國立政治大學

企業管理研究所(MBA 學位學程)

碩士學位論文

應用資料採礦技術於證券客戶風險偏好之分析
—以 P 公司為例

Applying Data Mining to Analyze Risk Appetite of Investors: The Case of P Securities Company

指導教授： 鄭宇庭 博士

研究生： 張紘嘉 撰

中華民國 一百一拾 年 陸 月

DOI:10.6814/NCCU202100509

謝誌

今年是個很不一樣的一年，當初想都沒想過研究所生涯會在疫情的肆虐下結束，雖然如此，還是很開心能夠順利畢業，取得碩士學位。這一路上受到許多人的幫助以及鼓勵，非常感謝家人支持我當初離開工作回校進修的這個決定，讓我沒有後顧之憂能專心於課業上。

另外，也非常感謝我的指導教授－鄭宇庭教授，當初修習鄭教授的巨量資料課程讓我對於數據分析產生興趣，因此毅然決然地往此領域鑽研並以此為我的論文研究題目，但一路上走得非常辛苦，一位商管背景的學生對於過去沒有碰過的知識以及量化的分析著實困難，鄭教授不僅補足我們在學識上的不足，也在他的引薦下讓我能在學生時期參與兩個相關的企業專案，讓我了解在實務上企業是如何利用資料科學解決問題，在教授以及團隊的努力下，很高興能在統一證券的 Fintech 專案中拿下評鑑第一名的殊榮；也很感謝典通數據公司讓我對於輿情分析以及情緒分析有了更深入的了解。

在此我也想特別感謝兩位研究所同學－黃子豪以及鄧絜心，數據分析領域對我們來說都很陌生，當初對於資料處理的方法以及演算法理論常常感到不知所措，感謝有你們的陪伴以及支持，在大家的努力下我們一起克服了很多困難，謝謝子豪在課業上的幫助，我們一起修習了許多外系的課程，也很順利的取得學分，另外，在寫程式的這條路上也感謝你與我一起努力。感謝絜心於碩一暑期時間一起到台大資訊班學習 Python，讓我對寫程式有了基礎的認識；也感謝你在課堂作業上的付出，那個時候常常為了分析報告忙到凌晨，在對資料處理手法不純熟的時期常常需要犧牲睡眠時間來完成課堂所要求的報告。非常慶幸在研究所這兩年中有志同道合的夥伴一同努力成長，也希望在未來的生涯中各位能夠鵬程萬里、一路順風，在各自的理想道路上持續前進！

張紘嘉　謹於

MBA 學位學程

中華民國 110 年 06 月

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100509

摘要

隨著科技的發展，Fintech 發展日益俱進，其中智能理財平台更是國內各大金融巨擘必爭之地。根據數據預估，2018 年至 2022 年全球自動化投資顧問資產規模每年以 38.2%的速度成長，到 2022 年時資產規模可達 1.45 兆美元，顯示越來越多的投資人選擇智能理財平台這項工具當作選擇投資標的之媒介。

因此本研究將利用 P 公司之客戶基本資料以及交易資料，運用資料採礦技術，輔以統計軟體，產出預測客戶風險偏好之模型。首先，依據交易金額以及購買的金融商品之報酬標準差加權平均出每一個客戶的加權平均標準差，依據此標準差將客戶分類成三群不同風險偏好：低風險群、中風險群以及高風險群；其次，結合客戶基本資料以及交易資料產出包含自變數以及應變數的分析資料，將此分析資料切分為訓練資料以及測試資料，訓練資料用於訓練機器演算法模型；測試資料則使用於檢驗該預測模型之準確率；最後，找出準確率最高之預測模型演算法。

由結果所產出之預測模型，未來可應用於在客戶還未實際投資的情況下，依據其所填寫之基本資料以及相關問題，產出該客戶之風險偏好，企業可依據此風險偏好推薦其相對應的投資組合產品，達到精準行銷之目的。

【關鍵字】 資料採礦、預測模型、風險偏好

DOI:10.6814/NCCU202100509

Abstract

With the development of technology, the development of Fintech is advancing day by day. Robo-advisor is a new wealth management platform that all the major financial company actively develop. According to estimates, the asset under management of robo-advisor will grow at a rate of 38.2% annually from 2018 to 20222, and the asset scale will reach US$1.45 trillion by 2022. This phenomenon showing that more and more investors are choosing this platform as a medium for selecting investment product.

Therefore, this research will use customer basic data and transaction data from P company, and supplemented by statistic software to produce a predicting model for predicting risk appetite from customer by data mining technology. Fist, based on the transaction amount and the standard deviation of the products purchased, calculated the weighted average standard deviation of each customer. Then according to this standard deviation, customers are classified into three groups with different risk appetite: low-risk group, medium-risk group, and high-risk group. Second, merge customer basic data and transaction data to produce analysis data including independent variable and dependent variable. Divide this data form into training data and testing data. The training data is used to train the machine learning model, and the testing data is used to evaluate the accuracy of the model. Finally, find the prediction model with the highest accuracy.

The prediction model selected by the research results can be used to predict customer's risk appetite before their actual investment behavior in the future. Based on the predict model results, produced by the basic information and related questions, the company can suggest corresponding investment portfolio to their customer according the predicted risk appetite and achieve the purpose of precise marketing.

DOI:10.6814/NCCU202100509

目　錄

摘要 ..... **I**

Abstract ..... **II**

目　錄 ..... **III**

表目錄 ..... **IV**

圖目錄 ..... **V**

第壹章　緒論 **1**

第一節　研究背景與動機 1

第二節　研究目的 4

第三節　研究流程 5

第貳章　文獻探討 **6**

第一節　智能理財平台概述 6

第二節　機器學習之文獻回顧 11

第參章　研究方法 **14**

第一節　資料來源 14

第二節　研究架構 14

第三節　操作性變數定義 16

第四節　分析方法 20

第肆章　實證分析 **24**

第一節　資料處理 24

第二節　模型建立 28

第三節　研究結果 30

第伍章　結論與建議 **37**

第一節　結論 37

第二節　建議 44

參考文獻 **46**

DOI:10.6814/NCCU202100509

表目錄

表 3-3-1 變數說明表.....16
表 4-1-1 客戶投資屬性問卷示意表.....24
表 4-1-2 客戶交易資料表.....26
表 5-1-1 混淆矩陣.....37
表 5-1-2 混淆矩陣(屬性偏好第一級).....38
表 5-1-3 風險屬性第一級召回率及精確率.....38
表 5-1-4 混淆矩陣(屬性偏好第二級).....39
表 5-1-5 風險屬性第二級召回率及精確率.....39
表 5-1-6 混淆矩陣(屬性偏好第三級).....40
表 5-1-7 風險屬性第三級召回率及精確率.....40
表 5-1-8 重要變數比較表.....38

DOI:10.6814/NCCU202100509

圖目錄

圖 1-1-1 Fintech 募資地圖 1
圖 1-1-2 基金風險報酬等級分類標準 2
圖 1-1-3 國人不願意投資理財的原因 3
圖 1-3-1 研究流程 5
圖 2-1-1 全球機器人理財發展歷程 6
圖 2-1-2 AUM 成長率 7
圖 2-1-3 AUM 規模 7
圖 2-1-4 機器人理財使用人數 8
圖 2-1-5 機器人理財之商業流程 9
圖 2-2-1 機器學習種類 12
圖 2-2-2 演算法時間軸 12
圖 3-2-1 研究架構流程 15
圖 3-4-1 類神經網路層次圖 23
圖 4-2-1 原始風險屬性分布 28
圖 4-2-2 平衡節點流程圖 28
圖 4-2-3 調整後風險屬性分布 29
圖 4-2-4 資料劃分結果圖 29
圖 4-2-5 建立模型流程圖 30
圖 4-3-1 曲線下面積圖 30
圖 4-3-2 類神經網路模型之矩陣圖 31
圖 4-3-3 類神經網路模型重要變數圖 32
圖 4-3-4 決策樹 CHAID 之矩陣圖 32
圖 4-3-5 決策樹 CHAID 模型重要變數圖 33
圖 4-3-6 決策樹 CHAID 之樹狀圖 33

DOI:10.6814/NCCU202100509

圖 4-3-7 決策樹 C5.0 之矩陣圖.....34
圖 4-3-8 決策樹 C5.0 模型重要變數圖.....34
圖 4-3-9 決策樹 CRT 之矩陣圖.....35
圖 4-3-10 決策樹 CRT 模型重要變數圖.....36
圖 4-3-11 決策樹 CRT 之樹狀圖.....36

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100509

第壹章　緒論

第一節 研究背景與動機

隨著科技的發展，電腦的運算能力提升造就人工智慧演算法的效用提高；各式各樣的互聯網行動裝置使的企業更容易取得相較以往更多更多元的數據；區塊鏈去中心化的發明使的既有的交易流程改變。這些創新不僅能讓人們生活更加便利，也能讓企業創造出更高的商業價值。

而這股科技創新風也吹向擁有百年歷史的金融業，金融科技(Financial technology，簡稱 FinTech)這一詞也成為了近幾年金融業的熱門關鍵詞，根據金融穩定理事會(FSB)的定義，金融科技是指科技技術為傳統金融服務所帶來的創新，它能創造新的商業模式、業務、流程與產品。在 2019 年 KPMG(Klynveld Peat Marwick Goerdeler，為四大會計事務所之一)所出具的「2019 年 FinTech100 金融科技創新報告」(2019 FinTech100 Report)指出，2019 年名列 100 名的金融科技公司，在過去 12 個月的募資金額已達 180 億美元以上；而募資超過 1 億美元的公司總計超過 32 間，這些企業遍佈世界各地，如圖 1-1-1 所示。

圖 1-1-1 Fintech 募資地圖（資料來源：KPMG）

DOI:10.6814/NCCU202100509

而台灣在 2017 年時由金管會通過《中華民國證券投資信託暨顧問商業同業公會證券投資顧問事業以自動化工具提供證券投資顧問服務（Robo-Advisor）作業要點》(以下簡稱 Robo-Advisor 作業要點)法規後，台灣的金融業者在符合適法性的規範下，也陸續推出許多智能理財平台的服務，期望可藉此平台吸引過去不易接觸到的客群，提高公司於財富管理業務上的收益。

根據過往傳統的財富管理業務流程，客戶與金融機構往來購買金融商品前，會先經過 KYC(Know Your Customer)的流程，需先填具一份投資屬性問卷，題目內容多為詢問客戶過往的投資經驗、預期報酬、預期持有期間等等相關問題，而每題答案皆有相對應的分數，最後將分數加總，依總分將客戶分類至不同的風險屬性中，風險屬性由低到高分為 RR1~RR6 級。另外，理財產品本身也有風險評級，主要是根據投資信託及投資顧問公會所公布的「基金風險報酬等級分類標準」，如圖 1-1-2，其分類標準是由基金類型、投資區域或投資產業由低到高區分為 RR1~RR5 級。

基金風險報酬等級分類標準

一、風險報酬等級說明：

投信投顧公會公佈「基金風險報酬等級分類標準」係依基金類型、投資區域或主要投資標的/產業，由低至高，區分為「RR1、RR2、RR3、RR4、RR5」五個風險報酬等級。

提醒投資人此等級分類係基於一般市場狀況反映市場價格波動風險，無法涵蓋所有風險，不宜作為投資唯一依據，投資人仍應注意所投資基金個別的風險。

二、風險報酬等級及主要基金類型：

(一)各類型基金之風險報酬等級以下表為原則

基金類型	投資區	主要投資標的/產業	風險報酬等級
股票型	全球	一般型(已開發市場)、公用事業、電訊、醫療健康護理	RR3
		一般型、中小型、金融、倫理/社會責任投資、生物科技、一般科技、資訊科技、工業、能源、替代能源、天然資源、週期性消費品及服務、非週期性消費品及服務、基礎產業、其他產業、未能分類	RR4
		黃金貴金屬	RR5
	區域或單一國家(已開發)	公用事業、電訊、醫療健康護理	RR3
		一般型、中小型、金融、倫理/社會責任投資、生物科技、一般科技、資訊科技、工業、能源、替代能源、天然資源、週期性消費品及服務、非週期性消費品及服務、基礎產業、其他產業、未能分類	RR4
		黃金貴金屬	RR5
	區域或單一國家(新興市場、亞洲、大中華、其他)	一般型(單一國家－臺灣)	RR4
		一般型、公用事業、電訊、醫療健康護理、中小型、金融、倫理/社會責任投資、生物科技、一般科技、資訊科技、工業、能源、替代能源、天然資源、週期性消費品及服務、非週期性消費品及服務、基礎產業、黃金貴金屬、其他產業、未能分類	RR5
債券型(固定收益型)	全球、區域或單一國家(已開發)	投資等級之債券	RR2
		高收益債券(非投資等級之債券) 可轉換債券 主要投資標的係動態調整為投資等級債券或非投資等級債券(複合式債券基金)	RR3
	區域或單一國家(新興市場、亞洲、大中華、其	投資等級之債券	RR2
		主要投資標的係動態調整為投資等級債券或非投資等級債券(複合式債券基金)	RR3
		他) 高收益債券(非投資等級之債券) 可轉換債券	RR4
保本型			按基金主要投資標的歸屬風險報酬等級
貨幣市場型			RR1
平衡型(混合型)			RR3 (偏股操作為 RR4 或 RR5)
金融資產證券化型	投資等級		RR2
	非投資等級		RR3
不動產證券化型	全球、區域或單一國家(已開發)		RR4
	區域或單一國家(新興市場、亞洲、大中華、其他)		RR5
指數型及指數股票型(ETF)			同指數追蹤標的之風險報酬等級
槓桿/反向之指數型及指數股票型(ETF)			以指數追蹤標的之風險等級，往上加一個等級
組合型基金			同主要投資標的風險報酬等級
其他型			同主要投資標的風險報酬等級

