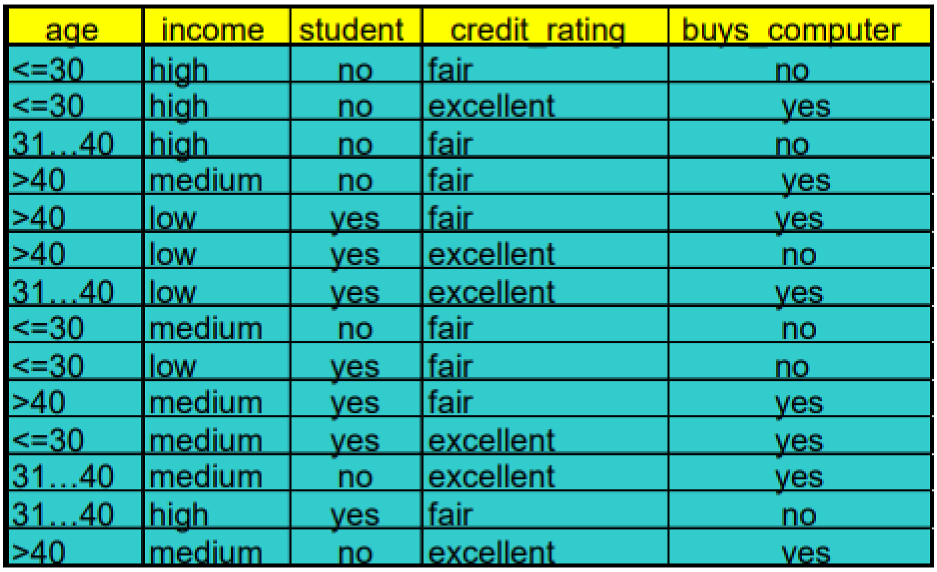


age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	yes
31…40	high	no	fair	no
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	no
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	no
>40	medium	no	excellent	yes

- We must determine whether we will make the first split, on age on income on student or on credit_rating.

We will use information gain.

Info(D)=I(8,6)=-8/4*$\log_2(\frac{8}{14})$-6/14*$\log_2(\frac{6}{14})$=0.985

Info$_{age}$(D)=5/14*I(2,3)+4/14*I(2,2)+5/14*I(4,1)=5/14*(-2/5*$\log_2(\frac{2}{5})$-3/5*$\log_2(\frac{3}{5})$)+

4/14*(-2/4*$\log_2\left(\frac{2}{4}\right)$-2/4*$\log_2(\frac{2}{4})$)+5/14*(-4/5*$\log_2(\frac{4}{5})$-1/5*$\log_2(\frac{1}{5})$)=0,8903

Info$_{student}$(D)=7/14*I(4,3)+7/14*I(4,3)=7/14*I(4,3)+7/14*I(4,3)=

7/14*(-4/7*$\log_2(\frac{4}{7})$-3/7*$\log_2(\frac{3}{7})$)+ 7/14*(-4/7*$\log_2(\frac{4}{7})$-3/7*$\log_2(\frac{3}{7})$)=0.985

Info$_{credit\_rating}$(D)=8/14*I(3,5)+6/14*I(5,1)=8/14*(-3/8*$\log_2(\frac{3}{8})$-5/8*$\log_2(\frac{5}{8})$)

+6/14*(-5/6*$\log_2(\frac{5}{6})$-1/6*$\log_2(\frac{1}{6})$)=0.823

Info$_{income}$(D)=4/14*I(2,2)+6/14*I(5,1)+4/14*I(1,3)=4/14*(-2/4*$\log_2(\frac{2}{4})$-2/4*$\log_2(\frac{2}{4})$)
+6/14*(-5/6*$\log_2(\frac{5}{6})$-1/6*$\log_2(\frac{1}{6})$)+ 4/14*(-1/4*$\log_2(\frac{1}{4})$-3/4*$\log_2(\frac{3}{4})$)=0.796

Gain(age)= Info(D)- Info$_{age}$(D)=0.0947

Gain(income)= Info(D)- Info$_{income}$(D)=0.192

Gain(student)= Info(D)- Info$_{student}$(D)=0

Gain(credit_rating)= Info(D)- Info$_{credit\_rating}$(D)=0.162

As a result, the first split on decision tree will be based on income.

