นางสาวนันทิชา วิชิต 623020526-8

Decision Tree Induction

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

หา Info (D)

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

$$= 0.94$$

หา $\text{Info}_{age}(D)$

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

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

$$I(4,0) = \frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\log_2\left(\frac{0}{4}\right) = 0$$

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

แทนค่า $\text{Info}_{age}(D) = \frac{5}{14}(0.971) + \frac{4}{14}(0) + \frac{5}{14}(0.971) = 0.694$

หา Gain(age)

$$\text{Gain(age)} = 0.94 - 0.694$$

$$= 0.246$$

หา $\text{Info}_{income}(D)$

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

$$I(2,2) = -\frac{2}{4}\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\log_2\left(\frac{2}{4}\right) = 1$$

$$I(4,2) = -\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right) = 0.918$$

$$I(3,1) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 0.811$$

แทนค่า $\text{Info}_{income}(D) = \frac{4}{14}(1) + \frac{6}{14}(0.918) + \frac{4}{14}(0.811) = 0.911$

ก) Gain(income)

$$Gain(income) = 0.94 - 0.911$$

$$= 0.029$$

ข) $Info_{student}(D)$

$$Info_{student}(D) = \underbrace{\frac{7}{14} I(6,1)}_{\text{Yes}\ (Y,N)} + \underbrace{\frac{7}{14} I(3,4)}_{\text{No}\ (Y,N)}$$

$$I(6,1) = -\frac{6}{7}\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right) = 0.592$$

$$I(3,4) = -\frac{3}{7}\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\log_2\left(\frac{4}{7}\right) = 0.985$$

แทนค่า $Info_{student}(D) = \frac{7}{14}(0.592) + \frac{7}{14}(0.985) = 0.789$

ค) Gain(student)

$$Gain(student) = 0.94 - 0.789$$

$$= 0.151$$

ง) $Info_{credit\_rating}(D)$

$$Info_{credit\_rating}(D) = \underbrace{\frac{8}{14} I(6,2)}_{\text{fair}\ (Y,N)} + \underbrace{\frac{6}{14} I(3,3)}_{\text{excellent}\ (Y,N)}$$

$$I(6,2) = -\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) = 0.8113$$

$$I(3,3) = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right) = 1$$

แทนค่า $Info_{credit\_rating}(D) = \frac{8}{14}(0.8113) + \frac{6}{14}(1) = 0.892$

จ) Gain(credit_rating)

$$Gain(credit\_rating) = 0.94 - 0.892$$

$$= 0.048$$

จาก Gain

Gain(age) = 0.246

Gain(income) = 0.029

Gain(student) = 0.151

Gain(credit_rating) = 0.048

เลือก Gain ที่มีค่ามากที่สุดมาเป็นตัวแรก ซึ่งในที่นี้คือ Gain(age)

age (<= 30)

หา Info(D) ของ age (<= 30)

$Info(D)_{age(<=30)} = I(2,3) = 0.971$

หา $Info_{income}(D)$ ของ age (<= 30)

$Info_{income}(D)_{age(<=30)} = \frac{2}{5} I(0,2) + \frac{2}{5} I(1,1) + \frac{1}{5} I(1,0)$ (high, medium, low)

$I(0,2) = -\frac{0}{2}\log_2(\frac{0}{2}) - \frac{2}{2}\log_2(\frac{2}{2}) = 0$

$I(1,1) = -\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) = 1$

$I(1,0) = -\frac{1}{1}\log_2(\frac{1}{1}) - \frac{0}{1}\log_2(\frac{0}{1}) = 0$

แทนค่า $Info_{income}(D)_{age(<=30)} = \frac{2}{5}(0) + \frac{2}{5}(1) + \frac{1}{5}(0) = 0.4$

หา Gain(income) ของ age (<= 30)

$Gain(income)_{age(<=30)} = 0.971 - 0.4 = 0.571$

หา $Info_{student}(D)$ ของ age (<= 30)

$Info_{student}(D)_{age(<=30)} = \frac{2}{5} I(2,0) + \frac{3}{5} I(0,3)$ (Yes, No)

สังเกต Yes → Yes (buy_computer), No → No (buy_computer)

เรียงแบ่งต่อ Student เพราะมากแบ่งแล้วได้ความแม่นยำ

age (>40)

on Info (D) vos age (>40)

$$\text{Info}(D)_{\text{vos age}(>40)} = I(3,2) = 0.971$$

on $\text{Info}_{\text{income}}(D)$ vos age (>40)

$$\text{Info}_{\text{income}}(D)_{\text{vos age}(>40)} = \underbrace{\frac{2}{5} I(2,1)}_{\text{medium}} + \underbrace{\frac{2}{5} I(1,1)}_{\text{low}}$$

$$I(2,1) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.918$$

$$I(1,1) = 1$$

$$\text{innoin } \text{Info}_{\text{income}}(D)_{\text{vos age}(>40)} = \frac{3}{5}(0.918) + \frac{2}{5}(1) = 0.959$$

on Gain (income) vos age (>40)

$$\text{Gain(income)}_{\text{vos age}(>40)} = 0.971 - 0.959 = 0.02$$

on $\text{Info}_{\text{student}}(D)$ vos age (>40)

$$\text{Info}_{\text{student}}(D)_{\text{vos age}(>40)} = \underbrace{\frac{3}{5} I(2,1)}_{\text{Yes}} + \underbrace{\frac{2}{5} I(1,1)}_{\text{No}}$$

$$I(2,1) = -\frac{2}{3}\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) = 0.918$$

$$I(1,1) = 1$$

$$\text{innoin } \text{Info}_{\text{student}}(D)_{\text{vos age}(>40)} = \frac{3}{5}(0.918) + \frac{2}{5}(1) = 0.951$$

on Gain (student) vos age (>40)

$$\text{Gain (student)}_{\text{vos age}(>40)} = 0.971 - 0.951$$

$$= 0.02$$

ต่อ $Info_{credit\_rating}(D)$ เมื่อ age (>40)

$Info_{credit\_rating}(D)$ เมื่อ age (>40) = $\frac{3}{5}I(3,0)$ + $\frac{2}{5}I(0,2)$

(fair, excellent)

ได้ผล fair → Yes (buy_computer) , excellent → No (buy_computer)

เรียกแบ่งต่อ credit_rating เพราะได้ผลการแบ่งบริสุทธิ์ที่สุด

สรุป