

信用卡營運 智慧分析系統

指導老師 黃登揚 老師、蔡智勇 老師

報告人 王國欽

組員 羅文隆、陳源成、李卉亭、
陳珮蓉、曾子容、林雯娟

前言

在消費型態改變與手機行動支付結合的推波助瀾之下；信用卡交易金額以每年約10%的幅度成長，銀行業者除提高信用卡使用率與金額的同時，如何降低盜刷與違約風險並保有獲利顯得格為重要。

有鑑於此，我們以玉山銀行公開競賽資料為基礎，透過大數據分析與模型建立；建立一個信用卡營運智慧分析系統；



報告大綱

01 信用卡交易流程與系統架構

- 信用卡交易流程
- 系統架構

02 資料觀察與處理

- 資料觀察
- 清洗資料
- 資料特性

03 盜刷預測

- 演算法
- 盜刷預測結果

04 系統介面展示

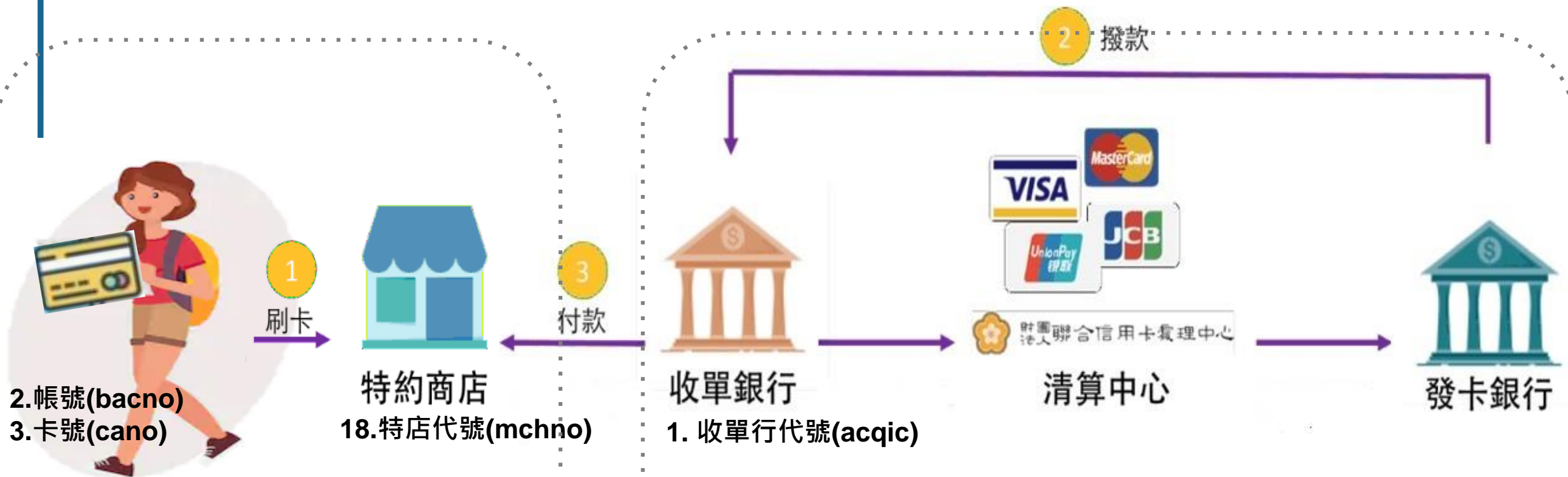
- 網頁畫面展示



信用卡盜刷風險 系統架構

- 信用卡交易流程
- 系統架構

信用卡交易流程



信用卡交易流程



7. 網路交易(ecfg) N/Y

21.消費地 國別(stocn) & 20.城市(scity) & 6.幣別(csmcu)

2. 撥款



4. 金額(conam)

15.授權日期(locdt) & 16.時間(loctm)

1

刷卡



特約商店

18.特店代號(mchno)

3

付款



收單銀行

1. 收單行代號(acqic)



清算中心



發卡銀行

2.帳號(bacno)

3.卡號(cano)

5. 交易類別(contp)

8. 交易型態(etymd)

17.交易商店類別(mcc)

22.狀態碼(stscd)

23.交易序號(txkey)

