1. 問題:誰會通過資料學習這門課
2. 屬性:智商、念書時間、出席、作業分數、性別、通過

rule1 由性別去區分0(女生)直接通過這門課，1(男生)則看rule2

rule2 由智商去做分類<100直接fail ，>150看rule3，其他看rule4

rule3 由出席次數判斷>3 pass其他fail

rule4 由讀書時間判斷>4 pass 其他看rule5

rule5 由作業成績判斷>60 pass，<60則fail

1. 產生資料

gender

IQ

attend

studyhour

homework

0

1

>150

100<x<150

<100

>60

>4

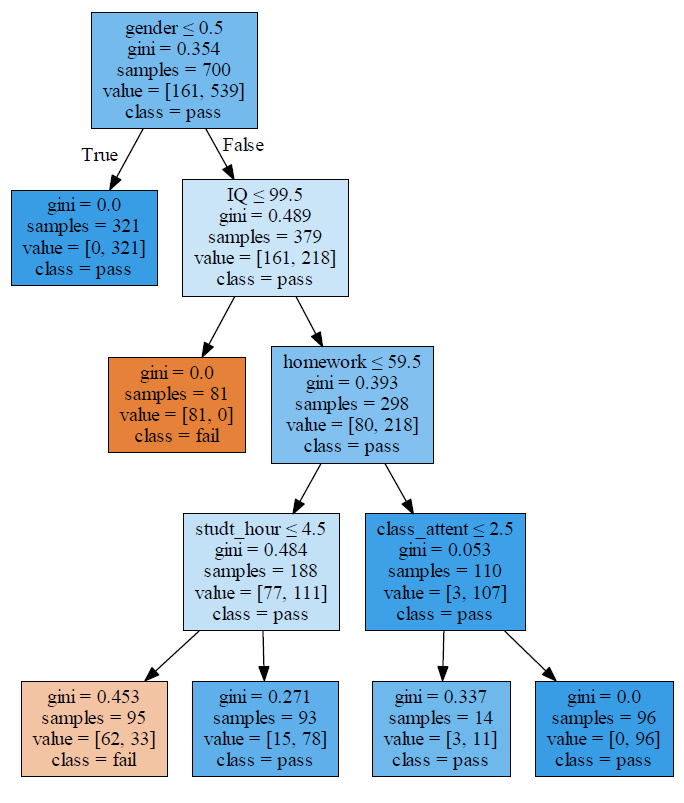
>3

<3

<4

<60

5. **decision tree**



Dicussion:

正確率0.967，算是很不錯的預測結果。並且所選的值做分類也滿接近預設值的。

覺得做出來的決策樹與上帝所建造的樹有所區隔，認為最主要的原因在於，決策樹是個二元樹，每一個node只能最多有兩個leaf。但是事實上真正的上帝規則可能不只是二分法，像是以智商來分類有分成<100、100<x<150、>150三個類別，但使用決策樹只能區分成兩個，因此在IQ這邊只分成<100與>100兩類，這可能也造成了後面class\_attent與homework的階層順序對調。

因為這個資料是使用亂數所產生出來的，因此不太會有兩筆資料相關係數高或是有相依性，如果兩個屬性具有相依性，可能在決策樹的順序上會有對調的可能。