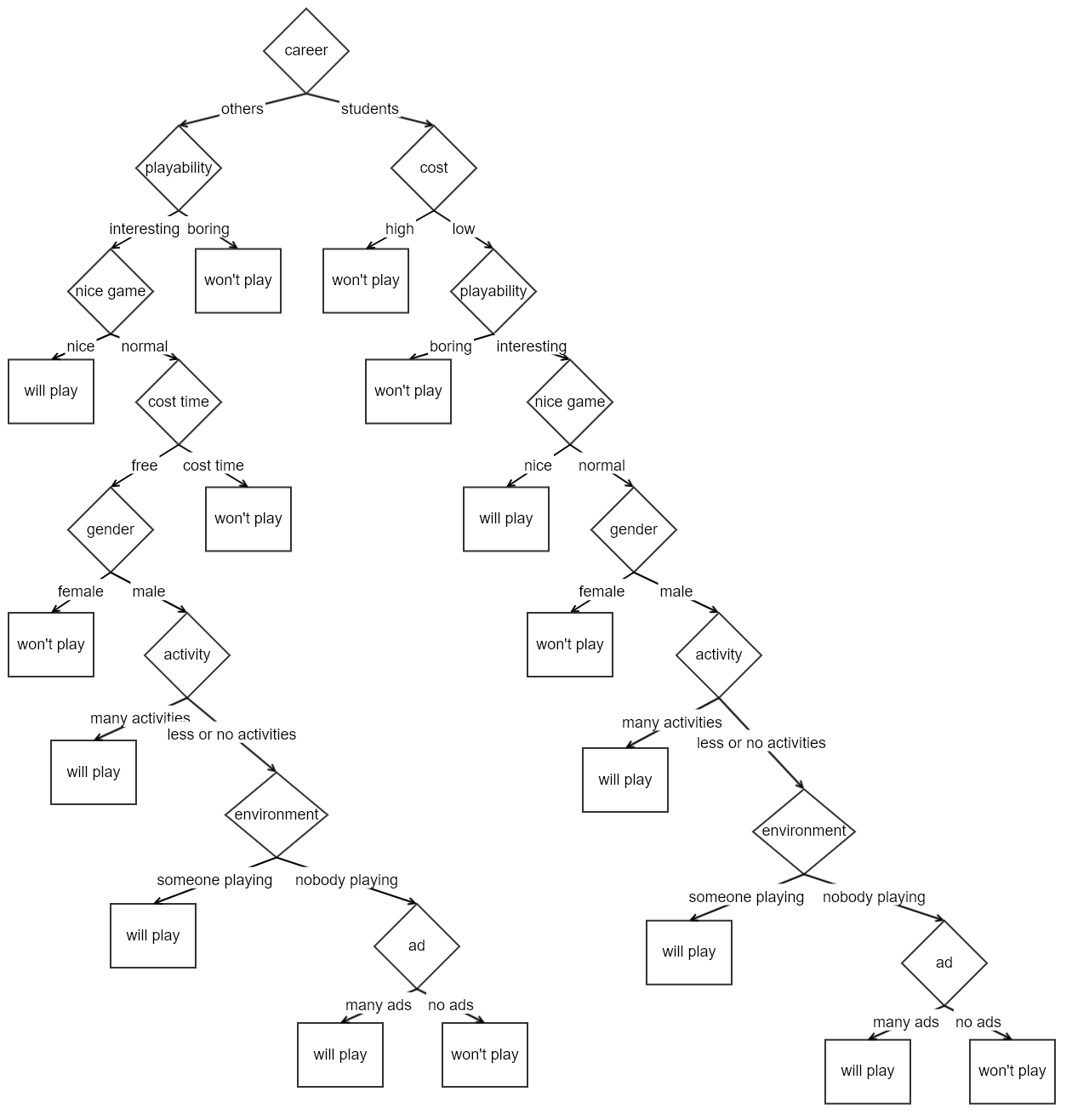
HW2 P76074046 方鈞麒

題目: 偏男性向手遊是否能使人去玩

Attribute:

1. Career(職業): student/others
2. Cost(需要花多少錢才能玩): low/high
3. Playability(遊戲性): boring/interesting
4. Nice game(是否佛心): normal/nice
5. Cost time(肝?): cost time/free
6. Gender(性別): female/male
7. Activity(合作活動): many activities/less or no activities
8. Environment(生活周遭有人在玩嗎): nobody playing/someone playing
9. Ad(廣告): no ads/many ads

Rules:



環境:

作業系統: Ubuntu 16.04

語言: Python 2.7

套件: sklearn、graphviz

編寫generate\_data.py用以產生10000筆資料

執行: python generate\_data.py

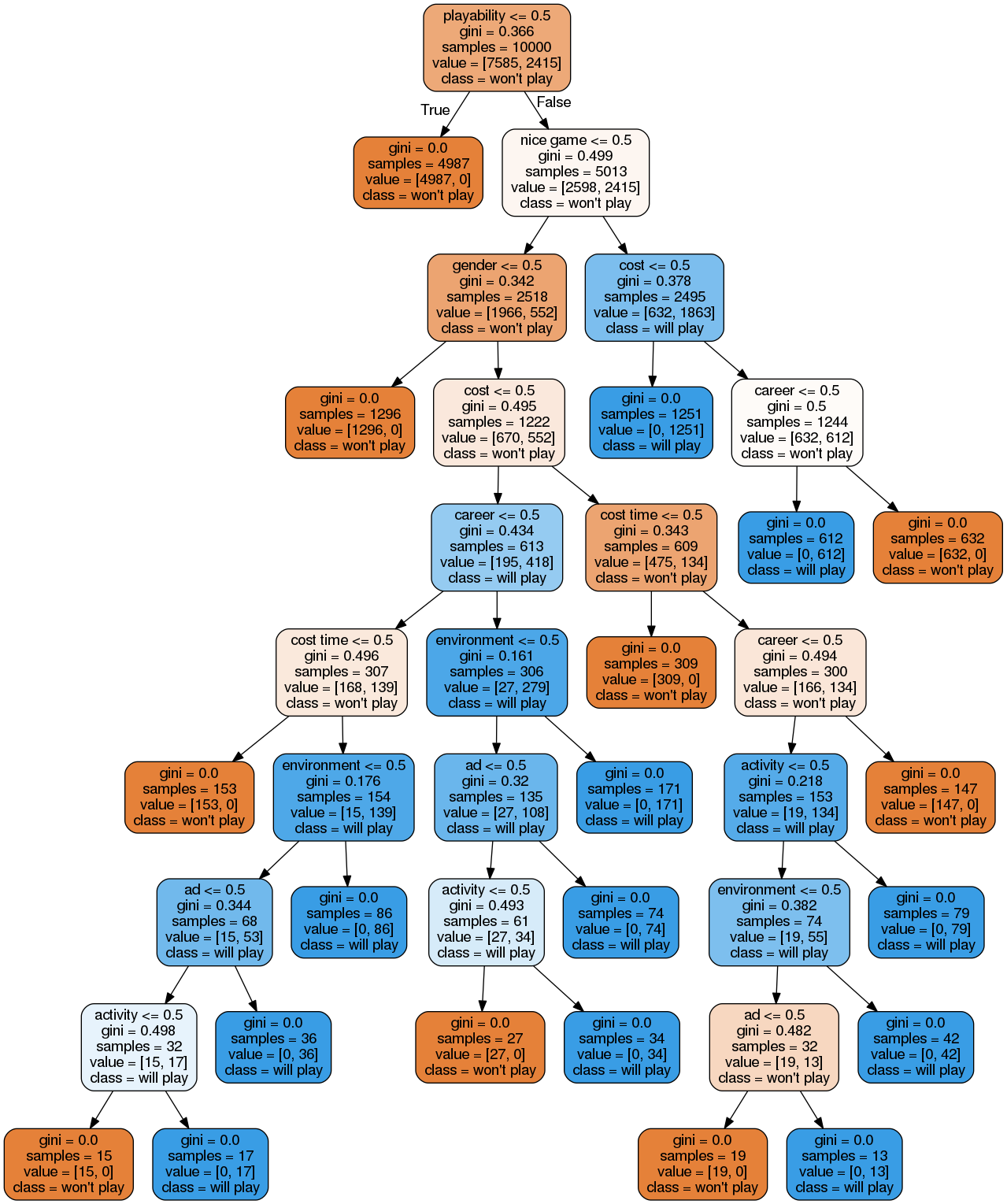
結果: 產生10000筆資料(9 attributes + 1 label)輸出至dataset.txt

編寫 hw2.py 用以產生decision tree

執行: python hw2.py

結果: 讀入dataset.txt使用sklearn套件產生decision tree、輸出tree.dot檔、並將此dot檔轉成png

Result decision tree:



Compare:

1. 結果圖明顯比我原本建立的rules來的較為平均分布
2. Attributes優先度明顯不同，我最優先考慮職業、花費、遊戲性、時間等，而結果圖從整體數據分布上得出優先考慮遊戲性、佛心、性別、時間等。
3. 最後產生的結果格子數不同，我的positive data 8個、negative data 8個，而結果圖的則是positive data 9個、negative data 11個。
4. 樹的高度不同，我的只有9層(因為學生沒有考慮花費時間、其他職業沒有考慮花費金額)，結果圖的卻達到10層(9 attributes+1 label)，此現象會導致分類時間不同。但同時，結果圖卻能將一半的data在第一步就直接歸類完，我的至少要跑到第二步才能判斷一部分，這情形卻又能幫忙減少不少分類時間。