CS 4445 B Term 2006
Homework 1 Solutions

Decision Tree Construction

?

We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*more*	*low*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*
vhigh	*vhigh*	*more*	*med*	*unacc*
med	*high*	*4*	*low*	*unacc*
high	*med*	*4*	*high*	*good*
low	*med*	*2*	*med*	*unacc*
low	*high*	*2*	*high*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*
med	*med*	*more*	*med*	*acc*
med	*vhigh*	*2*	*med*	*unacc*
med	*med*	*4*	*low*	*unacc*
med	*vhigh*	*more*	*low*	*unacc*
med	*low*	*4*	*med*	*acc*
high	*low*	*2*	*high*	*unacc*
high	*med*	*4*	*low*	*unacc*
med	*low*	*4*	*low*	*unacc*
high	*high*	*4*	*low*	*unacc*
low	*med*	*4*	*high*	*good*
low	*low*	*2*	*high*	*unacc*

We have multiple attributes to pick from and thus we need to find the one that will give us the maximum the information gain.

We calculate the entropy over all our instances:

$$\left(-\frac{5}{22}\log_2\frac{5}{22}\right)+\left(-\frac{2}{22}\log_2\frac{2}{22}\right)+\left(-\frac{15}{22}\log_2\frac{15}{22}\right)=1.177$$

The attributes we have to split the instances with are: **buying-price**, **maintenance**, **persons**, **safety**. We calculate the information gains after splitting the instances by each of those attributes:

- **buying-price** - We calculate the entropy of instances filtered by the value of their **buying-price** attribute:

– **buying-price**=*high* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
high	*med*	*4*	*high*	*good*
high	*low*	*2*	*high*	*unacc*
high	*med*	*4*	*low*	*unacc*
high	*high*	*4*	*low*	*unacc*

$$\left(-\frac{0}{4}\right)+\left(-\frac{1}{4}\log_2\frac{1}{4}\right)+\left(-\frac{3}{4}\log_2\frac{3}{4}\right)=0.811$$

– **buying-price**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*more*	*low*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*
med	*high*	*4*	*low*	*unacc*
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*med*	*more*	*med*	*acc*
med	*vhigh*	*2*	*med*	*unacc*
med	*med*	*4*	*low*	*unacc*
med	*vhigh*	*more*	*low*	*unacc*
med	*low*	*4*	*med*	*acc*
med	*low*	*4*	*low*	*unacc*

$$\left(-\frac{4}{11}\log_2\frac{4}{11}\right)+\left(-\frac{0}{11}\right)+\left(-\frac{7}{11}\log_2\frac{7}{11}\right)=0.946$$

– **buying-price**=*vhigh* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*more*	*med*	*unacc*
vhigh	*vhigh*	*4*	*med*	*unacc*

$$\left(-\frac{0}{2}\right)+\left(-\frac{0}{2}\right)+\left(-\frac{2}{2}\log_2\frac{2}{2}\right)=0.000$$

– **buying-price**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
low	*med*	*2*	*med*	*unacc*
low	*high*	*2*	*high*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
low	*med*	*4*	*high*	*good*
low	*low*	*2*	*high*	*unacc*

$$\left(-\frac{1}{5}\log_2\frac{1}{5}\right)+\left(-\frac{1}{5}\log_2\frac{1}{5}\right)+\left(-\frac{3}{5}\log_2\frac{3}{5}\right)=1.371$$

Thus the information gain if the split is to be made using the **buying-price** attribute is $1.177-\frac{4}{22}0.811+\frac{11}{22}0.946+\frac{2}{22}0.000+\frac{5}{22}1.371=0.245$.

- **maintenance** - We calculate the entropy of instances filtered by the value of their **maintenance** attribute:

– **maintenance**=*high* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*high*	*4*	*low*	*unacc*
low	*high*	*2*	*high*	*unacc*
high	*high*	*4*	*low*	*unacc*

$$\left(-\frac{0}{3}\right)+\left(-\frac{0}{3}\right)+\left(-\frac{3}{3}\log_2\frac{3}{3}\right)=0.000$$

– **maintenance**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
high	*med*	*4*	*high*	*good*
low	*med*	*2*	*med*	*unacc*
med	*med*	*more*	*med*	*acc*
med	*med*	*4*	*low*	*unacc*
high	*med*	*4*	*low*	*unacc*
low	*med*	*4*	*high*	*good*

$$\left(-\frac{1}{6}\log_2\frac{1}{6}\right)+\left(-\frac{2}{6}\log_2\frac{2}{6}\right)+\left(-\frac{3}{6}\log_2\frac{3}{6}\right)=1.459$$

– **maintenance**=*vhigh* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*more*	*low*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*
vhigh	*vhigh*	*more*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*
med	*vhigh*	*more*	*low*	*unacc*

$$\left(-\frac{3}{9}\log_2\frac{3}{9}\right)+\left(-\frac{0}{9}\right)+\left(-\frac{6}{9}\log_2\frac{6}{9}\right)=0.918$$

– **maintenance**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*low*	*4*	*med*	*acc*
high	*low*	*2*	*high*	*unacc*
med	*low*	*4*	*low*	*unacc*
low	*low*	*2*	*high*	*unacc*

$$\left(-\frac{1}{4}\log_2\frac{1}{4}\right)+\left(-\frac{0}{4}\right)+\left(-\frac{3}{4}\log_2\frac{3}{4}\right)=0.811$$

Thus the information gain if the split is to be made using the **maintenance** attribute is $1.177-\frac{3}{22}0.000+\frac{6}{22}1.459+\frac{9}{22}0.918+\frac{4}{22}0.811=0.256$.

- **persons** - We calculate the entropy of instances filtered by the value of their **persons** attribute:

  – **persons**=*2* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*2*	*med*	*unacc*
low	*med*	*2*	*med*	*unacc*
low	*high*	*2*	*high*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*
high	*low*	*2*	*high*	*unacc*
low	*low*	*2*	*high*	*unacc*

$$\left(-\frac{0}{6}\right)+\left(-\frac{0}{6}\right)+\left(-\frac{6}{6}\log_2\frac{6}{6}\right)=0.000$$

– **persons**=*4* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*high*	*4*	*low*	*unacc*
high	*med*	*4*	*high*	*good*
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*
med	*med*	*4*	*low*	*unacc*
med	*low*	*4*	*med*	*acc*
high	*med*	*4*	*low*	*unacc*
med	*low*	*4*	*low*	*unacc*
high	*high*	*4*	*low*	*unacc*
low	*med*	*4*	*high*	*good*

$$\left(-\frac{3}{11}\log_2\frac{3}{11}\right)+\left(-\frac{2}{11}\log_2\frac{2}{11}\right)+\left(-\frac{6}{11}\log_2\frac{6}{11}\right)=1.435$$

– **persons**=*more* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*more*	*low*	*unacc*
vhigh	*vhigh*	*more*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
med	*med*	*more*	*med*	*acc*
med	*vhigh*	*more*	*low*	*unacc*

$$\left(-\frac{2}{5}\log_2\frac{2}{5}\right)+\left(-\frac{0}{5}\right)+\left(-\frac{3}{5}\log_2\frac{3}{5}\right)=0.971$$

Thus the information gain if the split is to be made using the **persons** attribute is $1.177-\frac{6}{22}0.000+\frac{11}{22}1.435+\frac{5}{22}0.971=0.239$.

