Diane Nguyen Homework Assignment 2 MSBA 645 4/4/19



This assignment is based on Problem 9.1 on page 214. These are the items to turn in:





\*\_\_\_\_\_\* Node = 5\*\_\_\_\_\_\* if OpenPrice < 3.72 AND ClosePrice >= 3.645 or MISSING then Tree Node Identifier = 5Number of Observations = 333Predicted: Competitive\_=1 = 0.98Predicted: Competitive\_=0 = 0.02\*\_\_\_\_\_\* Node = 6\*\_\_\_\_\_\* if OpenPrice >= 3.72 or MISSING AND ClosePrice < 10.06 then Tree Node Identifier = 6Number of Observations = 243Predicted: Competitive\_=1 = 0.17Predicted: Competitive\_=0 = 0.83\*\_\_\_\_\_\* Node = 8\*\_\_\_\_\_\* if OpenPrice < 1.035 AND ClosePrice < 3.645 then Tree Node Identifier = 8Number of Observations = 63Predicted: Competitive\_=1 = 0.63Predicted: Competitive\_=0 = 0.37

\*\_\_\_\_\_\* Node = 9\*\_\_\_\_\_\* if OpenPrice < 3.72 AND OpenPrice >= 1.035 or MISSING AND ClosePrice < 3.645 then Tree Node Identifier = 9Number of Observations = 150Predicted: Competitive\_=1 = 0.21Predicted: Competitive\_=0 = 0.79\*\_\_\_\_\_\* Node = 14\*\_\_\_\_\_\* if OpenPrice < 10.345 AND OpenPrice >= 3.72 AND ClosePrice >= 10.06 or MISSING then Tree Node Identifier = 14Number of Observations = 122Predicted: Competitive =1 = 0.86Predicted: Competitive\_=0 = 0.14\*\_\_\_\_\_\* Node = 22\*\_\_\_\_\_\* if sellerRating < 562AND OpenPrice >= 10.345 or MISSING AND ClosePrice >= 10.06 or MISSING then Tree Node Identifier = 22Number of Observations = 64Predicted: Competitive\_=1 = 0.63Predicted: Competitive\_=0 = 0.38\_\_\_\_\_\* Node = 23\*\_\_\_\_\_ \_\_\_\_\_\* if sellerRating  $\geq 562$  or MISSING AND OpenPrice >= 10.345 or MISSING AND ClosePrice >= 10.06 or MISSING then Tree Node Identifier = 23Number of Observations = 207Predicted: Competitive\_=1 = 0.26Predicted: Competitive\_=0 = 0.74

- **B.** You must provide an explanation for your answer. Is this model practical for predicting the outcome of a new auction?

As seen below, the miscrassification rate for competitive is 12.39% and for non-competitive is 23.03	As seen	n below,	the m	nisclass	sification	rate t	for coi	npetitive	e is	12.39%	and	for non	-comp	etitive	is	23.	65°	%
--	---------	----------	-------	----------	------------	--------	---------	-----------	------	--------	-----	---------	-------	---------	----	-----	-----	---

Outcome Type	Data Role 🔻	Target Variable	Target Label	Target	Outcome	Correct Text	Target Percentage	Outcome Percentage	Frequency Count	Total	Correct
										Percentage	
PREDICTION	VALIDATE	Competitive	Competitive?	0	0	Correct	75.89499	87.60331	318	40.25316	0
PREDICTION	VALIDATE	Competitive	Competitive?	1	0	Incorrect	24.10501	23.6534	101	12.78481	1
PREDICTION	VALIDATE	Competitive	Competitive?	0	1	Incorrect	12.12938	12.39669	45	5.696203	1
PREDICTION	VALIDATE	Competitive	Competitive?	1	1	Correct	87.87062	76.3466	326	41.26582	0
PREDICTION	TRAIN	Competitive	Competitive?	0	0	Correct	79	87.29282	474	40.10152	0
PREDICTION	TRAIN	Competitive	Competitive?	1	0	Incorrect	21	19.71831	126	10.6599	1
PREDICTION	TRAIN	Competitive	Competitive?	0	1	Incorrect	11.85567	12.70718	69	5.837563	1
PREDICTION	TRAIN	Competitive	Competitive?	1	1	Correct	88.14433	80.28169	513	43.40102	0

