

新光人壽挪用 業務員偵測

第一組

110352008 金碩二 陳科穎 109701017 應數三 羅芷矜

110352010 金碩二 張辰煜 109701043 應數三 黃予嫻

110352028 金碩二 黃柏翔



大綱

1 問題介紹

2 期中回顧

3 模型選擇

4 模型結果及評估

5 提案建議

6 結論

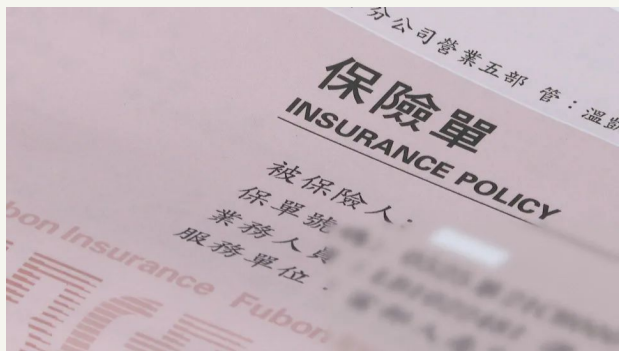




1. 問題介紹

- 研究動機
- 問題探討
- 預期目標

研究動機

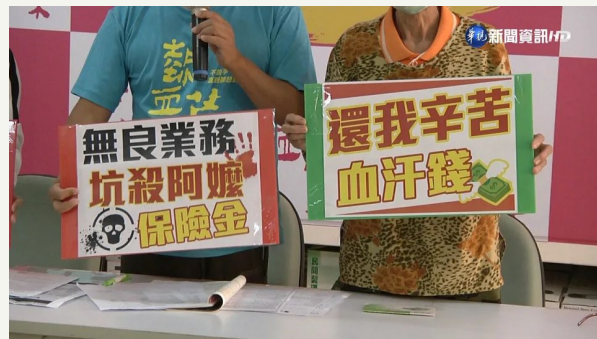


「黑心業務員騙客戶案例更是層出不窮。保經公司業務員尤小姐：『他們會有代客戶收取保費，實際比如說，收了一年、兩年、三年的保費，但其實沒有這張保單存在的狀況。』」

(來源:鏡新聞, 2022/08/01)

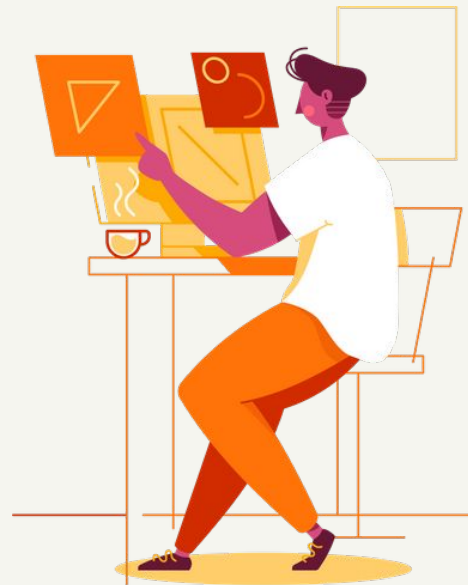
「台中一名 73 歲婦人，將工作辛苦存下來的 240 萬元，購買美元投資型保單，之後承辦的業務人員，卻在 107 年到 109 年間，以各種理由，要求她簽下不明文件，結果，竟然是業務員偷偷把款項轉進自己的戶頭...」

(來源:華視新聞, 2022/10/26)



問題探討

透過新光人壽給予的資料集，
從中萃取出重要的特徵，了解
哪些業務員的特徵，是能夠反映
挪用行為，藉此為公司制定
定期審核機制



預期目標

透過機器學習模型來協助新光人壽提升業務員挪用侵占保戶款項相關內控能力



順應ESG趨勢，強化公司治理(G)能力



提升風控能力並減少可能的財產或商譽損失



2. 期中回顧

- 填補缺失值
- 數據處理

缺失值統計

變數代號	缺失值比例	缺失值數量	變數名稱
dept_flg	0.05%	2	招攬部室代號
area_flg	0.05%	2	招攬區部代號
fyp_month_avg	32.94%	1324	月平均業績
persistence_prem_25	78.18%	3142	25個月繼續率
salary_ded_ratio	99.55%	4001	每月薪津扣款比例
age	0.05%	2	業務員現況年齡
bill_invalid_ratio	82.51%	3316	身為送金單領用人送金單作廢率(近一年平均)
bill_invalid_cnt	82.51%	3316	身為送金單領用人送金單作廢張數(近一年平均)
agent_own_plcy	70.04%	2815	業務員自有保單數
one_year_claim_plcy_cnt	70.04%	2815	保單生效後1年內申請賠償請求的保單數量
one_year_claim_plcy_ratio	70.04%	2815	保單生效後1年內申請賠償請求的保單數量佔比
addr_not_vld_rate	70.04%	2815	郵寄地址與居住地址不相符的保單張數佔比
comm_amount_1y_rate	70.04%	2815	近一年佣金收入與近三年佣金收入比
pc_in_60_days_rate	70.04%	2815	保單生效日後60天內保單變更的張數佔比
lapse_2y_rate	70.04%	2815	於理賠後兩年內保障停效之佔率

缺失值處理



刪除缺失 值超過八成的變數



SALARY_DED_RATIO(缺值4001筆)

→ 缺值補0



將**SERIAL_NO**、**CNTR_FLG**刪除

變數名稱	缺失值數量
BILL_INVALID_RATIO	3316
BILL_INVALID_CNT	
PERSISTENCE_PREM_25	3142
AGENT_OWN_PLCY	2815
ONE_YEAR_CLAIM_PLCY_CNT	
ONE_YEAR_CLAIM_PLCY_RATIO	
ADDR_NOT_VLD_RATE	
COMM_AMOUNT_1Y_RATE	
PC_IN_60_DAYS_RATE	
LAPSE_2Y_RATE	

數據處理



數值型資料: Min-Max Scaler



類別型資料: One-Hot Encoding



3. 模型選擇



1

羅吉斯迴歸

優點:

