

# 機器學習演算法應用於預測員工離職

## Machine learning Algorithm Applied to Predict Employee Turnover

作者：蔡逸帆

### 第壹章 緒論

#### 1-1 研究背景

員工離職 (Employee Turnover) 是指員工自願終止雇用關係 (Hom and Griffeth, 1995)，許多企業的管理者表示，員工離職率 (Employee Turnover Rates) 太高造成企業的人力成本損失，每一位員工因為離職而造成的人力成本損失，可以高達該員工年薪的 50% (Puni, Agyemang, and Asamoah, 2016)。

日本企業即使在終身雇用制和年功序列制的觀念之下，旅館業和餐飲業仍然有高達 14.7% 的離職率 (野地祐二 & 野々部惠美子, 2018)，日本是目前全球少子化和高齡化最嚴重的國家，勞動人口不足已經是不可避免的現象，再加上離職率的提高，讓總體國家勞動力更加惡化，日本政府為此提出許多擴大雇用人口的政策，例如鼓勵 65 歲以上已退休的高齡人口重新投入就業市場，貢獻自己的經驗和技術，另外日本政府也鼓勵已離開職場的家庭主婦以彈性工時的方式重返就業市場 (大湾秀雄, 2017)。而民間企業也加強徵才方式，例如在應屆畢業生畢業之前的就職活動 (類似台灣的就業博覽會或是單一企業的大型招募活動)，對於適合的學生提出保證畢業即就業的承諾，到職之後由資深員工一對一指導帶領，目的是防止新進員工因為適應不良而提早離職 (山本和史, 2017)，對於在職員工，則提供更好的薪資和福利制度，其目的也是防止員工可能因為薪資或福利和同業差異太大而跳槽，或是可能對薪資和福利制度不滿而離職。

並非只有中小企業會遭遇這個問題，即使是跨國的大型企業也面臨相同的問題，例如 IBM 意識到員工的離職無法被有效的掌握，IBM 的管理者不知道員工離職的真正原因，這也讓 IBM 的管理者不知道是否應該改善勞動條件，也無法評估誰是下一個適合的求職者 (Khan, Khan, and Muhammad, 2019)，IBM 提供的薪資待

遇已經高於同業，很顯然的高薪資水準仍然無法留住員工，某些原因讓 IBM 的員工寧願捨棄高薪也要離開公司。

台灣企業的也呈現高離職率的情況，根據勞動部在民國 108 年對國內 6,624 家企業進行第 4 次勞動調查，預計在 109 年 1 月底之前增加人力原因選項的企業，以“退離者之補充”最高，占 36.35%，而在 107 年「退離者之補充」選項為 35.95%，僅次於「需求市場擴大(含設備或部門擴充)」選項的 36.22% (中華民國勞動部, 2019)，顯示國內企業在近 2 年呈現高離職率的情況。

國內對於研究員工離職的文獻並不多，大多以特定企業為研究對象，例如對房仲業的離職率研究 (王怡文, 2019)，作者以信義房屋為個案對象，使用問卷調查對區主管的領導風格分數和新進人員晉升轉正率做回歸分析檢定，研究結果顯示區主管的領導風格分數和新進人員在第 30 天的離職率為負相關，但是這樣直接對員工進行的問卷調查有一個相當大的問題，那就是台灣職場文化就是不能說真話，即使對主管再怎樣不滿意，大多數人為了工作權仍然會給于滿意的評價，另一個問題是信義房屋主打「30 天工作鑑賞期，離職發 5 萬元離職金」的徵才方案，這個方案已經在求職市場被證實有人僅是為了這 5 萬元而去應徵，30 天到期即拿錢走人，而不是真的為了在信義房屋工作，當然也不會有人公開向公司承認應徵的目的是為了領這 5 萬元，這些問題讓人質疑問卷調查法和訪談法用於研究真實離職率的可信度。

