張昭憲、陳世軒 (2021),「以模型融合為基礎之線上拍賣詐騙偵測」,*資訊管理學報*,第二十八卷,第四期,頁 419-444。

以模型融合為基礎之線上拍賣詐騙偵測

張昭憲*
淡江大學資訊管理學系

陳世軒
淡江大學資訊管理學系

摘要

隨著金流與物流等基礎建設的成熟,電子商務的蓬勃發展有目共睹,而線上拍賣更是其中重要的一環。面對日益龐大交易金額,也引起不肖人士的覬覦,在拍賣平台中進行詐騙。有關線上拍賣詐騙偵測,已有許多方法已被提出,但對於日新月異的詐騙手法,其準確率仍有待提升。有鑑於此,本研究將運用模型融合(Model Fusion)概念,發展更有效的詐騙偵測方法。首先,我們分析單一模型應用在不同測試集之效能,發現當詐騙者與正常者比例未知時,單一模型的效能將受到限制。其次,本研究利用不同類型配比之訓練資料,探討如何產生有利於詐騙者與正常者之偵測模型。最後,運用多種不同特質之模型,分別以多階連續過濾及平衡過濾方式加以整合,以提升總體偵測效能。為驗證提出方法之有效性,我們採用 Yahoo!拍賣實際交易資料進行實驗。與各種單一偵測模型相較,本研究提出之連續過濾與平衡過濾法確能提升準確率,並提供更穩定的偵測結果。當使用連續過濾時,除獲得較高準確率外,也能對各階段之偵測精度進行分析,提升結果之實用性。此外,雖然模型融合時嘗試使用各種不同特質的單一模型可影響準確性,但我們發現在多階段過濾的流程下,對偵測效能之影響有限。由上述結果可知,本研究提出方法確有助於改善詐騙偵測之準確率,提供消費者更周全的交易防護。

關鍵詞:詐騙偵測、模型融合、分類、線上拍賣、電子商務

* 本文通訊作者。電子郵件信箱:jschang@mail.tku.edu.tw
2021/03/01 投稿;2021/06/29 修訂;2021/08/01 接受

Chang, J.S. & Chen, S.H. (2021). Online Auction Fraud Detection based on Model Fusion. *Journal of Information Management*, 28(4), 419-444.

Online Auction Fraud Detection based on Model Fusion

Jau-Shien Chang*
Department of Information Management, Tamkang University

Shih-Hsuan Chen
Department of Information Management, Tamkang University

Abstract

With the maturity of infrastructure such as cash flow and logistics, the booming development of e-commerce is obvious to all. However, facing such a large transaction amount, it also attracts many fraudsters to join e-commerce. Among the reported cases, online auction fraud undoubtedly forms a large proportion. Although a lot of detection methods have been proposed, the detection accuracy for the ever-changing fraud scheme still needs to be improved. To solve this problem, this study adopts the model fusion concept to develop more effective fraud detection methods. First, we analyzed the effectiveness of a single model in different test sets, and found that when the ratio of fraudsters to non-fraudsters is unknown, it is difficult for a single model to be effective. Secondly, this study uses different types of training data to explore how to generate a detection model that is beneficial to fraudsters and normal traders. Finally, a variety of models with different characteristics are used to integrate multi-stage successive filtering and balanced filtering to improve the overall performance. To verify the effectiveness of the proposed method, we use Yahoo! auction transaction data to conduct experiments. Compared with single detection models, the successive filtering and balanced filtering can improve the detection accuracy and provide more stable results. When using successive filtering, the precision of each stage can also be analyzed to enhance the practicability of the results. In addition, we found that changing the characteristics of each single model has a limited impact on the performance of the multi-stage filtering process. In summary, the proposed method can actually help improve the accuracy of fraud detection and provide a safer trading environment.

Keywords: Fraud detection, model fusion, classification, online auctions, e-commerce

* Corresponding author. Email: jschang@mail.tku.edu.tw

2021/03/01 received; 2021/06/29 revised; 2021/08/01 accepted

壹、緒論

電子商務的蓬勃發展有目共睹，配合金流、物流等基礎建設的成熟，大幅提升交易的便利性與時效性。由於不受時間與空間的限制，無論 B-to-B, B-to-C,或 C-to-C 等交易模式，均呈現逐年穩定成長。依據 eMarketer 市場研究機構調查顯示，2016 年全球網路零售交易為 1 兆 9,200 億美元，但在 2020 年預計將可超過 4 兆美元，成長幅度驚人 (eMarketer 2020)。而針對台灣的調查，在 2018 年至 2022 年未來五年間，電子商務每年預計也可維持近 7%的成長率。由上述資料可知，電子商務已成為未來的交易主流，並可能逐步取代實體通路。

面對如此龐大的交易金額，也引起不肖人士的覬覦，在電子商務平台中進行詐騙。以線上拍賣為例，常見的詐騙行為有收款不出貨、進行假交易、商品敘述不實、販賣偽劣貨等 (NW3C 2019; Tsang et al. 2014; Gavish & Tucci 2008)。但詐騙者亦可扮演買方，以進行付款詐欺、異常退款，甚至洗錢等不法行為(Chen et al. 2015)。為避免消費者懷疑，更有詐騙者以略低於市價販售商品，但巧妙隱藏商品規格細節(例如相機是否附原廠鏡頭)，售出後再以各種藉口規避退貨(Kim, Choi, & Park 2013)。由於網路的隱蔽性與便利性，讓這些不法行為在短期內便快速成長，若不加以抑制，將嚴重影響線上拍賣未來發展。為降低詐騙的發生，拍賣網站當局經常以二元名聲系統(binary reputation system)來協助評估使用者的信用。當買賣雙方交易完成時，可互相給予正評(+1)、普評(+0)、負評(-1)等三種評價，評價的給予可能受到貨時間、價格、售後服務等因素影響。此種評分機制雖然簡單，但累積的分數(名聲)卻足以影響他人是否願意與你進行交易的意願。例如，有經驗的消費者往往會在購買商品前，花費大量的時間查閱賣家的信用評價，以避免不必要的交易糾紛(Goes, Tu, & Tung 2009)。然而，因二元名聲系統過於簡單，許多不肖份子經常建立多個帳號進行假交易，以快速累積正評，待取得消費者信任後再進行詐騙。面對類似取巧行為，拍賣網站管理當局只能藉由停權來達到嚇阻效果，受害者仍須自行報案，才可能追回損失。期間可能花費大量的時間金錢，更讓許多求償的受害者心懷畏懼、裹足不前。凡此種種，均嚴重影響線上拍賣的長遠發展。

面對線上拍賣詐騙偵測，學界與業界莫不給予高度關注，並以積極方式來因應(West & Bhattacharya 2016)。Alford (2013)更認為現代企業都應發展智慧型詐騙偵測系統，檢視每日進行的所有交易，以維護電子商務之交易安全。為避免消費者受損失，學者也紛紛提出各種詐騙偵測方法。例如，Chau, Pandit, & Faloutsos (2006)利用價格異常做為偵測基礎，以分類樹方式建立偵測模型，藉以分辨詐騙者。Chang & Chang (2011, 2012)則提出詐騙預警概念，以階段切割法(phased-profiling)切割交易者生命週期，產生具有潛伏期詐騙者偵測能力之模型。Pandit et al. (2007)則提出二階段偵測概念，利用分類樹與 Markov Random Field 標示詐騙正犯與共犯，希望能找出詐騙者及其所屬共犯集團。Tsang et al. (2014)持續改進 Pandit 等人之方法，運用 Markov 方法計算帳號詐騙機率，以更精確方式標示

出詐騙共犯集團。為提升分類準確度，學者們亦嘗試透過集成學習(Ensemble Learning)來進行詐騙偵測。例如，Kumar et al. (2019)便透過 Random Forest 對信用卡詐欺進行偵測，而 Xuan et al. (2018)更改良使用 CART-based Random Forest，期能進一步提升偵測效能。除了學界外，業界對於詐騙偵測的投入也不遺餘力。例如，為有效遏止詐騙，阿里巴巴集團即發展了一套即時的詐騙防範與監控系統，監控的行為涵蓋不正常退款、多重帳號、盜取帳號，甚至洗錢等複雜的犯罪行為(Chen et al. 2015)。

線上拍賣詐騙偵測雖已獲得學界與業界高度關注，亦有許多方法被提出，但仍面臨諸多挑戰(Ahmed, Mahmood, & Islam 2016; West & Bhattacharya 2016)。其中，傳統分類技術(classification)雖然有效，但均具有效能瓶頸，需要進一步突破。為降低上述問題的影響，模型融合(Model Fusion)便成為可行的解決之道。各種詐騙偵測模型均有其特質，偵測效能也有所差異。模型融合(Model Fusion)是一種從多種模型中擷取有利於預測效能的流程。模型融合概念已應用在各種不同領域，例如，Li et al. (2012)同時使用貝式分類法與關聯規則探勘，歸納觀察詐騙者行為樣式，並配合專家判讀提升準確率。Chen et al. (2013)則利用貝氏方法進行模型融合，提升網路媒體語意擷取的正確性。Huang et al. (2015)則透過加權平均法整合各模型之預測結果，並藉由機器學習調整模型權重，以產生更精確商品推薦名單。面對電子商務詐騙，Chen et al. (2015)則利用 logistic regression 組合各種模型產生的可疑分數，以提升詐騙偵測效能，並發現決策樹與 Random Forest 能獲得較佳的偵測效能。為了解不平衡資料集所造成的影響，Makki et al. (2019)更對 8 種不同學習方法進行大規模實驗，找出適合於信用卡詐騙偵測之有效方法。根據上述文獻顯示，若能有效整合各種模型的特點，對於偵測的即時性與準確性將有很大助益。

