

Questions-1:

i. How many instances are there in the data file?

There are 800 instances in the data file

Current relation	
Relation: Mushroom-data-23	Attributes: 17
Instances: 800	Sum of weights: 800

ii. How many different values can ring-type take on?

Ring type can take on 4 different values

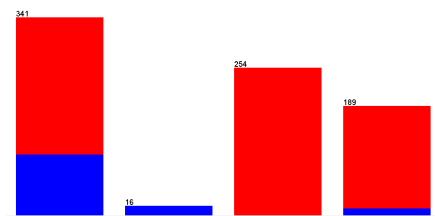
Selected attribute			
Name: ring-type		Type: Nominal	
Missing: 0 (0%)		Distinct: 4	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	e	341	341
2	f	16	16
3	l	254	254
4	p	189	189

iii. Which values of the ring-type attribute result in only edible mushrooms?

Ring-type F results in ONLY edible mushrooms

Selected attribute			
Name: ring-type		Type: Nominal	
Missing: 0 (0%)		Distinct: 4	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	e	341	341
2	f	16	16
3	l	254	254
4	p	189	189

Class: iv2status (Nom) Visualize All



Questions-2:

i. Use a text editor (such as TextEdit) to view the ARFF file representing the mush-room data. (This MUST be an editor that does not format the data — editors such as Word will not work.)

What does the first line tell you?

The first line is this “@RELATION Mushroom-data-23” which tells you the name of the dataset

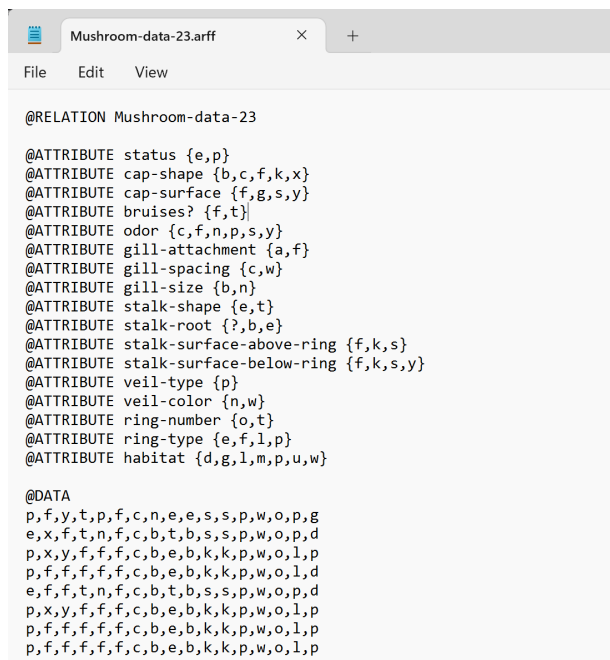
ii. What do the next 17 lines tell you?

The next 17 lines are the attributes and the values they can take on. For example, the “status” attribute is this “@ATTRIBUTE status {e,p}”

iii. What do the rest of the lines tell you?

The rest of the lines are the rows / data points. For example the first row is the mushroom with the values “p,f,y,t,p,f,c,n,e,e,s,s,p,w,o,p,g”