- 1.實做簡單
- 2.計算量小速度快, 儲存空間低

缺點:

- 1.特徵空間相當大時, 表現出的性能並不好
- 2.準確度不高

2

Support Vector Machine

優點:

- 可用kernel函數有效處理高維數據

缺點:

- 不適用資料量過大

3

隨機森林

優點:

- 1.準確率高, 能有效地在大數據上運行
- 2.引入隨機性, 不易過擬合

缺點:

- 訓練時所需要的空間和時間會很大, 導致模型越慢

4

XGBOOST

優點:

- 會修正之前訓練過之決策樹所產生的錯誤

缺點:

- 算法參數過多



4. 結果評估

- 研究流程
- 切分資料
- 參數調整

研究流程



▲ 切分資料

使用train_test_split套件進行訓練及測試資料切分

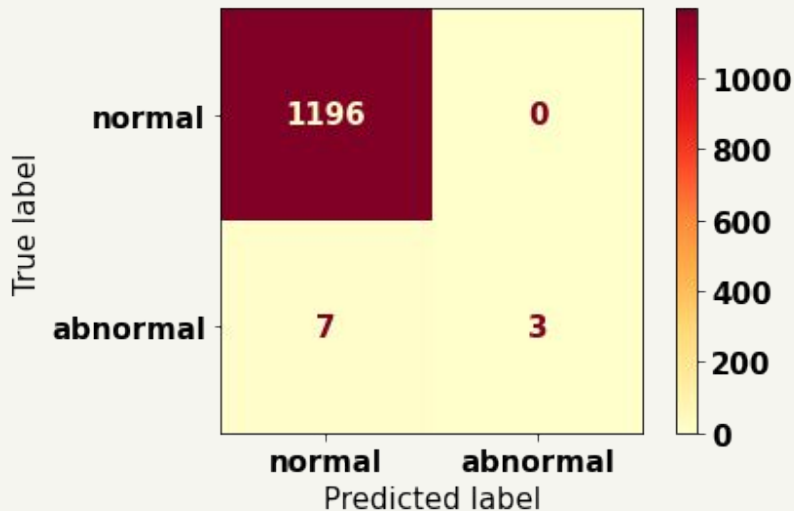
test_size = 0.3, 7成為訓練資料、3成為測試資料, random_state = 1234

