

# 機器學習與會計科目判斷—— 台灣半導體公司使用人工智慧提升行政效率個案 對會計教育之啟示

劉順仁 王俊凱 李季澄 黃宇翔

國立臺灣大學會計學系

通訊作者：劉順仁  
通訊地址：106216 臺北市羅斯福路四段 1 號  
E-mail：szliu@ntu.edu.tw  
投稿日期：2020 年 8 月 6 日；3 審後接受，接受日期：2021 年 8 月 11

## 摘 要

本文討論台灣某半導體公司（以下簡稱 F 公司）如何應用人工智慧技術建構自動判斷會計科目正確性的機器學習模型。本研究衍生自 F 公司財務處與某國立大學會計學系（以下簡稱 A 系）合作的「暑期人工智慧實習專案」，該專案期間為 2019 年 7 月 1 日到 2019 年 8 月 31 日。本文有兩個貢獻：(1) 探討在人工智慧時代，會計教育界與產業界如何合作發展數據分析 (data analytics) 與產業實作課程 (capstone course)，以協助會計教育的數位轉型。(2) 本文討論企業如何以機器學習技術，有效地降低成本並提升其會計作業行政流程效率。本研究期望能夠以此協助更多臺灣企業，更廣泛的運用機器學習技術來優化企業流程。

**關鍵詞：**產業實作課程、機器學習模型、會計科目判斷

---

本論文感謝兩位匿名評審的修正意見，文中論點由作者自行負責。  
數據可用性：本文使用的數據係個案公司機密資訊，無法由公開來源取得。



## 1. 前言

2017年5月23日到27日，由谷歌(Google)所研發的AlphaGo人工智慧程式，對弈世界圍棋排名第一的棋士柯潔，取得三戰全勝的戰績。從此，人工智慧的巨大潛力成為全球各界的關注焦點。迄今，人工智慧在自動駕駛、智慧醫療、財務工程等領域，已經出現諸多頗具潛力的初步應用(Google Developers 2017)。然而，人工智慧如何應用於會計實務中，仍然相當缺乏深度的討論與具體的應用。

本研究的主要目的，是敘述並討論F公司如何應用人工智慧技術建構自動判斷會計科目正確性的機器學習模型。本研究衍生自F公司財務處與A系合作的「暑期人工智慧實習專案」，期間為2019年7月1日到2019年8月31日。本研究期盼在兩個面向做出貢獻：(1)以F公司和A系協同合作完成此專案的經驗，探討會計教育界與產業界在人工智慧時代，如何發展更具有生產力的數據分析(Data Analytics)產學實作課程(Capstone Courses)，以加速會計教育的數位轉型。(2)以一個具體的會計實例，說明在一個大型企業中，機器學習如何有效地提升會計行政作業流程的效率，希望能藉此協助臺灣企業發想機器學習及其他相關人工智慧技術，未來如何更深入、更豐富地被運用於企業的流程優化。

本文結構如下：第2節介紹機器學習與其在會計上之應用；第3節敘述F公司之組織特性，並說明其財務處所欲解決的會計實務問題；第4節介紹解決本研究會計實務問題所使用的機器學習模型，說明所需要進行的資料清洗及分析，討論所經歷的組織互動流程，並報告機器學習模型經調整改善後最後所達到的成果。最後，第5節總結本研究對企業在財會領域應用機器學習技術之啟發，並討論會計教育在數位轉型中如何導入人工智慧之觀念及技術。此外，本節也將討論本研究的限制與未來研究方向。

## 2. 文獻探討——機器學習與其在會計上的應用

### 2.1 機器學習簡介

人工智慧(Artificial Intelligence)、機器學習(Machine Learning)、深度學習(Deep Learning)是本文中三個互相關聯的重要名詞。簡單來說，機器學習是一種人工智慧，而深度學習是機器學習中的一種<sup>1</sup>。人工智慧所指的範圍很廣，只要是讓電腦模擬人類的行為都可以

---

<sup>1</sup> 機器學習可以被廣義的分為監督式學習(Supervised Learning)與非監督式學習(Unsupervised Learning)，兩者最大的區別在於訓練的資料中是否有標籤(Label)。監督式學習是透過同時給予模型輸入值(Input)以及標籤，試圖讓模型學習輸入值與標籤之間的規則，分類(Classification)及回歸(Regression)是常見的手段。非監督式學習則是在只給予模型輸入值而不給予標籤的情況下，讓模型透過輸入值中的規則去找尋各個樣本之間的關係，分群(Clustering)即是這個類別中常見的一個技巧。本案例中的機器學習模型是採用監督式學習。

算是人工智慧；機器學習則是透過以下尋找最適函數的三個步驟模擬人類學習的過程。

### (1) 界定函數的數學式

以基礎的線性回歸為例，針對成對的  $(x, y)$  觀察值，我們希望能找到一個函數  $f(x) = a * x + b$ ，讓每個輸入  $x$  所得到的輸出  $f(x)$  和  $y$  的差異越小越好。那麼，我們所界定的數學式就是  $f(x) = a * x + b$ 。

### (2) 界定何謂最適函數

同樣以線性回歸為例，我們希望函數的輸出  $f(x)$  和真實  $y$  的差距越小越好。一般來說，我們使用最小平方法。當一個函數的  $f(x)$  與  $y$  的均方誤差總和越小，它就是個越好的函數。

### (3) 找出最適函數

界定完函數的數學式及好函數的定義後，電腦的任務便是在所有符合數學式的函數中，盡可能地找到一個最好的函數。值得注意的是，在多數情況下，符合該數學式的函數有無限多個，因此我們沒有辦法找到真正「最好的函數」，只能透過逼近的方式得到一個夠好的函數。

機器學習可以想像成一個尋找函數的過程，而這裡所指的函數，可以想像成一台高性能的電腦，只要給定一個輸入，它便能產生一個輸出。

而深度學習則專指機器學習中某一類的演算法，這類演算法在第一步驟（亦即界定函數的數學式時），會選用更為複雜的函數，並且用許多層的函數堆疊起來，讓上一層的函數運算的結果傳遞到下一層的函數，故稱為深度學習。

在傳統計算機科學領域中，以明確定義的演算過程設計程式，並針對輸入資料產生特定的輸出。相對起來，機器學習更擅長處理決策面的問題。以編製財務報表為例，傳統資訊領域傾向設計一套完整的系統，讓會計部門針對不同的情境選擇相對應的科目，輸入正確的數額，再根據這些數字，執行事先定義的程序，生成財務報表。

相對而言，機器學習則是透過大量的歷史資料，讓電腦學會如何判斷各類交易所歸屬的科目。例如，企業財會人員會透過交易明細（發票、收據等）判斷該筆交易要以哪個會計科目入帳。經過長期的累積，企業有了大量的交易明細及由財會人員判斷的相應科目。此時，我們就能透過機器學習的演算法，讓電腦去模擬專業人員的決策行為，使其能夠在收到單據時，自動判斷該筆開銷屬於哪個科目並入帳。而機器學習的演算法有很多種，通常我們會透過資料的性質來選擇合適的演算法。例如，類神經網路 (Neural Network) 擅於處理

影像及語音類別的資料，而傳統的演算法 [ 例如隨機森林 (Random Forest)]，則擅長處理其他分類問題。

對於商管領域的使用者來說，更重要的是去思考什麼樣的問題可以透過機器學習來解決。機器學習能做到的是提升解決專業問題的效率與正確性，而這兩者還是需要具有專業知識的人來定義。同時，作為對機器學習的潛在需求者，企業的管理階層需要在日常營運中蒐集足夠的人為判斷資料，將資料電子化儲存，以及做好資料管道 (Data Pipeline) 等預備工作。