根據上述討論，本研究將運用模型融合概念，發展有效的詐騙偵測方法，以提供消費者更安全的購物環境。首先，我們分析單一模型應用在不同測試集之效能，發現當詐騙者與正常者比例未知時，單一模型效能將受到限制。其次，本研究利用不同類型配比之訓練資料，探討如何產生有利於詐騙者與正常者之偵測模型。最後，運用多種不同特質之模型，分別以多階連續過濾及平衡過濾方式加以整合，以提升總體偵測效能。為驗證提出方法之有效性，我們採用 Yahoo!拍賣實際交易資料進行實驗。與各種單一偵測模型相較，本研究提出之融合模型確實能提升準確率，並提供更穩定的偵測結果。當使用連續過濾時，除可獲得較高準確率外，也能對各階段之偵測精度進行分析，提升結果之實用性。此外，雖然模型融合時嘗試組合各種不同特質的單一模型可影響準確性，但我們發現在多階段過濾的流程下，對於偵測效能之影響有限。由上述結果可知，本研究提出方法確有助於改善詐騙偵測準確率，提供消費者更周全的線上拍賣安全防護。

本論文後續章節如下：第二節介紹相關背景知識與技術，第三節為本研究提出過濾式模型融合方法與偵測流程，第四節為實驗結果及討論，第五節為結論與未來工作。

貳、背景知識與技術介紹

本節將介紹與本論文相關之背景知識、術語以及技術，以利後續章節之討論。

一、線上拍賣詐騙

電子商務帶來的經濟效益有目共睹，而線上拍賣更是其中重要的一環，但也引起了許多不肖份子的覬覦。根據美國 Internet Crime Complaint Center (IC3)報告指出，2018 年美國的網路詐騙申訴案件為 351,937 件，損失金額 27 億美金；2019 年申訴案件增加為 467,361 件，損失金額更高達 35 億美金(NW3C 2019)。上述數據顯示網路詐騙並沒有隨著政府的公開報告而降低，反而與日俱增。探究其原因，除了電子商務的便利性讓民眾戒心降低，與詐騙者不斷翻新的詐騙手法也有關聯。表 1 所列為數種常見的詐騙類型，其中以第一種收款不出貨最為常見。詐騙者在買家購買付款後，便以各種理由推託不出貨，最後消失無蹤。第二種方式則以不實圖文吸引消費者下標，再寄送品項不符之貨品或瑕疵品。若消費者進行申訴，則以交易糾紛為藉口，讓買方進入曠日廢時的仲裁流程。第三種方式則為發送假得標信或訊息給買方，聲稱需更改付款方式，讓消費者將帳款匯入詐騙者設定的帳號。第四種則最為惡劣，假冒買家截取寄件者貨品，將詐騙責任轉移給不知情的第三者。凡此種種，顯示相關當局需持續加以重視，並發展防範對策。

表 1：常見線上拍賣詐騙類型

詐騙類型	說明
收款不出貨	收款不出貨，累積足夠的受害者後，最後捲款潛逃。
圖文不符，矇騙消費者	消費者所購買商品與賣家刊登於交易平臺上所見不同。
寄發假得標信，攔截貨款	詐騙者從非法管道取得受害者之電子信箱，寄發假得標信，要求匯款至指定帳戶。
偽裝賣家回收貨品	詐騙者要求賣家商品必須指定寄貨地點以及收件人姓名，之後假裝寄件者，到指定地點要求取回貨品。

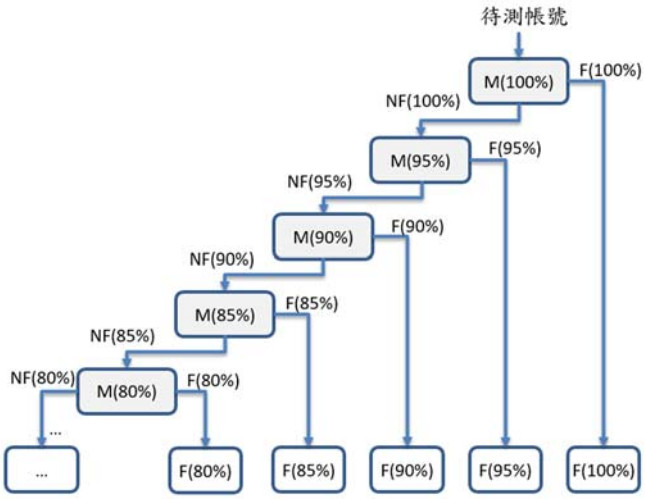
二、詐騙偵測與模型融合

詐騙偵測為異常偵測的一種，常見的做法為：(1)設計一組有效的偵測屬性集，(2)配合特定學習演算法產生偵測模型。在理想狀況下，若能事先得知測試資料之詐騙者與正常者比例(例如 1:2)，便可以相同比例訓練集來建構模型，以提升偵測準確率。但測試集中各類型資料配比實際上無法預先預知，因此很難產生萬用的單一模型。例如，針對 1000 個可疑帳號進行偵測，並無法事先知道其中詐騙者的比例(或許有 500 個，或許只有 1 個)。上述狀況造成的限制，便可透過模型融合來改善。

模型融合的要點為結合各種不同特性之單一模型，透過截長補短或相互配合的步驟，產生更精確的偵測結果。此外，為避免產生過適(overfitting)現象，模型融合流程對於訓練資料集的運用亦需更加精細。以下將介紹前人研究中提出的模型融合方法：

(1) 多階段模型融合

此種作法通常會建立多個分類模型($Model_1$~$Model_n$)，而後將待測資料逐一輸入各種模型，每個模型 $Model_i$ 會將上一階段模型($Model_{i-1}$)之偵測結果做為額外輸入，最後再整合產出偵測結果。針對此種作法，Chang & Chang (2012)曾提出連續過濾式之線上拍賣詐騙偵測方法。參考圖 1，過程中先以各種不同階段詐騙者特質產生模型(M(100%), M(95%), …, M(80%))，若任何一步驟可將待測帳號判定為詐騙者，流程便可中斷。此種作法雖然合理，但在建構偵測模型時，均假設待測資料中的詐騙者與正常者比例固定，因此採用相同配比的訓練集來塑模。如前所述，在實際狀況下，我們不可能事先預知測試集中詐騙者與正常者的比例。此外，流程中只以詐騙者為主，過程中除非被辨識為詐騙者，否則均需進行後續流程。此設計明顯忽略對於正常者判別的重要性，可能影響總體的預測準確率。

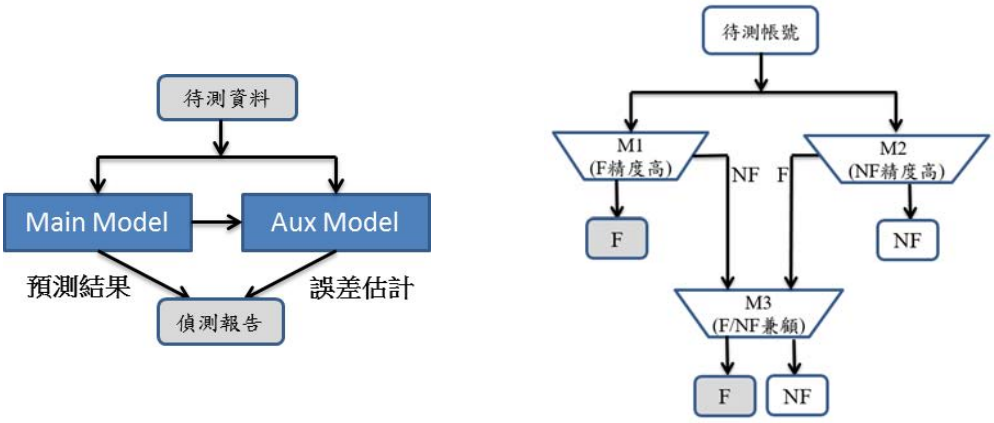


M(r%): 以訓練資料中帳戶前 r%交易歷史所建立之偵測模型
F(r%): 處於生命週期第 r%時間之詐騙帳號， NF(r%): 處於生命週期第 r%時間之正常帳號

圖 1：多階段連續過濾(Chang & Chang 2012)

(2) 互補式模型融合

此類型的基本形式由一個主模型(main model)配合一個輔助模型(Aux Model)而組成。參考圖 2(a)，其中主模型提供主要的偵測結果，輔助模型則估計主模型的偵測誤差，再產生最後的偵測結果(Darudi, Bashari, & Javidi 2015)。圖 2(b)則為利用互補概念所建構的線上拍賣詐騙偵測流程(Chang, Liu, & Lee 2020)，其中模型 M1 擅長偵測詐騙者，模型 M2 擅長偵測正常者，相互互補。當 M1, M2 無法決定時，再將待測帳號以第三個模型 M3 進行偵測。互補式模型融合雖然較複雜，但理論上可產生更精確的偵測結果。然而，此種作法的效能瓶頸在於最後階段的綜合模型之偵測能力。以上例而言，若 M3 無法對這些難以偵測之帳號進行有效分類，則之前 M1, M2 之準確性可能被抵銷。有鑑於此，運用此種融合流程時，可進一步設計更多樣化的模型並賦予不同任務，使待測帳號進入最後階段前能儘早被測出。



(a) 基本互補式融合流程(Darudi et al. 2015)　(b) 應用互補融合流程之研究(Chang et al. 2020)

圖 2：互補式模型融合流程

三、隨機森林(Random Forest)

本研究採用多模型融合方式進行詐騙偵測，過程中大都以隨機森林(Random Forest)來建構單一模型。因此，以下將概略介紹隨機森林的流程與特性。隨機森林的基本原理為結合多棵 Gini 索引法(Gini index)之決策樹，並加入隨機分配的訓練資料，大幅度提高運算的準確率。隨機森林之運作流程如下(Amrehn et al. 2018)：

