Documentation for Each Tool

SAS e-Miner:

Step1:

Before running the decision tree node, we first divide the data according to 70 (training): 30 (validation)**:**

Data Set Allocations	
Training	70.0
Validation	30.0
Test	0.0

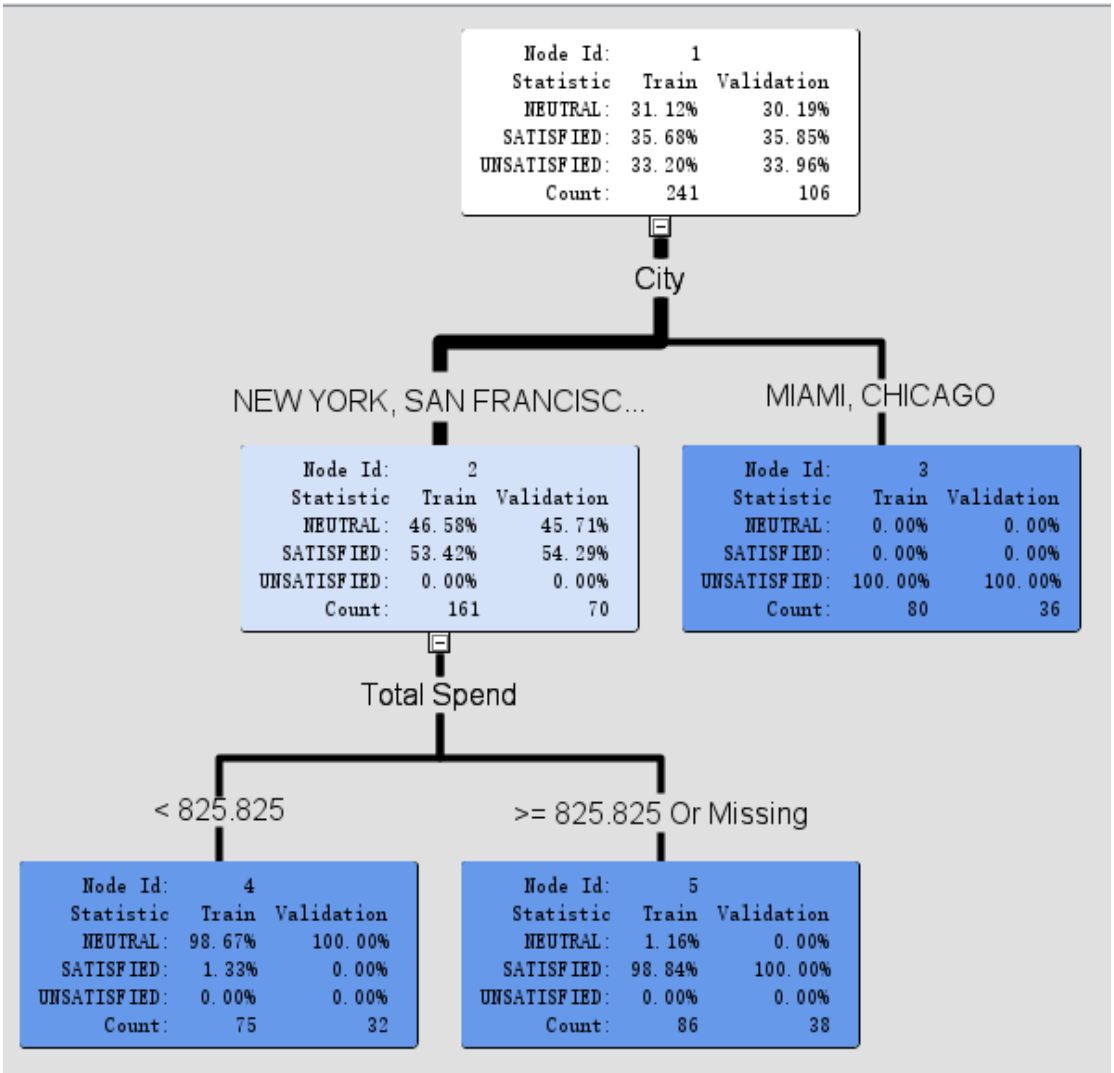
Step2:

Add decision tree node:

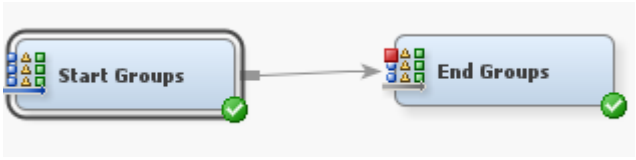
Configure decision tree parameters**:**

Property	Value
General	
Node ID	Tree
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Interactive	
Import Tree Model	No
Tree Model Data Set	
Use Frozen Tree	No
Use Multiple Targets	No
Splitting Rule	
Interval Target Criterio	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	2
Maximum Depth	6
Minimum Categorical Size	5
Node	
Leaf Size	5
Number of Rules	5
Number of Surrogate Rule	0
Split Size	.
Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Decision
Assessment Fraction	0.25

