



衛 生 福 利 部  
Ministry of Health and Welfare  
促進全民健康與福祉



# 兒少保護案件之精準派案

2018夏 資料英雄

2018/10/28

# CONTENTS

- 01 Team Introduction**
- 02 Problem Statement**
- 03 Analysis Methods**
- 04 Data Lists**
- 05 Findings**
- 06 Conclusions**

A large, abstract graphic on the left side of the slide features a grid of blue squares in various shades of light blue, medium blue, and dark blue. These squares are arranged in a staggered, non-uniform pattern that creates a sense of depth and movement.

01 | Team  
Introduction

# 01 | Team Introduction

## Mentors



Johnson



Jamie

## 資料英雄們



張赫麟



藍浩



吳奇倫



林瑜軒



張家豪



潘怡均



吳又建



陳姿君

A large, abstract graphic on the left side of the slide features a grid of blue squares in various shades of light blue, medium blue, and dark blue. The squares are arranged in a staggered, non-uniform pattern that creates a sense of depth and movement.

## 02 | Problem Statement

# 02 | Problem Statement

## 聚焦：一類兒少保護案件

分類 ➤ 家內 VS. 家外 VS. 其他

家內案件占35.29%：

討論聚焦在俗稱的家暴案件，對象限定在未滿18歲兒少。

分級 ➤ 緊急(4日) VS. 非緊急(30日)

