

# Documentation for Each Tool

## SAS e-Miner:

Step1:

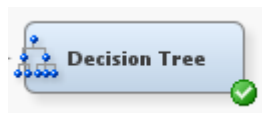
Before running the decision tree node, we first divide the data according to 70 (training): 30 (validation):



Data Set Allocations	
Training	70.0
Validation	30.0
Test	0.0

Step2:

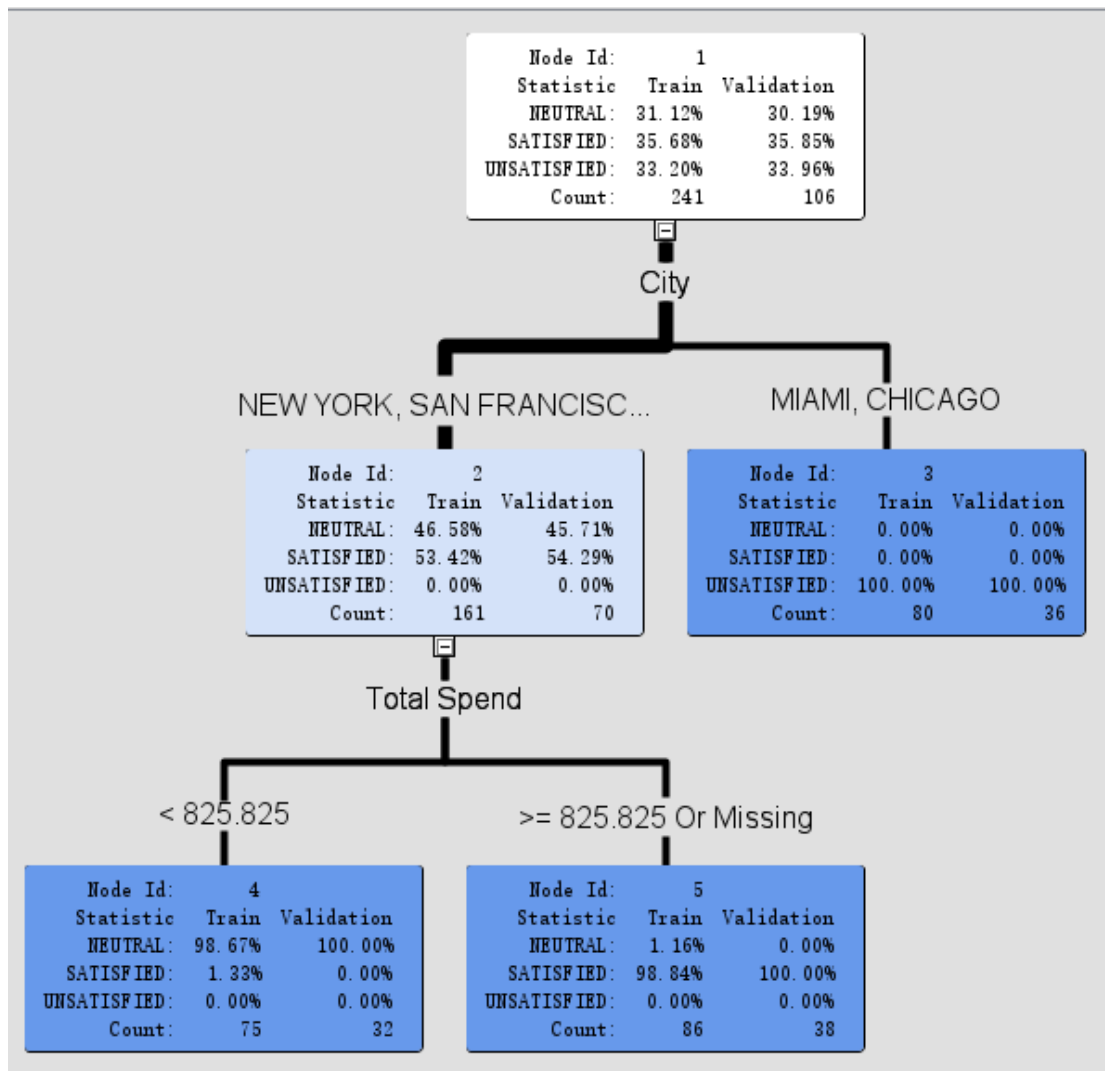
Add decision tree node:



Configure decision tree parameters:

Property	Value
<b>General</b>	
Node ID	Tree
Imported Data	...
Exported Data	...
Notes	...
<b>Train</b>	
Variables	...
Interactive	...
Import Tree Model	No
Tree Model Data Set	...
Use Frozen Tree	No
Use Multiple Targets	No
<input type="checkbox"/> 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	2
Maximum Depth	6
Minimum Categorical Size	5
<input type="checkbox"/> Node	
Leaf Size	5
Number of Rules	5
Number of Surrogate Rule	0
Split Size	.
<input type="checkbox"/> Split Search	
Use Decisions	No
Use Priors	No
Exhaustive	5000
Node Sample	20000
<input type="checkbox"/> Subtree	
Method	Assessment
Number of Leaves	1
Assessment Measure	Decision
Assessment Fraction	0.25

