

SAS Assignment 3 – Decision Tree

Whether customers enrolled in the loyalty program & provided with coupons for the organic products purchased organic items or not

JOO YONG YOON

MI 353

Dr. CEVIKPARMAK

2 March 2023

Answers to the questions presented (#4, 8, and 9)

* Please refer to the figure(s) provided for the question.

4.

a&c. Set up the roles and levels for the variables / Reject DemCluster & DemClusterGroup

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
DemAffl	Input	Interval	No		No	.	.
DemAge	Input	Interval	No		No	.	.
DemCluster	Rejected	Nominal	No		No	.	.
DemClusterGroup	Rejected	Nominal	No		No	.	.
DemGender	Input	Nominal	No		No	.	.
DemReg	Input	Nominal	No		No	.	.
DemTVReg	Input	Nominal	No		No	.	.
ID	ID	Nominal	No		No	.	.
PromClass	Input	Nominal	No		No	.	.
PromSpend	Input	Interval	No		No	.	.
PromTime	Input	Interval	No		No	.	.
TargetAmt	Rejected	Interval	No		No	.	.
TargetBuy	Target	Binary	No		No	.	.

Figure 1. Variable Settings

b. Examine the distribution of the target variable. What is the proportion of individuals who purchased organic products?

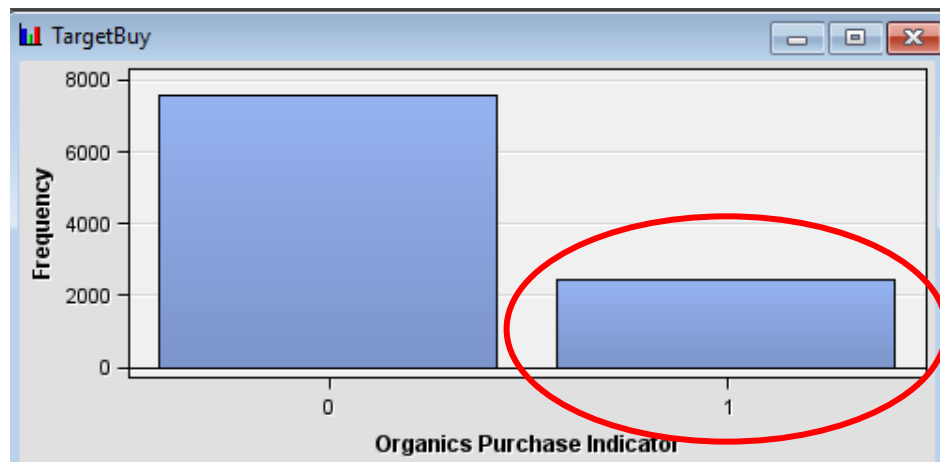


Figure 2. "TargetBuy" Distribution

According to Figure 2, about 24.77% of customers enrolled in the loyalty program & provided with coupons exclusively for the organic products purchased organic products.

d. Can TargetAmt be used as an input for a model used to predict TargetBuy? Why or why not?

TargetAmt cannot be used as input to predict TargetBuy because TargetAmt is the overlapped measure of TargetBuy. It is sufficient to predict organic product purchasers using TargetBuy. In addition, the purpose of the evaluation is not the amount of purchase per observation but rather the overall picture of buyers.

6. Partition Node (50% training, 50% validation)

Training	50.0
Validation	50.0
Test	0.0

Figure 3. Data Set Allocations

8. Decision Tree Model 1

Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Average Square Error
Assessment Fraction	0.25

Figure 4. Subtree Assessment Measure

a. How many leaves are in the optimal tree?