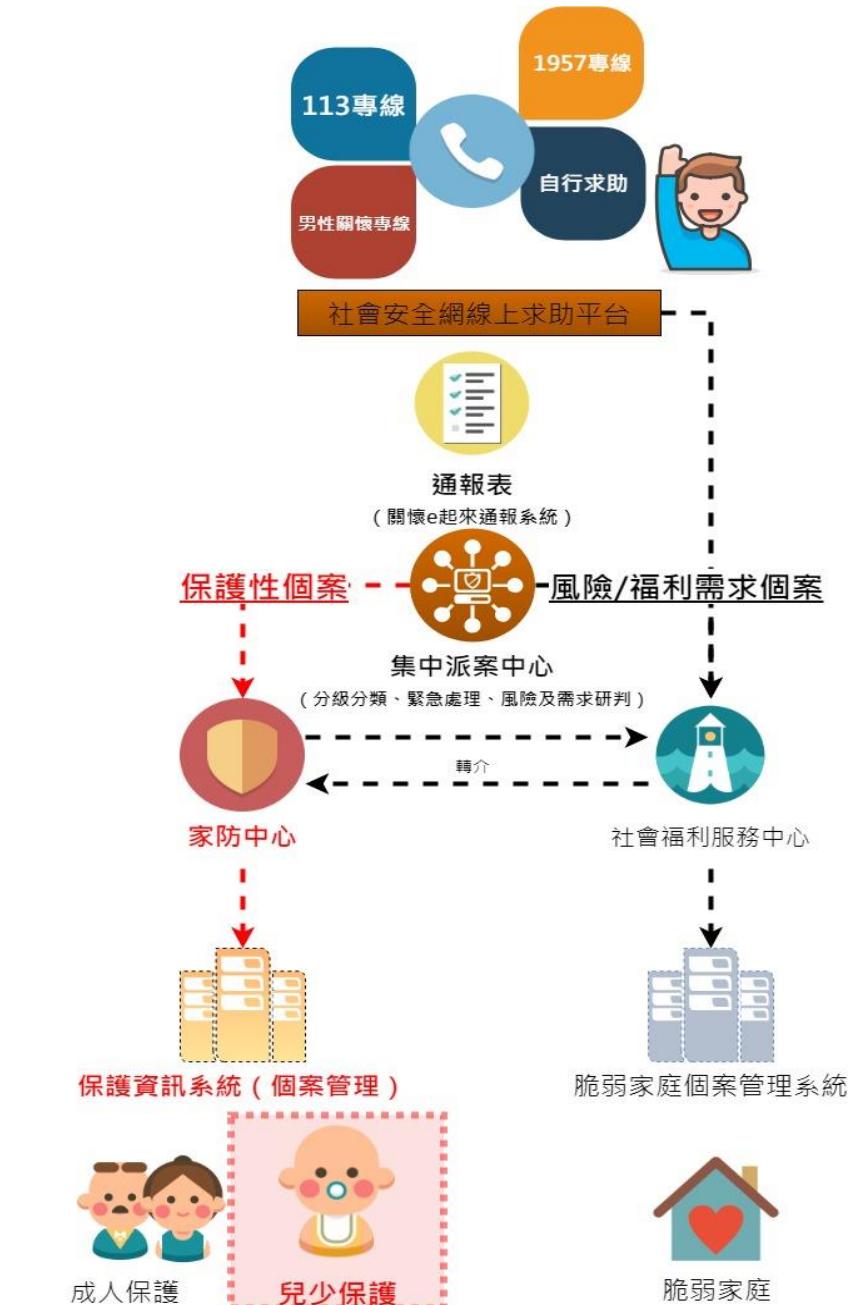
非緊急案件占93%：

亟需自現行通報及分級分類評估端所得資訊，以判斷案件處理的優先順序與投入程度。

服務 ➤ 共案（保護司+心口司） VS. 非共案（僅保護司服務）

非共同列管案件：

與共案有類似特徵，卻未有保護司、心口司同時介入服務，需有名冊預先提示兒少保社工提高服務密度及敏感度。



## 02 | Problem Statement

### 主題1：兒少保護通報開案預測

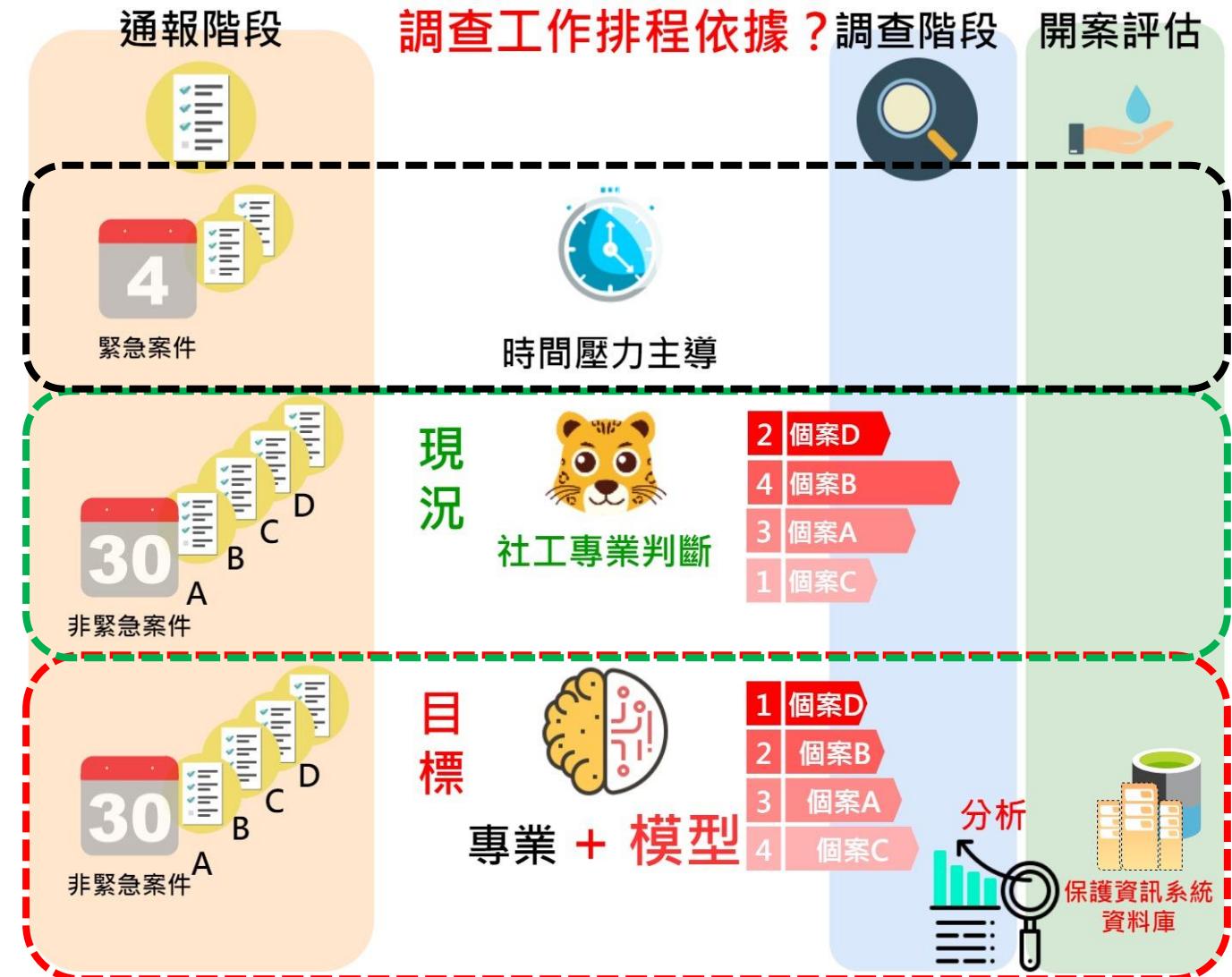
透過通報表資訊預測該案開案與否，提供兒少保社工進入調查階段前案件優先順序參考。

**緊急案件：**

分秒必爭，立即給予保護、安置或其他必要處置。

**非緊急案件：**

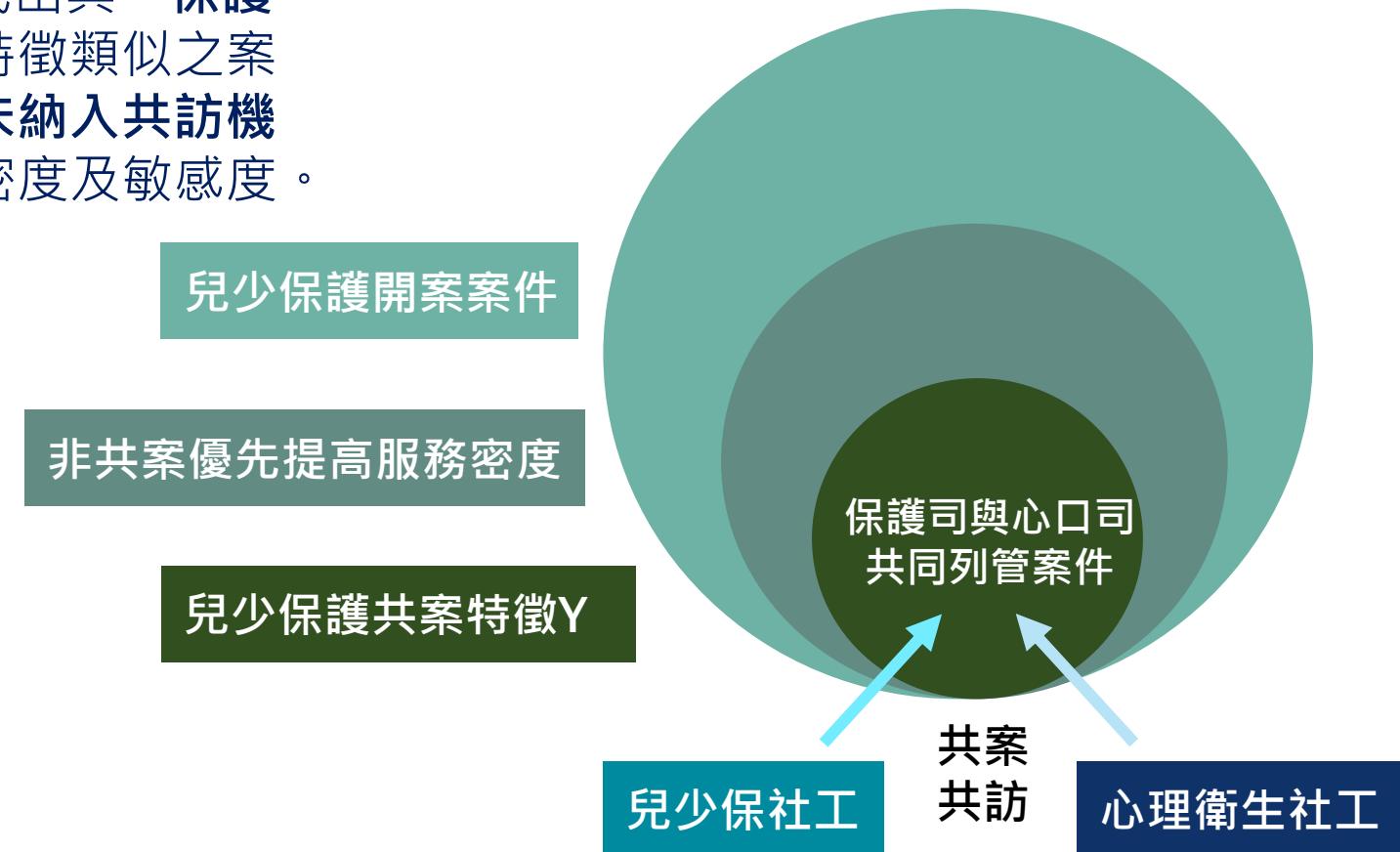
- 現況：調查順序仰賴專業判斷
- 目標：導入模型預測開案機率



## 02 | Problem Statement

### 主題2：共案特徵預測及訪視優先次序分析

兒少保護開案案件中，透過調查報告找出與「**保護司與心口司共同列管案件（共案）**」特徵類似之案件，但未有心理衛生社工介入服務（**未納入共訪機制**），預先提示兒少保社工提高服務密度及敏感度。



A large, abstract graphic on the left side of the slide features a grid of blue squares of varying shades and sizes, creating a pixelated or mosaic-like effect that tapers towards the top right.

# 03 | Analysis Methods

### 主題1：兒少保護通報開案預測

- 關聯分析
- 預測模型(Tree-Based Models)
- 案情陳述NLP

### 主題2：共案特徵預測及訪視優先次序分析

- 預測模型(XGBoost)

A large, abstract graphic on the left side of the slide features a grid of blue squares of varying shades, creating a pixelated or mosaic effect that tapers towards the top right.

# 04 | Data Lists

### 主題1：兒少保護通報開案預測

使用106年一類通報表資料(開案及未開案)及關係人調查表，共20,165筆。

### 主題2：共案特徵預測及訪視優先次序分析

使用105和106年一類被害人父母調查報告資料，共13,621筆。(開案)

A large, abstract graphic on the left side of the slide features a grid of blue squares of varying shades, creating a pixelated or mosaic effect. The squares are arranged in a roughly triangular shape pointing towards the center of the slide.

