

世新大學管理學院資訊管理學系
碩士論文

Department of Information Management
Shih Hsin University
Master Thesis

利用深度學習預測信用卡詐騙

Predicting Credit Card Fraud by Deep Learning

研究生： 陳佩妤 (CHEN, PEI-YU)

指導教授：胡碩誠 博士 (HU, SHUO-CHENG, Ph.D.)

中華民國 一零七年七月

July, 2018

世新大學管理學院資訊管理學系碩士班

論文題目：利用深度學習預測信用卡詐騙

研究生：陳佩妤

本文承蒙下列口試委員審議通過。

口試委員

許智輝

李家泰 李香良

指導教授

郭寶誠

中華民國 107 年 7 月 23 日

致謝

在碩士的這兩年當中，有歡樂有淚水。一開始相當徬徨，不知道該進哪個實驗室，後來加入了網路實驗室。感謝我的指導教授胡碩誠教授，願意接受程式基礎不好的我，在這兩年當中，老師給予相當多的幫助，不管是在論文研究上給予許多建言，也給予我擔任助教的機會，可以學習其他課程知識。平時可以跟老師一起聊天，一同吃喝玩樂，就像是朋友一樣，跟大家都很親近。實驗室空間雖然有點狹小，但是應有盡有，在實驗室與大家一起玩樂，可以射飛鏢、煮火鍋、投影看跨年演唱會。首先要感謝承樺，推薦我加入網路實驗室，也受到許多許多照顧。再來要感謝思頻，在大四及碩一的陪伴，一起修課一起吃晚餐。謝謝實驗室夥伴鴻偉、益維，一起度過在實驗室的日子。謝謝彥銓學長在有程式上的困難時的幫忙。最要感謝的是我的家人們，爸爸媽媽妹妹，總是無條件支持我任何決定，趕論文時期，時常晚歸也讓他們擔心了。最後要感謝自己，一路上跌跌撞撞，最後完成了論文研究。謝謝一路上幫助過我給予我鼓勵的大家。

陳佩妤 謹識

世新大學資訊管理研究所

中華民國 一零七年七月

摘要

隨著線上購物盛行，盜刷事件也隨之增加，根據統計，臺灣每年盜刷金額將近二十億。為防止盜刷問題持續發生，近年來有許多研究利用機器學習方法來偵測信用卡詐騙。這些方法遇到的挑戰之一，即是信用卡交易資料為不平衡資料集，詐欺與正常交易的比例相差懸殊。面對不平衡資料集，多數研究使用 Undersample，將正常交易筆數減少，使詐騙比例相對提高。此外機器學習需先進行資料前處理，挑選有利於訓練模型的特徵，但由於信用卡交易資料具有許多個人隱私資料，因此取得的資料集欄位可能已經過保密處理，對於機器學習建模有一定的困難度。然而建立深度學習模型前，不須挑選特徵，可使用所有資料欄位建模，對於處理如信用卡交易等具有個人隱私資料建模具有優勢。

本研究中，先將取得的信用卡交易資料進行 Undersample，調整詐騙交易與正常交易的比例為 50%, 15%, 10%, 5%, 1% 來進行比較。由於資料中的欄位名稱除了交易時間與交易金額外，都已經被處理過無法辨識，因此我們隨機挑選資料欄位作為特徵，使用 Logistic Regression、Random Forest、Support Vector Machine 三種機器學習建模。此外，我們用 Keras 建立深度學習模型，建立一個輸入層、三個隱藏層的模型，並進行 200 次訓練週期。在評估預測結果方面，考慮 Accuracy, Precision, Recall, F1-Score，以及 Matthews correlation coefficient 五項具代表性的指標。實驗的結果顯示，深度學習方法在五種不同比例的模型下所得到的預測結果都優於機器學習方法。

關鍵詞： 信用卡詐騙、不平衡資料集、機器學習、深度學習

Abstract

With the popularity of online shopping, cases of transaction fraud are also increasing. According to statistics, the amount of losses from credit card fraud is nearly 2 billion annually in Taiwan. In order to prevent fraud events, many studies have used machine learning methods to detect credit card fraud in recent years. One of the challenges of these methods is that the credit card transaction data is highly imbalanced, fraudulent transactions are largely outnumbered by genuine ones. A common strategy for dealing with the problem is to under-sample the legit transaction class in the training set. In addition, machine learning algorithms need to perform data pre-processing first, and select features that are beneficial to the training process. However, due to confidentiality issues, a publicly available credit card data set is usually encrypted or transformed which make it difficult to select proper attributes to train the classifier algorithms. However, there is no need to select features while establishing a deep learning model. The deep learning approach has advantages for processing confidential data such as credit card transactions.

In this study, the credit card transaction data obtained was first under-sampled and the ratio of fraud transactions to legit ones was adjusted to 50%, 15%, 10%, 5%, and 1% for comparison. Since the attribute names in the data set have been transformed except for transaction time and transaction amount. Therefore, we randomly selected the attributes to investigate the performance of Logistic Regression, Random Forest, and Support Vector

Machine. In addition, we used Keras to build a deep learning model which includes an input layer and three hidden layers, and conducted 200 times of training. Five metrics were used to report the performance of fraud detection classifiers including Accuracy, Precision, Recall, F1-Score, and Matthews correlation coefficient. The experimental results show that the deep learning method, in most circumstance, is outperform the machine learning methods.

