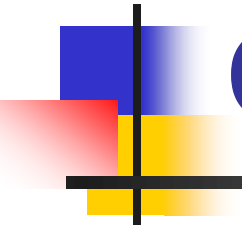


# COMP1942

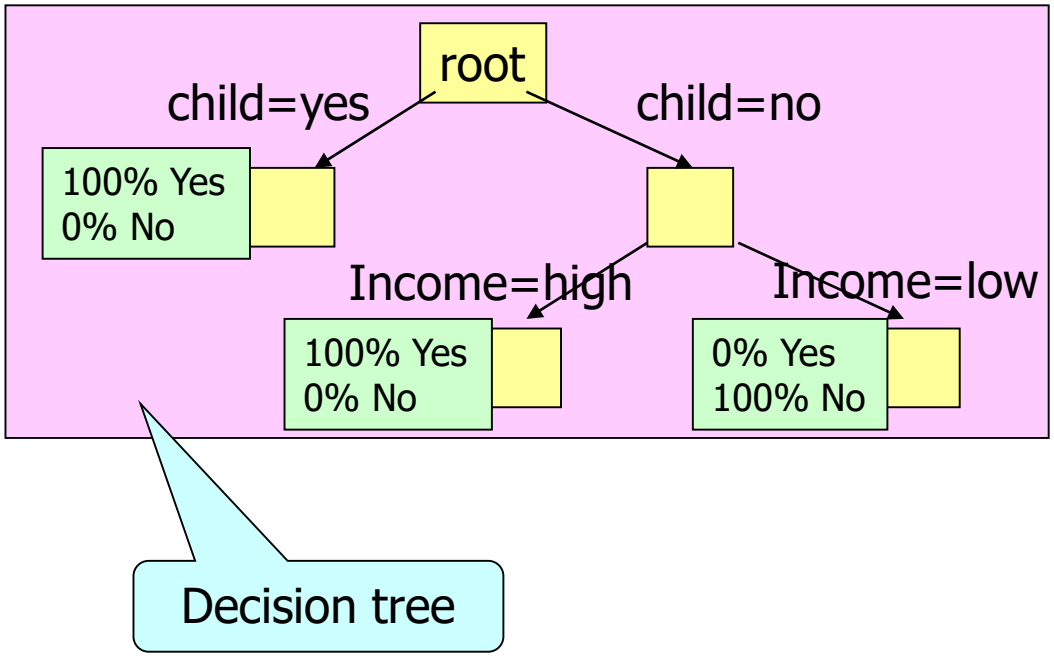
## Classification 1
## (Introduction and Decision Tree)

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 | ? |



# Classification

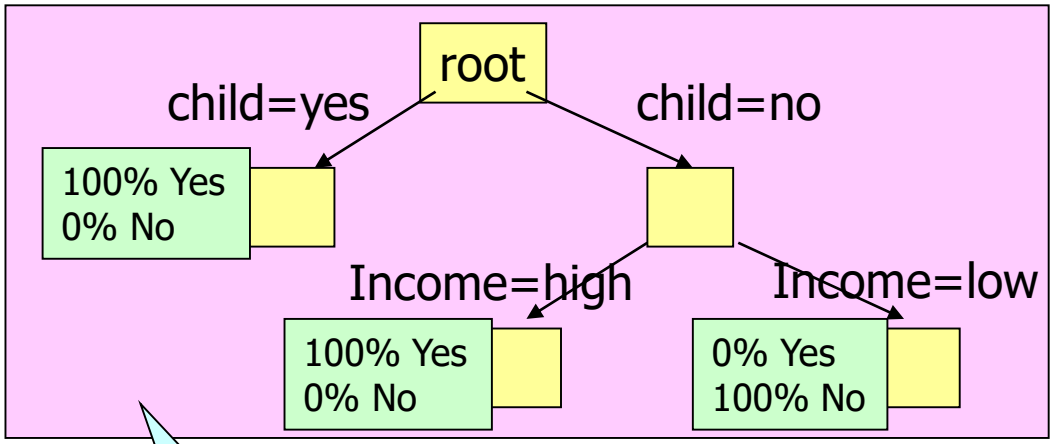
Suppose there is a person.

| Race | Income | Child | Insurance |
|---|---|---|---|
| white | high | no | ? |

New set

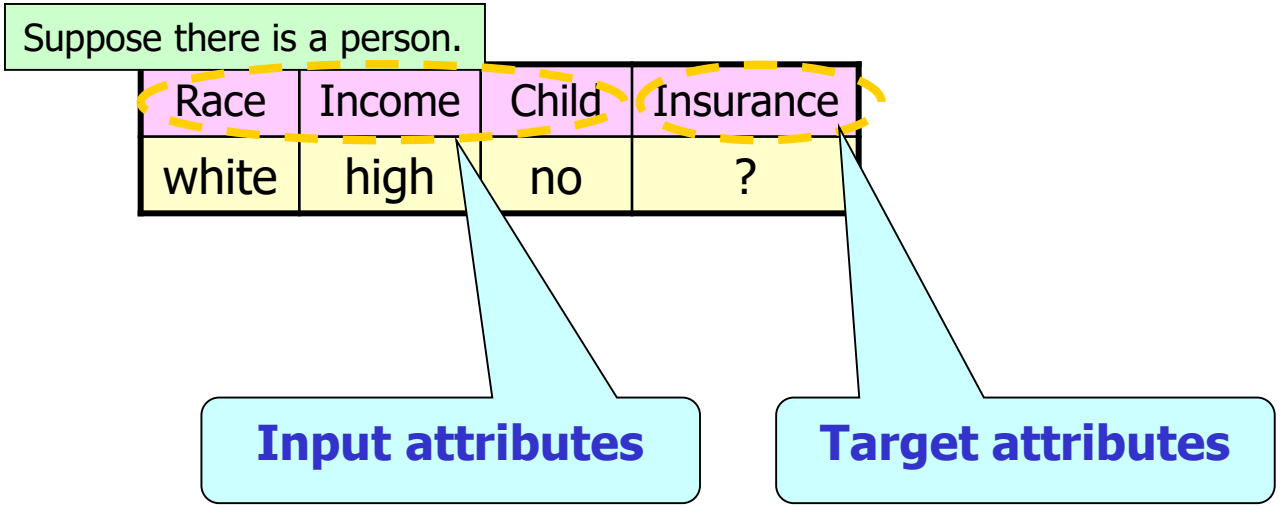
| Race | Income | Child | Insurance |
|---|---|---|---|
| black | high | no | yes |
| white | high | yes | yes |
| white | low | yes | yes |
| white | low | yes | yes |
| black | low | no | no |
| black | low | no | no |
| black | low | no | no |
| white | low | no | no |

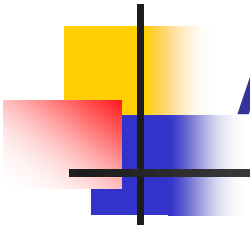
Training set




Decision tree

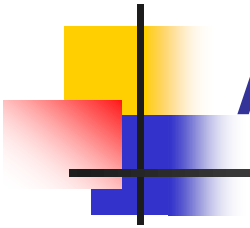
# Classification





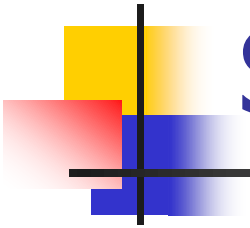
# Applications

- 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



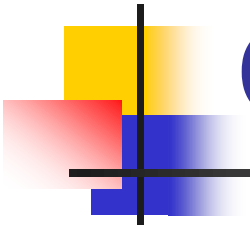
# Applications

- 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

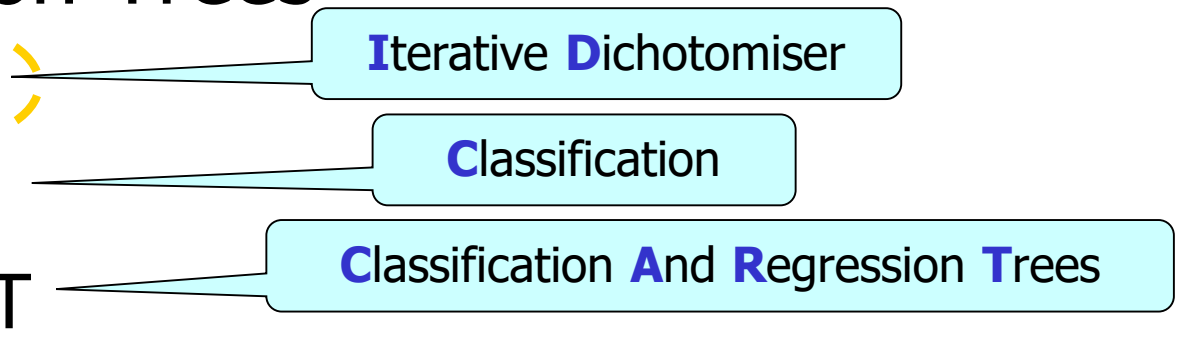


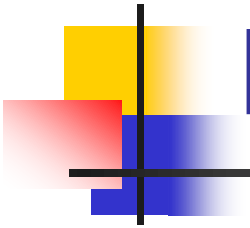
# Classification Methods

- Decision Tree
- Bayesian Classifier
- Nearest Neighbor Classifier

# Decision Trees

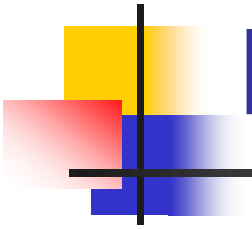
- Decision Trees
  - ID3
  - C4.5
  - CART
- Measurement
- How to use the data mining tool





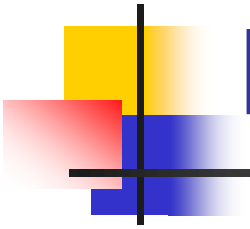
# Entropy

- 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 $P = (p_1, p_2, \ldots, p_n)$ then the **Information** conveyed by this distribution, also called the **Entropy** of P, is:
  $I(P) = - (p_1 \times \log p_1 + p_2 \times \log p_2 + \ldots + p_n \times \log p_n)$
- 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

| 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 |

Info(T) = - ½ log ½ - ½ log ½
= 1

For attribute Race,

$Info(T_{black}) = - ¾ \log ¾ - ¼ \log ¼ = 0.8113$

$Info(T_{white}) = - ¾ \log ¾ - ¼ \log ¼ = 0.8113$

$Info(Race, T) = ½ \times Info(T_{black}) + ½ \times Info(T_{white}) = 0.8113$

$Gain(Race, T) = Info(T) - Info(Race, T) = 1 - 0.8113 = 0.1887$

For attribute Race, Gain(Race, T) = 0.1887

# Entropy

| 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 |

Info(T) = - ½ log ½ - ½ log ½
= 1

For attribute Income,

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

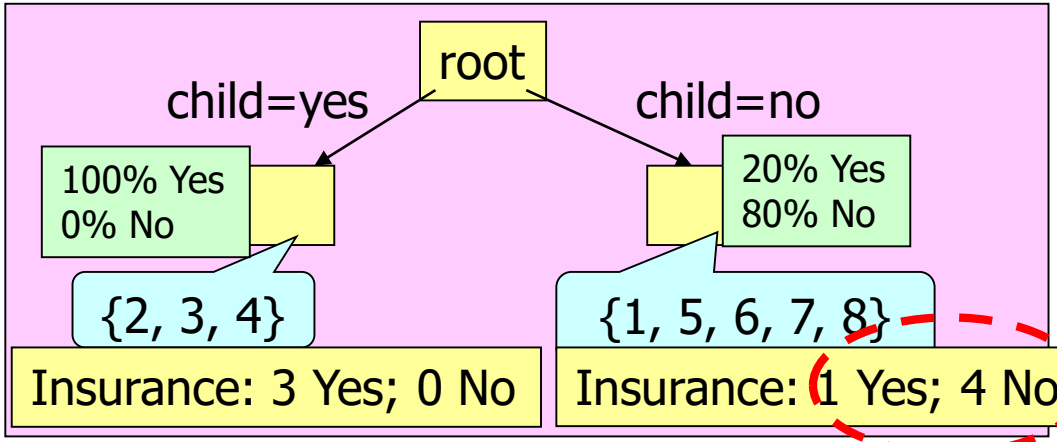
$Info(T_{low}) = - 1/3 \log 1/3 - 2/3 \log 2/3 = 0.9183$

$Info(Income, T) = ¼ \times Info(T_{high}) + ¾ \times Info(T_{low}) = 0.6887$

$Gain(Income, T) = Info(T) - Info(Income, T) = 1 - 0.6887 = 0.3113$

For attribute Race, Gain(Race, T) = 0.1887

For attribute Income, Gain(Income, T) = 0.3113




| | 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 |

Info(T) = - ½ log ½ - ½ log ½
= 1

For attribute Child,

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

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

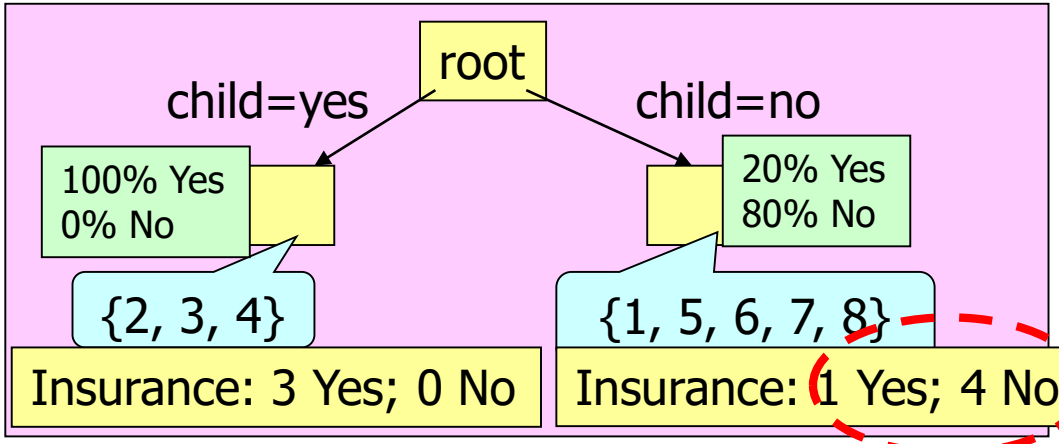
$Info(Child, T) = 3/8 \times Info(T_{yes}) + 5/8 \times Info(T_{no}) = 0.4512$

Gain(Child, T) = Info(T) – Info(Child, T) = 1 – 0.4512 = 0.5488

For attribute Race, Gain(Race, T) = 0.1887

For attribute Income, Gain(Income, T) = 0.3113

For attribute Child, Gain(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 |

Info(T) = - 1/5 log 1/5 – 4/5 log 4/5
= 0.7219

For attribute Race,

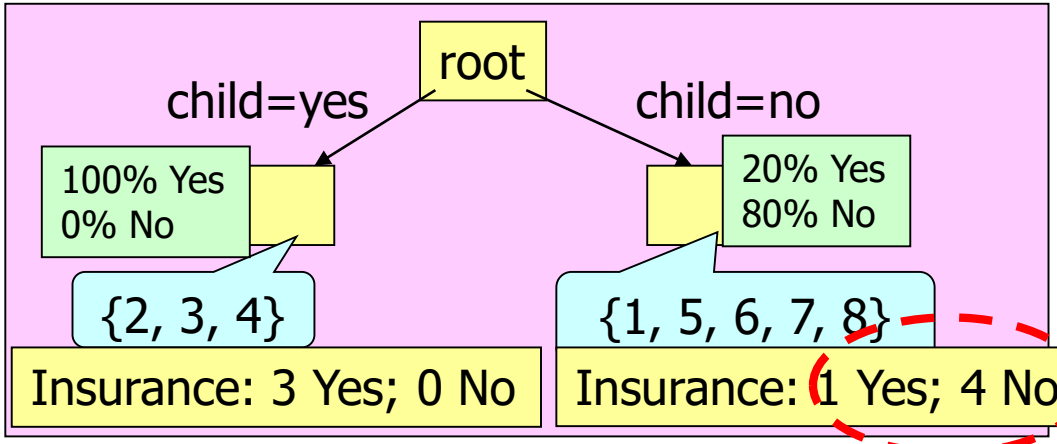
$Info(T_{black}) = - \frac{1}{4} \log \frac{1}{4} - \frac{3}{4} \log \frac{3}{4} = 0.8113$

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

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

Gain(Race, T) = Info(T) – Info(Race, T) = 0.7219 – 0.6490 = 0.0729

For attribute Race, Gain(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 |

$$Info(T) = - 1/5 \log 1/5 - 4/5 \log 4/5 = 0.7219$$

For attribute Income,

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

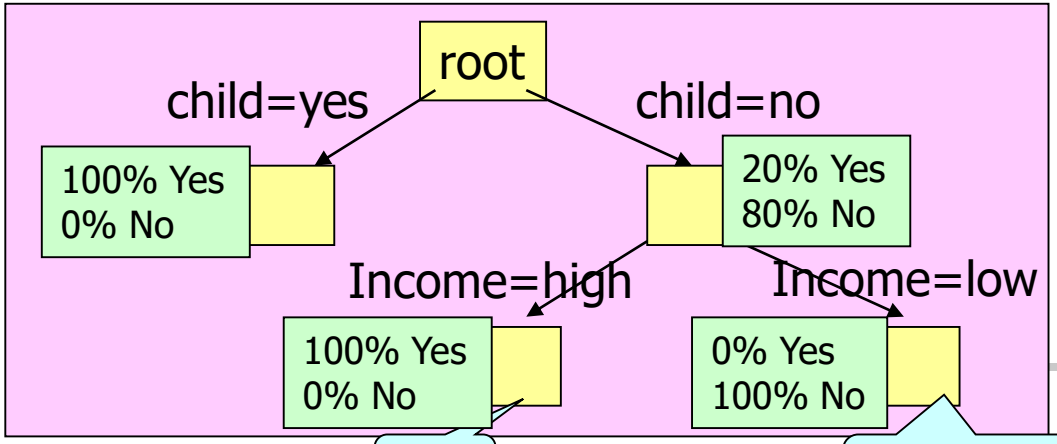
$$Info(T_{low}) = - 0 \log 0 - 1 \log 1 = 0$$

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

$$Gain(Income, T) = Info(T) - Info(Income, T) = 0.7219 - 0 = 0.7219$$

For attribute Race, Gain(Race, T) = 0.0729

For attribute Income, Gain(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 |

Info(T) = - 1/5 {1} /5 – 4/5 log 4/5 {5, 6, 7, 8}

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

For attribute Income,

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

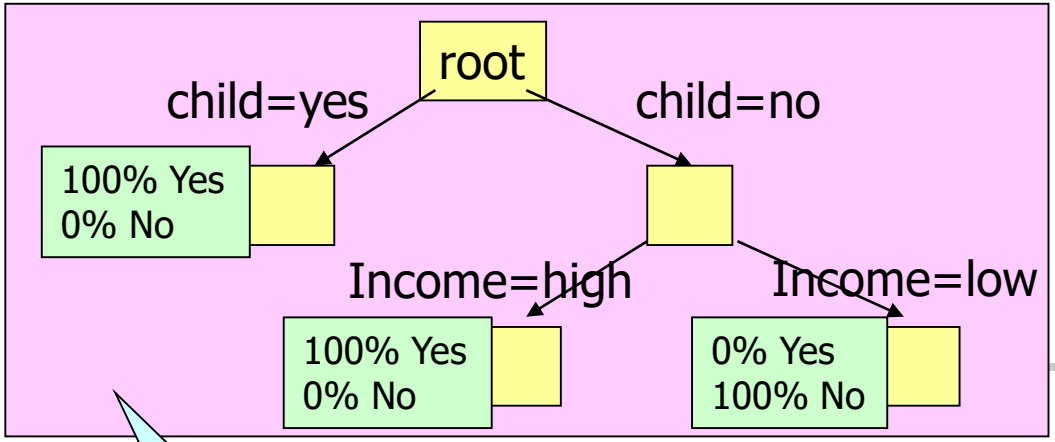
$Info(T_{low}) = - 0 \log 0 – 1 \log 1 = 0$

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

$Gain(Income, T) = Info(T) – Info(Income, T) = 0.7219 – 0 = 0.7219$

For attribute Race, Gain(Race, T) = 0.0729

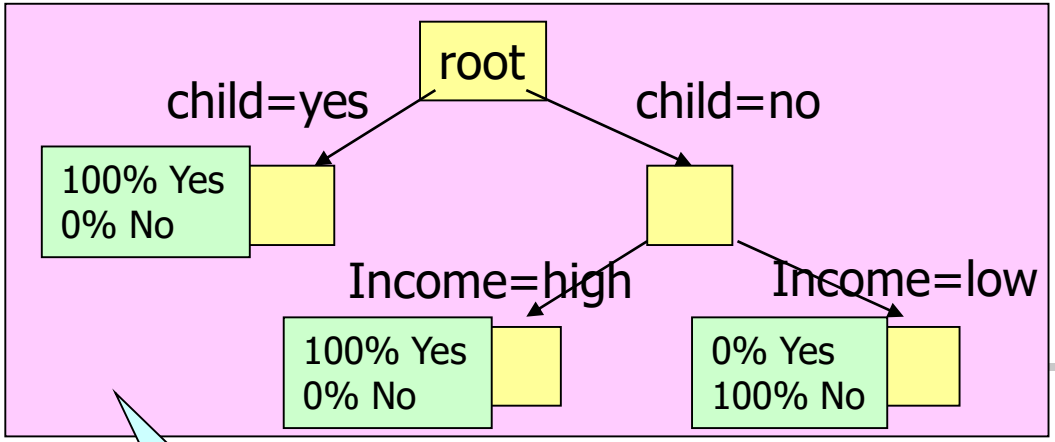
For attribute Income, Gain(Income, T) = 0.7219




| | 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 |

Suppose there is a new person.

| Race | Income | Child | Insurance |
|---|---|---|---|
| white | high | no | ? |

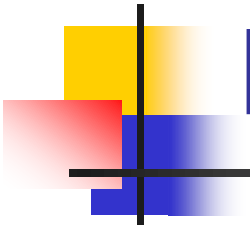



Decision tree

| | 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 |

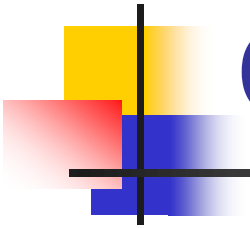
Termination Criteria?

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) = $-\Sigma_{v \in A}$ p(v) log p(v)

# Entropy

| 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 |

$Info(T) = - ½ \log ½ - ½ \log ½$
$= 1$

For attribute Race,

$Info(T_{black}) = - ¾ \log ¾ - ¼ \log ¼ = 0.8113$

$Info(T_{white}) = - ¾ \log ¾ - ¼ \log ¼ = 0.8113$

$Info(Race, T) = ½ \times Info(T_{black}) + ½ \times Info(T_{white}) = 0.8113$

$SplitInfo(Race) = - ½ \log ½ - ½ \log ½ = 1$

$Gain(Race, T) = (Info(T) – Info(Race, T))/SplitInfo(Race) = (1 – 0.8113)/1 = 0.1887$

For attribute Race, Gain(Race, T) = 0.1887

# Entropy

| 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 |

Info(T) = - ½ log ½ - ½ log ½
= 1

For attribute Income,

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

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

$Info(Income, T) = ¼ \times Info(T_{high}) + ¾ \times Info(T_{low}) = 0.6887$

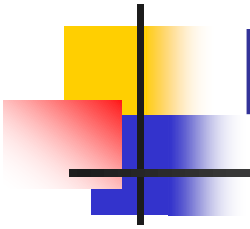
$SplitInfo(Income) = - 2/8 \log 2/8 - 6/8 \log 6/8 = 0.8113$

Gain(Income, T)= (Info(T)–Info(Income, T))/SplitInfo(Income) = (1–0.6887)/0.8113
= 0.3837

For attribute Race, Gain(Race, T) = 0.1887

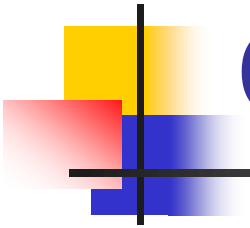
For attribute Income, Gain(Income, T) = 0.3837

For attribute Child, Gain(Child, T) = ?



# Decision Trees

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



# CART

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

# Gini

| 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 |

$$\text{Info}(T) = 1 - (½)^2 - (½)^2$$
$$= ½$$

For attribute Race,

$$\text{Info}(T_{black}) = 1 - (¾)^2 - (¼)^2 = 0.375$$

$$\text{Info}(T_{white}) = 1 - (¾)^2 - (¼)^2 = 0.375$$

$$\text{Info}(\text{Race}, T) = ½ \times \text{Info}(T_{black}) + ½ \times \text{Info}(T_{white}) = 0.375$$

$$\text{Gain}(\text{Race}, T) = \text{Info}(T) - \text{Info}(\text{Race}, T) = ½ - 0.375 = 0.125$$

For attribute Race, Gain(Race, T) = 0.125

# Gini

| 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 |

$$Info(T) = 1 - (½)^2 - (½)^2$$
$$= ½$$

For attribute Income,

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

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

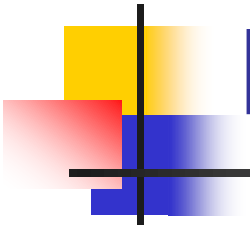
$$Info(Income, T) = 1/4 \times Info(T_{high}) + 3/4 \times Info(T_{low}) = 0.333$$

$$Gain(Income, T) = Info(T) - Info(Race, T) = ½ - 0.333 = 0.167$$

For attribute Race, Gain(Race, T) = 0.125

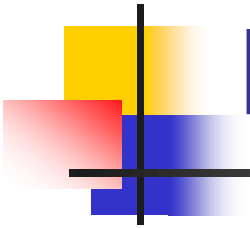
For attribute Income, Gain(Race, T) = 0.167

For attribute Child, Gain(Child, T) = ?



# 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

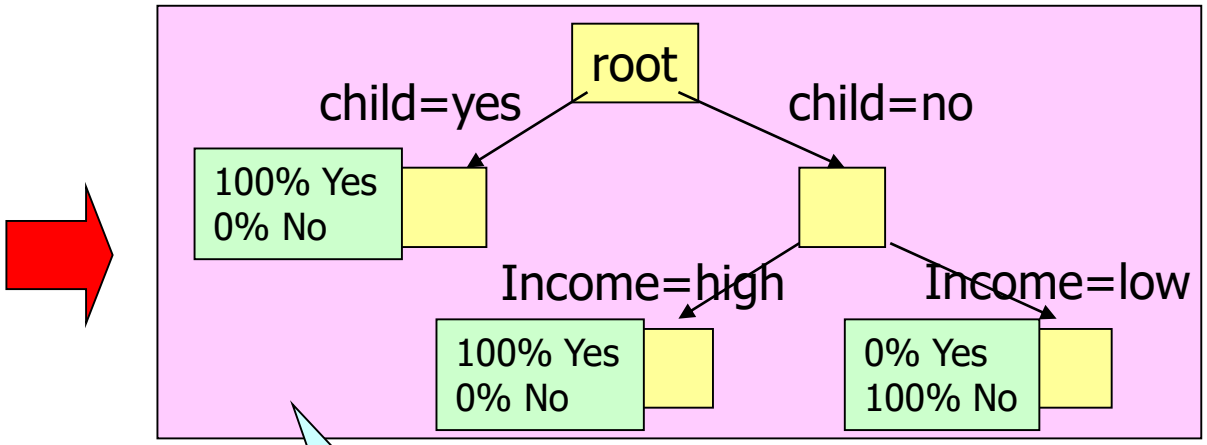
Suppose there is a person.

| 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

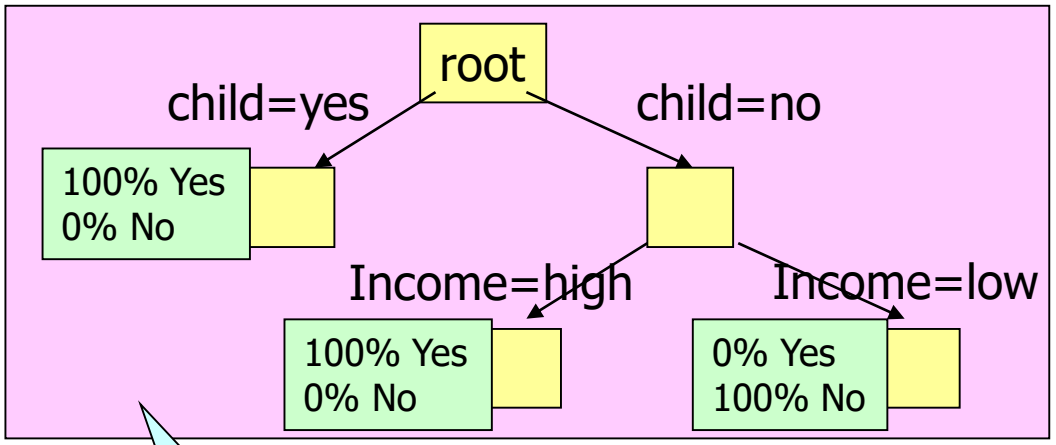



Decision tree

# 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

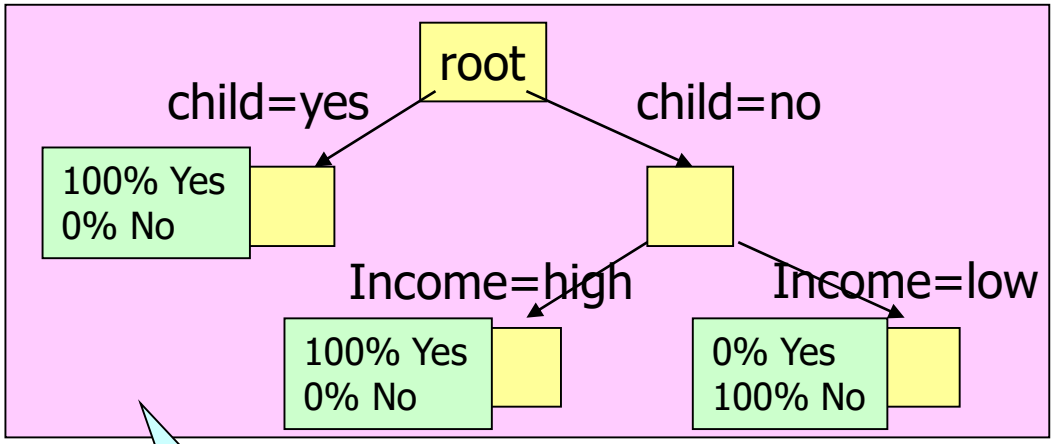
Confusion Matrix

| | Predicted Class | |
|---|---|---|
| Actual Class | Yes | No |
| Yes | 4 | 0 |
| No | 0 | 4 |

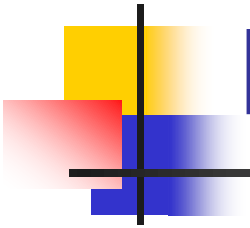
Is this decision tree "good"?

| 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

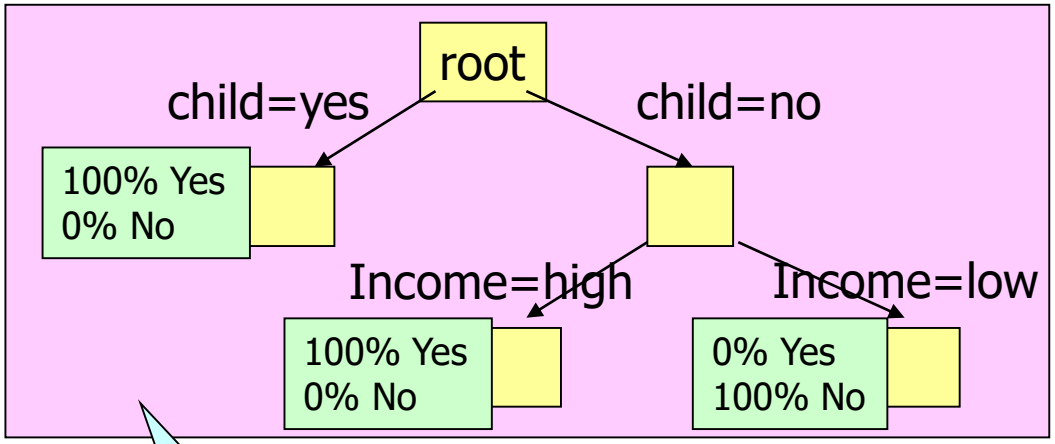
# Measure

Error Report

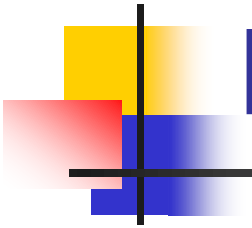
| Class | # Cases | # Errors | % Error |
|---|---|---|---|
| Yes | 4 | 0 | 0.00 |
| No | 4 | 0 | 0.00 |
| Overall | 8 | 0 | 0.00 |

| 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

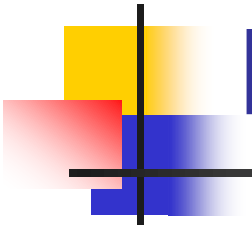



Decision tree



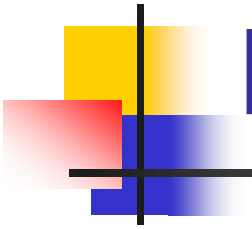
# Measurement

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



# Measurement - Lift Chart

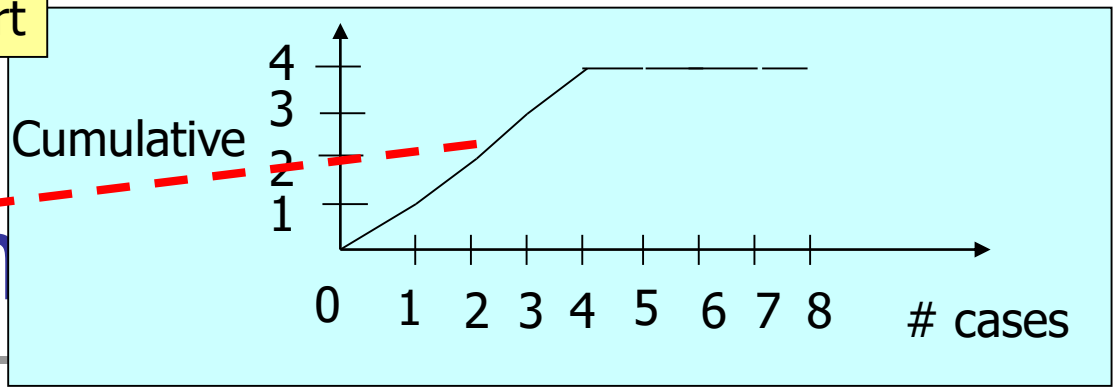
- Lift charts
  - visual aids for measuring model performance
  - consist of a lift curve and a baseline
  - The greater the area between the lift curve and the baseline, the better the model.



