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

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

$gini(avg\ weight\ age) = \left(0.48\times\frac{5}{14}\right)+\left(0\times\frac{4}{14}\right)+\left(0.48\times\frac{5}{14}\right) = 0.343$

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

$gini(avg\ weight\ income) = \left(0.5\times\frac{4}{14}\right)+\left(0.44\times\frac{6}{14}\right)+\left(0.375\times\frac{4}{14}\right)$

④ student_yes → "yes" = 6 , "no" = 0 | $gini(student\_yes) = 1-\left[\left(\frac{6}{7}\right)^2+\left(\frac{1}{7}\right)^2\right] = 0.245$
student_no → "yes" = 3 , "no" = 4 | $gini(student\_no) = 1-\left[\left(\frac{3}{7}\right)^2+\left(\frac{4}{7}\right)^2\right] = 0.489$

$gini(avg\ weight\ student) = \left(0.245\times\frac{7}{14}\right)+\left(0.489\times\frac{7}{14}\right) = 0.367$

⑤ credit_rating_fair → "yes" = 6 , "no" = 2 | $gini(fair) = 1-\left[\left(\frac{6}{8}\right)^2+\left(\frac{2}{8}\right)^2\right] = 0.375$
credit_rating_excellent → "yes" = 3 , "no" = 3 | $gini(excellent) = 1-\left[\left(\frac{3}{6}\right)^2+\left(\frac{3}{6}\right)^2\right] = 0.5$

$gini(avg\ weight\ credit\_rating) = \left(0.375\times\frac{8}{14}\right)+\left(0.5\times\frac{6}{14}\right) = 0.429$

→ จากการคำนวณค่า gini ของ age , income , student และ credit_rating แล้ว จะเห็นได้ว่า age มีค่า gini ต่ำที่สุด คือ 0.434 ดังนั้นจึงเลือก age เป็น decision node

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

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

$\therefore gini(\text{avg weight income}) = \left[\left(0 \times \frac{2}{5}\right) + \left(0.5 \times \frac{2}{5}\right) + \left(0 \times \frac{1}{5}\right)\right] = 0.2$

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

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

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

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

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

$gini(\text{avg weight credit\_ratings}) = \left[\left(0.4 \times \frac{3}{5}\right) + \left(0.5 \times \frac{2}{5}\right)\right] = 0.464$

กรณี age (31 - 40) , n = 4

income

- high = "yes" = 2 , "no" = 0

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

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

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

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

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

∴ gini(avg weight income) = $[(0 \times \frac{2}{4}) + (0 \times \frac{1}{4}) + (0 \times \frac{1}{4})] = 0$

student

- student_yes = "yes" = 2, "no" = 0

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

- student_no = "yes" = 2, "no" = 0

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

∴ gini(avg weight student) = $[(0 \times \frac{2}{4}) + (0 \times \frac{2}{4})] = 0$

credit_rating

- fair = "yes" = 2, "no" = 0

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

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

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

∴ gini(avg weight credit_rating) = $[(0 \times \frac{2}{4}) + (0 \times \frac{2}{4})] = 0$

กรณี age > 40, n = 5

income

- hight = "yes" = 0, "no" = 0

gini = $1 - [0 + 0] = 0$

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

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

- low = "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 incomes) = $\left[\left(0 \times \frac{0}{5}\right) + \left(0.44 \times \frac{3}{5}\right) + \left(0.5 \times \frac{2}{5}\right)\right] = 0.464$

student

- student_yes = "yes" = 2, "no" = 1

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

- student_no = "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 student) = $\left[\left(0.44 \times \frac{3}{5}\right) + \left(0.5 \times \frac{2}{5}\right)\right]$

credit_rating

- fair = "yes" = 3, "no" = 0

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

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

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

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

ที่ min_impurity_decrease มากกว่า = 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 → $\frac{14}{14} \times (0.459 - \frac{5}{14}(0.48) - 0 - \frac{5}{14}(0.48))$

$= 0.117$ ซึ่ง ≥ 0.1 จึงสามารถแยก node ต่อไป

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

$= 0.020$ ซึ่ง < 0.1 จึงไม่สามารถแยก node

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

$= 0.092$ ซึ่ง < 0.1 จึงไม่สามารถแยก node

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

$= 0$ ซึ่ง < 0.1 จึงไม่สามารถแยก node