

115年公務人員特種考試警察人員、一般警察人員、
國家安全局國家安全情報人員及移民行政人員考試試題

考試別：一般警察人員考試

等別：二等考試

類科組別：刑事警察人員犯罪分析組

科目：資料探勘技術（包括資料庫管理與運用、線上交易處理【OLTP】、資料倉儲【Data Warehouse】、資料探勘【Data Mining】）

考試時間：2 小時

座號：_____

※注意：(一)禁止使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

一、假設你在警檢調或司法機關，原來委外建置的關聯式資料庫有設計瑕疵：有些表格候選鍵（Candidate Key）沒有被識別、有些表格沒有達成第三正規化。請針對下列三個工作：「建立資料倉儲」、「操作 OLAP（線上分析處理）」、「進行資料探勘」，分別至少舉 2 例來說明會造成什麼問題？你可自行假設資料庫的表格、欄位、設計瑕疵狀況。（20 分）

二、假設某一個判斷是否為詐騙電子郵件的模型，在全部 10000 筆資料中，「True Positive 真的為詐騙、且模型判斷為詐騙」有 TP 筆、「False Negative 真的為詐騙、但模型判斷為沒有詐騙」有 FN 筆、「False Positive 不是詐騙、但模型判斷為詐騙」有 FP 筆、「True Negative 不是詐騙、且模型也判斷不是詐騙」有 TN 筆，其數值各如下：

TP=2000	FP=150
FN=2100	TN=5750

(一)判斷模型效力好壞有多個指標，請對下面五個指標，先列出公式，再計算出數值至小數 2 位：準確率（Accuracy）、精確率（Precision）、召回率（Recall）、特異度（Specificity）、F1 分數。（15 分）

(二)請以上述五個指標綜合判定該模型的好壞，並說明理由。（3 分）

(三)在警檢調、司法做資料探勘，不能只依賴「準確率」、「精確率」、「召回率」、「特異度」中單一指標，否則可能因樣本分布、探勘門檻設定等造成模型整體效力其實不佳的狀況。請假設四個情境來分別說明若只看該單一指標的問題。（12 分）

- 三、若你是司法人員，想找出慣竊之人格特徵與行為模式。請規劃一個完整的資料探勘專案流程，說明你會收集那些資料、希望找出的各種規則、對該規則使用那些技術，以及最後預期能達成什麼樣的施政幫助。(25 分)
- 四、若您是司法人員，想委外資訊廠商，找出慣竊之人格特徵與行為模式，請列出招標需求規範（RFP）應特別著重的核心規範要點。(25 分)