# 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



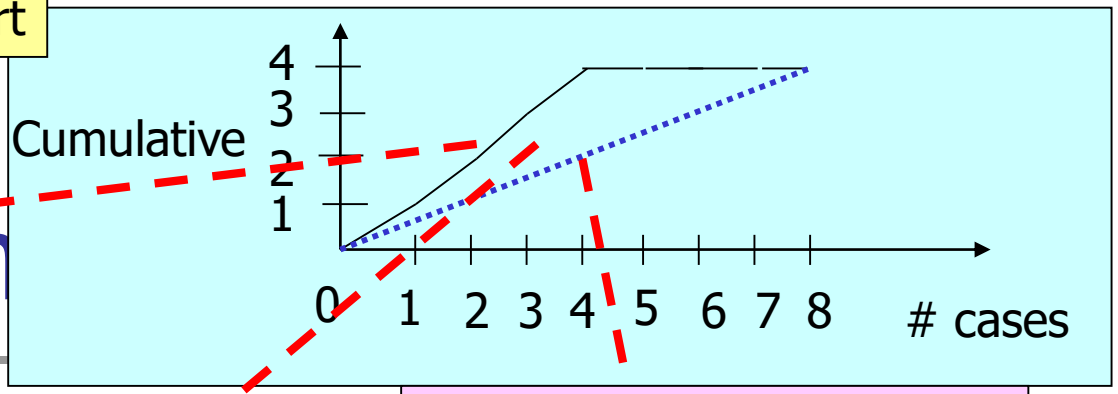

Cumulative Insurance of actual values (or Lift Curve)

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 |
|---|---|---|---|---|
| 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 |

| 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 |

Lift Chart




Cumulative Insurance of actual values (or Lift Curve)

Cumulative Insurance using Average (or Baseline)

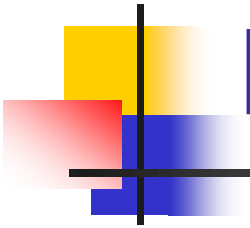
Sort the tuples according to the predicted value wh
sorted at a higher
value

Area between the lift curve and the baseline

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

| Race | Income | Child | Actual | Predicted |
|---|---|---|---|---|
| black | high | no | yes | yes |
| white | high | yes | yes | yes |
| wh | | | | |
| 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 |

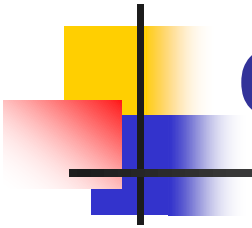
| Race | Income | Child | Actual | Predicted |
|---|---|---|---|---|
| black | high | no | yes | yes |
| white | high | yes | yes | yes |
| | | | yes | yes |
| | | | yes | yes |
| black | low | no | no | no |
| black | low | no | no | no |
| black | low | no | no | no |
| white | low | no | no | no |



# Measurement

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

# Measurement - Decile-wise lift chart

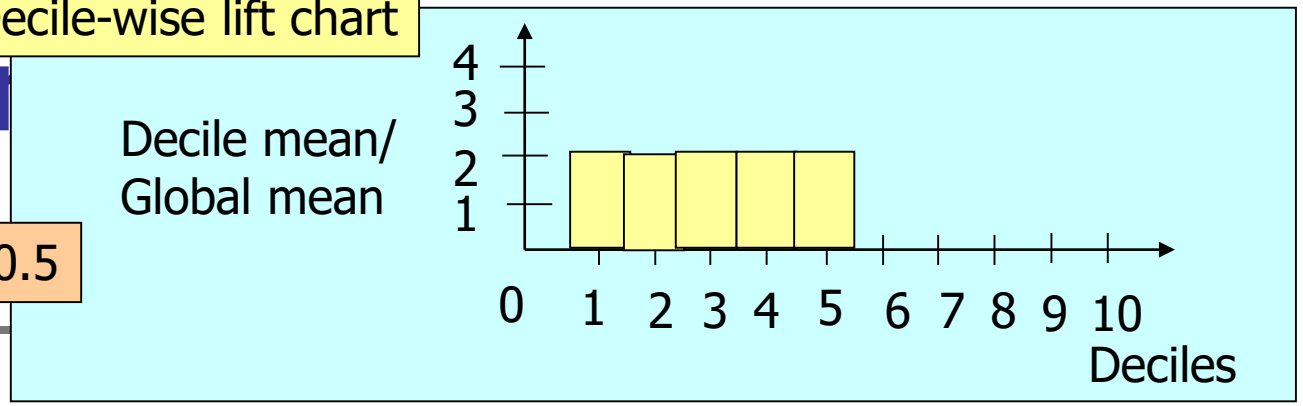


- A **decile** is any of the nine values that divide the sorted data into ten equal parts, so that each part represents 1/10 of the sample or population.
- E.g., 1st decile: the first 10% tuples
- E.g., 2nd decile: the second 10% tuples
- E.g., 3rd decile: the third 10% tuples

# Measur chart

Global mean = 4/8 = 0.5

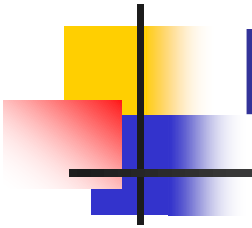
Decile-wise lift chart




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

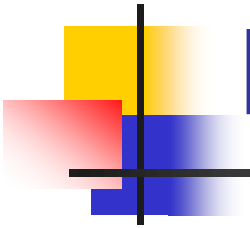
| Race | Income | Child | Actual | Predicted |
|---|---|---|---|---|
| black | high | Decile mean = 1.0 | | 1st Decile |
| white | high | Decile mean = 1.0 | | 2nd Decile |
| white | low | Decile mean = 1.0 | | 3rd Decile |
| white | low | Decile mean = 1.0 | | 4th Decile |
| black | low | Decile mean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

| Race | Income | Child | Actual | Predicted |
|---|---|---|---|---|
| blac | 12.5% | no | yes | yes |
| whit | 25% gh | yes | yes | yes |
| whit | 37.5% | yes | yes | yes |
| whit | 50% w | yes | yes | yes |
| blac | 62.5% | no | no | no |
| blac | 75% w | no | no | no |
| blac | 87.5% | no | no | no |
| whit | 100% v | no | no | no |



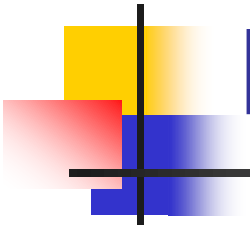
# Measurement

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



# Measurement - Others

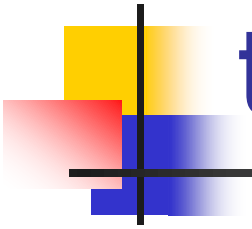
- We will discuss them later.
- E.g., Precision, Recall, Specificity, f1-score



# Decision Trees

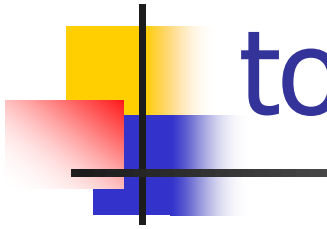
- Decision Trees
  - ID3
  - C4.5
  - CART
- Measurement
- How to use the data mining tool

# 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 | ? |

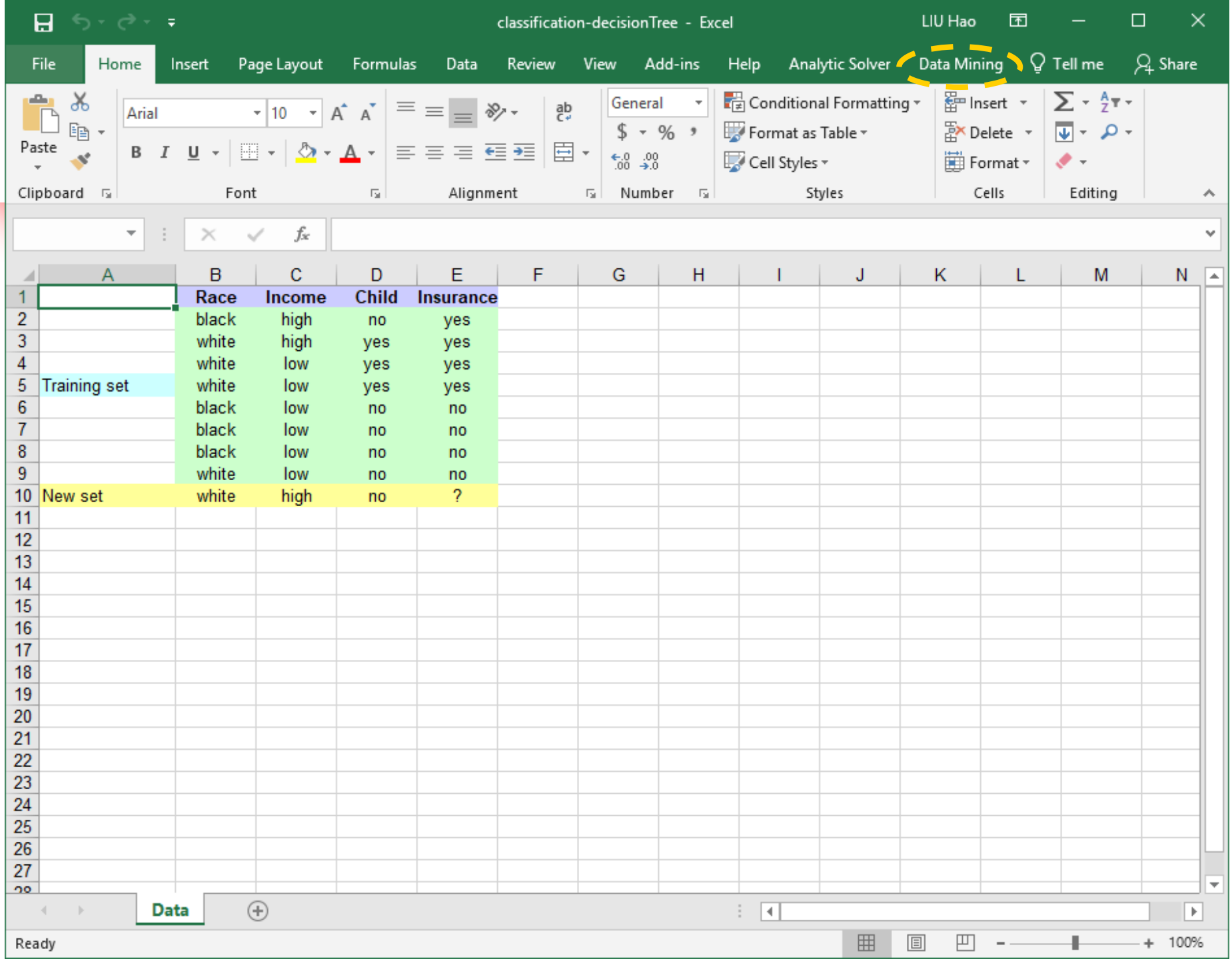
| Race | Income | Child | Insurance |
|---|---|---|---|
| 1 | 1 | 1 | yes |
| 2 | 1 | 2 | yes |
| 2 | 2 | 2 | yes |
| 2 | 2 | 2 | yes |
| 1 | 2 | 1 | no |
| 1 | 2 | 1 | no |
| 1 | 2 | 1 | no |
| 2 | 2 | 1 | no |

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

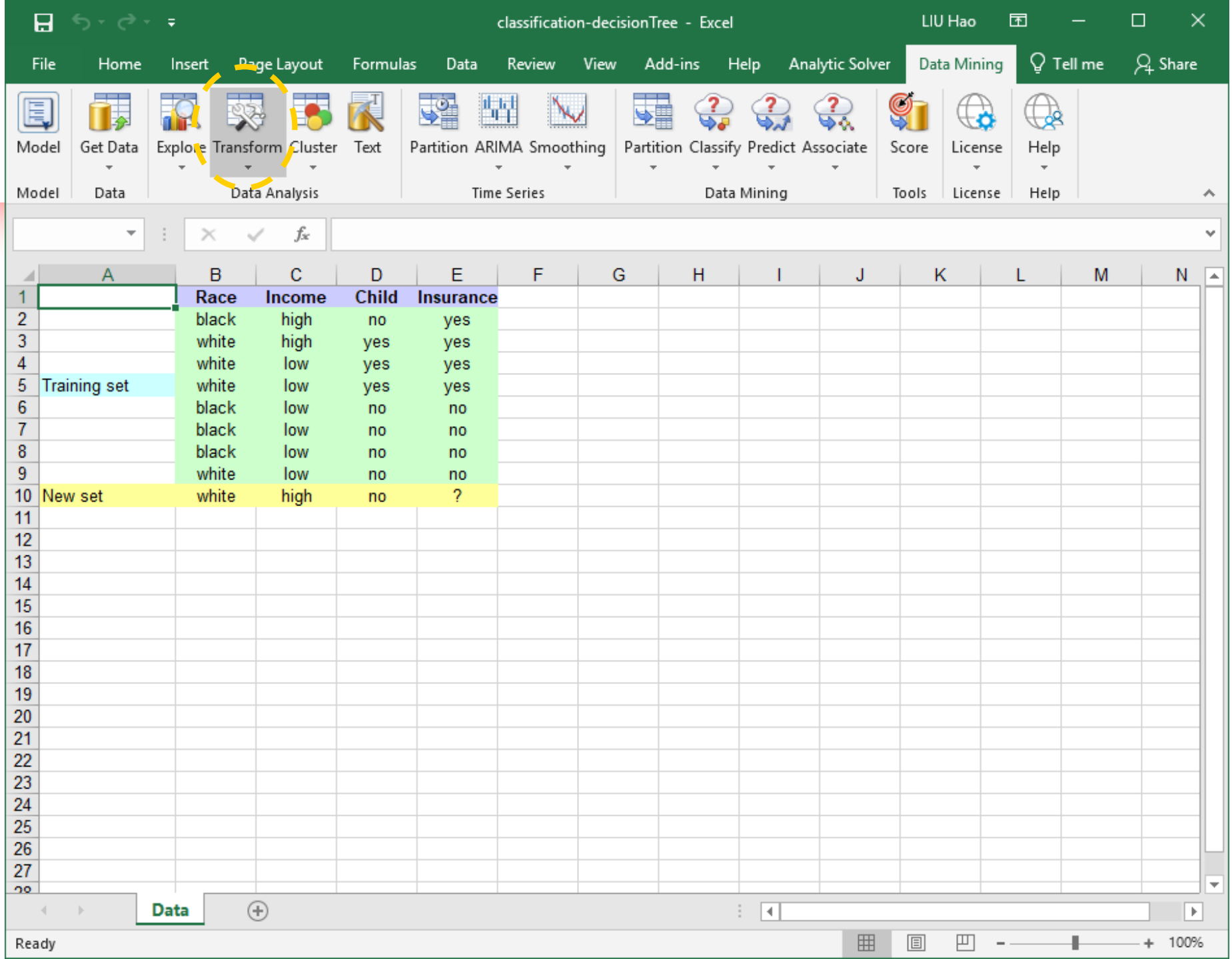
Input attributes:
- Categorical

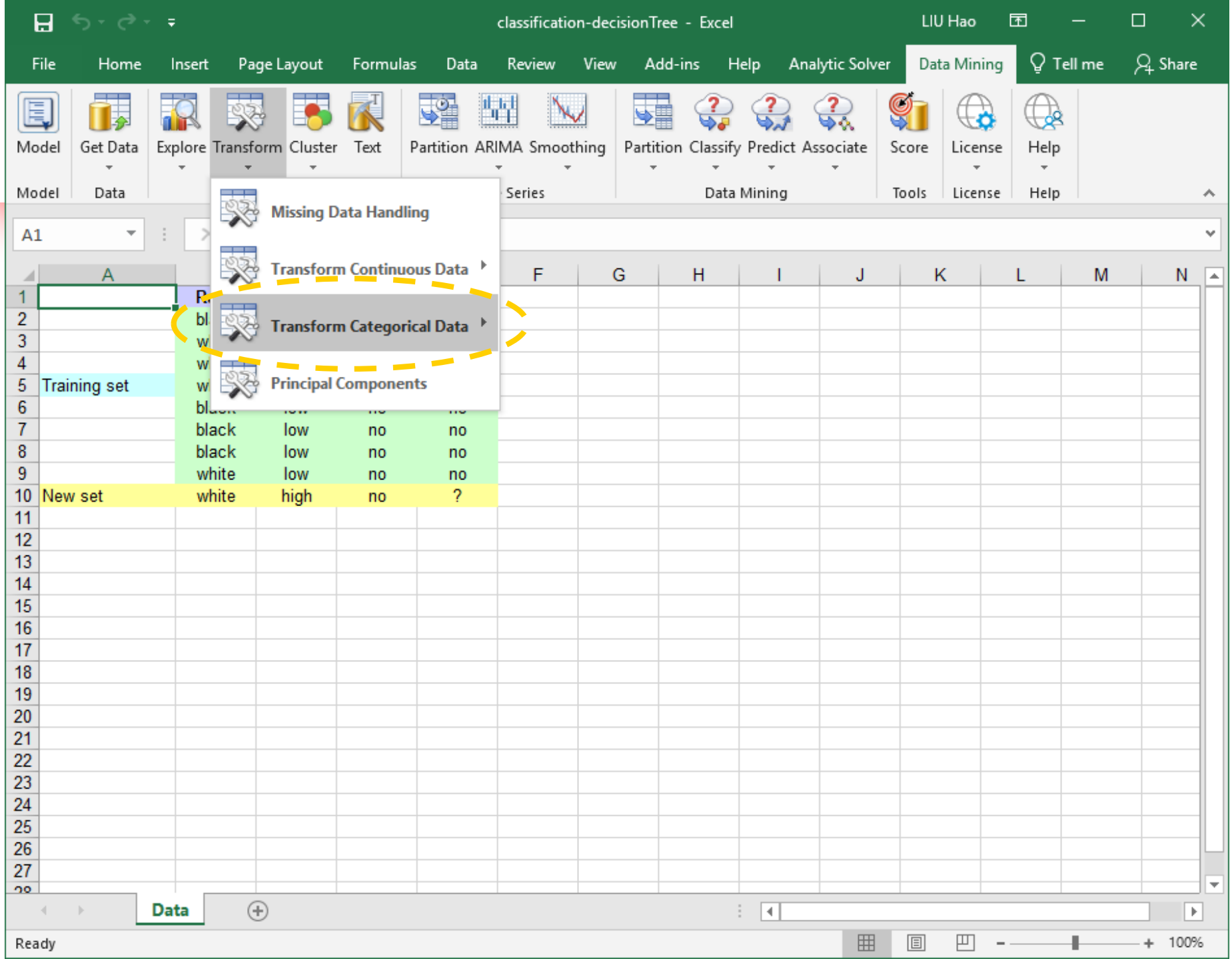
Target attribute:
- Categorical

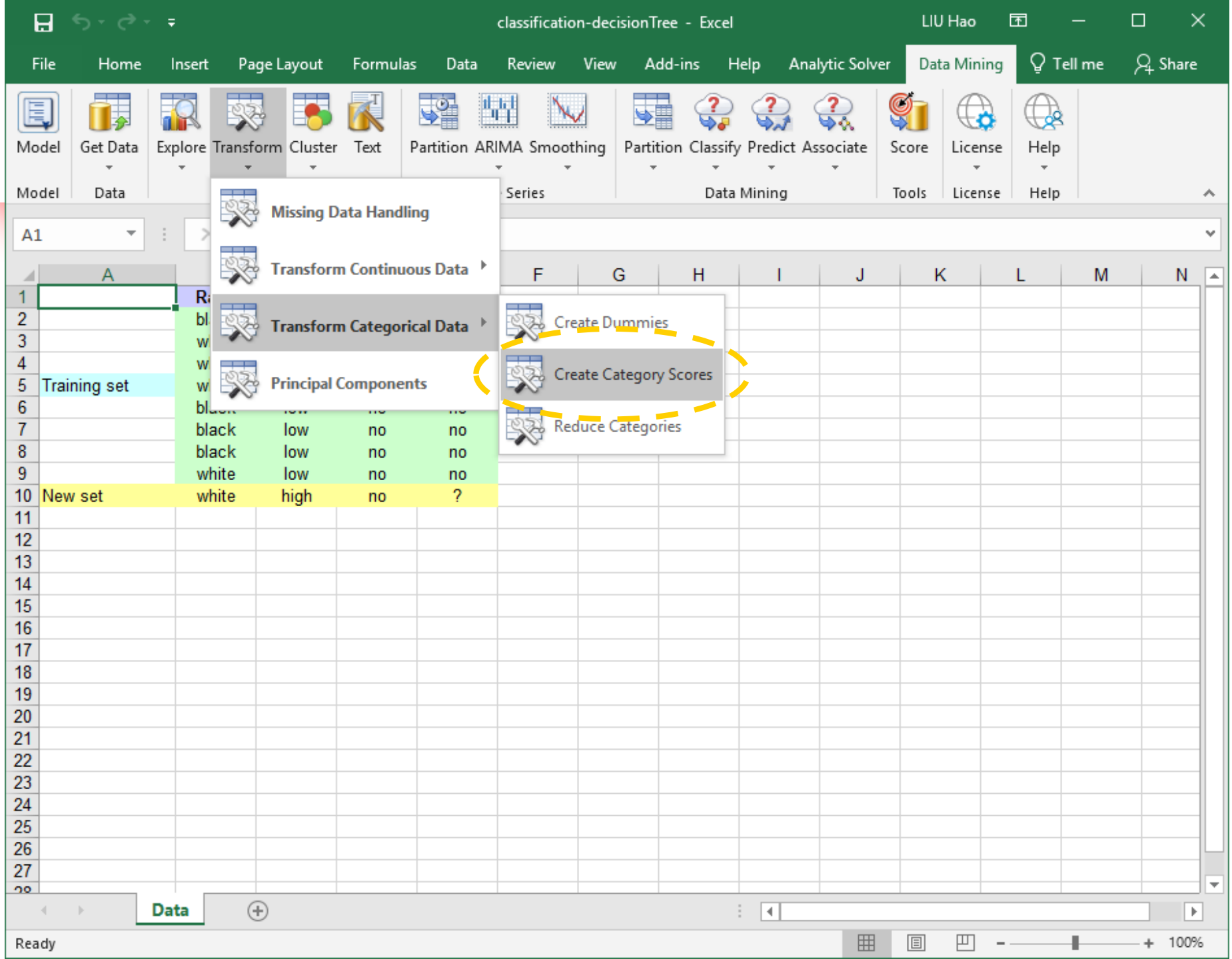
We can transform from "categorical" to "numeric" by XLMiner Transformation Tool first

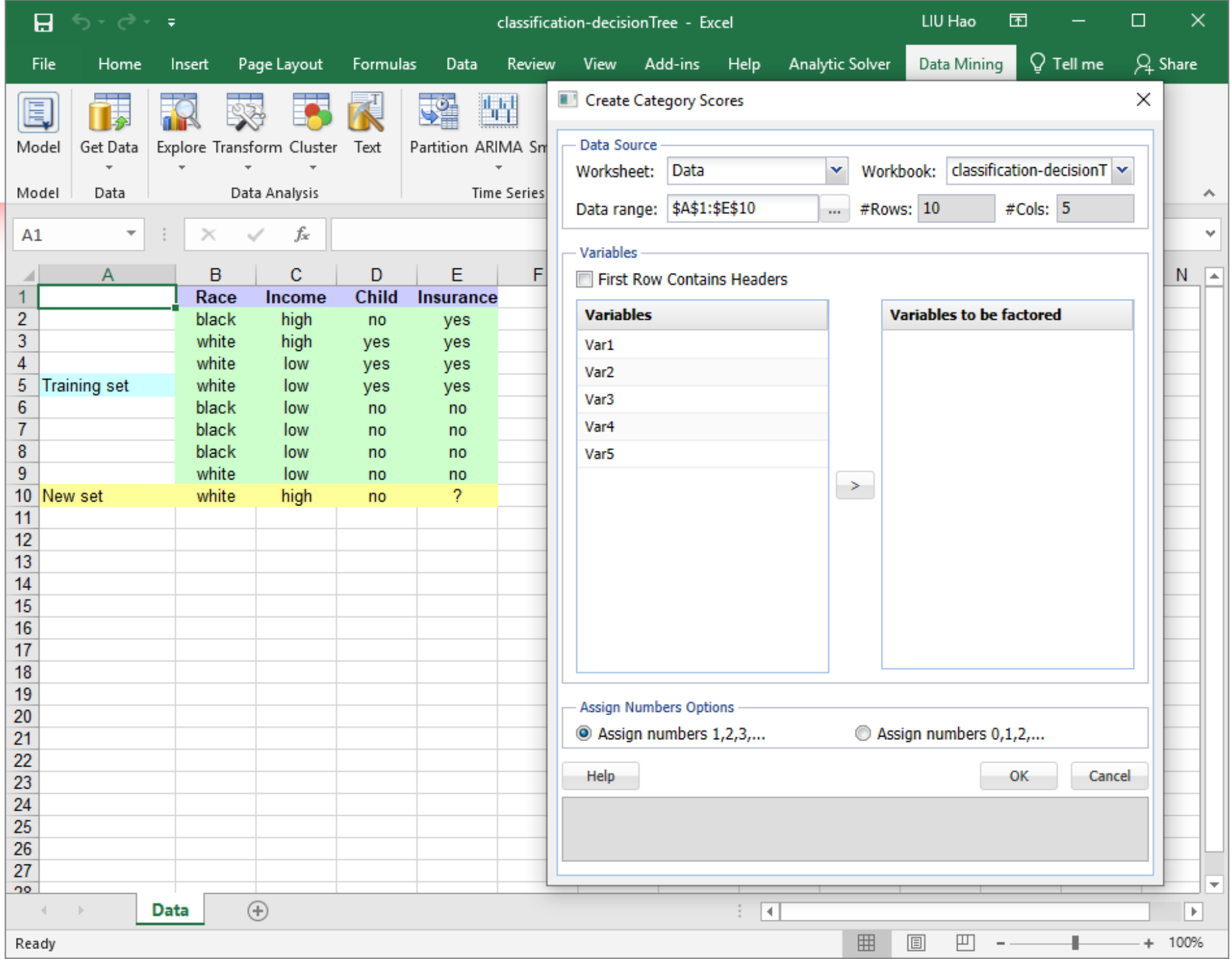


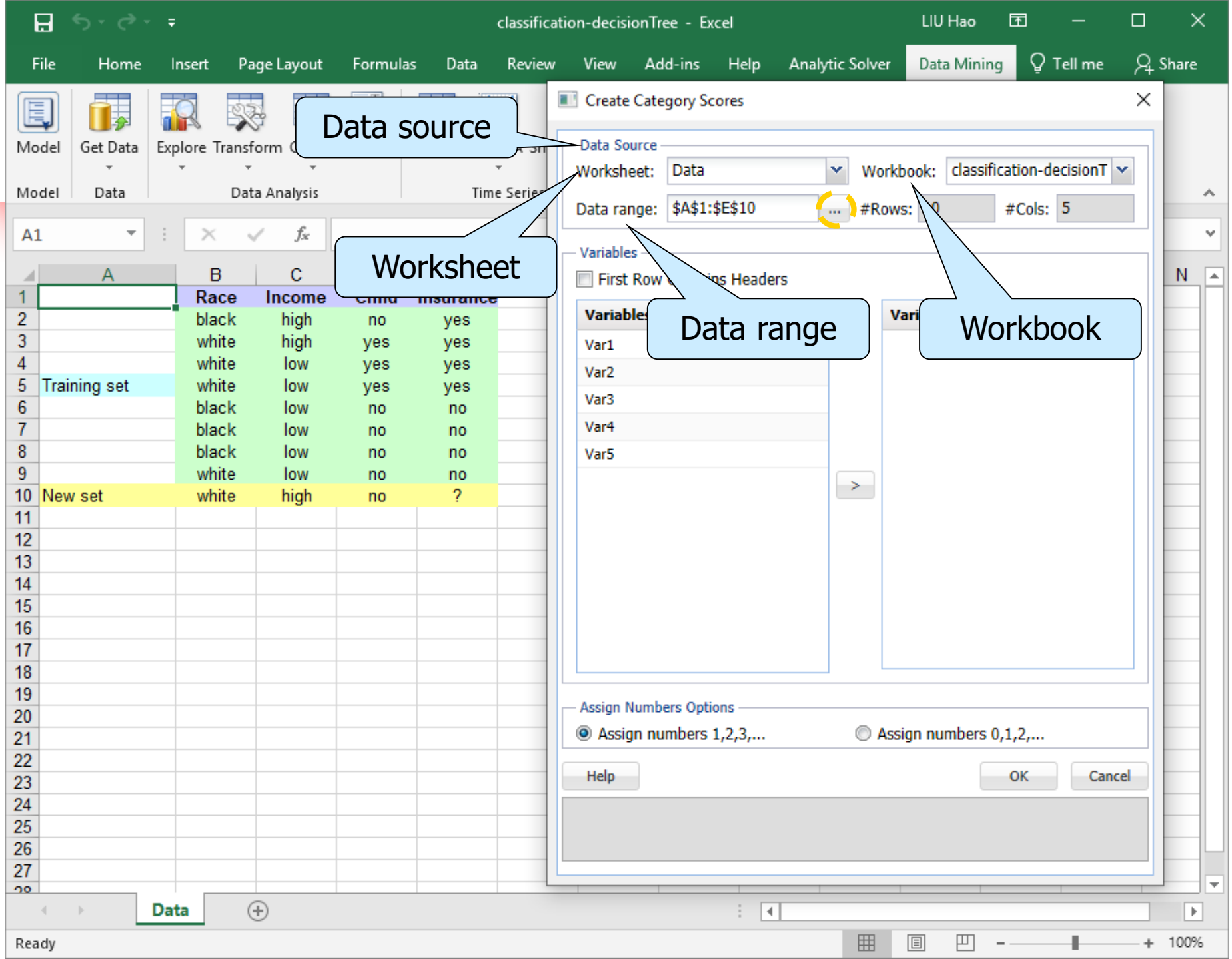

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | | Race | Income | Child | Insurance |
| 2 | | black | high | no | yes |
| 3 | | white | high | yes | yes |
| 4 | | white | low | yes | yes |
| 5 | Training set | white | low | yes | yes |
| 6 | | black | low | no | no |
| 7 | | black | low | no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


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
Race
Income
Child
Insurance
black high no yes
white high yes yes
white low yes yes
Training set
white low yes yes
black low no no
black low no no
black low no no
white low no no
New set
white high no ?
Data
Ready
100%

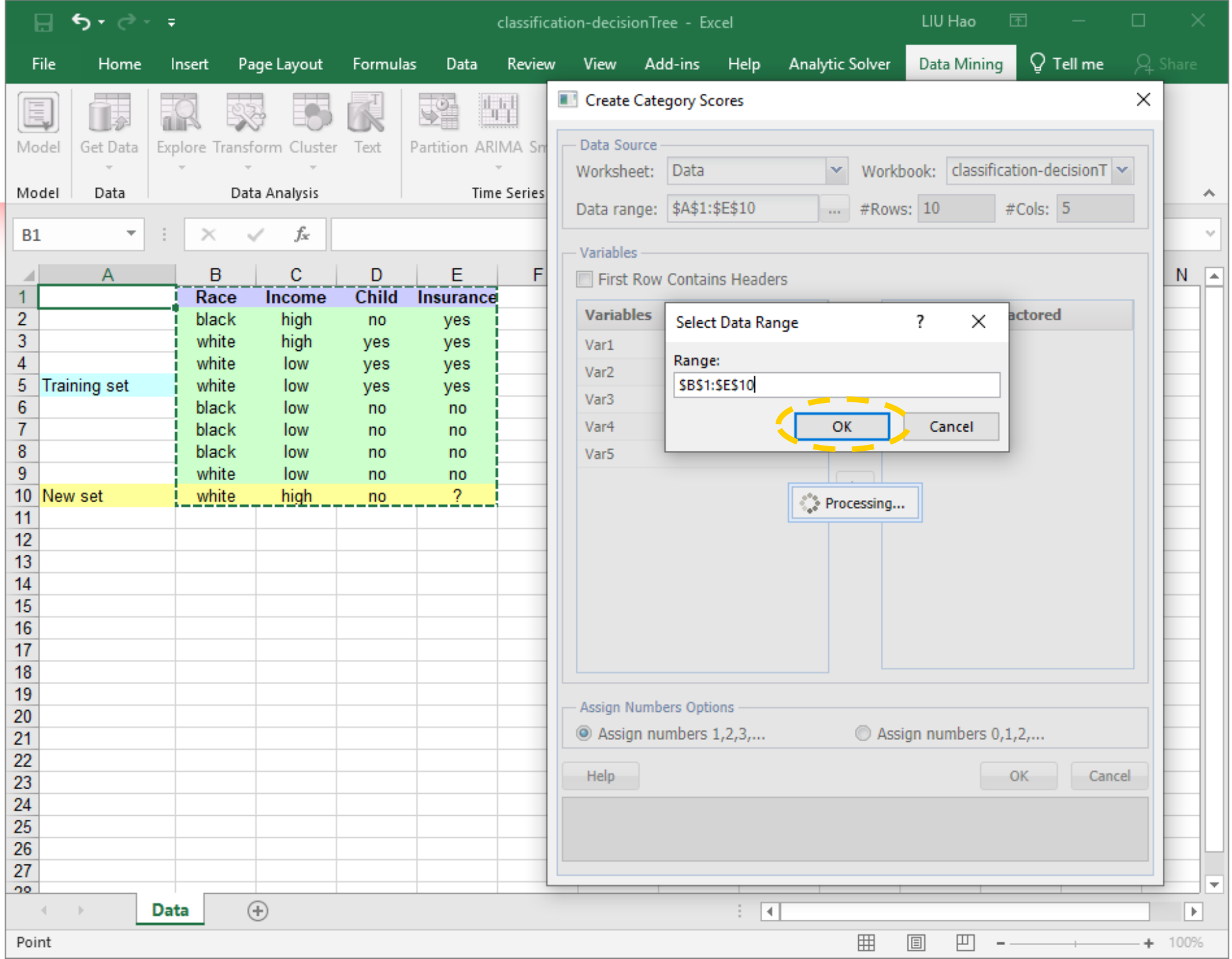

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Series
Data Mining
Tools
License
Help
Missing Data Handling
Transform Continuous Data
Transform Categorical Data
Principal Components
A1
Training set
New set
black low no no
black low no no
white low no no
white high no ?
Data
Ready
100%

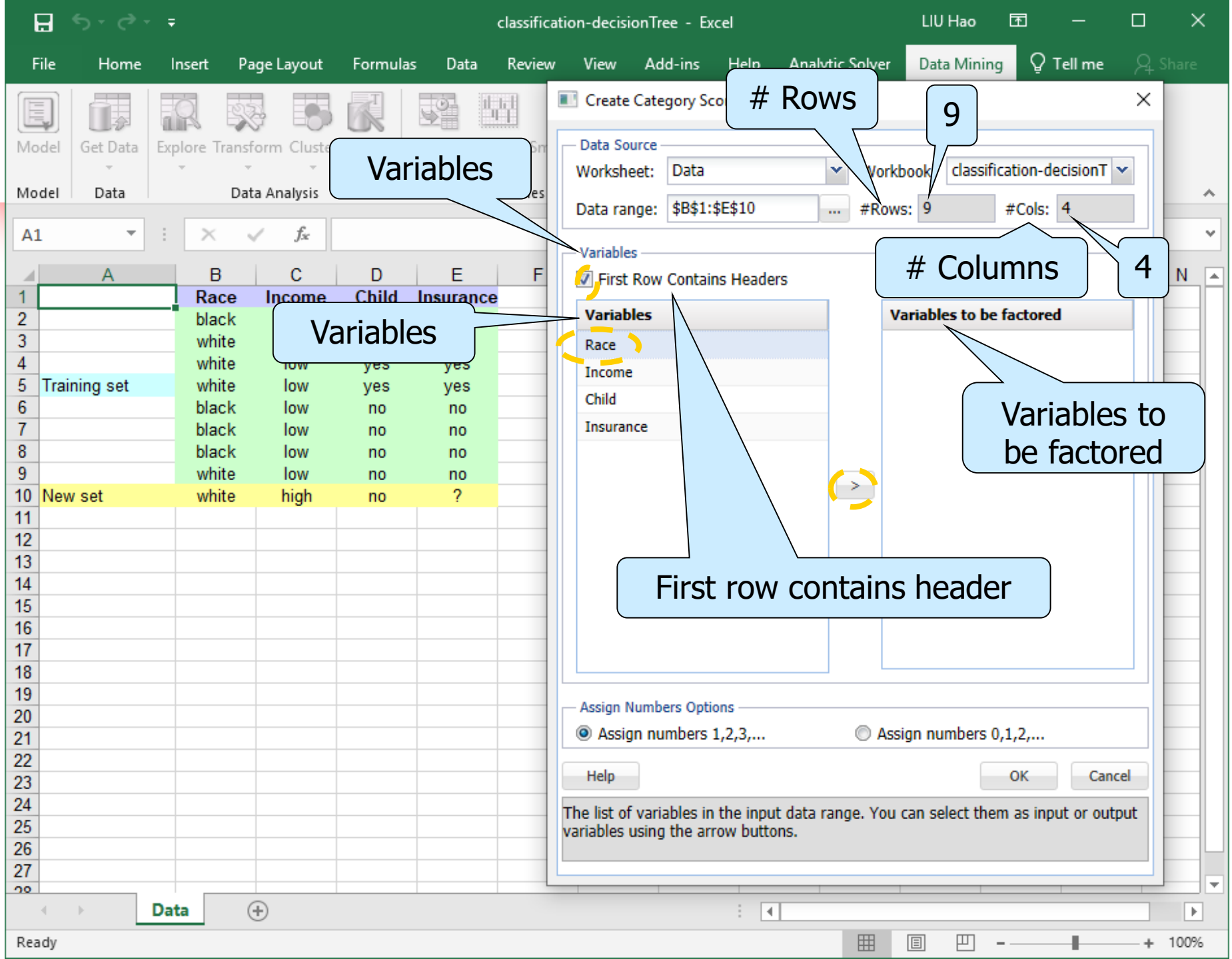

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Series
Data Mining
Tools
License
Help
Missing Data Handling
Transform Continuous Data
Transform Categorical Data
Principal Components
Create Dummies
Create Category Scores
Reduce Categories
A1
Training set
New set
black low no no
black low no no
white low no no
white high no ?
Data
Ready
100%

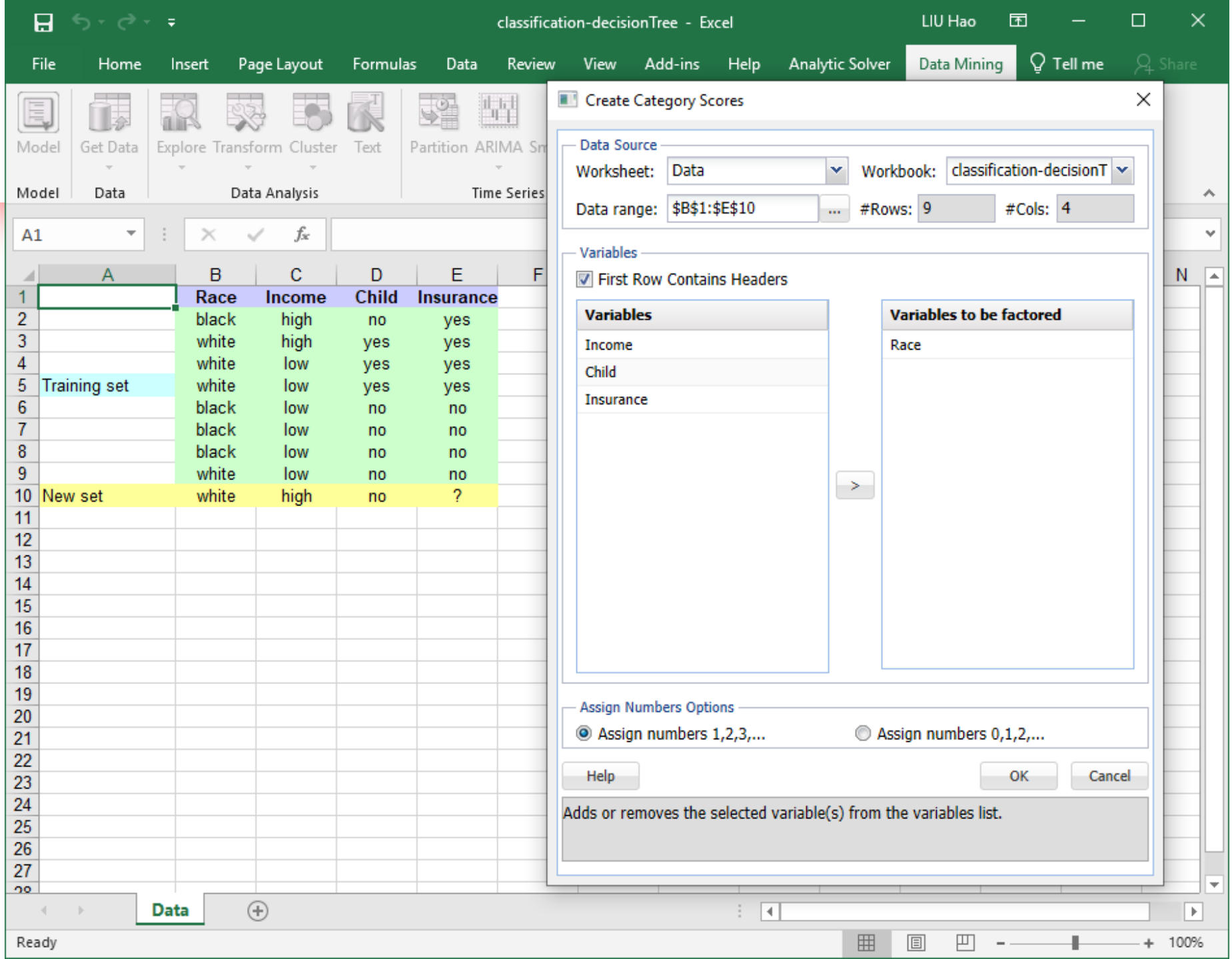

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Model
Data
Data Analysis
Time Series
A1
Race
Income
Child
Insurance
Training set
New set
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
white high no ?
Data
Ready
100%
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



