

# 三招帶你打進 Top 1% (StarRingChild, Ranked #11)

黃冠豪 (Wilson Huang)  
Data Analyst, Rayark Inc.  
github.com/KuanHaoHuang/

## 第一招：以 bacno 為中心的近遠期特徵 (+0.05)

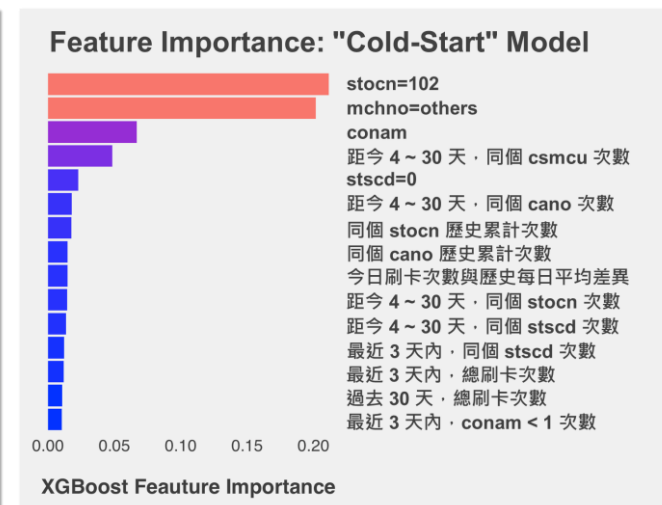
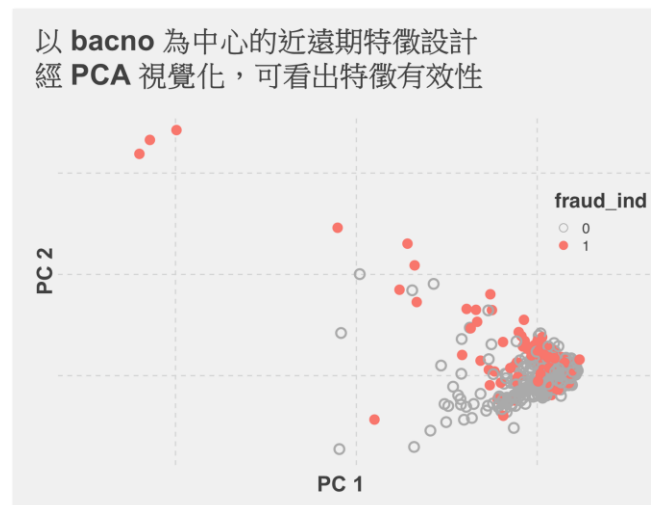
### 統計：

有多次盜刷紀錄的 bacno 之中，  
其中 85% 的人曾發生過相隔 3 天之內的連續盜刷

### 特徵工程：

- 該 bacno 近 3 天使用「這張卡」次數
- 該 bacno 距今第 4 ~ 30 天使用「這張卡」次數
- 該 bacno 所有使用「這張卡」累計次數

※「      」用 stscd、csmcu、contp 等特徵代換



## 第二招：Cold-Start & Warm-Start 兩種模型 (+0.03)

| 模型類型                  | Testing Data 內 bacno 選取 | 模型是否納入 fraud_ind 資訊    | Validation Data 選擇方式 |
|-----------------------|-------------------------|------------------------|----------------------|
| Cold-Start (baseline) | 在 Training Data 沒有刷卡紀錄  | 否。<br>模型會誤認為過去沒有盜刷紀錄。  |                      |
| Warm-Start            | 在 Training Data 曾有刷卡紀錄  | 是。<br>例如：近 3 天發生盜刷總次數。 |                      |

## 第三招：「稍笨」超參數(+0.01)

| 驗證集 F1 Score | 與 0.68 差距 |                       |
|--------------|-----------|-----------------------|
| 0.73         | 0.05      | Too Much Overfitting  |
| 0.71         | 0.03      |                       |
| 0.69         | 0.01      |                       |
| 0.68         | 0.00      | Ensemble (簡單平均)       |
| 0.67         | -0.01     |                       |
| 0.65         | -0.03     |                       |
| 0.63         | -0.05     | Too Much Underfitting |
| 0.58         | -0.10     |                       |