- **safety** - We calculate the entropy of instances filtered by the value of their **safety** attribute:

  – **safety**=*high* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
high	*med*	*4*	*high*	*good*
low	*high*	*2*	*high*	*unacc*
high	*low*	*2*	*high*	*unacc*
low	*med*	*4*	*high*	*good*
low	*low*	*2*	*high*	*unacc*

$$\left(-\frac{0}{5}\right)+\left(-\frac{2}{5}\log_2\frac{2}{5}\right)+\left(-\frac{3}{5}\log_2\frac{3}{5}\right)=0.971$$

– **safety**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*2*	*med*	*unacc*
vhigh	*vhigh*	*more*	*med*	*unacc*
low	*med*	*2*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*
med	*med*	*more*	*med*	*acc*
med	*vhigh*	*2*	*med*	*unacc*
med	*low*	*4*	*med*	*acc*

$$\left(-\frac{5}{10}\log_2\frac{5}{10}\right)+\left(-\frac{0}{10}\right)+\left(-\frac{5}{10}\log_2\frac{5}{10}\right)=1.000$$

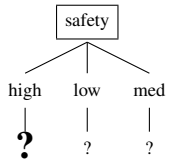
– **safety**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*more*	*low*	*unacc*
med	*high*	*4*	*low*	*unacc*
med	*med*	*4*	*low*	*unacc*
med	*vhigh*	*more*	*low*	*unacc*
high	*med*	*4*	*low*	*unacc*
med	*low*	*4*	*low*	*unacc*
high	*high*	*4*	*low*	*unacc*

$$\left(-\frac{0}{7}\right)+\left(-\frac{0}{7}\right)+\left(-\frac{7}{7}\log_2\frac{7}{7}\right)=0.000$$

Thus the information gain if the split is to be made using the **safety** attribute is $1.177-\frac{5}{22}0.971+\frac{10}{22}1.000+\frac{7}{22}0.000=0.502$.

We see that the best choice for the attribute to split with is **safety** with information gain of 0.502. We label our node with **safety**, create children nodes for each of its possible values (*high*, *med*, *low*), and proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
high	*med*	*4*	*high*	*good*
low	*high*	*2*	*high*	*unacc*
high	*low*	*2*	*high*	*unacc*
low	*med*	*4*	*high*	*good*
low	*low*	*2*	*high*	*unacc*

We have multiple attributes to pick from and thus we need to find the one that will give us the maximum the information gain.

We calculate the entropy over all our instances:

$$\left(-\frac{0}{5}\right)+\left(-\frac{2}{5}\log_2\frac{2}{5}\right)+\left(-\frac{3}{5}\log_2\frac{3}{5}\right)=0.971$$

The attributes we have to split the instances with are: **buying-price**, **maintenance**, **persons**. We calculate the information gains after splitting the instances by each of those attributes:

- **buying-price** - We calculate the entropy of instances filtered by the value of their **buying-price** attribute:
  - **buying-price**=*high* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
high	*med*	*4*	*high*	*good*
high	*low*	*2*	*high*	*unacc*

$$\left(-\frac{0}{2}\right)+\left(-\frac{1}{2}\log_2\frac{1}{2}\right)+\left(-\frac{1}{2}\log_2\frac{1}{2}\right)=1.000$$

  - **buying-price**=*med* - There are no instances with *med* for their **buying-price** attribute. Thus we use 0 for the entropy.
  - **buying-price**=*vhigh* - There are no instances with *vhigh* for their **buying-price** attribute. Thus we use 0 for the entropy.
  - **buying-price**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
low	*high*	*2*	*high*	*unacc*
low	*med*	*4*	*high*	*good*
low	*low*	*2*	*high*	*unacc*

$$\left(-\frac{0}{3}\right)+\left(-\frac{1}{3}\log_2\frac{1}{3}\right)+\left(-\frac{2}{3}\log_2\frac{2}{3}\right)=0.918$$

  Thus the information gain if the split is to be made using the **buying-price** attribute is $0.971-\frac{2}{5}1.000+\frac{0}{5}+\frac{0}{5}+\frac{3}{5}0.918=0.020$.

- **maintenance** - We calculate the entropy of instances filtered by the value of their **maintenance** attribute:
  - **maintenance**=*high* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
low	*high*	*2*	*high*	*unacc*

$$\left(-\frac{0}{1}\right)+\left(-\frac{0}{1}\right)+\left(-\frac{1}{1}\log_2\frac{1}{1}\right)=0.000$$

– **maintenance**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
high	*med*	*4*	*high*	*good*
low	*med*	*4*	*high*	*good*

$$\left(-\frac{0}{2}\right)+\left(-\frac{2}{2}\log_2\frac{2}{2}\right)+\left(-\frac{0}{2}\right)=0.000$$

– **maintenance**=*vhigh* - There are no instances with *vhigh* for their **maintenance** attribute. Thus we use 0 for the entropy.

– **maintenance**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
high	*low*	*2*	*high*	*unacc*
low	*low*	*2*	*high*	*unacc*

$$\left(-\frac{0}{2}\right)+\left(-\frac{0}{2}\right)+\left(-\frac{2}{2}\log_2\frac{2}{2}\right)=0.000$$

Thus the information gain if the split is to be made using the **maintenance** attribute is $0.971-\frac{1}{5}0.000+\frac{2}{5}0.000+\frac{0}{5}+\frac{2}{5}0.000=0.971$.

- **persons** - We calculate the entropy of instances filtered by the value of their **persons** attribute:

– **persons**=*2* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
low	*high*	*2*	*high*	*unacc*
high	*low*	*2*	*high*	*unacc*
low	*low*	*2*	*high*	*unacc*

$$\left(-\frac{0}{3}\right)+\left(-\frac{0}{3}\right)+\left(-\frac{3}{3}\log_2\frac{3}{3}\right)=0.000$$

– **persons**=*4* - The relevant instances are:

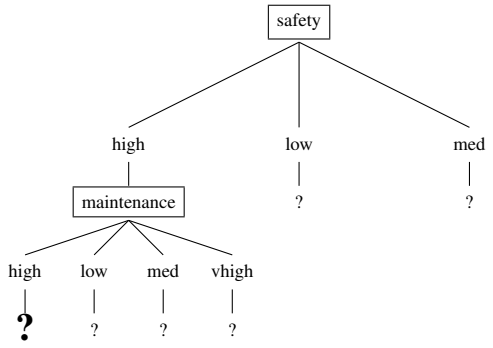
buying-price	maintenance	persons	safety	recommendation
high	*med*	*4*	*high*	*good*
low	*med*	*4*	*high*	*good*

$$\left(-\frac{0}{2}\right)+\left(-\frac{2}{2}\log_2\frac{2}{2}\right)+\left(-\frac{0}{2}\right)=0.000$$

– **persons**=*more* - There are no instances with *more* for their **persons** attribute. Thus we use 0 for the entropy.

Thus the information gain if the split is to be made using the **persons** attribute is $0.971-\frac{3}{5}0.000+\frac{2}{5}0.000+\frac{0}{5}=0.971$.

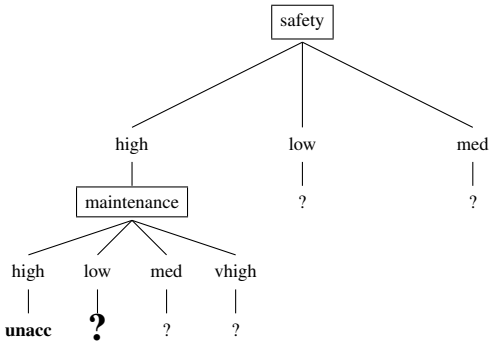
We see that the best choice for the attribute to split with is **maintenance** with information gain of 0.971. We label our node with **maintenance**, create children nodes for each of its possible values (*high*, *med*, *vhigh*, *low*), and proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
low	*high*	*2*	*high*	*unacc*

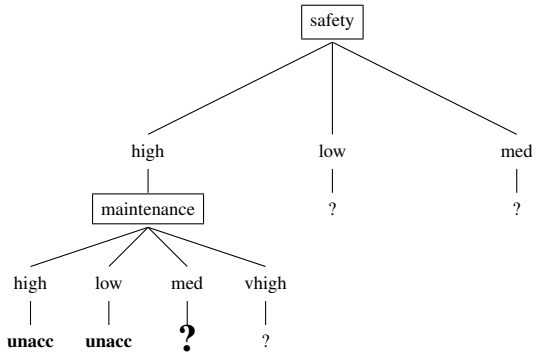
There is no entropy so we chose the only target value among these instances: **unacc**. We proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
high	*low*	*2*	*high*	*unacc*
low	*low*	*2*	*high*	*unacc*

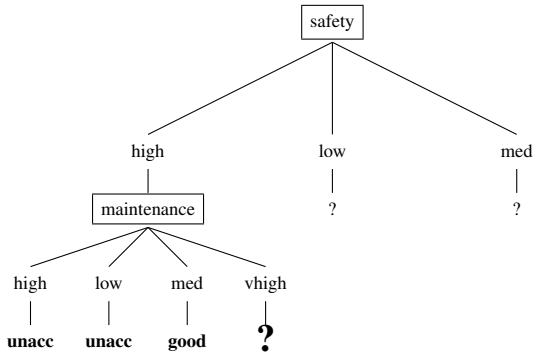
There is no entropy so we chose the only target value among these instances: **unacc**. We proceed to the next unexpanded node.