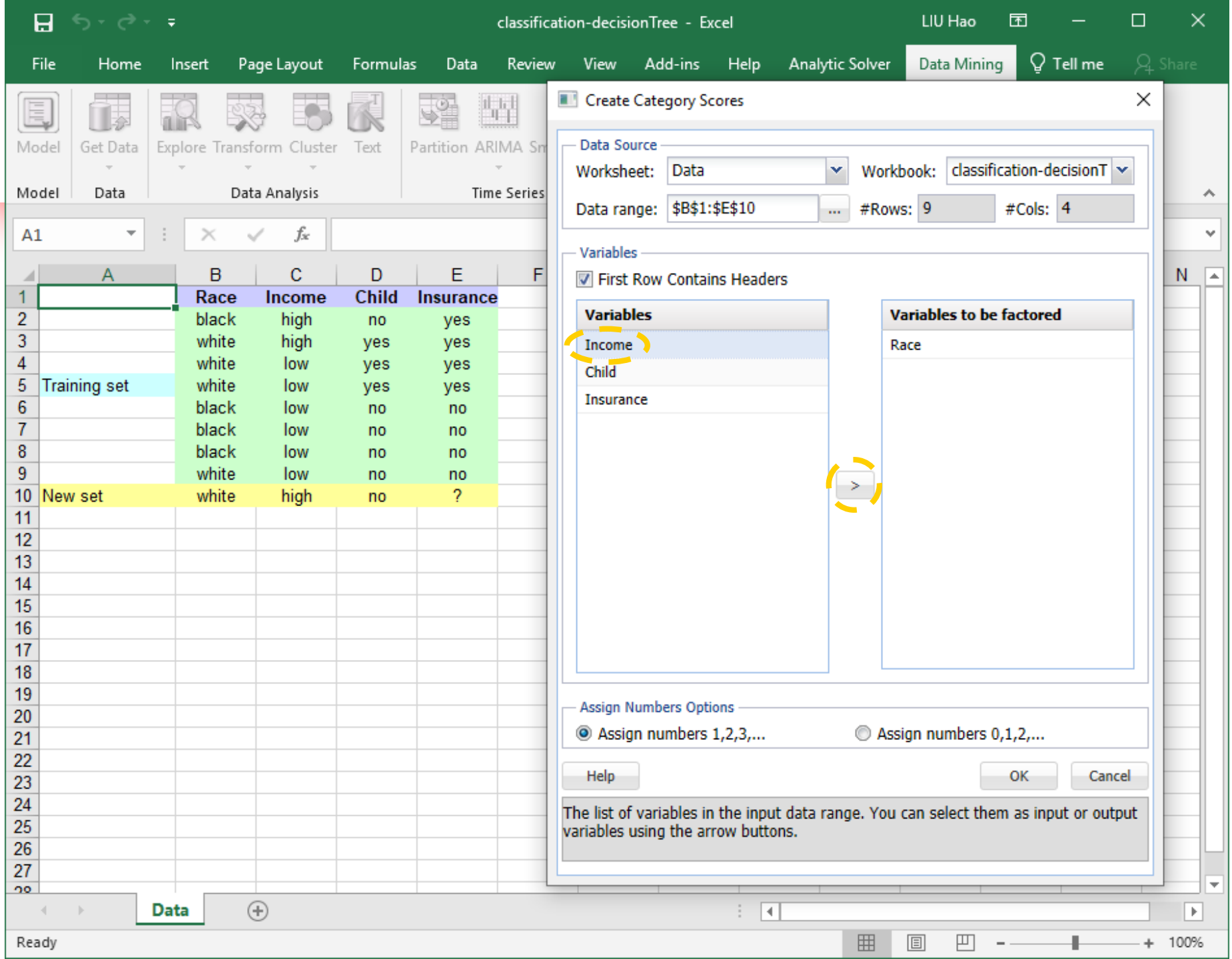

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | | Race | Income | Child | Insurance |
| 2 | | black | high | no | yes |
| 3 | | white | high | yes | yes |
| 4 | | white | low | yes | yes |
| 5 | Training set | white | low | yes | yes |
| 6 | | black | low | no | no |
| 7 | | black | low | no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

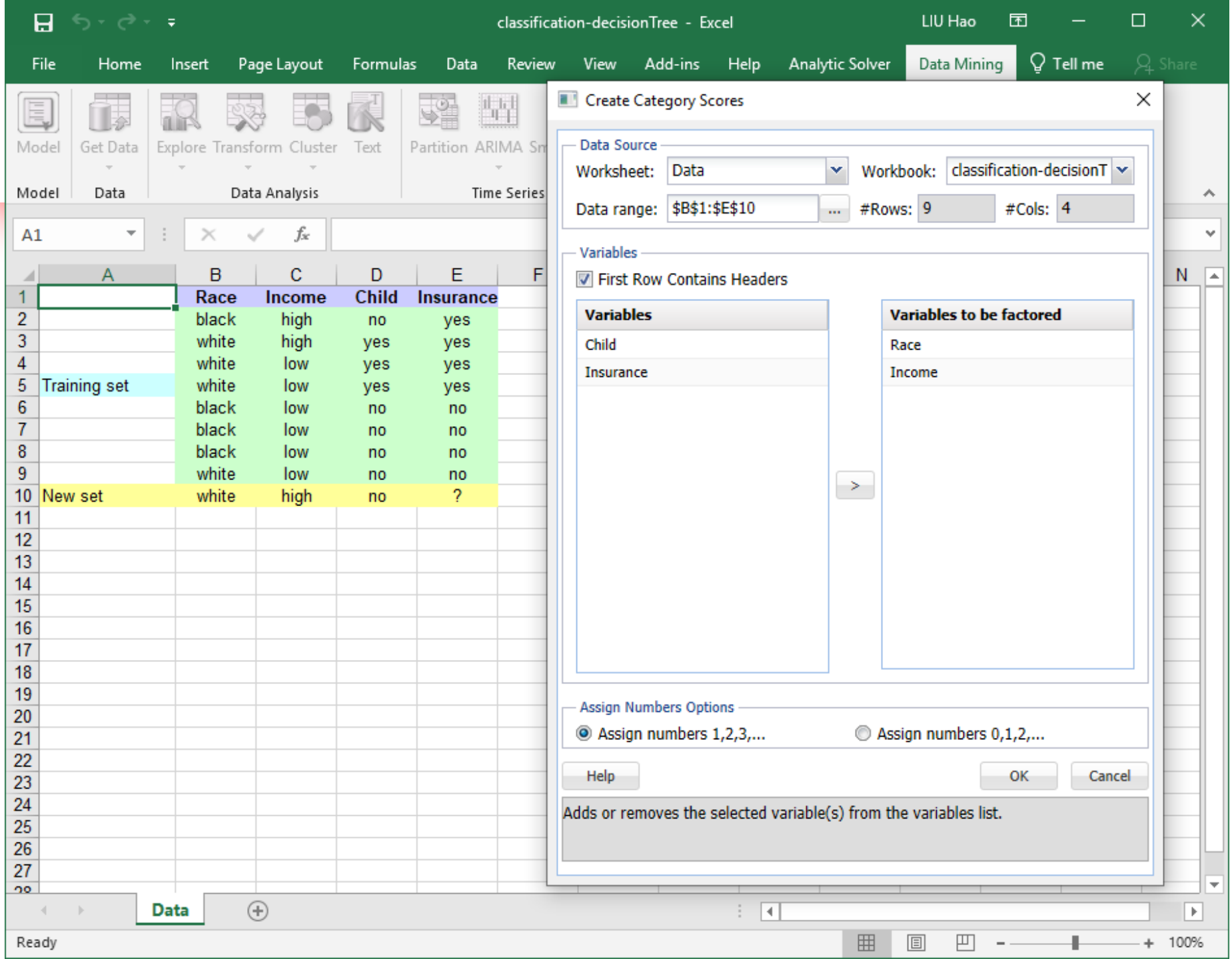

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Model
Data
Data Analysis
Time Series
B1
A
B
C
D
E
F
N
1 Race Income Child Insurance
2 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
Data
Point
100%
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
actored
Select Data Range
Range:
$B$1:$E$10
OK
Cancel
Processing...
Assign Numbers Options
Assign numbers 1,2,3,...
Assign numbers 0,1,2,...
Help
OK
Cancel


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
Data Analysis
A1
Race
Income
Child
Insurance
black
white
Training set
low
yes
no
New set
high
?
Data
Ready
100%
Create Category Sco
Data Source
Worksheet:
Data
Workbook:
classification-decisionT
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.
Variables
Variables
# Rows
9
# Columns
4
Variables to be factored
First row contains header

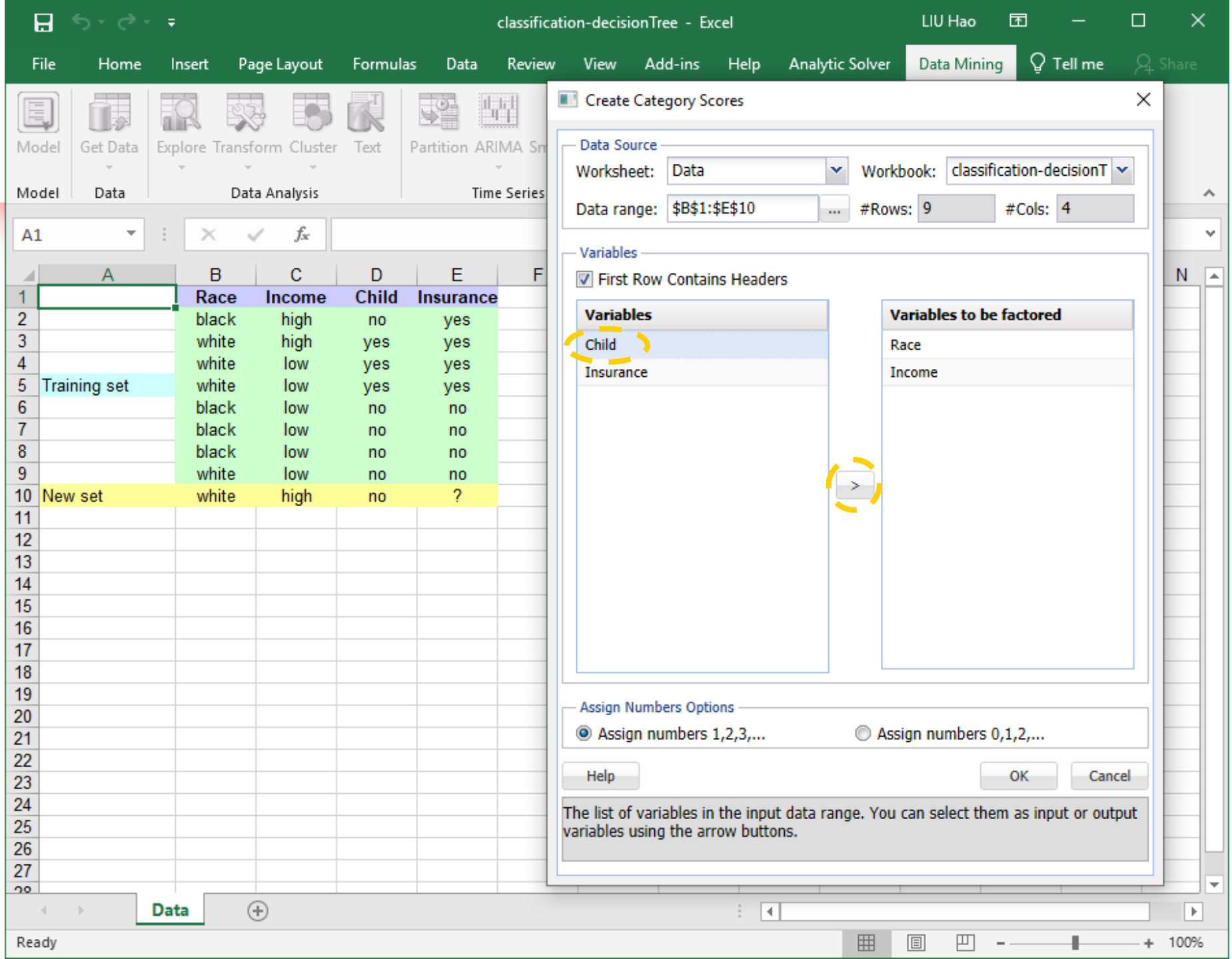


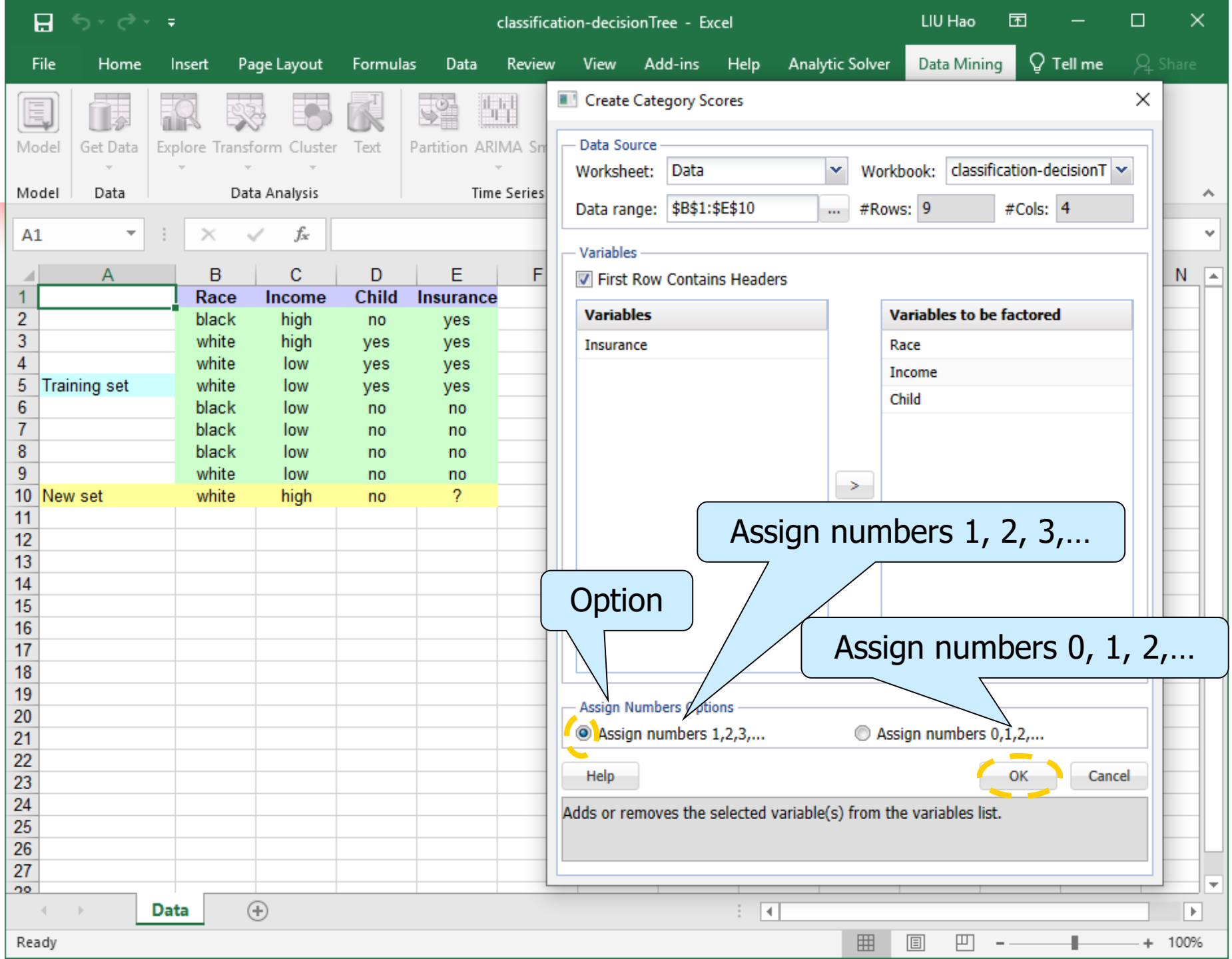

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | | Race | Income | Child | Insurance |
| 2 | | black | high | no | yes |
| 3 | | white | high | yes | yes |
| 4 | | white | low | yes | yes |
| 5 | Training set | white | low | yes | yes |
| 6 | | black | low | no | no |
| 7 | | black | low | no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

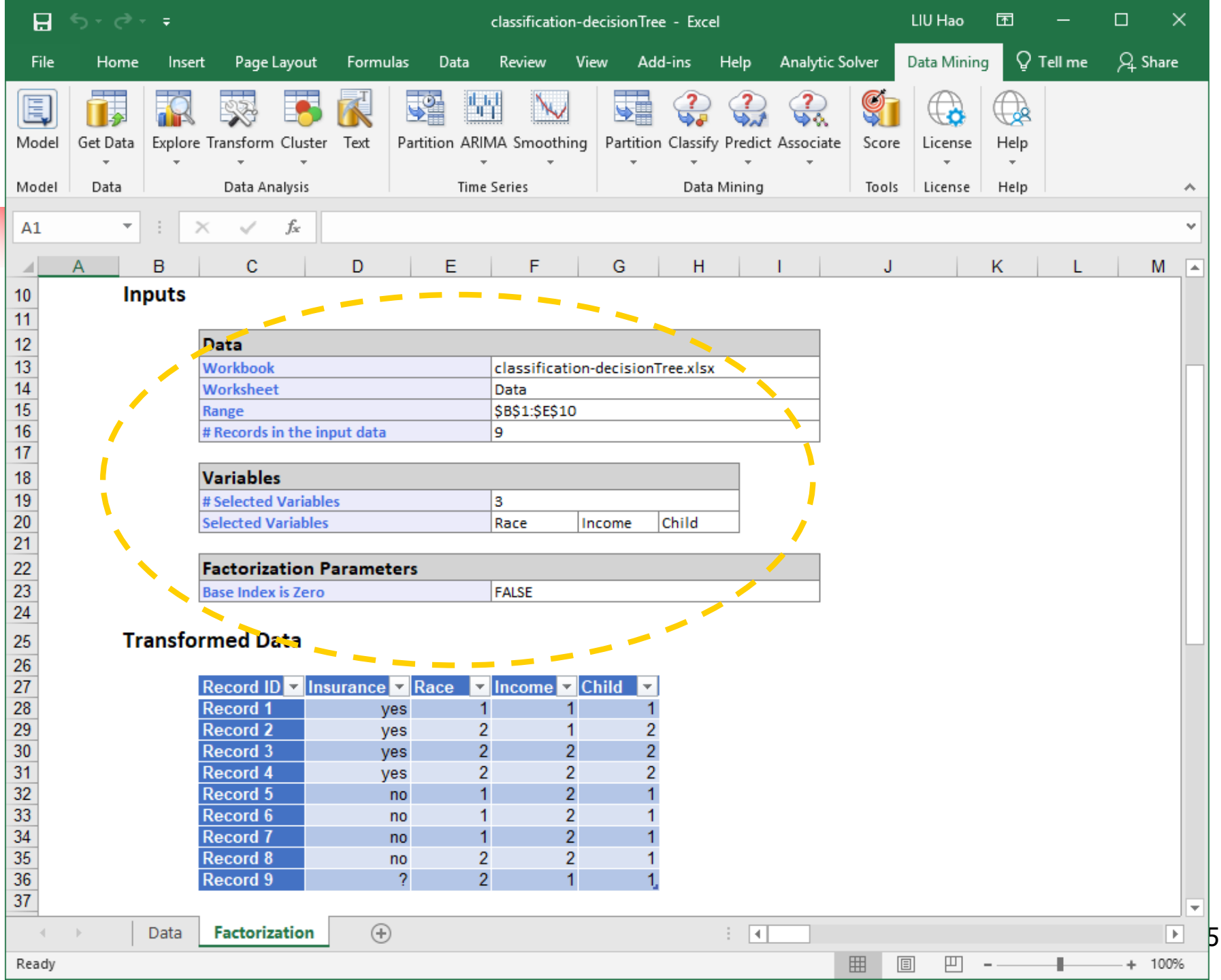

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA
Model Data Data Analysis Time Series
A1
Race Income Child Insurance
black high no yes
white high yes yes
white low yes yes
Training set white low yes yes
black low no no
black low no no
black low no no
white low no no
New set white high no ?
Data
Ready
100%
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.

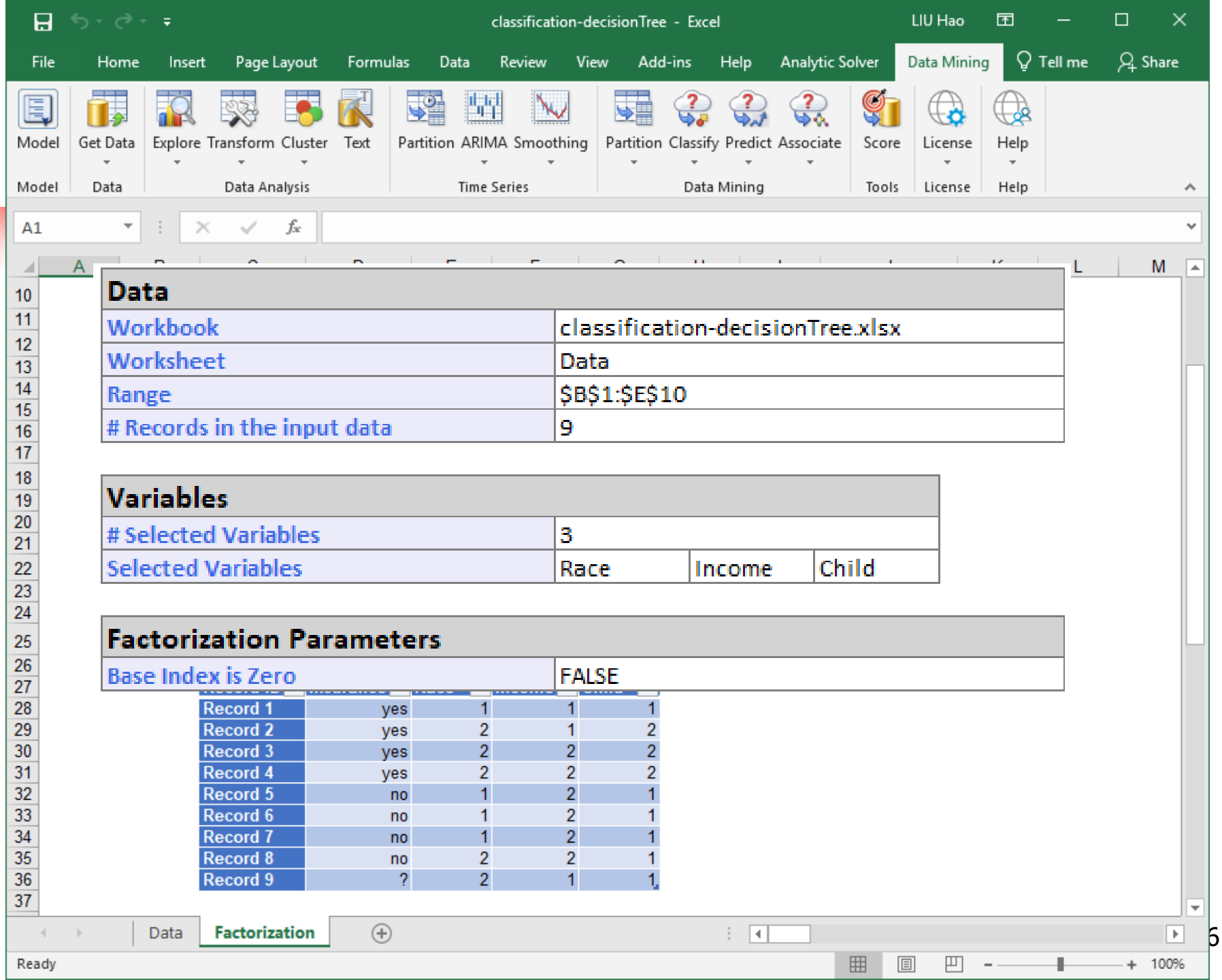


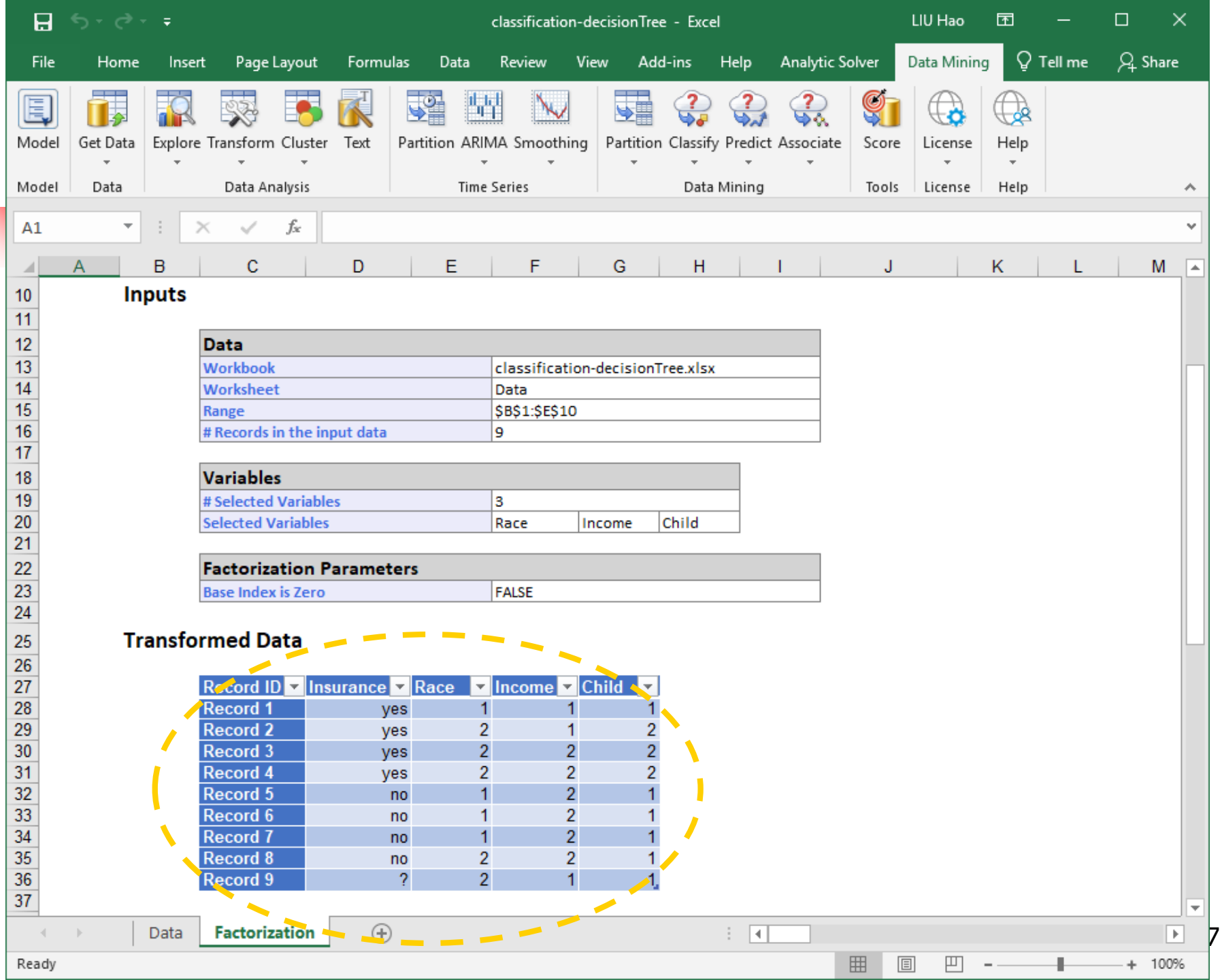

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | | Race | Income | Child | Insurance |
| 2 | | black | high | no | yes |
| 3 | | white | high | yes | yes |
| 4 | | white | low | yes | yes |
| 5 | Training set | white | low | yes | yes |
| 6 | | black | low | no | no |
| 7 | | black | low | no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

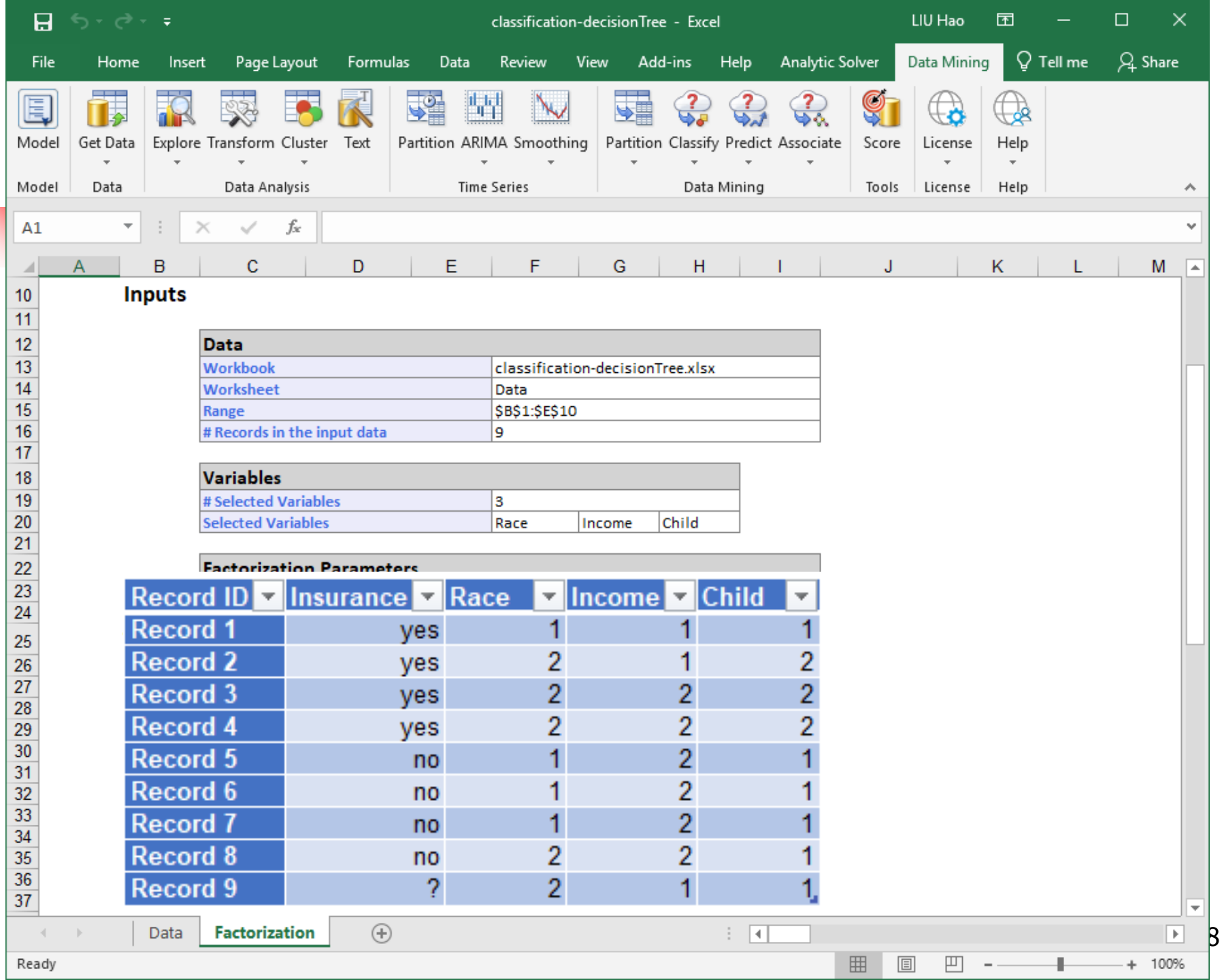

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA
Model Data Data Analysis Time Series
A1
Race Income Child Insurance
black high no yes
white high yes yes
white low yes yes
Training set white low yes yes
black low no no
black low no no
black low no no
white low no no
New set white high no ?
Data
Ready
100%
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.

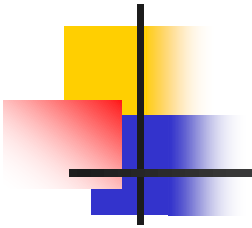

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Model
Data
Data Analysis
Time Series
A1
Race
Income
Child
Insurance
Training set
New set
Data
Ready
100%
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
Insurance
Variables to be factored
Race
Income
Child
>
Assign numbers 1, 2, 3,...
Option
Assign numbers 0, 1, 2,...
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.


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 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
Data Factorization
Ready
100%

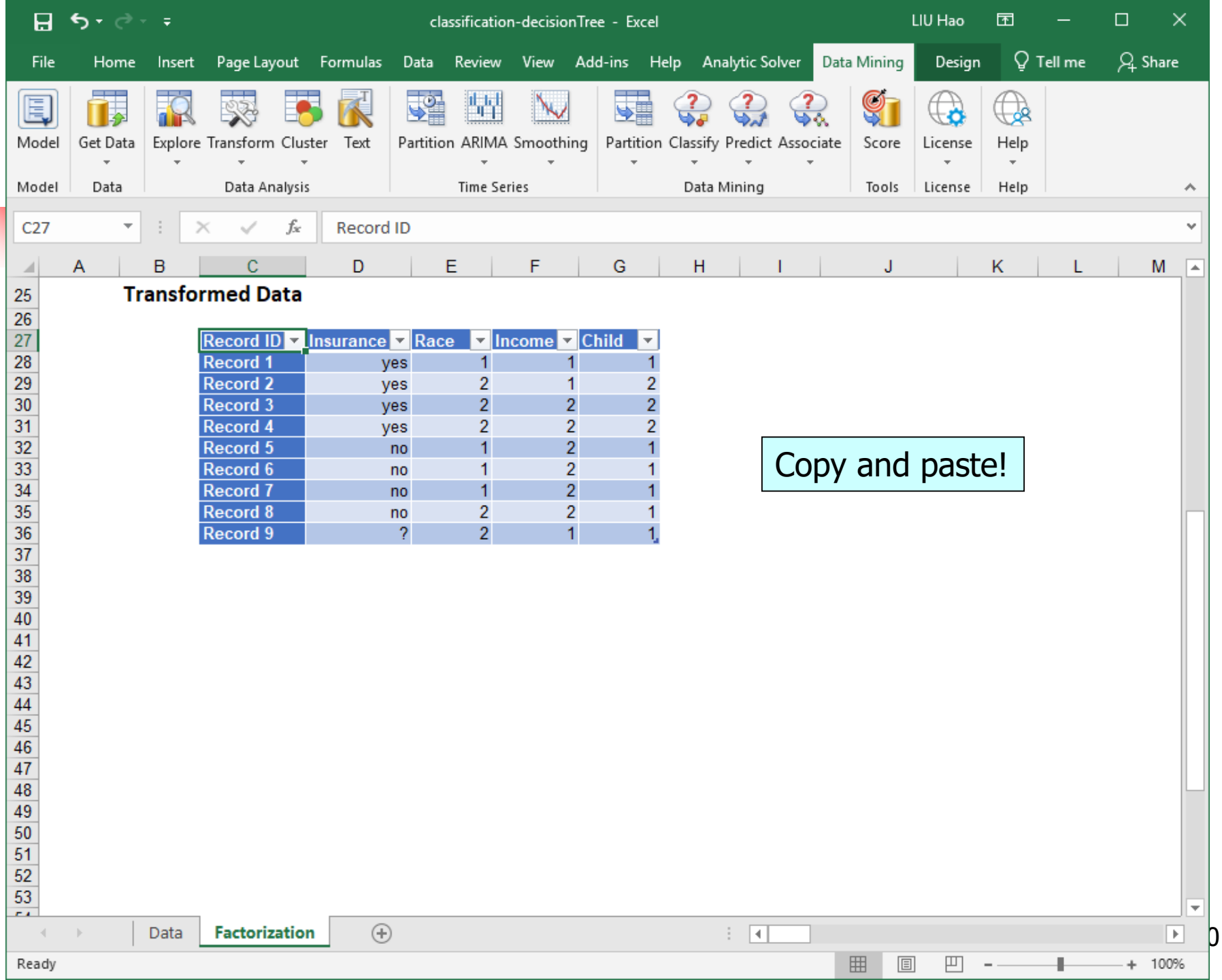

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
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
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
Ready
100%

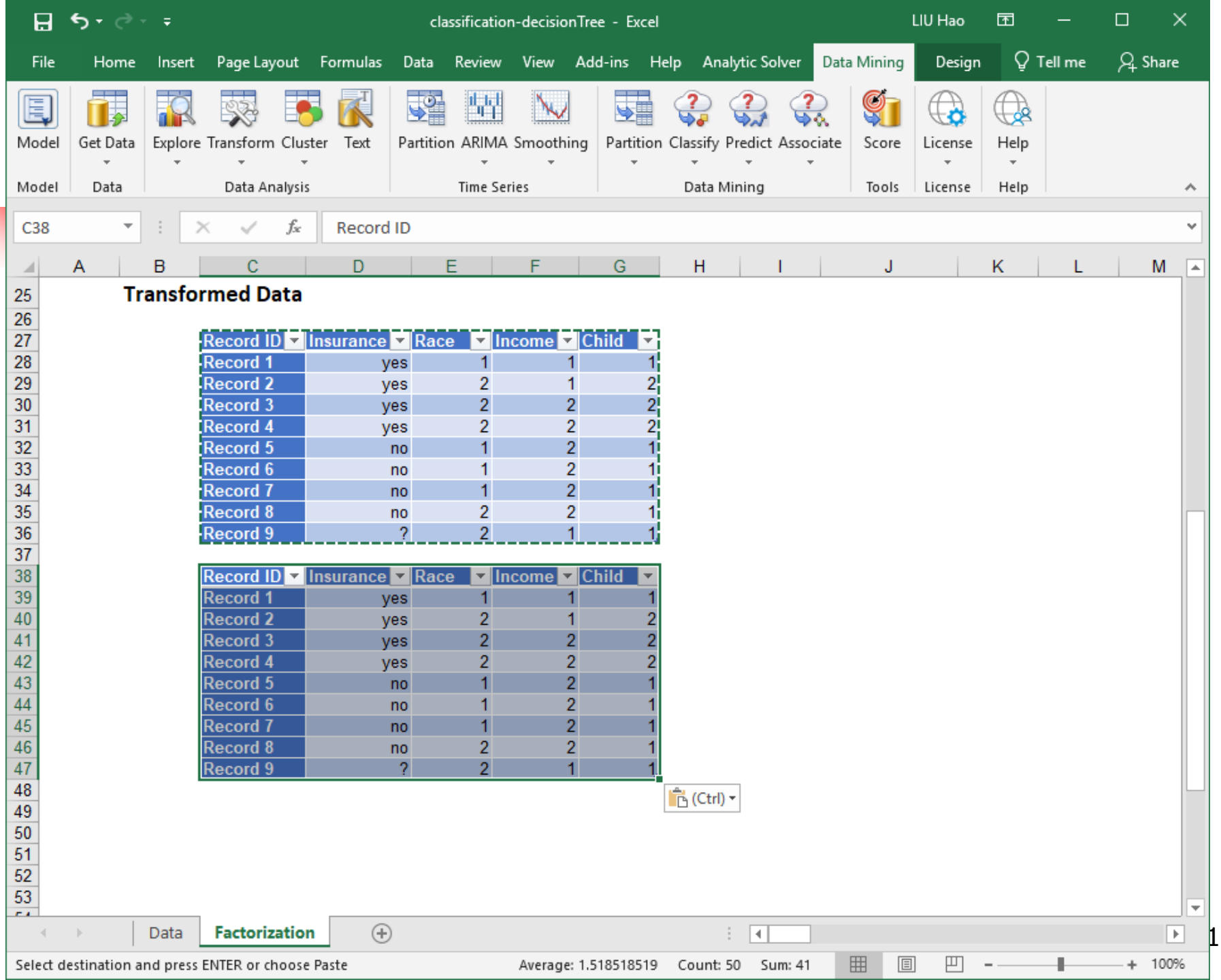

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook classification-decisionTree.xlsx
Worksheet 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
Data Factorization
Ready
100%