(1) 從資料集 X 中以 Bootstrap Sampling 方式抽選每棵樹決策樹 t_i 的所需之樣本集 B_i，一直到 B_i 和 X 的大小相等。
(2) 為每個樣本集 Bi 選擇一個隨機的特徵子集，並將其用於決策樹 t_i 的訓練。
(3) 挑選一種信息增益(Information gain)度量方法來協助產生未修剪之決策樹(unpruned decision trees)。
(4) 針對特定待測資料，根據各決策樹的輸出挑選最常出現的類別做為預測結果。

隨機森林雖然無法提供傳統決策樹之規則解釋能力，但仍擁有以下諸多優點：例如，對於各類型資料集，隨機森林能提供良好的分類準確度；資料集中若有遺失項目，仍可提供高準確度；對於不平衡的分類資料集，隨機森林可以自動平衡其誤差。此外，隨機森林亦可以被延伸應用在未標記的資料上，並進行非監督式學習。

參、以多模型融合為基礎之詐騙偵測方法

有關線上拍賣詐騙偵測，相關研究多使用單一模型進行預測，亦有學者採用階段性偵測，但效能仍有其限制。為發展更有效之偵測方法，本節將先仔細分析單一偵測模型之缺點，接著探討如何透過調整詐騙者以及正常者不同資料配比對，產生不同特性之模型。最後，透過模型融合概念，提出多階連續過濾偵測以及平衡過濾偵測流程，以多模型架構提升詐騙偵測的效能。

一、詐騙偵測屬性集

偵測模型的效能與所使用之屬性集息息相關，合適的屬性集能有效表示資料特質，產生高準確率的偵測模型。詐騙偵測屬性主要根據詐騙者的特徵來設計，希望能透過其交易歷史，分辨詐騙者與正常者之不同。根據前人研究(Chau & Faloutsos 2005; Chau et al. 2006; Chang & Chang 2012; 鄭孝儒 2010)，拍賣詐騙偵測屬性大致可分為以下三類：評價相關屬性(feedback-related features)、價格相關屬性(price-related features)與交易角色相關屬性(role-related features)，其用途與意義分述如下：

(1) 評價相關屬性：詐騙者經常急於在短時間內累積評價，以取得消費者信任。為避免曠日廢時，詐騙者可透過假交易來達成。然而，此舉將導致其評價累積的模式異於一般會員，因此可設計與評價累積方式相關之屬性來偵測。例如，若有新加入會員突然在短短幾天內，累積數十個甚至上百個正評價，便需密切加以觀察。

(2) 價格相關屬性：除假交易外，詐騙者亦可透過大量購買低價商品(如髮夾，手圈等)累積評價，以避免創建過多假帳號引起注意。俟獲得足夠評價後，再嘗試販售高價品進行詐騙。此種方式較假交易更不易察覺，但透過分析其交易歷史中的交易金額，便可獲得相關資訊。因此，可設計與交易價格變動之屬性來偵測。

(3) 角色相關屬性：詐騙者一開始可能扮演買方累積評價，後期則以賣方角色出現，因此造成明顯的角色轉換。對此特質，可根據其交易歷史，分析會員交易角色的變化，設計相關的屬性加以偵測。

考量上述詐騙者不同特質，為使偵測結果更為準確，本研究綜合前人所提出之各種偵測屬性(Chang & Chang 2009; Chang & Chang 2012; Chau & Faloutsos 2005; 鄭孝儒 2010)，並去除其中過於針對性之屬性(與特定機制相關，如 Yahoo 的安全賣家)，最後共獲得 52 種偵測屬性。參考表 2，其中列出本研究使用的偵測屬性之名稱、說明及其所屬類型。例如，DensityOfPos (編號 1)為其受測帳號在資料蒐集期間(單位:天)獲得正評價的密度(評價總數/天數)，為一種典型的「評價相關」屬性。有關所有屬性之詳細計算方式，礙於篇幅有限，請參閱表中最左欄之參考文獻。

表 2：本研究使用之 52 種詐騙偵測屬性

出處	編號	屬性名稱	說明	屬性類型
Chang& Chang (2009)	1	DensityOfPos	正評的密度	評價相關
	2	EndCloseToPos	平均獲得正評的時間	評價相關
	3	RatioOfPos	正評佔總評價的比例	評價相關
	4	RatioOfSToS	其他賣家給正評的比例	評價相關
	5	DensityOfNeg	負評的密度	評價相關
	6	LastNegCloseToCur	最後一筆負評到現在的時間	評價相關
	7	RatioOfNeg	負評佔總評價的比例	評價相關

Chau& Faloutsos (2005)	8	MeanBuyingFirst15	前 15 天平均買價	價格相關
	9	MeanBuyingFirst30	前 30 天平均買價	價格相關
	10	MeanBuyingLast15	後 15 天平均買價	價格相關
	11	MeanBuyingLast30	後 30 天平均買價	價格相關
	12	MeanSellingFirst15	前 15 天平均賣價	價格相關
	13	MeanSellingFirst30	前 30 天平均賣價	價格相關
	14	MeanSellingLast15	後 15 天平均賣價	價格相關
	15	MeanSellingLast30	後 30 天平均賣價	價格相關
	16	StdBuyingFirst15	前 15 天買價標準差	價格相關
	17	StdBuyingFirst30	前 30 天買價標準差	價格相關
	18	StdBuyingLast15	後 15 天買價標準差	價格相關
	19	StdBuyingLast30	後 30 天買價標準差	價格相關
	20	StdSellingFirst15	前 15 天賣價標準差	價格相關
	21	StdSellingFirst30	前 30 天賣價標準差	價格相關
	22	StdSellingLast15	後 15 天賣價標準差	價格相關
	23	StdSellingLast30	後 30 天賣價標準差	價格相關
	24	RatioOfBuyingRate	買東西比例	角色相關
Chang& Chang (2012)	25	MeanBuying	平均買價	價格相關
	26	MeanSelling	平均賣價	價格相關
	27	StdBuying	買價標準差	價格相關
	28	StdSelling	賣價標準差	價格相關
	29	GiveManRating	給評價者平均評價	評價相關
	30	Rating	評價分數	評價相關
	31	BuyingNumber	買東西數	角色相關
	32	BuyingNumberOfPos	買東西正評數	評價、角色相關
	33	SellingNumber	賣東西數	角色相關
	34	SellingNumberOfPos	賣東西正評數	評價、角色相關
	35	SellingNumberLast30	最後 30 天賣的商品數	角色相關
	36	SellingNegtiveNumberLast30	最後 30 天賣商品得到的負評	評價、角色相關
	37	NumberOfPostive	正評的數量	評價相關
	38	MeanSellingLast15-MeanSelling	後 15 天平均賣價-平均賣價	價格相關
	39	MeanSellingLast30-MeanSelling	後 30 天平均賣價-平均賣價	價格相關
	40	MeanSellingLast15-MeanSellingLast30	後 15 天平均賣價-過去 30 天平均賣價	價格相關
鄭孝儒 (2010)	41	StdSellingLast15-StdSelling	後 15 天賣價標準差-賣價標準差	價格相關
	42	StdSellingLast30-StdSelling	後 30 天賣價標準差-賣價標準差	價格相關
	43	StdSellingLast15-StdSellingLast30	後 15 天賣價標準差-過去 30 天賣價標準差	價格相關
	44	SellingLast15MeanTimeInterval	後 15 天平均交易時間間隔	價格相關
	45	SellingLast30MeanTimeInterval	後 30 天平均交易時間間隔	價格相關
	46	SellingMeanTimeInterval	賣東西平均交易時間間隔	評價相關
	47	SellingMeanTimeLastOneTwo	最後二筆交易之間的時間間隔	評價相關
	48	MAXSellingPrice	最大賣價	價格相關
	49	AllRating	所有評價的次數	評價相關
	50	SellingPrice_LastOneTwo	最後二筆交易之間的價格差	價格相關
	51	(SellingPrice/SellingTime)_LastOneTwo	最後兩筆交易價格差/交易時間差	價格相關
	52	LastSellingPrice	最後一筆交易賣價	價格相關

二、單一偵測模型之特性與效能瓶頸

在前人研究中，經常使用單一模型(如 Random Forest)進行詐騙偵測，雖可獲得不錯的結果(Xuan et al 2018; Kumar et al. 2019; Chau et al. 2006)，但會產生至少以下二項明顯的缺點：

(1) 實務上，若訓練集中各類型資料的比例可與測試資料相同，通常可獲得較高的偵測準確率。例如，前人研究中經常假設訓練集、測試集中的詐騙者與正常者比例均為 1:2 (以 F:NF=1:2 來表示)。然而，在實際狀況下，偵測資料之特性(類別比例)並無法事先得知，將導致模型的效能受到影響。例如，以 F:NF=1:2 配比建立之模型，對於 F:NF=1:8 測試集之偵測效能可能不如預期。

(2) 學習方法在訓練分類模型時，為產生較高的總體分類準確率，會有偏向比例較高的資料項目之傾向。例如，若訓練資料中的詐騙者(F)與正常者(NF)佔比為 1:5，一個 naïve 分類模型可將資料一律分類為 NF，則總體準確率仍可高達 5/6=83.333%。此偏差特性對於異常偵測而言，可能會產生非常負面之影響(容易將異常案例歸類為正常案例)。

