CIS 635 Data Mining Name \_\_Dhyansah Cakir\_

Homework 4

Hand in sheet

Write in or copy and paste the answers for the following:

### Part 1

1. gini (nominal)
   1. gender \_\_\_0.16\_\_ \_\_\_0.24\_\_
   2. ageGroup \_\_\_0.13\_\_ \_0.50\_\_\_\_ \_\_\_0\_\_
2. gini (numeric)
   1. pulse \_\_0.40\_\_\_ \_\_0.30\_\_\_ \_\_0.175\_\_\_ \_\_0.875\_\_ \_\_\_0.40\_\_ \_\_\_0.45\_\_ 0.475 \_\_\_0.45\_\_ \_\_0.45\_\_\_
3. which attribute \_\_\_gender: 0.40 age Group: 0.24\_\_\_\_\_\_pulse: 0.175

so based on these scores the best attribute is pulse

1. next step \_\_I think the next step is to determine if there needs to be another split. If yes then determine which attribute would be the best for the next split. If its no then keep the tree where it is.

### Part 2

1. class a \_\_no\_\_ b \_\_no\_\_ c \_\_no\_\_\_ d \_\_yes\_\_\_ e \_yes\_\_\_\_
2. \_\_\_patients with the virus are males in the age groups 2 or 3\_\_\_\_\_\_\_\_\_\_\_\_
3. The best way to avoid it is if you are in age group 3 have a pulse greater than 60. If you are in age group 3 have a pulse less than or equal to 60

### Part 3

1. e) hw04a \_\_\_\_\_\_\_\_none\_\_\_\_\_\_\_\_\_\_\_  
     
    hw04b \_\_\_\_\_\_\_\_\_none \_\_\_\_\_\_\_\_\_\_  
     
    hw04c \_\_\_\_\_\_couldn’t figure out what was going on, but after further investigation this data set has only 1 person with the virus. So there can’t really be a good decision tree made from it. At all. This data set should not be used. \_\_\_\_\_\_\_\_\_\_\_\_
2. \_\_\_\_\_\_\_\_\_I would use B because from the first branch off of pulse you can see that there is a 97% chance that they won’t have the virus. \_\_\_Then from the 3rd branch off of pulse then age group and pulse again you can see that there is a \_96% chance that they wont have the virus also. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. \_\_\_\_\_\_\_\_Because only one patient have the virus, therefore the algorithm cannot really make a tree, there are no positive cases to train from. There needs to be variability in the data set in order to get a better understanding of the trend.