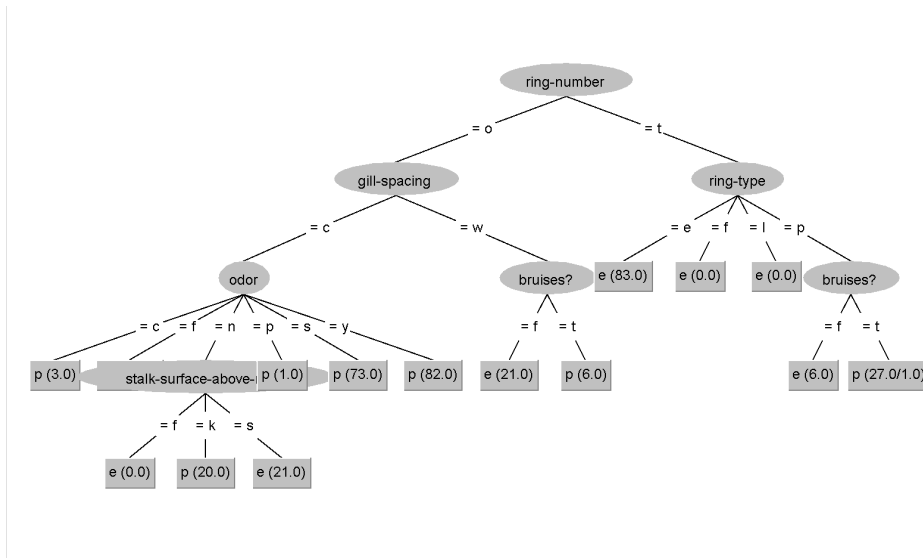
(I used Notepad because I own a windows machine I hope this is acceptable)



```
@RELATION Mushroom-data-23

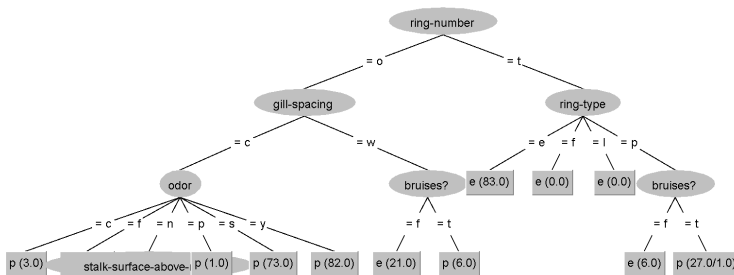
@ATTRIBUTE status {e,p}
@ATTRIBUTE cap-shape {b,c,f,k,x}
@ATTRIBUTE cap-surface {f,g,s,y}
@ATTRIBUTE bruises? {f,t}
@ATTRIBUTE odor {c,f,n,p,s,y}
@ATTRIBUTE gill-attachment {a,f}
@ATTRIBUTE gill-spacing {c,w}
@ATTRIBUTE gill-size {b,n}
@ATTRIBUTE stalk-shape {e,t}
@ATTRIBUTE stalk-root {?,b,e}
@ATTRIBUTE stalk-surface-above-ring {f,k,s}
@ATTRIBUTE stalk-surface-below-ring {f,k,s,y}
@ATTRIBUTE veil-type {p}
@ATTRIBUTE veil-color {n,w}
@ATTRIBUTE ring-number {o,t}
@ATTRIBUTE ring-type {e,f,l,p}
@ATTRIBUTE habitat {d,g,l,m,p,u,w}

@DATA
p,f,y,t,p,f,c,n,e,e,s,s,p,w,o,p,g
e,x,f,t,n,f,c,b,t,b,s,s,p,w,o,p,d
p,x,y,f,f,f,c,b,e,b,k,k,p,w,o,l,p
p,f,f,f,f,f,c,b,e,b,k,k,p,w,o,l,d
e,f,f,t,n,f,c,b,t,b,s,s,p,w,o,p,d
p,x,y,f,f,f,c,b,e,b,k,k,p,w,o,l,p
p,f,f,f,f,f,c,b,e,b,k,k,p,w,o,l,p
p,f,f,f,f,f,c,b,e,b,k,k,p,w,o,l,p
p,f,f,f,f,f,c,b,e,b,k,k,p,w,o,l,p
```



Questions-3:

i. Draw the first three levels of the decision tree that was developed.

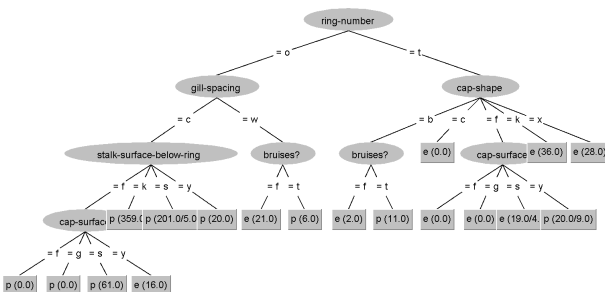


ii. Go back to the results screen and scroll through it to answer the following two questions. How many instances were classified correctly? How many incorrectly?

There were 799 correctly classified answers (99.875%)

There was 1 incorrectly classified answer (0.125%)

Questions-4:



i. What attribute is at the root of the decision tree?

Status

ii. Consider the following path in the decision tree:ring-number=o,gill-spacing=w,bruises=t. What is the class value assigned to instances that follow this path?

p (6.0)

iii. How many paths in the decision tree lead to a leaf node where some instances are classified incorrectly?

13 paths because there are 13 paths without 0.0

iv. Go back to the results screen to answer the following questions. How many instances were classified correctly? How many instances were classified incorrectly?

There were 782 correctly classified instances

There were 18 incorrectly classified instances

Correctly Classified Instances	782	97.75	%
Incorrectly Classified Instances	18	2.25	%

```
=== Confusion Matrix ===
      a    b   <-- classified as
118  14 |    a = e
  4 664 |    b = p
```

Questions-5:

i. How many instances are incorrectly classified?

There is 18 incorrectly classified instances

ii. What does the diagonal of the confusion matrix tell you?

The amount of false positives and false negatives. If you add these two together, you get the total amount of incorrectly classified instances.

iii. Which class (p or e) did the classifier get wrong most often?

It got edibles wrong most of the time, because there were 14 false positives while there was only 4 false negatives.