Fundamental of Data Mining
Due 10/30/2017
Assignment #3

Redo of Problem #3 ONLY

John Warlop

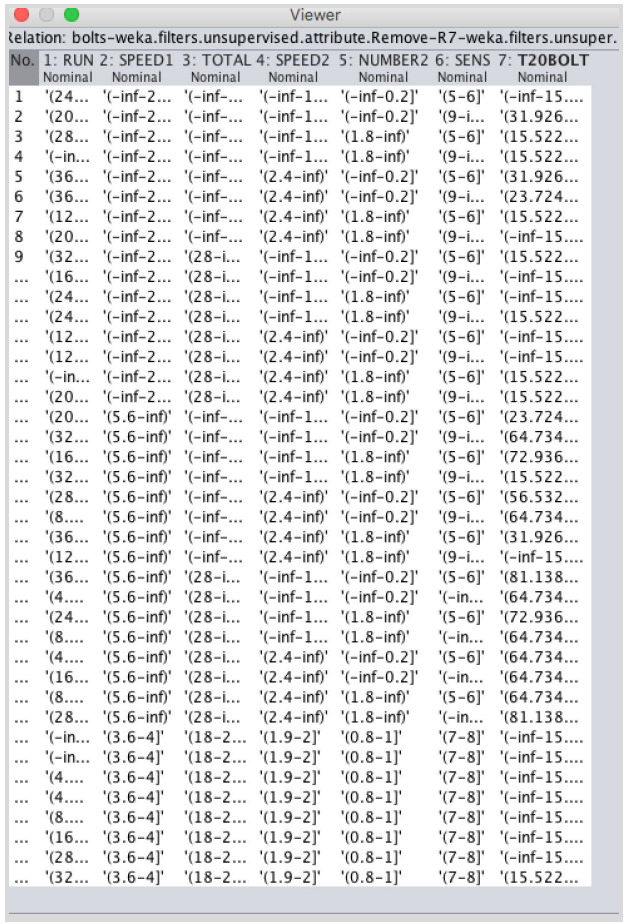
Denotes grading comment

1	NOTE DONE, REDO OF PROBLEM #3 ONLY q2: (-.5) What did you learn about the data? One thing to note is that the poor performance could be due to the small dataset (only 14 instances)
2	NOTE DONE, REDO OF PROBLEM #3 ONLY (-.5)What does it mean you could negate humidity. YOu could improve the model by modifying the J48 parameters. (for example use the unprunded tree)

3

Use the J48 Decision tree learning scheme to analyze the bolts data (bolts.arff without the TIME attribute). The dataset describes the time needed by a machine to produce and count 20 bolts. (More details can be found in the file containing the dataset, you can open the file using a file editor to read the comments)

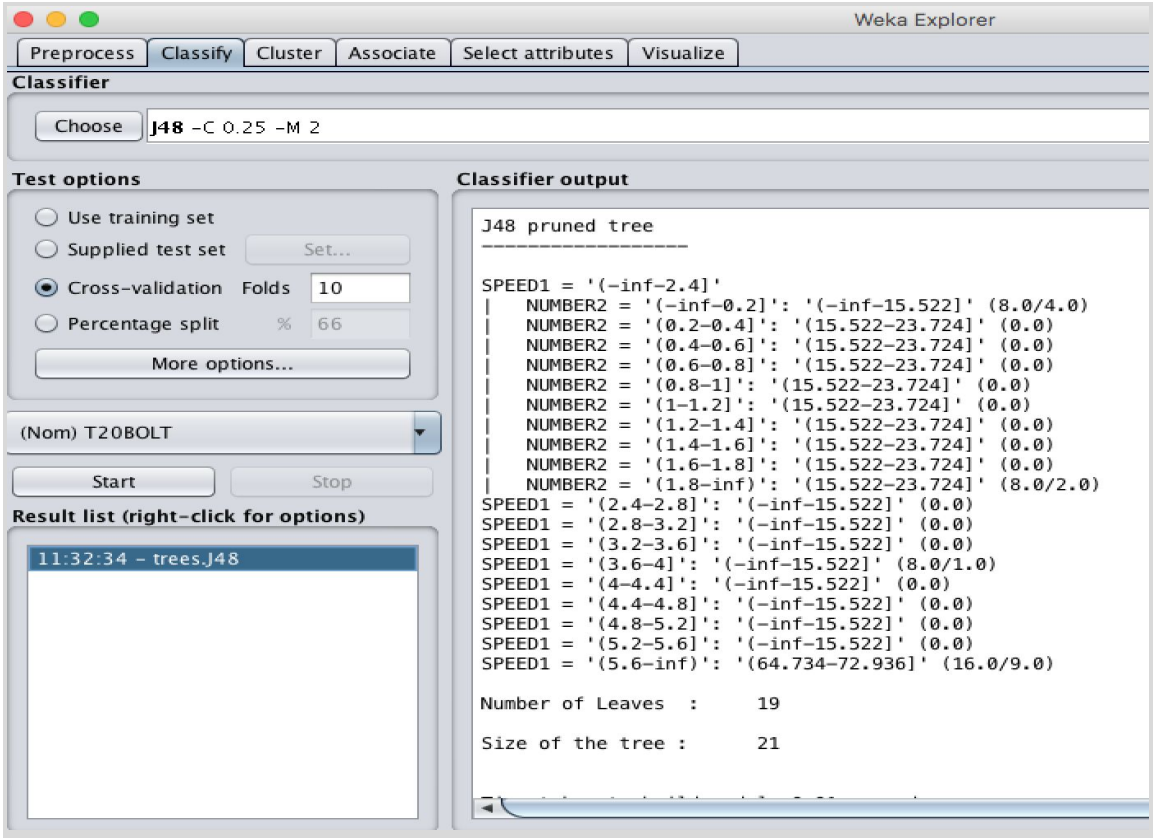
q3: (-2) Go back an revisit for full credit. The class value needs to be a nominal value to run J48. By default weka does not dicretize the class attribute. In order for the filter to be applied you need to toggle the paramter ignoreclass.