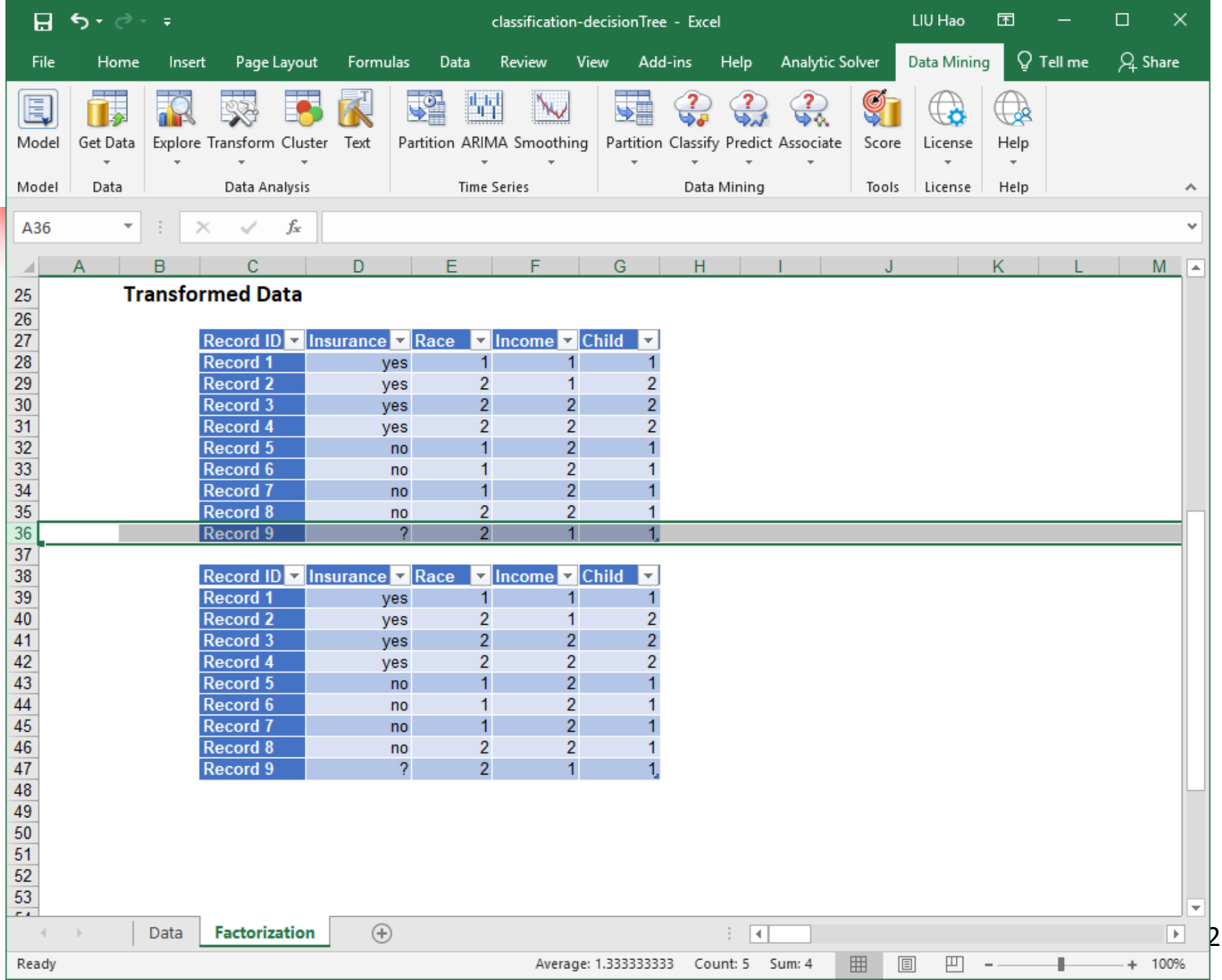

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Inputs
Data
Workbook classification-decisionTree.xlsx
Worksheet Data
Range $B$1:$E$10
# Records in the input data 9
Variables
# Selected Variables 3
Selected Variables Race Income Child
Factorization Parameters
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
Ready
100%

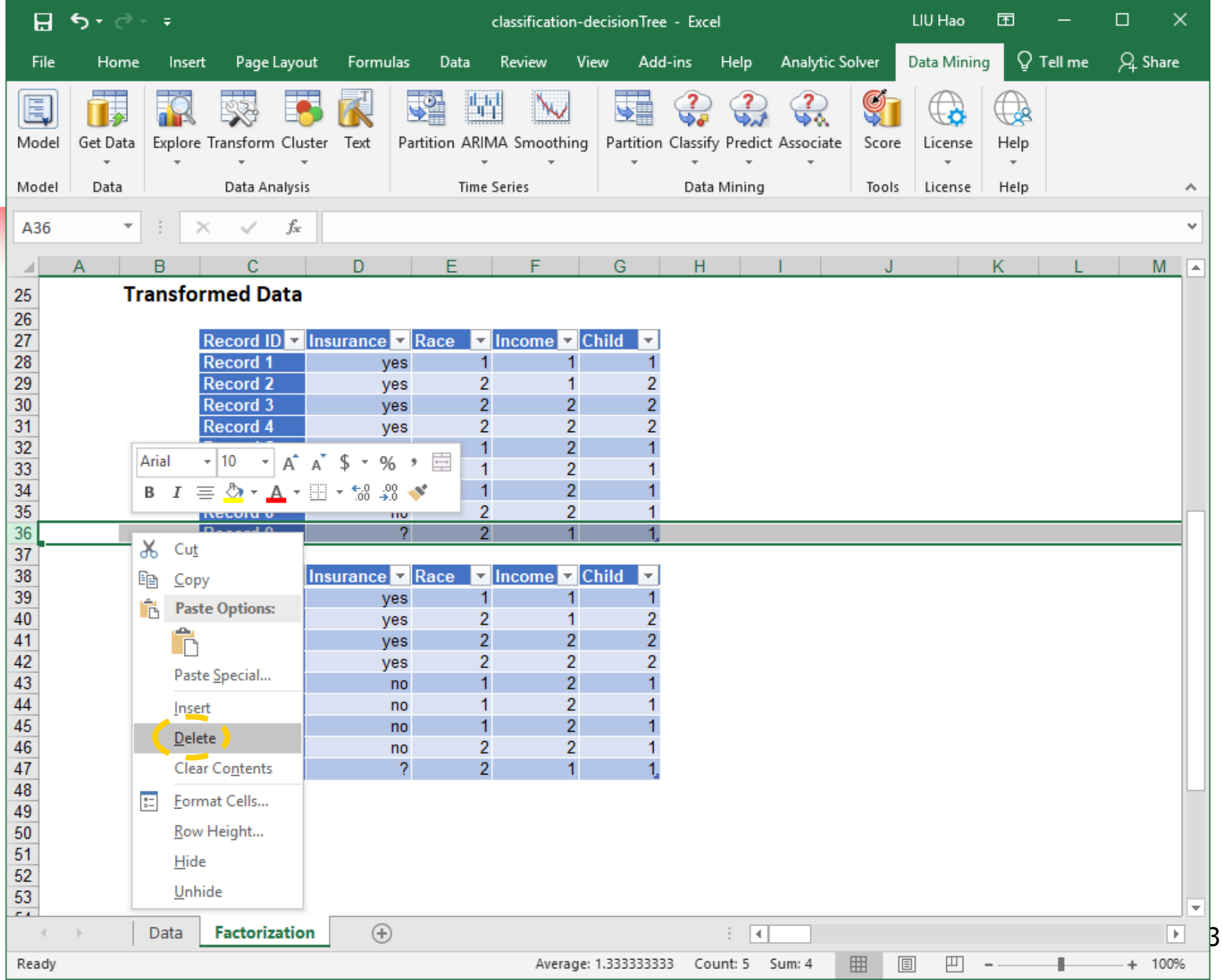


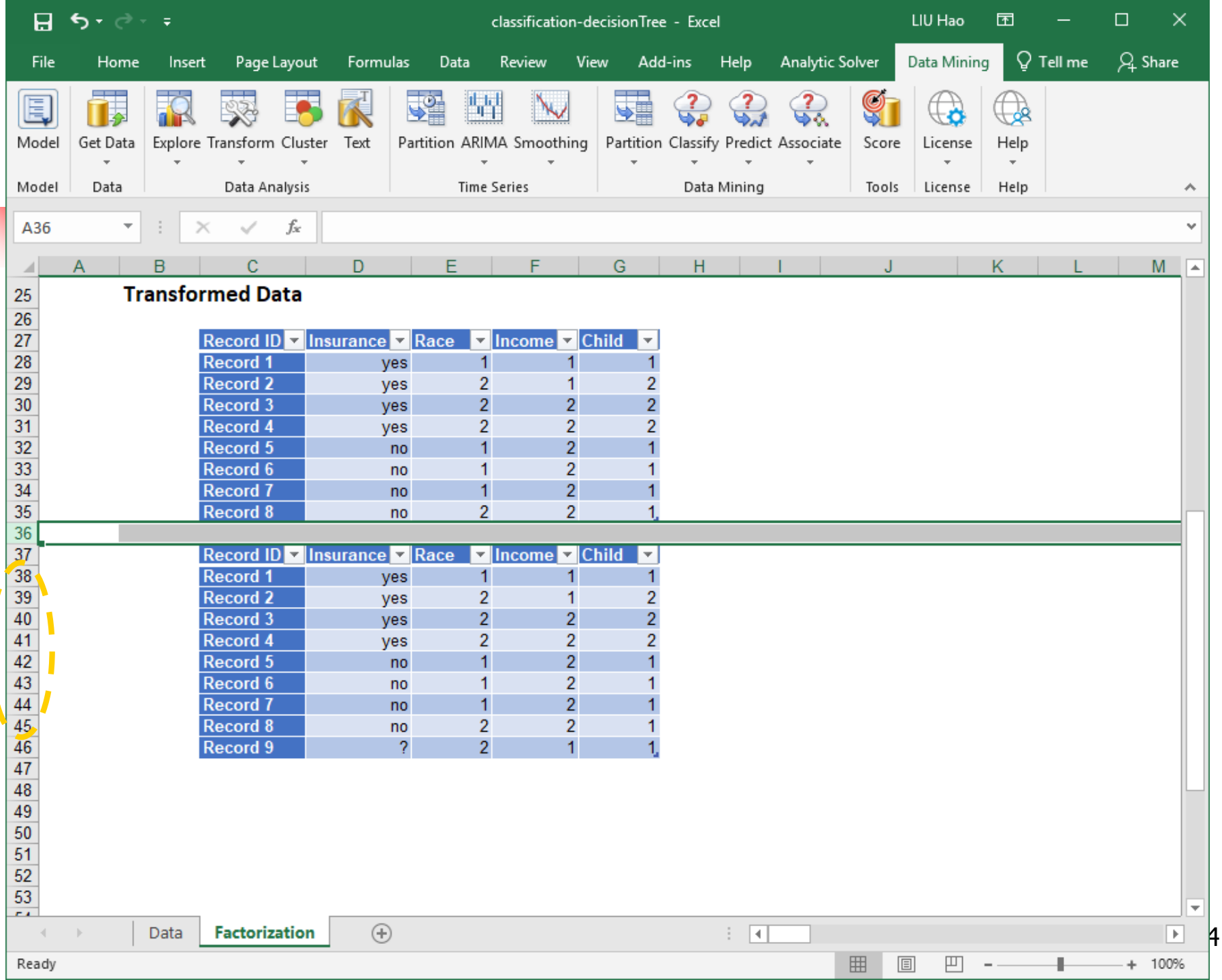
- We have finished the transformation.
- However, we want to make the classification process easier
- Thus, we want to “tidy up” the input format for the later process of classification now.

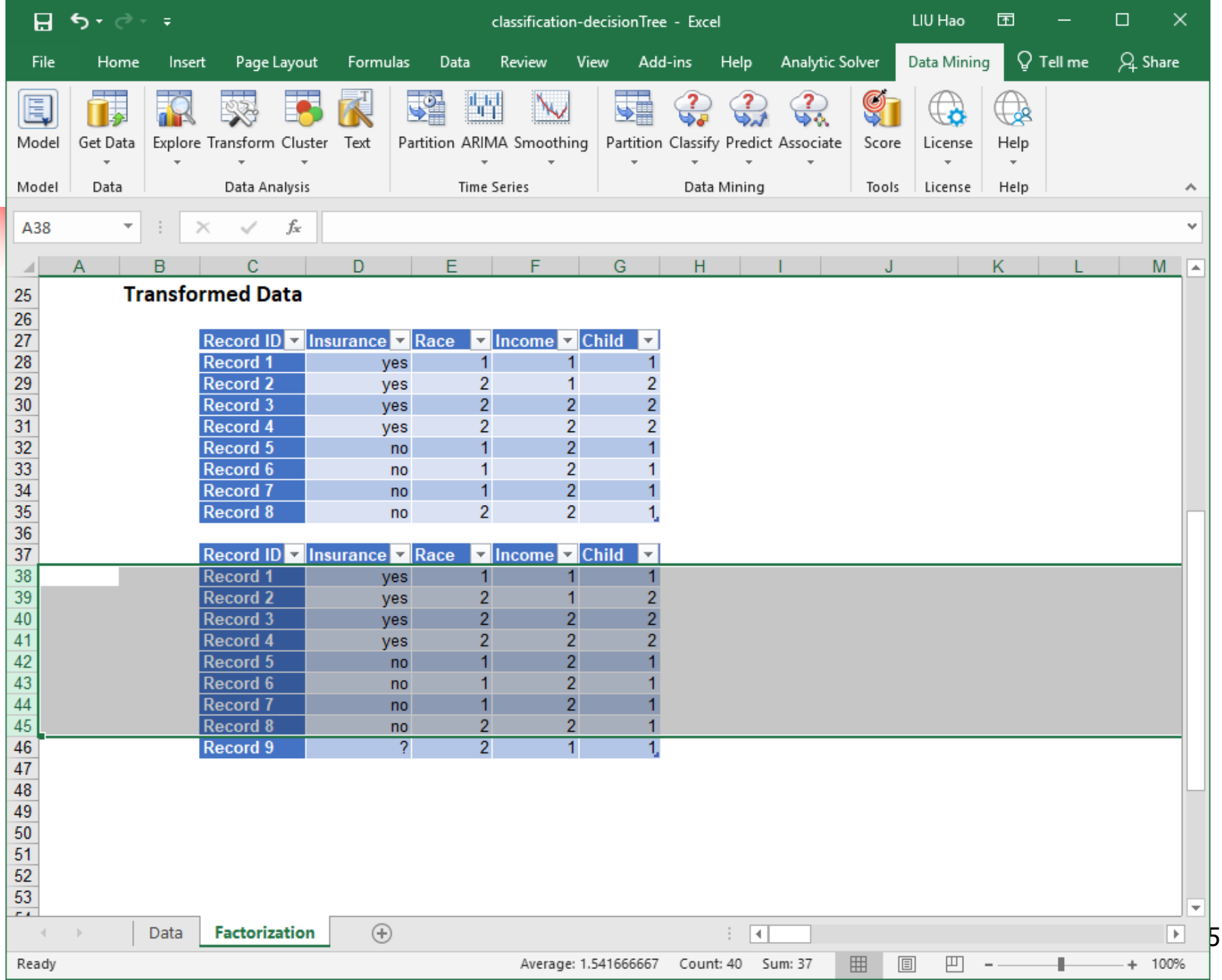

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Design Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
C27
Record ID
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
100%

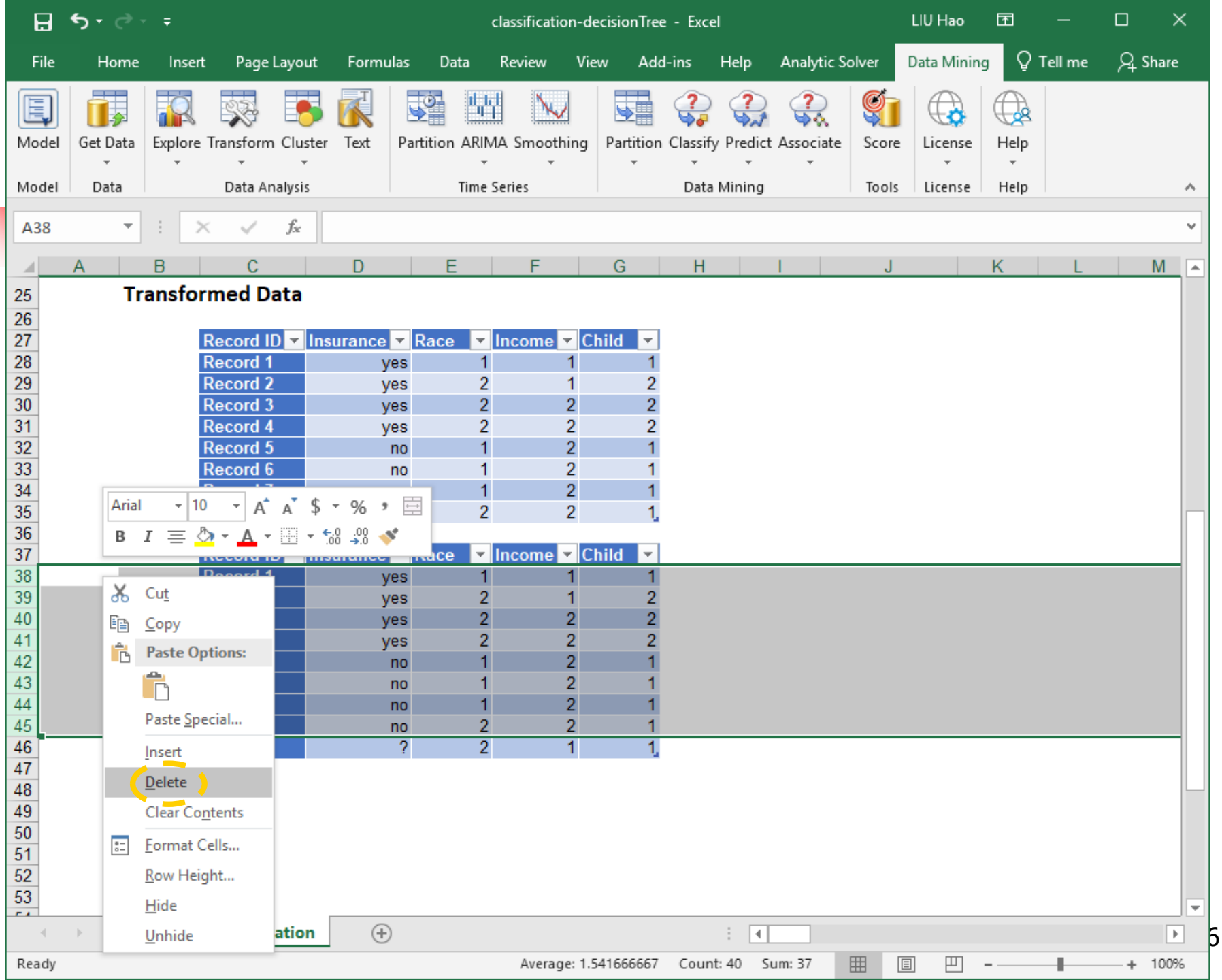

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Design Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
C38
Record ID
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
(Ctrl)
Data Factorization
Select destination and press ENTER or choose Paste
Average: 1.518518519 Count: 50 Sum: 41
100%

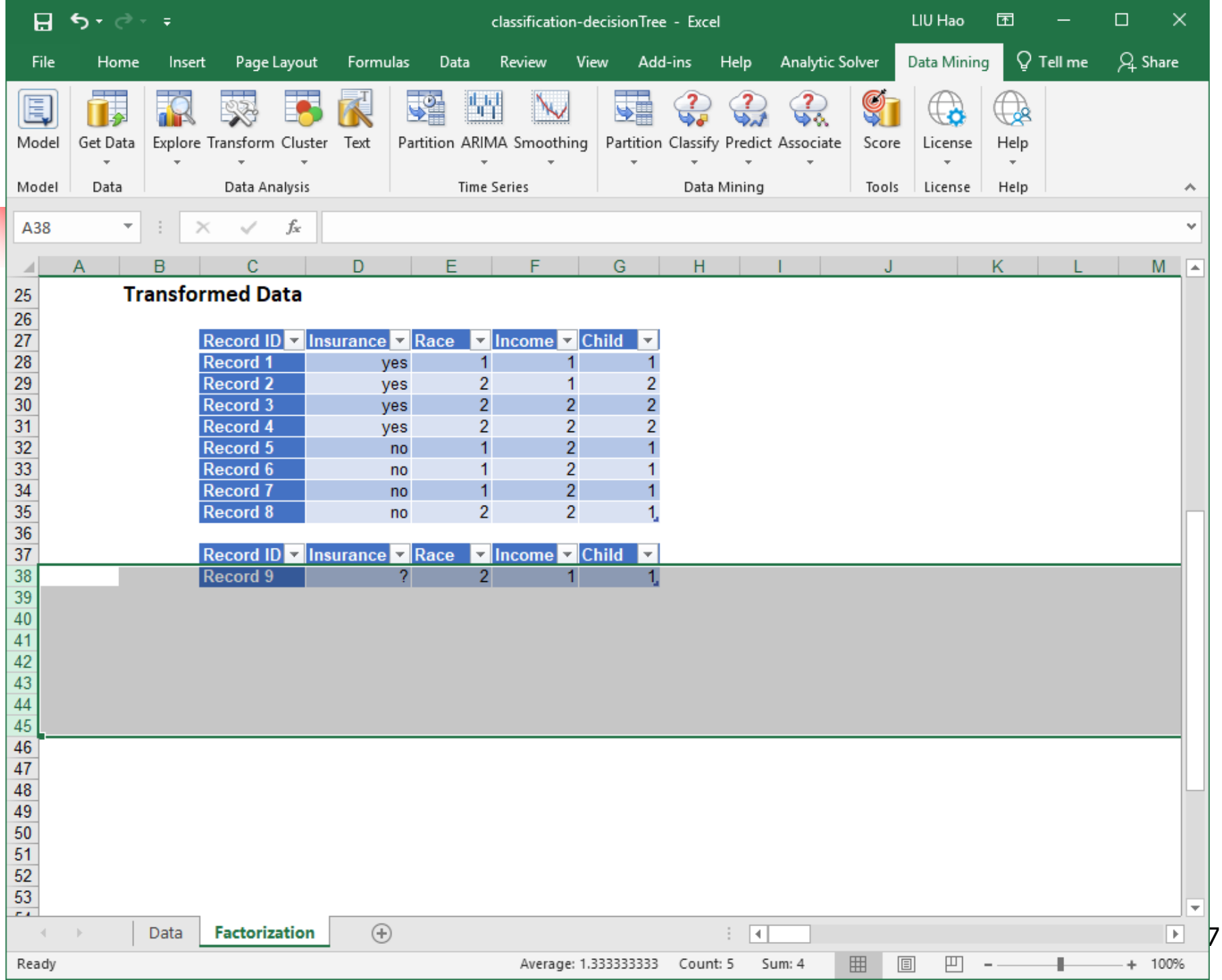

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A36
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
Ready
Average: 1.333333333 Count: 5 Sum: 4
100%

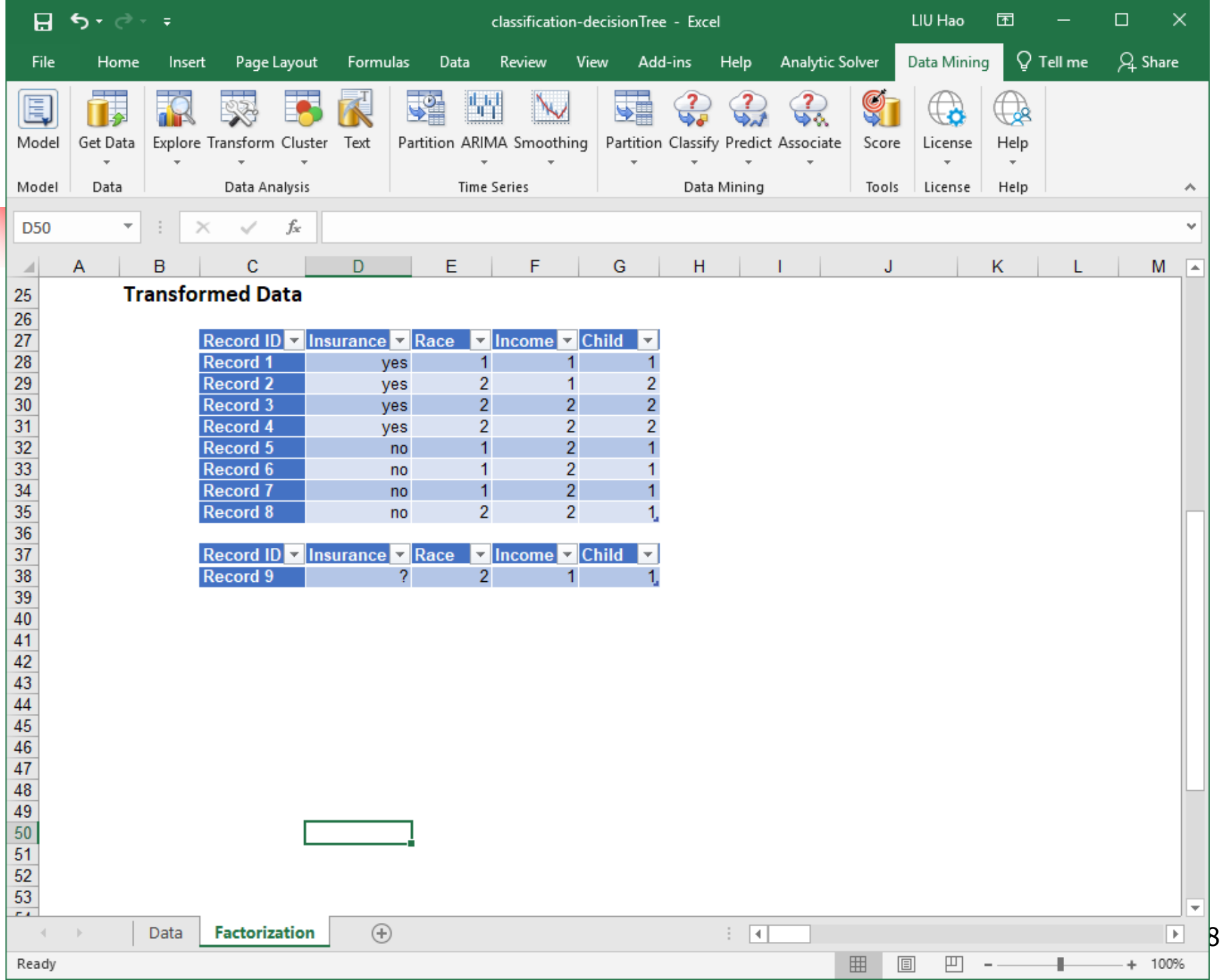

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A36
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
? 2 1 1
Arial 10
Cut
Copy
Paste Options:
Paste Special...
Insert
Delete
Clear Contents
Format Cells...
Row Height...
Hide
Unhide
Insurance Race Income Child
yes 1 1 1
yes 2 1 2
yes 2 2 2
yes 2 2 2
no 1 2 1
no 1 2 1
no 1 2 1
no 2 2 1
? 2 1 1
Data
Factorization
Ready
Average: 1.333333333
Count: 5
Sum: 4
100%

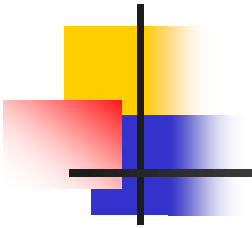

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A36
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
Data Factorization
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 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
A38
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
Data Factorization
Ready
Average: 1.541666667 Count: 40 Sum: 37
100%

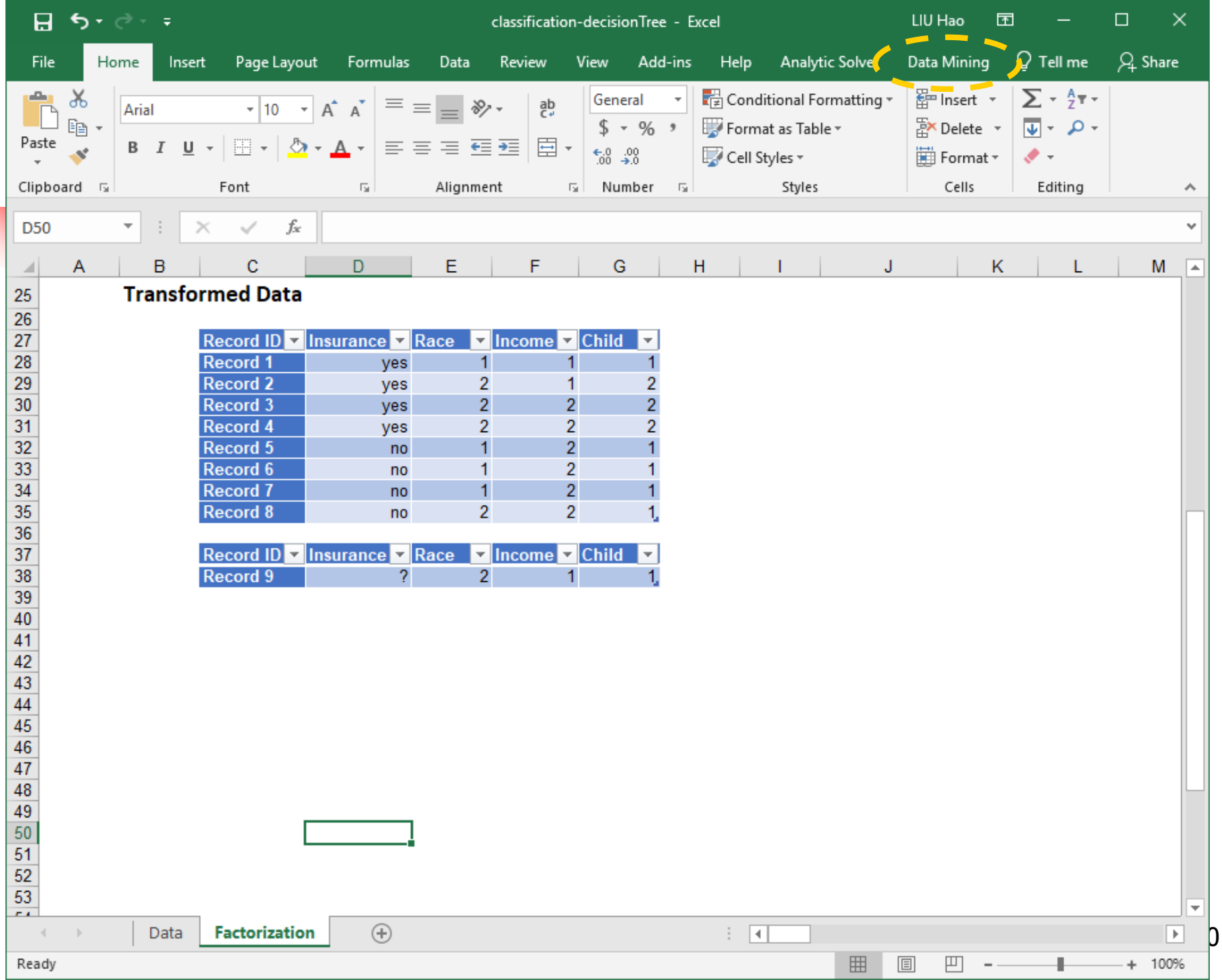

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A38
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
Arial 10
Cut
Copy
Paste Options:
Paste Special...
Insert
Delete
Clear Contents
Format Cells...
Row Height...
Hide
Unhide
Ready
Average: 1.541666667 Count: 40 Sum: 37
100%

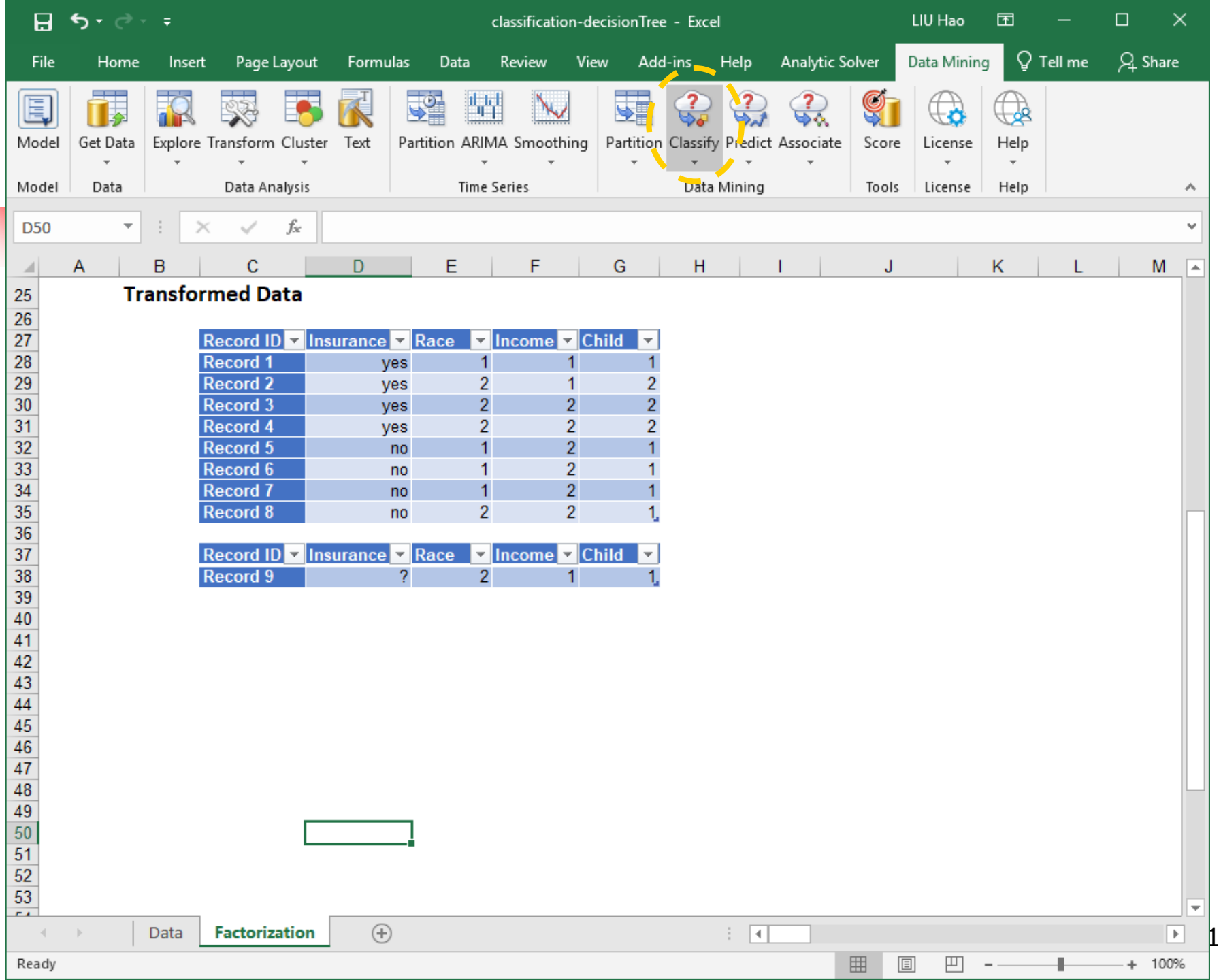

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A38
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
Average: 1.333333333 Count: 5 Sum: 4
100%

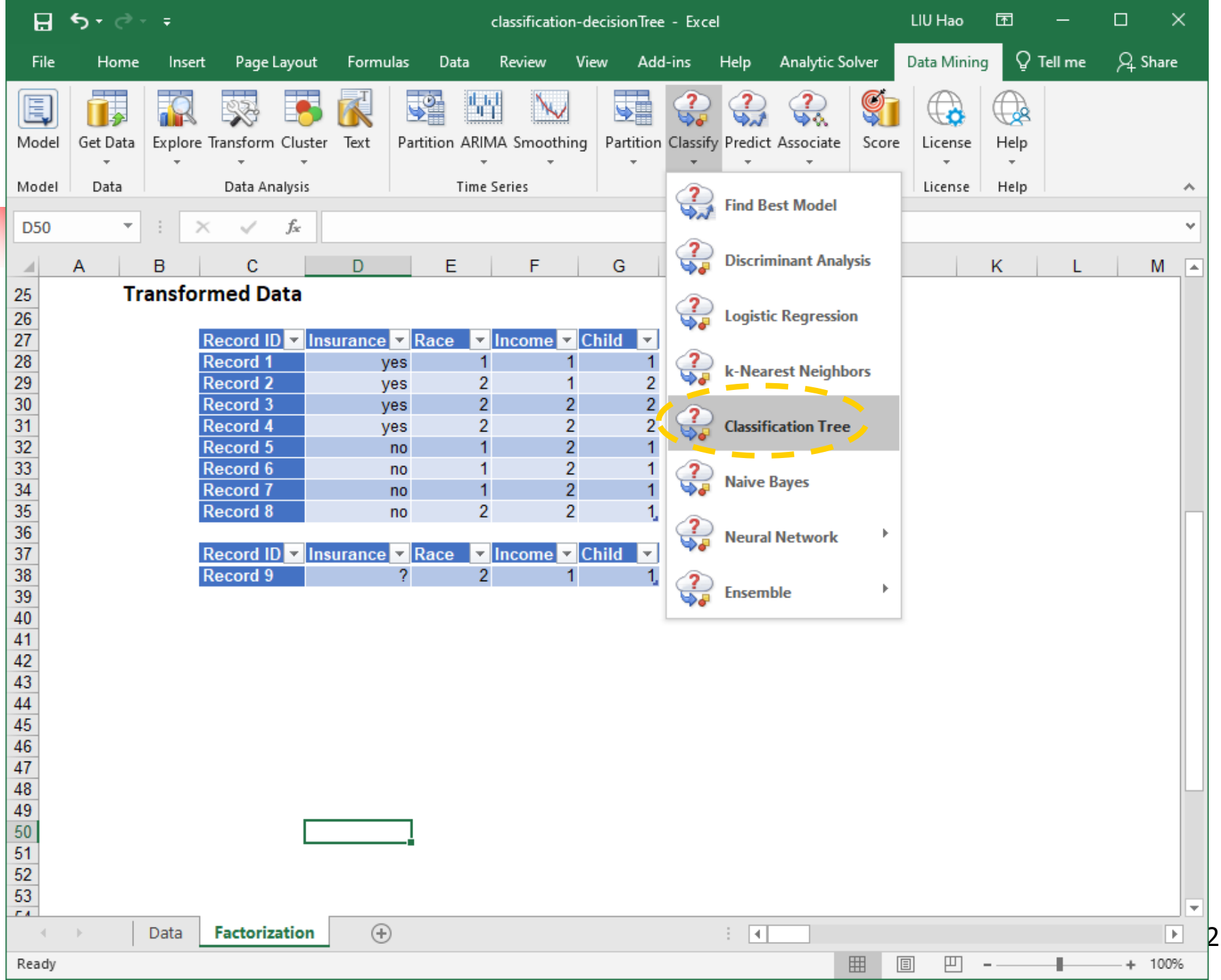

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
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
100%

