment) 作為文本分析之指標。其中，依據 Seebeck and Kaya (2023) 的定義，財務報告之可讀性 (Readability) 及語意 (Tone/Sentiment) 乃為財務報告溝通價值的衡量變數。是故，本研究將依照 Seebeck and Kaya (2023) 的定義，以年報的可讀性和語意作為年報文字基礎溝通價值的代理變數。

三、機器學習應用於信用風險事件預測效力之研究

機器學習與傳統的統計方法之相似處都是藉由過往大量的數據進行建模分析，與傳統的統計方法相比，機器學習更注重於「預測」之效力上，而統計方法則是注重於解釋於變數與變數之間的關係，近年來由於機器學習模型的興起，許多研究結果亦表明機器學習模型對於信評預測之效力比傳統統計模型還要來得佳，Tsai & Chen (2010) 亦發現提升信評預測準確率對於企業會有顯著的影響，其研究表明當企業信用評等之預測準確率提高 1%時，金融機構和政府的損失會顯著降低。因此，信評預測模型的建立方式是影響準確率非常重要的因素，而如何建立「即時」且「準確」的企業信用評等預測模型，是本研究所關注的重點之一，故本文在模型採用上會以機器學習為基礎模型，並進行企業信用評等之預測。

早期在信評預測上之模型大都以統計方法為基礎，例如 Fisher (1959) 的最小平方法 (OLS)、Pinches & Mingo (1973) 的多重判別分析 (MDA)、Laitinen (1999)；Reichert et al. (1983) 和 Thomas (2000) 的邏輯回歸 (Logistic Regression)

和線性判別分析 (LDA)、Friedman (1991) 的多元自適應回歸 (MARS)、Gogas et al. (2014) 和 Karminsky & Khromova (2016) 的序列機率回歸模型 (Ordered Probit Regression Model) 等，但由於統計模型有分配和假設上的侷限性，如變數之線性假設和資料分配假設等是。Huang et al. (2004) 和 Pai et al. (2015) 之研究亦證實這些統計上的假設會明顯侷限模型的預測效力，這是因為在真實世界中，資料不太可能為線性分佈；此外，近年來模型之投入變數也從結構化數據增加到非結構化數據，而非結構數據亦難以統計模型進行分析和預測，這都讓機器學習模型較之統計模型能應用於更廣泛的資料集且有更佳的模型泛化及預測能力。再者，機器學習模型不但可利用誤差進行反覆訓練、進行樣本內外之測試，其亦是一種非線性模型，可以打破傳統統計模型之線性假設。是故，機器學習比傳統統計方法將有更佳的預測效力。因此，近年來許多信評預測研究模型大都採用機器學習作為基礎模型，如 Golbayani et al. (2020) 及 Wang and Ku (2021)。

為了建立「即時」且「準確」的信評預測模型，本研究將利用隨機森林、極限梯度提升、支持向量機、及羅吉斯回歸等機器學習演算法來建立企業信評預測模型；然而，過往信評預測研究大都注重於準確率上的改善，而忽略了重要的型一和型二錯誤；然而，對於金融授信實務而言，型一型二錯誤和信評高估比例的改善將會比準確率改善的經濟意涵來得重要許多，並對外部利害關係人有更好的貢獻。因此，本研究會更注重在其他模型績效指標的改善上，如召回率 (Recall)、精確率 (Precision)、及 F1 分數 (F1-Score) 等是。

參、研究方法

一、研究架構

本文之研究流程主要分成五個步驟。第一，分別從 Compustat 和 SEC Analytic Suite 資料庫蒐集企業財務會計比率數據和年報文字基礎溝通價值變數 (可讀性、語意) 數據作為模型投入變數，並採用標準普爾公司 (S&P) 企業信用

評等資料作為目標值。第二，進行資料預處理，包含去除空缺值資料、將資料進行標準化、依照特徵重要性刪除相關性較高 (大於 0.7) 之變數、及處理資料不平衡問題等步驟。第三，以年份切割資料，並以逐年滾動方式 (Rolling Window Approach) 進行模型訓練。第四，建立各式機器學習分析模型。最後，利用相關模型績效指標評估預測結果。

二、資料與變數

（一）目標變數

本研究主要目的係建構一個即時且準確的企業信用評等模型，故本研究之目標變數採用標準普爾公司 (S&P) 之企業信用評等之資料，其資料範圍涵蓋 1994 年至 2017 年，S&P 之評等依等級高至低依序為 AAA、AA+、AA、AA-、A+、A、A-、BBB+、BBB、BBB-、BB+、BB、BB-、B+、B、B-、CCC+、CCC、CCC-、CC、SD、D 以細分各等級之信用風險高低，其中以 BBB-等級作為劃分，以上 (含) 為投資等級，以下則為投機等級。投資等級之評等代表企業面臨信用風險相對較低，亦代表企業有較健全的財務體質與營運狀況；而投機等級之評等，則表示企業面臨較嚴重之信用風險或其企業破產 (違約) 可能性較高。綜上所述，本研究之最終目標係以機器學習模型來預測企業未來的信用評等，並希望藉由模擬 S&P 之評等方法，以提供外部利害關係人更即時且精確的評等資訊。然為精簡預測類別，本研究將 S&P 信用評等分為六大類，分別為 Class A (AAA、AA+、AA、AA-)、Class B (A+、A、A-)、Class C (BBB+、BBB、BBB-)、Class D (BB+、BB、BB-)、Class E (B+、B、B-)、及 Class F (CCC+、CCC、CCC-、CC、SD、D)。其資料來源為 Compustat 及 Capital IQ 資料庫。

（二）年報文字基礎溝通價值投入變數

本研究依循 Seebeck and Kaya (2023) 及 Chen et al. (2023) 的定義，以年報可讀性 (Readability) 及內容語意 (Tone) 作為年報文字基礎溝通價值的代理變數。而上述年報文字特徵之數據乃來自 SEC Analytic Suite 資料庫，共計 37 個

文字特徵變數。其中，公司年報 Bog index 的資料乃自 Professor Brian P. Miller 的個人網站中取得(註 2)。

（三）財務會計比率投入變數

在過往信用風險預測文獻中，投入變數多以財務會計比率和股價相關變數為主要投入變量。因此，本研究依循 Mai et al.(2019) 和 Wang and Ku (2021) 所使用的財務變數，選用了其中 56 個財務會計及股價相關比率做為投入變數，且其資料皆來自於 Compustat 資料庫。

三、資料預處理

本研究之資料預處理之步驟包含評等資料分類、處理資料空缺值、資料正規化、依照特徵重要性刪除相關性較高之投入變數、資料不平衡處理、及採用逐年滾動方式切割訓練與測試資料。經由資料預處理步驟可以確保資料之完整性和一致性，以利後續將變數投入機器學習模型。本小節將詳細介紹本研究之資料預處理之步驟與方法。

（一）評等資料分類

本研究旨在於探討企業年度報告中之年報文字基礎溝通價值變數是否能增進公司信用評等預測效力。由於 S&P 評等資料包含 AAA 到 D 一共有 22 個評等類別，為加強機器學習模型之預測效力，本研究依循 Wang and Ku (2021) 的作法，將企業的信用評等進行分類，以減少預測誤差。然而，不同於其研究將信用評等分成四類，本研究欲進行更深入之分析，故以 S&P 信評之風險等級作為區分標準，並將其信用評等分為六大類(註 3)。因此，不同於過往企業破產二元分類之研究問題，本研究乃屬於多分類之問題。圖 1 及圖 2 為分類前後評等資料之分布，由此亦可以發現資料極度不平衡。

註 2：資料來源：https://sites.google.com/iu.edu/professorbrianpmiller/bog-data

註 3：如前所述，本研究將 AAA、AA+、AA、AA-歸類為 Class A ; A+、A、A-歸類為 Class B ; BBB+、BBB、BBB-歸類為 Class C ; BB+、BB、BB-歸類為 Class D ; B+、B、B-歸類為 Class E ; CCC+、CCC、CCC-、CC、SD、D 歸類為 Class F。

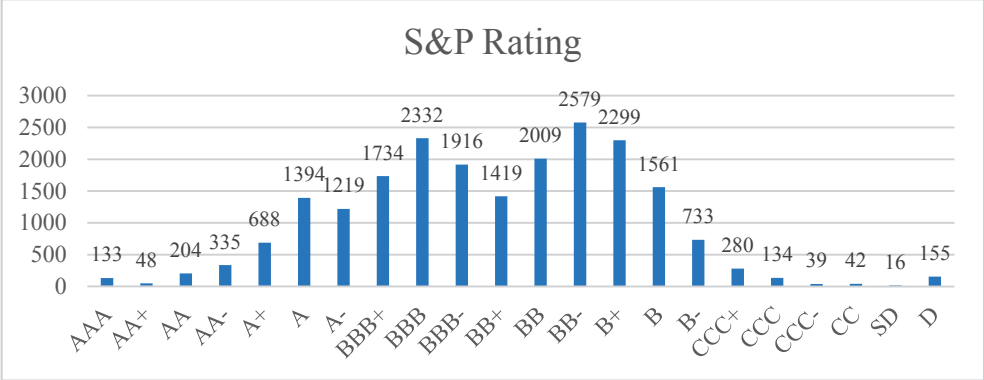

圖 1　分類前 S&P 評等資料之分布

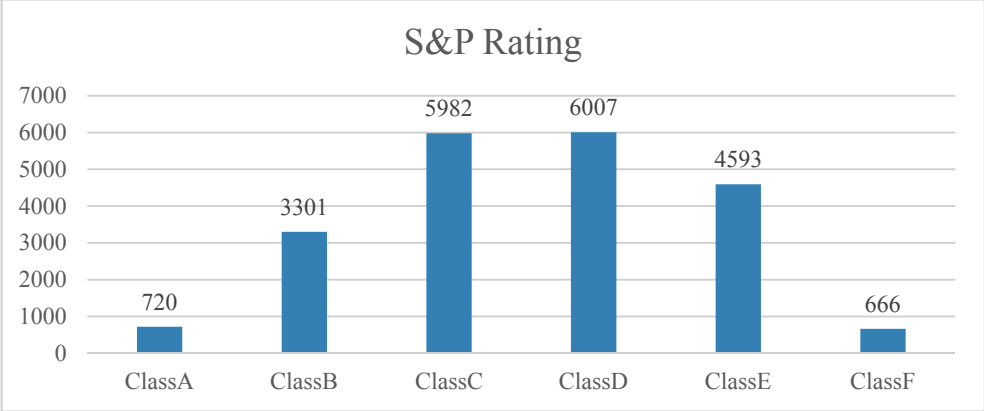

圖 2　分類後 S&P 評等資料之分布

（二）缺失值處理

由於 Compustat WRDS 和 SEC Analytic Suite 資料庫皆有些許缺失值，本研究首先將財務及會計資料、年報文字基礎溝通價值變數資料、和 S&P 評等資料合併，而為確保資料之完整性以及一致性，本研究將會計年度任何缺失值之資料完全刪除，處理後之資料將不會有缺失值之問題。

