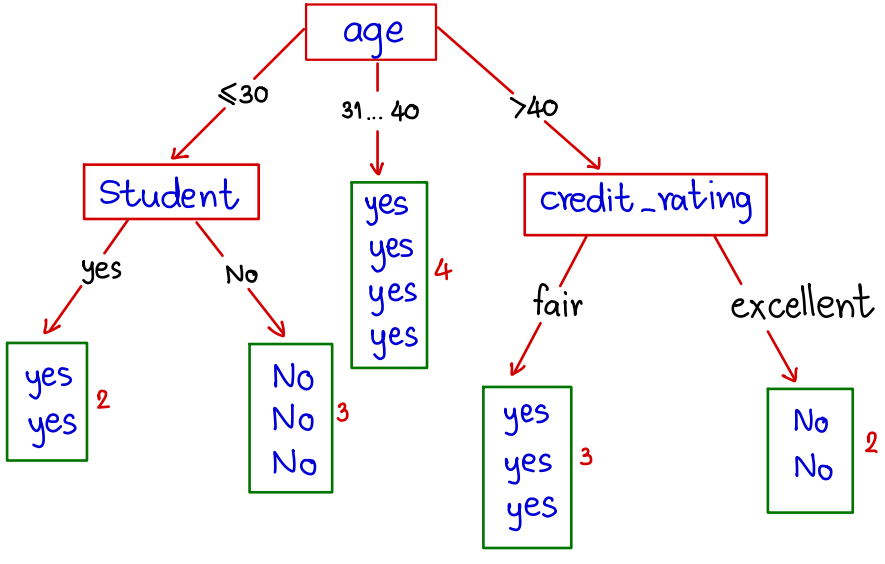
age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent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



min simple = 4

$$Info(D) = I(\overset{\text{yes}}{9}, \overset{\text{no}}{5}) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) = 0.940$$

$$Info_{age}(D) = \frac{5}{14}\overset{\leq 30}{I(2,3)} + \frac{4}{14}\overset{31\text{-}40}{I(4,0)} + \frac{5}{14}\overset{>40}{I(3,2)}$$

$$= \frac{5}{14}\left[-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right] +$$

$$\frac{4}{14}\left[-\frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\log_2\left(\frac{0}{4}\right)\right]^{0} +$$

$$\frac{5}{14}\left[-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right]$$

$$= 0.694$$

$$Info_{income}(D) = \frac{4}{14}\overset{\text{high}}{I(2,2)} + \frac{6}{14}\overset{\text{medium}}{I(4,2)} + \frac{4}{14}\overset{\text{low}}{I(3,1)}$$

$$= \frac{4}{14}\left[-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right] +$$

$$\frac{6}{14}\left[-\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right] +$$

$$\frac{4}{14}\left[-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right]$$

$$= 0.911$$

$$Info_{student}(D) = \frac{7}{14} \overset{yes}{I}(6,1) + \frac{7}{14} \overset{no}{I}(3,4)$$

$$= \frac{7}{14}\left[-\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right)\right] +$$

$$\frac{7}{14}\left[-\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right)\right]$$

$$= 0.788$$

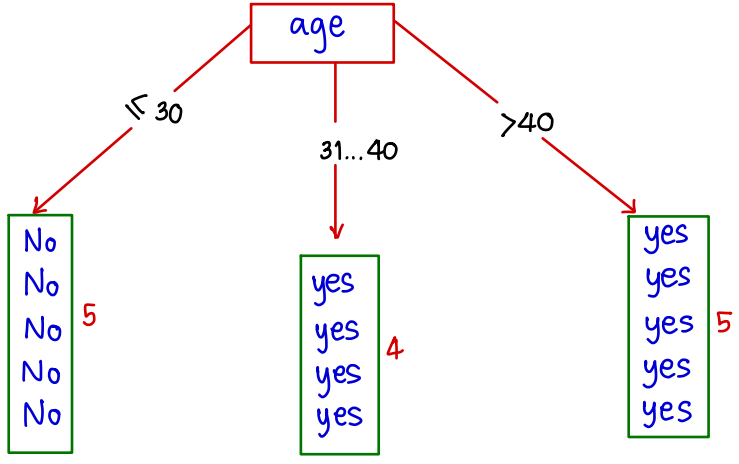
$$Info_{credit\_rating}(D) = \frac{6}{14} \overset{excellent}{I}(3,3) + \frac{8}{14} \overset{fair}{I}(6,2)$$

