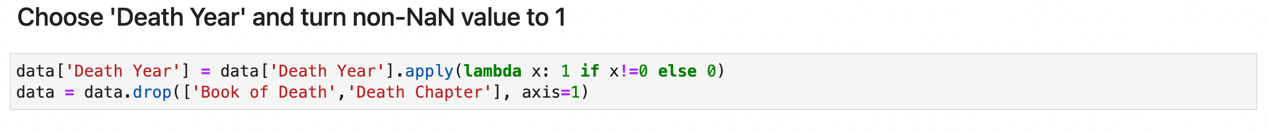
一張含有 桌 的圖片

自動產生的描述資料前處理的部分，我發現用median去填補 ‘Book Intro Chapter’ 能提升一點點準確度。

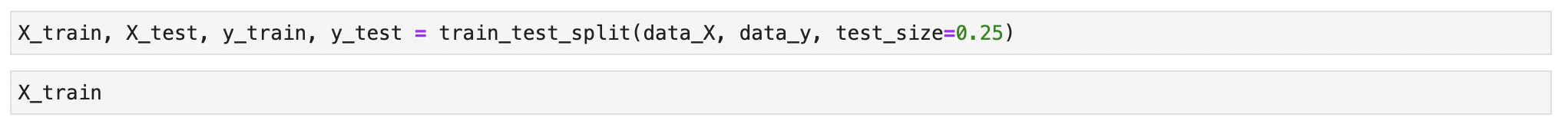