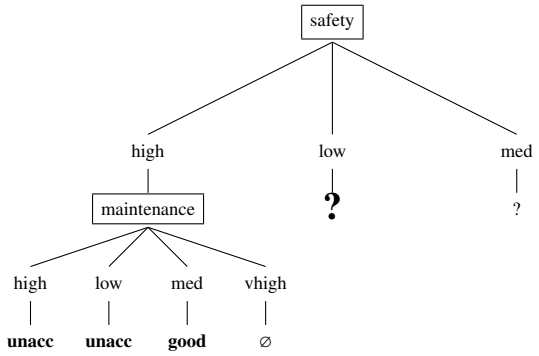
We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
high	*med*	*4*	*high*	*good*
low	*med*	*4*	*high*	*good*

There is no entropy so we chose the only target value among these instances: **good**. We proceed to the next unexpanded node.



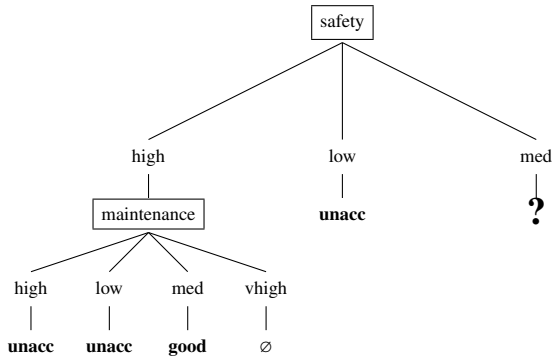
There are no instances relevant here. We label this node as ∅. Alternatively, we could have chosen the majority class over instances at the parent node: **unacc**. We continue with the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*more*	*low*	*unacc*
med	*high*	*4*	*low*	*unacc*
med	*med*	*4*	*low*	*unacc*
med	*vhigh*	*more*	*low*	*unacc*
high	*med*	*4*	*low*	*unacc*
med	*low*	*4*	*low*	*unacc*
high	*high*	*4*	*low*	*unacc*

There is no entropy so we chose the only target value among these instances: **unacc**. We proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*2*	*med*	*unacc*
vhigh	*vhigh*	*more*	*med*	*unacc*
low	*med*	*2*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*
med	*med*	*more*	*med*	*acc*
med	*vhigh*	*2*	*med*	*unacc*
med	*low*	*4*	*med*	*acc*

We have multiple attributes to pick from and thus we need to find the one that will give us the maximum the information gain.

We calculate the entropy over all our instances:

$$\left(-\frac{5}{10}\log_2\frac{5}{10}\right)+\left(-\frac{0}{10}\right)+\left(-\frac{5}{10}\log_2\frac{5}{10}\right)=1.000$$

The attributes we have to split the instances with are: **buying-price**, **maintenance**, **persons**. We calculate the information gains after splitting the instances by each of those attributes:

- **buying-price** - We calculate the entropy of instances filtered by the value of their **buying-price** attribute:
  - **buying-price**=*high* - There are no instances with *high* for their **buying-price** attribute. Thus we use 0 for the entropy.
  - **buying-price**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*2*	*med*	*unacc*
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*med*	*more*	*med*	*acc*
med	*vhigh*	*2*	*med*	*unacc*
med	*low*	*4*	*med*	*acc*

$$\left(-\frac{4}{6}\log_2\frac{4}{6}\right)+\left(-\frac{0}{6}\right)+\left(-\frac{2}{6}\log_2\frac{2}{6}\right)=0.918$$

  - **buying-price**=*vhigh* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*more*	*med*	*unacc*
vhigh	*vhigh*	*4*	*med*	*unacc*

$$\left(-\frac{0}{2}\right)+\left(-\frac{0}{2}\right)+\left(-\frac{2}{2}\log_2\frac{2}{2}\right)=0.000$$

  - **buying-price**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
low	*med*	*2*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*

$$\left(-\frac{1}{2}\log_2\frac{1}{2}\right)+\left(-\frac{0}{2}\right)+\left(-\frac{1}{2}\log_2\frac{1}{2}\right)=1.000$$

Thus the information gain if the split is to be made using the **buying-price** attribute is $1.000-\frac{0}{10}+\frac{6}{10}0.918+\frac{2}{10}0.000+\frac{2}{10}1.000=0.249$.

- **maintenance** - We calculate the entropy of instances filtered by the value of their **maintenance** attribute:
  - **maintenance**=*high* - There are no instances with *high* for their **maintenance** attribute. Thus we use 0 for the entropy.
  - **maintenance**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
low	*med*	*2*	*med*	*unacc*
med	*med*	*more*	*med*	*acc*

$$\left(-\frac{1}{2}\log_2\frac{1}{2}\right)+\left(-\frac{0}{2}\right)+\left(-\frac{1}{2}\log_2\frac{1}{2}\right)=1.000$$

  - **maintenance**=*vhigh* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*2*	*med*	*unacc*
vhigh	*vhigh*	*more*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*

$$\left(-\frac{3}{7}\log_2\frac{3}{7}\right)+\left(-\frac{0}{7}\right)+\left(-\frac{4}{7}\log_2\frac{4}{7}\right)=0.985$$

  - **maintenance**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*low*	*4*	*med*	*acc*

$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right)+\left(-\frac{0}{1}\right)+\left(-\frac{0}{1}\right)=0.000$$

Thus the information gain if the split is to be made using the **maintenance** attribute is $1.000-\frac{0}{10}+\frac{2}{10}1.000+\frac{7}{10}0.985+\frac{1}{10}0.000=0.110$.

- **persons** - We calculate the entropy of instances filtered by the value of their **persons** attribute:
  - **persons**=*2* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*2*	*med*	*unacc*
low	*med*	*2*	*med*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*

$$\left(-\frac{0}{3}\right)+\left(-\frac{0}{3}\right)+\left(-\frac{3}{3}\log_2\frac{3}{3}\right)=0.000$$

– **persons**=*4* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*
med	*low*	*4*	*med*	*acc*

$$\left(-\frac{3}{4}\log_2\frac{3}{4}\right)+\left(-\frac{0}{4}\right)+\left(-\frac{1}{4}\log_2\frac{1}{4}\right)=0.811$$

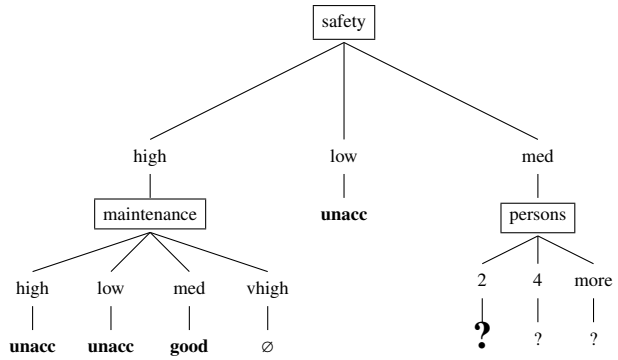
– **persons**=*more* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*more*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
med	*med*	*more*	*med*	*acc*

$$\left(-\frac{2}{3}\log_2\frac{2}{3}\right)+\left(-\frac{0}{3}\right)+\left(-\frac{1}{3}\log_2\frac{1}{3}\right)=0.918$$

Thus the information gain if the split is to be made using the **persons** attribute is $1.000-\frac{3}{10}0.000+\frac{4}{10}0.811+\frac{3}{10}0.918=0.400$.