$$= \frac{6}{14}\left[-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right] +$$

$$\frac{8}{14}\left[-\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right]$$

$$= 0.892$$

Gain age = 0.940 - 0.694 = 0.246
Gain income = 0.940 - 0.911 = 0.029
Gain student = 0.940 - 0.788 = 0.152
Gain credit_rating = 0.940 - 0.892 = 0.048



Max_feature = 2

สุ่ม 2 ฟีเจอร์ จาก age, income, student, credit_rating
สมมติได้ income และ credit_rating

$$\text{Info}(D) = I(9,5) = \frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) = 0.940$$

income

$$\text{Info}_{\text{income}}(D) = \frac{4}{14}\overset{\text{high}}{I}(2,2) + \frac{6}{14}\overset{\text{mediem}}{I}(4,2) + \frac{4}{14}\overset{\text{low}}{I}(3,1)$$

$$= \frac{4}{14}\left[-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right] + \frac{6}{14}\left[-\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right]$$

$$+ \frac{4}{14}\left[-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right]$$

$$= 0.911$$

credit_rating

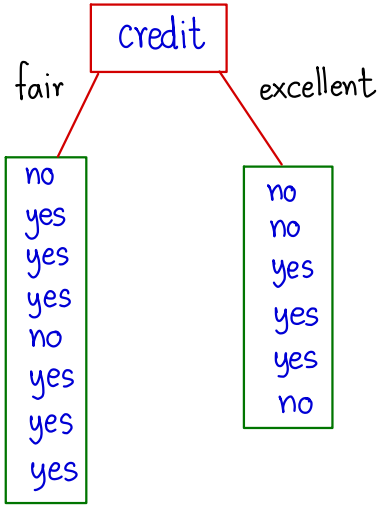
$$\text{Info}_{\text{credit\_rating}}(D) = \frac{6}{14}\overset{\text{excellent}}{I}(3,3) + \frac{8}{14}\overset{\text{fair}}{I}(6,2)$$

$$= \frac{6}{14}\left[-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right] + \frac{8}{14}\left[-\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right]$$

$$= 0.892$$

$$\text{Gain (income)} = \text{Info}(D) - \text{Info}_{\text{income}}(D) = 0.940 - 0.911 = 0.029$$

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



ใช้ credit_rating แบ่ง

credit_rating = fair

$$Info\ (D) = I(6,2) = -\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) = 0.811$$

สุ่ม 2 ฟีเจอร์ จาก age, income, student จะได้ income กับ age

$$info_{income}(D) = \frac{3}{8}\overset{high}{I}(2,1) + \frac{3}{8}\overset{mediem}{I}(2,1) + \frac{2}{8}\overset{low}{I}(2,0)$$

$$= \frac{3}{8}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] + \frac{3}{8}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right]$$

$$+ \frac{2}{8}\left[-\frac{2}{2}\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right]$$

$$= 0.396$$

$$info_{age}(D) = \frac{3}{8} I(1,2) + \frac{2}{8} I(2,0) + \frac{3}{8} I(3,0)$$

(≤ 30, $31\ldots40$, >40)