需注意的是，本研究之所以採取「刪除全部缺失值方式」的原因在於：某些特定財務變數乃為信評公司在評估信用評等時的重要考量因素，如流動比率、速動比率等。因此，採取刪除缺失值方式是較符合信評公司實際進行評等決策的作法。本研究依循過往相關文獻 (e.g. Wang and Ku, 2001) 蒐集與信用評

等相關之財務變數，而在原始擁有信用評等之 52,192 筆樣本資料觀察值中，除了流動資產及流動負債的缺失值比例為 28.57%外，其他主要的財務變數缺失值均不高 (約低於 10%)。而在刪除主要財務變數及年報文字基礎溝通價值變數的缺失值後的最終樣本資料數為 21,269 筆觀察值。

而在刪除變數缺失值前後的各評等群組樣本分布變化如下：在刪除缺失值前，各評等群組 (AA-以上、A、BBB、BB、B、CCC+以下) 佔原始總樣本的比例分別為 6%、21%、32%、20%、17%、3%；而在刪除缺失值後，各評等群組 (AA-以上、A、BBB、BB、B、CCC+以下) 佔刪除後樣本的比例分別為 3%、16%、28%、28%、22%、3%。因此，在刪除缺失值後，非投資級樣本比例稍增，而投資級樣本比例稍減。整體而言，各評等群組的分布在刪除缺失值前後的差異並不大，且各群組樣本在以「投資級與否」為分界下的分布亦更為對稱。

（三）資料正規化

在過往之機器學習分析研究時，不同的投入變數會可能會因為單位不同、數值大小不同、資料變異程度不同而有不同的代表意義，為避免此類問題影響模型訓練過程，本研究採用最小值與最大值正規化 (Min-Max Normalization) 將每一個投入變數依照其最小值和最大值調整至 0 和 1 之間。以下為小值與最大值正規化 (Min-Max Normalization) 之公式，其中 X_min 和 X_max 為其投入變數之最小值與最大值。

$$\boldsymbol{X_{nom}} = \frac{\boldsymbol{X - X_{min}}}{\boldsymbol{X_{max} - X_{min}}} \in [\mathbf{0}, \mathbf{1}] \tag{1}$$

（四）資料不平衡處理

在機器學習分類問題中，當資料集之目標值 (Target Value) 的資料類別比例差距過大 (Major Class vs. Minor Class) 或在處理多類別問題時，其多類別資料分布極度不平均，以上情況即可稱為「不平衡之資料 (Imbalanced Dataset)」。由上節中圖 1 和圖 2 可知，在分析企業信用評等時，評等較高之 Class A 與評等較低 Class F 佔所有樣本之比例來的極低，倘若不做任何處理，機器學習模

型將會傾向學習到資料比例較高的中間評等群組 (如 Class B、Class C、Class D、Class E)，導致過度擬合 (Overfitting) 之問題發生，進而影響模型之績效表現。故本研究採用 RandomOverSampling、SMOTE (Synthetic Minority Over-sampling Technique)、BalancedBaggingClassifier 作為資料不平衡之處理方法。

（五）資料切割

由於 S&P 信用評等資料具有時間序列的性質，本研究以時間點作為切割訓練集及測試集的基準，並採逐年滾動方式來切分訓練集與測試集資料。亦即，在本研究全樣本資料期間 (1994 年至 2017 年) 中，乃以 1994 年至 2011 年當作訓練資料，2012 年當作測試資料；其後以 1994 年至 2012 年當作訓練資料，2013 年當作測試資料，再以此類推，直到最後測試資料年份為 2017 年。在此法下，每一個機器學習模型將會有六個結果 (2012 年至 2017 年六個時間段)，最後以平均的方式來計算未來一年信用評等預測的結果。較之隨機切割分組，本研究資料切割方式的優點如下：(1)可避免以未來資訊預測過去的不合理現象發生；(2)可有效補抓研究變數在時間趨勢上的軌跡，未來亦可進一步延伸至深度學習模型。

（六）特徵選取

在過往機器學習研究中，一個資料集可會有非常多的投入變量，如何選擇與模型目標值最相關之特徵變量以避免維度詛咒 (Curse of Dimensionality) 是近年來機器學習關注的議題。本研究採用兩種特徵選取作法，一為隨機森林特徵重要性，二為極限梯度提升特徵重要性。在兩種做法下，隨機森林最後篩選出之變數涵蓋較多文字特徵變數，其模型最終效力也較極限梯度提升特徵重要性來的佳，故本研究最終選擇隨機森林演算法來對財務變數與年報文字基礎溝通價值特徵變數進行重要性排序，進一步篩選而得最終的模型投入變數。其中，隨機森林是先計算每個特徵變量在每棵樹的貢獻度 (以資訊增益 (Information Gain) 來衡量)，其後計算每棵樹之平均值而得到每個特徵變量的相對貢獻度。

本研究之特徵變量 (投入變數) 包含 56 個財務會計比率變數和 37 個年報文

字基礎溝通價值特徵變數，共 93 個投入變數。為避免維度詛咒與共線性問題，本研究首先針對相關係數大於 0.7 之投入變數，依照隨機森林之特徵重要性刪除較不重要之特徵變數，最後得出 25 個財務會計比率變數與 15 個年報文字基礎溝通價值特徵變數，共 40 個特徵變數。表 1 及表 2 分別為隨機森林特徵重要性分佈圖與本研究最終挑出之變數與其定義。

表 1　隨機森林特徵選取模型篩選結果：財務特徵變數

財務特徵變數	變數定義
GP_REVT	Gross Profit / Revenue
DVPSX_C	Dividends per Share - Ex-Date - Calendar
BKVLPS	Book Value Per Share
EMP	Number of Employees
EPS	Earnings Per Share from Operations
FINCF	Financing Activities - Net Cash Flow
CF	Net Cash Flow
CHE_AMV	(Cash and Short-term Investment) /（Market Equity + Total Liabilities）
CH_LCT	Cash / Current Liabilities
EBIT_AT	EBIT / Total Asset
Debt_AT	Total Debts / Total Assets
INVT_SALE	Inventories / Sales
LCT_CH_AT	（Current Liabilities - Cash）/ Total Asset
LCT_LT	Current Liabilities / Total Liabilities
LT_ AMV	Total Liabilities /（Market Equity + Total Liabilities）
LOG_AT	Log（Total Assets）
PB	Market-to-Book ratio
NI_AT	Net Income / Total Asset
LOG_PRCC	Log（Price）
RE_AT	Retained Earnings / Total Asset
WC_AT	Working Capital / Total Assets
CH	Cash
DCL	Debt in Current Liabilities - Total
RE	Retained Earnings
MKVALT	Market Equity

註：表 1 為本研究使用隨機森林特徵選取模型後，依其重要性刪除相關性大於 0.7 之財務特徵 (FIN) 變數，總計 25 個，亦為最終投入四項機器學習模型之 FIN 變數。

表 2　隨機森林特徵選取模型篩選結果：年報文字基礎溝通價值特徵變數

TCV 變數	變數定義
FSIZE	文本文件檔案之大小（單位為 megabytes）
FK_Ease	Flesch_Reading_Ease：文本中子句長度、字尾字首總數、100 個單字中人稱字（Personal References）字數總和計算之文章難易程度（閱讀之舒適度），可讀性指標
CL	Coleman_liau_Index：文本中每一百個單字平均字母數和平均句子數進行計算之可讀性指標，又稱柯爾曼—劉指數
A_WPS	averageWordsPerSentence：文本中平均每句句子有多少單字字數
A_WPP	averageWordsPerParagraph：文本中平均每段章節有多少單字字數
FT_Neg	FinTerms_Negative：文本中來自 Loughran and McDonald 字典內負面情緒相關單字之比例，語意指標
FT_Mweak_C	FinTerms_ModalWeak_count：文本中來自 Loughran and McDonald 字典內語氣微弱相關單字之總數，語意指標
FT_Pos	FinTerms_Positive：文本中來自 Loughran and McDonald 字典內正面情緒相關單字之比例，語意指標
Har_Neg	HarvardIV_Negative：文本中來自 Harvard General Inquirer 字典內負面情緒相關單字之比例，語意指標
FT_LITI_C	FinTerms_Litigious_count：文本中來自 Loughran and McDonald 字典內有關法律訴訟相關單字之總數，語意指標
FT_Mweak	FinTerms_ModalWeak：文本中來自 Loughran and McDonald 字典內有關語氣微弱相關單字之總數，語意指標
FT_LITI	FinTerms_Litigious：文本中來自 Loughran and McDonald 字典內有關法律訴訟相關單字之比例，語意指標
LM_DICT	LM_Master_Dictionary：文本中屬於 Loughran and McDonald 字典相關單字之比例，語意指標
FT_Mstrong	FinTerms_ModalStrong：文本中來自 Loughran and McDonald 字典內有關語氣強勢相關單字之比例，語意指標
BOG	Bog index：文本中以平均句子長度除以最大句長、關乎用詞（專有名詞、主被動語態、隱藏動詞等等）、表達形式字數總和計算之可讀性指標

註：表 2 為本研究使用隨機森林特徵選取模型後，依其重要性刪除相關性大於 0.7 之年報文字基礎溝通價值 (TCV) 變數，總計 15 個，亦為最終投入四項機器學習模型之 TCV 變數。

為進一步瞭解年報文字基礎溝通價值變數相對於財務變數於信用評等預測的重要性為何，本研究將年報文字基礎溝通價值變數加入財務變數後一同以隨機森林方式進行特徵選取，實證結果如表 3 所示。由表 3 的結果可知：前 26

高特徵重要性 (資訊含量) 的變數有 7 項年報文字基礎溝通價值變數，依其重要性排序分別為：FSIZE、FT_Mweak_C、FT_Mweak、FT_Neg、CL、FT_LITI、及 Har_Neg。其中包含 2 項可讀性變數 (FSIZE、CL) 及 5 項弱語氣、負面詞、訴訟性詞語意變數 (FT_Mweak_C、FT_Mweak、FT_Neg、FT_LITI、及 Har_Neg)。結果顯示年報可讀性變數的重要性排序較語意變數為前，惟重要性在前 26 高的語意變數個數較可讀性變數為多。

表 3　財務與年報文字基礎溝通價值變數之特徵選取結果

Rank	Group	feature	Feature_IMP	Rank	Group	feature	Feature_IMP
1	FIN	RE	0.068657	14	FIN	EBIT_AT	0.024709
2	FIN	DVPSX_C	0.062718	15	TCV	FT_Mweak_C	0.024131
3	FIN	MKVALT	0.054966	16	TCV	FT_Mweak	0.023801
4	FIN	RE_AT	0.044395	17	FIN	NI_AT	0.022958
5	FIN	EPS	0.040640	18	FIN	GP_REVT	0.022480
6	FIN	LOG_AT	0.037588	19	FIN	LCT_LT	0.021190
7	FIN	LT_AMV	0.034118	20	TCV	FT_Neg	0.021031
8	FIN	LOG_PRCC	0.031797	21	TCV	CL	0.019612
9	FIN	EMP	0.030618	22	FIN	WC_AT	0.019609
10	FIN	Debt_AT	0.030209	23	TCV	FT_LITI	0.019346
11	FIN	BKVLPS	0.026106	24	FIN	FINCF	0.019324
12	TCV	FSIZE	0.025937	25	FIN	INVT_SALE	0.019075
13	FIN	DCL	0.024892	26	TCV	Har_Neg	0.018273

