Jacob Perrone

Assignment 5 Report

Part 1

Model 1 screenshots

Diagram

Description automatically generated

Chart, waterfall chart

Description automatically generated

Graphical user interface, table

Description automatically generated

Model 2

Diagram

Description automatically generated

Chart, waterfall chart

Description automatically generated

Graphical user interface, table

Description automatically generated

Scoring on model 2 for Rapid miner

Table

Description automatically generated

Scoring design

Diagram, schematic

Description automatically generated

Explanation:

First, I loaded the model data into RapidMiner, and then applied the data to the data handling to replace the missing data. Then added the partitioning which is cross validation using 10 folds. Accuracy and Classification matrix were the chosen metrics to display as results. The model 1 processed was saved and copied to make model 2. Model 2 process has a changed cross validation to 5 folds. Model 2 had the better accuracy, so it was used for scoring. To score, I imported the new data and added it to the process.

Part 2

Fully Grown Tree

Diagram

Description automatically generated with low confidence

Best pruned tree

Diagram

Description automatically generated

Min Error tree

Diagram

Description automatically generated

Can the best pruned tree be used for dimension reduction? How many predictors did it reduce?

* Based on the best pruned tree and fully grown tree, both of mine are the same for my best model so no dimensions were reduced. The minimum error tree is also the same as well. The trees are dimension reduced from the original dataset as only 5 of the 13 are used in each of the trees above.

Scoring on new data

Table

Description automatically generated

The scoring predictions came from the predictors color intensity, proline, alcohol, flavonoids, and ash alcalnity.