Key words : Credit Card Fraud、Imbalanced Data、Machine Learning、Deep Learning

目錄

致謝	i
摘要	ii
Abstract.....	iii
目錄	v
圖目錄	vi
表目錄	vii
第一章 緒論	1
1.1 研究動機.....	1
1.2 研究分析.....	1
1.3 論文架構.....	3
第二章 文獻探討	4
2.1 處理不平衡資料集之方法	4
2.2 偵測信用卡詐騙相關研究	5
2.3 深度學習簡介	6
第三章 研究方法與實驗	7
3.1 Undersample.....	7
3.2 三種機器學習方法	8
3.3 深度學習 – Keras.....	9
第四章 實驗結果	10
4.1 評估指標.....	10
4.2 預測結果.....	11
第五章 結論	18
參考文獻	19

圖目錄

圖 1 Keras 多層感知器模型.....	9
圖 2 機器學習的 Accuracy.....	12
圖 3 機器學習的 Precision	12
圖 4 機器學習的 F1-Score	13
圖 5 機器學習的 Recall.....	13
圖 6 機器學習的 MCC.....	14
圖 7 Keras 與 RF 的 Accuracy	14
圖 8 Keras 與 RF 的 Precision	15
圖 9 Keras 與 RF 的 Recall	15
圖 10 Keras 與 RF 的 MCC.....	16
圖 11 Keras 與 RF 的 F1-Score	16

表目錄

表 1 訓練資料集與測試資料集	7
表 2 Confusion Matrix	10

第一章 緒論

1.1 研究動機

根據 ”尼爾國際市場調查公司” 調查結果顯示，2014 年約有 36% 的臺灣民眾使用信用卡取代現金交易且平均每個人有 2.16 張信用卡[3]。盜刷事件也時有所聞，一般信用卡盜刷事件可分為盜用身分、偽卡詐騙和多帳單詐騙等，隨著線上購物交易盛行，盜刷事件也更為猖獗，2017 年在臺灣盜刷金額即高達 18.8 億[5]。為防止盜刷問題，金融監督管理委員會建立防範信用卡網路交易盜刷機制，包含發卡機構傳送消費簡訊或發現異常交易行為以電話通知持卡人[4]。目前也有許多使用機器學習的實例，像是攔截垃圾郵件、語音辨識等，知名訂房網站 Airbnb 也利用機器學習判斷信用卡盜刷事件[6]。

1.2 研究分析

近年來也有相關研究提出機器學習方法來偵測詐騙，而經常被研究者提出來使用的方法有 Logistic Regression(LR)[7][9][12] 、Naïve Bayes(NB) [7][12][13] 、Random Forest(RF) [9][12] 、K Nearest Neighbor(KNN)[7][13] 、Decision Tree(DT)[13] 、Support Vector Machine(SVM)[9][12] [13] 等多種方法。

在機器學習模型建立前，需進行資料前處理，挑選出重要的欄位後，

再訓練模型並預測結果。要如何選出具有影響力的特徵欄位，首先需知道資料集主題所要預測的結果為何，進而依照貼近資料集主題的專業知識挑選出對於預測結果有幫助的資料欄位。但由於信用卡交易資料中包含許多重要的個人資料，例如持卡人姓名、信用卡號、安全碼、交易日期、金額、信用卡類別等，資料欄位名稱都被隱藏且資料數值都已經過處理，所以在不知道欄位名稱的情況下，無法挑選出對於機器學習建模較有影響的資料欄位。

在信用卡交易資料中，相較於正常交易的資料筆數，詐騙交易資料筆數極少，兩者資料量相差甚大，我們則稱之為「不平衡資料集」。在資料量不平衡的情況下，直接建模及訓練，會影響預測結果。因此在建模之前，須將不平衡資料集進行處理，有以下三種常見的方法 :Oversample[11] 、 Undersample [8][9][11][13] 及 SMOTE(Synthetic Minority Over-Sampling Technique)[11]，讓正常交易資料與詐騙交易資料取得較為平衡的比例。

本研究實驗中將機器學習與深度學習進行比較。機器學習使用 LR、RF、SVM 三種方法，並使用 Undersample 將詐騙資料與正常交易取得不同比例，再隨機挑選資料欄位進行建模。而深度學習使用 Keras 建立多層感知器模型，輸入要訓練的資料、隱藏層數量及所要的輸出結果，完成建模及訓練。

1.3 論文架構

本篇論文共有五章。第二章將介紹處理不平衡資料集方法及偵測信用卡詐騙相關研究；第三章詳述本實驗不平衡資料集處理，LR、RF 和 SVM 三種機器學習方法以及深度學習 Keras；第四章為實驗結果及分類效果評估，利用以下五種評估指標，Accuracy、Precision、Recall、F1-Score、Matthews correlation coefficient (MCC)評估預測結果；最後第五章為此研究的結論。

第二章 文獻探討

2.1 處理不平衡資料集之方法

在整份資料集中，不同類別的資料差距甚大，稱之為「不平衡資料集」，大部分的處理方法是將多數類別減少，少數類別增多[9]，如何處理不平衡資料集和挑選特徵及選擇機器學習方法對於預測結果都影響極大[7]。

為改善不平衡資料集，有以下三種方法，第一種 Oversample 是將數量較少的類別資料增加，從數量少的類別資料中隨機選擇後複製並新增多筆資料，但由於此方法只是將少數類別資料直接複製，並不是增加新的資料，所以對於模型訓練幫助並不大。第二種 Undersample 則是將數量多的類別資料減少，是從數量多的類別資料隨機刪除，研究顯示遇到不平衡資料集時，刪除數量多的類別資料對於整體資料平衡有幫助且不會影響到數量少的類別資料，因此大部分研究都使用此方法。最後一種是 SMOTE，是屬於 Oversample 的一種，利用數量少的類別資料進行合成，從數量少的資料中找出幾個鄰近點，生成新的資料，以達到資料平衡[11]。

2.2 偵測信用卡詐騙相關研究

發生信用卡詐騙，通常都是非法獲取信用卡資料或是使用者資料外洩，才使得信用卡遭到盜刷[13]。信用卡詐騙一直存在的原因有兩個，第一是詐騙交易的型態變化快速，第二是詐騙交易資料在整體交易紀錄中佔有少數[7]，因此為了防範信用卡詐騙，偵測詐騙系統的技術也必須隨之進步。

偵測信用卡詐騙最重要的就是分析持卡人消費行為[7][10]，像是在某個地區、每筆資料交易日期時間、消費類別、持卡人基本資料等[9][13]，篩選出對於結果有意義的欄位。