我用’Death Year’當作結果並把其餘兩者drop

一張含有 桌 的圖片

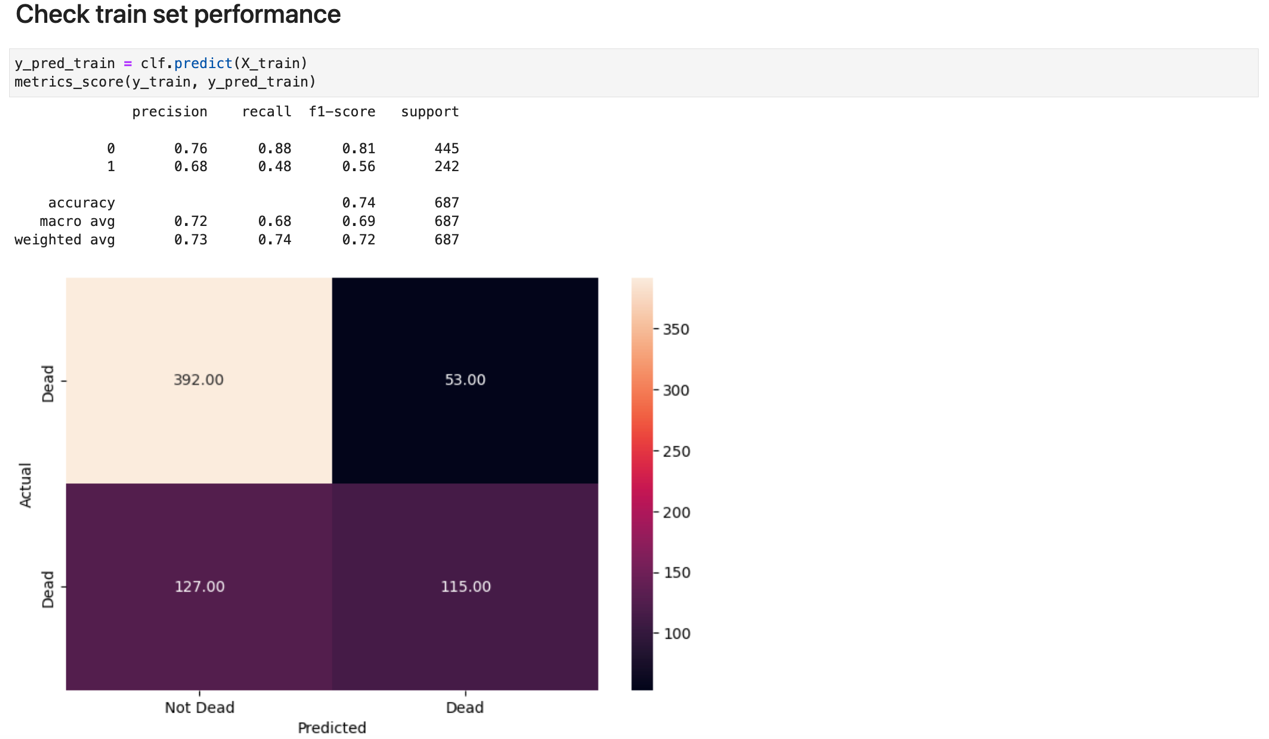
自動產生的描述

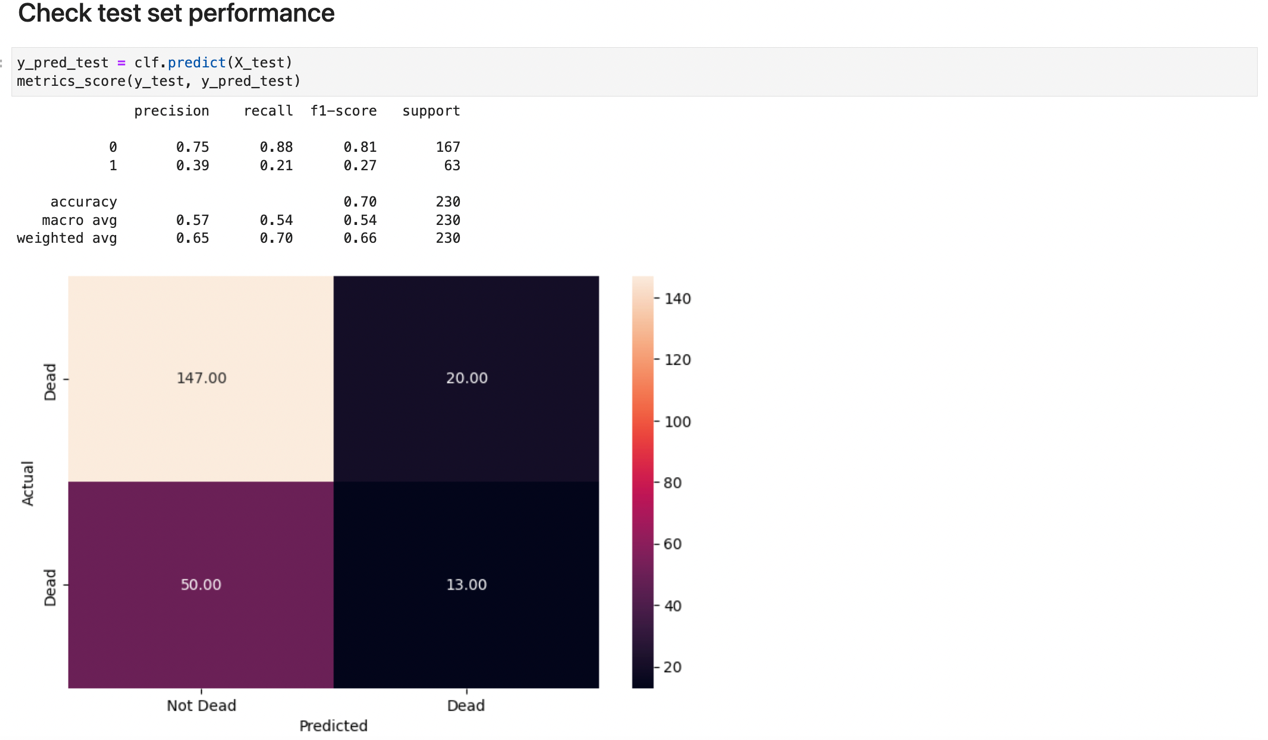
轉換成dummy的過程。先做dummy再與原本的data結合。



