



Classification 1

(Introduction and Decision Tree)

Prepared by Raymond Wong

The examples used in Decision Tree are borrowed from LW Chan's notes
XLMiner Screenshot captured by Hao Liu

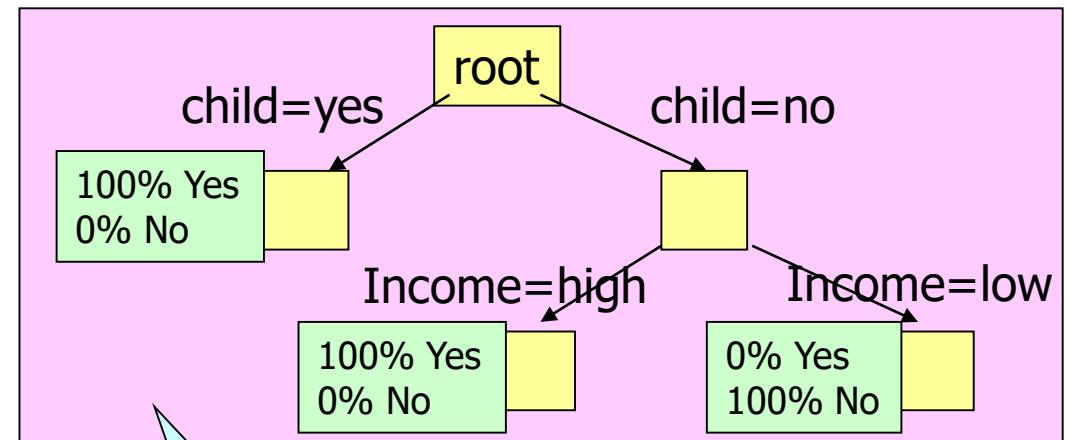
Presented by Raymond Wong

raywong@cse

Classification

Suppose there is a person.

Race	Income	Child	Insurance
white	high	no	?



Decision tree

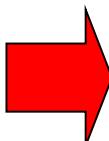
Classification

Suppose there is a person.

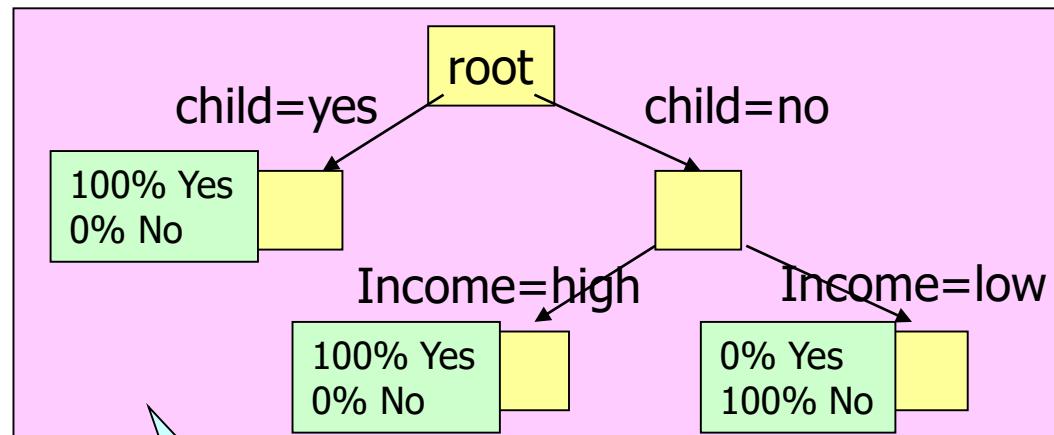
Race	Income	Child	Insurance
white	high	no	?

New set

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no



Training set



Decision tree

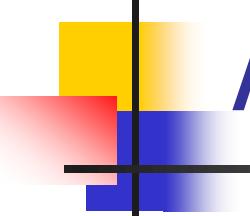
Classification

Suppose there is a person.

Race	Income	Child	Insurance
white	high	no	?

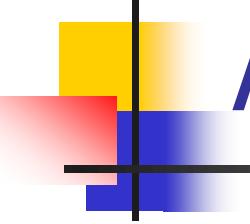
Input attributes

Target attributes



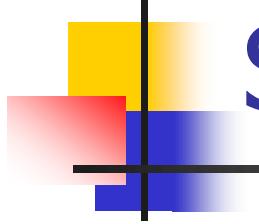
Applications

- Insurance
 - According to the attributes of customers,
 - Determine which customers will buy an insurance policy
- Marketing
 - According to the attributes of customers,
 - Determine which customers will buy a product such as computers
- Bank Loan
 - According to the attributes of customers,
 - Determine which customers are “risky” customers or “safe” customers



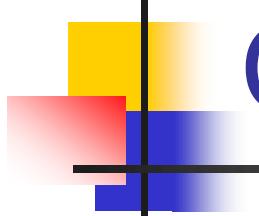
Applications

- Network
 - According to the traffic patterns,
 - Determine whether the patterns are related to some “security attacks”
- Software
 - According to the experience of programmers,
 - Determine which programmers can fix some certain bugs



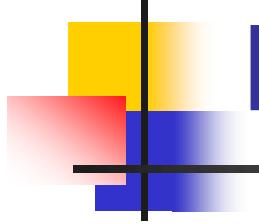
Same/Difference

- Classification
- Clustering



Classification Methods

- Decision Tree
- Bayesian Classifier
- Nearest Neighbor Classifier



Decision Trees

- Decision Trees

- ID3

Iterative Dichotomiser

- C4.5

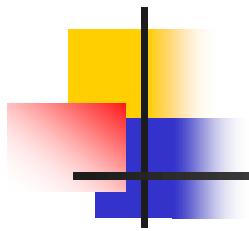
Classification

- CART

Classification And Regression Trees

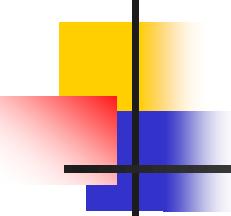
- Measurement

- How to use the data mining tool



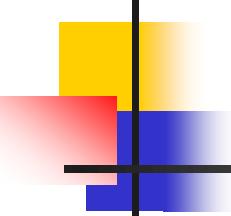
Entropy

- Example 1
 - Consider a random variable which has a uniform distribution over 32 outcomes
 - To identify an outcome, we need a label that takes 32 different values.
 - Thus, 5 bit strings suffice as labels



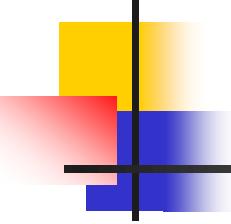
Entropy

- **Entropy** is used to measure how informative is a node.
- If we are given a probability distribution $P = (p_1, p_2, \dots, p_n)$ then the **Information** conveyed by this distribution, also called the **Entropy** of P , is:
$$I(P) = - (p_1 \times \log p_1 + p_2 \times \log p_2 + \dots + p_n \times \log p_n)$$
- All logarithms here are in base 2.



Entropy

- For example,
 - If P is $(0.5, 0.5)$, then $I(P)$ is 1.
 - If P is $(0.67, 0.33)$, then $I(P)$ is 0.92,
 - If P is $(1, 0)$, then $I(P)$ is 0.
- The **entropy** is a way to measure the amount of information.
- The smaller the entropy, the more informative we have.



Entropy

$$\text{Info}(T) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1$$

For attribute Race,

$$\text{Info}(T_{\text{black}}) = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} = 0.8113$$

$$\text{Info}(T_{\text{white}}) = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} = 0.8113$$

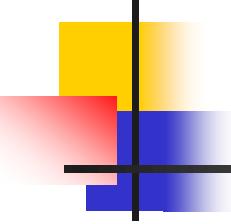
$$\text{Info}(\text{Race}, T) = \frac{1}{2} \times \text{Info}(T_{\text{black}}) + \frac{1}{2} \times \text{Info}(T_{\text{white}}) = 0.8113$$

$$\text{Gain}(\text{Race}, T) = \text{Info}(T) - \text{Info}(\text{Race}, T) = 1 - 0.8113 = 0.1887$$

For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.1887$$

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no



Entropy

$$\text{Info}(T) = - \frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} \\ = 1$$

For attribute Income,

$$\text{Info}(T_{\text{high}}) = - 1 \log 1 - 0 \log 0 = 0$$

$$\text{Info}(T_{\text{low}}) = - \frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} = 0.9183$$

$$\text{Info}(\text{Income}, T) = \frac{1}{4} \times \text{Info}(T_{\text{high}}) + \frac{3}{4} \times \text{Info}(T_{\text{low}}) = 0.6887$$

$$\text{Gain}(\text{Income}, T) = \text{Info}(T) - \text{Info}(\text{Income}, T) = 1 - 0.6887 = 0.3113$$

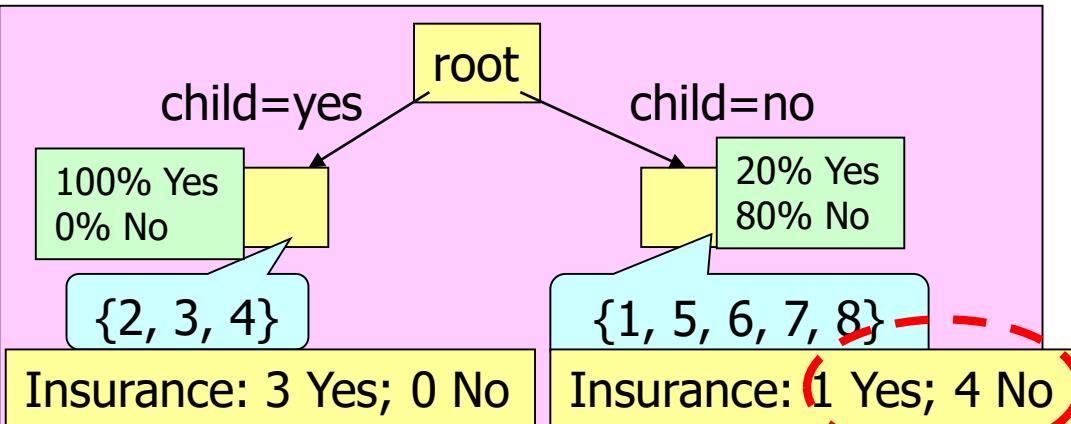
For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.1887$$

For attribute Income,

$$\text{Gain}(\text{Income}, T) = 0.3113$$

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no



$$\begin{aligned}\text{Info}(T) &= - \frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} \\ &= 1\end{aligned}$$

For attribute Child,

$$\text{Info}(T_{\text{yes}}) = - 1 \log 1 - 0 \log 0 = 0$$

$$\text{Info}(T_{\text{no}}) = - \frac{1}{5} \log \frac{1}{5} - \frac{4}{5} \log \frac{4}{5} = 0.7219$$

$$\text{Info}(\text{Child}, T) = \frac{3}{8} \times \text{Info}(T_{\text{yes}}) + \frac{5}{8} \times \text{Info}(T_{\text{no}}) = 0.4512$$

$$\text{Gain}(\text{Child}, T) = \text{Info}(T) - \text{Info}(\text{Child}, T) = 1 - 0.4512 = 0.5488$$

For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.1887$$

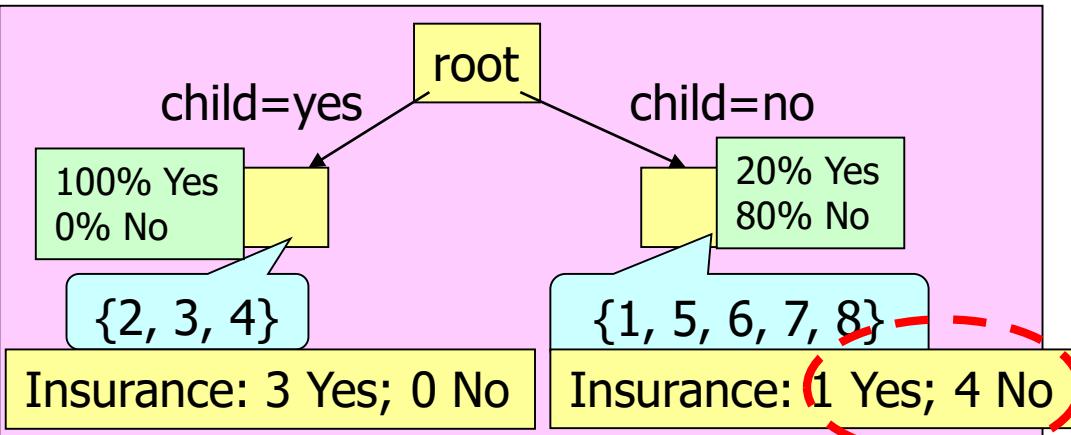
For attribute Income,

$$\text{Gain}(\text{Income}, T) = 0.3113$$

For attribute Child,

$$\text{Gain}(\text{Child}, T) = 0.5488$$

	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no



$$\begin{aligned}\text{Info}(T) &= - \frac{1}{5} \log \frac{1}{5} - \frac{4}{5} \log \frac{4}{5} \\ &= 0.7219\end{aligned}$$

For attribute Race,

$$\text{Info}(T_{\text{black}}) = - \frac{1}{4} \log \frac{1}{4} - \frac{3}{4} \log \frac{3}{4} = 0.8113$$

$$\text{Info}(T_{\text{white}}) = - 0 \log 0 - 1 \log 1 = 0$$

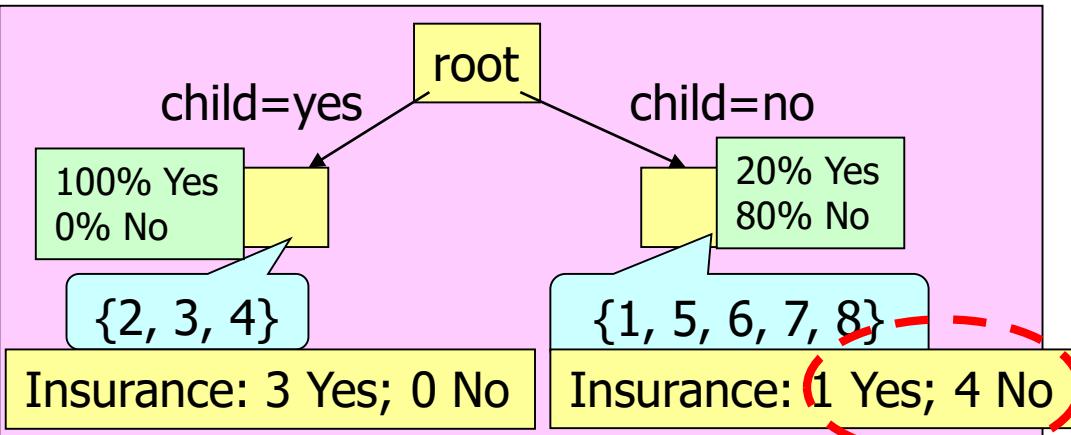
$$\text{Info}(\text{Race}, T) = \frac{4}{5} \times \text{Info}(T_{\text{black}}) + \frac{1}{5} \times \text{Info}(T_{\text{white}}) = 0.6490$$

$$\text{Gain}(\text{Race}, T) = \text{Info}(T) - \text{Info}(\text{Race}, T) = 0.7219 - 0.6490 = 0.0729$$

For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.0729$$

	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no



$$\begin{aligned}\text{Info}(T) &= - \frac{1}{5} \log \frac{1}{5} - \frac{4}{5} \log \frac{4}{5} \\ &= 0.7219\end{aligned}$$

For attribute Income,

$$\text{Info}(T_{\text{high}}) = - 1 \log 1 - 0 \log 0 = 0$$

$$\text{Info}(T_{\text{low}}) = - 0 \log 0 - 1 \log 1 = 0$$

$$\text{Info}(\text{Income}, T) = \frac{1}{5} \times \text{Info}(T_{\text{high}}) + \frac{4}{5} \times \text{Info}(T_{\text{low}}) = 0$$

$$\text{Gain}(\text{Income}, T) = \text{Info}(T) - \text{Info}(\text{Income}, T) = 0.7219 - 0 = 0.7219$$

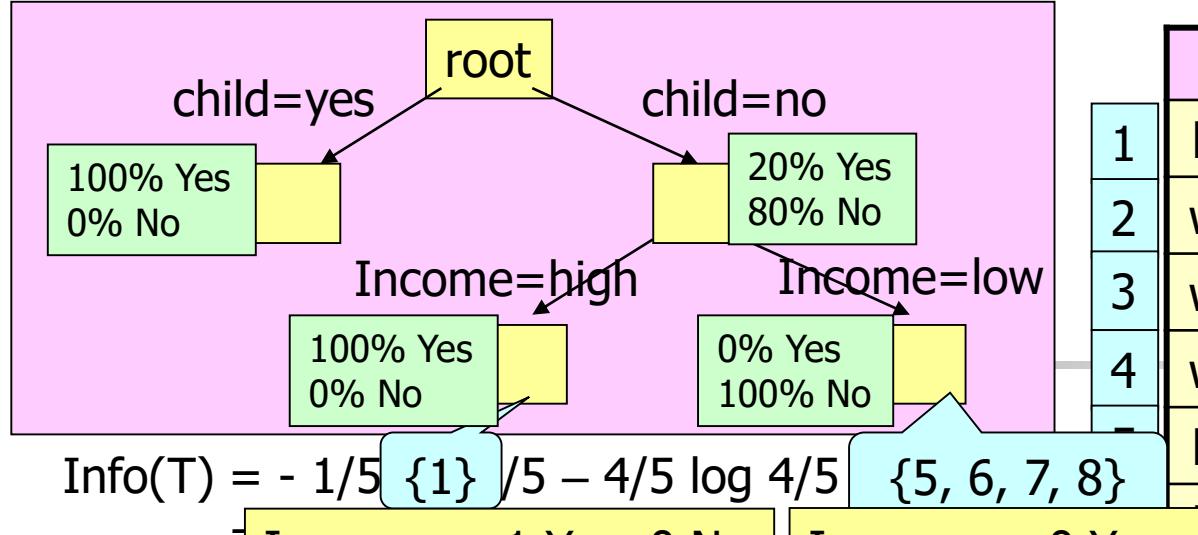
For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.0729$$

For attribute Income,

$$\text{Gain}(\text{Income}, T) = 0.7219$$

	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no



$$\text{Info}(T) = - \frac{1}{5} \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5}$$

= Insurance: 1 Yes; 0 No Insurance: 0 Yes; 4 No

For attribute Income,

$$\text{Info}(T_{\text{high}}) = -1 \log 1 - 0 \log 0 = 0$$

$$\text{Info}(T_{\text{low}}) = -0 \log 0 - 1 \log 1 = 0$$

$$\text{Info}(\text{Income}, T) = \frac{1}{5} \times \text{Info}(T_{\text{high}}) + \frac{4}{5} \times \text{Info}(T_{\text{low}}) = 0$$

$$\text{Gain}(\text{Income}, T) = \text{Info}(T) - \text{Info}(\text{Income}, T) = 0.7219 - 0 = 0.7219$$

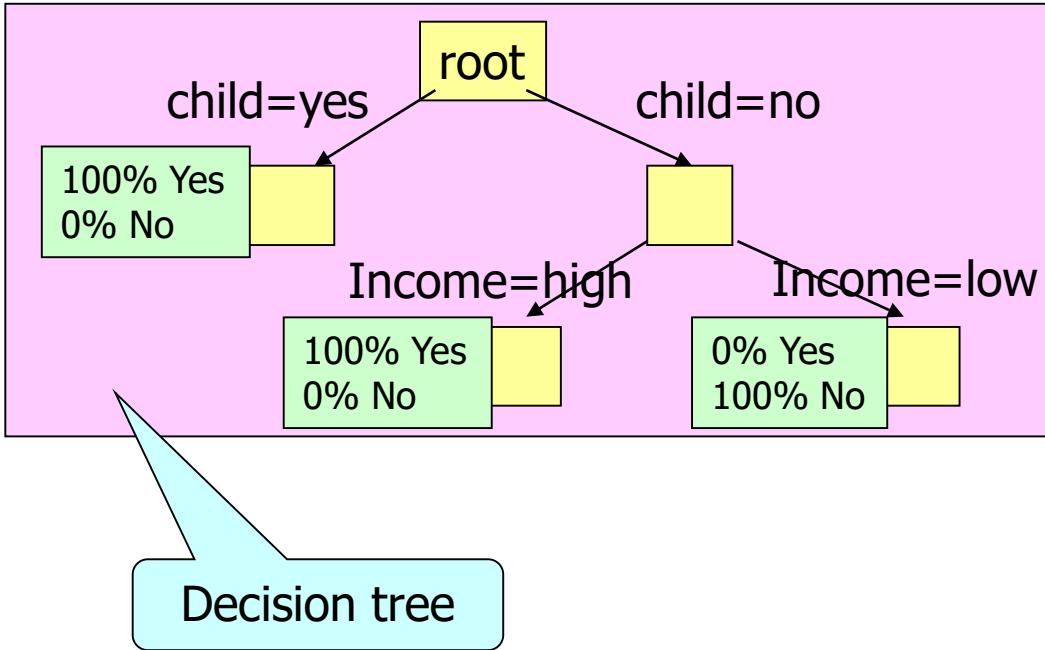
For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.0729$$

For attribute Income,

$$\text{Gain}(\text{Income}, T) = 0.7219$$

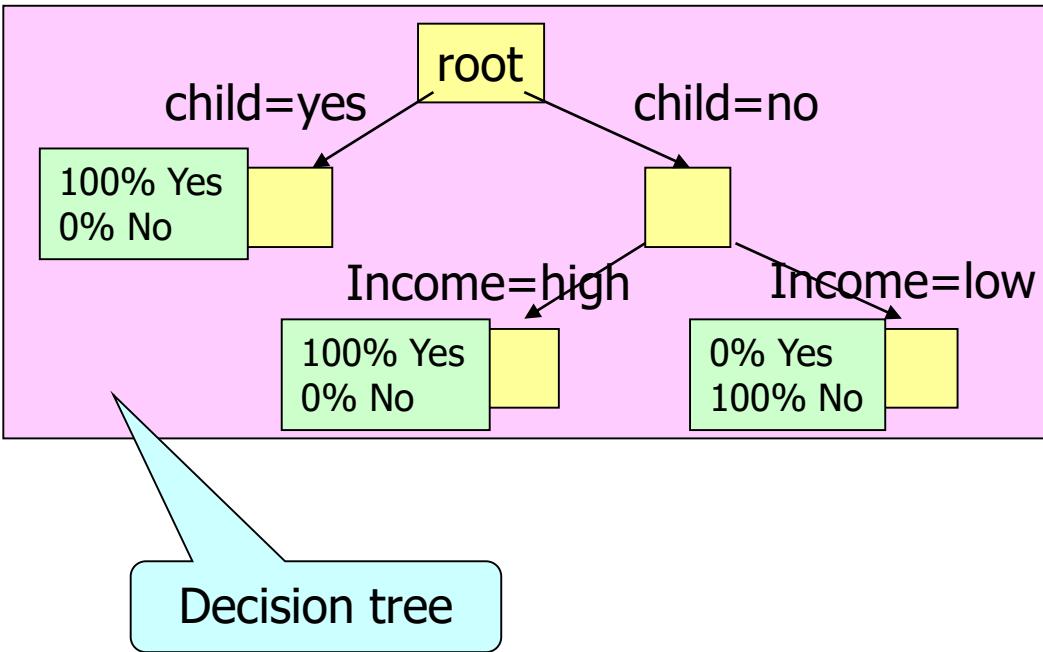
	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
	black	low	no	no
		low	no	no
7	black	low	no	no
8	white	low	no	no



	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no

Suppose there is a new person.

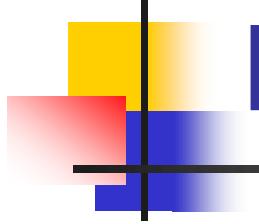
Race	Income	Child	Insurance
white	high	no	?



