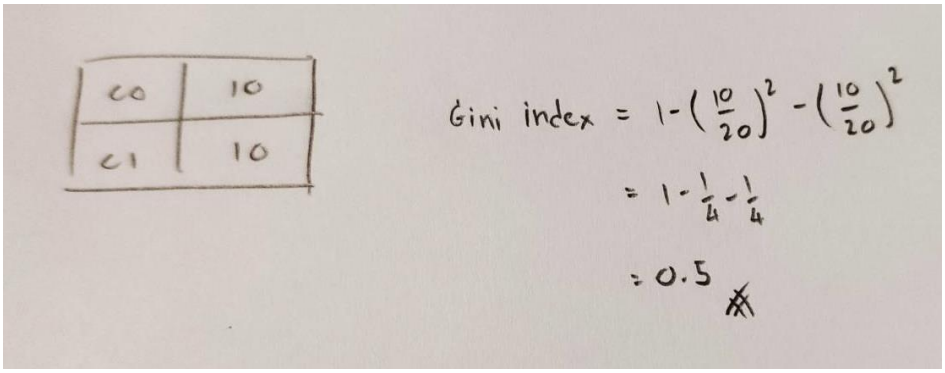
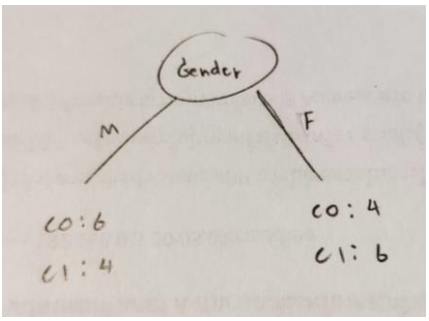
Homework 5

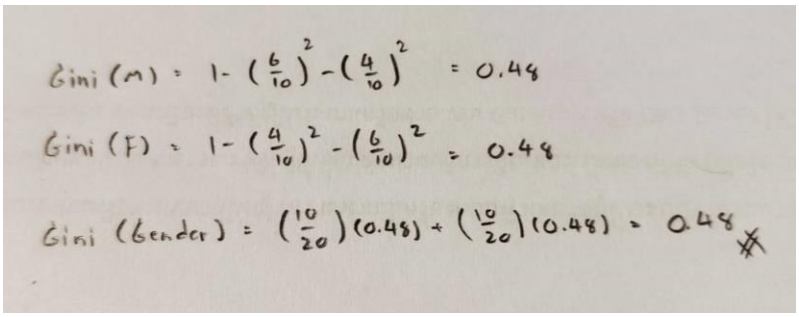
Exercise 1 (Feature selection):

(a) Compute the Gini index for the overall collection of training examples.

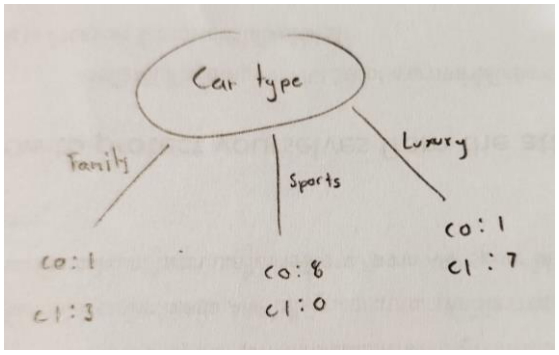


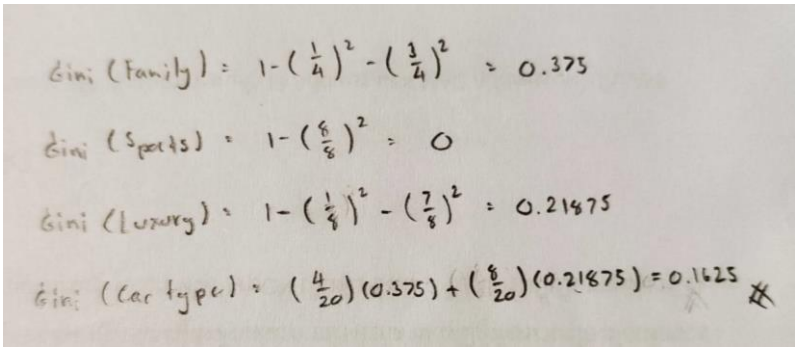
(b) Compute the Gini index for the Gender attribute.