註：Feature_IMP 代表特徵重要性 (資訊含量)。FIN 及 TCV 分別代表財務面及年報文字基礎溝通價值面之類別。各變數定義請參閱表 1 及表 2。

此外，本研究亦發現在加入年報文字基礎溝通價值變數後，財務變數的重要性排序發生改變，如在 Wang and Ku (2021) 的重要性排序前三名的變數為 Debt_AT (負債比率)、MKVALT (權益市值)、及 EBI (息前淨利)，而在加入年報文字基礎溝通價值變數後，重要性排序前三名的變數為 RE (保留盈餘)、DVPSX_C (每股股利)、及 MKVALT (權益市值)。此外，部分的財務變數的重要性排序亦因加入年報文字基礎溝通價值變數後而往後移動。然而，表 3 的結果亦顯示：對信用評等各群組擁有較高資訊含量者大多仍為財務特徵變數。這是因為 S&P 的企業信用評等準則架構，乃是先根據財務面指標 (經營風險及財務風險) 先定錨一個初步的信用評等水準後，再經過其他非財務資訊調整後而

得最終的信用評等。因此，表 3 所呈現的財務變數與年報文字基礎溝通價值變數之特徵重要性排序現象乃與 S&P 的企業信用評等準則架構相符。

表 4 及表 5 分別為本研究所使用之財務變數與年報文字基礎溝通價值變數之敘述統計量表 (未正規化)。結果顯示：部份財務變數 (如 BKVLPS、FINCF、CF、PB、CH、DCL、RE、MKVALT) 及年報文字基礎溝通價值變數 (FSIZE、FT_LITI_C、A_WPP) 有極端值現象，顯示正規化有其必要性，故本研究將投入變數進行正規化處理可避免極端值對「非決策樹基礎 (non-tree based) 機器學習模型」的實證結果產生干擾影響。

表 4　財務變數之敘述性統計量

Variable	Obs	Mean	Std	Min	Max
GP_REVT	21269	0.320	1.737	-189.486	1.000
DVPSX_C	21269	0.535	1.344	0.000	95.500
BKVLPS	21269	16.432	1029.866	-71137.000	132253.000
EMP	21269	27.295	75.875	0.000	2300.000
EPS	21269	1.580	5.586	-229.670	505.086
FINCF	21269	-233.465	1842.928	-57705.000	33443.400
CF	21269	51.164	918.622	-42930.000	50435.000
CHE_AMV	21269	0.057	0.070	0.000	1.133
CH_LCT	21269	0.432	0.964	0.000	61.719
EBIT_AT	21269	0.081	0.118	-5.537	1.066
Debt_AT	21269	0.370	0.244	0.000	8.150
INVT_SALE	21269	0.100	0.584	0.000	82.250
LCT_CH_AT	21269	0.152	0.172	-0.920	6.337
LCT_LT	21269	0.346	0.193	0.001	1.000
LT_ AMV	21269	0.472	0.218	0.009	1.000
LOG_AT	21269	3.509	0.620	1.197	5.647
PB	21269	2.967	90.257	-4027.244	7071.350
NI_AT	21269	0.022	0.174	-5.779	6.754
LOG_PRCC	21269	1.372	0.490	-3.000	3.240
RE_AT	21269	0.070	0.595	-16.862	1.983
WC_AT	21269	0.123	0.188	-6.157	0.927
CH	21269	596.490	1683.757	0.000	53528.000
DCL	21269	448.012	4055.526	0.000	257764.000
RE	21269	2085.724	11386.530	-102362.000	398278.000
MKVALT	21269	10749.201	30363.533	0.007	790050.098

註：以上為本研究所使用之財務變數的敘述性統計量 (未正規化)，經資料預處理後無包含缺失值，共計 21,269 筆資料。變數定義請參閱表 1。

表 5　年報文字基礎溝通價值變數之敘述性統計量

Variable	Obs	Mean	Std	Min	Max
FSIZE	21269	7969793	13102868	56010	434657609
FK_Ease	21269	25.279	3.930	-15.803	46.921
CL	21269	22.413	0.852	19.306	31.541
A_WPS	21269	24.796	3.350	8.302	73.510
A_WPP	21269	133.445	3903.846	10.816	475998
FT_Neg	21269	0.015	0.005	0.002	0.050
FT_MWeak_C	21269	223.271	173.128	1	2343
FT_Pos	21269	0.008	0.001	0.0010	0.026
Har_Neg	21269	0.041	0.007	0.0063	0.066
FT_LITI_C	21269	585.481	785.813	15	18018
FT_MWeak	21269	0.005	0.002	0.0002	0.035
FT_LITI	21269	0.012	0.006	0.002	0.065
LM_DICT	21269	1	0	1	1
FT_Mstrong	21269	0.002	0.0009	0.0002	0.018
BOG	21269	84.278	7.427	51	139

註：以上為本研究所使用之年報文字基礎溝通價值變數的敘述性統計量 (未正規化)，經資料預處理後無包含缺失值，共計 21,269 筆資料。變數定義請參閱表 2。

四、機器學習模型

根據過往信用風險預測的相關文獻 (Barboza et al., 2017; Mai et al., 2019; Wang and Ku, 2021; Chen et al., 2023) 可知，隨機森林、支持向量機、羅吉斯回歸、極限梯度提升等模型為最經常使用的機器學習模型。因此，本研究依循過往相關研究，使用上述四類機器學習模型作為信用評等預測的基礎模型。以下將介紹四類機器學習模型的概念。

（一）隨機森林(Random Forest, RF)

隨機森林演算法 (Breiman,2001) 乃由多種決策樹 (Decision Tree) 所建構而成之模型，並在各棵決策樹生成結果後投票。而在每一層做決策時，皆會計算每個節點的資訊增益 (Information gain，註 4)，並以資訊增益較大的特徵作為

註 4：資訊增益(Information gain)乃定義為資訊量(Entropy)扣減分類完的加權平均資訊量。

分類依據。而隨機森林乃基於 Bagging (Bootstrap Aggregation) 的概念來重複取樣並建構資料子集及建模預測，最終以多數決的方式決定分類結果。而其主要優點為可在不調整超參數下作出模型之合理預測。

（二）支援向量機(Support Vector Machine, SVM)

SVM (Support Vector Machine, Vapnik, 1963) 屬於監督式的機器學習模型，其概念在於找出可區分兩種類別之間的最大化之決策邊界 (Decision Boundary)。一般來說，SVM 可以是以線性 (二維) 或超平面 (多維) 的方式來分割資料。若欲分析的資料集為線性不可分時，可採用 SVM-RBF (Radial Based Function，高斯核函數) 來處理。而在破產預測相關文獻中，使用 SVM-RBF 來區分破產及非破產公司的成效均優於 SVM-Linear (Min and Lee, 2005；Barboza, et al., 2017)，故本研究依循以往文獻利用 SVM-RBF 來對信用評等分類預測進行分析。

（三）羅吉斯回歸(Logistic Regression)

羅吉斯回歸乃基於統計上的最大概似法，其目標變數可為二元或多元類別變數。本研究中所使用的類別變數即為信用評等類別變數 (亦即信用評等由高至低分別以 1 至 6 表示，註 5)，而相較其他機器學習模型，其算法較為易懂，而其生成結果亦較容易解釋。其模型可如式(2)及(3)所示：

$$f(x)_\theta = \frac{1}{1+e^{-x}} \tag{2}$$

$$\mathrm{P}(h(x)_\theta = 1,2,\dots,6|\theta_0,\dots.,\theta_n) = \frac{1}{1+e^{-(\theta_0+\theta_1 x1+\cdots..+\theta_N x_N)}} \tag{3}$$

在式(2)及(3)中，x 及θ分別代表各項投入變數及其權重值；而式(3)乃基於 Sigmoid function 下所得的各類別信用評等預測之機率值(P)。此外，透過損失函數之建立，並藉由誤差之最小化可求得最佳參數 θ。

註 5：信用評等類別變數之變數值以 1 至 6 分別代表 AA-以上、A、BBB、BB、B、CCC+以下等六大信用評等類別。

（四）極限梯度提升(XGBoost)

極限梯度提升(Extreme Gradient Boosting (XGBoost), Chen and Guestrin, 2016) 乃基於梯度提升決策樹 (Gradient Boosted Decision Tree, GBDT)，屬於 Boosting 算法之一，其原理即為將許多弱分類器聚合為一個強分類器；並會在過程中提高舊分類器的資料錯誤權重，接著再訓練新分類器。故過程中被錯誤分類資料之特性可被新分類器及後續的訓練學習，進而達到改善的效果。XGBoost 模型的特性包含：過程中持續以梯度下降的方式來極小化殘差、加入正則項當作懲罰值以避免發生模型過擬和 (Overfitting) 的問題。其樹模型為 CART 迴歸樹模型，如式(4)所示。其中，$y\tilde{\imath}$ 代表模型預測結果，N 及 f_n 分別為決策樹總個數及第幾棵決策樹。

$$\hat{y}_i = \sum_{n=1}^{N} f_n(x_i) \tag{4}$$

極限梯度提升模型之目標函數如(5)所示。其中，k 及 p 分別表示訓練集中樣本之數量及建構的第 p 棵樹。而式(5)函數的前半部 (ℓ) 為損失函數 (Loss function)，可衡量實際及預測結果之差異。而後半部為正則化項，可透過超參數 γ 及 λ 之調整來決定懲罰值水準，避免過擬和 (Overfitting) 之問題。其中，T 及 w 分別代表樹之規模大小 (葉子節點個數) 及葉子節點之權重。而 $\hat{y}_i$ 涵蓋 $f_p(x_i)$ (目前樹模型之值) 及 $\hat{y}_i^{(p-1)}$ (之前樹的預測結果)，可修正過去每一棵樹的殘差和新加入樹之殘差。

$$Obj^{(p)} = \sum_{i=1}^{k} \ell(y_i, \hat{y}_i^{(p-1)} + f_p(x_i)) + \Omega(f_p) \tag{5}$$

$$\text{Where } \Omega(f_p) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \text{ ，}$$

