

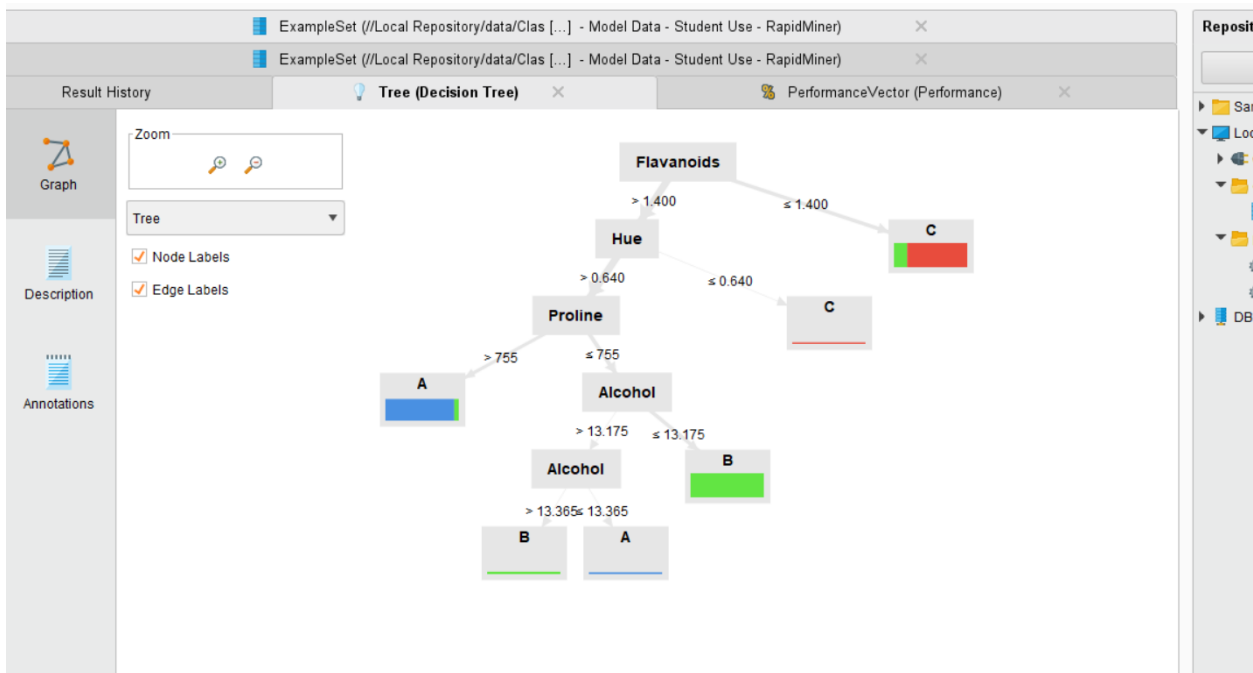
Jacob Perrone

Assignment 5 Report

Part 1

Model 1 screenshots

The screenshot displays the RapidMiner Process Designer interface. The main workspace shows a workflow starting with an input node 'inp' connected to 'Retrieve Class Assi...', followed by 'Replace Missing Val...', 'Set Role', and finally 'Cross Validation'. The 'Cross Validation' node is highlighted with an orange border. To the right, the 'Parameters' panel for 'Cross Validation' is visible, showing settings for 'leave one out' (unchecked), 'number of folds' (10), and 'sampling type' (automatic). Below the parameters, there are links for 'Show advanced parameters' and 'Change compatibility (9.10.000)'. A 'Help' panel is also visible at the bottom right, providing information about 'Cross Validation' and its concurrency.