	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no

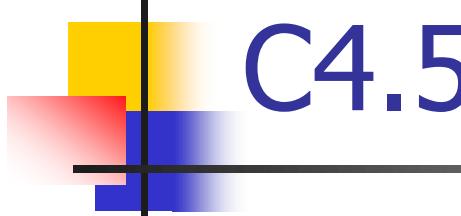
Termination Criteria?

e.g., height of the tree
e.g., accuracy of each node



Decision Trees

- Decision Trees
 - ID3
 - C4.5
 - CART
- Measurement
- How to use the data mining tool



C4.5

- ID3

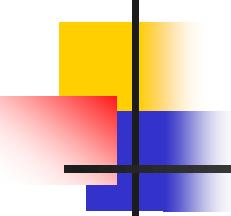
- Impurity Measurement

- Gain(A, T)
= Info(T) – Info(A, T)

- C4.5

- Impurity Measurement

- Gain(A, T)
= (Info(T) – Info(A, T))/SplitInfo(A)
 - where SplitInfo(A) = $-\sum_{v \in A} p(v) \log p(v)$



Entropy

$$\begin{aligned}\text{Info}(T) &= -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} \\ &= 1\end{aligned}$$

For attribute Race,

$$\text{Info}(T_{\text{black}}) = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} = 0.8113$$

$$\text{Info}(T_{\text{white}}) = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} = 0.8113$$

$$\text{Info}(\text{Race}, T) = \frac{1}{2} \times \text{Info}(T_{\text{black}}) + \frac{1}{2} \times \text{Info}(T_{\text{white}}) = 0.8113$$

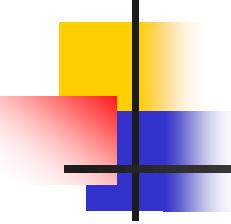
$$\text{SplitInfo}(\text{Race}) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1$$

$$\text{Gain}(\text{Race}, T) = (\text{Info}(T) - \text{Info}(\text{Race}, T)) / \text{SplitInfo}(\text{Race}) = (1 - 0.8113) / 1 = 0.1887$$

For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.1887$$

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no



Entropy

$$\text{Info}(T) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} \\ = 1$$

For attribute Income,

$$\text{Info}(T_{\text{high}}) = -1 \log 1 - 0 \log 0 = 0$$

$$\text{Info}(T_{\text{low}}) = -\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} = 0.9183$$

$$\text{Info}(\text{Income}, T) = \frac{1}{4} \times \text{Info}(T_{\text{high}}) + \frac{3}{4} \times \text{Info}(T_{\text{low}}) = 0.6887$$

$$\text{SplitInfo}(\text{Income}) = -\frac{2}{8} \log \frac{2}{8} - \frac{6}{8} \log \frac{6}{8} = 0.8113$$

$$\text{Gain}(\text{Income}, T) = (\text{Info}(T) - \text{Info}(\text{Income}, T)) / \text{SplitInfo}(\text{Income}) = (1 - 0.6887) / 0.8113 \\ = 0.3837$$

For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.1887$$

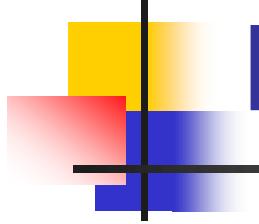
For attribute Income,

$$\text{Gain}(\text{Income}, T) = 0.3837$$

For attribute Child,

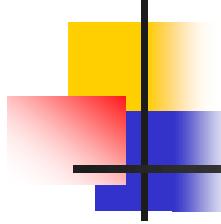
$$\text{Gain}(\text{Child}, T) = ?$$

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no



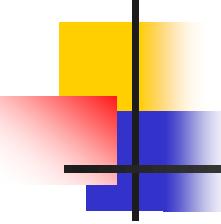
Decision Trees

- Decision Trees
 - ID3
 - C4.5
 - CART
- Measurement
- How to use the data mining tool



CART

- Impurity Measurement
 - Gini
- $$I(P) = 1 - \sum_j p_j^2$$



Gini

$$\text{Info}(T) = 1 - (\frac{1}{2})^2 - (\frac{1}{2})^2 = \frac{1}{2}$$

For attribute Race,

$$\text{Info}(T_{\text{black}}) = 1 - (\frac{3}{4})^2 - (\frac{1}{4})^2 = 0.375$$

$$\text{Info}(T_{\text{white}}) = 1 - (\frac{3}{4})^2 - (\frac{1}{4})^2 = 0.375$$

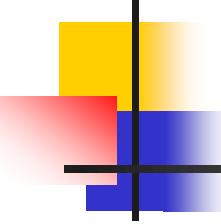
$$\text{Info}(\text{Race}, T) = \frac{1}{2} \times \text{Info}(T_{\text{black}}) + \frac{1}{2} \times \text{Info}(T_{\text{white}}) = 0.375$$

$$\text{Gain}(\text{Race}, T) = \text{Info}(T) - \text{Info}(\text{Race}, T) = \frac{1}{2} - 0.375 = 0.125$$

For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.125$$

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no



Gini

$$\text{Info}(T) = 1 - (\frac{1}{2})^2 - (\frac{1}{2})^2 \\ = \frac{1}{2}$$

For attribute Income,

$$\text{Info}(T_{\text{high}}) = 1 - 1^2 - 0^2 = 0$$

$$\text{Info}(T_{\text{low}}) = 1 - (1/3)^2 - (2/3)^2 = 0.444$$

$$\text{Info}(\text{Income}, T) = 1/4 \times \text{Info}(T_{\text{high}}) + 3/4 \times \text{Info}(T_{\text{low}}) = 0.333$$

$$\text{Gain}(\text{Income}, T) = \text{Info}(T) - \text{Info}(\text{Race}, T) = \frac{1}{2} - 0.333 = 0.167$$

For attribute Race,

$$\text{Gain}(\text{Race}, T) = 0.125$$

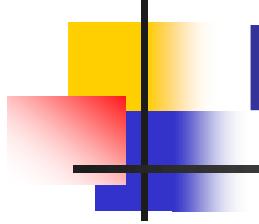
For attribute Income,

$$\text{Gain}(\text{Race}, T) = 0.167$$

For attribute Child,

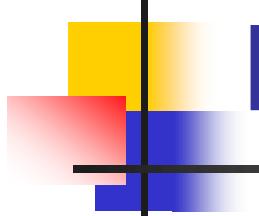
$$\text{Gain}(\text{Child}, T) = ?$$

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no



Decision Trees

- Decision Trees
 - ID3
 - C4.5
 - CART
- Measurement
- How to use the data mining tool



Measurement

Confusion Matrix

- Error Report
- Lift Chart
- Decile-wise lift chart
- Others

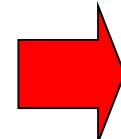
Measurement – Confusion Matrix

Suppose there is a person.

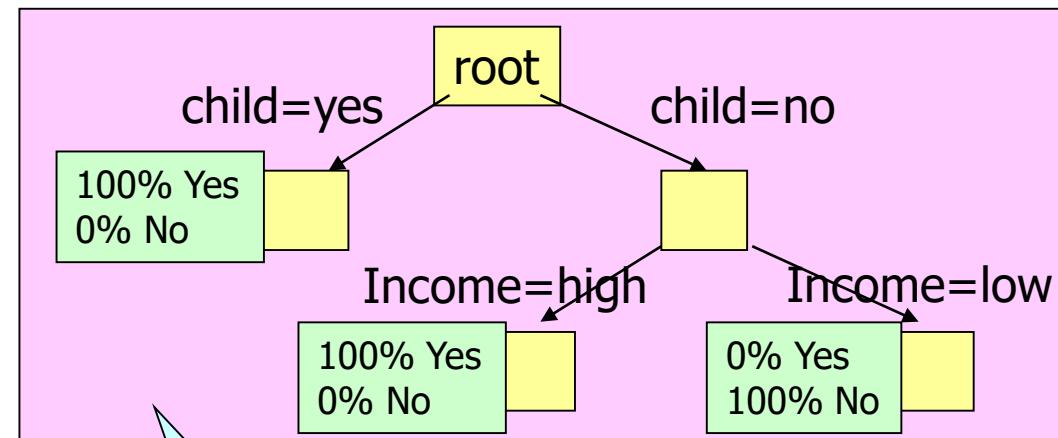
Race	Income	Child	Insurance
white	high	no	?

New set

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no



Training set

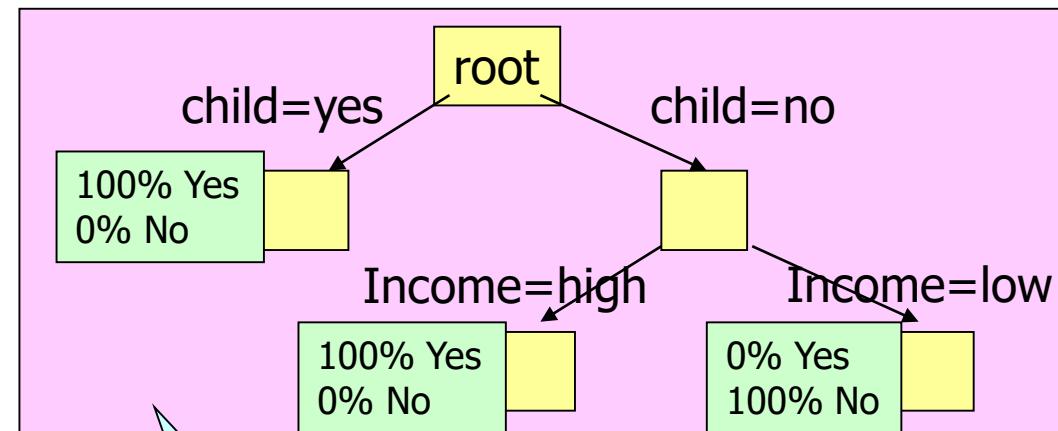


Decision tree

Measurement – Confusion Matrix

Race	Income	Child	Insurance	Predicted
black	high	no	yes	yes
white	high	yes	yes	yes
white	low	yes	yes	yes
white	low	yes	yes	yes
black	low	no	no	no
black	low	no	no	no
black	low	no	no	no
white	low	no	no	no

Training set



Decision tree

Measurement – Confusion

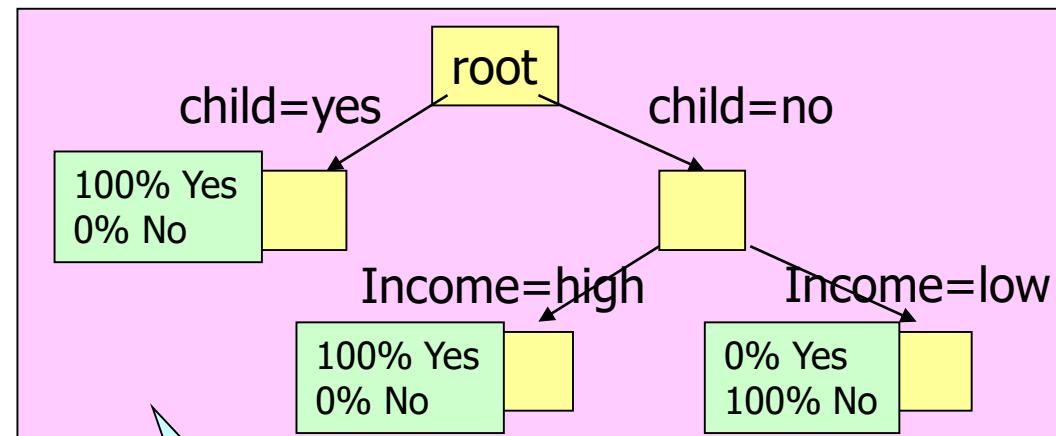
Matrix

Confusion Matrix

		Predicted Class		
		Actual Class	Yes	No
		Yes	4	0
		No	0	4

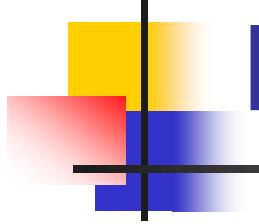
Race	Income	Child	Actual	Predicted
black	high	no	yes	yes
white	high	yes	yes	yes
white	low	yes	yes	yes
white	low	yes	yes	yes
black	low	no	no	no
black	low	no	no	no
black	low	no	no	no
white	low	no	no	no

Is this decision tree “good”?



Training set

Decision tree



Measurement

- Confusion Matrix
- Error Report
- Lift Chart
- Decile-wise lift chart
- Others

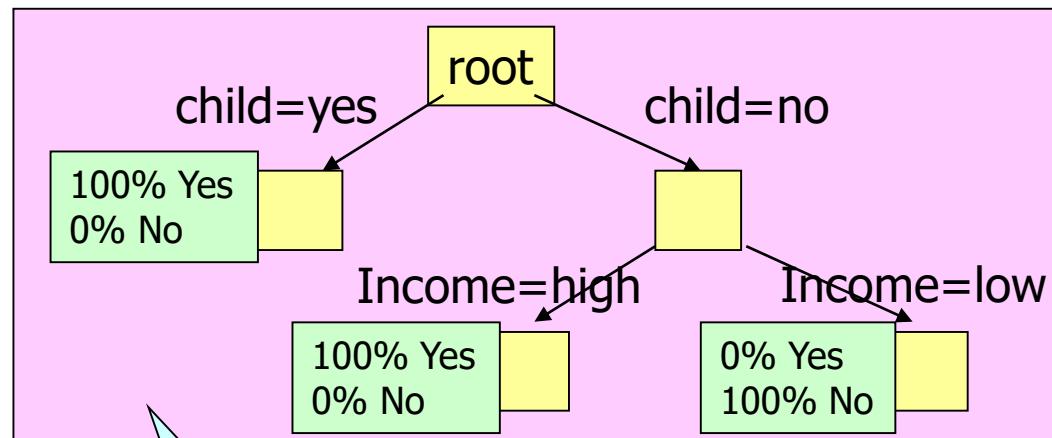
Measures

Error Report

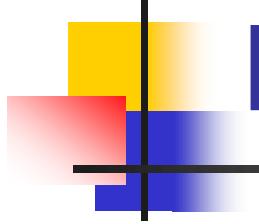
Class	# Cases	# Errors	% Error
Yes	4	0	0.00
No	4	0	0.00
Overall	8	0	0.00

Race	Income	Child	Actual	Predicted
black	high	no	yes	yes
white	high	yes	yes	yes
white	low	yes	yes	yes
white	low	yes	yes	yes
black	low	no	no	no
black	low	no	no	no
black	low	no	no	no
white	low	no	no	no

Training set

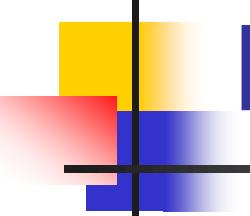


Decision tree



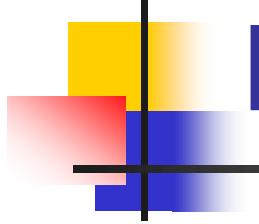
Measurement

- Confusion Matrix
- Error Report
- Lift Chart
- Decile-wise lift chart
- Others



Measurement - Lift Chart

- Lift charts
 - visual aids for measuring model performance
 - consist of a lift curve and a baseline
- The greater the area between the lift curve and the baseline, the better the model.



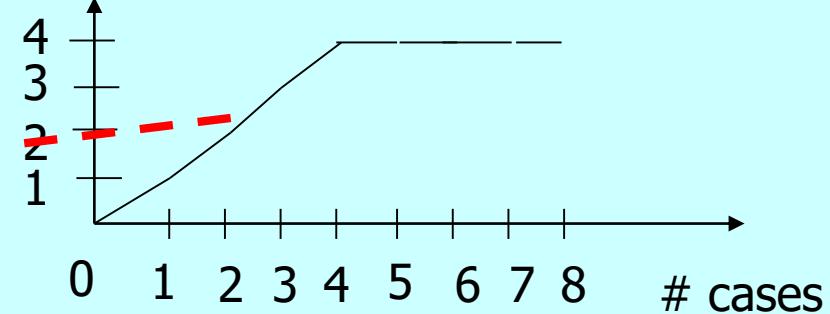
Measurement - Lift Chart

- Lift charts
 - We need to define which value in the target attribute is a “success”
 - In our running example, we can treat “Yes” as a success

Lift Chart

Cumulative

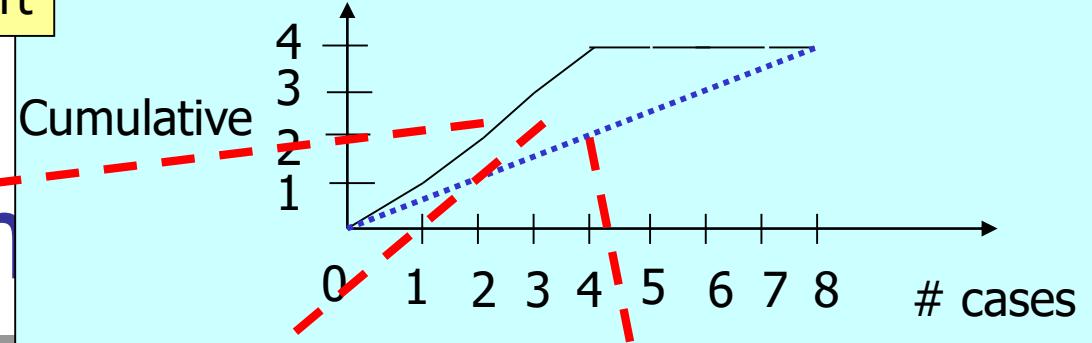
Cumulative Insurance of
actual values (or Lift Curve)



Sort the tuples according to the predicted value where “success” (or “yes”) are sorted at a higher priority and then the actual value

Race	Income	Child	Actual	Predicted	Race	Income	Child	Actual	Predicted
black	high	no	yes	yes	black	high	no	yes	yes
white	high	yes	yes	yes	white	high	yes	yes	yes
white	low	yes	yes	yes	white	low	yes	yes	yes
white	low	yes	yes	yes	white	low	yes	yes	yes
black	low	no	no	no	black	low	no	no	no
black	low	no	no	no	black	low	no	no	no
black	low	no	no	no	black	low	no	no	no
white	low	no	no	no	white	low	no	no	no

Lift Chart



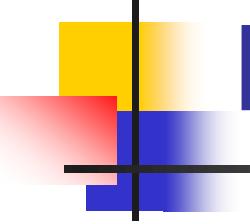
Sort the tuples according to the predicted value which is sorted at a higher rank.

Area between the lift curve values and the baseline

Race	Income	Child	Actual	Predicted
black	high	no	yes	yes
white	high	yes	yes	yes
white	low	yes	yes	yes
black	low	no	no	no
black	low	no	no	no
black	low	no	no	no
white	low	no	no	no

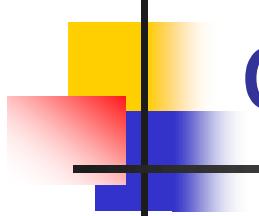
Race	Income	Child	Actual	Predicted
black	high	no	yes	yes
white	high	yes	yes	yes
white	low	yes	yes	yes
white	low	yes	yes	yes
black	low	no	no	no
black	low	no	no	no
black	low	no	no	no
white	low	no	no	no

The larger the area is, the better the classifier is.



Measurement

- Confusion Matrix
- Error Report
- Lift Chart
- Decile-wise lift chart
- Others



Measurement - Decile-wise lift chart

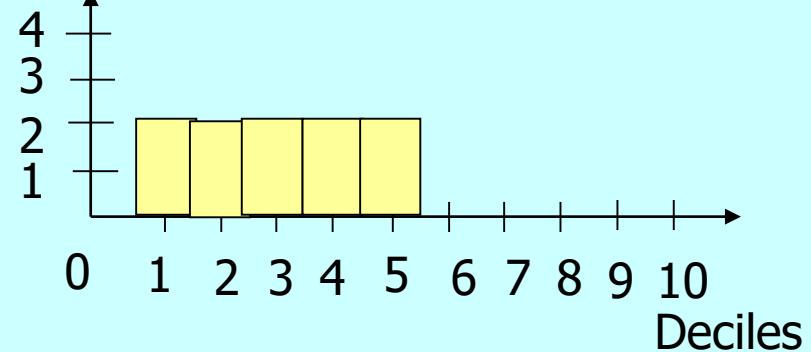
- A **decile** is any of the nine values that divide the sorted data into ten equal parts, so that each part represents 1/10 of the sample or population.
- E.g., 1st decile: the first 10% tuples
- E.g., 2nd decile: the second 10% tuples
- E.g., 3rd decile: the third 10% tuples

Measuring chart

Decile-wise lift chart

Decile mean/
Global mean

$$\text{Global mean} = 4/8 = 0.5$$

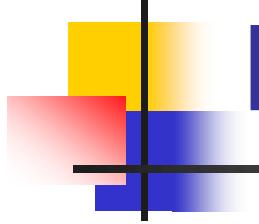


Sort the tuples according to the predicted value where “success” (or “yes”) are sorted at a higher priority and then the actual value

Race	Income	Child	Actual	Predicted
black	high		Decile mean = 1.0	1 st Decile
white	high		Decile mean = 1.0	2 nd Decile
white	low		Decile mean = 1.0	3 rd Decile
white	high		Decile mean = 1.0	4 th Decile
black	high		Decile mean = 0.0	5 th Decile
black	low		Decile mean = 0.0	6 th Decile
black	low		Decile mean = 0.0	7 th Decile
black	low		Decile mean = 0.0	8 th Decile
white	low		Decile mean = 0.0	9 th Decile
				10 th Decile

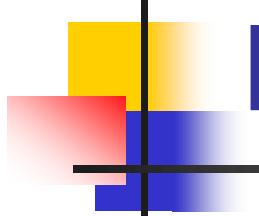
Race	Income	Child	Actual	Predicted
black	12.5%	no	yes	yes
white	25%	gh	yes	yes
white	37.5%	yes	yes	yes
white	50%	w	yes	yes
black	62.5%	no	no	no
black	75%	w	no	no
black	87.5%	no	no	no
white	100%	v	no	no

COMP



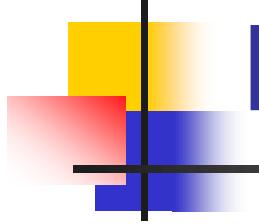
Measurement

- Confusion Matrix
- Error Report
- Lift Chart
- Decile-wise lift chart
- Others



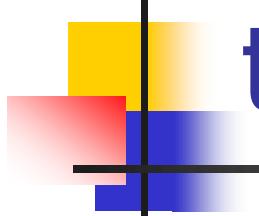
Measurement - Others

- We will discuss them later.
- E.g., Precision, Recall, Specificity, f1-score



Decision Trees

- Decision Trees
 - ID3
 - C4.5
 - CART
- Measurement
- How to use the data mining tool



How to use the data mining tool

- We can use XLMiner for classification (Decision Tree, CART)

How to use the data mining tool

- We have the following 2 versions.
 - ➡ ■ XLMiner Desktop (installed in either the CSE lab machine or your computer)
 - XLMiner Cloud (installed as a plugin in your Office 365 Excel)

Suppose there is a person.

	Race	Income	Child	Insurance
	white	high	no	?

- XLMiner requires that the input data should have the following format.

- Input attributes
 - Numeric
- Target attribute
(or output attribute)
 - Categorical

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Input attributes:

- Categorical

Target attribute:

- Categorical

We can transform from “categorical” to “numeric” by XLMiner Transformation Tool first

Suppose there is a person.

	Race	Income	Child	Insurance
	2	1	1	?

- XLMiner requires that the input data should have the following format.

- Input attributes
 - Numeric
- Target attribute
(or output attribute)
 - Categorical

Race	Income	Child	Insurance
1	1	1	yes
2	1	2	yes
2	2	2	yes
2	2	2	yes
1	2	1	no
1	2	1	no
1	2	1	no
2	2	1	no

Input attributes:

- Categorical

Target attribute:

- Categorical

We can transform from “categorical” to “numeric” by XLMiner Transformation Tool first

Suppose there is a person.

	Race	Income	Child	Insurance
	2	1	1	?