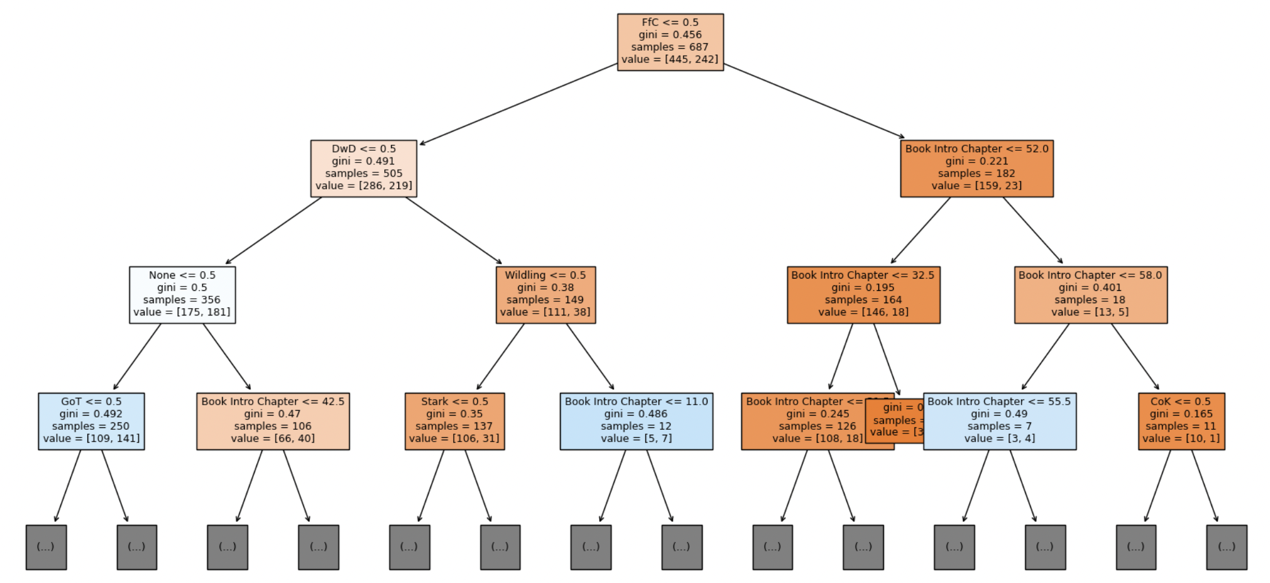
將資料切成75% training 和 25% testing。

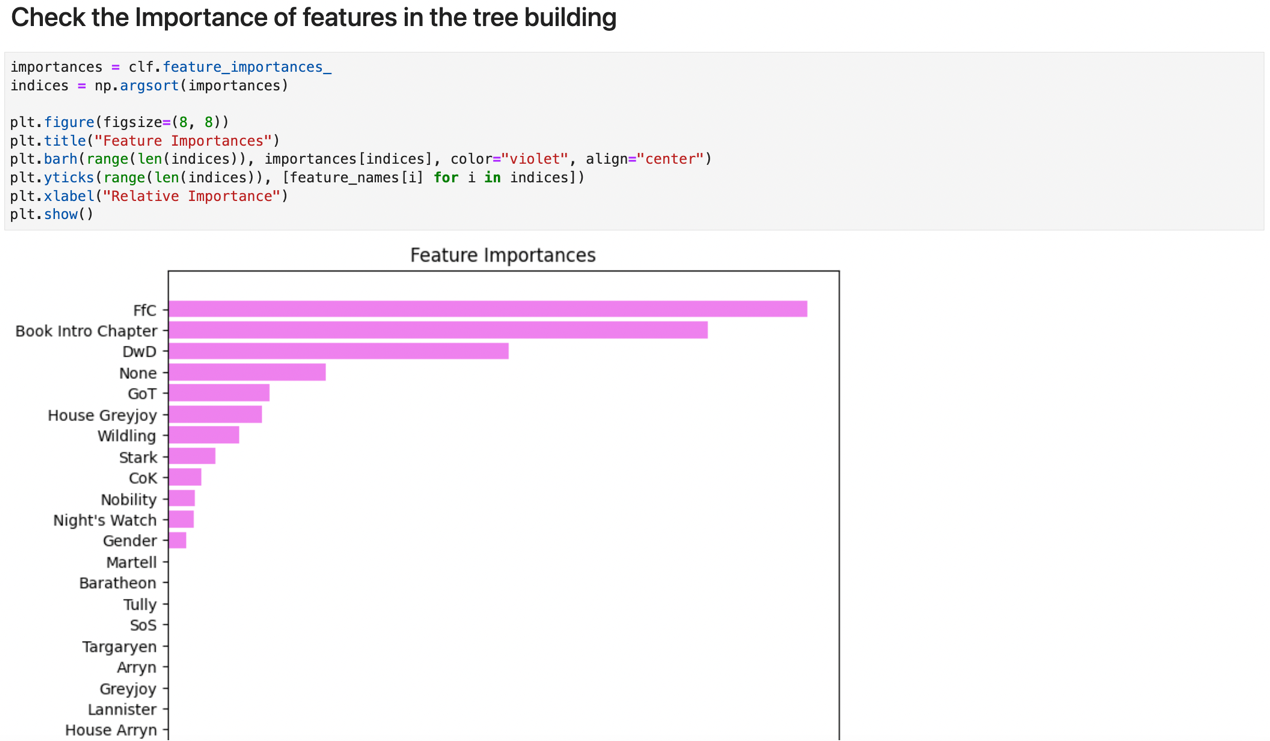
一張含有 文字 的圖片

自動產生的描述

Define ‘metric\_score’的function，並把Training