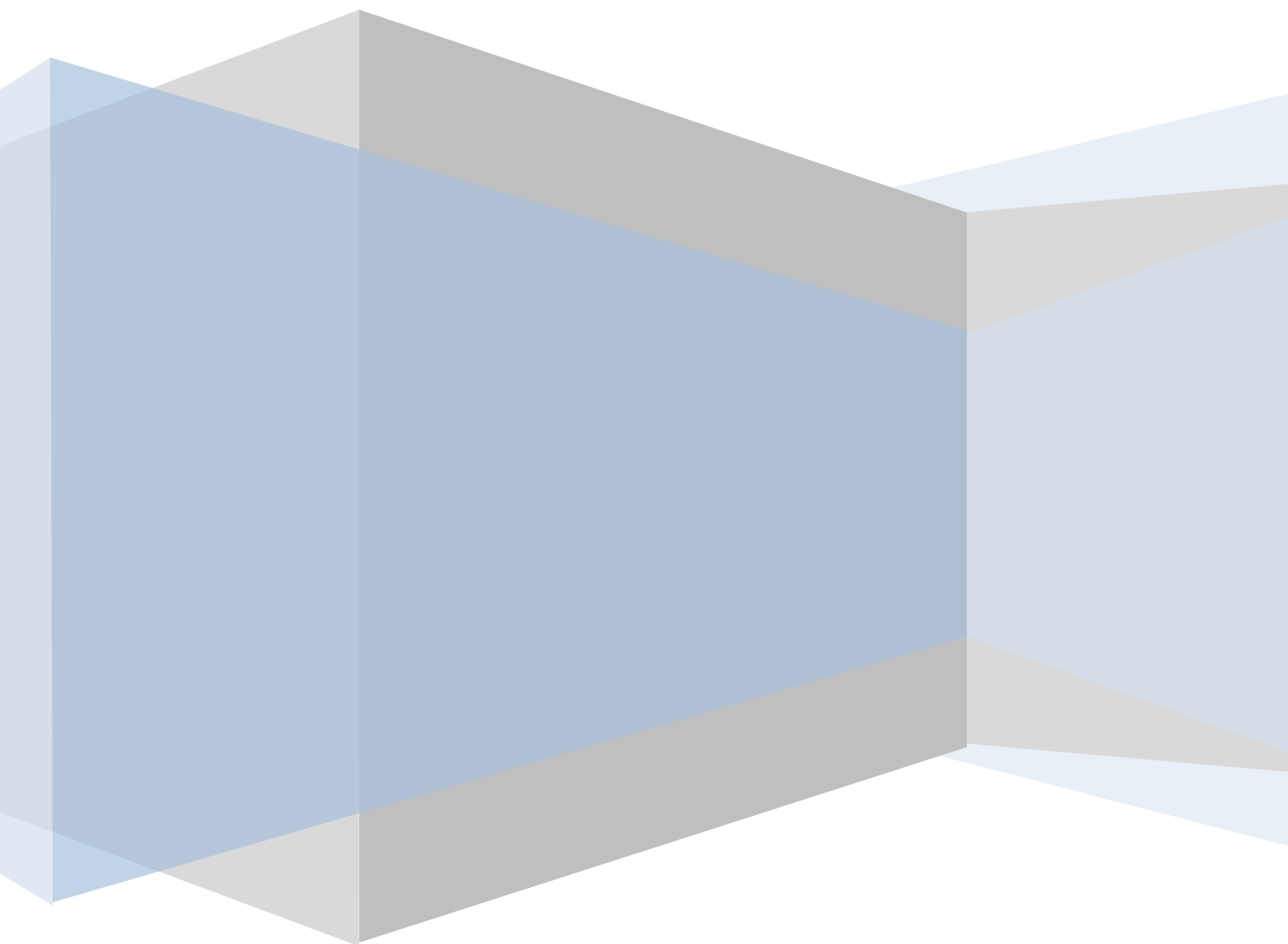


商用數據應用師

試題指南



中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (B) 1. 商管人最需要培養的資料探勘技能是：
- (A) 了解演算法
 - (B) 理解商業意涵
 - (C) 懂得程式設計
 - (D) 學會系統分析。
- (D) 2. 資料輸入不包含哪個程序：
- (A) 資料欄位選擇
 - (B) 前置作業
 - (C) 轉換資料格式
 - (D) 操作軟體工具。
- (A) 3. 將不必要的欄位刪除或不納入分析這是程序上那一個環節？
- (A) 精簡維度
 - (B) 調整編碼方式
 - (C) 調整顆粒度
 - (D) 資料型態轉換。
- (C) 4. 將原始資料改變為較能理解的格式這是程序上那一個環節？
- (A) 精簡維度
 - (B) 資料型態轉換
 - (C) 調整編碼方式
 - (D) 調整顆粒度。
- (B) 5. 「決定資料是要分析到每日還是每月？」這是程序上那一個環節？
- (A) 調整編碼方式
 - (B) 調整顆粒度
 - (C) 精簡維度
 - (D) 資料型態轉換。
- (D) 6. 連續值與離散值轉換這是程序上那一個環節？
- (A) 調整顆粒度
 - (B) 精簡維度
 - (C) 調整編碼方式
 - (D) 資料型態轉換。
- (D) 7. 解釋探勘後的資訊成為知識無法透過下列那一種方式完成？
- (A) 訪談
 - (B) 次級資料分析
 - (C) 觀察
 - (D) 解譯程式碼。
- (A) 8. 監督式與非監督的差異，以下何者為正確
- (A) 監督式需先制定特定目標
 - (B) 非監督式需先制定特定目標
 - (C) 關聯法則是監督式
 - (D) 決策樹是非監督式。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (B) 9. 關於資料探勘以下何者為非：
- (A) 挖掘出有用的東西
 - (B) 挖掘出奇點
 - (C) 具備意涵
 - (D) 對應商管目的。
- (B) 10. 以下關於挖礦與資料探勘何者為非
- (A) 知道要挖金礦這是監督式探勘
 - (B) 能從挖到的結果歸納一些可能有用的東西這是監督式探勘
 - (C) 資料探勘分為監督式與非監督式探勘
 - (D) 資料探勘就是從資料找尋有用的結果。
- (C) 11. 資料探勘上，所謂的「挖哪邊」以下何者為非？
- (A) 就是痛點
 - (B) 痛點就是對商管而言有意義的點
 - (C) 痛點可以視為替代品
 - (D) 痛點可以視為必需品。
- (A) 12. 如果遇到資料找到不多通常可以從下列哪一點去著手改進？
- (A) 我們資料找的還不夠仔細，我們去找找是否可以找到顆粒度較細的資料
 - (B) 請該單位準備好我們所需要的資料給我們
 - (C) 就用少量數據去分析
 - (D) 直接跑演算法。
- (A) 13. 關於資料探勘？
- (A) 是一種事後分析，需要有資料才能進行
 - (B) 不需要被驗證
 - (C) 只有社會科學可以但自然科學不可以使用
 - (D) 是一種事前分析，可以透過推衍得知結果。
- (A) 14. 「如何確認資料可用」包含了
- (A) 確認是使用什麼演算法
 - (B) 精準定義結論
 - (C) 應用管理工具
 - (D) 確認分析軟體。
