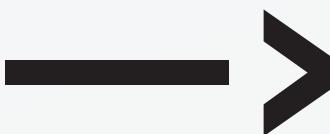




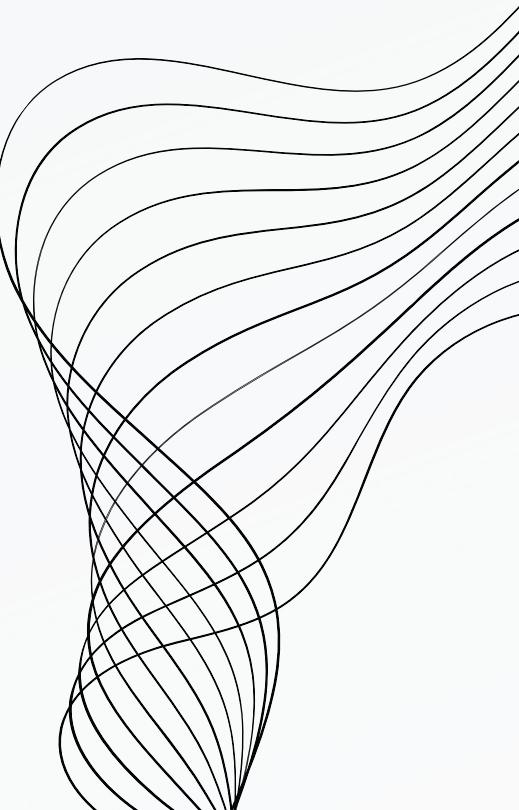
# **DECISION TREE PARAMETER**



## **MIN IMPURITY DECREASE**

**PRESENT BY**

**แต่งโดยมาดแมนหุนปาน daraโก้ & ช่วยด้วย**



# ຕາຣາពຂໍອມູນ

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

# สูตรการคำนวณ GINI

$$GINI = 1 - [(P)^2 + (1-P)^2]$$

เมื่อ P คือความน่าจะเป็นของ CLASS  
โดยในที่นี่ค่าของ CLASS คือ YES, NO

# K ROOT NODE

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  $31 \dots 40 \rightarrow$  "yes" = 4 , "No" = 0

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

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

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

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

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

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

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

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

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

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

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

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

# ଆ ROOT NODE

④

student\_yes → "yes" = 6 , "No" = 0

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

student\_No → "yes" = 3 , "No" = 4

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

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

⑤

credit\_rating\_fair → "yes" = 6 , "No" = 2

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

credit\_rating\_excellent → "yes" = 3 , "No" = 3

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

$$\therefore \text{Gini}(\text{Avg weight credit\_rating}) = \left( 0.375 \times \frac{8}{14} \right) + \left( 0.5 \times \frac{6}{14} \right) = 0.429 \ast$$