# 05 | Findings

### 主題1：兒少保護通報開案預測

1. 通報表關聯分析
2. 通報表開案預測模型
3. 通報表案情陳述NLP

### 主題2：共案特徵預測及訪視優先次序分析

1. 兒少保開案共案特徵預測

### 主題1：兒少保護通報開案預測

- 1.通報表關聯分析
- 2.通報表開案預測模型
- 3.通報表案情陳述NLP

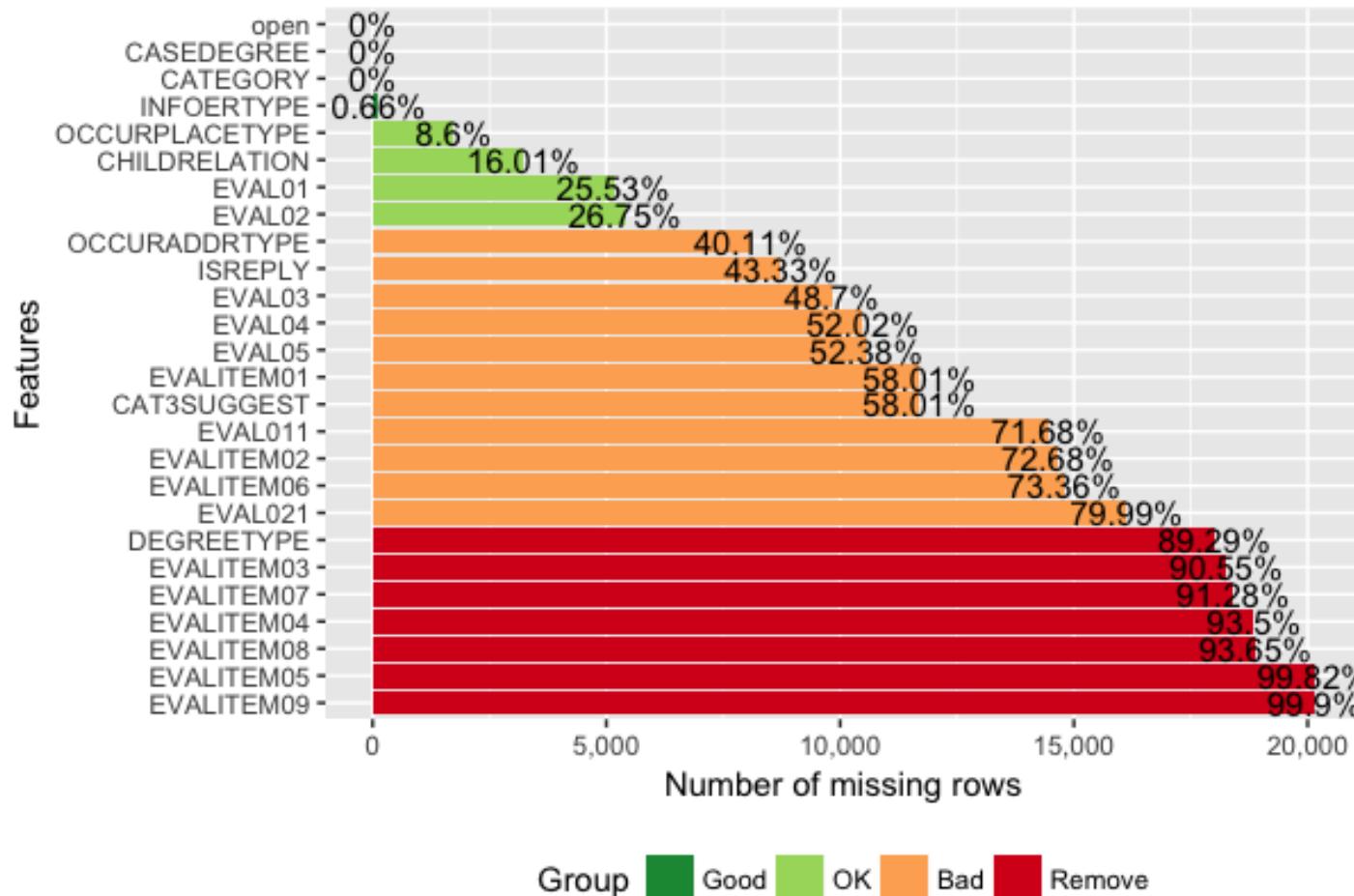
### 主題2：共案特徵預測及訪視優先次序分析

- 1.兒少保開案共案特徵預測

### 1. 通報表關聯分析

- 資料缺失性嚴重
- 使用CART決策樹進行遺漏值預測
- 關聯分析結果

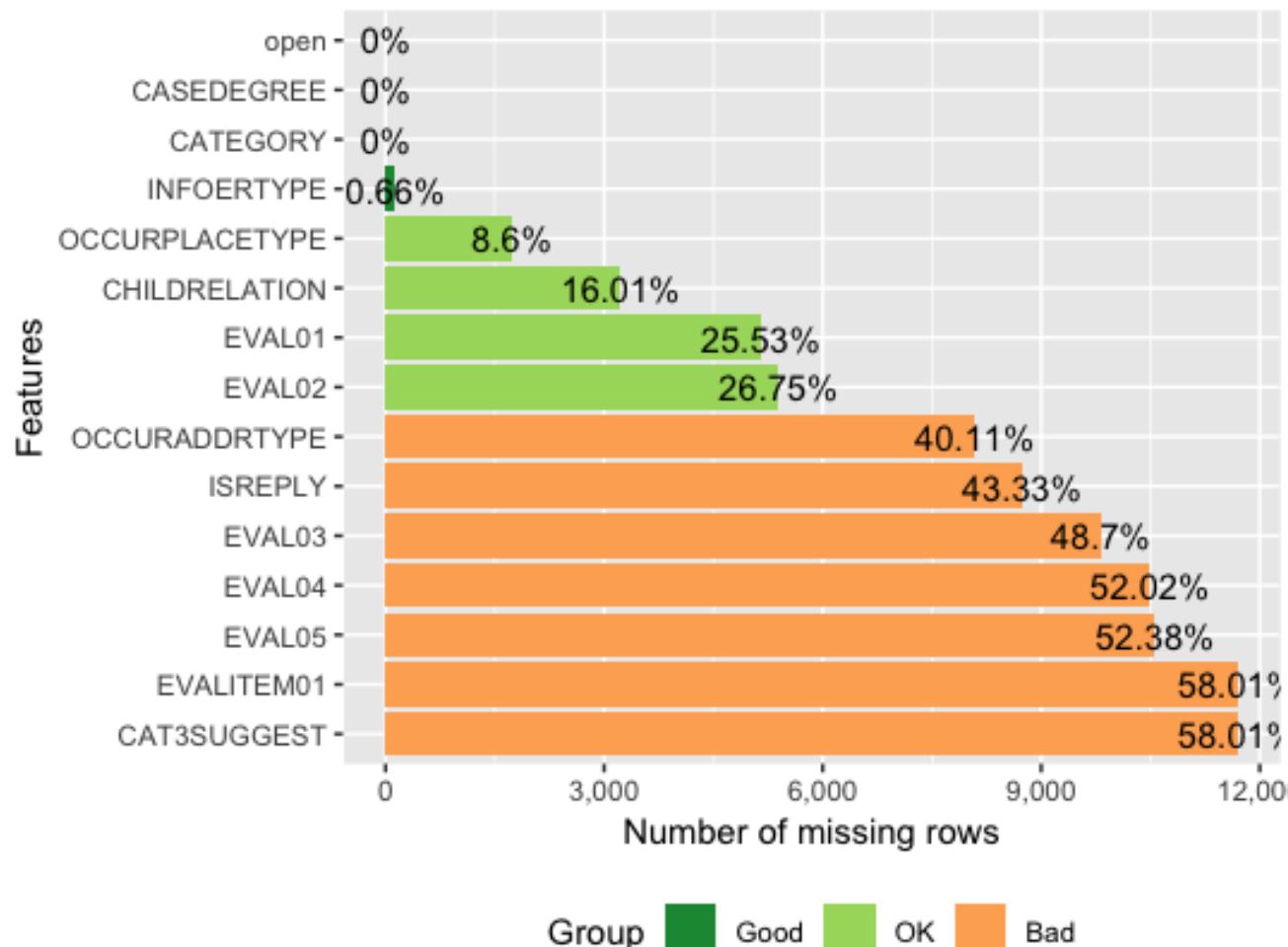
## 1. 通報表關聯分析



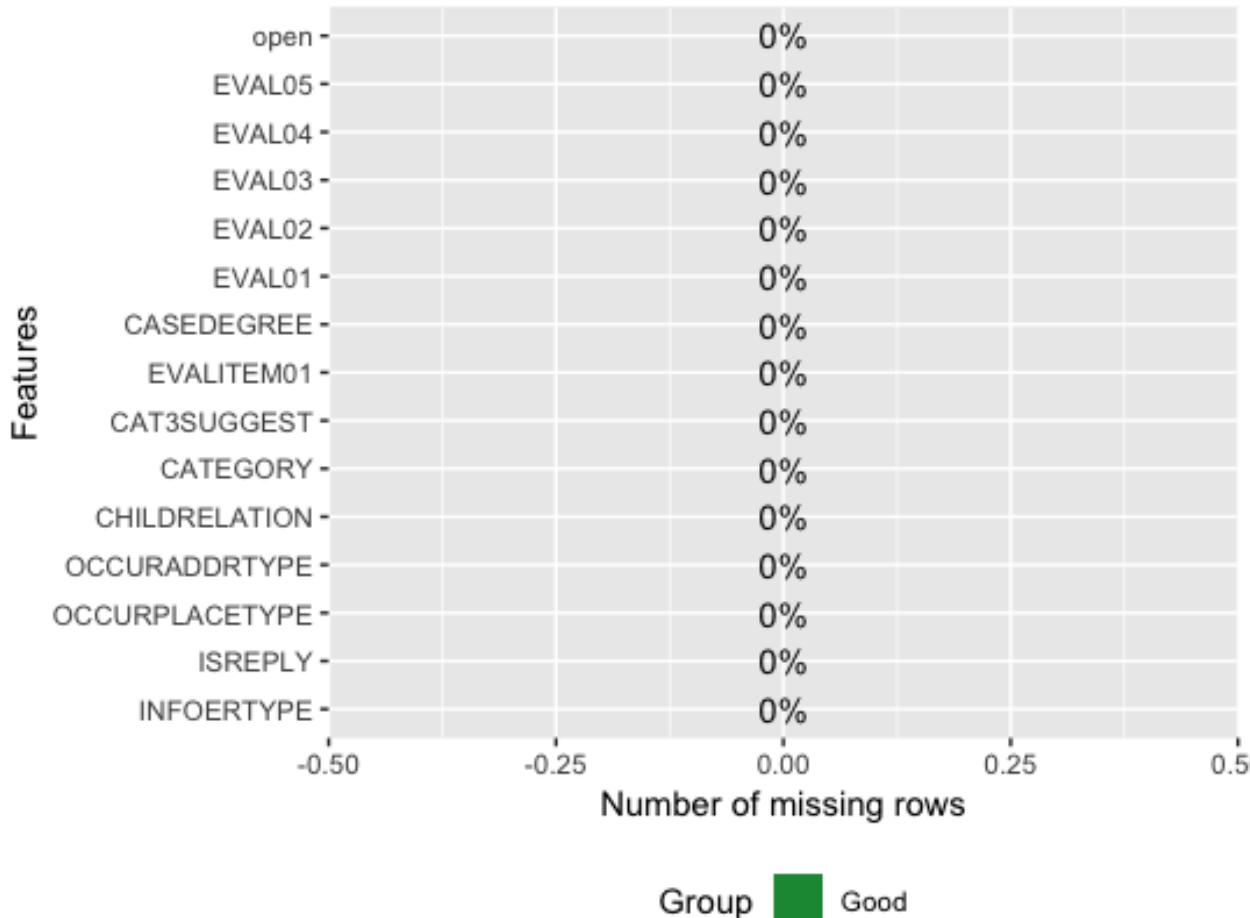
資料缺失性嚴重

20,165 obs.

## 1. 通報表關聯分析



## 1. 通報表關聯分析



## 1. 通報表關聯分析

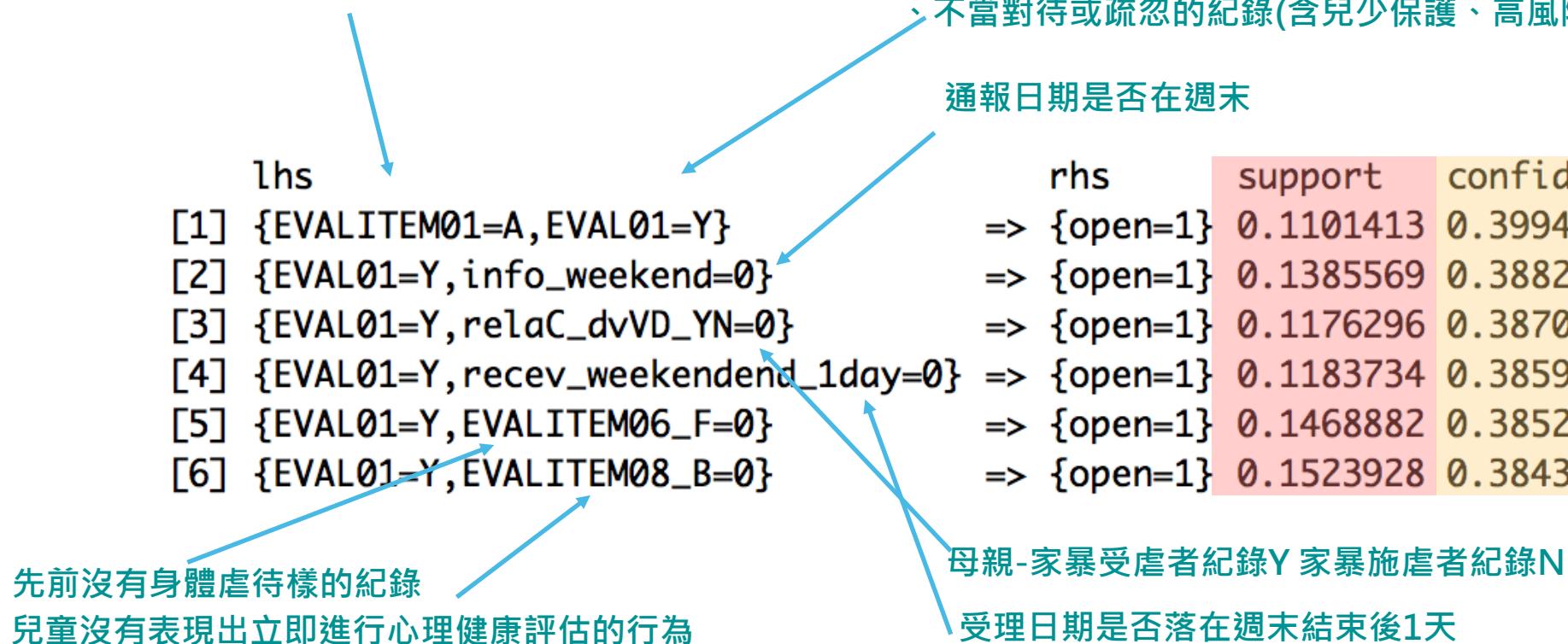
SUPPORT: 組合出現的比例

CONFIDENCE: 組合中通報的比例

LIFT: 預測結果時能比隨機發生的機會好多少

平均開案率：25%

EVALITEM01: 第三評估指標第一層 ( A: 身體虐待樣態 )      EVAL01: 兒少先前是否有遭受父母、監護人、實際照顧之人或其他家庭成員虐待  
 、不當對待或疏忽的紀錄(含兒少保護、高風險家庭通報)？ ( Y: 是 )

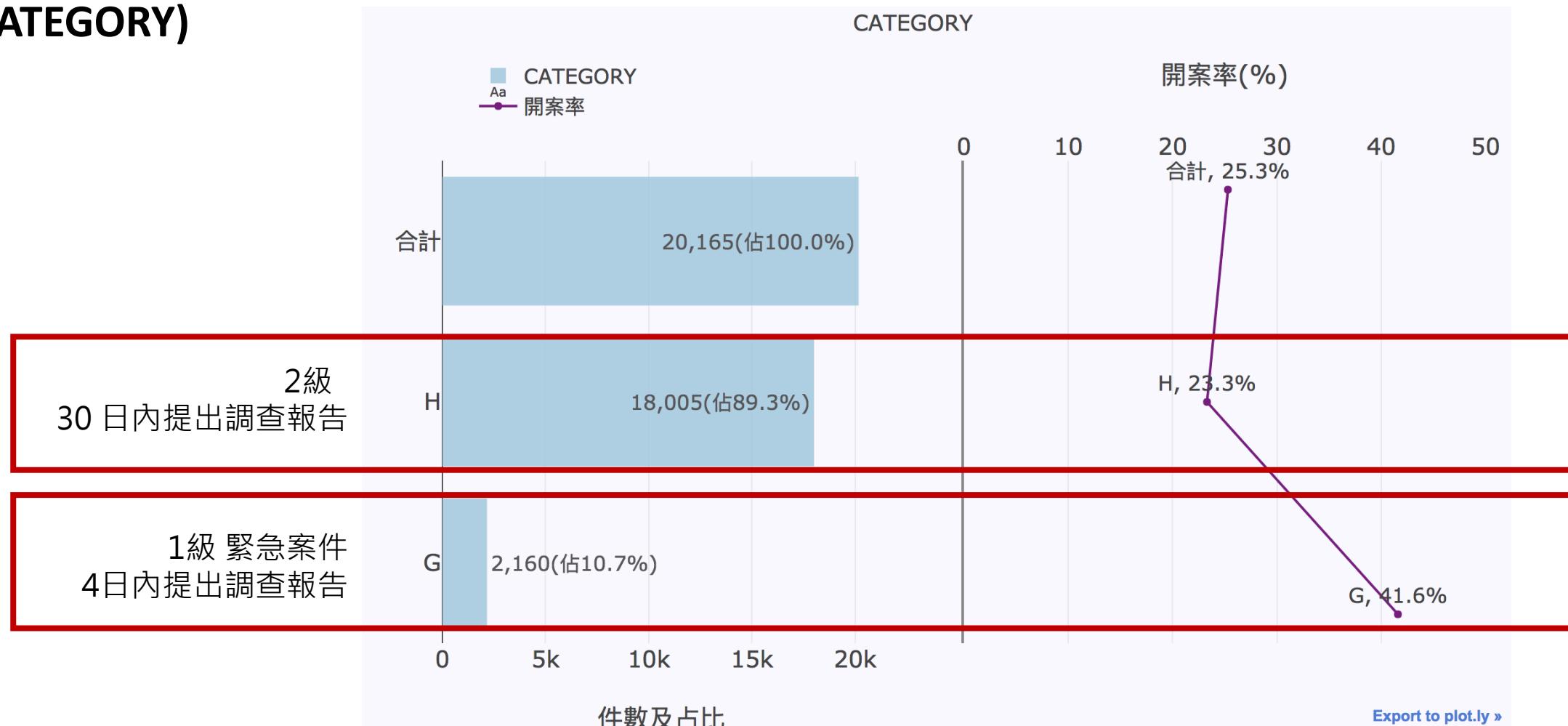


rhs	support	confidence	lift	count
=> {open=1}	0.1101413	0.3994604	1.580055	2221
=> {open=1}	0.1385569	0.3882713	1.535796	2794
=> {open=1}	0.1176296	0.3870126	1.530818	2372
=> {open=1}	0.1183734	0.3859337	1.526550	2387
=> {open=1}	0.1468882	0.3852758	1.523948	2962
=> {open=1}	0.1523928	0.3843652	1.520346	3073

### 2.通報表開案預測模型

- 通報表評估類別
- 模型開發結果 - 兩個子模型
- 模型重要特徵變數
- 模型評估
- 小結

## 2. 通報表開案預測模型

通報表評估類別  
(CATEGORY)

## 2. 通報表開案預測模型

## 模型開發結果 – 隨機森林 (Random Forest)

Model	Accuracy	Recall	Precision	F1 - Score
G Model (通報後 4天內須處理)	0.826	0.737	0.809	0.771
H Model (通報後30天內須處理)	0.900	0.574	0.988	0.726

