JONATHAN PETERSEN
A01236750
CS5800 - DATABASE SYSTEMS

Associative Data Mining

On the Mushroom Data Set

1. What color mushrooms should you avoid eating?

Using a preprocess step, I removed all attributes that did not have a color associated with them, alongside veil-color since almost every mushroom has a white veil-color. I then set up an Apriori rule builder with mostly default values, since they seemed apropriate for the nominal rules I was searching for. The only default I changed was the car attibute in WAKA, telling the builder to search only for rules which led to a specific class value. Finally, I filtered the search to only show rules leading to poisonous mushrooms.

I generated 19 rules based on color, and gill-color b was associated with poison at 100% confidence, alongside spore-print-color h at 97% confidence. Finally, spore-print-color w and cap-colors e and n also showed at least a 90% confidence rating. These rules are interesting because it showes that that gill-color is the best indicator of poison, followed by spore-print-color and then cap-color.

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

To locate interesting rules for edible mushrooms, I first filtered out the bruising and veil-color in a preprocess step. I then ran a Select Attributes search in WAKA to locate attributes that were closely correlated to class. A genetic search with consistency subset revealed that odor, stalk-surface-above-ring, ring-number, ring-type, and spore-print-color were all closely related to edibility.

I ran a default Apriori rule builder with the same settings as in question 1, only this time searching for edible mushrooms instead of poisonous ones. There were several rules that showed 100% confidence, including odor=n gill-size=b ring-number=o, odor=n stalk-shape=t, and odor=n gill-attachment=f stalk-shape=t. The first 37 rules all had odor = n, so a lack of odor seems like a good indicator of edibility. The most significant rule without odor as a component was gill-size=b stalk-surface-above-ring=s stalk-surface-below-ring=s ring-number=o, which seems too specific to be conclusive.

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

For this search, I filtered all of the attributes down to cap-color, odor, gill-color, stalk-color-above-ring, stalk-color-below-ring, spore-print-color, and habitat. I then ran another Select Attributes test with the same settings as in quesiton 2. This was less helpful, since it seemed like

all of the attributes were related. So I simply ran the Apriori rule builder with all-defaults and no filters, and it showed 24 rules.

gill-color=b 1728 ==> spore-print-color=w, stalk-color-below-ring=w habitat=g 1716 ==> stalk-color-above-ring=w, and stalk-color-above-ring=w habitat=g 1716 ==> stalk-color-below-ring=w were the top three rules, all at 100% confidence. The only rules without 100% confidence were spore-print-color=h 1632 ==> odor=f 1584 conf:(0.97), cap-color=e spore-print-color=w 924 ==> gill-color=b 864 conf:(0.94), and spore-print-color=k habitat=d 960 ==> odor=n 864 conf:(0.9).

In summary of all the rules, it would appear that gill color and spore print color are related, stalk color and habitat are related, odor and habitat are related, cap color and stalk color are related, and overall it would seem that these attributes are well-related. This suggests that color, habitat, and odor are all closely related.