## As seen below the overall misclassification rate is 18.48%.

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Competitive	Competitive?	NOBS	Sum of Frequencies	1182	790	1
Competitive	Competitive?	MISC	Misclassification Rate	0.164975	0.18481	
Competitive	Competitive?	MAX	Maximum Absolute Error	0.984985	0.984985	j –
Competitive	Competitive?	SSE	Sum of Squared Errors	295.8465	220.2479	1
Competitive	Competitive?	ASE	Average Squared Error	0.125147	0.139397	1
Competitive	Competitive?	RASE	Root Average Squared Error	0.353761	0.37336	i i
Competitive	Competitive?	DIV	Divisor for ASE	2364	1580	1
Competitive	Competitive?	DFT	Total Degrees of Freedom	1182		

Based on the misclassification rates above for the validation set, it only misclassified the predicting outcome 18.481% of the time. Given this information I would say that this model is practical for predicting the outcome of a new auction.

- C. Name one rule that surprises you and explain why.



The rule that surprised me is that if the OpenPrice was <10.345, then they would receive no sellerRating. Given that Amazon gives all of its sellers a rating regardless of price you would think that eBay would do the same to ensure that all of its sellers offer a good product regardless of the price point. OpenPrice is the only thing that the seller can control that would make a difference in if their product is competitive or not.

- D. Fit another classification tree (using the best-pruned tree, with a minimum number of records per terminal node = 50 and maximum allowed number of displayed levels), this time only with predictors that can be used for predicting the outcome of a new auction. Describe the resulting tree in terms of rules. Make sure to report the smallest set of rules required for classification. Provide the overall, competitive, and non-competitive misclassification rates.

```
Node = 4

*______*

if OpenPrice < 1.035

then

Tree Node Identifier = 4

Number of Observations = 190

Predicted: Competitive_=1 = 0.88

Predicted: Competitive_=0 = 0.12

*______*

Node = 7

*______*

if sellerRating >= 660.5 or MISSING

AND OpenPrice >= 3.72 or MISSING

then
```

\*

```
Tree Node Identifier = 7
Number of Observations = 484
Predicted: Competitive_=1 = 0.31
Predicted: Competitive_=0 = 0.69
*_____*
Node = 8
*_____*
if sellerRating < 2365.5 or MISSING
AND OpenPrice < 3.72 AND OpenPrice >= 1.035 or MISSING
then
Tree Node Identifier = 8
Number of Observations = 268
Predicted: Competitive =1 = 0.74
Predicted: Competitive_=0 = 0.26
*_____*
Node = 9
*_____*
if sellerRating \geq 2365.5
AND OpenPrice < 3.72 AND OpenPrice >= 1.035 or MISSING
then
Tree Node Identifier = 9
Number of Observations = 88
Predicted: Competitive_=1 = 0.38
Predicted: Competitive_=0 = 0.63
*_____*
Node = 10
*_____*
if sellerRating < 660.5
AND OpenPrice < 9.97 AND OpenPrice >= 3.72
then
Tree Node Identifier = 10
Number of Observations = 61
Predicted: Competitive_=1 = 0.49
Predicted: Competitive =0 = 0.51
*_____*
Node = 11
*_____*
if sellerRating < 660.5
AND OpenPrice >= 9.97 or MISSING
then
Tree Node Identifier = 11
Number of Observations = 91
Predicted: Competitive =1 = 0.67
Predicted: Competitive_=0 = 0.33
```

- The smallest set of rules required for classification is:



Node = 7

if sellerRating >= 660.5 or MISSING AND OpenPrice >= 3.72 or MISSING then Tree Node Identifier = 7 Number of Observations = 484 Predicted: Competitive\_=1 = 0.31 Predicted: Competitive\_=0 = 0.69

