Jason Melnik

09/25/2021

Homework 4

* 1. Mean Percentage:
     1. Rl: 1.52%
     2. Na: 13.41%
     3. Mg: 2.69%
     4. Al: 1.44%
     5. Si: 72.65%
     6. K: 0.50%
     7. Ca: 8.96%
     8. Ba: .18%
     9. Fe: 0.06%
     10. The most common element is Si and the least common element is Fe.
  2. It seems like with the graph below that Na can help predict what kind of glass type it is since “head” type has mostly Na while “bwnf” has the least amount. This can help make a tree that helps decide what the glass type is based of the Na percentage. Chart, box and whisker chart

     Description automatically generated
  3. The graph below is seeing the percent Si in glass type since Si has the highest mean percentage in the data. It seems like Si would not help decide glass type since it is pretty even throughout the glass types with no outliers.

Chart, box and whisker chart

Description automatically generatedThe graph below shows the percent Fe in each glass type and based of the graph it can help with building the tree. Fe seems like it’s rarely in “cont”, “head”, and “table” which can help remove half the types for the tree.Chart, box and whisker chart

Description automatically generated

1. Prediction Accurary
   1. Decision Tree prediction accuracy is 65 Percent.
   2. Random forest prediction accuracy is 81 Percent which is much more successful than the decision tree accuracy.
2. Rerunning the code without setting the seed causes the methods to have different percentages with C50 have a prediction accuracy of 74 percent while the random forest had an accuracy of 69 percent which makes C50 better this time. After running the program 10 times these are the results(C50, RandomTree): (59, 75), (71,73), (75, 71), (59, 77), (65, 81), (71, 72), (69, 71), (69, 81), (65, 76), (67, 75). This shows us that most of the time RandomForest has a higher prediction accuracy with the off chance that C50 has a higher accuracy. It is important to rerun both functions to get a better understanding on which method to use. In this case you would want to use RandomForest since it had the highest accuracy of 81% during two of the runs while C50 had the highest accuracy of 75%.