9. Fallback註記(flbnk) N/Y

10.3DS註記 (flg_3dsmk) N/Y

12.支付型態(hcefg)



13.分期交易註記(insfg) N/Y

14.分期期數(item)

19.超額註記(ovrlt) N/Y

11.盜刷註記(fraud_ind) 1=盜刷

系統架構



Data
資料觀察

- 檢視欄位內容意涵
- 觀察資料集的分佈
- 檢驗欄位值與空值

 python™

 pandas

 matplotlib

Clean
資料清理

- 補空值
- 合併並新增欄位
- 特徵篩選
- 利用KMEANS數字轉類別

 SciPy

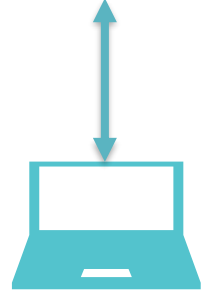
ML
機器學習

- 資料集平衡
- 模型測試
- 模型選定

 dmlc XGBoost

 scikit learn

Model
預測模型





信用卡交易 資料觀察與處理

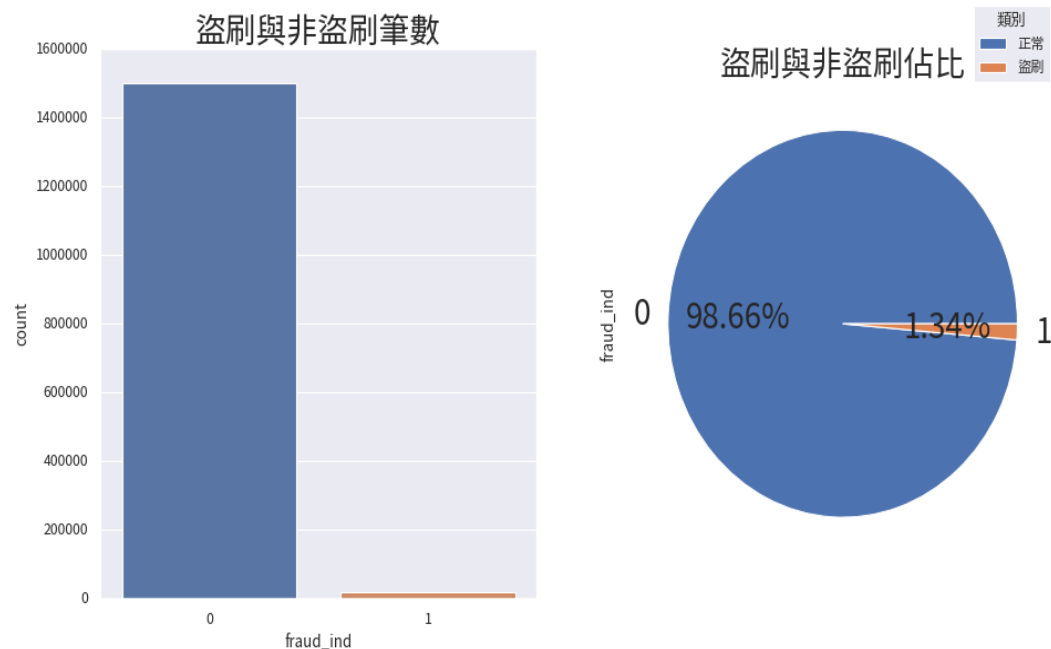
- 資料觀察
- 清洗資料
- 資料特性

資料觀察 – 欄位說明

編號	欄位	說明	項目數	備註
1	acqic	收單行代碼	6051	消費商店合作銀行的代碼
2	bacno	歸戶帳號	95214	持卡人帳戶
3	cano	交易卡號	129413	帳戶下可以有很多組卡號
4	conam	交易金額-台幣 (經過轉換)	72961	
5	contp	交易類別	7	例；正向交易
6	csmcu	消費地幣別	72	使用哪個幣別進行消費
7	ecfg	網路交易註記	2	交易是否在網路上完成
8	etymd	交易型態	11	交易方式，例：刷磁條機
9	flbmk	Fallback 註記	3	銀行懷疑不正常交易
10	flg_3dsmk	3DS 交易註記	3	有進行手機驗證碼之交易
11	fraud_ind	盜刷註記	2	是否有盜刷(應變數)
12	hcefg	支付形態	9	虛擬卡交易註記，其型態可能為APPLE PAY等

