



## Predicting Default Payments of Credit Card Clients by Using XGBoost with Particle Swarm Optimization

---

Lee Jyh-Yuan and Hung Jui-Chung

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 24, 2021

# XGBoost 結合最佳化粒子群演算法預測信用卡客戶的繳款狀況

李智遠 洪瑞鍾\*

Department of Computer Science University of Taipei, Taipei 100, Taiwan  
juichung@gmail.com

## 摘要

金管會統計臺灣電子支付的使用人數已經正式突破 1,000 萬人，行政院主計處的國情統計通報中，109 年 12 月底信用卡總流通卡數 5,012 萬張，累計 1~12 月信用卡簽帳金額 3 兆 196 億元，其中超過 3 個月未繳足最低應繳的逾期帳款比率為 0.15%，由此可知違約率的預測是非常重要的議題。

極限梯度提升法(Extreme Gradient Boosting, XGBoost)是一種整合性演算法，將模型中的目標函數加入參數的複雜度，並以泰勒展開式的近似方式來減少過度擬合及提升其分類效能，由於同時須調整損失函數、樹的深度及樣本的權重，是一個複雜的問題，因此本論文將使用最佳化粒子群演算法(Particle Swarm Optimization, PSO)來篩選最佳參數。信用卡客戶的違約記錄是屬於不平衡的資料，且違約率是一重要議題，但所佔的比例很少，因此無法透過準確率(Accuracy)真正呈現，本論文以準確率(Accuracy)並加入召回率(Recall)為評估的準則，在可以接受的準確率(Accuracy)範圍內，來提升召回率(Recall)。

## 1 簡介

表 1.1 為信用卡統計業務，截至 109 年 12 月底信用卡總流通卡數 5,012 萬張，最近六個月有消費紀錄之有效卡數 3,300 萬張，累計 109 年 1~12 月信用卡簽帳金額 3 兆 196 億元，循環信用餘額 1,056 億元。

圖 1.1 為信用卡整體逾期放款比率趨勢圖，逾期放款比率是指借款人積欠本金或利息超過清償期 3 個月，占總放款的比率已達至 0.15% 的新低水位，顯示持卡人透過銀行的發卡、審核及協商機制，不再任意並有限度的使用信用卡交易。其中違約戶與正常戶呈現極不平衡的資料分佈，因此本論文使用準確率(Accuracy)及召回率(Recall)為評估的準則，並使用在不平衡資料有良好效果的 XGBoost，相關參數由 PSO 來搜尋，以獲得較好的預測模型。

表 1.1：信用卡統計業務

信用卡	109 年 12 月	較上月之 增減變動		較去年同期之 增減變動	
		數量	百分比	數量	百分比
總流通卡數 (萬張)	5,012	15	0.30%	273	5.76%
總有效卡數 (萬張)	3,300	18	0.55%	108	3.38
循環信用餘 額(億元)	1,056	9	0.86%	-83	-7.29%
簽帳總金額 (億元)	2,958	363	13.99%	49	1.68%
逾期帳款比 率(%)	0.15	-0.01	/	-0.06	/
民國年、月	簽帳金額			預借現金	
	合計	國內	國外	國內	國外
109 年 1-12 月簽帳金額 (億元)	30,196	28,403	1,793	220	2.07

(資料來源：金管會銀行局[25])



(圖片來源：金管會銀行局[25])

圖 1.1：信用卡整體逾期放款比率趨勢圖

## 2 文獻探討

### 2.1 信用卡交易及風險評估

信用卡為現代人主要的消費方式之一，其中每月信用卡帳單上揭露的訊息內容包含帳單結帳日、繳款截止日、本期應繳總金額、本期最低應繳金額、循環信用利率，其中若未於繳款截止日前繳足本期最低應繳金額，則會被銀行收取 100~500 元不等的違約金額。