- (D) 15. 關於 CSV 檔案，以下何者為非？
- (A) comma separated values
 - (B) CSV 檔案優點在於可以儲存幾乎無窮大
 - (C) CSV 檔案缺點就在於一律以文字方式儲存
 - (D) CSV 檔案中的 NULL 就是「空」。
- (B) 16. 何謂有品質的資料？
- (A) 漂亮
 - (B) 可用
 - (C) 大量
 - (D) 少量。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (C) 17. 資料有效性是指
- (A) 時間正確
 - (B) 選對列
 - (C) 沒有缺漏或空值
 - (D) 選對欄位。
- (A) 18. 關於探勘可用的資料表
- (A) 多以二維表示
 - (B) 超過二維以上不要刪除
 - (C) 超過二維以上可以透過直接刪除來降維
 - (D) 直接跑樞紐分析結果表就好。
- (B) 19. 集群分析從資料上看是
- (A) 垂直面的集合
 - (B) 水平面的集合
 - (C) 水平與垂直面的整合
 - (D) 隨機選取。
- (C) 20. 集群就是
- (A) 垂直面整合
 - (B) 取個體之間的最大差異點來整合
 - (C) 取共通性來整合
 - (D) 最大化差異。
- (B) 21. 集群的作法不包含
- (A) 看個別欄位的內容是否接近
 - (B) 喜不喜歡這個物件
 - (C) 選擇適當的屬性（欄位）並比較相似處
 - (D) 肘方法來判別集群數。
- (D) 22. 電腦要判別人長相是否接近不會透過……
- (A) 眉毛圖像的形狀的資料庫
 - (B) 嘴唇圖像的資料庫
 - (C) 眼睛圖像的資料庫
 - (D) 聲音的資料庫。
- (A) 23. 關於歐幾里德距離計算
- (A) 丟入隨機參數，以選定的幾個（比如：三個）單筆資料為主，看其他資料與該種子資料的距離進行運算
 - (B) 固定選取三個單筆資料為主，看其他資料與該種子資料的距離進行運算
 - (C) 餵入判別的欄位資料是依序由小到大選取的
 - (D) 餵入判別的欄位資料是依序由大到小選取的。。
- (D) 24. 透過人腦要判別一個人的長相是否是美女/帥哥不包含
- (A) 觀察眼睛
 - (B) 觀察頭髮
 - (C) 觀察耳朵
 - (D) 觀察書寫文字。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (A) 25. 關於集群分析何者為非
- (A) 所有資料都要納入
 - (B) 不合規範的要拿掉
 - (C) 不能觀察的要拿掉
 - (D) 因顆粒度無法納入的可視狀況調整後納入分析。
- (B) 26. 關於 K-means 集群分析何者為非
- (A) 電腦針對有限的隨機種子與計算
 - (B) 每一個物件都必須要擔任種子
 - (C) 電腦可以大量計算不同的隨機種子與其他物件的距離
 - (D) 電腦可以大量計算不同的質心與其他物件的距離。
- (B) 27. 以下何者為不為集群分析
- (A) 同羽毛的鳥就是會聚一起
 - (B) 整理過往資料建立模型判別進入門內的是男生或女生
 - (C) 非監督式探勘
 - (D) 組內差異最小組間差異最大。
- (A) 28. 資料探勘五類分析方法中的集群分析其英文名稱為下列哪一個選項
- (A) Clustering
 - (B) Classification
 - (C) Association Rule