五、模型績效指標

本研究依循過往文獻 (e.g. Chen et al., 2023) 使用混淆矩陣 (Confusion Matrix) 來評估機器學習分類問題的模型績效。由於本研究將信用評等依照 S&P 風險等級分為六大類評等 (Class A～Class F)，故本研究之混淆矩陣為 6 x 6 之矩陣，如表 6 所示。其矩陣每一維度中有一樣的資料目標值之類別，其結果可分為真陽性 (True Positive, TP)、偽陽性 (False Positive, FP)、真陰性 (True Negative, TN)、偽陰性 (False Negative, FN) 四種類別。而其中偽陽性 (False Positive, FP) 和偽陰性 (False Negative, FN) 又分別為型一錯誤 (Type I Error) 和型二錯誤 (Type II Error)，透過混淆矩陣之評估指標為準確率 (Accuracy)、精確率 (Precision)、召回率 (Recall)、F1 分數 (F1-Score)，如以下列各式所示。

表 6　多類別混淆矩陣示意圖

Predict Class / Actual Class	$C_0, \dots, C_{k-1}$	C_k	$C_{k+1}, \dots, C_n$
$C_0, \dots, C_{k-1}$	True Negative	False Positive	True Negative
C_k	False Negative	True Positive	False Negative
$C_{k+1}, \dots, C_n$	True Negative	False Positive	True Negative

（一）準確率(Accuracy)

分類模型所有預測正確的結果佔總資料集之比例。

$$Accuracy = \frac{(TP + TN)}{(TP + FN + FP + TN)}$$

（二）精確率(Precision)

分類模型預測是 Positive 的所有結果中，被分類模型預測正確之比例。

$$Precision = \frac{TP}{TP + FP}$$

（三）召回率(Recall)

實際值是 Positive 的所有結果中，被分類模型預測正確之比例。

$$Recall = \frac{TP}{TP + FN}$$

（四）F1 分數(F1-Score)

F1-Score 為精確率與召回率之調和平均數，為一種加權平均數。

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

由於本研究為多分類之研究議題，每一種類別都有其精確率 (Precision)、召回率 (Recall) 和 F1 分數 (F1-Score) 之數值，為考慮機器學習模型之綜合效能 (同時包含每個類別)，故本研究採用三種不同加權方法作為評估模型績效基準，分別為 Weighted-Average、Micro-Average 和 Macro-Average，而後續研究結果將以 Macro-Average 方法為主，其原因說明如下：由於本研究之預測對象「評等群組」樣本資料分布過於不平衡，亦即評等好及評等差兩群組的樣本數較少，且中間評等群組的樣本數較多。若以 Weighted-Average 來設算績效指標，則模型預測效力結果將為中間評等群組所決定，而無法充分反映評等好及評等差等兩群組的預測效力結果。而 Macro-Average 乃為簡單平均，故能充分反映評等好及評等差等兩群組的預測效力結果。因此，本研究將採用 Macro-Average 作為主要模型績效評比方式。

再者，由於本研究將信用評等分類為 6 個群組，錯誤預估評等類別的樣本可能被預測到其他 5 組信用評等類別，且由於類別間存在優劣順序之分，故此時的錯誤預估亦將存在高估或低估的現象。而不同信用評等群組的高低估嚴重性亦不一 (如將低評等的公司誤判為高評等的嚴重性較之將高評等的公司誤判為低評等的嚴重性為高)。由於 F1 score 為各信用評等群組的精確率 (Precision) 及召回率 (Recall) 的調和平均數，故同時考量了目標信用評等群組及其他信用

評等群組的正確分類及錯誤分類之相關資訊，因此，本研究以 F1 score 作為信用評等分類預測的績效評估指標具有其適當性。而為了進一步解析模型高低估的現象，本研究亦進一步利用混淆矩陣各評等群組的正確預測及錯誤預測比例資訊來進行補充比較分析。

肆、研究結果

本研究乃利用 1994 年至 2017 年的美國公司信用評等資料來探討機器學習模型在加入年報文字基礎溝通價值變數後是否能提升原模型 (僅考慮財務變數) 的預測效力。其中，本研究所使用的機器學習模型分別為隨機森林、支持向量機、極限梯度提升、和羅吉斯回歸；而資料不平衡處理方法包含 RandomOverSampling、SMOTE 及 BalanceBaggingClassifier。而為強化分析結果的穩健性，本研究採用逐年滾動方式來訓練集與測試集資料之切分，亦即以 1994 年至 2011 年為訓練組，逐年滾動式地來預測未來一年的公司信用評等，最後將上述六項結果平均而得最終預測結果。

以下將分別介紹在兩種資料不平衡處理方法下四種機器學習模型的績效結果。其中，混淆矩陣數值以百分比表示，表示在該實際類別下，有多少百分比被預測為該評等 (混淆矩陣斜對角線) 或其他評等 (非斜對角線)。因此，混淆矩陣每一列之數值加總為 100%，且對角線的數值越高，代表預測正確比例越高，亦即模型績效越好；而若非對角線數值越低，代表被錯誤預測比例越低，模型績效較佳。

一、RandomOverSampling 方法處理 S&P 評等之資料

表 7 及表 8 分別為以 RandomOverSampling 方法來處理資料不平衡下之各機器學習模型的分析結果與混淆矩陣。以 Accuracy 來說，在加入年報文字基礎溝通價值(可讀性及語意)變數後，四個模型皆有顯著的提升，隨機森林模型提升 4.56%、支持向量機模型提升 12%、極限梯度提升模型提升 5.44%、羅吉斯回歸模型提升 15.05%，其中準確率最高之模型為隨機森林和極限梯度提升

模型。且其再加入年報文字基礎溝通價值變數後，準確率分別可以高達 76.35%和 76.89%。除了準確率提升以外，其他指標如 Precision、Recall 和 F1-Score，皆有顯著的提升，以 Macro-Average 方法來看，四個模型 (隨機森林、支持向量機、極限梯度提升、羅吉斯回歸) 的 Precision 分別提升 5.20%、7.60%、6.56%、9.54%，Recall 分別提升 2.32%、8.06%、3.02%、8.59%。Precision 和 Recall 的比率提高，直接代表了 F1-Score 的改善，這是因為 F1-Score 為前兩者之調和平均數，代表機器學習模型的泛化能力。其中，四個模型 (隨機森林、支持向量機、極限梯度提升、羅吉斯回歸) 的 Macro F1-Score 分別為 74.32%、54.17%、75.09%、及 51.53%，亦即加入年報文字基礎溝通價值變數後可分別提升 4.14%、10.15%、4.77%、及 12.3%。

除上述模型績效指標外，本研究更關注於評等被錯誤分類的情況，亦即實際上評等較低的類別會被模型預測為評等較高的類別之情況。尤其是實際為投機級評等 (Class D 以下等級)，但會被機器學習模型預測為投資級評等 (Class C 以上等級) 之情況，此種高估情況在信用評等預測上屬於比較嚴重的錯誤。本研究乃以模型的混淆矩陣斜對角線左下角區域來進行觀察分析，亦即實際評等被模型預測高估的部分 (預測評等高於實際評等)。由表 8 的結果可知，若以四種機器學習模型分別來看，在加入年報文字基礎溝通價值變數前，隨機森林模型的 Class D 有 1.99%的比例會被預測為 Class B，有 23.84%的比例會被預測成 Class C。而在加入年報文字基礎溝通價值變數後，其錯誤預測比例分別下降為 1%和 18.6%；此外，在加入年報文字基礎溝通價值變數後，隨機森林模型的 Class E 被錯誤高估為 Class C 的比率由 3.33%降至 0.96%。而在支持向量機模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class A、Class B、Class C 的比率分別由 0.66%、7.28%、43.38%分別降至 0%、2.33%、29%；而在極限梯度提升模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class B、Class C 的比率分別由 1.66%、23.26%分別降至 0.66%、16.94%；而在羅吉斯回歸模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class A、Class B、Class C 的比率分別由 1%、13.95%、40.2%分別降至 0%、2.66%、29.57%。

表 7　各項機器學習模型之信用評等預測分析結果：
資料不平衡處理方法採用 RandomOversampling

機器學習模型	績效指標	FIN 變數	FIN & TCV 變數
隨機森林 Random Forest	Accuracy	71.79%	76.35%
	Weighted_Precision	72.57%	76.74%
	Weighted_Recall	71.79%	76.35%
	Weighted_F1_score	71.44%	76.03%
	Micro_Precision	71.79%	76.35%
	Micro_Recall	71.79%	76.35%
	Micro_F1_score	71.79%	76.35%
	Macro_Precision	74.64%	79.84%
	Macro_Recall	69.88%	72.20%
	Macro_F1_score	70.18%	74.32%
支持向量機 SVM	Accuracy	44.92%	56.92%
	Weighted_Precision	50.63%	58.58%
	Weighted_Recall	44.92%	56.92%
	Weighted_F1_score	46.05%	57.18%
	Micro_Precision	44.92%	56.92%
	Micro_Recall	44.92%	56.92%
	Micro_F1_score	44.92%	56.92%
	Macro_Precision	44.32%	51.92%
	Macro_Recall	52.95%	61.01%
	Macro_F1_score	44.02%	54.17%
極限梯度提升 XGBoost	Accuracy	71.45%	76.89%
	Weighted_Precision	72.05%	77.11%
	Weighted_Recall	71.45%	76.89%
	Weighted_F1_score	71.29%	76.67%
	Micro_Precision	71.45%	76.89%
	Micro_Recall	71.45%	76.89%
	Micro_F1_score	71.45%	76.89%
	Macro_Precision	72.12%	78.68%
	Macro_Recall	70.74%	73.76%
	Macro_F1_score	70.32%	75.09%
羅吉斯回歸 Logistic Regression	Accuracy	39.97%	55.02%
	Weighted_Precision	47.70%	56.92%
	Weighted_Recall	39.97%	55.02%
	Weighted_F1_score	41.59%	55.24%
	Micro_Precision	39.97%	55.02%
	Micro_Recall	39.97%	55.02%
	Micro_F1_score	39.97%	55.02%
	Macro_Precision	39.95%	49.49%
	Macro_Recall	50.44%	59.03%
	Macro_F1_score	39.24%	51.53%

註：本研究將 S&P 信用評等分為六大類，分別為 Class A (AAA、AA+、AA、AA-)、Class B (A+、A、A-)、Class C (BBB+、BBB、BBB-)、Class D (BB+、BB、BB-)、Class E (B+、B、B-)、及 Class F (CCC+、CCC、CCC-、CC、SD、D)。而財務特徵 (FIN) 變數及年報文字基礎溝通價值 (TCV) 特徵變數分別為表 1 及表 2 所列示之 25 項及 15 項變數。

表 8　各機器學習模型分類預測之混淆矩陣結果 (RandomOverSampling)
（FIN 變數 / FIN & TCV 特徵變數）

Panel A. 隨機森林模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	90.91% / 81.82%	9.09% / 18.18%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	3.91% / 1.57%	82.81% / 81.89%	12.50% / 16.54%	0.78% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class C	0.00% / 0.00%	15.61% / 10.30%	79.73% / 84.72%	4.32% / 4.65%	0.33% / 0.33%	0.00% / 0.00%
Class D	0.00% / 0.00%	1.99% / 1.00%	23.84% / 18.60%	65.23% / 70.10%	8.94% / 10.30%	0.00% / 0.00%
Class E	0.00% / 0.00%	0.00% / 0.00%	3.33% / 0.96%	26.19% / 22.01%	69.52% / 76.08%	0.95% / 0.96%
Class F	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	3.03% / 0.00%	63.64% / 63.64%	33.33% / 36.36%