Random Forest – H model

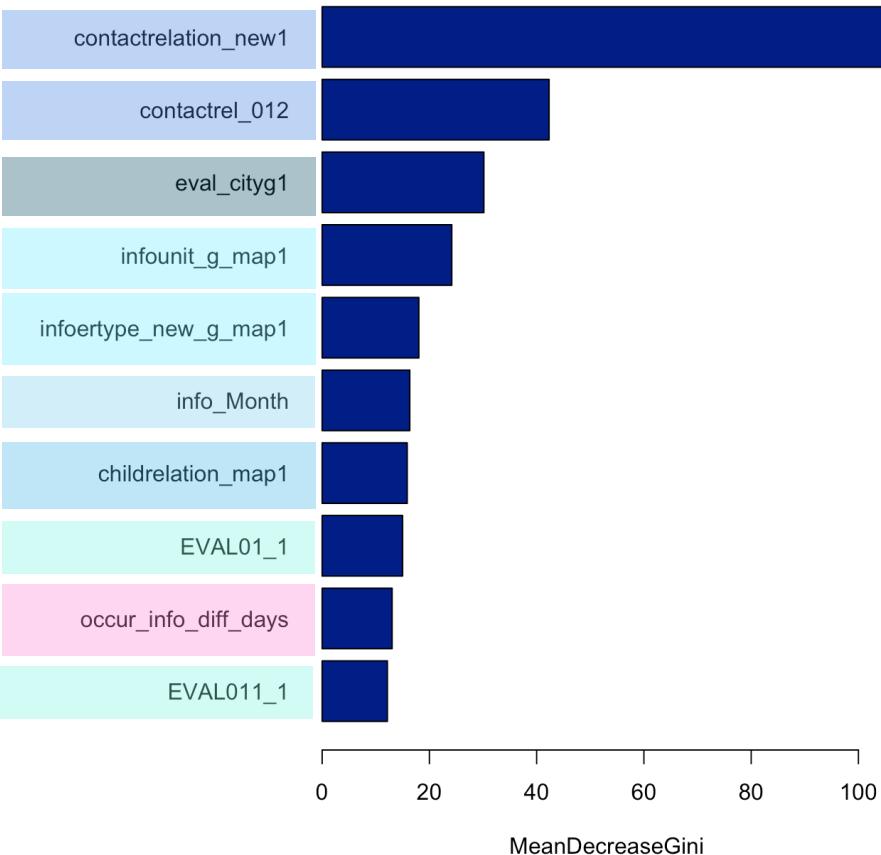
測試樣本 (test)		預測		合計
		未開案	開案	
實際	未開案	4,148	9	4,157
	開案	530	715	1,245
合計		4,678	724	5,402

- Accuracy 預測正確/全部樣本數  
 $= (4148+715) / 5402 = 0.900$
- Recall 正確預測開案/實際開案  
 $= 715 / 1245 = 0.574$
- Precision 正確預測為開案/預測開案  
 $= 715 / 724 = 0.988$
- F1-Score = 
$$\frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

F1-Score 數值介於0~1  
(1代表最完美的Recall及Precision)

## 模型重要特徵變數

1級 緊急案件  
4日內提出調查報告  
**Feature Importance - G model**



安全聯絡人關係是否有值  
其他可聯絡親友關係欄位是否有值

評估者縣市區域

通報單位類型  
通報人員類型

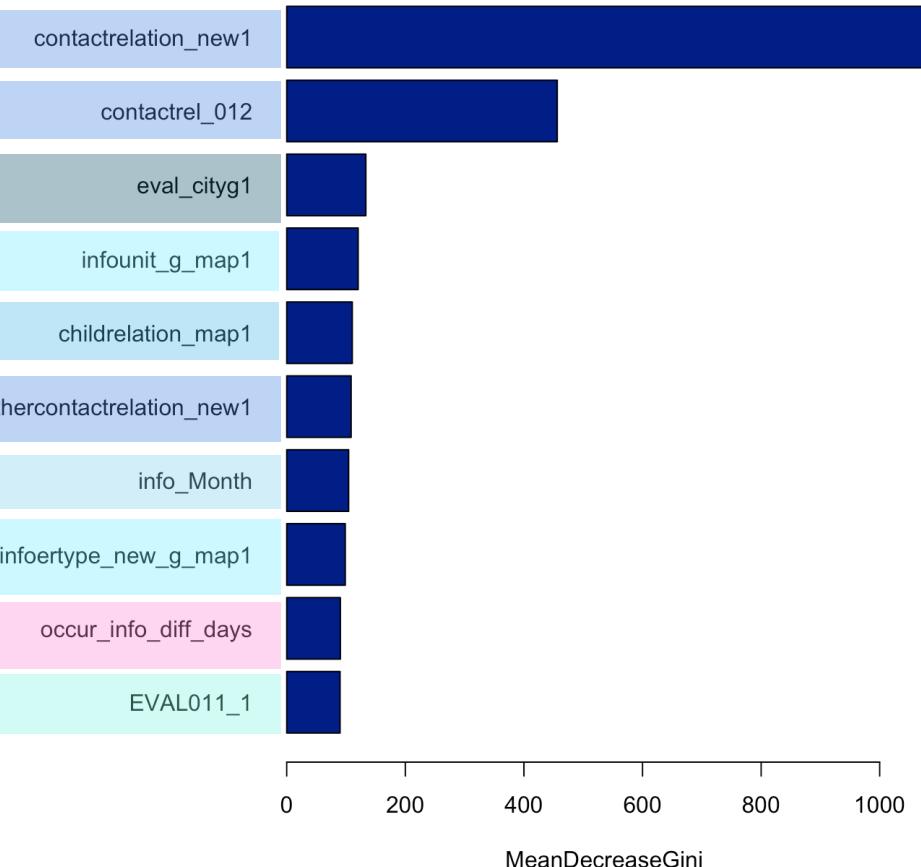
兒少保兩造關係

通報月份

發生時間與通報時間相差日數

兒少先前是否有遭受父母、監護人、  
實際照顧之人或其他家庭成員虐待、  
不當對待或疏忽的紀錄

2級 非緊急案件  
30日內提出調查報告  
**Feature Importance - H model**



G Model  
是/否  
發生次數

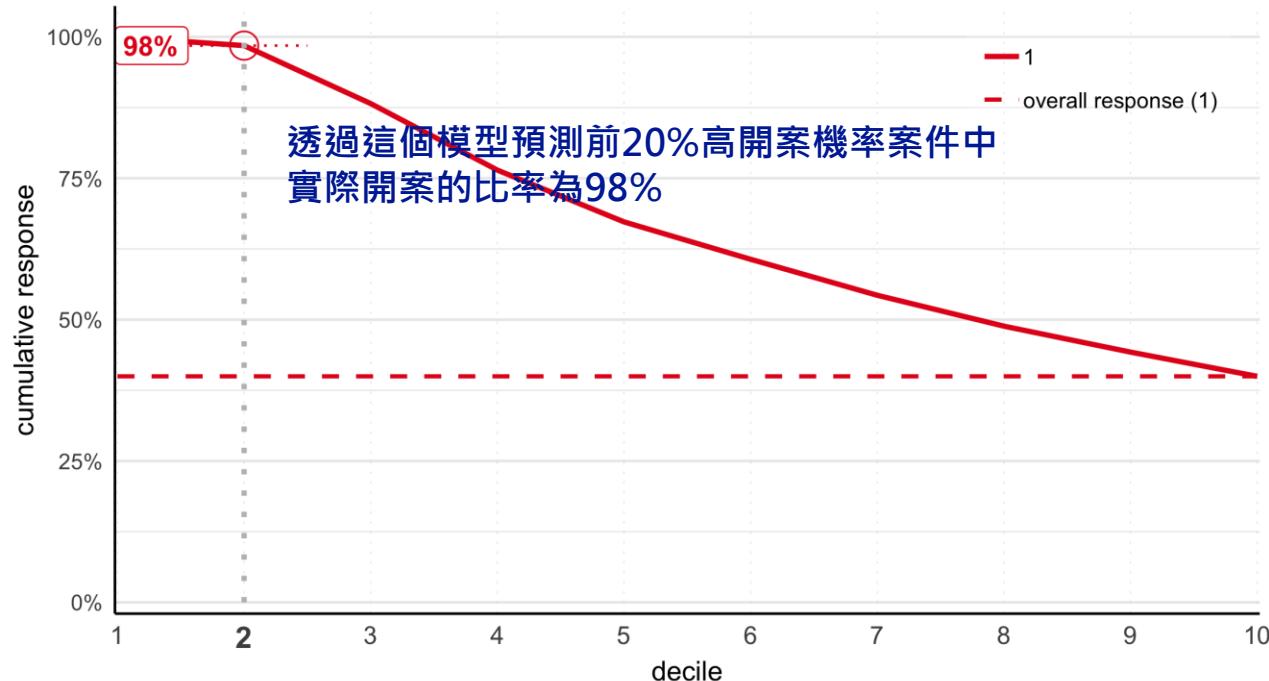
H Model  
發生次數

## 2. 通報表開案預測模型

## 模型評估 – G model

Y軸：  
開案比率

**Cumulative response**  
model: Random Forest & dataset: test data" & target class: 1 (開案)

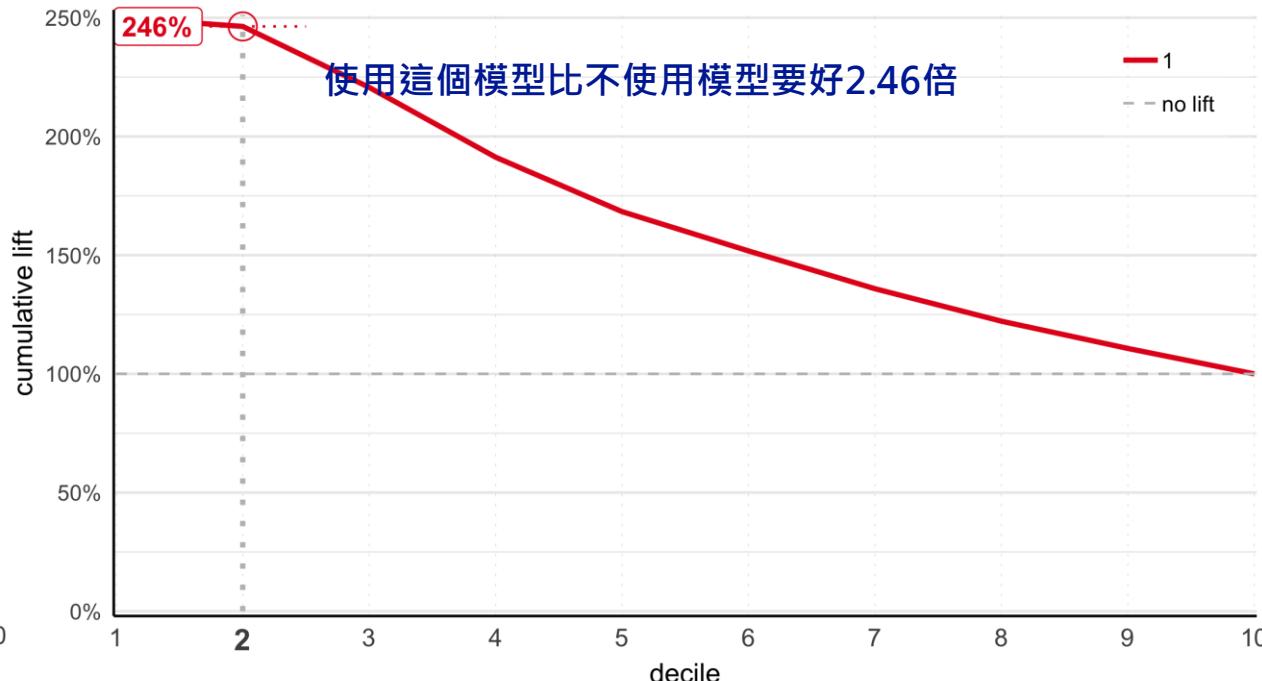


When we select deciles 1 until 2 according to model Random Forest in dataset test data the % of 1 cases in the selection is 98%.