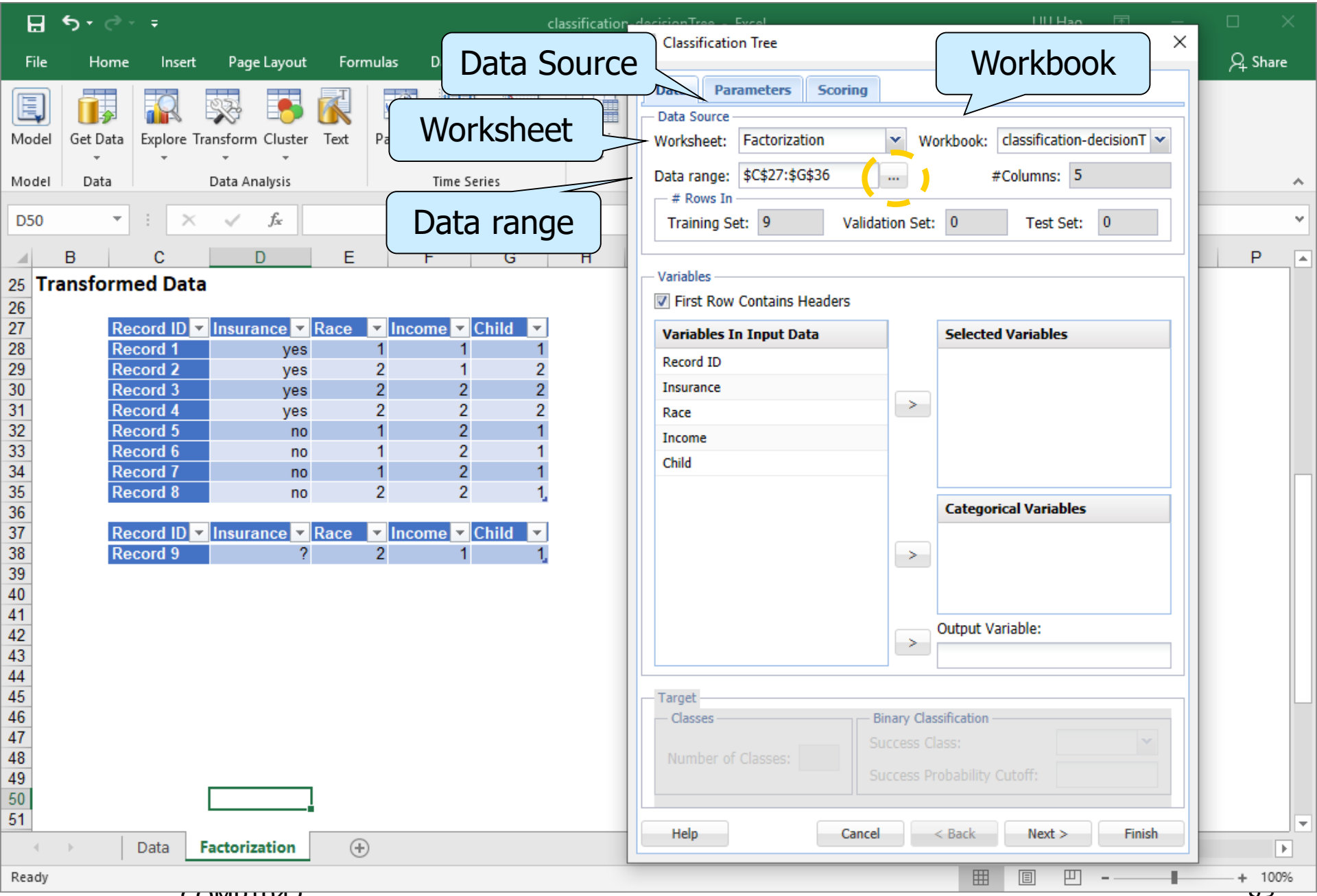


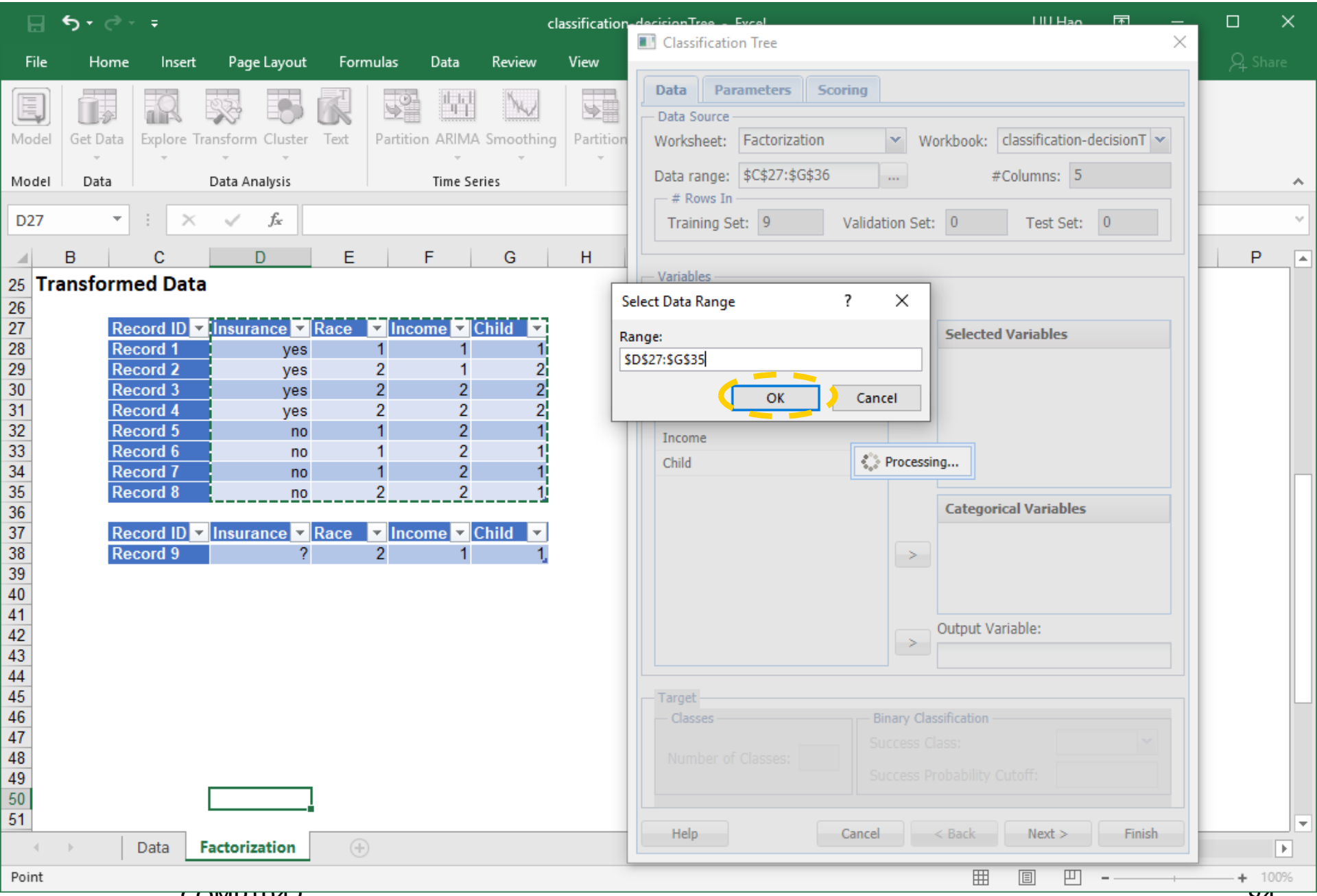
- Now, we understand how to perform the transformation.
- We also “tidied up” the data for the process of classification
- Next, we need to perform the data mining task for classification (Decision Tree, CART)

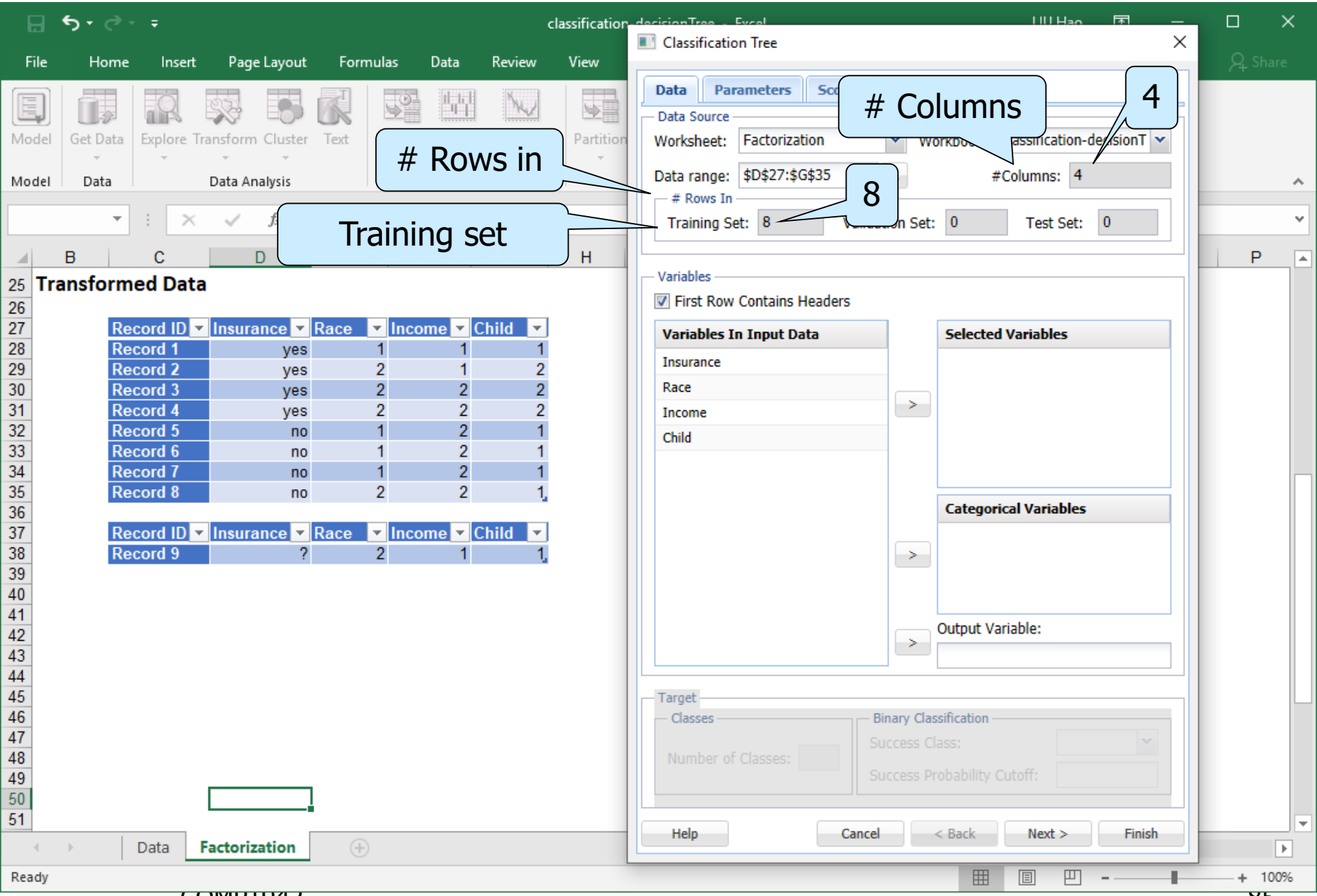

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solve
Data Mining
Tell me
Share
Paste
Clipboard
Arial
10
Font
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
Editing
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
100%

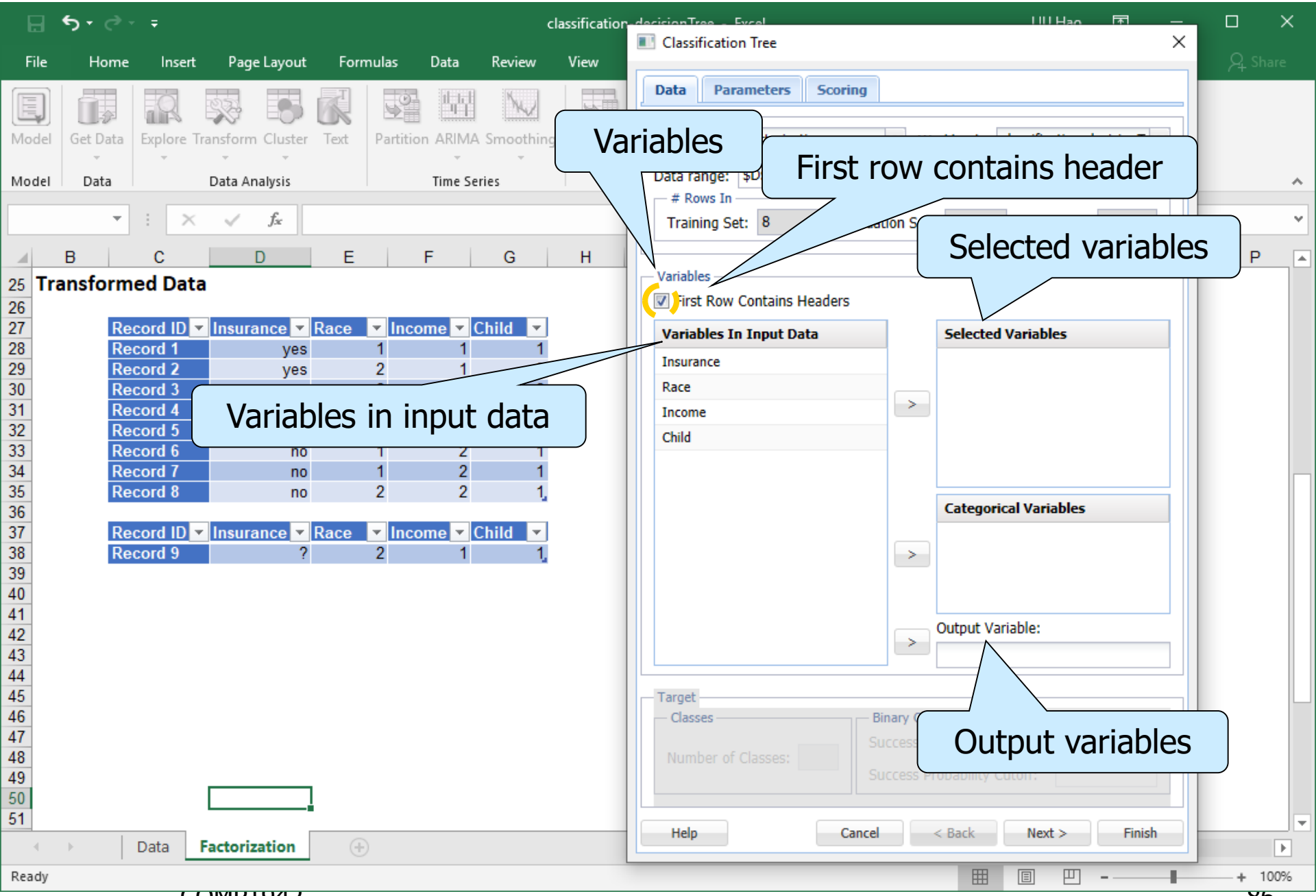

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
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
100%

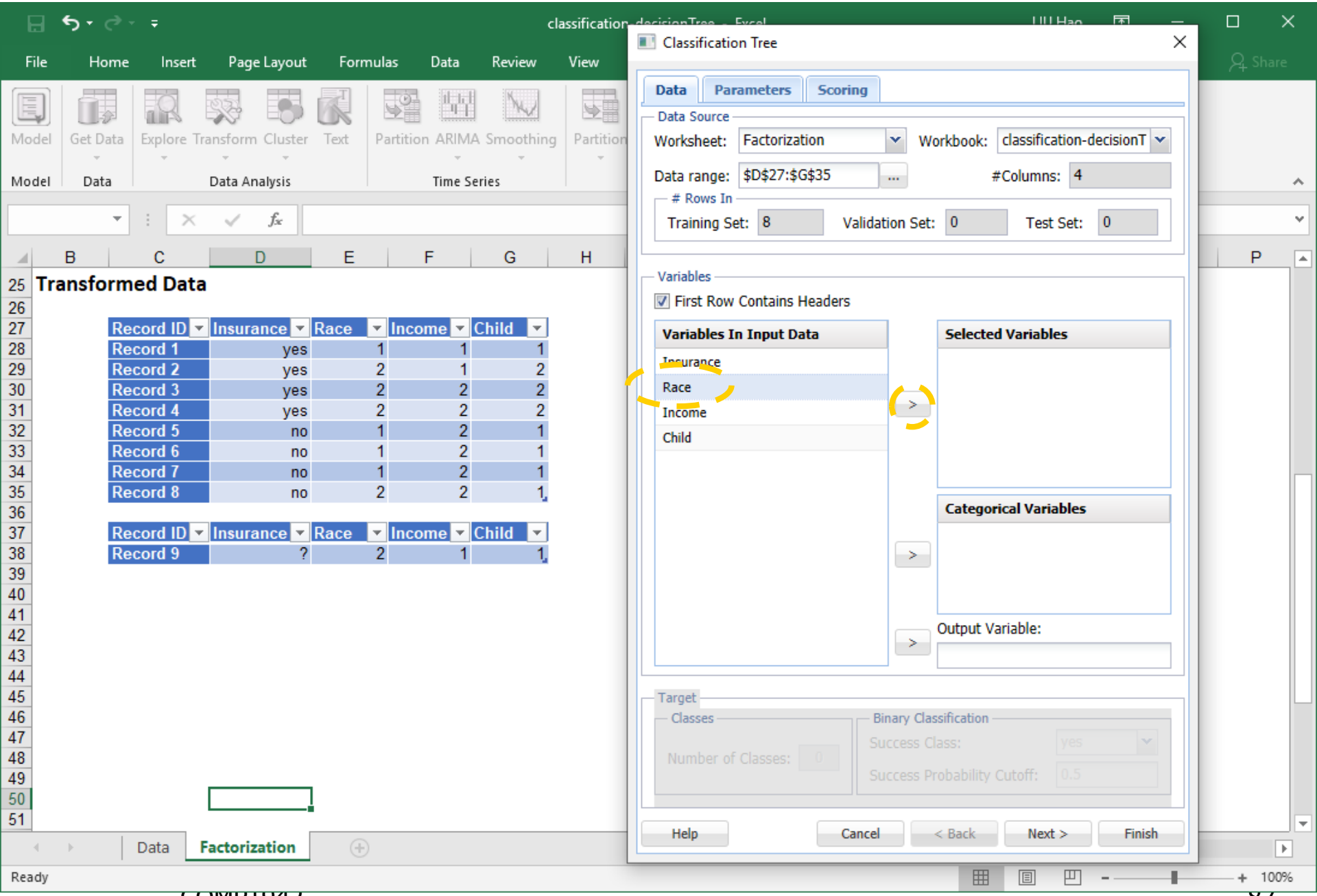

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
License
Help
Find Best Model
Discriminant Analysis
Logistic Regression
k-Nearest Neighbors
Classification Tree
Naive Bayes
Neural Network
Ensemble
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
100%

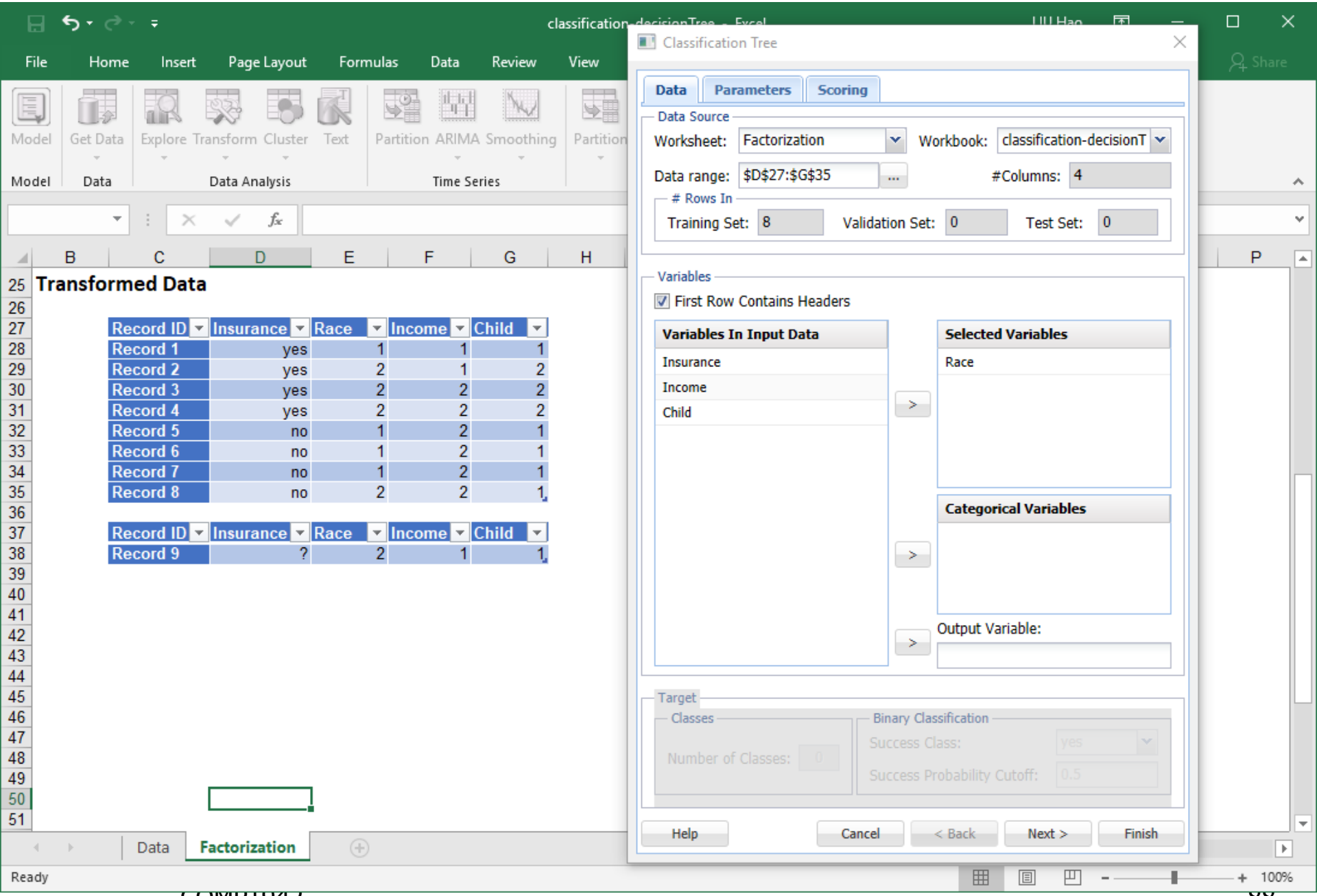

Data Source
Workbook
Worksheet
Data range
Classification Tree
Parameters
Scoring
Data Source
Worksheet: Factorization
Workbook: classification-decisionT
Data range: $C$27:$G$36
#Columns: 5
# Rows In
Training Set: 9
Validation Set: 0
Test Set: 0
Variables
First Row Contains Headers
Variables In Input Data
Record ID
Insurance
Race
Income
Child
Selected Variables
Categorical Variables
Output Variable:
Target
Classes
Number of Classes:
Binary Classification
Success Class:
Success Probability Cutoff:
Help
Cancel
< Back
Next >
Finish
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

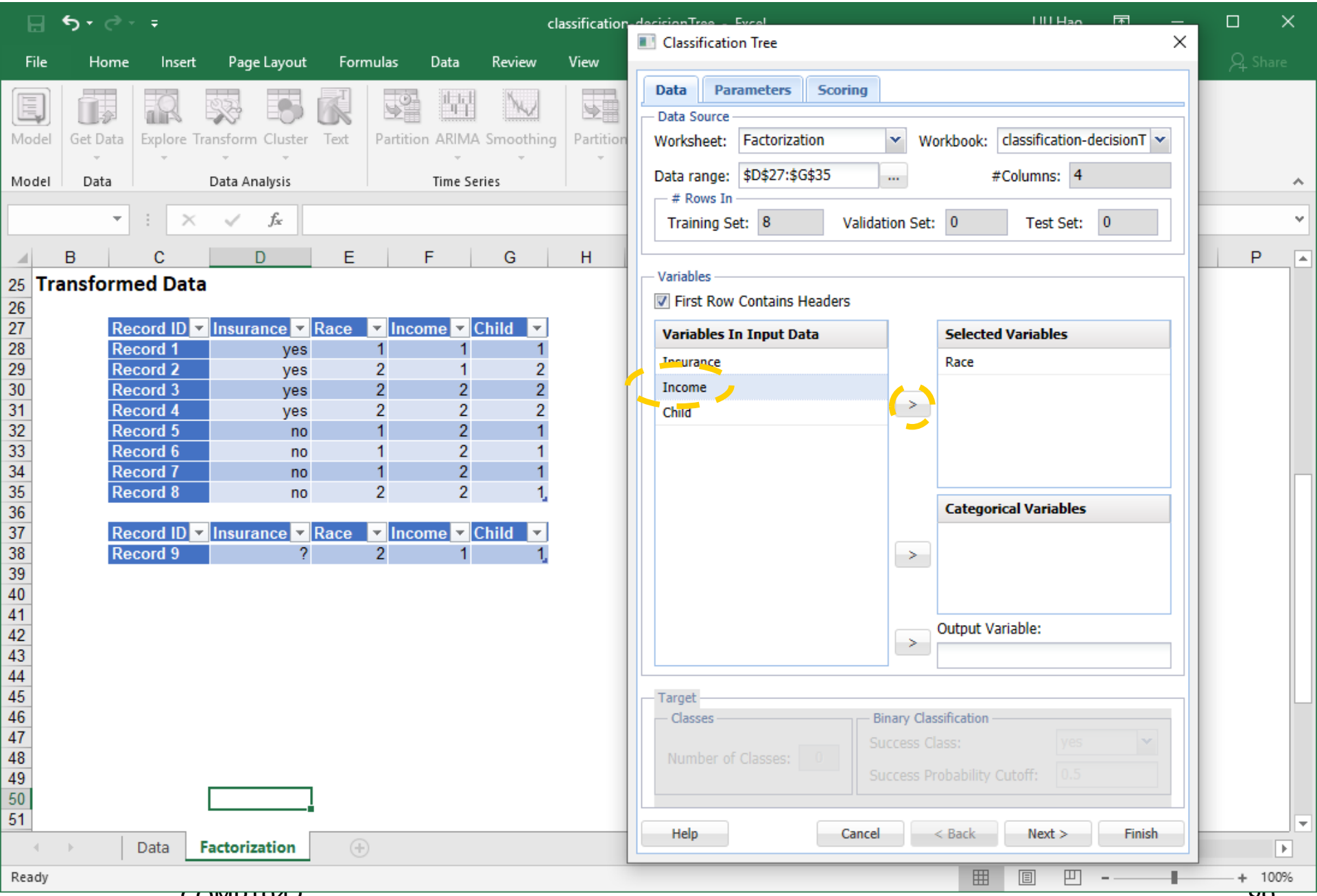

classification-decisionTree - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
D27
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 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
Point
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
Number of Classes:
Binary Classification
Success Class:
Success Probability Cutoff:
Help
Cancel
< Back
Next >
Finish

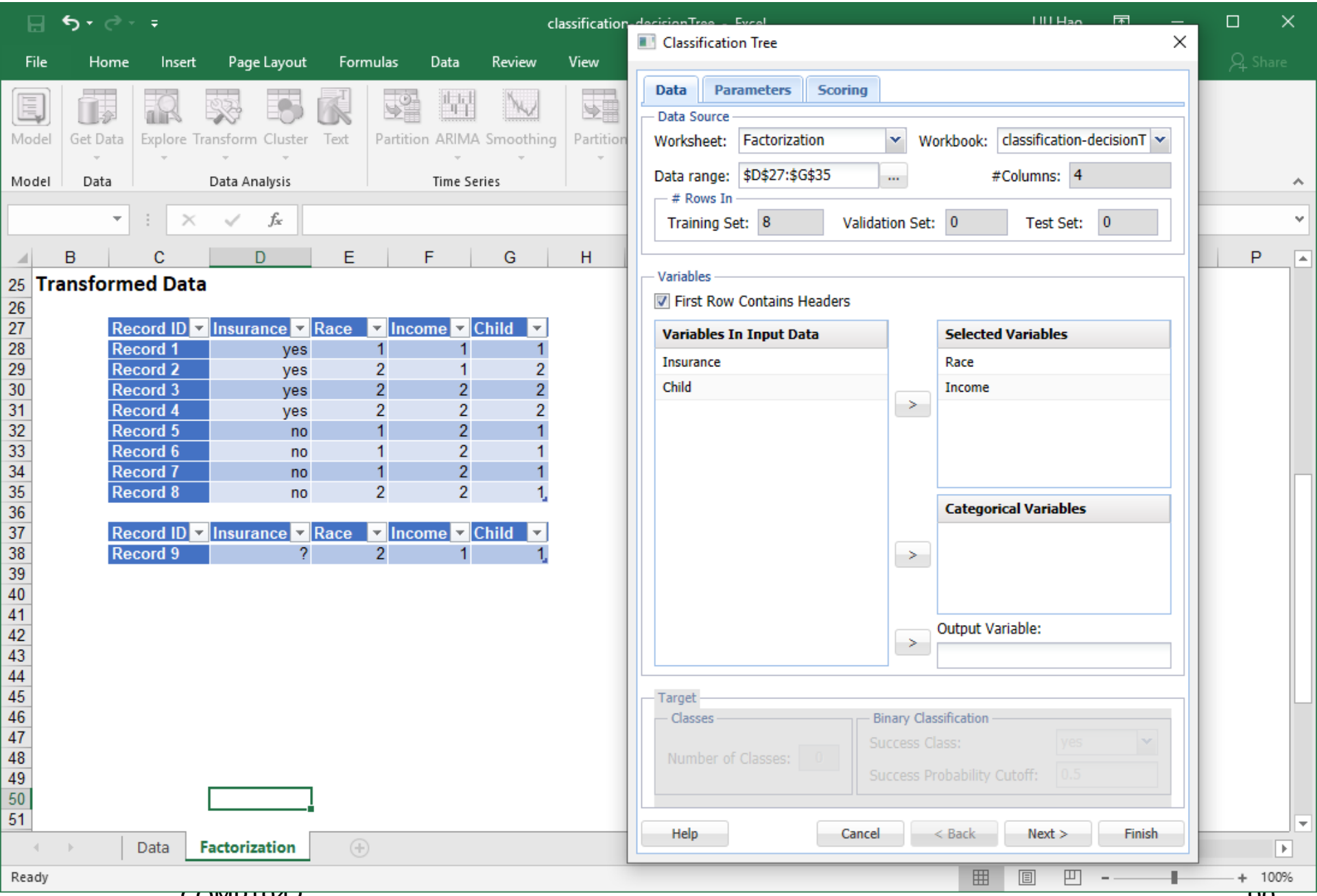

Classification Tree
Data
Parameters
# Columns
4
Data Source
Worksheet: Factorization
Data range: $D$27:$G$35
#Columns: 4
# Rows In
Training Set: 8
Test Set: 0
# Rows in
8
Training set
Variables
First Row Contains Headers
Variables In Input Data
Insurance
Race
Income
Child
Selected Variables
Categorical Variables
Output Variable:
Target
Classes
Number of Classes:
Binary Classification
Success Class:
Success Probability Cutoff:
Help
Cancel
< Back
Next >
Finish
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
100%

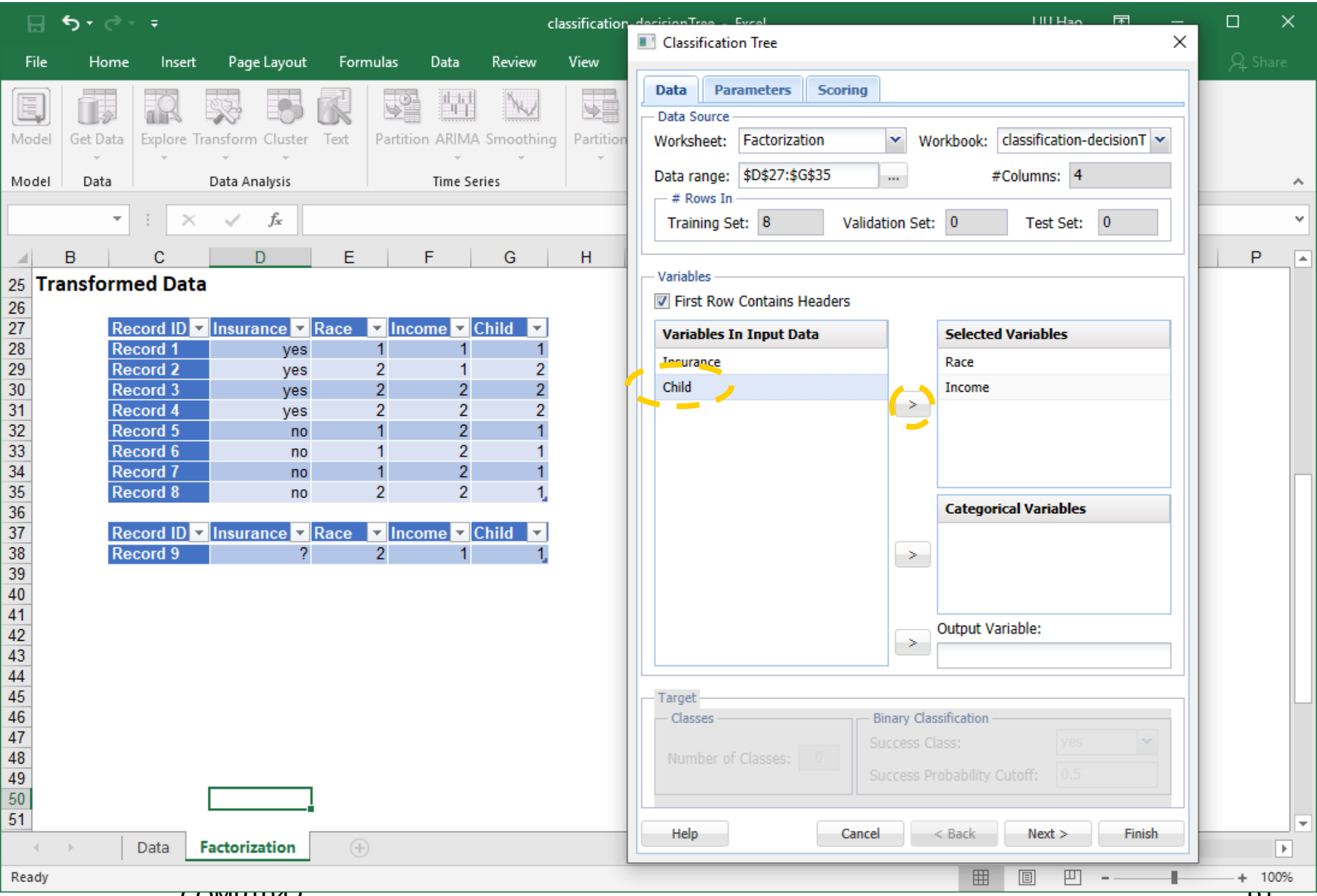

Variables
First row contains header
Selected variables
Variables in input data
Output variables
Transformed Data
Record ID Insurance Race Income Child
Record 1 yes 1 1 1
Record 2 yes 2 1
Record 3
Record 4
Record 5
Record 6 no 1 2 1
Record 7 no 1 2 1
Record 8 no 2 2 1
Record ID Insurance Race Income Child
Record 9 ? 2 1 1
Classification Tree
Data
Parameters
Scoring
Data range:
# Rows In
Training Set: 8
Variables
First Row Contains Headers
Variables In Input Data
Insurance
Race
Income
Child
Selected Variables
Categorical Variables
Output Variable:
Target
Classes
Number of Classes:
Help
Cancel
< Back
Next >
Finish
Data
Factorization
Ready

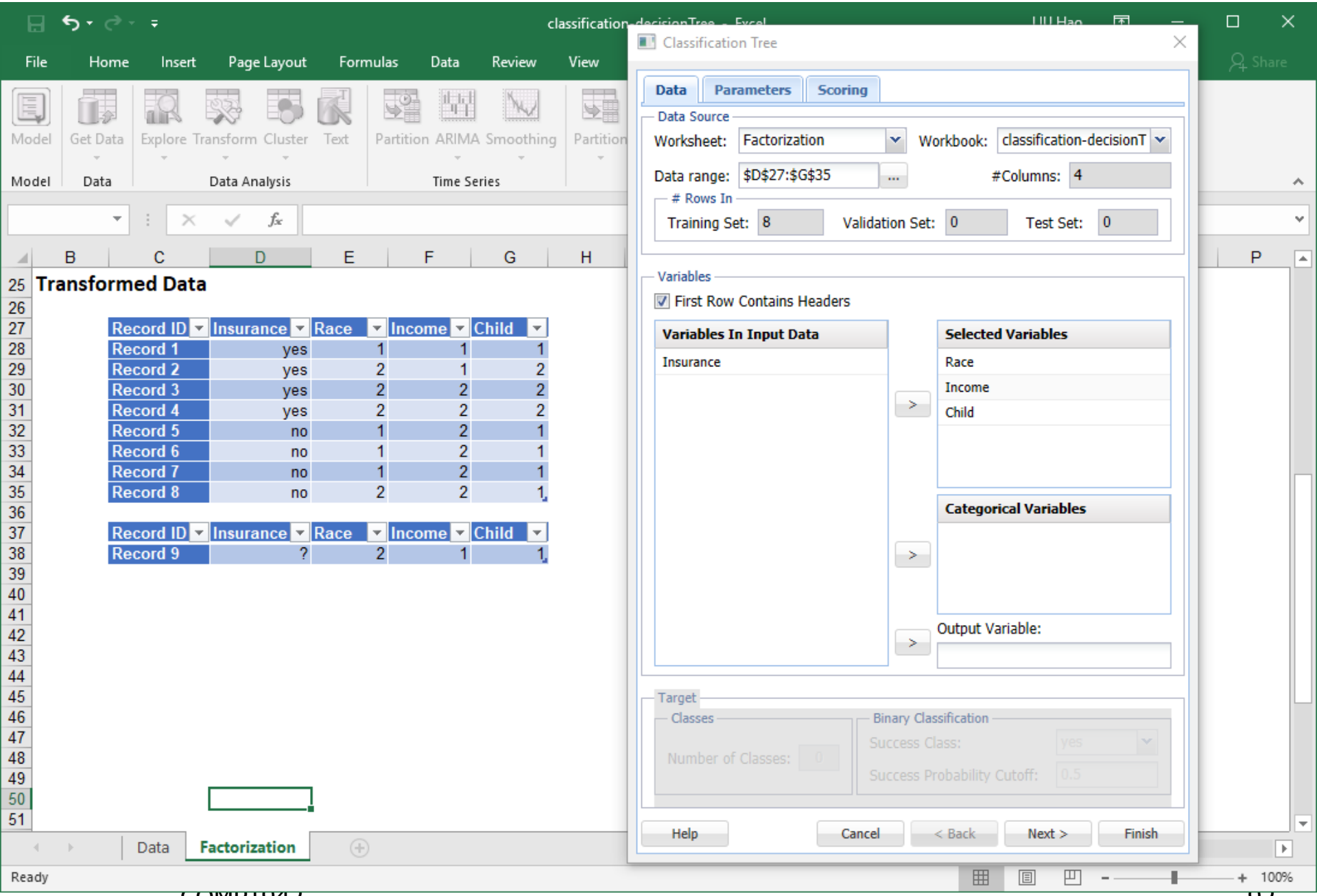

File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 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
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
Number of Classes:
Binary Classification
Success Class: yes
Success Probability Cutoff: 0.5
Help
Cancel
< Back
Next >
Finish