	0	1
Train	2785	28
Test	1196	10



羅吉斯迴歸 預設參數

	Precision	Recall	F1-score	Support
0	0.99	1.00	1.00	1196
1	1.00	0.30	0.46	10

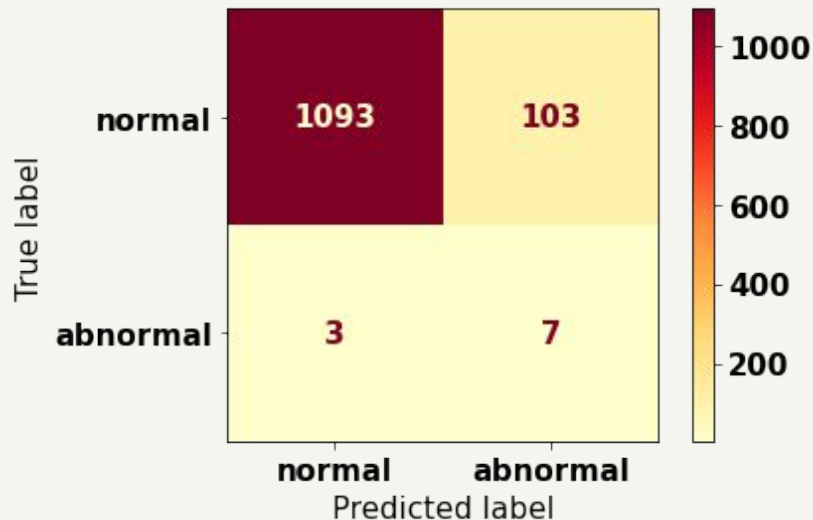




羅吉斯迴歸 調整參數

	C	class_weight
F1	1000	0:0.001,1:0.999

	Precision	Recall	F1-score	Support
0	1.00	0.91	0.95	1196
1	0.06	0.70	0.12	10

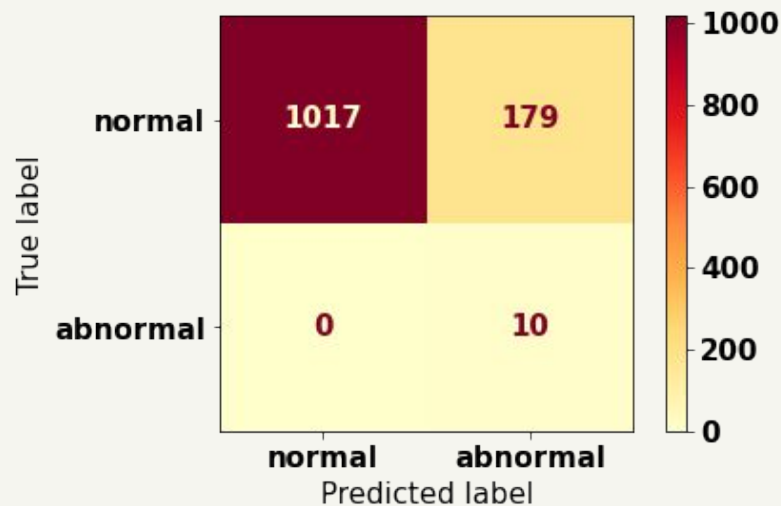




羅吉斯迴歸 調整參數

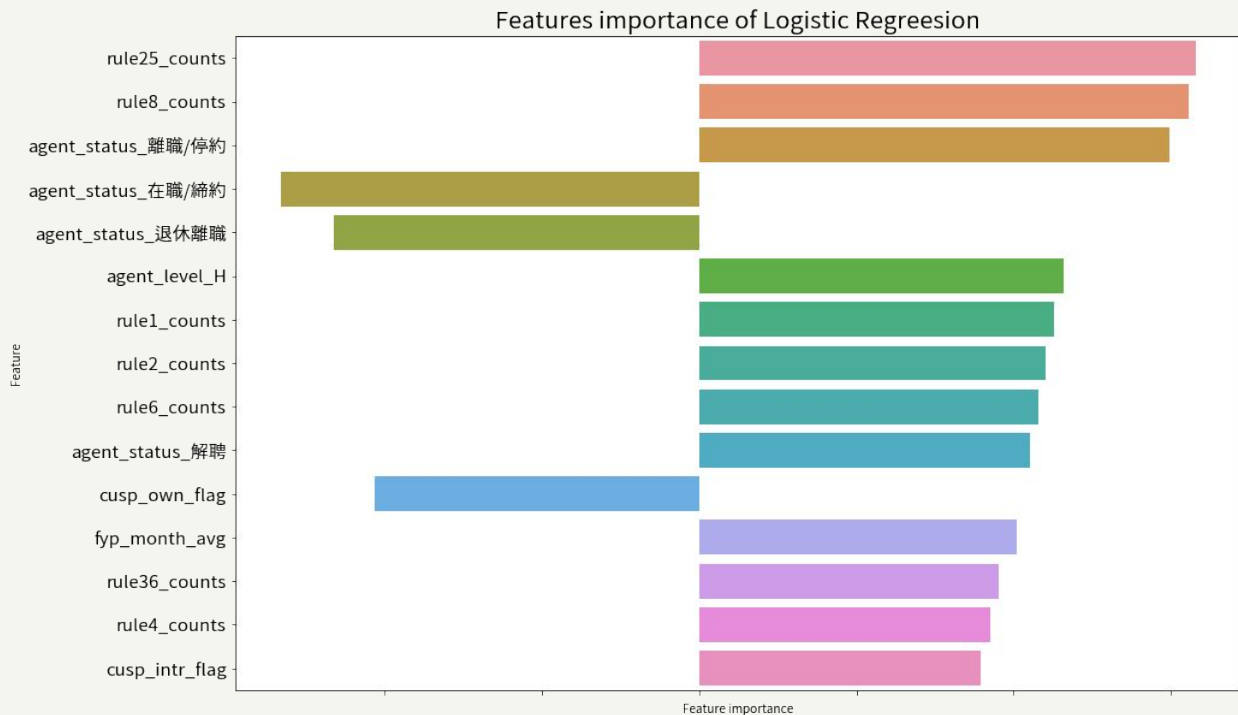
	C	class_weight
Recall	100	0:0.001,1:0.999