編號	欄位	說明	項目數	備註
13	insfg	分期交易註記	2	交易是否有分期
14	iterm	分期期數	9	
15	locdt	授權日期	90	刷卡時向發卡銀行提出授權之日期
16	loctm	授權時間	84810	刷卡時向發卡銀行提出授權之時間
17	mcc	MCC_CODE	434	消費子類別代碼
18	mchno	特店代號	89316	消費商店之代碼
19	ovrlt	超額註記碼	2	累積使用額度超過額度上限
20	scity	消費城市	5698	消費商店登記城市
21	stocn	消費地國別	103	消費商店登記國家
22	stscd	狀態碼	5	例：停卡、掛失、逾期等
23	txkey	交易序號	1521787	一筆交易產生一組交易序號(唯一值)

資料觀察 - 類別資料佔比



已結束

玉山人工智慧公開挑戰賽2019秋季賽
真相只有一個 - 『信用卡盜刷偵測』 >

9/6/2019 開始 11/22/2019 結束

23 萬元(NTD) TOTAL REWARD 1366 TEAMS

一卡在手，妙用無窮！
在台灣，20歲以上持有信用卡人數超過六成。因信用卡具備高回饋、延遲付款以及...
(More)

- 玉山提供150萬筆的資料，23個欄位
- 其中1欄為盜刷註記
- 盜刷註記資料比例為98.66 : 1.34 ; 資料極度不平衡

	盜刷交易	正常交易
筆數	20,355	1,501,432
佔比	98.66%	1.34%

資料處理- 缺值

總缺值筆數:

False 1509206

True 12581

盜刷缺值筆數:

False 20160

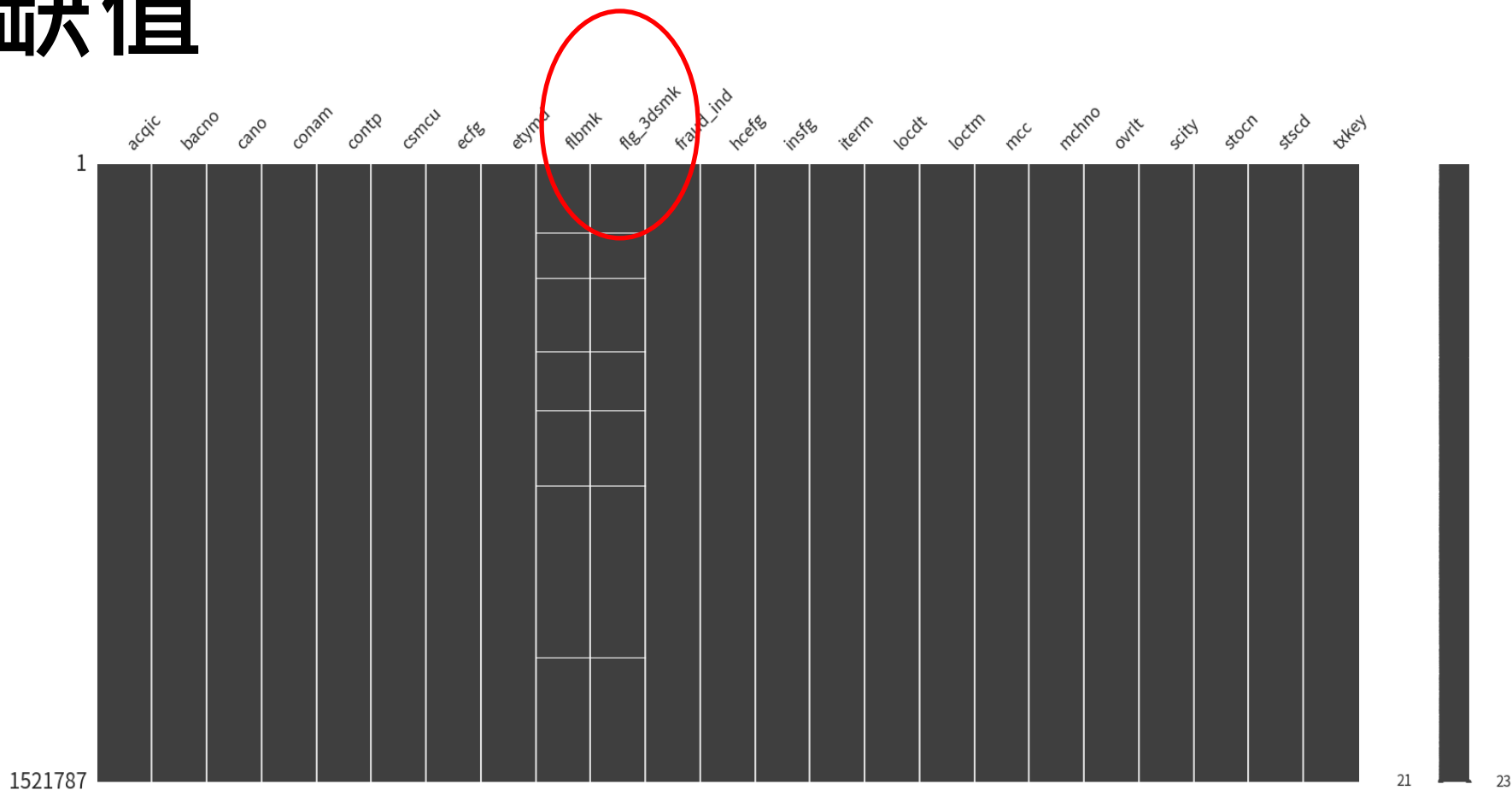
True 195

正常交易缺值筆數:

False 1489046

True 12386

- FlbmK與flg_3dsmk ,
- 會同時缺值。
- 佔盜刷資料將近1% , 我們無法損失這1%的錯誤; 因為盜刷資料佔比也僅1.34%。
- 故我們以 999值 填空值



資料處理 – 擴充欄位(時間)

15	locdt	授權日期	90	刷卡時向發卡銀行提出授權之日期
16	loctm	授權時間	84810	刷卡時向發卡銀行提出授權之時間



編號	欄位	說明
24	locdt_tm (timestamp)	賦予實際時間， 以2019-03-01為起算日期 Ex. 2019-03-01 16:36:37



25	hour	時
26	hour_0_6	(時間0 - 6點)
27	time_diff	與上次交易間隔小時
28	time_diff_0h	是否(一小時內)連續交易

資料處理 – 欄位合併(原始資料延展)

21	stocn	國家代碼		
20	scity	城市代碼		



17	mcc	商店類別代碼		
18	mchno	商店代碼		



編號	欄位	項目數	說明
30	stocn_scity		取代原本的城市代碼
29	mcc_mchno		取代原本的商店代碼

資料處理 – 新增欄位(原始資料延展)

1. 因欄位項目過多，但資料僅集中於部分項目

(1) stocn 變更為 stocn_name

參考台灣108年信用卡海外交易國家排名資料，
定義stocn之消費筆數前10名國家，餘為others

stocn_name			
stocn	stocn_name	stocn	stocn_name
102	Taiwan	20	Japan
38	USA	52	Singapore
46	Ireland	93	China
104	Netherlands	42	South Korea
75	England	44	Hong Kong