Panel B. 支持向量機模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	90.91% / 81.82%	9.09% / 18.18%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	35.43% / 18.90%	40.16% / 48.03%	20.47% / 28.35%	3.15% / 3.94%	0.79% / 0.79%	0.00% / 0.00%
Class C	6.33% / 1.67%	39.00% / 22.67%	48.33% / 62.33%	6.00% / 13.00%	0.33% / 0.33%	0.00% / 0.00%
Class D	0.66% / 0.00%	7.28% / 2.33%	43.38% / 29.00%	38.41% / 53.00%	9.60% / 15.00%	0.66% / 0.67%
Class E	0.48% / 0.00%	0.95% / 0.47%	8.57% / 3.32%	34.29% / 25.59%	47.62% / 58.77%	8.10% / 11.85%
Class F	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	6.06% / 0.00%	39.39% / 37.50%	54.55% / 62.50%

Panel C. 極限梯度提升模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	86.96% / 86.96%	13.04% / 13.04%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	4.72% / 2.36%	78.74% / 81.10%	16.54% / 16.54%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class C	0.00% / 0.00%	16.00% / 8.67%	78.00% / 85.00%	6.00% / 6.33%	0.00% / 0.00%	0.00% / 0.00%
Class D	0.00% / 0.00%	1.66% / 0.66%	23.26% / 16.94%	65.78% / 71.76%	9.30% / 10.63%	0.00% / 0.00%
Class E	0.00% / 0.00%	0.47% / 0.00%	2.83% / 1.43%	23.58% / 19.52%	70.75% / 77.14%	2.36% / 1.90%
Class F	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	3.03% / 0.00%	54.55% / 56.25%	42.42% / 43.75%

Panel D. 羅吉斯回歸模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	91.30% / 81.82%	8.70% / 18.18%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	42.19% / 20.47%	34.38% / 44.09%	19.53% / 29.92%	3.91% / 3.94%	0.00% / 1.57%	0.00% / 0.00%
Class C	11.59% / 2.66%	41.72% / 21.26%	38.74% / 60.80%	7.28% / 14.29%	0.66% / 1.00%	0.00% / 0.00%
Class D	1.00% / 0.00%	13.95% / 2.66%	40.20% / 29.57%	33.55% / 49.17%	10.63% / 17.94%	0.66% / 0.66%
Class E	0.48% / 0.00%	1.90% / 0.00%	13.33% / 4.31%	26.19% / 22.97%	45.24% / 59.81%	12.86% / 12.92%
Class F	0.00% / 0.00%	0.00% / 0.00%	2.94% / 0.00%	11.76% / 6.06%	26.47% / 33.33%	58.82% / 60.61%

註：本研究將 S&P 信用評等分為六大類，分別為 Class A (AAA、AA+、AA、AA-)、Class B (A+、A、A-)、Class C (BBB+、BBB、BBB-)、Class D (BB+、BB、BB-)、Class E (B+、B、B-)、及 Class F (CCC+、CCC、CCC-、CC、SD、D)。而財務特徵 (FIN) 變數及年報文字基礎溝通價值 (TCV) 特徵變數分別為表 1 及表 2 所列示之 25 項及 15 項變數。而表格內的「數字比例/數字比例」，前者代表「FIN+TCV 變數作為投入變數下各信用評等群組實際數被預測為六類信用評等群組的個數比例分布」，後者代表「FIN 變數作為投入變數下各信用評等群組實際數被預測為六類信用評等群組的個數比例分布」。

綜合上述各面向分析結果可知，在加入年報文字基礎溝通價值變數 (可讀性及語意) 後，四種機器學習模型在各項績效指標上有明顯的改善，尤其是在高估評等的情況下，預測錯誤比例有下降的趨勢。此類評等高估之預測錯誤的改善對整體模型提升預測效力有重要的貢獻。惟混淆矩陣對角線的 Class A 其比例有少許下降的趨勢，代表再減少高估比例的同時，會造成等級較高的評等被低估的情形，亦指混淆矩陣斜對角右上方之比例增加。其中，極限梯度提升模型在加入年報文字基礎溝通價值變數 (可讀性及語意) 變數後，Class A 預測準確的比例無改變，且低估之比例下降幅度最低，顯示其為最穩定、泛化能力最出色的模型。

二、SMOTE 方法處理 S&P 評等之資料

表 9 及表 10 分別為以 SMOTE (Synthetic Minority Over-sampling Technique) 作為資料不平衡處理方法下，各機器學習模型的分析結果與混淆矩陣。以 Accuracy 來說，在加入年報文字基礎溝通價值 (可讀性及語意) 變數後，隨機

森林、支持向量機、極限梯度提升、及羅吉斯回歸等模型分別提升 6.38%、11.70%、6.50%、及 13.47%，其中準確率最高之模型為隨機森林和極限梯度提升。且其再加入年報文字基礎溝通價值變數後，準確率分別可以高達 76.56%和 76.23%。除了準確率提升以外，其他指標如 Precision、Recall 和 F1-Score，皆有顯著的提升，以 Macro-Average 方法來看，四個模型 (隨機森林、支持向量機、極限梯度提升、羅吉斯回歸) 的 Precision 分別提升 7.55%、7.03%、8.24%、7.85%，Recall 分別提升 3.84%、7.29%、4.34%、7.47%。Precision 和 Recall 的比率提高，直接代表了 F1-Score 的改善，這是因為 F1-Score 為前兩者之調和平均數，代表機器學習模型的泛化能力。其中，四個模型 (隨機森林、支持向量機、極限梯度提升、羅吉斯回歸) 的 Macro F1-Score 分別為 75.63%、53.51%、74.86%、及 50.32%，亦即加入年報文字基礎溝通價值變數後可分別提升 6.18%、9.43%、6.93%、及 10.50%。

此外，由表 10 的結果可知，若以四種機器學習模型分別來看，在加入年報文字基礎溝通價值變數前，隨機森林模型的 Class D 有 1.99%的比例會被預測為 Class B，有 23.59%的比例會被預測成 Class C。而在加入年報文字基礎溝通價值變數後，其錯誤預測比例分別下降為 1%和 17.94%；此外，在加入年報文字基礎溝通價值變數後，隨機森林模型的 Class E 被錯誤高估為 Class C 的比率由 3.35%降至 1.42%。而在支持向量機模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class A、Class B、Class C 的比率分別由 0.66%、7.97%、42.86%分別降至 0%、2.65%、28.81%；而在極限梯度提升模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class B、Class C 的比率分別由 2.33%、23.00%分別降至 0.66%、17.28%；而在羅吉斯回歸模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class A、Class B、Class C 的比率分別由 0.99%、15.23%、38.08%分別降至 0%、2.99%、28.24%。

表 9　各項機器學習模型之信用評等預測分析結果：
資料不平衡處理方法採用 SMOTE

機器學習模型	績效指標	FIN 變數	FIN & TCV 變數
隨機森林 Random Forest	Accuracy	0.7018	0.7656
	Weighted_Precision	0.7107	0.7684
	Weighted_Recall	0.7018	0.7656
	Weighted_F1_score	0.7008	0.7640
	Micro_Precision	0.7018	0.7656
	Micro_Recall	0.7018	0.7656
	Micro_F1_score	0.7018	0.7656
	Macro_Precision	0.6885	0.7640
	Macro_Recall	0.7207	0.7591
	Macro_F1_score	0.6945	0.7563
支持向量機 SVM	Accuracy	0.4493	0.5663
	Weighted_Precision	0.5063	0.5848
	Weighted_Recall	0.4493	0.5663
	Weighted_F1_score	0.4606	0.5697
	Micro_Precision	0.4493	0.5663
	Micro_Recall	0.4493	0.5663
	Micro_F1_score	0.4493	0.5663
	Macro_Precision	0.4422	0.5125
	Macro_Recall	0.5312	0.6041
	Macro_F1_score	0.4409	0.5351
極限梯度提升 XGBoost	Accuracy	0.6973	0.7623
	Weighted_Precision	0.7046	0.7628
	Weighted_Recall	0.6973	0.7623
	Weighted_F1_score	0.6964	0.7608
	Micro_Precision	0.6973	0.7623
	Micro_Recall	0.6973	0.7623
	Micro_F1_score	0.6973	0.7623
	Macro_Precision	0.6719	0.7543
	Macro_Recall	0.7053	0.7487
	Macro_F1_score	0.6794	0.7486
羅吉斯回歸 Logistic Regression	Accuracy	0.4058	0.5405
	Weighted_Precision	0.4828	0.5629
	Weighted_Recall	0.4058	0.5405
	Weighted_F1_score	0.4225	0.5438
	Micro_Precision	0.4058	0.5405
	Micro_Recall	0.4058	0.5405
	Micro_F1_score	0.4058	0.5405
	Macro_Precision	0.4036	0.4821
	Macro_Recall	0.5078	0.5825
	Macro_F1_score	0.3983	0.5032

表 10　各機器學習模型分類預測之混淆矩陣結果 (SMOTE)
（FIN 變數 / FIN & TCV 特徵變數）

Panel A. 隨機森林模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	91.30% / 86.96%	8.70% / 13.04%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	6.30% / 3.13%	80.31% / 82.03%	12.60% / 14.84%	0.79% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class C	0.66% / 0.00%	18.60% / 10.96%	75.75% / 84.39%	4.65% / 4.32%	0.33% / 0.33%	0.00% / 0.00%
Class D	0.00% / 0.00%	1.99% / 1.00%	23.59% / 17.94%	64.12% / 70.10%	10.30% / 10.96%	0.00% / 0.00%
Class E	0.00% / 0.00%	0.00% / 0.00%	3.35% / 1.42%	24.40% / 19.91%	68.42% / 75.36%	3.83% / 3.32%
Class F	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	51.52% / 45.45%	48.48% / 54.55%

Panel B. 支持向量機模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	90.91% / 78.26%	9.09% / 21.74%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	34.65% / 19.69%	40.94% / 48.03%	20.47% / 27.56%	3.15% / 3.94%	0.79% / 0.79%	0.00% / 0.00%
Class C	6.31% / 2.00%	38.54% / 23.33%	48.17% / 61.33%	6.31% / 13.00%	0.66% / 0.33%	0.00% / 0.00%
Class D	0.66% / 0.00%	7.97% / 2.65%	42.86% / 28.81%	38.54% / 52.65%	9.63% / 15.23%	0.33% / 0.66%
Class E	0.48% / 0.00%	0.96% / 0.47%	8.61% / 3.79%	33.49% / 24.17%	47.85% / 59.24%	8.61% / 12.32%
Class F	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	6.06% / 0.00%	39.39% / 37.50%	54.55% / 62.50%