Y軸：

使用模型

比不使用模型比較 model: Random Forest &amp; dataset: test data" &amp; target class: 1 (開案)

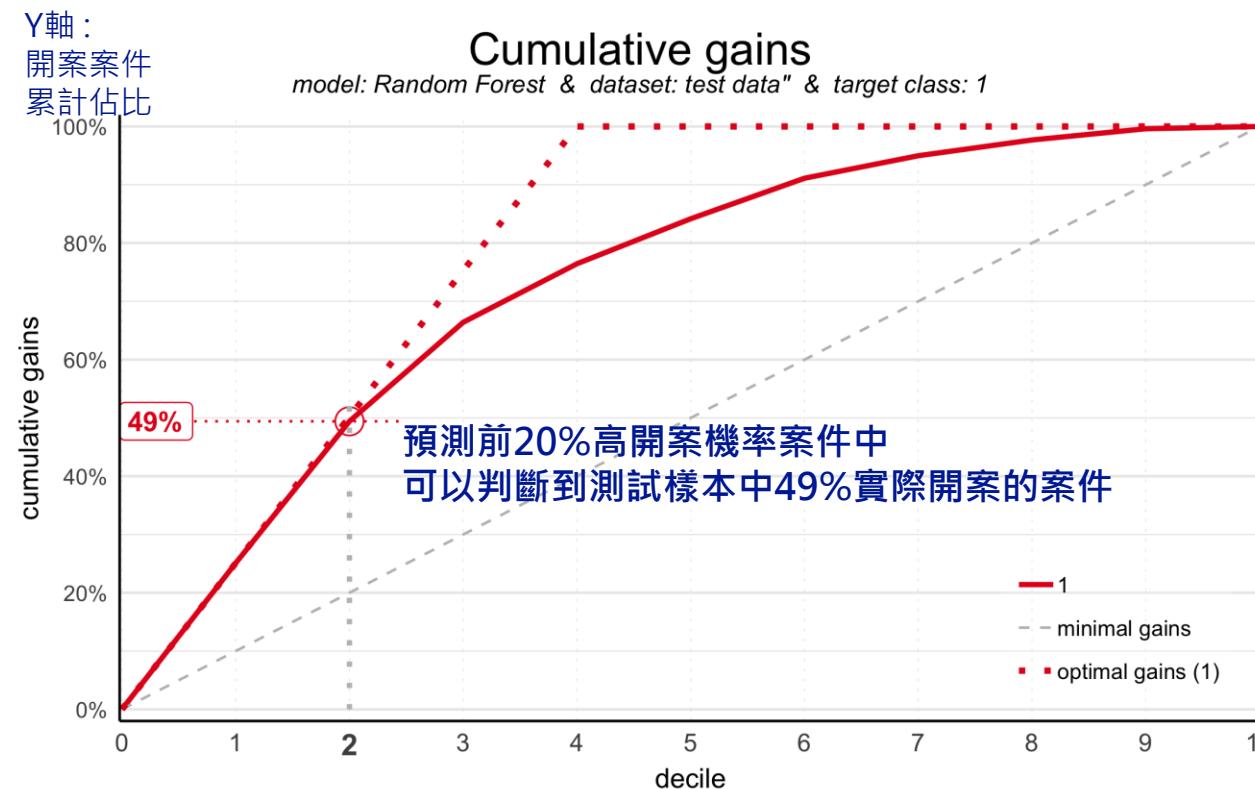
**Cumulative lift**

When we select 20% with the highest probability according to model Random Forest in test data, this selection for 1 cases is 2.5 times better than selecting without a model.

X軸(decile)：將測試樣本依模型預測為1(開案)之機率自高到低排序並切為10等分 (即1為預測機率最高的10%測試樣本)

## 2. 通報表開案預測模型

## 模型評估 – G model

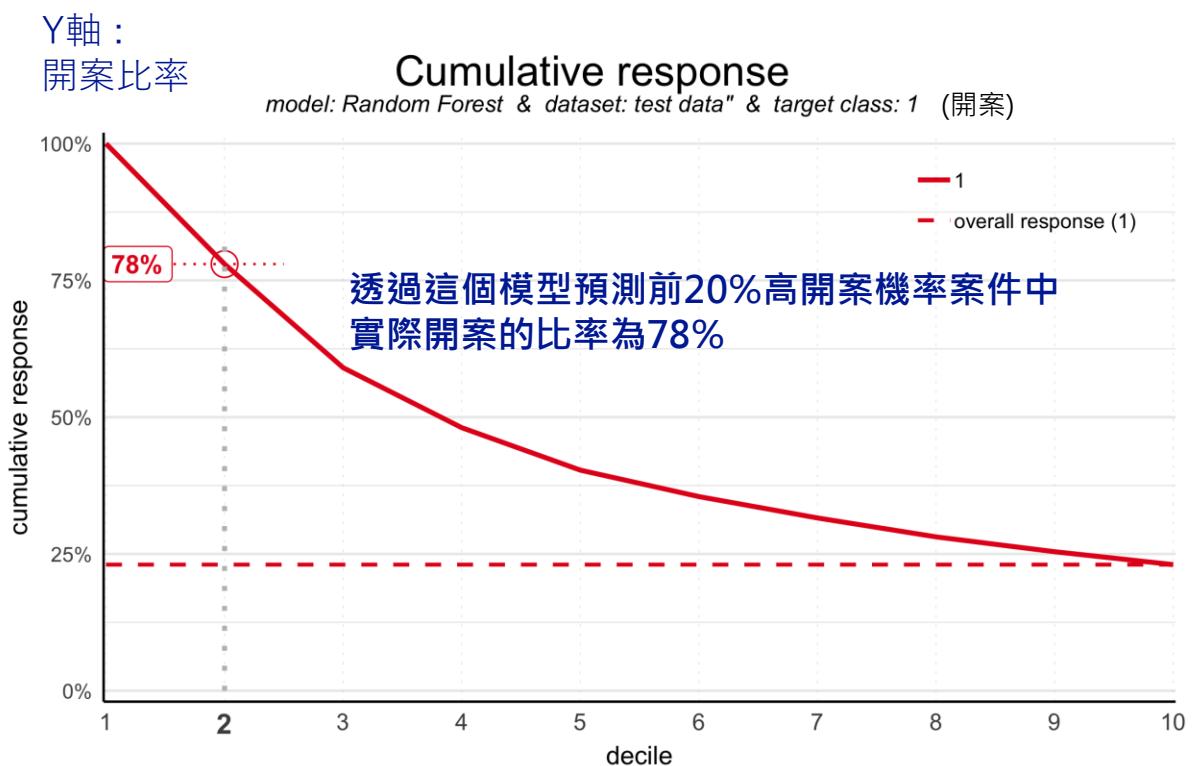


When we select 20% with the highest probability according to Random Forest, this selection holds 49% of all 1 cases in test data.

X軸(decile)：將測試樣本依模型預測為1(開案)之機率自高到低排序並切為10等分 (即1為預測機率最高的10%測試樣本)

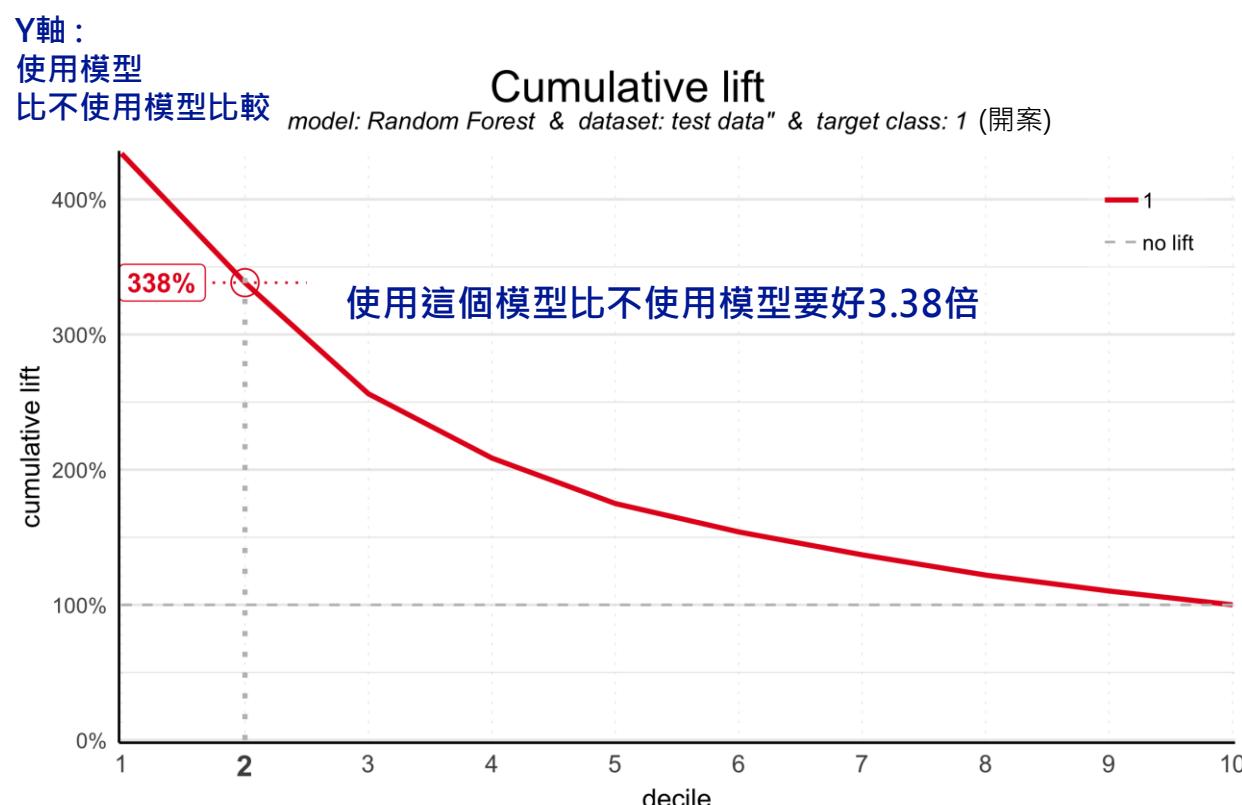
## 2. 通報表開案預測模型

## 模型評估 – H model



When we select deciles 1 until 2 according to model Random Forest in dataset test data the % of 1 cases in the selection is 78%.