- How can XLMiner perform the transformation?
- Open “classification-decisionTree.xlsx” in MS Excel in a CSE lab machine

Race	Income	Child	Insurance
1	1	1	yes
2	1	2	yes
2	2	2	yes
2	2	2	yes
1	2	1	no
1	2	1	no
1	2	1	no
2	2	1	no

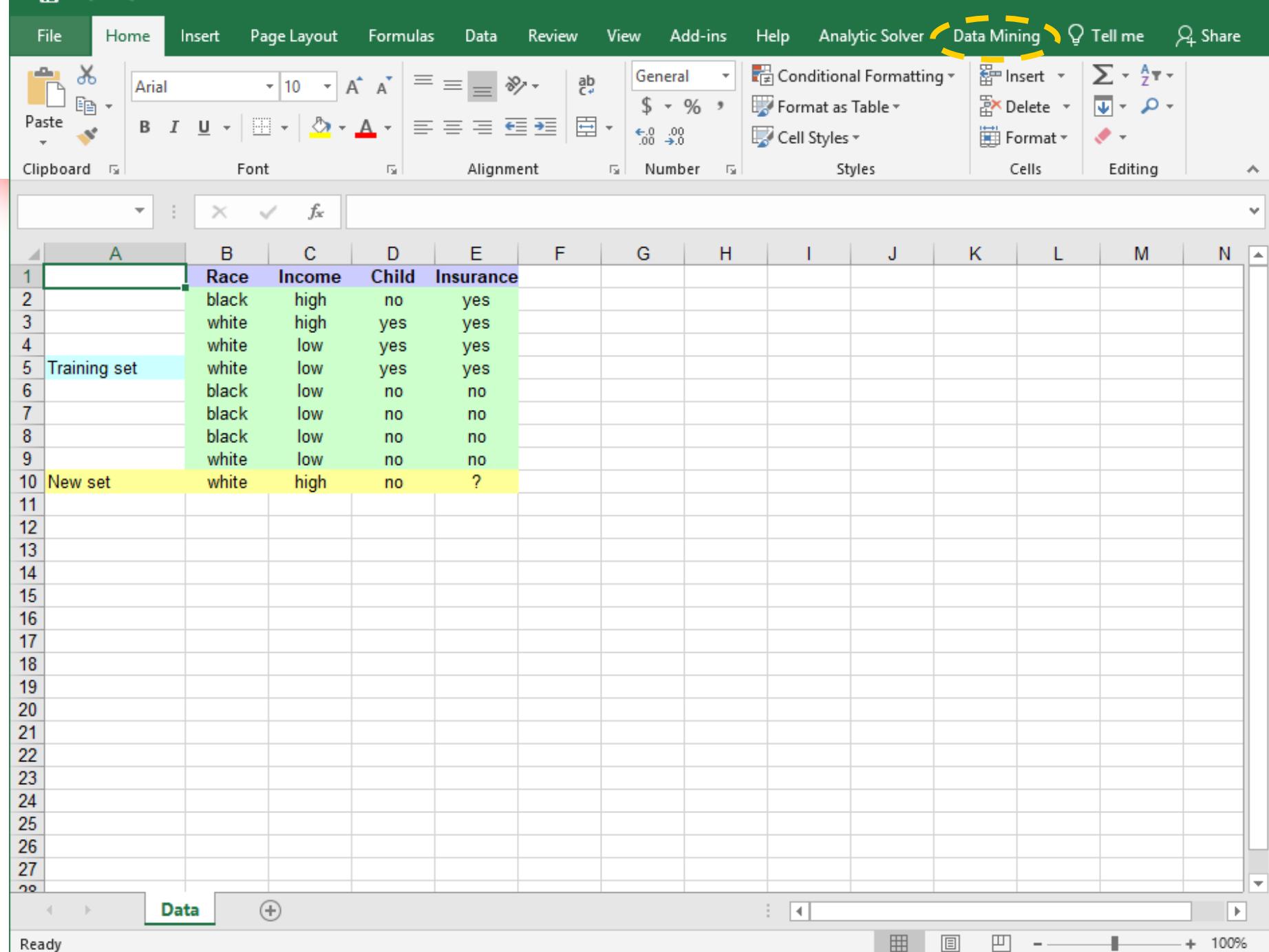
Input attributes:

- Categorical

Target attribute:

- Categorical

We can transform from “categorical” to “numeric” by XLMiner Transformation Tool first





File Home Insert **Page Layout** Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Time Series Partition Classify Predict Associate Data Mining Tools License Help

Model Data Data Analysis

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Race	Income	Child	Insurance									
2		black	high	no	yes									
3		white	high	yes	yes									
4		white	low	yes	yes									
5	Training set	white	low	yes	yes									
6		black	low	no	no									
7		black	low	no	no									
8		black	low	no	no									
9		white	low	no	no									
10	New set	white	high	no	?									
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														



File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Series Data Mining Tools License Help

Model Data

A1

Missing Data Handling

Transform Continuous Data

Transform Categorical Data

Principal Components

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2														
3														
4														
5	Training set	bl	w	w	w									
6		black	low	no	no									
7		black	low	no	no									
8		black	low	no	no									
9		white	low	no	no									
10	New set	white	high	no	?									
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														

Data Mining Tools License Help

A1

F G H I J K L M N

Ready



A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	black	low	no	no										
2	black	low	no	no										
3	black	low	no	no										
4	white	low	no	no										
5	Training set													
6	black	low	no	no										
7	black	low	no	no										
8	white	low	no	no										
9														
10	New set	white	high	no	?									
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														

Data Mining

Missing Data Handling

Transform Continuous Data

Transform Categorical Data

Principal Components

Create Dummies

Create Category Scores

Reduce Categories

A1

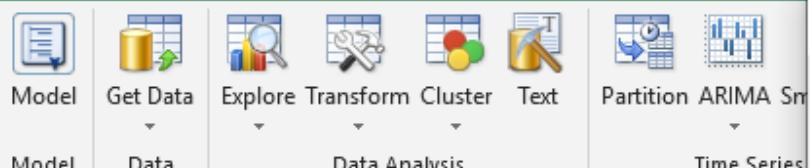
F G H I J K L M N

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

Ready

Data

100%



A1	B	C	D	E	F
1	Race	Income	Child	Insurance	
2	black	high	no	yes	
3	white	high	yes	yes	
4	white	low	yes	yes	
5	white	low	yes	yes	Training set
6	black	low	no	no	
7	black	low	no	no	
8	black	low	no	no	
9	white	low	no	no	
10	New set	white	high	no	?
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					

Create Category Scores

Data Source

Worksheet: Data Workbook: classification-decisionT

Data range: \$A\$1:\$E\$10 #Rows: 10 #Cols: 5

Variables

First Row Contains Headers

Variables
Var1
Var2
Var3
Var4
Var5

>

Variables to be factored

Assign Numbers Options

Assign numbers 1,2,3,... Assign numbers 0,1,2,...

Help OK Cancel

	A	B	C	D	E
1		Race	Income	Child	Insurance
2		black	high	no	yes
3		white	high	yes	yes
4		white	low	yes	yes
5	Training set	white	low	yes	yes
6		black	low	no	no
7		black	low	no	no
8		black	low	no	no
9		white	low	no	no
10	New set	white	high	no	?
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					

Data source

Worksheet

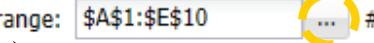
Create Category Scores

Data Source

Worksheet: Data

Workbook: classification-decisionT

Data range: \$A\$1:\$E\$10



#Rows: 0 #Cols: 5

Variables

First Row Contains Headers

Variables

Var1

Var2

Var3

Var4

Var5

Data range

Workbook

Assign Numbers Options

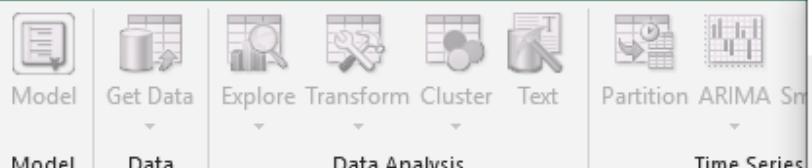
Assign numbers 1,2,3,...

Assign numbers 0,1,2,...

Help

OK

Cancel



	A	B	C	D	E	F
1	Race	Income	Child	Insurance		
2	black	high	no	yes		
3	white	high	yes	yes		
4	white	low	yes	yes		
5	white	low	yes	yes		
6	black	low	no	no		
7	black	low	no	no		
8	black	low	no	no		
9	white	low	no	no		
10	white	high	no	?		
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						

Create Category Scores

Data Source

Worksheet: Data Workbook: classification-decisionT

Data range: \$A\$1:\$E\$10 #Rows: 10 #Cols: 5

Variables

First Row Contains Headers

Variables Select Data Range ? X

Range: \$B\$1:\$E\$10

OK Cancel

Processing...

Assign Numbers Options

Assign numbers 1,2,3,... Assign numbers 0,1,2,...

Help OK Cancel

Variables

	Race	Income	Child	Insurance
1	black	low	yes	yes
2	white	low	yes	yes
3	white	low	no	no
4	white	low	no	no
5	black	low	no	no
6	black	low	no	no
7	white	low	no	no
8	white	high	no	?
9				
10				

Variables

Rows 9

Columns 4

Variables to be factored

First row contains header

Create Category Score

Worksheet: Data

Data range: \$B\$1:\$E\$10

#Rows: 9 #Cols: 4

Variables -

First Row Contains Headers

Variables:

- Race
- Income
- Child
- Insurance

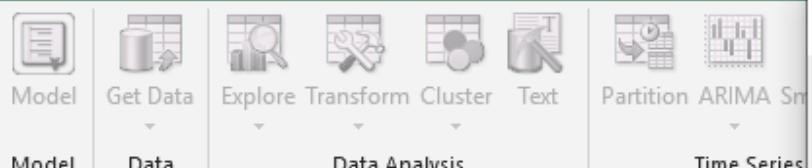
Variables to be factored

Assign Numbers Options

Assign numbers 1,2,3,... Assign numbers 0,1,2,...

Help OK Cancel

The list of variables in the input data range. You can select them as input or output variables using the arrow buttons.



A1	B	C	D	E	F
1	Race	Income	Child	Insurance	
2	black	high	no	yes	
3	white	high	yes	yes	
4	white	low	yes	yes	
5	white	low	yes	yes	Training set
6	black	low	no	no	
7	black	low	no	no	
8	black	low	no	no	
9	white	low	no	no	
10	white	high	no	?	New set
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					

Create Category Scores

Data Source

Worksheet: Data Workbook: classification-decisionT

Data range: \$B\$1:\$E\$10 #Rows: 9 #Cols: 4

Variables

First Row Contains Headers

Variables
Income
Child
Insurance

Variables to be factored
Race

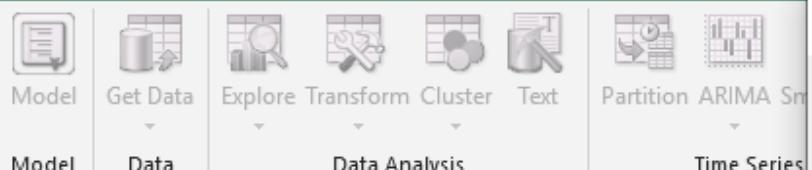
>

Assign Numbers Options

Assign numbers 1,2,3,... Assign numbers 0,1,2,...

Help OK Cancel

Adds or removes the selected variable(s) from the variables list.



A1	B	C	D	E	F
1	Race	Income	Child	Insurance	
2	black	high	no	yes	
3	white	high	yes	yes	
4	white	low	yes	yes	
5	white	low	yes	yes	Training set
6	black	low	no	no	
7	black	low	no	no	
8	black	low	no	no	
9	white	low	no	no	
10	white	high	no	?	New set
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					

Create Category Scores

Data Source

Worksheet: Data Workbook: classification-decisionT

Data range: \$B\$1:\$E\$10 #Rows: 9 #Cols: 4

Variables

First Row Contains Headers

Variables
Income
Child
Insurance

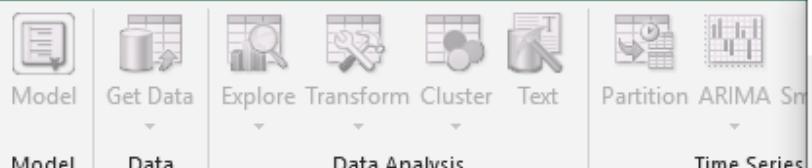
Variables to be factored
Race

Assign Numbers Options

Assign numbers 1,2,3,... Assign numbers 0,1,2,...

Help OK Cancel

The list of variables in the input data range. You can select them as input or output variables using the arrow buttons.



A1	B	C	D	E	F
1	Race	Income	Child	Insurance	
2	black	high	no	yes	
3	white	high	yes	yes	
4	white	low	yes	yes	
5	white	low	yes	yes	Training set
6	black	low	no	no	
7	black	low	no	no	
8	black	low	no	no	
9	white	low	no	no	
10	white	high	no	?	New set
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					

Create Category Scores

Data Source

Worksheet: Data Workbook: classification-decisionT

Data range: \$B\$1:\$E\$10 #Rows: 9 #Cols: 4

Variables

First Row Contains Headers

Variables
Child
Insurance

Variables to be factored
Race
Income

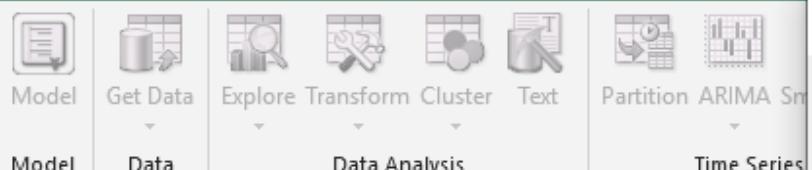
>

Assign Numbers Options

Assign numbers 1,2,3,... Assign numbers 0,1,2,...

Help OK Cancel

Adds or removes the selected variable(s) from the variables list.



A1	B	C	D	E	F
1	Race	Income	Child	Insurance	
2	black	high	no	yes	
3	white	high	yes	yes	
4	white	low	yes	yes	
5	white	low	yes	yes	Training set
6	black	low	no	no	
7	black	low	no	no	
8	black	low	no	no	
9	white	low	no	no	
10	white	high	no	?	New set
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					

Create Category Scores

Data Source

Worksheet: Data Workbook: classification-decisionT

Data range: \$B\$1:\$E\$10 #Rows: 9 #Cols: 4

Variables

First Row Contains Headers

Variables
Child
Insurance

Variables to be factored
Race
Income

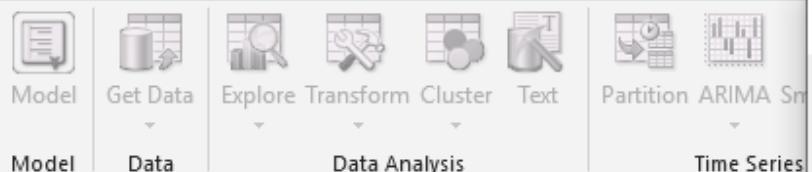
>

Assign Numbers Options

Assign numbers 1,2,3,... Assign numbers 0,1,2,...

Help OK Cancel

The list of variables in the input data range. You can select them as input or output variables using the arrow buttons.



A1	B	C	D	E	F
1	Race	Income	Child	Insurance	
2	black	high	no	yes	
3	white	high	yes	yes	
4	white	low	yes	yes	
5	white	low	yes	yes	Training set
6	black	low	no	no	
7	black	low	no	no	
8	black	low	no	no	
9	white	low	no	no	
10	white	high	no	?	New set
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					

Create Category Scores

Data Source: Worksheet: Data Workbook: classification-decisionT

Data range: \$B\$1:\$E\$10 #Rows: 9 #Cols: 4

Variables: First Row Contains Headers

Variables to be factored: Race, Income, Child

Assign numbers 1, 2, 3, ...

Assign numbers 0, 1, 2, ...

Option

Assign Numbers Options:

- Assign numbers 1,2,3,...
- Assign numbers 0,1,2,...

Help OK Cancel

Adds or removes the selected variable(s) from the variables list.

Inputs

Data	
Workbook	classification-decisionTree.xlsx
Worksheet	Data
Range	\$B\$1:\$E\$10
# Records in the input data	9

Variables

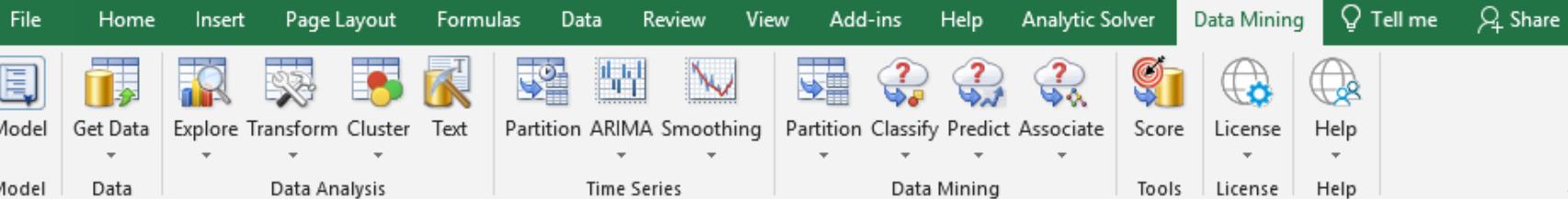
# Selected Variables	3		
Selected Variables	Race	Income	Child

Factorization Parameters

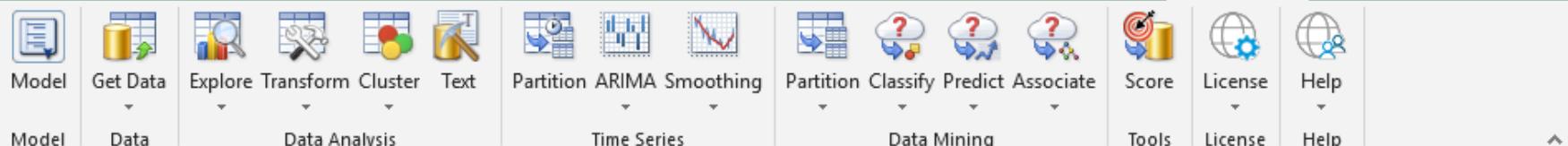
Base Index is Zero	FALSE	
--------------------	-------	--

Transformed Data

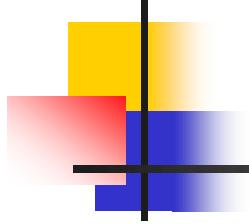
Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

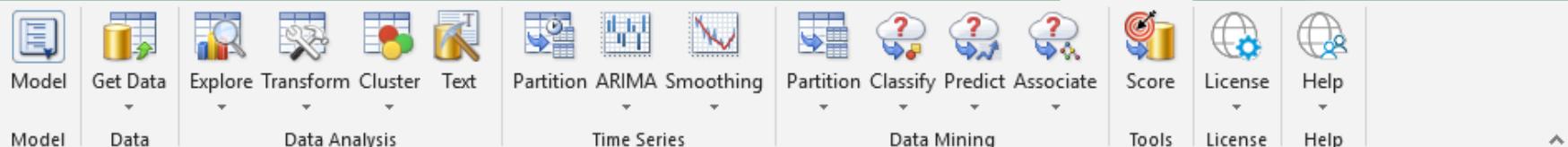


A1									
10	Data								
11	Workbook								classification-decisionTree.xlsx
12	Worksheet								Data
13	Range								\$B\$1:\$E\$10
14	# Records in the input data								9
15									
16									
17									
18	Variables								
19	# Selected Variables								3
20	Selected Variables					Race	Income	Child	
21									
22									
23									
24									
25	Factorization Parameters								
26	Base Index is Zero					FALSE			
27	Record 1	yes	1	1	1				
28	Record 2	yes	2	1	2				
29	Record 3	yes	2	2	2				
30	Record 4	yes	2	2	2				
31	Record 5	no	1	2	1				
32	Record 6	no	1	2	1				
33	Record 7	no	1	2	1				
34	Record 8	no	2	2	1				
35	Record 9	?	2	1	1				
36									
37									



A1	B	C	D	E	F	G	H	I	J	K	L	M
10	Inputs											
11												
12												
13	Data											
14	Workbook	classification-decisionTree.xlsx										
15	Worksheet	Data										
16	Range	\$B\$1:\$E\$10										
17	# Records in the input data	9										
18												
19	Variables											
20	# Selected Variables	3										
21	Selected Variables	Race Income Child										
22												
23	Factorization Parameters											
24	Record ID	Insurance	Race	Income	Child							
25	Record 1	yes	1	1	1							
26	Record 2	yes	2	1	2							
27	Record 3	yes	2	2	2							
28	Record 4	yes	2	2	2							
29	Record 5	no	1	2	1							
30	Record 6	no	1	2	1							
31	Record 7	no	1	2	1							
32	Record 8	no	2	2	1							
33	Record 9	?	2	1	1							
34												
35												
36												
37												

- 
- We have finished the transformation.
 - However, we want to make the classification process easier
 - Thus, we want to “tidy up” the input format for the later process of classification now.



C27 : Record ID

A B C D E F G H I J K L M

25 **Transformed Data**

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

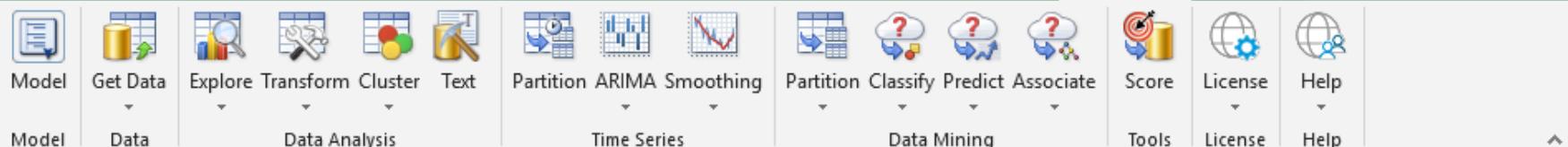
Copy and paste!

Data Factorization

Ready

0

100%



C38 : Record ID

A B C D E F G H I J K L M

Transformed Data

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

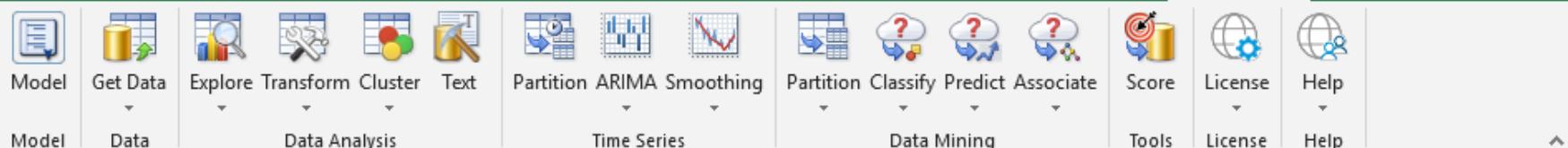


Data Factorization

Select destination and press ENTER or choose Paste

Average: 1.518518519 Count: 50 Sum: 41

100%



A36 : X ✓ f_x

A B C D E F G H I J K L M

25 Transformed Data

26

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

27

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

28

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

29

30

31

32

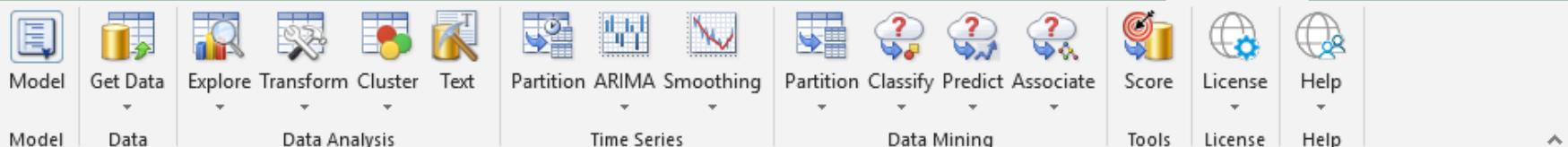
33

34

35

36

37



A36	B	C	D	E	F	G	H	I	J	K	L	M
Transformed Data												
25												
26												
27												
28												
29												
30												
31												
32												
33												
34												
35												
36												
37												
38												
39												
40												
41												
42												
43												
44												
45												
46												
47												
48												
49												
50												
51												
52												
53												
54												
55												
56												
57												
58												
59												
60												
61												
62												
63												
64												
65												
66												
67												
68												
69												
70												
71												
72												
73												
74												
75												
76												
77												
78												
79												
80												
81												
82												
83												
84												
85												
86												
87												
88												
89												
90												
91												
92												
93												
94												
95												
96												
97												
98												
99												
100												

Transformed Data

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

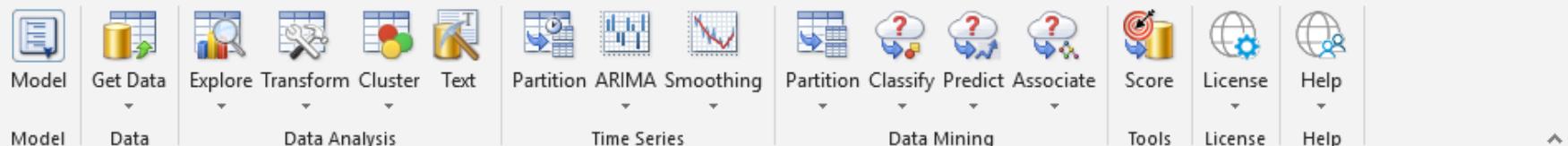
A38 ▾

f_x

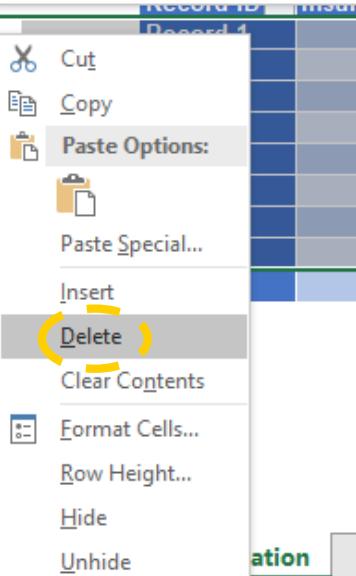
Transformed Data

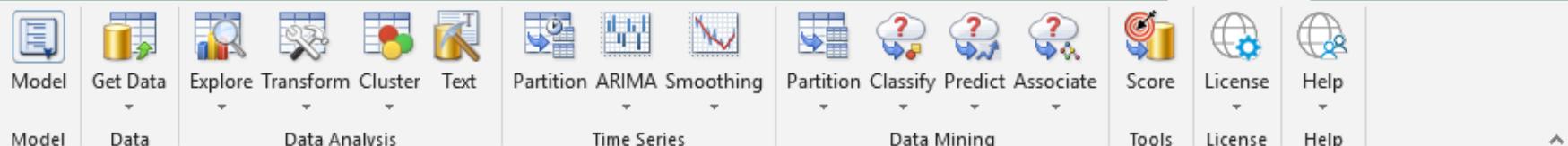
Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1



A38	:	X	✓	fxc											
A	B	C	D	E	F	G	H	I	J	K	L	M			
25	Transformed Data														
26	Record ID	Insurance	Race	Income	Child										
27	Record 1	yes	1	1	1										
28	Record 2	yes	2	1	2										
29	Record 3	yes	2	2	2										
30	Record 4	yes	2	2	2										
31	Record 5	no	1	2	1										
32	Record 6	no	1	2	1										
33			1	2	1										
34			2	2	1										
35															
36															
37															
38	Record 1	yes	1	1	1										
39		yes	2	1	2										
40		yes	2	2	2										
41		yes	2	2	2										
42		no	1	2	1										
43		no	1	2	1										
44		no	1	2	1										
45		no	2	2	1										
46		?	2	1	1										
47															
48															
49															
50															
51															
52															
53															