model放入，檢查performance。

檢查test的結果。發現表現相較training的結果並不太好。



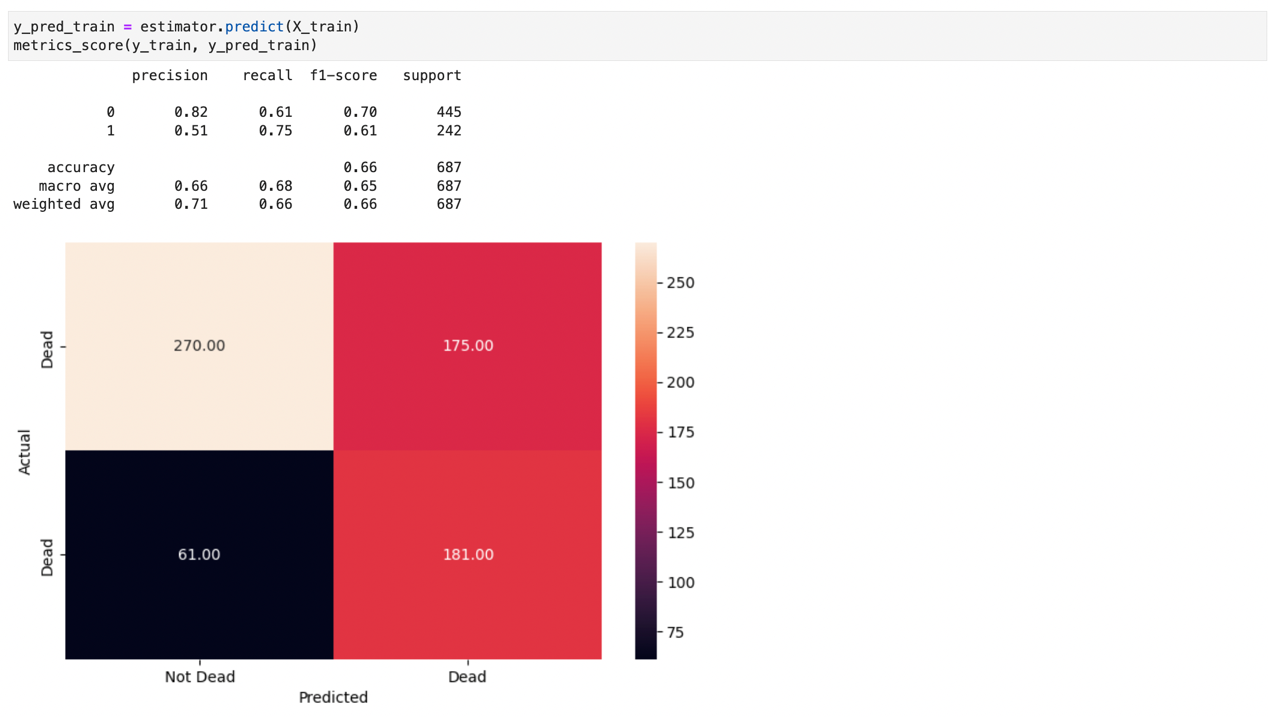
產生決策樹的圖。

可以看到各個features的重要性比較。