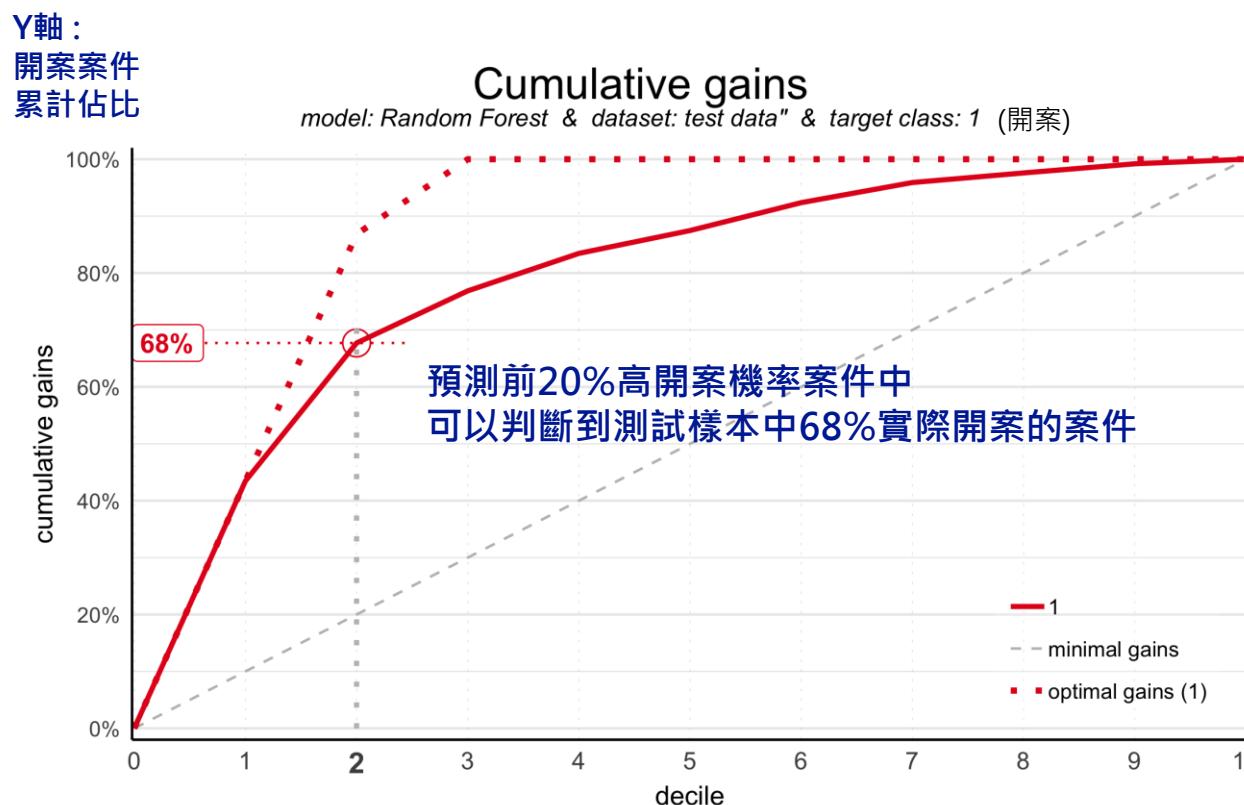
X軸(decile)：將測試樣本依模型預測為1(開案)之機率自高到低排序並切為10等分 (即1為預測機率最高的10%測試樣本)



When we select 20% with the highest probability according to model Random Forest in test data, this selection for 1 cases is 3.4 times better than selecting without a model.

## 2. 通報表開案預測模型

## 模型評估 – H model



When we select 20% with the highest probability according to Random Forest, this selection holds 68% of all 1 cases in test data.

X軸(decile)：將測試樣本依模型預測為1(開案)之機率自高到低排序並切為10等分 (即1為預測機率最高的10%測試樣本)

## 2. 通報表開案預測模型

## 小結

- ✓ 兩個子模型測試樣本中，
  - G模型預測最高前20%能預測到約49%的開案案件，且前20%的案件開案率高達98%。
  - H模型預測最高前20%能預測到約68%的開案案件，且前20%的案件開案率78%。
- ✓ 建議未來可作為社工安排案件調查順序之輔助參考，特別是在非緊急案件(H model)可在有限時間內(30天內須提出調查報告)做更有效率之安排。

1000個通報案件中，模型預測前200個高開案案件中，可以抓到近一半實際開案件，且這200個高開案中，開案率高達98%。

### 3. 通報表案情陳述NLP (Natural Language Processing)

- 中文語意斷詞 (Jieba)
- TF-IDF
- 文字分群(加入預測變因)
- WordCloud

## 3. 通報表案情陳述NLP

## 中文語意斷詞

## Jieba 結巴斷詞演算法概觀



## 3. 通報表案情陳述NLP

## 中文語意斷詞

	idx	text
3382	3471	[今早, 送往, 保母, 保母, 發現, 僵直, 抽蓄, 情況, 撥打, 送入, 本院, 急...
1145	1191	[口角, 遭案夫, 筆電, 感到, 不適, 忠心, 不穩, 眼睛, 失焦, 主述, 三起, ...
4375	4469	[之前, 有案, 生活, 穩穩定, 前幾日, 入獄, 哥哥, 打理, 生活, 事宜, 同住, ...
987	1031	[.., 預約, 進行, 早療, 服務, 初次, 訪視, 家訪, 評估, 訪視, 時案, 外...
1069	1114	[之一, 自己, 經常, 看起來, 瘦骨如柴, 虛弱, 無精打采, 難以, 集中, 注意力, ...
4164	4257	[主因, 情緒, 不穩, 發脾氣, 情緒, 失控, 徒手, 招案, 脖子, 致案, 頸部, ...
4302	4396	[晚上, 哭鬧, 影響, 睡覺, 摔到, 床上, 上唇, 一直, 流血, 就醫, 治療, 驗...

權威詞			
父親 年級			
個案 兒少 母親 父母 臉頰			
案母 個案 案父 兒少 小孩 案弟 個案及案弟 案母表示 案母			
案主 案主兄弟 案鄰居 案家 案母 個資 台北市 年紀 聲音 家			
案鄰居 案母 飭罵 案家 舉報身 案主 孩童 聲響 聲音 聽聞案			
)			
案主 案父 案母 施虐 派案 案家			
母親 此兒童 情緒 社工			
小孩			
案父 案主 持皮帶 案家 物品 概況 要中心 表述 帶走案主 至			
案父 案主 案母 通報表 派案 家暴 維護案主			
)			
案主 案父母 現況 專線			
)			
打工 其父 鑰匙 情事 打工地點 新台幣 向其			

#### 3. 通報表案情陳述NLP

##### 文字分群

目的：找出案情陳述內特定文字群並設為變因，提高預測開案精準率

將通報表案情陳述中的文字斷詞後，依照詞頻、文章距離分為 9 群。

## 3. 通報表案情陳述NLP

## TF-IDF

抽菸 0.702601

公分 0.667158

大雅 0.655029

虎口 0.651425

德蘭 0.605058

寄養 0.588546

大伯 0.583665

词频(*tf*)、逆文档频率(*idf*)公式

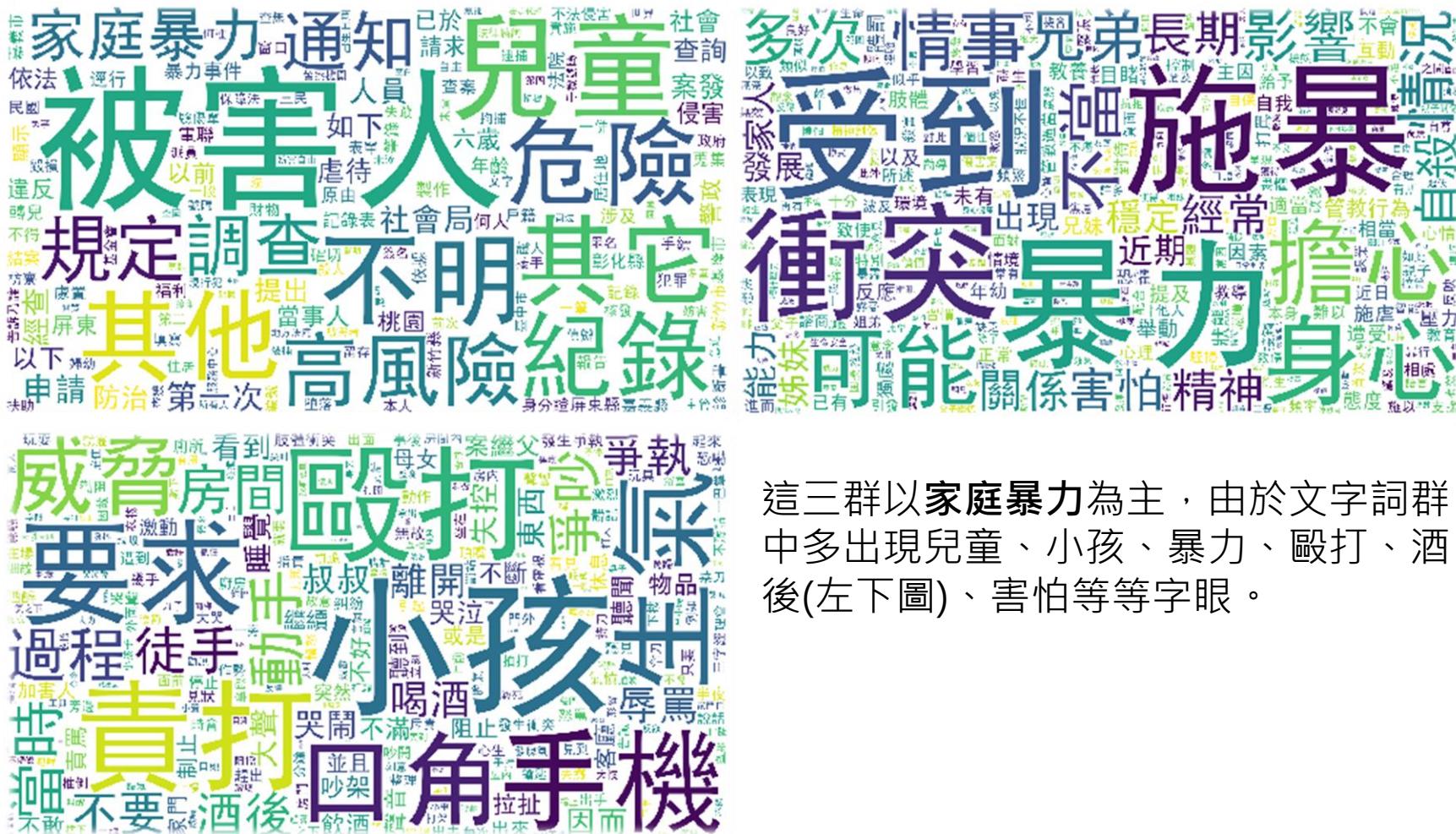
$$tf - idf = -n * \log\left(\frac{D_W}{D_T}\right)$$

