

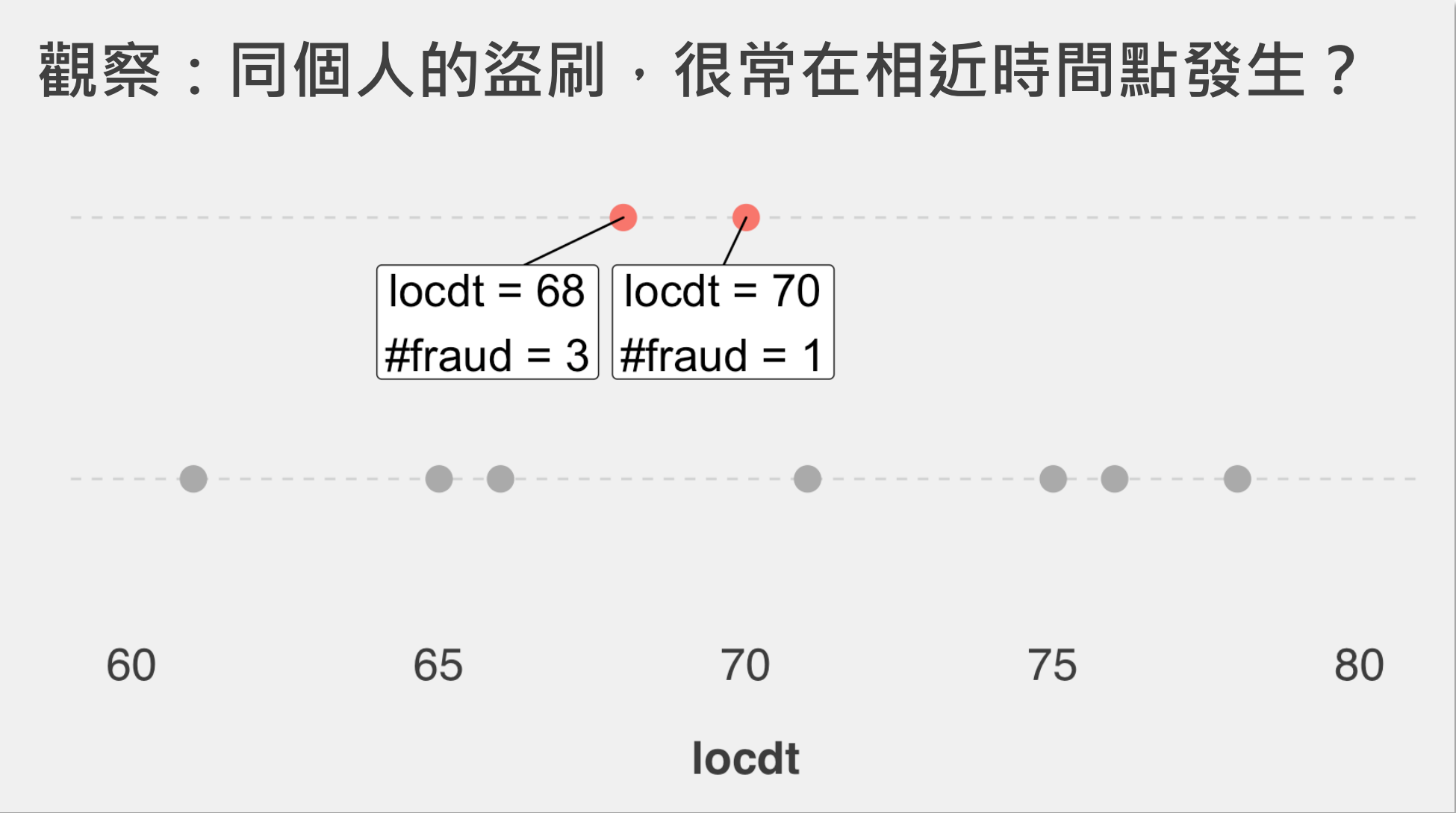


三招帶你打進 Top 1%

StarRingChild · Ranked #11

第一招：以 bacno 為中心的近遠期特徵