為說明上述概念，我們以實際交易資料來進行分析。資料集中共有 1500 個拍賣帳號，其中詐騙者與正常者分別為 500 與 1000 筆，訓練集與測試集之比例為 2:1。參考表 3 之結果，每列代表使用不同類別比例所建構之分類模型(此處使用 Random Forest)，例如 Model(1:1)代表訓練資料中詐騙者與正常者之比例為 1:1。為了說明不平衡測試資料對於偵測模型的影響，此處則採用 F:NF=1:8 的測試集。由表中數據可看出，當以 Model(1:8)來進行偵測時，可獲得最高的準確率(Accuracy)，與上述第(1)點推論相符。然而，由表中另可看出，詐騙者的召回率(recall_F)卻逐次遞減。使用 Model(1:1)進行偵測有 0.856，但在 Model(1:8)時只剩 0.602。參考圖 3 中的趨勢圖，可看出詐騙者的召回率隨著訓練資料中詐騙者(F)比例下降而下降。至終，Model(1:1)與 Model(1:8)的詐騙者召回率(recall_F)相差竟高達 25.4%(=0.856-0.602)。換言之，如上述第(1)點所述，Model(1:8)模型傾向將大多數不容易判別的案例判斷為 NF，以獲得較高的總體準確率。如此一來，Model(1:8)雖具有高準確率(93%)，但實際應於詐騙偵測時，卻無法獲得應有期望效果。相同的結果也發生在不同資料配比的資料集中(如 F:NF=1:4, 1:1)，但礙於篇幅，此處並不列出詳細結果。

綜合上述觀察發現，若偵測模型(Model(1:n))與測試集(Test Set(1:m))中的類別資料配比相同(n=m)，可獲最高的準確率。但由於 m 無法事先得知，因此使 n 的選擇成為難題。此外，詐騙者召回率均可能隨著 n 的增加而明顯降低。換言之，單一模型顯然無法對比例未知之不平衡資料集進行有效偵測。有鑑於此，本研究將提出不同的方法，以改善傳統單一偵測模型所遭遇的困難。

表 3：以不同資料比例建立偵測模型對不平衡測試資料之偵測結果

偵測模型 Model(F:NF[1])	Test Set(F:NF=1:8)				
	Accuracy	prec_NF	recall_NF	prec_F	recall_F
Model(1:1)[2]	0.863	0.973	0.864	0.512	0.856
Model(1:2)	0.908	0.959	0.932	0.656	0.764
Model(1:4)	0.927	0.950	0.965	0.773	0.698
Model(1:6)	0.927	0.940	0.978	0.826	0.628
Model(1:8)	0.930	0.937	0.985	0.878	0.602

[1] F:NF 為資料集中 Fraud(詐騙者)與 Non-Fraud(正常者)資料筆數之比例

[2] Model(1:1)表示以 F:NF=1:1 的比例產生之訓練資料來建立模型

[3] prec(precision):精度, recall:召回率

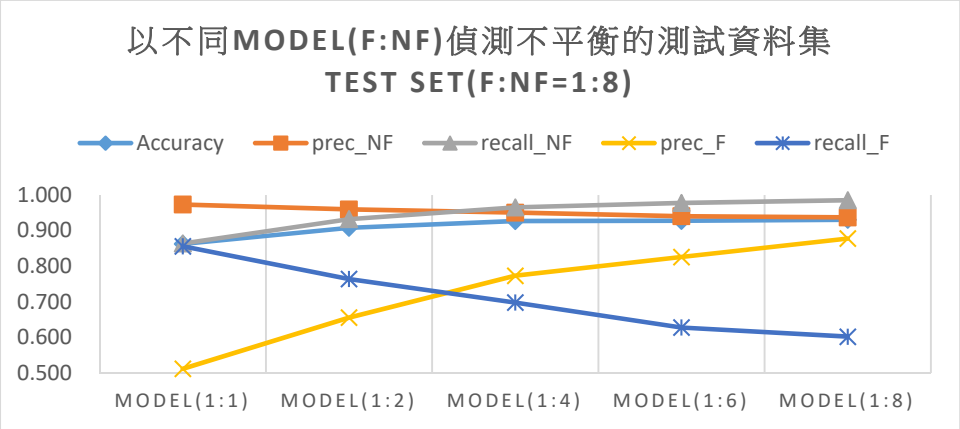


圖 3：以不同資料比例塑模對不平衡測試集(F:NF=1:8)之偵測結果

根據以上說明，使用單一模型進行詐騙偵測顯然有待改進，因此本研究將根據模型融合概念，結合多個不同特性之模型，以提升總體偵測效能。為建立有效的多模型偵測架構，以下將探討各種不同類別配比的訓練資料，如何影響偵測模型的效能。以下分析仍採用與表 3 相同之 1500 筆交易記錄，訓練集與測試集之 F:NF=1:2，訓練集與測試集資料數量為 2:1。我們使用 Random Forest, AdaBoost, Decision Tree, Multi-Layer Perception 與 Nearest Neighbor 等單一分類模型，在不同詐騙者與正常者配比下(F:NF)塑模，並進行偵測效能比較(請參考附錄 A)。其中，以 Random Forest 模型表現最佳。因此以下將以 Random Forest 分別針對訓練集之 NF:F=1:5, 1:3, 1:2, 1:1, 2:1, 3:1, 5:1 塑模，說明各種偵測指標的變化。參考表 4 之結果，可看出偵測準確率仍以 2:1 之模型最高(0.869)，但對於正常者(NF)與詐騙者(F)之偵測，則可看出不同資料配比產生之差異：

(1) 對正常者(NF)而言，當 F:NF=5:1 時，產生模型具有極高之精度(prec_NF=0.977)。當 F:NF=3:1 時，prec_NF 略有下降(0.959)，但召回率卻提升至 0.743。總體而言，正常者的偵測精度隨著 F 之比例增加而增加，但召回率則反之。

(2) 對詐騙者(F)而言，當 F:NF 1:5 時，偵測精度可達 0.900。當 F:NF=1:3 時，

prec_F 略降至 0.882，但召回率(recall_F)提升至 0.693。依照表中的趨勢可發現，詐騙者的偵測精度(prec_F)隨著 NF 的比例增加而增加，但召回率則反之。

根據上述觀察，使用不同資料配比產生之模型，具有不同特質。若能有效加以串接，便有機會改善整體的偵測效能。

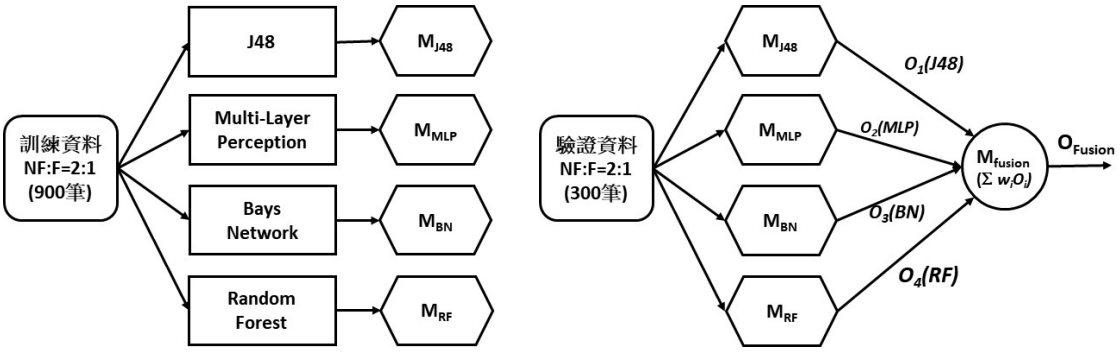
表 4：在不同 NF:F 配比下塑模之偵測結果(使用 Random Forest 塑模)

偵測模型	Test Set(F:NF=1:2)				
RandomForest	Accuracy	prec_NF	recall_NF	prec_F	recall_F
F:NF =5:1	0.768	0.977	0.669	0.594	0.969
F:NF =3:1	0.806	0.959	0.743	0.644	0.935
F:NF =2:1	0.829	0.943	0.793	0.686	0.903
F:NF =1:1	0.862	0.910	0.881	0.775	0.825
F:NF =1:2	0.869	0.877	0.934	0.848	0.739
F:NF =1:3	0.867	0.862	0.954	0.882	0.693
F:NF =1:5	0.849	0.834	0.966	0.900	0.614

如何設計多模型偵測架構需有特別考量，以有效整合各模型之優點。面對此問題，以下先介紹一種最直覺的融合方式：

(1) 針對相同訓練集，以不同學習方法(如 J48, MLP, Bays Networks, Random Forest 等)產生多個偵測模型(參考圖 4(a))。

(2) 針對待測資料，以線性迴歸方式整合各種模型的輸出，彙總後產生偵測結果(參考圖 4(b))。



(a) 建立單一模型　　(b) 建立融合模型

圖 4：詐騙偵測模型融合方法

上述方法看似簡單合理，但並未仔細考慮各種模型的特性(對不同 F:NF 測試集之偵測精度、召回率等)。若只單純將各模型輸出以線性迴歸方式來組合，可能產生與單一模型相同的問題(參考本節之前的討論)。因此，本研究將以不同方式來產生多模型架構，希望能先依照需求設計各種不同特點之模型，之後再以合適的方式加以組合。

三、以模型融合概念建立多階段詐騙偵測模型

根據上述討論，我們發現在不同訓練資料配比下，確實會對分類器造成顯著的預測精度影響。因此，本研究將利用此特性設計更有效的融合模型。為配合不同模型融合架構，我們將先建置如下之五種偵測模型(M1~M5)，其特質如下：

表 5：本研究模型融合時使用之各種單一模型

模型	訓練資料配比	模型特點	用途
M1	F:NF=5:1	對 Non-Fraud 具有高偵測精度	用於判別正常者
M2	F:NF=1:5	對 Fraud 具有高偵測精度	用於判別詐騙者
M3	F:NF=3:1	對 Non-Fraud 具有較高偵測精度	用於判別正常者
M4	F:NF=1:3	對 Fraud 具有較高偵測精度	用於判別詐騙者
M5	F:NF=1:2	對 Fraud,Non-Fraud 具有均衡偵測能力	用於判別正常者與詐騙者

(一) 連續過濾法

參考圖 5(a)，本研究提出之第一個多模型偵測架構為連續過濾 (Successive Filtering)，透過組合五個不同特質之模型(M1~M5)，對待測帳號 acc 進行過濾偵測。架構中的各模型具有不同偵測精度，且偵測對象不同，運作之詳細步驟如下(參考圖 5(b))：

