作業概述：

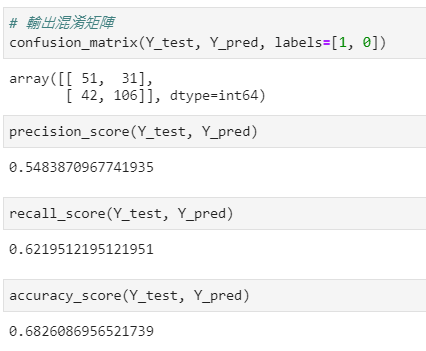
本作業是採用決策樹進行分析，讀取character-deaths.csv檔後，

將['Death Year', 'Book of Death', 'Death Chapter']這三者選擇'Book of Death'當作Y值，然後所有column扣掉'Name'以及['Death Year', 'Book of Death', 'Death Chapter']這三者後當作X值。其中'Allegiances'這一項的值因為不是數值資料，所以將其轉成one-hot-encoding的形式。

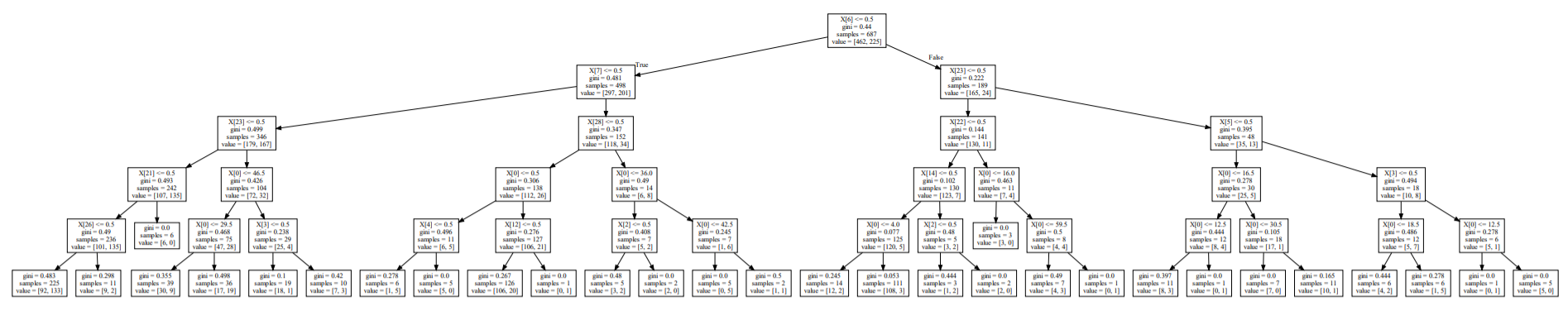
做完前處理後，接著將資料切割成75%/25%，75%當訓練集、剩餘的當測試集

接著帶入決策樹模型，因為避免圖形過大以及過擬和，選擇超參數樹的深度為5。接著跑混淆矩陣、計算accuracy/Precision/Recall這些值，接著使用graphviz, pydotplus等套件來視覺化決策樹。

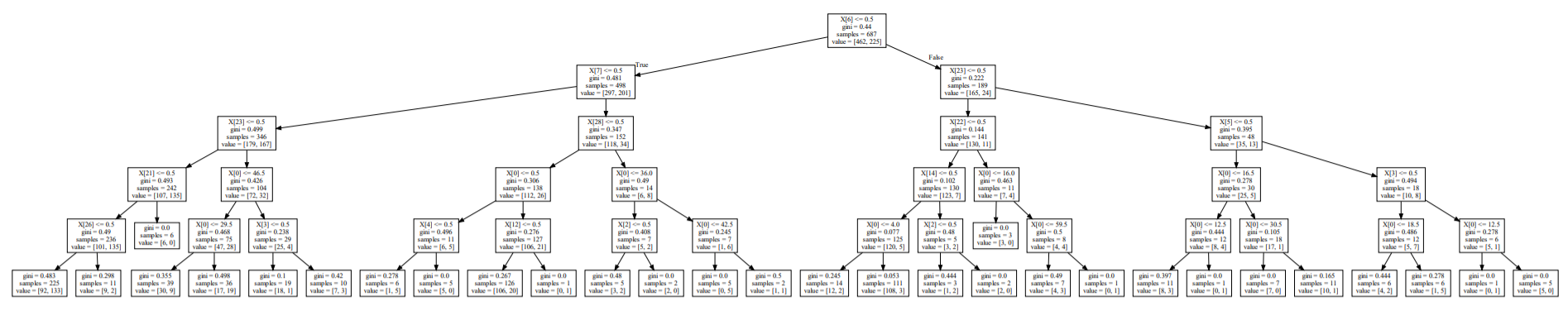
輸出結果：



決策樹的圖形：



由於層數有點多，所以可能看的不是很清楚，因此下面一張有90度的放大圖，請見諒。



結果分析：

最後跑出的accuracy/Precision/Recall並不高，大約在五成到七成間徘徊，初步分析可能原因是深度太淺導致欠擬和，但本人試過其他更大的數字發現表現並沒有更好的跡象。第二層原因是可能超參數設的不夠好，可以用python內建的gridsearchcv來尋找最好的超參數。若仍然沒有更好的表現，猜測此資料集不適合用決策樹來擬和抑或是決策樹已經是其中最好的方法。

雖然在處理graphviz這個問題時發生了許多困難，不過上網找了許多資料後，問題就一一解決了，真心覺得要加強資料搜尋的能力，以後處理事情或是寫code都能事半功倍。