notebook

Attribute selection

with Information gain

age	p_i	n_i	$I(p_i, n_i)$
<=30	2	3	0.971
31..40	4	0	0
>40	3	2	0.971

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

Class P : buys_computer = "yes"

Class N : buys_computer = "no"

Yes: 9

No: 5

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

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

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

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

$$= 0.694$$

Age

<=30 Y:2/N:3

31-40 Y:4/N:0

>40   Y:3/N:2

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

$$= 0.911$$

Income

High Y:2/N:2

Medium Y:4/N:2

Low Y:3/N:1

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

$$= 0.789$$

Student,

Yes Y:6/N:1

No Y:3/N:4

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

$$= 0.892$$

Credit

Fair   Y:6/N:2

Excellent Y:3/N:3

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

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

$$Gain(student) = Info(D) - Info_{student}(D) = 0.940 - 0.789 = 0.151$$

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

F 1

age	income	Student	Credit	buy
<=30	high	no	fair	no
<=30	high	no	excellent	no
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
<=30	medium	yes	excellent	yes

Yes: 2 No = 3

Income
High Y:0/N:2
Medium Y:1/N:1
Low Y:1/N:0

Student,
Yes Y:2/N:0
No Y:0/N:3

Credit
Fair   Y:1/N:2
Excellent Y:1/N:1

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

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

$$= \frac{2}{5}\left(-\frac{2}{2} \log_2 \frac{2}{2}\right) + \frac{2}{5}\left(-\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2}\right) + \frac{1}{5}\left(-\frac{1}{1} \log_2 \frac{1}{1}\right) = 0.4$$

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

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

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

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

Gain (income) = Info (D) - Info$_{income}$(D) = 0.971 - 0.4 = 0.571
Gain (student) = Info (D) - Info$_{student}$(D) = 0.971 - 0 = 0.971 ✓
Gain (credit_rating) = Info(D) - Info$_{credit}$(D) = 0.971 - 0.951 = 0.02

age	income	Student	Credit	buy
31 - 40	high	no	fair	yes
31 - 40	Low	yes	excellent	yes
31 - 40	Medium	no	Excellent	yes
31 - 40	High	yes	fair	yes

Yes :4

No :0

Income
High Y:2/N:0

Medium Y:1/N:0

Low Y:1/N:0

Student,
Yes Y:2/N:0

No Y:2/N:0

Credit
Fair Y:2/N:0

Excellent Y:2/N:0

age	income	Student	Credit	buy
>40	medium	no	fair	yes
>40	Low	yes	fair	yes
>40	low	yes	excellent	no
>40	medium	yes	fair	yes
>40	medium	no	excellent	no

Yes :3

No :2

Income
High Y:0/N:0
Medium Y:2/N:1
Low Y:1/N:1

Student,
Yes Y:2/N:1
No Y:1/N:1

Credit
Fair   Y:3/N:0
Excellent Y:0/N:2

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

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

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

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

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

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

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

Gain (income) = Info (D) - Info$_{income}$ (D) = 0.971 - 0.951 = 0.02
Gain (student) = Info (D) - Info$_{student}$ (D) = 0.971 - 0.951 = 0.02
Gain (credit_rating) = Info(D) - Info$_{credit}$ (D) = 0.971 - 0 = 0.971

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

age	...	buy
<=30		no
<=30		no
<=30		no
<=30		yes
<=30		yes

age	...	buy
31 - 40		yes
31 - 40		yes
31 - 40		yes
31 - 40		yes

age	. .	buy
>40		yes
>40		yes
>40		no
>40		yes
>40		no

Grain student = 0.971

Buy

Grain credit_rating = 0.971

Student	buy
no	no
no	no
no	no
yes	yes
yes	yes

Credit	buy
fair	yes
fair	yes
excellent	no
fair	yes
excellent	no

No — Not buy

Yes — Buy

excellent — Not buy

fair — Buy