Run decision tree node, the results are as follows:

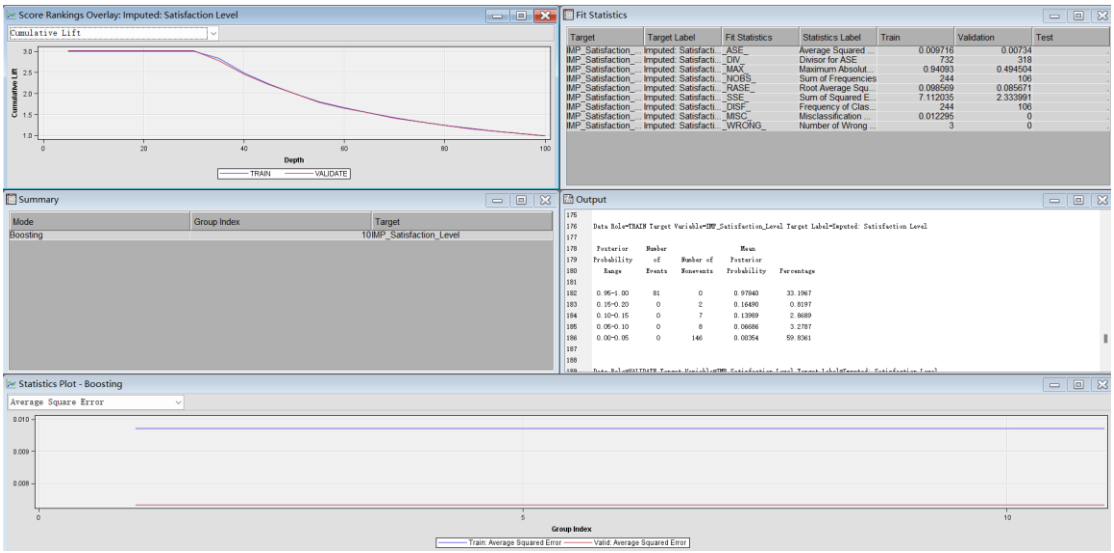


Add Start Groups and End Groups, and under the Start Groups node, set the mode to Boosting.



General	
Mode	Boosting
Target Group	No
Index Count	10
Minimum Group Size	10

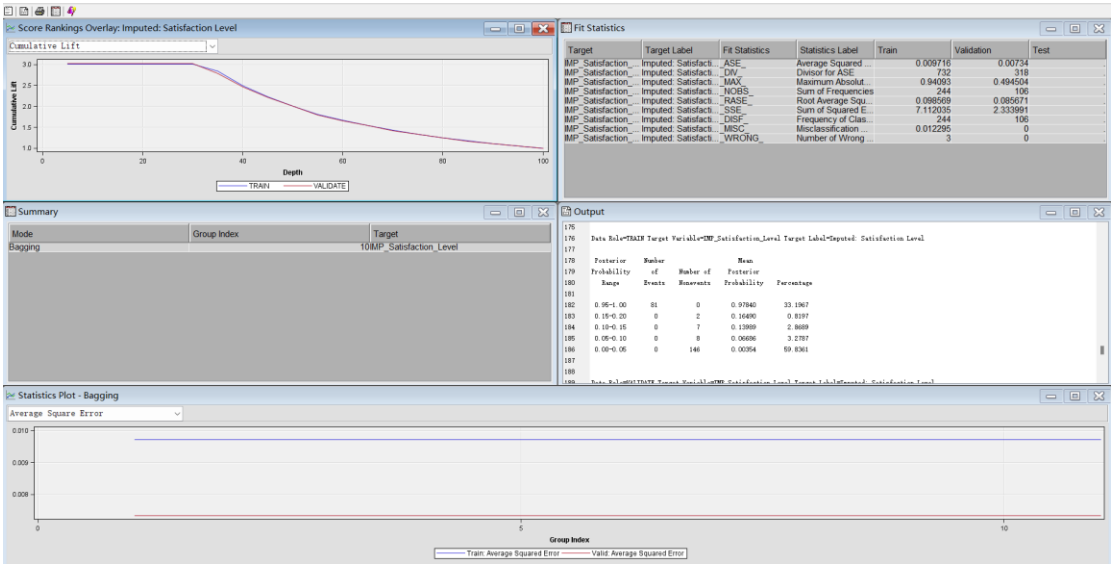
The result of End Groups in Boosting:



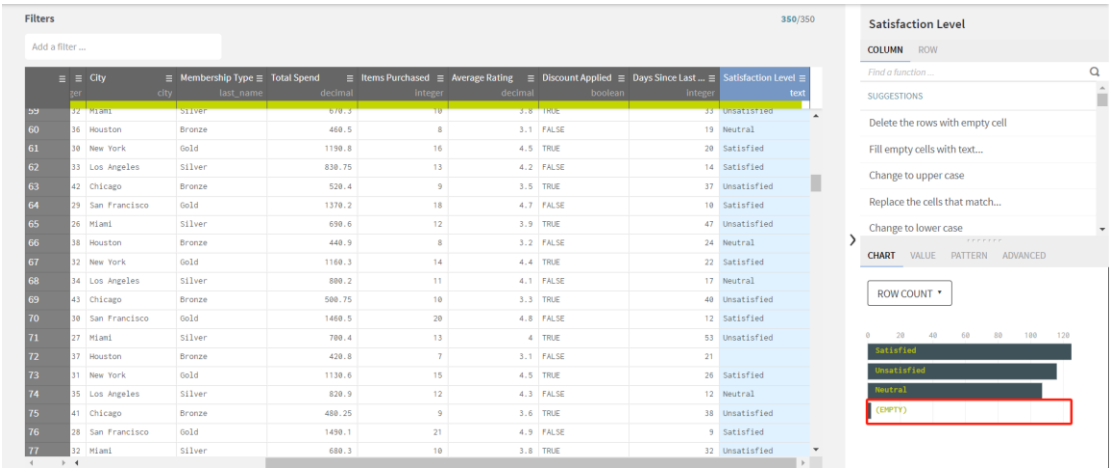
Add Start Groups and End Groups, and under the Start Groups node, set the mode to Bagging.

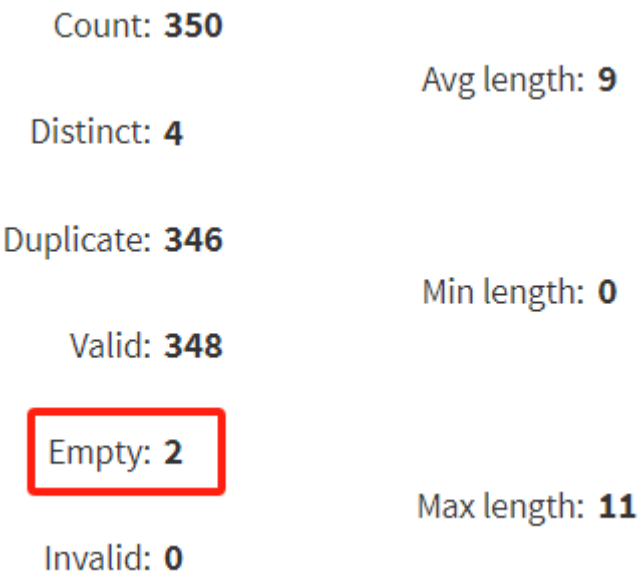
General	
Mode	Bagging
Target Group	No
Index Count	10
Minimum Group Size	10
Bagging	
Type	Percentage
Observations	.
Percentage	10.0
Random Seed	12345

The result of End Groups in Bagging:

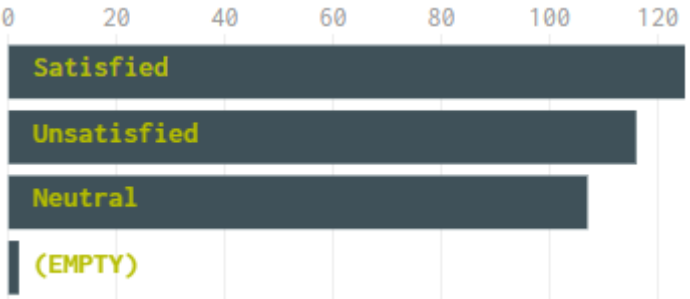


Talend Data Prep:

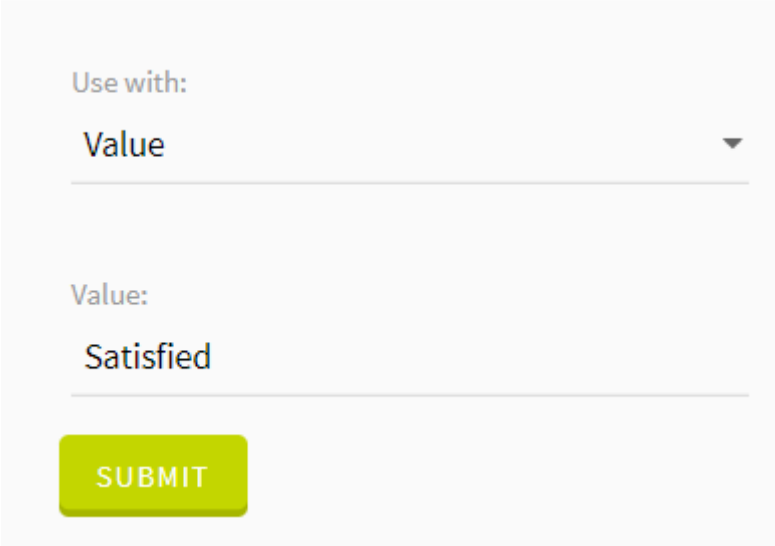




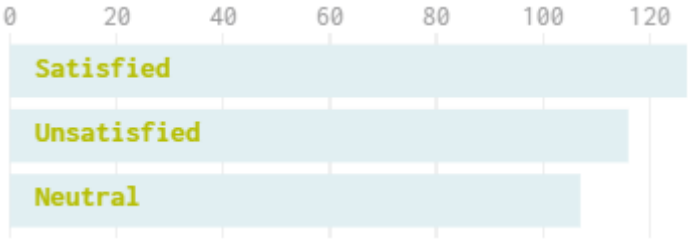
Since the most common category is satisfaction, the two missing values in the satisfaction_level are filled as "Satisfied".

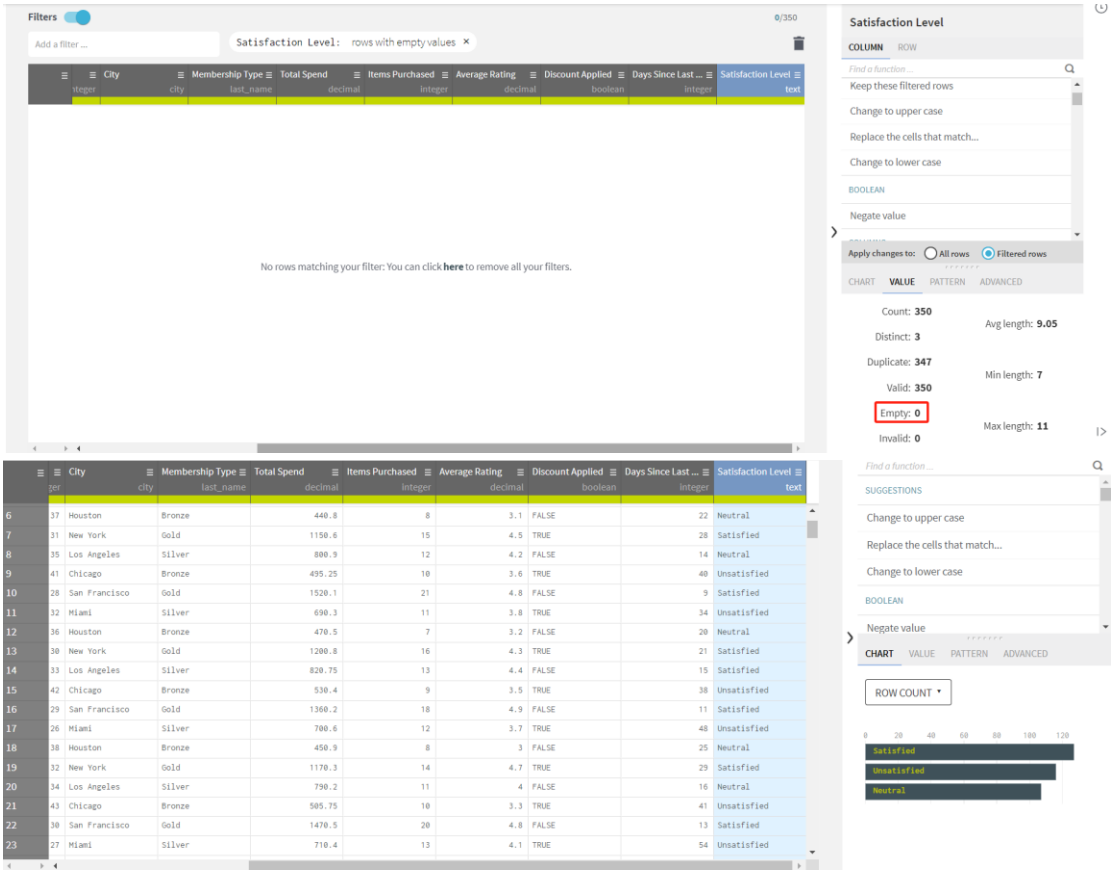


	City	Membership Type	Total Spend	Items Purchased	Average Rating	Discount Applied	Days Since Last ...	Satisfaction Level
	city	last_name	decimal	integer	decimal	boolean	integer	text
72	Houston	Bronze	420.8	7	3.1	FALSE	21	
144	Houston	Bronze	430.8	7	3.4	FALSE	23	



After filling in the missing values, empty no longer exists.





Missing value handling in SAS Miner:

The number of occurrences of Satisfied before missing value processing is 125.