Homework 2

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

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

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

$$= \frac{5}{14}\left[-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right] + \frac{4}{14}\left[-\frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\log_2\left(\frac{0}{4}\right)\right] +$$

$$\frac{5}{14}\left[-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right] = 0.694$$

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

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

$$\frac{4}{14}\left[-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right] = 0.911$$

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

$$= \frac{7}{14}\left[-\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right)\right] + \frac{7}{14}\left[-\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right)\right] = 0.788$$

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

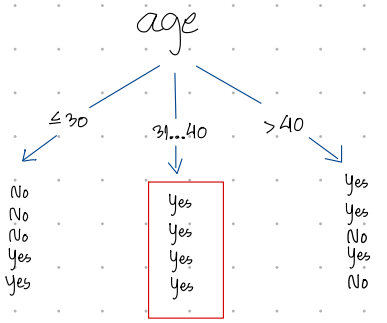
$$= \frac{6}{14}\left[-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right] + \frac{8}{14}\left[-\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right] = 0.892$$

$Gain_{age} = 0.940 - 0.694 = 0.246$ *

$Gain_{income} = 0.940 - 0.911 = 0.029$

$Gain_{student} = 0.940 - 0.788 = 0.152$

$Gain_{credit\_rating} = 0.940 - 0.892 = 0.048$



นรีกานต์ ปัญญารักษ์ 643020505-8

Age ≤ 30

$$Info_{age \le 30}(D) = I(2,3) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.971$$

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

$$= \frac{2}{5}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] +$$

$$\frac{1}{5}\left[-\frac{1}{1}\log_2\left(\frac{1}{1}\right) - \overset{0}{\cancel{\frac{0}{1}\log_2\left(\frac{0}{1}\right)}}\right] = 0.4$$

$$Info_{student}(D) = \overset{yes}{\frac{2}{5}}I(2,0) + \overset{no}{\frac{3}{5}}I(0,3)$$

$$= \frac{2}{5}\left[-\frac{2}{2}\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right)\right] + \frac{3}{5}\left[-\frac{0}{3}\log_2\left(\frac{0}{3}\right) - \frac{3}{3}\log_2\left(\frac{3}{3}\right)\right] = 0$$

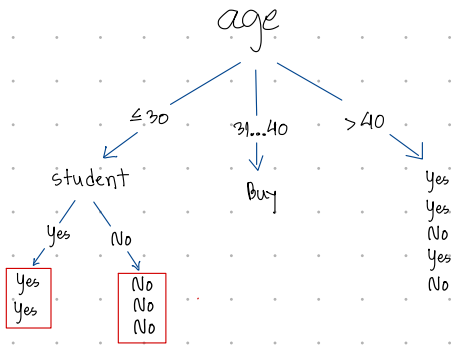
$$Info_{credit\_rating}(D) = \overset{excellent}{\frac{2}{5}}I(1,1) + \overset{fair}{\frac{3}{5}}I(1,2)$$

$$= \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] + \frac{3}{5}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right] = 0.951$$

$Gain_{income} = 0.971 - 0.4 = 0.571$

$Gain_{student} = 0.971 - 0 = 0.971$ *

$Gain_{credit\_rating} = 0.971 - 0.951 = 0.020$



Age > 40

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

$$Info_{income}(D) = \overset{medium}{\frac{3}{5}I(2,1)} + \overset{low}{\frac{2}{5}I(1,1)}$$

$$= \frac{3}{5}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] = 0.951$$

$$Info_{student}(D) = \overset{yes}{\frac{3}{5}I(2,1)} + \overset{no}{\frac{2}{5}I(1,1)}$$

$$= \frac{3}{5}\left[-\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right] = 0.951$$

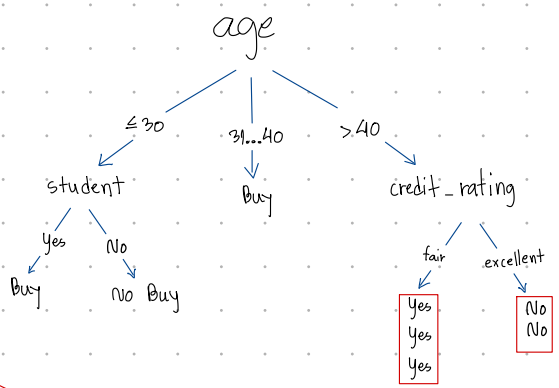
$$Info_{credit\_rating}(D) = \overset{excellent}{\frac{2}{5}I(0,2)} + \overset{fair}{\frac{3}{5}I(3,0)}$$

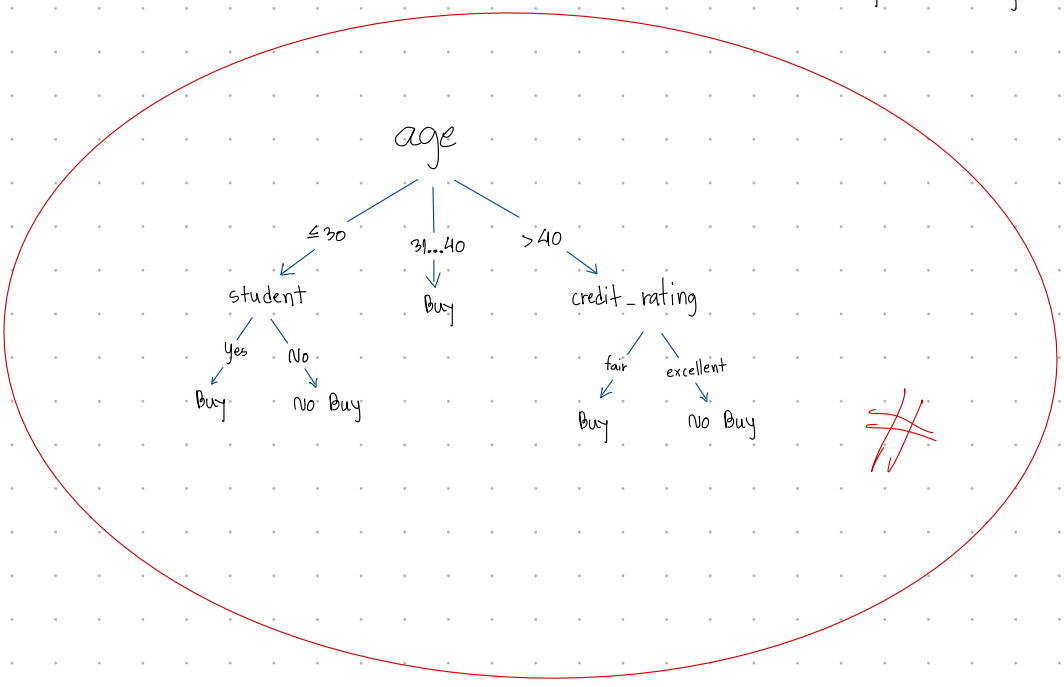
$$= \frac{2}{5}\left[-\frac{0}{2}\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\log_2\left(\frac{2}{2}\right)\right] + \frac{3}{5}\left[-\frac{3}{3}\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\log_2\left(\frac{0}{3}\right)\right] = 0$$

$Gain_{income} = 0.971 - 0.951 = 0.020$

$Gain_{student} = 0.971 - 0.951 = 0.020$

$Gain_{credit\_rating} = 0.971 - 0 = 0.971$ *





#

นรีกานต์ ปัญญารักษ์ 643020505-8