(2) mcc 變更為 mcc_adj

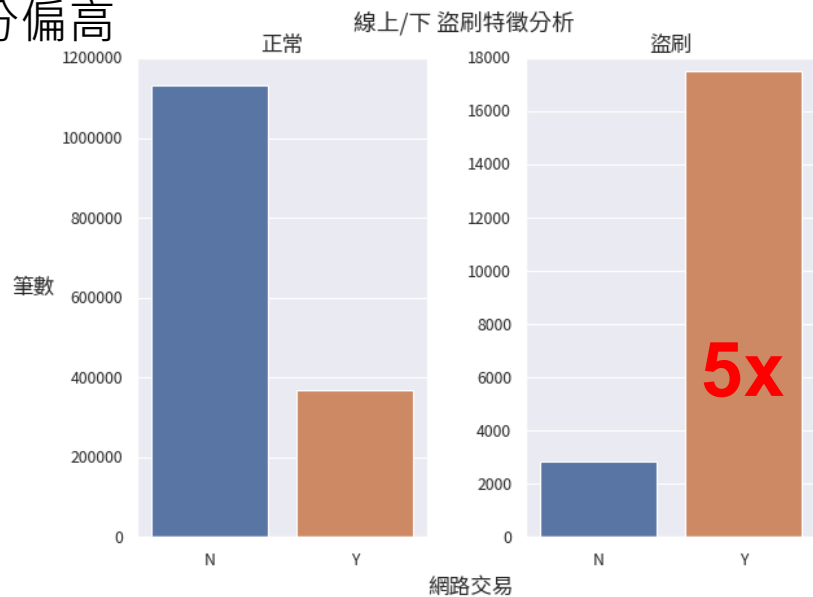
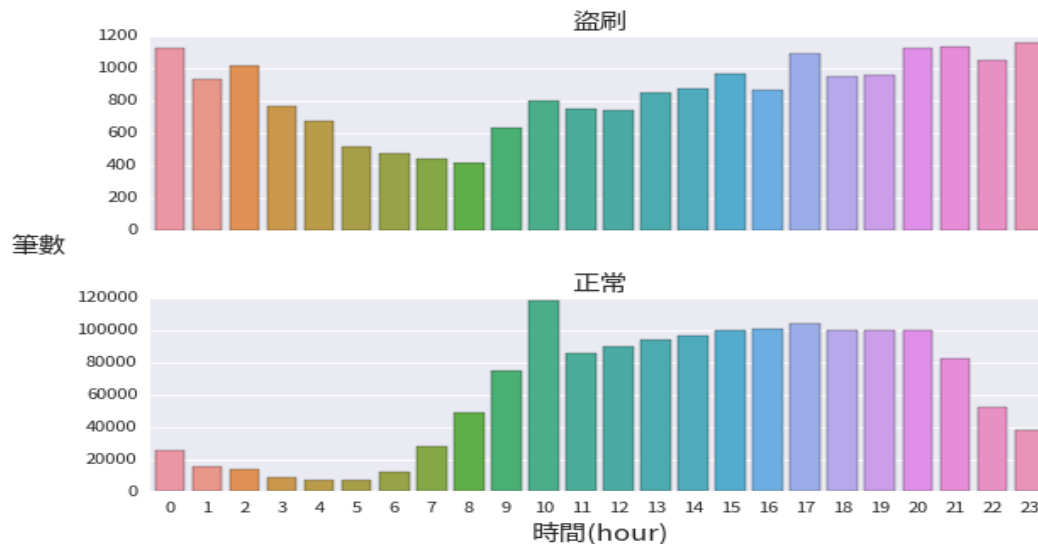
篩選出前10高交易筆數之mcc，餘為others

mcc_adj			
mcc	mcc_adj	stocn	mcc_adj
288	Video Game Archades	192	Drug Stores and Pharmacies
289	Cigar Store	247	Transportation Services
380	Computer Network Services	251	Medical Services
457	Restaurant	292	Hotels
263	Miscellaneous Retail		Others

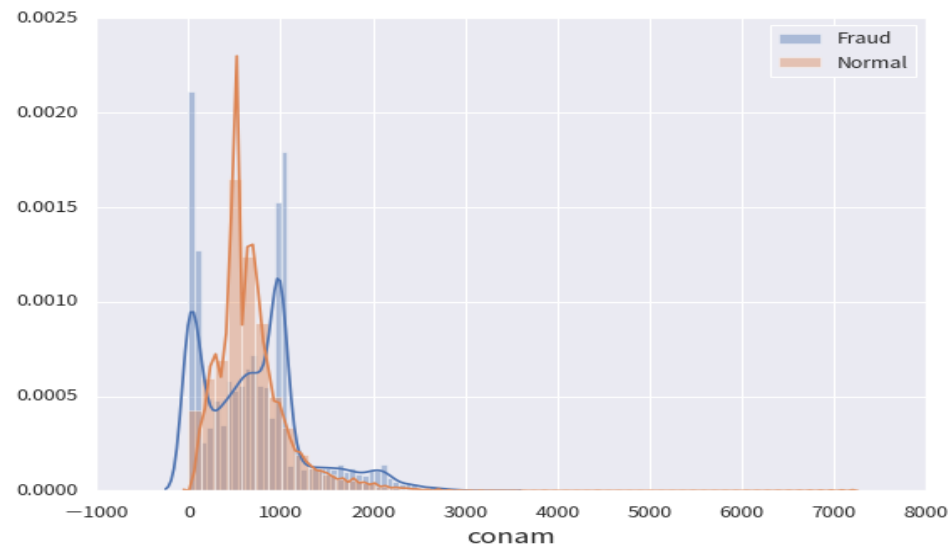
資料觀察

- 盜刷時間
 - 出現在凌晨時段
- 線上/實體店
 - 盜刷70%都是在網上
- 金額
 - 盜刷的金額有兩個高峰而正常交易資料常態分佈
 - 盜刷小額交易的部分偏高

