電子商務技術 HW3：Regression and Classification

第一部分：醫療費用預測，預測某人的醫療花費是多少 (target：charges 欄位)

請使用 Python 完成以下題目 (30%)

1. 載入 insurance.csv 為 pandas DataFrame 格式 (0%)
2. 資料前處理 (e.g., feature encoding) (10%)
3. 將資料集劃分為訓練集與測試集 (2%)
4. 建立並訓練以下三種迴歸模型：Linear Regression、SVM (SVR)、Decision Tree (10%)
5. 請以 R2 (coefficient of determination)*、RSME (root mean square error)、MAE (mean absolute error)評估/比較三個模型在訓練資料及測試資料上的表現。(8%)

*註：請查看 scikit-learn 的 score( )

第二部分：顧客流失分析，判斷顧客是否會繼續消費 (target：churn 欄位)

請使用 Python 完成以下題目 (45%)

1. 載入 customer_churn.csv 為 pandas DataFrame 格式 (0%)
2. 列出資料筆數、屬性數量、每個欄位的空值個數以及各類別(target)的資料筆數 (4%)
3. 資料集中有部分資料重覆出現，請刪除重覆多餘的資料 (僅保留一筆)，並列出剩餘的資料筆數。 (2%)
4. 填補空值及其他資料前處理 (e.g., feature encoding) (10%)
5. 將資料集劃分為訓練集與測試集 (1%)
6. 訓練與測試 SVM、Logistic Regression、Decision Tree 模型 (10%)

7. 請以 Accuracy 評估同一模型在不同超參數設定下於訓練資料及測試資料上的表現 (至少比較兩組超參數) (6%)

8. 請以 Accuracy 比較三個模型的表現 (6%)

9. 請列出三個模型的 Confusion matrix，並簡述其意義 (6%)

請使用 Weka 完成以下題目，並將操作步驟截圖 (25%)

1. 載入 customer_churn.csv，列出資料筆數、屬性數量以及每個欄位的空值個數 (5%)

2. 請刪除重覆多餘的資料 (僅保留一筆)，並列出剩餘的資料筆數 (5%)

3. 資料前處理 (5%)

4. 訓練、測試 SVM、Logistic Regression、Decision Tree 模型，請以 Accuracy 評估模型表現 (10%)

作業繳交說明

- 繳交期限：3/30 (三) 中午 12:00
- Python 題請繳交.ipynb 檔、Weka 題請繳交 pdf 檔，檔名為 ECT_HW3_學號。
  - ipynb 檔可分成兩個檔案繳交 (請在檔名註明檔案內容，e.g., ECT_HW3_學號_第一部分)
  - 程式中請以註解或文字方塊標示題號
- 上傳至 ee-class 作業區，遲交一天扣該次作業得分 5%，最多扣 50%。

附件：資料集屬性

1. insurance.csv

- age：年紀
- sex：性別
- bmi：身體質量指數
- children：撫養幾個小孩
- smoker：是否會抽菸
- region：居住區域
- charges：醫療花費

2. customer_churn.csv

- CustomerID：識別顧客 (unique)
- Churn：是否為流失的顧客 (1：是、0：否)
- Tenure：顧客多久以前開始在此商店消費
- PreferredLoginDevice：顧客習慣用何種裝置登入此網路商店
- CityTier：顧客居住的城市等級
- WarehouseToHome：倉庫到顧客家的距離
- PreferredPaymentMode：顧客慣用付款方式
- Gender：顧客性別
- HourSpendOnApp：顧客每天花幾小時瀏覽網路商店
- NumberOfDeviceRegistered：顧客在幾個裝置上登入
- PreferedOrderCat：顧客過去一個月偏好的商品類型
- SatisfactionScore：顧客滿意度
- MaritalStatus：顧客的婚姻狀態
- NumberOfAddress：顧客的地址數量
- Complain：顧客過去一個月內是否客訴過
- OrderAmountHikeFromlastYear：相較於去年此顧客的消費增長百分比
- CouponUsed：顧客過去一個月內用過幾次優惠卷
- OrderCount ：顧客過去一個月內下訂單的次數
- DaySinceLastOrder：距離上次顧客消費的天數
- CashbackAmount：顧客過去一個月的平均現金回饋