(二)區域(亞洲、大中華)股票型基金係指投信投顧公會委請之專家學者、理柏(Lipper)、晨星(Morningstar)或晶實資訊(股)公司等基金評鑑機構所作評比資料之基金類型為亞洲或大中華區域者。

(三)不動產證券化型基金包括不動產證券化型之投信基金，與主要投資不動產證券化商品之股票型境外基金。

(四)基金無法歸類為上述第 1 項所列主要基金類型者，證券投資信託事業、總代理人及基金銷售機構應自行衡量基金屬性、投資策略及風險特性揭露該基金風險報酬。

圖 1-1-2 基金風險報酬等級分類標準

（資料來源：中華民國證券投資信託暨顧問商業同業公會）

DOI:10.6814/NCCU202100509

在經過投資屬性問卷評估後，傳統理財專員會依據其結果推介相對應的理財產品，例如：針對投資屬性 RR2 的投資人，業務人員可能就會推薦全球投資等級之債券型基金。然而，經由數據分析比對後發現，大部分的投資人風險屬性會與其購買的金融商品風險屬性不符，以 P 公司為例，不符比例約達 61.15%。根據「行為財務學在機器人理財顧問的運用、實務與挑戰」(2018, 曹書華)一文指出，可能的原因為此問卷為制式化問題，並且單純以分數加總的方式進行計分及分類，並無法完整呈現客戶的實際投資需求。

另外，根據《Cheers 雜誌》於 2019 年的網路調查，國人不願意投資理財的原因中，第一名為資金不足(56.1%)；第二名為沒時間研究商品或市場(46.8%)；接續為害怕虧損(32.4%)以及認為申辦流程繁瑣(14.0%)。

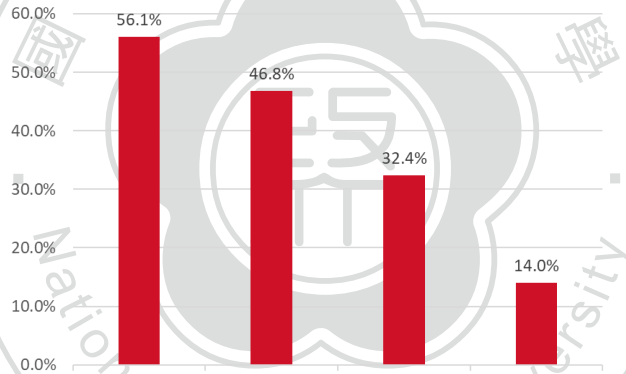

圖1-1-3 國人不願意投資理財的原因(資料來源：Cheers 2019網路調查)

而金融服務機構可以嘗試解決這些問題，讓客戶的投資意願上升，進而提升企業的獲利。

因此綜述以上幾點，本研究欲以消費者的需求為出發點，嘗試以非傳統既有的模式找出方法解決流程繁瑣之問題，並提高建議之投資組合符合客戶風險偏好屬性的比例。

DOI:10.6814/NCCU202100509

第二節 研究目的

近年來，台灣金融業者各自打造自家的智能理財平台，投資者會面對如此眾多的選擇，哪個平台能解決投資者的問題就越能得到青睞。而開發此智能理財平台之最終目的即是促使客戶於平台上購買公司之金融商品，進而提升營收。因此發展方向應從投資者的角度出發，針對其痛點進行改善。金融業者可以針對智能理財平台之商業模式中的各個階段進行改良，例如：利用改善演算法提高投資者獲利的勝率；或者改善平台交易流程，縮短執行交易階段的時間等等。

本研究採用 P 證券公司之客戶基本資料、投資屬性問卷資料以及交易資料，以客戶購買之金融商品報酬標準差作為應變數；其餘客戶資料作為自變數，建立預測模型，並檢視其模型準確率。本研究主要目的如下：

1. 建立預測力良好之模型，未來只要輸入相對應的客戶資料即可產出該客戶的風險偏好，P 公司即可運用此結果推薦客戶相對應的投資組合，達到精準行銷的目的，增加銷售成功率。
2. 運用模型自動化的特性，縮短客戶建檔階段至資產配置階段的時間，相較於傳統理財業務的流程，投資者可大幅減少其申辦時間，不僅可節省時間，也可提高客戶體驗。

DOI:10.6814/NCCU202100509

第三節 研究流程

本研究之整體研究流程，如圖 1-3-1 所示，且本文的研究架構安排如下：第一章為緒論，第二章為文獻探討，第三章為研究方法，第四章為實證分析，第五章為結論與建議。

圖 1-3-1 研究流程 (資料來源：本研究整理)

DOI:10.6814/NCCU202100509

第貳章　文獻探討

第一節 智能理財平台概述

金融科技在近幾年中發展迅速，目前所觸及之領域大致可分為：支付清算、借貸融資、財富管理、零售銀行、保險以及交易結算此六大領域。其中，財富管理領域的相關科技運用將是本研究主要討論的議題。2008 年金融海嘯發生後，投資人對於銀行失去信心與信任感，在評價模型演算法以及大數據分析的支持下，美國各家金融機構開始推出機器人理財相關平台(Robo-advisor)，到目前為止，美國的機器人理財規模也是全球最大的。而全球機器人理財的發展歷程如圖 2-1-1。

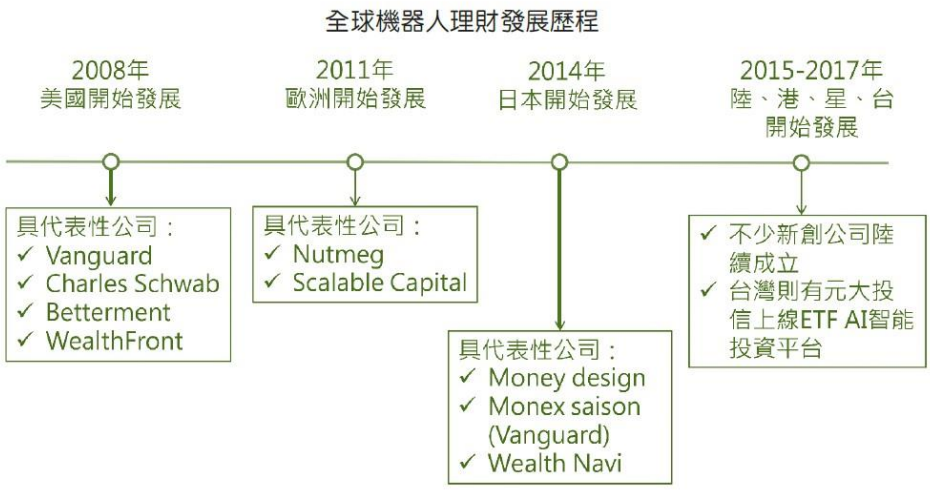

圖 2-1-1 全球機器人理財發展歷程

（資料來源：2019 證券暨期貨月刊 第三十七卷 第一期 專題二，張明珠）

新興亞洲例如台灣、香港也相繼在 2015 年開始發展機器人理財，時至今日，可以看到各大金融機構都有推出自己的智能理財平台或類似概念的相關產品。

而根據數據預估，2018 年至 2024 年全球自動化投資顧問資產規模每年成長率雖逐漸下滑但還是有雙位數的成長(圖 2-1-2)；而整體資產規模(Assets Under Management)到 2024 年時資產規模可達 2.7 兆美元(圖 2-1-3)。

DOI:10.6814/NCCU202100509

圖 2-1-2 AUM 成長率 (資料來源：Statista)

圖 2-1-3 AUM 規模 (資料來源：Statista)

由此顯示，在成長率趨於平穩以及整體資產持續上升之下，機器人理財對於傳統金融業財富管理業務上的重要性也會越來越高。再者，從使用該服務的人口來看(圖 2-1-4)會發現，使用人數從 2017 年一路成長，未來於 2024 年可望達到 4 億多人口使用機器人理財的服務，年成長率約 40%，因此不管是從質(使用人數)還是量(管理資產)來看，機器人理財或智能理財平台會是金融業在財富管理業務領

DOI:10.6814/NCCU202100509

域中下一個紅海市場。

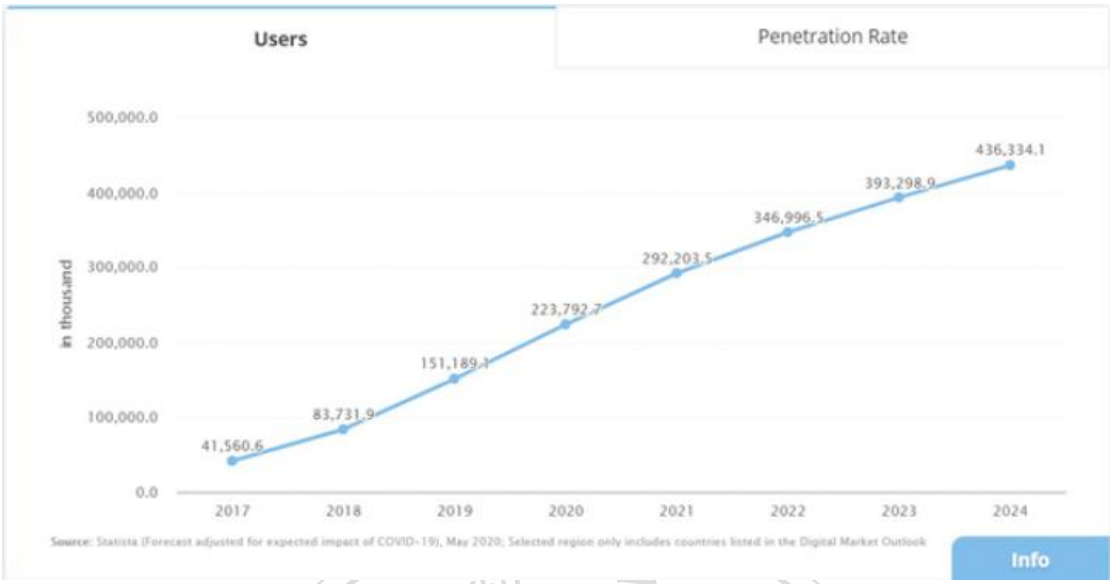

圖 2-1-4 機器人理財使用人數 (資料來源：Statista)

而會造成這樣的風潮，其中的原因可以歸因於機器人理財可以消除常見的投資人偏誤，部分投資人於投資市場經常出現不理性的行為發生，然而機器人理財是基於演算法決定買賣時機，因此較不易出現追高殺低的情況發生。另外，根據「機器人理財服務是否可有效消除投資人偏誤」一文所述(2020, 康力中)，機器人理財不存在處分效果，目前市面上的機器人理財平台並沒有顯著傾向賣掉持股中獲利的資產，留下虧損的資產之情形，投資人經常會對虧損的部位產生惜售的心態，對於有獲利的部位則會毫不猶豫地賣出，這也使的實際損益經常為正，但實質上的帳戶損益卻是負的。再者，機器人理財也較不易出現賭資效應以及蛇咬效應，這兩種效應也為一般投資人較常在投資市場上出現的心理現象，賭資效應為投資人容易將獲利的部分視為額外的收入，因此對獲利的處置以及投資方式會跟原始本金不同，對於獲利的部位會表現出較高的風險偏好，投資於高風險的金融商品，對於可能的潛在損失所表現出來的在意程度也會較原本本金低；至於蛇咬效應則是投資人與金融市場失利時，常會找理由來掩飾自己的失敗，例如：投資本來就是賭博、投資可能不適合自己等等的言論，利用這些理由來讓自己的心情平復。上述這兩種心理現象都可以利用機器人理財平台來消除，由於此平台是

DOI:10.6814/NCCU202100509

由一系列演算法以及數學模型所組成，人為的操作影響較小，大部分為半自動化的投資流程，因此較能去除人的非理性選擇，進而用更理性的方式來面對投資市場。

因此對於企業來說，機器人理財服務對於其客戶是有正面的影響的，根據「金融機構導入理財機器人關鍵成功因素之研究」(2019, 陳雅婷)，企業在選擇導入此服務會有以下重要考慮因素：投入成本、機會影響、績效期望以及易用預期，藉此可知道企業在引進新的應用時所注重的還是在於成本以及效益的考量。效益的提升可以藉由增加使用者以及提高單位管理資產來達成，而解決消費者的需求問題即是一個例子，上述對於消費者不願意投資的原因可作為改善的出發點，從消費者的觀點出發來構想機器人理財平台可達成的預定目標；另外在成本方面，企業可以藉由改善流程進而縮減建置成本，機器人理財的商業流程基本上切分為七大項目(如圖 2-1-5)：

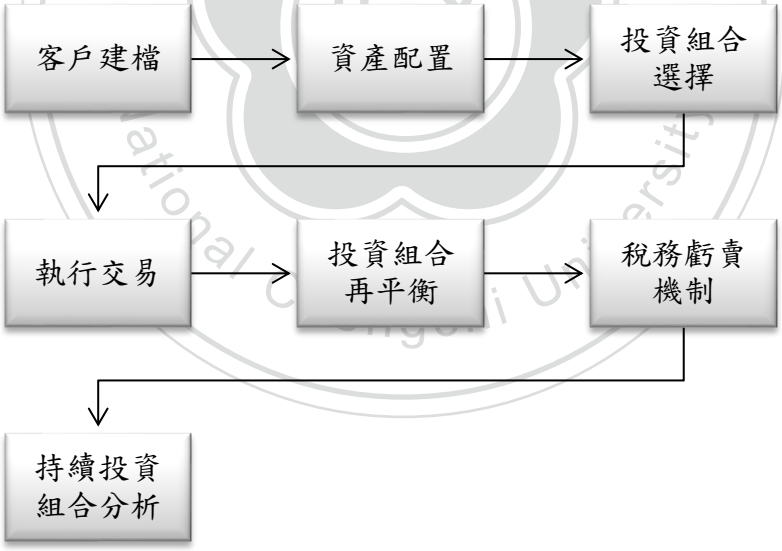

圖 2-1-5 機器人理財之商業流程

(資料來源：2019，理財機器人於台灣理財市場之適用性探討，莊子郁)

