專案作業二 訓練模型預測數值

* 作業內容包括：
  + 第一個實驗為Adult資料集，adult.train.txt當作訓練資料集、adult.test.txt當作測試資料集。預測測試資料集中hours-per-week之欄位值。請使用KNN、SVR、RandomForest、XGBoost四個演算法，並比較不同演算法之預測績效以及所需要的計算時間。
  + 第二個實驗資料集Boston Housing dataset: 506 samples and 13 feature variables，預測的目標為MEDV欄位的數值。
    1. 本實驗請使用XGBoost演算法及K-fold cross validation，K設為5，並且輸出每一個fold的預測績效以及5 folds的平均績效。
    2. 使用XGBoost的feature\_importance計算特徵重要性，對欄位特徵進行篩選，比較特徵欄位刪減前後績效之差異。
  + 績效指標：MAPE、RMSE及R2
  + 撰寫報告，請遵循學術論文寫作格式，內容至少必須包括下列要項：

**摘要** (學術論文摘要通常不分段，內容簡要地涵蓋動機、目的、方法與實驗結果)

**一、緒論**

**1.1動機 (**描述本研究之動機)

**1.2目的** (想進行哪種探勘，期望了解甚麼？)

**二、方法**

(1. 簡單說明所選用的探勘演算法或所採用的方法，2. 程式操作方法，無法操作者，以不正確論。)

**三、實驗**

**3.1資料集** (簡介資料集：名稱、筆數、欄位、型態…、並以資料表格呈現部份資料內容)

**3.2前置處理 (**如果有的話，例如缺值與雜訊處理、種類型屬性轉換、值域正規化、…**)**

**3.3實驗設計** (實驗如何進行、參數如何設定、績效如何評估等）

**3.4實驗結果** (分析比較不同模型、不同參數設定的預測績效)

**四、結論**

**五、參考文獻** (參考論文、網站程式碼、資料，請註明引用文獻或來源)

* 評分標準

內容(難度、豐富性、方法正確性)：60%，論文格式：20%，寫作技巧(文筆通順)：20%

* **任何複製剪貼網站資料或學長姐以前作業，皆屬抄襲，嚴重者零分。引用網站或他人資料，應註明引用來源。**
* 此份報告最多15頁，請加註頁碼。請遵循投稿論文寫作格式，使用一欄式，無需製作論文目錄。可以觀摩 ACM Transactions on Knowledge Discovery from Data期刊論文的寫作技巧與格式。
* 紙本報告於上課前繳交或直接放入許老師的系辦信箱。專案程式直接上傳GitHub，專案名稱：hw2+組長學號，並新增助教為專案成員，以利下載評分。缺紙本或程式皆視為未完成繳交作業。
* 繳交紙本時，在左上角訂一訂書針即可，無須其他任何裝訂。
* 每遲交一天，扣該次作業成績10分。
* 繳交日期：112/11/23（四）