	Precision	Recall	F1-score	Support
0	1.00	0.85	0.92	1196
1	0.05	1.00	0.10	10





羅吉斯迴歸 調整參數-特徵重要性

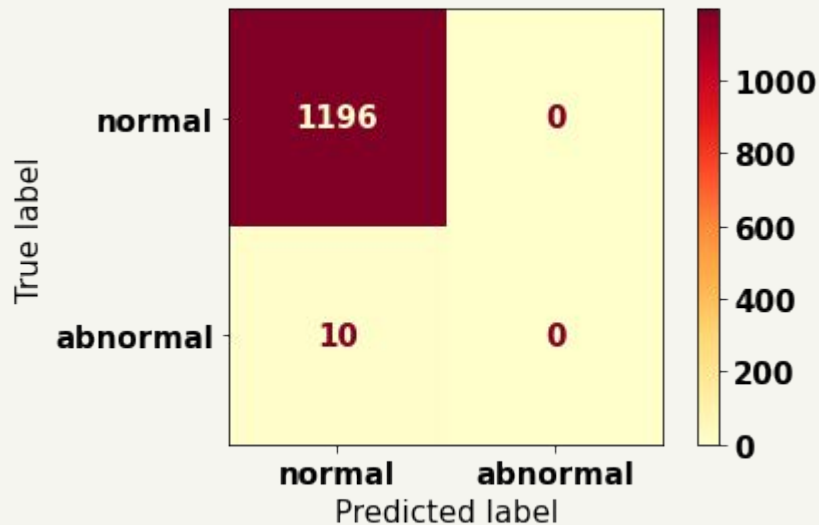


	C	class_weight
Recall	100	0:0.001,1:0.999



SVM 預設參數

	Precision	Recall	F1-score	Support
0	0.99	1.00	1.00	1196
1	0	0	0	10

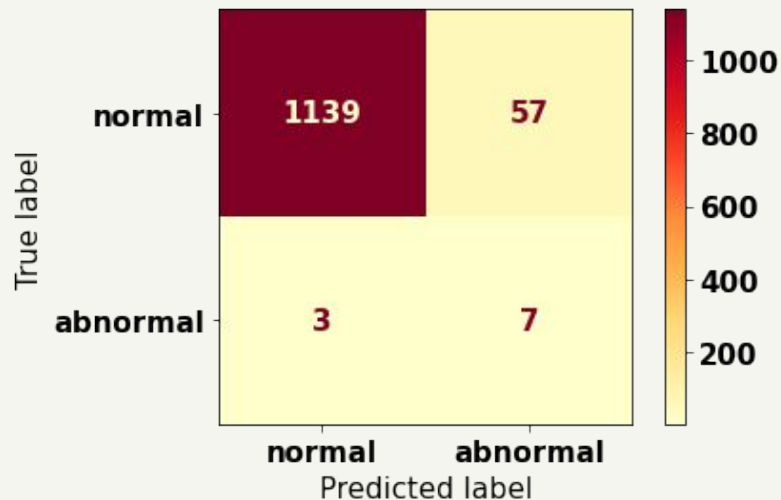




SVM 調整參數

	kernel	C	class_weight
F1	linear	1000	0:0.001,1:0.999

	Precision	Recall	F1-score	Support
0	1.00	0.95	0.97	1196
1	0.11	0.70	0.19	10

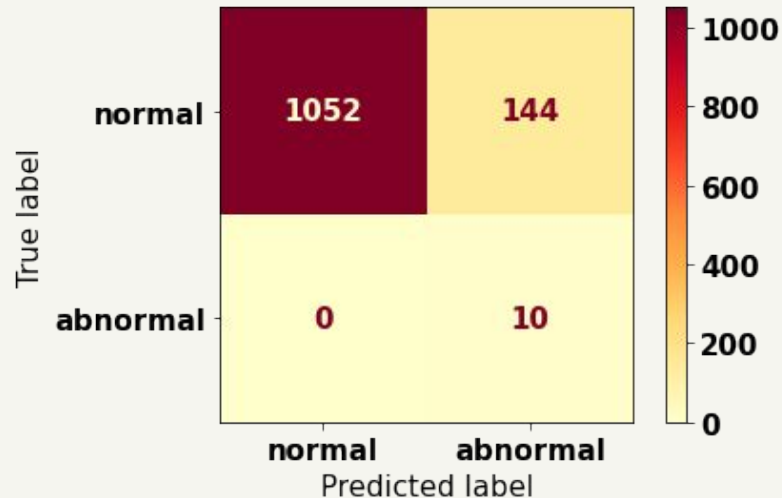




SVM 調整參數

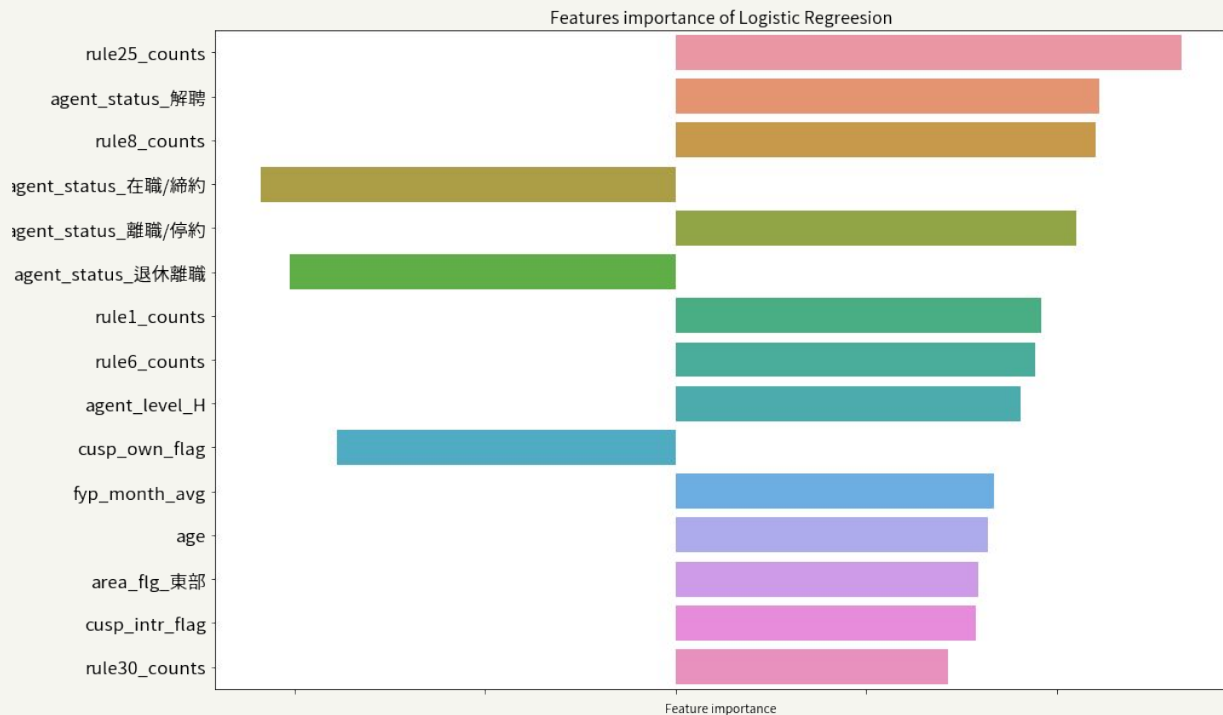
	kernel	C	class_weight
Recall	linear	20	0:0.001,1:0.999