Panel C. 極限梯度提升模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	90.91% / 86.96%	9.09% / 13.04%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	7.81% / 3.10%	76.56% / 77.52%	14.84% / 18.60%	0.78% / 0.78%	0.00% / 0.00%	0.00% / 0.00%
Class C	0.66% / 0.00%	17.94% / 9.00%	75.08% / 83.67%	5.98% / 7.33%	0.33% / 0.00%	0.00% / 0.00%
Class D	0.00% / 0.00%	2.33% / 0.66%	23.00% / 17.28%	64.00% / 70.43%	10.67% / 11.30%	0.00% / 0.33%
Class E	0.00% / 0.00%	0.00% / 0.00%	2.84% / 1.90%	23.22% / 18.10%	70.14% / 76.67%	3.79% / 3.33%
Class F	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	50.00% / 50.00%	50.00% / 50.00%

Panel D. 羅吉斯回歸模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	91.30% / 78.26%	8.70% / 21.74%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	39.84% / 22.48%	35.94% / 42.64%	19.53% / 29.46%	3.91% / 3.88%	0.78% / 1.55%	0.00% / 0.00%
Class C	10.96% / 2.98%	41.86% / 23.84%	39.20% / 57.62%	7.31% / 14.57%	0.66% / 0.99%	0.00% / 0.00%
Class D	0.99% / 0.00%	15.23% / 2.99%	38.08% / 28.24%	33.77% / 48.50%	11.26% / 19.27%	0.66% / 1.00%
Class E	0.47% / 0.00%	2.37% / 0.00%	12.80% / 4.31%	25.12% / 20.57%	46.45% / 59.81%	12.80% / 15.31%
Class F	0.00% / 0.00%	0.00% / 0.00%	3.03% / 0.00%	9.09% / 3.13%	30.30% / 34.38%	57.58% / 62.50%

註：本研究將 S&P 信用評等分為六大類，分別為 Class A (AAA、AA+、AA、AA-)、Class B (A+、A、A-)、Class C (BBB+、BBB、BBB-)、Class D (BB+、BB、BB-)、Class E (B+、B、B-)、及 Class F (CCC+、CCC、CCC-、CC、SD、D)。而財務特徵 (FIN) 變數及年報文字基礎溝通價值 (TCV) 特徵變數分別為表 1 及表 2 所列示之 25 項及 15 項變數。而表格內的「數字比例/數字比例」，前者代表「FIN+TCV 變數作為投入變數下各信用評等群組實際數被預測為六類信用評等群組的個數比例分布」，後者代表「FIN 變數作為投入變數下各信用評等群組實際數被預測為六類信用評等群組的個數比例分布」。

三、BalanceBaggingClassifier 方法處理 S&P 評等之資料

表 11 及表 12 分別為以 BalanceBaggingClassifier 作為資料不平衡處理方法下，各機器學習模型的分析結果與混淆矩陣。在表 11 中，在加入年報文字基礎溝通價 (可讀性及語意) 變數後，隨機森林、支持向量機、極限梯度提升、及羅吉斯回歸等模型的 Accuracy 分別提升 6.62%、15.04%、6.73%、及 14.75%，其中準確率最高之模型為極限梯度提升。且其再加入年報文字基礎溝通價值變數後，準確率可以達到 67.65%。除了準確率提升以外，其他指標如 Precision、Recall 和 F1-Score，皆有顯著的提升，以 Macro-Average 方法來看，四個模型 (隨機森林、支持向量機、極限梯度提升、羅吉斯回歸) 的 Precision 分別提升 5.87%、8.03%、5.98%、8.25%，Recall 分別提升 3.52%、8.05%、3.70%、8.33%。Precision 和 Recall 的比率提高，直接代表了 F1-Score 的改善，這是因為 F1-Score 為前兩者之調和平均數，代表機器學習模型的泛化能力。其中，四個模型 (隨機森林、支持向量機、極限梯度提升、羅吉斯回歸) 的 Macro F1-Score 分別為 64.53%、51.41%、66.55%、及 49.35%，亦即加入年報文字基礎溝通價值變數後可分別提升 6.38%、11.74%、6.12%、及 11.59%。

表 11　各項機器學習模型之信用評等預測分析結果：
資料不平衡處理方法採用 BalancedBaggingClassifier

機器學習模型	績效指標	FIN 變數	FIN & TCV 變數
隨機森林 Random Forest	Accuracy	0.5807	0.6469
	Weighted_Precision	0.6155	0.6635
	Weighted_Recall	0.5807	0.6469
	Weighted_F1_score	0.5850	0.6482
	Micro_Precision	0.5807	0.6469
	Micro_Recall	0.5807	0.6469
	Micro_F1_score	0.5807	0.6469
	Macro_Precision	0.5616	0.6203
	Macro_Recall	0.6702	0.7054
	Macro_F1_score	0.5815	0.6453
支持向量機 SVM	Accuracy	0.3891	0.5395
	Weighted_Precision	0.4755	0.5612
	Weighted_Recall	0.3891	0.5395
	Weighted_F1_score	0.4093	0.5436
	Micro_Precision	0.3891	0.5395
	Micro_Recall	0.3891	0.5395
	Micro_F1_score	0.3891	0.5395
	Macro_Precision	0.4134	0.4937
	Macro_Recall	0.5019	0.5824
	Macro_F1_score	0.3967	0.5141
極限梯度提升 XGBoost	Accuracy	0.6092	0.6765
	Weighted_Precision	0.6373	0.6894
	Weighted_Recall	0.6092	0.6765
	Weighted_F1_score	0.6127	0.6779
	Micro_Precision	0.6092	0.6765
	Micro_Recall	0.6092	0.6765
	Micro_F1_score	0.6092	0.6765
	Macro_Precision	0.5781	0.6379
	Macro_Recall	0.6876	0.7246
	Macro_F1_score	0.6043	0.6655
羅吉斯回歸 Logistic Regression	Accuracy	0.3786	0.5261
	Weighted_Precision	0.4756	0.5488
	Weighted_Recall	0.3786	0.5261
	Weighted_F1_score	0.3999	0.5280
	Micro_Precision	0.3786	0.5261
	Micro_Recall	0.3786	0.5261
	Micro_F1_score	0.3786	0.5261
	Macro_Precision	0.3886	0.4711
	Macro_Recall	0.5086	0.5919
	Macro_F1_score	0.3776	0.4935

此外，由表 12 的結果可知，若以四種機器學習模型分別來看，在加入年報文字基礎溝通價值變數前，隨機森林模型的 Class D 有 5.32%的比例會被預測為 Class B，有 30.90%的比例會被預測成 Class C。而在加入年報文字基礎溝通價值變數後，其錯誤預測比例分別下降為 2.66%和 25.91%；此外，在加入年報文字基礎溝通價值變數後，隨機森林模型的 Class E 被錯誤高估為 Class C 的比率由 3.81%降至 1.90%。而在支持向量機模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class A、Class B、Class C 的比率分別由 0.66%、15.89%、39.74%分別降至 0%、3.99%、30.56%；而在極限梯度提升模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class B、Class C 的比率分別由 4.33%、27.33%分別降至 1.99%、22.59%；而在羅吉斯回歸模型中，加入年報文字基礎溝通價值變數後，Class D 被錯誤高估為 Class A、Class B、Class C 的比率分別由 0.67%、17.33%、36.67%分別降至 0%、3.31%、30.79%。根據上述分析可知，在使用 BalanceBaggingClassifier 作為資料不平衡處理方法下的實證結果仍與其他兩種資料不平衡處理方法的結論一致。

表 12　各機器學習模型分類預測之混淆矩陣結果 (BalancedBaggingClassifier)
（FIN 變數 / FIN & TCV 特徵變數）

Panel A. 隨機森林模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	100.00% / 95.65%	0.00% / 4.35%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	19.53% / 10.94%	67.97% / 73.44%	11.72% / 15.63%	0.78% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class C	2.66% / 0.99%	35.88% / 26.49%	55.48% / 65.23%	5.65% / 6.62%	0.33% / 0.66%	0.00% / 0.00%
Class D	0.00% / 0.00%	5.32% / 2.66%	30.90% / 25.91%	51.16% / 58.14%	12.29% / 13.29%	0.33% / 0.00%
Class E	0.00% / 0.00%	0.00% / 0.00%	3.81% / 1.90%	26.67% / 24.76%	62.38% / 65.71%	7.14% / 7.62%
Class F	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	36.36% / 36.36%	63.64% / 63.64%

Panel B. 支持向量機模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	95.65% / 78.26%	4.35% / 21.74%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	46.09% / 20.31%	33.59% / 46.09%	15.63% / 27.34%	3.13% / 4.69%	1.56% / 1.56%	0.00% / 0.00%
Class C	10.33% / 1.99%	48.00% / 25.83%	36.00% / 58.61%	5.67% / 13.25%	0.00% / 0.33%	0.00% / 0.00%
Class D	0.66% / 0.00%	15.89% / 3.99%	39.74% / 30.56%	35.10% / 50.83%	7.95% / 13.95%	0.66% / 0.66%
Class E	0.47% / 0.00%	2.37% / 0.48%	11.37% / 5.24%	34.12% / 28.57%	43.13% / 54.29%	8.53% / 11.43%
Class F	0.00% / 0.00%	0.00% / 0.00%	3.03% / 0.00%	12.12% / 6.06%	30.30% / 33.33%	54.55% / 60.61%

Panel C. 極限梯度提升模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	95.65% / 91.67%	4.35% / 8.33%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	16.41% / 10.94%	71.88% / 73.44%	10.94% / 15.63%	0.78% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class C	2.66% / 0.67%	32.56% / 22.33%	57.81% / 69.33%	6.64% / 7.67%	0.33% / 0.00%	0.00% / 0.00%
Class D	0.00% / 0.00%	4.33% / 1.99%	27.33% / 22.59%	56.33% / 62.13%	11.67% / 12.96%	0.33% / 0.33%
Class E	0.00% / 0.00%	0.00% / 0.00%	2.86% / 1.44%	27.62% / 22.01%	61.90% / 68.90%	7.62% / 7.66%
Class F	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	3.03% / 0.00%	33.33% / 33.33%	63.64% / 66.67%

Panel D. 羅吉斯回歸模型

Actual \ Predict	Class A	Class B	Class C	Class D	Class E	Class F
Class A	95.65% / 86.96%	4.35% / 13.04%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%	0.00% / 0.00%
Class B	48.03% / 25.78%	33.07% / 38.28%	14.17% / 30.47%	3.15% / 3.91%	1.57% / 1.56%	0.00% / 0.00%
Class C	12.62% / 3.99%	47.84% / 21.26%	33.55% / 59.47%	5.65% / 14.29%	0.33% / 1.00%	0.00% / 0.00%
Class D	0.67% / 0.00%	17.33% / 3.31%	36.67% / 30.79%	33.33% / 45.70%	11.33% / 18.87%	0.67% / 1.32%
Class E	0.48% / 0.00%	2.86% / 0.48%	12.38% / 5.71%	24.29% / 21.90%	43.81% / 57.14%	16.19% / 14.76%
Class F	0.00% / 0.00%	0.00% / 0.00%	3.13% / 0.00%	9.38% / 5.88%	25.00% / 29.41%	62.50% / 64.71%