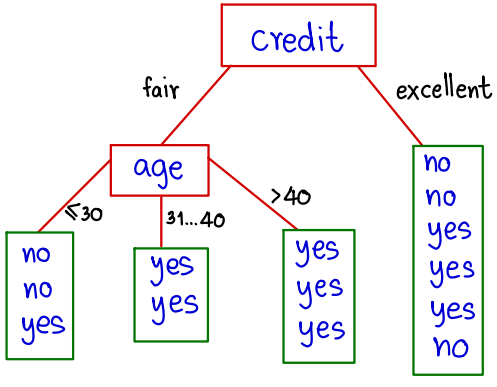
$$= \frac{3}{8}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right] + \frac{2}{8}\left[-\frac{2}{2}\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right]^{0}$$

$$+ \frac{3}{8}\left[-\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right]^{0}$$

$$= 0.344$$

$$Gain(income) = Info(D) - Info_{income}(D) = 0.811 - 0.396 = 0.415$$

$$Gain(age) = Info(D) - Info_{age}(D) = 0.811 - 0.344 = 0.467$$



กรณี age ≤ 30

ใช้ student , income แบ่ง

credit = fair , age ≤ 30

$$Info\ (D) = I(1,2) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.918$$

$$Info_{student}(D) = \frac{2}{3}\overset{no}{I}(0,2) + \frac{1}{3}\overset{yes}{I}(1,0)$$

$$= \frac{2}{3}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right]^{0} + \frac{1}{3}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right)\right]^{0}$$

$$= 0$$

$$Info_{income}(D) = \frac{1}{3}\overset{high}{I}(0,1) + \frac{1}{3}\overset{medium}{I}(0,1) + \frac{1}{3}\overset{low}{I}(1,0)$$

$$= \frac{1}{3}\left[-\frac{0}{1}\log_2\left(\frac{0}{1}\right) - \frac{1}{1}\log_2\left(\frac{1}{1}\right)\right]^{0} + \frac{1}{3}\left[-\frac{0}{1}\log_2\left(\frac{0}{1}\right) - \frac{1}{1}\log_2\left(\frac{1}{1}\right)\right]^{0}$$

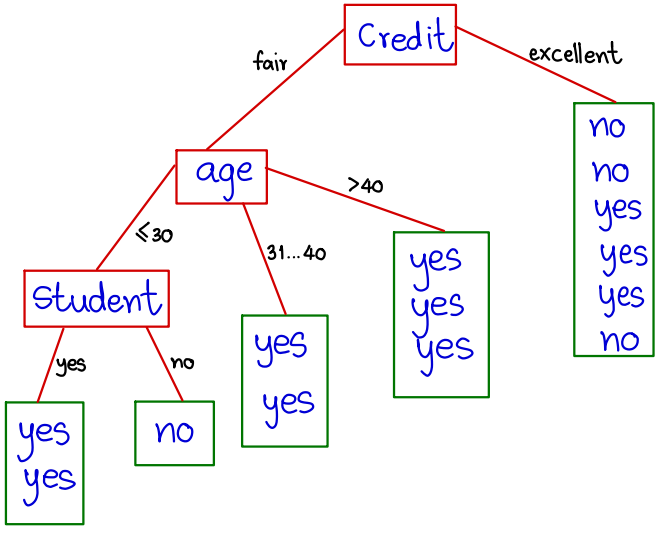
$$+ \frac{1}{3}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right)\right]^{0}$$

$$= 0$$

$$Gain\ (student) = Info(D) - Info_{student}(D) = 0.918 - 0 = 0.918$$

$$Gain\ (income) = Info(D) - Info_{income}(D) = 0.918 - 0 = 0.918$$

สามารถเลือกใช้อะไรก็ได้



credit_rating = excellent

$$\text{Info}(D) = I(3,3) = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right) = 1$$

ลูก 2 ตัวเลือก จาก age, income, student สมมติ income กับ student

$$\text{Info}_{\text{student}}(D) = \frac{3}{6}\overset{\text{yes}}{I}(1,2) + \frac{3}{6}\overset{\text{no}}{I}(1,2)$$

$$= \frac{3}{6}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right] + \frac{3}{6}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right]$$

$$= 0.918$$

$$\text{Info}_{\text{income}}(D) = \frac{1}{6}\overset{\text{high}}{I}(0,1) + \frac{3}{6}\overset{\text{medium}}{I}(2,1) + \frac{2}{6}\overset{\text{low}}{I}(1,1)$$