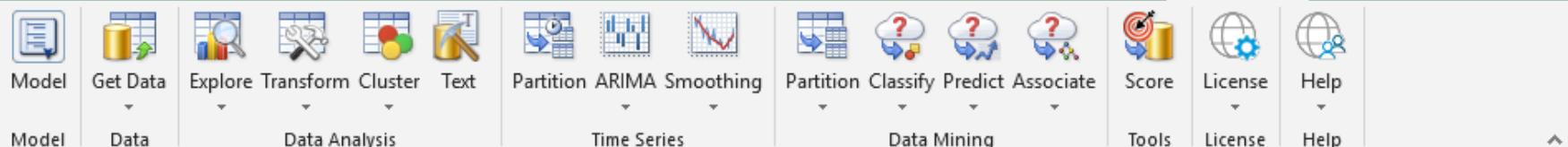
A38 : X ✓ f_x

A B C D E F G H I J K L M

Transformed Data

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1

Record ID	Insurance	Race	Income	Child
Record 9	?	2	1	1



D50

X ✓ f_x

A B C D E F G H I J K L M

Transformed Data

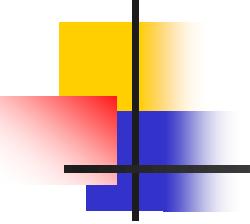
Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1

Record ID	Insurance	Race	Income	Child
Record 9	?	2	1	1

Data

Factorization

+

- 
- Now, we understand how to perform the transformation.
 - We also “tidied up” the data for the process of classification
 - Next, we need to perform the data mining task for classification (Decision Tree, CART)

File

Home

Insert

Page Layout

Formulas

Data

Review

View

Add-ins

Help

Analytic Solve

Data Mining

Conditional Formatting

Format as Table

Cell Styles

Insert

 \sum

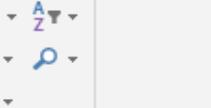
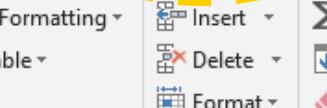
A-Z

Delete

Format



Paste



Editing

D50

Transformed Data

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1

Record ID	Insurance	Race	Income	Child
Record 9	?	2	1	1

Data

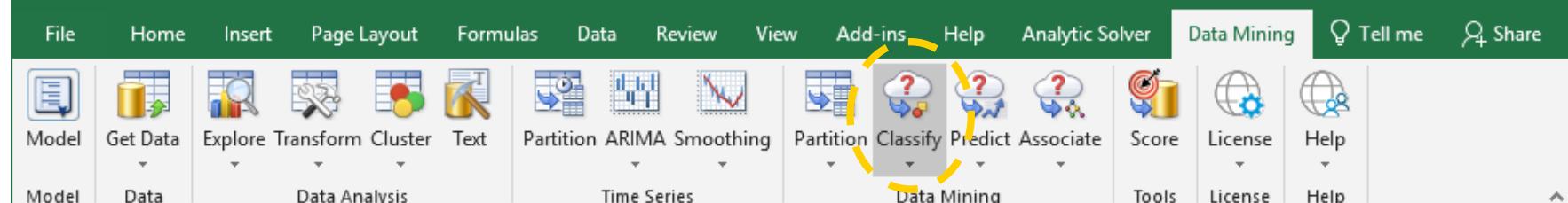
Factorization



Ready

0

100%



D50

Transformed Data

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1

Record ID	Insurance	Race	Income	Child
Record 9	?	2	1	1



Model



Get Data



Explore Transform Cluster



Text



Partition ARIMA Smoothing



Partition



Classify



Predict



Associate



Score



License



Help

License

Help

D50



A

B

C

D

E

F

G

K

L

M

Transformed Data

Record ID	Insurance	Race	Income	Child
Record 1	yes	1	1	1
Record 2	yes	2	1	2
Record 3	yes	2	2	2
Record 4	yes	2	2	2
Record 5	no	1	2	1
Record 6	no	1	2	1
Record 7	no	1	2	1
Record 8	no	2	2	1
Record 9	?	2	1	1

- Find Best Model
- Discriminant Analysis
- Logistic Regression
- k-Nearest Neighbors
- Classification Tree
- Naive Bayes
- Neural Network
- Ensemble

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

Data

Factorization



This screenshot shows the Data Analysis ribbon tab in Microsoft Excel, specifically the Classification Tree dialog box. The dialog is titled "Classification Tree" and contains several tabs: Data, Parameters, and Scoring. The "Data" tab is selected.

The "Data Source" section includes fields for "Worksheet" (set to "Factorization"), "Workbook" (set to "classification-decisionT"), "Data range" (\$C\$27:\$G\$36), "#Columns" (5), "# Rows In" (9), "Training Set" (9), "Validation Set" (0), and "Test Set" (0). A yellow arrow points from the "Data range" field to the "..." button.

The "Variables" section includes a checkbox for "First Row Contains Headers" which is checked. It also contains two lists: "Variables In Input Data" (Record ID, Insurance, Race, Income, Child) and "Selected Variables".

The "Target" section includes fields for "Classes" (Binary Classification), "Number of Classes" (1), "Success Class" (1), and "Success Probability Cutoff" (0.5).

At the bottom of the dialog are buttons for "Help", "Cancel", "< Back", "Next >", and "Finish".

The main Excel window shows a worksheet titled "Transformed Data" containing two tables of data. The first table has 8 rows labeled Record 1 through Record 8, and the second table has 1 row labeled Record 9. The columns are Record ID, Insurance, Race, Income, and Child.

Annotations with callouts point to the "Data Source" (top left), "Worksheet" (top center), "Workbook" (top right), and "Data range" (middle center).

	Record ID	Insurance	Race	Income	Child
Record 1		yes	1	1	1
Record 2		yes	2	1	2
Record 3		yes	2	2	2
Record 4		yes	2	2	2
Record 5		no	1	2	1
Record 6		no	1	2	1
Record 7		no	1	2	1
Record 8		no	2	2	1

	Record ID	Insurance	Race	Income	Child
Record 9		?	2	1	1

classification_decisionTree - Excel

Liu Hao

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$C\$27:\$G\$36 #Columns: 5

Rows In

Training Set: 9 Validation Set: 0 Test Set: 0

Variables

Select Data Range

Range: \$D\$27:\$G\$35

OK Cancel

Selected Variables

Income Child Processing...

Categorical Variables

Output Variable:

Target

Classes Binary Classification

Number of Classes: Success Class: Success Probability Cutoff:

Help Cancel < Back Next > Finish

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Point

COMPI1942

84

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition

Model Data Data Analysis

Rows in Training set

Columns

Worksheet: Factorization

Data range: \$D\$27:\$G\$35

Rows In: 8

#Columns: 4

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Insurance
Race
Income
Child

Selected Variables

Categorical Variables

Output Variable:

Target

Classes

Binary Classification

Number of Classes:

Success Class:

Success Probability Cutoff:

Help Cancel < Back Next > Finish

Ready

Data Factorization

Classification Tree

Transformed Data

	Record ID	Insurance	Race	Income	Child
27	Record 1	yes	1	1	1
28	Record 2	yes	2	1	2
29	Record 3	yes	2	2	2
30	Record 4	yes	2	2	2
31	Record 5	no	1	2	1
32	Record 6	no	1	2	1
33	Record 7	no	1	2	1
34	Record 8	no	2	2	1
35	Record 9	?	2	1	1

85

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Time Series

Data Analysis

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1

30 Record 3

31 Record 4

32 Record 5

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Data Parameters Scoring

Variables

First row contains header

Selected variables

Variables in input data

Output variables

Variables In Input Data

Insurance
Race
Income
Child

Selected Variables

Categorical Variables

Output Variable:

Target

Classes
Number of Classes:

Binary
Success
Success Probability Cut-off:

Help Cancel < Back Next > Finish

Share

100%

86

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Ready

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$27:\$G\$35 #Columns: 4

Rows In

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Insurance
Race
Income
Child

>

Selected Variables

Categorical Variables

>

Output Variable:

>

Target

Classes Binary Classification

Number of Classes: 0 Success Class: yes

Success Probability Cutoff: 0.5

Help Cancel < Back Next > Finish

classification_decisionTree - Excel

Liu Hao

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

100%

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$27:\$G\$35 #Columns: 4

Rows In

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Insurance
Income
Child

>

Selected Variables

Race

>

Categorical Variables

>

Output Variable:

Target

Classes Binary Classification

Number of Classes: 0 Success Class: yes

Success Probability Cutoff: 0.5

Help Cancel < Back Next > Finish

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Ready

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$27:\$G\$35 #Columns: 4

Rows In

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Insurance
Income
Child

Race

Categorical Variables

Output Variable:

Target

Classes Binary Classification

Number of Classes: 0 Success Class: yes

Success Probability Cutoff: 0.5

Help Cancel < Back Next > Finish

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Ready

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$27:\$G\$35 #Columns: 4

Rows In

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Insurance
Child

>

Selected Variables

Race
Income

>

Categorical Variables

>

Output Variable:

>

Target

Classes

Binary Classification

Number of Classes: 0 Success Class: yes Success Probability Cutoff: 0.5

Help Cancel < Back Next > Finish

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Ready

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$27:\$G\$35 #Columns: 4

Rows In

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Insurance
Child

Selected Variables

Race
Income

Categorical Variables

Output Variable:

Target

Classes

Binary Classification

Number of Classes: 0 Success Class: yes Success Probability Cutoff: 0.5

Help Cancel < Back Next > Finish

classification_decisionTree - Excel

Liu Hao

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Ready

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$27:\$G\$35 #Columns: 4

Rows In

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Insurance

>

Selected Variables

Race
Income
Child

>

Categorical Variables

>

Output Variable:

>

Target

Classes

Binary Classification

Number of Classes: 0 Success Class: yes Success Probability Cutoff: 0.5

Help Cancel < Back Next > Finish

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Ready

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$27:\$G\$35 #Columns: 4

Rows In

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Insurance

Selected Variables

Race
Income
Child

Categorical Variables

Output Variable:

Target

Classes

Binary Classification

Number of Classes: 0 Success Class: yes Success Probability Cutoff: 0.5

Help Cancel < Back Next > Finish

classification_decisionTree - Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

	Record ID	Insurance	Race	Income	Child
27	Record 1	yes	1	1	1
28	Record 2	yes	2	1	2
29	Record 3	yes	2	2	2
30	Record 4	yes	2	2	2
31	Record 5	no	1	2	1
32	Record 6	no	1	2	1
33	Record 7	no	1	2	1
34	Record 8	no	2	2	1
35	Record 9	?	2	1	1

Target

Number of classes:

Classification Tree

Data Parameters Scoring

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$27:\$G\$35 #Columns: 4

Rows In

Training Set: 8 Validation Set: 0 Test Set: 0

Variables

First Row Contains Headers

Variables In Input Data

Race
Income
Child

Selected Variables

Categorical Variables

Binary Classification

Success class

Yes

Target Classes

Number of Classes: 2

Binary Classification Success Class: yes Success Probability Cutoff: 0.5

Help < Back Next > Finish

Success probability cutoff

0.5

94

Ready

Factorization

COMPI1942

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smooth

Model Data Data Analysis

B C D E F G

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Preprocessing

Partition Data Rescale Data

Decision Tree: Fitting Growth

Nodes Splits Records in Terminal Nodes

Prune (Using Validation Set) Tree for Scoring

Show Feature Importance

Tree Diagram

Maximum Number of Levels: 7

Trees to Display

Help Cancel < Back Next > Finish

Specifies the minimum number of records any terminal node contains to limit tree growth

100%

Share

Preprocessing

Partition data

Rescale data

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data

B C D E H

25 Transformed Data

26

27 Record ID Insurance Race Income

28 Record 1 yes 1

29 Record 2 yes 2 1

30 Record 3 yes 2 2

31 Record 4 yes 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Tree Growth

Limit number of:

Levels

Nodes

Splits

Limit value

Records in terminal nodes

Decision tree: Display

Show Feature Importance

Tree Diagram

Maximum Number of Levels: 7

Trees to Display

Help Cancel < Back Next > Finish

Specifies the minimum number of records any terminal node contains to limit tree growth

100%

96

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Data Parameters Scoring

Preprocessing

Partition Data Rescale Data

Decision Tree: Fitting

Tree Growth

Limit Number of:

- Levels
- Nodes
- Splits
- Records in Terminal Nodes

Limit Value:

- 10
- 20
- 50
- 1

Advanced

Prior Probability

1

Decision Tree: Model

Prune (Using Validation Set)

Tree for Scoring

Decision tree: Display

Show Feature Importance

Tree Diagram

Maximum Number of Levels: 7

Trees to Display

Help Cancel < Back Next > Finish

Specifies the minimum number of records any terminal node contains to limit tree growth

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Data Parameters Scoring

Preprocessing Partition Data Rescale Data

Decision Tree: Fitting Tree Growth Advanced

Limit Number of: Limit Value:

- Levels 10
- Nodes 20
- Splits 50
- Records in Terminal Nodes 1

Prior Probability

Decision Tree: Model

- Prune (Using Validation Set)
- Tree for Scoring

Decision tree: Display

- Show Feature Importance

Tree Diagram

Maximum Number of Levels: 5

Trees to Display

Help Cancel < Back Next > Finish

Specifies the maximum number of levels in the tree.

5

98

Maximum number of levels (to display)

classification decisionTree ... Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Data Parameters Scoring

Preprocessing

Partition Data Rescale Data

Decision Tree: Fitting

Tree Growth Advanced

Limit Number of: Limit Value:

Levels 10
 Nodes 20
 Splits
 Records in T Prior Probability

Select Trees to Display

Tree Type

Fully Grown
 Best Pruned
 Minimum Error
 User Specified

Decision Tree: Model

Prune (Using V)
Tree for Scoring

Decision Tree: Display

Show Feature

Number of Decision Nodes:

Tree Diagram

Maximum Number of Levels: 5

Trees to Display

Done

If checked, the full tree will be used.

Help Cancel < Back Next > Finish

100%

99

classification decisionTree ... Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization +

Classification Tree

Data Parameters Scoring

Preprocessing

Partition Data Rescale Data

Decision Tree: Fitting

Tree Growth

Limit Number of: Levels Nodes Splits Records in Terminal Nodes

Limit Value: 10 20 50 1

Advanced Prior Probability

Decision Tree: Model

Prune (Using Validation Set) Tree for Scoring

Decision tree: Display

Show Feature Importance

Tree Diagram

Maximum Number of Levels: 5 Trees to Display

Help Cancel < Back Next > Finish

100%

Share

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Score Training Data

Detailed Report

Summary Report

Lift Charts

In Worksheet In Database

Help Cancel < Back Next > Finish

If checked, scoring can be performed on new data from worksheet.

100% 101

Score training data

Detailed report

Summary report

Lift charts

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Data Parameters

Score Training Data

Detailed Report

Summary Report

Lift Charts

Score New Data

In Worksheet

Score new data

In worksheet

Help Cancel < Back Next > Finish

If checked, scoring can be performed on new data from worksheet.

100%

102

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text

Model Data

Data Analysis

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Data Parameters Scoring New Data (WS)

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$C\$27:\$G\$36 #Rows: 9 #Cols: 5

Variables

First Row Contains Headers

Variables In New Data

Record ID
Insurance
Race
Income
Child

Scale Variables In Input Data

Race
Income
Child

Categorical Variables In Input Data

Match Selected
Unmatch Selected
Unmatch All
Match By Name
Match Sequentially

Help Cancel < Back Next > Finish

The worksheets in the currently selected workbook.

100%

Share

Data source

Worksheet

Data range

Workbook

classification_decisionTree - Excel

Liu Hao

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

D37

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Data Parameters Scoring New Data (WS)

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$C\$27:\$G\$36 #Rows: 9 #Cols: 5

Variables

First Row Contains Headers

Variables In New Data

Select Data Range

Range: \$D\$37:\$G\$38

OK Cancel

Scale Variables In Input Data

Categorical Variables In Input Data

Match Selected Unmatch Selected Unmatch All Match By Name Match Sequentially

Help Cancel < Back Next > Finish

The worksheets in the currently selected workbook.

Point

100%

104

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

New Data

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$37:\$G\$38 #Rows: 1 #Cols: 4

Variables

First Row Contains Headers

Variables In New Data

- Insurance
- Race
- Income
- Child

Scale Vari Input Data

Match Selected Unmatch Selected Unmatch All Match By Name Match Sequentially

Help Cancel < Back Next > Finish

The worksheets in the currently selected workbook.

Rows in data 1

Columns in data 4

First row contains headers

100%

105

classification decisionTree Excel

File Home Insert Page Layout Formulas Data Review View

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition

Model Data Data Analysis Time Series

B C D E F G H

25 Transformed Data

26

27 Record ID Insurance Race Income Child

28 Record 1 yes 1 1 1

29 Record 2 yes 2 1 2

30 Record 3 yes 2 2 2

31 Record 4 yes 2 2 2

32 Record 5 no 1 2 1

33 Record 6 no 1 2 1

34 Record 7 no 1 2 1

35 Record 8 no 2 2 1

36

37 Record ID Insurance Race Income Child

38 Record 9 ? 2 1 1

39

40

41

42

43

44

45

46

47

48

49

50

51

Data Factorization

Classification Tree

Data Parameters Scoring New Data (WS)

Data Source

Worksheet: Factorization Workbook: classification-decisionT

Data range: \$D\$37:\$G\$38 #Rows: 1 #Cols: 4

Variables

First Row Contains Headers

Variables In New Data

Insurance

Scale Variables In Input Data

Race<-->Race
Income<-->Income
Child<-->Child

Categorical Variables In Input Data

Match Selected Unmatch Selected Unmatch All Match By Name Match Sequentially

Help Cancel < Back Next > Finish

Matches all the same name variables from the new data variable list to input data variable list.

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

Inputs

Data
Workbook classification-decisionTree.xlsx
Worksheet Factorization
Data Range \$D\$27:\$G\$35
Records 8

Variables
Variables 3
Scale Variables Race Income Child
Categorical Variables
Output Variable Insurance

Rescaling: Fitting Parameters
Rescale Data? FALSE

Decision Tree: Fitting Parameters
Limit min # records in leaves 1

Decision Tree Classification: Fitting Parameters
Prior Probability Calculation EMPIRICAL

Decision Tree: Model Parameters
Prune? FALSE
Scoring tree type Fully grown

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

108

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

10

Data

Workbook	classification-decisionTree.xlsx
Worksheet	Factorization
Data Range	\$D\$27:\$G\$35
# Records	8

11

12

13

14

15

16

17

18

Variables

# Variables	3		
Scale Variables	Race	Income	Child
Categorical Variables			
Output Variable	Insurance		

19

20

21

22

23

24

25

26

27

Decision Tree: Fitting Parameters

Limit min # records in leaves	1
-------------------------------	---

28

29

30

Decision Tree Classification: Fitting Parameters

Prior Probability Calculation	EMPIRICAL
-------------------------------	-----------

31

32

33

34

35

Decision Tree: Model Parameters

Prune?	FALSE
Scoring tree type	Fully grown

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

109

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 : X ✓ fx

A B C D E F G H I J K L M N

10 Inputs

11

12 **Data**

Workbook	classification-decisionTree.xlsx
Worksheet	Factorization
Data Range	\$D\$27:\$G\$35
# Records	8

13

14

15

16

17

18 **Variables**

# Variables	3		
Scale Variables	Race	Income	Child
Categorical Variables			
Output Variable	Insurance		

19

20

21

22

23

24 **Rescaling: Fitting Parameters**

Rescale Data?	FALSE
---------------	-------

25

26

27 **Decision Tree: Fitting Parameters**

Limit min # records in leaves	1
-------------------------------	---

28

29

30 **Decision Tree Classification: Fitting Parameters**

Prior Probability Calculation	EMPIRICAL
-------------------------------	-----------

31

32

33 **Decision Tree: Model Parameters**

Prune?	FALSE
Scoring tree type	Fully grown

34

35

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

110

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 : X ✓ fx

A B C D E F G H I J K L M N

10 Inputs

11

12 **Data**

Workbook	classification-decisionTree.xlsx
Worksheet	Factorization
Data Range	\$D\$27:\$G\$35
# Records	8

13

14

15

16

17

18 **Variables**

# Variables	3
Scale Variables	Race Income Child
Categorical Variables	

19

20

21

22 **Rescaling: Fitting Parameters**

23 Rescale Data? FALSE

24

25

26

27 **Decision Tree: Fitting Parameters**

28 Limit min # records in leaves 1

29

30 Prior Probability Calculation EMPIRICAL

31

32

33 **Decision Tree: Model Parameters**

34 Prune? FALSE

35 Scoring tree type Fully grown

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

COMPI1942 III

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 : X ✓ fx

A B C D E F G H I J K L M N

10 Inputs

11

12 **Data**

Workbook	classification-decisionTree.xlsx
Worksheet	Factorization
Data Range	\$D\$27:\$G\$35
# Records	8

13

14

15

16

17

18 **Variables**

# Variables	3		
Scale Variables	Race	Income	Child
Categorical Variables			
Output Variable	Insurance		

19

20

21

22

23

24 **Rescaling: Fitting Parameters**

Rescale Data?	FALSE
---------------	-------

25

26

27 **Decision Tree: Fitting Parameters**

Limit min # records in leaves	1
-------------------------------	---

28

29

30 **Decision Tree Classification: Fitting Parameters**

Prior Probability Calculation	EMPIRICAL
-------------------------------	-----------

31

32

33 **Decision Tree: Model Parameters**

Prune?	FALSE
Scoring tree type	Fully grown

34

35

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

112

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
28			Limit min # records in leaves 1											
29														
30			Decision Tree Classification: Fitting Parameters											
31			Prior Probability Calculation EMPIRICAL											
32														
33			Decision Tree: Model Parameters											
34			Prune? FALSE											
35			Scoring tree type Fully grown											
36														
37			Decision Tree Classification: Model Parameters											
38			# Classes 2											
39			Success Class yes											
40			Success Probability 0.5											
41														
42			Decision Tree: Reporting Parameters											
43			Trees to draw Fully grown											
44			# Max levels to display 5											
45			Show feature importance? FALSE											
46														
47			Output Options											
48			Summary report of scoring on training data											
49			Detailed report of scoring on training data											
50			Lift charts on training data											
51			New worksheet data scores											
52														
53														

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

113

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

28
29
30 Prior Probability Calculation EMPIRICAL
31
32
33 Decision Tree Classification: Fitting Parameters
34
35 Prune? FALSE
36 Scoring tree type Fully grown
37
38 # Classes 2
39 Success Class yes
40 Success Probability 0.5
41
42 Decision Tree: Reporting Parameters
43 Trees to draw Fully grown
44 # Max levels to display 5
45 Show feature importance? FALSE
46
47 Output Options
48 Summary report of scoring on training data
49 Detailed report of scoring on training data
50 Lift charts on training data
51 New worksheet data scores
52
53

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

114

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
28			Limit min # records in leaves		1									
29														
30	Decision Tree Classification: Fitting Parameters													
31	Prior Probability Calculation	EMPIRICAL												
32														
33	Decision Tree: Model Parameters													
34	Prune?	FALSE												
35	Scoring tree type	Fully grown												
36														
37	Decision Tree Classification: Model Parameters													
38	# Classes	2												
39	Success Class	yes												
40	Success Probability	0.5												
41														
42	Decision Tree: Reporting Parameters													
43	Trees to draw	Fully grown												
44	# Max levels to display	5												
45	Show feature importance?	FALSE												
46														
47	Output Options													
48	Summary report of scoring on training data													
49	Detailed report of scoring on training data													
50	Lift charts on training data													
51	New worksheet data scores													
52														
53														

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

COMPI1942

115

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
28			Limit min # records in leaves		1									
29			Decision Tree Classification: Fitting Parameters											
30			Prior Probability Calculation		EMPIRICAL									
31			Decision Tree: Model Parameters											
32			Decision Tree Classification: Model Parameters											
33			# Classes		2									
34			Success Class		yes									
35			Success Probability		0.5									
36			Decision Tree: Reporting Parameters											
37			Trees to draw		Fully grown									
38			# Max levels to display		5									
39			Show feature importance?		FALSE									
40			Detailed report of scoring on training data											
41			Lift charts on training data											
42			New worksheet data scores											
43			Data Factorization	CT_Output	CT_FullTree	CT_TrainingScore	...	+	:					
44			Ready											

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
28			Limit min # records in leaves		1									
29														
30	Decision Tree Classification: Fitting Parameters													
31	Prior Probability Calculation	EMPIRICAL												
32														
33	Decision Tree: Model Parameters													
34	Prune?	FALSE												
35	Scoring tree type	Fully grown												
36														
37	Decision Tree Classification: Model Parameters													
38	# Classes	2												
39	Success Class	yes												
40	Success Probability	0.5												
41														
42	Decision Tree: Reporting Parameters													
43	Trees to draw	Fully grown												
44	# Max levels to display	5												
45	Show feature importance?	FALSE												
46														
47	Output Options													
48	Summary report of scoring on training data													
49	Detailed report of scoring on training data													
50	Lift charts on training data													
51	New worksheet data scores													
52														
53														