註：本研究將 S&P 信用評等分為六大類，分別為 Class A (AAA、AA+、AA、AA-)、Class B (A+、A、A-)、Class C (BBB+、BBB、BBB-)、Class D (BB+、BB、BB-)、Class E (B+、B、B-)、及 Class F (CCC+、CCC、CCC-、CC、SD、D)。而財務特徵 (FIN) 變數及年報文字基礎溝通價值 (TCV) 特徵變數分別為表 1 及表 2 所列示之 25 項及 15 項變數。而表格內的「數字比例/數字比例」，前者代表「FIN+TCV 變數作為投入變數下各信用評等群組實際數被預測為六類信用評等群組的個數比例分布」，後者代表「FIN 變數作為投入變數下各信用評等群組實際數被預測為六類信用評等群組的個數比例分布」。

綜合上述分析，本研究可得如下結論：在加入年報文字基礎溝通價值變數後，四個機器學習模型的績效皆有提升，此與 Duffie and Lando (2001) 信用風險理論模型之不完全資訊為信用風險的重要影響因子概念一致。再者，本研究亦發現年報文字基礎溝通價值變數可改善評等分類預測中將非投資級公司誤判為投資級公司的比例。

此外，本研究亦針對上述額外加入年報文字基礎溝通價值變數後的 F1 score 改善程度之經濟顯著性及統計顯著性作進一步的分析。在經濟顯著性方面，本研究以過往文獻 Wang and Ku (2021) 的預測力結果作為基準並與本研究實證結果進行比較，以補充說明信用評等分類預測效力是否能有效提升的經濟顯著性。Wang and Ku (2021) 提出 Parallel ANNs (Artificial Neural Network) 模型，實證發現該模型較 ANN 模型在 F1-score 指標上可提升 1.66% (68.11% ~69.77%) ~ 2.37% (68.93%~71.3%)；而在 Accuracy 部份可提升 2.20% (67.75% ~69.95%) ~ 2.81% (68.51%~71.32%)。而本研究實證結果發現：在加入年報文字基礎溝通價值變數後，隨機森林及極限梯度上升模型在結合 SMOTE 及 BalancedBaggingClassifier 等資料不平衡處理方法下，Macro F1-score 的增加程度約為 6.12%~6.93%。與 Wang and Ku (2021) 的實證結果相較，本研究加入年報文字基礎溝通價值變數所提升的 F1-score (6.12%~6.93%) 約為 Wang and Ku (2021) 優化模型後所提升的 F1-score (1.66%~2.37%) 的 2.58~4.17 倍。故在加入年報文字基礎溝通價值變數後所提升的信用評等分類預測效力 (6.12%~6.93%) 具有一定程度的經濟顯著性。

而在統計顯著性方面，本研究在測試期間 (2012~2017) 內以逐年滾動方式來進行各時點下之未來一年的信用評等分類預測效力分析，並進一步檢定在額外加入年報文字基礎溝通價值變數後，較之財務變數是否能有顯著的預測效力提升，以驗證其統計顯著性。實證結果如表 13 所示。而表 13 的模型效力差異 T 檢定結果顯示：本研究所使用的四種機器學習模型，在結合 RandomOverSampling、SMOTE、及 BalancedBaggingClassifier 等資料不平衡處理方法下，額外加入年報文字基礎溝通價值變數後均能使未來一年內的模型預測效力 (以 Macro F1-score 衡量) 顯著地提升。這亦顯示年報文字基礎溝通價值變數對信用評等分類預測效力的提升效果有達到統計顯著性。

表 13　未來一年信用評等分類預測效力差異比較分析

Panel A. Combining RandomOverSampling

Model	績效指標	FIN & TCV 變數	FIN 變數	Difference	T-statistics
RF	Weighted_F1_score	0.7603	0.7144	0.0458	3.9392
	Micro_F1_score	0.7635	0.7179	0.0455	3.9035
	Macro_F1_score	0.7432	0.7018	0.0414	4.0237
SVM	Weighted_F1_score	0.5718	0.4605	0.1112	13.7563
	Micro_F1_score	0.5692	0.4492	0.1199	13.0225
	Macro_F1_score	0.5417	0.4402	0.1015	9.5854
XGBoost	Weighted_F1_score	0.7667	0.7129	0.0538	5.8825
	Micro_F1_score	0.7689	0.7145	0.0544	6.0047
	Macro_F1_score	0.7509	0.7032	0.0477	4.5596
Logistic	Weighted_F1_score	0.5524	0.4159	0.1365	14.1245
	Micro_F1_score	0.5502	0.3997	0.1506	13.5476
	Macro_F1_score	0.5153	0.3924	0.1230	16.1807

Panel B. Combining SMOTE

Model	績效指標	FIN & TCV 變數	FIN 變數	Difference	T-statistics
RF	Weighted_F1_score	0.7640	0.7008	0.0632	12.6943
	Micro_F1_score	0.7656	0.7018	0.0639	12.1221
	Macro_F1_score	0.7563	0.6945	0.0618	6.1974
SVM	Weighted_F1_score	0.5697	0.4606	0.1091	13.1558
	Micro_F1_score	0.5663	0.4493	0.1170	13.0088
	Macro_F1_score	0.5351	0.4409	0.0943	10.5636
XGBoost	Weighted_F1_score	0.7608	0.6964	0.0644	5.5539
	Micro_F1_score	0.7623	0.6973	0.0651	5.8719
	Macro_F1_score	0.7486	0.6794	0.0693	7.4062
Logistic	Weighted_F1_score	0.5438	0.4225	0.1214	13.4503
	Micro_F1_score	0.5405	0.4058	0.1347	12.2776
	Macro_F1_score	0.5032	0.3983	0.1050	10.7305

Panel C. Combining BalancedBaggingClassifier

Model	績效指標	FIN & TCV 變數	FIN 變數	Difference	T-statistics
RF	Weighted_F1_score	0.6482	0.5850	0.0633	20.6563
	Micro_F1_score	0.6469	0.5807	0.0662	20.8628
	Macro_F1_score	0.6453	0.5815	0.0638	14.9921
SVM	Weighted_F1_score	0.5436	0.4093	0.1343	14.3521
	Micro_F1_score	0.5395	0.3891	0.1504	13.8007
	Macro_F1_score	0.5141	0.3967	0.1174	13.9786
XGBoost	Weighted_F1_score	0.6779	0.6127	0.0651	8.4849
	Micro_F1_score	0.6765	0.6092	0.0674	8.4254
	Macro_F1_score	0.6655	0.6043	0.0612	6.9689
Logistic	Weighted_F1_score	0.5280	0.3999	0.1281	10.1698
	Micro_F1_score	0.5261	0.3786	0.1475	10.5838
	Macro_F1_score	0.4935	0.3776	0.1159	11.2394

註：FIN 及 TCV 分別代表財務面及年報文字基礎溝通價值面之類別。RF、SVM、XGBoost、Logistic 分別代表隨機森林、支援向量機、極限梯度提升、羅吉斯迴歸等機器學習模型。

綜合上列結果及討論可知，在加入年報文字基礎溝通價值變數 (可讀性及語意) 後，機器學習模型在各項績效指標上有明顯的改善，特別是高估信用評等的情況；此外，極限梯度提升模型在加入年報文字基礎溝通價值變數後，其預測效力最穩定且泛化能力最為出色。再者，本研究亦發現年報文字基礎溝通價值變數與傳統財務變數兩者對信用評等的分類預測效力具有互補性，其可能的原因如下：對信用品質較好的公司而言，隱含其資訊不完全程度較低，故財務變數的可信度較高 (觀測到的財務變數值與實際的財務變數值差異不大)，也因此年報文字基礎溝通價值變數所能額外補抓到的資訊效果有限；而對信用品質較差的公司而言，隱含其資訊不完全程度較高，故財務變數的可信度較低，也因此年報文字基礎溝通價值變數所能額外補抓到的資訊效果相對較高，此與 Duffie and Lando (2001) 的不完全資訊觀點亦一致。

伍、研究結論與建議

實務上，由於公司的信用評等有資訊落後的問題，且並非每家企業皆能負擔巨額的信評評估成本 (亦即非每家企業皆有評等資訊)，故本研究希望利用機器學習模型以嘗試模擬 S&P 之信評方法來預測企業的信用評等，希望能提供外部資訊使用者即時且準確的信評資訊。除了以財務會計相關變數作為基礎投入變數之外，本研究亦額外引入年報文字基礎溝通價值變數於機器學習模型之預測中，實證結果顯示年報文字基礎溝通價值變數除了能提高各類別信用評等的預測準確度外，尚能減少將非投資級信用評等公司高估為投資級公司的比率。因此，年報文字基礎溝通價值變數所隱含之資訊確實能捕獲傳統財務會計變數無法涵蓋的信用風險相關之不完全資訊，此亦代表本研究模型能捕獲 S&P 信用評等方法在基於財務會計資訊以外的其他調整機制。是故，本研究模型可合理應用於無信評之公司，進而能提供給外部利害關係人關於該公司信用風險之參考依據。

近年來，機器學習應用於企業信用風險事件之研究愈來愈廣泛，其原因包含如下：第一，不同於傳統統計模型著重在相關性的討論，機器學習模型更注

重於對事件的「預測分析」；其二，不同於傳統統計模型對變數間的線性關係假設，機器學習模型可以解釋變數間的非線性關係，此亦較符合現實世界中的資料型態。本研究實證結果顯示：在以財務變數為標竿模型下 (e.g. Wang and Ku, 2021)，額外引入年報文字基礎溝通價值變數可明顯改善機器學習模型績效表現，如準確率 (Accuracy)、精確率 (Precision)、召回率 (Recall)、F1 分數 (F1-Score) 等是。而對比 Wang and Ku (2021) 僅投入財務會計變數之隨機森林模型的準確率最高可達 68.92%，本研究在額外引入年報文字基礎溝通價值變數後，隨機森林模型的準確率可達 76.56%，顯示本研究所建立之模型有更精確的解釋能力。此外，本研究亦發現在額外引入年報文字基礎溝通價值變數後，投機級評等 (Class D 以下等級) 被預測為投資級評等 (Class C 以上等級) 之比例有明顯的減少，這顯示年報文字基礎溝通價值變數能捕抓到信評公司的非量化因素調整之資訊內涵。此部份亦能實際幫助到金融機構在進行授信業務或債券投資時，可減少其遇到違約事件發生的機率。如此不只能提高經營績效，也能減少所面臨的違約風險，並避免嚴重錯誤的判斷所造成的損失。上述結果亦與 Mayew et al.(2015)、Campbell et al. (2014)、Loughran and Mcdonald (2011)、Chen and Tseng (2021) 的看法一致，亦即年報中的 MD&A (管理階層對財務狀況和經營成果的討論與分析) 及 Notes (財務報表附註) 等內容包含企業信用風險相關的前瞻性資訊。最後，本研究結果亦顯示年報文字基礎溝通價值資訊對信用評等有不同於傳統財務變數的額外解釋能力，特別是對非投資級公司的評等分類有更高的預測效力。