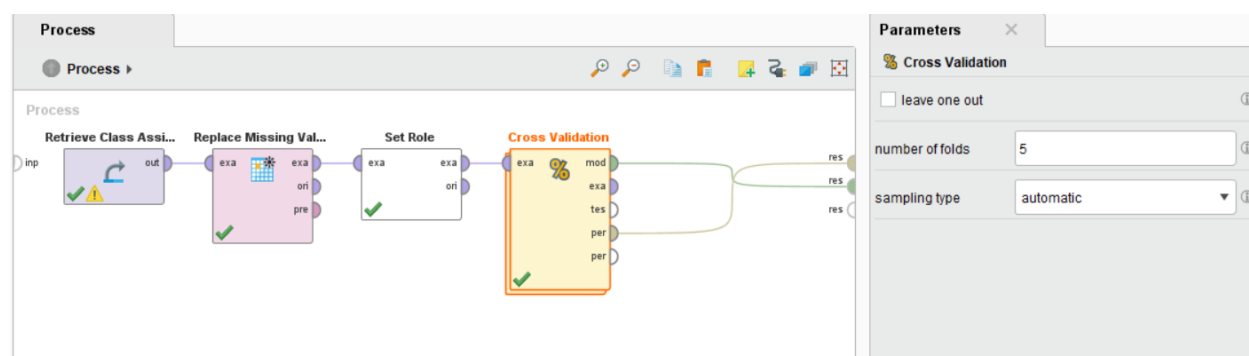


☒ Table View
 ☐ Plot View

accuracy: 92.09% +/- 6.10% (micro average: 92.10%)



	true A	true B	true C	class precision
pred. A	90	8	1	90.91%
pred. B	3	124	2	96.12%
pred. C	2	10	89	88.12%
class recall	94.74%	87.32%	96.74%	

