Training data set: Who buys computer?

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

2. Income

$Gain_{income} = Info(D) - Info_{in}(D)$

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

$= \frac{4}{14}\left[-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right] + \frac{6}{14}\left[-\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right] + \frac{4}{14}\left[-\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right]$

$= 0.232 + 0.399 + 0.285$

$= 0.911$

$\therefore Gain_{inc} = 0.940 - 0.911 = 0.029$ ✱

1. age ก่อนอัด - หลังอัด

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

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

หลัง $Info_{age}(D)$

$= \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2)$

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

$= 0.694$

$\therefore$

age	p_i	n_i	$I(p_i, n_i)$
≤ 30	2	3	0.971
30 - 40	4	0	0
>40	3	2	0.971

$\therefore Gain_{age} = 0.940 - 0.694 = 0.246$ ✱

3. student

$Gain_{stu} = Info(D) + Info_{stu}(D)$

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

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

$= 0.296 + 0.493$

$= 0.789$

$\therefore Gain_{stu} = 0.940 - 0.789 = 0.151$ ✱

4. credit_rating

$Gain_{cre} = Info(D) - Info_{cre}(D)$

$Info_{cre}(D) = \frac{8}{14} I(6,2) + \frac{6}{14} I(3,3)$

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

$= 0.464 + 0.429$

$= 0.893$

$\therefore Gain_{cre} = 0.940 - 0.893 = 0.047$ ✱