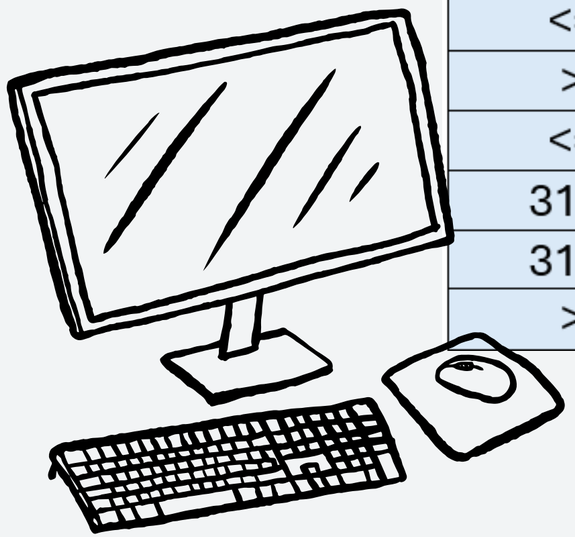
DECISION TREE PARAMETER

—>

MIN IMPURITY DECREASE

ตารางข้อมูล

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



กำหนด MIN IMPURITY DECREASE = 0.1

การคำนวณค่า GINI

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

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

② age ≤ 30 → "yes" = 2 , "No" = 3 | $gini(\leq 30) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right] = 0.48$

age 31...40 → "yes" = 4 , "No" = 0 | $gini(31...40) = 1 - \left[\left(\frac{4}{4}\right)^2 + \left(\frac{0}{4}\right)^2\right] = 0$

age >40 → "yes" = 3 , "No" = 2 | $gini(>40) = 1 - \left[\left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right] = 0.48$

$$\therefore Gini(Avg\ weight\ Age) = (0.48 \times \tfrac{5}{14}) + (0 \times \tfrac{4}{14}) + (0.48 \times \tfrac{5}{14})$$
$$= 0.343$$

③ income = high → "yes" = 2 , "No" = 2 | $gini(high) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 0.5$

income = medium → "yes" = 4 , "No" = 2 | $gini(medium) = 1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right] = 0.44$

income = low → "yes" = 3 , "No" = 1 | $gini(low) = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = 0.375$

$$\therefore Gini(Avg\ weight\ income) = (0.5 \times \tfrac{4}{14}) + (0.44 \times \tfrac{6}{14}) + (0.375 \times \tfrac{4}{14})$$
$$= 0.440$$

④

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

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

② age ≤ 30 → "yes" = 2 , "No" = 3 | $gini(\leq 30) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right] = 0.48$

age 31...40 → "yes" = 4 , "No" = 0 | $gini(31...40) = 1 - \left[\left(\frac{4}{4}\right)^2 + \left(\frac{0}{4}\right)^2\right] = 0$

⑤ age >40 → "yes" = 3 , "No" = 2 | $gini(>40) = 1 - \left[\left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right] = 0.48$

$$\therefore Gini(Avg\ weight\ Age) = (0.48 \times \tfrac{5}{14}) + (0 \times \tfrac{4}{14}) + (0.48 \times \tfrac{5}{14})$$
$$= 0.343$$

∴

→ ③ income = high → "yes" = 2 , "No" = 2 | $gini(high) = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 0.5$

income = medium → "yes" = 4 , "No" = 2 | $gini(medium) = 1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right] = 0.44$

income = low → "yes" = 3 , "No" = 1 | $gini(low) = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = 0.375$

$$\therefore Gini(Avg\ weight\ income) = (0.5 \times \tfrac{4}{14}) + (0.44 \times \tfrac{6}{14}) + (0.375 \times \tfrac{4}{14})$$
$$= 0.440$$

AGE<=30,N=5

จาก age (≤ 30), $n = 5$

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

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

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

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

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

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

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

$= 0.2$ ✗

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

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

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

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

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

$= 0$ ✗

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

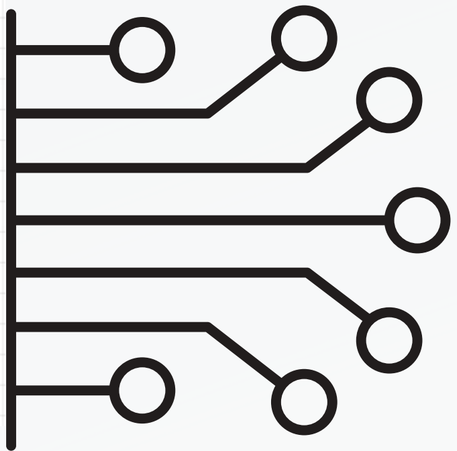
$gini = 1-\left[\left(\frac{1}{3}\right)^2+\left(\frac{2}{3}\right)^2\right] = 0.4$