信用卡風險評估為近年來的熱門議題，例如透過決策樹、隨機森林、神經網路集，利用 UCI 公開的信用卡資料集進行模型的建立，研究結果顯示教育、婚姻、還款狀況與帳單金額是提升準確率的重要因素[16]。利用 RapidMiner 軟體，將資料進行特徵優化選擇及交叉驗證的預處理，使用決策樹、隨機森林、類神經網路、支持向量機、倒轉遞類神經網路等演算法，透過模型參數最佳化，有效的提升準確率[21]。採用資料探勘常用的演算法，將信用卡的繳款行為資料透過集群方法進行資料分割並快速處理，輔以涵蓋曲線值(Area Under the Curve, AUC)進行評估，研究結果顯示 F-score 有所提升[17]。上述研究皆採用平衡資料集的方式進行評估，容易導致準確率高，但對於違約戶的預測無顯著的提升，即連續 3 個月符合上述「未於繳款截止日前繳足本期最低應繳金額」者，因為佔所有信用卡交易戶的比例非常低，屬於不平衡資料集，因此有學者提出相關不平衡資料集的研究方法如下：

表 2.1：不平衡信用數據相關研究

研究學者	年份	資料來源	研究結果
T. M. Alam <i>et al.</i>	2020	臺灣、德國、比利時信用及欺詐數據集	使用梯度提升決策樹(Gradient Boosted Decision Tree, GBDT)模型，透過資料集進行 K-means SMOTE 重採樣，有效的提升準確率[4]。
V. Arora <i>et al.</i>	2020	信用卡欺詐檢測及信用卡客戶數據集	信用卡資料透過不平衡資料集處理技術(Random Under Sampling Boost, RUSBoost)，並與決策樹、邏輯回歸、多層感知器、K 近鄰、隨機森林和支持向量機比較，獲得較佳的預測結果[5]。
陳佩妤	2018	信用卡交易資料	透過減少多數法(Undersample)設定成 5 種不同比例，並透過深度學習計算出 5 種預測結果，結果顯示整體預測結果皆有良好的表現[18]。

綜合以上研究，由於本論文使用的資料具有不平衡性，因此將資料進行前處理，利用在不平衡資料有良好效果的 XGBoost 演算法[23]建立預測模型，並透過準確率(Accuracy)及召回率(Recall)同時進行多目標的評估。

## 2.2 極限梯度提升演算法探討

信用卡交易違約率僅佔信用卡整體比例 0.15%，屬於不平衡資料集，對於不平衡資料集一般的處理方式有 2 種，將資料重新抽樣及更新優

化演算法，而資料抽樣的方法主要有 3 種[19]：1. 過取樣(Oversampling)，依資料樣本的比例設定，隨機複製數量較少的類別資料，使資料分佈平衡。2. 欠取樣(Undersampling)，依資料樣本的比例設定，隨機將數量較多的類別資料減少，使資料分佈平衡。3. 合成少數過取樣技術(Synthetic Minority Over-Sampling Technique, SMOTE) [7]，改良過取樣(Oversampling)的抽樣方法，從數量較少的樣本中找出幾個鄰近點，生成新的少數組樣本，使資料分佈平衡。

由於以不平衡資料集建立的預測模型會導致準確率的失真，因此本論文以相同比例的少數組與多數組，隨機合併成多份資料，並同時使用準確率(Accuracy)及召回率(Recall)做為評估準則，利用 XGBoost 建立預測模型。XGBoost [2]是一種整合性演算法，由於訓練效率高、準確率高，屬於熱門的機器學習演算法，其中因為演算法中使用了多維的參數，因此如何調整參數並提升模型效果是一複雜的問題，下列為使用 XGBoost 的相關研究：

表 2.2：XGBoost 的相關研究

研究學者	年份	參數設定方法	研究結果
李宜珊	2020	基因演算法(Genetic Algorithm, GA) [1]	使用 Rossmann Store Sales 資料集利用 XGBoost 建立預測模型，並透過 GA 來篩選參數，有效的降低均方根百分比誤差[15]。
劉順富	2019	R-MLR 套件[6]	以 2004 年 7 月至 2018 年 9 月的盤後資料，透過 XGBoost 預測漲跌，結果顯示與分類迴歸樹(Classification And Regression Tree, CART)皆有獲利，但 XGBoost 績效較優[20]。
M. Wang <i>et al.</i>	2018	Grid Search [8]	以 UCI 提供的 1994 年德國信用資料集，利用 XGBoost 取得重要特徵並結合邏輯迴歸(Logistic regression, LR)，有效提升 AUC 值[13]。