## 2.2 機器學習之會計應用

目前在會計領域中，已有不少應用機器學習輔助預測的例子 (柯旂 2017；李鎧婷 2020；王鈞茹 2020)。以國外文獻為例，Guo, Shi and Tu 利用機器學習中的類神經網路 (Neural Network) 模型處理大量的市場消息 (新聞、電話會議記錄及各類財務報告等文字資料等)，訓練模型判斷這些文字背後隱含的情緒，加速對大量資料的判讀與分類 (2016, 153-170)。Li and Wang 利用傳統統計方法選擇指標並搭配專業領域知識過濾，將篩選出的指標作為機器學習模型的輸入，用以預測中國上市公司的破產情形 (2018, 1-20)。Ding, Kexing, Baruch Lev, Xuan Peng, Ting Sun, and Miklos A. Vasarhelyi 使用保險業損失準備金估計與實際資料，驗證搭配機器學習的估計方法相較傳統的估計方法，更能準確做出會計估計，從而扮演協助會計師與管理階層決策的角色 (2020, 1098-1134)。Li, Bin, Julia Yu, Jie Zhang, Bin Ke, and Boa Yang 利用從財報中透過會計專業挑選的變量搭配機器學習技術建立偵測舞弊的模型，並解決了過往關於舞弊的資料較少而使得模型訓練較為困難的問題 (2020, 199-235)。Siano and Wysocki 則透過機器學習中遷移學習 (Transfer Learning) 的方法，先使用現有的大資料集訓練模型，再針對資料較小的任務微調模型，讓模型在不易取得特定資料的情況下仍得以運作，改善原先需要大量特定資料才可以訓練模型的問題 (2020, 217-244)。

以國內文獻為例，王若樸指出，銀行面對法人與個人申請金融業務時，需要衡量客戶的倒帳風險，傳統上靠人為的統計分析來推估，準確率約八成 (2019)。玉山銀行建構機器學習模型，利用各種工具分析大量挑選出的用戶資料，現在倒帳風險預測準確率已經來到九成五。此外，Faggella 指出，KPMG 會計師事務所建置自有的人工智慧工具——KPMG Ignite，整合事務所內的資訊科技資源，並利用機器學習中自然語言處理相關技術，擷取客戶電話會議中重要的資訊、判讀會議的情緒，並用以作為會計師預測未來事件、決策的依據 (2020)。

總結來說，當企業擁有大量資料，且面對沒有明確答案、需要一定程度的人為判斷，

具有某種潛在規律的問題，就是機器學習可以發揮最大效益的場域。人工智慧時代的商管教育，應該加強訓練學生思考企業內部是否有符合這類情境的問題。

### 3. F公司特性與會計實務問題界定

#### 3.1 研究背景介紹

近年來，F公司積極推動數位轉型，內部已有多項透過大數據分析與人工智慧技術提升工作流程自動化程度的應用。首先是搬運系統，F公司一天一座工廠每分鐘至少會發派六百萬到八百萬不等的指令，建立自動智慧化搬運系統後，可在一分鐘之內計算出最佳生產排列組合。在導入人工智慧系統後，F公司十二吋晶圓廠的機台生產力提升了62%，同時減少了65%的人力負荷。另外，空調系統在F公司節能策略上一直是關鍵角色。半導體廠廠務用電平均占全廠用電近47%，其中又以空調設備耗為最大宗，占廠務用電近60%。目前F公司使用的是雙溫度冰水冷卻系統，冰機運轉負載更低，並取消鍋爐以熱回收系統替代，並應用大數據分析，在不同季節及環境下，動態調整水溫（黃敬哲 2020）。而目前人工智慧技術已能將溫度精確控制在正負2°C，節能效率上看19%，並且在今年初成功導入全球12座的12吋晶圓廠。

而除了製造部門透過機器學習判斷生產品質之優劣，F公司財務處也導入相關技術以優化財會流程及提升管理決策，本研究即為典型例子。本研究中包括兩個會計帳務上的實務問題：

##### (1) 零用金的會計處理

由於金額較小，零用金的花費並不需要經過嚴格的請採購流程，只要各部門人員填妥單據及相關交易科目後，便能繳交給會計部門，經過會計專業人員審查後入帳。然而，F公司每年的零用金交易約有十萬多筆，為了確保這些科目的分類正確，需要聘請數位專業會計人員，除了入帳程序極為耗時影響會計流程的進行外，也產生了一筆額外的開銷。

##### (2) 請採購流程的會計處理

除了零用金之外，F公司還有許多需要經過請採購流程的高額交易項目，像是建置廠房、機台維修及保養等。這類科目的交易金額高，因此需要經過嚴格的請採購流程。過去，F公司為了確保這些交易的科目能被正確分類，便在請採購流程中設計了一套「PR六階」系統。所謂的PR六階系統，是透過文字探勘的方式，歸納出各個科目的常用詞彙，再搭配專業人員的知識將會計科目的分類歸納成約2500條規則，並將程式邏輯埋在六道關卡

的篩選中讓各部門人員選出最合適的會計科目。然而，在了解各科目的高頻詞彙後，還得以人工的方式整理並歸納出數千個分類規則，必然會產生邏輯上的漏洞，也使得財務處每年仍需要分派人力確保這些交易的分類正確性。

雖然 F 公司財務處近年透過人工方式盡量確保這些會計科目的正確性，但面對數十萬筆交易，要確認每一筆的分類正確並不符合成本效益，多數時候只能透過抽查重要交易的方式進行。然而，那些沒有被檢查到的錯誤會計科目分類，將會導致財務報表的不實表達、管理決策的錯誤以及稅務法規的違反。

以上兩個會計實務問題，都是很典型的分類問題，且擁有龐大的歷史資料，因此很適合透過機器學習模型來解決。由於 F 公司財務處主管具備機器學習應用的素養，乃思考透過最先進的技術解決其管理問題。

### 3.2 專案準備

要執行一項成功的產學合作專案，勢必需要實務界與學術界共同努力。為此，F 公司財務長及 A 系系主任在專案正式執行前約十個月就開始進行準備。準備內容包括：確認專案目的與範圍、選擇執行時間、決定學生待遇；A 系在內部大力介紹此專案並提供獎學金後進行初步學生選拔；F 公司人資部門對 A 系推派學生進行面試及決定最後人選等。

在開始建構模型之前，F 公司財務處會計資訊組的員工向兩位 A 系暑期實習生（一位為研究所二年級同學，另一位為大四同學）詳細的介紹整個計畫（如表 1 所示）。內容包

表 1 F 公司在模型建構前的準備

建構模型前的準備	說明
目前遇到的問題	因交易複雜性以及工廠人員對會計的不了解所造成會計科目的錯誤分類
問題的嚴重性	錯誤的會計分類將導致財報、稅務上以及管理決策上的損失
計劃目的	透過產學合作的順利進行，讓業界能夠與學術接軌，而學生能運用學校所學的知識實際落地，雙方都有所學習，並從中獲益
數據的準備	平時電子化記錄各種可能有價值的數據，提供 F 公司歷年下來累積的龐大資料集
可以利用的資源	提供足以支撐學生欲使用的模型的硬體資源等
模型方向	提供學生關於模型的輸入、輸出以及分類模型的選擇建議
欲達成的目標	建立一個模型用來自動分類交易的會計科目，並提供一個正確率為 90% 的參考標準及目標
專案的價值	針對零用金交易可以處理每年約十幾萬筆交易為財會部門節省至少 5 個人力，而針對經過請、採購流程的交易將可處理每年近十萬筆交易以節省財會部門至少 3 個人力。另外，兩種交易的分類正確性皆提升約 10%