Model 2



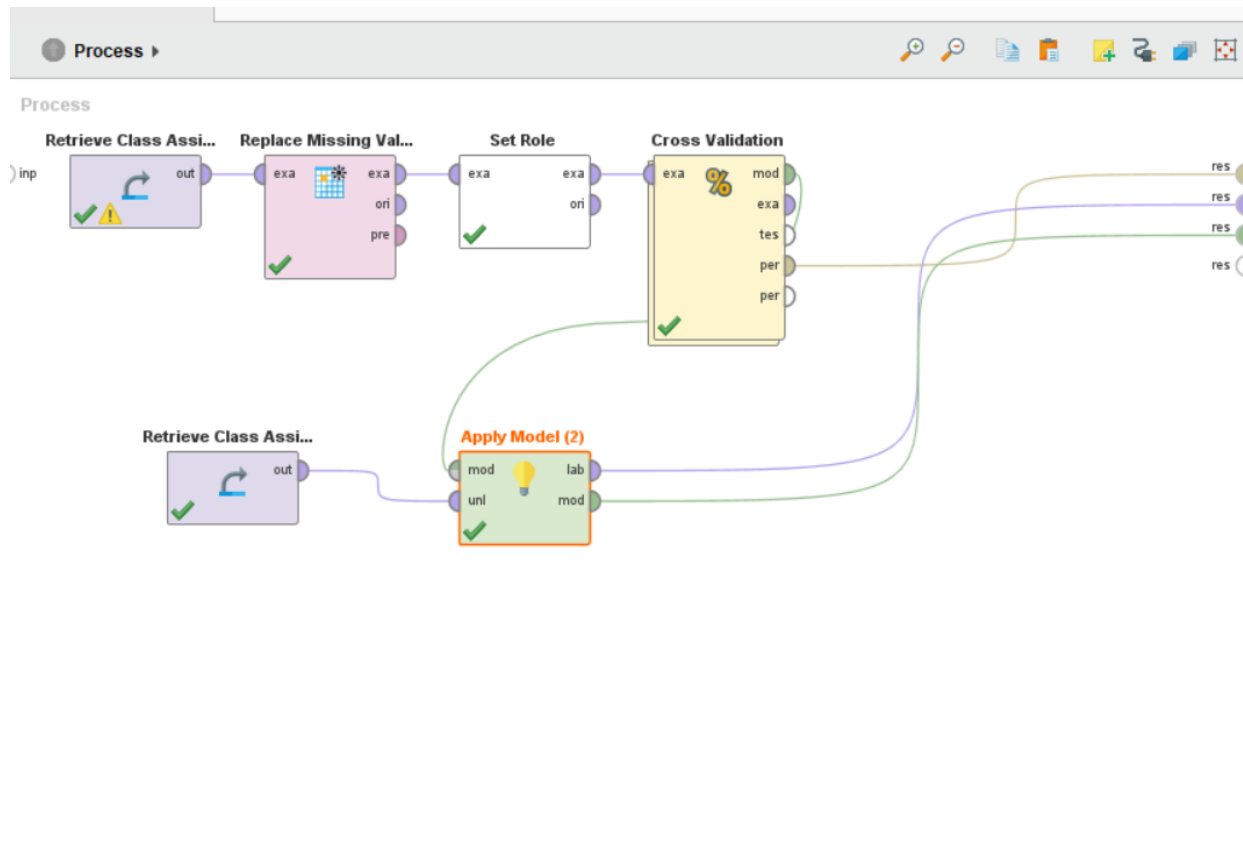


Scoring on model 2 for Rapid miner

Open in  Turbo Prep  Auto Model Filter (14 / 14 examples): all

Row No.	prediction(T...	confidence(A)	confidence(B)	confidence(C)	Alcohol	Malic_Acid	Ash	Ash_Alcalinity	Magnesium	Tota
1	C	0	0.182	0.818	13.730	4.360	2.260	22.500	88	1.28
2	C	0	0.182	0.818	13.450	3.700	2.600	23	111	1.70
3	C	0	0.182	0.818	12.820	3.370	2.300	19.500	88	1.48
4	B	0	1	0	12.520	2.430	2.170	21	88	2.55
5	B	0	1	0	11.760	2.680	2.920	20	103	1.75
6	B	0	1	0	11.410	0.740	2.500	21	88	2.48
7	B	0	1	0	12.080	1.390	2.500	22.500	84	2.56
8	B	0	1	0	11.030	1.510	2.200	21.500	85	2.46
9	B	0	1	0	11.820	1.470	1.990	20.800	86	1.98
10	A	0.939	0.061	0	13.280	1.640	2.840	15.500	110	2.60
11	A	0.939	0.061	0	13.050	1.650	2.550	18	98	2.45
12	A	0.939	0.061	0	13.070	1.500	2.100	15.500	98	2.40
13	A	0.939	0.061	0	14.220	3.990	2.510	13.200	128	3
14	A	0.939	0.061	0	13.560	1.710	2.310	16.200	117	3.15