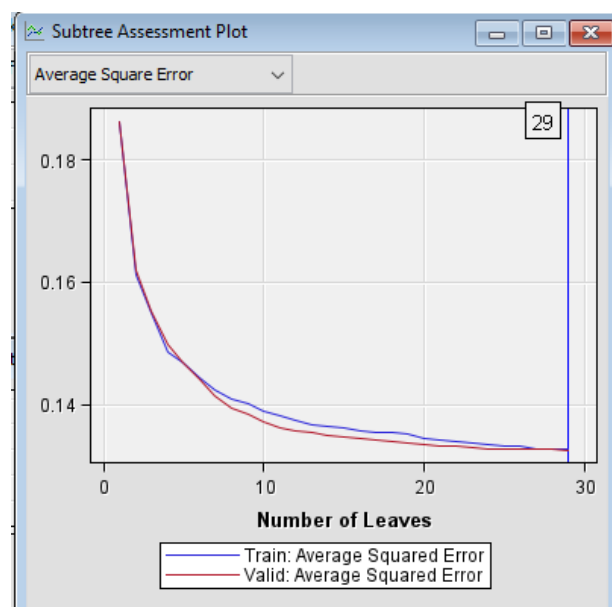


Figure 5. Model 1, # of Leaves

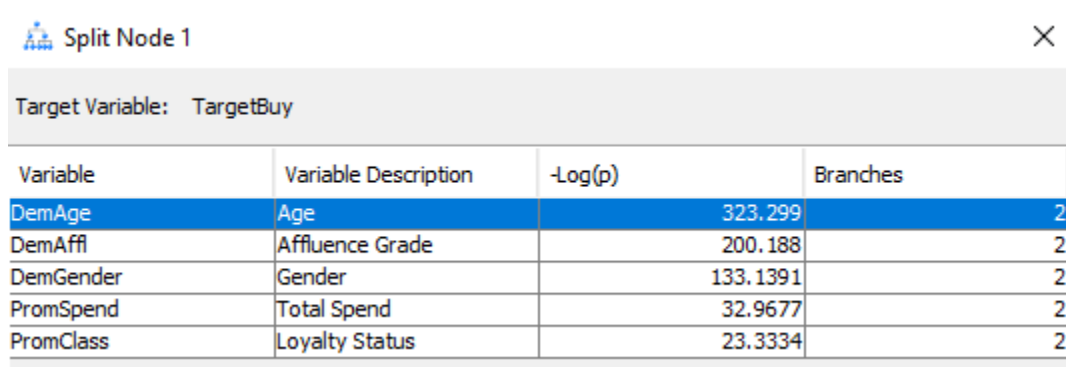
There are 29 leaves in the optimal tree for model 1 decision tree.

b. Which variable was used for the first split?

According to Figure 6 and 7, variable DemAge (Age, in years) was used for the first split.

c. What are the competing splits for this first split?

Competing splits are DemAffl (affluence grade), DemGender (gender), PromSpend (total spend) and PromClass (loyalty status).



Split Node 1

Target Variable: TargetBuy

Variable	Variable Description	-Log(p)	Branches
DemAge	Age	323.299	2
DemAffl	Affluence Grade	200.188	2
DemGender	Gender	133.1391	2
PromSpend	Total Spend	32.9677	2
PromClass	Loyalty Status	23.3334	2