在[9]的研究中，根據已知的信用卡交易資料統整後，結合多個欄位後創建出新的資料欄位，再利用這些欄位將曾有詐騙交易的信用卡紀錄以及所有交易中有發生過詐騙的交易紀錄，隨機選出正常與詐騙交易資料進行訓練和測試。但也有許多研究取得的信用卡交易資料，資料欄位是被隱藏的，而在不知道欄位名稱的情況下，很難取出對於建立模型有用的欄位，在[13]的研究中則是找出有四個重複數值的欄位，將原本 20 個未知欄位留下 16 個，而在[10]是使用資訊增益的方法，從 41 個未知的欄位中挑選出前 24 個。

而在[13]的研究中，將正常交易與詐騙交易比例調整為 50:50；在[10]的研究中比例為 75:25；較為特別的是在[8][9]的研究中，調整成多種不同的比例建立模型及訓練，再使用同一筆測試資料去做預測。遇到不平衡資料集時，透過 Undersample 處理過後，訓練模型對於準確度有一定的提升。

機器學習的方法有相當多種，如 DT、RF、KNN、LR、NB、SVM 等方法，僅需簡單的設定參數，就可完成機器學習的建模。在[12]的研究中，將兩個以上機器學習方法結合，使用 Majority Voting 選出預測結果較好的模型。

實驗的預測結果需要依靠評估指標，判斷預測結果的好壞。利用 Confusion Matrix 為基礎，列出以下幾種評估指標 Accuracy、Precision、F1-Score、False Positive Rate(FPR)、Balanced Classification Rate(BCR)、Sensitivity (Recall / True Positive Rate(TPR)) 、Matthews Correlation Coefficient(MCC) [7][8][9][10][12][13]。

2.3 深度學習簡介

機器學習與深度學習最大的不同，就是深度學習能處理高維度特徵資料，並使用 GPU 圖形處理器，處理及運算大量資料，相較於 CPU，GPU 能加快運算時間。深度學習會有一個輸入層、一個輸出層加上自行設定多個隱藏層。在深度學習模型的建立步驟上，需先將資料預處理，再開始建模，訓練模型，最後進行預測。Keras 是一個以 TensorFlow 運行的高階程式，為模型式的深度學習，能夠簡單的建模、訓練及預測，為較容易入手的深度學習模型[2]。

第三章 研究方法與實驗

本研究總資料筆數為 284807 筆，正常交易為 284315 筆，詐騙交易為 492 筆，詐騙比例為 0.17%，總資料欄位有 31 個，包含 28 個未知欄位(V1-V28)、交易時間(Time)、金額(Amount)以及是否為詐騙的欄位(Class)，正常交易為 0、詐騙交易為 1。

3.1 Undersample

實驗會將所有資料切分為兩份，一份進行訓練，另一份進行測試。訓練資料集的詐騙比例分別為 50%、15%、10%、5%、1%，測試資料集詐騙比例 0.17%，如表 1，將資料比例設為多種，可以比較出不同比例建模後的測試結果。

表 1 訓練資料集與測試資料集

	Train Data					Test Data
DataSet	DataSet 1	DataSet 2	DataSet 3	DataSet 4	DataSet 5	Test Data
Fraud	246	246	246	246	246	246
Normal	246	1394	2214	4674	24354	144459
Total	492	1640	2460	4920	24600	144705
Fraud Rate	50%	15%	10%	5%	1%	0.17%

3.2 三種機器學習方法

LR 是最廣為人知用來偵測信用卡詐騙的方法[9][12]，此方法常用於二元或多類別分類，可採用一個或多個特徵，利用 Sigmoid Function 算出機率值，機率值介於 0 到 1 之間，利用得到的機率值去做分類，若機率值大於 0.5 則判斷為 1，小於 0.5 則判斷為 0 [7]。

SVM 是線性分類器，取最大化邊界，獲得最小誤差，將資料分類；若是遇到非線性分類問題，則使用 Kernel SVM，能將高維度特徵做線性分離，將資料做分類[1][9][12][13]。

RF 是基於 DT 的一個分類模型，DT 會利用特徵數值做層層的分類，而 RF 是由多棵 DT 組成，設定好要訓練 DT 的棵數，實務上，設定越多棵 DT，分類效果越好，但需視資料集大小需求設定適合的棵數，最後藉由 Majority Voting 的方法留下分類效果較好的 DT 形成一個 RF 模型[1][12]。

本研究使用以上三種機器學習方法建立模型，而實驗中取得的資料集欄位名稱都已隱藏，因此將隨機挑選 14 個及 19 個資料欄位進行訓練，了解預測後的分類效果是否良好。

3.3 深度學習 – Keras

本研究設定輸入層的神經元為 30 個(特徵)，隱藏層神經元設定為 500 個，並使用三個隱藏層，輸出層神經元為 1 個輸出值(即是否為詐騙)，本研究 Keras 多層感知器模型架構，如圖 1。設定 200 次訓練週期，並將每次訓練的誤差及準確度記錄下來。預測結果利用 Confusion Matrix 為基礎，計算出五種評估指標，評估分類效果是否良好，並與機器學習方法得出的結果進行比較。

