

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

กลุ่ม แบ่งได้ตามขนาดแผนผังป่าชอากั & อยดว้ย

Parameter \rightarrow min_impurity_decrease

① Class P : buy_computer = "yes" = 9
Class N : buy_computer = "No" = 5

$$\text{gini}(\text{buy_computer}) = 1 - \left[\left(\frac{9}{14} \right)^2 + \left(\frac{5}{14} \right)^2 \right] = 0.459$$

② age $\leq 30 \rightarrow$ "yes" = 2 , "No" = 3

age 31-40 \rightarrow "yes" = 4 , "No" = 0

age $> 40 \rightarrow$ "yes" = 3 , "No" = 2

$$\text{gini}(\leq 30) = 1 - \left[\left(\frac{2}{5} \right)^2 + \left(\frac{3}{5} \right)^2 \right] = 0.48$$

$$\text{gini}(31-40) = 1 - \left[\left(\frac{4}{4} \right)^2 + \left(\frac{0}{4} \right)^2 \right] = 0$$

$$\text{gini}(> 40) = 1 - \left[\left(\frac{3}{5} \right)^2 + \left(\frac{2}{5} \right)^2 \right] = 0.48$$

$$\therefore \text{Gini}(\text{Avg weight Age}) = \left(0.48 \times \frac{5}{14} \right) + \left(0 \times \frac{4}{14} \right) + \left(0.48 \times \frac{5}{14} \right)$$

$$= 0.343 \quad \#$$

③ income = high \rightarrow "yes" = 2 , "No" = 2

income = medium \rightarrow "yes" = 4 , "No" = 2

income = low \rightarrow "yes" = 3 , "No" = 1

$$\text{gini}(\text{high}) = 1 - \left[\left(\frac{2}{4} \right)^2 + \left(\frac{2}{4} \right)^2 \right] = 0.5$$

$$\text{gini}(\text{medium}) = 1 - \left[\left(\frac{4}{6} \right)^2 + \left(\frac{2}{6} \right)^2 \right] = 0.44$$

$$\text{gini}(\text{low}) = 1 - \left[\left(\frac{3}{4} \right)^2 + \left(\frac{1}{4} \right)^2 \right] = 0.375$$

$$\therefore \text{Gini}(\text{Avg weight income}) = \left(0.5 \times \frac{4}{14} \right) + \left(0.44 \times \frac{6}{14} \right) + \left(0.375 \times \frac{4}{14} \right)$$

$$= 0.440 \quad \#$$

④ student_yes \rightarrow "yes" = 6 , "No" = 0

student_No \rightarrow "yes" = 3 , "No" = 4

$$\text{gini}(\text{student_yes}) = 1 - \left[\left(\frac{6}{7} \right)^2 + \left(\frac{1}{7} \right)^2 \right] = 0.245$$

$$\text{gini}(\text{student_No}) = 1 - \left[\left(\frac{3}{7} \right)^2 + \left(\frac{4}{7} \right)^2 \right] = 0.489$$

$$\therefore \text{Gini}(\text{Avg weight student}) = \left(0.245 \times \frac{7}{14} \right) + \left(0.489 \times \frac{7}{14} \right) = 0.367 \quad \#$$

⑤

credit_rating_fair \rightarrow "yes" = 6, "No" = 2

$$\text{gini(fair)} = 1 - \left[\left(\frac{6}{8} \right)^2 + \left(\frac{2}{8} \right)^2 \right] = 0.375$$

credit_rating_excellent \rightarrow "yes" = 3, "No" = 3

$$\text{gini(excellent)} = 1 - \left[\left(\frac{3}{6} \right)^2 + \left(\frac{3}{6} \right)^2 \right] = 0.5$$

$$\therefore \text{Gini(Avg weight credit_rating)} = (0.375 \times \frac{8}{14}) + (0.5 \times \frac{6}{14}) = 0.429 \quad \times$$

\rightarrow จากการคำนวณค่า gini ของ age, income, student และ credit_rating แล้ว
จะเห็นว่า age มีค่า gini ต่ำที่สุด คือ 0.343 ดังนั้นจึงเลือก age ให้อยู่ใน Decision Node.

