This document was prepared using R version 4.0.1
Depending on R version, the output may be slightly different, and that's acceptable.

Session 8:

8 Learning Objective

Understand how the cutoff values can be adapted to suit the distribution of the two classes in a binary classification

Assignment 8 (15 marks total)

Steps

1. **Load the auction data ("auction.csv") on Rattle and define variable appropriately (won.auction is the target variable)**
2. **Use the partitioning scheme – 95/5/0**
3. **Run Random Forest model with default options**
4. **Save probability scores on validate data in a an excel/ csv file**
5. **Create a scoring method in Excel where given a cutoff probability it determines the error matrix**
6. **Finally, choose a cutoff probability at which errors are minimized (try cutoff probabilities from 0.1 to 0.9, increasing by 0.1)**

Deliverables 8

Deliverables:

1. **Deliverable part 8_1**: Follow the instruction and provide screenshots of each step and the final results for the RF model and the optimum cutoff value for probability
2. **Deliverable part 8_2**: Briefly explain what you have done in each step and interpret your final result.

Rubric for MOD 8 Peer Review Assignment

For Deliverable 8_1: (5 points)

- 5 points: Screen shot(s) provided for all steps and the final results.

  - 3 points: Screen shot(s) are missing for 1 or 2 steps.
  - 1 point: Screen shot(s) are missing most of the steps.
  - 0 points: No screen shots are provided, or irrelevant screen shots are provided, or no assignment submission.

This document was prepared using R version 4.0.1
Depending on R version, the output may be slightly different, and that's acceptable.

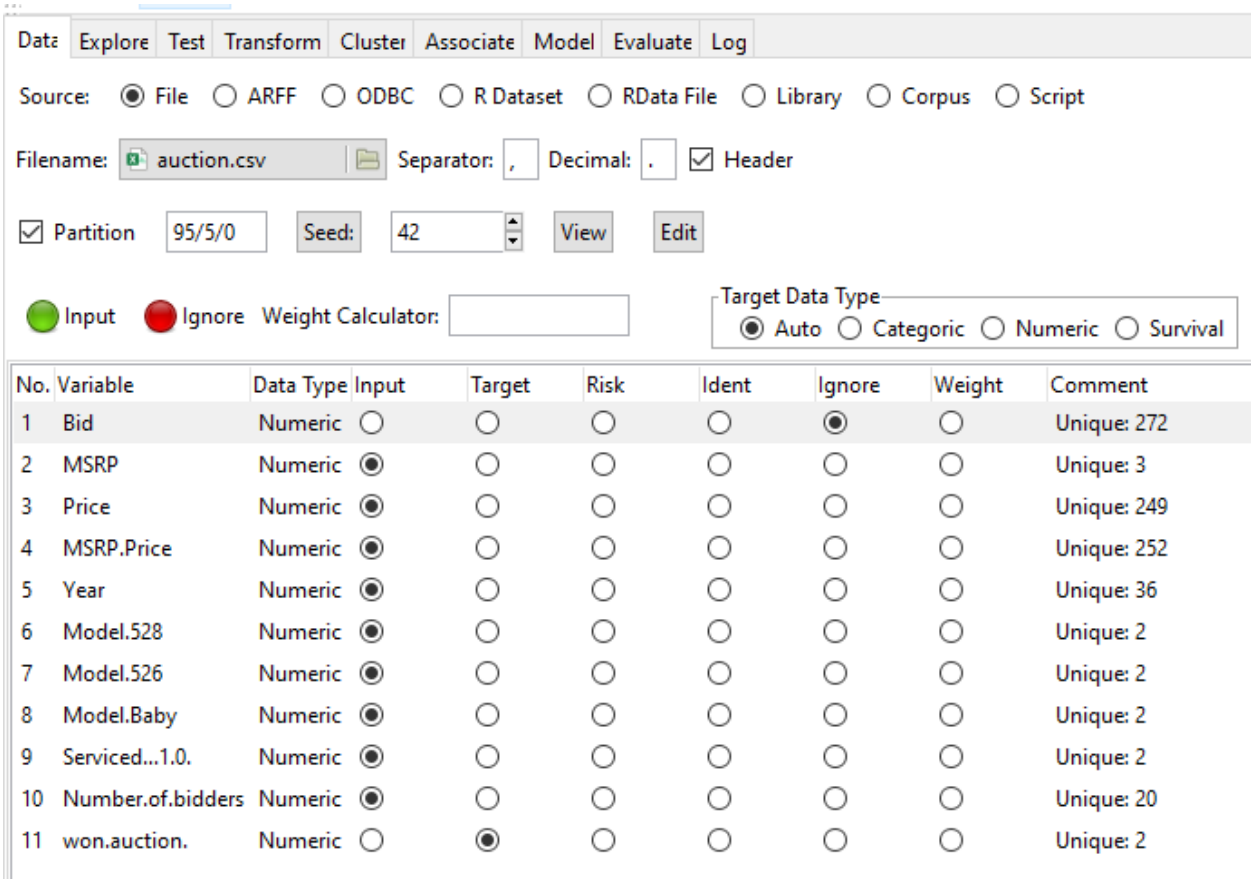
For Deliverable 8_2: (10 points)

- 10 points: Each step has been explained and verified, and findings are interpreted.
- 8 points: Explanation or verification of 1 or 2 steps are missing
- 5 point: Explanation or verification of multiple steps are missing, or findings are not interpreted.
- 0 points: No explanations, verification, or interpretations are given, or no assignment submission.