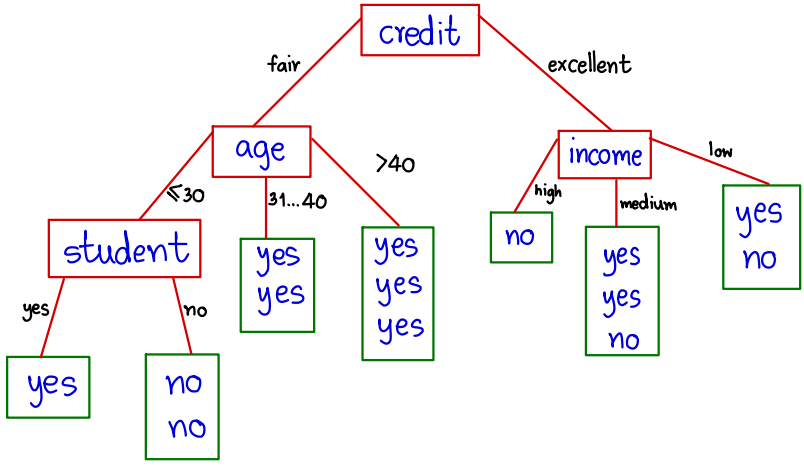
$$= \frac{1}{6}\left[-\frac{0}{1}\log_2\left(\frac{0}{1}\right) - \frac{1}{1}\log_2\left(\frac{1}{1}\right)\right]^{0} + \frac{3}{6}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right]$$

$$+ \frac{2}{6}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.459$$

Gain (student) = $Info(D) - Info_{student}(D) = 1 - 0.918 = 0.082$

Gain (income) = $Info(D) - Info_{income}(D) = 1 - 0.759 = 0.241$



กรณี medium

ใช้ student, age แทน

credit = excellent , Income = medium

$$Info(D) = I(2,1) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.918$$

$$Info_{student}(D) = \frac{1}{3}\overset{yes}{I}(1,0) + \frac{2}{3}\overset{no}{I}(1,1)$$

$$= \frac{1}{3}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right)\right]^{0} + \frac{2}{3}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 0.6$$

$$Info_{age}(D) = \frac{1}{3} I(1,0)^{\leq 30} + \frac{1}{3} I(1,0)^{31\ldots40} + \frac{1}{3} I(0,1)^{>40}$$

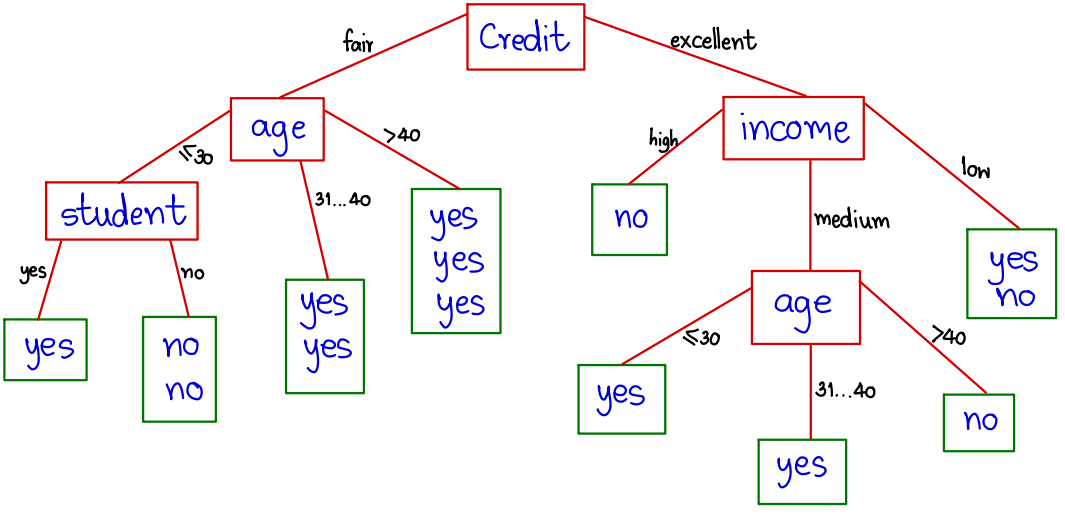
$$= \frac{1}{3}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right)\right]^{0} + \frac{1}{3}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log_2\left(\frac{0}{1}\right)\right]^{0}$$