括：目前遇到的問題、問題嚴重性、計畫目的、數據的準備、可利用的資源、模型方向、欲達成的目標以及專案價值，這使得對於實務運作不熟悉的會計系學生，能夠很容易地了解自己的定位以及方向。F 公司也事先建構出一個最基礎的模型作為參考，使對於機器學習模型較不熟悉的學生不必由零開始，能夠很快地在現有的基礎上進行修正，朝著明確的目標努力。

此外，在實習期間學生會被配對一位專門協助實習生的員工（稱為實習生的 Buddy），這位 Buddy 會不時的來關心實習生的進度以及狀況，並給予適當的協助及建議。在面對不熟悉的交易項目時，Buddy 會協助實習生找到相關部門的人員快速解答問題。例如，當實習生發現有兩筆「購買滑鼠」的交易分別被歸類在雜項購置以及電腦周邊這兩類會計科目時，Buddy 就會請負責辦公設備的人員協助實習生了解什麼樣的情況會把交易歸在雜項購置，什麼樣的情況又會歸在電腦周邊。而當實習生遇到模型優化的瓶頸或是其他技術上的問題時，Buddy 也會協助實習生聯絡大數據處的專家，給予建議以及方向。

最後，在實習期間 F 公司給予實習生很大的發揮空間，讓實習生多方去嘗試各種不一樣的模型，讓實習生有自我學習的動力，進而找到更好的方法去解決問題，F 公司的這一系列準備與作為，對於完成這項產學合作計畫至關重要。

## 4. 以機器學習優化會計流程的方法及成效

### 4.1 模型建置流程

#### 4.1.1 機器學習模型介紹

本研究選擇使用類神經網路及隨機森林兩種機器學習模型，茲簡單說明如下：

##### 4.1.1.1 類神經網路

類神經網路為利用層層的非線性函數（稱之為神經元）來描繪出資料的規則。此專案中，使用的類神經網路模型為長短期記憶模型 (Long-short Term Memory, LSTM) 的改良版「雙向長短期記憶模型 (Bidirectional Long-short Term Memory, BiLSTM) 加上注意力機制 (Attention Mechanism)」<sup>2</sup> (Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins 1999, 2451-2471; Huang, Zhiheng, Wei Xu, and Kai Yu 2015; Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin 2017, 5998-6008)。

<sup>2</sup> 注意力機制，是在每個時間點的輸出向量上加上一個可以被學習的權重，讓模型可以學習根據不同的輸入決定應該要取用多少比例的其他時間點的輸出，從而達到模仿人類注意力機制的效果。

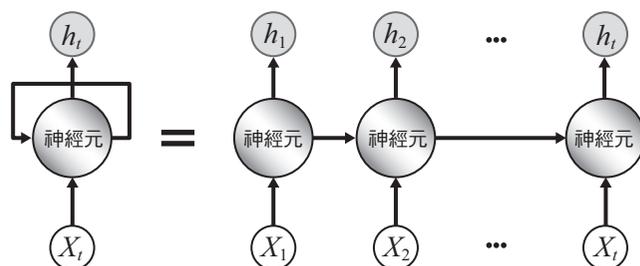


圖 1 RNN 示意圖 (其中  $X$  為時間  $t$  的輸入,  $h$  為時間  $t$  輸出以及時間  $t+1$  的輸入)

LSTM 模型與其他類神經網路模型最大的差異在於該模型會考慮資料的順序性, 因此在很多序列性資料中是非常受歡迎的, 例如: 時間序列分析、自然語言處理以及語音辨識等。LSTM 模型源自於循環神經網路 (Recurrent Neural Network, RNN) 模型, 並加以改良, 解決 RNN 模型對於較久遠訊息遺忘的現象 (Mikolov, Tomas, Stefan Kombrink, Lukas Burget, Jan Cernocky, and Sanjeev Khudanpur 2011, 5528-5531)。LSTM 模型以及 RNN 模型都是將前一個神經元的輸出視為下一個神經元的輸入, 藉由這種機制, 每個神經元都會學習到前一個資料的特徵, 如圖 1 所示, 而 LSTM 將 RNN 的神經元作進一步的改良。由於 LSTM 只會將前面神經元的輸出當作下一層的輸入, 亦即該模型只會顧慮前文的資訊, 然而文字是有前後關聯的, 而任何一個詞都可以由前後的字來提供暗示。舉例來說, 一個交易的內容為「購買 fab12 的電腦以及相關周邊軟體」。現在把電腦兩個字遮住而變成「購買 fab12 的○○以及相關周邊軟體」, 若只從○○之前進行判斷, 我們絕對無法知道○○是代表什麼, 然而若結合○○之後的資訊, 我們可以從「周邊軟體」這兩個字得知○○可能代表電腦。為此, 本案例中我們使用了 BiLSTM, 此模型因為是雙向模型, 因此會同時考慮前後文的資訊來進行判斷。使用注意力機制, 即是利用矩陣運算讓模型的參數學習資料中需要重點關注的目標區域並提高注意力於該區域, 這可以使模型更迅速的找到一個新文本的重點, 不僅可以加快模型的運算效果, 更可以提升模型的正確性, 此機制是目前自然語言處理中前沿技術 (如: Transformer) 的基礎。

#### 4.1.1.2 隨機森林

隨機森林是一種整合多棵決策樹的演算法, 而所謂的決策樹便是透過二分法不斷交叉畫出的樹 (Liaw, Andy, and Matthew Wiener 2002, 18-22)。以本專案為例, 電腦學習時能發現成本中心是否為 FAB、單價低於或高於一千萬, 及請購單文字敘述等有助於會計科目分類的資訊。將這些二分法的規則記錄下來, 就形成圖 2 的決策樹。

而既然我們已經有一棵能幫助分類的決策樹, 為什麼又會需要隨機森林呢? 原因就在

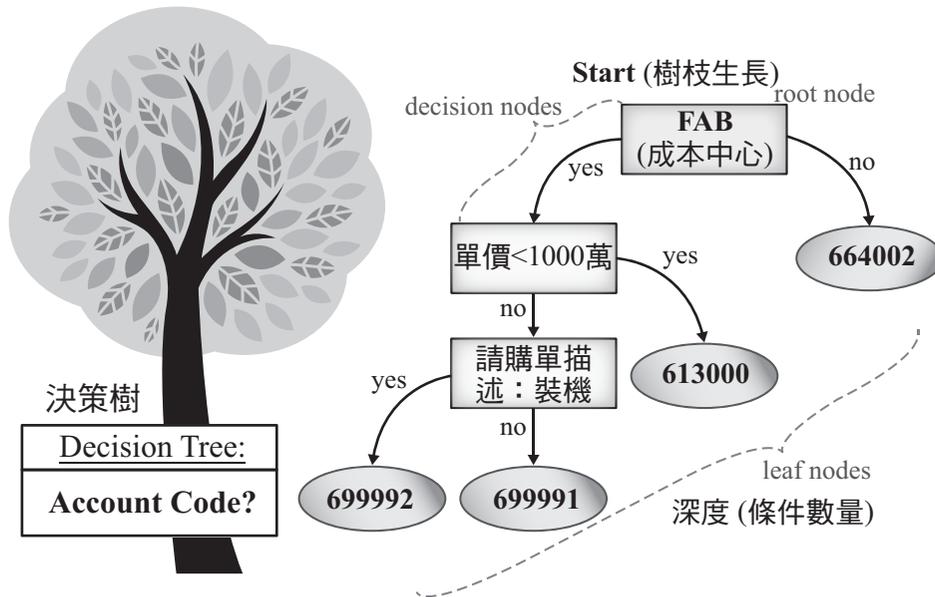


圖 2 決策樹示意圖

於我們每次生成決策樹時，除了只抽取母體資料中的部分樣本之外，也不會用到所有的因變數，因此每次生成的決策樹都會不太一樣。由於單看一棵決策樹的分類結果太過偏頗與果斷，我們便將所有決策樹的結果記下，並選擇最多決策樹認同的結果作為最終答案，如圖 3 所示。