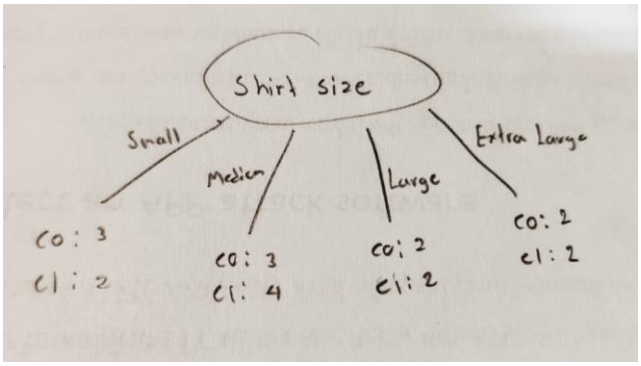


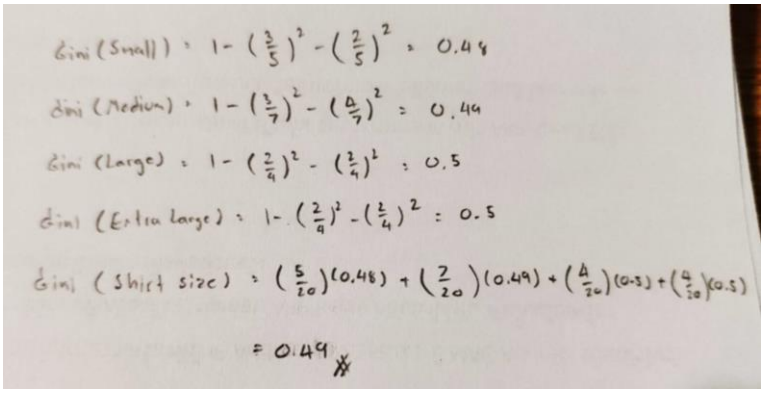
(d) Compute the Gini index for the Car Type attribute using multi-way split.





(e) Compute the Gini index for the Shirt Size attribute using multi-way split.





(f) Which attribute is better, Gender, Car Type, or Shirt Size? Why?

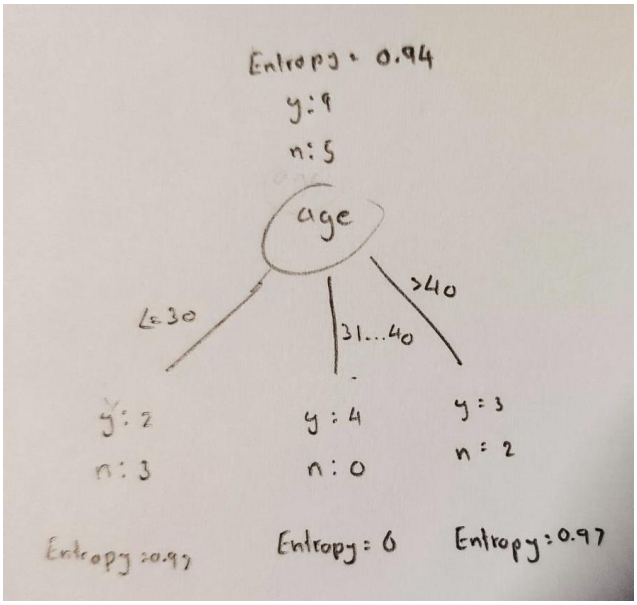
Attribute ที่ดีที่สุดคือ Car Type เนื่องจากมีค่า Gini Index ต่ำที่สุด (Gini Index = 0.1625)

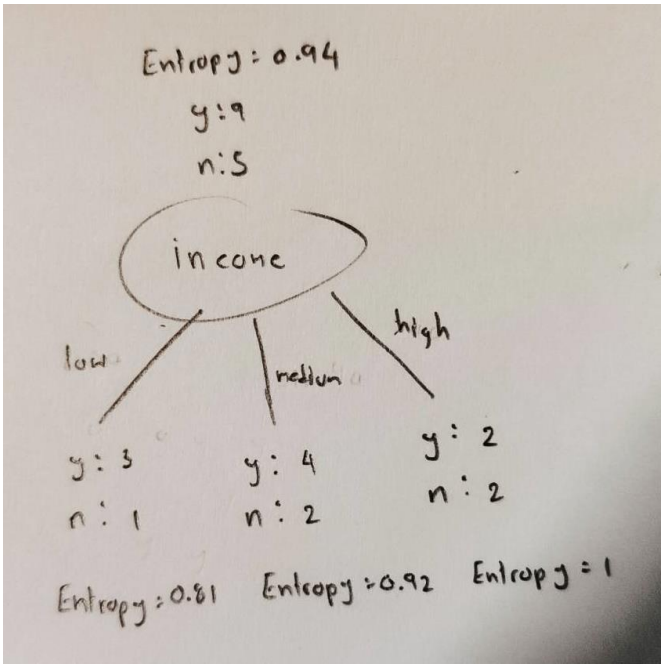
(g) Explain why Customer ID should not be used as the attribute test condition even though it has the lowest Gini.