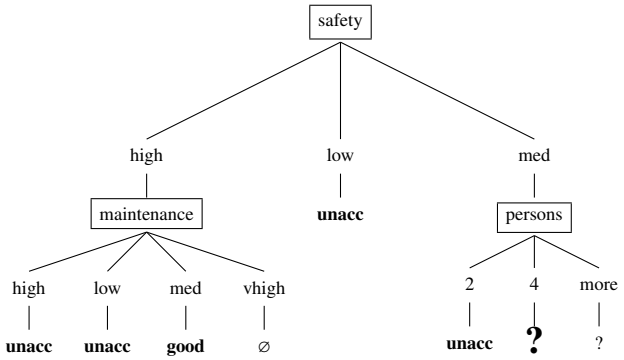
We see that the best choice for the attribute to split with is **persons** with information gain of 0.400. We label our node with **persons**, create children nodes for each of its possible values (*2*, *4*, *more*), and proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*2*	*med*	*unacc*
low	*med*	*2*	*med*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*

There is no entropy so we chose the only target value among these instances: **unacc**. We proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*
med	*low*	*4*	*med*	*acc*

We have multiple attributes to pick from and thus we need to find the one that will give us the maximum the information gain.

We calculate the entropy over all our instances:

$$\left(-\frac{3}{4}\log_2\frac{3}{4}\right)+\left(-\frac{0}{4}\right)+\left(-\frac{1}{4}\log_2\frac{1}{4}\right)=0.811$$

The attributes we have to split the instances with are: **buying-price**, **maintenance**. We calculate the information gains after splitting the instances by each of those attributes:

- **buying-price** - We calculate the entropy of instances filtered by the value of their **buying-price** attribute:
  - **buying-price**=*high* - There are no instances with *high* for their **buying-price** attribute. Thus we use 0 for the entropy.
  - **buying-price**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*low*	*4*	*med*	*acc*

$$\left(-\frac{3}{3}\log_2\frac{3}{3}\right)+\left(-\frac{0}{3}\right)+\left(-\frac{0}{3}\right)=0.000$$

  - **buying-price**=*vhigh* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*4*	*med*	*unacc*

$$\left(-\frac{0}{1}\right)+\left(-\frac{0}{1}\right)+\left(-\frac{1}{1}\log_2\frac{1}{1}\right)=0.000$$

  - **buying-price**=*low* - There are no instances with *low* for their **buying-price** attribute. Thus we use 0 for the entropy.

  Thus the information gain if the split is to be made using the **buying-price** attribute is $0.811 - \frac{0}{4} + \frac{3}{4}0.000 + \frac{1}{4}0.000 + \frac{0}{4} = 0.811$.

- **maintenance** - We calculate the entropy of instances filtered by the value of their **maintenance** attribute:

  - **maintenance**=*high* - There are no instances with *high* for their **maintenance** attribute. Thus we use 0 for the entropy.

  - **maintenance**=*med* - There are no instances with *med* for their **maintenance** attribute. Thus we use 0 for the entropy.

  - **maintenance**=*vhigh* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
vhigh	*vhigh*	*4*	*med*	*unacc*

$$\left(-\frac{2}{3}\log_2\frac{2}{3}\right)+\left(-\frac{0}{3}\right)+\left(-\frac{1}{3}\log_2\frac{1}{3}\right)=0.918$$

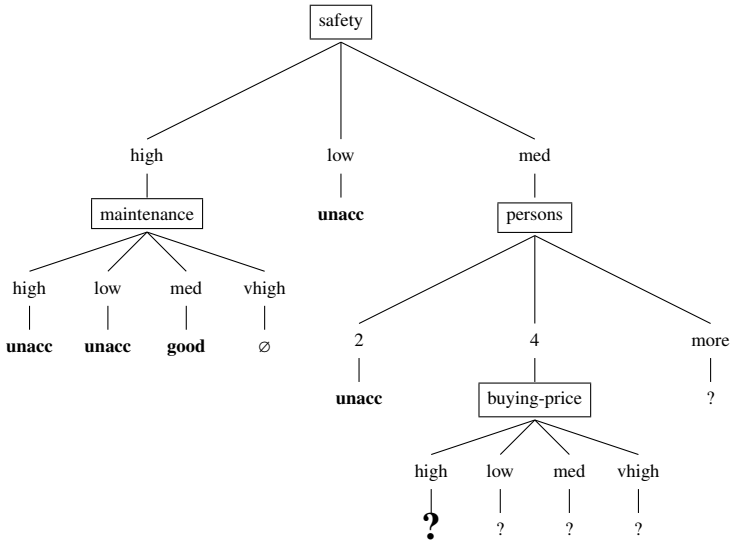
  - **maintenance**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*low*	*4*	*med*	*acc*

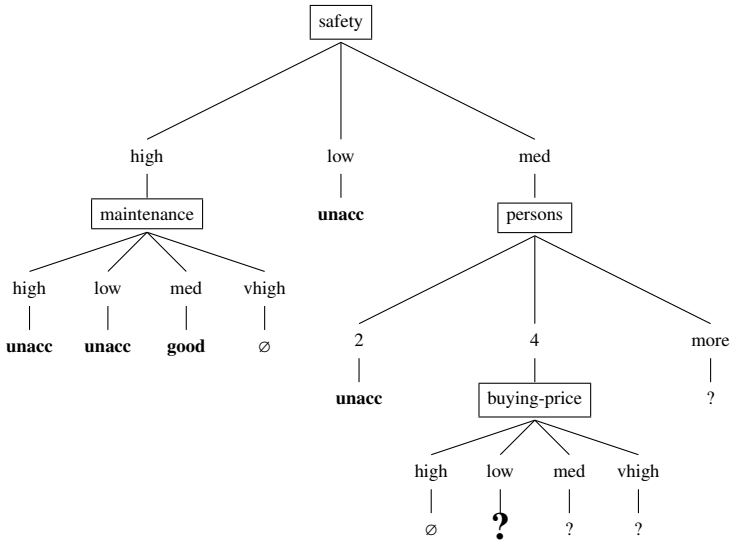
$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right)+\left(-\frac{0}{1}\right)+\left(-\frac{0}{1}\right)=0.000$$

  Thus the information gain if the split is to be made using the **maintenance** attribute is $0.811 - \frac{0}{4} + \frac{0}{4} + \frac{3}{4}0.918 + \frac{1}{4}0.000 = 0.123$.

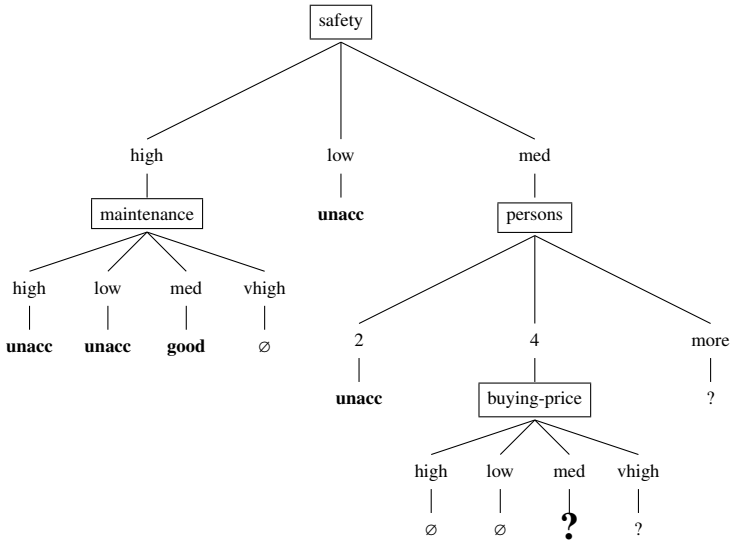
We see that the best choice for the attribute to split with is **buying-price** with information gain of 0.811. We label our node with **buying-price**, create children nodes for each of its possible values (*high*, *med*, *vhigh*, *low*), and proceed to the next unexpanded node.