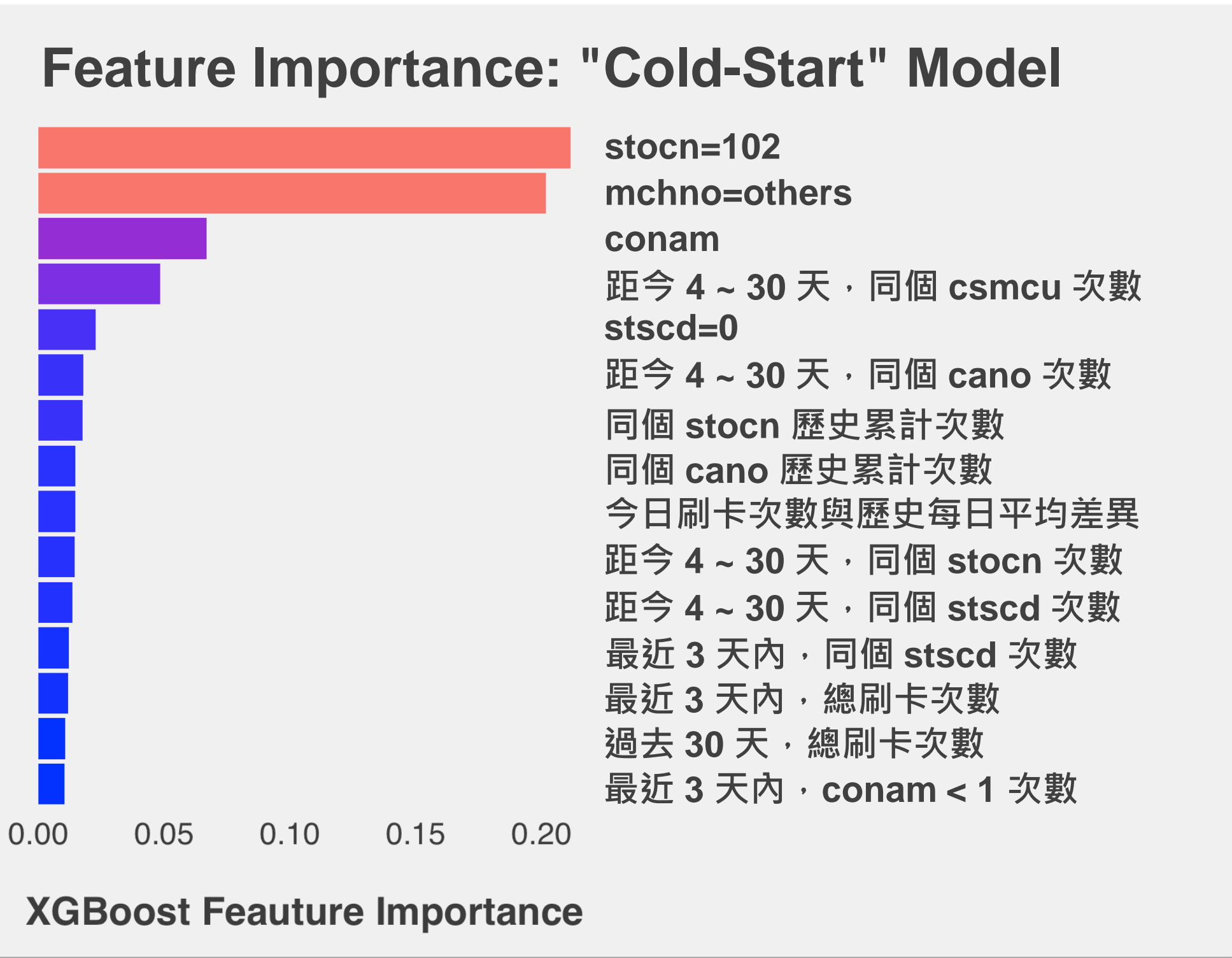
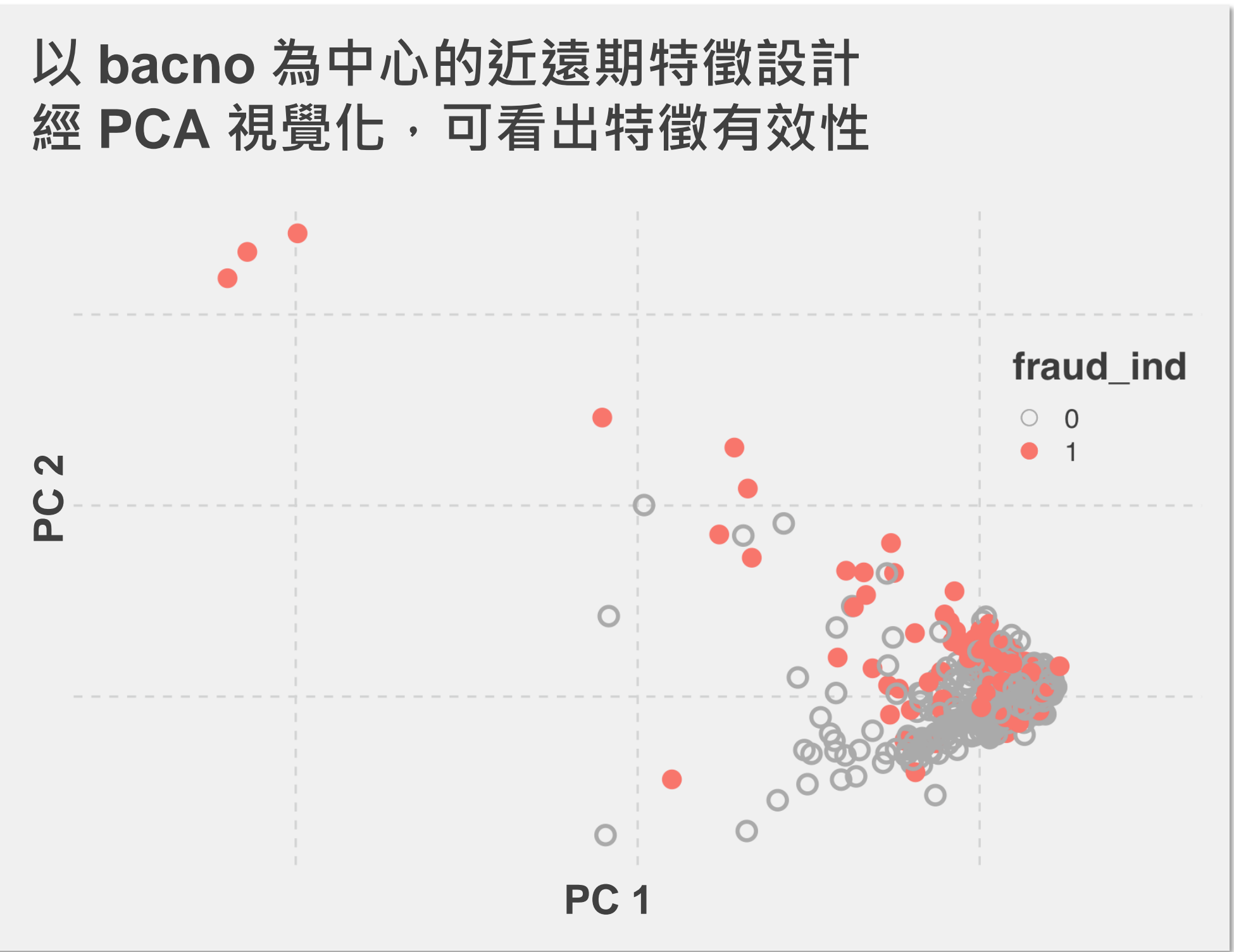
F1_Score +0.05