上述研究中皆採用演化式學習方法設定 XGBoost 的參數，如 GA、R-MLR 套件、Grid Search 等，本論文調整 XGBoost 的參數項目較少，並具有明確的範圍，因此採用 PSO 的群體搜尋方法，依其特性可快速獲得最佳化參數。

## 2.3 最佳化粒子群演算法探討

PSO 是一種演化式演算法[3]，調整方式主要是透過先前行為的慣性、個人最佳的經驗及整體最佳的經驗，對於參數的搜尋具有簡單、速度快、

較少參數設定、易於收斂的特質，相關研究如下表：

表 2.3：演算法結合 PSO 的相關研究

研究學者	年份	演算法	研究結果
謝婷亘	2020	支援向量迴歸(Support Vector Regression, SVR)	使用環保署開放平臺的空氣品質監測資料，SVR 建立預測模型，並透過 PSO 篩選最佳化參數，相較於未使用 PSO 的方法，均方根誤差(Root Mean Squared Error, RMSE)降低 19.567%[22]。
H. Jiang <i>et al.</i>	2020	XGBoost	以 NSL-KDD 數據集為基準，評估原生 XGBoost 與其他分類器，研究顯示可有效提高網路入侵檢測的準確率，特別是少數種類(U2R 和 R2L)的攻擊[9]。
K. Song <i>et al.</i>	2020	XGBoost	在材料科學中，鋼的拉伸強度和塑性分別具有抵抗斷裂和變形的能力，是二個重要性能，研究中透過 XGBoost 及 PSO 篩選模型參數以預測具有 27 個特徵的拉伸強度和塑性，可有效的優化此二性能[12]。
L. T. Le <i>et al.</i>	2019	XGBoost	研究中分析了 837 項對建築物的調查，並利用 XGBoost 建立模型及 PSO 優化性能，研究顯示熱負荷估計優於其它 5 項預測模型，並發現多項關鍵特徵[10]。

### 3 研究方法

#### 3.1 資料集與資料的前處理

本論文的資料來源為 Kaggle 平臺上 UCI 機器學習提供的 2005 年 4 月至 2005 年 9 月的臺灣地區信用卡客戶的資料集，包含 25 個特徵欄位(附錄一)，共有 30,000 筆資料，其中正常戶 23,364 筆，違約戶 6,636 筆，如圖 3.1。

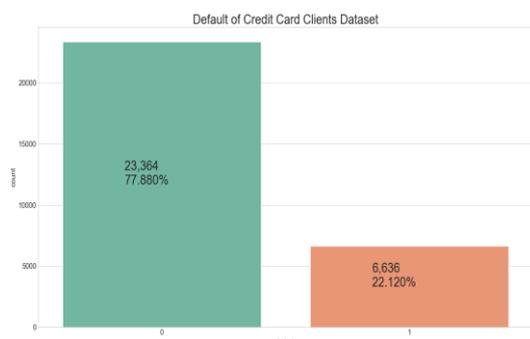


圖 3.1：UCI 信用卡客戶資料集

由於資料中不同的特徵欄位有各自的度量單位，因此預先將所有數值型特徵欄位利用標準均一化(1)，指定至 0 與 1 的範圍內，圖 3.2 為經過標準化後的部份欄位所呈現出來的信用卡客戶資料樣本集。

$$x_{nor} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

式子(1)中  $x$  表示為特徵屬性的值， $x_{nor}$  為標準化後的數值， $x_{min}$  為特徵屬性最小值， $x_{max}$  為特徵屬性最大值。