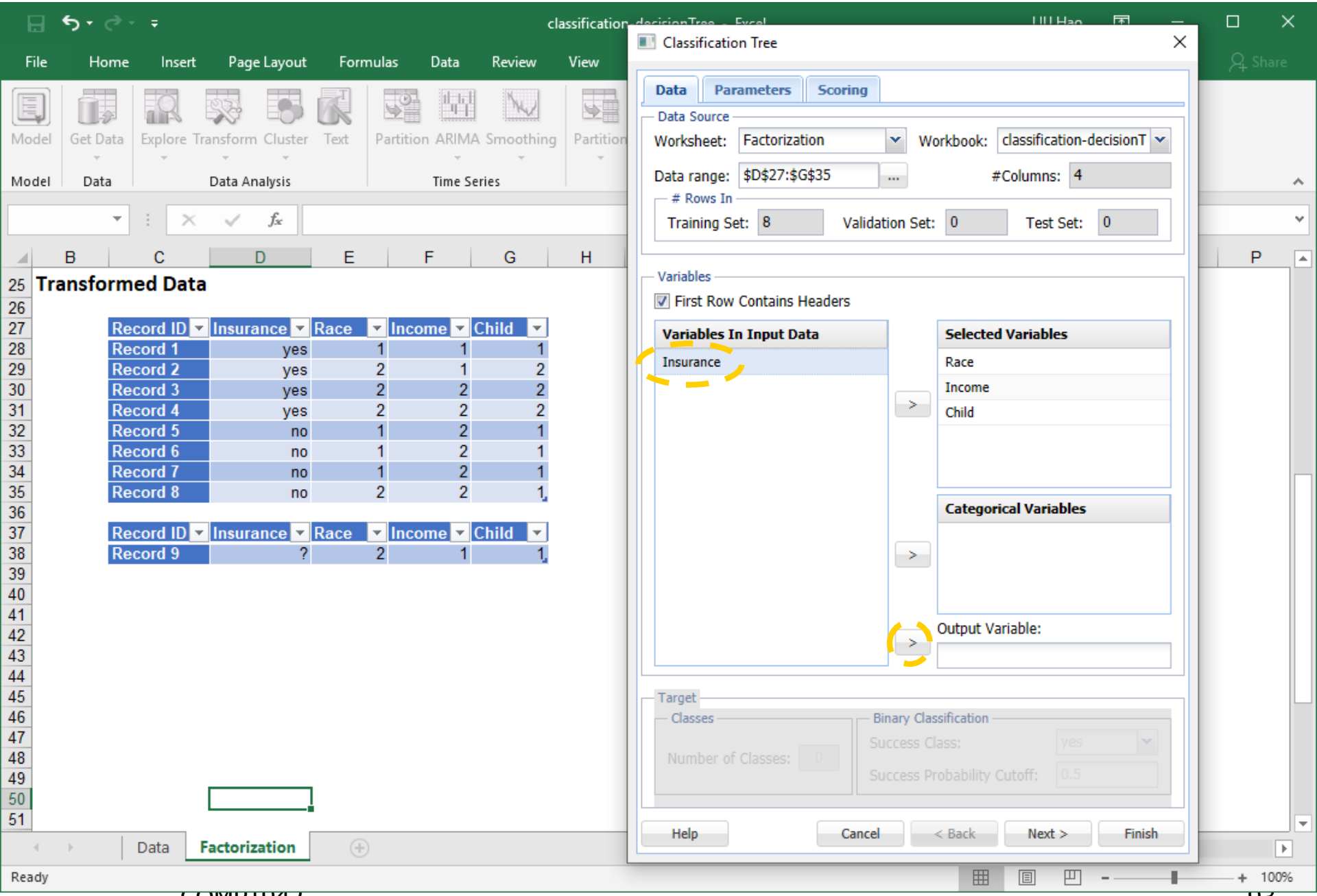

classification-decisionTree - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 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
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
Selected Variables
Race
Categorical Variables
Output Variable:
Target
Classes
Number of Classes: 0
Binary Classification
Success Class: yes
Success Probability Cutoff: 0.5
Help
Cancel
< Back
Next >
Finish
100%

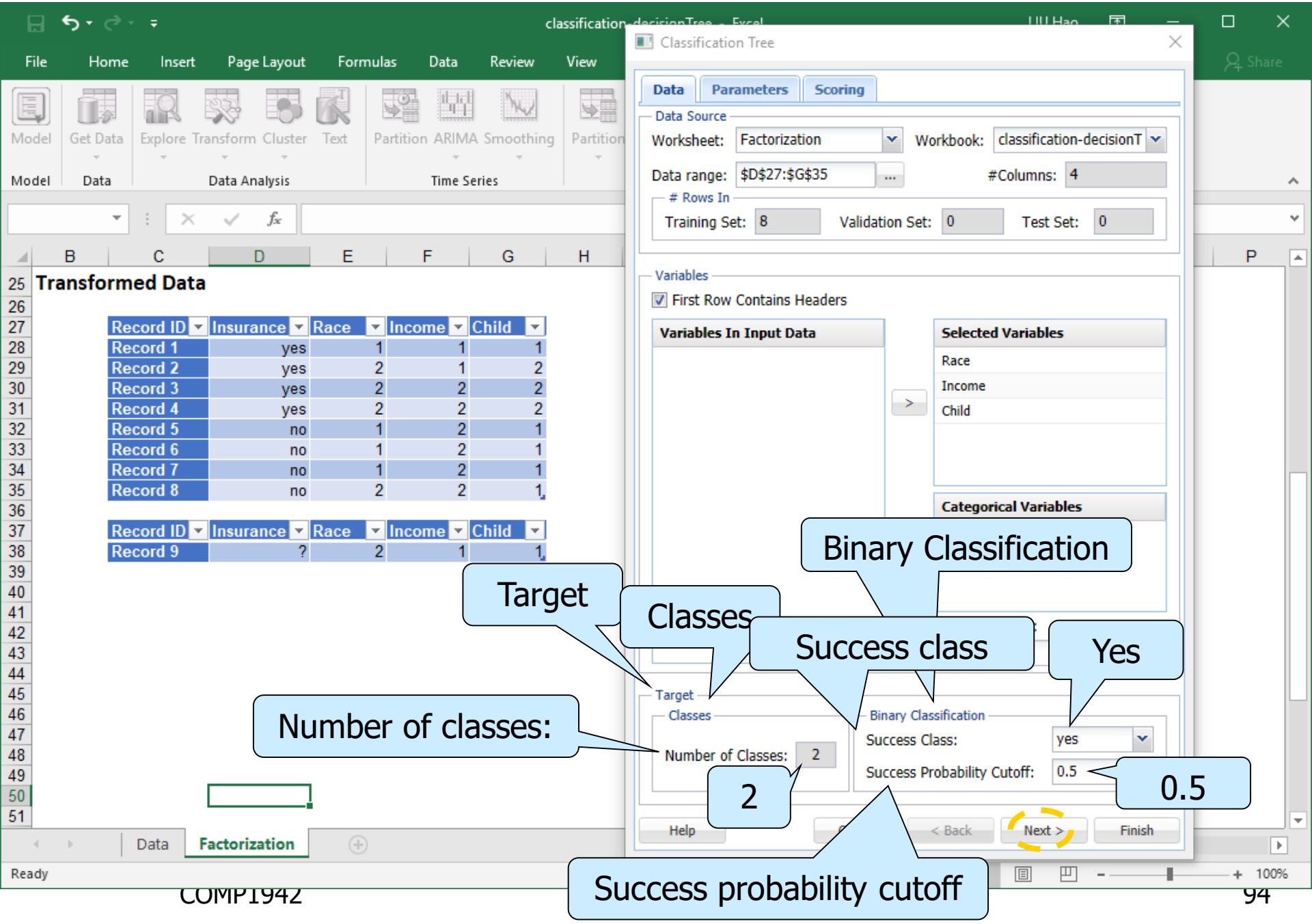

classification-decisionTree - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 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
Data
Factorization
Ready
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
Number of Classes: 0
Binary Classification
Success Class: yes
Success Probability Cutoff: 0.5
Help
Cancel
< Back
Next >
Finish

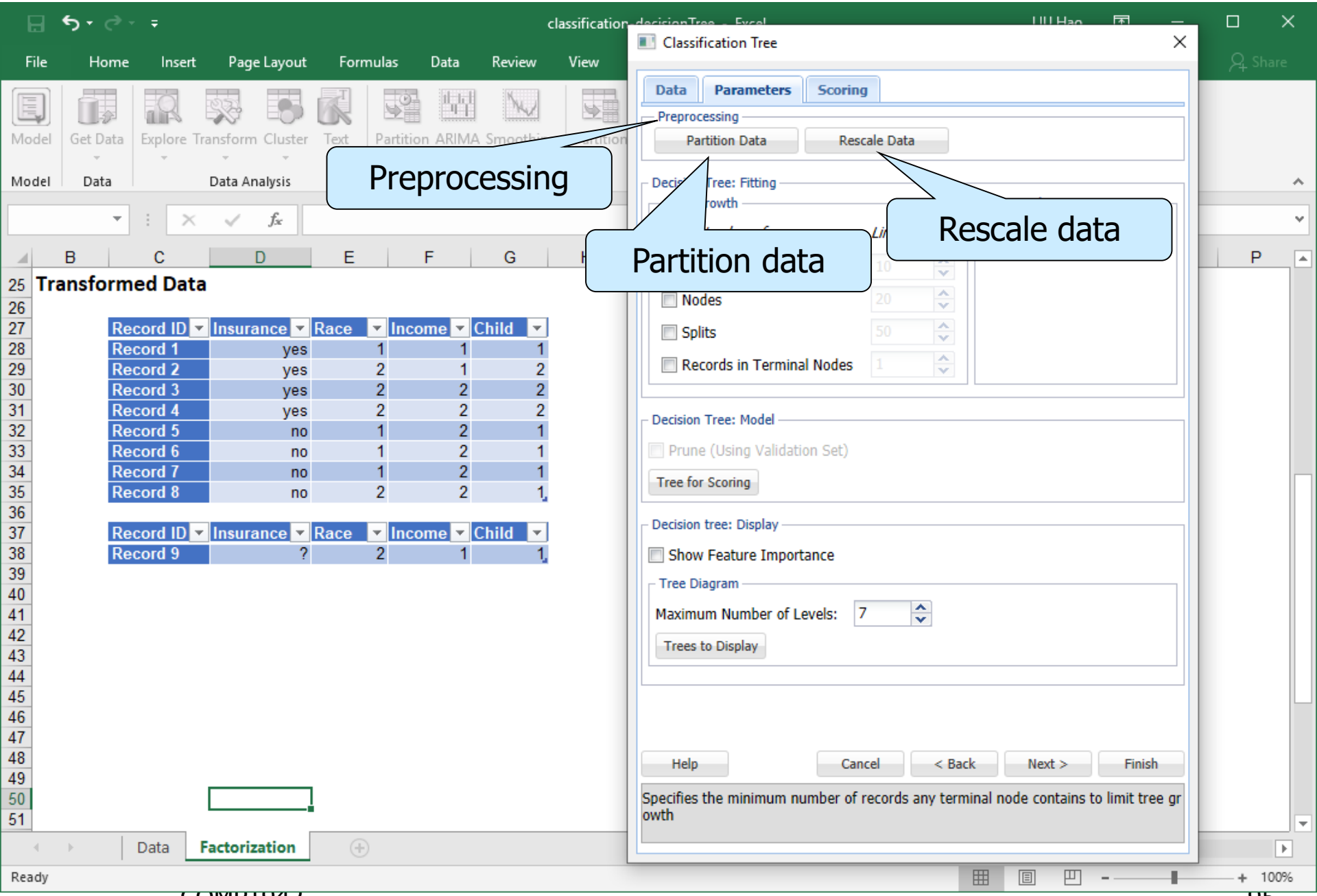

File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 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
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
Number of Classes: 0
Binary Classification
Success Class: yes
Success Probability Cutoff: 0.5
Help
Cancel
< Back
Next >
Finish
100%

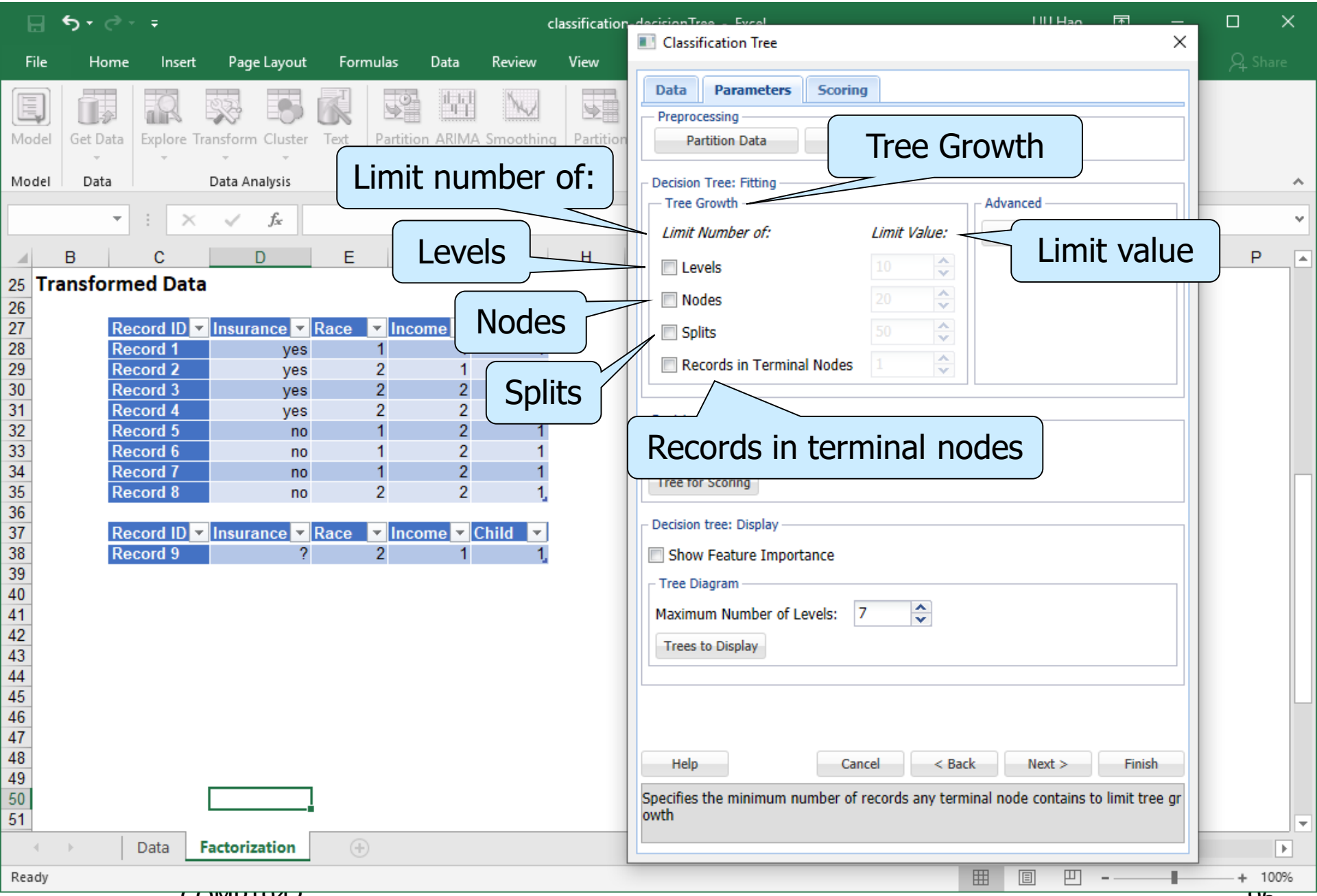

File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 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
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
Number of Classes: 0
Binary Classification
Success Class: yes
Success Probability Cutoff: 0.5
Help
Cancel
< Back
Next >
Finish

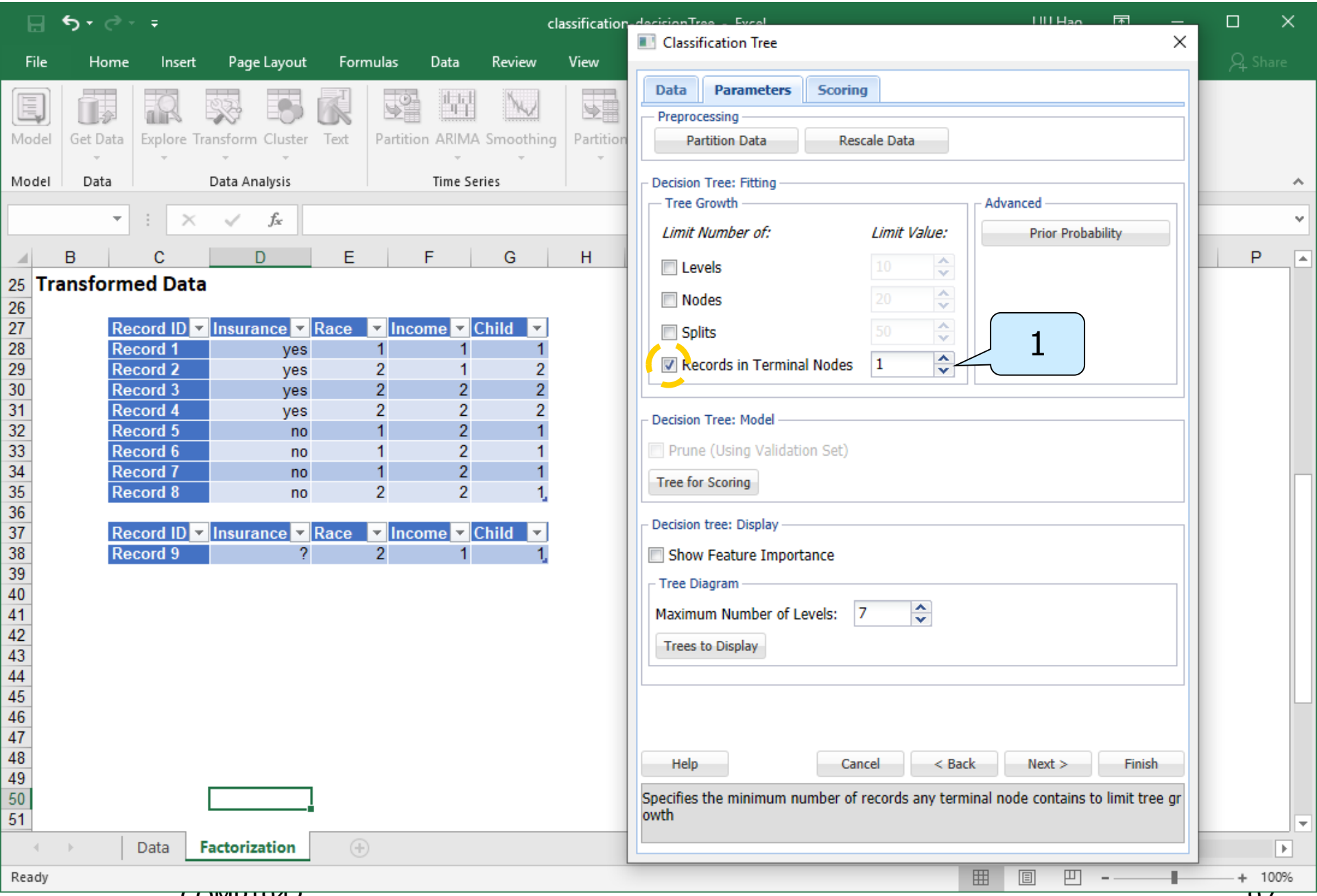

classification-decisionTree - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 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
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
Number of Classes: 0
Binary Classification
Success Class: yes
Success Probability Cutoff: 0.5
Help
Cancel
< Back
Next >
Finish
100%

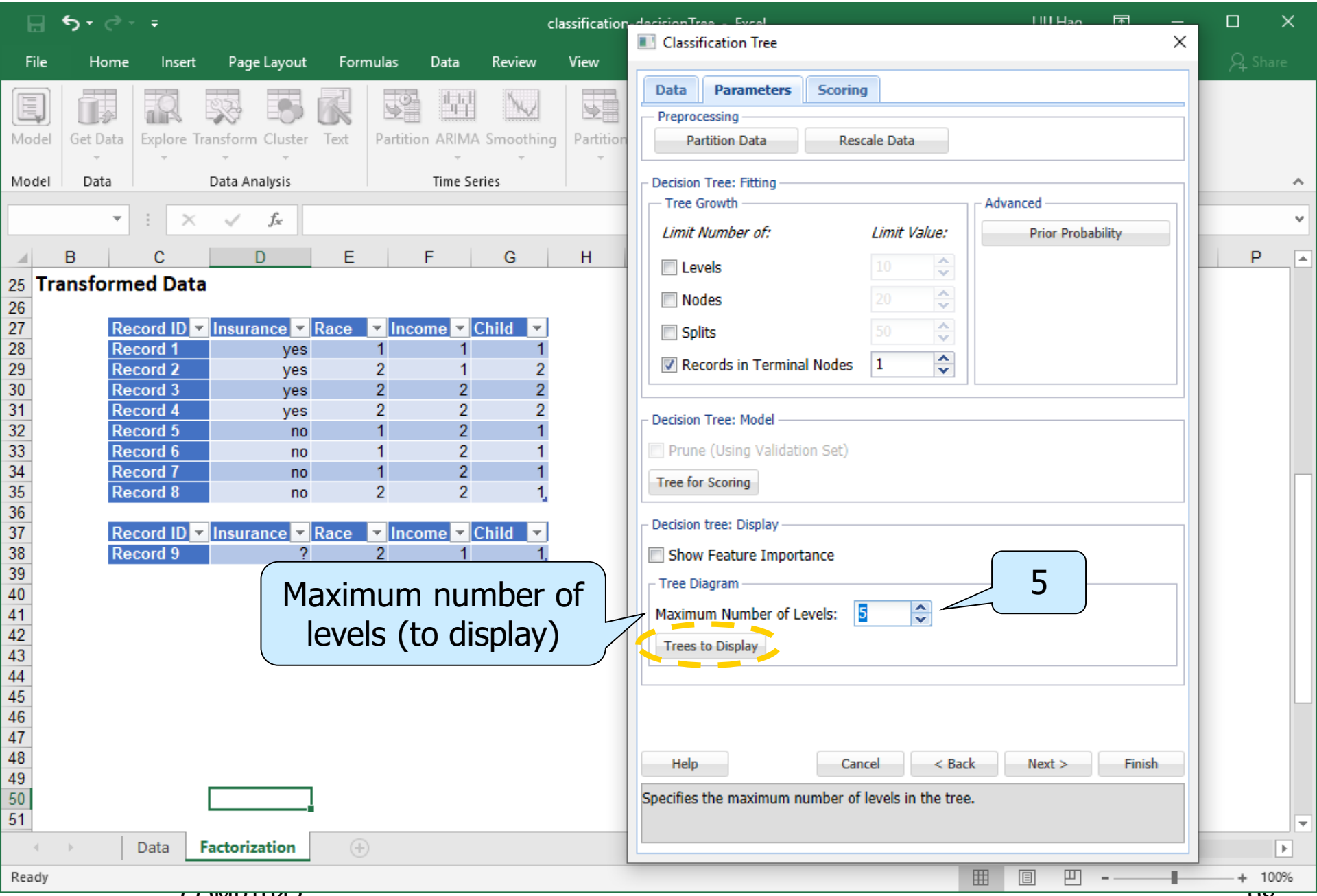

File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 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
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
Number of Classes: 0
Binary Classification
Success Class: yes
Success Probability Cutoff: 0.5
Help
Cancel
< Back
Next >
Finish

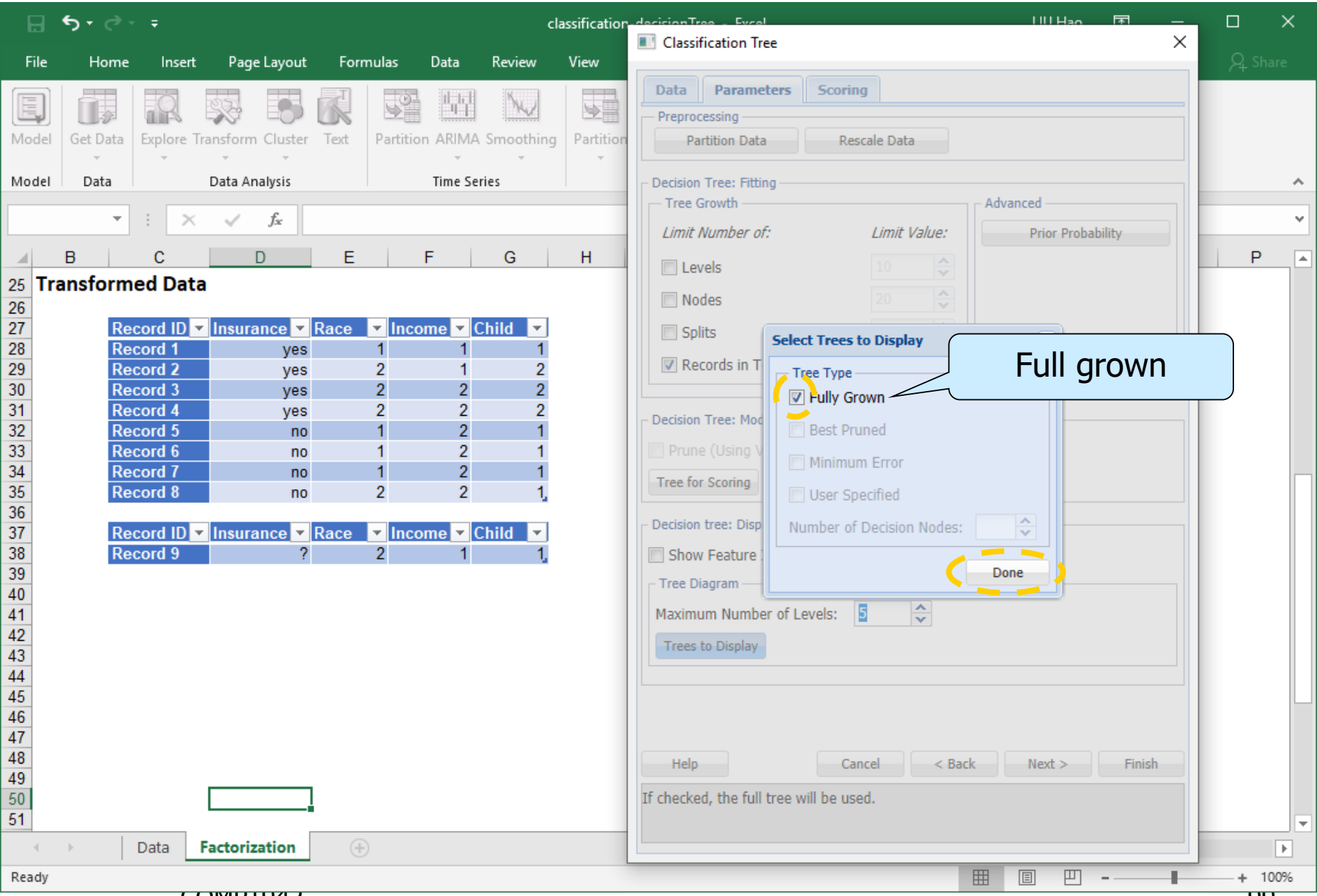

Transformed Data
Record ID Insurance Race Income Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 2 2
Record 5 no 1 2 1
Record 6 no 1 2 1
Record 7 no 1 2 1
Record 8 no 2 2 1
Record 9 ? 2 1 1
Classification Tree
Data
Parameters
Scoring
Data Source
Worksheet: Factorization
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
Selected Variables
Race
Income
Child
Categorical Variables
Target
Classes
Number of Classes: 2
Binary Classification
Success Class: yes
Success Probability Cutoff: 0.5
Help
< Back
Next >
Finish
Binary Classification
Target
Classes
Success class
Yes
Number of classes:
2
0.5
Success probability cutoff
Data
Factorization
Ready

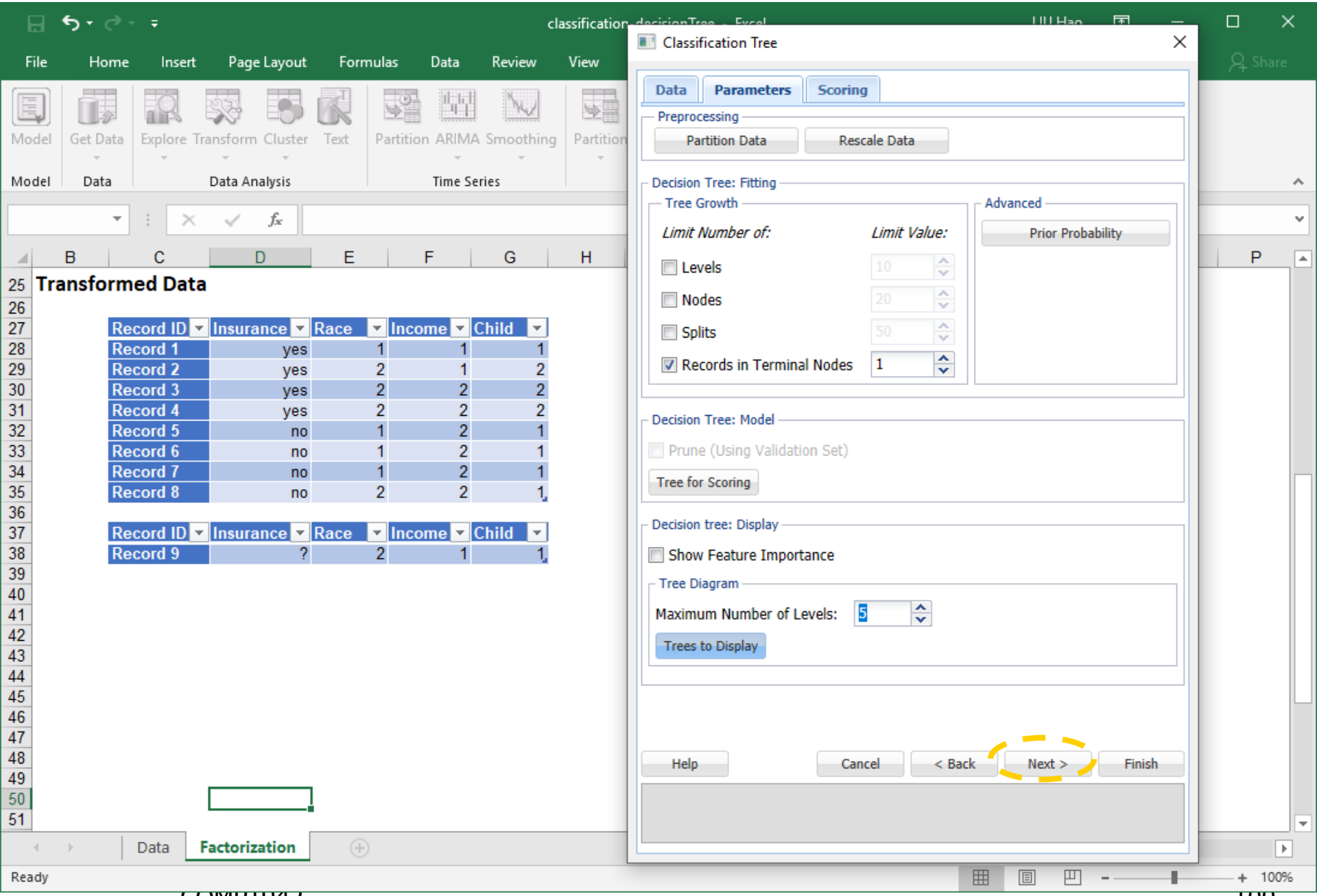

Preprocessing
Partition data
Rescale data
Classification Tree
Data
Parameters
Scoring
Partition Data
Rescale Data
Nodes
Splits
Records in Terminal Nodes
Decision Tree: Model
Prune (Using Validation Set)
Tree for Scoring
Decision tree: Display
Show Feature Importance
Tree Diagram
Maximum Number of Levels:
Trees to Display
Help
Cancel
< Back
Next >
Finish
Specifies the minimum number of records any terminal node contains to limit tree gr owth
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

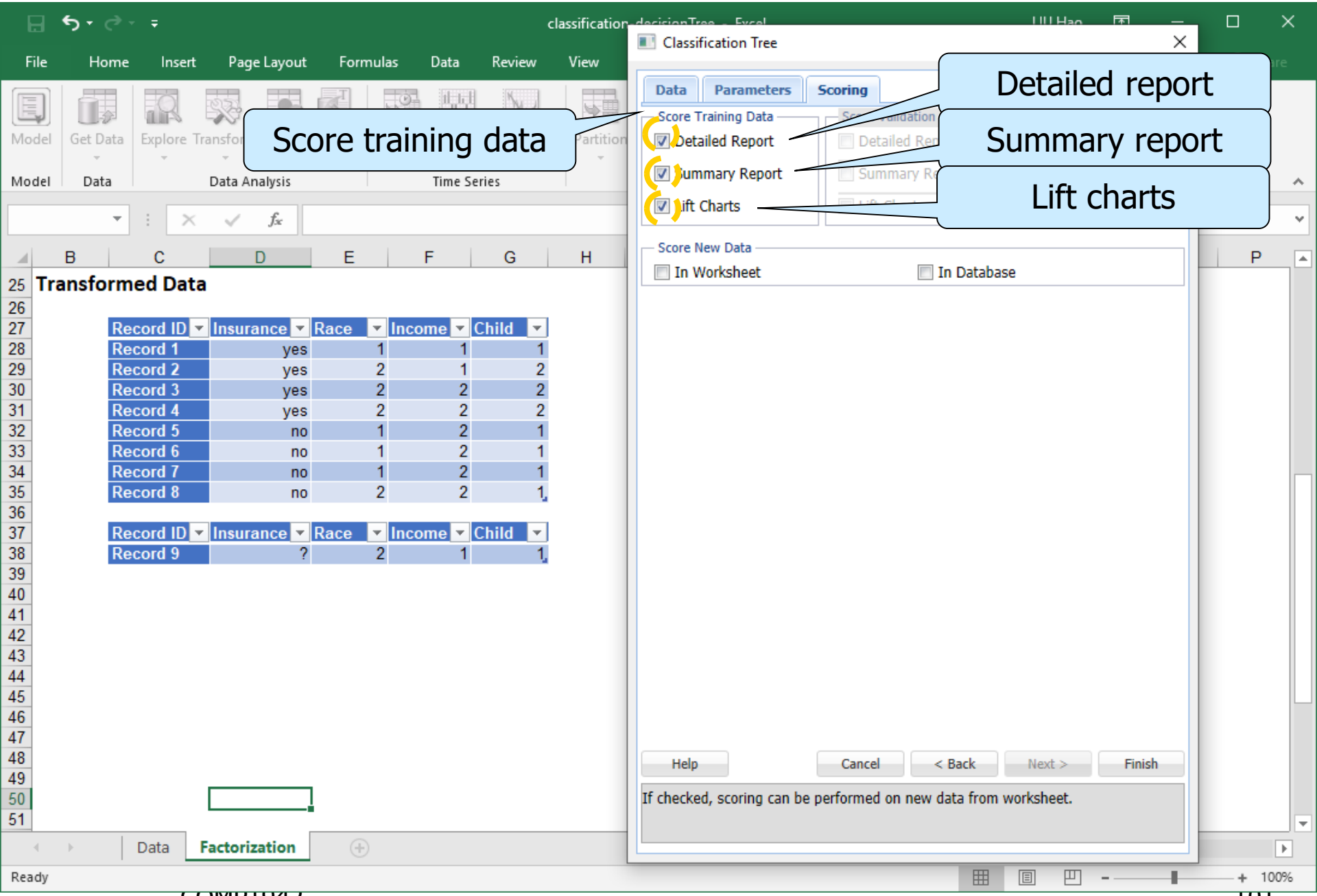

Limit number of:
Levels
Nodes
Splits
Tree Growth
Limit value
Records in terminal nodes
Classification Tree
Data
Parameters
Scoring
Preprocessing
Partition Data
Decision Tree: Fitting
Tree Growth
Limit Number of:
Limit Value:
Advanced
Levels
Nodes
Splits
Records in Terminal Nodes
Tree for Scoring
Decision tree: Display
Show Feature Importance
Tree Diagram
Maximum Number of Levels:
Trees to Display
Help
Cancel
< Back
Next >
Finish
Specifies the minimum number of records any terminal node contains to limit tree growth
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1
Record 2
Record 3
Record 4
Record 5
Record 6
Record 7
Record 8
Record 9
Data
Factorization
Ready

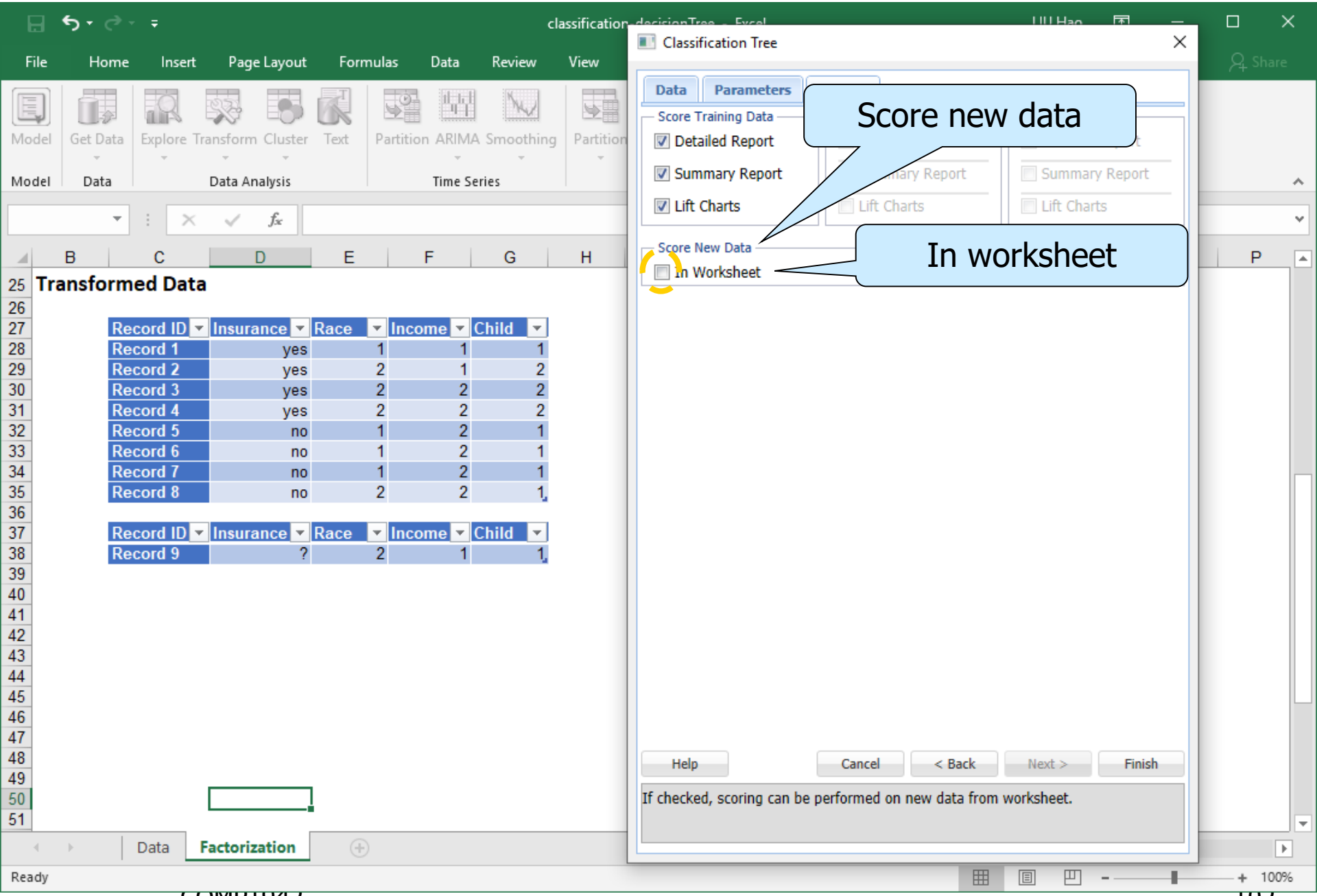

File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 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
Data
Factorization
Ready
Classification Tree
Data
Parameters
Scoring
Preprocessing
Partition Data
Rescale Data
Decision Tree: Fitting
Tree Growth
Limit Number of:
Limit Value:
Levels
10
Nodes
20
Splits
50
Records in Terminal Nodes
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

