電子商務技術 HW5：Comparing learning schemes

資料集：customer_churn.csv

任務：顧客流失分析，判斷顧客是否會繼續消費 (target：churn 欄位)

以下題目請使用 python 完成

1. 針對 churn 欄位使用 Stratified sampling 從原本的資料集中取 60%的資料
2. 列出取樣後各類別的資料數量
3. 資料前處理，並以 10 folds cross-validation 建立 Logistic Regression 及 SVM 模型
4. 針對測試資料印出兩個模型的平均 Accuracy
5. 重複 1~4 題 30 次，並印出兩種模型最終的平均 Accuracy
6. 根據模型於 30 次 10 folds cross-validation 的 Accuracy，以 paired t-test 比較兩種模型，並說明結論

以下題目請使用 Weka 完成，並將操作步驟與結果截圖，並在截圖上圈出能滿足題目要求的設定(即使是預設值)

1. 使用 Stratified sampling 從原本的資料集中取 60%的資料
2. 顯示取樣後各類別的資料數量
3. 資料前處理，並以 repeated 10 folds cross-validation (重複 10 次) Paired t-test 比較 Logistic Regression 及 SVM 模型
4. 根據 weka 的輸出說明結論

** 若資料量太大以致於 weka 無法運作，在附上截圖證明後，可降低抽樣數量

** weka t-test 功能在 Experimenter 中 (可參考 weka manual chapter 6 - experimenter)

作業繳交說明

- 繳交期限：5/11 (三) 中午 12:00
- Python 題請繳交.ipynb 檔、Weka 題請繳交 pdf 檔，檔名 ECT_HW5_學號。
  - 程式中請以註解或文字方塊標示題號
  - **需確保程式執行上傳至 ee-class 作業區那一版本的資料集不會出錯**
- 上傳至 ee-class 作業區，遲交一天扣該次作業得分 5%，最多扣 50%。