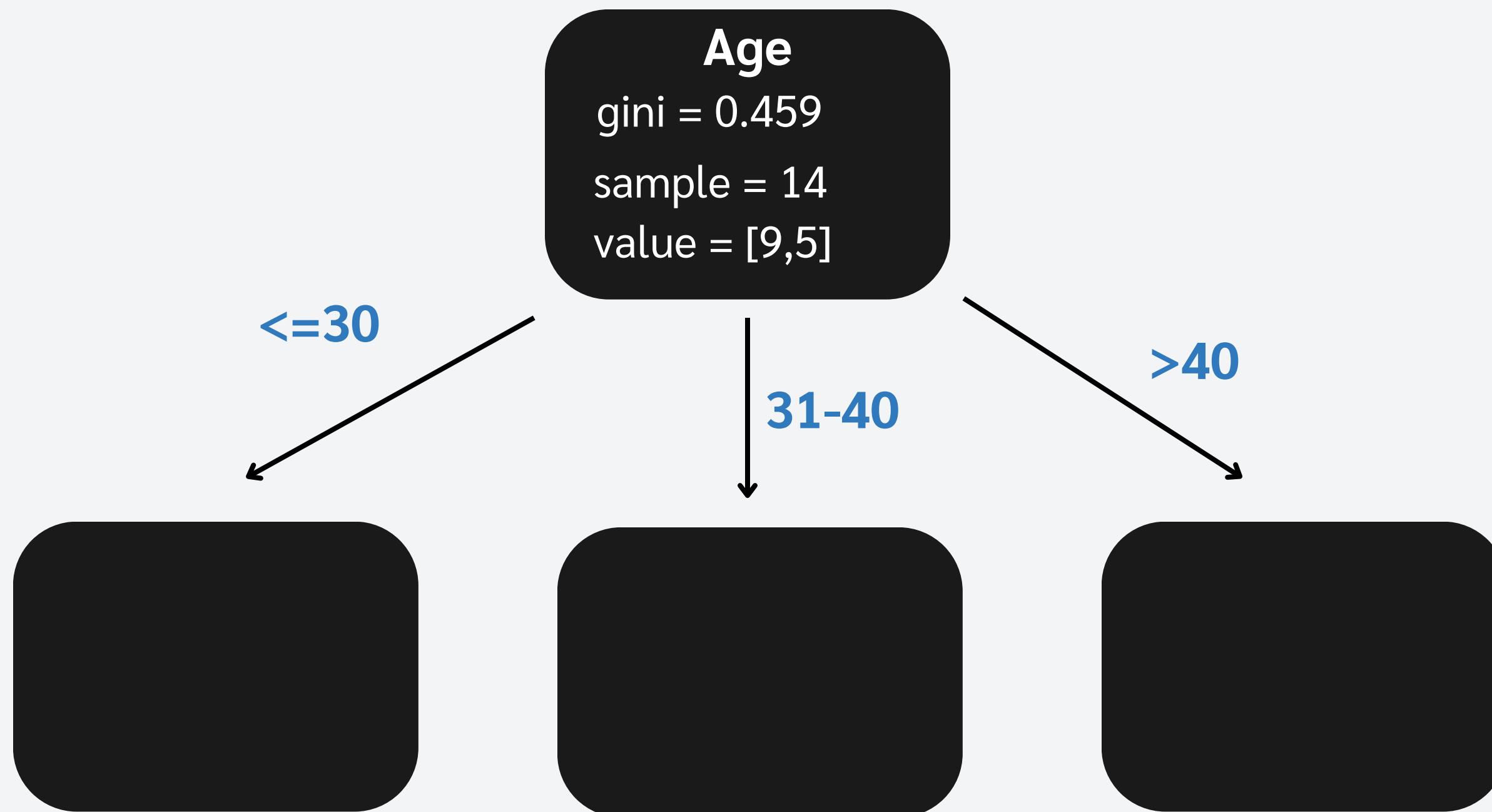
จากการคำนวณค่า gini ของ buy computer = 0.495

และค่า gini ของ age,income,student และ credit rating

จะเห็นได้ว่าค่า gini ของ age มีค่าต่ำที่สุดคือ 0.343

จึงเลือก age ให้เป็น Decision Node

# DECISION NODE



# AGE<=30, N=5

Age	Income	Student	Credit Rating	Buys Computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes

?? age ( $\leq 30$ ),  $n = 5$

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

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

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

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

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

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

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

$$= 0.2 \times$$

# AGE<=30,N=5

Age	Income	Student	Credit_Rating	Buys_Computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes

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" = 0 , "No" = 3

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

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

$$= 0 \quad \text{X}$$

# AGE<=30,N=5

Age	Income	Student	Credit_Rating	Buys_Computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes

credit\_rating - fair = "yes" = 1 , "No" = 2

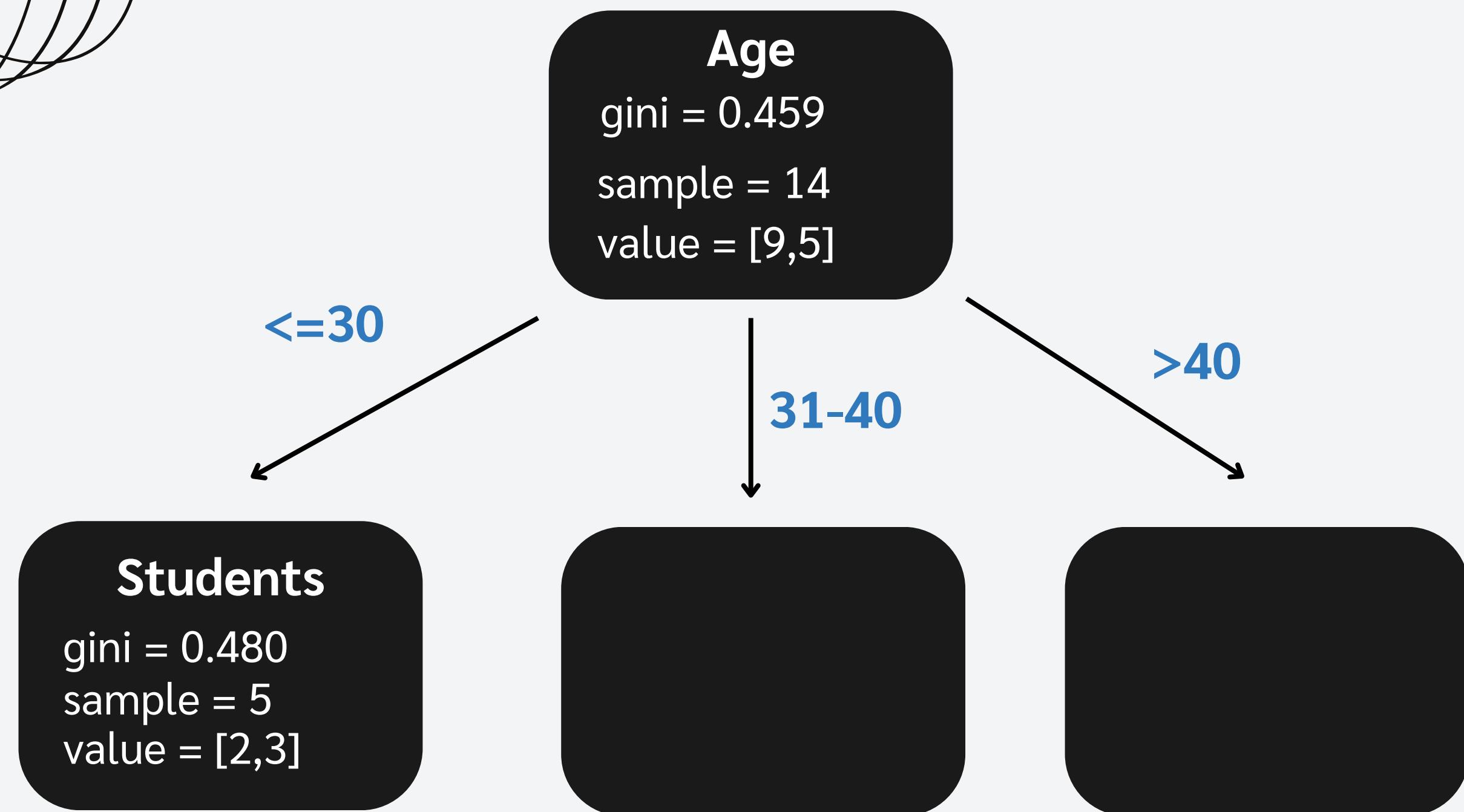
$$\text{gini} = 1 - \left[ \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right] = 0.4$$

