HW 4

นางสาวบงกช ใจหาญ 6630202841-0 SC-SIDS

	age	income	student	credit_rating	buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31…40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31…40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31…40	medium	no	excellent	yes
13	31…40	high	yes	fair	yes
14	>40	medium	no	excellent	no

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

$$= 0.940$$

$$\text{Info}_{age}(D) = \frac{5}{14}I(2,3) + \frac{4}{14}I(4,0) + \frac{5}{14}I(3,2)$$

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

$$= 0.694$$

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

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

$$= 0.911$$

$$\text{Info}_{student}(D) = \frac{7}{14}(6,1) + \frac{7}{14}(3,4)$$

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

$$= 0.789$$

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

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

$$= 0.892$$

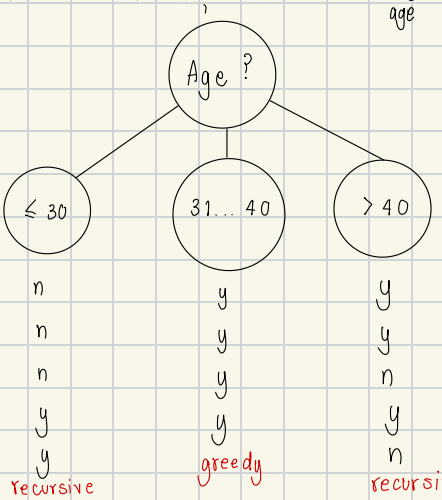
$$\text{Gain}_{age}(D) = 0.940 - 0.694 = 0.246$$

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

$$\text{Gain}_{student}(D) = 0.940 - 0.789 = 0.151$$

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

เลือก Gain (D) ที่มีค่ามากที่สุด ได้แก่ $\text{Gain}_{age}(D)$



กรณี : Age ≤ 30

	age	income	student	credit_rating	buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31…40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31…40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31…40	medium	no	excellent	yes
13	31…40	high	yes	fair	yes
14	>40	medium	no	excellent	no

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

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

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

$$= 0.4$$

$$Info_{student}(D) = \frac{2}{5}I(2,0) + \frac{3}{5}I(0,3)$$

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

$$= 0$$

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

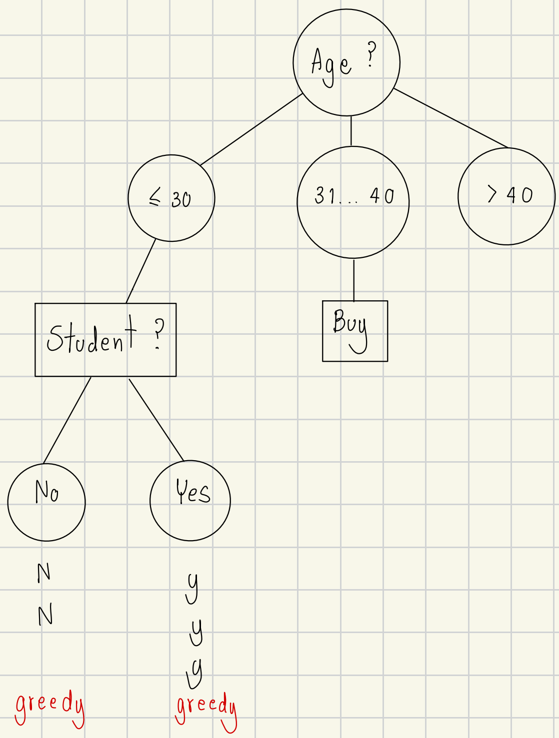
$$= \frac{3}{5}\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right] + \frac{2}{5}\left[-\frac{1}{5}\log_2\left(\frac{1}{2}\right) - \frac{1}{5}\log_2\left(\frac{1}{2}\right)\right]$$

$$=$$

$$Gain_{income}(D) = 0.971 - 0.4 = 0.571$$

$$Gain_{student}(D) = 0.971 - 0 = 0.971$$ → Dicision Node

$$Gain_{Credit}(D) = 0.971 - 0.951 = 0.020$$



	age	income	student	credit_rating	buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31…40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31…40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31…40	medium	no	excellent	yes
13	31…40	high	yes	fair	yes
14	>40	medium	no	excellent	no

$$Info(D) = I(3,2) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$

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

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

$$= 0.951$$

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

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

$$= 0.951$$

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

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

$$= 0$$

$$Gain_{income}(D) = 0.971 - 0.951 = 0.020$$

$$Gain_{student}(D) = 0.971 - 0.951 = 0.020$$

$$Gain_{credit}(D) = 0.971 - 0 = 0.971$$ → Dicision Node