 - (D) Classifier。
- (A) 29. 集群分析又可以稱為
- (A) 聚類分析
 - (B) 因素分析
 - (C) 因果分析
 - (D) 主成份分析。
- (B) 30. 關於 RFM 模型的英文的基本名稱，正確選項為
- (A) F (Fluently)
 - (B) R (Recency)
 - (C) M (Mental)
 - (D) M (Many)。
- (A) 31. 關於 K-Means 演算法的觀念，下列正確選項為
- (A) K 是指 K 個分群數量的意思
 - (B) 計算隨機種子與其他資料的因果關係
 - (C) K 值永遠為 5
 - (D) K 值可以設定為 0。
- (B) 32. 關於 Python 語言中 pandas 模組套件中的何種函數是可以讀取 csv 檔案
- (A) read_pdf()
 - (B) read_csv()
 - (C) read_doc()
 - (D) read_gif()。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (D) 33. 關於 Colab 敘述何者正確
- (A) 是微軟公司的產品
 - (B) 是 Collaboratory 的縮寫
 - (C) 是與 Visual Studio 很接近的工具
 - (D) 不用在自己電腦上面安裝 python 就可以執行的程式。
- (B) 34. 關於下列這行指令 `oringnaldata = pd.read_csv('clustering_Ex1.csv')` 的描述何者正確
- (A) 等號左邊的運算結果給予右邊
 - (B) 等號右邊的運算結果給予左邊
 - (C) 等號右邊是變數
 - (D) oringnaldata 變數名稱不可以隨便更改。
- (D) 35. 集群分析結束要將 oringnaldata 資料匯出到 OutPut.csv 可以輸入的指令為
- (A) `oringnaldata.add_csv("OutPut.csv")`
 - (B) `oringnaldata.join_csv("OutPut.csv")`
 - (C) `oringnaldata.put_csv("OutPut.csv")`
 - (D) `oringnaldata.to_csv("OutPut.csv")`。
- (B) 36. 讓消費習慣有差異的客戶可以分別集結到不同的客戶組群中，同時也將有相類似消費習慣的客戶集結到相對應的客戶組群中，這就是
- (A) 獨熱編碼
 - (B) 集群分析
 - (C) 關聯規則
 - (D) 主題模型分析。
- (A) 37. 關聯規則分析又稱為
- (A) 購物車分析
 - (B) 集群分析
 - (C) 序列分析
 - (D) 隨機森林分析。
- (B) 38. 關聯規則是屬於下列哪一種演算法
- (A) supervised algorithm
 - (B) unsupervised algorithm
 - (C) random algorithm
 - (D) non-sense algorithm。
- (D) 39. 實體商店與電子商務差異以下何者為非
- (A) 實體商店可以觀察貨架
 - (B) 電子商務下消費者不容易一直逛貨架
 - (C) 電子商務的彈出式廣告效果有限
 - (D) 實體商店可以賣出更多商品。
- (C) 40. 何謂關聯規則分析以下何者為非
- (A) 關聯就是關係
 - (B) 獨立存在的物件之間產生關係