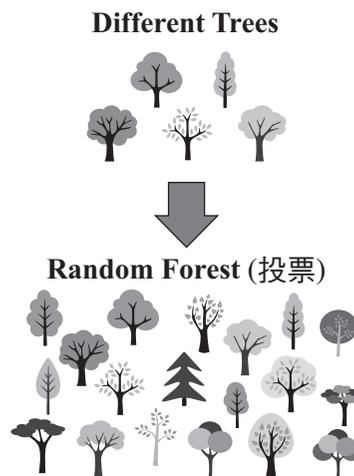


圖 3 隨機森林示意圖

#### 4.1.2 模型選擇原因

結合上述對於模型的介紹，在此解釋於本專案中選擇這兩個模型的原因。就類神經網路而言一共有三個理由。第一，一般來說，當資料量大到一個程度時，類神經網路模型的表現會超越傳統的機器學習模型。而隨著資料量的增加，傳統的機器學習會有局限性使得正確率無法再增加，然而類神經網路所受的局限性較小。此專案中所使用的資料高達數十萬筆，根據經驗這樣的數據量是夠大的。第二，許多新穎的人工智慧技術都是透過類神經網路模型實作，因此有許多可以嘗試的方向。第三，由於資料中包含文字的資料，使用類神經網路中的 RNN 以及 LSTM<sup>3</sup> 可以考量文字中詞語之間的前後關係，並藉此作為判斷的參考依據。

而就選擇隨機森林演算法而言，主要原因有四個。第一，隨機森林在傳統機器學習模型中算是非常強大的分類模型，利用集成學習 (Ensemble Learning)<sup>4</sup> 的方式可以排除其他傳統機器學習過於武斷的缺點。第二，與類神經網路相比，因為隨機森林是從決策樹所集成而來，因此在模型的解釋上相對容易。而一般的迴歸模型雖然也可以被容易的解釋，然而其模型的複雜度卻無法達到隨機森林的程度，因此也無法比隨機森林更好的刻畫資料的規則。第三，隨機森林能夠很容易的看出各個變數的重要性，即特徵重要性 (Feature Importance)，這可以提示我們模型的缺點以及資料的問題並加以改進。最後，由於該專案的問題和傳統的分類問題接近，因此此模型能夠達到很好的效果。

基於以上原因，本研究採取兩種不同的方法進行模型建構，期望能從比較新穎的深度學習與傳統的機器學習兩個角度切入，提高解決問題的機率。藉由隨機森林的特徵重要性的挖掘，了解模型及資料的重點，並加以提高類神經網路模型的正確性，而類神經網路的複雜度，也可以為隨機森林提供改進的方向。兩個模型相輔相成，讓本研究更有可能去完成預定的目標。類神經網路模型的複雜度及難以解釋程度一直造成大家對於其的恐懼，而現今可解釋機器學習 (Explainable ML) 領域中的其中一個分支就是採用隨機森林等較易解釋的樹狀模型，去模仿比較複雜且難以解釋的類神經網路 (往往被戲稱為黑盒子)，藉此確定複雜模型執行適當的學習。

---

<sup>3</sup> LSTM 與一般 RNN 不同之處在於加入了遺忘閥 (Forget Gate)、輸入閥 (Input Gate)、輸出閥 (Output Gate) 的機制，使得模型得以學習如何根據序列資料的不同而決定要不要提取累積至前一個時間點的資訊，以及要不要將現有的資訊加入累積的資訊之中，從而使模型可以學習到較長序列中前後文的訊息。所謂的遺忘閥、輸入閥、輸出閥，指的是三種可以被學習的函數，它們根據不同的輸入，輸出一個 0 到 1 之間的值，利用這個輸出的值乘以模型前一個時間點的輸入以及這個時間點的輸出，模型得以決定要不要提取儲存的資訊，以及把當下時間點的資訊往下一個時間點傳遞。

<sup>4</sup> 可解釋機器學習是機器學習新興的領域，旨在透過視覺化模型參數等方式，讓人類可以理解機器所學習到的資料特徵，從而改良模型或是做出決策。集成學習指的是透過不同模型的組合，來提高整體任務表現的一種機器學習的方式。例如，透過讓不同模型「投票」以決定最後的預測結果，即是一種集成學習。

### 4.1.3 變數分析與資料清洗

本專案中輸入的解釋變數包含交易原因(文字資料)、成本中心(分類型變數)、交易金額(連續性變數)等等,而應變量則為會計科目。在進行後續的分析之前,先對變數進行描述性統計以及探索性資料分析(exploratory data analysis),瞭解各個變數之間的關係,包含:會計科目的分佈、高頻字與會計科目的關係以及變相之間的相關性等等<sup>5</sup>。

而這項專案中的模型和一般統計模型最大的不同,在於文字將作為模型的因變量之一。F公司財務處規定,各部門人員在提交交易明細給會計部門時,需附上文字敘述說明該筆交易的用途。而會計專業人員判斷科目時,這些文字敘述往往是判斷的關鍵。也就是說,若能將這些文字轉為電腦能夠理解的形式,模型的能力也會顯著提升。在此,本研究使用一個基礎但實用的文字轉向量技術 One-Hot Encoding。

所謂的 One-Hot Encoding,就是將所有可能出現的性質用一個由 1 及 0 組成的陣列表示。舉例來說,有兩個句子分別為「購置新型號機台」及「購置廠房清潔用品」。這兩個句子中出現的詞彙分別為「購置」、「新」、「型號」、「機台」及「購置」、「廠房」、「清潔」、「用品」,本研究假設一個句子中只有可能出現這些單詞,並用建置一個 One-Hot 陣列,依序代表「購置」、「新」、「型號」、「機台」、「廠房」、「清潔」及「用品」。此時,便能用 [1, 1, 1, 1, 0, 0, 0] 表示第一個句子,因為句中只出現了前面四個詞;另一方面,也能用 [1, 0, 0, 0, 1, 1, 1] 來表示第二個句子,因為句中並沒有出現「新」、「型號」及「機台」三個詞彙。透過此方式,文字轉換成了由 1 與 0 組成的陣列,也是電腦能夠理解並分析的形式。另外,在建立 One-Hot 陣列之前,我們會用到 Python 中的 Jieba 套件來協助我們斷詞,將「購置新型號機台」自動斷成「購置」、「新」、「型號」、「機台」(Sun 2012)。上述 Jieba 斷詞的原理乃是透過其內建的文字資料庫判斷詞彙的斷點,然有許多會計專有名詞及 F 公司內部常用詞彙並不會出現在原始的資料庫中,例如「折舊」、「裝機」等。為了讓 Jieba 的斷詞更加精確,我們會將這些詞彙加入原始的資料庫中。此外,我們也會修正常見的錯別字及英文拼音,避免意義相同的文字被判斷為不同的詞彙。

值得注意的是,在這個例子中只以兩個句子為例,但在本研究中,每筆交易便有一筆交易敘述。也就是說,本研究會使用數十萬個句子建成陣列,這樣一來只會遺漏極少量的詞彙。另外,在建立 One-Hot 陣列的過程中,還需要把句子切割成詞彙,這個部分可以透過 Python 中的套件 Jieba 來實現。

<sup>5</sup> 由於與 F 公司簽有保密協議,有關實驗數據的各類統計資料圖表無法在本研究眾進行彙報。然而,進行描述性統計以及探索性資料分析是資料分析中不可或缺的過程。

除了文字之外，交易明細中還有許多變量有助於會計科目分類，而變量的選擇除了透過既有技術篩選外也要仰賴會計專業人員的判斷。在此也列舉幾個重要的變量：

廠商	不同廠商所販售的物品並不相同，因此直接影響了交易發生時的會計科目。
物品單價	物品的價格會影響物品的種類，舉例來說，機台設備價格較高，辦公室雜物等物品價格較低，而不同種類的物品也有其相應的會計科目。
廠房 / 單位	不同單位購買的物品不盡相同，因此會直接影響會計科目。
採購人員	各採購人員固定負責某類型的交易，因此對會計科目會有影響。