Run decision tree node, the results are as follows:

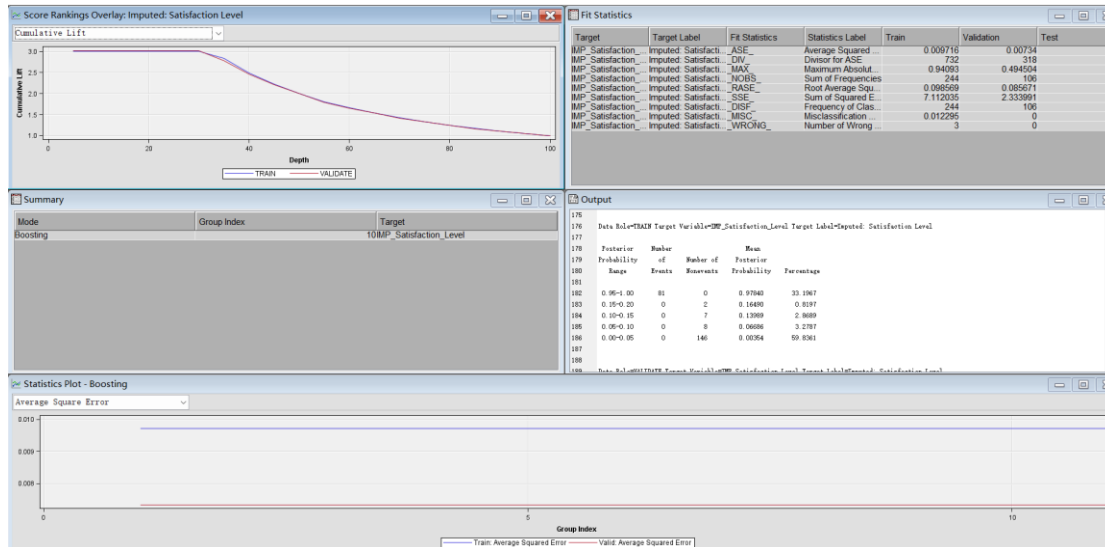


Add Start Groups and End Groups, and under the Start Groups node, set the mode to Boosting.



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

The result of End Groups in Boosting:

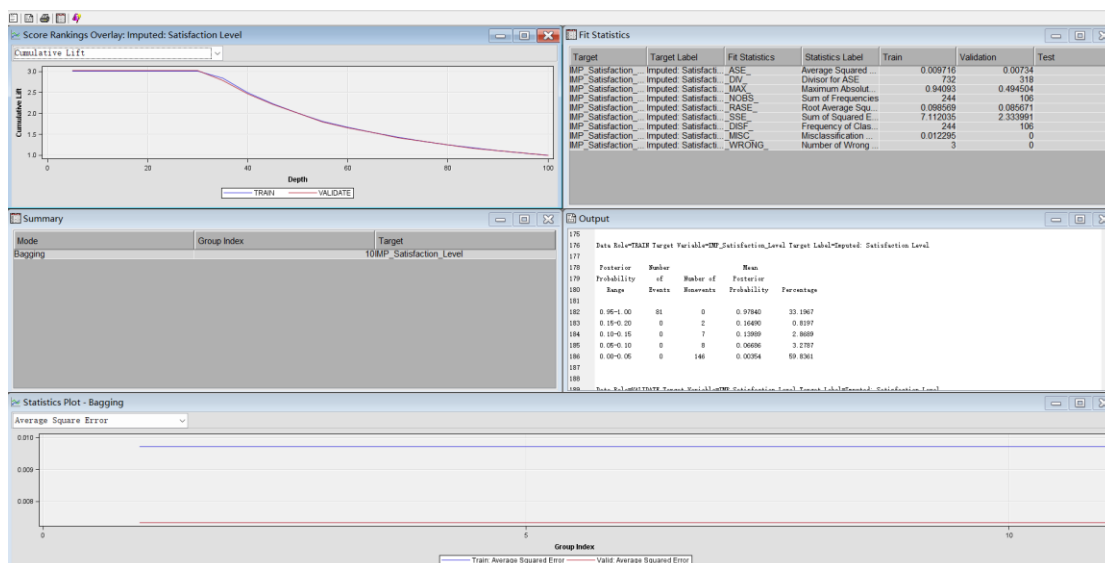


Add Start Groups and End Groups, and under the Start Groups node, set the mode to Bagging.

