

The picture above shows the current decision tree, the leaves are the instances classified by the node values. I will keep constructing the decision tree step by step, and use the same scheme to show the decision tree at each step.

Step 1. Choose a node when “visibility = yes”:

Instance	Stability	Error	Wind	Visibility	Class
2	xstab	LX	head	yes	noauto
5	stab	SS	head	yes	auto
6	xstab	SS	tail	yes	noauto
7	stab	MM	head	yes	noauto
10	xstab	MM	head	yes	noauto
11	stab	MM	tail	yes	auto
12	stab	MM	tail	yes	noauto
13	stab	MM	tail	yes	noauto
14	stab	LX	head	yes	noauto
15	stab	SS	head	yes	auto
16	stab	MM	head	yes	noauto
18	stab	MM	tail	yes	auto
20	xstab	SS	head	yes	noauto
21	xstab	LX	tail	yes	noauto
22	xstab	LX	tail	yes	noauto

The Entropy of the dataset in this step is:

$$Entropy = \underbrace{-\frac{11}{15} * \log_2 \frac{11}{15}}_{noauto} - \underbrace{\frac{4}{15} * \log_2 \frac{4}{15}}_{auto} = 0.837$$

$$Entropy(Stability) = \underbrace{\frac{6}{15} * \left[-\left(\frac{6}{6}\right) \log_2 \left(\frac{6}{6}\right) \right]}_{Stability=xstab} + \underbrace{\frac{9}{15} * \left[-\frac{5}{9} \log_2 \left(\frac{5}{9}\right) - \frac{4}{9} * \log_2 \left(\frac{4}{9}\right) \right]}_{Stability=stab}$$

$$= 0.595$$

$$Gain(Stability) = Entropy - Entropy(Stability) = 0.242$$

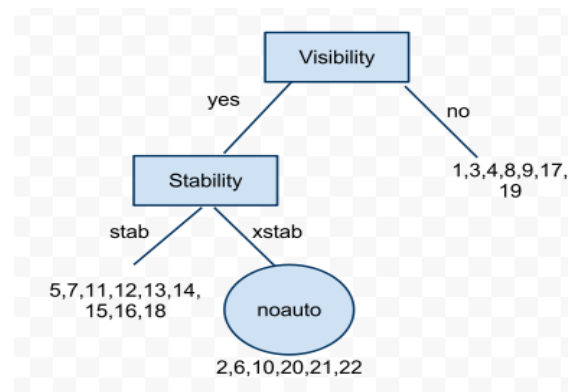
$$\begin{aligned}
 Entropy(Error) &= \frac{4}{15} * \left[-\underbrace{\left(\frac{4}{4}\right) \log_2 \left(\frac{4}{4}\right)}_{\substack{\text{noauto} \\ \text{Error=LX}}} \right] + \frac{4}{15} * \left[-\underbrace{\frac{2}{4} \log_2 \left(\frac{2}{4}\right)}_{\text{noauto}} - \underbrace{\frac{2}{4} * \log_2 \left(\frac{2}{4}\right)}_{\text{auto}} \right] \\
 &\quad + \frac{7}{15} * \left[-\underbrace{\left(\frac{5}{7}\right) \log_2 \left(\frac{5}{7}\right)}_{\text{noauto}} - \underbrace{\left(\frac{2}{7}\right) \log_2 \left(\frac{2}{7}\right)}_{\text{auto}} \right] \\
 &= 0.669
 \end{aligned}$$

$$Gain(Error) = Entropy - Entropy(Error) = 0.168$$

$$\begin{aligned}
 Entropy(Wind) &= \frac{8}{15} * \left[-\underbrace{\left(\frac{6}{8}\right) \log_2 \left(\frac{6}{8}\right)}_{\text{noauto}} - \underbrace{\left(\frac{2}{8}\right) \log_2 \left(\frac{2}{8}\right)}_{\text{auto}} \right] + \frac{7}{15} * \left[-\underbrace{\frac{5}{7} \log_2 \left(\frac{5}{7}\right)}_{\text{noauto}} - \underbrace{\frac{2}{7} * \log_2 \left(\frac{2}{7}\right)}_{\text{auto}} \right] \\
 &= 0.835
 \end{aligned}$$

$$Gain(Wind) = Entropy - Entropy(Wind) = 0.002$$

Thus, "Stability" is selected to subdivide "Visibility=yes". Following is the current decision tree; a circle node means that all instances in this branch belong to one class.



Step 2. Choose next node when "Visibility=yes && Stability=stab"

Instance	Stability	Error	Wind	Visibility	Class
5	stab	SS	head	yes	auto
7	stab	MM	head	yes	noauto
11	stab	MM	tail	yes	auto
12	stab	MM	tail	yes	noauto
13	stab	MM	tail	yes	noauto
14	stab	LX	head	yes	noauto
15	stab	SS	head	yes	auto
16	stab	MM	head	yes	noauto
18	stab	MM	tail	yes	auto