(1) M_1 對於正常者(NF)具有最高之偵測精度，因此只要待測帳號 acc 於 M_1 被判定為正常者(NF)，即將其加入 NF_List 並結束偵測。若非如此，繼續進行下一階段之測試(M_2)。

(2) M_2 對於詐騙者(F)具有高精度預測，因此只要於 M_2 被檢測為詐騙者，即將其加入 F_List 並結束偵測。若非如此，則進行下一階段之測試(M_3)。

(3) M_3 對於正常者(NF)之偵測精度略低於 M_1，但具有較高之召回率，若 acc 為正常者，可降低其被錯過的機會。與(1)相同，若 acc 於 M_3 被判定為正常者(NF)，即將其加入 NF_List 並結束偵測，否則將繼續進行下一階段之測試(M_4)。

(4) M_4 對於詐騙者(F)之偵測精度略低於 M_2，但具有較高之召回率，若 acc 為詐騙者，可降低其被錯過的機會。若 acc 於 M_4 被判定詐騙者(F)，即將其加入 F_List 並結束偵測，否則將繼續進行下一階段之測試(M_5)。

(5) 最後，如果都無法檢測出 acc 之類別，則交由 M_5 進行最後判定(M_5 之偵測效能較為平衡，對於詐騙者與正常者並無特別偏重)。

(二) 平衡式過濾

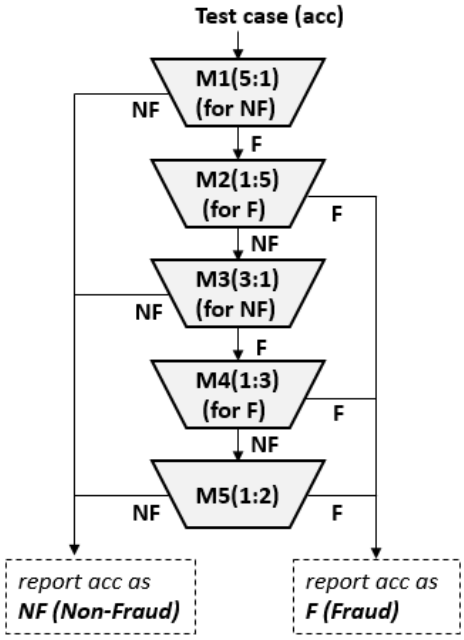
上述連續過濾之優點為逐步運用高精度的模型，過濾出較確定之帳號分類(F 或 NF)。然而，當 M_1 與 M_2 或者 M_3 與 M_4 之判斷不同時，可能導致誤判。因此，本研究再提出平衡式過濾法，希望能將具有爭議的帳號，留待後續模型進行偵測。詳細之偵測流程圖與虛擬碼如圖 6 所示，其運作步驟如下：

(1) 當待測帳號 acc 進入平衡式偵測流程時，同時由 M_1(對正常者擁有高偵測精度)進與 M_2(對詐騙者擁有高偵測精度)進行偵測，而非依序由 M_1 與 M_2 進行偵測。當 M_1 與 M_2 判斷結果相左，無法對 acc 之類別做出一致性決定，便會繼續進行下一階段偵測。

(2) 在此階段，則將 acc 由 M_3 與 M_4 同時進行判別，其流程與 M_1, M_2 之運用相同。

(3) 最後，將 M_1~M_4 無法分辨之帳號交由 M5 進行最後判定。

由上述流程可看出，平衡式過濾法具有雙重特點，既能有效運用高精度模型進行偵測，也能避免過早決爭議性帳號，或可藉此提升偵測準確率。上述由本研究提出之二種多模型詐騙偵測流程，將在第四節中對其效能進行探討。

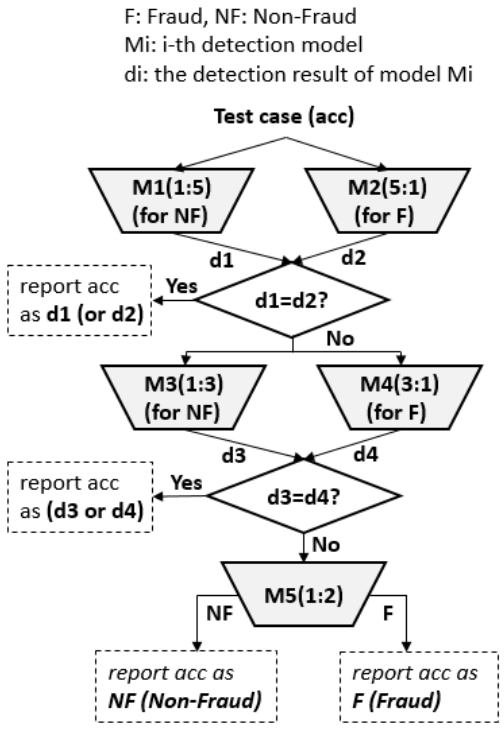


(a) 偵測架構流程圖

```
procedure SuccessiveFiltering
input     M1~M5: 模型集合
          TestSet: 待測帳號集合
output    F_List: 儲存被判斷為詐騙者(F)之帳號
          NF_List: 儲存被判斷為詐騙者(NF)之帳號
begin
    // F: Fraud, NF: non-Fraud
    for each account acc in TestSet do
        if (M1.detect(acc)==NF)
        // 若 M1 判斷 acc 為 NF, 便將其標示為 NF
            NF_List.add(acc)
        else if (M2.detect(acc)==F)
        //若 M2 判斷 acc 為 F, 便將其標示為 F
            F_List.add(acc)
        else if (M3.detect(acc)==NF)
        //若 M3 判斷 acc 為 NF, 便將其標示為 NF
            NF_list.add(acc)
        else if (M4. detect(acc)==F)
        //若 M4 判斷 acc 為 F, 便將其標示為 F
            F_List.add(acc)
        else if(M5.detect(acc)==NF)
            NF_List.add(acc)
        else
            F_list.add(acc)
end procedure
```

(b) 偵測流程之虛擬碼

圖 5：運用連續過濾進行詐騙偵測



(a) 偵測架構流程圖

```
procedure BalancedFiltering
input     M1~M5: 模型集合
          TestSet: 待測帳號集合
output    F_List: 儲存被判斷為詐騙者(F)之帳號
          NF_List: 儲存被判斷為詐騙者(NF)之帳號
begin
    for each account acc in TestSet do
        d1=M1.detect(acc); d2=M2.detect(acc)
        if (d1==d2) {
            if (d1==NF)
                NF_list.add(acc)
            else
                F_list.add(acc)
        } else {
            d3=M3.detect(acc)
            d4=M4.detect(acc)
            if (d3==d4) {
                if (d3==NF)
                    NF_list.add(acc)
                else
                    F_list.add(acc)
            } else {
                if(M5.detect(acc)==NF)
                    NF_List.add(acc)
                else
                    F_list.add(acc)
            }
        }
end procedure
```

(b) 偵測流程之虛擬碼

圖 6：平衡式過濾偵測流程

肆、實驗結果

為驗證本研究提出方法之有效性，我們以 Yahoo!拍賣網站蒐集之實際交易資料進行測試，比較融合模型與傳統偵測模型的差異。此外，透過分析多模型架構中各模型的偵測精度，探討是否能提供使用者更具實用性之偵測結果。此外，我們也將嘗試使用不同的資料配比來建立多模型架構，並討論其對於總體效能的影響。

一、實驗設定

本研究採用文獻中(Chang et al. 2020)所蒐集之 Yahoo!拍賣資料做為資料集，

共包含 1500 個帳號之交易資料 ，其中詐騙者為 500 位，正常者為 1000 位(比例為 1:2)。為進行實驗，再將其中 1000 筆做為訓練集(333 位詐騙者、667 位正常者)，500 筆做為測試集(167 位詐騙者、333 位正常者)。實驗時，會根據所需的資料類型比例組合所需的資料集。例如，若需建立 F:NF=1:4 之偵測模型，則會從 1000 筆訓練資料中隨機抽取 166 位詐騙者與所有 667 位正常者組成訓練集。實驗結果均為 10 次實驗之平均，資料集則依照所需比例隨機抽選產生。實作時，本研究採用 Keras 機器學習套件，並配合 Python 來撰寫。

表 6：混淆矩陣(Confusion Matrix)

Classification		Predictive Classes	
		Positive	Negative
Actual Classes	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

為比較各種模型或方法之效能，以下介紹實驗結果所使用的評量指標(參考本段下方公式)。參考表 6 之 2x2 混淆矩陣(Confusion Matrix)，假設分類結果僅有 Positive 與 Negative 二種(Positive 可代表詐騙偵測中的詐騙者)。由表中可知，全部受檢案例共有 N=TP+TN+FP+FN，其中被正確檢出的個數有 TP+TN，因此分類器的準確率(Accuracy)可定義為(TP+TN)/N。除準確率外，還可針對各種分類產生精度(Precision)與召回率(Recall)二種指標。以 Precision(Positive)表示經分類器判別為 Positive 的案例中，共有多少真正為 Positive 之比率；而 Recall(Positive)則表示測試資料集中所有陽性案例被此分類器檢出的比率。除個別效能外，相關研究中亦常使用 F-Measure 與 MCC(Matthews correlation coefficient) 來呈現分類器的綜合效能。F-measure 同時納入特定分類之 precision 與 recall 為變數，並計算此二個變數的調和平均數(harmonic mean)，其值將會介於[0,1]，越接近 1 為佳。MCC 則同時考量 Positive 與 Negative 類別的偵測結果，因此應用更為廣泛。MCC 值會介於[-1,+1]之間，越接近 1 表示分類效果越佳。