1. 客戶建檔：主要進行客戶的 KYC(Know Your Customer)，針對客戶的基本資料、投資知識、投資經驗、財務狀況以及風險承受度等相關問題進行了解，根據我國 Robo-Advisor 作業要點，該評估方式須取得足夠之資訊，對於問

DOI:10.6814/NCCU202100509

題表達方式須明確，另外也需要有防呆機制，避免客戶的答案有矛盾的情形產生。

2. 資產配置：利用演算法產生建議之投資組合，而各家業者對於演算法皆有不同的選擇方式，較常見的有隨機森林、蒙地卡羅方法等等。國內法規對於演算法也有相關的規定，主要著重在初期的審核以及後續定期的監管。
3. 投資組合選擇：客戶可以從不同的投資組合中做選擇，目前有些國內業者會先利用演算法產出各種不同風險程度的投資組合，例如將建議的投資組合分類成積極型、成長型以及保守型這三大類，消費者可以依照自己的喜好以及目的去選擇適合的投資組合。
4. 執行交易：此階段須將機器人理財平台與原有的交易系統平台去做串接，讓客戶可以在選擇完投資組合後，系統即可為客戶進行交易，將投資組合內的所有金融產品依照建議比例購置於客戶的帳上。
5. 投資組合再平衡：由於在投資期間中，各資產的報酬率皆不相同，使的原本所建議的資產比例會有調整的必要，以期達成原始設定的目標或者降低投資組合的風險。而國內的法規針對此部分也有規範，主要為資訊揭露的相關規定以及執行期間、執行方式等等。
6. 稅務虧賣機制：有些服務業者例如 Betterment 會有此功能，意即利用演算法的偵測，將帳上虧損的部位賣掉，降低整體的資本利得，達到減稅的機制，但此方法只適用於徵收資本利得相關稅務的國家。
7. 持續投資組合分析：金融市場瞬息萬變，因此需要透過不斷的分析調整來平衡改善原本的投資組合，針對可能的突發狀況制定調整策略，在風險最小化的前提下提高效益。

綜述以上，金融服務機構可以利用不同的切入點來建構自有的機器人理財平台服務。由於台灣國人的數位化金融服務比率較低，根據數據顯示其比率僅為 32.6%，世界排名第 32 名，最大的差異在於高收入族群與低收入族群(2021，陳鴻)，高收入族群相較於低收入族群較常使用到金融服務，因此對於數位化的服務該族群

DOI:10.6814/NCCU202100509

的接受度也較高。由此可知，金融數位科技在國內仍是具有發展潛力的領域，在這片藍海市場中要如何找尋服務以及產品的定位，提高民眾的使用率，也許是企業可以思考的方向，以實現「普惠金融」之目標。

第二節 機器學習之文獻回顧

預測模型為利用統計技術以及大量的歷史數據資料，針對未來可能發生的事件、數量、金額等變量進行預判，拜科技發展所賜，電腦的算力提升、AI 人工智慧的發展，都讓現今的預測模型能夠更加準確。而機器學習(Machine Learning, ML)則是 AI 人工智慧領域中在商業分析發展應用最廣的技術，根據 Tom Mitchell 的定義，機器學習為「一個程式在執行某種任務時被認為具備學習的能力，那麼這個程式一定需要展現出兩點：1. 利用現有的經驗。2. 不斷改善並完成任務的特性及性能。」(Tom Mitchell，1997，Machine Learning)，因此可以發現機器學習有兩個要點，一個是過去的經驗；另一個則是不斷的改善，過去的經驗某種程度上來說可以影射為數據資料，也就是訓練機器所需要的素材；而不斷的改善則是訓練的過程。

時至今日，科技的發展使得我們每天都會產生大量的數據，而數據的取得也變得更加方便以及快速，這讓企業可以利用這些素材進行機器學習，用來產生更好的商業價值。廣義上來說，機器學習可分為監督式學習(Supervised Learning)、非監督式學習(Unsupervised Learning)、深度學習(Deep Learning)以及強化學習(Reinforcement Learning)，監督以及非監督式學習差別在於對於資料有無標籤，意即是否有將資料定義及說明，較常見的監督式學習演算法有邏輯回歸、決策樹、SVM 等等；而非監督式學習演算法則有 K-means 集群、階層式分群等等。但近幾年來最大的應用則是在深度學習方面，舉凡自動駕駛中的影像辨識系統又或者智慧型手機裡的語音助理都是利用深度學習的演算法，根據定義，深度學習則是讓機器擁有自我學習的能力(李宏毅，2016)，就像是人類的大腦，藉由函數集去模擬人類的神經元，當變數輸入後會透過複雜的網絡產出相對應的結果。

DOI:10.6814/NCCU202100509

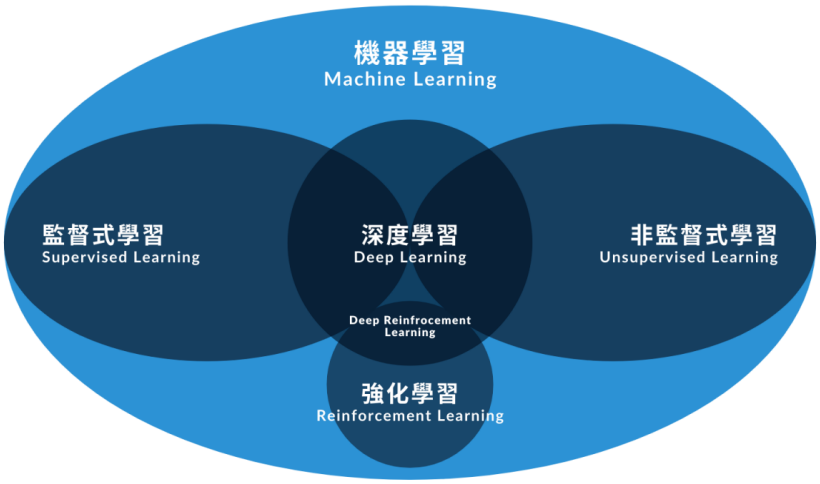

圖 2-2-1 機器學習種類（資料來源：OOSGA，2021）

然而另外有趣的是，機器學習的演算法其實早在 20 世紀初就開始出現了，於 1959 年 IBM 就提出了會學習的西洋棋程式，經過了多年的發展，後續也出現許多重要的演算法以及學習方式。一些常見的演算法及發展年代如圖 2-2-2：

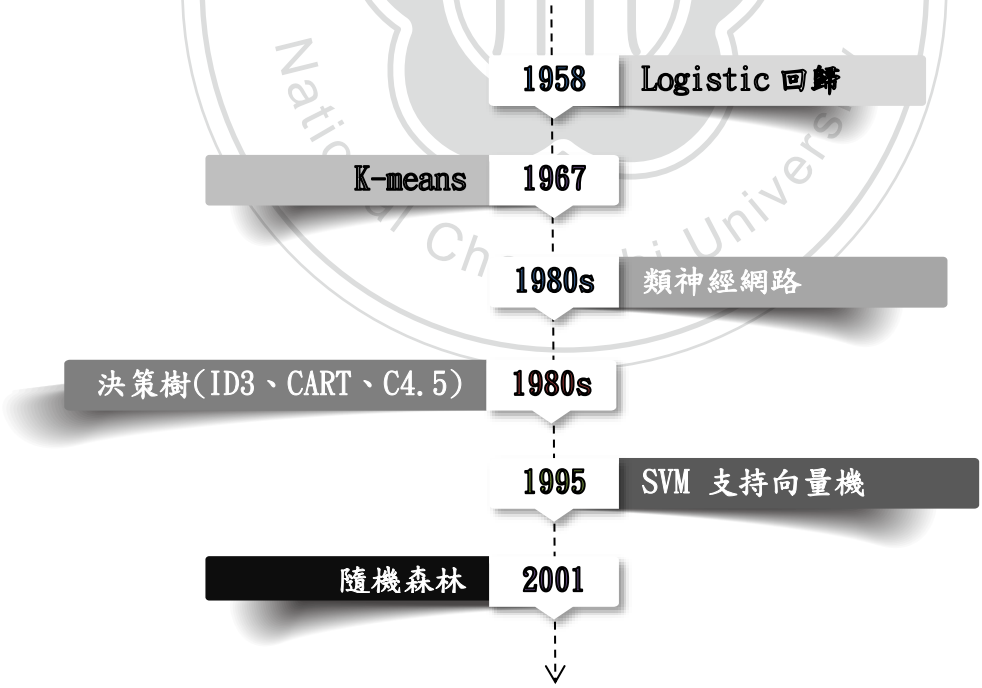

圖 2-2-2 演算法時間軸（資料來源：本研究整理）

機器學習發展至今應用的相當廣泛，各行各業都有利用機器學習來解決商業

DOI:10.6814/NCCU202100509

問題的例子。根據「2019 中國人工智能獨角獸白皮書」(新智元，2019)中，在中國大陸估值 1,000 億人民幣以上的 AI 獨角獸企業共有八家，例如：螞蟻金服、字節跳動，這些企業廣泛分布於金融業、文化媒體業、運輸業等等。字節跳動利用機器學習發展人工智能相機甚至於自然語言處理(Nature Language Processing, NLP)領域當中，現今在抖音上看到的各式各樣濾鏡都是利用人臉辨識技術，將畫面中的眼、鼻、口等重要器官辨識出來後再美化構圖；而自然語言處理則可發展文字探勘(Text Mining)以及情緒分析(Sentiment Analysis)技術，此技術廣泛應用於網路社群上的聲量分析以及群眾情緒走向。

而在金融業的應用中，以螞蟻金服為例，該企業運用大規模的深度學習去判斷一個帳戶被盜的風險，提升風險控管的能力；另外，在財富管理業務上，該平台使用神經網路的組合進行輿情分析以及情緒分析，判斷市場對於投資標的的輿論正負值，多方輿論為正值；空方輿論為負值，藉此來輔助投資方面的決策，此為較新型態的分析方法，有鑑於網路社群的蓬勃發展，人們在社群平台上討論投資理財的議題也越來愈多，因此可以收集到大量的文本資料進行分析。而在量化投資領域也大量使用機器學習的技術作為買賣的判斷標準，國內外也有許多關於這類型的研究，例如使用公司財報中的數據變量作為變數，利用 SVM 支持向量機模型來預測台股未來的走向，實證研究後發現該方法確實在賺取超額報酬方面有一定的作用(卓越，2019)；或者利用 K 線等技術指標作為變數，以類神經網路作為預測模型，預測大陸股票市場的後續走勢，以此方法所建構的投資組合及策略大在多數的年度，其年化報酬率皆優於大盤(張力元，2018)。

由此可知，機器學習在投資理財上的應用是相當的純熟，此應用技術的優勢在於機器人的〝理性〞，模型在判斷買賣訊號以及價位時所運用的是實際的數據，而傳統的投資人在做投資決策時往往會參雜個人的情緒心理因素，導致偏誤的發生，這也可以說明為何量化投資漸漸變成現今金融業主流的投資決策。

DOI:10.6814/NCCU202100509

第參章 研究方法

本章節分為四個小節，第一節主要揭露本研究的資料來源以及資料期間；第二節為本研究的架構，利用 Data Mining 的資料處理流程來進行資料的預處理以及後續的建模架構，資料預處理流程分為遺漏值處理、無效值處理以及資料合併；第三節為研究變數的說明，資料經過處理後在進行分析前，需先定義各項變數的意義及內容；最後一節則為本研究的方法理論，主要介紹決策樹以及類神經網路，這兩種演算法為大數據分析常使用到的模型。

第一節 資料來源

本研究使用 P 公司的資料作為研究以及實證資料，使用的資料如下：

1. 客戶投資屬性問卷，資料期間為：2017 年~2020 年，原始資料共計 400,925 筆資料。
2. 客戶交易資料，資料期間為 2016 年~2020 年，原始資料共計約 420 萬筆。
3. 金融商品基本資料表
4. 境內外基金淨值表，資料期間為 2005 年~2020 年。

由於保密條款的關係，客戶投資屬性問卷以及客戶交易資料中具有識別客戶身分之資料皆已刪除，例如：姓名、身分證字號、地址、電話等等。每筆客戶資料皆用編號去識別化。

第二節 研究架構

本研究目標為利用 Data Mining 之技術建構預測力良好的模型，將此模型應用於商業場景中，而預測力越高越能準確的提供相對應風險之金融商品，達到精準行銷的目的。因此，研究架構著重於資料的處理以及模型的建構，本研究架構流程如下：

DOI:10.6814/NCCU202100509

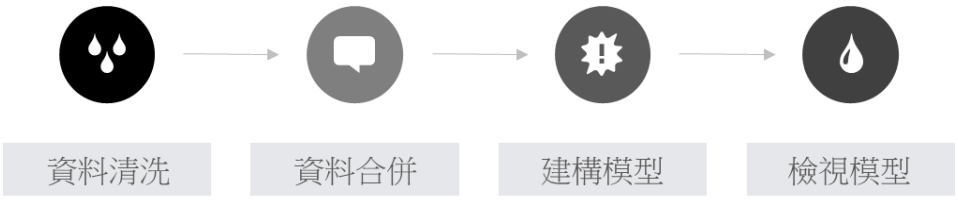

圖 3-2-1 研究架構流程(資料來源：本研究整理)