盜刷時間高峰比較圖



正常/盜刷交易金額分布圖



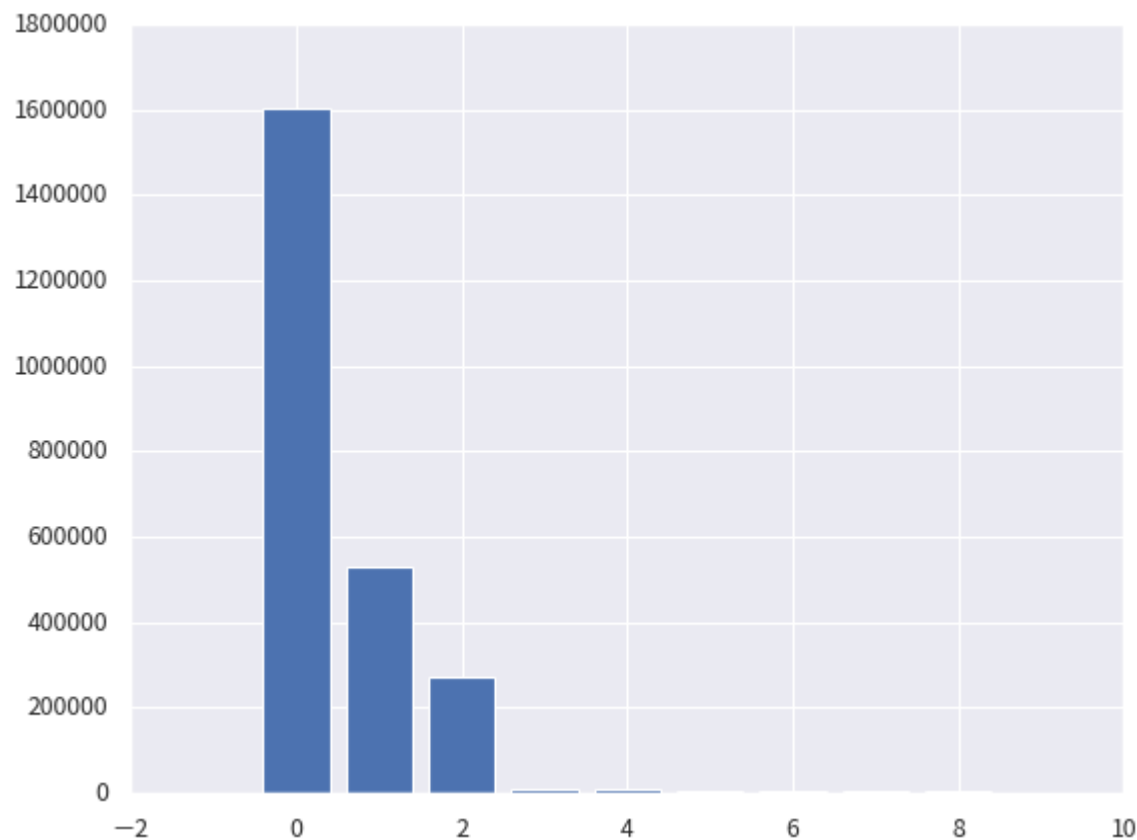


信用卡盜刷預測

- 演算法
- 盜刷預測結果

資料觀察- 篩選欄位

- 運用皮爾森卡方檢定挑選出9個與Y相關的類別變數

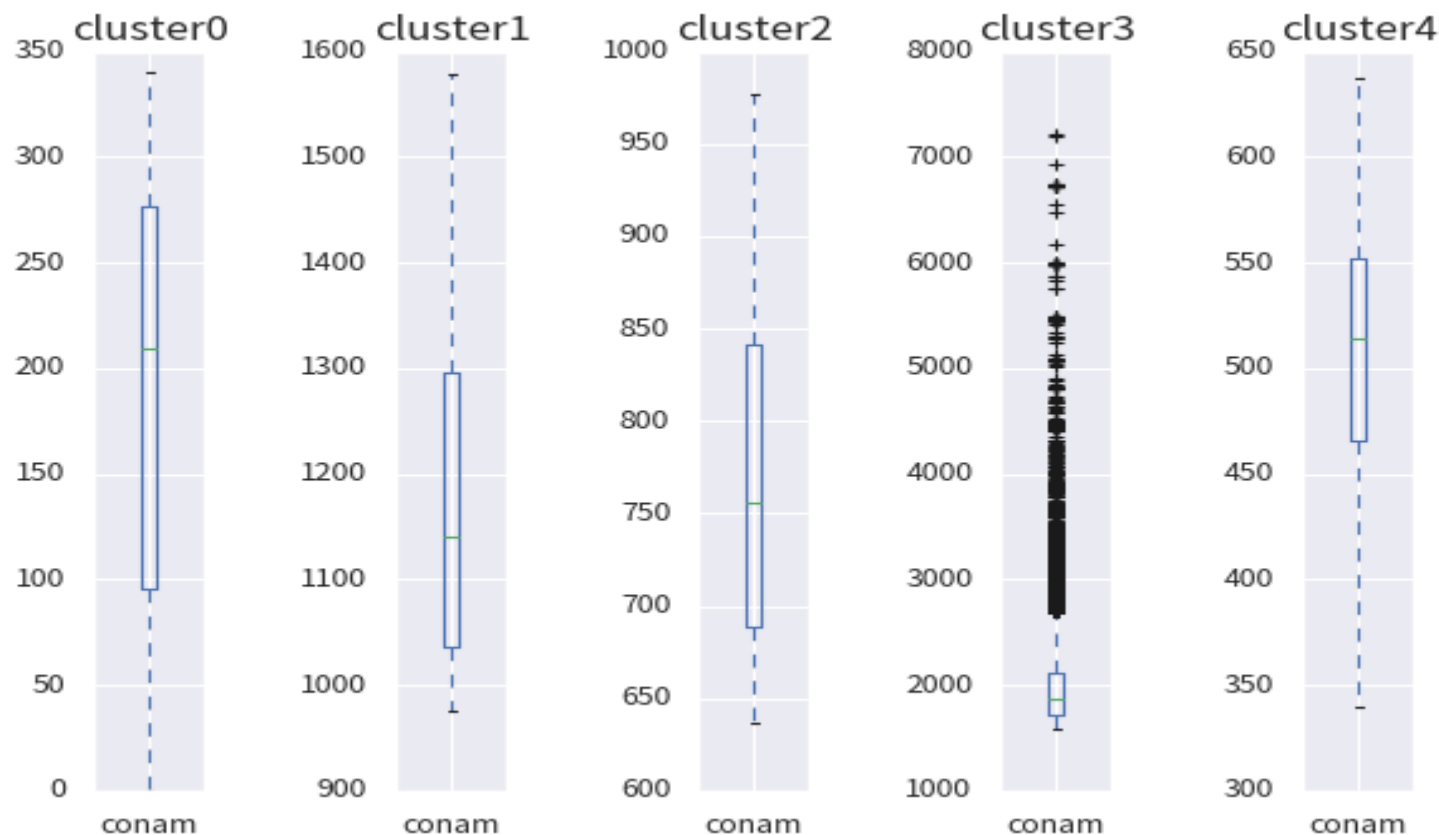


Index	FName	Name	Score
0	mcc_mchno	商店代碼	1604107
1	acqic	收單行	530045
2	time_diff	交易時間距	269837
3	stocn_adj	國家	9218
4	stscd	狀態	7188
5	ecfg	網路交易	4778
6	stocn_scity	城市	4392
7	mcc_adj	交易分類	3620
8	time_diff_0h	連續交易	2848

資料觀察- 集群

- 我們並不清楚，消費金額的區隔什麼樣叫做高中低。
- 利用KMEANS分群，讓他幫我們自動切割將消費金額做五個等級的區別。

Conam統計資料	
mean	654.722
std	402.808
min	0
25%	437.06
50%	591.9
75%	807.34
max	7208.77