Accuracy = (TP+TN)/(TP+TN+FP+FN) = (TP+TN)/N

Precision(Positive) = TP/(TP+FP), Precision(Negative) = TN/(TN+FN)

Recall(Positive) = TP/(TP+FN), Recall(Negative) = TN/(TN+FP)

$$\mathrm{F-Measure} = 2 \times \frac{precision \times recall}{(precision + recall)}$$

$$\mathrm{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

二、多模型詐騙偵測方法之效能測試

本節將比較單一模型、連續過濾法、平衡式偵測法之詐騙偵測效能。其中，除了連續過濾法之 M3, M4 模型外(使用 AdaBoost 塑模)，其餘所有模型均採用

Random Forest 做為塑模方法。產生各個模型時，其訓練資料之類型比例與數量如下：

表 7：本研究使用之多模型架構中各模型之訓練資料類型配比

模型	訓練集類型比例	詐騙者數量(個)	正常者數量(個)
M1	F:NF=5:1	333	67
M2	F:NF=1:5	133	667
M3	F:NF=3:1	333	111
M4	F:NF=1:3	222	667
M5	F:NF=1:2	333	667

表 8 為使用單一模型、連續過濾法、平衡式偵測法對 F:NF=1:2 測試集之偵測結果。其中，Model(m:n)列為使用單一模型之偵測結果，但由於無法事先知道測試集的資料配比，因此最後以平均值來評估其效能(參考 Average(Single)列)。由表中可看出，針對此測試集，連續過濾法(Model(successive))之準確率最高(0.871)，平衡式偵測法(Model(Balanced))次之(0.866)，單一模型最低。當考量偵測結果之綜合表現，(F-Measure 與 MCC)，可發現連續過濾法之表現仍為最佳。值得注意的是，單一模型之綜合表現(MCC)依照塑模之資料配比產生大幅變化(0.611~0.705)。由此可知，在測試資料配比未知狀況下，單一模型確實無法提供穩定的偵測結果。

表 8：單一模型、連續過濾法、平衡式偵測法之效能比較，Test Set(F:NF=1:2)

Detection Model	Test Set (F:NF=1:2)							
	Accuracy	prec_NF	recall_NF	prec_F	recall_F	F-Measure(F)	F-Measure	MCC
Single Model(1:1)	0.861	0.917	0.871	0.766	0.842	0.802	0.849	0.705
Single Model(1:2)	0.870	0.876	0.939	0.858	0.733	0.791	0.851	0.677
Single Model(1:4)	0.861	0.846	0.969	0.913	0.646	0.757	0.842	0.642
Single Model(1:6)	0.856	0.831	0.985	0.954	0.597	0.734	0.839	0.627
Single Model(1:8)	0.849	0.822	0.989	0.962	0.570	0.716	0.832	0.611
Average(Single)*	0.859	0.858	0.950	0.891	0.677	0.760	0.842	0.652
Model(Successive)	**0.871**	0.880	0.935	0.851	0.744	**0.794**	**0.852**	**0.688**
Model(Balanced)	0.866	0.871	0.937	0.853	0.723	0.783	0.846	0.672

* Average(Single)為表中五種單一模型偵測結果之平均值。

為進一步驗證三種方法之優劣，本研究也對各種不同測試集進行比較(Test Set(F:NF=1:1, 1:2, 1:4, 1:8))，結果如表 9 所示。由表中可發現，其結果與表 8 類似，均以連續過濾法為最佳，平衡式偵測法居次。由上述結果可知，在測試資料配比狀況未知時，本研究提之多模型偵測法確可獲得較佳的準確率，產生較穩定的偵測結果。

表 9：單一模型、連續過濾法、平衡式偵測法之效能比較

Test Set(F:NF=1:1, 1:2, 1:4, 1:8)

Detection Model	Test Set (F:NF=1:1)							
	Accuracy	prec_NF	recall_NF	prec_F	recall_F	F-Measure(F)	F-Measure	MCC
Average(Single)	0.827	0.769	0.946	0.935	0.707	0.800	0.839	0.674
Model(Successive)	**0.856**	0.807	0.937	0.925	0.774	**0.843**	**0.861**	**0.714**
Model(Balanced)	0.843	0.786	0.943	0.929	0.742	0.825	0.850	0.695
	Test Set (F:NF=1:2)							
Detection Model	Accuracy	prec_NF	recall_NF	prec_F	recall_F	F-Measure(F)	F-Measure	MCC
Average(Single*)	0.859	0.858	0.950	0.891	0.677	0.760	0.842	0.652
Model(Successive)	**0.871**	0.880	0.935	0.851	0.744	**0.794**	**0.852**	**0.688**
Model(Balanced)	0.866	0.871	0.937	0.853	0.723	0.783	0.846	0.672
Detection Model	Test Set (F:NF=1:4)							
	Accuracy	prec_NF	recall_NF	prec_F	recall_F	F-Measure(F)	F-Measure	MCC
Average(Single)	0.890	0.917	0.945	0.821	0.708	0.748	0.912	**0.675**
Model(Successive)	**0.901**	0.933	0.938	0.791	0.776	**0.783**	**0.919**	0.633
Model(Balanced)	0.897	0.924	0.944	0.799	0.741	0.769	0.916	0.632
	Test Set (F:NF=1:8)							
Detection Model	Accuracy	prec_NF	recall_NF	prec_F	recall_F	F-Measure(F)	F-Measure	MCC
Average(Single)	0.911	0.952	0.945	0.729	0.710	0.702	0.923	0.628
Model(Successive)	**0.915**	0.962	0.938	0.677	0.776	**0.723**	**0.925**	**0.633**
Model(Balanced)	0.911	0.959	0.937	0.669	0.76	0.712	0.922	0.609

在正常運作的拍賣網站中，詐騙者(Fraud 分類)比例應明顯低於一般交易者(Non-Fraud 分類)。但為避免詐騙事件影響消費者使用意願，運用偵測系統協助辨識詐騙者，可降低消費者疑慮、提升交易安全性。然而，無論偵測系統的使用者為網站管理者或終端消費者，考量蒐集會員交易歷史並進行分析所耗費之網路與運算成本，針對「可疑者」進行篩檢(而非「普篩」)，應較合理也較符合成本效益。如此一來，偵測系統所面對之待測資料，其詐騙者與正常者比例差距便可能因此縮小。經驗豐富之使用者所認為的「可疑者」確為「詐騙者」之比例應該較高(例如 F:NF=1:1 或 1:2)；而較生疏的新手所認為之「可疑者」之詐騙比例則可能較低(例如 F:NF=1:4 或 1:8)。

然而，為探究極度不平衡資料集對各種偵測方法之影響，以下我們分別以詐騙者與正常者配比(F:NF)為 1:20、1:30 與 1:50 之測試資料進行實驗。參考表 10 之結果，在三種測試集中，五種單一模型之平均準確率(Average(Single))均獲得些微領先(參考 Accuracy 欄)。然而，當考慮詐騙者之召回率(recall_F)時，可發現單一模型(Average(Single))之表現明顯低於本研究提出之連續過濾法(Model(Successive))。在三種測試集中，單一模型平均可獲得 0.674, 0.679,與 0.708

之詐騙者召回率；而連續過濾法則為 0.745, 0.757 與 0.762 之召回率，明顯優於前者。由此可見，面對極度不平衡之測試集，使用連續過濾模型仍可維持良好的總體準確率，並獲得較佳之詐騙者召回率。而五種單一模型的平均偵測準確率雖較高，但其本質上缺點如前所述：由於無法事先得知測試資料的成分比例，因此並不知該使用何種資料配比所建構之單一模型來偵測。

表 10：單一模型、連續過濾法、平衡式偵測法之效能比較

Test Set(F:NF=1:20, 1:30, 1:50)

Detection Model	Test Set (F:NF=1:20)							
	Accuracy	prec_NF	recall_NF	prec_F	recall_F	F-Measure(F)	F-Measure	MCC
Average(Single)*	**0.927**	0.977	0.945	0.543	0.674	**0.602**	**0.784**	0.495
Model(Successive)	0.926	0.981	0.938	0.459	**0.748**	0.569	0.777	**0.509**
Model(Balanced)	0.923	0.981	0.937	0.453	0.738	0.561	0.773	0.509
Detection Model	Test Set (F:NF=1:30)							
	Accuracy	prec_NF	recall_NF	prec_F	recall_F	F-Measure(F)	F-Measure	MCC
Average(Single)	**0.933**	0.985	0.945	0.46	0.679	**0.548**	**0.764**	0.423
Model(Successive)	0.931	0.988	0.939	0.369	**0.757**	0.496	0.754	**0.449**
Model(Balanced)	0.927	0.987	0.937	0.353	0.729	0.476	0.743	0.449
Detection Model	Test Set (F:NF=1:50)							
	Accuracy	prec_NF	recall_NF	prec_F	recall_F	F-Measure(F)	F-Measure	MCC
Average(Single)	**0.939**	0.992	0.945	0.353	0.708	**0.471**	**0.742**	0.336
Model(Successive)	0.934	0.993	0.939	0.252	**0.762**	0.379	0.719	**0.385**
Model(Balanced)	0.931	0.993	0.937	0.237	0.738	0.359	0.709	0.320