1. 資料清洗(Data Cleaning)：資料清洗是指發現並糾正數據中可識別的錯誤之流程，包括檢查數據一致性，處理無效值和遺漏值等。

(1) 一致性檢查(consistency check)

是根據每個變數的合理取值範圍和相互關係，檢查數據是否合乎要求，發現超出正常範圍、邏輯上不合理或者相互矛盾的數據。

(2) 無效值和遺漏值的處理

由於調查、編碼和紀錄誤差，數據中可能存在一些無效值和缺失值，需要給予適當的處理。本研究之處理方法有：估算、變數刪除和整列刪除。

- 估算(estimation)：利用某個變數的樣本平均值、中位數或眾數代替無效值和遺漏值。這種方法較為簡單，但沒有充分考慮數據中已有的訊息，誤差可能較大。另一種方法就是根據調查對象對其他問題的答案，通過變數之間的相關分析或邏輯推論進行估計。例如：某筆資料於婚姻欄位為遺漏值，但在子女數欄位有值，因此可以合理推斷該筆資料客戶為已婚，並將此值填入婚姻欄位。
- 變數刪除(variable deletion)：如果某一變數的無效值和遺漏值很多，而且該變數對於所研究的問題不是特別重要，則可以考慮將該變數刪除。這種做法減少了供分析用的變數數目，但沒有改變樣本量。
- 整列刪除(row deletion) 刪除含有遺漏值的樣本。由於很多問卷都可能存在遺漏值，這種做法的結果可能導致有效樣本量大大減少，

DOI:10.6814/NCCU202100509

無法充分利用已經收集到的數據。因此，只適合關鍵變數缺失，或者含有無效值或遺漏值的樣本比重很小的情況。

2. 資料合併(Data Merging)：將不同來源的資料依照相同的排列格式進行合併，進而產生出由客戶編號為列資料，後續以欄的方式呈現該客戶的各項變數，例如：年齡、年收入、風險問卷之填答答案等等。將此資料作為建構模型使用。

3. 建構模型(Building Model)：使用統計軟體，將資料切分為訓練資料以及測試資料，訓練資料用於機器學習，測試資料用於驗證模型。演算法的挑選以軟體內建為主，主要是以類神經網路(Neural Network, NN)以及決策樹(Decision Tree)來建構預測模型。

4. 檢視模型(Model Evaluation)：利用軟體中 Evaluation 節點以及混淆矩陣(confusion matrix)來檢視模型的預測能力，評估每一風險偏好族群中的預測正確率，以每一族群正確率皆較高以及總預測正確率較高者為主。另外，觀察每一模型中的重要變數，以了解模型是運用哪些變數將客戶分類，此變數可以提供企業未來在改善既有流程上的依據。

第三節 操作性變數定義

所有資料表經過資料清洗以及合併後產出訓練模型資料表，以客戶風險偏好為依變數，其餘變數為自變數建構模型。每一列皆為一個客戶，不重複，每一欄位則顯示自變數，以下為變數說明：

表 3-3-1 變數說明表

變數名稱	說明
編號	去識別化後的代號，每一客戶的編號是唯一的。
學歷	1：博士；2：碩士；3：大學；4：大專；5：高中(職)；

DOI:10.6814/NCCU202100509

	6：國中；7：小學；8：小學以下。
婚姻狀況	1：未婚；2：已婚。
持有重大傷病卡	1：是；2：否。
職位別	1：負責人；2：中/高階主管；3：基層主管；4：一般職員；5：其他。
行業別	A：農林漁牧礦業；B：軍警消公教業；C：製造業；D：醫療保健及社會工作服務業；E：批發及零售業；F：住宿及餐飲業；G：藝術娛樂及休閒服務業；H：金融業；I：資訊及科技業；J：營建業；K：律師、會計師、不動產經紀業；L：家管、退休、學生、自由業；M：珠寶、銀樓、典當業；N：酒家、舞廳、電玩等特種行業；O：博弈、軍火商；P：公證人、記帳士、地政士及報稅代理人。
個人平均年收入	1：50 萬以下；2：50~99 萬；3：100~299 萬；4：300~499 萬；5：500~999 萬；6：1000 萬以上。
家庭平均年收入	1：100 萬以下；2：100~299 萬；3：300~499 萬；4：500~999 萬；5：1000 萬以上。
不動產	為個人不動產總值扣除貸款後之價值。1：無；2：500 萬以下；3：500~999 萬；4：1000~2999 萬；5：3000~4999 萬；6：5000 萬以上。
資產總值	為個人扣除貸款後不動產總值+其他資產。1：300 萬以下；2：300~499 萬；3：500~999 萬；4：1000~2999 萬；5：3000~4999 萬；6：5000 萬以上。

DOI:10.6814/NCCU202100509

投資資金來源	A1-薪水	0：否；1：是。
	A2-退休金	0：否；1：是。
	A3-閒置資金	0：否；1：是。
	A4-投資收益	0：否；1：是。
	A5-其他基金轉入	0：否；1：是。
	A6-其他	0：否；1：是。
預計可供規劃資產		1：100 萬以下；2：100~299 萬；3：300~499 萬；4：500~999 萬；5：1000 萬以上。
投資目的	B1-財富累積	0：否；1：是。
	B2-子女教育基金	0：否；1：是。
	B3-退休規劃	0：否；1：是。
	B4-節稅規劃	0：否；1：是。
	B5-資產轉移	0：否；1：是。
	B6-其他	0：否；1：是。
投資收益入帳頻率		某些金融商品(例如：債券、開放式基金)會固定配發利息，頻率會因商品架構有所不同。1：每月固定；2：每季固定；3：每半年固定；4：每年固定；5：不定期；6：其他。
年齡		1：71 歲以上或未滿 20 歲；2：61~70 歲；3：51~60 歲；4：41~50 歲；5：20~40 歲。
備用金狀況		當遇到突發狀況，家庭備用金足夠支應多久的開銷。1：無備用金；2：3 個月以下；3：3~6 個月；4：6~9 個月；5：9 個月以上。

DOI:10.6814/NCCU202100509

投資經驗		過去投資具有價值波動性的金融商品有多少年之經驗。1：無；2：3 年以下；3：3~6 年；4：7~10 年；5：10 年以上。
目前持有商品	C1-現金型	0：無；1：有。
	C2-債券型	0：無；1：有。
	C3-外幣型	0：無；1：有。
	C4-股票型	0：無；1：有。
	C5-衍生性商品	0：無；1：有。
風險性商品占比		投資資產中有多少比例屬於價值波動性產品。1：0%；2：0~10%；3：10~25%；4：25~50%；5：50%以上。
風險承受度		過往經驗中可接受之價格波動。1：介於-5%與+5%之間；2：介於-10%與+10%之間；3：介於-15%與+15%之間；4：介於-20%與+20%之間；5：±20%以上。
資金運用偏好		每月收入有多少比例可用於投資或儲蓄。1：5%以下；2：5%~10%；3：10%~25%；4：25%~50%；5：50%以上。
金融商品持有期間		投資於價格波動性產品時，通常願意持有多久時間。1：1 年以下；2：1 年~3 年；3：4 年~5 年；4：6 年~10 年；5：10 年以上。

資料來源：本研究整理。

DOI:10.6814/NCCU202100509

第四節 分析方法

本研究使用決策樹以及類神經網路兩個演算法建立預測模型，預測客戶的風險偏好是屬於高風險、中風險還是低風險的客群。以下為各個演算法的基本原理以及介紹：

一、決策樹(Decision Tree)：

決策樹為一個預測模型；所代表的為對象屬性與對象值之間的一種映射關係。樹中每個節點表示某個屬性，而每個分叉路徑則代表某個可能的屬性值，而每個葉節點則對應從根節點到該葉節點所經歷的路徑所表示的值。每個決策樹都表述了一種樹型結構，它由它的分支來對該類型的對象依靠屬性進行分類。每個決策樹可以依靠對資料庫的分割進行數據測試。這個過程可以遞歸式的對樹進行修剪。當不能再進行分割或一個單獨的類別可以被應用於某一分支時，遞歸過程就完成。較常見的算法如下：

1. C5.0
2. CHAID(CHi-squared Automatic Interaction Detector)
3. CRT(Classification and Regression Trees)

而每種算法的差異取決於使用不同的指標來分割往下增生，分割的主要原則是能得到最大的資訊增益(Information Gain, IG)

資訊增益(Information Gain, IG)：

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^{m} \frac{N_j}{N_p} I(D_p)$$

$I(D_p)$：原本的資訊量

$\sum_{j=1}^{m} \frac{N_j}{N_p} I(D_p)$：分割後的資訊量

為了使獲得的資訊量最大化，因此經由分割的資訊量要越小越好，而衡

DOI:10.6814/NCCU202100509

量資訊量常見的有兩種方法：

1. 基尼不純度(Gini Impurity)：

基尼不純度表示一個隨機選中的樣本在子集中分類錯誤的可能性。基尼不純度為這個樣本被選中的概率乘以分類錯誤的機率。當一個節點中所有樣本都是一個類時，基尼不純度為零。假設 y 的可能取直有 C 個類別，另$i \in \{1,2,\ldots\ldots,C\}$，$P_i$表示被標定為第 i 類的機率，則基尼不純度的計算為：

$$I_G(t) = \sum_{i=1}^{c} p(i|t)(1 - p(i|t)) = 1 - \sum_{i=1}^{c} p(i|t)^2$$

2. 熵(Entropy)：

熵表示對隨機變量不確定度的度量，當 X 為一個離散隨機變量時，其熵 H（X）定義為：

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i$$

而條件熵 H（Y|X）表示在已知隨機變量 X 的條件下隨機變量 Y 的不確定性。隨機變量 X 給定的條件下隨機變量 Y 的條件熵 H(Y|X)定義為：

$$H(Y \mid X) = -\sum_{i=1}^{n} p_i H(Y \mid X = x_i)$$

二、類神經網路(Neural Network)：

為一種模仿生物神經網路的結構和功能的數學計算模型，用於估計以及預測。神經網路由大量的人工神經元連結進行計算，具備學習能力。而現代神經網路是一種非線性統計資料建模工具，透過基於數學統計學

DOI:10.6814/NCCU202100509

類型的學習方法達到最佳化。一般來說，類神經網路擁有下列三個部分：

- 結構(Architecture)：決定網路的樣貌以及變數間的關係，常見的有單層神經元網路及多層神經元網路。
- 激發函式(Activation Rule)：每一個神經元皆為一個激發函式，當對一個神經元輸入一個輸入值(input)後，經過激發函式的運算，產出輸出值(output)，而這個輸出值就會再傳入下一個神經元，成為該神經元的輸入值。
- 學習規則(Learning Rule)：為該神經網路調整變數權重以降低誤差值的方法。其學習過程可分為下列兩個步驟：

  1. 前向傳播(Forward-Propagation)：

     將特徵向量經由數個神經元由左至右傳遞，產出輸出值，為單一方向的資訊傳遞。例如：將車子圖片中的特徵向量透過神經網路傳遞，中間會經過數個神經元上，藉由激發函式的運算產出最終的預測答案。

  2. 反向傳播(Backward-Propagation)：

     於前向傳播的概念中加入了更正錯誤的方法，即為代價函數(Cost function)。代價函數是預測結果與真實結果之間的差距。如果預測準確則代價函數為 0；如果兩者間的差距越大，則代價函數的值也越大。因此，為了使預測結果與真實結果的差距越小，意即代價函數最小化，神經網路會將誤差結果由右至左回傳，藉此調整神經元的權重以找到代價函數的最小值。經過重複動作後，直到錯誤率最低，該網路即可正確預測。

而一般多層次的網路結構是由三個部分所組成，如圖 3-4-1：

- 輸入層(Input layer)：眾多神經元於此層接受大量非線性資料，輸入的資料稱為輸入向量。
- 隱藏層(Hidden layer)：介於輸入層與輸出層之間，由眾多神經元以

DOI:10.6814/NCCU202100509

及連結所組成，可以為一層或多層。於此層中的節點越多代表此神經網路的非線性關係越顯著。

- 輸出層(Output layer)：資訊於眾多神經元中傳遞、分析、最後形成結果，而輸出的資訊稱之為輸出向量。

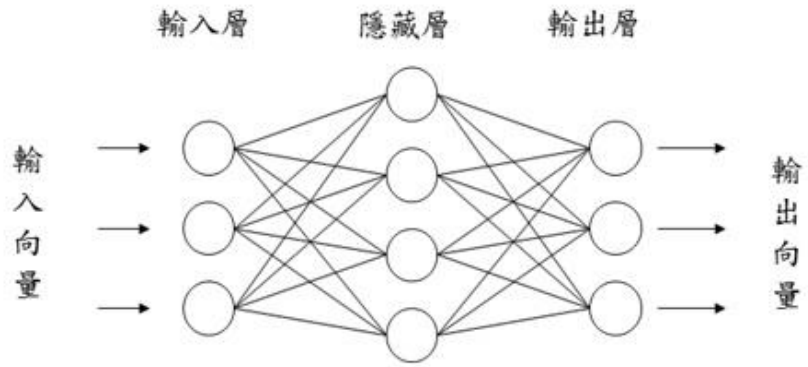

圖 3-4-1 類神經網路層次圖（資料來源：SAS）

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100509

第肆章　實證分析

本章分為三個部分，第一節介紹各個資料表的處理方式以及過程，為了產出訓練模型用的資料集，需要將資料表經過清洗、篩選、刪除、合併等步驟產出含有依變數以及自變數的資料表；第二節介紹模型的建構過程，建立模型使用統計軟體，所使用的模型演算法為決策樹以及類神經網路；第三節則針對各個模型的結果進行分析及評估。