\*\_\_\_\_\_\*

Outcome Type	Data Role 🔻	Target Variable	Target Label	Target	Outcome	Correct Text	Target Percentage	Outcome	Frequency Count	Total Percentage	Correct
								Percentage			
PREDICTION	VALIDATE	Competitive	Competitive?	0	0	Correct	67.86571	77.96143	283	35.82278	3 0
PREDICTION	VALIDATE	Competitive	Competitive?	1	0	Incorrect	32.13429	31.38173	134	16.96203	3 1
PREDICTION	VALIDATE	Competitive	Competitive?	0	1	Incorrect	21.44772	22.03857	80	) 10.12658	3 1
PREDICTION	VALIDATE	Competitive	Competitive?	1	1	Correct	78.55228	68.61827	293	37.08861	0
PREDICTION	TRAIN	Competitive	Competitive?	0	0	Correct	66.50869	77.53223	421	35.6176	3 0
PREDICTION	TRAIN	Competitive	Competitive?	1	0	Incorrect	33.49131	33.17684	212	2 17.9357	7 1
PREDICTION	TRAIN	Competitive	Competitive?	0	1	Incorrect	22.22222	22.46777	122	2 10.32149	) 1
PREDICTION	TRAIN	Competitive	Competitive?	1	1	Correct	77 77778	66 82316	42	36 12521	0

The misclassification rate for competitive is 31.38% and the misclassification rate for non-competitive is 22.038%.

\*

\_\_\_\_\_

	÷	· · · · · · · · · · · · · · · · · · ·				_
Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Tes
Competitive	Competitive?	NOBS	Sum of Frequencies	1182	790	
Competitive	Competitive?	MISC	Misclassification Rate	0.282572	0.270886	
Competitive	Competitive?	MAX	Maximum Absolute Error	0.878947	0.878947	
Competitive	Competitive?	SSE	Sum of Squared Errors	461.1236	301.3923	
Competitive	Competitive?	ASE	Average Squared Error	0.195061	0.190755	
Competitive	Competitive?	RASE	Root Average Squared Error	0.441657	0.436755	
Competitive	Competitive?	DIV	Divisor for ASE	2364	1580	
Competitive	Competitive?	DFT	Total Degrees of Freedom	1182		

The overall misclassification rate is 28.25%.

- E. List the variables in descending of variable importance and number of splits they are used for. View→models→ variable importance

Variable importance for decision tree 1:

Variable Name	Label	Number of Splitting Rules	Importance V	Validation Importance	Ratio of Validation to Training
					Importance
ClosePrice	ClosePrice		2 1.000	1.000	0 1.0000
OpenPrice	OpenPrice		3 0.9698	3 0.939	0 0.9682
sellerRating	sellerRating		0.304	3 0.425	5 1.3959
Category	Category	(	0.000	0.000	0.
currency	currency	(	0.000	0.000	0.
Duration	Duration	(	0.000	0.000	0.
endDav	endDav	(	0.000	0.000	0 .

## Variable importance for decision tree 2:

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
OpenPrice	OpenPrice		3 1.00	0 1.000	1.0000
sellerRating Variable Name	sellerRating		2 0.65	6 0.722	1.1103
Category	Category		0.00	0.000	) .
Duration	Duration		0.00	0.000	)
currency	currency		0.00	0.000	)

F. Based on second decision tree you have created (see d), what can you conclude from these data about the chances of an auction obtaining at least two bids and its relationship to the auction settings set by the seller (duration, opening price, ending day, currency)? What would you recommend for a seller as the strategy that will most likely lead to a competitive auction?

You can conclude that nothing matters except for opening price as indicated from gathering the importance of variables in the previous questions. I would recommend that a seller focus on setting a lower opening price in order to make the product they are selling seeming more enticing to people who may want to bid on it. This will lead to a competitive auction.