	Precision	Recall	F1-score	Support
0	1.00	0.88	0.94	1196
1	0.06	1.00	0.12	10





SVM 調整參數-特徵重要性

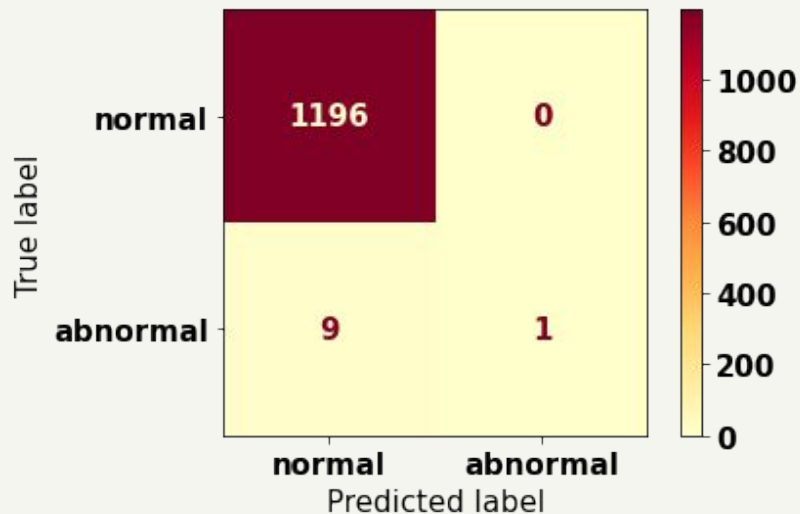


	kernel	C	class_weight
Recall	linear	20	0:0.001,1:0.999



隨機森林 預設參數

	Precision	Recall	F1-score	Support
0	0.99	1.00	1.00	1196
1	1.00	0.10	0.18	10

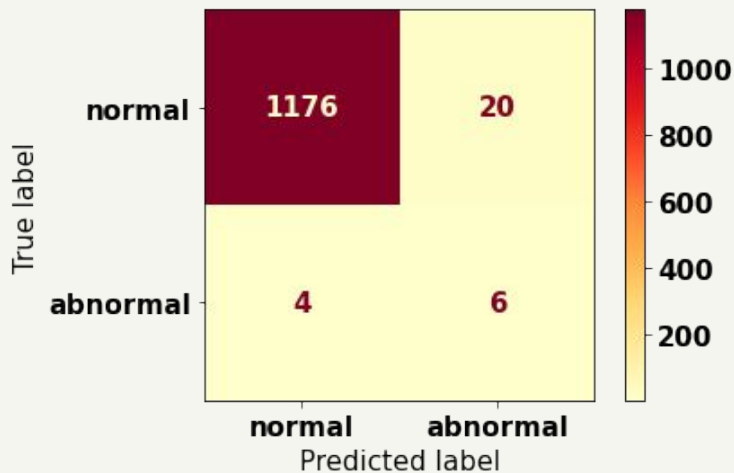




隨機森林 調整參數

	criterion	class_weight	n_estimator	max_depth
F1	entropy	0:0.001,1:0.999	400	6

	Precision	Recall	F1-score	Support
0	1.00	0.98	0.99	1196
1	0.23	0.60	0.33	10

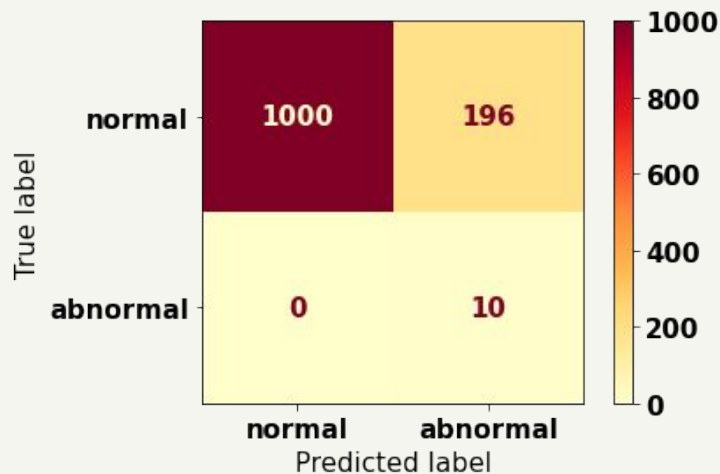




隨機森林 調整參數

	criterion	class_weight	n_estimator	max_depth
Recall	entropy	0:0.001,1:0.999	300	4

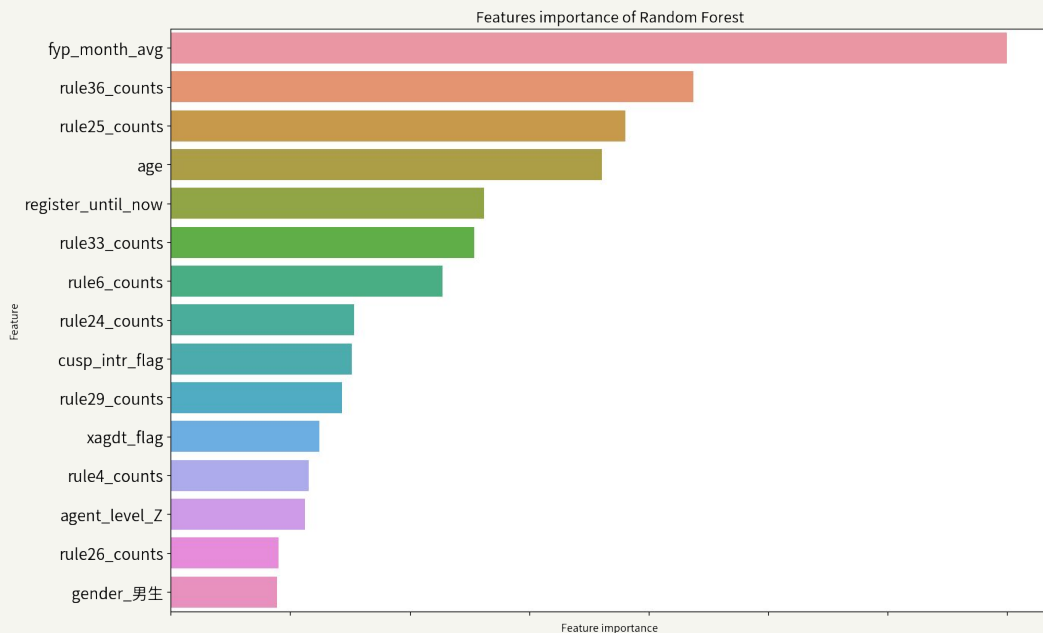
	Precision	Recall	F1-score	Support
0	1.00	0.84	0.91	1196
1	0.05	1.00	0.09	10





隨機森林 調整參數-特徵重要性

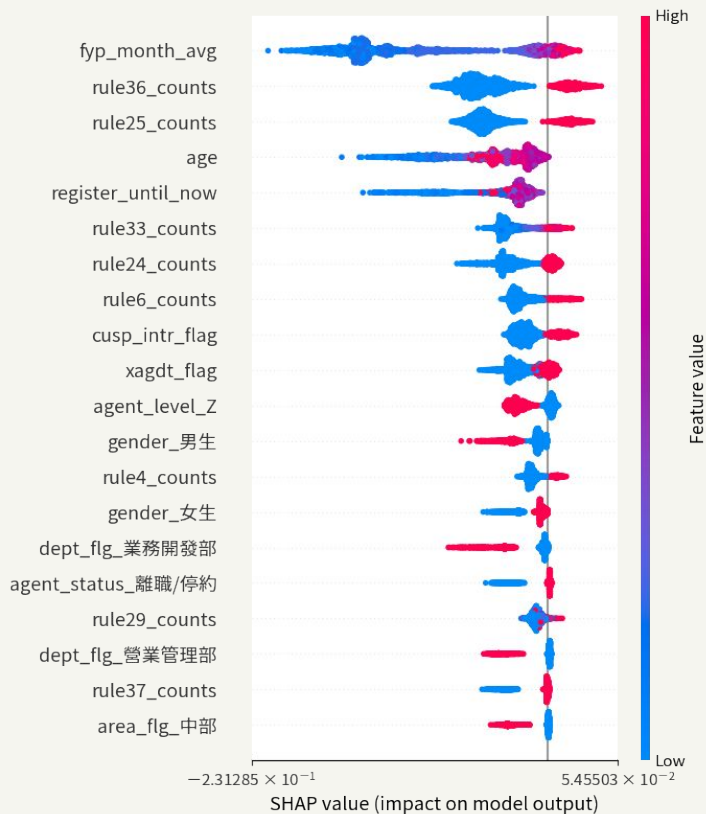
	criterion	class_weight	n_estimator	max_depth
Recall	entropy	0:0.001,1:0.999	300	4





