HW2補做 Report

本次作業二重做，資料來源由原先法源網改為司法院資料檢索系統https://law.judicial.gov.tw/FJUD/default\_AD.aspx

Scrapy.ipynb 爬蟲該資料網的判決書，本次作業爬蟲民國100/1/1到民國108/6/1地方法院民法第1052條的判決結果用以實作Decision Tree。

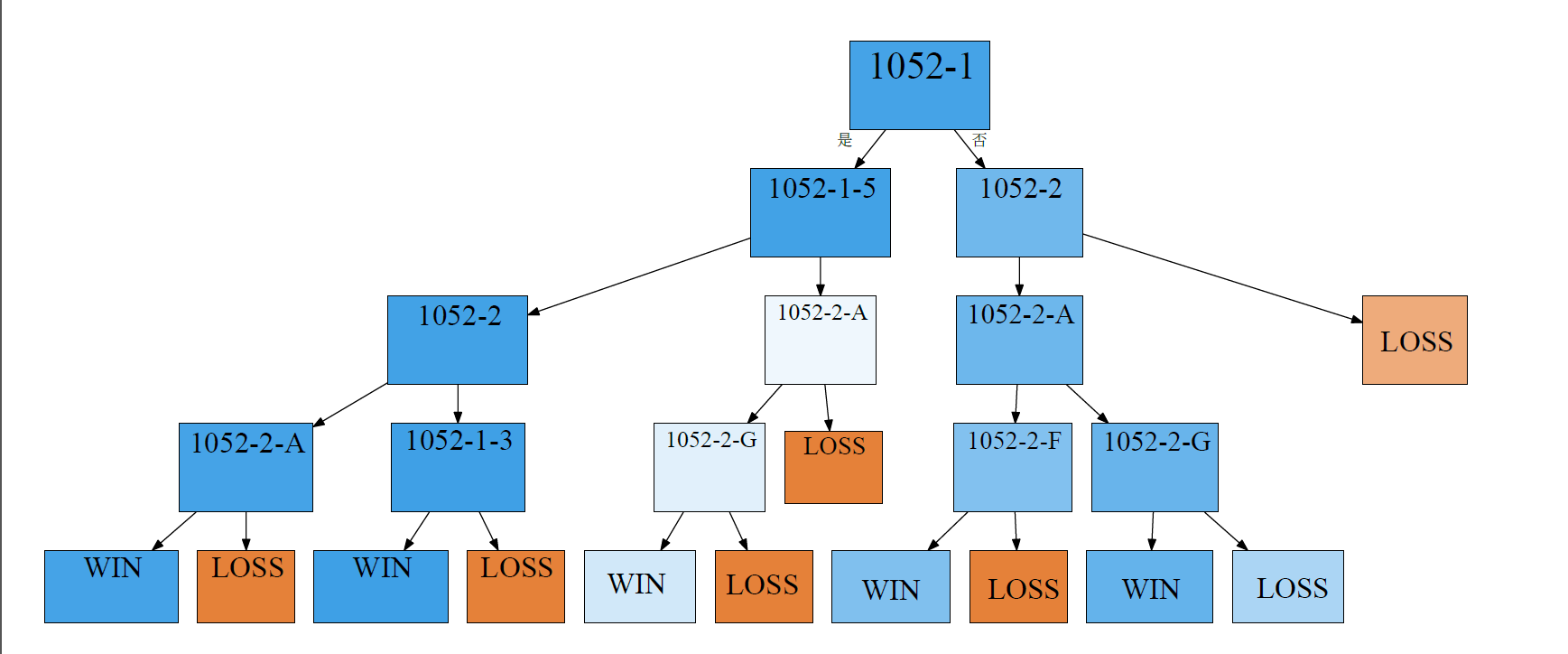
Data\_Preprocessing.ipynb 過濾爬取出來的資料，首先濾掉與離婚判決無關的資料，並且整理出judten.csv檔提供分析。

DataANA.ipyng 包含分析data以及實作decision tree

Data\_Generate rule:本次data來源為爬蟲，而民法第1052條關於離婚訴訟的判決，其中包含第1052條第1項第1到10款、第1052條第2項A~G款(A~G為自訂的原因附於CODE中說明)，資料中有符合某一款則此款欄位為1否則為0。