Scoring design

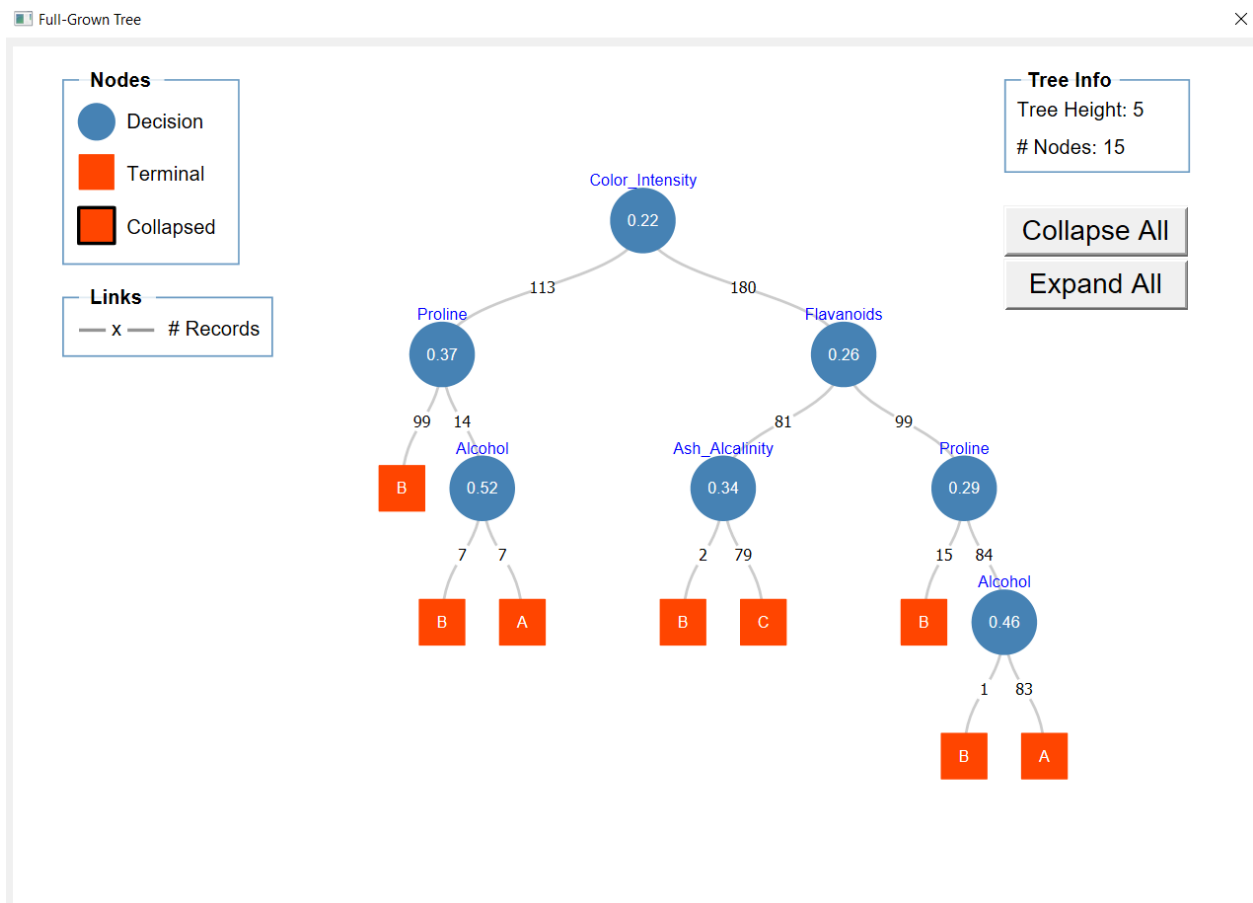


Explanation:

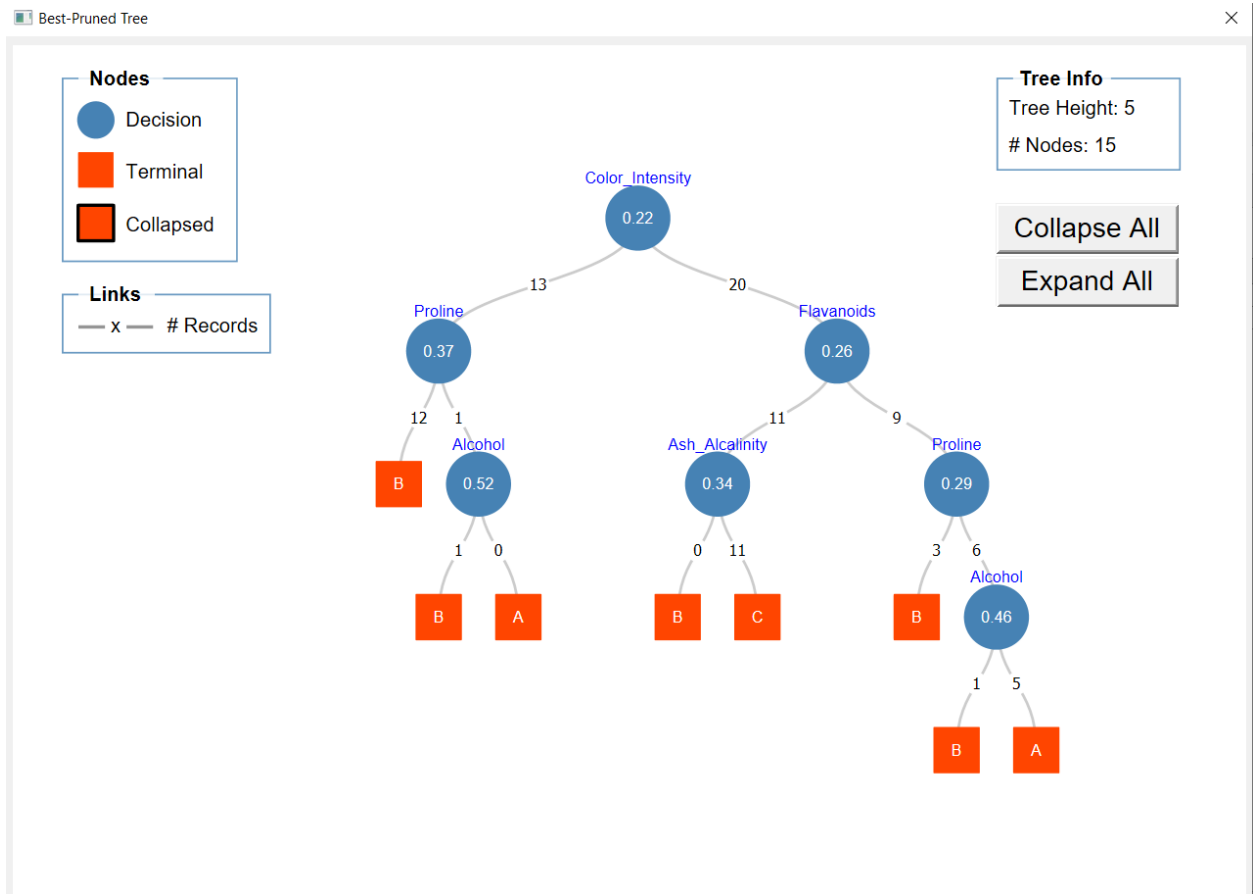
First, I loaded the model data into RapidMiner, and then applied the data to the data handling to replace the missing data. Then added the partitioning which is cross validation using 10 folds. Accuracy and Classification matrix were the chosen metrics to display as results. The model 1 processed was 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