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

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

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

$$= 0.464$$



# AGE = 31..40, N=4

Age	Income	Student	Credit Rating	Buys Computer
31...40	high	no	fair	yes
31...40	low	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes

in age (31-40) n = 4

$$\text{income - high} = \text{"yes"} = 2, \text{"No"} = 0$$

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

$$\text{- medium} = \text{"yes"} = 1, \text{"No"} = 0$$

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

$$\text{- low} = \text{"yes"} = 1, \text{"No"} = 0$$

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

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

# AGE =31..40, N=4

Age	Income	Student	Credit_Rating	Buys_Computer
31...40	high	no	fair	yes
31...40	low	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes

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

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

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

$$= 0 \quad \cancel{x}$$

# AGE = 31..40, N = 4

Age	Income	Student	Credit_Rating	Buys_Computer
31...40	high	no	fair	yes
31...40	low	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes

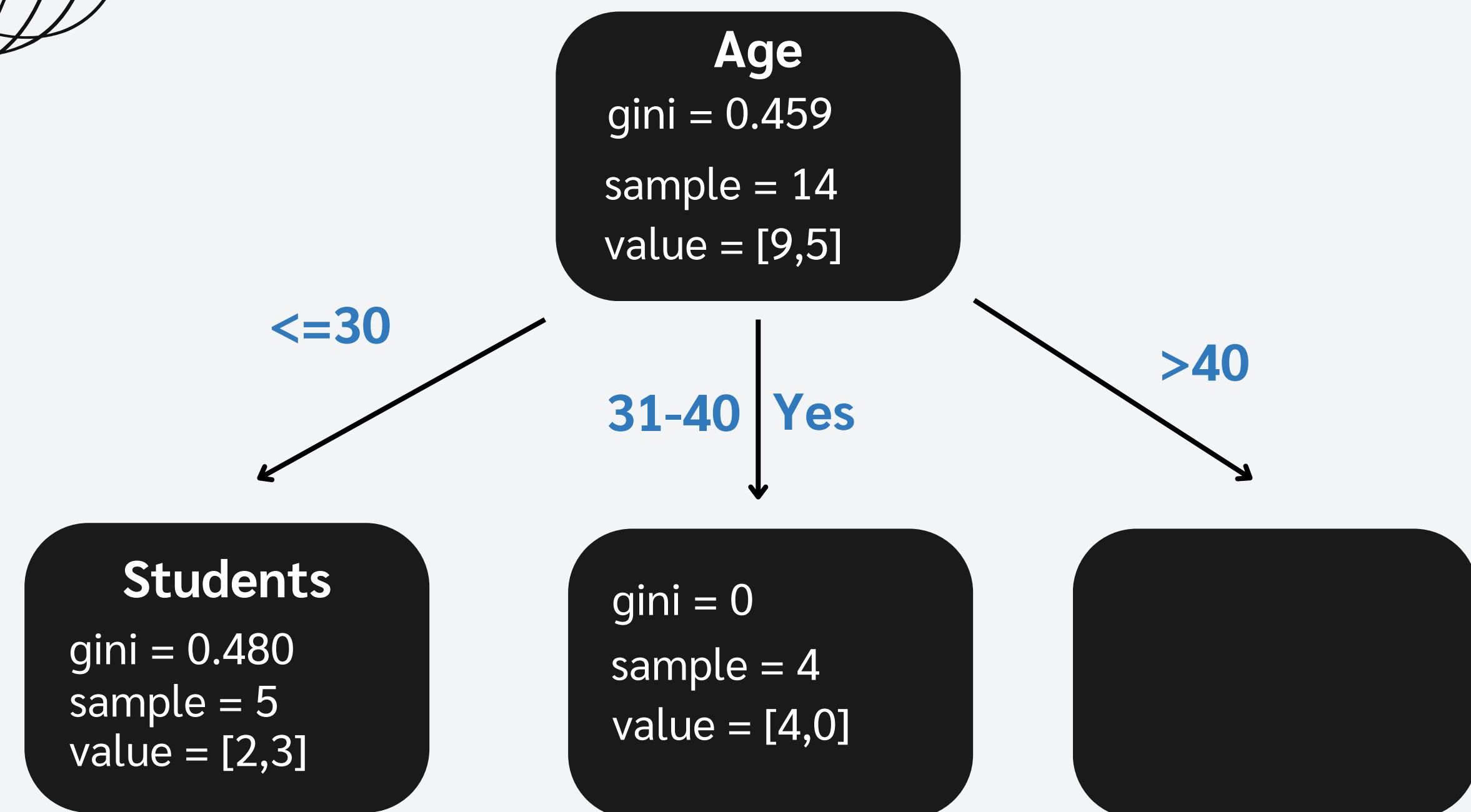
credit\_rating - fair = "yes" = 2, "No" = 0