隨機森林 調整參數-特徵重要性

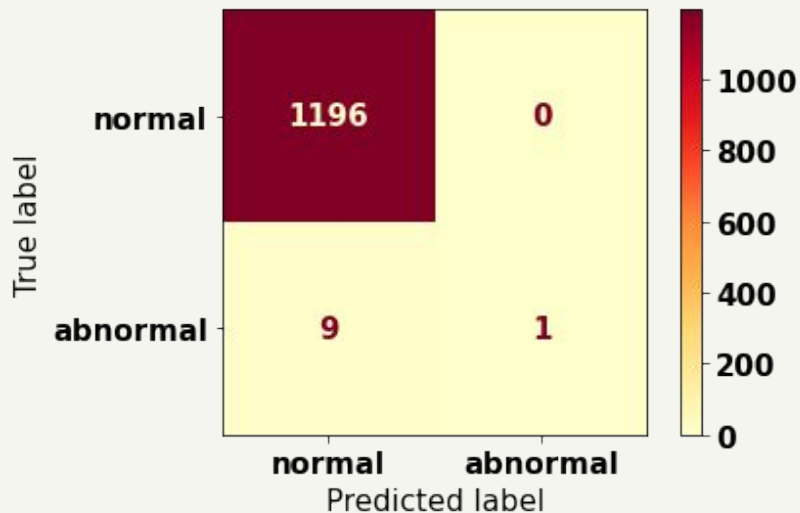
	Recall
criterion	entropy
class_weight	0:0.001,1:0.999
n_estimators	300
max_depth	4





XGBOOST 預設參數

	Precision	Recall	F1-score	Support
0	0.99	1.00	1.00	1196
1	1.00	0.10	0.18	10

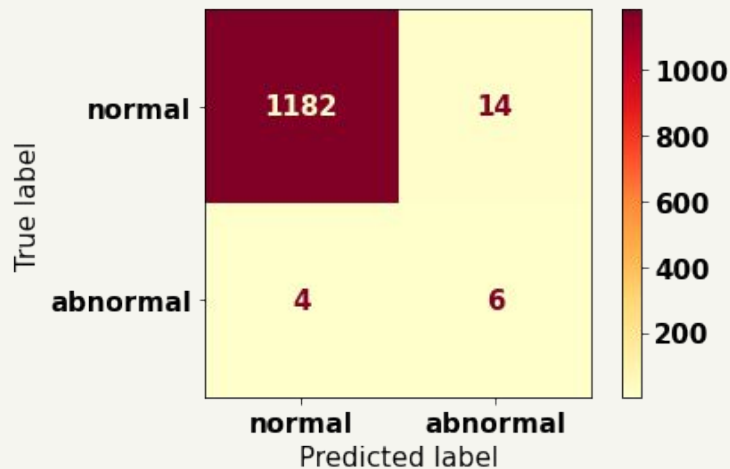




XGBOOST 調整參數

	class_weight	n_estimators	max_depth	learning_rate	scale_pos_weight
F1	0:0.001,1:0.999	80	4	0.2	2000

	Precision	Recall	F1-score	Support
0	1.00	0.99	0.99	1196
1	0.30	0.60	0.40	10

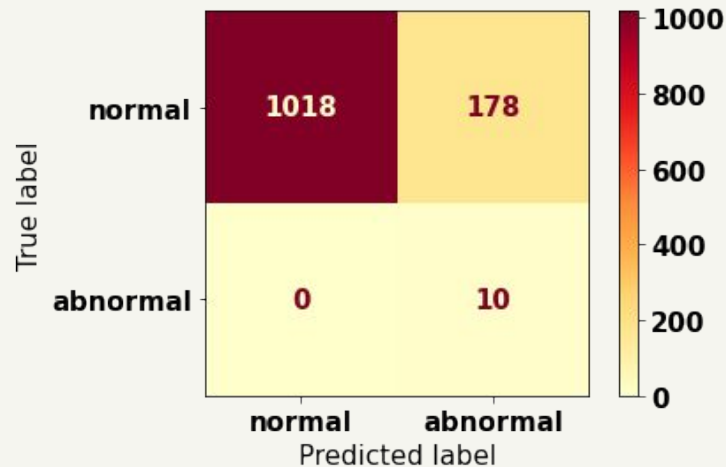




XGBOOST 調整參數

	class_weight	n_estimators	max_depth	learning_rate	scale_pos_weight
Recall	0:0.001,1:0.999	20	4	0.15	3500