Figure 6. Split Node 1, Model 1

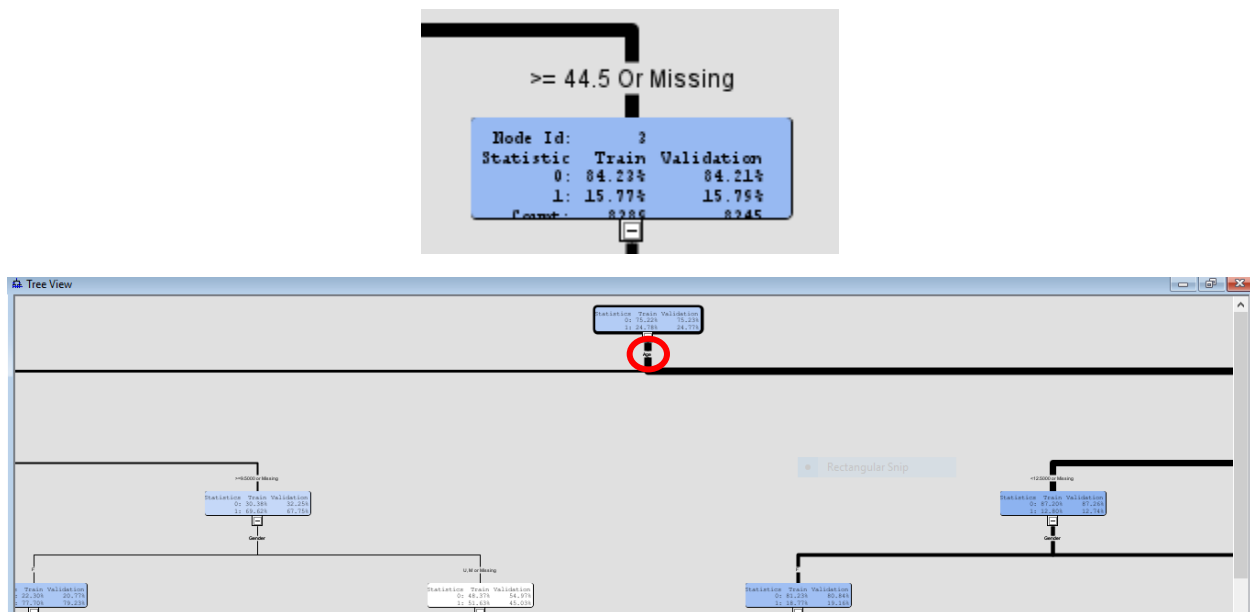


Figure 7. First Split, Model 1

9. Decision Tree Model 2

a. Change maximum number of branched to allow for three-way splits.

Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	3
Maximum Depth	6
Minimum Categorical Size	5

Figure 8. Maximum Branch 3

b. Subtree Assessment Measure to average square error.

Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Average Square Error
Assessment Fraction	0.25

Figure 9. Assessment Measure, Model 2

c. How many leaves are in the optimal tree?

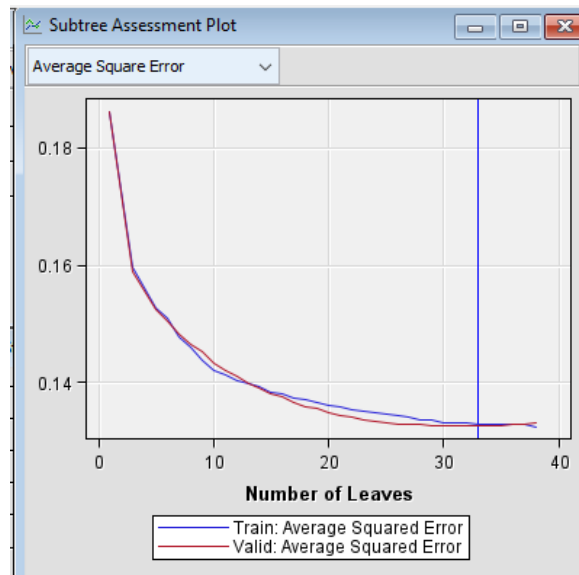
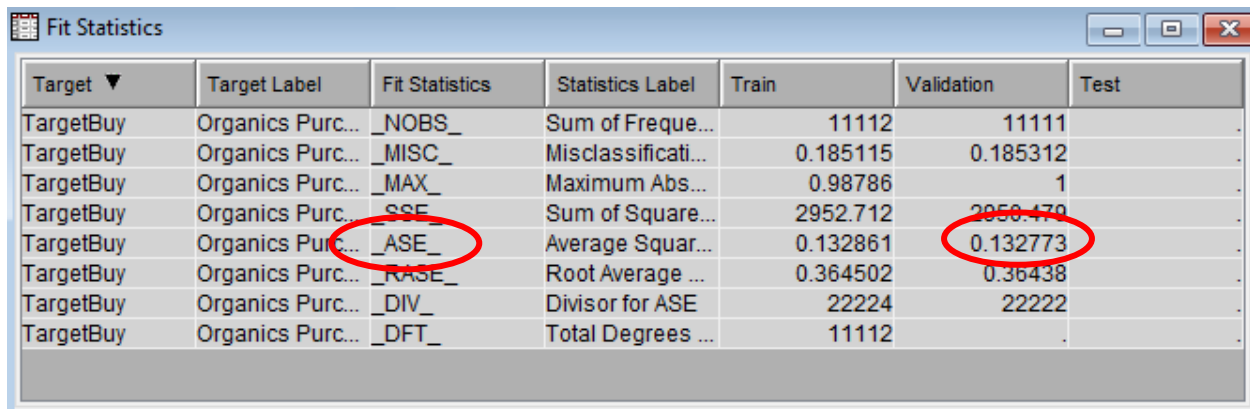


Figure 10. Model 2, # of Leaves

There are 33 leaves in the optimal tree for decision tree model 2.