	LIMIT_BAL	PAY_0	BILL_AMT1	PAY_AMT1
0	0.010101	0.400000	0.149982	0.000000
1	0.111111	0.100000	0.148892	0.000000
2	0.080808	0.200000	0.172392	0.001738
3	0.040404	0.200000	0.188100	0.002290
4	0.040404	0.100000	0.154144	0.002290
5	0.040404	0.200000	0.203506	0.002862
6	0.494949	0.200000	0.472126	0.062961
7	0.090909	0.200000	0.157028	0.000435
8	0.131313	0.200000	0.156505	0.003811
9	0.010101	0.000000	0.146519	0.000000

圖 3.2：標準化後信用卡客戶資料樣本集

資料集內容中，目標值 0 表示正常戶，1 表示違約戶，並以隨機方式切分為訓練資料集 70% 及測試資料集 30%，其中測試資料集共 9,000 筆，包含正常戶 7,009 筆，違約戶 1,991 筆，訓練資料集共 21,000 筆，包含正常戶 16,355 筆，違約戶 4,645 筆，本論文將採 2 種方式進行訓練，首先，如圖 3.3，以一般方式進行訓練，其次為多模組平衡集方式(Multi Model Balance Set, MMBS)進行訓練，如圖 3.4，以訓練資料集的目標值進行資料的切分，其中 0 與 1 的比例為 3.521：1 屬於不平衡資料集，因此會將訓練資料集進行前處理作業，透過一般資料探勘常用的方法，以目標值 1 的資料共 4,645 筆，隨機取得 3 組相同筆數，目標值 0 的資料，且資料間彼此互斥不重覆，再重新與目標值 1 的資料合併成為 3 組資料集，因此可得到 3 組近似平衡訓練集資料均為 9,290 筆，再利用 XGBoost 建立 3 個預測模型，並依 3 個模型的預測結果，以多數決的方式產生新的預測結果，如下表 3.1，再透過 PSO 同時以準確率(Accuracy)及召回率(Recall)做為篩選最佳參數的條件。

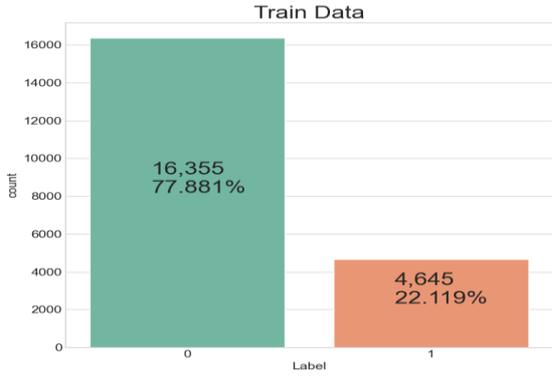


圖 3.3：一般訓練資料集

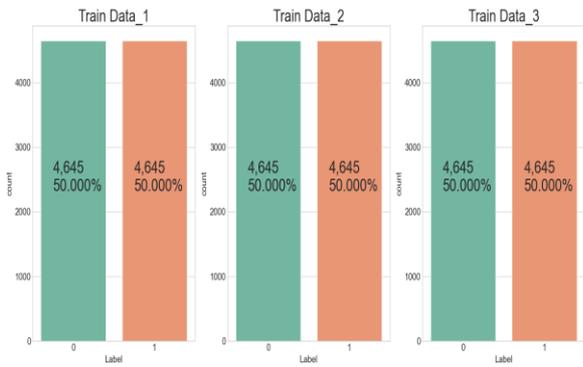


圖 3.4：多模組平衡訓練資料集

表 3.1：調整後預測結果模型示例

序號	預測結果			最後結果
	模型 1	模型 2	模型 3	
0	0	0	0	0
1	0	0	1	0
2	0	1	0	0
3	0	1	1	1
4	1	0	0	0
5	1	0	1	1
6	1	1	0	1
7	1	1	1	1

表 3.2 為混淆矩陣，可用於計算準確率 (Accuracy)、精確率 (Precision)、召回率 (Recall) 等值，主要分為 4 個值：TN (True Negative)、TP (True Positive)、FN (False Negative)、FP (False Positive)，其中 TN 及 TP 值越高表示分類效果越好，本論文將透過準確率 (Accuracy) 及召回率 (Recall) 做為 PSO 的評估準則，其中準確率 (Accuracy) 的評估式如下：