There are no instances relevant here. We label this node as ∅. Alternatively, we could have chosen the majority class over instances at the parent node: **acc**. We continue with the next unexpanded node.



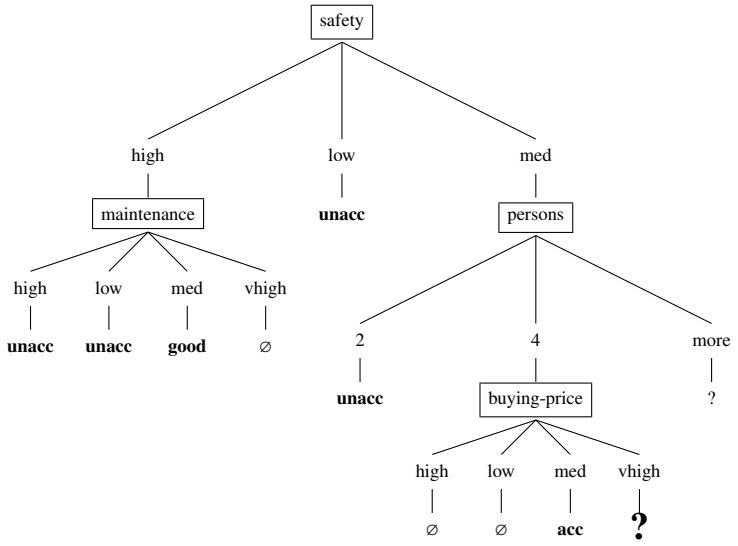
There are no instances relevant here. We label this node as ∅. Alternatively, we could have chosen the majority class over instances at the parent node: **acc**. We continue with the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
med	*vhigh*	*4*	*med*	*acc*
med	*vhigh*	*4*	*med*	*acc*
med	*low*	*4*	*med*	*acc*

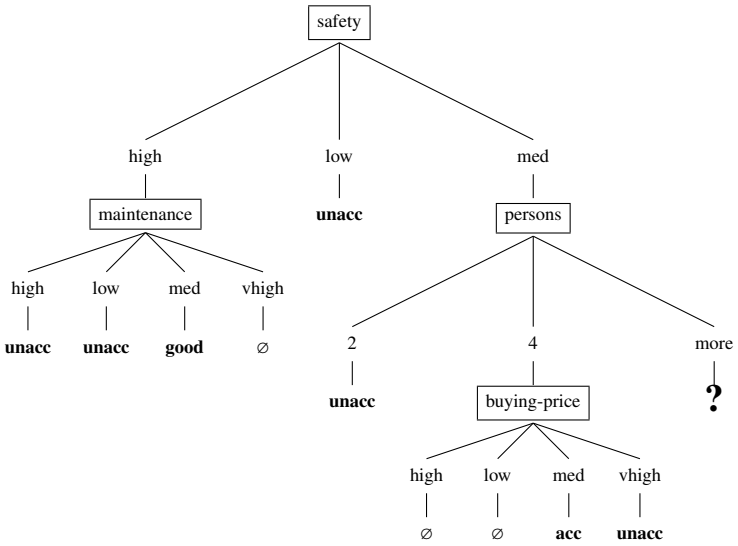
There is no entropy so we chose the only target value among these instances: **acc**. We proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*4*	*med*	*unacc*

There is no entropy so we chose the only target value among these instances: **unacc**. We proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*more*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*
med	*med*	*more*	*med*	*acc*

We have multiple attributes to pick from and thus we need to find the one that will give us the maximum the information gain.

We calculate the entropy over all our instances:

$$\left(-\frac{2}{3}\log_2\frac{2}{3}\right)+\left(-\frac{0}{3}\right)+\left(-\frac{1}{3}\log_2\frac{1}{3}\right)=0.918$$

The attributes we have to split the instances with are: **buying-price**, **maintenance**. We calculate the information gains after splitting the instances by each of those attributes:

- **buying-price** - We calculate the entropy of instances filtered by the value of their **buying-price** attribute:
  - **buying-price**=*high* - There are no instances with *high* for their **buying-price** attribute. Thus we use 0 for the entropy.
  - **buying-price**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*med*	*more*	*med*	*acc*

$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right)+\left(-\frac{0}{1}\right)+\left(-\frac{0}{1}\right)=0.000$$

  - **buying-price**=*vhigh* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*more*	*med*	*unacc*

$$\left(-\frac{0}{1}\right)+\left(-\frac{0}{1}\right)+\left(-\frac{1}{1}\log_2\frac{1}{1}\right)=0.000$$

  - **buying-price**=*low* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
low	*vhigh*	*more*	*med*	*acc*

$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right)+\left(-\frac{0}{1}\right)+\left(-\frac{0}{1}\right)=0.000$$

  Thus the information gain if the split is to be made using the **buying-price** attribute is $0.918-\frac{0}{3}+\frac{1}{3}0.000+\frac{1}{3}0.000+\frac{1}{3}0.000=0.918$.

- **maintenance** - We calculate the entropy of instances filtered by the value of their **maintenance** attribute:
  - **maintenance**=*high* - There are no instances with *high* for their **maintenance** attribute. Thus we use 0 for the entropy.
  - **maintenance**=*med* - The relevant instances are:

buying-price	maintenance	persons	safety	recommendation
med	*med*	*more*	*med*	*acc*

$$\left(-\frac{1}{1}\log_2\frac{1}{1}\right)+\left(-\frac{0}{1}\right)+\left(-\frac{0}{1}\right)=0.000$$

– **maintenance**=*vhigh* - The relevant instances are:

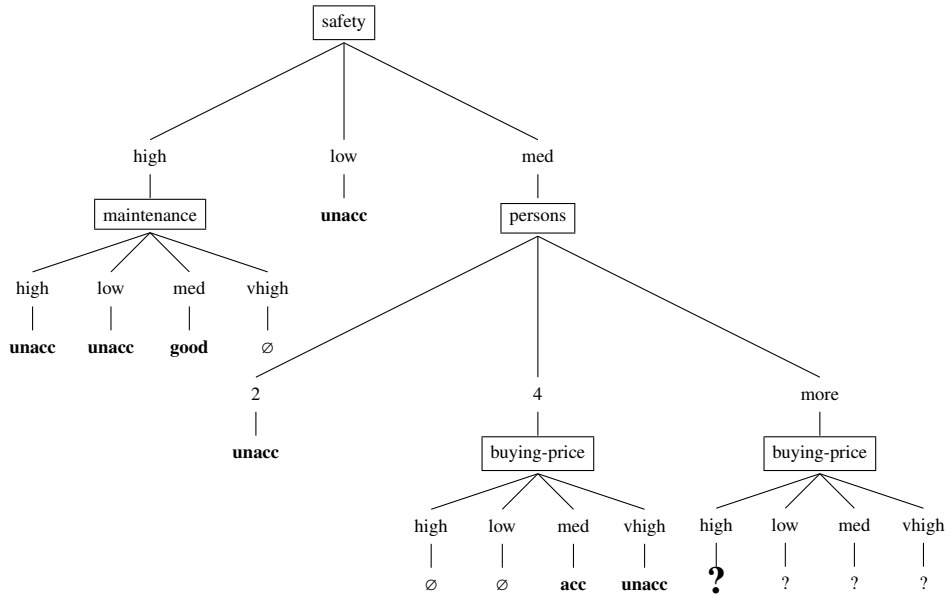
buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*more*	*med*	*unacc*
low	*vhigh*	*more*	*med*	*acc*

$$\left(-\frac{1}{2}\log_2\frac{1}{2}\right)+\left(-\frac{0}{2}\right)+\left(-\frac{1}{2}\log_2\frac{1}{2}\right)=1.000$$