## 1-2 研究動機

在過去的 100 年中，西方學術界發表了 2,000 篇以上有關於員工離職的研究文獻，相對的，現在有許多研究者也回顧了過去有關員工離職的理論和研究 ( Lee, Hom, Eberly, Li, and Mitchell, 2017 )，透過比較過去的文獻和現在的職場環境，人們廣泛認為人才市場需求已取代員工忠誠度，而忠誠度是影響雇主與員工之間關係的決定性因素之一。同樣明顯的是，這種根本性的轉變導致雇主與員工之間的忠誠度和信任度下降 (Goldberg, 2014)。這個轉變也實現在大多數的台灣企業對於員工離職的態度，即使是優秀的員工離職，大多數的管理者普遍抱持著「你不做後面還有多人搶著做」的心態，而大多數的中小企業 HR 也沒有人力資源管理的相關知識，類似會計兼 HR，或是餐飲科系等和人力資源管理完全無關的人一出校門即擔任人資的工作，這些荒唐的情況多不勝數，更謬論對於員工離職的管控，儘管有極少數的大型企業會對員工做離職訪談，或是在離職後進行問卷式的調查，但是在員工已決定離職或是已離職的情況下，這些事前事後的問卷和訪談已經不具有意義。

員工離職有其優點和缺點。優點是低素質的員工可以被替換，提高組織的創造力，靈活性和適應性。缺點是人力成本將會增加，這些成本包括顯性成本（招募，培訓，生產力損失等），隱性成本（士氣低落，企業聲譽降低，職位鏈中斷，機會損失等）。員工離職的形式分為兩種：自願離職和非自願離職。自願離職是指由員工做出離開公司的決定，有非常多的因素讓員工決定離開公司，非自願離職是指雇主作出的離職決定，例如新冠肺炎導致公司經營困難而資遣員工，不適任解僱等形式。對於管理者來說，非自願離職是可以預測和控制的，但是自願離職無法事先預測。不同類型的員工離職對企業的影響是不同的，但是過多的自願離職對於企業來說絕對是不利的 (Zhang, 2016)。轉換另一個觀點，如果能預測可能離職或是即將離職的員工，那麼企業將可事先查覺而予以慰留，或是分析找出不適任的員工，因此本研究計畫的目標是基於數據，研究可預測員工離職的實際應用方法。

## 第貳章 文獻回顧

在回顧相關文獻時，考慮到不同國家地區的企業文化是不同的，因此本章節分為歐美地區和亞洲地區的相關文獻回顧。

### 2-1 歐美地區的員工離職相關研究

在歐美地區的管理領域中，對於員工離職的研究已經有百年歷史 (Hom, Lee, Shaw and Hausknecht, 2017)，在早期的文獻中記錄了雇主如何通過加薪手段來阻止員工辭職 (Local, 1917)，初期的研究通常是討論薪資和勞動環境對員工離職的影響。根據美國人口普查局在 2012 年的統計數據，有 80% 的餐飲業者雇用低於 20 名員工，他們面臨健康和人身安全問題，包括職場暴力、兇殺風險增加、工作量過多、休息時間不足以及長期站立，從而導致高離職率。美國餐飲業員工的高離職率表明對於經濟和產業層面的影響，在經濟層面上，離職的員工未必可以立刻找到下個工作，在失業期間領取失業救濟金的人數被視為經濟景氣指標之一，部份失業者因為無力負擔房租等生活開銷而淪落為遊民，形成都市裡的治安危險人口。在產業層面上，Hancock., Allen, Bosco, McDaniel, and Pierce, (2013) 等人的研究表明高離職率的企業比低離職率的企業生產效率更低，高離職率的企業薪資水準通常低於或等於最低薪資標準，這透露出另一個資訊，降低員工離職率將在某種程度上減少失業人數和國民的經濟壓力。