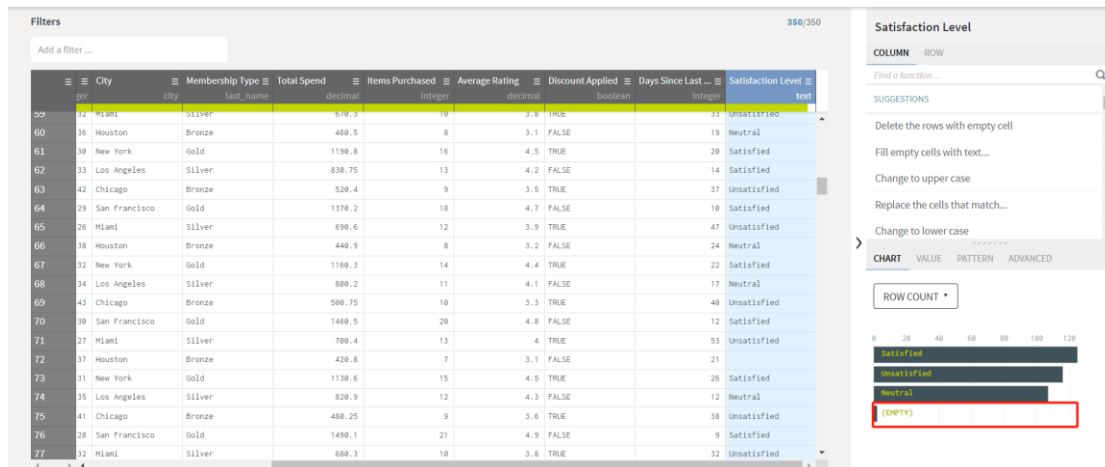


General	
Mode	Bagging
Target Group	No
Index Count	10
Minimum Group Size	10
Bagging	
Type	Percentage
Observations	.
Percentage	10.0
Random Seed	12345

The result of End Groups in Bagging:



## Talend Data Prep:



Count: **350**

Avg length: **9**

Distinct: **4**

Duplicate: **346**

Min length: **0**

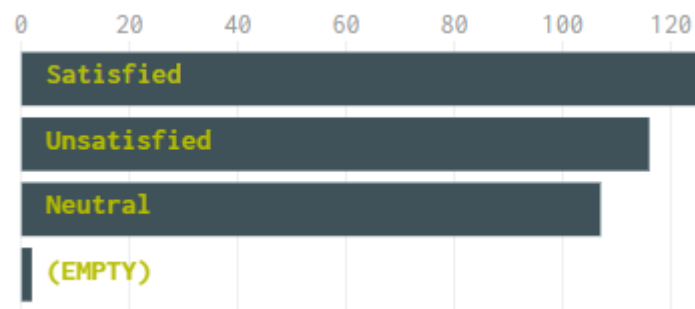
Valid: **348**

Empty: **2**

Max length: **11**

Invalid: **0**

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
integer	city	last_name	decimal	integer	decimal	boolean	integer	text
72	37	Houston	Bronze	420.8	7	3.1	FALSE	21
144	37	Houston	Bronze	430.8	7	3.4	FALSE	23

Use with:

Value

---

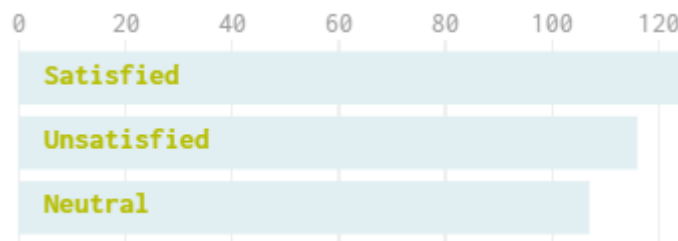
Value:

Satisfied

---

**SUBMIT**

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



Filters 0/350

Add a filter ... Satisfaction Level: rows with empty values

	City	Membership Type	Total Spend	Items Purchased	Average Rating	Discount Applied	Days Since Last ...	Satisfaction Level
	string	string	decimal	integer	decimal	boolean	integer	text
No rows matching your filter. You can click <a href="#">here</a> to remove all your filters.								

**Satisfaction Level**

COLUMN ROW

Find a function ...

Keep these filtered rows

Change to upper case

Replace the cells that match...

Change to lower case

BOOLEAN

Negate value

Apply changes to: ☐ All rows ☒ Filtered rows

CHART VALUE PATTERN ADVANCED

Count: 350 Avg length: 9.05

Distinct: 3

Duplicate: 347

Valid: 350

**Empty: 0**

Invalid: 0

Min length: 7

Max length: 11

Find a function ...

SUGGESTIONS

Change to upper case

Replace the cells that match...

Change to lower case

BOOLEAN

Negate value

CHART VALUE PATTERN ADVANCED

ROW COUNT

Satisfaction Level	Count (approx.)
Satisfied	120
Unsatisfied	100
Neutral	90

Missing value handling in SAS Miner:

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