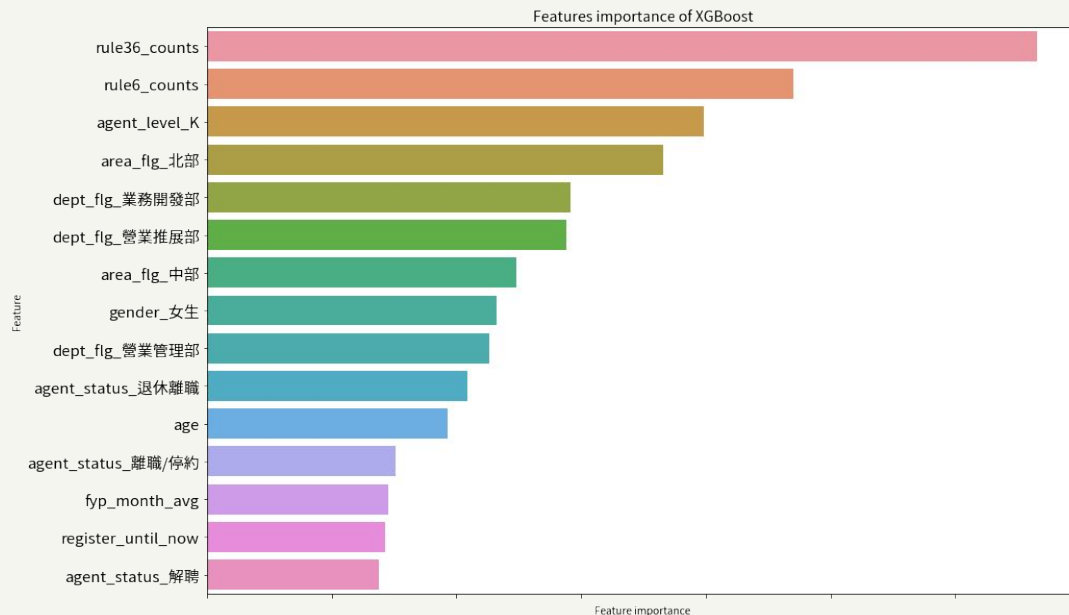
	Precision	Recall	F1-score	Support
0	1.00	0.85	0.92	1196
1	0.05	1.00	0.10	10





XGBOOST 調整參數-特徵重要性

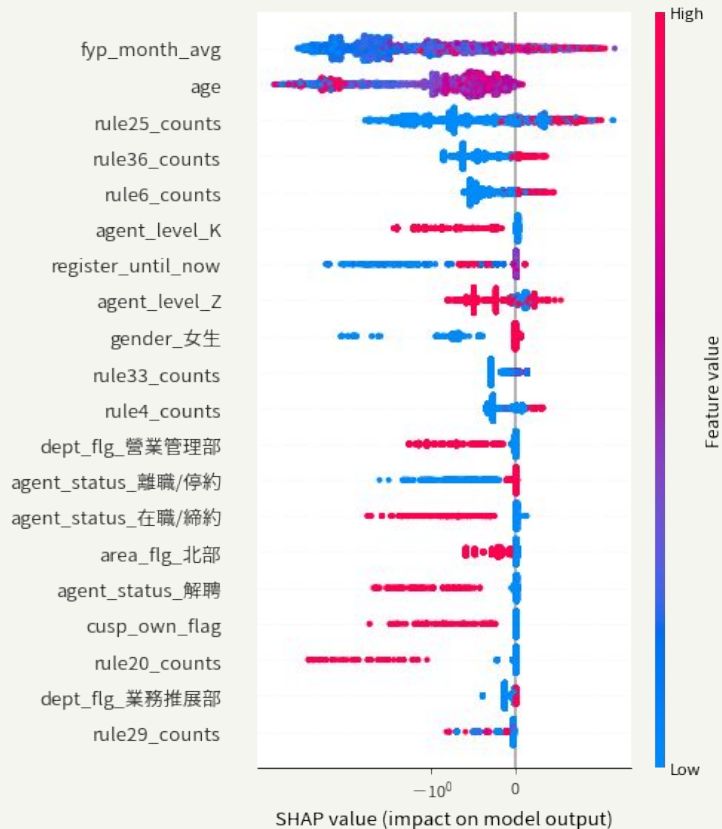
	class_weight	n_estimators	max_depth	learning_rate	scale_pos_weight
Recall	0:0.001,1:0.999	20	4	0.15	3500