The Entropy of the dataset in this step is:

$$Entropy = \underbrace{-\frac{5}{9} * \log_2 \frac{5}{9}}_{noauto} - \underbrace{\frac{4}{9} * \log_2 \frac{4}{9}}_{auto} = 0.991$$

$$Entropy(Error) = \frac{2}{9} * \left[\underbrace{-\left(\frac{2}{2}\right) \log_2 \left(\frac{2}{2}\right)}_{Error=SS} \right] + \frac{1}{9} * \left[\underbrace{-\left(\frac{1}{1}\right) \log_2 \left(\frac{1}{1}\right)}_{Error=LX} \right] + \frac{6}{9} * \left[\underbrace{-\left(\frac{4}{6}\right) \log_2 \left(\frac{4}{6}\right)}_{noauto} - \underbrace{\left(\frac{2}{6}\right) \log_2 \left(\frac{2}{6}\right)}_{auto} \right]_{Error=MM}$$

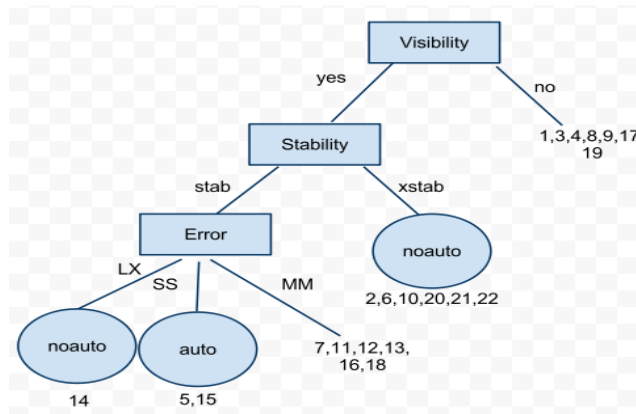
$$= 0.612$$

$$Gain(Error) = Entropy - Entropy(Error) = 0.379$$

$$Entropy(Wind) = \frac{5}{9} * \left[\underbrace{-\frac{3}{5} \log_2 \left(\frac{3}{5}\right)}_{noauto} - \underbrace{\frac{2}{5} \log_2 \left(\frac{2}{5}\right)}_{auto} \right]_{Wind=head} + \frac{4}{9} * \left[\underbrace{-\left(\frac{2}{4}\right) \log_2 \left(\frac{2}{4}\right)}_{noauto} - \underbrace{\left(\frac{2}{4}\right) \log_2 \left(\frac{2}{4}\right)}_{auto} \right]_{Wind=tail} = 0.984$$

$$Gain(Wind) = Entropy - Entropy(Wind) = 0.007$$

So “Error” is selected to subdivide “Visibility=yes && Error=SS”, and yields the current decision tree;



Step 3. Choose next node when “Visibility=yes && Stability=stab && Error=MM”

Instance	Stability	Error	Wind	Visibility	Class
7	stab	MM	head	yes	noauto
11	stab	MM	tail	yes	auto
12	stab	MM	tail	yes	noauto
13	stab	MM	tail	yes	noauto
16	stab	MM	head	yes	noauto
18	stab	MM	tail	yes	auto

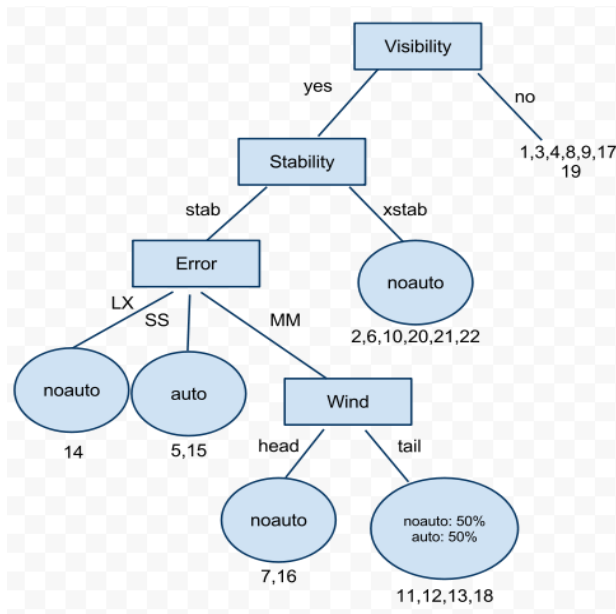
$$Entropy = \underbrace{-\frac{4}{6} * \log_2 \frac{4}{6}}_{noauto} - \underbrace{\frac{2}{6} * \log_2 \frac{2}{6}}_{auto} = 0.918$$

$$Entropy(Wind) = \frac{2}{6} * \left[-\left(\frac{2}{2}\right) \log_2 \left(\frac{2}{2}\right) \right] + \frac{4}{6} * \left[-\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right) \right] = 0.667$$

$\underbrace{\hspace{10em}}_{Wind=head}$
 $\underbrace{\hspace{10em}}_{Wind=tail}$

$$Gain(Wind) = Entropy - Entropy(Wind) = 0.251$$

There is only one attribute left (“Wind”), so it is selected to subdivide the branch at this step. It can be also observed that two of four instances in this branch belong to class “noauto”, while the other two instances belong to class “auto”, so I choose to output that the instances in this branch have 50% chance of being “noauto” and 50% chance of being “auto”. Following is the current decision tree:



Step 4. Choose next node when “Visibility=no”

Instance	Stability	Error	Wind	Visibility	Class
1	xstab	MM	tail	no	auto
3	stab	LX	head	no	auto
4	xstab	SS	head	no	auto
8	xstab	MM	head	no	auto
9	xstab	SS	head	no	auto
17	xstab	MM	tail	no	auto
19	xstab	SS	tail	no	auto

Fortunately, all the instances in this branch belong to class “auto”, so the tree do not need to split any more. The final decision tree is:

