

等 別：二等考試

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、請定義 relational database 裡的 equi-join 及 natural join 運算並舉例說明之。另請說明 hash-join algorithm 如何運作，以及為甚麼 hash-join algorithm 只能用在處理 equi-join 及 natural join 而不能處理一般 join？（25 分）
- 二、請考慮 frequent itemset 探勘。何謂 Apriori principle？請描述 Apriori algorithm 並指出 Apriori principle 如何用在 Apriori algorithm 上。（25 分）
- 三、請舉例說明以下資料探勘技術：(一) clustering，(二) sequential pattern discovery，(三) anomaly detection。（25 分）
- 四、請就以下 frequent 3-sequences： $\langle\{1,2,3\}\rangle, \langle\{1,2\}\{3\}\rangle, \langle\{1\}\{2,3\}\rangle, \langle\{1,2\}\{4\}\rangle, \langle\{1,3\}\{4\}\rangle, \langle\{1,2,4\}\rangle, \langle\{2,3\}\{3\}\rangle, \langle\{2,3\}\{4\}\rangle, \langle\{2\}\{3\}\{3\}\rangle, \langle\{2\}\{3\}\{4\}\rangle,$ （25 分）
 - (一)列出使用 GSP algorithm 在 candidate generation 步驟後所產生的所有 candidate 4-sequences。
 - (二)列出使用 GSP algorithm 在 candidate pruning 步驟後所有被去除的 candidate 4-sequences。