จาก age (≤ 30), $n=5$

income - high \rightarrow "yes" = 0, "No" = 2

$$\text{gini} = 1 - \left[\left(\frac{0}{2} \right)^2 + \left(\frac{2}{2} \right)^2 \right] = 0$$

- medium \rightarrow "yes" = 1, "No" = 1

$$\text{gini} = 1 - \left[\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2 \right] = 0.5$$

- low \rightarrow "yes" = 1, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{1}{1} \right)^2 + \left(\frac{0}{1} \right)^2 \right] = 0$$

$$\therefore \text{Gini(Avg weight income)} = \left[\left(0 \times \frac{2}{5} \right) + \left(0.5 \times \frac{2}{5} \right) + \left(0 \times \frac{1}{5} \right) \right]$$

$$= 0.2 \quad \times$$

student - student_yes \rightarrow "yes" = 2, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{2}{2} \right)^2 + \left(\frac{0}{2} \right)^2 \right] = 0$$

- student_no \rightarrow "yes" = 0, "No" = 3

$$\text{gini} = 1 - \left[\left(\frac{0}{3} \right)^2 + \left(\frac{3}{3} \right)^2 \right] = 0$$

$$\therefore \text{Gini(Avg weight student)} = \left[\left(0 \times \frac{2}{5} \right) + \left(0 \times \frac{3}{5} \right) \right]$$

$$= 0 \quad \times$$

credit_rating - fair = "yes" = 1, "No" = 2

$$\text{gini} = 1 - \left[\left(\frac{1}{3} \right)^2 + \left(\frac{2}{3} \right)^2 \right] = 0.4$$

- excellent = "yes" = 1, "No" = 1

$$\text{gini} = 1 - \left[\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2 \right] = 0.5$$

$$\therefore \text{Gini}(\text{Avg weight credit_rating}) = \left[\left(0.4 \times \frac{3}{5} \right) + \left(0.5 \times \frac{2}{5} \right) \right] \\ = 0.464$$

age (31-40) $n = 4$

income - high = "yes" = 2, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{2}{2} \right)^2 + \left(\frac{0}{2} \right)^2 \right] = 0$$

- medium = "yes" = 1, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{1}{1} \right)^2 + \left(\frac{0}{1} \right)^2 \right] = 0$$

- low = "yes" = 1, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{1}{1} \right)^2 + \left(\frac{0}{1} \right)^2 \right] = 0$$

$$\therefore \text{Gini}(\text{Avg weight income}) = \left[\left(0 \times \frac{2}{4} \right) + \left(0 \times \frac{1}{4} \right) + \left(0 \times \frac{1}{4} \right) \right] \\ = 0 //$$

student - Student_yes = "yes" = 2, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{2}{2} \right)^2 + \left(\frac{0}{2} \right)^2 \right] = 0$$

- Student_no = "yes" = 2, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{2}{2} \right)^2 + \left(\frac{0}{2} \right)^2 \right] = 0$$

$$\therefore \text{Gini}(\text{Avg weight student}) = \left[\left(0 \times \frac{2}{4} \right) + \left(0 \times \frac{2}{4} \right) \right] \\ = 0 //$$

credit_rating - fair = "yes" = 2, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{2}{2} \right)^2 + \left(\frac{0}{2} \right)^2 \right] = 0$$

- excellent = "yes" = 2, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{2}{2} \right)^2 + \left(\frac{0}{2} \right)^2 \right] = 0$$

$$\therefore \text{Gini(Avg weight credit_rating)} = \left[(0 \times \frac{2}{4}) + (0 \times \frac{2}{4}) \right] \\ = 0 \quad \#$$

age > 40, n = 5

income - high = "yes" = 0, "No" = 0

$$\text{gini} = 1 - [0 + 0] = 0$$

- medium = "yes" = 2, "No" = 1

$$\text{gini} = 1 - \left[\left(\frac{2}{3} \right)^2 + \left(\frac{1}{3} \right)^2 \right] = 0.44$$

- low = "yes" = 1, "No" = 1

$$\text{gini} = 1 - \left[\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2 \right] = 0.5$$

$$\therefore \text{Gini(Avg weight income)} = \left[(0 \times \frac{0}{5}) + (0.44 \times \frac{3}{5}) + (0.5 \times \frac{2}{5}) \right] \\ = 0.464 \quad \#$$

student - student_yes = "yes" = 2, "No" = 1

$$\text{gini} = 1 - \left[\left(\frac{2}{3} \right)^2 + \left(\frac{1}{3} \right)^2 \right] = 0.44$$

- student_no = "yes" = 1, "No" = 1

$$\text{gini} = 1 - \left[\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2 \right] = 0.5$$

$$\therefore \text{Gini(Avg weight student)} = \left[(0.44 \times \frac{3}{5}) + (0.5 \times \frac{2}{5}) \right] \\ = 0.464 \quad \#$$

credit_rating - fair = "yes" = 3, "No" = 0

$$\text{gini} = 1 - \left[\left(\frac{3}{3} \right)^2 + \left(\frac{0}{3} \right)^2 \right] = 0$$

- excellent = "yes" = 0, "No" = 2

$$\text{gini} = 1 - \left[\left(\frac{0}{2} \right)^2 + \left(\frac{2}{2} \right)^2 \right] = 0$$

$$\therefore \text{Gini(Avg weight credit_rating)} = \left[\left(0 \times \frac{3}{5} \right) + \left(0 \times \frac{2}{5} \right) \right] \\ = 0 \quad \#$$

ค่า min_impurity_decrease

กำหนดให้ min_impurity_decrease = 0.1

จากสูตร

The weighted impurity decrease equation is the following:

$$N_t / N * (\text{impurity} - N_{t_R} / N_t * \text{right_impurity} - N_{t_L} / N_t * \text{left_impurity})$$

where N is the total number of samples, N_t is the number of samples at the current node, N_{t_L} is the number of samples in the left child, and N_{t_R} is the number of samples in the right child.

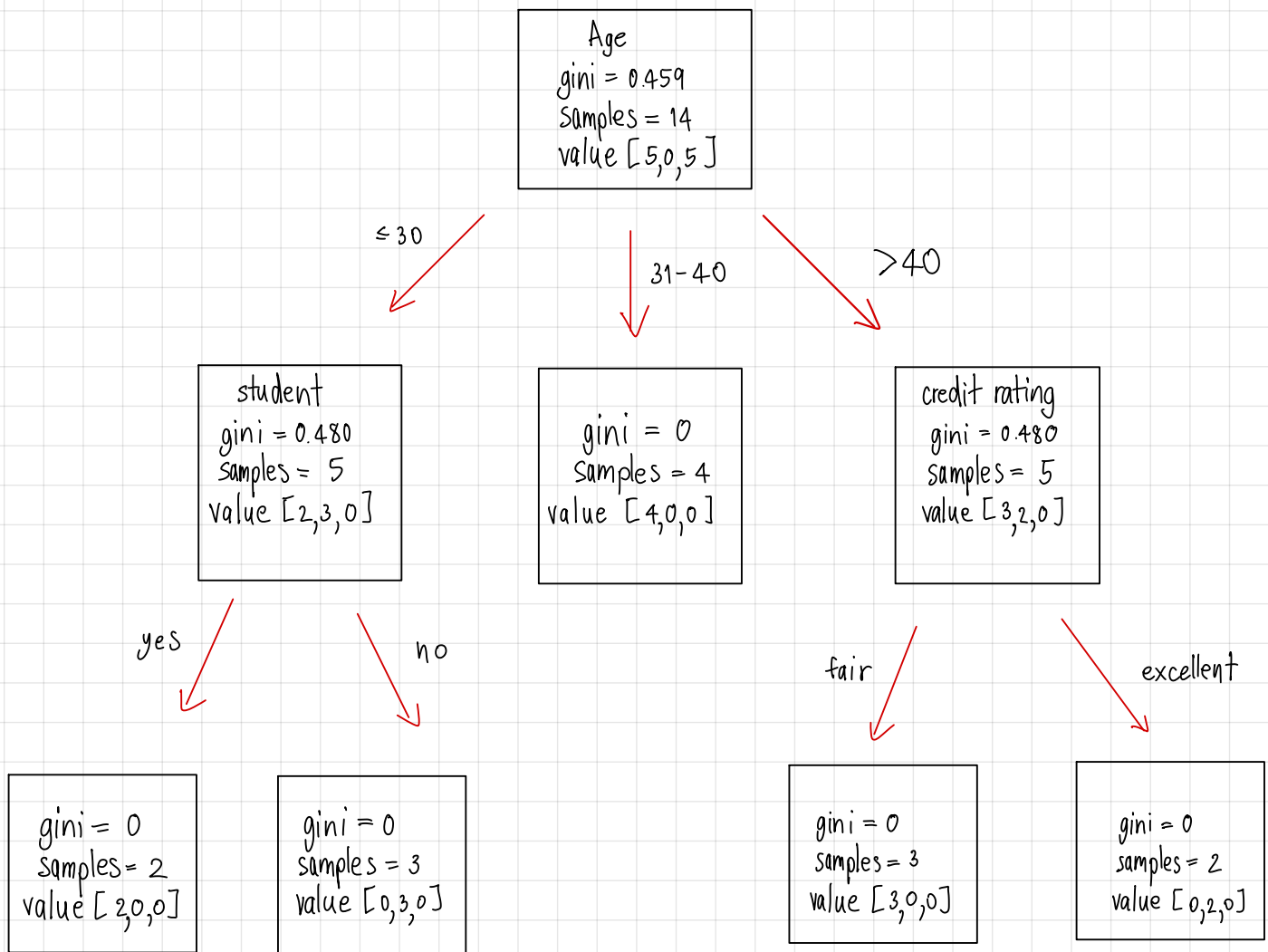
จ.ไอ ✓ Age $\rightarrow \frac{14}{14} \times \left(0.459 - \frac{5}{14} (0.48) - 0 - \frac{5}{14} (0.48) \right)$
 $= 0.117$ ซึ่ง ≥ 0.1 จึงไม่มีการแยก Node นี้ออก $\#$

income $\rightarrow \frac{14}{14} \times \left(0.459 - \frac{4}{14} (0.5) - \frac{6}{14} (0.44) - \frac{4}{14} (0.375) \right)$
 $= 0.020$ ซึ่ง < 0.1 จึงไม่มีการแยก Node $\#$

student $\rightarrow \frac{14}{14} \times \left(0.459 - \frac{7}{14} (0.245) - \frac{7}{14} (0.489) \right)$
 $= 0.092$ ซึ่ง < 0.1 จึงไม่มีการแยก Node $\#$

credit_rating $\rightarrow \frac{14}{14} \times \left(\frac{8}{14} \times (0.375) - \frac{6}{14} (0.5) \right)$
 $= 0$ ซึ่ง < 0.1 จึงไม่มีการแยก Node $\#$

วิธีได้ Decision tree ดังนี้



- เนื่องจากใน age < 30 student มีค่า gini = 0 ซึ่งมีความน้อยที่สุด จึงไม่มีการแยก Node ต่อ
- เนื่องจากใน age 31-40 income, student และ credit_rating มีค่าเป็น 0 ดังนั้นจึงไม่มีการแยก Node ต่อ
- เนื่องจากใน age > 40 credit rating มีค่า gini = 0 ซึ่งมีความน้อยที่สุด จึงไม่มีการแยก Node ต่อ

สรุป

- จากข้อมูล การซื้อคอมพิวเตอร์ (buy computer) ที่มีตัวอย่าง คือ 14 คน
- โดยที่ - ในอายุ น้อยกว่า 30 จะมีนักเรียน ซื้อคอมพิวเตอร์ 2 คน และคนที่ไม่ใช่ นักเรียน
จะไม่ซื้อคอมพิวเตอร์ 3 คน
- ในอายุ 31 ถึง 40 จะซื้อคอมพิวเตอร์ 7 คน
- ในอายุ 40 ปีขึ้นไป ผู้ที่มี credit rating ในระดับ fair จะซื้อคอมพิวเตอร์ 3 คน
และ ผู้ที่มี credit rating ในระดับ excellent จะไม่ซื้อคอมพิวเตอร์ 2 คน

สมาชิก

กลุ่ม แต่งไฮมาดเมहनุ่นปานอร่าโก้ & กลุ่ม ช่วยด้วย

1. นางสาวทิพย์วัลย์ ลู่โพธิ์ รหัส 643020062-6
2. นางสาวกรวรรณ อู่จอหอ รหัส 643020495-5
3. พายชนบถี่ ภูชมศรี รหัส 643020502-4
4. นางสาวลลลัธรรดา แดงสูงเนิน รหัส 643020518-9
5. หายศิริโชค ศิริวิชา รหัส 643020522-8
6. นางสาวศานันท์ ทวีพงษ์มีพนาศัล รหัส 643021272-0
7. นางสาวสุชาภา พิลาก รหัส 643021276-2
8. นางสาวอึ้งอร พลพาล รหัส 643021279-6