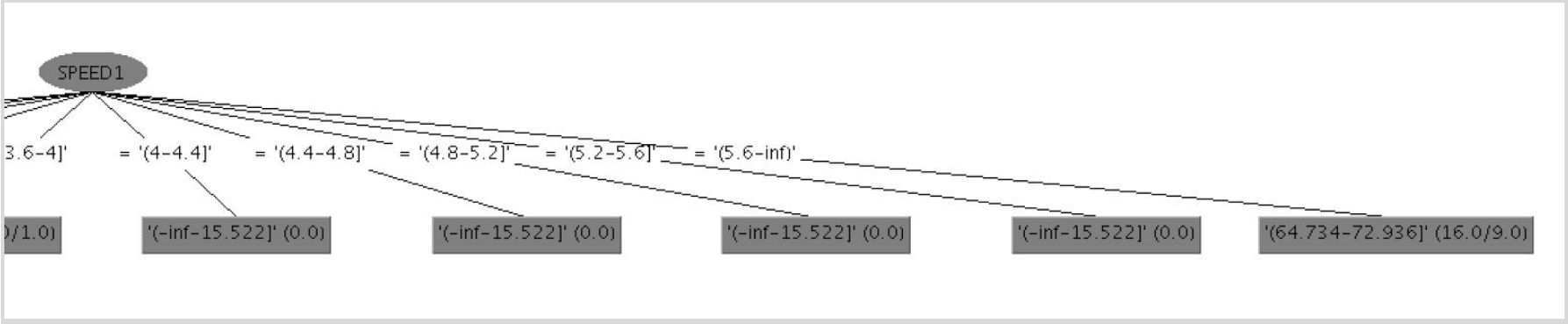

```
% SUMMARY:
%
% Data from an experiment on the affects of machine adjustments on
% the time to count bolts.  Data appear as the STATS (Issue 10) Challenge.
%
% DATA:
A manufacturer of automotive accessories provides hardware, e.g. nuts, bolts, washers and
screws, to fasten the accessory to the car or truck. Hardware is counted and packaged
automatically.  Specifically, bolts are dumped into a large metal dish.  A plate that forms
the bottom of the dish rotates counterclockwise.  This rotation forces bolts to the outside of
the dish and up along a narrow ledge.  Due to the vibration of the dish caused by the spinning
bottom plate, some bolts fall off the ledge and back into the dish.  The ledge spirals up to a
point where the bolts are allowed to drop into a pan on a conveyor belt.  As a bolt drops, it
passes by an electronic eye that counts it.  When the electronic counter reaches the preset
number of bolts, the rotation is stopped and the conveyor belt is moved forward.

There are several adjustments on the machine that affect its operation.  These include; a
speed setting that controls the speed of rotation (SPEED1) of the plate at the bottom of the
dish, a total number of bolts (TOTAL) to be counted, a second speed setting (SPEED2) that is
used to change the speed of rotation (usually slowing it down) for the last few bolts, the
number of bolts to be counted at this second speed (NUMBER2), and the sensitivity of the
electronic eye (SENS).  The sensitivity setting is to insure that the correct number of bolts
are counted.  Too few bolts packaged causes customer complaints.  Too many bolts packaged
increases costs.  For each run conducted in this experiment the correct number of bolts was
counted.  From an engineering standpoint if the correct number of bolts is counted, the
sensitivity should not affect the time to count bolts.  The measured response is the time
(TIME), in seconds, it takes to count the desired number of bolts.  In order to put times on a
equal footing the response to be analyzed is the time to count 20 bolts (T20BOLT).
Below are the data for 40 combinations of settings.  RUN is the order in which the data were
collected.
%
Analyze the data.  What adjustments have the greatest effect on the time to count 20 bolts?
How would you adjust the machine to get the shortest time to count 20 bolts?  Are there any
unusual features to the data?
```

a. Analyze the model produced. What adjustments (if you were to make any) would have the greatest effect on the time to count 20 bolts (attribute: T20Bolt) (i.e. what is the most important/selective attribute/value pair in the tree)?