第一節 資料處理

本節主要介紹各個資料表的預處理過程，包含：資料清洗、遺漏值處理以及變數刪除等等，將各個資料表處理完後就可進行資料表合併產出訓練資料集，再藉由此資料集建立模型。

一、　客戶投資屬性問卷資料

此份資料日期為 2017 年至 2020 年，各年份原始資料量如下：2017 年 58,116 筆；2018 年 164,227 筆；2019 年 104,806 筆；2020 年 73,776 筆；共計 400,925 筆。內容主要紀錄客戶的帳號、填答日期、各題目的填答答案。題目分為 A、B、C 三個部分，A 部分共 13 題，主要為客戶的基本資料，例如：年齡、性別、職業別、等等；B 部分則為法人戶的專業題目；C 部分共 8 題，內容則是詢問客戶對於投資理財的一些經驗以及偏好。示意表格如下：

表 4-1-1 客戶投資屬性問卷示意表

異動日期	帳號	題目類型	題目題號	答案內容
20200102	0000000056	A	1	02
20200102	0000000056	A	2	02
20200102	0000000056	A	3	04
…				
20200102	0000000056	A	12	01,03
20200102	0000000056	A	13	05
20200102	0000000056	C	1	05

DOI:10.6814/NCCU202100509

20200102	0000000056	C	2	02
20200102	0000000056	C	3	01
20200102	0000000056	C	4	01,02,03,04,06
…				
20200102	0000000056	C	8	02
20200102	0000000105	A	1	01
20200102	0000000105	A	2	05
…				

資料來源：本研究整理

經由檢查後發現此份資料並無遺漏值，但由於 B 部分為法人戶資料，並不是本研究所要探討的目標客群，因此將 B 部分的資料全數刪除，另外，C 部分之第 9 題、第 10 題也是針對法人戶的問題，因此整列刪除。刪除後剩餘資料量按照年份為：2017 年 41,847 筆；2018 年 118,587 筆；2019 年 75,579 筆；2020 年 53,235 筆；總計 289,248 筆。

由於原始資料為長資料格式，為了後續合併資料需求，將此資料表轉置為寬資料格式，以每列表示為一個客戶，欄位則表示各題目的答案。另外，A 部分第 10 題、第 12 題以及 C 部分的第 4 題分別詢問客戶的投資資金來源、投資目的以及目前所持有的金融商品，皆為複選題，因此會出現有多個答案的狀況。將這三題轉換成虛擬變數(Dummy Variable)，例如：以 0、1 的方式表達客戶是否持有債券型基金。

二、　基金淨值表

運用爬蟲程式將 CMoney 網站上的境內外基金淨值資料抓取下來，區間為 2005 年至 2020 年，共計 15 年。利用淨值計算出報酬標準差，公式如下：

$$\text{平均值：}\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N}\mu_i$$

$$\text{標準差：}s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mu_i - \bar{\mu})^2}$$

DOI:10.6814/NCCU202100509

該標準差即可於後續用來判斷客戶風險偏好的依據，如果客戶所購買的產品報酬標準差越高，則該客戶則偏向高風險；反之，則為低風險。計算出所有基金的報酬標準差後，即可利用國際證券識別碼(ISIN Code)將標準差資料合併於客戶交易資料中，ISIN Code 為國際辨別有價證券的唯一編碼，共有十二位，包含字母與數字。

三、　客戶交易資料

此份資料為 2016 年~2020 年期間所有的金融商品交易資料，共計五年。原始資料量約為 423 萬筆，欄位包括：交易日期、交易帳號、介紹人員編、商品代號、商品名稱等 24 個變數。

1.　資料清洗：

由於此份資料詳細記載了該公司於每筆交易中的損益以及歸屬單位，例如：該筆交易的手續費、毛利；銷售員的員工編號以及歸屬單位等等。這些變數不是本研究所要探討的，因此全數刪除，僅留下交易日期、客戶帳號、商品代號、交易金額等變數，示意表格如下：

表 4-1-2 客戶交易資料表

交易日期	客戶帳號	證券代號	商品名稱	交易金額
20160513	0000000056	A010001	統一強棒貨幣型基金	320,000
20160515	0000000056	A050052	元大加權股價指數基金	1,000,000
20160607	0000000103	A010009	統一龍馬基金	560,000
20170104	0000000023	A080026	野村全球生技醫療基金	2,300,000
20170323	0000000298	A080032	野村巴西基金	750,000
…				

資料來源：本研究整理

DOI:10.6814/NCCU202100509

刪除與研究無關的變數後，剩下的資料經由檢視並無遺漏值，總資料量約為 217 萬筆。其中，該公司銷售的商品中有些為債券附買回交易，由於此架構屬於衍生性金融商品，其淨值計算方式以及價格波動方式都與一般常見的境內外基金或者債券不同，因此予以刪除，此類型交易約有 37 筆。

2. 加入商品報酬標準差：

由於基金淨值表中的基金代號是用 ISIN Code 做為區別，因此需將此資料表與商品分類總表利用水平合併的方式，以證券代號將 ISIN Code 合併於交易資料表中。可利用 Excel 中的 VLOOKUP 函數完成此步驟。

3. 加權計算客戶的報酬標準差：

清洗過後的資料為長資料的格式，但由於客戶對於每檔商品的申購金額皆不同，因此為了方便計算加權後的報酬標準差，利用樞紐分析的功能將資料轉置為寬資料的格式，並標示出每檔商品的報酬標準差，以申購金額用加權平均法計算出每位客戶的數值，公式如下：

$$a = \frac{\sum_{t=1}^{n} M_t \times s_t}{\sum_{t=1}^{n} M_t}$$

a：客戶風險偏好

M：客戶申購金額

s：金融商品報酬標準差

四、 資料合併

將處理後的客戶交易資料以及客戶投資屬性問卷資料進行合併，由於有些客戶於此期間只有屬性問卷資料並沒有交易，因此使用左聯合併的方式，以客戶交易資料為主進行合併。合併後將客戶加權過後的報酬標準差由低到高分成三級，第一級為標準差較低的客群，其風險偏好較低，投資操作類別較為保守；第二級為風險偏好適中的客群，其投資操作較偏向為成長型資產為主；第三級為標準差較高的客群，其風險偏好較高，投資操作類別多為風險性資產，屬於積極型投資

DOI:10.6814/NCCU202100509

人。

合併後的資料表即為後續建立模型所使用的資料表，其中包含自變數與依變數，自變數包含了生物特徵變數、對於投資過往經驗及投資現況的問卷資料變數，共 36 個；而依變數即是客戶的風險偏好。

第二節 模型建立

1. 資料匯入及型態定義

將資料匯入統計軟體中，並且定義資料型態，自變數皆為輸入變數(Input)，依變數為輸出變數(Output)。將所有虛擬變數型態設為布林值(Boolean)；學歷、職位、行業別為集合(Set)；C 部分問卷資料設定為次序集合(Ordered Set)。

2. 平衡節點

此節點用來均衡資料的不均勻性，由於客戶風險屬性的分布不太平均，如圖 4-2-1 所示：

Value	Proportion	%
1		2.97
2		17.67
3		79.36

圖 4-2-1 原始風險屬性分佈(資料來源：本研究整理)

第一級客戶占了 2.97%；第二級為 17.67%；第三級占了大宗為 79.36%，因此為了提高預測效果，使用資料增生(Boost)的方式平衡資料的分布，過程如圖 4-2-2：

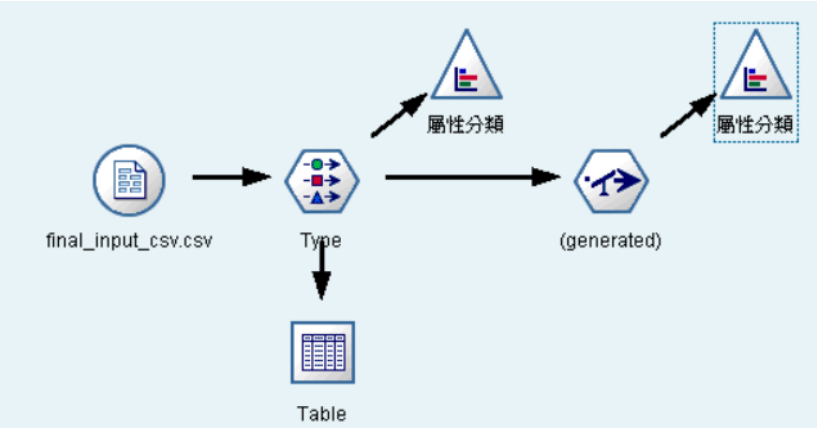

圖 4-2-2 平衡節點流程圖(資料來源：本研究整理)

DOI:10.6814/NCCU202100509

平衡後的資料分佈如圖 4-2-3 所示，第一級占比為 33.4%；第二級占比為 33.26%；第三級占比則為 33.34%，整體相較未調整前更為均勻。

Value	Proportion	%
1		33.4
2		33.26
3		33.34

圖 4-2-3 調整後風險屬性分佈(資料來源：本研究整理)

3. 資料劃分

使用軟體中 partition 節點將資料切分為訓練資料集(training)及測試資料集(testing)，訓練資料用於產生最適化的模型，而測試資料並不會用於模型的建構，主要用途為檢測建構完成的模型。本研究所使用的切割比率為 70%訓練資料、30%測試資料，如圖 4-2-4 所示，可以看到每列資料後面會有劃分註記。

	型_結構型	衍生性	風險占比C5	風險承受C6	投資占比C7	持有期間C8	風險等級	加權商品等級	是否相等	屬性分類	Partition
1	1	0	4	1	1	2	4	2	0	1	1_Training
2	1	0	4	1	1	2	4	2	0	1	1_Training
3	1	0	4	1	1	2	4	2	0	1	1_Training
4	1	0	4	1	1	2	4	2	0	1	2_Testing
5	1	0	4	1	1	2	4	2	0	1	1_Training
6	1	0	4	1	1	2	4	2	0	1	1_Training
7	1	0	4	1	1	2	4	2	0	1	1_Training
8	1	0	4	1	1	2	4	2	0	1	1_Training
9	1	0	4	1	1	2	4	2	0	1	1_Training
10	1	0	4	1	1	2	4	2	0	1	1_Training
11	1	0	4	1	1	2	4	2	0	1	1_Training
12	1	0	4	1	1	2	4	2	0	1	1_Training
13	1	0	4	1	1	2	4	2	0	1	2_Testing
14	1	0	4	1	1	2	4	2	0	1	1_Training
15	1	0	4	1	1	2	4	2	0	1	2_Testing
16	1	0	4	1	1	2	4	2	0	1	2_Testing
17	1	0	4	1	1	2	4	2	0	1	1_Training
18	1	0	4	1	1	2	4	2	0	1	1_Training
19	1	0	4	1	1	2	4	2	0	1	1_Training
20	1	0	4	1	1	2	4	2	0	1	1_Training
21	1	0	4	1	1	2	4	2	0	1	1_Training
22	1	0	4	1	1	2	4	2	0	1	2_Testing
23	1	0	4	1	1	2	4	2	0	1	1_Training
24	1	0	4	1	1	2	4	2	0	1	1_Training
25	1	0	4	1	1	2	4	2	0	1	1_Training
26	1	0	4	1	1	2	4	2	0	1	1_Training
27	1	0	4	1	1	2	4	2	0	1	2_Testing
28	1	0	3	5	5	4	5	4	0	3	2_Testing
29	1	0	3	3	3	3	4	3	0	2	1_Training
30	1	0	3	3	3	3	4	3	0	2	1_Training

圖 4-2-4 資料劃分結果圖(資料來源：本研究整理)

4. 建立模型

使用統計軟體中的模型節點建立模型，所使用的演算法有類神經網路、決策樹 CHAID、決策樹 C5.0 以及決策樹 CRT。整體流程如圖 4-2-5：

DOI:10.6814/NCCU202100509

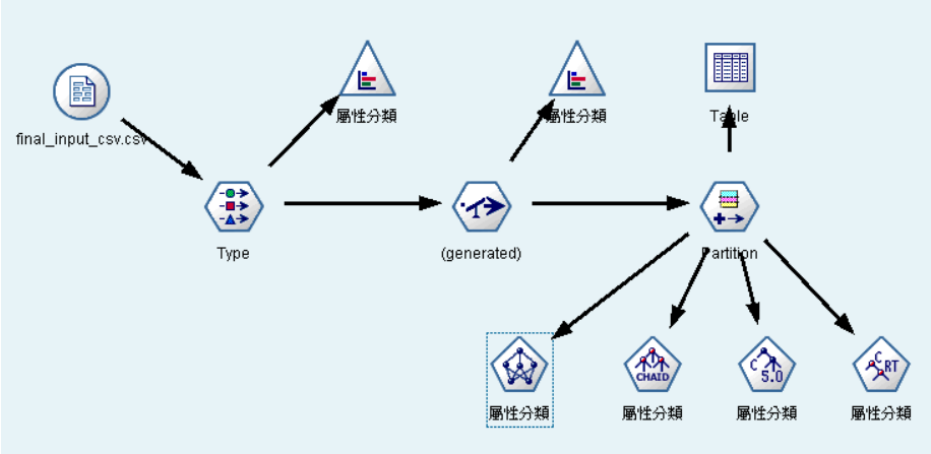

圖 4-2-5 建立模型流程圖(資料來源：本研究整理)

所有模型皆以屬性分類作為輸出變數，其餘變數為輸入變數，執行後即可點擊節點檢視結果。

第三節 研究結果

本節使用矩陣(Matrix)以及曲線下面積圖(AUC, Area under the Curve)來檢視、評估模型結果。將 Evaluation 節點接至模型後即可產生曲線下面積圖，如圖 4-3-1：