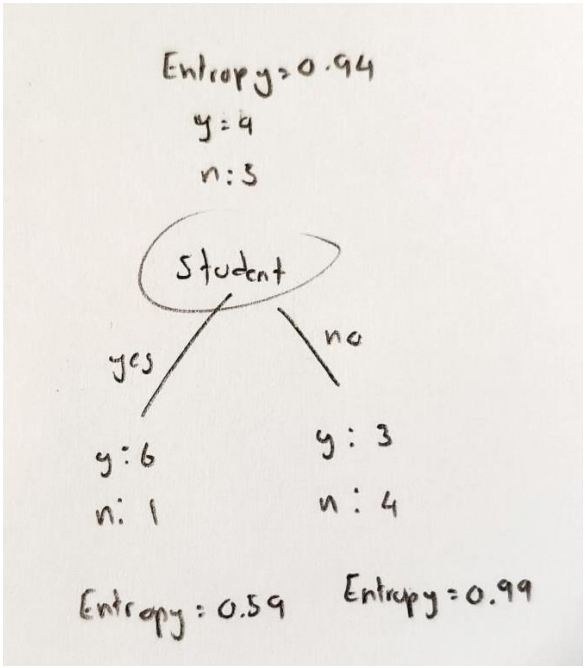
ที่ไม่เลือก Customer ID มาเป็นหนึ่งใน attribute ทั้งที่มี Gini Index ต่ำเนื่องจาก Customer ID เป็น Primary key ในตารางนี้ ทำให้มีความแตกต่างกันในทุกๆแถวและเป็นค่า Unique ของแต่ละแถว

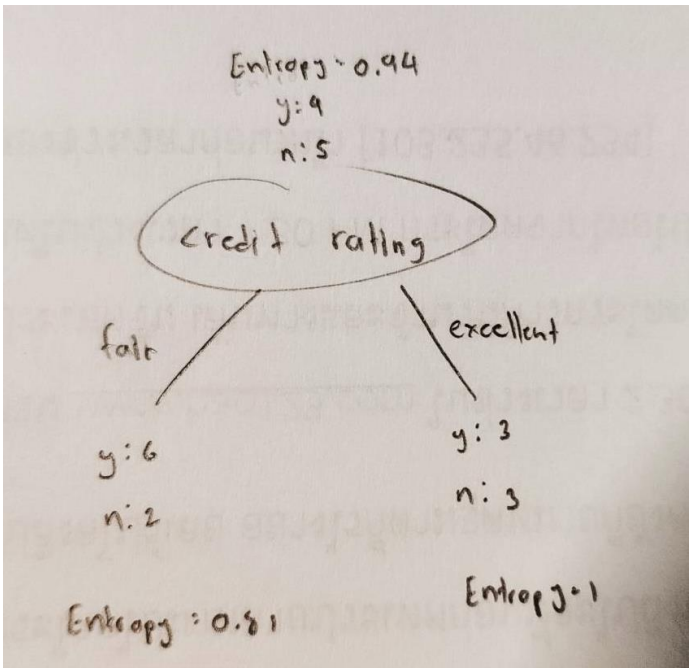
Exercise 2 (Building decision tree):

(a) Step by step, build a decision tree using Information Gain based on Entropy.

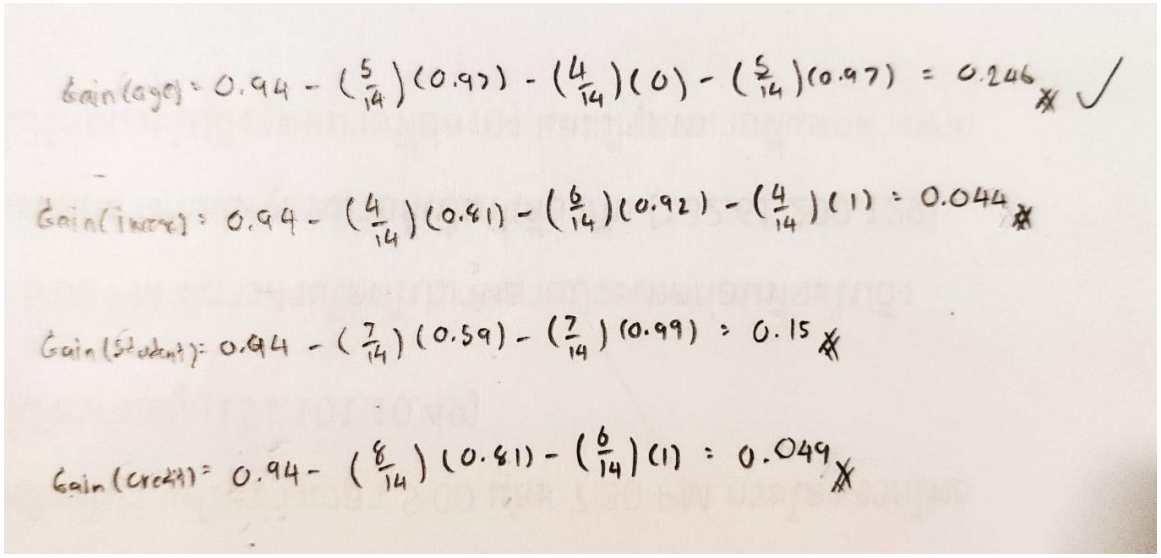


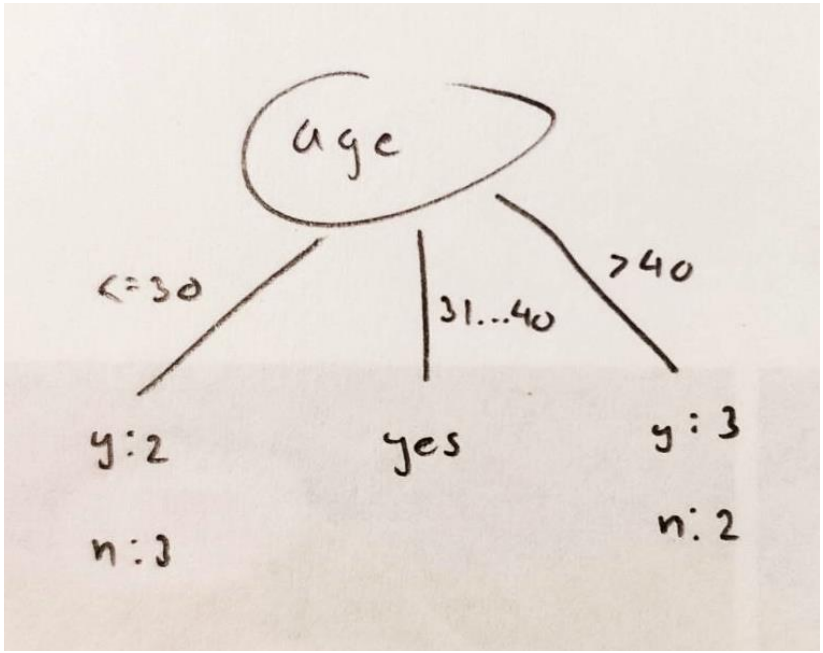


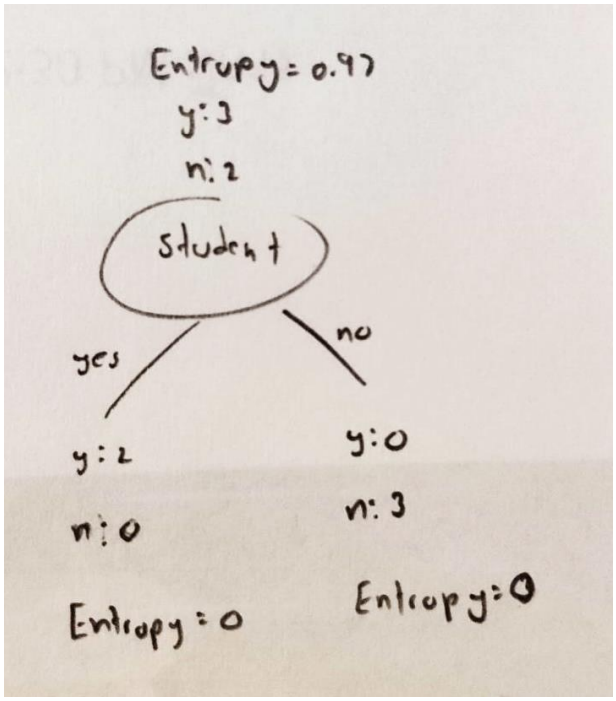


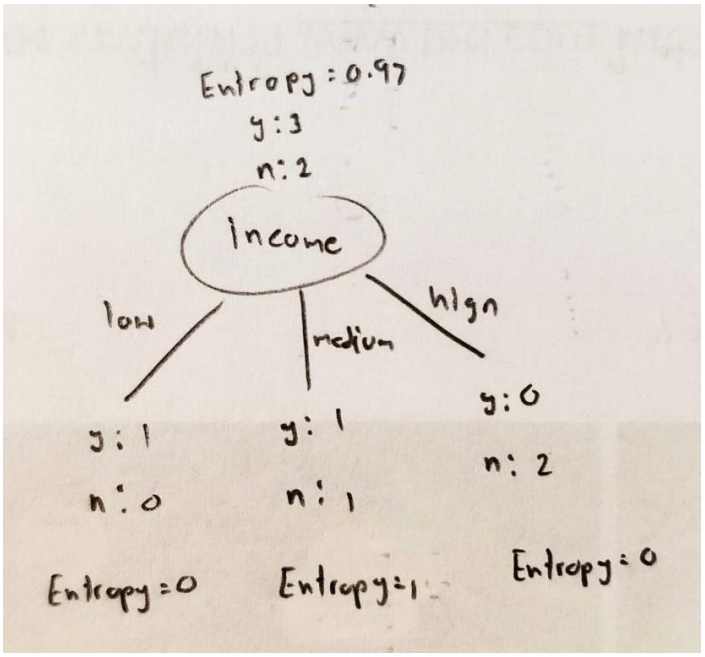


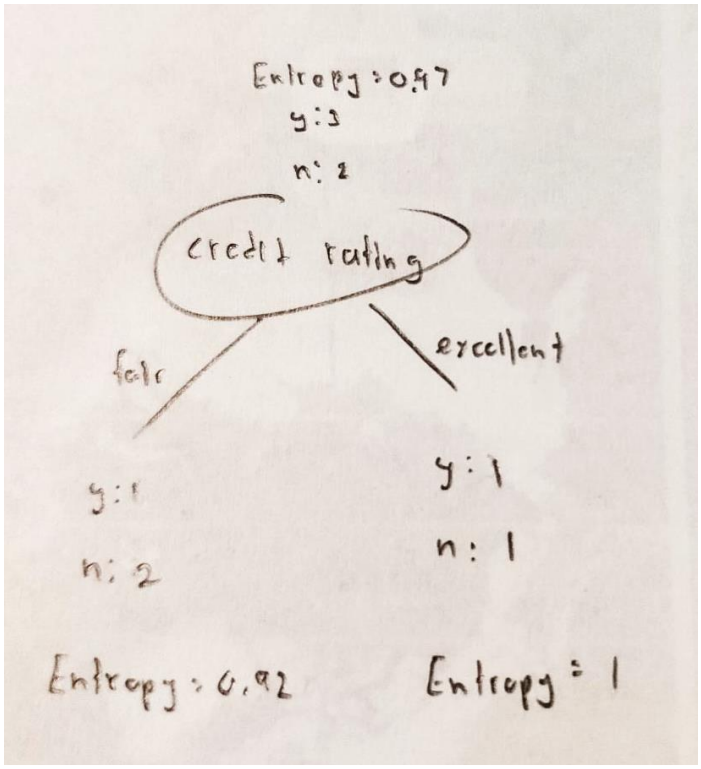
เลือก age เนื่องจากมีค่า Gain สูงที่สุดแล้วหา node ต่อไป



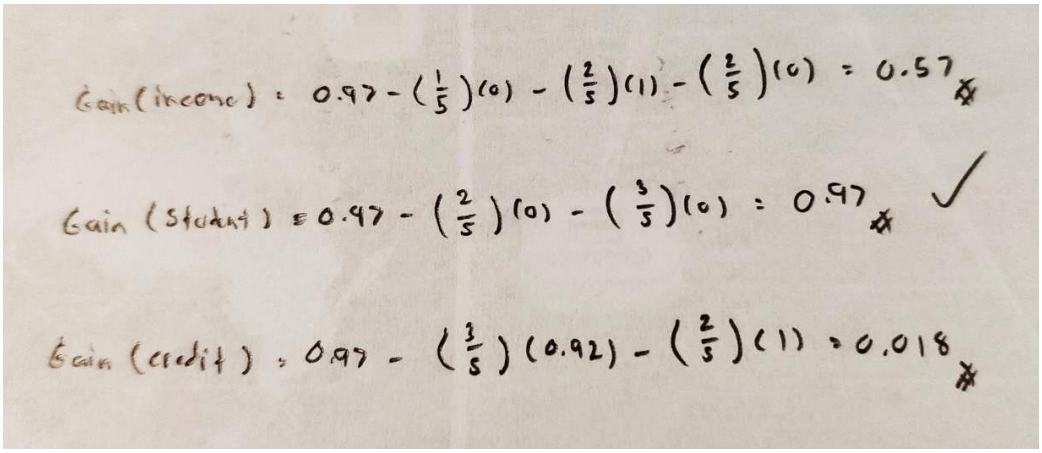


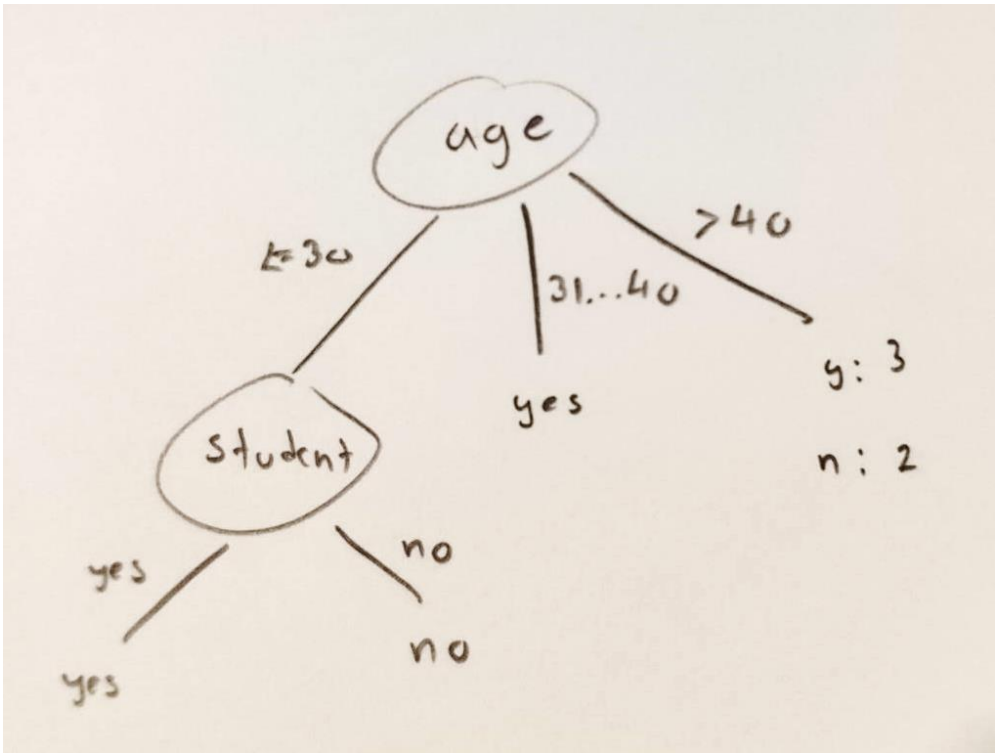

Entropy = 0.97
y: 3
n: 2
student
yes
no
y: 2
n: 0
y: 0
n: 3
Entropy = 0
Entropy = 0