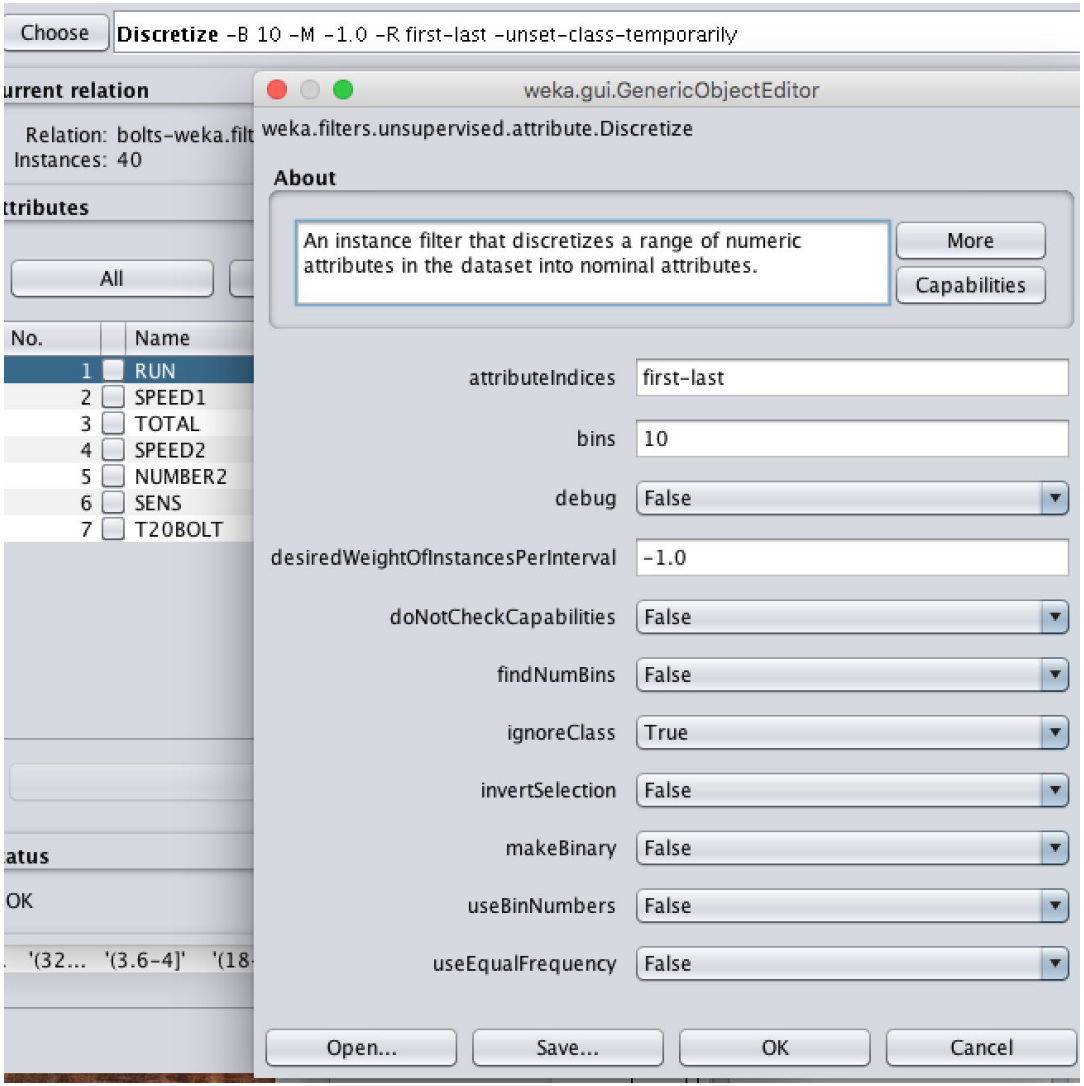


As I look at the decision tree built by J48 I see that adjusting Speed1 to (5.6-inf) and waiting between (64-72) seconds will yield 16 bolts with count of 20. Of course, doing this would take the most time to count 20 bolts.



b. According to the classifier, how would you adjust the machine(the other attributes) to get the shortest time to count 20 bolts?

As I look at the pruned J48 tree I see that the shortest time(T20BOLT -inf-15.5) to count 20 bolts would be to adjust SPEED1 to (3.6-4). On the tree, this gave 8 instances.



Note: When I first did this problem I did not know how to make the T20BOLT attribute be nominal. Once I checked the IgnoreClass option to True, this fixed it so I could use J48.

Appendix - Weka Output