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (2)$$

召回率 (Recall) 的評估式如下：

$$Recall = TP / (TP + FN) \quad (3)$$

表 3.2：混淆矩陣

	預測為假	預測為真
實際為假	TN (True Negative)	FP (False Positive)
實際為真	FN (False Negative)	TP (True Positive)

本論文使用資料集為一不平衡資料集，Accuracy 及 Recall 呈現對應的關係，因此搭配 Sigmoid 函數進行準則的平衡，Sigmoid 函數是一種常見的邏輯函數(4)，可將輸入的資料映射到 [0, 1] 之間，使用此函數可調整輸入的 Accuracy 及 Recall，在接近最大值時減緩其增加速度，利用 S 形映射值[14]改進 PSO 以獲得最優解，如圖 3.5 所示。

$$f(x) = \frac{1}{1 + e^{-ax}}, \quad 1 \leq a \leq 10 \quad (4)$$

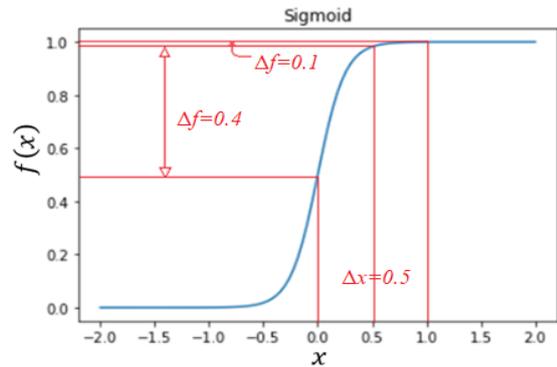


圖 3.5：Sigmoid 函數

其中  $a$  表示為 Sigmoid 函數的曲度係數項，本論文中以 1~10 進行評估， $\Delta$  表示為  $f(x)$  對應相同等距的輸入值  $x$ ，所增加的值。

### 3.2 極限梯度提升 (Extreme Gradient Boosting, XGBoost)

XGBoost 是一種整合式學習法，將許多樹的預測結果整合在一起以實現更好的分類效能，屬於監督式學習演算法，其目標函數表示如下式[23]：

$$Obj(\theta) = L(\theta) + \Omega_K(\theta) \quad (5)$$

其中  $\theta$  為訓練模型的參數， $L$  為損失函數， $\Omega_K$  為模型複雜度的懲罰項，若預設資料樣集為  $(x_i, y_i)$ ，其中  $x_i \in R^n, y_i \in R$ ，表示  $x_i$  有  $n$  個特徵， $y_i$  表示為  $x_i$  的目標值，其模型可表示如(6)：

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (6)$$

$\hat{y}_i$  表示為模型的預測值， $i = 1, 2, \dots, N$  為資料的筆數， $K$  為樹的數目， $f_k(x_i)$  表示為樹的模型結構，式子 (5) 的損失函數本論文中，定義為使用

*Accuracy* 與 *Recall* 為評估的準則，可表示如式子(7)：

$$L(\theta) = N((1 - Accuracy) + (1 - Recall)) \quad (7)$$

及搭配使用 Sigmoid 函數評估的式子(8)：

$$L(\theta) = N \left( \left( 1 - \frac{1}{1+e^{-aAccuracy}} \right) + \left( 1 - \frac{1}{1+e^{-aRecall}} \right) \right) \quad (8)$$

式子(5)中  $\Omega_K(\theta)$  為樹的複雜度，表示如(9)：

$$\Omega_k(\theta) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (9)$$

式子(9)主要分為葉子節點數及懲罰項因子，其中  $T$  為葉子的節點數目， $\gamma$  為控制葉子節點數目的比重， $\lambda$  為懲罰項的比重，用於避免過擬合， $\omega$  為葉子節點的權重，為了優化  $f_k(x_i)$  因此 XGBoost 透過泰勒展開式(Taylor Expansion)的二階近似[23]，將目標函數進行優化，表示如式子(10)：

$$Obj^{(k)} \cong \sum_{i=1}^N \left[ m_i f_k(x_i) + \frac{1}{2} c_i f_k^2(x_i) \right] + \sum_{k=1}^K \left( \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \right) \quad (10)$$