另外，在該表中，廠商、廠房及採購人員這三個變數為類別型變數。也就是說，這些變數的值並非連續的數字而是離散的類別。對於這類變數，我們也能夠透過 One-Hot Encoding 的方式將變數轉為電腦能夠理解的向量。

#### 4.1.3.1 組織互動流程

最初的模型建立完畢後，正確率約為 86%。然而，這樣的正確率距離主管所要求的 90% 還有一段距離。為了改善模型，實習生每週都會和會計部門中其他組別的同事開會，向他們請教各個會計科目的專業知識，並了解現行分類下的問題，進而調整資料清洗及模型建立的細節。在會計專業知識的部分，舉例來說，隨著時序推移，許多過去常用的會計科目至今已無太大的分類意義。在沒有定期刪除以及新增科目的情況下，將會出現相同性質的事件被分類為不同的會計科目以及不同性質的事件都被分類為相同科目等情形（也就是所謂的雜訊），將會造成模型判斷上的困難。因此，科目存在的必要性也是專案會議的重點項目之一。而反覆執行開會並修正模型的流程數次後，最終模型的成效來到了 92.7%。

#### 4.1.4 模型的調整

此專案中採用的兩種模型各自有不同的特點，因此也可以達到互相幫助的效果。在建構出最基本的模型後，透過比較容易解釋的隨機森林模型，可以清楚地看到文本中哪些用詞對於模型的判斷有著重要的影響，即特徵重要性。透過了解關鍵的詞語，檢查進行斷詞時是否有出現斷錯的問題，並將重要的詞語加入 Jieba 的字典中（降低斷錯詞語的機率）。如此一來，資料清洗的正確可以幫助類神經網路在判斷上達到更好的效果。此外，在模型產出的結果，可以繪製出混淆矩陣 (Confusion Matrix)，將兩個模型的矩陣對比，可以找出輸入的資料本身就on容易判斷錯誤的會計科目以及交易說明，而不會只是單一模型的偏誤，接著再針對這些會計科目進行分析以及與部門人員討論。模型的調整中，透過加入人工的樣本 (data augmentation)，以及不斷的嘗試不同的模型架構以及參數，而達到最終的成效。

另外，在調整模型的過程中，我們將資料以 7:2:1 的比例拆分為訓練集、驗證集與測試集。接下來，我們透過模型在驗證集上的表現調整模型的細節，並利用測試集來衡量模型的最終能力。以下詳述模型調整細節：

針對類神經網路在訓練過程的超參數以及函數的選擇，一共包括迭代次數 (Iteration Number)、批量大小 (Batch Size)、損失函數 (Loss Function)、激活函數、最優化算法、學習率以及初始化方式。迭代次數為 200 次，大約在 150 次左右即達到收斂。另外，我們也在損失函數中加入了 L2 正則化 (Lambda 為 0.01) 調整項，避免過度擬合 (Overfitting) 的現象。Batch 的大小會影響訓練的穩定度及過擬合程度，分別以 64、128 及 256 實驗後發現 128 的正確率最佳。而由於是分類問題，我們選擇交叉熵 (Cross-entropy) 作為損失函數，並使用 Adam 的最優化算法找尋最佳解 (學習率設定為 0.003)。各層的激活函數都選擇 ReLu，並採用 Xavier 均勻分布初始化。另外，除了一般的完全連接層 (Fully-connected layer) 外，我們也加入了 Dropout 層 (設定為 0.5) 以及 Batch Normalization 層。

相較於類神經網路，隨機森林的參數較為簡易，我們主要調整樹的數量及各樹的最大深度。在隨機森林中，樹的數量越多就越能避免模型武斷的情形，但也同時會使運算時間增加。而各樹的深度越深則會影響模型的擬合能力，若深度過淺模型將無法模仿複雜的函數，深度過深則會有過度擬合的問題。在反覆測試後，我們選擇以 400 棵深度最大為 200 層的樹組成最終的隨機森林模型。

#### 4.1.5 模型經調整改善後所達到的結果 (應用及成效)

在模型建立完畢後，本研究對於兩個流程有不同的應用方式：

##### (1) 零用金交易

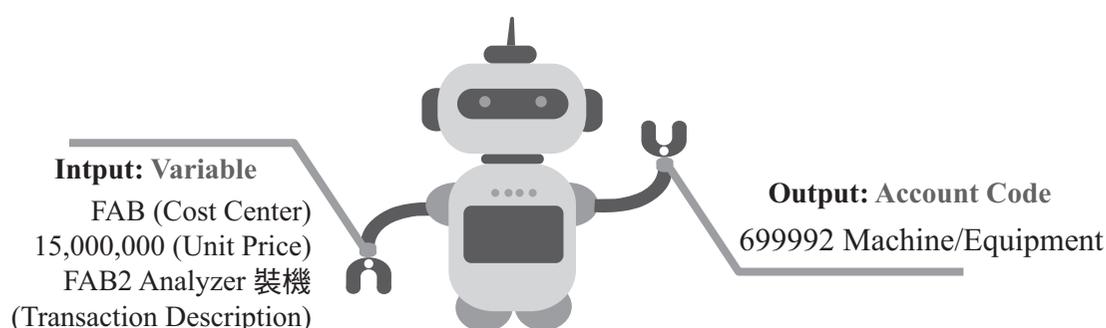
由於不需要經過請採購流程，各部門人員只需要在交易完成提供交易明細。在填寫交易資料明細的過程中，模型會透過填寫內容分析出合適的幾個會計科目，即採用推薦使用者的模式。各部門人員選擇會計科目後，便能將交易資料及會計科目送交至會計部門。而當最後選擇的科目並非模型推薦的最佳科目時，會計人員便能將該交易列為重點檢驗項目，仔細檢驗分類的正確性，此為驗證式模式。

##### (2) 請採購交易

請採購交易的應用可以分為兩個階段。首先是在請購階段，有許多交易的細項還無法確認，包含廠商及物品單價等，此時會透過一個較基礎的模型 (使用的變數較少，正確率約 91%)，初步分類該交易的會計科目，及早確認分類正確性。而在採購完畢後，由於交易細項都已確認，便能使用最終模型 (完善的模型，正確率約 92.7%)，再次確保分類結果無誤。

值得注意的是，這些檢驗的結果都會被逐一記錄，當提交之會計科目有誤時，會計部門便能分析模型的弱點，觀察其為何沒有推薦最正確的會計科目，進而優化模型。舉例來說，若模型在分類機台維護與機台購置上的表現不佳，便需要檢視訓練時的原始資料，是否有文字清洗不乾淨抑或是資料量不足等問題。若有，將透過資料清洗的步驟增加及搜集更多相關訓練資料的方式優化模型。而若提交錯誤是人為造成的，則可以分析各部門人員填寫錯誤會計科目之原因，以加強這方面的教育訓練。

#### 4.1.5.1 實例



註：上圖中的資料涉及保密協議，僅能以少量、非敏感的文字呈現。

在取得完整的交易資料後，工廠人員可以將成本中心、單位價格、交易敘述等資料輸入至模型中，此時模型會將最有可能的會計科目輸出給會計部門。

在多數情況下，實際的會計科目分類規則遠比以規則為基礎 (Rule-Based) 的「PR 六階系統」所設計出的複雜，造成舊系統在面對「裝機」、「Fab」等辭彙時會錯誤地將該筆資料分類為「機台相關雜項成本」、「機台維修費用」等科目 (F 公司內部有許多與機台相關且十分相近的科目)。然而，新的模型能透過機器學習演算法從大量的數據中找到真正關鍵的分類原則，將該筆交易分類至正確的科目。