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

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

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

$= 0.464$



AGE =31..40,N=4

จาก age (31-40) n = 4

income – high = "yes" = 2 , "No" = 0

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

– medium = "yes" = 1 , "No" = 0

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

– low = "yes" = 1 , "No" = 0

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

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

= 0 //

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

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

– Student_no = "yes" = 2 , "No" = 0

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

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

= 0 //

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

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

– excellent = "yes" = 2 , "No" = 0

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

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

= 0 //

AGE >40,N=5

จาก age > 40 , n = 5

income
- high = "yes" = 0 , "No" = 0
  gini = $1-\left[0+0\right] = 0$
- medium = "yes" = 2 , "No" = 1
  gini = $1-\left[(\frac{2}{3})^2+(\frac{1}{3})^2\right] = 0.44$
- low = "yes" = 1 , "No" = 1
  gini = $1-\left[(\frac{1}{2})^2+(\frac{1}{2})^2\right] = 0.5$

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

$= 0.464$ ✳

student
- student_yes = "yes" = 2 , "No" = 1
  gini = $1-\left[(\frac{2}{3})^2+(\frac{1}{3})^2\right] = 0.44$
- student_no = "yes" = 1 , "No" = 1
  gini = $1-\left[(\frac{1}{2})^2+(\frac{1}{2})^2\right] = 0.5$

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

$= 0.464$ ✳

credit_rating
- fair = "yes" = 3 , "No" = 0
  gini = $1-\left[(\frac{3}{3})^2+(\frac{0}{3})^2\right] = 0$
- excellent = "yes" = 0 , "No" = 2
  gini = $1-\left[(\frac{0}{2})^2+(\frac{2}{2})^2\right] = 0$