 - (C) 關聯必須是「連續」而非「離散」的資料
 - (D) 信心度是用來產生關聯規則的依據。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (D) 41. 發現規則之後以下何者為非
- (A) 搭配質化分析結果規劃行銷企劃
 - (B) 不要貿然直接看數據說故事
 - (C) 量化分析與質化分析需要一起看
 - (D) 可以直接應用。
- (B) 42. 關聯看的是資料表格上的
- (A) Y 軸的關聯
 - (B) X 軸的關聯
 - (C) 資料的關係
 - (D) 交叉分析。
- (C) 43. Apriori 演算法
- (A) 只需找出最頻繁項目集
 - (B) 找出好用的項目集
 - (C) 一定要給定支持度與信心度
 - (D) 可以快速且節省資源的生成規則。
- (B) 44. 關於支持度以下何者為是
- (A) 決策變數在聯集中所出現的比例
 - (B) 就是商品組合在納入模型計算的樣本中出現的總次數
 - (C) 可以看成統計裡面的條件機率
 - (D) 決策變數在交集中所出現的比例。
- (D) 45. 關於「買啤酒(Beer)也會買尿布(Diaper)」的故事當時是發生在
- (A) Apple 公司
 - (B) 亞馬遜公司
 - (C) SAP 公司
 - (D) 沃爾瑪(Walmart)大型零售賣場中。
- (B) 46. 「啤酒(Beer)與尿布(Diaper)」的故事是採用下列哪一類演算法
- (A) 決策樹分類法
 - (B) 關聯規則
 - (C) 集群分析
 - (D) 貝氏分類法。
- (B) 47. 購物籃(Shopping basket)分析所採用方法是
- (A) 隨機森林
 - (B) 關聯規則
 - (C) 集群分析
 - (D) kNN 分類法。
- (A) 48. 分析購物籃的採購內容就等同是在分析客戶的
- (A) 消費行為(Consumer behavior)
 - (B) 倉庫管理行為
 - (C) 財務分析行為
 - (D) 預期心理行為。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (B) 49. 關於 Market basket analysis 的中文意思為
- (A) 出貨分析
 - (B) 購物籃分析
 - (C) 投籃命中分析
 - (D) 進貨分析。
- (C) 50. 購物籃(車)中的內容可以提供許多珍貴的資訊，這對賣場中商品的下列哪一項活動非常重要
- (A) 出貨
 - (B) 會計
 - (C) 補貨(Replenishment)
 - (D) 財務。
- (D) 51. 關於 Apriori 方法下列何者正確
- (A) 是一種決策樹分類法
 - (B) 是一種貝氏分類法
 - (C) 是一種集群分析方法
 - (D) 是一種關聯規則方法。
- (A) 52. 因為當交易資料量很大的時候常常會遇到計算上瓶頸，因此需要將原來的 list 資料型態轉為二維表格的資料框(DataFrame)，而這個資料框的內容填滿真(True)與假(False)的邏輯值，這些真、假值也就是什麼值
- (A) 布林值
 - (B) 焦點值
 - (C) 奇異值
 - (D) 閾值。
- (B) 53. 關於決策樹的基本原理以下何者為非
- (A) 監督式學習
 - (B) 非監督式學習
 - (C) 有目標
 - (D) 用以分類。
