

100年公務人員特種考試一般警察人員考試、
100年公務人員特種考試警察人員考試及
100年特種考試交通事業鐵路人員考試試題

代號：20360

全一頁

等 別：二等一般警察人員考試

類 科：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。

一、請就(一)資料模型，(二)查詢方式，(三)查詢複雜度，(四)資料變動頻率四方面比較資料倉儲與傳統資料庫之差異。(20分)

二、請回答以下有關關聯法則 (Association Rule) 分析之問題：

(一)請舉例說明如何運用關聯法則於犯罪分析上。(10分)

(二)對於一資料集，我們進行關聯法則探勘後可能會產生非常多的關聯法則，有些關聯法則雖內容不同但彼此之間可能具有潛在之高度重複性 (redundancy)。請說明如何分析找出具有潛在高度重複性之關聯法則。(15分)

三、分群 (Clustering) 是資料探勘中常用到的一種技術，請回答下列問題：

(一)請說明 k-means 分群方法之運作方式。(15分)

(二)請說明如何評估一個分群結果之好壞。(10分)

四、請回答以下有關分類 (Classification) 之問題：

(一)決策樹 (decision tree) 與支持向量機 (support vector machine, 簡稱 SVM) 為分類中常用之技術，請比較其優缺點。(15分)

(二)請說明如何利用 K-fold 交叉驗證 (K-fold cross validation) 來驗證分類模型之有效性。(15分)