Decision tree:

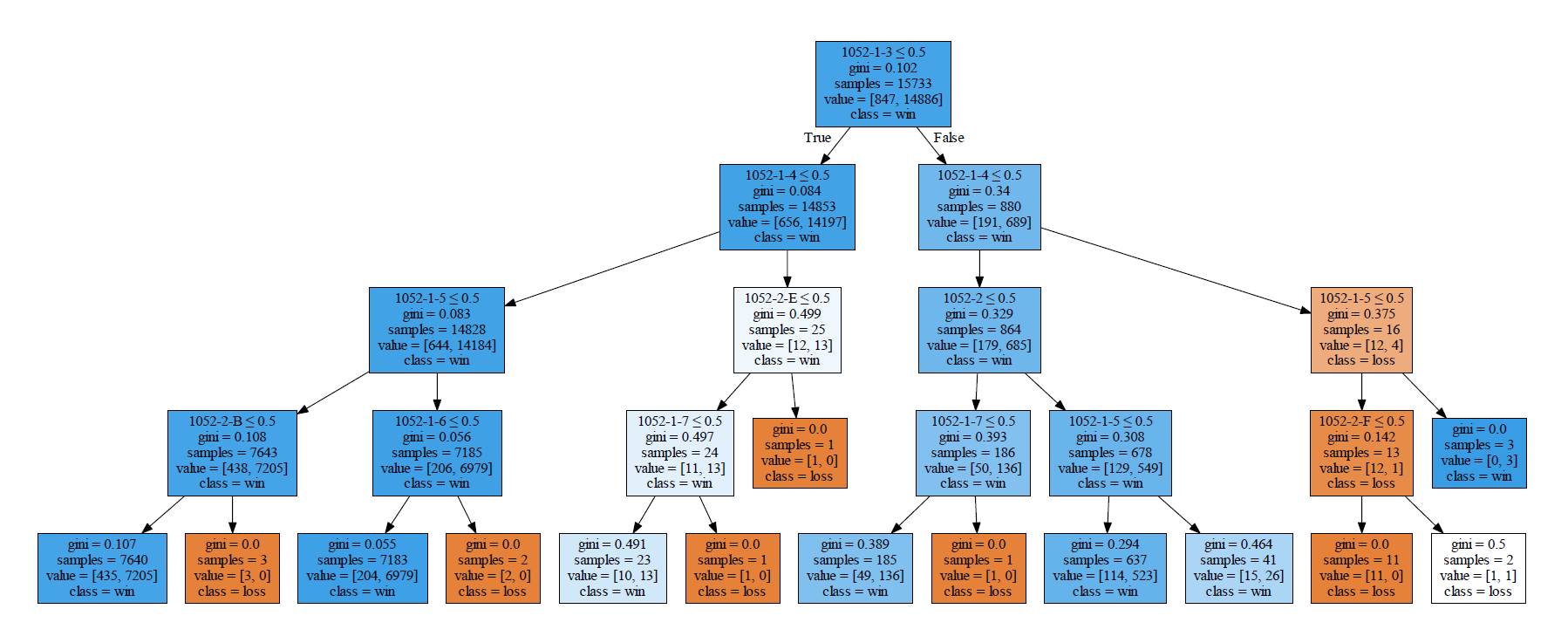
**原始的假設tree**



其中的判別規則如下:

因為第1052條第1項第1~10款為明確原因、行為，所以先行判斷有無，再來第1052條第2項自己訂的原因中，A與F與G算是勝率較高且較多案件數所以加入判斷，而第1052條第1項第3、5款基本上符合也一定勝訴，所以也納入判斷，其餘的狀況皆為敗訴。

**利用sklearn decision tree 做的 tree**



由演算法Decision Tree跑出來的結果可發現，對於第1052條第1項的第3、5款判別與預期一樣，但是對於6、7款也有著墨，而第二項基本上只探討有無符合即為勝訴，後來去調查法律事實，所謂民法第1052條第2項的意義其實是:除了在民法第1052條第1項中的其他原因造成離婚條件，基本上若能提出具體事實即可獲判勝訴，所以演算法的判斷多為正確。