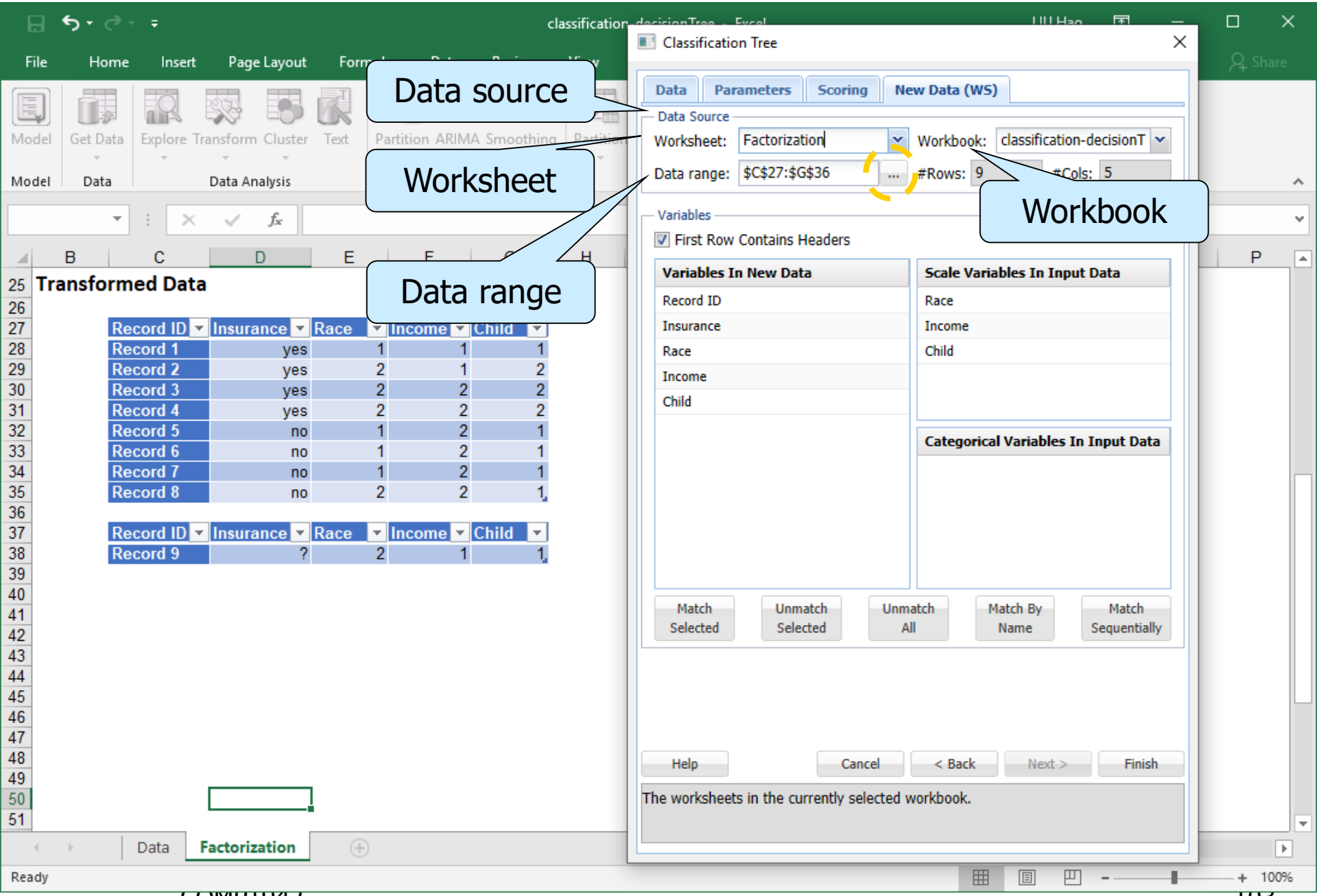

Transformed Data
Record ID Insurance Race Income Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 2 2
Record 5 no 1 2 1
Record 6 no 1 2 1
Record 7 no 1 2 1
Record 8 no 2 2 1
Record ID Insurance Race Income Child
Record 9 ? 2 1 1
Classification Tree
Data
Parameters
Scoring
Preprocessing
Partition Data
Rescale Data
Decision Tree: Fitting
Tree Growth
Limit Number of:
Limit Value:
Levels
Nodes
Splits
Records in Terminal Nodes
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:
Trees to Display
Help
Cancel
< Back
Next >
Finish
Specifies the maximum number of levels in the tree.
Maximum number of levels (to display)
5
Data
Factorization
Ready

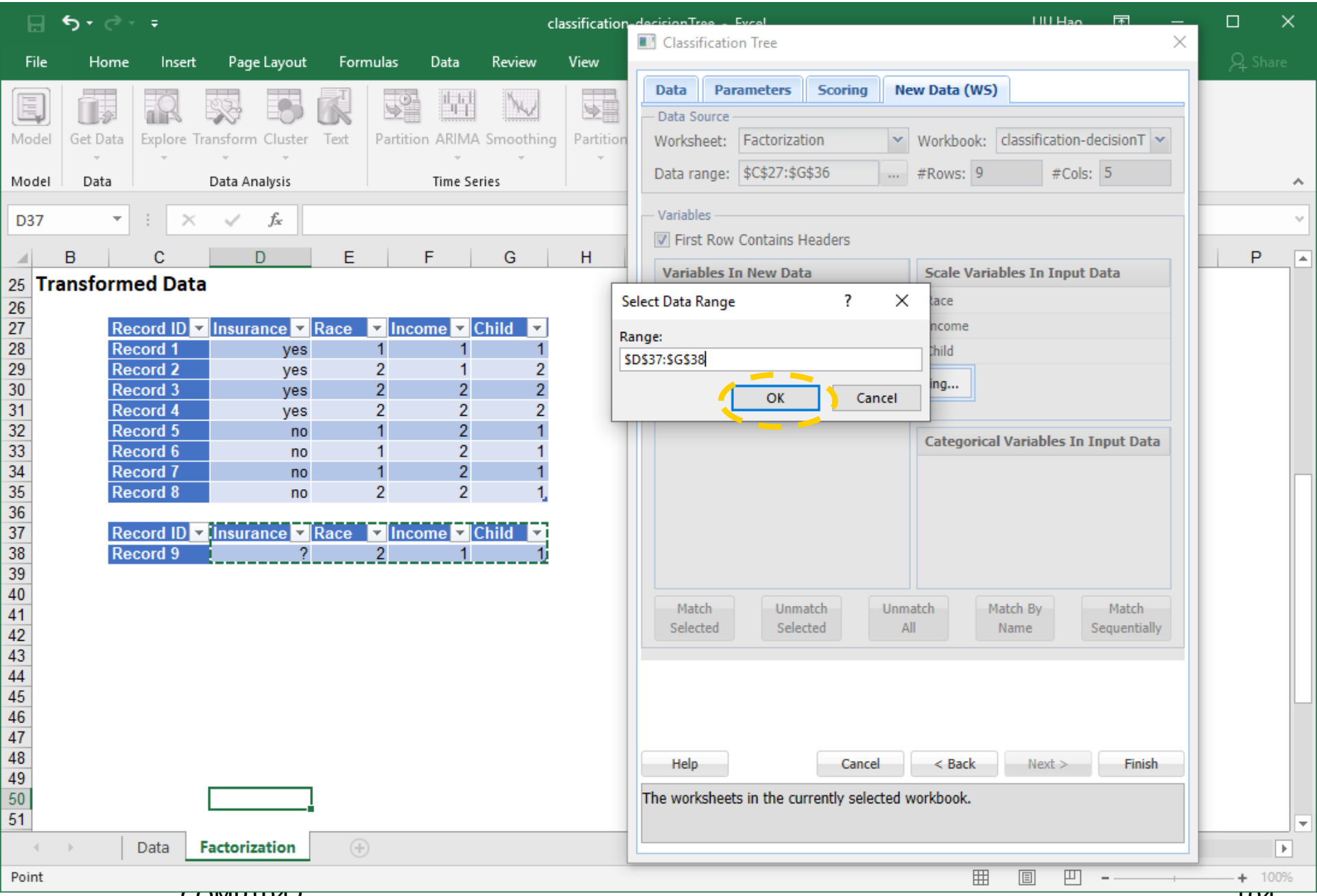

Transformed Data
Record ID | Insurance | Race | Income | Child
Record 1 | yes | 1 | 1 | 1
Record 2 | yes | 2 | 1 | 2
Record 3 | yes | 2 | 2 | 2
Record 4 | yes | 2 | 2 | 2
Record 5 | no | 1 | 2 | 1
Record 6 | no | 1 | 2 | 1
Record 7 | no | 1 | 2 | 1
Record 8 | no | 2 | 2 | 1
Record ID | Insurance | Race | Income | Child
Record 9 | ? | 2 | 1 | 1
Classification Tree
Data
Parameters
Scoring
Preprocessing
Partition Data
Rescale Data
Decision Tree: Fitting
Tree Growth
Limit Number of:
Limit Value:
Levels
Nodes
Splits
Records in T
Advanced
Prior Probability
Select Trees to Display
Tree Type
Fully Grown
Best Pruned
Minimum Error
User Specified
Number of Decision Nodes:
Done
Full grown
Tree for Scoring
Show Feature
Tree Diagram
Maximum Number of Levels:
Trees to Display
Help
Cancel
< Back
Next >
Finish
If checked, the full tree will be used.
Data
Factorization
Ready

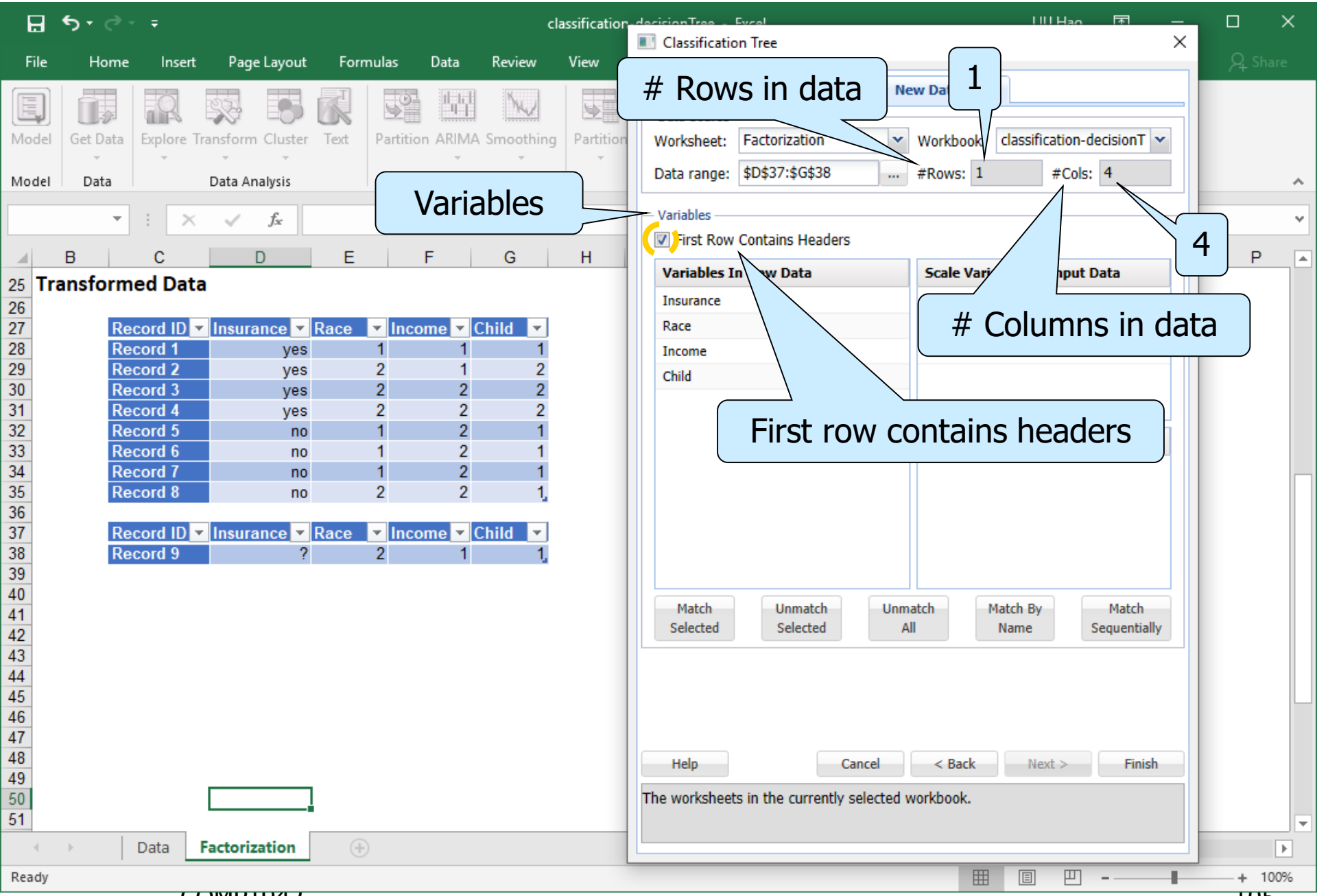

File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 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
Data
Factorization
Ready
Classification Tree
Data
Parameters
Scoring
Preprocessing
Partition Data
Rescale Data
Decision Tree: Fitting
Tree Growth
Limit Number of:
Limit Value:
Levels
Nodes
Splits
Records in Terminal Nodes
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:
Trees to Display
Help
Cancel
< Back
Next >
Finish

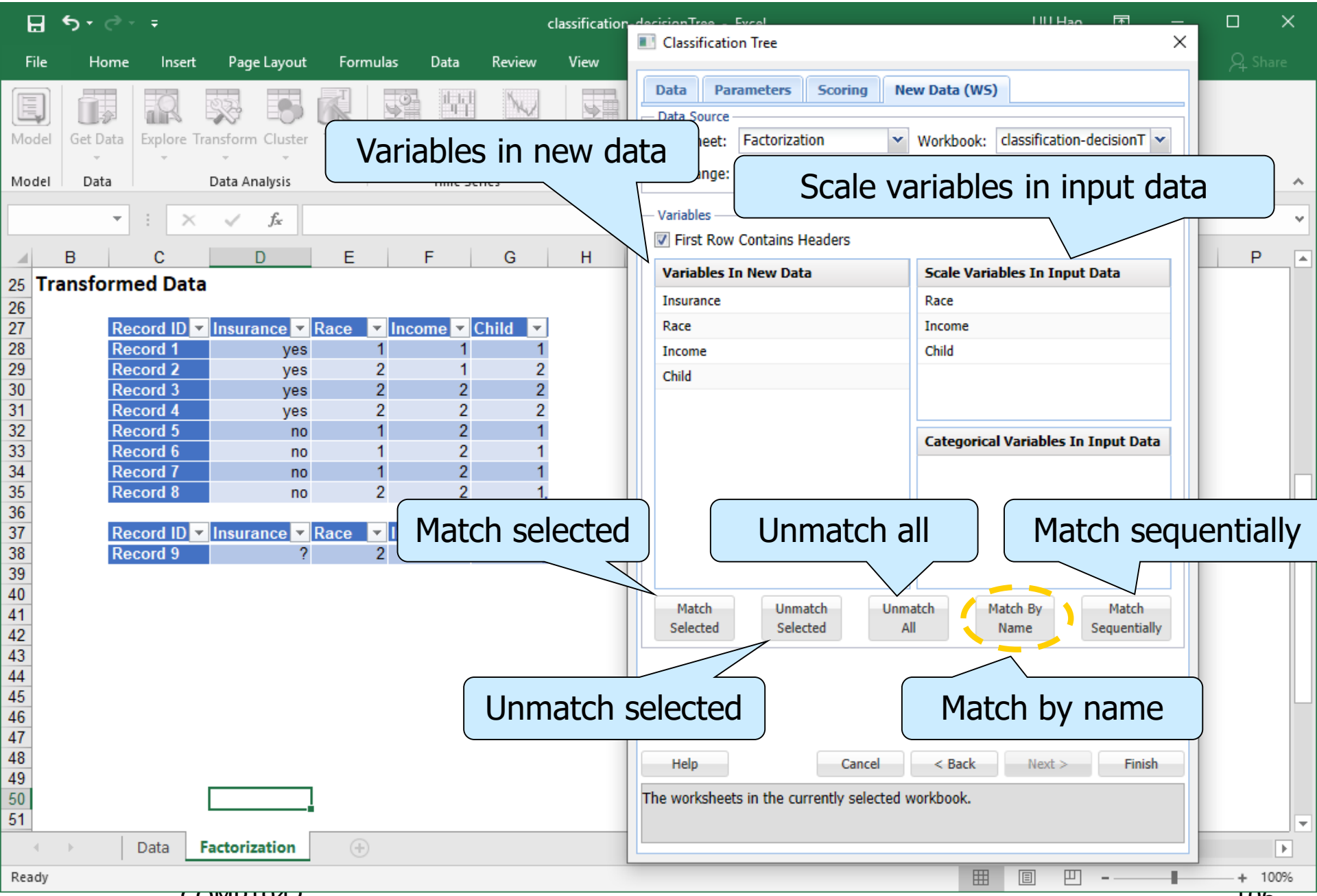

Score training data
Detailed report
Summary report
Lift charts
Classification Tree
Data
Parameters
Scoring
Score Training Data
Detailed Report
Summary Report
Lift Charts
Score New Data
In Worksheet
In Database
Help
Cancel
< Back
Next >
Finish
If checked, scoring can be performed on new data from worksheet.
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

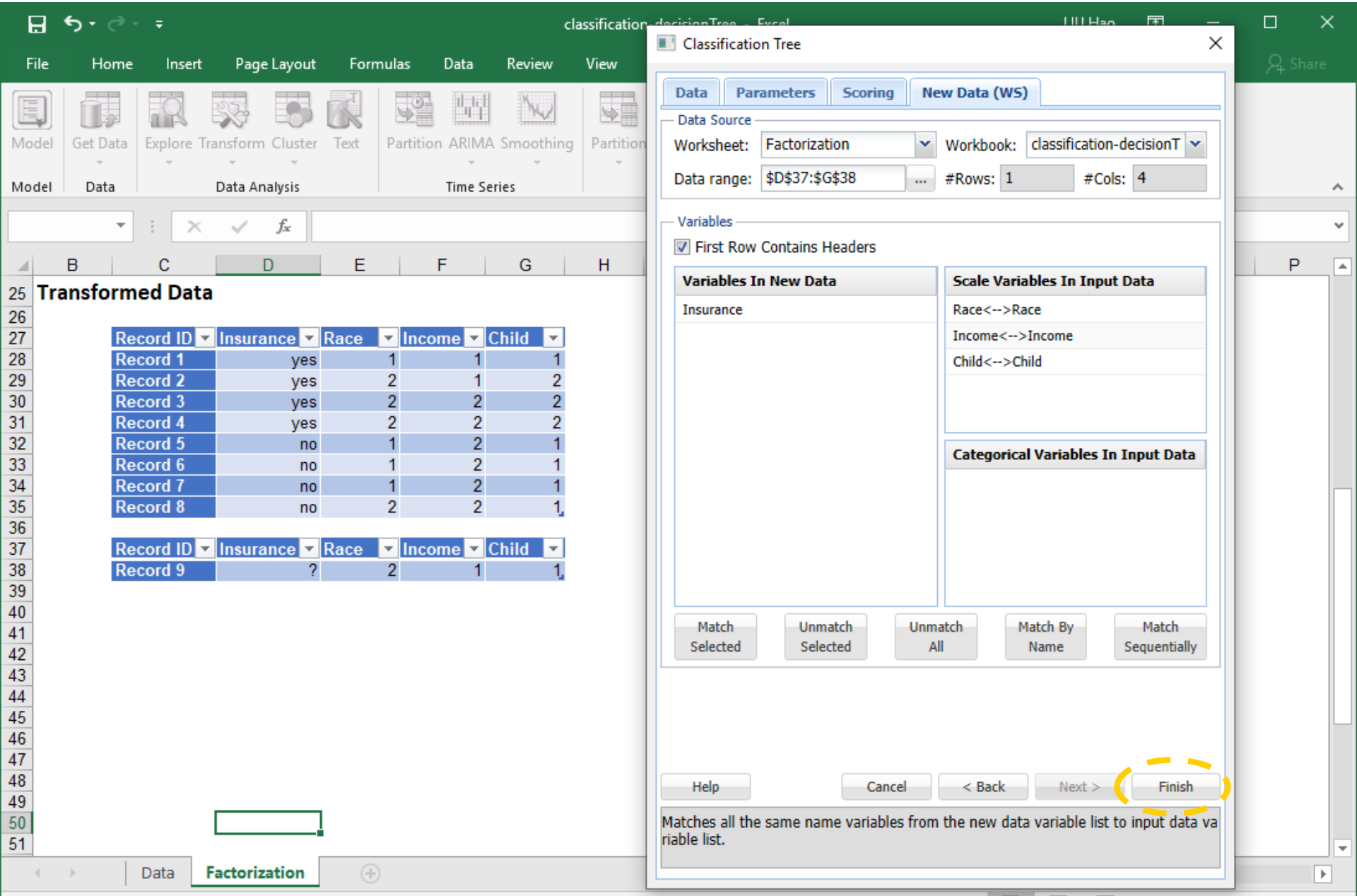

Classification Tree
Data
Parameters
Score Training Data
Detailed Report
Summary Report
Lift Charts
Score New Data
In Worksheet
Score new data
In worksheet
Help
Cancel
< Back
Next >
Finish
If checked, scoring can be performed on new data from worksheet.
Transformed Data
Record ID | Insurance | Race | 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

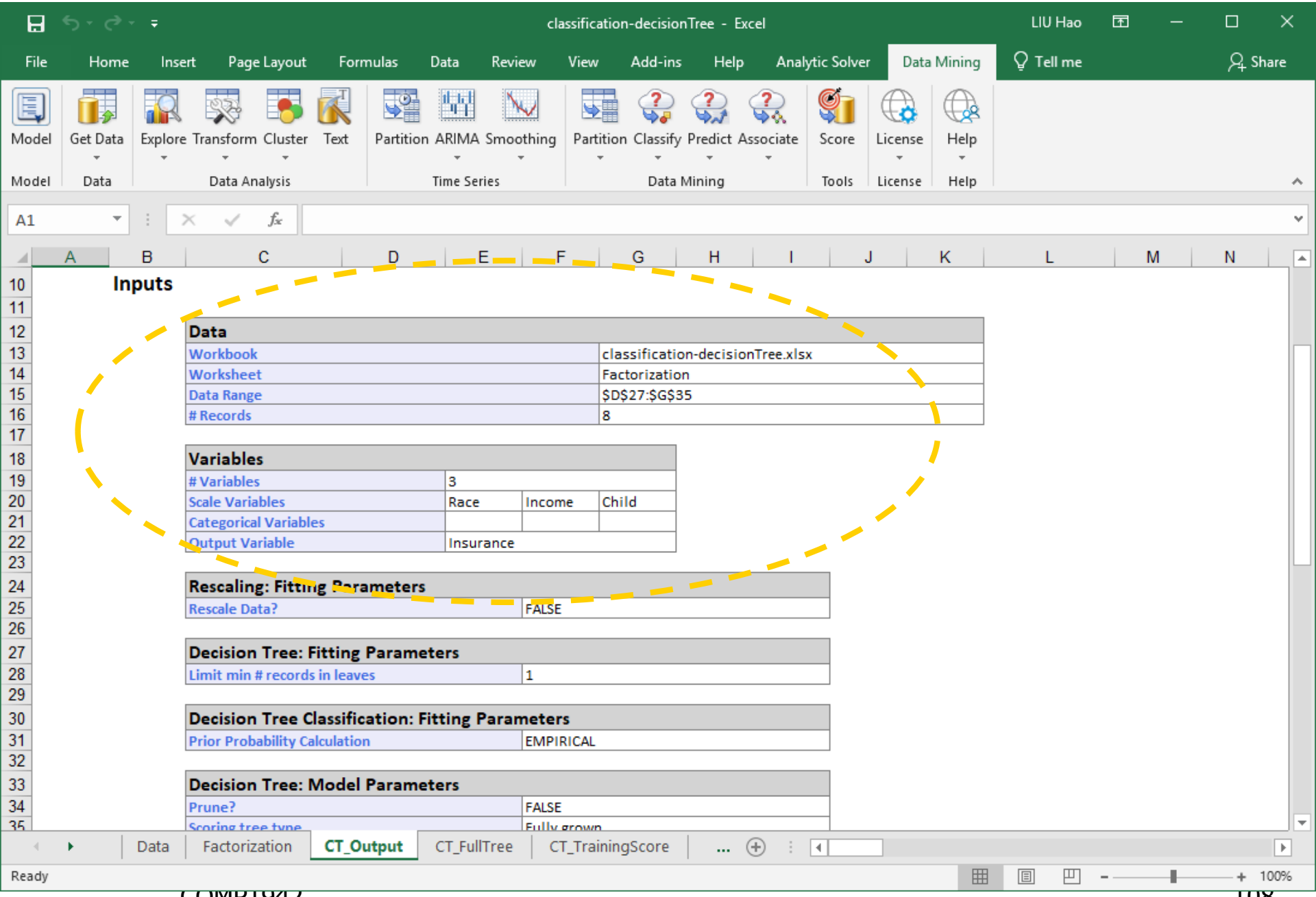

Data source
Worksheet
Data range
Workbook
Classification Tree
Data
Parameters
Scoring
New Data (WS)
Data Source
Worksheet: Factorization
Workbook: classification-decisionT
Data range: $C$27:$G$36
#Rows: 9
#Cols: 5
Variables
First Row Contains Headers
Variables In New Data
Record ID
Insurance
Race
Income
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.
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

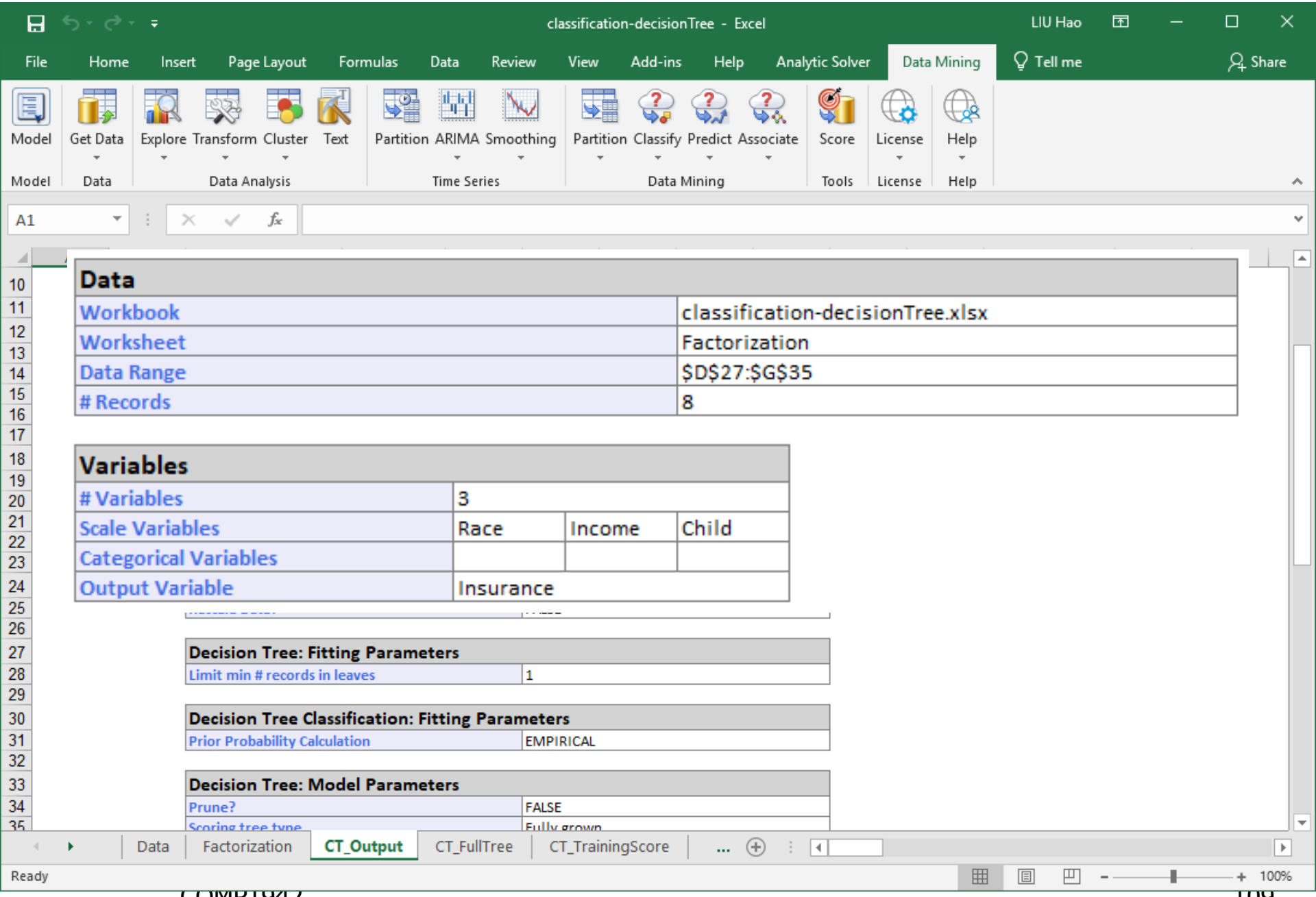

Transformed Data
Record ID Insurance Race Income Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 2 2
Record 5 no 1 2 1
Record 6 no 1 2 1
Record 7 no 1 2 1
Record 8 no 2 2 1
Record ID Insurance Race Income Child
Record 9 ? 2 1 1
Classification Tree
Data
Parameters
Scoring
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
Scale Variables In Input Data
Categorical Variables In Input Data
Match Selected
Unmatch Selected
Unmatch All
Match By Name
Match Sequentially
Select Data Range
Range:
$D$37:$G$38
OK
Cancel
Help
Cancel
< Back
Next >
Finish
The worksheets in the currently selected workbook.
Data
Factorization
Point

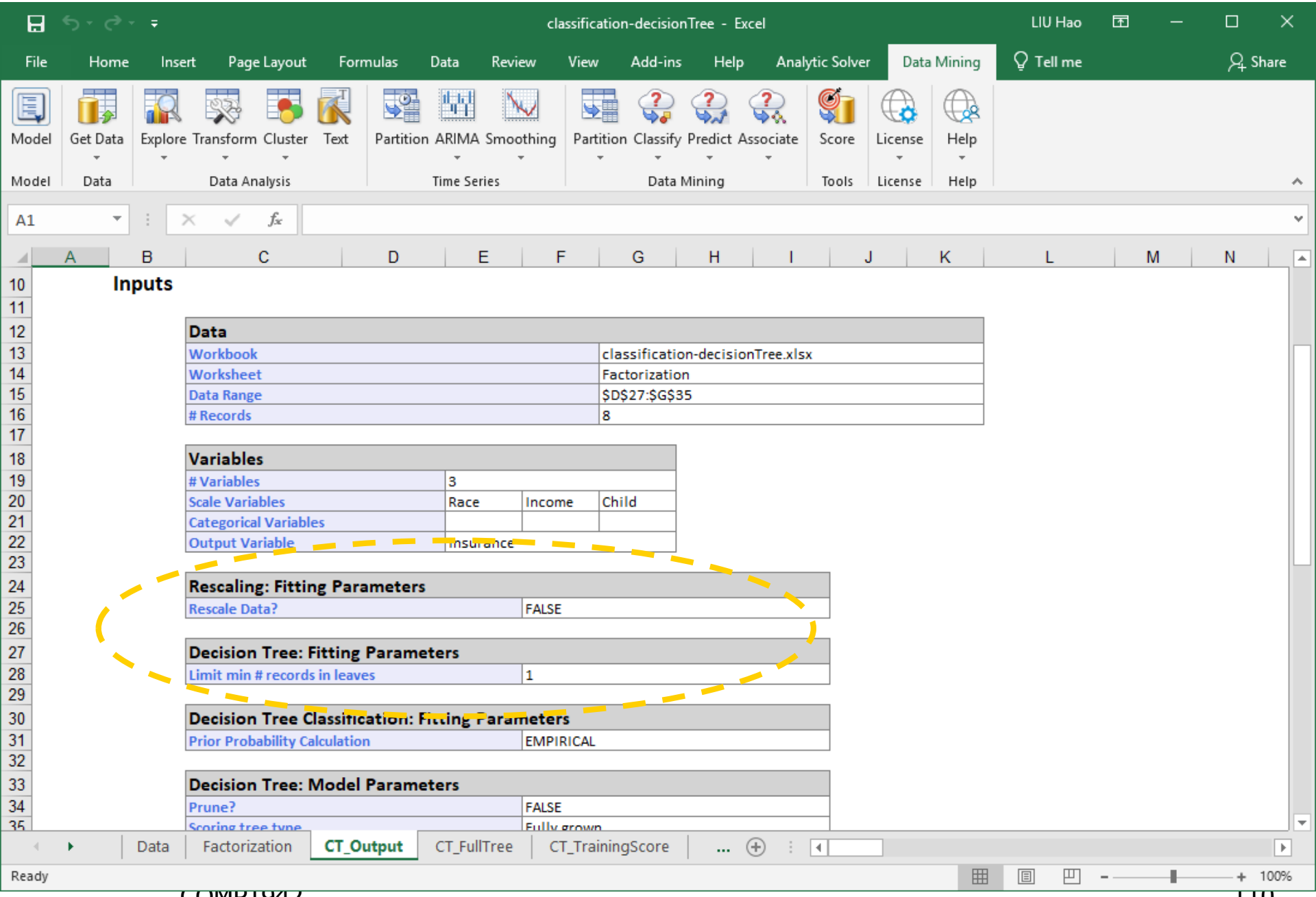

Classification Tree
# Rows in data
1
Worksheet: Factorization
Workbook: classification-decisionT
Data range: $D$37:$G$38
#Rows: 1
#Cols: 4
Variables
First Row Contains Headers
Variables In
Scale Vari
Insurance
Race
Income
Child
Match Selected
Unmatch Selected
Unmatch All
Match By Name
Match Sequentially
# Columns in data
First row contains headers
Help
Cancel
< Back
Next >
Finish
The worksheets in the currently selected workbook.
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

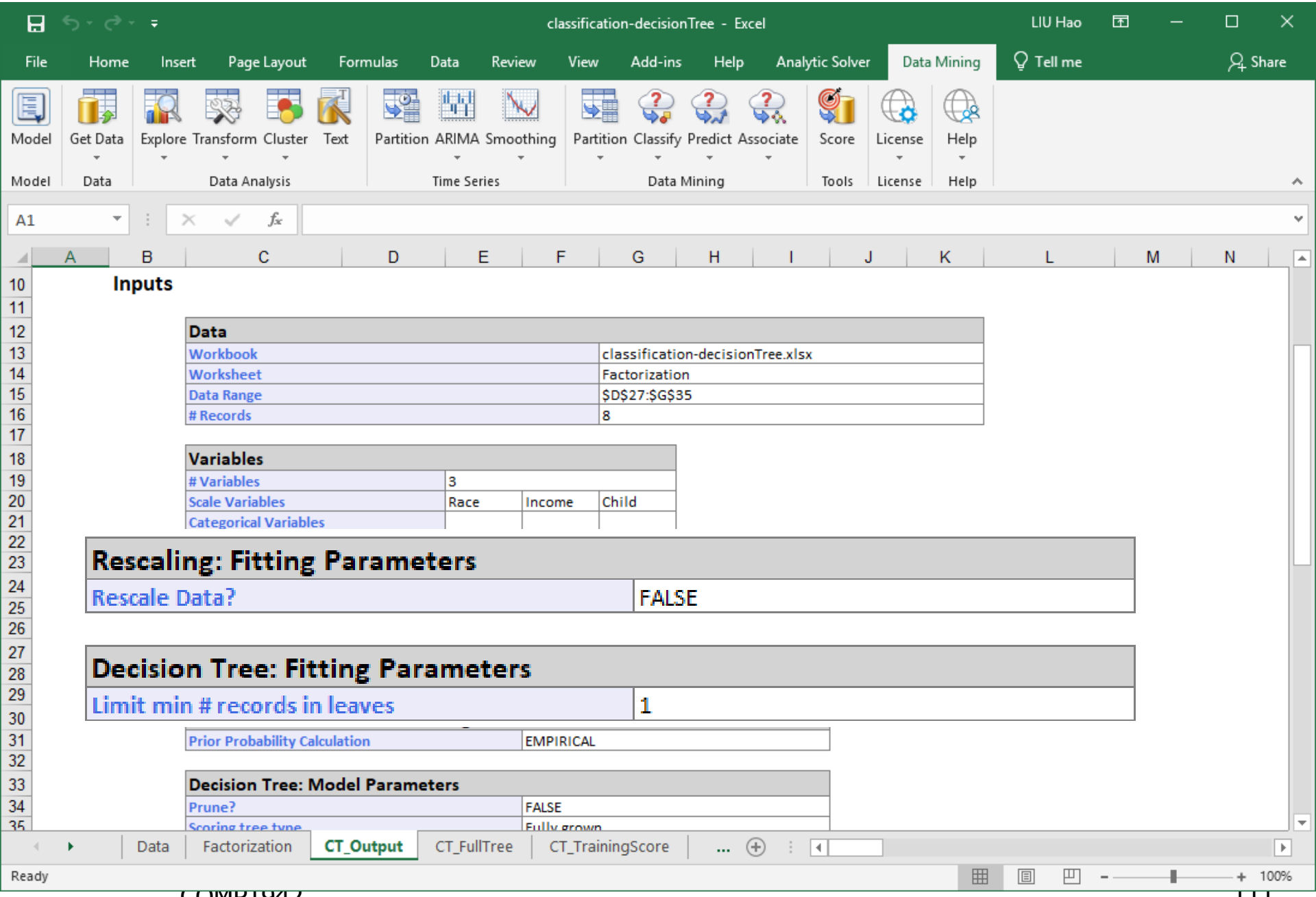

Variables in new data
Scale variables in input data
Match selected
Unmatch all
Match sequentially
Unmatch selected
Match by name
Classification Tree
Data
Parameters
Scoring
New Data (WS)
Data Source
Factorization
classification-decisionT
Variables
First Row Contains Headers
Variables In New Data
Insurance
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.
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
Data
Factorization
Ready

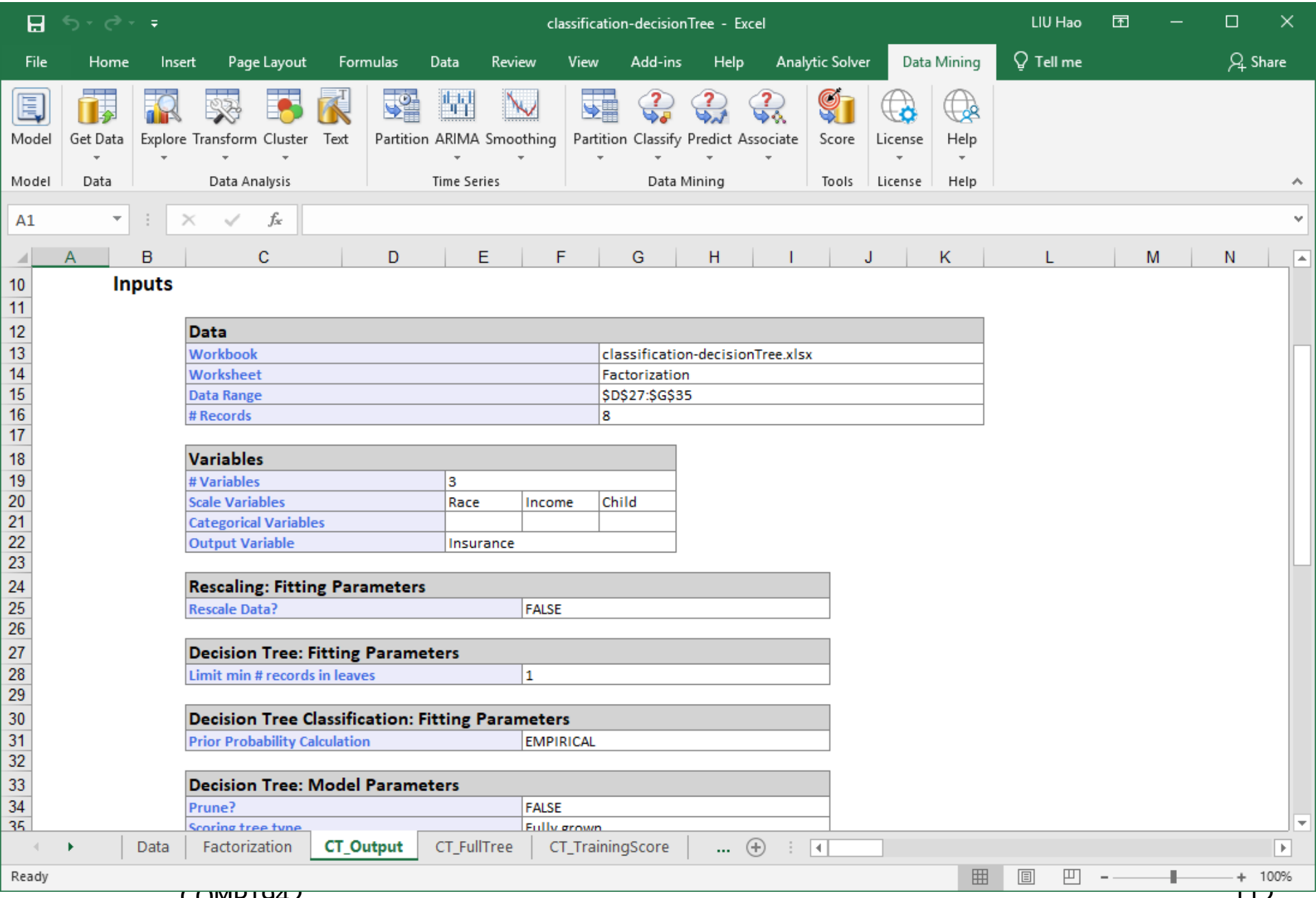

File
Home
Insert
Page Layout
Formulas
Data
Review
View
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Model
Data
Data Analysis
Time Series
Transformed Data
Record ID
Insurance
Race
Income
Child
Record 1 yes 1 1 1
Record 2 yes 2 1 2
Record 3 yes 2 2 2
Record 4 yes 2 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
Data
Factorization
Ready
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.