另一個被忽視的問題是兼職員工和計時員工的離職。兼職員工和計時員工參與的工時直接影響他們的收入，有關工時不足的研究表明從個人層面的直接影響擴展到組織層面的調節因素，因此，它們提供了直接的經驗證據，人力資源專業人員可以用組織變革來解決工時不足引起的離職問題 (Wang, 2018)。在歐美企業，高離職率仍然是普遍存在的現象，可能會對任何企業的運營和獲利能力產生負面影響。數十年來，管理人員、企業領導者和學術界人士已投入大量時間和資源來理解和減少這種現象。而由於頻繁的招募和員工離職引起的勞動力短缺，企業營運績

效和獲利能力因為高離職率而下降甚至導致中斷運營 (Smith, 2018)，這對於企業產生了實質的負面影響，De Winne, Marescaux, Sels, Van Beveren, and Vanormelingen (2019) 等人的研究表明，在複雜的非線性關係中，組織的勞動生產率在較低的離職率情況下將會增加，在某個時間點達到峰值，然後以負衰減的方式下降，離職率的波動與勞動生產率直接呈現負相關的關係。

工作滿意度調查的人員數據和資訊常被歐美國家的企業用來預測員工辭職 (Frederiksen, 2017)，雇主在用於管理和人力資源分析的調查中確保員工的匿名性，將工作滿意度調查的實施委託給外部諮詢公司的簡單機制可能是最佳的，這個方法解除了員工對於公司的不信任感，讓問卷調查變的更有可信度。

## 2-2 亞洲地區的員工離職相關研究

Chen, Chuang and Chen (2018)等人的研究指出，人力資源管理對於不斷增長的東亞經濟至關重要，東亞國家正努力在經濟上實現多樣化，轉向高附加價值的產品市場，以確保經濟的持續增長。而台灣的人力資源管理以儒家思想為基礎，同時受到西方文化的管理研究和實踐的影響，近年來正經歷著重大的典範式轉變。隨著台灣經濟的快速增長以及其產業和組織的轉型，台灣的人力資源管理已發展成為應對競爭日益激烈，瞬息萬變的全球商業環境。經濟部中小企業管理局將中小企業定義為員工人數少於 200 人且年營業額不超過新台幣 1 億元的組織。中小企業在全球幾個國家的總就業增長中佔很大比例，而中小企業在這些國家的國內生產總值增長中佔有很大份額。Cheng and Chang (2019)的研究指出，經濟部中小企業管理局將中小企業定義為員工人數少於 200 人且年營業額不超過新台幣 1 億元的組織。中小企業在全球幾個國家的總就業增長中佔很大比例，而中小企業在這些國家的國內生產總值增長中佔有很大份額。台灣的中小企業在 2015 年的表現，特別是中小企業的數量，達到創紀錄的 1,383,981 家，佔台灣所有企業的 97.69%。此外，台灣中小企業的就業人數增至近年最高水平的 8,759,000 人，佔台灣所有就

業人數的 78.22%，人力資源學者對中小企業員工的敬業度問題進行了充分的討論，因為它對組織和個人產生了積極而重大的影響。員工敬業度直接影響員工的離職傾向。

日本企業主要由男性主導，女性管理者比例極少，日本社會嚴重的「男尊女卑」文化讓日本女性在職場的地位低落，女性離職的比例遠高於男性，Hasegawa and Hasegawa (2017) 的研究探討日本女性員工離職意願的 5 種決定因素，研究表明影響日本女性員工離職意願的最大因素是「積極的思維方式」，這項因素的影響遠超過「對當前公司的職業機會的感知」、「職業前景」、「當前的技能類型或企業經營方式」和「當前公司的未來前景」，這表示日本女性員工在職場地位低落的環境中產生強烈的危機意識，她們更願意思考如何改變現況，這和日本男性員工安於現況的心態形成強烈的對比，而這種「積極的思維方式」無法被預測和事前得知。

印度的勞動力人口在亞洲僅次於中國，印度企業同樣面對高離職率的問題，Singh(2018)的研究指出，「無意義的工作」、「組織信任」、「組織承諾」、「學習能力」、「道德感」、「不公平的人力資源管理方法」、「心理契約的履行」、「工作滿意度」和「溝通」等因素，對於全球各種不同產業和規模的員工離職都會產生負面影響，在印度製造業中，「無意義的工作」、「組織信任」和「溝通」是印度員工產生離職意圖的最主要因素。但是迄今為止，印度尚未開展任何研究來共同研究這些變量，而印度雇主正嘗試努力保留優秀的勞動力，特別是印度的軟體產業人才。