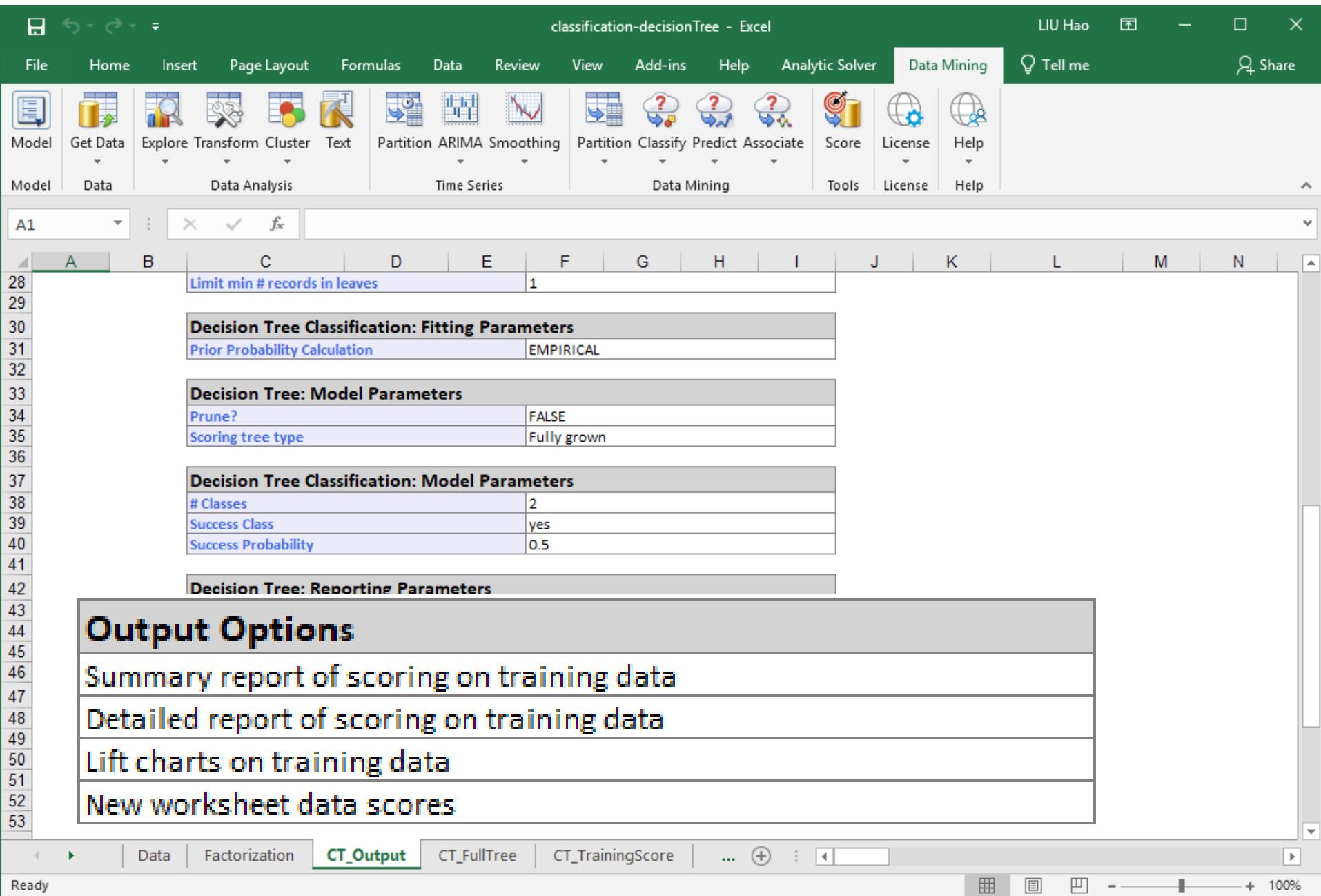
Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

COMPI1942

100%

117



classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
28			Limit min # records in leaves		1									
29			Decision Tree Classification: Fitting Parameters											
30			Prior Probability Calculation		EMPIRICAL									
31			Decision Tree: Model Parameters											
32			Prune?		FALSE									
33			Scoring tree type		Fully grown									
34			Decision Tree Classification: Model Parameters											
35			# Classes		2									
36			Success Class		yes									
37			Success Probability		0.5									
38			Decision Tree: Reporting Parameters											
39			Trees to draw		Fully grown									
40			# Max levels to display		5									
41			Show feature importance?		FALSE									
42			Output Options											
43			Summary report of scoring on training data											
44			Detailed report of scoring on training data											
45			Lift charts on training data											
46			New worksheet data scores											
47														
48														
49														
50														
51														
52														
53														

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... + Ready

100%

COMPI1942 119

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

A B C D E F G H I J K L M N

46

47 **Output Options**

Summary report of scoring on training data

Detailed report of scoring on training data

Lift charts on training data

New worksheet data scores

48

49

50

51

52

53

54 **Training Log (Growing the full tree using training data)**

55

56 # Decision Nodes ▾ Error Rate ▾

57 0 0.5

58 1 0.125

59 2 2.77556E-17

60

61

62

63

64

65

66

67

68

69

70

71

72

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

120

COMPI1942

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

A B C D E F G H I J K L M N

46

47 **Output Options**

48 Summary report of scoring on training data

49 Detailed report of scoring on training data

50 Lift charts on training data

51 New worksheet data scores

52

53 # Decision Nodes ▾ Error Rate ▾

54

# Decision Nodes	Error Rate
0	0.5
1	0.125
2	2.77556E-17

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

121

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

A B C D E F G H I J K L M N

46

47 **Output Options**

48 Summary report of scoring on training data

49 Detailed report of scoring on training data

50 Lift charts on training data

51 New worksheet data scores

52

53

54 **Training Log (Growing the full tree using training data)**

55

56 # Decision Nodes ▾ Error Rate ▾

57 0 0.5

58 1 0.125

59 2 2.77556E-17

60

61

62

63

64

65

66

67

68

69

70

71

72

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

122

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing

Model Data Data Analysis Time Series

Partition Classify Predict Associate Score License Help

Tools License Help

F58	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
2															
3															
4	New: Classification Detail	Inputs		PMML Model											
5	Training: Classification St	Training: Classification													
6															
7															
8															
9															
10															
11															
12															
13				classification-decisionTree.xlsx											
14				Factorization											
15				\$D\$27:\$G\$35											
16				8											
17															
18															
19				3											
20				Race	Income	Child									
21															
22				Insurance											
23															
24	parameters														
25				FALSE											
26															
27	ing Parameters														

Elapsed Times in Milliseconds			
Data Reading Time	Algorithm Time	Report Time	Total
1	5	5	11

classification-decisionTree - Excel

LIU, Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing

Model Data Data Analysis Time Series

Partition Classify Predict Associate Score License Help

Data Mining Tools License Help

F58

D E F G H I J K L M N O P Q R

2

3

4 New: Classification Detail Inputs PMML Model

5 Training: Classification St Training: Classification

6

7

8

9

10

11

12

13 classification-decisionTree.xlsx

14 Factorization

15 \$D\$27:\$G\$35

16 8

17

18

19 3

20 Race Income Child

21

22 Insurance

23

24 parameters

25 FALSE

26

27 Parameters

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

100%

124

COMPI194Z

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing

Model Data Data Analysis Time Series

Partition Classify Predict Associate Score License Help

Tools License Help

F58												
	D	E	F	G	H	I	J	K	L	M	N	O
2												
3												
4	New: Classification Detail	Inputs		PMML Model								
5	Training: Classification St	Training: Classification										
6												
7												
8												
9												
10												
11												
12												
13				classification-decisionTree.xlsx								
14				Factorization								
15				\$D\$27:\$G\$35								
16				8								
17												
18												
19				3								
20		Race	Income	Child								
21												
22		Insurance										
23												
24	parameters											
25				FALSE								
26												
27	ing Parameters											
	Data	Factorization	CT_Output	CT_FullTree		CT_TrainingScore	...	+	:			

classification-decisionTree - Excel

LIU Hao

Data Mining Tell me Share

File Home Insert Page Layout Formulas Data Review View Add-ins Analytic Solver Data Mining

Model Get Data Explore Model Data A1

A B Data M Output Fully Grown Training: C

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Full-Grown Tree

Nodes

- Decision (Blue Circle)
- Terminal (Orange Square)
- Collapsed (Orange Square with black border)

Links

— x — # Records

Tree Info

Tree Height: 3
Nodes: 5

Child 1.5

Income 1.5

yes

no

1 4

Fully Grown

ases Response Node

ases	Response	Node
0	yes	
0	no	
0	yes	
0	yes	
0	no	

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... +

Ready

126

COMP1942

```
graph TD; Income((Income 1.5)) -- 1 --> Yes1[yes]; Income -- 4 --> No1[no]; Yes1 --- Val5[5]; No1 --- Val3[3];
```

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share

Full-Grown Tree

Nodes

Child

Child_{ord} <= 1.5

Child_{ord} > 1.5

of cases in this branch = 5

of cases in this branch = 3

Income

Income_{ord} <= 1.5

Income_{ord} > 1.5

of cases in this branch = 1

of cases in this branch = 4

Tree Info

Collapse All

Expand All

Fully Grown Training: C

Elaps Data I

Response Node

yes 0

no 0

yes 0

yes 0

no 0

Full

1.5

5

3

1.5

yes

1

4

yes

no

127

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... + : ← →

Ready

COMP1942

100%

The diagram illustrates a classification decision tree. The root node is labeled 'Child' with value '1.5'. It splits into two branches: 'Child_{ord} <= 1.5' (left) and 'Child_{ord} > 1.5' (right). The left branch leads to a terminal node labeled '# of cases in this branch = 5'. The right branch leads to another node labeled 'Income' with value '1.5', which further splits into 'Income_{ord} <= 1.5' (left) and 'Income_{ord} > 1.5' (right). The leftmost branch from this node leads to a terminal node labeled '# of cases in this branch = 1'. The rightmost branch leads to a terminal node labeled '# of cases in this branch = 4'. Each terminal node contains an orange square labeled 'yes' or 'no'.

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Analytic Solver Data Mining Tell me... Share

Model Get Data Explore Model Data A1 A B Data M Output Fully Grown Training: C

Full-Grown Tree

Nodes

- Decision (Blue Circle)
- Terminal (Orange Square)
- Collapsed (Orange Square with black border)

Links

— x — # Records

Tree Info

Tree Height: 3
Nodes: 5

Child 1.5

Income 1.5

yes 5

yes 1

no 4

yes 3

Fully Grown

ases Response Node

ases	Response	Node
0	yes	
0	no	
0	yes	
0	yes	
0	no	

Data Factorization CT_Output CT_FullTree CT_TrainingScore ... + Ready

100% 128

COMP1942

```
graph TD; Root((Income 1.5)) -- 5 --> Yes1[yes]; Root -- 4 --> No1[no]; Yes1 -- 5 --> Yes2[yes]; Yes1 -- 3 --> No2[no]; No1 -- 1 --> Yes3[yes]; No1 -- 4 --> No3[no];
```

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 B C D E F G H I J K L M

Data Mining: Classification Tree

Output Navigator

Fully Grown Tree Rules	New: Classification Details	Inputs	PMM Model
Training: Charts	Training: Classification Summary	Training: Classification Details	

Elapsed Times in Milliseconds

Data Reading Time	Algorithm Time
45	37

Fully Grown Tree Rules (using Training Data)

Node ID	Parent ID	Left Child ID	Right Child ID	Split Var	Split Value/Set	Training Cases	Validation Cases	Response	Node Type
1	N/A	2	3	Child	1.5	8	0	yes	Decision
2	1	4	5	Income	1.5	5	0	no	Decision
3	1	N/A	N/A	N/A	N/A	3	0	yes	Terminal
4	2	N/A	N/A	N/A	N/A	1	0	yes	Terminal
5	2	N/A	N/A	N/A	N/A	4	0	no	Terminal

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Tra ...

Ready 100%

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Data Analysis

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing

Data Partition Classify Predict Associate Score License Help

Time Series

Data Mining

A1 B C D E F G H I J K L M

1 Data Mining: Classification Tree

2

3 Output Navigator

4 Fully Grown Tree Rules New: Classification Details Inputs PMML Model

5 Training: Charts Training: Classification Summa Training: Classification Details

Elapsed Times in Milliseconds

Data Reading Time Algorithm Time

45 37

6

7

8

9

10 Fully Grown Tree Rules (Using Training Data)

11

Node ID	Parent ID	Left Child ID	Right Child ID	Split Var	Split Value/Set	Training Cases	Validation Cases	Response	Node Type
1	N/A	2	3	Child	1.5	8	0	yes	Decision
2	1	4	5	Income	1.5	5	0	no	Decision
3	1	N/A	N/A	N/A	N/A	3	0	yes	Terminal
4	2	N/A	N/A	N/A	N/A	1	0	yes	Terminal
5	2	N/A	N/A	N/A	N/A	4	0	no	Terminal

19

20

21

22

23

24

25

26

27

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ...

Ready

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 B C D E F G H I J K L M

1 Data Mining: Classification Tree

2

3 Output Navigator

4 Fully Grown Tree Rules New: Classification Details Inputs PMML Model

5 Training: Charts Training: Classification Summary Training: Classification Details

Elapsed Times in Milliseconds

Data Reading Time Algorithm Time

45 37

6

7

8

9

10 Fully Grown Tree Rules (Using Training Data)

Node ID	Parent ID	Left Child ID	Right Child ID	Split Var	Split Value/Set	Training Cases	Validation Cases	Response	Node Type
1	N/A	2	3	Child	1.5	8	0	yes	Decision
2	1	4	5	Income	1.5	5	0	no	Decision
3	1	N/A	N/A	N/A	N/A	3	0	yes	Terminal
4	2	N/A	N/A	N/A	N/A	1	0	yes	Terminal
5	2	N/A	N/A	N/A	N/A	4	0	no	Terminal

11

12

13

14

15

16

17

18

19

20

21

22

23

24

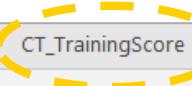
25

26

27

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ...

Ready



classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

A B C D E F G H I J K L M

9

10 Training: Classification Summary

11

12 Confusion Matrix

13 Actual Predicted no yes

14 no 4 0

15 yes 0 4

16

17 Error Report

18 Class # Cases # Errors % Error

19 no 4 0 0

20 yes 4 0 0

21 Overall 8 0 0

22

23 Metrics

24 Metric Value

25 Accuracy (#correct) 8

26 Accuracy (%correct) 100

27 Specificity 1

28 Sensitivity (Recall) 1

29 Precision 1

30 F1 score 1

31 Success Class yes

32 Success Probability 0.5

33

34 Training: Classification Details

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ...

Ready

100%

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

Confusion Matrix

Actual\Predicted	no	yes
no	4	0
yes	0	4

Error Report

Class	# Cases	# Errors	% Error
no	4	0	0
yes	4	0	0
Overall	8	0	0

Accuracy (#correct)	8
Accuracy (%correct)	100
Specificity	1
Sensitivity (Recall)	1
Precision	1
F1 score	1
Success Class	yes
Success Probability	0.5

Training: Classification Details

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ...

Ready

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 A B C D E F G H I J K L M

9

10 Training: Classification Summary

11

12 Confusion Matrix

Actual	Predicted	no	yes
no		4	0
yes		0	4

13

14

15

16

17 Error Report

Class	# Cases	# Errors	% Error
no	4	0	0
yes	4	0	0
Overall	8	0	0

18

19

20

21

22

23 Metrics

Metric	Value
Accuracy (#correct)	8
Accuracy (%correct)	100
Specificity	1
Sensitivity (Recall)	1
Precision	1
F1 score	1
Success Class	yes
Success Probability	0.5

24

25

26

27

28

29

30

31

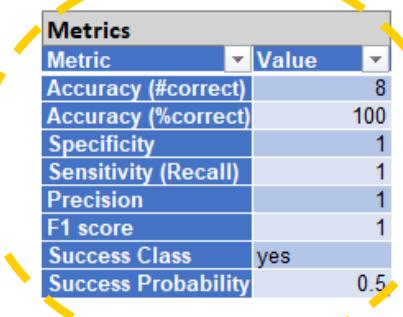
32

33

34 Training: Classification Details

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ...

Ready



100%

The screenshot shows a Microsoft Excel spreadsheet titled "classification-decisionTree - Excel". The ribbon menu is visible at the top, with the "Data Mining" tab selected. The main content area displays three tables under the heading "Training: Classification Summary".

Confusion Matrix:

Actual\Predicted		no	yes
no	4	0	
yes	0	4	

Error Report:

Class	# Cases	# Errors	% Error
0	0	0	0
0	0	0	0
0	0	0	0

Metrics:

Metric	Value
Accuracy (#correct)	8
Accuracy (%correct)	100
Specificity	1
Sensitivity (Recall)	1
Precision	1
F1 score	1
Success Class	yes
Success Probability	0.5

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 : X ✓ fx

A B C D E F G H I J K L M

9

10 Training: Classification Summary

11

12 Confusion Matrix

Actual	Predicted	no	yes
no		4	0
yes		0	4

13

14

15

16

17 Error Report

Class	# Cases	# Errors	% Error
no	4	0	0
yes	4	0	0
Overall	8	0	0

18

19

20

21

22

23 Metrics

Metric	Value
Accuracy (#correct)	8
Accuracy (%correct)	100
Specificity	1
Sensitivity (Recall)	1
Precision	1
F1 score	1
Success Class	yes
Success Probability	0.5

24

25

26

27

28

29

30

31

32

33

34 Training: Classification Details

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ...

Ready 100%

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

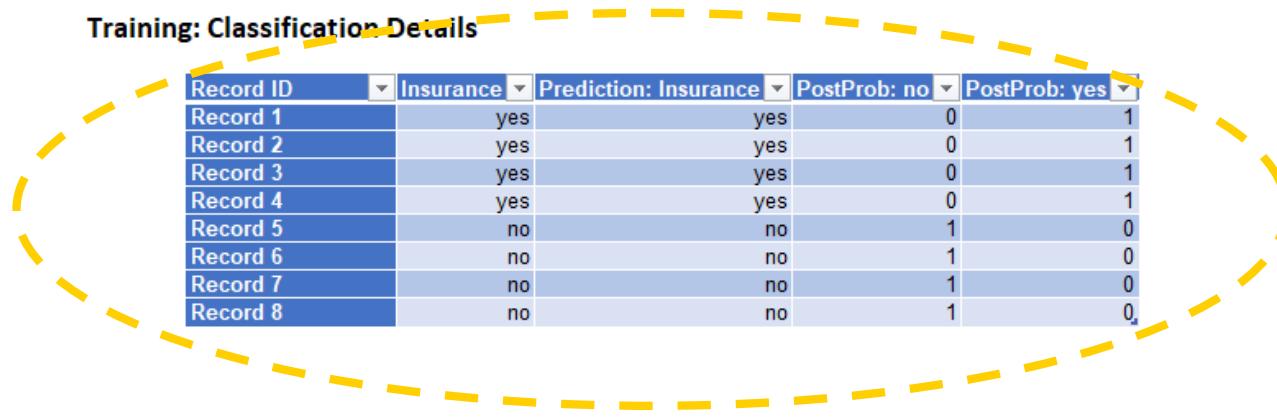
Model Data Data Analysis Time Series Data Mining Tools License Help

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M
27			Specificity	1									
28			Sensitivity (Recall)	1									
29			Precision	1									
30			F1 score	1									
31			Success Class	yes									
32			Success Probability	0.5									
33													
34			Training: Classification Details										
35													
36			Record ID	Insurance	Prediction: Insurance	PostProb: no	PostProb: yes						
37			Record 1	yes	yes	0	1						
38			Record 2	yes	yes	0	1						
39			Record 3	yes	yes	0	1						
40			Record 4	yes	yes	0	1						
41			Record 5	no	no	1	0						
42			Record 6	no	no	1	0						
43			Record 7	no	no	1	0						
44			Record 8	no	no	1	0						
45													
46													
47													
48													
49													
50													
51													
52													
53													

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ... + : < >

Ready



classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

	A	B	C	D	E	F	G	H	I	J	K	L	M
27			Specificity	1									
28			Sensitivity (Recall)	1									
29			Precision	1									
30			F1 score	1									
31			Success Class	yes									
32			Success Probability	0.5									
33	Record ID	Insurance	Prediction: Insurance		PostProb: no		PostProb: yes						
34	Record 1	yes		yes	0		1						
35	Record 2	yes		yes	0		1						
36	Record 3	yes		yes	0		1						
37	Record 4	yes		yes	0		1						
38	Record 5	no		no	1		0						
39	Record 6	no		no	1		0						
40	Record 7	no		no	1		0						
41	Record 8	no		no	1		0						
42													
43													
44													
45													
46													
47													
48													
49													
50													
51													
52													
53													

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ... + : < >

Ready 100%

classification-decisionTree - Excel LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 : X ✓ f_x

27	C	D	E	F	G	H	I	J	K	L	M
28	Specificity	1									
29	Sensitivity (Recall)	1									
30	Precision	1									
31	F1 score	1									
32	Success Class	yes									
33	Success Probability	0.5									

34 Training: Classification Details

Record ID	Insurance	Prediction: Insurance	PostProb: no	PostProb: yes
Record 1	yes	yes	0	1
Record 2	yes	yes	0	1
Record 3	yes	yes	0	1
Record 4	yes	yes	0	1
Record 5	no	no	1	0
Record 6	no	no	1	0
Record 7	no	no	1	0
Record 8	no	no	1	0

Data Factorization CT_Output CT_FullTree CT_TrainingScore CT_Trai ... +

Ready 100%

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

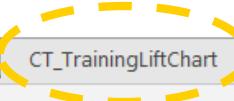
A1

27	C	D	E	F	G	H	I	J	K	L	M
28	Specificity	1									
29	Sensitivity (Recall)	1									
30	Precision	1									
31	F1 score	1									
32	Success Class	yes									
33	Success Probability	0.5									

