น.ส. อัทรวดี นาดี รหัส 6330210222 SC-SIDS

Homework 5

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

Attribute Selection with Information Gain

1. $$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

$$Info(D) = I(9,5) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.940$$
Yes(9), No(5)

$$Info_{age}(D) = \frac{5}{14}I(2,3) + \frac{4}{14}I(4,0) + \frac{5}{14}I(3,2) = 0.694$$
>=30(5), 31...40(4), >40(5)

$$Info_{Income}(D) = \frac{4}{14}I(2,2) + \frac{6}{14}I(4,2) + \frac{4}{14}I(3,1) = 0.911$$
high(4), medium(6), low(4)

$$Info_{student}(D) = \frac{7}{14}I(6,1) + \frac{7}{14}I(3,4) = 0.789$$
Yes(7), No(7)

$$Info_{cradit}(D) = \frac{8}{14}I(6,2) + \frac{6}{14}I(3,3) = 0.892$$
fair(8), excellent(6)

หา Gain ที่มีค่ามากที่สุด

$$Gain(A) = Info(D) - Info_A(D)$$

$Gain(age) = Info(D) - Info_{age}(D) = 0.940 - 0.694 = 0.246$ มีค่า Gain มากสุด
$Gain(Income) = Info(D) - Info_{income}(D) = 0.940 - 0.911 = 0.029$
$Gain(student) = Info(D) - Info_{student}(D) = 0.940 - 0.789 = 0.048$
$Gain(Cradit\_rating) = Info(D) - Info_{cradit}(D) = 0.940 - 0.892 = 0.048$

2. Age <=30

age	income	student	cradit_rating	buys_computer
<=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

$$Info(D) = I(2,3) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) = 0.971$$
Yes(2), No(3)

$$Info_{income}(D) = \frac{2}{5}I(0,2) + \frac{2}{5}I(1,1) + \frac{1}{5}I(1,0) = 0.4$$
high(2), medium(1), low(2)

$$Info_{student}(D) = \frac{2}{5}I(2,0) + \frac{3}{5}I(0,3) = 0$$
Yes(2), No(3)

$$Info_{cradit}(D) = \frac{3}{5}I(1,2) + \frac{2}{5}I(1,1) = 0.951$$
fair(3), excellent(2)

หา Gain ที่มีค่ามากที่สุด

Gain (income) = $Info(D) - Info_{income}(D) = 0.971 - 0.4 = 0.571$

Gain (student) = $Info(D) - Info_{student}(D) = 0.971 - 0 = 0.971$ มีค่ามากสุด

Gain (cradit_rating) = $Info(D) - Info_{cradit} = 0.971 - 0.951 = 0.02$

3. Age 31...40

age	income	student	cradit_rating	buys_computer
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

ถ้าอายุอยู่ในช่วง 30...40 เป็น yes

4. Age > 40

age	income	student	cradit_rating	buys_computer
>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

$$Info(D) = I(3,2) = -\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right) = 0.971$$
Yes(3), No(2)

$$Info_{income}(D) = \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1) = 0.951$$
medium(3), low(2)

$$Info_{student}(D) = \frac{3}{5}I(2,1) + \frac{2}{5}I(1,1) = 0.951$$
Yes(3), No(2)

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

หา Gain ที่มีค่ามากสุด

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(cradit_rating) = $Info(D) - Info_{cradit}(D)$ = 0.971 - 0 = **0.971** มีค่ามากสุด

สรุปการคำนวณหา Information gain ได้ Dicision tree ดังนี้