從文獻來看，員工離職已經被視為一項重要的議題，並被人力資源從業者廣泛視為企業的主要威脅 (Porter and Steers, 1973)，而員工的離職意圖在促使他離開公司的思考過程中起著重要作用。這種離職意圖可以解釋為「關於在不久的將來離開組織可能性的個人的主觀估計」，因此，預測員工的離職意圖對於人力資源從業者來說是必須的 (Mowday, Porter and Steers, 2013)。

## 第參章 研究方法

### 3-1 機器學習( Machine learning )

機器學習的初步概念在 1995 年已經被提出 (Bishop, 1995)，常見的機器學習定義為：「機器學習是一門人工智慧的科學，該領域的主要研究物件是人工智慧，特別是如何在經驗學習中改善具體演算法的效能，電腦程式可以從經驗 (E) 中學習有關某類任務 (T) 和性能度量 (P) 的資訊，電腦對任務 (T) 的性能(由 P 衡量)隨著經驗 (E) 提高而跟著提高」。機器學習是人工智慧 AI (Artificial Intelligence) 的一個分支。人工智慧的研究有著一條從「推理」為起點，「數據」為中繼點，再到以「學習」為終結點的清晰脈絡。機器學習是實現人工智慧的一個途徑，也就是以機器學習為手段解決人工智慧中更高深的問題。機器學習目前已發展為一門多領域的科學，涉及概率論、統計學、逼近論、計算複雜性理論等多門科學。機器學習理論主要是設計和分析讓電腦可以自動「學習」的演算法。機器學習演算法是從資料中自動分析獲得一項規律，並利用這項規律對未知的資料進行預測的演算法。因為機器學習演算法中涉及了大量的統計學理論，機器學習與推斷統計學的關係尤為密切，也被稱為統計學習理論。在演算法設計方面，機器學習理論關注於可實現的，有效的機器學習演算法。許多推論問題並無方程式或模型可循，所以部分的機器學習研究是以開發容易處理的近似演算法為目的。機器學習目前已經被廣泛應用於資料探索、影像辨識、自然語言處理、生物特徵辨識、搜尋引擎、醫學診斷、信用卡詐騙檢測、證券市場分析、DNA 序列測序、語音與手寫辨識和神經網絡機器人等領域。許多科技大廠，包含 Google、IBM、Amazon 和 Microsoft 都在發展機器學習演算法，使用雲端運算或是後台運算讓機器學習的執行更快速，只要導入數據和程式包，機器學習即可完成所有的計算，使用者僅需要了解基本的程式語言函數即可，例如目前最受歡迎的 R 或是 Python，而無需任何事先的程式編寫經驗或是資訊工程背景。

最常見的機器學習演算法種類有「構造間隔理論分布」、「構造條件概率」

和「近似推斷技術」，國外已經有研究者使用數據科學的方式研究影響員工離職的因素，例如 Ajit (2016) 的研究使用機器學習演算法的方式，評估何種機器學習演算法有最高的準確率，在此之前並未有任何人力資源管理領域的研究使用機器學習演算法做分析，作者使用 XGBoost、Logistic Regression、Naïve Bayesian、Random Forest (Depth controlled)、SVM (RBF kernel)、LDA 和 KNN (Euclidean distance) 共 8 種機器學習演算法，在 Apple MacBook OS 10.10.5 版本作業系統和 Python 2.7 運行，使用的數據來自某一跨國零售商的人力資源資訊系統 (HRIS)，作者的結論為 XGBoost 是最準確的演算法。但是該項研究僅表示 XGBoost 演算法最準確，而且沒有說明變數，這可能和企業機密有關，企業不希望太多的內部數據被公開。隨著科技進步，機器學習演算法已經進入另一個世代，使用機器學習算法預測員工離職可能性成為一種先進的研究方法。本研究在 IBM Watson Machine Learning 建立模型架構，並使用最新版的 Python 3.7 進行數據分析。