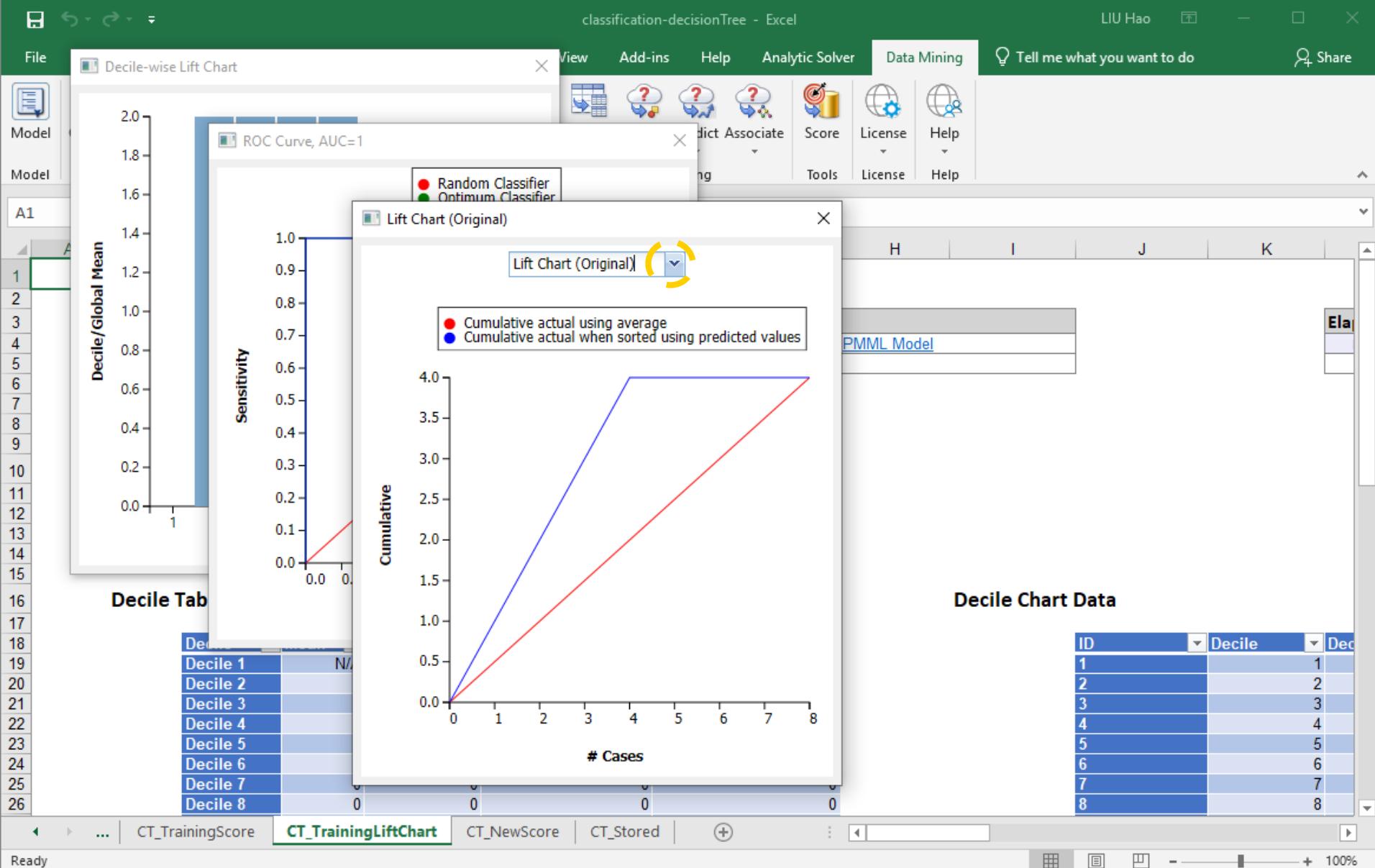
34 Training: Classification Details

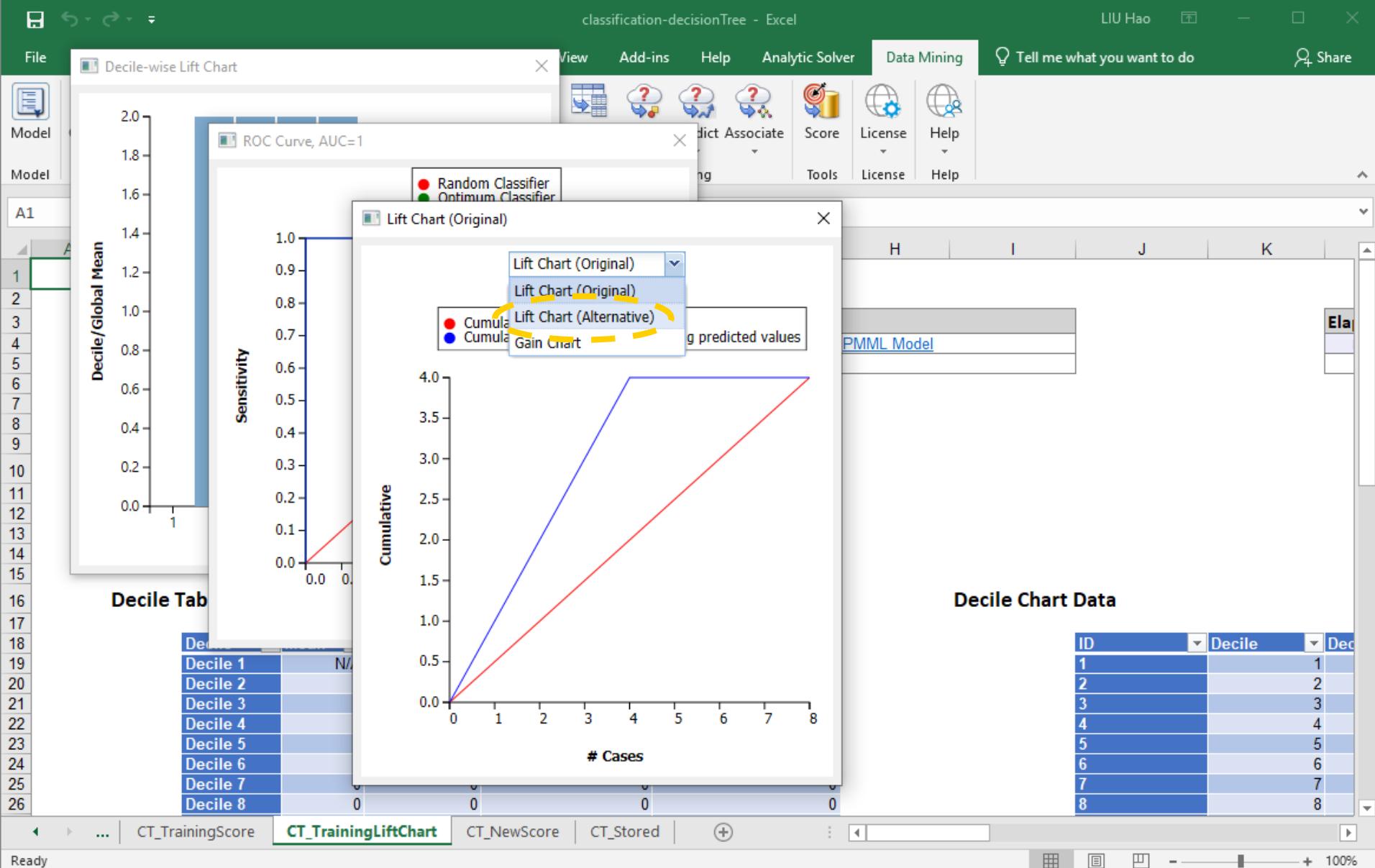
Record ID	Insurance	Prediction: Insurance	PostProb: no	PostProb: yes
Record 1	yes	yes	0	1
Record 2	yes	yes	0	1
Record 3	yes	yes	0	1
Record 4	yes	yes	0	1
Record 5	no	no	1	0
Record 6	no	no	1	0
Record 7	no	no	1	0
Record 8	no	no	1	0

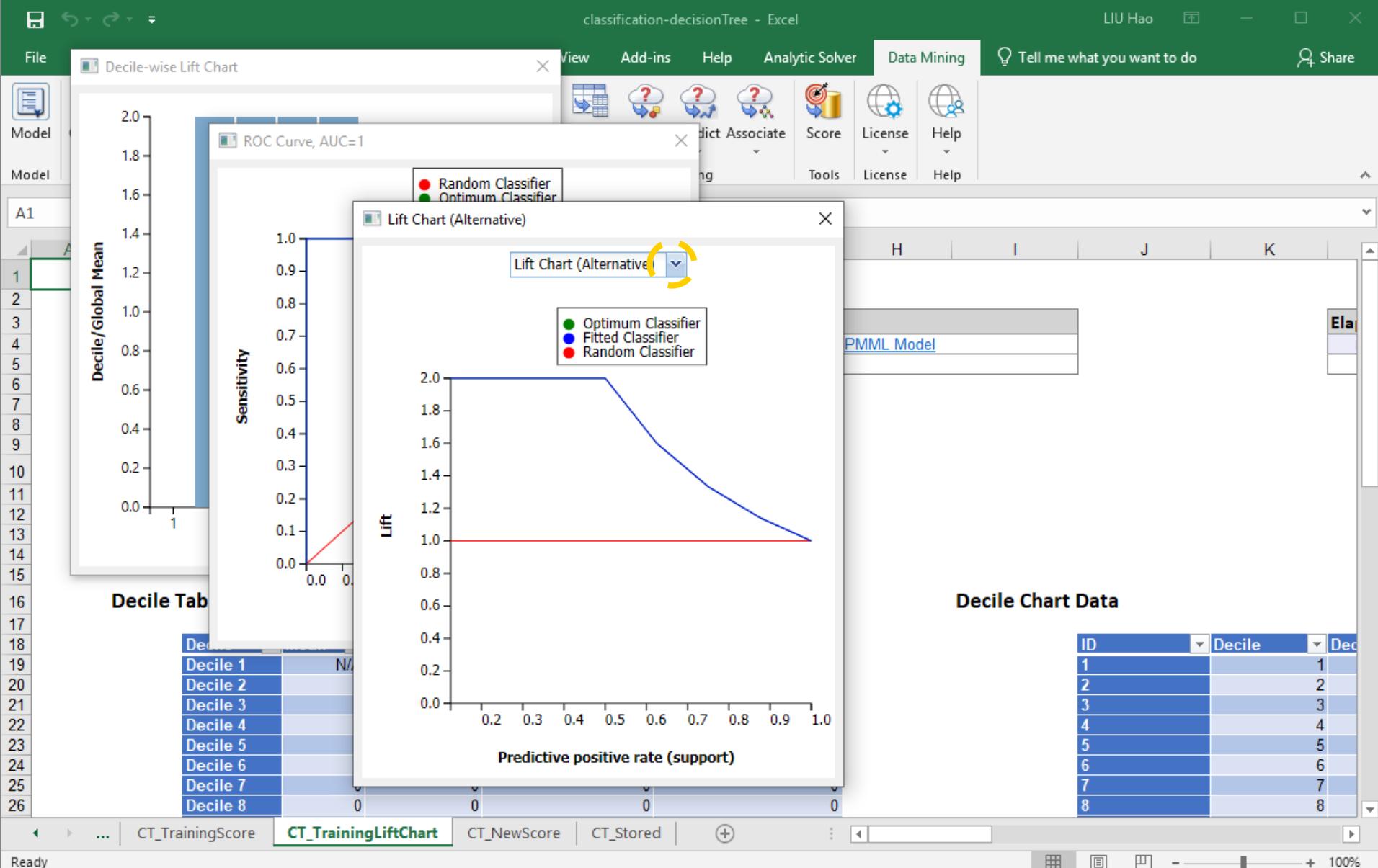
CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored

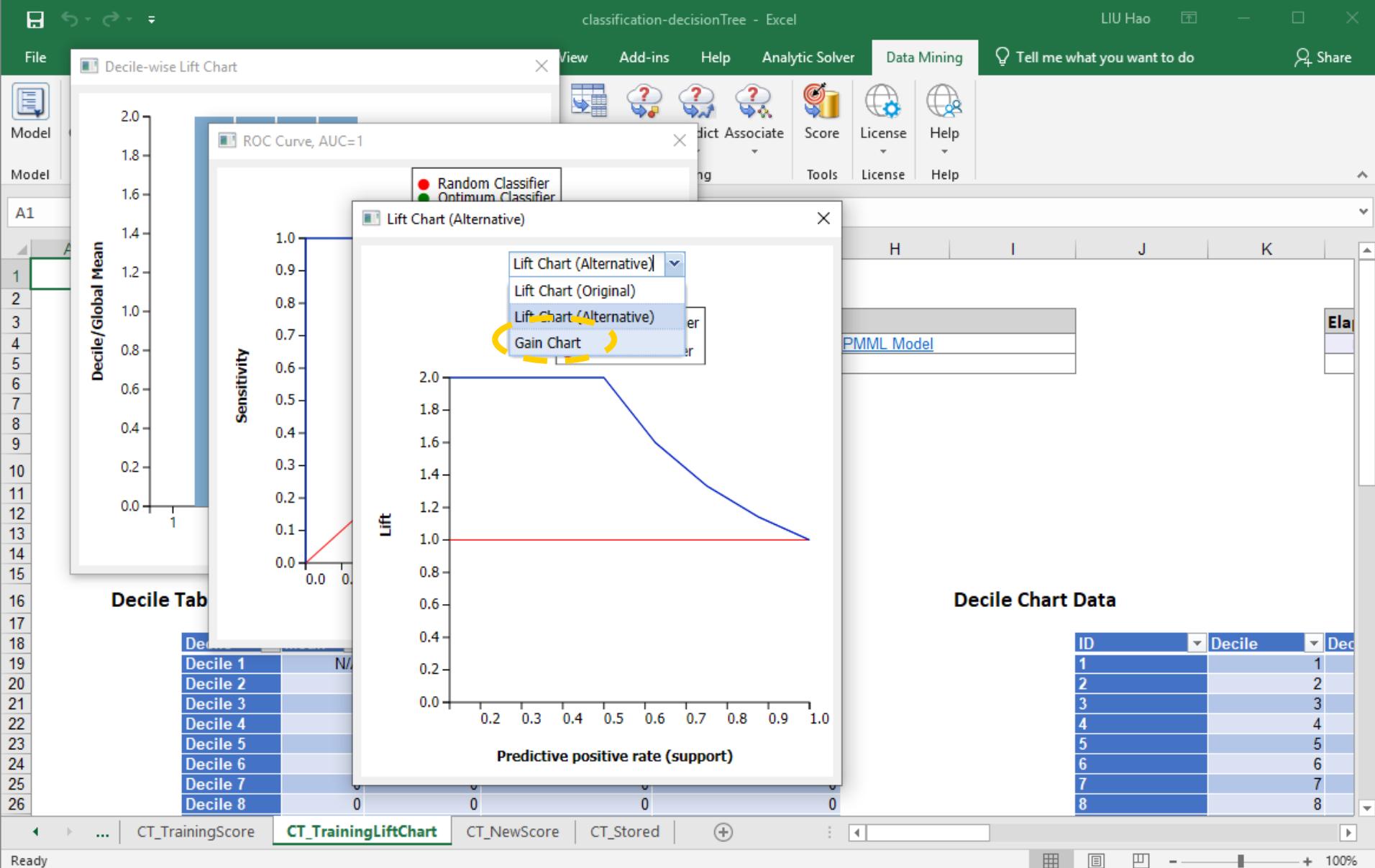


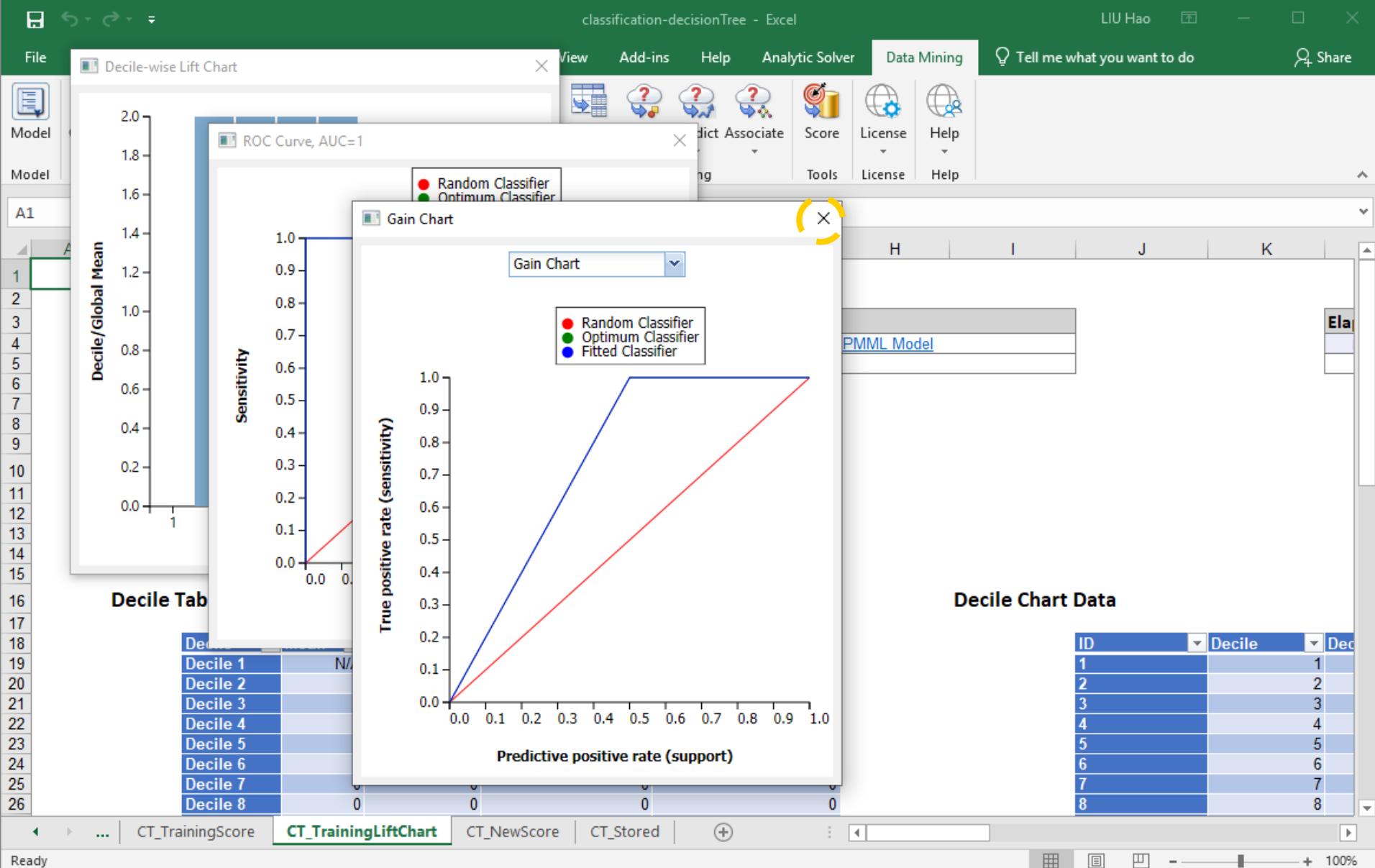
Ready 100%

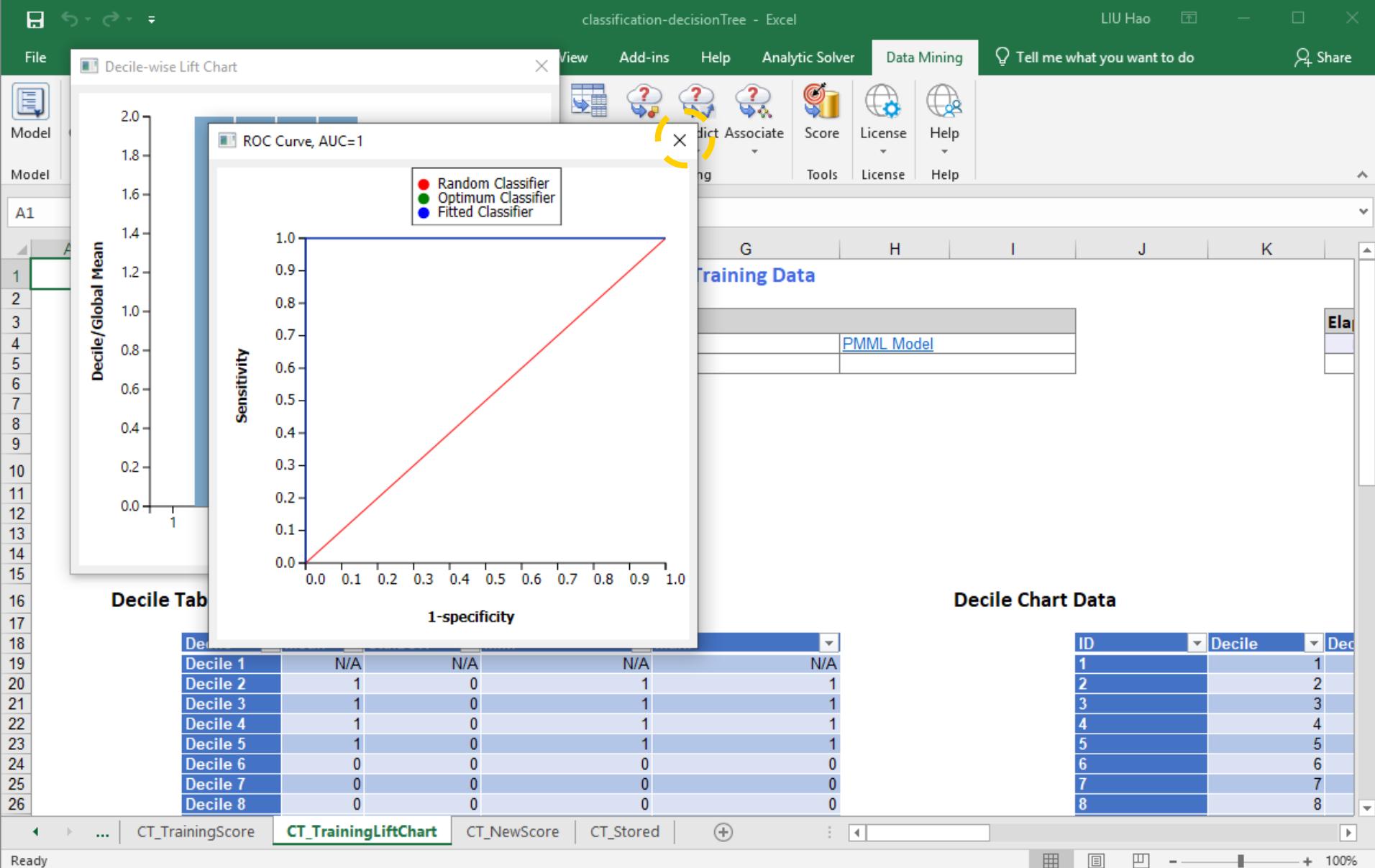


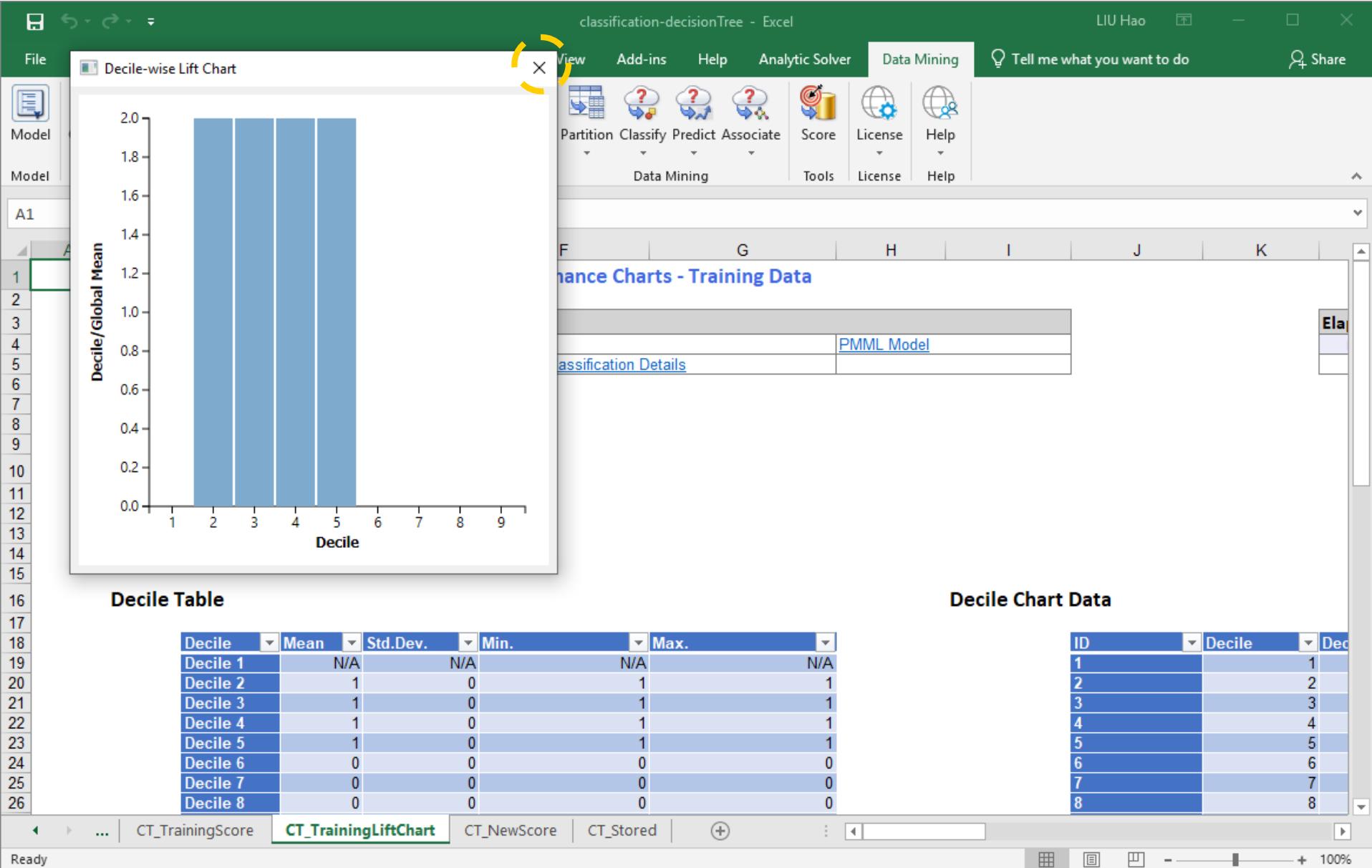












File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Data Analysis

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing

Time Series

Partition Classify Predict Associate Score License Help

Data Mining

Tools License Help

A1 A B C D E F G H I J K

7
8
9
10 **Terminology**
11
12 PPR Predictive Positive Rate (Support)
13 TPR True Positive Rate (Sensitivity, Recall)
14 FPR False Positive Rate (1-Specificity)
15

Decile Table

Decile	Mean	Std.Dev.	Min.	Max.
Decile 1	N/A	N/A	N/A	N/A
Decile 2	1	0	1	1
Decile 3	1	0	1	1
Decile 4	1	0	1	1
Decile 5	1	0	1	1
Decile 6	0	0	0	0
Decile 7	0	0	0	0
Decile 8	0	0	0	0
Decile 9	0	0	0	0

Decile Chart Data

ID	Decile	Decile
1		1
2		2
3		3
4		4
5		5
6		6
7		7
8		8
9		9

Lift/Gain Chart Data

# Records	# Cases	PPR	Cumulative (Model)	Cumulative (Random)	Lift (Model)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR
0	0	0	0	0	0	0	0	0	0

CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 : X ✓ f_x

A B C D E F G H I J K

7
8
9
10 Terminology
11
12 PPR Predictive Positive Rate (Support)
13 TPR True Positive Rate (Sensitivity, Recall)
14
15 Decile Mean Std.Dev. Min. Max.
16 Decile 1 N/A N/A N/A N/A
17 Decile 2 1 0 1 1
18 Decile 3 1 0 1 1
19 Decile 4 1 0 1 1
20 Decile 5 1 0 1 1
21 Decile 6 0 0 0 0
22 Decile 7 0 0 0 0
23 Decile 8 0 0 0 0
24 Decile 9 0 0 0 0
25
26
27
28
29
30
31 # Records # Cases PPR Cumulative (Model) Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model) TPR
32 0 0 0 0 0 0 0 0 0 0

... CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored +

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1	B	C	D	E	F	G	H	I	J	K
7										
8										
9										
10	Terminology									
11	PPR	Predictive Positive Rate (Support)								
12	TPR	True Positive Rate (Sensitivity, Recall)								
13	FPR	False Positive Rate (1-Specificity)								
14										
15										
16	Decile Table									
17	Decile	Mean	Std.Dev.	Min.	Max.					
18	Decile 1	N/A	N/A	N/A	N/A					
19	Decile 2	1	0	1	1					
20	Decile 3	1	0	1	1					
21	Decile 4	1	0	1	1					
22	Decile 5	1	0	1	1					
23	Decile 6	0	0	0	0					
24	Decile 7	0	0	0	0					
25	Decile 8	0	0	0	0					
26	Decile 9	0	0	0	0					
27										
28										
29	Lift/Gain Chart Data									
30	# Records	# Cases	PPR	Cumulative (Model)	Cumulative (Random)	Lift (Model)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR
31	0	0	0	0	0	0	0	0	0	0
32										

Decile	Mean	Std.Dev.	Min.	Max.
Decile 1	N/A	N/A	N/A	N/A
Decile 2	1	0	1	1
Decile 3	1	0	1	1
Decile 4	1	0	1	1
Decile 5	1	0	1	1
Decile 6	0	0	0	0
Decile 7	0	0	0	0
Decile 8	0	0	0	0
Decile 9	0	0	0	0

ID	Decile	Dec
1		1
2		2
3		3
4		4
5		5
6		6
7		7
8		8
9		9

Lift/Gain Chart Data

# Records	# Cases	PPR	Cumulative (Model)	Cumulative (Random)	Lift (Model)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR
0	0	0	0	0	0	0	0	0	0

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 G H I J K L M N O P Q

7
8
9
10
11
12
13
14
15
16 Decile Chart Data
17

ID	Decile	Decile/Global Mean
1	1	N/A
2	2	2
3	3	2
4	4	2
5	5	2
6	6	0
7	7	0
8	8	0
9	9	0

18 Max. N/A
19 1
20 1
21 1
22 1
23 1
24 0
25 0
26 0
27 0

28
29
30
31 Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model) TPR (Random) TPR (Optimum)
32 0 0 0 0 0 0 0

... CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored +

Ready

100%

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 G H I J K L M N O P Q I

7
8
9
10
11
12
13 ID Decile Decile/Global Mean
14 1 1 N/A
15 2 2 2
16 3 3 2
17 4 4 2
18 Max. 5 2
19 N/A
20 1
21 1
22 1
23 1
24 0
25 0
26 0
27 0
28
29
30
31 Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model) TPR (Random) TPR (Optimum)
32 0 0 0 0 0 0 0

CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored +

Ready

classification-decisionTree - Excel

LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 G H I J K L M N O P Q

7
8
9
10
11
12
13
14
15
16 Decile Chart Data

18 Max. N/A

19 1

20 1

21 1

22 1

23 1

24 0

25 0

26 0

27 0

18 ID Decile Decile/Global Mean

19 1 N/A

20 2 2

21 3 2

22 4 2

23 5 2

24 6 0

25 7 0

26 8 0

27 9 0

31 Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model) TPR (Random) TPR (Optimum)

32 0 0 0 0 0 0 0

Ready

CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored +

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Data Analysis

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Mining Tools License Help

A1

TPR True Positive Rate (Sensitivity, Recall)
FPR False Positive Rate (1-Specificity)

Decile Table

Decile	Mean	Std.Dev.	Min.	Max.
Decile 1	N/A	N/A	N/A	N/A
Decile 2	1	0	1	1
Decile 3	1	0	1	1
Decile 4	1	0	1	1
Decile 5	1	0	1	1
Decile 6	0	0	0	0
Decile 7	0	0	0	0
Decile 8	0	0	0	0
Decile 9	0	0	0	0

Decile Chart Data

ID	Decile	Decile/Global Mean
1	1	N/A
2	2	2
3	3	2
4	4	2
5	5	2
6	6	0
7	7	0
8	8	0
9	9	0

Lift/Gain Chart Data

# Record	Case	PPR	Cumulative (Mode)	Cumulative (Random)	Lift (Mode)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR (Random)	TPR (Optimum)
0	0	0	0	0	0	0	0	0	0	0
1	1	0.125	1	0.5	2	1	2	0.25	0.125	0.25
2	2	0.25	2	1	2	1	2	0.5	0.25	0.5
3	3	0.375	3	1.5	2	1	2	0.75	0.375	0.75
4	4	0.5	4	2	2	1	2	1	0.5	1
5	5	0.625	4	2.5	1.6	1	1.6	1	0.625	1
6	6	0.75	4	3	1.333333333	1	1.333333333	1	0.75	1
7	7	0.875	4	3.5	1.142857143	1	1.142857143	1	0.875	1
8	8	1	4	4	1	1	1	1	1	1

ROC Curve Data, AUC=1

# Record	FPR	TPR (Model)	TPR (Random)	TPR (Optimum)
0	0	0	0	0
1	0.125	0.25	0.125	0.25
2	0.25	0.5	0.25	0.5
3	0.375	0.75	0.375	0.75
4	0.5	1	0.5	1
5	0.625	1.6	0.625	1
6	0.75	2.5	0.75	1
7	0.875	3.5	0.875	1
8	1	4	1	1

CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1	B	C	D	E	F	G	H	I	J	K	L	M	N
TPR	True Positive Rate (Sensitivity, Recall)												
FPR	False Positive Rate (1-Specificity)												
Decile Table													
	Decile	Mean	Std.Dev.	Min.	Max.								
19	Decile 1	N/A	N/A	N/A	N/A								
20	Decile 2	1	0	1	1								
21	Decile 3	1	0	1	1								
22	Decile 4	1	0	1	1								
23	Decile 5	1	0	1	1								
24	Decile 6	0	0	0	0								
25	Decile 7	0	0	0	0								
26	Decile 8	0	0	0	0								
27	Decile 9	0	0	0	0								
28													
29													
	Lift/Gain Chart Data												
# Records	# Cases	PPR	Cumulative (Model)	Cumulative (Random)	Lift (Model)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR (Random)	TPR (Optimum)			
0	0	0	0	0	0	0	0	0	0	0			
1	1	0.125	1	0.5	2	1	2	0.25	0.125	0.25			
2	2	0.25	2	1	2	1	2	0.5	0.25	0.5			
3	3	0.375	3	1.5	2	1	2	0.75	0.375	0.75			
4	4	0.5	4	2	2	1	2	1	0.5	1			
5	5	0.625	4	2.5	1.6	1	1.6	1	0.625	1			
6	6	0.75	4	3	1.333333333	1	1.333333333	1	0.75	1			
7	7	0.875	4	3.5	1.142857143	1	1.142857143	1	0.875	1			
8	8	1	4	4	1	1	1	1	1	1			
42	ROC Curve Data, AUC=1												
43													
44	# Record	FPR	TPR (Model)	TPR (Random)	TPR (Optimum)								
45	n	0	0	0	0								

CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored

Ready

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Data Analysis

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Mining Tools License Help

A1

TPR True Positive Rate (Sensitivity, Recall)
FPR False Positive Rate (1-Specificity)

Decile Table

Decile	Mean	Std.Dev.	Min.	Max.
Decile 1	N/A	N/A	N/A	N/A
Decile 2	1	0	1	1
Decile 3	1	0	1	1
Decile 4	1	0	1	1
Decile 5	1	0	1	1
Decile 6	0	0	0	0
Decile 7	0	0	0	0
Decile 8	0	0	0	0
Decile 9	0	0	0	0

Decile Chart Data

ID	Decile	Decile/Global Mean
1	1	N/A
2	2	2
3	3	2
4	4	2
5	5	2
6	6	0
7	7	0
8	8	0
9	9	0

Lift/Gain Chart Data

# Record	# Case	PPR	Cumulative (Mode)	Cumulative (Random)	Lift (Mode)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR (Random)	TPR (Optimum)
0	0	0	0	0	0	0	0	0	0	0
1	1	0.125	1	0.5	2	1	2	0.25	0.125	0.25
2	2	0.25	2	1	2	1	2	0.5	0.25	0.5
3	3	0.375	3	1.5	2	1	2	0.75	0.375	0.75
4	4	0.5	4	2	2	1	2	1	0.5	1
5	5	0.625	4	2.5	1.6	1	1.6	1	0.625	1
6	6	0.75	4	3	1.333333333	1	1.333333333	1	0.75	1
7	7	0.875	4	3.5	1.142857143	1	1.142857143	1	0.875	1
8	8	1	4	4	1	1	1	1	1	1

ROC Curve Data, AUC=1

# Record	FPR	TPR (Model)	TPR (Random)	TPR (Optimum)
0	0	0	0	0
1	0.125	0.25	0.125	0.25
2	0.25	0.5	0.25	0.5
3	0.375	0.75	0.375	0.75
4	0.5	1	0.5	1
5	0.625	1.6	0.625	1
6	0.75	2.5	0.75	1
7	0.875	3.5	0.875	1
8	1	4	1	1

CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

Lift/Gain Chart Data

# Records	# Cases	PPR	Cumulative (Model)	Cumulative (Random)	Lift (Model)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR (Random)
0	0	0	0	0	0	0	0	0	0
1	1	0.125	1	0.5	2	1	2	2	0.25
2	2	0.25	2	1	2	2	2	2	0.5
3	3	0.375	3	1.5	2	1	2	2	0.75
4	4	0.5	4	2	2	1	2	2	1
5	5	0.625	4	2.5	1.6	1	1.6	1	
6	6	0.75	4	3	1.333333333	1	1.333333333	1	
7	7	0.875	4	3.5	1.142857143	1	1.142857143	1	
8	8	1	4	4	1	1	1	1	1

ROC Curve Data, AUC=1

# Records	FPR	TPR (Model)	TPR (Random)	TPR (Optimum)
0	0	0	0	1
1	0	1	0	1
2	1	1	1	1

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 : X ✓ fx

A B C D E F G H I J K

28

29 Lift/Gain Chart Data

30

# Records	# Cases	PPR	Cumulative (Model)	Cumulative (Random)	Lift (Model)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR
0	0	0	0	0	0	0	0	0	0
1	1	0.125	1	0.5	2	1	2	2	0.25
2	2	0.25	2	1	2	1	2	2	0.5
3	3	0.375	3	1.5	2	1	2	2	0.75
4	4	0.5	4	2	2	1	2	2	1
5	5	0.625	4	2.5	1.6	1	1.6	1	
6	6	0.75	4	3	1.333333333	1	1.333333333	1	
7	7	0.875	4	3.5	1.142857143	1	1.142857143	1	
8	8	1	4	4	1	1	1	1	1

41

# Records	FPR	TPR (Model)	TPR (Random)	TPR (Optimum)
0	0	0	0	1
1	0	1	0	1
2	1	1	1	1

42

43

44

45

46

47

48

49

50

51

52

53

54

CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored +

Ready

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1

Lift/Gain Chart Data

# Records	# Cases	PPR	Cumulative (Model)	Cumulative (Random)	Lift (Model)	Lift (Random)	Lift (Optimum)	TPR (Model)	TPR
0	0	0	0	0	0	0	0	0	0
1	1	0.125	1	0.5	2	1	2	2	0.25
2	2	0.25	2	1	2	2	2	2	0.5
3	3	0.375	3	1.5	2	1	2	2	0.75
4	4	0.5	4	2	2	1	2	2	1
5	5	0.625	4	2.5	1.6	1	1.6	1	
6	6	0.75	4	3	1.333333333	1	1.333333333	1	
7	7	0.875	4	3.5	1.142857143	1	1.142857143	1	
8	8	1	4	4	1	1	1	1	1

ROC Curve Data, AUC=1

# Records	FPR	TPR (Model)	TPR (Random)	TPR (Optimum)
0	0	0	0	1
1	0	1	0	1
2	1	1	1	1

CT_TrainingLiftChart

classification-decisionTree - Excel LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 A B C D E F G H I J K L M N Date: 29

Data Mining: Classification Tree - Prediction of New Data

Output Navigator

Fully Grown Tree Rules	New: Classification Details	Inputs	PMML Model
Training: Charts	Training: Classification Summary	Training: Classification Detail	

Elapsed Times in Milliseconds

Data Reading Time	45
Algorithm Time	37
Report Time	

New: Classification Details

Record ID	Prediction: Insurance	PostProb: no	PostProb: yes
Record 1	yes	0	1

CT_TrainingScore CT_TrainingLiftChart **CT_NewScore** CT_Stored +

Ready 100%

classification-decisionTree - Excel LIU Hao

File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share

Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help

Model Data Data Analysis Time Series Data Mining Tools License Help

A1 : X ✓ fx

A B C D E F G H I J K L M N Date: 29

1 Data Mining: Classification Tree - Prediction of New Data

2

3 Output Navigator

4 Fully Grown Tree Rules ([New: Classification Details](#)) Inputs PMML Model

5 Training: Charts [Training: Classification Summary](#) [Training: Classification Detail:](#)

6 Elapsed Times in Milliseconds

7 Data Reading Time Algorithm Time Report Time

8 45 37

9

10 New: Classification Details

11 Record ID ▾ Prediction: Insurance ▾ PostProb: no ▾ PostProb: yes ▾

12 Record 1 yes 0 1

13

14

15

16

17

18

19

20

21

22

23

24

25

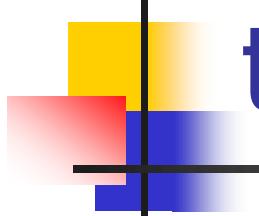
26

27

... CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored +

Ready 100%

This screenshot shows a Microsoft Excel spreadsheet titled "classification-decisionTree - Excel". The ribbon has the "Data Mining" tab selected. The main area displays a classification tree prediction interface. Row 1 contains the title "Data Mining: Classification Tree - Prediction of New Data". Row 2 has columns A through N. Row 3 contains the "Output Navigator" section with links to "Fully Grown Tree Rules", "Inputs", "PMML Model", "Training: Charts", "Training: Classification Summary", and "Training: Classification Detail:". Row 4 shows elapsed times in milliseconds for data reading, algorithm time, and report time. Rows 10 and 11 show the "New: Classification Details" section with dropdown menus for "Record ID", "Prediction: Insurance", "PostProb: no", and "PostProb: yes". Row 12 shows a single record with "Record 1" in the first column, "yes" in the second, "0" in the third, and "1" in the fourth. The bottom navigation bar includes tabs for "CT_TrainingScore", "CT_TrainingLiftChart", "CT_NewScore" (which is active), and "CT_Stored". The status bar at the bottom right shows "100%".

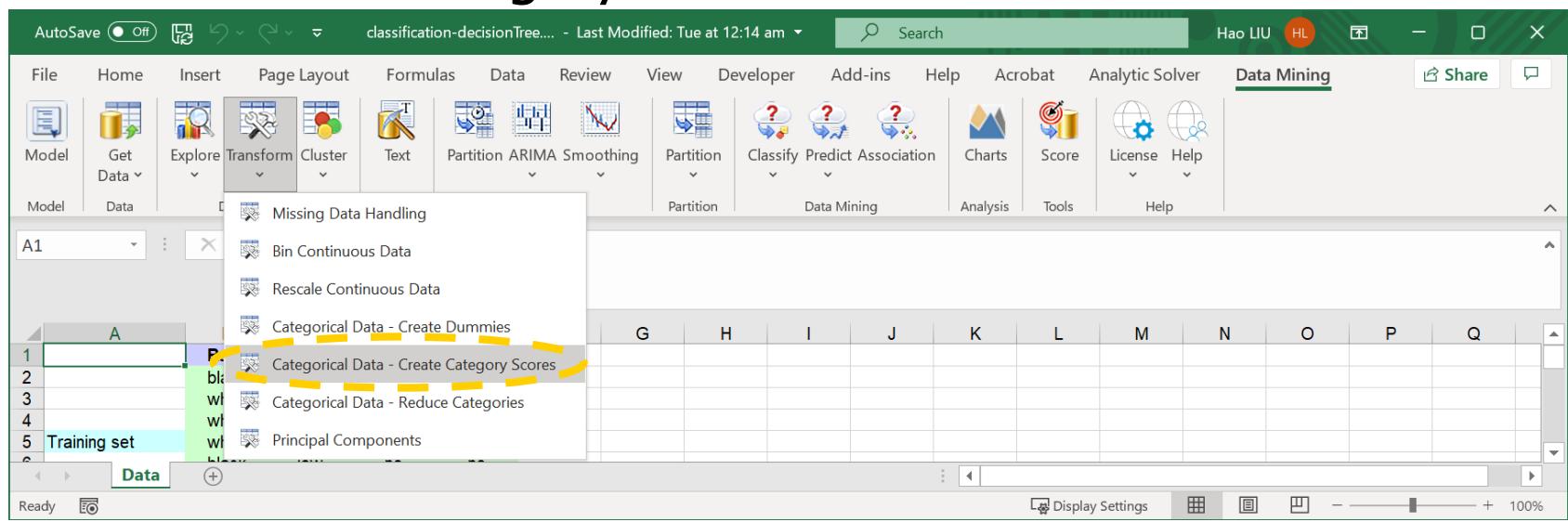


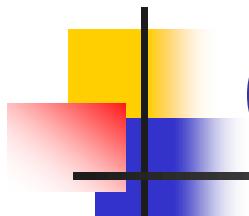
How to use the data mining tool

- We have the following 2 versions.
 - XLMiner Desktop (installed in either the CSE lab machine or your computer)
 - ➡ ■ XLMiner Cloud (installed as a plugin in your Office 365 Excel)

How to use the data mining tool (XLMiner Cloud)

- The way of opening “Create Category Scores” in XLMiner Cloud plugin in your Office 365 Excel
 - “Data Mining” Tag → Transform → Categorical Data - Create Category Scores





How to use the data mining tool (XLMiner Cloud)

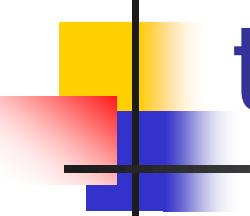
- The steps and output format of “create category scores” in XLMiner Cloud are similar to the steps in XLMiner Desktop.
- The transformation result of XLMiner Cloud Platform is the same as that from XLMiner Desktop.

How to use the data mining tool (XLMiner Cloud)

- The way of opening “classification tree” in XLMiner Cloud plugin in your Office 365 Excel
 - “Data Mining” Tag → Classify → Classification Tree

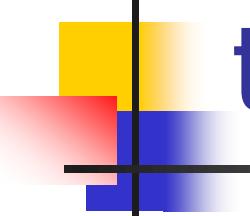
A screenshot of Microsoft Excel showing the Data Mining ribbon tab selected. The ribbon tabs include File, Home, Insert, Page Layout, Formulas, Data, Review, View, Developer, Add-ins, Help, Acrobat, Analytic Solver, and Data Mining. The Data Mining tab is highlighted with a green background. Below the ribbon, there are several data analysis tools: Model (Model, Get Data), Data Analysis (Explore, Transform, Cluster), Text (Text, Partition, ARIMA, Smoothing), Time Series (Partition), and Partition (Partition). The Data Mining tab has three main sections: Classify (Find Best Model, Discriminant Analysis, Logistic Regression, k-Nearest Neighbors, Classification Tree, Naive Bayes), Predict (Score, License), and Association (Help). A dropdown menu for Classification Tree is open, showing options: Find Best Model, Discriminant Analysis, Logistic Regression, k-Nearest Neighbors, Classification Tree, Naive Bayes, Automatic Neural Network, Manual Neural Network, Ensemble - Bagging, Ensemble - Boosting, and Ensemble - Random Trees. In the Excel worksheet, row 5 is labeled "Training set" and row 10 is labeled "New set". The "Classification Tree" option is highlighted with a yellow dashed arrow.

	A	B	C	D	E	F	G	H
1		Race	Income	Child	Insurance			
2		black	high	no	yes			
3		white	high	yes	yes			
4		white	low	yes	yes			
5	Training set	white	low	yes	yes			
6		black	low	no	no			
7		black	low	no	no			
8		black	low	no	no			
9		white	low	no	no			
10	New set	white	high	no	?			



How to use the data mining tool (XLMiner Cloud)

- The steps of performing “classification tree” in XLMiner Cloud is similar to the steps in XLMiner Desktop.
- The decision tree and classification result of XLMiner Cloud is the same as that from XLMiner Desktop.



How to use the data mining tool (XLMiner Cloud)

- The output format of XLMiner Cloud is similar to the output in XLMiner Desktop.
- However, to display the classification tree and the lift charts, you need to call out the “Charts” window.

AutoSave



File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Search Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Partition Classify Predict Association Charts Score License Help Model Data Data Analysis Text Time Series Partition Data Mining Analysis Tools Help

J20

A B C D E F G H I J K L

Data Mining: Classification Tree

Output Navigator				Elapsed Times in Milliseconds
Fully Grown Tree Rules (Using Train)	New: Classification Details	Inputs	PMML Model	Data Reading Time
Training: Charts	Training: Classification Su	Training: Classificati		

Inputs

Data	
Workbook	classification-decisionTree-score.xlsx
Worksheet	Factorization
Data Range	D27:G35
# Records	8

Variables			
# Variables	3		
Scale Variables	Race	Income	Child
Categorical Variables			
Output Variable	Insurance		

Rescaling: Fitting Parameters			
Factorization	CT_Output	CT_FullTree	CT_TrainingScore

The screenshot shows the Microsoft Excel ribbon with the 'Data Mining' tab selected. The ribbon includes tabs for File, Home, Insert, Page Layout, Formulas, Data, Review, View, Help, Analytic Solver, Data Mining, Design, Search, and License. Under the Data Mining tab, there are groups for Model, Data, Data Analysis, Text, Time Series, Partition, Data Mining, and Tools. The 'Charts' icon in the Analysis group is highlighted with a yellow oval.

C12 : Node ID

A B C D E F G H I J K

1 Data Mining: Classification Tree

2

3 Output Navigator

4 Fully Grown Tree Rules (New: Classification Details) Inputs PMML Model

5 Training: Charts Training: Classification Summar Training: Classification Details

6

7

8

9

10 Fully Grown Tree Rules (Using Training Data)

11

Node ID	Parent ID	Left Child ID	Right Child ID	Split Var	Split Value/Set	Training Cases	Validation Cases	Respon
1	N/A	2	3	Child	1.5	8	0	
2	1	4	5	Income	1.5	5	0	
3	1	N/A	N/A	N/A	N/A	3	0	
4	2	N/A	N/A	N/A	N/A	1	0	
5	2	N/A	N/A	N/A	N/A	4	0	

12

13

14

15

16

17

18

19

20

21

22

23

24

Factorization CT_Output CT_FullTree CT_TrainingScore C ... + : ▶

Average: 2.36 Count: 60 Sum: 59

100%

105

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Data

Model

Data

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349</div

AutoSave (● off)



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Model

Get Data

Model

Data

C12

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

C12

Worksheet:

CT_FullTree

Chart:

Full Classification Tree

Data

Factorization

CT_Output

CT_FullTree

CT_TrainingScore

CT_TrainingLiftChart

CT_NewScore

CT_Stored

LINKS

Records

Full Classification Tree

x

Records

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Model



Get Data

Model

C44

A

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search

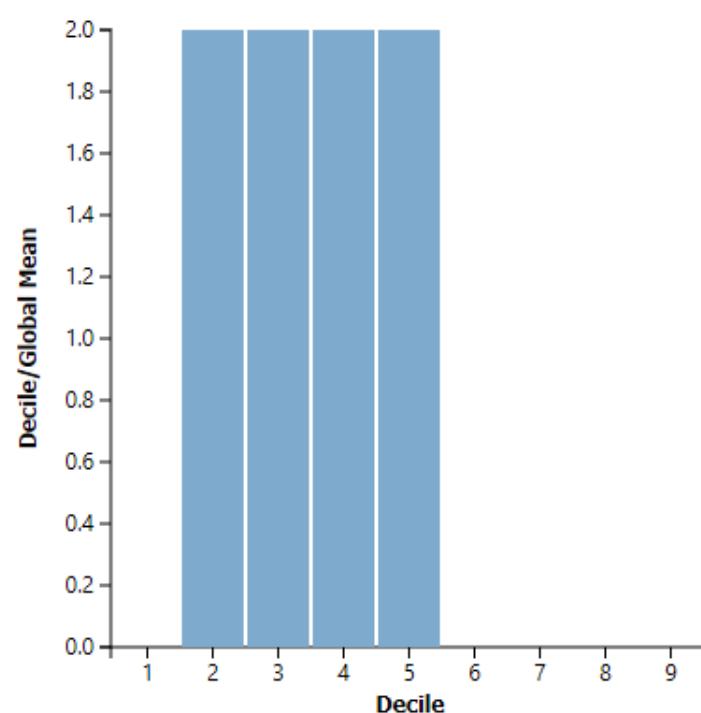
Analytic Solver Data Mining - <https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1>

Worksheet:

CT_TrainingLiftChart

Chart:

Decile Chart



Decile 6

0

0

0

0

0

CT_FullTree

CT_TrainingScore

CT_TrainingLiftChart

CT_NewSc ...