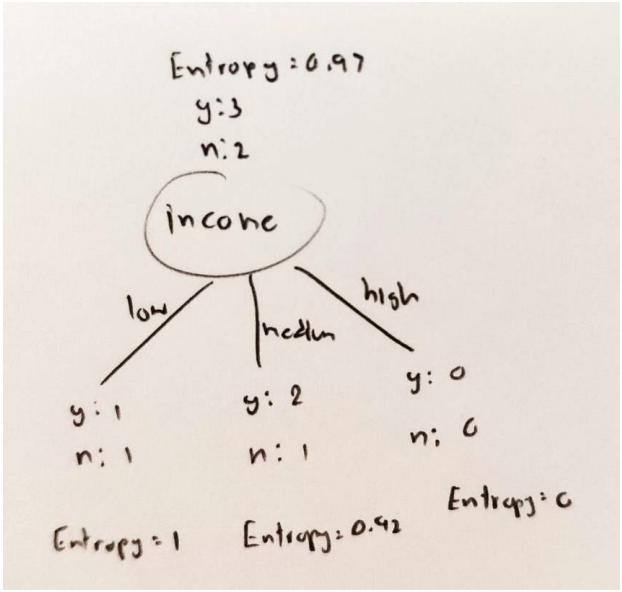

Entropy = 0.97
y: 3
n: 2
Income
low
medium
high
y: 1
n: 0
y: 1
n: 1
y: 0
n: 2
Entropy = 0
Entropy = 1
Entropy = 0

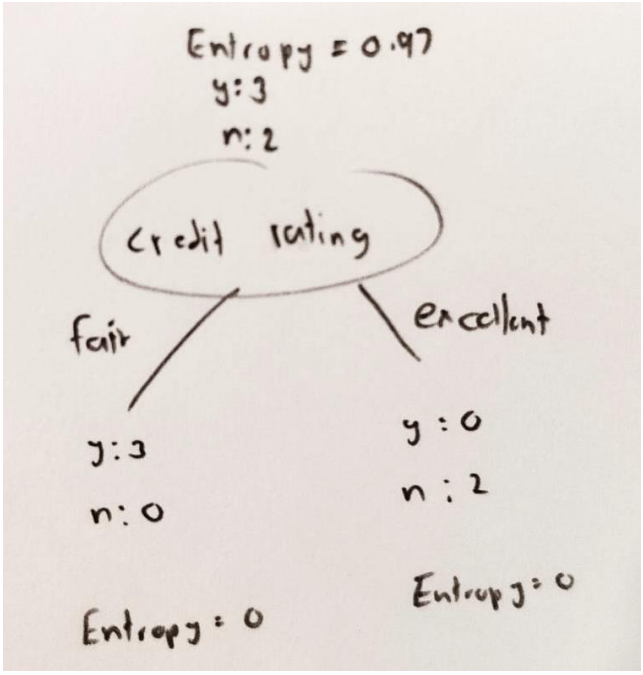

Entropy = 0.97
y: 3
n: 2
credit rating
fair
excellent
y: 1
n: 2
y: 1
n: 1
Entropy = 0.92
Entropy = 1