- (B) 54. 關於決策樹以下何者為非
- (A) 樹狀的結構
 - (B) 向上長的樹
 - (C) 有樹根
 - (D) 可能會過度擬合。
- (D) 55. 決策樹分枝的判別標準可以有
- (A) 熵 (Gini Coefficient)
 - (B) 吉尼係數 (Entropy)
 - (C) 商 (Entropy)
 - (D) 吉尼係數 (Gini Coefficient)。
- (A) 56. 決策樹分枝的判別標準
- (A) 熵越低越好
 - (B) 吉尼係數越高越好
 - (C) 熵越高越好
 - (D) 隨機性越高越好。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (D) 57. CART 決策樹特色以下何者為非
- (A) 可以分類連續變數與類別變數
 - (B) 已用過的變數可以重複使用
 - (C) 透過計算吉尼係數來判別分類
 - (D) 不能分類連續變數。
- (A) 58. 關於事後剪枝
- (A) 在樹已經生成之後，才對過過度擬合的部分進行修剪
 - (B) 生成樹的時候就對樹的生長進行限制
 - (C) 預先給定標準
 - (D) 多為主觀剪枝。
- (A) 59. 關於決策樹的描述何者正確
- (A) 監督式方法
 - (B) 非監督式方法
 - (C) 兼容監督與非監督方法
 - (D) 加強非監督式方法。
- (D) 60. 關於決策樹的描述何者錯誤
- (A) 有目標變數
 - (B) 有自變數
 - (C) 運用自變數去預測目標變數
 - (D) 自變數就是要預測的未來資料。
- (D) 61. 關於決策樹的預測的驗證議題，下列何者錯誤
- (A) 使用訓練資料集來建立模型
 - (B) 將資料拆成訓練與測試組，前者訓練模型，後者用以驗證
 - (C) 測試組資料量不要多於訓練組資料量
 - (D) 測試組資料量一定要是訓練組資料量的九倍。
- (A) 62. `np.array()` 函數(numpy 引入程式後命名為 `np`) 的意義是
- (A) 將目前資料轉換成陣列
 - (B) 將陣列的資料轉換成資料框資料
 - (C) 將 `list` 轉為 `table`
 - (D) 將 `list` 轉為資料框。
- (C) 63. 使用分類六大步驟完成建立分類與預測決策樹不包含
- (A) 資料載入與準備
 - (B) 訓練與測試資料切割
 - (C) 產生訓練資料集的預測結果
 - (D) 預測未知類別的新資料。
- (A) 64. 關於 `DecisionTreeClassifier()` 中第一個參數 `criterion` 如果設定為 `gini` 表示
- (A) 使用 CART 演算法
 - (B) 表示 ID3 演算法
 - (C) 表示隨機樹演算法
 - (D) 表示 kNN 演算法。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (A) 65. 在 Python 語言中所使用 `DecisionTreeClassifier()` 指令參數 `criterion` 可以設定為 `entropy` 做為計算，請問 `entropy` 的意思為下列哪一個選項
- (A) 熵值
 - (B) 極端值
 - (C) 奇異值
 - (D) 平均值。
- (D) 66. 在建立模型的過程中，需要將原始資料切割成兩部分，第一部分是訓練資料集，第二部分是
- (A) 交叉資料集
