Topic: ***Decision Tree***

Description:

* 本次作業為預測角色是否死亡，資料集在以下網址: <https://www.kaggle.com/mylesoneill/game-of-thrones>，其中第二份資料的character-deaths.csv或下載附件檔案。其中三個欄位 **Death Year , Book of Death , Death Chapter 取其中一個欄位當預測目標用即可**，請將欄位的空值轉成**0(代表存活)，有數值的轉成1(代表死亡)**。
* 作業流程&要求 (※請使用python處理，可以使用任何套件)

1)將資料讀取進來(可用pandas套件)

2)資料前處理

2-1把空值以0替代

2-2Death Year , Book of Death , Death Chapter三者取一個，將有數值的轉成1

2-3將Allegiances轉成dummy特徵(底下有幾種分類就會變成幾個特徵，值是0或1，**本來的資料集就會再增**

**加約20種特徵**)

2-4亂數拆成訓練集(75%)與測試集(25%)

3)使用scikit-learn的**DecisionTreeClassifier**進行預測(可以先試著將網頁範例(iris)跑出來在使用這次作業的資料集)

4)**做出Confusion Matrix，並計算Precision, Recall, Accuracy** (提示: 可使用sklearn.metrics)

5)**產出決策樹的圖** (p.s 記得限制樹的深度，以免結果無法顯示)

* 請將(4)(5)的結果截圖上傳(word 或 PDF)，並上傳python程式碼。可以嘗試改善資料前處理部分&調整決策樹參數以增加Accuracy
* 要上傳的東西:

1.Word/PDF(截圖跟簡短說明、討論等等)

2.Python程式碼

Reference:

1.推薦用Anaconda安裝python環境，包含Jupyter Notebook、Spyder等IDE以及相關套件

2. scikit-learn上的決策樹教學：http://scikit-learn.org/stable/modules/tree.html

3. 若(5)graphviz發生問題，請試試看conda install graphviz並重啟IDE

https://stackoverflow.com/questions/28312534/graphvizs-executables-are-not-found-python-3-4