### 3-2 變數設計

在許多研究員工離職的文獻，最常被使用的變數是員工滿意度，這項滿意度指的是員工對於當前的工作環境、薪資水準等的綜合評價，被視為是主流的研究變數，另一個是雇主對員工的績效評分，這 2 項變數的差距可以視為雇主和員工之間的信任度，當 2 項分數越接近，表示雇主和員工之間對於彼此有高信任度和默契，但是有一項個別的現象，就是雇主和員工都給彼此相同的分數，或是同一個分數出現在多位員工，此時必須將數據標準化，標準化之後的數據將會轉換成一個介於 1 到 0 之間的數字，此時即可看出相同的分數其實是不同的，標準化的操作在許多統計軟體可以被輕易完成，例如 SAS，SPSS 等。另外一種變數是客觀的實際數據，這些數據是員工在任職過程中產生的歷史數據，例如薪資、工作年資和績效等，這些不可變的歷史數據是客觀的，沒有經過人為修改的，機器學習的特點是需要大量的變數，變數種類越多則準確度越高，因此企業紀錄的人事資料是極為重要的數據，



可以使用以下或更多的變數：

Satisfaction level：員工滿意度（0-1）

Last evaluation Evaluation of employee performance (0-1)：最近一次的員工績效評估（0-1）。

Number project：工作上負責的項目數量。

Average monthly hours at the workplace：平均每月工作時間。

Time spend company Number of years spent in the company：工作年資

Work accident：員工是否發生過工安事故(0 或 1)。

Left：員工是否已經離職（0 或 1）。

Departments：工作部門。

Salary: 薪資（可以僅區分低中高或是實際數據）。

根據以上的數據使用機器學習演算法分析，過程中使用多種的演算法交叉驗證，最後提出準確度最高的演算法，輸出的結果是影響員工離職的最可能變數和最可能離職的員工，這項數據介於 1 到 0 之間，越趨近 1 則越可能離職。

#### 第肆章 預期的研究貢獻

根據研究結果所顯示的數據，提供企業用於預測員工離職可能性的衡量方法。企業應該與人力資源專家合作討論如何改善勞動條件和勞動環境，儘管許多企業的管理者宣稱公司的離職率太高而且無法被控制，但是合理的數據分析將提供管理者和人力資源從業人員一個量化的評估方法，除了預測可能的員工離職意圖之外，也提供企業在徵才時的篩選條件，企業可以審視過去的就業環境與現在的勞動條件是否有落差。基於機器學習得到的研究結果對於研究者或是企業來說都極為寶貴且重要的，研究者必須洞察研究結果並提出結論，企業再將研究結果實際應用於管理層面，改善企業在人力資源管理上的缺點，研究者和企業可採取緊密的合作關係，持續發掘企業的管理問題。

另一個研究貢獻是，許多企業對於內部員工教育訓練越來越重視，但是企業內部沒有適合的或是具備專業知識的人才，這讓企業常仰賴外部的教育資源和人才，當教育訓練項目被當成變數，且研究結果顯示教育訓練的不足或是不合時宜的教育訓練內容可能影響員工離職時，企業就必須考慮並設計適合的教育訓練課程，企業可以考慮和人力資源專家或是職業教育專家合作，這是一個長期的合作關係，而研究者必須提出適合的解決方案建議，無論是專家或是研究者，同時具備專業知識和豐富的工作經驗是必須。

## 参考文献

### 一、中文部份

中華民國勞動部. (2019). 勞動統計專網--勞動統計調查.

### 二、日文部份

大湾秀雄. (2017). 働き方改革と女性活躍支援における課題：人事経済学の視点から. RIETI ポリシーディスカッションペーパー, ページ: 17-P (2017): 006.

山本和史. (2017). 中小企業における新卒採用行動に関する実証分析. 日本労務学会誌, 18(1), 4-20.