其中  $m_i$  為泰勒展開一階的係數項， $c_i$  則為二階係數項，本論文中須同時調整  $K$ 、 $\gamma$ 、 $\lambda$ ，因此採用 PSO 調整目標函數至最小值。

### 3.3 最佳化粒子群演算法 (Particle Swarm Optimization, PSO)

PSO 是一種演化式演算法，源自於簡化鳥類群聚社會的模擬，每個粒子皆代表鳥群的個體，主要是透過先前行為的慣性，個體的最佳經驗及群體的最佳經驗，並經過多次的迭代，更新個體及群體的最佳位置、速度，直到尋找到式子(11)的最小值為其最佳解，每個粒子的速度及位置更新如下式[3]：

$$\begin{aligned} v_{i,d}(t+1) &= \beta * v_{i,d}(t) \\ &+ c_1 * rand_1 * (pbest_{i,d} - x_{i,d}(t)) \\ &+ c_2 * rand_2 * (gbest_d - x_{i,d}(t)) \end{aligned} \quad (11)$$

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1) \quad (12)$$

其中  $i = 1, 2, \dots, P$  為粒子群的數量； $d = 3$  表示參數的維度； $\beta$  表示慣性權重； $c_1$  及  $c_2$  為個體及群體的學習因子； $rand_1$  及  $rand_2$  表示在  $(0, 1)$  之間均勻分佈的 2 個隨機變數； $t$  為迭代的次數； $x_{i,d}$  為第  $i$  個

粒子在  $d$  維度的搜尋位置； $v_{i,d}$  表示第  $i$  個粒子在  $d$  維度的速度； $pbest_{i,d}$  表示第  $i$  個粒子在  $d$  維度經歷過的最佳位置； $gbest_d$  表示群體在  $d$  維度中經歷過的最佳位置。本論文將參考表 3.2 的參數設定，避免  $c_1$ 、 $c_2$  及  $\beta$  的設定不良，導致搜尋最佳化參數時無法獲得最適合的結果。

表 3.3：PSO 搜尋參數範圍[11]

學習因子		慣性權重	迭代次數
$c_1$	$c_2$	$\beta$	
2.1	2.1	0.9	50

## 4 研究結果

本論文透過 2 種方式進行訓練並產生預測結果的比較，如下表 4.1，結果顯示，不平衡資料集中，在  $Accuracy \geq 0.7$  的可接受範圍內，以一般方式使用 XGBoost 進行訓練與搭配 PSO 以  $Accuracy$  與  $Recall$  為評估準則， $Recall$  可由 0.37 提升到 0.59，若以 MMBS 方式進行訓練，並搭配 PSO 及 Sigmoid 函數，則  $Recall$  由 0.65 提升至 0.74。

表 4.1：訓練結果表

訓練方式	訓練集	測試集	準確率	召回率	PSO	Sigmoid 函數
一般	21,000	9,000	0.82	0.37		
一般	21,000	9,000	0.78	0.59	V	
MMBS	9,290	9,000	0.73	0.65		
MMBS	9,290	9,000	0.71	0.67	V	
MMBS	9,290	9,000	0.71	0.74	V	V

## 5 結論

信用卡為現代人的主要支付消費方式，發卡機構如何提前預測客戶違約，減少逾期放款所帶來的損失是一重要議題，本論文利用多模組平衡訓練方式，並透過 PSO 尋找 XGBoost 的最佳參數，數據顯示可以有效的提升預測不平衡資料集中的違約客戶。

## 參考文獻

- [1] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.
- [2] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the*

