

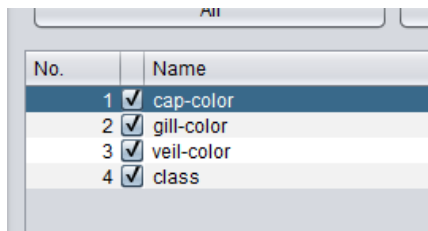
1. What color mushrooms should you *avoid* eating?

The class distribution tells us that we have 3916 poisonous mushrooms out of 8214.

Now we have been asked to associate poisonous mushrooms with the color of the mushrooms.

We can do this by associating all the color related attributes of the mushrooms with class=p

I selected the following attributes to generate association rules-



| No. | Name |
|-----|--|
| 1 | <input checked="" type="checkbox"/> cap-color |
| 2 | <input checked="" type="checkbox"/> gill-color |
| 3 | <input checked="" type="checkbox"/> veil-color |
| 4 | <input checked="" type="checkbox"/> class |

Now when I run the rule miner using apriori algorithm, I get the following list of rules-

1. cap-color=g 1840 ==> veil-color=w 1840 <conf:(1)> lift:(1.03) lev:(0.01) [45] conv:(45.3)
2. gill-color=b 1728 ==> veil-color=w 1728 <conf:(1)> lift:(1.03) lev:(0.01) [42] conv:(42.54)
3. gill-color=b 1728 ==> class=p 1728 <conf:(1)> lift:(2.07) lev:(0.11) [895] conv:(895.05)
4. gill-color=b class=p 1728 ==> veil-color=w 1728 <conf:(1)> lift:(1.03) lev:(0.01) [42] conv:(42.54)
5. gill-color=b veil-color=w 1728 ==> class=p 1728 <conf:(1)> lift:(2.07) lev:(0.11) [895] conv:(895.05)
6. gill-color=b 1728 ==> veil-color=w class=p 1728 <conf:(1)> lift:(2.08) lev:(0.11) [896] conv:(896.76)
7. cap-color=e 1500 ==> veil-color=w 1500 <conf:(1)> lift:(1.03) lev:(0) [36] conv:(36.93)
8. gill-color=p 1492 ==> veil-color=w 1492 <conf:(1)> lift:(1.03) lev:(0) [36] conv:(36.73)
9. class=p 3916 ==> veil-color=w 3908 <conf:(1)> lift:(1.02) lev:(0.01) [88] conv:(10.71)
10. class=e 4208 ==> veil-color=w 4016 <conf:(0.95)> lift:(0.98) lev:(-0.01) [-88] conv:(0.54)

What we are looking for here are rules that would tell us which mushrooms fall under the poisonous category?

We look at the highlighted rules-

Given the number of poisonous mushrooms(which is 3916), 3908 have a veil-color=w. Now this is possible because 7924 mushrooms have veil-color=w, which is more than 90% of the mushrooms.

We need to look at more characteristics. We will look at gill-color. We can see from the rules that there are a total of 1728 mushrooms with gill-color=b and given these mushrooms in the whole data set, all of them are class=p which is poisonous. So, we can be sure that a mushroom with gill-color=b is definitely poisonous.

We can try narrowing down the results by associating veil-color with gill-color.

We find two rules-

class=p 3916 ==> veil-color=w 3908 <conf:(1)> lift:(1.02) lev:(0.01) [88] conv:(10.71) gill-color=b
class=p 1728 ==> veil-color=w 1728 <conf:(1)> lift:(1.03) lev:(0.01) [42] conv:(42.54)

This suggests that all the mushrooms with gill-color=b also have veil-color=w

Therefore, you should avoid eating a mushroom with a gill-color as buff.

2. What are the properties or characteristics of edible mushrooms?

I found the following rules that helped me find some properties of edible mushrooms-

odor=n 3528 ==> class=e 3408 <conf:(0.97)> lift:(1.86) lev:(0.19) [1580] conv:(14.05)
gill-spacing=w 1312 ==> class=e 1200 <conf:(0.91)> lift:(1.77) lev:(0.06) [520] conv:(5.6)
odor=n gill-spacing=w 1120 ==> class=e 1104 <conf:(0.99)> lift:(1.9) lev:(0.06) [523] conv:(31.76)

Therefore we can say that an edible mushroom has no odor, has a crowded gill spacing. These were the best possible properties of edible mushrooms.

3. Are there any interesting observations that relate the odor, color, and/or habitat of a mushroom?

I extracted the following rules by associating color related properties, odor and/or habitat-

1. spore-print-color=h 1632 ==> odor=f 1584 <conf:(0.97)> lift:(3.65) lev:(0.14) [1150] conv:(24.45)

In this rule we can see, if spore-print-color=h then it is very very likely that the odor will be f.

The below rules suggest that if a mushroom has a habitat as d,g or p then the veil-color is always w.

1. habitat=d 3148 ==> veil-color=w 3148 <conf:(1)> lift:(1.03) lev:(0.01) [77] conv:(77.5)
2. habitat=g 2148 ==> veil-color=w 2148 <conf:(1)> lift:(1.03) lev:(0.01) [52] conv:(52.88)
3. habitat=p 1144 ==> veil-color=w 1144 <conf:(1)> lift:(1.03) lev:(0) [28] conv:(28.16)

The below rule suggests that if a mushroom has a gill-color=b then it will definitely have a spore-print-color as w-

4. gill-color=b 1728 ==> spore-print-color=w 1728 <conf:(1)> lift:(3.4) lev:(0.15) [1220] conv:(1220.06)

Also, I tried associating habitat with order, but due to our min limit on confidence level which is set to 0.9, there were no rules associating habitat with order found.

Describe how you used the association rule builder, why you chose the builder you did and why the rule is interesting.

I used weka for associative rule mining, because weka is a GUI application that provides the feasibility to go through a large dataset, with a click of a button. This could have been done through R or python but all the complexities are abstracted when the need is to do some kind of data analysis on a data set. I used apriori algorithm for all the questions because it is simple to understand and gives robust results.