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

$= 0$ ✳

กำหนดให้ค่า MIN_IMPURITY_DECREASE =0.1

ค่า min_impurity_decrease

กำหนดให้ min_impurity_decrease = 0.1

จากสูตร The weighted impurity decrease equation is the following:

```
N_t / N * (impurity - N_t_R / N_t * right_impurity
                    - N_t_L / N_t * 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.

จะได้ Age → $\frac{14}{14} \times \left(0.459 - \frac{5}{14}(0.48) - 0 - \frac{5}{14}(0.48)\right)$

= 0.117 ซึ่ง ≥ 0.1 จึงมีการแยก Node ต่อไป ✓

income → $\frac{14}{14} \times \left(0.459 - \frac{4}{14}(0.5) - \frac{6}{14}(0.44) - \frac{4}{14}(0.375)\right)$

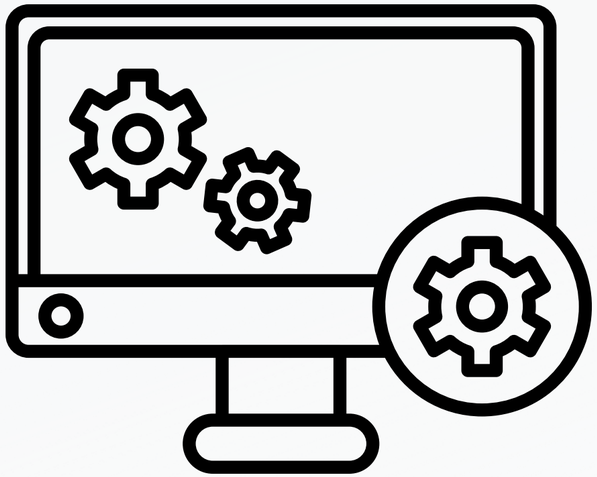
= 0.020 ซึ่ง < 0.1 จึงไม่มีการแยก Node ✗

student → $\frac{14}{14} \times \left(0.459 - \frac{7}{14}(0.245) - \frac{7}{14}(0.489)\right)$

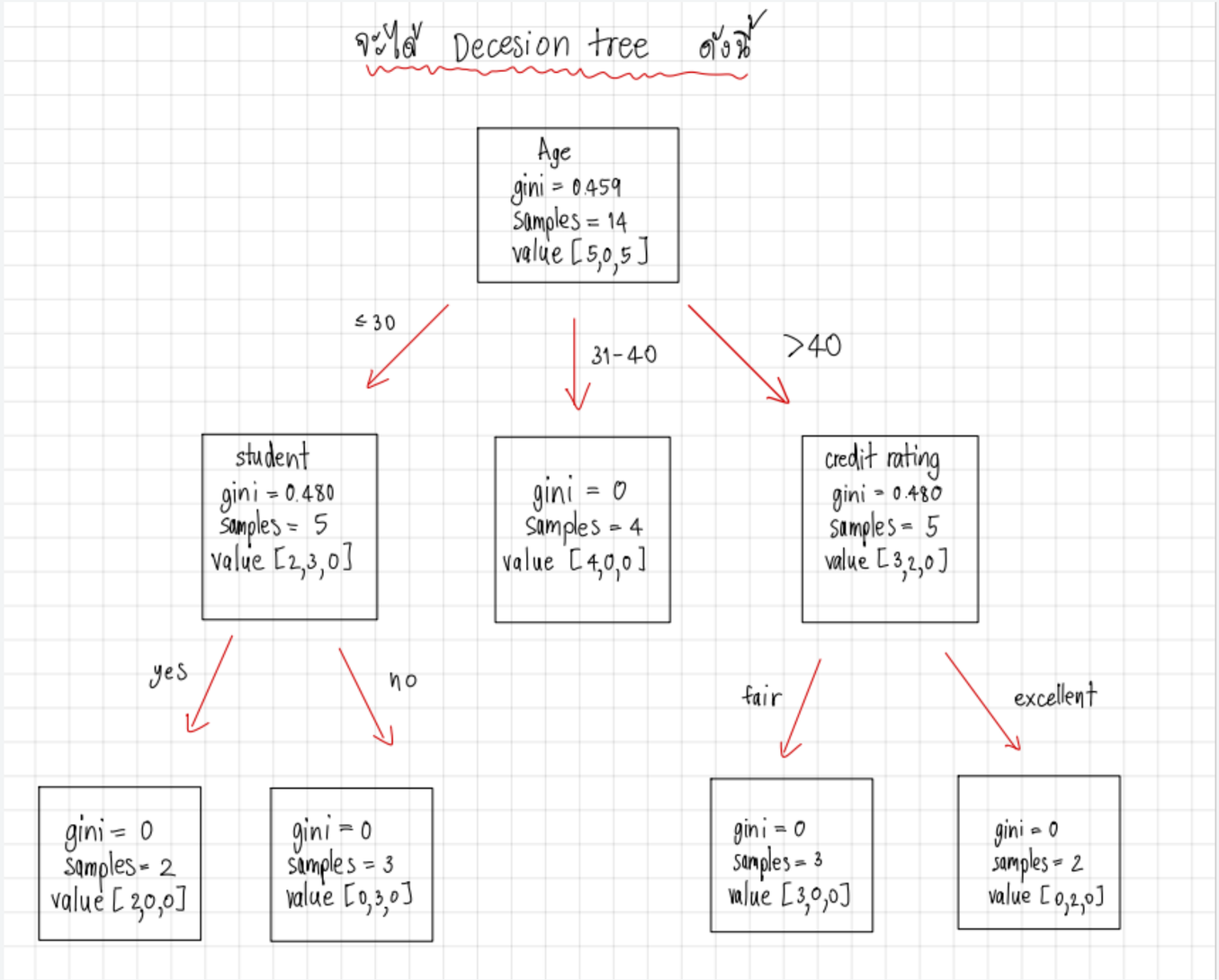
= 0.092 ซึ่ง < 0.1 จึงไม่มีการแยก Node ✗

credit_rating → $\frac{14}{14} \times \left(\frac{8}{14} \times (0.375) - \frac{6}{14}(0.5)\right)$

= 0 ซึ่ง < 0.1 จึงไม่มีการแยก Node ✗



DECISION TREE



- เนื่องจากใน age < 30 student มีค่า gini = 0 ซึ่งมีค่าน้อยที่สุด 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 ซึ่งมีค่าน้อยที่สุด จึงไม่มีการแยก Node ต่อ

สรุป

จากข้อมูลการซื้อคอมพิวเตอร์ (buy computer) ที่มีตัวอย่าง คือ 14 คน

โดยที่
- ในอายุน้อยกว่า 30 จะมีนักเรียนซื้อคอมพิวเตอร์ 2 คน และคนที่ไม่ใช่นักเรียนจะไม่ซื้อคอมพิวเตอร์ 3 คน
- ในอายุ 31 ถึง 40 จะซื้อคอมพิวเตอร์ทุกคน
- ในอายุ 40 ปีขึ้นไป ผู้ที่มี credit rating ในระดับ fair จะซื้อคอมพิวเตอร์ 3 คน และ ผู้ที่มี credit rating ในระดับ excellent จะไม่ซื้อคอมพิวเตอร์ 2 คน

MEMBER

นางสาวกรวรรณ อู่จอหอ	643020495-5
นายธนบดี ภูชมศรี	643020502-4
นางสาวลภัสรดา แดงสูงเนิน	643020518-9
นายศิริโชค ศิริวิชา	643020522-8
นางสาวศดานันท์ ทรัพย์มีมหาศาล	643021272-0
นางสาวอิงอร พลพาล	643021279-6
นางสาวทิพย์วัลย์ สุโพธิ์	643020062-6
นางสาวสุชานาถ พิลาภ	643021276-2

THANK YOU