Solution 8

1. **Load the auction data ("auction.csv") on Rattle and define variable appropriately (won.auction is the target variable)**

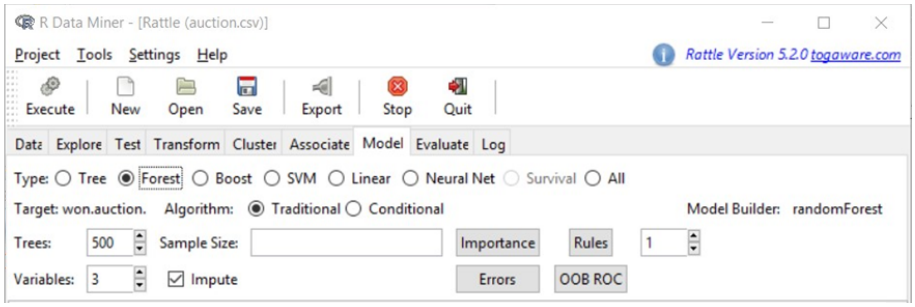
Answer



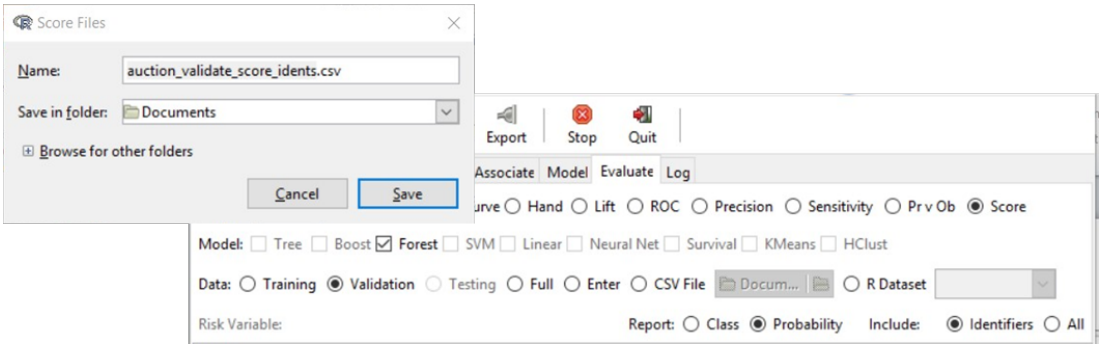
No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	Bid	Numeric	○	○	○	○	●	○	Unique: 272
2	MSRP	Numeric	●	○	○	○	○	○	Unique: 3
3	Price	Numeric	●	○	○	○	○	○	Unique: 249
4	MSRP.Price	Numeric	●	○	○	○	○	○	Unique: 252
5	Year	Numeric	●	○	○	○	○	○	Unique: 36
6	Model.528	Numeric	●	○	○	○	○	○	Unique: 2
7	Model.526	Numeric	●	○	○	○	○	○	Unique: 2
8	Model.Baby	Numeric	●	○	○	○	○	○	Unique: 2
9	Serviced...1.0.	Numeric	●	○	○	○	○	○	Unique: 2
10	Number.of.bidders	Numeric	●	○	○	○	○	○	Unique: 20
11	won.auction.	Numeric	○	●	○	○	○	○	Unique: 2

2. **Run Random Forest model with default options**

This document was prepared using R version 4.0.1
Depending on R version, the output may be slightly different, and that's acceptable.



3. Save probability scores on validate data in a an excel/ csv file



won.auction	rf
0	0.172
0	0.138
0	0.132
0	0.366
0	0.89
0	0.25
0	0.092
0	0.024
0	0.05
0	0.12
1	0.418
0	0.064
0	0.102
1	0.24

These are the probability scores for fourteen observations in the validate dataset. The higher probability means more likely to win the auction.

We can compare the actual outcome ("won") with the predicted outcome (by specifying a cutoff probability ("rf") value such that if the "rf" is above the cutoff, the predicted outcome is "won"=1. Then we can compare the actual outcome with the predicted outcome to see how accurate our predictions are.

This document was prepared using R version 4.0.1
Depending on R version, the output may be slightly different, and that's acceptable.

4. **Create a scoring method in Excel where given a cutoff probability it determines the error matrix**

   Note this may not be the real solution-
   you may or may not get similar solution. However, you need to report a similar table.

won.auction	rf	Predict	0,0	0,1	1,0
0	0.172	0	1	0	0
0	0.138	0	1	0	0
0	0.132	0	1	0	0
0	0.366	0	1	0	0
0	0.89	1	0	1	0
0	0.25	0	1	0	0
0	0.092	0	1	0	0
0	0.024	0	1	0	0
0	0.05	0	1	0	0
0	0.12	0	1	0	0
1	0.418	0	0	0	1
0	0.064	0	1	0	0
0	0.102	0	1	0	0
1	0.24	0	0	0	1
			11	1	2
Cutoff	0.8				

		Predict		
		0	1	Error
Actual	0	11	1	0.0833333
	1	2	0	1

Cutoff	Error	Sensitivity	Specificity	Precision
	0.2142857	0	0.917	0.000
0.1	0.5714286	1	0.333	0.200
0.2	0.2142857	1	0.750	0.400
0.3	0.2142857	0.5	0.833	0.333
0.4	0.1428571	0.5	0.917	0.500
0.5	0.2142857	0	0.917	0.000
0.6	0.2142857	0	0.917	0.000
0.7	0.2142857	0	0.917	0.000
0.8	0.2142857	0	0.917	0.000

The error is minimized at 0.4 value of cutoff value.