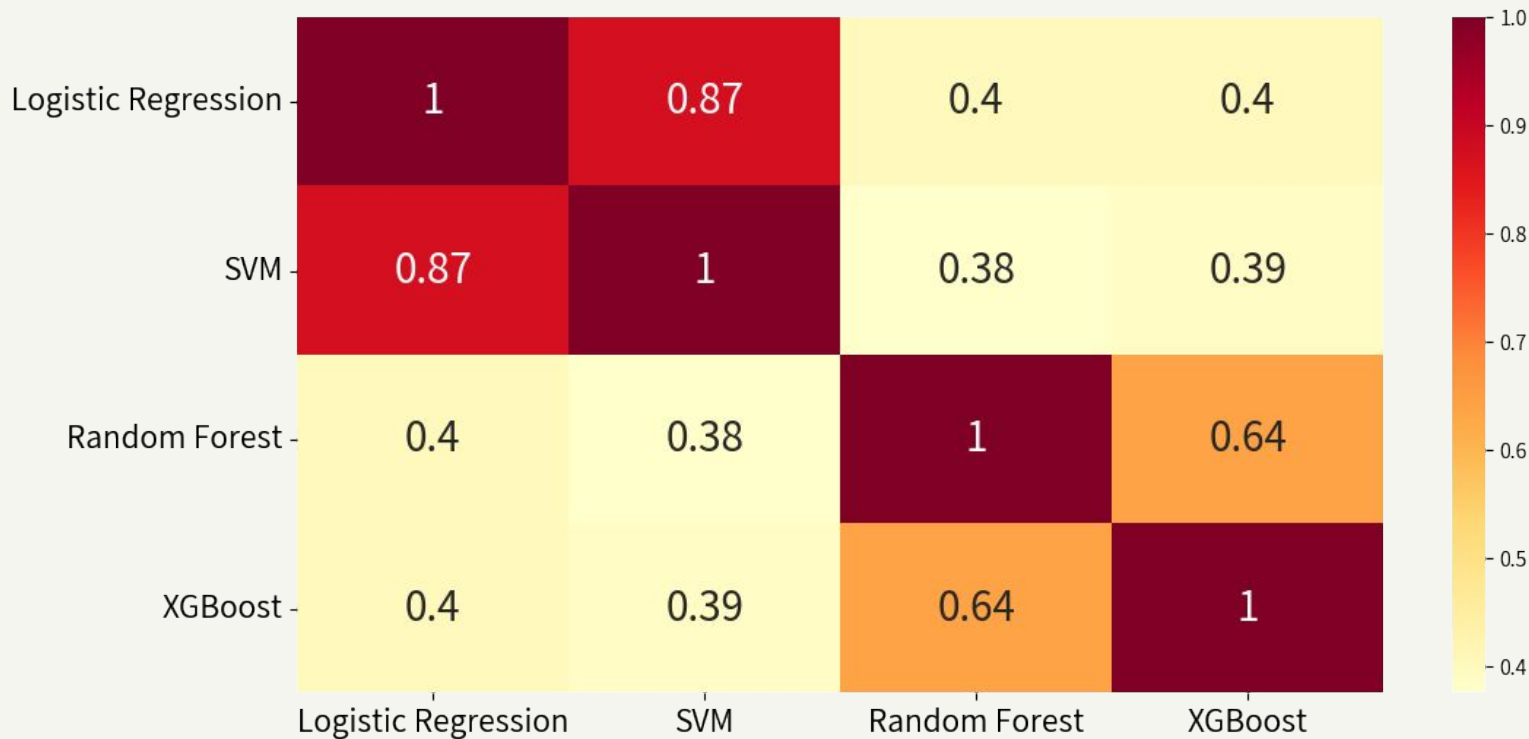
XGBOOST 調整參數-特徵重要性

	Recall
class_weight	0:0.001,1:0.999
n_estimators	20
max_depth	4
learning_rate	0.15
scale_pos_weight	3500





調整參數後模型相關性比較





模型結果

- 線性模型
- 樹狀模型
- 綜合評估

結果評估 - 線性模型

	特徵
職等	agent level H (三級區經理、代理區經理、資深區經理、二級區經理、三級區經理、代行區經理、試用區經理)
在職狀況	離職/停約 解聘
表現	月平均業績 業務員為放款介紹人註記
行為	規則 25 業務員年資大於等於2年 規則 8 收費員送金單作廢率大於8%且張數大於3 規則 1 業務員職務為業務襄理或業務副理或區經理者 規則 2 催告退回比率大於3%(地址異動) 規則 6 催告退回比率大於3%(主招) 規則 36 收費客戶支票取消禁背的件數大於5 規則 4 業務員手上招攬件的保單狀況為停效、失效、自動墊繳、繳清的比例大於 30% 規則 30 郵寄地址與居住地址不相符的保單張數占比

結果評估 - 樹狀模型

	特徵
職等	-
在職狀況	離職/停約
表現	月平均業績 業務員為放款介紹人註記 業務員欠款註記
行為	規則36 收費客戶支票取消禁背的件數大於5 規則25 業務員年資大於等於2年 規則33 要保人購買新保單當曆月及前6個曆月或後6個曆月，又辦理其他張保單的貸款或縮小保額或繳清或部贖交易或停效 規則24 業務員自身有結欠款或法扣 規則6 催告退回比率大於3%(主招) 規則4 業務員手上招攬件的保單狀況為停效、失效、自動墊繳、繳清的比例大於 30% 規則29 要保人購買新保單當曆月及前6個曆月或後6個曆月，又辦理其他張保單的解約 規則37 業務員年齡大於等於40歲

結果評估 - 綜合評估

線性、樹狀、綜合

	特徵
職等	agent level H (三級區經理、代理區經理、資深區經理、二級區經理、三級區經理、代行區經理、試用區經理)
在職狀況	離職/停約 解聘
表現	月平均業績 業務員為放款介紹人註記 業務員欠款註記
行為	規則25 業務員年資大於等於2年 規則6 催告退回比率大於3%(主招) 規則36 收費客戶支票取消禁背的件數大於5 規則4 業務員手上招攬件的保單狀況為停效、失效、自動墊繳、繳清的比例大於30% 規則8 收費員送金單作廢率大於8%且張數大於3 規則1 業務員職務為業務襄理或業務副理或區經理者 規則2 催告退回比率大於3%(地址異動) 規則30 郵寄地址與居住地址不相符的保單張數占比 規則29 要保人購買新保單當曆月及前6個曆月或後6個曆月，又辦理其他張保單的解約 規則37 業務員年齡大於等於40歲 規則33 要保人購買新保單當曆月及前6個曆月或後6個曆月，又辦理其他張保單的貸款或縮小保額或繳清或部贖交易或停效 規則24 業務員自身有結欠款或法扣



5 提案建議

- 重點關注對象
- 季度審核
- 半年度審核
- 年度審核

重點關注對象



業務員職務為**業務襄理**或**業務副理**或**區經理**者



離職/停約以及**解聘**的業務員



業務員年資大於等於**2年**，年齡大於**40歲**



季度審核

每季審核有以下表現的業務員：

- 月平均業績異常高或異常低
- 業務員為放款介紹人註記
- 業務員欠款註記

每季審核有以下行為的業務員：

- 催告退回比率大於 3%(主招)
- 收費客戶支票取消禁背的件數大於 5
- 業務員手上招攬件的保單狀況為停效、失效、自動墊繳、繳清的比例大於 30%



半年度審核

每半年審核有以下行為的業務員：

- 收費員送金單作廢率大於 8%且張數大於 3
- 催告退回比率大於 3%(地址異動)
- 郵寄地址與居住地址不相符的保單張數占比數高



年度審核

每年審核有以下行為的業務員：

- 要保人購買新保單當曆月及前 6 個曆月或後 6 個曆月，又辦理其他張保單的解約
- 要保人購買新保單當曆月及前 6 個曆月或後 6 個曆月，又辦理其他張保單的貸款或縮小保額或繳清或部贖交易或停效
- 業務員自身有結欠款或法扣



結論

稽核室

負責查核及評估第一二道防線之 內控機制是否有效運作
並適時提供建議

管理單位

管理監督及控制，協助並監督第一道防線管理風險與執行
情形

業務單位

直線單位的控管與制衡，管理營運活動所 產生的相關風險
最初的風險意識、評估與控制





結論

新光反貪腐風險評估：

1. 自評機制

在新光既有的作業風險控制自評機制下，額外新增AI特徵篩選的審核計畫

2. 評估狀況、矯正狀況

挪用業務員的特徵可能因為技術與時間的演進而有所不同，透過新資料的訓練，可以萃取出新的重要特徵，藉此優化並改善前述審核計畫，達到最好的效果



感謝聆聽