## 4.2 模型價值與成效

有了模型的應用方式，F 公司將能創造出更多價值。在建構出機器學習模型之前，因為工廠的員工對於會計並不熟悉，容易錯誤分類會計科目，F 公司則必須以人工的方式進行會計科目與交易細項的核對。採用模型前，公司的會計人員每年都要針對上萬筆的資料進行覆核，即使採用抽查的方式進行，還是需要花費大量的時間，且效果並不佳 (根據內部員工的敘述，人工覆核的正確率不到 80%)。為什麼 F 公司要如此重視會計科目的正確性呢？要回答這兩個問題，就要從三個方面去解釋。

第一，對於需要經過請採購流程的交易來說，金額通常是相對龐大的，而對於零用金交易而言，金額雖然較小，筆數卻很多，導致總額也很大，因此兩種交易皆需要重視。從財務報導的角度來看，錯誤分類會計科目的失誤將會造成財務報表的重大不實表達，一個最簡單也最容易發生錯誤的例子就是將本來應該資本化的交易列成了費用，而造成當期的費用虛增，進而低估當期利潤。這樣的結果不論對於 F 公司或是投資者來說，都是不希望見到的事情。

第二，從法規監管層面來說，錯誤的會計科目可能會導致違法或是引來不必要的訴訟，最常見的例子就是在稅法層面上，雖然錯誤的將會計科目列為費用可以使當期的應納稅額減少，然而卻有違法之虞，容易引來不必要的爭議。

最後，從管理會計的角度出發，資訊、制度以及管理者決策與行為是構成管理會計的核心三要素，三個要素不僅相互之間影響，也會深深地影響著一家公司的成敗。若將交易錯誤分類，影響的是三要素中的資訊要素，這將會進而影響管理者的決策。舉例來說，工廠人員很容易將應列為產品成本的支出列為當期費用，這樣的錯誤分類將導致產品成本的低估，進而影響產品的定價決策以及造成其他錯誤的管理決策。由此可知，交易分類的正確性是極為重要的，潛在的損失遠大於人工成本的支出，這也是為什麼 F 公司需要花費大量人力成本來處理這個問題。

然而，這樣的人工成本並不會為企業帶來額外的價值。因此，我們需要建構這樣的一個自動分類會計科目的模型。根據業界的慣例，這類的模型至少需要達到 90% 的正確率，管理層才有足夠的信心去使用。最終建構出的自動分類會計科目模型正確率針對零用金交易達到 90.7%，而針對經過請採購流程的交易達到 92.7%，皆符合管理層的要求。由於不同模型的效果在不同任務上各有差異，最終是以類神經網路模型處理零用金交易，以隨機森林模型處理請採購流程的交易（如表 2 所示）。除了正確率以外，本專案也參考了其他指標包括：精確率 (Precision)、召回率 (Recall) 以及 F1 分數<sup>6</sup>，以確保模型在學習規則中並沒有

---

<sup>6</sup> 為了方便理解，我們先從二元分類的例子著手。給定類別為正負兩類，考慮分類正確的情況：若模型將類別為正的結果預測為正，則稱之為 True Positive (TP)，若模型將類別為負的結果預測為負，則稱之為 False Negative (FN)；反之，考慮預測錯誤的情況，若模型將類別為正的樣本預測為負，則稱之為 True Negative (TN)，若模型將類別為負的樣本預測為正，則稱之為 False Positive (FP)，給定這四類的預測結果，我們可以定義精確率、召回率、F1 分數如下：

$$\text{精確率} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{召回率} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1 分數} = 2 \times \text{精確率} \times \text{召回率} / (\text{精確率} + \text{召回率})$$

其中，精確率衡量的是對模型回報正類別時的準確程度，召回率衡量的則是正確分類的樣本中網羅正樣本的比例，F1 分數則可以視作兩者的調和平均。若進一步推廣到多分類的情形，可以將正樣本輪流設為各類別，從而得到各類別的精確率、召回率、F1 分數。

表 2 兩模型正確率比較

	隨機森林	類神經網路
零用金交易	90.5%	90.7%
請採購流程交易	92.7%	91.3%

嚴重的偏誤。最終的模型在精確率、召回率以及 F1 分數皆為 90%。這樣一個對於工廠員工輸出會計科目進行驗證的模型，若針對零用金交易可以處理每年十幾萬筆交易為財會部門節省至少 5 個人力，而針對經過請採購流程的交易將可處理每年近十萬筆交易以節省財會部門至少 3 個人力 (此節省人力資料來自公司內部員工的估計)。

若將模型的用途變成推薦式模型，亦即在工廠人員針對零用金的交易選擇會計科目時，在人員填完應填的資訊後，模型馬上推薦最可能的三個會計科目，這樣的用途能讓工廠人員不需要在眾多複雜的會計科目中尋找，可大大的減少他們浪費掉的時間。以零用金交易為例，依此用途，模型可以取前三名分數高者作為模型的輸出 (模型在運算時最後會給每個科目都一個分數)。將正確率的定義為只要正確的會計科目有在推薦之中，則算正確的比率。如此一來，模型正確率針對零用金交易可以達到 95.9% (如圖 4 所示)。

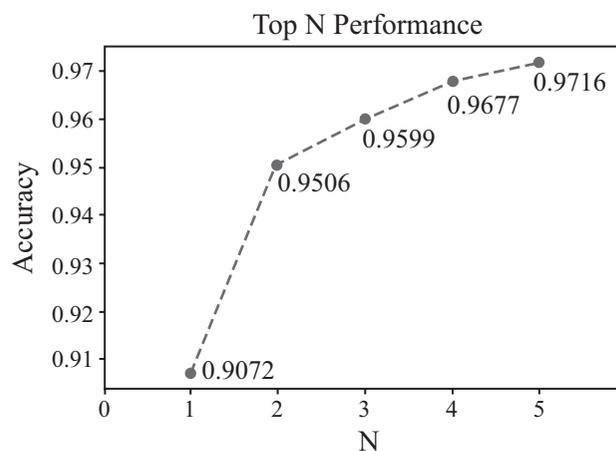


圖 4 模型推薦個數與正確率關係圖

## 5. 總結，研究限制與未來方向

### 5.1 總結

本研究敘述並討論 F 公司如何應用人工智慧技術建構自動判斷會計科目正確性的機器

學習模型。經由 F 公司財務處與 A 系合作的「2019 年暑期人工智慧實習專案」，本專案中的兩位實習生優化 F 公司既有的機器學習模型，建構出正確率超過 90%，具有實用價值的模型。在本研究中，我們觀察到：(1) 未來任何一個企業，都必須大量地使用人工智慧相關技術，以加快數位轉型速度，加寬數位轉型的範圍。透過本研究具體的會計實例，我們希望協助臺灣企業發想更多機器學習及其他相關人工智慧技術的應用。(2) 以 A 系和 F 公司協同合作完成此專案的經驗而言，在人工智慧時代會計教育的數位轉型，在會計課程中大量加入數據分析 (data analytics) 的元素，已經是刻不容緩。本研究希望能協助教育界與產業界緊密合作，以發展更具有生產力的數據分析產學合作教育模式。

會計作業中的交易分類，理想上是透過對會計準則的正確理解而推論 (deductive decisions) 得來的。然而，實務上經常由請購方業務人員做初步判斷，再由會計專業人員進行審核 (review)。由於請購方業務人員對會計準則的理解未必正確，而會計專業人員在交易數量龐大下難免有所疏忽，因此總是存在相當比例的錯誤。因此，當公司累積大量交易資料時，以數據導向 (data driven) 機器學習模型進行歸納性的快速分類判斷，乃成為提升行政效率的有力工具。