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

F1_Score +0.03

Testing Data 中，4.4% 的 bacno 在 Training Data 也有資料

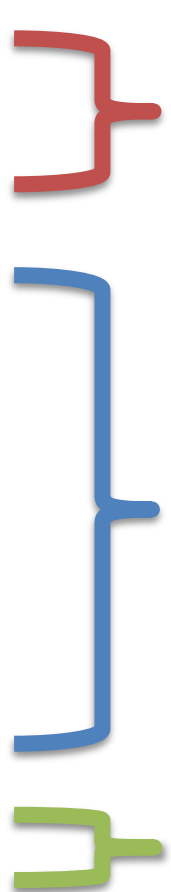
| 模型類型 | Testing Data 內 bacno 選取 | 模型是否納入 fraud_ind 資訊 | Validation Data 選擇方式 |
|-----------------------|-------------------------|--|----------------------|
| Cold-Start (baseline) | 在 Training Data 沒有刷卡紀錄 | 否。 模型會誤認為過去沒有盜刷紀錄。 | |
| Warm-Start | 在 Training Data 曾有刷卡紀錄 | 是。 例如：該 cano 過去曾被盜刷次數，近 3 天發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

約 600 筆 Cold-Start 模型認為非盜刷的資料點，Warm-Start 模型更積極預測為盜刷

第三招：「稍笨」超參數、納入 Ensemble

F1_Score +0.01

| 隨機選擇超參數 | 驗證集 F1 Score | 與 0.68 差距 |
|---------|--------------|-----------|
| XGB 1 | 0.73 | 0.05 |
| XGB 2 | 0.71 | 0.03 |
| XGB 3 | 0.69 | 0.01 |
| XGB 4 | 0.68 | 0.00 |
| XGB 5 | 0.67 | -0.01 |
| XGB 6 | 0.65 | -0.03 |
| XGB 7 | 0.63 | -0.05 |
| XGB 8 | 0.58 | -0.10 |



Too Much Overfitting

Ensemble (Simple Average)

Too Much Underfitting

黃冠豪 (Wilson Huang)
Data Analyst, Rayark Inc.



R Code 與海報：
github.com/KuanHaoHuang/