- 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785-794.
- [3] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-International Conference on Neural Networks*, 1995, vol. 4: IEEE, pp. 1942-1948.
- [4] T. M. Alam *et al.*, "An Investigation of Credit Card Default Prediction in the Imbalanced Datasets," *IEEE Access*, vol. 8, pp. 201173-201198, 2020.
- [5] V. Arora, R. S. Leekha, K. Lee, and A. Kataria, "Facilitating User Authorization from Imbalanced Data Logs of Credit Cards Using Artificial Intelligence," *Mobile Information Systems*, vol. 2020, 2020.
- [6] B. Bischl *et al.*, "mlr: Machine Learning in R," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 5938-5942, 2016.
- [7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321-357, 2002.
- [8] J. Y. Hesterman, L. Caucci, M. A. Kupinski, H. H. Barrett, and L. R. Furenlid, "Maximum-likelihood estimation with a contracting-grid search algorithm," *IEEE transactions on nuclear science*, vol. 57, no. 3, pp. 1077-1084, 2010.
- [9] H. Jiang, Z. He, G. Ye, and H. Zhang, "Network intrusion detection based on PSO-XGBoost model," *IEEE Access*, vol. 8, pp. 58392-58401, 2020.
- [10] L. T. Le, H. Nguyen, J. Zhou, J. Dou, and H. Moayed, "Estimating the heating load of buildings for smart city planning using a novel artificial intelligence technique PSO-XGBoost," *Applied Sciences*, vol. 9, no. 13, p. 2714, 2019.
- [11] S. Nath, J. K. Sing, and S. K. Sarkar, "Performance comparison of PSO and its new variants in the context of VLSI global routing," *Particle Swarm Optimization with Applications*, vol. 61, 2018.
- [12] K. Song, F. Yan, T. Ding, L. Gao, and S. Lu, "A steel property optimization model based on the XGBoost algorithm and improved PSO," *Computational Materials Science*, vol. 174, p. 109472, 2020.
- [13] M. Wang, J. Yu, and Z. Ji, "Credit Fraud Risk Detection Based on XGBoost-LR Hybrid Model," 2018.
- [14] 李浩君, 劉中鋒, and 王萬良, "採用弧形映射函數的二進製粒子群優化算法," *小型微型計算機系統*, vol. 38, no. 12, pp. 2637-2640, 2017.
- [15] 李宜珊, "XGBoost 結合基因演算法預測銷售額—以 Kaggle Rossmann Store Sales 問題為例," 碩士, 資訊科學系, 臺北市立大學, 臺北市, 2020.
- [16] 洪偉欣, "信用卡違約付款分析預測," 碩士, 管理學院資訊管理學程, 國立交通大學, 新竹市, 2019.
- [17] 張金燦, "運用分群技術於信用卡風險評估模型之研究," 碩士, 資訊工程學系碩士班, 銘傳大學, 臺北市, 2012.
- [18] 陳佩好, "利用深度學習預測信用卡詐騙," 碩士, 資訊管理學研究所(含碩專班), 世新大學, 臺北市, 2018.
- [19] 陳逸真, "不平衡資料分類方法之比較," 碩士, 統計學系, 國立臺北大學, 新北市, 2017.
- [20] 劉順富, "用 XGBOOST 演算法預測臺灣指數期貨," 碩士, 財務金融系, 國立雲林科技大學, 雲林縣, 2019.
- [21] 謝宗螢, "以資料探勘預測信用卡違約風險," 碩士, 財務金融學系碩士班, 國立屏東大學, 屏東縣, 2018.
- [22] 謝婷亘, "支援向量迴歸結合粒子群演算法建立 PM2.5 濃度預測系統之實作," 碩士, 資訊科學系, 臺北市立大學, 臺北市, 2020.
- [23] x. developers. "XGBoost Tutorials." <https://xgboost.readthedocs.io/en/latest/tutorials/index.html>.
- [24] Kaggle. "Default of Credit Card Clients Dataset." <https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset>.
- [25] 金管會銀行局. "109 年 12 月份信用卡、現金卡、電子票證及電子支付機構業務資訊." [https://www.banking.gov.tw/ch/home.jsp?id=540&parentpath=0,524,539&mcustomize=news\\_view.jsp&dataserno=202102040003&dt=News](https://www.banking.gov.tw/ch/home.jsp?id=540&parentpath=0,524,539&mcustomize=news_view.jsp&dataserno=202102040003&dt=News).