演算法比較

- 在原始資料的情況底下，隨機森林與XGBoost會是一樣的得分。
- 如果使用新增的欄位並做LBEC的話，XGBoost整體的準確率會較高。
- 因此我們選定XGBoost來做後續的演算法來做為後續預測。

Algorithm	Features	Dimension reduction	Type	Accuracy	Precision	Recall	F1-score
Random Forest	Original	None	Tree	0.987	0.758	0.504	0.504
XGBoost	Original	None	Tree	0.987	0.758	0.504	0.504
Decision Tree	Original	None	Tree	0.828	0.5	0.502	0.465
Random Forest	21	刪除重複功能性欄位		0.921857	0.561769	0.872283	0.588785
XGBoost	21	刪除重複功能性欄位		0.981016	0.644107	0.651375	0.647646

預測結果

Algorithm	Features	Dimension reduction	Type	Accuracy	Precision	Recall	F1-score
Decision Tree	Original	None	Tree	0.828	0.5	0.502	0.465
XGBoost	21	刪除重複相關性欄位	Tree	0.981016	0.644107	0.651375	0.647646
XGBoost	9	chi square 檢定	Tree	0.975557	0.616246	0.685853	0.642405
XGBoost	+ conam		Tree	0.972633	0.618959	0.742305	0.657912
XGBoost	+conam_cluster		Tree	0.979704	0.650475	0.705237	0.673376



系統介面展示

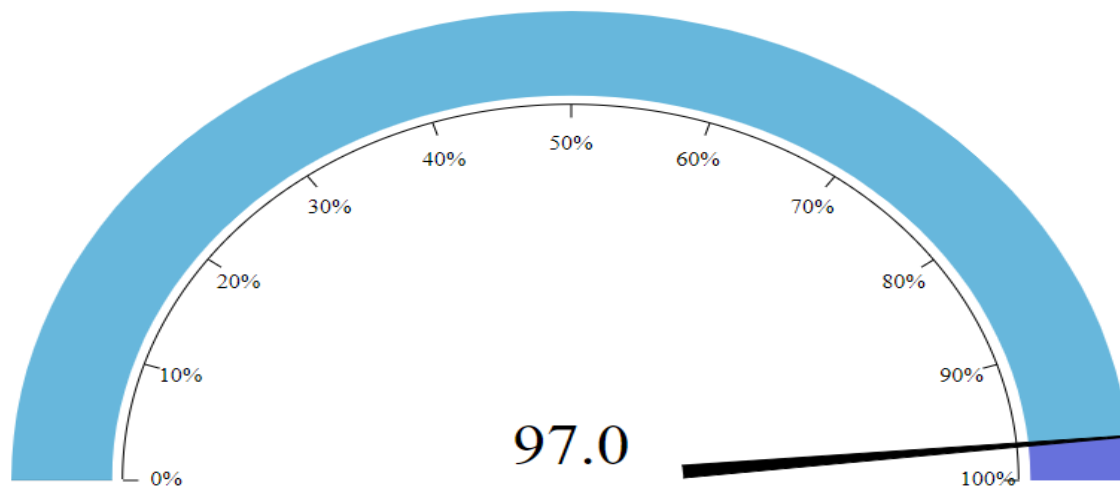
- 網頁畫面展示

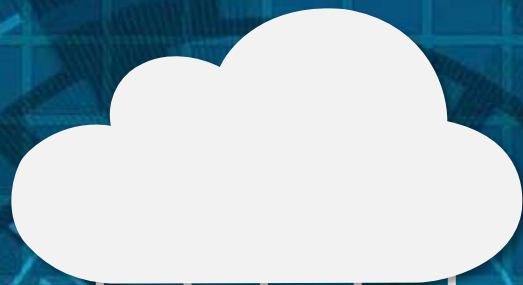
示意畫面

隨機交易資料

類別	特店代號	收單行	交易時間間隔(小時)	消費地	狀態碼	網路交易	國別城市	消費類別	交易時段為0-6時	交易金額
	28670836	6716	0.0	Taiwan	0	N	1025813	Others	1	776

此筆為 非詐欺 交易，預測模型的準確率為 97 %





THANK YOU