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

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

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

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



# AGE >40, N=5

Age	Income	Student	Credit_Rating	Buys_Computer
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
>40	medium	yes	fair	yes
>40	medium	no	excellent	no

in 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

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

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

# AGE >40, N=5

Age	Income	Student	Credit_Rating	Buys_Computer
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
>40	medium	yes	fair	yes
>40	medium	no	excellent	no

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

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

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

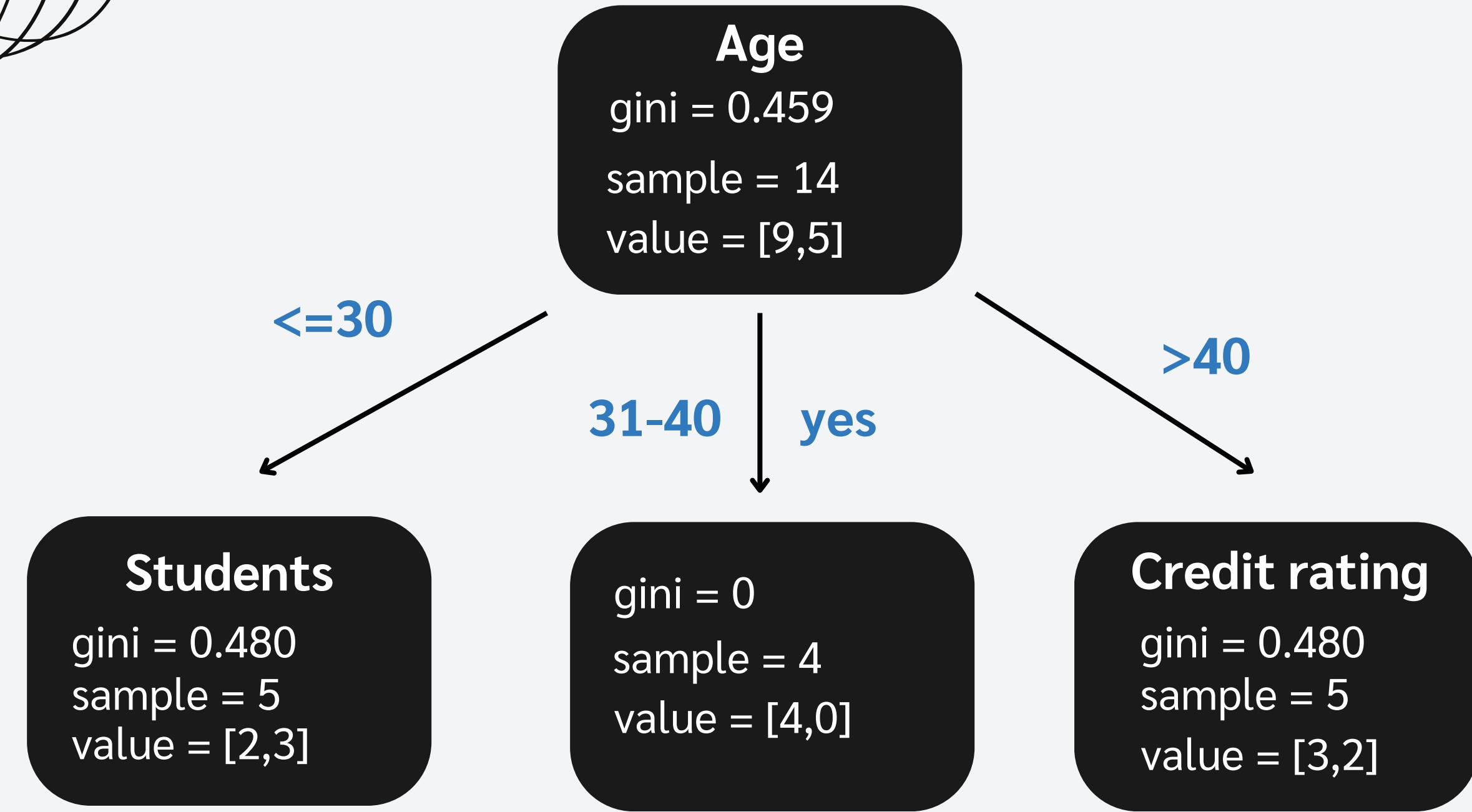
# AGE >40, N=5

Age	Income	Student	Credit_Rating	Buys_Computer
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
>40	medium	yes	fair	yes
>40	medium	no	excellent	no

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

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



# הParameter MIN\_IMPURITY\_DECREASE

The weighted impurity decrease equation is the following:

$$\frac{N_t}{N} * (\text{impurity} - \frac{N_{t\_R}}{N_t} * \text{right\_impurity} - \frac{N_{t\_L}}{N_t} * \text{left\_impurity})$$

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.