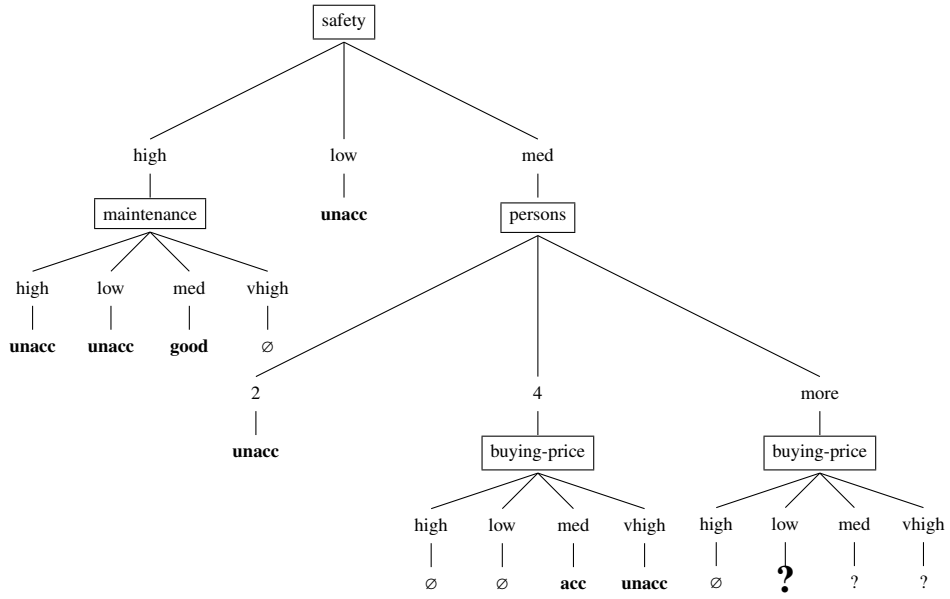
– **maintenance**=*low* - There are no instances with *low* for their **maintenance** attribute. Thus we use 0 for the entropy.

Thus the information gain if the split is to be made using the **maintenance** attribute is $0.918-\frac{0}{3}+\frac{1}{3}0.000+\frac{2}{3}1.000+\frac{0}{3}=0.252$.

We see that the best choice for the attribute to split with is **buying-price** with information gain of 0.918. We label our node with **buying-price**, create children nodes for each of its possible values (*high*, *med*, *vhigh*, *low*), and proceed to the next unexpanded node.



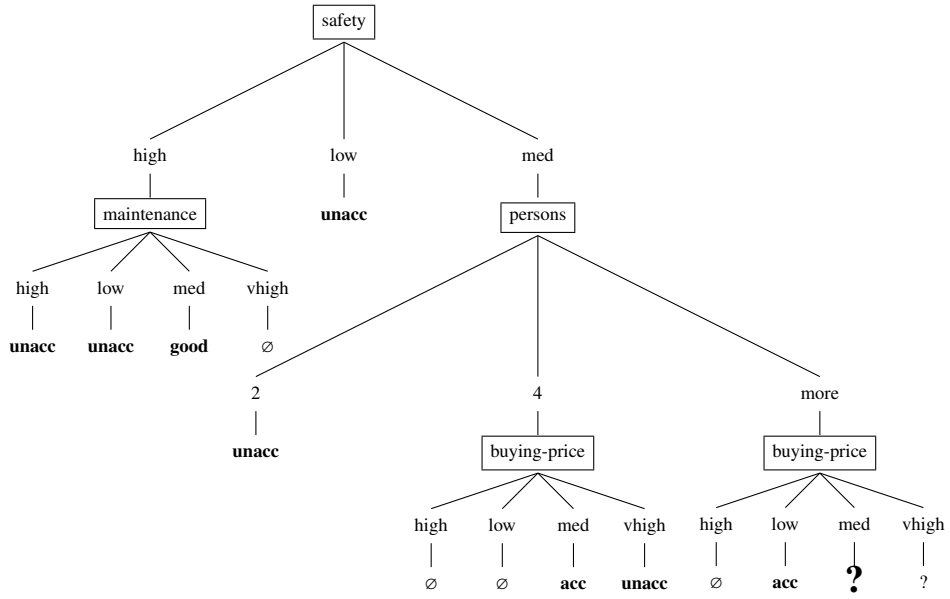
There are no instances relevant here. We label this node as ∅. Alternatively, we could have chosen the majority class over instances at the parent node: **acc**. We continue with the next unexpanded node.



We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
low	*vhigh*	*more*	*med*	*acc*

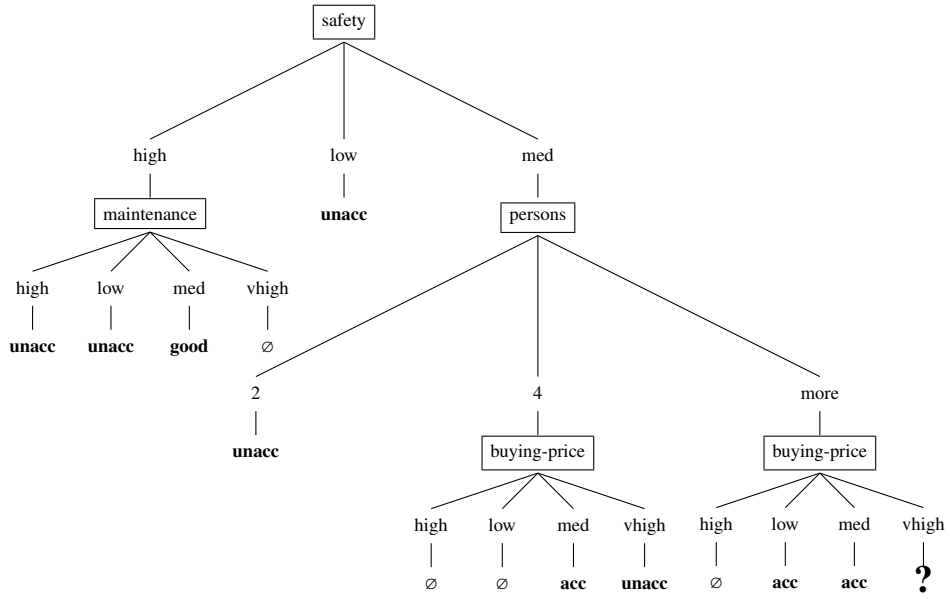
There is no entropy so we chose the only target value among these instances: **acc**. We proceed to the next unexpanded node.



We have the following instances at this point:

buying-price	**maintenance**	**persons**	**safety**	**recommendation**
med	*med*	*more*	*med*	*acc*

There is no entropy so we chose the only target value among these instances: **acc**. We proceed to the next unexpanded node.