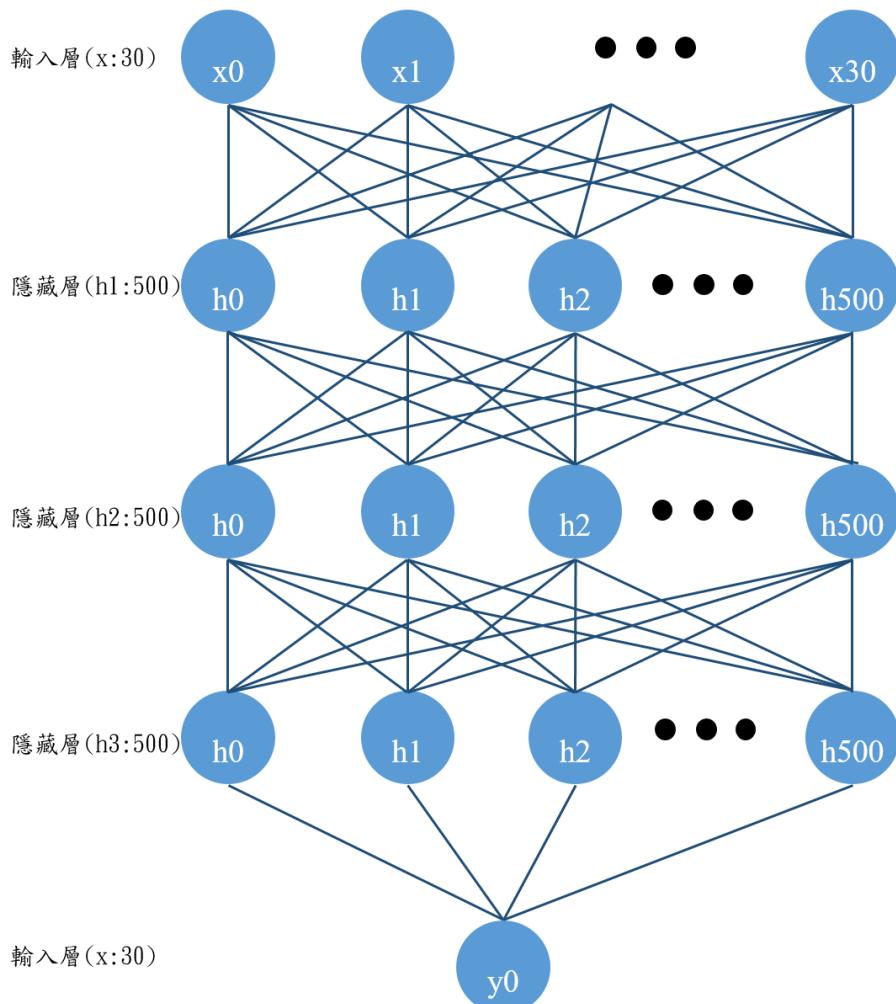


圖 1 Keras 多層感知器模型

第四章 實驗結果

4.1 評估指標

本研究利用五種評估指標評估模型的預測結果，以下表 2 為 Confusion Matrix，利用 Confusion Matrix 為基礎，列出以下五種評估指標公式 Accuracy 第(1)式 、 Precision 第(2)式 、 Recall 第(3)式、 F1-Score 第(4)式 、 MCC 第(5)式。

表 2 Confusion Matrix

	Prediction 0	Prediction 1
True 0	True Negetives(TN)	False Positives(FP)
True 1	False Negetives(FN)	True Positives(TP)

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (4)$$

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

4.2 預測結果

利用五種評估指標，比較機器學習與深度學習所建立的模型結果。使用三種機器學習以五個評估指標表現預測結果。機器學習的 Accuracy，如圖 2；機器學習的 Precision，如圖 3；機器學習的 Recall，如圖 4；機器學習的 F1-Score，如圖 5；機器學習的 MCC，如圖 6。深度學習也依照五個評估指標，查看預測結果，並與機器學習中的 RF 進行比較。Keras 與 RF 的 Accuracy，如圖 7；Keras 與 RF 的 Precision，如圖 8；Keras 與 RF 的 Recall，如圖 9；Keras 與 RF 的 F1-Score，如圖 10；Keras 與 RF 的 MCC，如圖 11。

利用評估指標得到以下結果，所有實驗中，Accuracy 都高於 95% 以上，是因為在每一種比例下，實際是正常交易與預測是正常交易的筆數佔大多數，因此 Accuracy 會隨著正常交易筆數所牽動，對於詐騙交易是否被正確偵測出來無法直接明確的得知。

Precision 則代表預測為詐騙資料中，有多少筆數是正確預測，由圖 3、圖 8 可知詐騙交易資料佔有比例越少，Precision 則提升，表示判斷為詐騙交易的正確率提高，錯誤率減少。

Recall 代表有多少詐騙交易實際被偵測出來，由圖 4、圖 9 得知當詐騙交易佔有比例減少，Recall 則降低，表示在正常交易與詐騙交易比例相差越大，詐騙交易越不易被預測出來。