野地祐二, 野々部 恵美子. (2018). 平成 30 年上半期雇用動向調査結果の概況. 日本国厚生労働省.

### 三、英文部份

Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms. *algorithms*, 4(5), C5.

BishopMC. (1995). Neural networks for pattern recognition. Oxford university press.

Chen, S. Y., Chuang, C. H., & Chen, S. J. (2018). A conceptual review of human resource management research and practice in Taiwan with comparison to select economies in East Asia. *Asia Pacific Journal of Management*, 35(1), 213-239.

Cheng, K. T., & Chang, K. (2019). Enhancing employee engagement for small and medium enterprises in Taiwan. *Information Resources Management Journal (IRMJ)*, 32(1), 28-47.

De Winne, S., Marescaux, E., Sels, L., Van Beveren, I., & Vanormelingen, S. (2019). The impact of employee turnover and turnover volatility on labor productivity: a flexible non-linear approach. *The International Journal of Human Resource Management*, 30(21), 3049-3079.

DESA, U. (2019). *World Population Prospects 2019: Highlights*. New York (US): United Nations Department for Economic and Social Affairs.

- Frederiksen, A. (2017). Job Satisfaction and Employee Turnover: A firm-level perspective. *German Journal of Human Resource Management*, 31(2), 132-161.
- Goldberg, J. (2014). Trends in Employee Turnover and Retention. *Graziadio Business Review*, 17(3).
- Hancock, J. I., Allen, D. G., Bosco, F. a., McDaniel, K. R., & Pierce, C. a. (2013). Metaanalytic review of employee turnover as a predictor of firm performance. *Journal of Management*, 39, 573–603.
- Hasegawa, R., & Hasegawa, S. (2017). Exploring the determinants of female workers' turnover intention in Japan. DAITO BUNKA UNIVERSITY, INSTITUTE OF BUSINESS RESEARCH.
- Hom, P. W., & Griffeth, R. W. (1995). Structural equations modeling test of a turnover theory: Cross-sectional and longitudinal analyses. *Journal of Applied Psychology*, 76, 350–366.
- Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, 102(3), 530.
- Khan, E. A., Khan, H., & Muhammad, S. (2019). Factors Affecting Employee Attrition and Predictive Modelling Using IBM HR Data. *Journal of Computational and Theoretical Nanoscience*, 16(8), 3379-3383.
- Lee, T. W., Hom, P. W., Eberly, M. B., Junchao (Jason) Li, & Mitchell, T. R. (2017). On the next decade of research in voluntary employee turnover. *Academy of Management Perspectives*, 31(3), 201-221.
- Local, C. (1917). News from cities and towns south of Tehachapi's top—Los Angeles county items. *Los Angeles Times*, 1886 –1922.
- Mowday, R. T., Porter, L. W., & Steers, R. (2013). *Employee—organization linkages: The psychology of commitment, absenteeism, and turnover*. Academic press.

- Porter, L. W., & Steers, R. M. (1973). Organizational, work, and personal factors in employee turnover and absenteeism. *Psychological Bulletin*, 80, 151-176.
- Puni, A., Agyemang, C. B., & Asamoah, E. S. (2016). Leadership styles, employee turnover intentions and counterproductive work behaviours. *International Journal of innovative research and development*, 5(1), 1-7.
- Singh, N. (2018). Factors Affecting High Employee Turnover in Indian Manufacturing Sector. In *n Harnessing Human Capital Analytics for Competitive Advantage* ( 209-225). IGI Global.
- Smith, J. D. (2018). *Successful Strategies for Reducing Employee Turnover in the Restaurant Industry*. Walden University.
- Wang, J. (2018). Hours underemployment and employee turnover: the moderating role of human resource practices. *The International Journal of Human Resource Management*, 29(9), 1565-1587.
- Zhang, Y. (2016). A review of employee turnover influence factor and countermeasure. *Journal of Human Resource and Sustainability Studies*, 4(2), 85-91.