- Regarding the low case in income we must determine where we must split it om age, on student or on credit_rating

Info(D)=I(2,2)=-2/4$*\log_2(\frac{2}{4})$-2/4$*\log_2(\frac{2}{4})$=1

Info_{age}(D)=1/4*I(0,1)+1/4*I(1,0)+2/4*I(1,1)=2/4*(-1/2*$\log_2(\frac{1}{2})$-1/2*$\log_2(\frac{1}{2})$)=0.5

$\mathrm{Info}_{student}$(D)=I(2,2)=-2/4$*\log_2(\frac{2}{4})$-2/4$*\log_2(\frac{2}{4})$=1

$\mathrm{Info}_{credit\_rating}$(D)=2/4*I(2,0)+2/4*I(1,1)= 2/4*(-1/2*$\log_2(\frac{1}{2})$-1/2*$\log_2(\frac{1}{2})$)=0.5

Gain(age)=Info(D)- Info_{age}(D)=0.5

Gain(student)=Info(D)- $\mathrm{Info}_{student}$(D)=0

Gain(credit)=Info(D)- $\mathrm{Info}_{credit\_rating}$(D)=0.5

As the info gain is the same for age and credit_rating we make the split wherever we want.

So we split by age.

- Regarding the medium case in income we must determine whether we will split it on age, on student or on credit_rating

Info(D)=I(5,1)=-5/6*$\log_2(\frac{5}{6})$-1/6*$\log_2(\frac{1}{6})$=0.65

Info_{age}(D)=2/6*I(1,1)+1/6*I(1,0)+3/6*I(3,0)=2/6*(-1/2*$\log_2(\frac{1}{2})$-1/2*$\log_2(\frac{1}{2})$)=0.333

$\mathrm{Info}_{student}$(D)=2/6*I(2,0)+4/6*I(3,1)=4/6*(-3/4*$\log_2(\frac{3}{4})$-1/4*$\log_2(\frac{1}{4})$)=0.54

$\mathrm{Info}_{credit\_rating}$(D)=3/6*I(2,1)+3/6*I(3,0)=3/6*(-2/3*$\log_2(\frac{2}{3})$-1/3*$\log_2(\frac{1}{3})$)=0.459

Gain(age)=Info(D)- Info_{age}(D)=0.317

Gain(student)=Info(D)- $\mathrm{Info}_{student}$(D)=0.11

Gain(credit_rating)=Info(D)- $\mathrm{Info}_{credit\_rating}$(D)=0.191

As a result, we will split the medium case on income based on age

- Regarding the high case in income we must determine whether we will split it on age, on student or on credit_rating

Info(D)=I(2,2)=-2/4$*\log_2(\frac{2}{4})$-2/4$*\log_2(\frac{2}{4})$=1

Info_{age}(D)=2/4*I(1,1)+2/4*I(0,2)=2/4*(-1/2*$\log_2(\frac{1}{2})$-1/2*$\log_2(\frac{1}{2})$)=0.5

$\mathrm{Info}_{student}$(D)=1/4*I(0,1)+3/4*I(0,2)=3/4*(-1/3*$\log_2(\frac{1}{3})$-2/3*$\log_2(\frac{2}{3})$)=0.688

$\mathrm{Info}_{credit\_rating}$(D)=3/4*I(0,3)+1/4*I(1,0)=0

Gain(age)=Info(D)- Info_{age}(D)=0.5

Gain(student)=Info(D)- $\mathrm{Info}_{student}$(D)=0.312

Gain(credit_rating)=Info(D)- $\mathrm{Info}_{credit\_rating}$(D)=1

As a result, we will split the high income based on credit_rating.

- In low income:

As for the ≤30 age in low income it does not need further split because there is one such instance.

Regarding the 31…40 age in low income it does not need further split because there is one instance.

Regarding the age >40 in low income we must define where we will split it on student or on credit_rating.

Info(D)=I(1,1)=-1/2*$\log_2(\frac{1}{2})$-1/2*$\log_2(\frac{1}{2})$=1

$\mathrm{Info}_{student}$(D)=I(1,1)=1

$\mathrm{Info}_{credit\_rating}$(D)=1/2*I(1,0)+1/2*I(0,1)=0

Gain(student)=Info(D)- $\mathrm{Info}_{student}$(D)=0

Gain(credit_rating)=Info(D)- $\mathrm{Info}_{credit\_rating}$(D)=1

As a result, we will split the age >40 in low income based on credit_rating.