$$+ \frac{1}{3}\left[-\frac{0}{1}\log_2\left(\frac{0}{1}\right) - \frac{1}{1}\log_2\left(\frac{1}{1}\right)\right]^{0}$$

$$= 0$$

$$Gain(student) = Info(D) - Info_{student}(D) = 0.918 - 0.6 = 0.318$$

$$Gain(age) = Info(D) - Info_{age}(D) = 0.918 - 0 = 0.918$$



กรณี Income = low

ดูแค่ student , age แล้ว

Credit = excellent , Income = low

$$Info(D) = I(1,1) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1$$

$$Info_{age}(D) = \frac{1}{2}\overset{31...40}{I}(1,0) + \frac{1}{2}\overset{>40}{I}(0,1)$$

$$= \frac{1}{2}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \frac{0}{1}\log\left(\frac{0}{1}\right)\right] + \frac{1}{2}\left[-\frac{0}{1}\log_2\left(\frac{0}{1}\right) - \frac{1}{1}\log_2\left(\frac{1}{1}\right)\right]$$

(0 และ 0)

$$= 0$$

$$Info_{student}(D) = \frac{2}{2}\overset{yes}{I}(1,1)$$

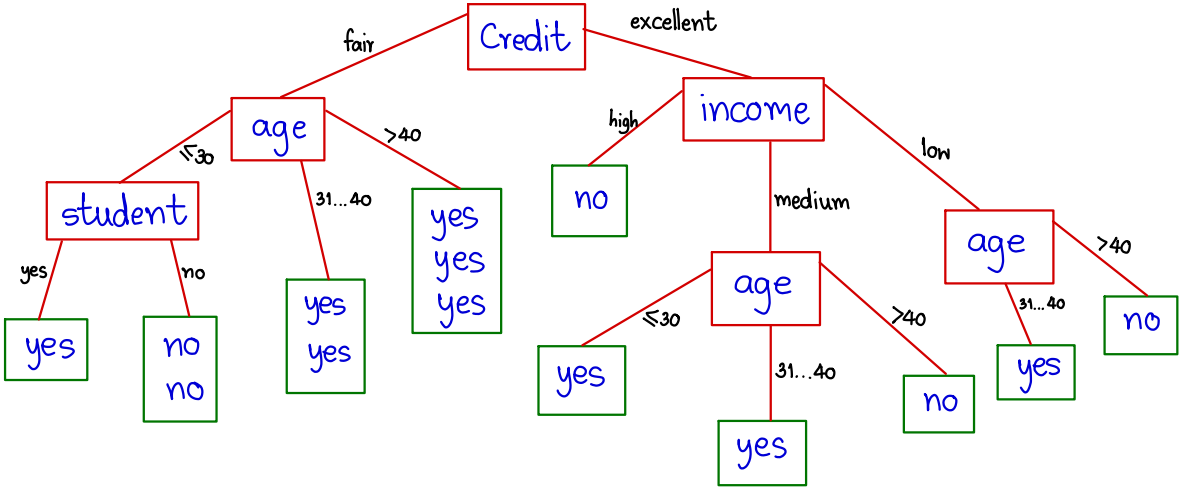
$$= \frac{2}{2}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right]$$

$$= 1$$

$$Gain(age) = Info(D) - Info_{age}(D) = 1 - 0 = 1$$

$$Gain(student) = Info(D) - Info_{student}(D) = 1 - 1 = 0$$

final


Credit
fair
excellent
age
income
student
yes
yes
yes
no
no
yes
yes
yes
yes
yes
no
high
medium
low
no
age
age
yes
yes
no
yes
no
>40
≤30
31...40
>40
≤30
31...40
>40
31...40