F1-Score 是將 Precision 和 Recall 做調和平均的計算，讓實際是詐騙與預測為詐騙的資料平均得到一個較為均衡的數值。

MCC 可以得知整體資料中被預測為正常交易與詐騙交易的正確率。

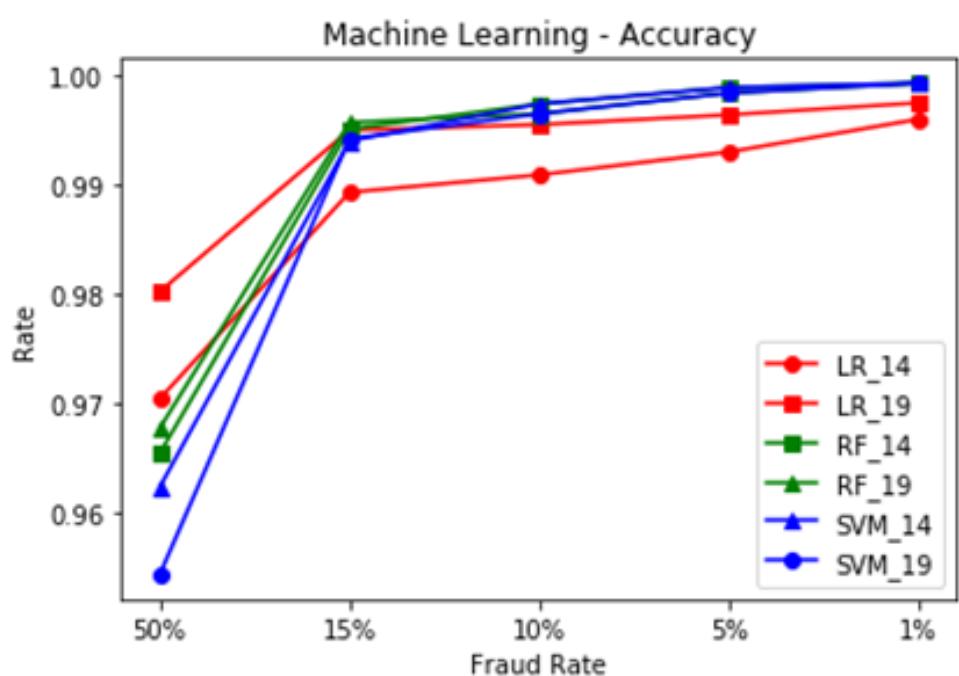


圖 2 機器學習的 Accuracy

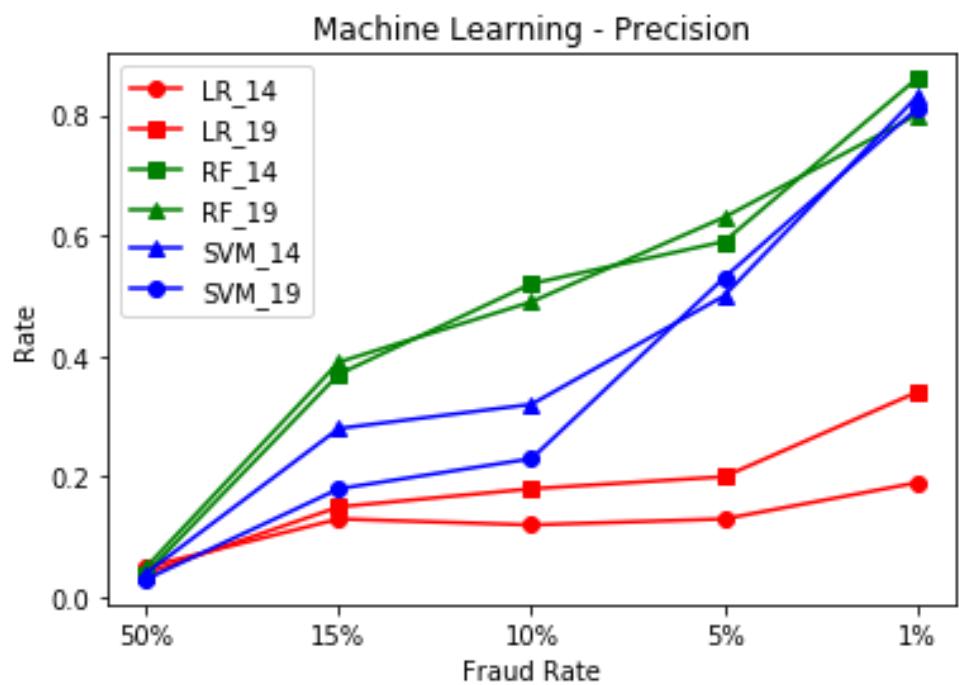


圖 3 機器學習的 Precision

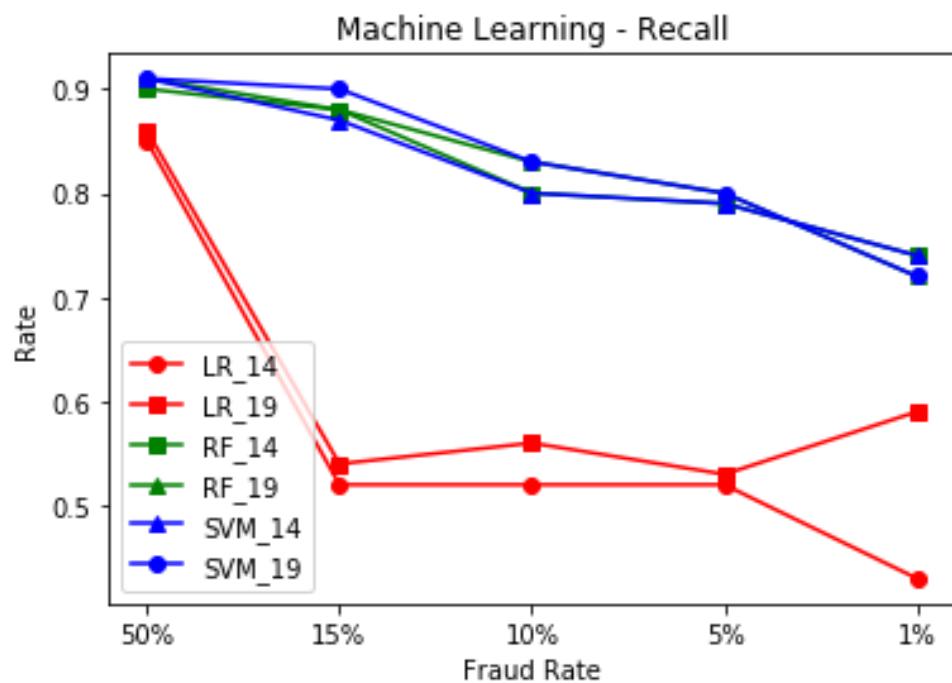


圖 4 機器學習的 Recall

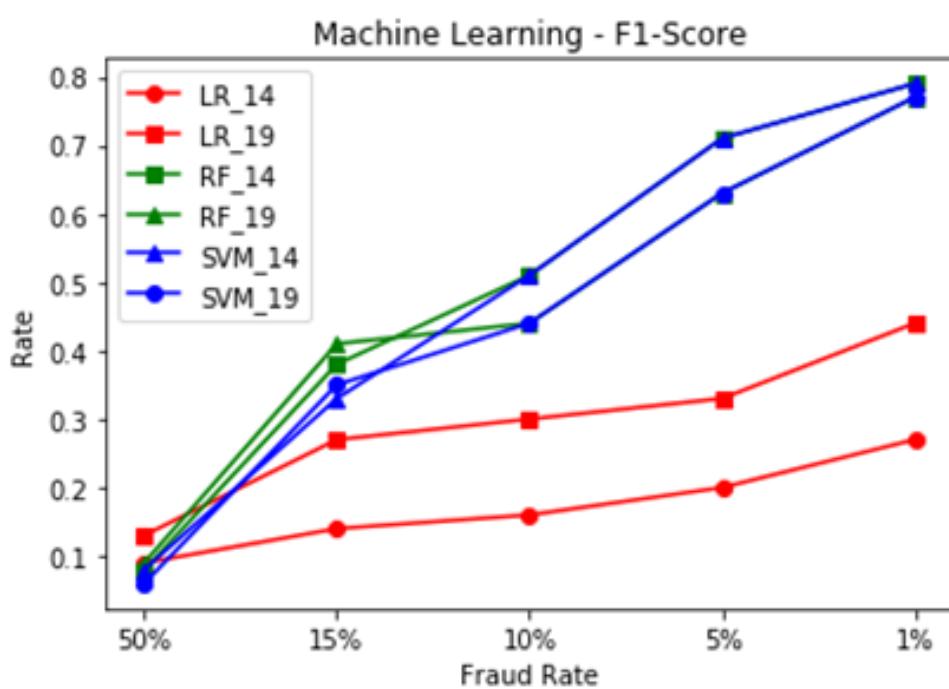


圖 5 機器學習的 F1-Score

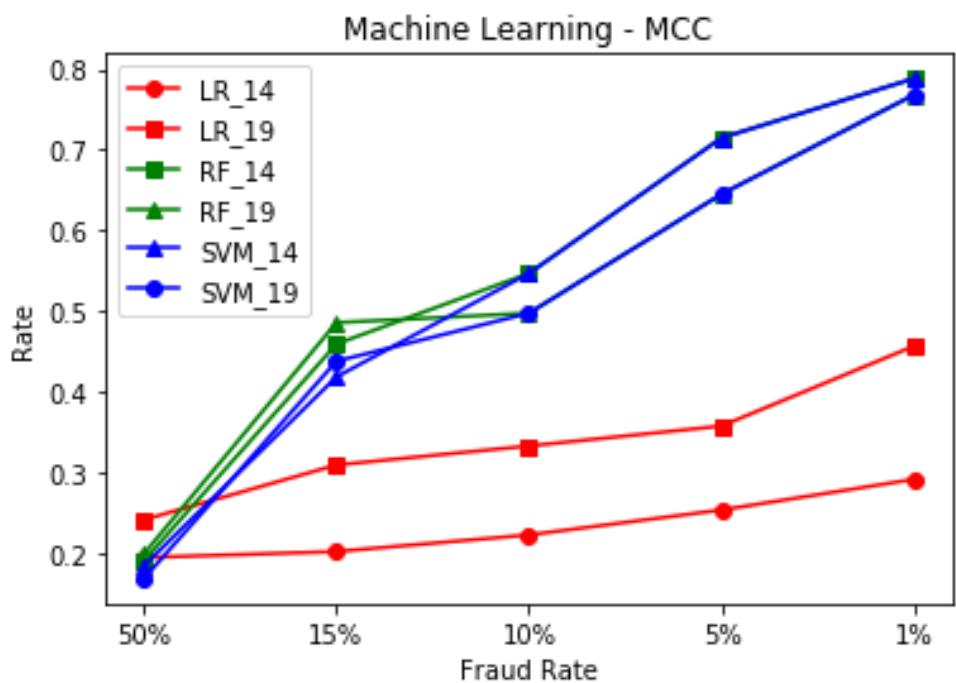


圖 6 機器學習的 MCC

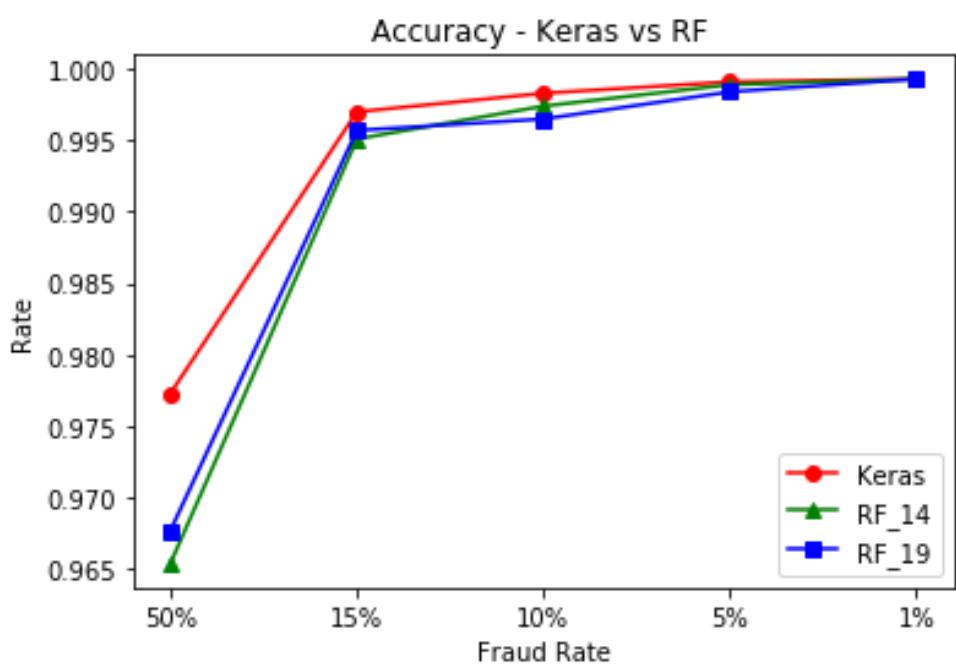


圖 7 Keras 與 RF 的 Accuracy

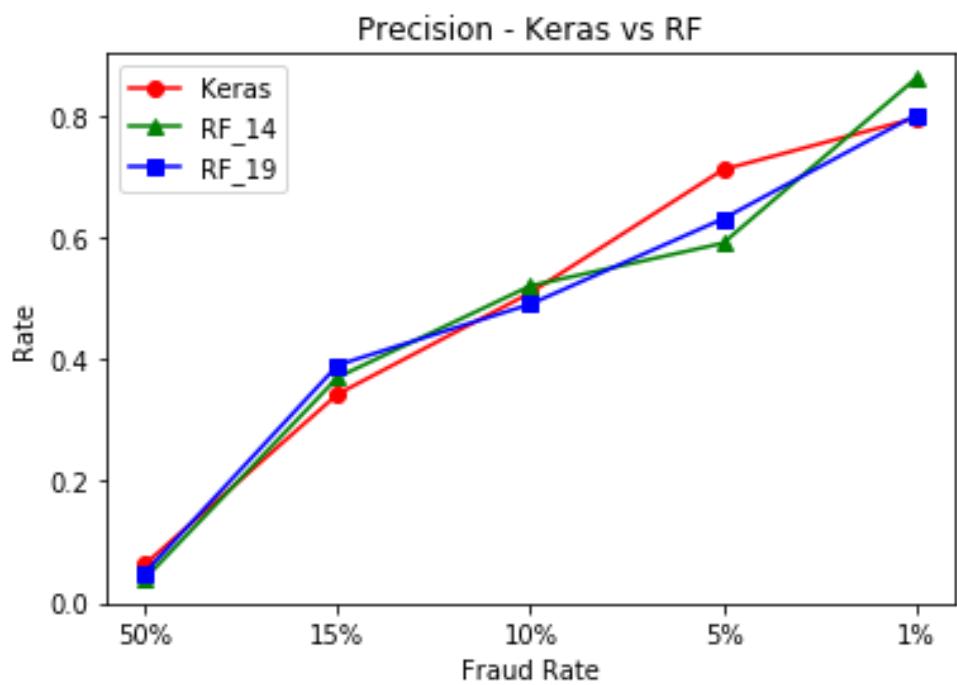


圖 8 Keras 與 RF 的 Precision

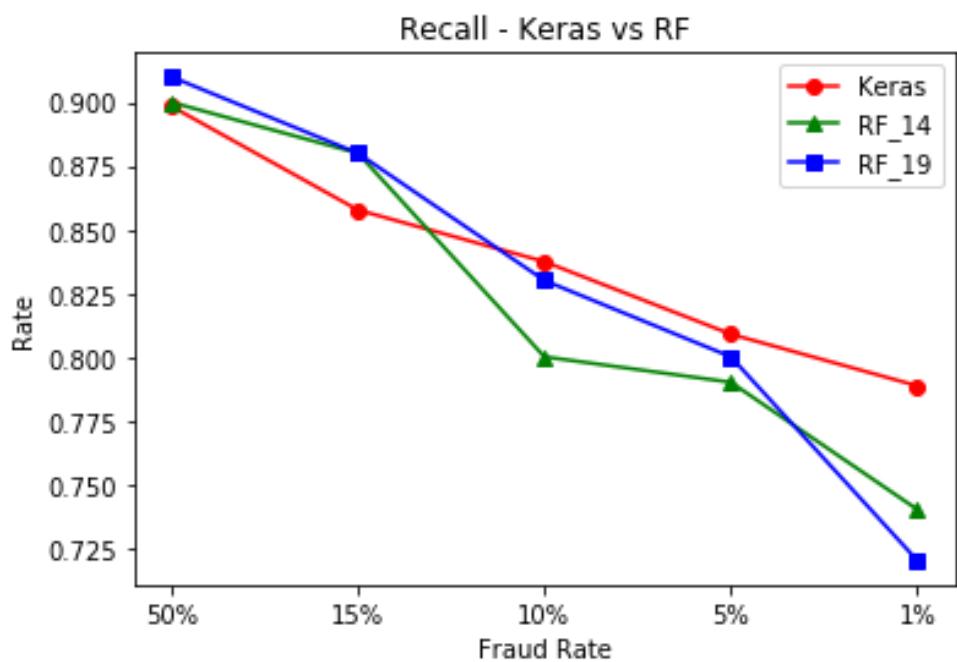


圖 9 Keras 與 RF 的 Recall

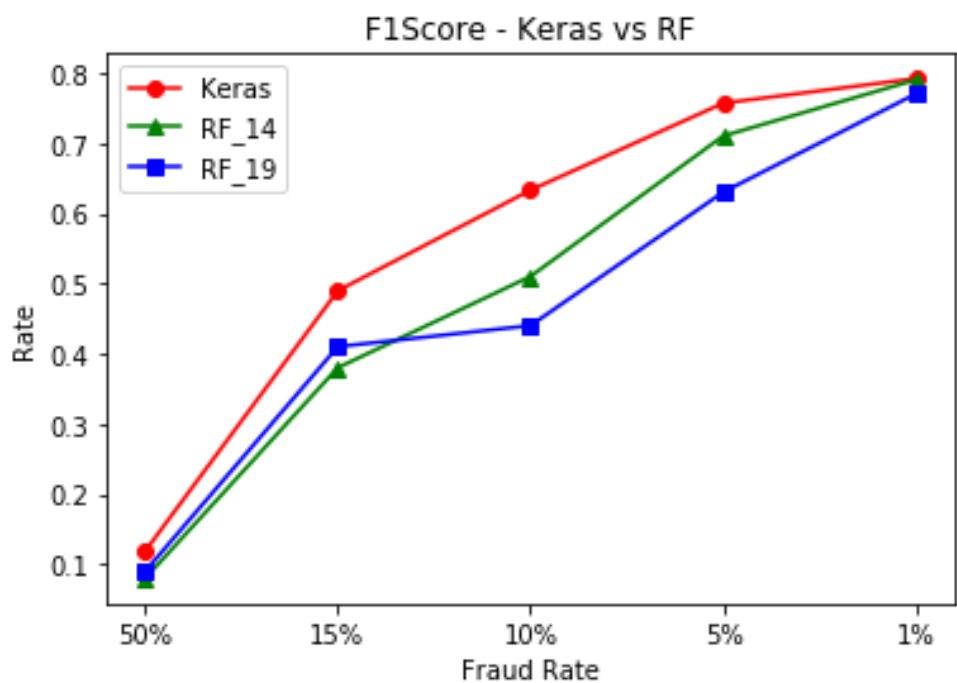


圖 10 Keras 與 RF 的 F1-Score

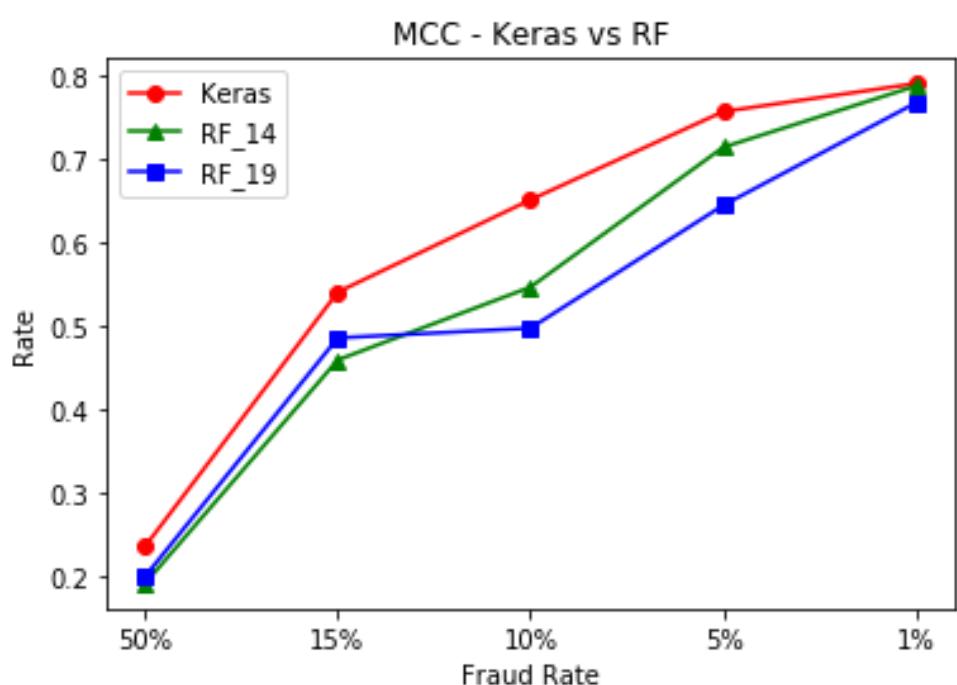


圖 11 Keras 與 RF 的 MCC

根據圖 2 到圖 5，三種機器學習方法的預測結果得知，利用 RF 方法建模，相較於另外兩種機器學習方法，每一種評估指標的表現都較為良好，因此將 RF 所建立的兩種模型，提出預測結果與深度學習 Keras 進行比較。如圖 7 到圖 11 所示，Keras 的整體表現大都優於 RF。