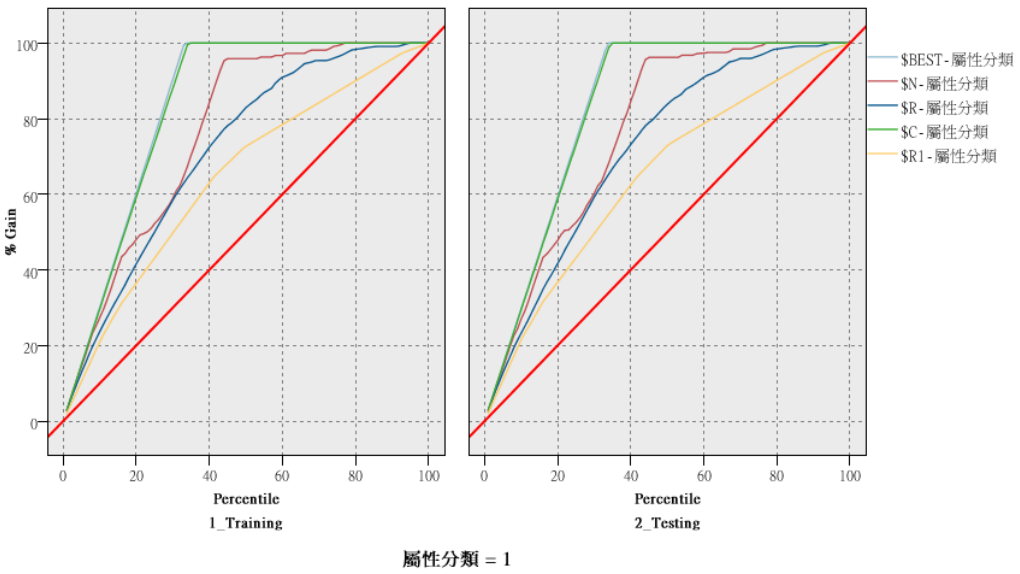

圖 4-3-1 曲線下面積圖(資料來源：本研究整理)

藉由圖形判斷，所有模型的 AUC 皆大於 0.5，表示所有模型的分類效果皆優

DOI:10.6814/NCCU202100509

於隨機猜測。另外，不管是訓練組還是測試組，每種模型的曲線幾乎都一樣，因此可以判斷模型不會隨著資料集的不同而使其結果會有所不同。預測力最好的模型皆為決策樹 C5.0，該模型的曲線幾乎貼近最佳線(Best line)，其餘接續為類神經網路、決策樹 CHAID 以及決策樹 CRT。

一、　類神經網路

利用矩陣檢視此模型的預測準確率，如圖 4-3-2，欄位為客戶偏好風險分類之實際值；橫列為模型的預測值。此模型整體預測準確率約為 66.12%，其中，實際客戶偏好屬性為第一級，模型也分類為第一級的資料占了總體的 31.56%；第二級則為 14.91%；第三級為 19.64%，經由換算，第一級的預測正確準確率約為 94.5%；第二級約為 44.8%；第三級則為 58.91%。因此，此模型對於預測第一級之客戶其準確率較高；在第二、三級的部分較低，約有 15.65%的第二級客戶被預測為第三級，另外，有 10.09%的第三級客戶被預測為第二級，而第一級客群被分類錯誤的比率僅有個位數百分比，藉由以上判斷此模型對於預測出第一級的客戶較有鑑別度，但比較無法鑑別出第二、三級之客戶。

		屬性分類		
$N-屬性分類		1	2	3
1	Count	5629	478	653
	Total %	31.563	2.680	3.662
2	Count	107	2660	1800
	Total %	0.600	14.915	10.093
3	Count	213	2791	3503
	Total %	1.194	15.650	19.642

圖 4-3-2 類神經網路模型之矩陣圖(資料來源：本研究整理)

圖 4-3-3 則為此模型在預測時所使用的重要變數及其重要性，依序為：行業別、預計可供規劃資產、資產總值、不動產、風險承受度等等。

DOI:10.6814/NCCU202100509

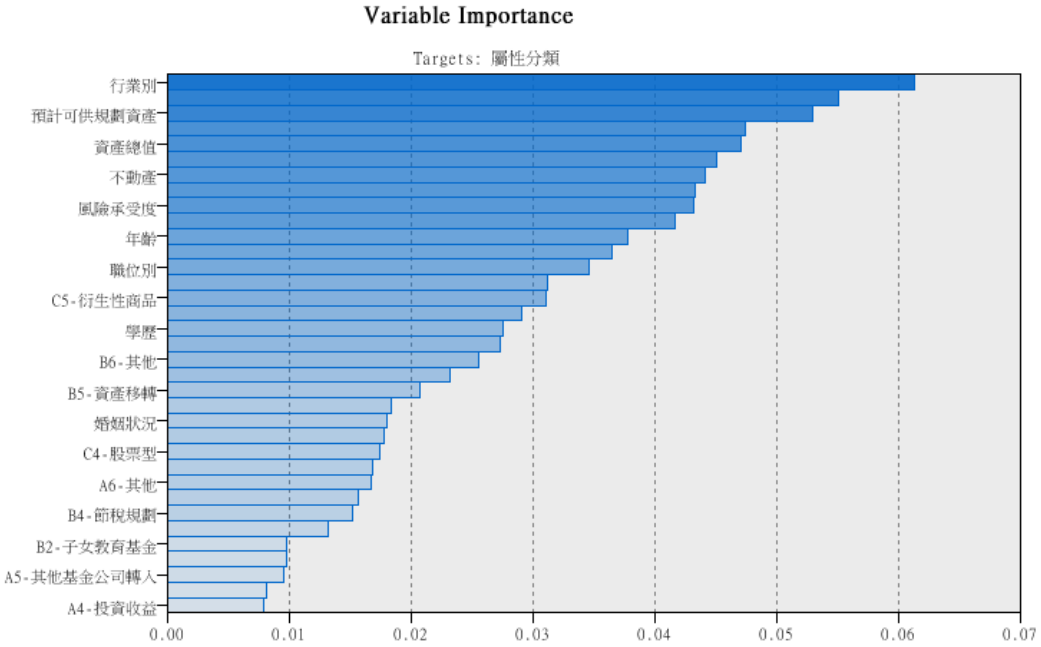

圖 4-3-3 類神經網路模型重要變數圖(資料來源：本研究整理)

二、 決策樹 CHAID

其矩陣結果如圖 4-3-4，整體的預測準確率約為 53.14%，其中，預測正確之資料筆數在各分級中的占比為：第一級 25.10%、第二級 8.12%、第三級 19.92%。經由計算，各偏好屬性的預測準確率分別為：第一級 75.32%、第二級 24.36%、第三級 59.71%，此模型區別第一級客戶時較有預測力，僅將個位數之百分比的客戶預測為第二或第三級；而第三級則有過半的預測力，預測錯誤的比率也只僅有個位數百分比，但在區分第二級客戶時其準確率明顯偏低，會將大部分的客戶預測為第一或第三級。

		屬性分類		
$R-屬性分類		1	2	3
1	Count	4487	1919	1509
	Total %	25.105	10.737	8.443
2	Count	611	1451	887
	Total %	3.419	8.118	4.963
3	Count	859	2590	3560
	Total %	4.806	14.491	19.918

圖 4-3-4 決策樹 CHAID 之矩陣圖(資料來源：本研究整理)

DOI:10.6814/NCCU202100509

此模型的重要變數及其重要性依序為：預計可供規劃資產、行業別、資金運用偏好、個人年平均收入、資產總值等等，如圖 4-3-5：

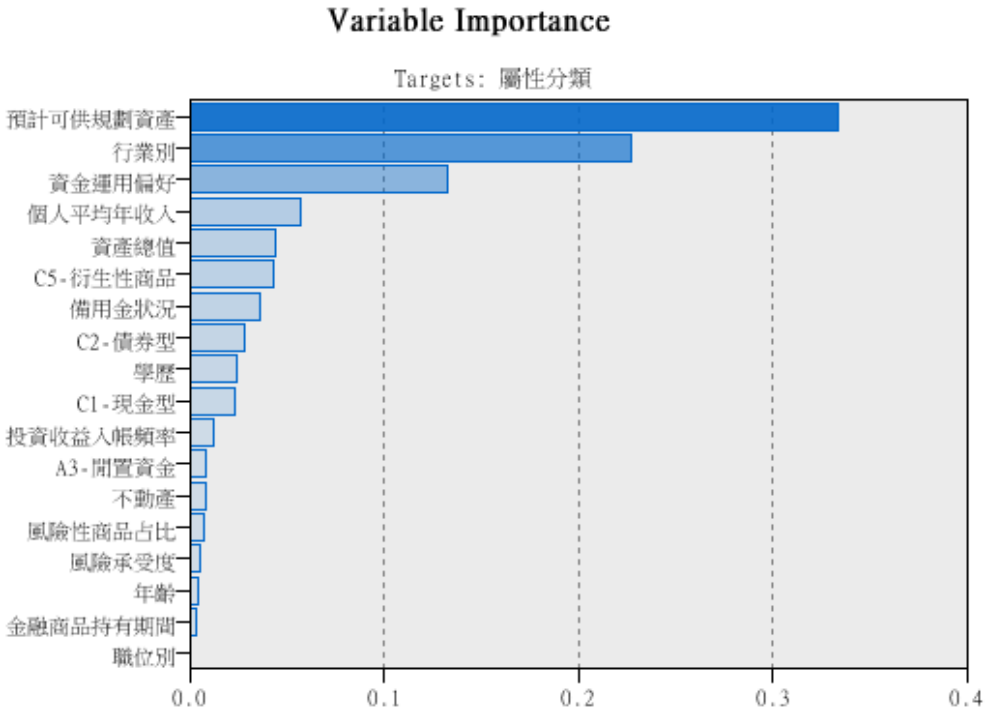

圖 4-3-5 決策樹 CHAID 模型重要變數圖(資料來源：本研究整理)

從整體樹狀圖(圖 4-3-6)來看，大致上可歸類出一些變數對預測結果的影響，例如：預計可供規劃資產越多的客戶，其風險偏好越小；金融業從業人員其風險偏好也較小。而有些變數的影響則會因上層樹狀結構的不同而有所不同，例如：金融從業人員在資金用偏好上不管一個月使用多少百分比的資金進行投資，其風險偏好皆較小；然而，製造業、零售業或住宿餐飲業的從業人員，其資金運用偏好中百分比越高的人其風險偏好也越高，反之則越低。

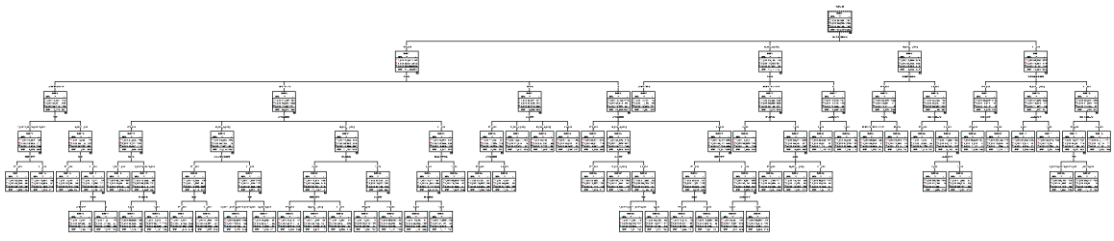

圖 4-3-6 決策樹 CHAID 之樹狀圖(資料來源：本研究整理)

DOI:10.6814/NCCU202100509

三、　　決策樹 C5.0

圖 4-3-7 為此模型的矩陣圖，欄位為實際值，橫列為預測值。整體預測準確度約為 92.75%，各分級中預測正確的資料占比為：第一級 33.35%；第二級 31.58%；第三級 27.82%，而準確率分別為：第一級 100%；第二級 94.78%；第三級 83.46%，此模型在各分級的預測度皆優於隨機預測，並且都達到八成以上的水準。

		屬性分類		
$C-屬性分類		1	2	3
1	Count	5960	27	122
	Total %	33.352	0.151	0.683
2	Count	0	5643	863
	Total %	0.000	31.578	4.829
3	Count	0	284	4971
	Total %	0.000	1.589	27.818

圖 4-3-7 決策樹 C5.0 之矩陣圖(資料來源：本研究整理)

此模型所使用到的重要變數相較其他模型較少，依其重要性分別為：行業別、備用金狀況、C3-外幣型、個人平均年收入、C4-股票型等等。而行業別為最重要的預測依據，其重要性遠高於其他變數，如圖 4-3-8：

