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

$$Info(D) = I(9,5) = -\tfrac{9}{14}\log_2(\tfrac{9}{14}) - \tfrac{5}{14}\log_2(\tfrac{5}{14}) = 0.940$$

(yes: 9, No: 5)

$$Info_{age}(D) = \tfrac{5}{14} I(2,3) + \tfrac{4}{14} I(4,0) + \tfrac{5}{14} I(3,2) = 0.694$$

(<=30: (2,3), 31..40: (4,0), >40: (3,2))

$$Info_{income}(D) = \tfrac{4}{14} I(2,2) + \tfrac{4}{14} I(3,1) + \tfrac{6}{14} I(4,2)$$

(hight: (2,2), low: (3,1), medium: (4,2))

$$= \tfrac{4}{14}\left(-\tfrac{2}{4}\log_2(\tfrac{2}{4}) - \tfrac{2}{4}\log_2(\tfrac{2}{4})\right) + \tfrac{4}{14}\left(-\tfrac{3}{4}\log_2(\tfrac{3}{4}) - \tfrac{1}{4}\log_2(\tfrac{1}{4})\right)$$

$$+ \tfrac{6}{14}\left(-\tfrac{4}{6}\log_2(\tfrac{4}{6}) - \tfrac{2}{6}\log_2(\tfrac{2}{6})\right) = 0.911$$

$$Info_{student}(D) = \tfrac{7}{14} I(4,3) + \tfrac{7}{14} I(6,1)$$

(no: (4,3), yes: (6,1))

$$= \tfrac{7}{14}\left(-\tfrac{4}{7}\log_2(\tfrac{4}{7}) - \tfrac{3}{7}\log_2(\tfrac{3}{7})\right) + \tfrac{7}{14}\left(-\tfrac{6}{7}\log_2(\tfrac{6}{7}) - \tfrac{1}{7}\log_2(\tfrac{1}{7})\right)$$

$$= 0.788$$

$$Info_{credit\_rating}(D) = \tfrac{8}{14}\left(-\tfrac{6}{8}\log_2(\tfrac{6}{8}) - \tfrac{2}{8}\log_2(\tfrac{2}{8})\right) + \tfrac{6}{14}\left(-\tfrac{3}{6}\log_2(\tfrac{3}{6}) - \tfrac{3}{6}\log_2(\tfrac{3}{6})\right)$$

$$= \tfrac{8}{14} I(6,2) + \tfrac{6}{14} I(3,3) = 0.892$$

(fair: (6,2), excellent: (3,3))

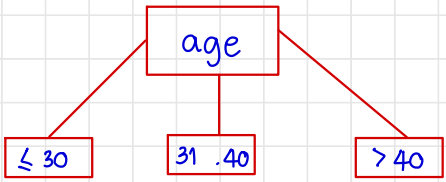
$$Gain(age) = Info(D) - Info_{age}(D) = 0.940 - 0.694 = 0.246$$

$$Gain(income) = Info(D) - Info_{income}(D) = 0.940 - 0.911 = 0.029$$

$Gain(Student) = Info(D) - Info_{student}(D) = 0.940 - 0.788 = 0.152$

$Gain(credit\_rating) = Info(D) - Info_{credit\_rating}(D) = 0.940 - 0.892 = 0.048$

สรุป คือ Gain ที่มากที่สุด คือ age ที่มีค่าเท่ากับ 0.246 ถ้า age ≤ 30



$Info_{age:\leq 30}(D) = I(2,3) = -\frac{2}{5}\log_2(\frac{2}{5}) - \frac{3}{5}\log_2(\frac{3}{5}) = 0.971$ (yes, No)

$Info_{income}(D) = \frac{1}{5}I(1,0) + \frac{2}{5}I(1,1) + \frac{2}{5}I(0,2)$

$= \frac{1}{5}(-\frac{1}{1}\log_2(\frac{1}{1}) - 0) + \frac{2}{5}(-\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}))$

$+ \frac{2}{5}(-\frac{2}{5}\log_2(\frac{2}{2}) - 0) = 0.4$

$Info_{student}(D) = \frac{3}{5}(0,3) + \frac{2}{5}(2,0) = \frac{3}{5}(-\frac{3}{3}\log_2(\frac{3}{3}) + \frac{2}{5}(-\frac{2}{2}\log_2(\frac{2}{2})) = 0$ (No, yes)

$Info_{credit\_rating}(D) = \frac{3}{5}I(1,2) + \frac{2}{5}(1,1) = \frac{3}{5}(-\frac{1}{3}\log_2(\frac{1}{3}) - \frac{2}{3}\log_2(\frac{2}{3}))$ (fair, excellenz)

$+ \frac{2}{5}(-\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log(\frac{1}{2})) = 0.951$

$Gain_{(income)} = Info_{age:\leq 30}(D) - Info_{income}(D) = 0.971 - 0.4 = 0.571$

$Gain_{(student)} = Info_{age:\leq 30}(D) - Info_{student}(D) = 0.971 - 0 = 0.971$

$Gain_{(credit\_rating)} = Info_{age\leq 30}(D) - Info_{credit\_rating}(D) = 0.971 - 0.951 = 0.020$

ดังนั้น decision node แรก คือ student เพราะค่า Gain สูงสุด

$$Info_{age=31...40}(D) = I(4,0) = -\tfrac{4}{4}\log_2(\tfrac{4}{4}) - \tfrac{0}{4}\log_2(\tfrac{0}{4}) = 0$$

ดังนั้น 31...40 ไม่มี decision node แต่เราสามารถ $I(4,0)$ ที่ใช้ age:31...40 = ซื้อทั้งหมด

$$Info_{age:>40}(D) = I(3,2) = -\tfrac{3}{5}\log_2(\tfrac{3}{5}) - \tfrac{2}{5}\log_2(\tfrac{2}{5}) = 0.971$$

$$Info_{income}(D) = \tfrac{2}{5}(I(1,1)) + \tfrac{3}{5}I(2,1)$$

$$= \tfrac{2}{5}(-\tfrac{1}{2}\log_2(\tfrac{1}{2}) - \tfrac{1}{2}\log_2(\tfrac{1}{2})) + \tfrac{3}{5}(-\tfrac{2}{3}\log_2(\tfrac{2}{3}) - (\tfrac{1}{3})\log_2(\tfrac{1}{3})) = 0.951$$

$$Info_{(student)}(D) = \tfrac{3}{5}I(2,1) + \tfrac{2}{5}I(1,1) = \tfrac{3}{5}(-\tfrac{2}{3}\log_2(\tfrac{2}{3}) - \tfrac{1}{3}\log_2(\tfrac{1}{3}))$$

$$+ \tfrac{2}{5}(-\tfrac{1}{2}\log_2(\tfrac{1}{2}) - \tfrac{1}{2}\log_2(\tfrac{1}{2})) = 0.951$$

$$Info_{credit\_rating}(D) = \tfrac{3}{5}I(3,0) + \tfrac{2}{5}I(0,2)$$

$$= \tfrac{3}{5}(-\tfrac{3}{3}\log_2(\tfrac{3}{3}) + \tfrac{2}{5}(-\tfrac{2}{2}\log_2(\tfrac{2}{2}) = 0$$

$$Gain_{(income)} = Info_{age:>40}(D) - Info_{student}(D) = 0.971 - 0.951 = 0.02$$

$$Gain_{(student)} = Info_{age:>40}(D) - Info_{student}(D) = 0.971 - 0.951 = 0.02$$