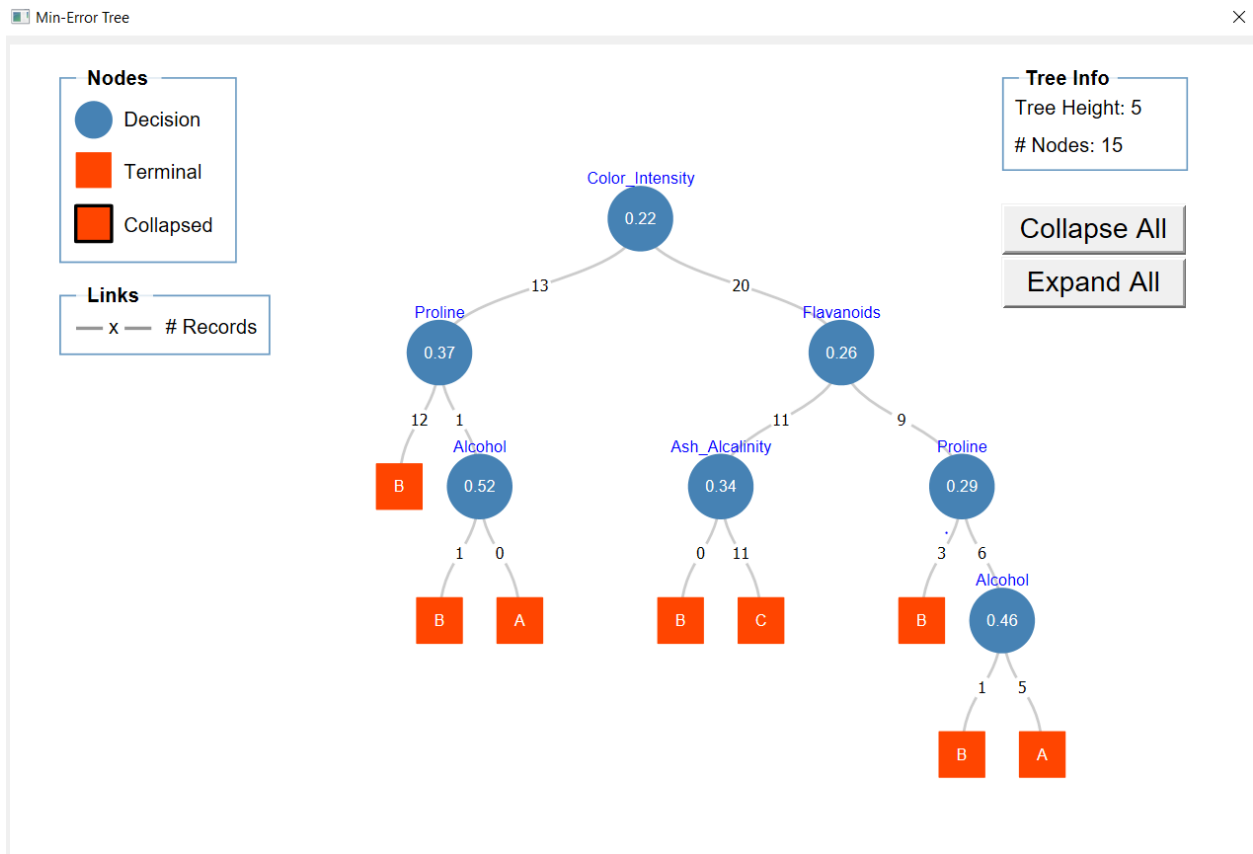
Fully Grown Tree



Best pruned tree



Min Error tree



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

22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	

Scoring

Record ID	Prediction: Type	PostProb: A	PostProb: B	PostProb: C
Record 1	C	0	0	1
Record 2	C	0	0	1
Record 3	C	0	0	1
Record 4	B	0	1	0
Record 5	B	0	1	0
Record 6	B	0	1	0
Record 7	B	0	1	0
Record 8	B	0	1	0
Record 9	B	0	1	0
Record 10	A	1	0	0
Record 11	A	1	0	0
Record 12	A	1	0	0
Record 13	A	1	0	0
Record 14	A	1	0	0

Scoring_DecisionTree

STDPartitionCT_OutputCT_FullTreeCT_BestTreeCT_MinErrorTreeCT_T

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