由此情況可知在不知道資料欄位的情況下，機器學習隨機挑選資料欄位建模，但每一次隨機選擇資料欄位都不同，得到的結果可能也會有些許的差異，雖然 RF 在機器學習當中表現較好，但卻不知其資料欄位，相對於深度學習，則是在訓練時自行挑選出重要的欄位，進行訓練及建模，得到較好的預測結果。

機器學習的 Recall 表現都較差，當詐騙比例逐漸降低，Recall 也隨之降低，原因為機器學習時常將大量正常交易誤判成詐騙交易。相較於機器學習，深度學習 Keras 的 Recall 大都超過 80%，代表能正確判斷詐騙交易，也能夠正確判斷正常交易。而在 Precision 的部分，不管是機器學習或深度學習的預測結果都表現不錯，能在詐騙交易比例減少的情況下，依然能詐騙交易正確判斷出來，且準確度持續提升。

第五章 結論

本論文討論當考慮保護個人隱私的因素，以至於取得的資料包含無法辨識的資料欄位與內容。在此情況之下，無法挑選出適合且重要的特徵，若要透過機器學習的方法進行建模及訓練，將會有一定的困難度，且建立好的模型不見得預測效果良好，因此選擇使用深度學習模型來作為預測信用卡詐騙的方法。

我們使用三種機器學習方法 Logistic Regression、Random Forest、Support Vector Machine 與深度學習 Keras 進行信用卡詐騙偵測的比較，並將資料集利用 Undersample 調整成五種比例，再由五種評估指標評估結果。Keras 在整體表現較為良好，Accuracy 都在 97%以上，Precision 最高將近 80%，Recall 平均都在 80%以上，F1-Score 最好能達到 80%，而 MCC 代表整體預測精度，最好的結果能達到 80%。

參考文獻

1. Sebastian Raschka. (2017) · Python 機器學習(初版九刷)(劉立民、吳建華) · 臺灣：博碩文化股份有限公司 · (原著出版年:2015)。
2. 林大貴。 (2017) · TensorFlow+Keras 深度學習人工智慧實務應用(初版) · 臺灣:博碩文化股份有限公司。