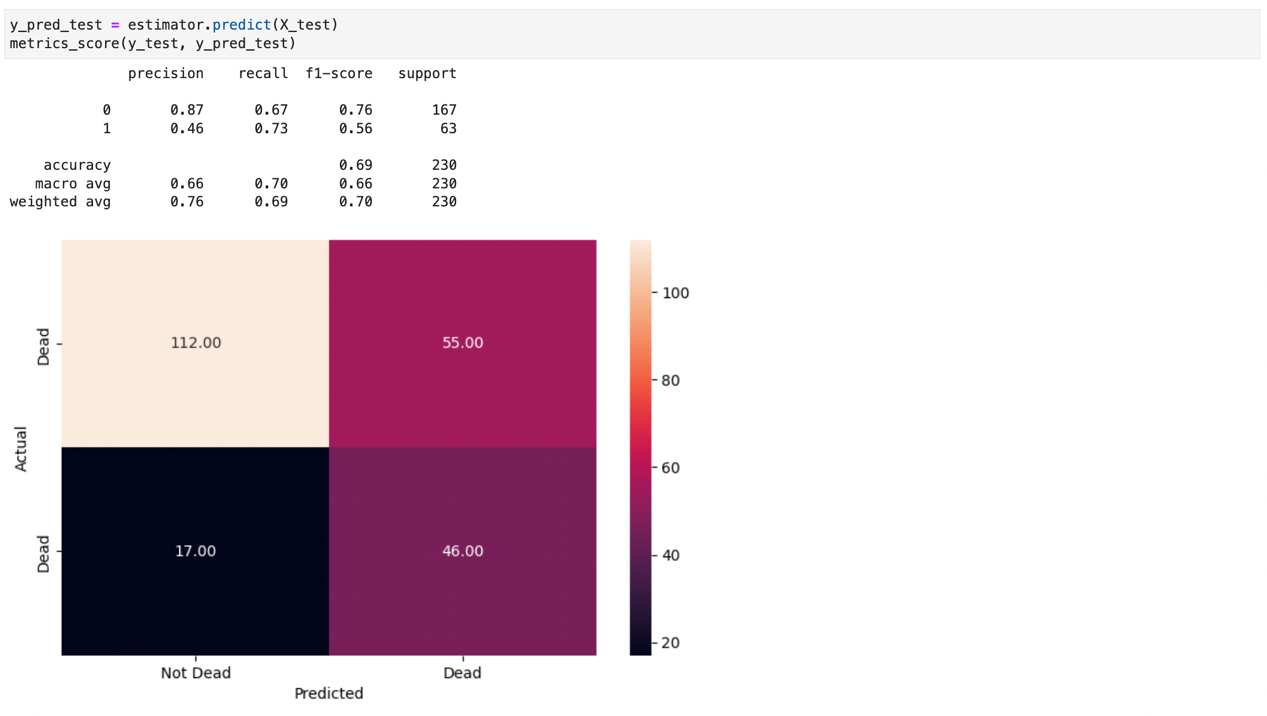
一張含有 文字 的圖片

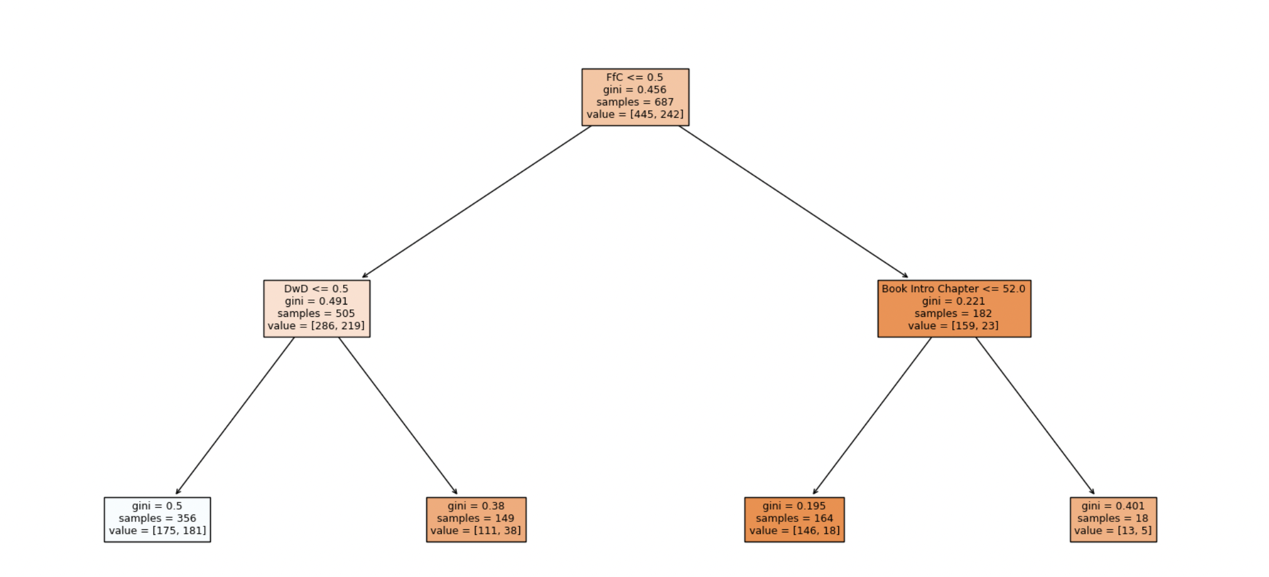
自動產生的描述

因為test的結果可以再更好，我使用Grid search去幫我找到decision tree下最好的參數。可以看到下面的結果就是找出來最佳的參數組合。