The fair and excellent credit_rating in >40 age in low income does not need further split because they contain only one instance.

- In medium income:

As for the age ≤30 in medium income we must determine whether we must split it on student or on credit_rating.

Info(D)=I(1,1)=1

$\mathrm{Info}_{credit\_rating}$(D)=1/2*I(0,1)+1/2*I(1,0)=0

$\mathrm{Info}_{student}$(D)=1/2*I(0,1)+1/2*I(1,0)=0

Gain(credit_rating)=1

Gain(student)=1

As the information gain is the same for credit_rating and student we will split it wherever we want.

So we will split by student.

As for the case student=no in age ≤30 in medium income it does not need any further split because it contains one instance.

Also, the case student=yes in age ≤30 in medium income it does not need any further split because it contains one instance.

As for the age >40 in medium income it does not need any further split because it contains only the instance buy_computer=yes.

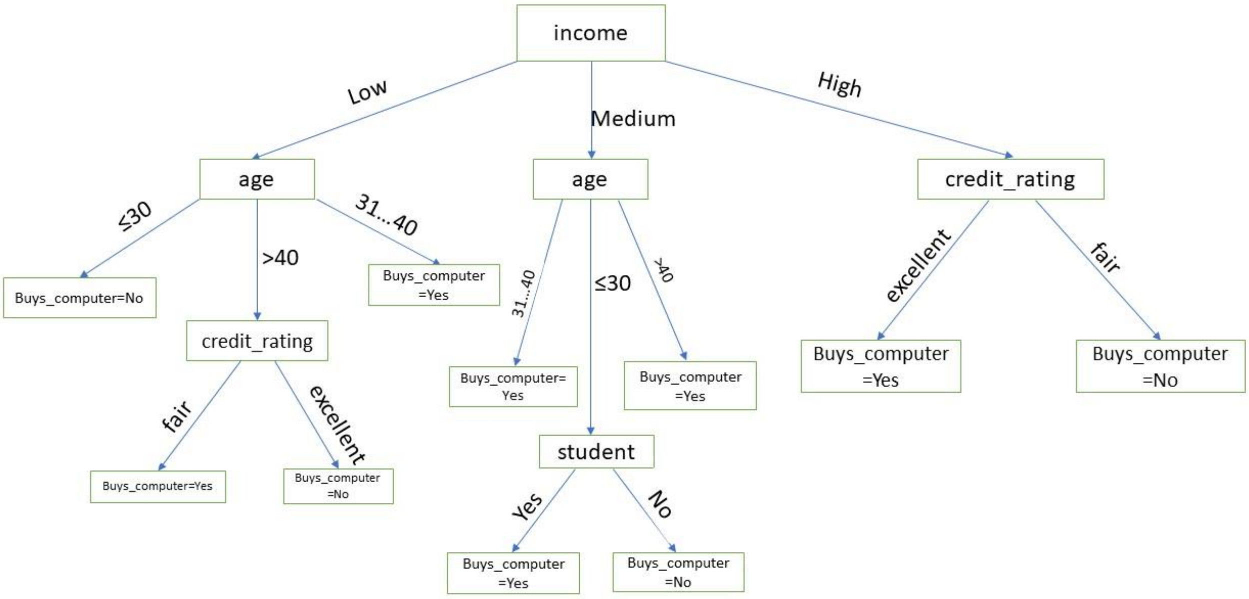
As for the age 31…40 in medium income it does not need to be split because it contains only one instance.

- In high income:

As for the excellent credit_rating in high income it does not need αvυ further split because it contains only one instance.

As for the fair credit_rating in high income it does not need any further split because it contains only the case buys computer=no.

Final Decision Tree:


income
Low
Medium
High
age
≤30
>40
31…40
Buys_computer=No
Buys_computer =Yes
credit_rating
fair
excellent
Buys_computer=Yes
Buys_computer =No
age
31…40
≤30
>40
Buys_computer= Yes
Buys_computer =Yes
student
Yes
No
Buys_computer =Yes
Buys_computer =No
credit_rating
excellent
fair
Buys_computer =Yes
Buys_computer =No