Average: 0.666666667 Count: 20 Sum: 10



100%

COT 15/12

172

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design

Search



Model

Get Data

Model

Data

C44

...

A

B

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

Data

Output

Fully C

Trainin

Term

PPR

TPR

FPR

Deci

Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1

Worksheet:

CT_TrainingLiftChart

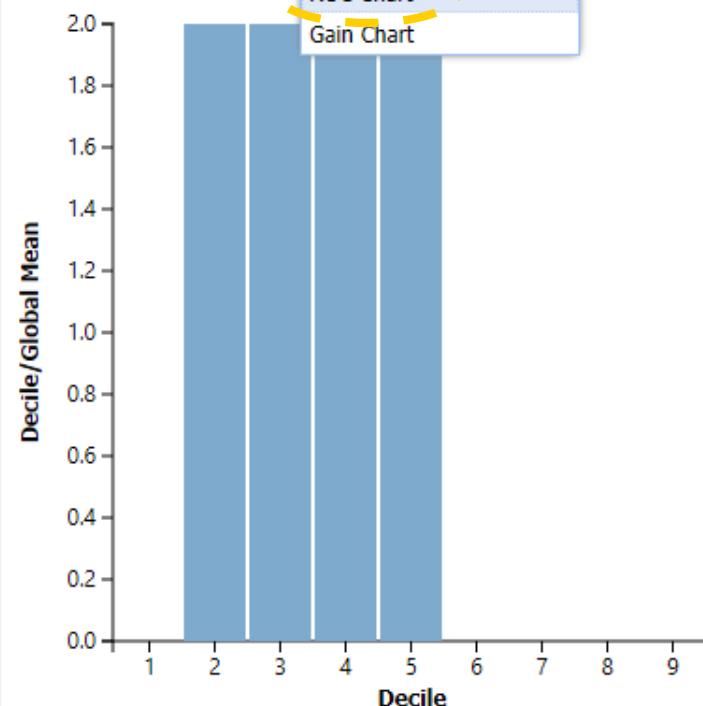
Chart:

Decile Chart

Decile Chart

ROC Chart

Gain Chart



Decile 6

0

0

0

0

CT_FullTree CT_TrainingScore

CT_TrainingLiftChart

CT_NewSc ...

+

:

<

>

:

-

-

+

100%

Average: 0.666666667 Count: 20 Sum: 10



COT 15/12

175

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Model

Data

C44

A

B

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

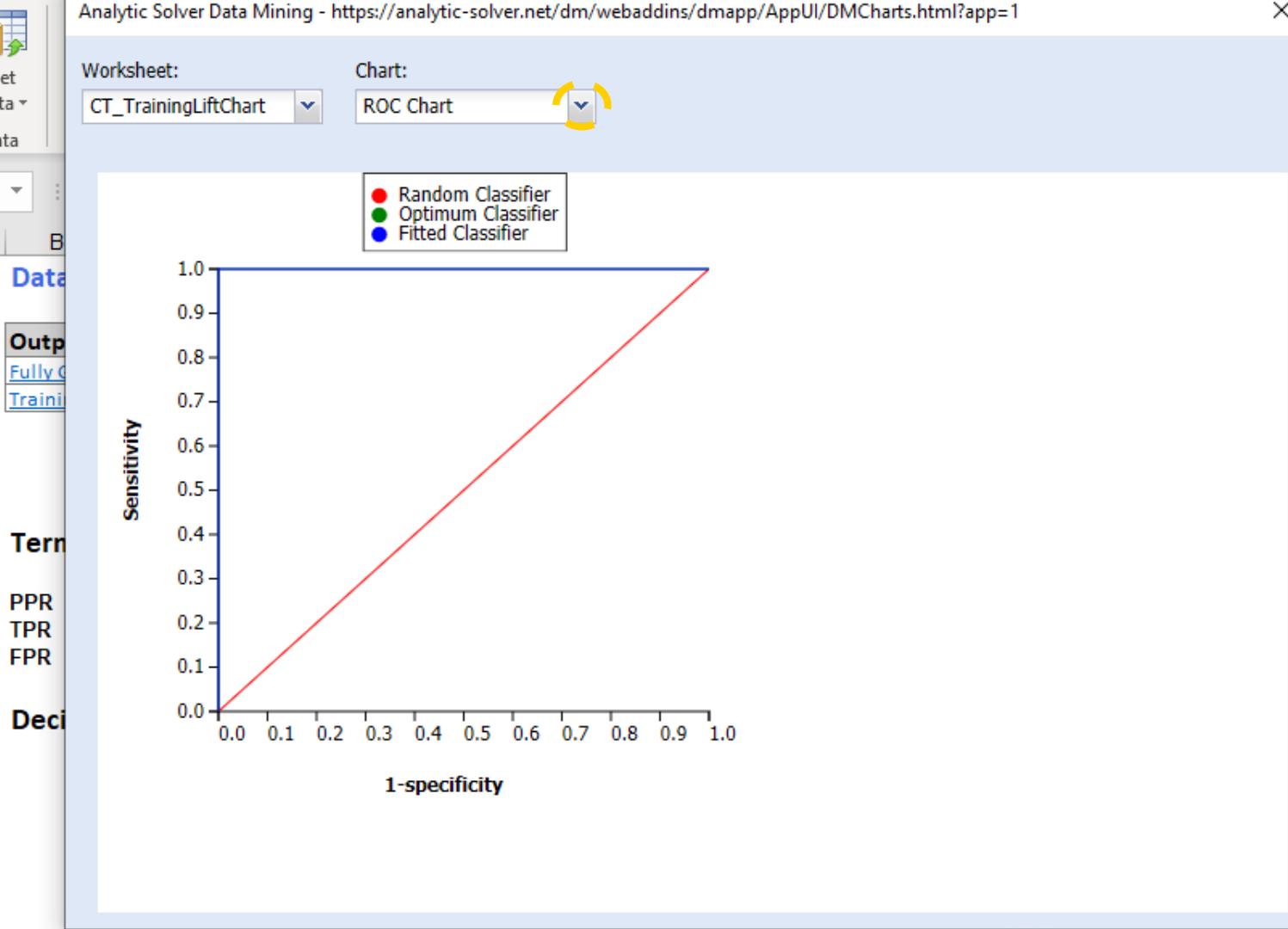
20

21

22

23

24



Decile 6

0

0

0

0

CT_FullTree | CT_TrainingScore |

CT_TrainingLiftChart

CT_NewSc ...



+ 100%

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Model

Data

C44

A

B

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

Data

Output

Fully C

Trainin

Tern

PPR

TPR

FPR

Deci

6

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

AutoSave

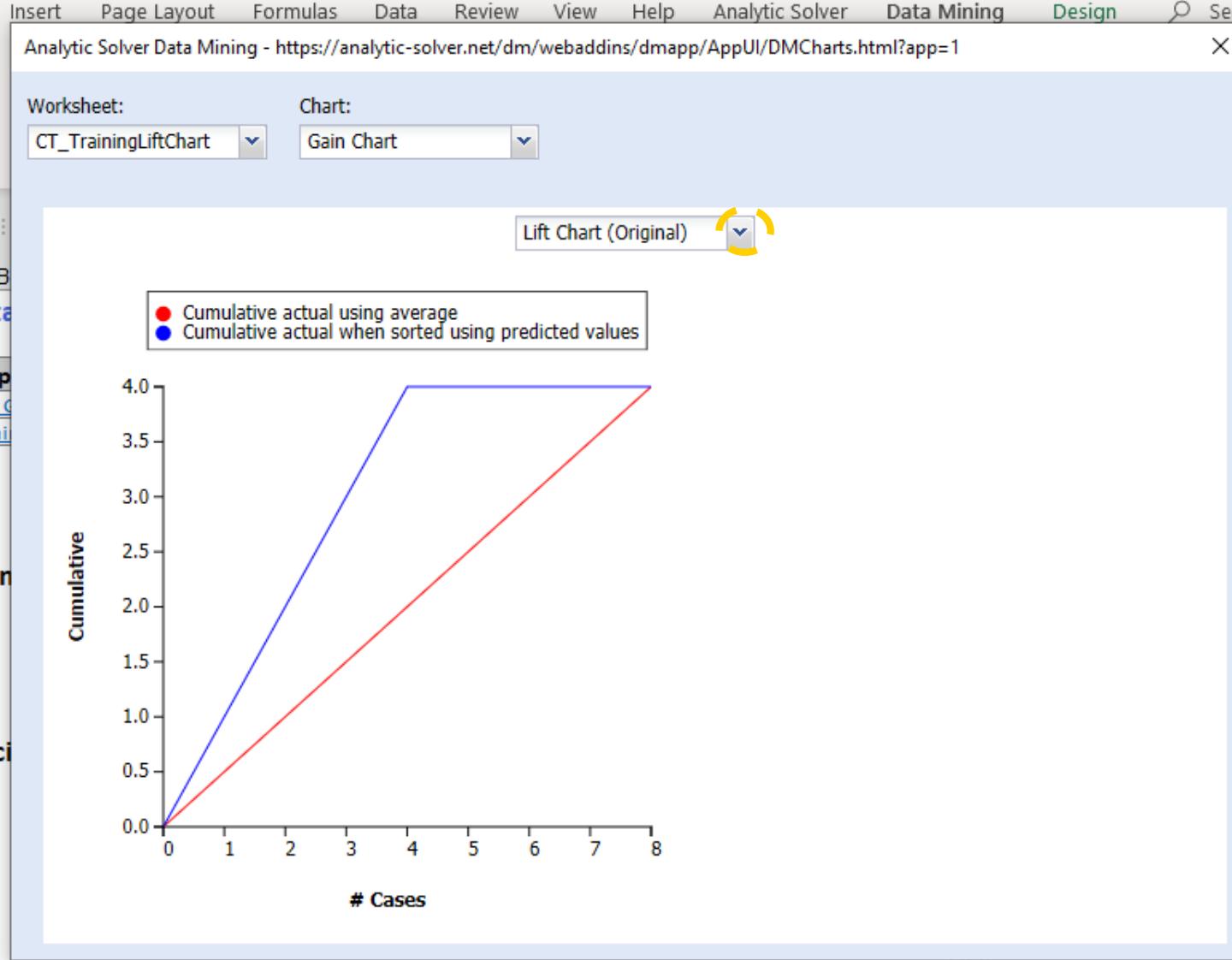


Data

Model

Data

C44



Decile 6

CT_TrainingLiftChart

CT_NewSc ...

0

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Model

Data

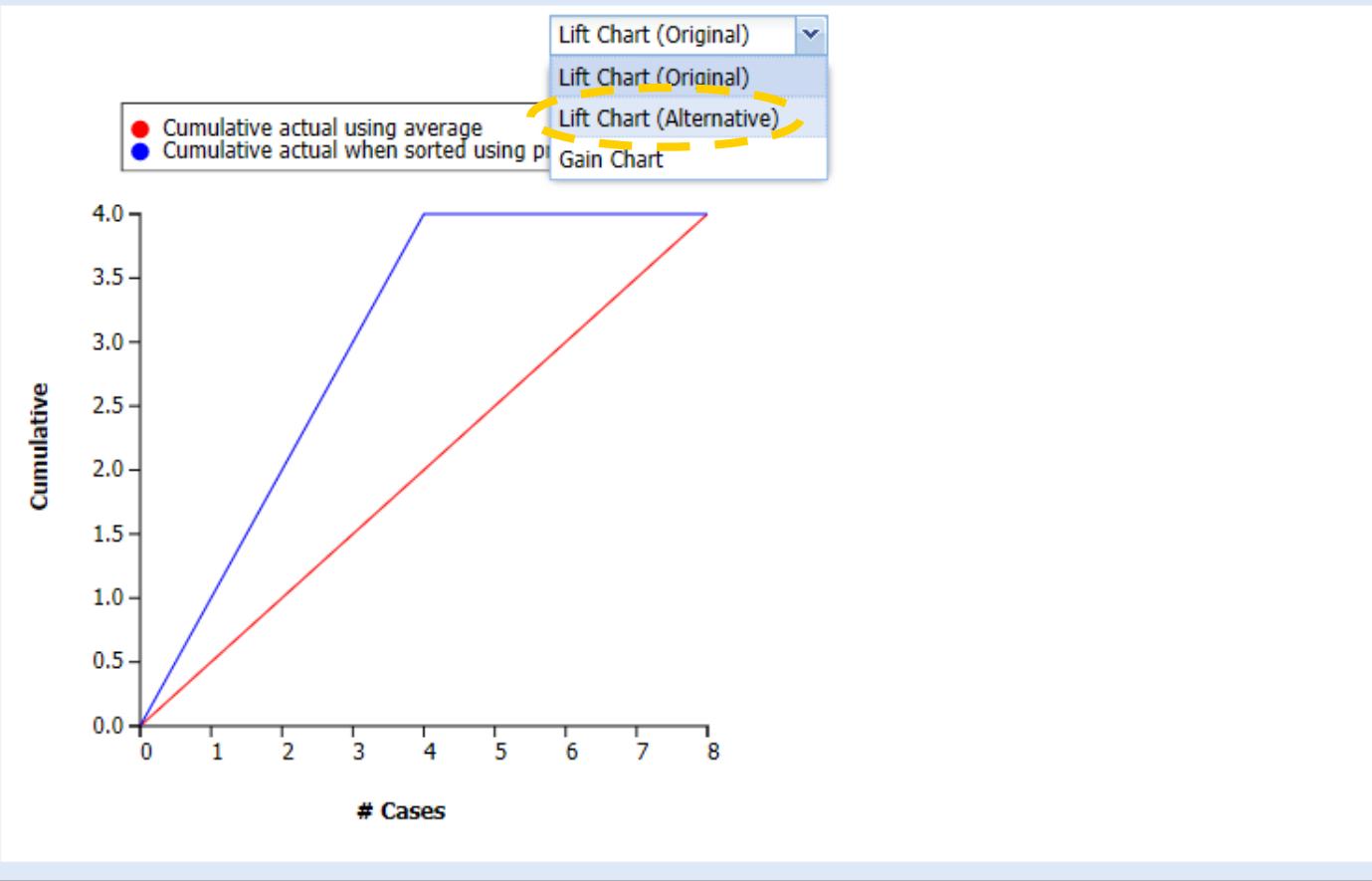
C44

Worksheet:

CT_TrainingLiftChart

Chart:

Gain Chart



Decile 6

CT_FullTree

CT_TrainingScore

CT_TrainingLiftChart

CT_NewSc ...

0

0

0

0

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Model

Data

C44

A

B

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

Data

Output

Fully C

Trainin

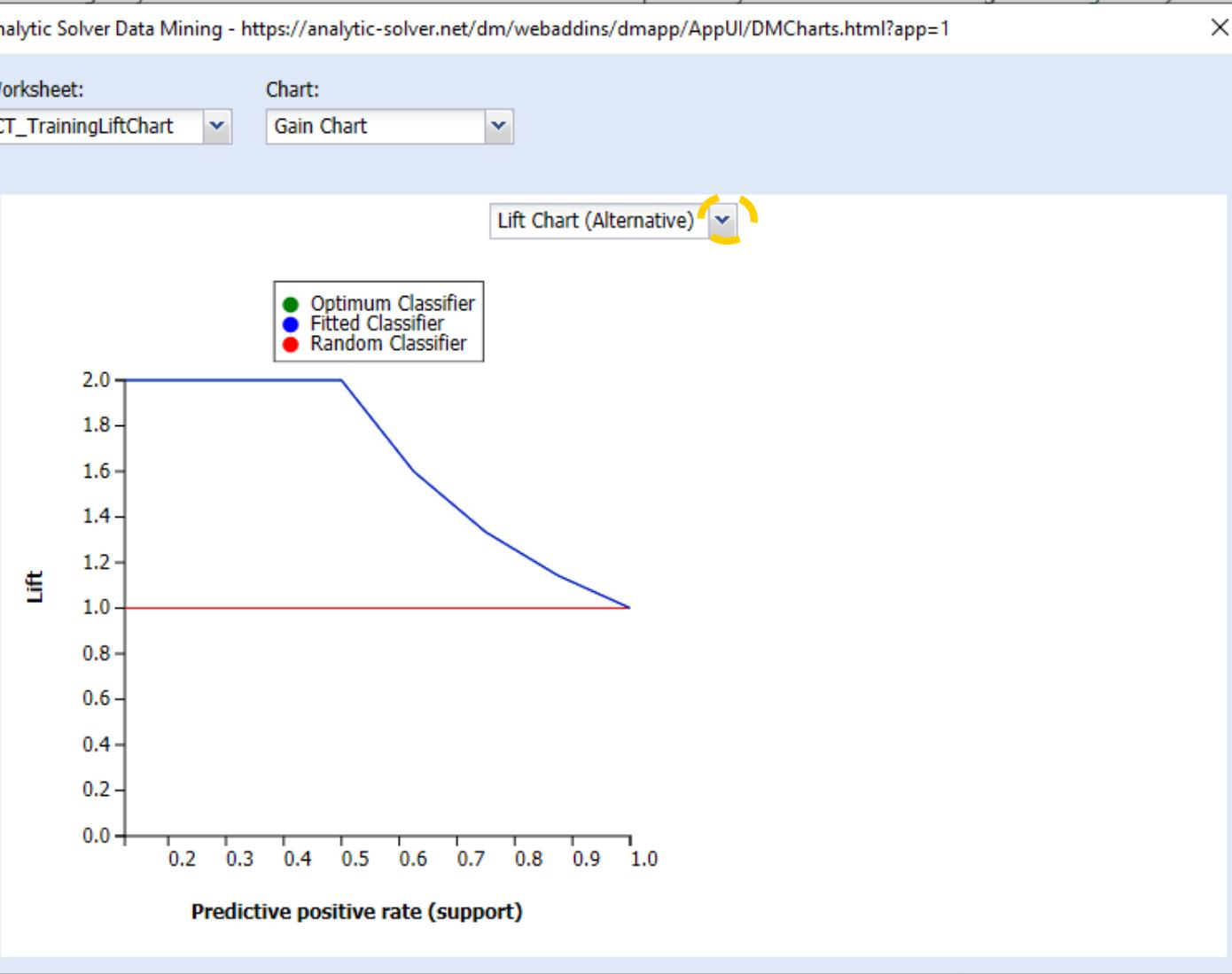
Term

PPR

TPR

FPR

Deci



Decile 6

CT_TrainingLiftChart

CT_NewSc ...

0

0

Average: 0.666666667 Count: 20 Sum: 10



100%

C011 15/12

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Model

Data

C44

A

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

Data

Outp

Fully C

Trainin

Term

PPR

TPR

FPR

Deci

Worksheet:

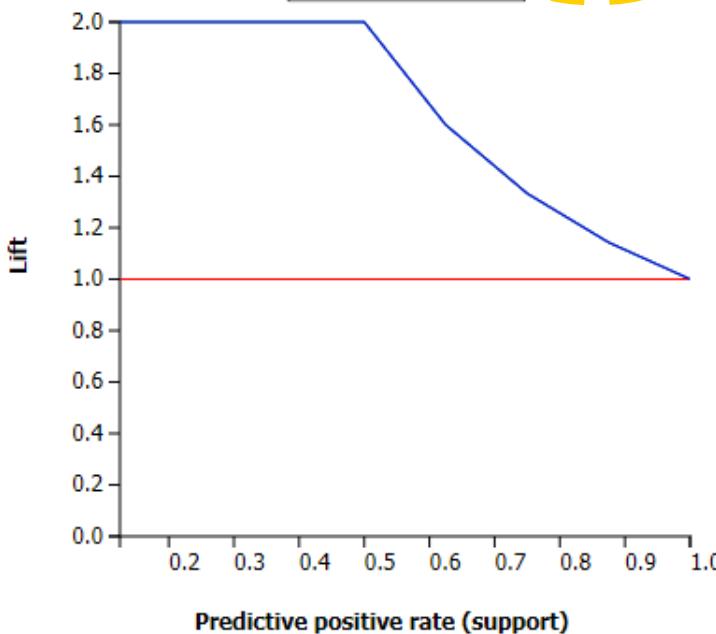
CT_TrainingLiftChart

Chart:

Gain Chart

- Lift Chart (Alternative)
- Lift Chart (Original)
- Lift Chart (Alternative)
- Gain Chart

- Optimum Classifier
- Fitted Classifier
- Random Classifier



Decile 6

0

0

0

0

CT_FullTree CT_TrainingScore

CT_TrainingLiftChart

CT_NewSc ...

ID

1

2

3

4

5

6

Average: 0.666666667 Count: 20 Sum: 10

100%

AutoSave



File

Home

Insert

Page Layout

Formulas

Data

Review

View

Help

Analytic Solver

Data Mining

Design



Search



Model

Data

C44

A

B

Data

Outp

Fully C

Traini

Tern

PPR

TPR

FPR

Deci

Worksheet:

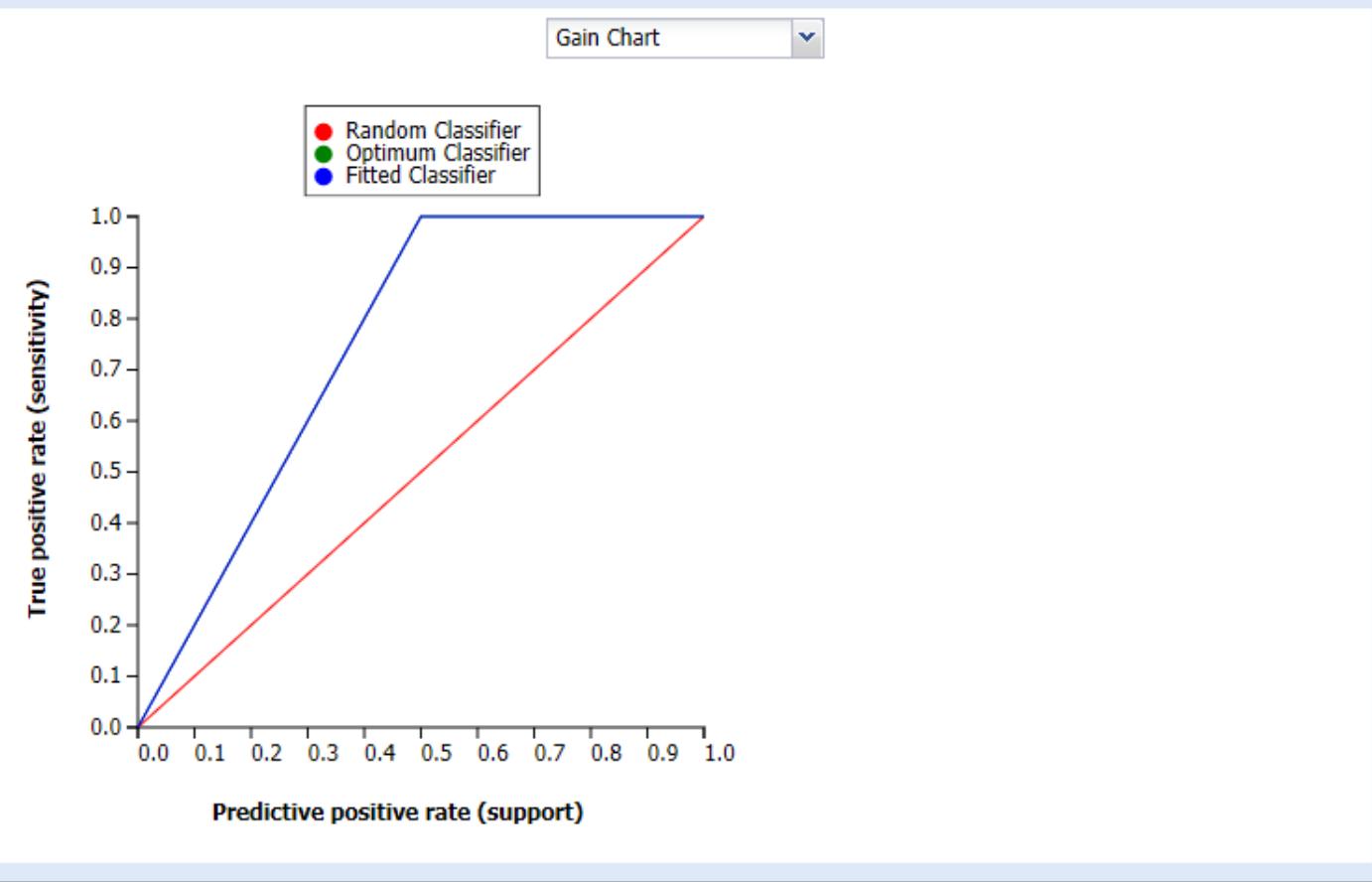
CT_TrainingLiftChart

Chart:

Gain Chart



Help



Decile 6

CT_TrainingLiftChart

CT_NewSc ...

