- Class P: buys_computer = "yes"
- Class N: buys_computer = "no"

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

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

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

$\frac{5}{14}I(2,3)$ means "age <=30" has 5 out of 14 samples, with 2 yes'es and 3 no's.

Hence

$$Gain(age) = Info(D) - Info_{age}(D) = 0.246$$

Similarly, we can get

$$Gain(income) = 0.029$$
$$Gain(student) = 0.151$$
$$Gain(credit\_rating) = 0.048$$

Info(D) = I(8,4) = -log(8/12)log2(8/12)-(4/12)log2(4/12)=0.9183

Infoage(D)

Age	pi	ni	I(pi,ni)
<=30	2	2	1
31…40	3	0	0
>40	3	2	0.737

Infoage(D): 4/12(I(2,2))+3/12(I(3,0))+5/12(I(3,2))=0.64

Gain(age)=0.918-0.64=0.278

Infoincome(D)

Income	Pi	Ni	I(pi,ni)
High	2	2	1
Medium	4	1	0.722
Low	2	1	0.918

Infoincome(D)=0.864

Gain(student)=0.918-0.864=0.054

Infostudent

student	Pi	Ni	I(pi,ni)
Yes	5	1	0.65
No	3	3	1

Infostudent(D)=0.825

Gain(student) = 0.918-0.825=0.093

Infocredit

Credit_rating	Pi	Ni	I(pi,ni)
Fair	6	1	0.592
Excellent	2	3	0.971

Infocredit = 0.749

Gain(credit)= 0.169

Age มีค่าGainมากที่สุดจึงเลือก age เป็น root node