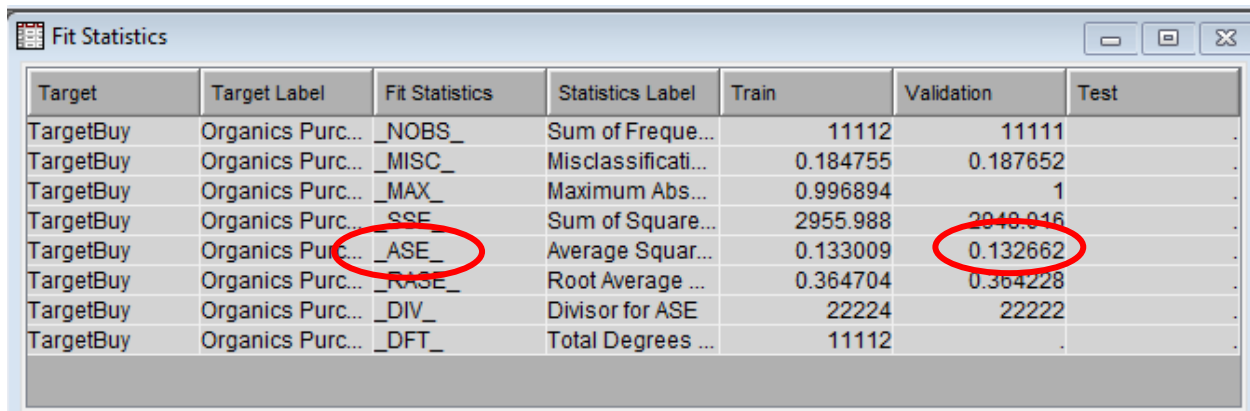
d. Based on average square error, which of the decision tree model appears to be better?



Fit Statistics

Target ▼	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
TargetBuy	Organics Purc...	_NOBS_	Sum of Freque...	11112	11111	.
TargetBuy	Organics Purc...	_MISC_	Misclassificati...	0.185115	0.185312	.
TargetBuy	Organics Purc...	_MAX_	Maximum Abs...	0.98786	1	.
TargetBuy	Organics Purc...	_SSE_	Sum of Square...	2952.712	2950.479	.
TargetBuy	Organics Purc...	_ASE_	Average Squar...	0.132861	0.132773	.
TargetBuy	Organics Purc...	_RASE_	Root Average ...	0.364502	0.36438	.
TargetBuy	Organics Purc...	_DIV_	Divisor for ASE	22224	22222	.
TargetBuy	Organics Purc...	_DFT_	Total Degrees ...	11112	.	.

Figure 11. Fit Statistics, Model 1



Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
TargetBuy	Organics Purc...	_NOBS_	Sum of Freque...	11112	11111	.
TargetBuy	Organics Purc...	_MISC_	Misclassificati...	0.184755	0.187652	.
TargetBuy	Organics Purc...	_MAX_	Maximum Abs...	0.996894	1	.
TargetBuy	Organics Purc...	_SSE_	Sum of Square...	2955.988	2948.016	.
TargetBuy	Organics Purc...	_ASE_	Average Squar...	0.133009	0.132662	.
TargetBuy	Organics Purc...	_RASE_	Root Average ...	0.364704	0.364228	.
TargetBuy	Organics Purc...	_DIV_	Divisor for ASE	22224	22222	.
TargetBuy	Organics Purc...	_DFT_	Total Degrees ...	11112	.	.

Figure 12. Fit Statistics, Model 2

According to Figure 11 & 12, the second decision tree (w/ three-way splits) has lower validation average square error than the first decision tree.