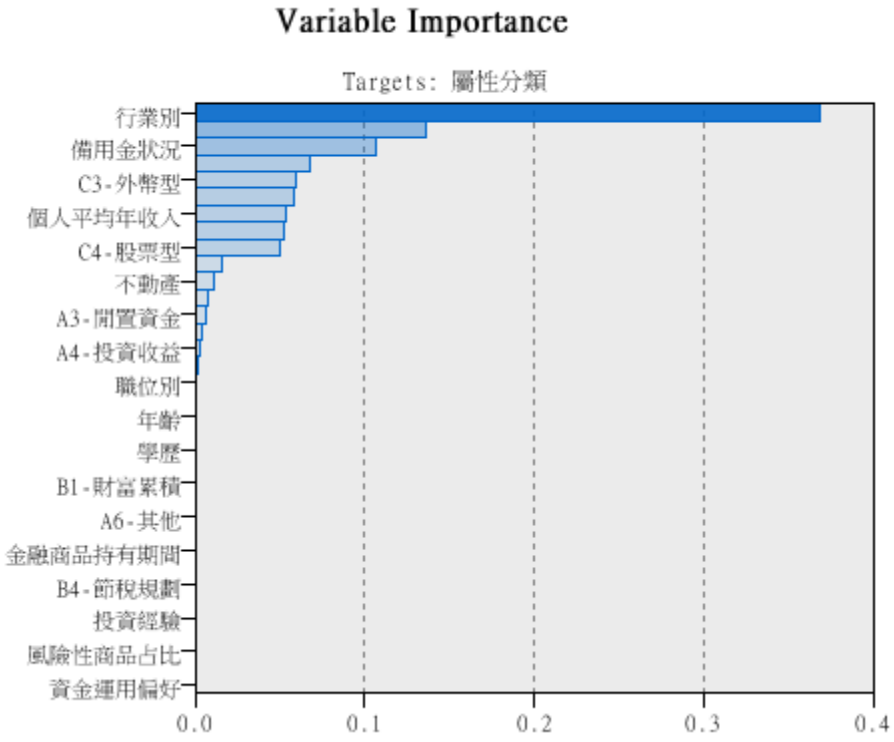

圖 4-3-8 決策樹 C5.0 模型重要變數圖(資料來源：本研究整理)

DOI:10.6814/NCCU202100509

此模型整體的樹狀結構共有 44 層，相較其他三種決策樹模型更為複雜，大致上可歸類出一些變數對預測結果的影響，例如：可供規劃資產越高者，其風險偏好越小；而備用金越高的客戶也越偏好積極的操作，風險偏好較高；金融業從業人員的風險偏好也較其他職業小。

四、　決策樹 CRT

此模型的矩陣如圖 4-3-9，欄位為客戶風險屬性偏好實際值，橫列則為預測值，整體預測準確率約為 47.99%，其中各分級中預測正確的資料占比分別為：第一級 24.21%、第二級 8.57%、第三級 15.20%；而經由計算後，各分級的預測準確率分別為：第一級 72.69%、第二級 25.66%、第三級 45.64%。這三級中唯獨只有第一級的預測準確率高於隨機預測，其餘準確率皆低於隨機預測，在第三級的分類結果中，約有 10.55%的資料模型預測為第一級，但由於第三級及第一級是相差甚遠的等級，第三級為風險偏好較高的積極型投資人，後續的投資建議組合多為股票類型；然而，第一級為風險偏好最低的保守型投資人，投資建議會偏向以政府公債等保守型金融商品為主，因此在此模型無法正確預測這兩類型的投資人情況下，該模型並不適合在實際商業流程上使用。

		屬性分類		
$R1-屬性分類		1	2	3
1	Count	4326	2534	1884
	Total %	24.214	14.183	10.545
2	Count	560	1531	1356
	Total %	3.134	8.569	7.590
3	Count	1071	1888	2716
	Total %	5.995	10.568	15.202

圖 4-3-9 決策樹 CRT 之矩陣圖(資料來源：本研究整理)

此模型所使用的重要變數以及其重要性如圖 4-3-10，依序為：預計可供規劃資產、行業別、金融商品持有期間、風險承受度、年齡等等。而主要使用的變數為預計可供規劃資產以及行業別，其重要性遠高於其他變數。

DOI:10.6814/NCCU202100509

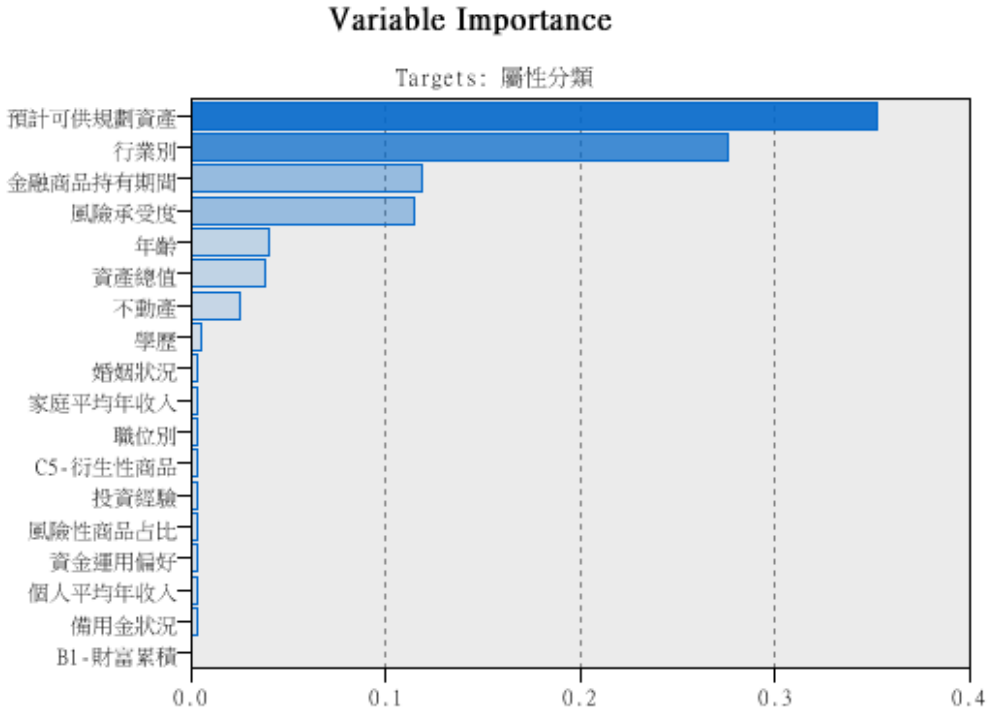

圖 4-3-10 決策樹 CRT 模型重要變數圖(資料來源：本研究整理)

此模型的整體樹狀結構如圖 4-3-11，相較於其他決策樹模型，此結構相對簡單，共有 5 層。經由比對，大致上可供規劃的資產越多其風險偏好越小；金融業以及特種行業從業人員其風險偏好較小，其他行業的人風險偏好較大。

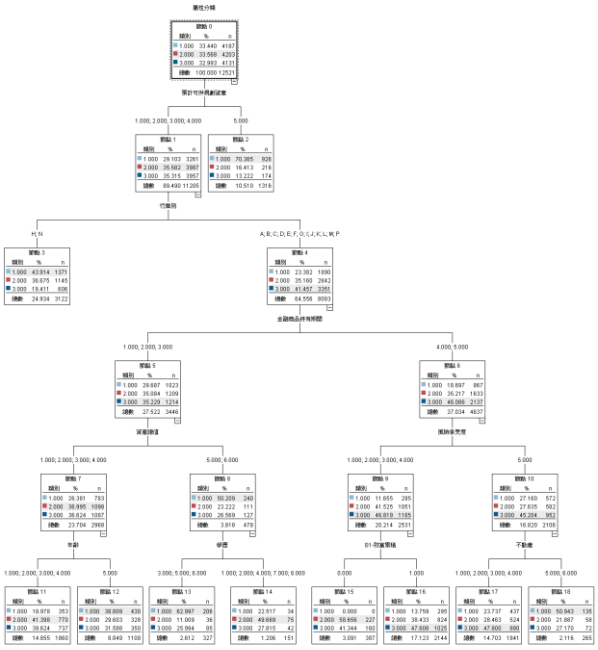

圖 4-3-11 決策樹 CRT 之樹狀圖(資料來源：本研究整理)

DOI:10.6814/NCCU202100509

第伍章　結論與建議

第一節 結論

一、　召回率(Recall)及精確率(Precision)

從上述的研究結果可以發現每種模型的風險偏好屬性之預測準確率以及所選擇的重要變數皆不盡相同，針對每一種客群(保守型、成長型、積極型)的辨識力也因模型而異，因此使用召回率(Recall)以及精確率(Precision)來評估各個模型中不同風險偏好族群的預測情況。而此兩項指標都是使用混淆矩陣(Confusion Matrix)中的數值計算，此矩陣分別有四個數值，分別為：

TP(True Positive)：實際值為是，預測值也為是。

FP(False Positive)：實際值為否，預測值為是。

TN(True Negative)：實際值為否，預測值也為否。

FN(False Negative)：實際值為是，預測值為否。

如表 5-1-1 所示：

表 5-1-1 混淆矩陣

	實際 Yes	實際 No
預測 Yes	TP	FP
預測 No	FN	TN

資料來源：本研究整理

而召回率公式為：$\frac{TP}{TP+FN}$，意即模型於實際值為是的資料中能預測出多少比例為是；而精確率公式為：$\frac{TP}{TP+FP}$，意即模型預測出為是的資料中有多少比例實際值也為是。這兩種指標的高低並不代表該模型的好壞，模型的選擇取決於實際應用的商業場景需求為何。

1. 風險偏好屬性第一級客戶(保守型)：

以實際值為第一級的資料為正向，第二、三級為負向，產出各模型的混淆

DOI:10.6814/NCCU202100509

矩陣，結果如下表 5-1-2：

表 5-1-2 混淆矩陣(屬性偏好第一級)

類神經網路

	實際正向	實際負向
預測正向	5,630	1,140
預測負向	316	10,773

決策樹 CHAID

	實際正向	實際負向
預測正向	4,485	3,446
預測負向	1,471	8,456

決策樹 C5.0

	實際正向	實際負向
預測正向	5,968	150
預測負向	0	11,750

決策樹 CRT

	實際正向	實際負向
預測正向	4,334	4,421
預測負向	1,634	7,471

資料來源：本研究整理

藉由以上混淆矩陣即可利用公式計算出各模型的召回率及精確率，計算結果如表 5-1-3：

表 5-1-3 風險屬性第一級召回率及精確率

	類神經網路	決策樹 CHAID	決策樹 C5.0	決策樹 CRT
Recall	94.69%	75.30%	100%	72.62%
Precision	83.16%	56.55%	97.55%	49.49%

資料來源：本研究整理

綜合比較後，不管是召回率還是精確率由高至低皆為：決策樹 C5.0＞類神經網路＞決策樹 CHAID＞決策樹 CRT，因此決策樹 C5.0 在預測風險偏好第一級的客戶時有較精準的預測力，未來假使應用之商業場景需要找尋操作金融商品較保守的客戶時，可使用此模型來做預測。而決策樹 CRT 在精確率不到 50%的情況下並不適合使用，因為此模型會所預測出的第一級客戶有超過一半實質上在金融商品操作上會比較偏向積極、高風險的商品。

DOI:10.6814/NCCU202100509

2. 風險偏好屬性第二級客戶(成長型)：

以實際值為第二級的資料為正向，第一級、第三級資料為負向，利用模型結果統計出混淆矩陣，如表 5-1-4 所示：

表 5-1-4 混淆矩陣(屬性偏好第二級)

類神經網路

	實際正向	實際負向
預測正向	2,660	1,907
預測負向	3,269	9,998

決策樹 CHAID

	實際正向	實際負向
預測正向	1,451	1,498
預測負向	4,509	10,415

決策樹 C5.0

	實際正向	實際負向
預測正向	5,643	863
預測負向	311	11,053

決策樹 CRT

	實際正向	實際負向
預測正向	1,531	1,916
預測負向	4,422	9,997

資料來源：本研究整理

利用上面數字計算出各模型的召回率以及精確率，如下表所示：

表 5-1-5 風險屬性第二級召回率及精確率

	類神經網路	決策樹 CHAID	決策樹 C5.0	決策樹 CRT
Recall	44.86%	24.35%	94.78%	32.19%
Precision	58.24%	49.20%	86.74%	44.42%

資料來源：本研究整理

模型之召回率由高至低依序為：決策樹 C5.0＞類神經網路＞決策樹 CRT＞決策樹 CHAID；而精確率由高至低則為：決策樹 C5.0＞類神經網路＞決策樹 CHAID＞決策樹 CRT。綜合評估後，決策樹 C5.0 為該級最佳的預測模型；決策樹中 CHAID 以及 CRT 的方法所產出之模型，其召回率及精確率皆不及五成，低於隨機預測，因此於實際商業場景中不適合使用。

DOI:10.6814/NCCU202100509

3. 風險偏好屬性第三級客戶(積極型)：

以實際值為 3(第三級)的資料為正向，1(第一級)、2(第二級)資料為負向，計算出個模型的混淆矩陣，如表 5-1-6：

表 5-1-6 混淆矩陣(屬性偏好第三級)

類神經網路

	實際正向	實際負向
預測正向	3,503	3,004
預測負向	2,453	8,874

決策樹 CHAID

	實際正向	實際負向
預測正向	3,560	3,449
預測負向	2,396	8,468

決策樹 C5.0

	實際正向	實際負向
預測正向	4,971	284
預測負向	985	11,630

決策樹 CRT

	實際正向	實際負向
預測正向	2,716	2,959
預測負向	3,240	8,951

資料來源：本研究整理

經由計算後，各模型的召回率及精確率整理如下表：

表 5-1-7 風險屬性第三級召回率及精確率

	類神經網路	決策樹 CHAID	決策樹 C5.0	決策樹 CRT
Recall	58.81%	59.77%	83.46%	45.60%
Precision	53.83%	50.79%	94.60%	45.46%

資料來源：本研究整理

綜合比較後，召回率由高至低的模型為：決策樹 C5.0＞決策樹 CHAID＞類神經網路＞決策樹 CRT；而精確率由高至低為：決策樹 C5.0＞類神經網路＞決策樹 CHAID＞決策樹 CRT。整體來說，決策樹 C5.0 在這兩種指標皆優於其他模型，為最佳模型；另外，決策樹 CRT 模型在這兩種指標皆低於五成，低於隨機預測，因此較不適合作為商業用途。