其中

*n*：文章中包含该词的个数(*tf*)*D<sub>W</sub>*：包括该词的所有文档数*D<sub>T</sub>*：所有文档数

## 3. 通報表案情陳述NLP

## 文字雲



### 3.通報表案情陳述NLP

# 文字雲



這兩群以**受傷程度及部位**的文字為主，文字詞群中多出現頭部、傷痕、急診、就醫、外傷、舊傷等等明顯受傷字眼。

## 3. 通報表案情陳述NLP

## 文字雲



這四群以學校學習行為及家庭狀況為主，其中有詞群中多出現案父母、案外祖母等等主在表達家庭關係與狀況(右上圖)，較無明顯傷案字眼。

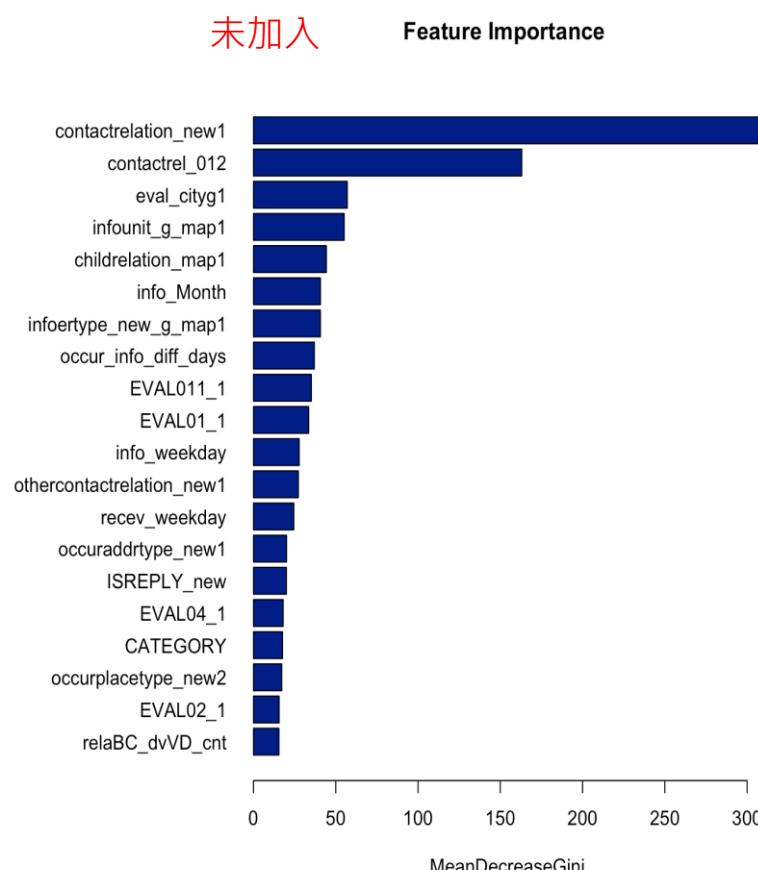
# 05 | Findings

## 主題1：兒少保護通報開案預測

通報表預測納入案情陳述文字分群模型結果

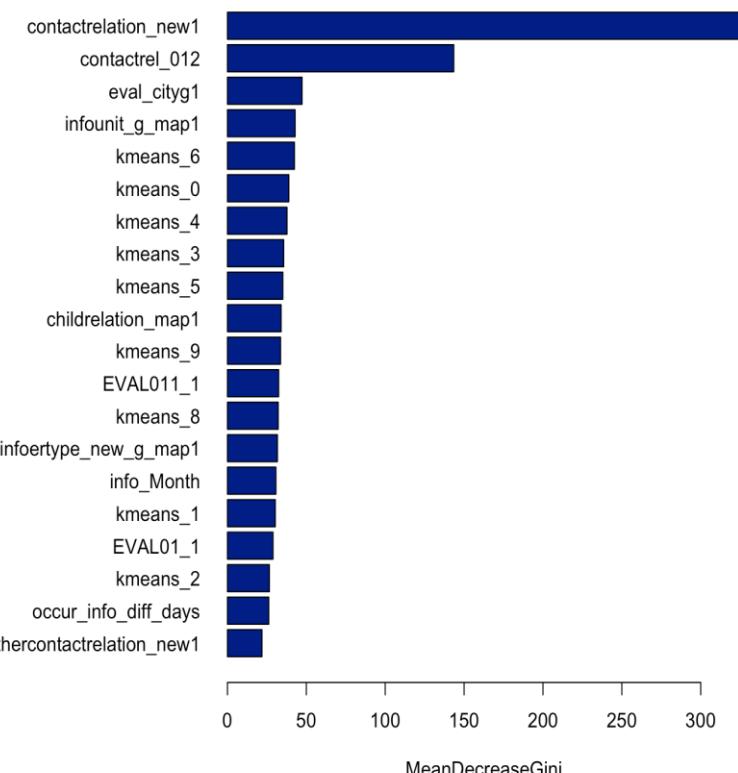
### 模型評估

Model: Random Forest  
樣本數: 500筆

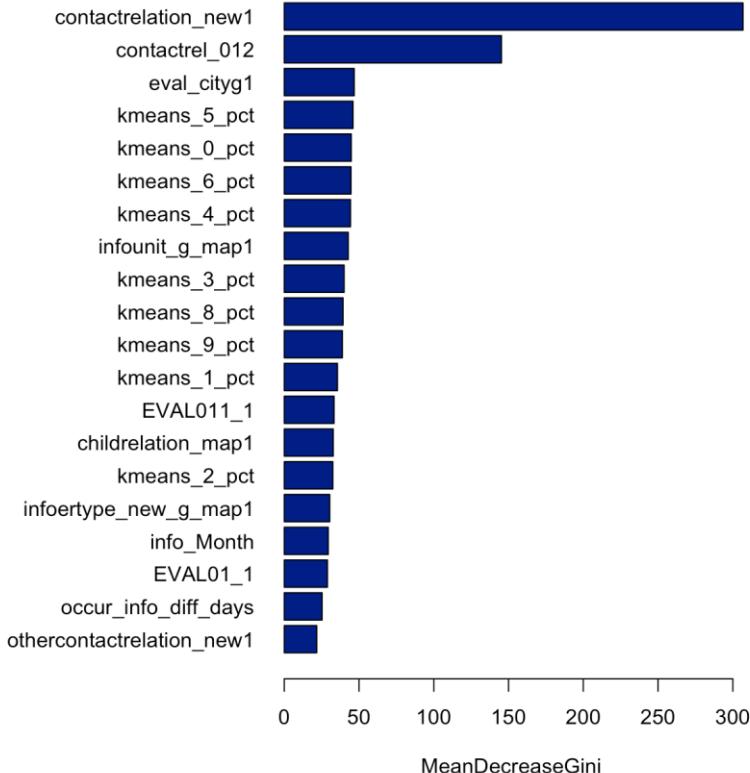


Model	Accuracy	Recall	Precision	F1 - Score
未加入	0.827	0.622	0.929	0.745
加入文字分群(9群詞頻)	0.829	0.633	0.924	0.751
加入文字分群(9群佔比)	0.826	0.620	0.929	0.744

**加入文字分群 (9群詞頻) Feature Importance - Text Clusters**



**加入文字分群 (9群佔比) Feature Importance - Text Clusters pct**



### 主題1：兒少保護通報開案預測

- 1.通報表關聯分析
- 2.通報表開案預測模型
- 3.通報表案情陳述NLP

### 主題2：共案特徵預測及訪視優先次序分析

- 1.兒少保開案共案特徵預測

#### 1. 兒少保開案共案特徵預測

### 緣由

兒少保護開案案件中，透過調查報告找出與「**保護司與心口司共同列管案件（共案）**」特徵類似之案件，但未有心理衛生社工介入服務（**未納入共訪機制**），預先提示兒少保社工提高服務密度及敏感度。

#### 1. 兒少保開案共案特徵預測

## 資料

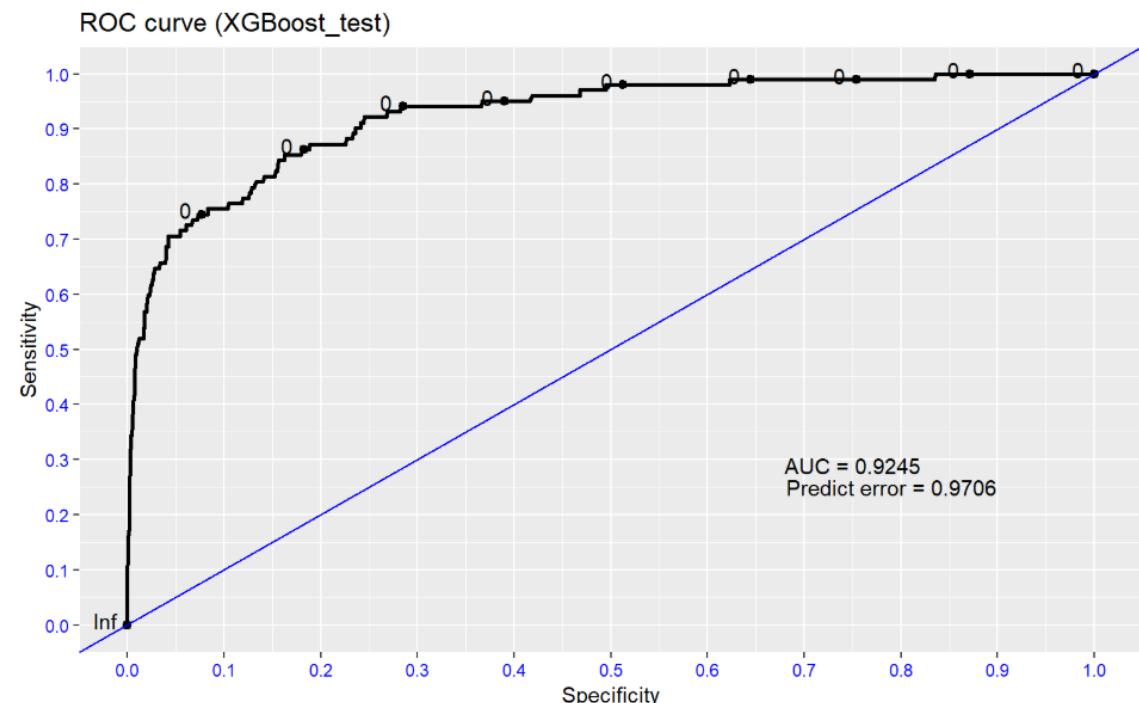
- 使用2016和2017年一類被害人父母調查報告資料進行分析：
- 目標：找出應提高服務敏感度之案件（與共案特徵相似之案件）。
- 目標特徵：兒少保案件應進行共案管理者。
- 特徵：建檔月份、年齡、與案主關係、教育程度、國籍別、現住地址（縣市）、是否與案主同住、居住狀況、身心障礙、婚姻狀況、職業、親子關係、性別、自殺紀錄、家暴施虐紀錄、家暴受虐紀錄。

## 1. 兒少保開案共案特徵預測

## 模型(XGBoost)

- 資料分割：將13621筆資料切分為80%訓練資料、20%測試資料
- 混淆矩陣：測試資料的預測狀況

	預測為 沒有共案特徵	預測為 有共案特徵
實際為 沒有共案特徵	2617	6
實際為 有共案特徵	74	28

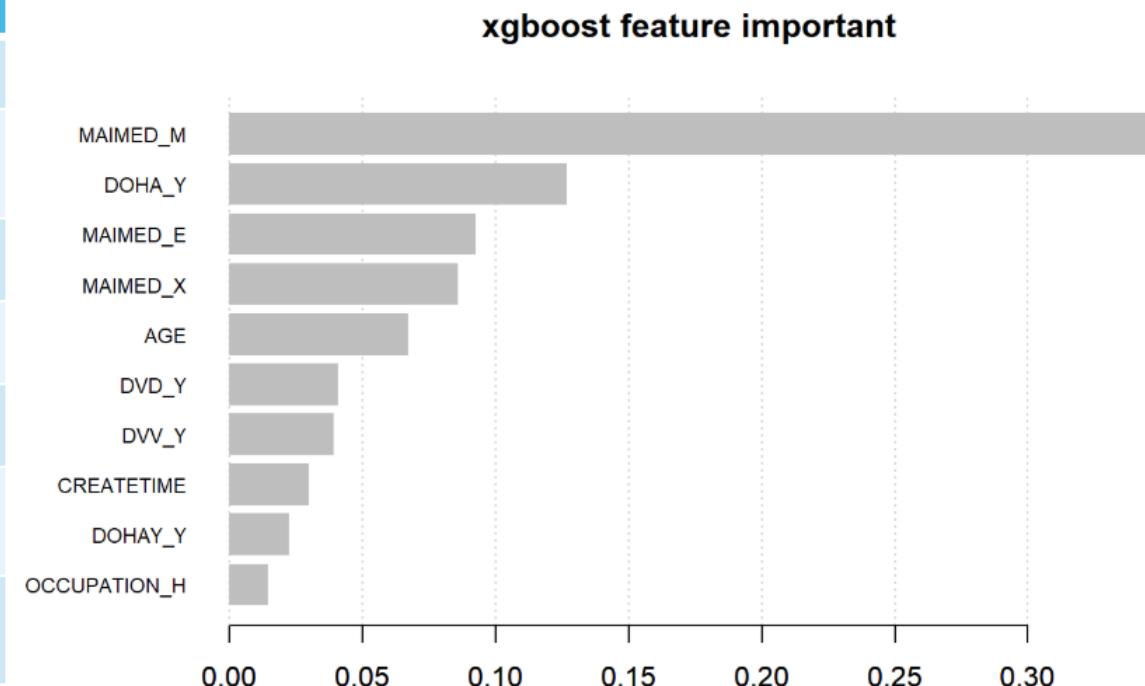


## 1. 兒少保開案共案特徵預測

## 特徵重要性

- 依照XGBoost的gain值(drop-one-loss)找出影響最大的特徵

欄位名稱	類別敘述
MAIMED (身心障礙)	領有身心障礙手冊 - 精神障礙 : M
DOHA (曾經列管自殺防治紀錄)	Y:有;N:無
MAIMED (身心障礙)	疑似身心障礙者 - 精神障礙 : E
MAIMED (身心障礙)	非身心障礙者 : X
AGE	年齡
DVD (有無家暴施虐者紀錄)	Y:有;N:無
DVV (有無家暴受虐者紀錄)	Y:有;N:無
建檔時間	月份
OCCUPATION (職業)	無工作 : H