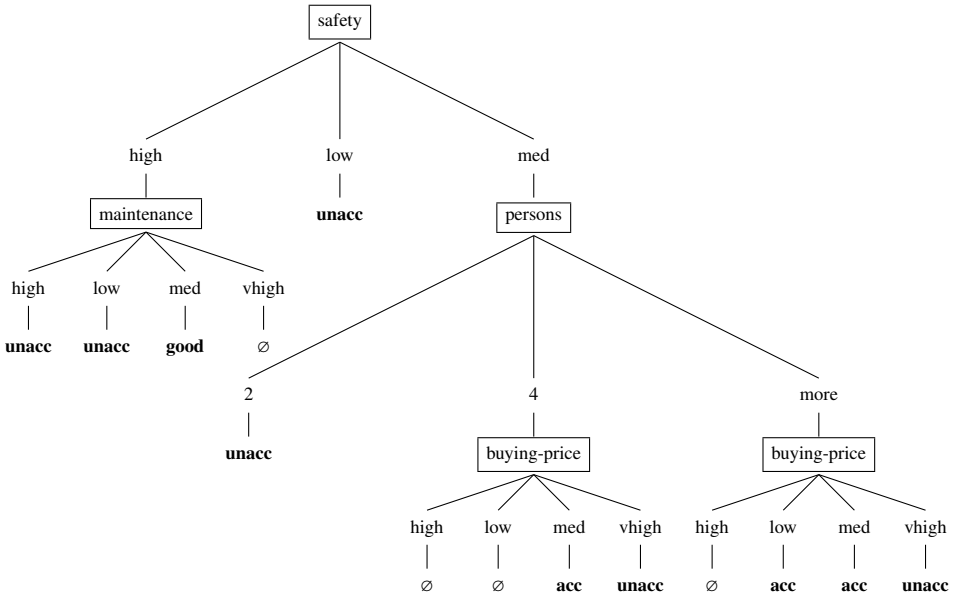


We have the following instances at this point:

buying-price	maintenance	persons	safety	recommendation
vhigh	*vhigh*	*more*	*med*	*unacc*

There is no entropy so we chose the only target value among these instances: **unacc**. We proceed to the next unexpanded node.

All nodes have been expanded as much as possible and thus we have our final tree:



Accuracy over Test Instances

Now we can use this tree to make predictions. We have the following results:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*unacc*

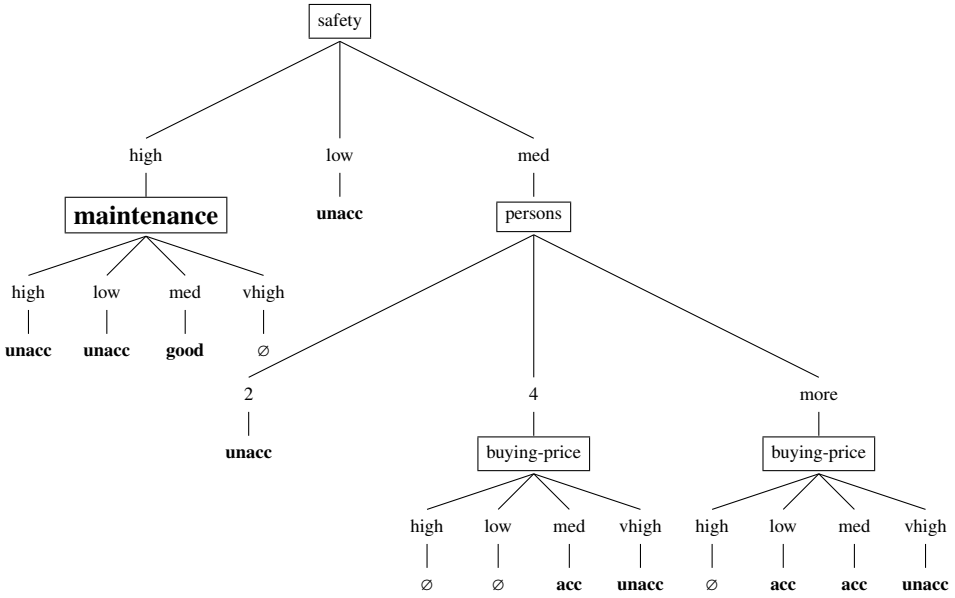
We see that the prediction was correct 9 out of 11 times and hence our accuracy is **81.82 %**.
Also we have the following confusion matrix (actual values on the left and predicted on the top):

	acc	**good**	**unacc**
acc	2	0	1
good	0	0	1
unacc	0	0	7

Subtree Replacement Prunning

We arbitrarily decide to consider potential nodes in a left-to-right manner. Thus the first node from the left that has only leafs as children is our first candidate. In our tree this is the left-most node labeled **maintenance**. The pruning technique is described for this node and the rest in detail below.

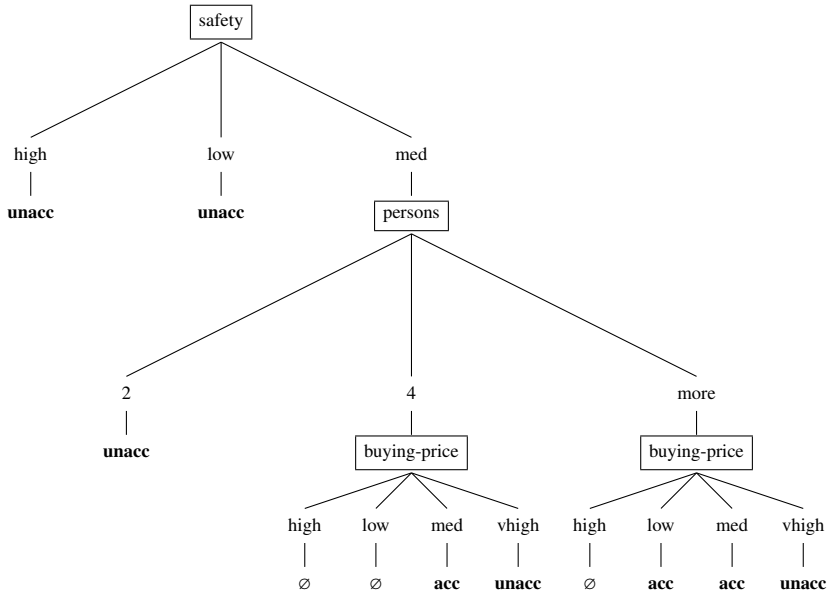
We consider pruning the tree at the following (bolded) node:



We determine the accuracy of the tree as is. The test instances are classified in the following manner:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*unacc*

Thus the accuracy of the tree over the training data with the node in question left alone is 81.818 %. We next consider the pruned tree. The node would be replaced with the majority class over training instances applicable for the node or *unacc* in this case.

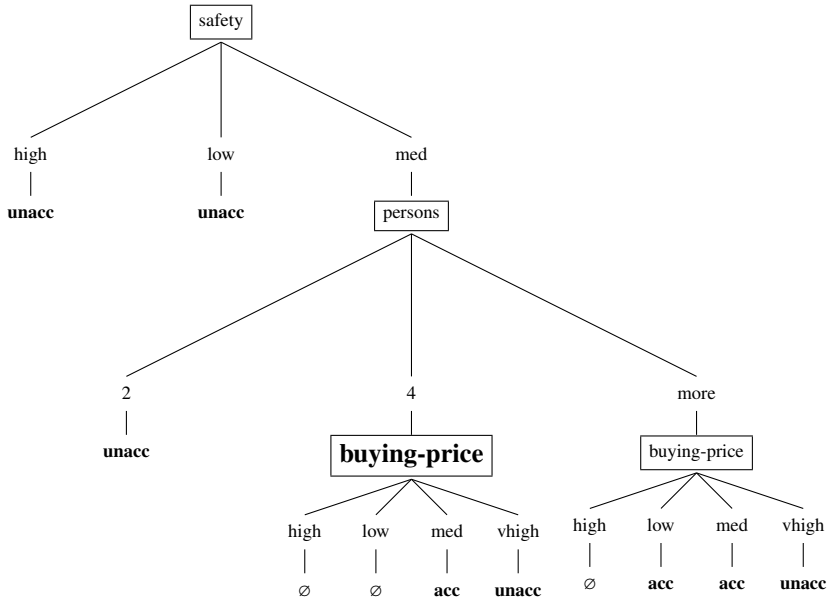


The test instances are classified in the following manner:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*unacc*

Thus if the node were to be replaced with the leaf with the decision *unacc* then the accuracy of the resulting tree would be 81.818 %. Since the new accuracy is at least as good as the old, we replace the node in question with the decision *unacc* and proceed to the next prune candidate.