在 F 公司的責任中心制度中，財務處係屬「費用中心」，其主要任務就是在達到預期目標下盡可能減少支出。對於財會主管或會計人員而言，在引入模型後，可以花費較少時間在審核會計科目的正確性上，並將此部分人力配置到其他工作中，對於被歸類為費用中心的財會部門來說，乃是重要績效。此外，由於 F 公司非常重視創新，能導入機器學習的創新技術來達到減少支出的目的，並且為未來人工智慧技術更廣泛的應用打下基礎，也是財務處另一個重要的績效。然而，隨著每年交易量的增大以及新交易種類的產生，新的交易可能會影響模型的正確率。因此，針對系統後續的維護以及持續改善而言，若能定期的更新模型，亦即將新的數據加入原本的數據中一同訓練現有的機器學習模型，將可以使模型更符合管理需求。

## 5.2 研究限制與未來方向

本研究的機器學習模型受限於暑期只有兩個月的建構修正時間，仍然存在不少未來可以繼續精進之處。例如：

### (1) 本研究可以使用更新的模型及演算法

目前在自然語言處理領域中，最廣泛被使用的模型是 BERT (全名為 Bidirectional Encoder Representations from Transformers) 以及其延伸的模型，例如：XLNet、RoBERTa、DistilBERT、ALBERT 以及 ELECTRA 等 (Devlin 2018; Yang 2019; Liu 2019; Sanh 2019; Lan 2019;

Clark 2020)。該模型是基於 Transformer 所建立的模型 (為 Transformer 的 Encoder 的部分)，於 2018 年由 Google 提出，其在很多表現上都比以往的其他模型效果要佳，若專案時間更長，可以嘗試使用 BERT 來進行模型的分類，可以適當的解決一些原本使用 BiLSTM + Attention 模型的問題 (例如：必須一個詞一個詞處理，無法在 GPU 上平行化運算，導致運算時間可能會拉得很長)。此外，BERT 為 Google 使用維基百科及其他資料共 33 億個語料所訓練出來的模型，若能對該模型進行遷移學習 (Transfer learning) 將能大幅度提升將文本向量化後，各個向量之間 (各個詞之間) 的順序關係。

## (2) 本研究可以使用更新的套件

在本專案中使用 Jieba 進行中文的斷詞，而在去年年底臺灣中研院釋出新的斷詞套件名為 Cktagger，其斷詞效果比 Jieba 要佳，斷詞是自然語言處理的基礎，若能在斷詞上作改善，模型的效果將會有所提升 (Chen 1994, 111-126)。

## (3) 本研究資料可以進行更嚴格的清洗

所謂「Garbage-in-garbage-out」，若輸入模型的資料是非常雜亂的，模型自然不會有好的結果。本專案所使用的數據中仍然有許多錯誤的標記 (即錯誤的會計科目)，若有更多時間進行資料的清理以及釐清，將有助於模型的訓練。

最後，我們認為本研究的經驗可以擴大應用到會計其他領域及問題。除了本研究中的會計科目分類外，會計中仍有許多能透過機器學習改善的問題。例如，勤業 (Deloitte) 會計師事務所運用透過機器學習中的自然語言處理 (Natural Language Processing) 分析交易及契約，進而了解背後的風險 (William D. Eggers, Neha Malik, and Matt Gracie 2019)；安永 (EY) 會計師事務所也建立機器學習模型幫助會計師尋找財務資料中的異常數據，繼而出造假之交易 (Ernst & Young, n.d.)。另外，中央研究院的團隊也透過文字探勘技術分析財報中文字與企業財務風險之關係，期望能透過機器學習更深入地使用財報中的資訊。過去，要達成這些目標，往往需要耗費很多人力與時間，但現在有了機器學習的幫助，只需要幾個資料科學家、專業領域人員以及歷史資料便能達成這些任務，這是劃時代的突破，也是值得會計實務及會計教育未來持續探索的新方向。

## 參考文獻

王若樸，「玉山金控近百項金融 AI 專案新進展，靠 ML 評估顧客 3 年倒帳風險比人還要準。」iThome，2019 年 8 月 28 日，<https://www.ithome.com.tw/news/132698>。

- 王鈞茹，「【專欄】淺談人工智慧於財務文字資料分析之應用：中研院訊」，中研院訊，2021年8月26日，<https://newsletter.sinica.edu.tw/22950/>。
- 李鎧婷，「與會計相關的國際趨勢或金融資訊」2020年4月23日，<https://www.angle.com.tw/accounting/cross/post.aspx?ipost=4222>。
- 柯旂，「理財不靠運氣！機器學習分析財報『軟資訊』，出現這些字財務風險特別高：大數聚 Big Data Group」，大數聚 - Big Data Group，2017年11月7日。<https://group.daily-view.tw/article/detail/615/>。
- 黃敬哲，「台積電冷卻系統再進化，引進 AI 增加晶圓良率」，TechNews 科技新報，2020年3月6日，<https://technews.tw/2020/03/06/tsmcs-cooling-system-evolves-again-introducing-ai-to-increase-wafer-yield/>。
- Chen, Keh Jiann, Shing Huan Liu, Li Ping Chang, and Yeh Hao Chin. "A practical tagger for Chinese corpora." In Proceedings of Rocling VII Computational Linguistics Conference VII (August, 1994): pp. 111-126. <https://aclanthology.org/O94-1005>.
- Clark, Kevin, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. "ELECTRA: Pre-Training Text Encoders as Discriminators Rather than Generators." ArXiv [Cs.CL] (2020). <http://arxiv.org/abs/2003.10555>.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of deep bidirectional Transformers for language understanding." ArXiv [cs.CL] (2018). <http://arxiv.org/abs/1810.04805>.
- Ding, Kexing, Baruch Lev, Xuan Peng, Ting Sun, and Miklos A. Vasarhelyi. "Machine Learning Improves Accounting Estimates: Evidence from Insurance Payments." *Review of Accounting Studies* 25, no. 3 (2020): 1098-1134. <https://doi.org/10.1007/s11142-020-09546-9>.
- EY. "How an AI Application Can Help Auditors Detect Fraud." EY. EY, January 7, 2019. <https://go.ey.com/2PphF1D>.
- Faggella, Daniel. "Ai in the Accounting Big Four - Comparing Deloitte, PWC, KPMG, and EY." *Emerj*. Emerj, April 3, 2020. <https://emerj.com/ai-sector-overviews/ai-in-the-accounting-big-four-comparing-deloitte-pwc-kpmg-and-ey/>.
- Gers, Felix A. "Learning to Forget: Continual Prediction with LSTM." 9th International Conference on Artificial Neural Networks: ICANN '99, 1999. <https://doi.org/10.1049/cp:19991218>.
- Google Developers "How Machine Learning with TensorFlow Enabled Mobile Proof-of-Purchase at Coca-Cola." Google Developers Blog, September 21, 2017. <https://developers.googleblog.com/2017/09/how-machine-learning-with-tensorflow.html>.