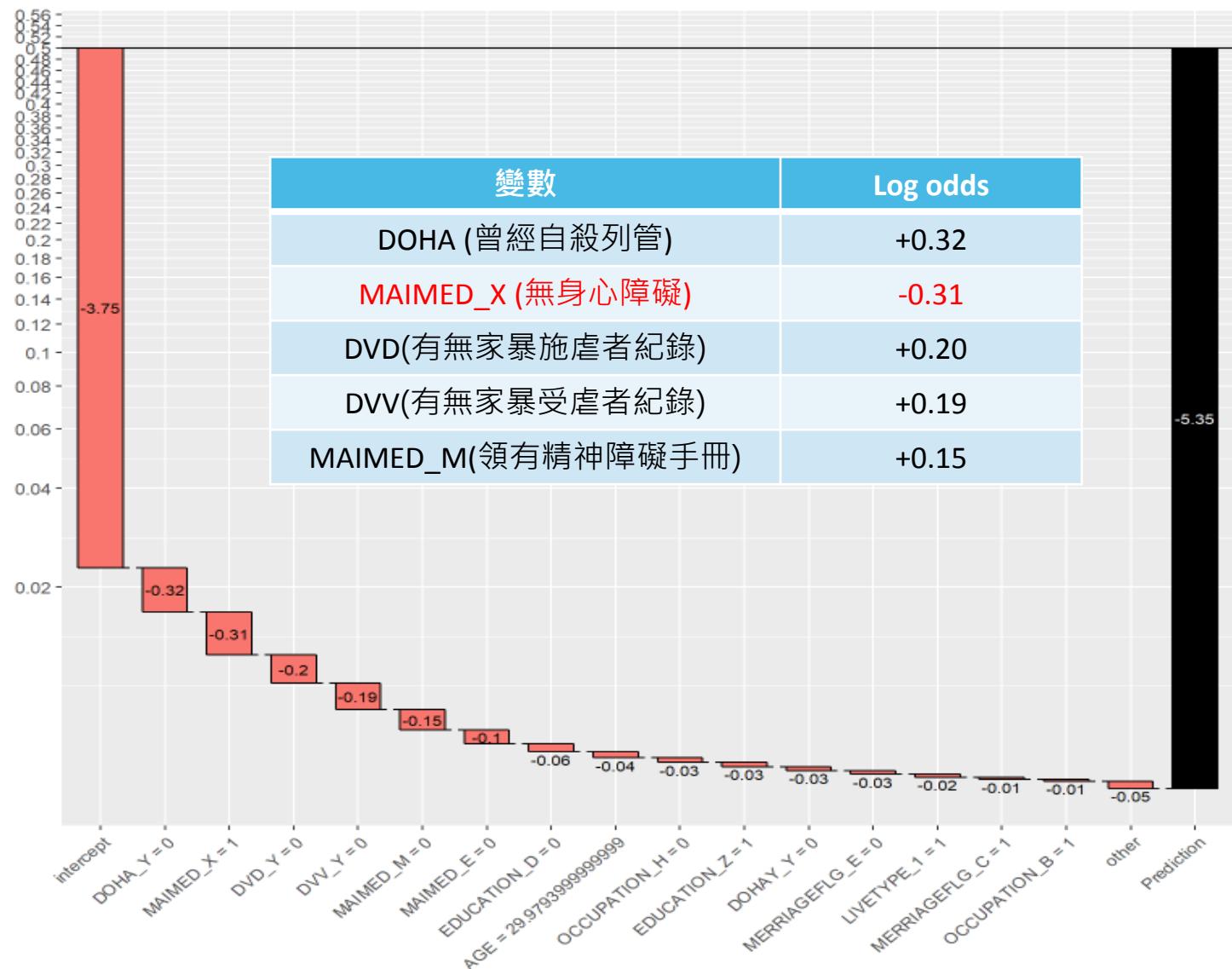
Q : 影響的正負相關性？

## 1. 兒少保開案共案特徵預測

## 模型解釋(個案分析1)

- 預測為共案特徵機率最低 (0.0047)
- 可看出 DOHA,DVD,DVV,MAIMED\_M 為對共案特徵有較顯著的正相關特徵
- MAIMED\_X 為對共案特徵有較顯著負相關特徵

※ 圖中顯示的數值為log odds

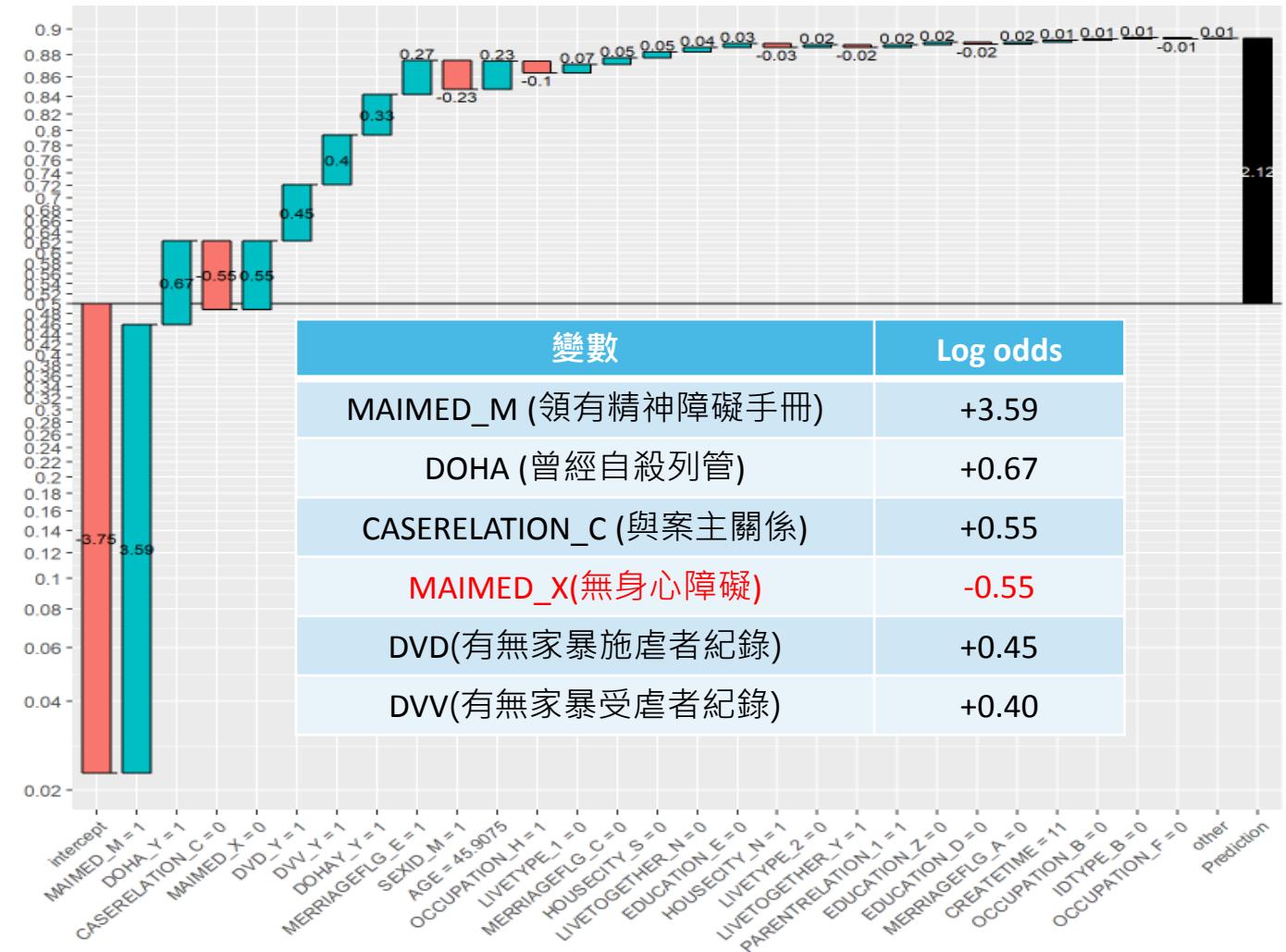


## 1. 兒少保開案共案特徵預測

## 模型解釋(個案分析2)

- 預測為共案特徵機率最高 (0.8932)
- 可看出 DOHA,DVD,DVV,MAIMED\_M, CASERELATION\_C為對共案特徵有較顯著的正相關特徵
- MAIMED\_X為對共案特徵有較顯著的負相關特徵

※ 圖中顯示的數值為log odds



# 05 | Findings

## 主題2：共案特徵預測及訪視優先次序分析

## 1. 兒少保開案共案特徵預測

需提高服務密度與敏感度族群

案號	性別 (0:女,1:男)	精神障礙 (領有手冊)	自殺列管 (曾經)	精神障礙 (疑似)	無身心障礙	年齡	家暴施虐 紀錄	家暴受虐 紀錄	建檔月份	自殺列管 (目前)	無工作者	共案特徵 預測機率值
個案1	1	1	1	0	0	46	1	0	10	0	0	0.7056358
個案2	0	1	0	0	0	41	0	1	11	0	0	0.6862443
個案3	0	1	1	0	0	39	0	1	3	0	1	0.6707948
個案4	0	1	0	0	0	40	1	1	7	0	0	0.6670761
個案5	0	1	0	0	0	36	1	1	7	0	1	0.6460465
個案6	0	1	0	0	0	39	0	1	4	0	0	0.618697
個案7	1	1	1	1	1	47	1	1	7	0	0	0.5948444
個案8	0	0	1	1	1	47	1	1	9	1	1	0.586201
個案9	0	1	1	1	1	47	1	1	9	0	0	0.5492619
個案10	0	1	0	1	1	47	1	1	10	0	1	0.5475327
個案11	0	1	0	0	0	37	0	1	7	0	0	0.544766
個案12	1	1	1	0	0	47	1	1	3	0	0	0.5416907
個案13	0	1	1	0	0	33	0	0	5	0	0	0.5317094

## 小結

- ✓ 在模型中可以找出影響共案特徵較顯著的變數，供往後兒少保社工在拿到通報表後能夠做出初步的篩選。
  
- ✓ 建議未來可作為雙邊的社工安排案件訪視順序之輔助參考，特別是在案件數過多的情形，可在有限時間內做更有效率之安排。

A large, abstract graphic on the left side of the slide features a grid of blue squares of varying shades and sizes, creating a pixelated or mosaic-like effect that tapers towards the top right.

# 06 | Conclusions

### □ 通報表預測：

- 未來可作為兒少保社工安排案件調查順序之輔助參考，特別在非緊急案件(H model)，預期提升3.4倍效率，紓解非緊急案件處理順序問題。

### □ 共案特徵預測：

- 協助兒少保社工在處遇階段，利用共案特徵預測，找到潛在高關注個案，及其影響因子，提高服務密度與敏感度。

# Thanks.

Data for Social Good  
2018夏 資料英雄



衛 生 福 利 部  
Ministry of Health and Welfare  
促進全民健康與福祉