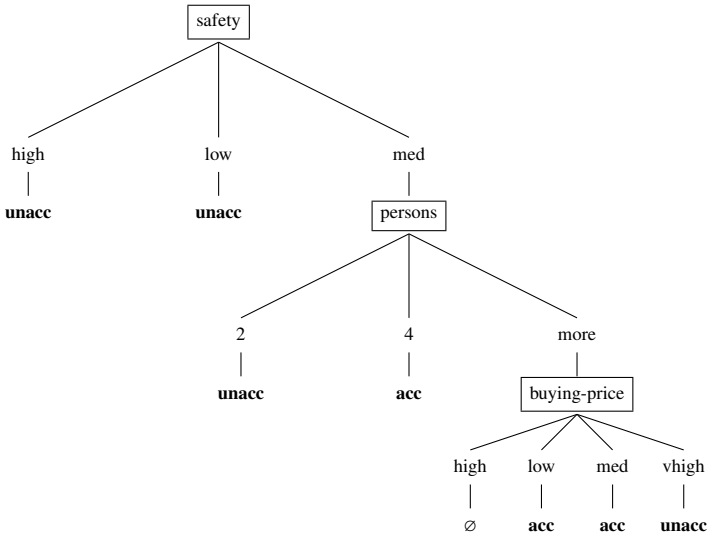
We consider pruning the tree at the following (bolded) node:



We determine the accuracy of the tree as is. The test instances are classified in the following manner:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*unacc*

Thus the accuracy of the tree over the training data with the node in question left alone is 81.818 %. We next consider the pruned tree. The node would be replaced with the majority class over training instances applicable for the node or *acc* in this case.

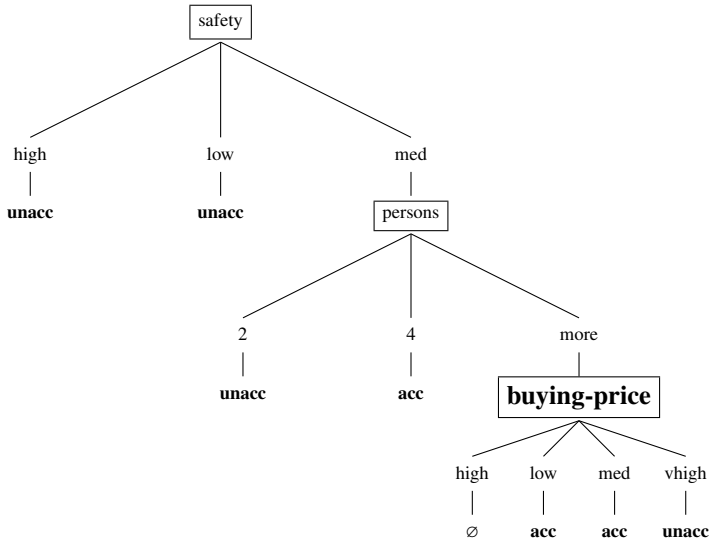


The test instances are classified in the following manner:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*acc*

Thus if the node were to be replaced with the leaf with the decision *acc* then the accuracy of the resulting tree would be 90.909 %. Since the new accuracy is at least as good as the old, we replace the node in question with the decision *acc* and proceed to the next prune candidate.

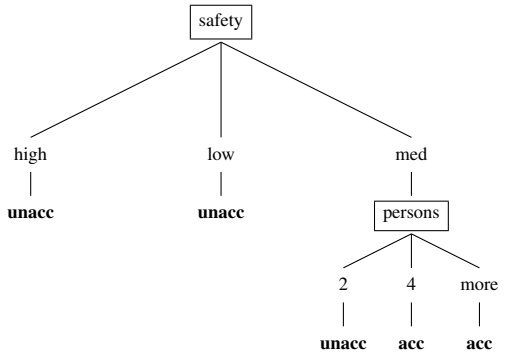
We consider pruning the tree at the following (bolded) node:



We determine the accuracy of the tree as is. The test instances are classified in the following manner:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*acc*

Thus the accuracy of the tree over the training data with the node in question left alone is 90.909 %. We next consider the pruned tree. The node would be replaced with the majority class over training instances applicable for the node or *acc* in this case.

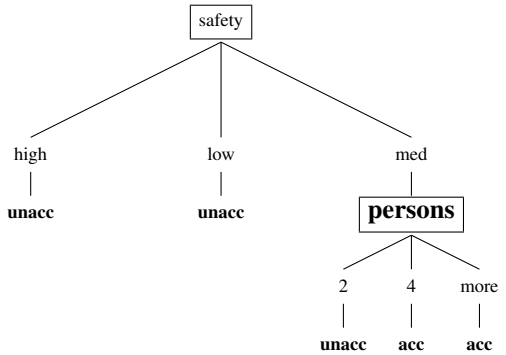


The test instances are classified in the following manner:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*acc*

Thus if the node were to be replaced with the leaf with the decision *acc* then the accuracy of the resulting tree would be 90.909 %. Since the new accuracy is at least as good as the old, we replace the node in question with the decision *acc* and proceed to the next prune candidate.

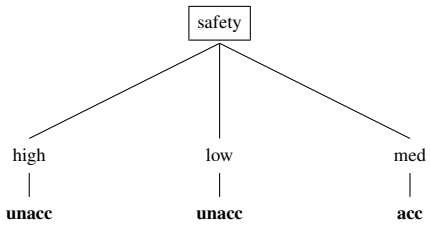
We consider pruning the tree at the following (bolded) node:



We determine the accuracy of the tree as is. The test instances are classified in the following manner:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*acc*

Thus the accuracy of the tree over the training data with the node in question left alone is 90.909 %. We next consider the pruned tree. The node would be replaced with the majority class over training instances applicable for the node or *acc* in this case.

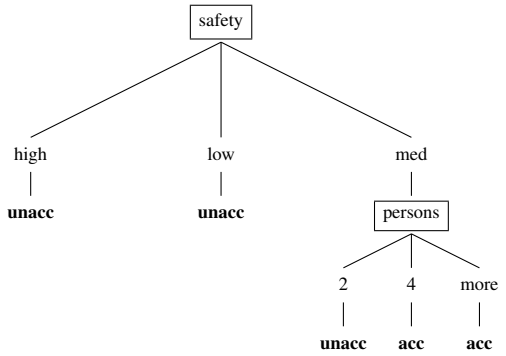


The test instances are classified in the following manner:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*acc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*acc*

Thus if the node were to be replaced with the leaf with the decision *acc* then the accuracy of the resulting tree would be 81.818 %. The original accuracy was better than the one with the pruned tree so we leave this node as is and proceed to the next prune candidate.

We have checked all the nodes for subtree replacement. We thus have a new (and improved tree):



Now we can use this tree to make predictions. We have the following results:

buying-price	maintenance	persons	safety	recommendation	prediction
med	*low*	*4*	*med*	*acc*	*acc*
high	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*4*	*high*	*good*	*unacc*
med	*vhigh*	*more*	*med*	*acc*	*acc*
low	*high*	*4*	*low*	*unacc*	*unacc*
med	*med*	*2*	*low*	*unacc*	*unacc*
med	*low*	*2*	*low*	*unacc*	*unacc*
med	*med*	*4*	*low*	*unacc*	*unacc*
med	*vhigh*	*2*	*med*	*unacc*	*unacc*
low	*high*	*2*	*low*	*unacc*	*unacc*
vhigh	*low*	*4*	*med*	*acc*	*acc*

We see that the prediction was correct 10 out of 11 times and hence our accuracy is **90.91 %**.
Also we have the following confusion matrix (actual values on the left and predicted on the top):

	acc	**good**	**unacc**
acc	3	0	0
good	0	0	1
unacc	0	0	7