เลือก student เนื่องจากมีค่า Gain สูงที่สุดแล้วหา node ต่อไป

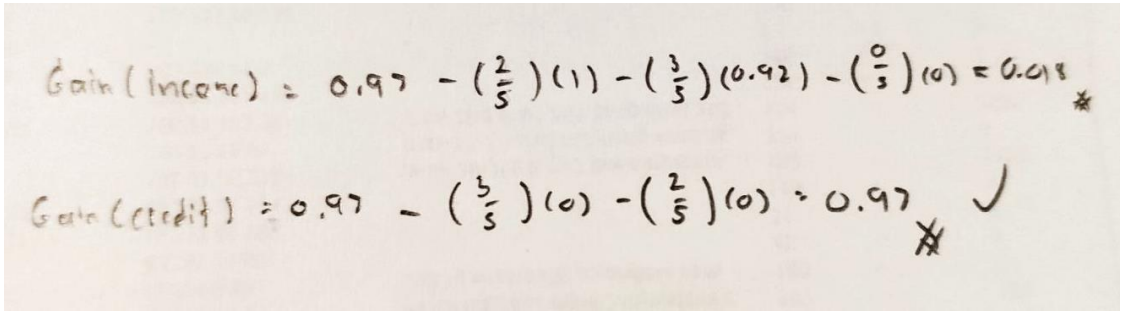


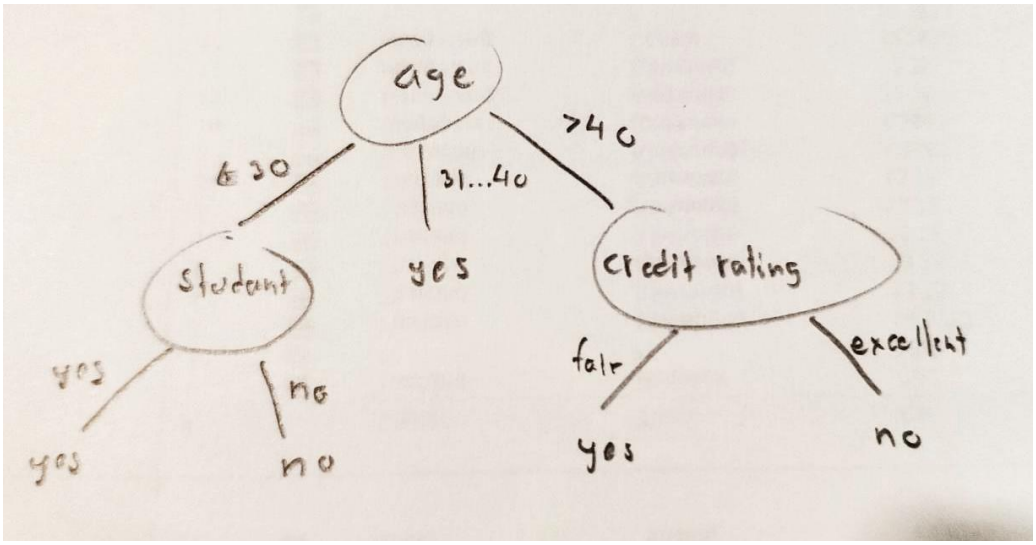






เลือก credit rating เนื่องจากมีค่า Gain สูงที่สุด หลังจากนั้นจะได้ decision tree ที่สมบูรณ์

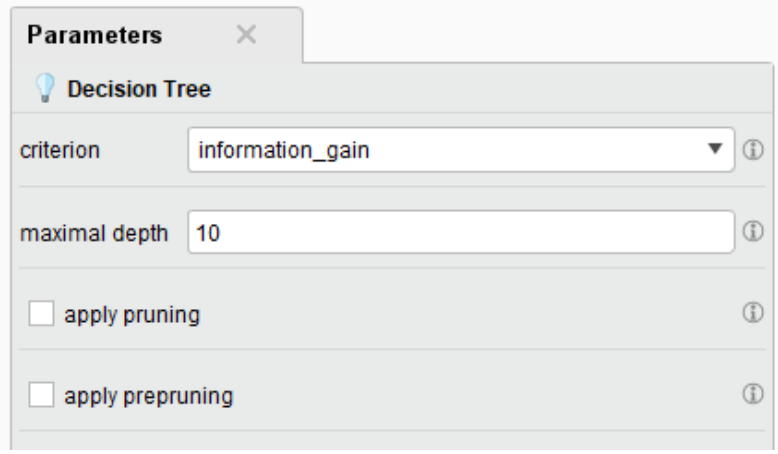


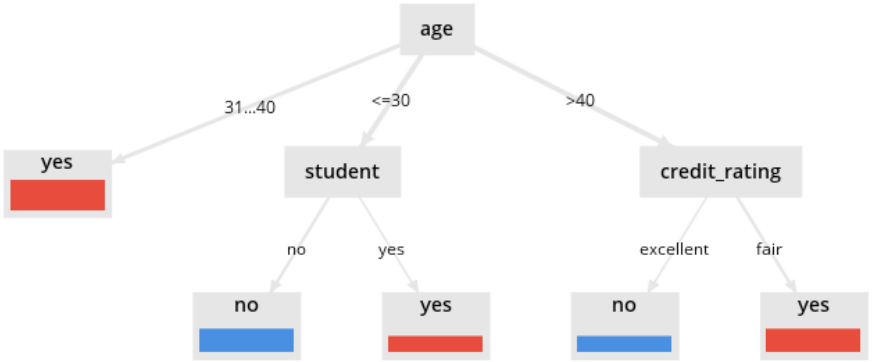


(b) Use the constructed tree to predict the class of the following new example: age<=30, income=medium, student=yes, credit-rating=fair.

Predict: yes

Exercise 3 (Building decision tree using software):





Decision tree ที่สร้างเองเหมือนกันกับที่สร้างจาก software (RapidMiner)