- Guo, Li, Feng Shi, and Jun Tu. "Textual Analysis and Machine Learning: Crack Unstructured Data in Finance and Accounting." *The Journal of Finance and Data Science* 2, no. 3 (2016): 153-70. <https://doi.org/10.1016/j.jfds.2017.02.001>.
- Huang, Zhiheng, Wei Xu, and Kai Yu. "Bidirectional LSTM-CRF models for sequence tagging." ArXiv [cs.CL] (2015). <http://arxiv.org/abs/1508.01991>.
- Lan, Zhenzhong, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. "ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations." ArXiv [Cs.CL] (2019). <http://arxiv.org/abs/1909.11942>.
- Li, Bin, Julia Yu, Jie Zhang, Bin Ke, Boa Yang. "Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach." *Journal of Accounting Research* 58, no. 1 (2020): 199-235. <https://doi.org/10.1111/1475-679x.12292>.
- Li, Yachao, and Yufa Wang. "Machine Learning Methods of Bankruptcy Prediction Using Accounting Ratios." *Open Journal of Business and Management* 06, no. 01 (2018): 1-20. <https://doi.org/10.4236/ojbm.2018.61001>.
- Liaw, Andy, and Matthew Wiener. Classification and regression by randomForest. R news (2002), 2(3), 18-22.
- Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." ArXiv [Cs.CL] (2019). <http://arxiv.org/abs/1907.11692>.
- Mikolov, Tomas, Stefan Kombrink, Lukas Burget, Jan Cernocky, and Sanjeev Khudanpur. "Extensions of Recurrent Neural Network Language Model." 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011. <https://doi.org/10.1109/icassp.2011.5947611>.
- Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. "DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter." ArXiv [Cs.CL] (2019). <http://arxiv.org/abs/1910.01108>.
- Siano, Federico, and Peter D. Wysocki. "Transfer Learning and Textual Analysis of Accounting Disclosures: Applying Big Data Methods to Small(Er) Data Sets." *SSRN Electronic Journal*, 2020. <https://doi.org/10.2139/ssrn.3560355>.
- Sun, Jian. Jieba chinese word segmentation tool. 2012. <https://github.com/fxsjy/jieba>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In Advances in neural information processing systems (2017): 5998-6008. <https://arxiv.org/abs/1706.03762>.

William D. Eggers, Neha Malik, and Matt Gracie “Using AI to Unleash the Power of Unstructured Government Data.” Deloitte Insights, 16 January 2019. <https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/natural-language-processing-examples-in-government-data.html>.

Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. “XLNet: Generalized Autoregressive Pretraining for Language Understanding.” ArXiv [Cs. CL] (2019). <http://arxiv.org/abs/1906.08237>.

# Machine Learning and Account Classification Judgements: Educational Implications of the Application of Artificial Intelligence to Accounting Procedures of a Taiwan Semiconductor Company

**Shuen-Zen Liu Chun-Kai Wang Chi-Cheng Lee Yu-Shiang Huang**

Department of Accounting, National Taiwan University

Corresponding author : Shuen-Zen Liu

Address : No. 1, Sec. 4, Roosevelt Rd., Da' an Dist., Taipei City 106216, Taiwan (R.O.C.)

E-mail : szliu@ntu.edu.tw

Received : August 6, 2020; After 3 rounds of review, Accepted: August 11, 2021

## Abstract

This study discusses how a Taiwan semiconductor company (F company hereafter) applies artificial intelligence (AI) techniques to build machine learning models for automatic account classification judgements. The study builds on a 2019 summer internship project collaborated by the finance department of F company and the department of accounting of a national university in Taiwan. The study contributes to the accounting literature in two ways. (1) The study provides implications for accounting educators on how to collaborate with industry practitioners in terms of conducting AI problem-solving projects in building data-driven capstone courses. Thus, the study helps expedite the digital transformation of accounting education. (2) Additionally, the study shows how a large company builds machine learning models to save personnel costs and enhance operational efficiency in its accounting procedures. The study aims to motivate Taiwanese companies to explore opportunities of optimizing their operational efficiency through AI techniques.

**Keywords:** Capstone courses, Machine learning models, Account classification judgements

---

The authors acknowledge helpful comments of two anonymous reviewers, and take sole responsibility for their views.

Data availability: Data used in this study are corporate confidential data and thus are not available upon request.



## 1. Research Issues

The study is built on a 2019 artificial intelligence (AI) summer internship project collaborated by the finance department of a Taiwanese semiconductor company (F company hereafter) and the department of accounting (A department hereafter) of a national university in Taiwan. The internship project aims to explore how to build machine learning models for automatic account classification judgements. Two types of judgements are investigated: (1) petty cash account judgements and (2) capital expenditure account judgements. Although F company's petty cash transactions involve small amounts, the company often has more than 100,000 transactions annually, spreading in five different sub-petty cash funds. Correct account classifications require the sound judgements of purchasing personnel at the beginning and subsequent valid audits of accounting staff. In addition, as a large semiconductor manufacturer, F company incurs a significant dollar amount of equipment purchases and maintenance. Thus, F company has designed complex account classification rules based on keywords contained in purchase orders of business units. Such activities involve significant auditing manpower but still result in numerous errors because of logic flaws in the classification rules. F company advocates AI technique application in the entire company to improve efficiency and create values. The manufacturing activities of the company have significantly progressed, whereas its supporting functions have made slower progress in terms of AI application. Thus, the CFO of F company considers the AI internship an opportunity to obtain new ideas from students and professors in the A department to gain momentum of its digital transformation in staff functions.

## 2. Research Methods

The study does not adopt typical archive empirical research methods, which often involve hypothesis testing through statistical models. Consistent with the research issues, the study adopts two types of machine learning models to build effective models for automatic account classification judgements. These models mainly aim to increase the accuracy rate of classifications and thus save auditing costs. The machine learning models under investigation are (1) bidirectional long/short-term memory model in neural networks and (2) random forest model. Model 1 considers data sequences to explore their implications, whereas model 2 uses multiple levels of decision trees to improve the accuracy of classification. In addition, model 2 can observe the feature importance of variables through decision trees. These two models are popular in the current machine learning literature and have convenient references.

After selecting the machine learning models for the study, we conducted variable definitions and data cleaning that are essential to data analytics. The independent variables include transaction reasons (text data in the purchase orders), type of cost centers, the dollar amount of transactions, etc., whereas the dependent variable includes accounts of transactions.

Text analysis is a relevant research method to the study because F company requires purchasing parties to submit descriptions of the intended use of the purchase. The description provides useful keywords that subsequently form machine-readable matrixes with 1 or 0 as elements (the so-called one-hot encoding). In addition, the study has to clean various “noises” in the dataset, which consist of missing data, the same transactions classified as different accounts, different transactions classified as the same accounts, etc. Data cleaning also requires F company’s executives to help provide their domain knowledge of data and transactions for the identification of different types of noises.

The study adjusts the two machine learning models by splitting data into a 7:2:1 proportion for the training, validation, and test datasets, respectively. Various detailed processes of model adjustments are employed, such as the number of iteration (approximately 200 times), the number of batch size (128, to obtain the highest accuracy), the choice of cross-entropy as a loss function, and the selected ReLu as an activation function. The above-mentioned processes ensure that the two machine learning models improve their accuracy rates to levels that executives agree to apply in practice.

### 3. Research Findings and Contributions

Before the internship project begins, the experts of F company had built preliminary machine learning models for petty cash transaction and capital expenditure classifications with accuracy rates of approximately 86%. F company’s executives require that accuracy rates exceed 90% to put the models in practice. This study found that, if carefully “trained,” the machine learning models’ accuracy rates can be improved to approximately 92% in petty cash and capital expenditure account classification judgements. In addition, the study confirms the importance of organizational support to ensure the success of the internship project. F company provides two important resources: (1) domain knowledge experts and (2) big data division’s experts to clarify the need of the finance department and sources of noises and to provide the expertise in building machine learning models, respectively. We conclude that careful mentoring in the research site is critical for capstone projects to succeed.

This study contributes to the accounting literature in the following two ways. (1) The study helps accounting educators expand their understanding of how to collaborate with industry practitioners in terms of conducting AI problem-solving projects in building data-driven capstone courses. Thus, the study provides an important implication on how to expedite the digital transformation of accounting education. (2) Additionally, the study shows how a large company builds machine-learning models to save personnel costs and enhance operational efficiency in its accounting procedures. The study aims to motivate several Taiwanese companies to explore opportunities of optimizing their operational efficiency through AI techniques.