DOI:10.6814/NCCU202100509

召回率意即在實際正向資料中，模型可以預測出多少比例的正向資料，因此企業可以將召回率高的模型應用至智能理財平台的商業流程中，在第一階段客戶建檔時留存各項變數的資料，將這些變數資料輸入模型，產出的結果即可為後續投資組合建議使用，於傳統財富管理業務中，理財專員可利用預測出來的結果，推介相對應的投資組合，假使結果為第一級，即可推介信用評等較高之債券基金組合；假使結果為第三級，則可推介以股票型基金為主的投資組合。另外，於新型態智能理財平台上，該模型所預測出的風險屬性偏好是以報酬標準差區分，因此可以使用此標準差於後續的商業流程中，在投資組合選擇的流程中，作為模型的輸入變數，以此輸入變數所產生之投資組合結果會較符合投資者自身的風險偏好，達到精準行銷的目的。

另外，精確率為模型所預測的正向資料中，有多少比例於實際中也為正向。因此企業可使用精確率高的模型於行銷策略上，例如：投資信託公司發行一檔以政府公債為主的債券型基金，欲推廣給保守的客戶，即可利用於風險屬性偏好第一級精準率較高的模型，由模型預測出的第一級客戶作為推廣的對象。

因此模型指標的選擇會依實際的商業目的而有所不同，企業可依其需求選擇指標較高的模型作使用。然而，綜合比較上述各個風險屬性偏好分級之召回率及精確率後可以發現，決策樹 C5.0 的模型在這兩種指標的表現皆優於其他模型。因此在此個案中，P 公司在預測客戶的風險屬性偏好上，最佳的模型皆為決策樹 C5.0。

二、　預測模型的重要變數

於前述研究結果中發現眾多輸入變數中，每種模型所使用到的重要變數都不盡相同，為了整合並了解重要變數對於預測結果的影響趨勢，因此於表 5-1-8 整理出每種預測模型所使用的前十大重要變數：

DOI:10.6814/NCCU202100509

表 5-1-8 重要變數比較表

類神經網路	決策樹 CHAID	決策樹 C5.0	決策樹 CRT
行業別	預計可供規劃資產	行業別	預計可供規劃資產
投資收益入帳頻率	行業別	風險承受度	行業別
預計可供規劃資產	資金運用偏好	備用金狀況	金融商品持有期間
投資經驗	個人平均年收入	B2-子女教育基金	風險承受度
資產總值	資產總值	C3-外幣型	年齡
風險性商品占比	C5-衍生性商品	婚姻狀況	資產總值
不動產	備用金狀況	個人平均年收入	不動產
資金運用偏好	C2-債券型	B3-退休規劃	學歷
風險承受度	學歷	C4-股票型	婚姻狀況
家庭平均年收入	C1-現金型	A2-退休金	職位別

資料來源：本研究整理

綜合比較後發現，行業別、預計可供規劃資產、資產總值以及風險承受度這四種變數為模型幾乎都有使用到的重要變數，經由交叉比對得出以下結論：

1. 預計可供規劃資產越多，客戶之風險屬性偏好越小，意即模型預測為第一級(保守型)之客戶。客戶欲投資的金額越高，其在金融商品操作上較偏保守，例如：投資於政府公債基金或者貨幣市場基金這些報酬標準差較小的產品。而預計可供規劃資產較少的客戶，其風險偏好就越大，金融商品操作上較積極。
2. 資產總值越小，客戶之風險屬性偏好越大，模型預測為第三級(積極型)之客戶；反之資產總值越大，風險屬性偏好就越小。此變數與上點所述的趨勢相似，差別在於一個是客戶本身自有資產總值，另一個為客戶欲投資的金額。推測原因為金額越高，面對波動度的承受度就越小，只要波動度有較大的起伏，其影響的實際金額也就越大，然而金額較小的的話，就算波動度較

DOI:10.6814/NCCU202100509

大，其影響的實際金額也會較小。

3. 風險承受度越高，其風險屬性偏好也越大，反之則越小。意即客戶過去可以接受的商品波動度越高，其在未來也越有可能操作積極型的金融商品，但如果客戶過去只能接受較低的價格波動，而他後續也會偏保守操作，高波動的金融投資商品占比不高。

4. 在行業別的影響中，金融業從業人員的風險屬性偏好較低；珠寶業、銀樓、典當業、酒家、舞廳以及電玩等特種行業的從業人員其風險屬性偏好較高。推測原因為跟行業屬性有關，由於金融業屬於風險管控嚴格的行業，從業人員多半都有風險管理的概念，因此反映在自身的投資策略及理念上，金融商品的操作上就顯得較為保守，另外，從事金融業的人對於金融產品較為熟悉，了解其架構以及背後運作的原理，相較於其他產業人員，更容易辨識商品之優缺點，對於可能的損失(風險)也就格外注重。然而，珠寶、典當等特種行業由於其本身即為高風險的產業，因此反映到從業人員自身的投資策略上就較其他行業的人積極，偏好高風險高報酬的金融商品。

上面所述的重要變數對於預測結果的趨勢影響，可以運用於 P 公司縮短商業流程上的依據，傳統財富管理業務在進行客戶 KYC(Know Your Customer)時都會填具一份客戶風險屬性評量表，而這份評量表以 P 公司為例，多達 23 題，填寫完後再加總各題所得到的分數，以總分區分出客戶的風險屬性，理財專員在依此屬性推薦相對應的金融商品，因此在整個流程中，客戶會消耗大量時間在填具文件上，進而造成客戶認為跟金融機構進行投資理財的流程繁瑣的問題。

P 公司可利用這些重要變數改善傳統的理財業務流程，將紙本問卷重要性較低的變數刪除，留下重要性較高的。以召回率及精確率最高的決策樹 C5.0 模型為例，該模型所使用的重要性較大的變數約有八個，因此可使用此八個變數重新設計投資屬性問卷，即可減少約 70%的流程時間，而評估方式則以預測模型的方式而非傳統加總計分的方式，這樣即可加速整個評估流程。

DOI:10.6814/NCCU202100509

第二節 建議

本研究是以P公司為例，實驗數據及商業流程也以P公司實際情況為主，因此實驗模型及結果僅適用於P公司，其他類似產業之公司以相同方法及模型所產出之結果應會與本研究不盡相同，但研究方法及概念是可以加以利用的。以下提出幾點為未來可以加以改善的地方：

1. 數據面：本研究所使用的交易資料日期區間大約五年，基於大數據的概念，如果可以取得更多的交易資料，對於模型的建構會更加完善，其結果也會更加精確。
2. 屬性偏好區間：本研究將客戶的風險偏好以常見的方式區分為三級，保守型、成長型、積極型。然而可嘗試將客戶分為更多的層級，但前提為資料量要足夠，每種類別的樣本數要達到一定的數量，這樣對於後續的統計分析以及建構模型才有顯著性。
3. 時間軸問題：本研究為橫斷面研究，為研究某一時間中的客戶資料，並不包含同一客戶的生命週期以及產品的生命週期，每一客戶在人生中的不同階段對於投資理財皆有可能抱持著不同的風險偏好態度；另外，金融產品在不同的階段也會有不同的報酬標準差表現，例如：面對同一檔的股票型基金，在金融海嘯購買的客戶以及景氣擴張時期購買的客戶，這兩種客戶所表現出來的風險偏好應是不同的，利用長時間的報酬標準差會抹平該檔標的在不同時期表現出的價格波動。

科技的發展使的許多過往的商業流程以及模式發生改變，過去僅能以相關業務人員的經驗判斷一位新客戶適合哪種金融商品，但現今由於數據的取得更加方便以及技術上的進步，可以發展預測模型來達成此一目的，而此模型不僅能使用於線上的智能理財平台外，也可當作實體通路的工具，讓業務人員先依據此模型的結果來判斷適合客戶的商品，進而進行銷售推廣。然而，業務發展的重要關鍵不僅是模型的精準度而已，”人”的角色在業務連結上也扮演著重要的一環，在

DOI:10.6814/NCCU202100509

傳統的理財業務流程中，這些工具都是輔助的角色，而關鍵在於使用這些工具的人，根據「行為財務學在機器人理財顧問的應用、實務與挑戰」(2018, 曹書華)一文所述，於金融機構進行投資理財的關鍵取決於信任，而信任是由人與人之間的關係所產生，這是機器很難取代的。因此不管是傳統的線下模式還是新興的線上模式，還需從業人員善加利用工具，發掘客戶問題，進而提出解決問題的方法。

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100509

參考文獻

一、中文文獻

1. 吳岱芸，2017，運用 RFM 模型結合資料採礦預測潛在顧客提升行銷效益－以 Y 藥局為例，政治大學企業管理研究所碩士論文。
2. 李宏毅，2017，什麼是深度學習－機器學習的捲土重來，數理人文雜誌。
3. 李怡，2017，資料探勘應用於航空公司之顧客分群研究，政治大學企業管理研究所碩士論文。
4. 卓越，2019，基於支持向量回歸的選股模型實證研究－以台股市場為例，政治大學金融學系研究所碩士論文。
5. 康力中，2020，機器人理財是否可有效消除投資人行為偏誤，臺北大學企業管理學系碩士論文。
6. 張力元，2018，深度學習應用於股價走勢之研究：以大陸市場為例，政治大學風險管理與保險學系研究所。
7. 張明珠，2019，淺談機器人理財在台灣未來之發展，證券暨期貨月刊 第三十七卷 第一期，頁 14-25。
8. 張雅鈞，2016，人壽客群與商品搭售分析－以 C 人壽資料為例，政治大學企業管理研究所碩士論文。
9. 曹書華，2018，行為財務學在機器人理財顧問的運用、實務與挑戰，政治大學經營管理碩士學程碩士論文。
10. 莊子郁，2018，理財機器人於台灣理財市場之適用探討，政治大學風險管理與保險學系碩士論文。
11. 陳力新，2017，智能理財產業分析，政治大學經營管理碩士學程碩士論文。
12. 陳雅婷，2019，金融機構導入理財機器人關鍵成功因素之研究，臺北科技大學資訊與財金管理 EMBA 專班碩士論文。
13. 陳鴻，2021，臺灣普惠金融發展之研究：以韓國為鑑，政治大學企業管理研

DOI:10.6814/NCCU202100509

究所碩士論文。

14. 彭士嘉，2013，應用資料採礦技術於消費者網路使用行為之研究，政治大學企業管理研究所碩士論文。
15. 新智元，2019，2019 中國人工智能獨角獸白皮書。
16. 謝邦昌、鄭宇庭、蘇志雄，2009，Data Mining 概述—以 Clementine 12.0 為例，中華資料採礦協會。
17. 韓傳祥，2018，計量財務金融〈金融科技〉，新陸書局。

二、英文文獻

1. Alex J.Smola, Bemgard Scholkopf, 2004, A tutorial on Support Vector Regression, Statistics and Computing, 199-222.
2. Alexander Kunst, 2020, RoboAdvisor usage in the U.S. 2020, Statista.
3. Berry, Michael J.A., Linoff, Gordon S., 1997, Data Mining Techniques: For Marketing, Sales, and Customer Support, U.S.A., John Wiley & Sons, Inc.
4. Betterment, 2021, Retrieved from https://www.betterment.com/
5. EY, 2015, Advice Goes Virtual: How New Digital Investment Services Are Changing the Wealth Management Landscape.
6. Krzystof, C., P. Witold & S. Roman, 1998, Data Mining: Methods for Knowledge Discovery. Kluwer Academic Publishers, Boston.
7. Michael Nielsen, 2015, Neural Networks and Deep Learning,
8. Statista Research Department, 2021, Assets under management of selected robo-advisors globally 2021, Statista.
9. Statista, 2021, Retrieved from https://www.statista.com/
10. Tom Mitchell, 1997, Machine Learning
11. WEALTHFRONT, 2021, Retrieved from https://www.wealthfront.com/

DOI:10.6814/NCCU202100509

三、網路文獻

1. David Huang，2016，資料專案的完整流程 2：Information process，出自：https://taweihuang.hpd.io/2016/09/19/%E8%B3%87%E6%96%99%E5%B0%88%E6%A1%88%E7%9A%84%E5%AE%8C%E6%95%B4%E6%B5%81%E7%A8%8B-2%EF%BC%9Ainformation-process/
2. KPMG，2019，2019 年 FinTech 100 金融科技創新者報告，出自：https://home.kpmg/tw/zh/home/insights/2019/11/ventures-kpmg-fintech100-2019.html
3. OOSGA，2021，機器學習－定義、演算法以及商業應用，出自：https://oosga.com/pillars/machine-learning/
4. SIGAI，2018，機器學習發展歷史回顧，出自：https://kknews.cc/zh-tw/tech/n389r88.html
5. 吳志峯，2018，會導致投資虧損的三種心理效應，別讓他左右我們的腦袋，出自：https://medium.com/@iam54f/%E6%9C%83%E5%B0%8E%E8%87%B4%E6%8A%95%E8%B3%87%E8%99%A7%E6%90%8D%E7%9A%84%E4%B8%89%E7%A8%AE%E5%BF%83%E7%90%86%E6%95%88%E6%87%89-%E5%88%A5%E8%AE%93%E5%AE%83%E5%B7%A6%E5%8F%B3%E6%88%91%E5%80%91%E7%9A%84%E8%85%A6%E8%A2%8B-ecdbcaf4d545
6. 沈庭安，2017，為何能吸引民眾託付百億資產？美國理財機器人雙巨投策略大剖析，出自：https://www.ithome.com.tw/news/116555
7. 林哲緯，2019，別以為年輕人都不理財！獨家青世代理財調查：於 6 成有在規劃，最愛用的工具是「」，Cheers 雜誌，出自：https://www.cheers.com.tw/article/article.action?id=5095747&page=2

DOI:10.6814/NCCU202100509