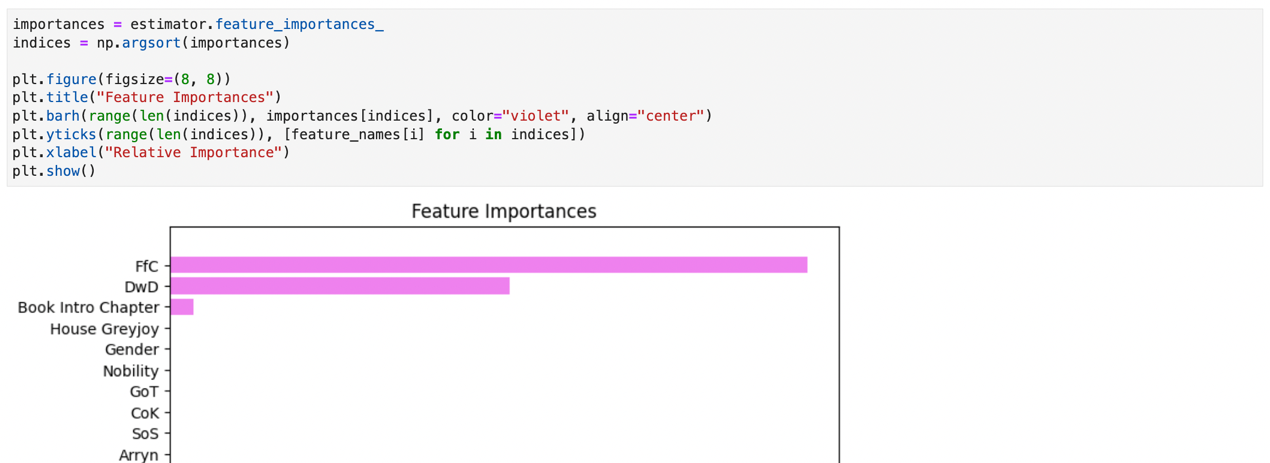


可以看到training結果，新的model跟舊的相比，並沒有比較好。



但是在test可以發現有明顯提升，代表新的model無overfitting。

新的決策樹圖。



新的model中，許多features變的沒那麼重要。