 - (B) 缺失值資料集
 - (C) 常數值資料集
 - (D) 測試資料集。
- (A) 67. 隨機森林是衍生自於
- (A) 決策樹
 - (B) 類神經
 - (C) 關聯規則
 - (D) 集群分析。
- (B) 68. 隨機森林的可能用途
- (A) 判別顧客年紀
 - (B) 推算顧客重要性
 - (C) 確定員工薪資
 - (D) 找出關聯規則。
- (A) 69. kNN 演算法的 `k` 的條件以下何者為非
- (A) 用來集群
 - (B) `k` 是指最近鄰居
 - (C) 隨機找 `k` 個
 - (D) `k` 不可以是 1。
- (D) 70. 關於 kNN 演算法以下何者為非
- (A) `k` 不可以太小
 - (B) `k` 需要是奇數
 - (C) 選擇的鄰居針對特定議題的意見居多數者勝出
 - (D) 與 k-Means 相同。
- (C) 71. 關於驗證
- (A) `k-fold` 不適用於隨機森林
 - (B) `k-fold` 不可以用來驗證訓練出的決策樹的好壞
 - (C) 在隨機森林中沒被抽到的樣本數可以當作驗證資料用
 - (D) 在隨機森林中沒被抽到的樣本數不可以當作驗證資料用。
- (B) 72. 關於隨機森林
- (A) 就是少數決
 - (B) 隨機森林的樣本抽樣設計是採取抽取之後還可以放回的方法
 - (C) 我的意見只會出現在一棵樹裡面
 - (D) 隨機森林的樣本抽樣設計是採取抽取之後不可以放回。

中華企業資源規劃學會 專業認證

商用數據應用師-參考題型

- (A) 73. 隨機森林是衍生自於
- (A) 決策樹
 - (B) 類神經
 - (C) 關聯規則
 - (D) 集群分析。
- (A) 74. 關於 kNN 預測分析以下何者正確
- (A) 設定 k 必須要是奇數
 - (B) 只能餵入偶數筆資料
 - (C) 預測結果可以有多個答案
 - (D) 設定 k 必須要設定成 0。
- (A) 75. 分類模型又可以稱為
- (A) 分類器
 - (B) 分歧器
 - (C) 橢圓曲線加密器
 - (D) 迴歸解密器。
- (B) 76. 所謂「近朱者赤，近墨者黑」的觀念類似下列哪一種商業數據分析的技術
- (A) 決策樹
 - (B) k 最近鄰演算法
 - (C) 簡單貝氏法
 - (D) 奇異值分解法。
- (B) 77. 建立模型過程中，基本上會將資料集切割成兩部分，其一為測試資料集，另一為
- (A) 開放資料集
 - (B) 訓練資料集
 - (C) 結案資料集
 - (D) 半折資料集。
- (A) 78. 整個建立模型過程中的啟動點在於資料的取得與區分，如果將測試資料集與訓練資料集合併在一起就等同於
- (A) 原始資料集
 - (B) 開放資料集
 - (C) 驗證資料集
 - (D) 鑑證資料集。
- (C) 79. 如果使用 kNN 建立分類模型，則其原理與 CART 決策樹分類模型
- (A) 相同
 - (B) 偽陽性資料量過半則相同
 - (C) 不同
 - (D) 偽陰性資料量過半則相同。
- (D) 80. 使用 Python 建立 kNN 模型過程中使用這一個指令 `KNeighborsClassifier(n_neighbors = 3)` 可以輕易建立 kNN 分類模型，其中 `n_neighbors` 參數代表意義為
- (A) 最壞狀況鄰居數量
 - (B) 平均鄰居數量
 - (C) 最遠鄰居數量
 - (D) 最近鄰居數量。