3. 尼爾森：不到四成台灣消費者每天消費偏好塑膠貨幣付款。Retrieved from <http://www.nielsen.com/tw/zh/press-room/2014/newsTaiwanPayment0225.html>
4. 金融監督管理委員會 - 金融監督管理委員會-防範信用卡網路交易盜刷機制。

Retrieved from

https://www.fsc.gov.tw/ch/home.jsp?id=96&parentpath=0,2&mcustomize=news_view.jsp&dataserno=201802130007&aplistdn=ou&dtable=News

5. 財團法人聯合信用卡處理中心公開資料 - 各發卡機構通報之詐欺金額。

Retrieved from

https://www.nccc.com.tw/wps/wcm/connect/zh/home/openinformation/Business?WC_M_PI=1&WCM_Page.bcd178c-e429-4c73-b2bb-43c4fbe133e5=1

6. Fighting Financial Fraud with Targeted Friction – Airbnb Engineering & Data Science – Medium. Retrieved from <https://medium.com/airbnb-engineering/fighting-financial-fraud-with-targeted-friction-82d950d8900e>

7. Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *2017 International Conference on Computing Networking and Informatics (ICCNI)*.
8. Bahnsen, A. C., Stojanovic, A., Aouada, D., & Ottersten, B. (2013). Cost Sensitive Credit Card Fraud Detection Using Bayes Minimum Risk. *2013 12th International Conference on Machine Learning and Applications*.
9. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602-613.
10. Kho, J. R., & Vea, L. A. (2017). Credit card fraud detection based on transaction behavior. *TENCON 2017 - 2017 IEEE Region 10 Conference*.
11. Pozzolo, A. D., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating Probability with Undersampling for Unbalanced Classification. *2015 IEEE Symposium Series on Computational Intelligence*.
12. Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., & Nandi, A. K. (2018). Credit Card Fraud Detection Using AdaBoost and Majority Voting. *IEEE Access*, 6, 14277-14284.

13. M. Fahmi, A. Hamdy and K. Nagati, "Data Mining Techniques for Credit Card Fraud Detection: Empirical Study," Sustainable Vital Technologies in Engineering & Informatics, pp. 1-9, 2016.