此外，本研究以時間點作為切割訓練集及測試集的基準，並採逐年滾動方式來逐年延展訓練集應有之資訊。較之隨機切割分組，除了能避免以未來資訊預測過去的不合理現象發生外，亦能有效補抓研究變數在時間趨勢上的軌跡，未來亦可進一步延伸至深度學習模型。然而，由於目前本研究在資料預處理後資料量不夠龐大，故在應用深度學習模型的效力將會有所限制。因此，在未來資料量足夠下，建議可嘗試使用深度學習的方法如卷積神經網路 (Convolutional Neural Network)、遞迴神經網路 (Recurrent Neural Network) 或長短期記憶模型 (Long Short-Term Memory) 等。這是因為會計和財務變數也可

描述長期和短期的企業體質特性，故透過深度學習，應可更有效率的學習財務和會計變數的資訊內容。最後，亦建議可透過其他特徵選取模型來更精確地萃取出相對重要的投入變數，以及其他超參數調整的方法，以達到優化機器學習模型預測效力之目的。

參考文獻

Altman, E.I. (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The Journal of Finance*, Vol. 23, No. 4, 589-609.

Baesens, B., R. Setiono, C. Mues and J. Vanthienen (2003), "Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation," *Management Science*, Vol. 49, No.3, 312-329.

Barboza, F., H. Kimura and E. Altman (2017), "Machine Learning Models and Bankruptcy Prediction," *Expert Systems with Applications*, Vol. 83, 405-417.

Bernanke, B.S. (1981), "Bankruptcy, Liquidity, and Recession," *The American Economic Review*, Vol. 71, No.2, 155–159.

Breiman, L. (2001), "Random Forests," *Machine Learning*, Vol. 45, No.1, 5-32.

Campbell, J.L., H. Chen, D.S. Dhaliwal, H. Lu and L.B. Steele (2014), "The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings," *Review of Accounting Studies*, Vol. 19, No.1, 396-455.

Chen, T., and C. Guestrin (2016), "XGBoost: A Scalable Tree Boosting System," *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

Chen, T.K. and Y. Tseng (2021), "Readability of Notes to Consolidated Financial Statements and Corporate Bond Yield Spread," *European Accounting Review*, Vol. 30, No.1, 83-113.

Chen, T.K., H.H. Liao, G.D. Chen, W.H. Kang and Y.C. Lin (2023), "Bankruptcy Prediction Using Machine Learning Models with the Text-based Communicative

Value of Annual Reports," *Expert Systems with Applications* Vol. 233, 120714.

Donovan, J., J. Jennings, K. Koharki and J. Lee (2021), "Measuring Credit Risk Using Qualitative Disclosure," *Review of Accounting Studies*, Vol. 26, No.2, 815-863.

Duffie, D. and D. Lando (2001), "Term Structures of Credit Spreads with Incomplete Accounting Information," *Econometrica*, Vol. 69, No. 3, 633-664.

Fisher, L. (1959), "Determinants of Risk Premiums on Corporate Bonds," *Journal of Political Economy*, Vol. 67, No. 3, 217-237.

Friedman, J.H. (1991), "Multivariate Adaptive Regression Splines," *The Annals of Statistics*, Vol.19, No.1, 1-67.

Ganguin, B., J. Bilardello (2005), *Fundamentals of Corporate Credit Analysis*, McGraw-Hill Education

Gogas, P., T. Papadimitriou and A. Agrapetidou (2014), "Forecasting Bank Credit Ratings," *The Journal of Risk Finance*, Vol. 15, No.2, 195-209.

Golbayani, P., I. Florescu and R. Chatterjee (2020), "A Comparative Study of Forecasting Corporate Credit Ratings Using Neural Networks, Support Vector Machines, and Decision Trees," *The North American Journal of Economics and Finance*, Vol. 54, 101251.

Hajek, P., V. Olej and O. Prochazka (2017), "Predicting Corporate Credit Ratings Using Content Analysis of Annual Reports–A Naïve Bayesian Network Approach," In: Feuerriegel, S., Neumann, D. (eds) Enterprise Applications, Markets and Services in the Finance Industry. FinanceCom 2016. Lecture Notes in Business Information Processing, Vol. 276. Springer, Cham.

Huang, Z., H. Chen, C.-J. Hsu, W.-H. Chen and S. Wu (2004), "Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study," *Decision Support Systems*, Vol. 37, No.4, 543-558.

Irmatova, E. (2016), "Relarm: A Rating Model Based on Relative PCA Attributes and K-Means Clustering," Papers 1608.06416, arXiv.org.

Karminsky, A. M. and E. Khromova (2016), "Extended Modeling of Banks' Credit

Ratings," *Procedia Computer Science*, Vol. 91, 201-210.

Kim, K. J. and H. Ahn (2012), "A Corporate Credit Rating Model Using Multi-Class Support Vector Machines with an Ordinal Pairwise Partitioning Approach," *Computers & Operations Research*, Vol. 39, No.8, 1800-1811.

Laitinen, E.K. (1999), Predicting a Corporate Credit Analyst's Risk Estimate by Logistic and Linear Models," *International Review of Financial Analysis*, Vol. 8, No.2, 97-121.

Lee, Y. C. (2007), "Application of Support Vector Machines to Corporate Credit Rating Prediction," *Expert Systems with Applications*, Vol. 33, No.1, 67-74.

Loughran, T. and B. Mcdonald (2011), "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, Vol. 66, No.1, 35-65.

Mai, F., S. Tian, C. Lee and L. Ma (2019), "Deep Learning Models for Bankruptcy Prediction Using Textual Disclosures," *European Journal of Operational Research*, Vol. 274, No.2, 743-758.

Martens, D., T. Van Gestel, M. De Backer, R. Haesen, J. Vanthienen and B. Baesens (2010), "Credit Rating Prediction Using Ant Colony Optimization," *Journal of the Operational Research Society*, Vol. 61, No.4, 561-573.

Mayew, W. J., M. Sethuraman and M. Venkatachalam (2015), "MD&A Disclosure and the Firm's Ability to Continue as a Going Concern," *The Accounting Review*, Vol. 90, No.4, 1621-1651.

Merton, R.C. (1974), "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *The Journal of Finance*, Vol. 29, No.2, 449-470.

Ohlson, J.A. (1980), "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, Vol. 18, No.1, 109-131.

Pai, P. F., Y. S. Tan and M.F. Hsu (2015), "Credit Rating Analysis by the Decision-Tree Support Vector Machine with Ensemble Strategies," *International Journal of Fuzzy Systems*, Vol. 17, No.4, 521-530.

Pinches, G.E. and K.A. Mingo (1973), "A Multivariate Analysis of Industrial Bond Ratings," *The Journal of Finance*, Vol. 28, No.1, 1-18.

Reichert, A. K., C.-C. Cho and G.M. Wagner (1983), "An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models," *Journal of Business & Economic Statistics*, Vol. 1, No.2, 101-114.

Seebeck, A. and D. Kaya (2023). The Power of Words: An Empirical Analysis of the Communicative Value of Extended Auditor Reports. *European Accounting Review*, Vol. 32, No. 5, 1185-1215.

Thomas, L.C. (2000), "A Survey of Credit and Behavioural Scoring: Forecasting Financial Risk of Lending to Consumers," *International Journal of Forecasting*, Vol. 16, No.2, 149-172.

Tsai, C. F. and M. L. Chen (2010), "Credit Rating by Hybrid Machine Learning Techniques," *Applied Soft Computing*, Vol. 10, No. 2, 374-380.

Vapnik, V. (1963), "Pattern Recognition Using Generalized Portrait Method," *Automation and Remote Control*, Vol. 24, 774-780.

Wang, M. and H. Ku (2021), "Utilizing Historical Data for Corporate Credit Rating Assessment," *Expert Systems with Applications*, Vol. 165, 113925.

Yu, L., S. Wang and K. K. Lai (2008), "Credit Risk Assessment with A Multistage Neural Network Ensemble Learning Approach," *Expert Systems with Applications*, Vol. 34, No.2, 1434-1444.

Text-Based Communicative Value of Annual Reports and Corporate Credit Rating Predictions: Using Machine Learning Models

Tsung-Kang Chen

Department of Management Science, National Yang Ming Chiao Tung University & Center for Research in Econometric Theory and Applications, National Taiwan University

Hsien-Hsing Liao

Department of Finance, National Taiwan University

Hao Kuan

Department of Management Science, National Yang Ming Chiao Tung University

Yu-Chun Lin

Department of Management Science, National Yang Ming Chiao Tung University

Yun Hao*

Department of Management Science, National Yang Ming Chiao Tung University

Different from the previous literature, this study employs American firm credit rating data from 1994 to 2017 to explore whether additionally introducing the text-based communicative value (TCV) variables of annual reports (e.g. readability and tones) improves the effectiveness of credit rating predictions based on the machine learning models only with financial characteristic variables as input ones. Empirical results show that after additionally introducing the TCV variables of annual reports,

* Corresponding author. Yun Hao, E-mail: peterhao.c@nycu.edu.tw; Tsung-Kang Chen, E-mail: vocterchen @nycu.edu.tw; Hsien-Hsing Liao, E-mail: hliao@ntu.edu.tw; Hao Kuan, E-mail: kuanhao.861029 @gmail.com; Yu-Chun Lin, E-mail: magic.mg09@nycu.edu.tw.

We are greatly indebted to Prof. Huimin Chung (the editor), two anonymous reviewers, and the conference discussant (Professor Chung-Ying Yeh) and participants at 2023 International Conference of Taiwan Finance Association for their insightful suggestions and comments on earlier drafts of the paper. We gratefully acknowledge the financial support from the Center for Research in Econometric Theory and Applications (Grant no. 112L900201) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

the model prediction effectiveness has a certain improvement, and the random forest and XGBoost models perform the best overall (e.g. F1 score reaches 0.76-0.77, which increases by about 6%). This shows that the TCV information of annual reports has an additional explanatory power for credit ratings compared to the traditional financial variables. In addition, this study also finds that the TCV information of annual reports can further reduce the ratio of non-investment grade firms being misjudged as investment grade ones. That is, TCV information of annual reports has the greater predictive power for non-investment grade firms. Therefore, this study verifies that the TCV information of annual reports captures the information contents of the non-quantitative factor adjustments of credit rating agencies.

Key Words: Text-based communicative value of annual reports, Credit rating, Machine learning, Incomplete information.