# 機器學習導論

## Homework #4
## Due 2022 Oct 31 11:00PM

**(一)** 針對員工離職率(left)進行預測

資料檔案：*HW3_hr-analytics.csv*

作業要求：

1. 讀入資料，並判斷出那些數據格式不是數字，或是有缺失值。
2. 將非數字類型的資料進行必要的編碼。
3. 若有缺失值請填補。
4. 將資料切割成訓練集 70%，預測集 30%。
5. 利用 Decision tree 及神經網路模型進行預測。
6. 請與前次 Logistic Regression 的預測準確率進行比較。請探討那個模型比較適合，其可能原因為何?

**(二)** 針對信用卡交易資料，預測是否為詐騙的交易（class==1）

資料檔案：*HW4_creditcard.csv*

作業要求：

1. 讀入資料、切割資料（測試集佔 30%，訓練集佔 70%）
2. 利用 Decision tree 及神經網路模型進行預測，計算出 Accuracy, Recall, Precision, F1-Score。
3. 統計 class==0 及 class==1 的資料比數，看是否類別間資料數量是否有很不平衡的現象。
4. 為了要提高 recall 的數值，請：
   - 改變 Decision tree 及神經網路模型中類別權重或訓練權重，計算新的結果，與之前結果比較。
   - 利用 imbalanced-learn 套件中 SMOTE 的方法來增量資料，計算新的結果，與之前結果比較。

繳交說明：請繳交 jupyter notebook 之檔案。若有討論部分也利用 jupyter notebook 說明。