* Average(Single)為表中五種單一模型偵測結果之平均值，其塑模配比 F:NF 分別為 1:1, 1:2, 1:4, 1:6 與 1:8

連續過濾的優點除能獲得高偵測準確率外，更可在不同階段提供不同的偵測精度。表 11 為使用連續過濾之各模型偵測結果，其中透過 M1 過濾出的正常者共有 118 位(117+1)，其中 117 位確實為 NF，精度高達 99.15%。換言之，若有待測帳號在此階段被判斷為 NF，則將具有很高的可信度。同樣地，M2 對詐騙者(F類別)的偵測精度高達 98.63%，在此階段被判斷為詐騙帳號，同樣具有很高的可信度。偵測精度隨著過濾步驟增加漸漸下降，M3 仍有 91.94%，M4 則為 71.64%。最後，M5 之 NF 偵測精度僅剩 72.29%。縱使如此，流程之總體偵測準確率(Accuracy)仍高達 87.1%。綜合上述結果，顯示本研究提出方法與單一模型具有重大差異。同樣具有 87.1%之單一模型，當有帳號被判定為 F 或 NF 時，並無法對此結果提供可信度資料。若以訓練集塑模結果來斷定，實無太大參考價值。但對於多模型連續過濾而言，卻可因待測帳號在哪一階段被偵測出來，概略了解其偵測結果之可信度(越早被偵測出來之帳號，其可信度可能越高)。

表 11：連續過濾偵測流程之各階段偵測精度

分類/模型	M1	M2	M3	M4	M5
NF(true)	117	0	114	0	79
F(true)	0	72	0	48	4
NF(false)	1	0	10	0	31
F(false)	0	1	0	19	1
prec_NF	99.15%		91.94%		72.82%
prec_F		98.63%		71.64%	80.0%

三、多模型偵測方法之個別模型資料配比實驗結果

本節將探討多模型偵測方法中個別模型資料配比之不同，對於偵測結果之影響。參考表 12，其中針對連續過濾法五個模型(M1~M5)之不同資料配比進行比較，測試資料之 F:NF 比值為 1:8。由結果可看出，與原始設定相較(No.1)，No.2, No.3 可獲得略好的結果(+0.5%)。但若將 M5 由 F:NF=1:2 改為 1:1，則偵測準確率將大幅下降(0.868)。此結果顯示，M5 的資料配比若與測試集差異太大，將嚴重影響偵測準確率。有關測試資料集(F:NF=1:4)的結果則顯示於表 13 中，其結果與表 12 類似。換言之，以本研究提出架構為基礎，若僅小幅改變訓練資料配比，對偵測效能之影響有限。

表 12：連續過濾法中各模型使用不同資料配比之偵測效能比較

Test Size(F:NF)=1:8

No	Model Setting	Accuracy	prec_NF	recall_NF	prec_F	recall_F	MCC
1	(1,5)(5,1)(1,3)(3,1)(1,2)	0.915	0.962	0.938	0.677	0.776	0.610
2	(1,8),(8,1),(1,4),(4,1),(1,2)	0.920	0.961	0.945	0.700	0.772	0.660
3	(1,6)(6,1)(1,4)(4,1)(1,2)	0.920	0.961	0.945	0.700	0.772	0.660
4	(1,6)(6,1)(1,3)(3,1)(1,1)	0.868	0.972	0.872	0.526	0.85	0.559

表 13：連續過濾法中各模型使用不同資料配比之偵測效能比較

Test Size(F:NF)=1:4

No	Model Setting	Accuracy	prec_NF	recall_NF	prec_F	recall_F	MCC
1	(1,5)(5,1)(1,3)(3,1)(1,2)	0.901	0.933	0.938	0.791	0.776	0.644
2	(1,8),(8,1),(1,4),(4,1),(1,2)	0.903	0.931	0.945	0.806	0.767	0.657
3	(1,6)(6,1)(1,4)(4,1)(1,2)	0.902	0.931	0.943	0.803	0.767	0.651
4	(1,6)(6,1)(1,3)(3,1)(1,1)	0.868	0.953	0.872	0.668	0.857	0.563

在上述結果中，各種連續過濾模型(No.1~No.3)總體準確率雖可超過 90%，也具有很高之正常者精度(Prec_NF)與召回率(prec_F)，但對詐騙者之召回率

(recall_F)均未達 80%(約 76%~77%)。當以「找出詐騙者」為主要考量時，實有改善空間。觀察表 12, 13 中的結果，推測可能原因如下：當待測帳號透過連續過濾法之 M1~M4 模型均無法判別其類型時，便使用 M5 模型做最後判定(塑模配比為 F:NF=1:2)。此配比乃假設待測資料中詐騙者可能少於正常者，且為避免過於偏頗任一類型，因此將其比例設定為 1:2。如此設計有利總體準確率的提升，但可能導致「正常者」較受重視。因此，若將最後一個模型(M5)之塑模配比改為 F:NF=1:1，應可改善此一現象。

表 14 比較連續過濾法之最後一個模型 M5 分別使用 F:NF=1:2 與 1:1 塑模，所產生之效能差異。由結果可看出，對於各種測試集，M5(F:NF=1:1)均可獲得約 85%的詐騙者召回率，較之前有明顯改善(參考表中之 M_{suc}(M5(1:1))列)。然而，提升詐騙者召回率亦可能產生以下的缺點：(1) 與 M_{suc}(M5(1:2))相較，其詐騙者之偵測精度(prec_F)明顯降低(請參考 prec_F 欄)，也就是偽陽性比例增多，(2) 除第一種測試集(F:NF=1:1)外，M_{suc}(M5(1:1))之總體準確率(Accuracy)亦均下降；當測試資料 F:NF=1:8 時，其準確率僅有 0.870，而原連續過濾模型則有 0.915。由上述結果可知，透過調整塑模資料配比確可增加詐騙者的召回率，避免產生漏網之魚，提升系統的實用性。然而，由於偽陽性比例增多，亦可能讓部分正常交易者受到無謂干擾。有鑑於上述特性，在實際使用時，可根據自訂目標選取不同的偵測模型，以有效找出詐騙者，保障正常交易者的權益。

表 14：改變連續過濾法最後一個模型(M5)資料配比對於各種測試集之效能影響

Test Set	Successive Model	Model Setting	Accuracy	prec_NF	recall_NF	prec_F	recall_F
F:NF**=1:1	M_{suc}(M5(1:2))	(1:5)(5:1)(1:3)(3:1)(1:2)	0.856	0.807	0.937	0.925	0.774
	M_{suc}(M5(1:1))*	(1:5)(5:1)(1:3)(3:1)(1:1)	**0.863**	0.855	0.877	0.874	**0.850**
F:NF=1:2	M_{suc}(M5(1:2))	(1:5)(5:1)(1:3)(3:1)(1:2)	**0.871**	0.880	0.935	0.851	0.744
	M_{suc}(M5(1:1))	(1:5)(5:1)(1:3)(3:1)(1:1)	0.867	0.922	0.874	0.775	**0.852**
F:NF=1:4	M_{suc}(M5(1:2))	(1:5)(5:1)(1:3)(3:1)(1:2)	**0.901**	0.933	0.938	0.791	0.776
	M_{suc}(M5(1:1))	(1:5)(5:1)(1:3)(3:1)(1:1)	0.867	0.952	0.871	0.666	**0.853**
F:NF=1:8	M_{suc}(M5(1:2))	(1:5)(5:1)(1:3)(3:1)(1:2)	**0.915**	0.962	0.938	0.677	0.776
	M_{suc}(M5(1:1))	(1:5)(5:1)(1:3)(3:1)(1:1)	0.870	0.973	0.873	0.528	**0.852**

* M_{suc}(M5(1:1)): 將原連續過濾法最後一個模型(M5)之塑模資料配比由 F:NF=1:2 改為 1:1

** F:NF: 測試集中詐騙者(Fraud)與正常者(Non-Fraud)之數量配比

伍、結論與未來工作

近年來，電子商務已成為現代人生活得一部分，讓電子商務的交易金額年年攀升。發展至今，全球電子商務銷售額預計在 2021 年將達到 4.8 兆美元。隨著線上消費行為的普及，全球將真正進入數位經濟時代。面對如此龐大的交易金額，

也引起不肖人士的覬覦，在電子商務平台中進行詐騙，其中又以線上拍賣詐騙為大宗。若不加以抑制，將不利於電子商務的長遠發展。有關線上拍賣詐騙偵測，已有許多方法已被提出，但對於日新月異的詐騙手法，其準確率仍有待提升。

為解決此問題，本研究提出一套過濾式模型融合方法，以提升線上拍賣詐騙偵測之準確性。首先，透過分析單一模型應在不同測試集之效能，我們發現當詐騙者與正常者比例未知時，單一模型的效能受到限制。其次，本研究靈活運用不同類型配比之訓練資料，產生有利於詐騙者與正常者之偵測模型。最後，針對這些不同特質之模型，分別以多階連續過濾及平衡過濾方式加以整合，以改善單一模型的弱點。為驗證提出方法之有效性，我們採用實際拍賣交易資料進行實驗。與各種單一偵測模型相較，本研究提出之連續過濾與平衡過濾法確能提升準確率，並提供更穩定的偵測結果。當使用連續過濾時，除可獲得較高準確率外，也能對各階段之偵測精度進行分析，提升結果之實用性。此外，雖然模型融合時嘗試建各種不同特質的單一模型可影響準確性，但我們發現在多階段過濾的流程下，對於偵測效能之影響有限。由上述結果可知，本研究提出方法確有助於改善詐騙偵測準確率，提供消費者更周全的交易防護。

雖然本研究提出之方法能協助避免詐騙情事的發生，但在使用時仍有以下限制：

(1) 本研究假設網站會員的交易資料均可自由下載，若交易網站日後對此進行限制，則可能因交易歷史不足，無法準確預測。
(2) 詐騙者除自行創建帳號外，亦可能以竊取他人帳戶或購買高評價帳號做為掩護。對此類型之詐騙者，由於無法在其交易歷史中找出相關特質，可能會發生誤判狀況。
(3) 本研究提出之方法僅針對詐騙者與正常者進行分類，並無對其嚴重程度與類型進行分析。
(4) 交易雙方在交易後可互留文字評價訊息，其中可能會透漏許多重要訊息，本研究發展的方法並無對此進行分析。

本研究可能之未來工作如下：雖然使用 52 種偵測屬性，但對於稀有詐騙類型仍有其瓶頸，若能發展新的詐騙屬性(或針對特定詐騙類型設計特定偵測屬性)，將有助於提升整體詐騙偵測準確率。有關此主題，亦可考慮使用 Convolutional Neural Network 等方法來進行屬性轉換與縮減。其次，本研究均採用傳統如分類樹之學習模型，未來可嘗試使用深度學習慣用之 RNN 或 LSTM，驗證是否能進一步提升偵測準確性。最後，若能將發展之方法應用拍賣網站之實際交易流程中，將有助於管理當局早期發現詐騙情事，儘早進行監控，以維護電子商務之健全發展。

參考文獻

鄭孝儒 (2010)，「線上拍賣潛伏期詐騙者之有效偵測」，碩士論文，淡江大學資訊管理研究所，新北市。

Alford M. (2013). Intelligent fraud detection: a comparison of neural and Bayesian methods. *Computer Fraud & Security*, 14-16.

Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems*, 55, 278-288.

Amrehn, M., Mualla, F., Angelopoulou, E., Steidl, S., & Maier, A. (2018, December 19). *The Random Forest Classifier in WEKA: Discussion and New Developments for Imbalanced Data*. https://deepai.org/publication/the-random-forest-classifier-in-weka-discussion-and-new-developments-for-imbalanced-data.

Chang, J. S. & Chang, W. H. (2009) An early fraud detection mechanism for online auctions based on phased modeling. *Proceedings of the 2009 International Workshop on Mobile Systems E-Commerce and Agent Technology*, Taipei, Taiwan, December 3–5.

Chang, W. H. & Chang, J. S. (2011). A novel two-stage phased modeling framework for early fraud detection in online auctions. *Expert Systems with Applications*, 38, 11244–11260.

Chang, W. H. & Chang, J. S. (2012). An effective early fraud detection method for online auctions. *Electronic Commerce Research and Applications*, 11(4), 346-360

Chang, J. S., Liu, Y. H., & Lee, C. F. (2020). Developing Effective Fraud Detection Methods for Online Auction. *TANET 2020 臺灣網際網路研討會*, 台灣大學。

Chau, D. H. & Faloutsos, C. (2005). Fraud detection in electronic auction. *Proceedings of European Web Mining Forum at ECML/PKDD 2005*.

Chau, D. H., Pandit, S., & Faloutsos, C. (2006). Detecting fraudulent personalities in networks of online auctioneers. *Proceedings of PKDD 2006*, 103-144.

Chen, C., Zhu, Q., Lin, L., & Shyu M. L. (2013). Web Media Semantic Concept Retrieval via Tag Removal and Model Fusion. *ACM Transactions on Intelligent Systems and Technology*, 4(4), 1-22.

Chen, J., Tao, Y., Wang, H., & Chen, T. (2015). Big Data based fraud risk management at Alibaba. *The Journal of Finance and Data Science*, 1(1), 1-10.

Darudi, A., Bashari, M., & Javidi, H. (2015). Electricity price forecasting using a new data fusion algorithm. *IET Generation, Transmission & Distribution*, 2015, 9, 1382-1390.

eMarketer (2020). Retail & Ecommerce report. Retrieved on Mar. 1, 2020, https://www.emarketer.com/topics/topic/retail-ecommerce.

Gavish, B. & Tucci, C. (2008). Reducing Internet Auction Fraud. *Communications of the ACM*, 51(5), 89-97.

Goes, P. B., Tu, Y., & Tung, A. (2009). Online Auctions Hidden Metrics. *Communications of the ACM*, 52(4), 147-149.

Huang, S., Ma, J., Cheng, P., & Wang, S. (2015). A Hybrid Multigroup Co-clustering Recommendation Framework Based on Information Fusion. *ACM Transactions on Intelligent Systems and Technology*, 6(2), 1-22.

Kim, K., Choi, Y., & Park, J. (2013), Price fraud detection in online shopping malls using a finite mixture model. *Electronic Commerce Research and Applications*, 12, 195-207.

Kumar, M. S., Soundarya, V., Kavitha, S., Keerthika, E.S., & Aswini, E. (2019). Credit Card Fraud Detection Using Random Forest Algorithm. *Proceedings of IEEE 3rd International Conference on Computing and Communication Technologies (ICCCT)*, 149-155.

Li, S. H., Yen, D. C., Lu, W. H., & Wang, C. (2012). Identifying the signs of fraudulent accounts using data mining techniques. *Computers in Human Behavior*, 28, 1002–1013.

Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M.-S., & Makki H. Z. (2019). An Experimental Study with Imbalanced Classification Approaches for Credit Card Fraud Detection. *IEEE Access*, 7, 93010-93022.

National White Collar Crime Center (NW3C) (2019, March 1). 2019 Internet Crime Report. https://pdf.ic3.gov/2019_IC3Report.pdf

Pandit, S., Chau, D.H., Wang, S., & Faloutsos, C. (2007). Netprobe: a fast and scalable system for fraud detection in online auction networks. *Proceedings of the 16th international conference on World Wide Web*, 201-210.

Tsang, S., Koh, Y.-S., Dobbie, G., & Alam, S. (2014). SPAN: Finding collaborative frauds in online auctions. *Knowledge-based systems*, 71, 389-408.

West J. & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computer & Security*, 57, 47-66.

Xuan, S., Liu, G., Li, Z., Zheng, L., & Wang S., (2018). Random Forest for Credit Card Fraud Detection. *Proceedings of IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 27-29.

附錄: 各種單一分類模型之效能比較

本附錄比較 Random Forest, Ada Boost, Decision Tree, Multi-Layer Perception (MLP)與 Nearest Neighbor 等單一分類模型，在不同詐騙者與正常者配比下(F:NF)之偵測效能，實驗設定與 3.2 節中之表 4 相同。實驗結果如表 15~19 所示，其中以 Random Forest 表現最佳，可獲得最高之平均準確率(Accuracy)。

表 15：以 Random Forest 塑模，在不同 NF:F 配比下塑模之偵測結果

偵測模型 Random Forest	Test Set(F:NF=1:2)				
	Accuracy	prec_NF	recall_NF	prec_F	recall_F
F:NF =5:1	0.768	0.977	0.669	0.594	0.969
F:NF =3:1	0.806	0.959	0.743	0.644	0.935
F:NF =2:1	0.829	0.943	0.793	0.686	0.903
F:NF =1:1	0.862	0.910	0.881	0.775	0.825
F:NF =1:2	0.869	0.877	0.934	0.848	0.739
F:NF =1:3	0.867	0.862	0.954	0.882	0.693
F:NF =1:5	0.849	0.834	0.966	0.900	0.614
Average	0.836	0.909	0.849	0.761	0.811

表 16：以 Ada Boost 塑模，在不同 NF:F 配比下塑模之偵測結果

偵測模型 Ada Boost	Test Set(F:NF=1:2)				
	Accuracy	prec_NF	recall_NF	prec_F	recall_F
F:NF =5:1	0.712	0.961	0.591	0.538	0.952
F:NF =3:1	0.768	0.953	0.686	0.598	0.931
F:NF =2:1	0.808	0.941	0.759	0.653	0.903
F:NF =1:1	0.844	0.917	0.842	0.730	0.848
F:NF =1:2	0.871	0.896	0.912	0.818	0.791
F:NF =1:3	0.874	0.880	0.938	0.858	0.746
F:NF =1:5	0.870	0.862	0.959	0.895	0.695
Average	0.821	0.916	0.813	0.727	0.838

表 17： 以 Decision Tree 塑模，在不同 NF:F 配比下塑模之偵測結果

偵測模型 Decision Tree	Test Set(F:NF=1:2)				
	Accuracy	prec_NF	recall_NF	prec_F	recall_F
F:NF =5:1	0.727	0.941	0.630	0.555	0.920
F:NF =3:1	0.750	0.928	0.677	0.581	0.895
F:NF =2:1	0.772	0.913	0.728	0.613	0.861
F:NF =1:1	0.800	0.894	0.794	0.664	0.813
F:NF =1:2	0.827	0.874	0.865	0.735	0.751
F:NF =1:3	0.820	0.860	0.872	0.736	0.717
F:NF =1:5	0.816	0.838	0.897	0.761	0.656
Average	0.787	0.893	0.780	0.664	0.802

表 18：以多層感知機(MLP)塑模，在不同 NF:F 配比下塑模之偵測結果

偵測模型 MLP	Test Set(F:NF=1:2)				
	Accuracy	prec_NF	recall_NF	prec_F	recall_F
F:NF =5:1	0.695	0.923	0.594	0.528	0.896
F:NF =3:1	0.685	0.903	0.591	0.518	0.869
F:NF =2:1	0.749	0.893	0.709	0.589	0.827
F:NF =1:1	0.797	0.887	0.797	0.666	0.797
F:NF =1:2	0.819	0.871	0.856	0.724	0.746
F:NF =1:3	0.805	0.838	0.878	0.733	0.662
F:NF =1:5	0.800	0.824	0.893	0.752	0.617
Average	0.765	0.877	0.760	0.644	0.773

表 19：以 Nearest Neighbor 方法塑模，在不同 NF:F 配比下塑模之偵測結果

偵測模型 Nearest Neighbor	Test Set(F:NF=1:2)				
	Accuracy	prec_NF	recall_NF	prec_F	recall_F
F:NF =5:1	0.548	0.936	0.347	0.423	0.951
F:NF =3:1	0.660	0.919	0.539	0.496	0.904
F:NF =2:1	0.727	0.901	0.663	0.560	0.855
F:NF =1:1	0.797	0.873	0.814	0.673	0.763
F:NF =1:2	0.825	0.844	0.905	0.778	0.667
F:NF =1:3	0.817	0.822	0.927	0.804	0.600
F:NF =1:5	0.812	0.803	0.953	0.850	0.533
Average	0.741	0.871	0.735	0.655	0.753