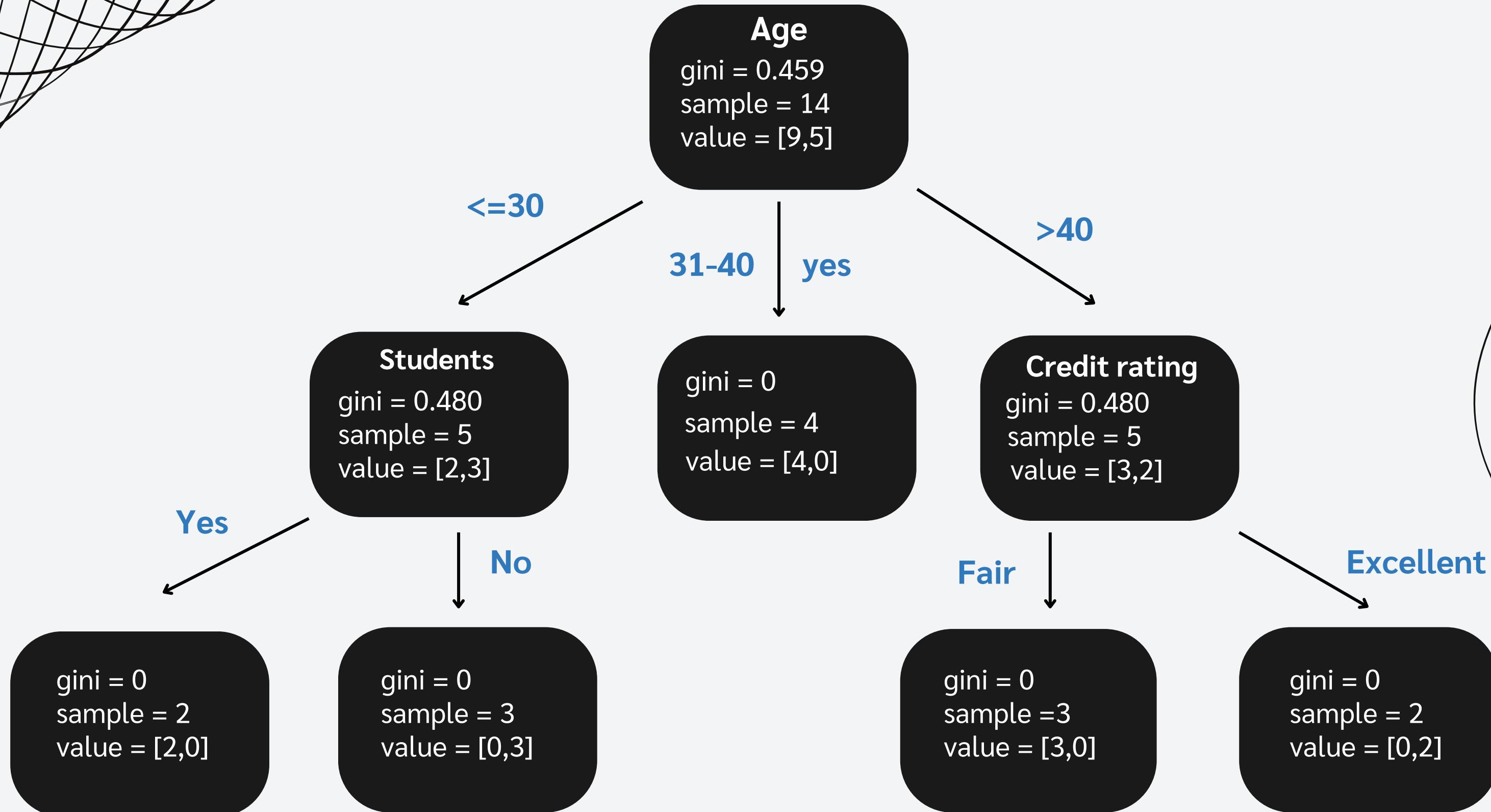
REF

[HTTPS://SCIKIT-LEARN.ORG/STABLE/MODULES/GENERATED/SKLEARN.TREE.DECISIONTREECLASSIFIER.HTML](https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html)

# จากการแทนค่า

จุดเด่น Age →  $\frac{14}{14} \times \left( 0.459 - \frac{5}{14} (0.48) - 0 - \frac{5}{14} (0.48) \right)$   
= 0.117 ซึ่ง  $\geq 0.1$  จึงมีการแยก Node ต่อไป ✗

# FINAL TREE



# CONCLUSION

សម្រេច  
បាន

ជាក់អ្នក ការទិន្នន័យកម្មពិន័យ (buy computer) កំណត់ពាណិជ្ជកម្ម គឺ 14 រោង

- តើយក
- វិនាទាយ អំពី 30 ទំនើស ដែលបានទិន្នន័យកម្មពិន័យ 2 រោង និង កំណត់ពាណិជ្ជកម្ម ជាប្រចាំឆ្នាំ 3 រោង
  - វិនាទាយ 31 ពី 40 ទំនើស កម្មពិន័យ ក្នុងក្រុង
  - វិនាទាយ 40 វិនាទាយ ផ្តល់នូវ credit rating ទិន្នន័យលើ fair ទិន្នន័យកម្មពិន័យ 3 រោង និង ផ្តល់នូវ credit rating ទិន្នន័យលើ excellent ទិន្នន័យកម្មពិន័យ 2 រោង

# MEMBER

นางสาวกรวรรณ อุ่จ้อหอ	643020495-5
นายธนบดี ภูชน์ศรี	643020502-4
นางสาวลักษณดา แแดงสูงเนิน	643020518-9
นายศิริโชค ศิริวิชา	643020522-8
นางสาวศدانันท์ ทรัพย์มีมหาศาล	643021272-0
นางสาวอิงอร พลพาล	643021279-6
นางสาวกิพย์วัลย์ สุโพรี	643020062-6
นางสาวสุชาบนาถ พิลาກ	643021276-2

**THANK YOU**