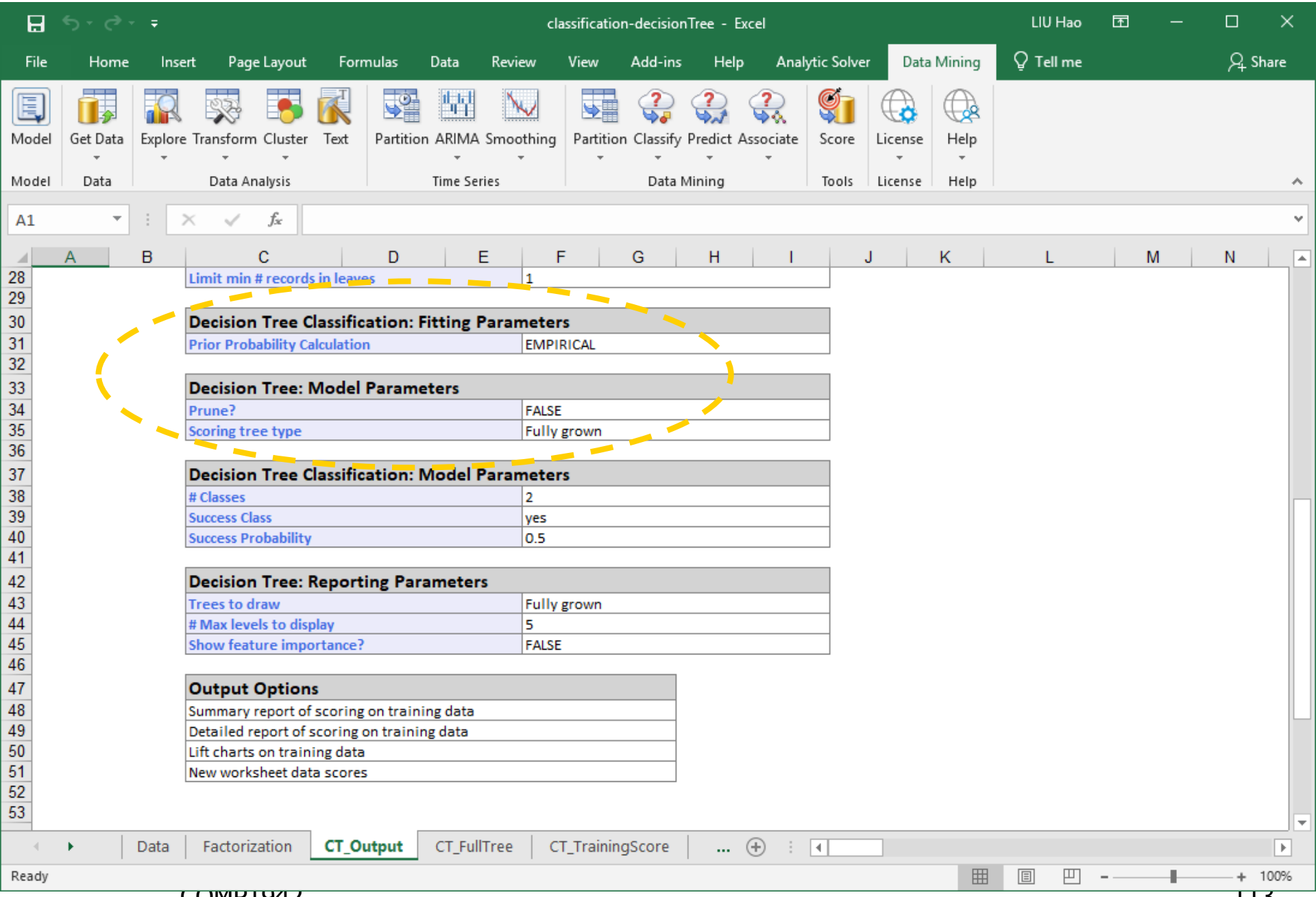

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis 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
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

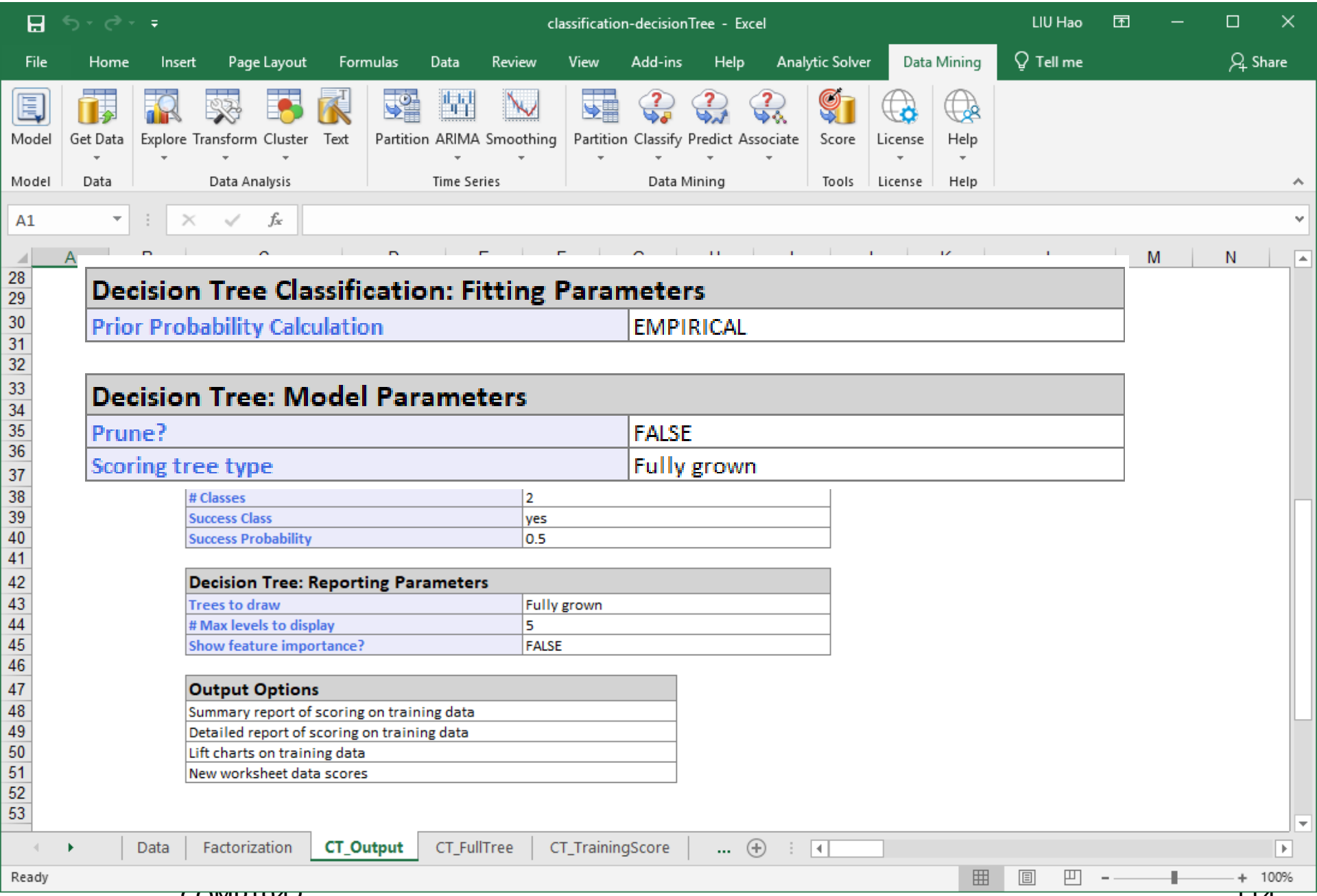

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
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
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
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

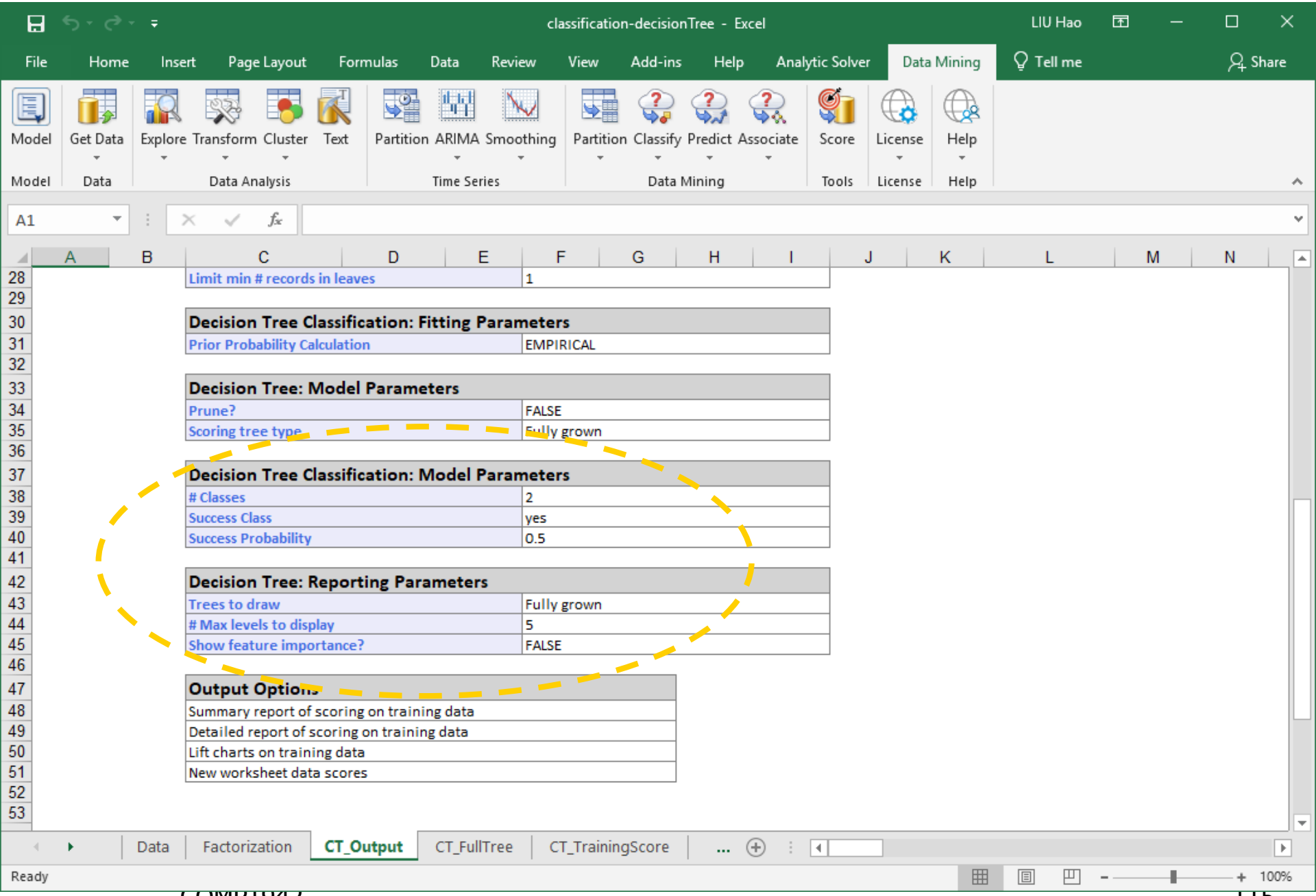

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis 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%

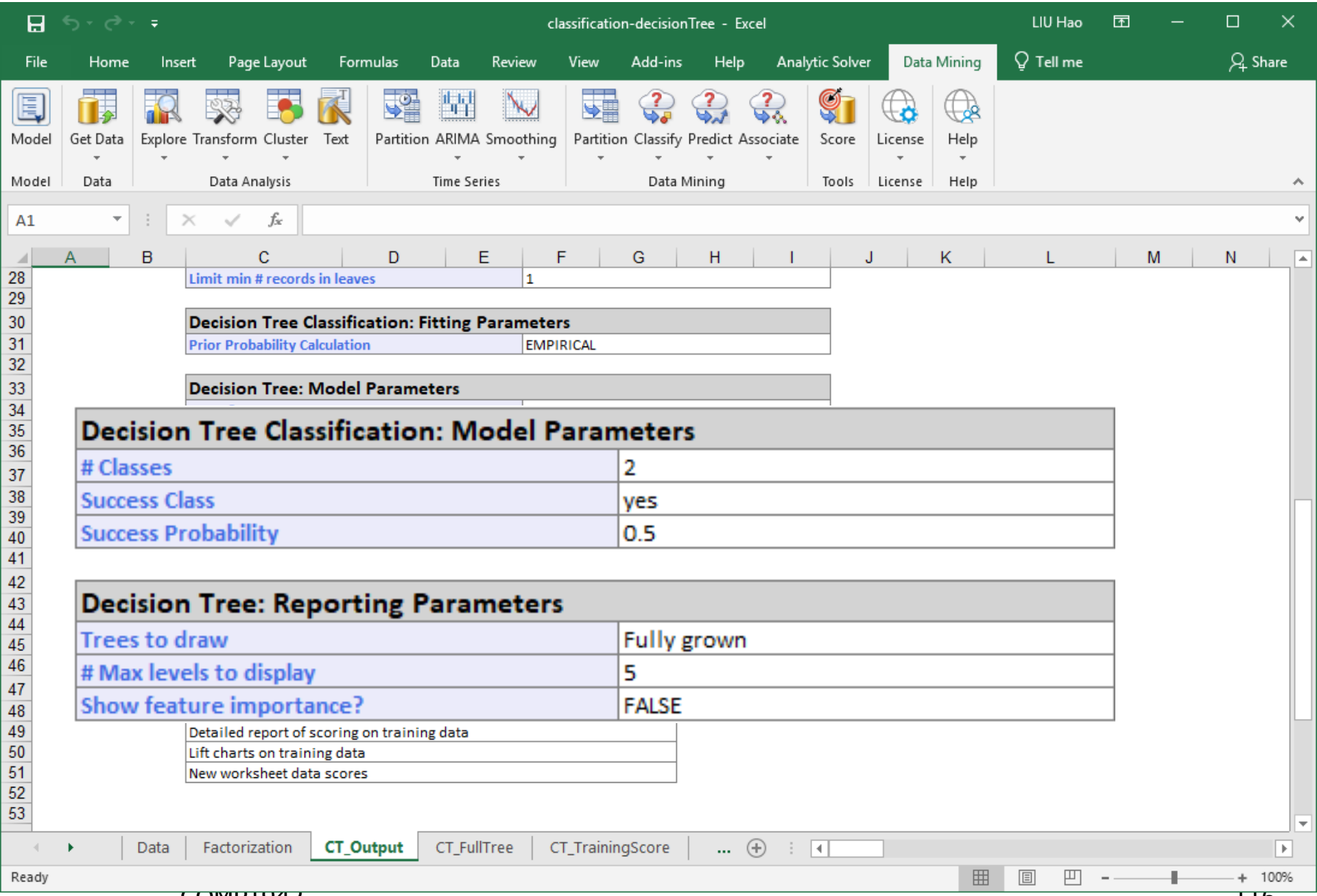

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis 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
Rescaling: Fitting Parameters
Rescale Data? FALSE
Decision Tree: Fitting Parameters
Limit min # records in leaves 1
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%

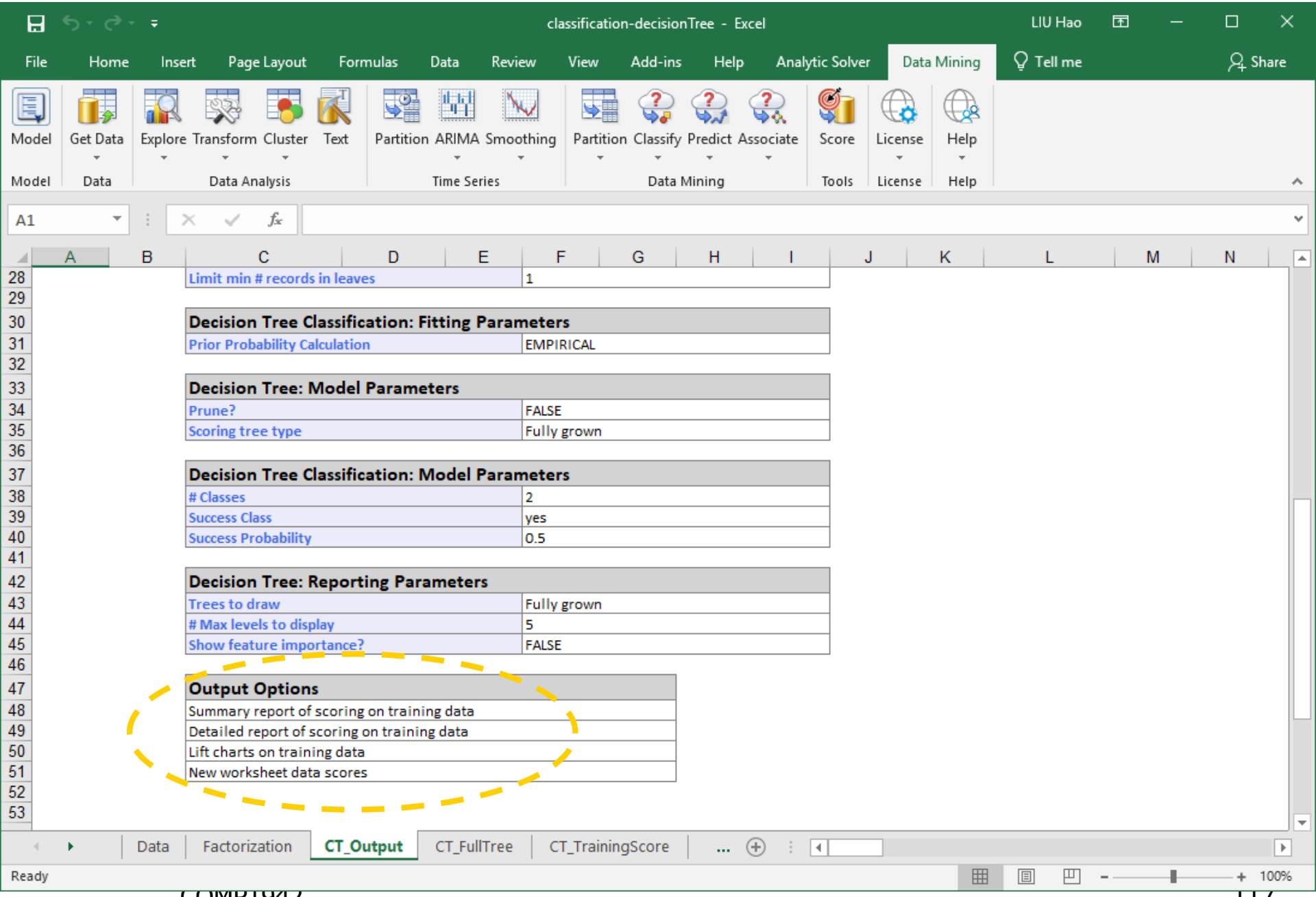

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis 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%

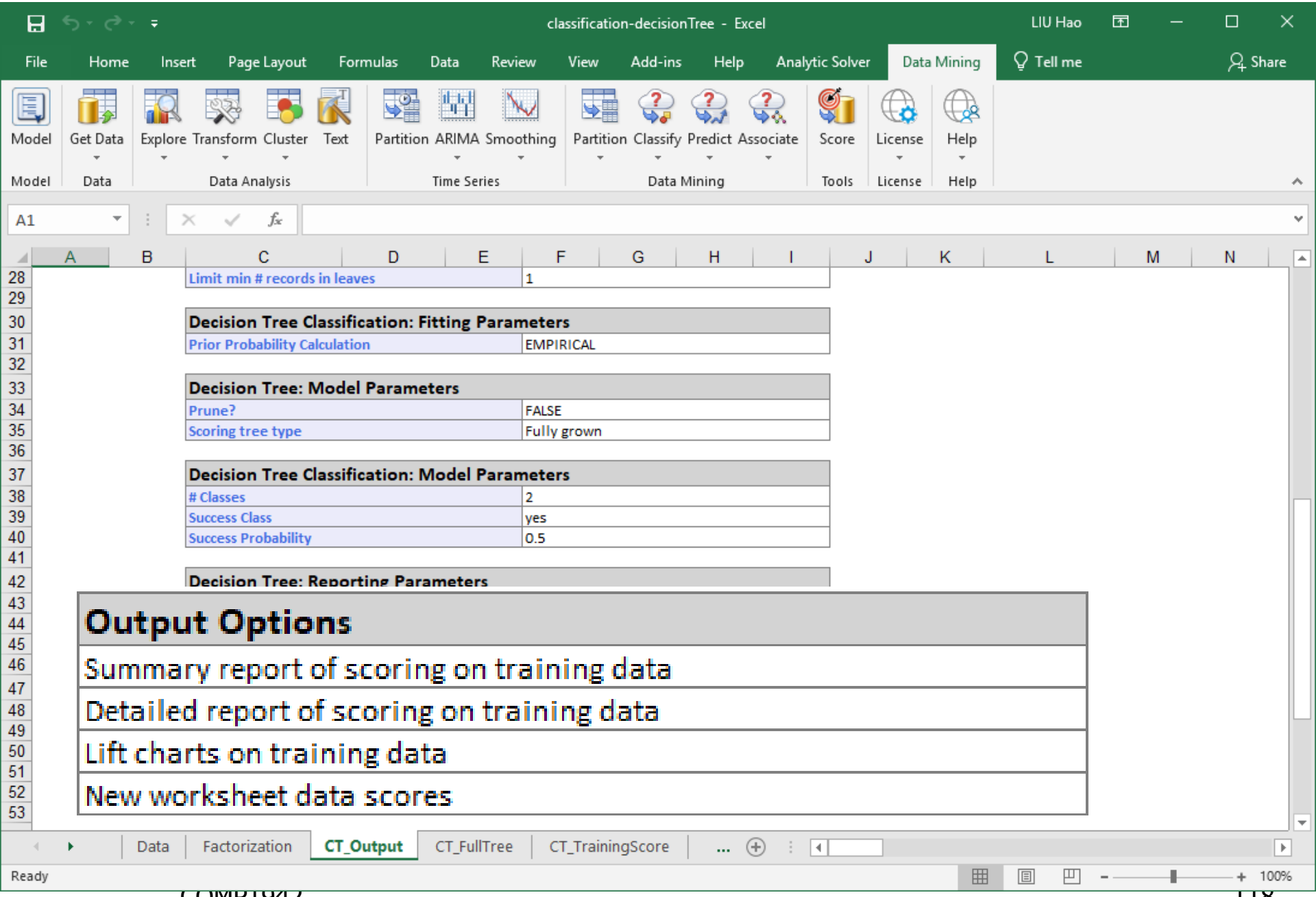

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
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
Decision Tree Classification: Model Parameters
# Classes 2
Success Class yes
Success Probability 0.5
Decision Tree: Reporting Parameters
Trees to draw Fully grown
# Max levels to display 5
Show feature importance? FALSE
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
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

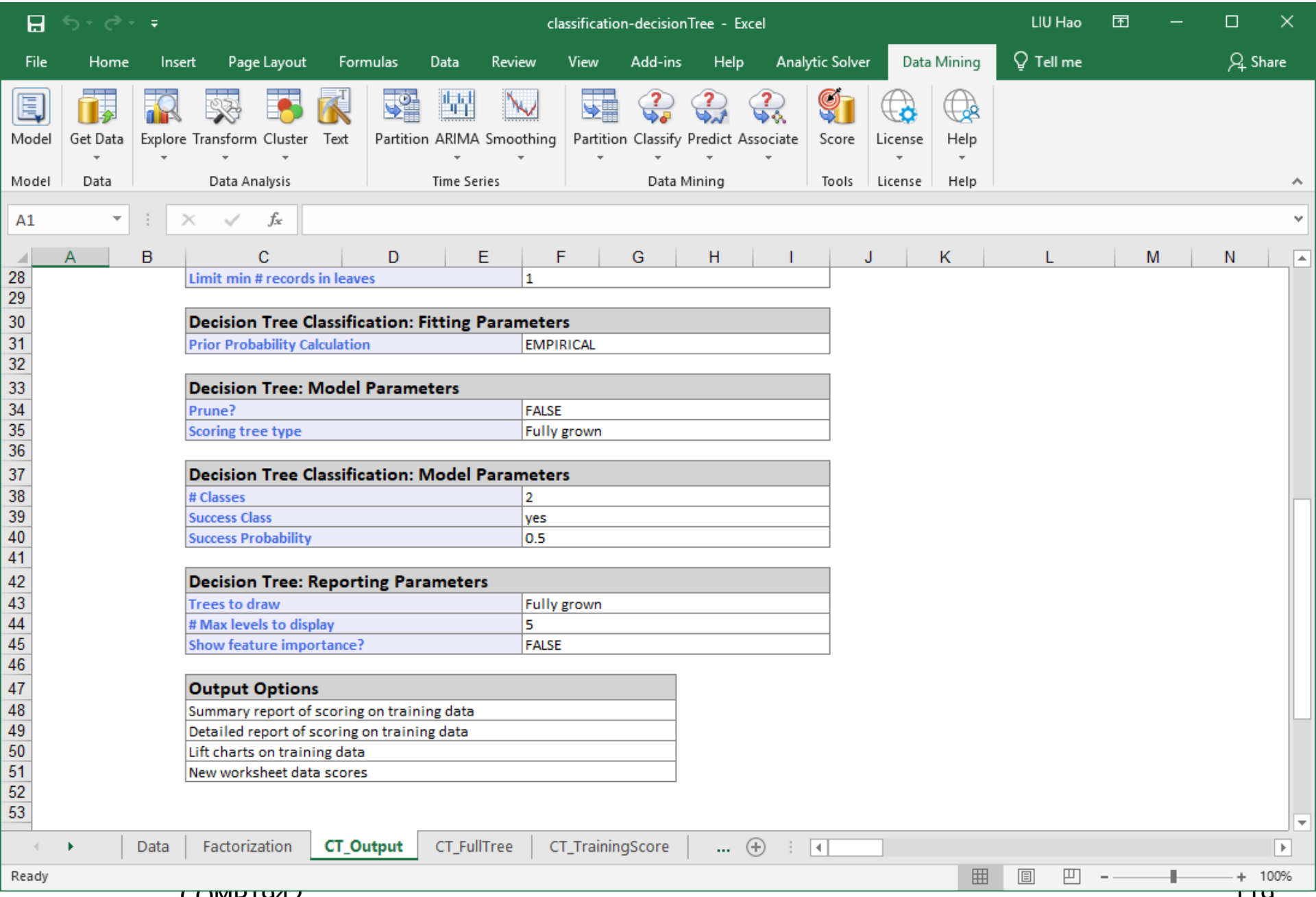

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
Decision Tree Classification: Fitting Parameters
Prior Probability Calculation
EMPIRICAL
Decision Tree: Model Parameters
Prune?
FALSE
Scoring tree type
Fully grown
# Classes
2
Success Class
yes
Success Probability
0.5
Decision Tree: Reporting Parameters
Trees to draw
Fully grown
# Max levels to display
5
Show feature importance?
FALSE
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
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Ready
100%

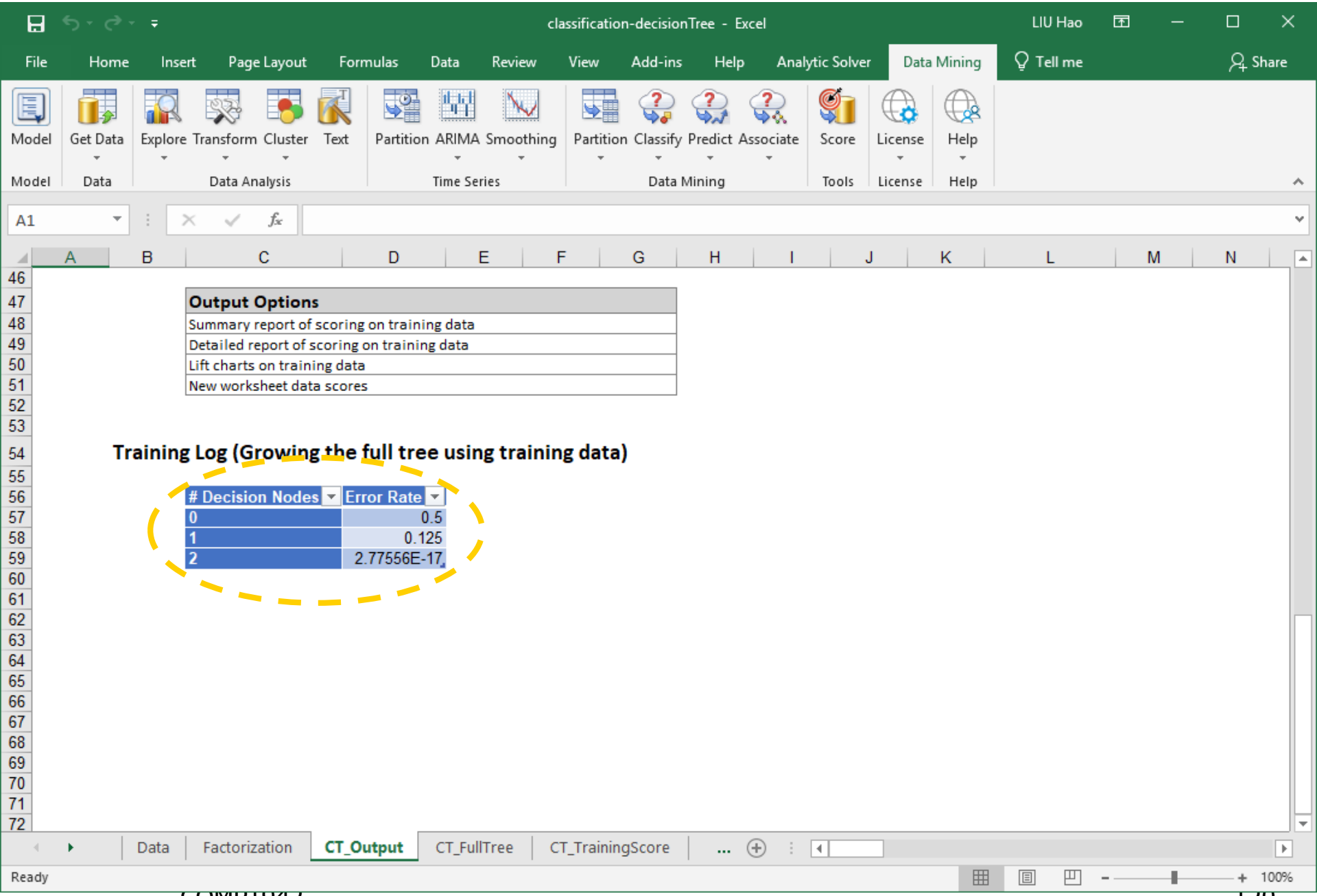

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
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
Decision Tree Classification: Model Parameters
# Classes 2
Success Class yes
Success Probability 0.5
Decision Tree: Reporting Parameters
Trees to draw Fully grown
# Max levels to display 5
Show feature importance? FALSE
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
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

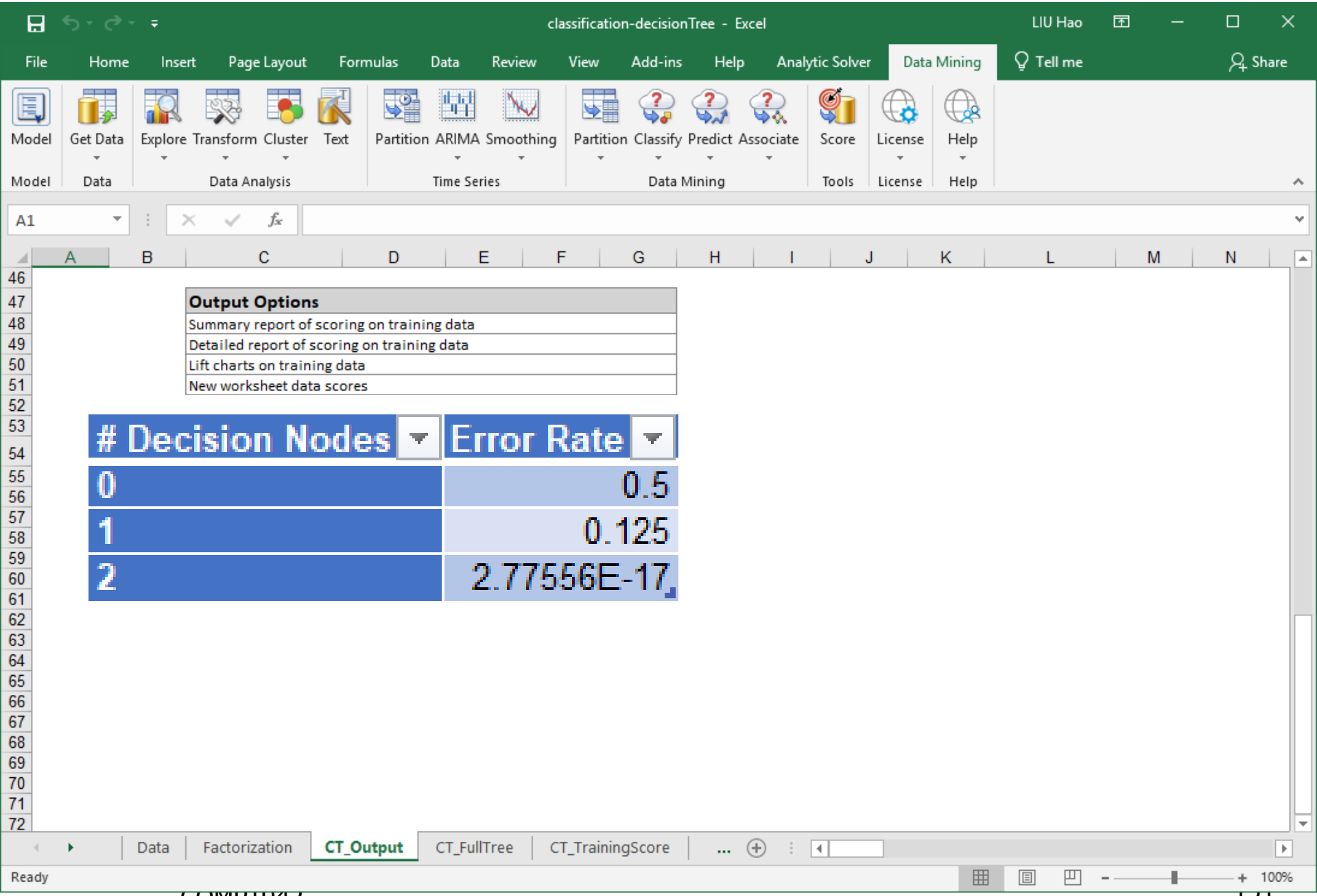

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Limit min # records in leaves 1
Decision Tree Classification: Fitting Parameters
Prior Probability Calculation EMPIRICAL
Decision Tree: Model Parameters
Decision Tree Classification: Model Parameters
# Classes 2
Success Class yes
Success Probability 0.5
Decision Tree: Reporting Parameters
Trees to draw Fully grown
# Max levels to display 5
Show feature importance? FALSE
Detailed report of scoring on training data
Lift charts on training data
New worksheet data scores
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

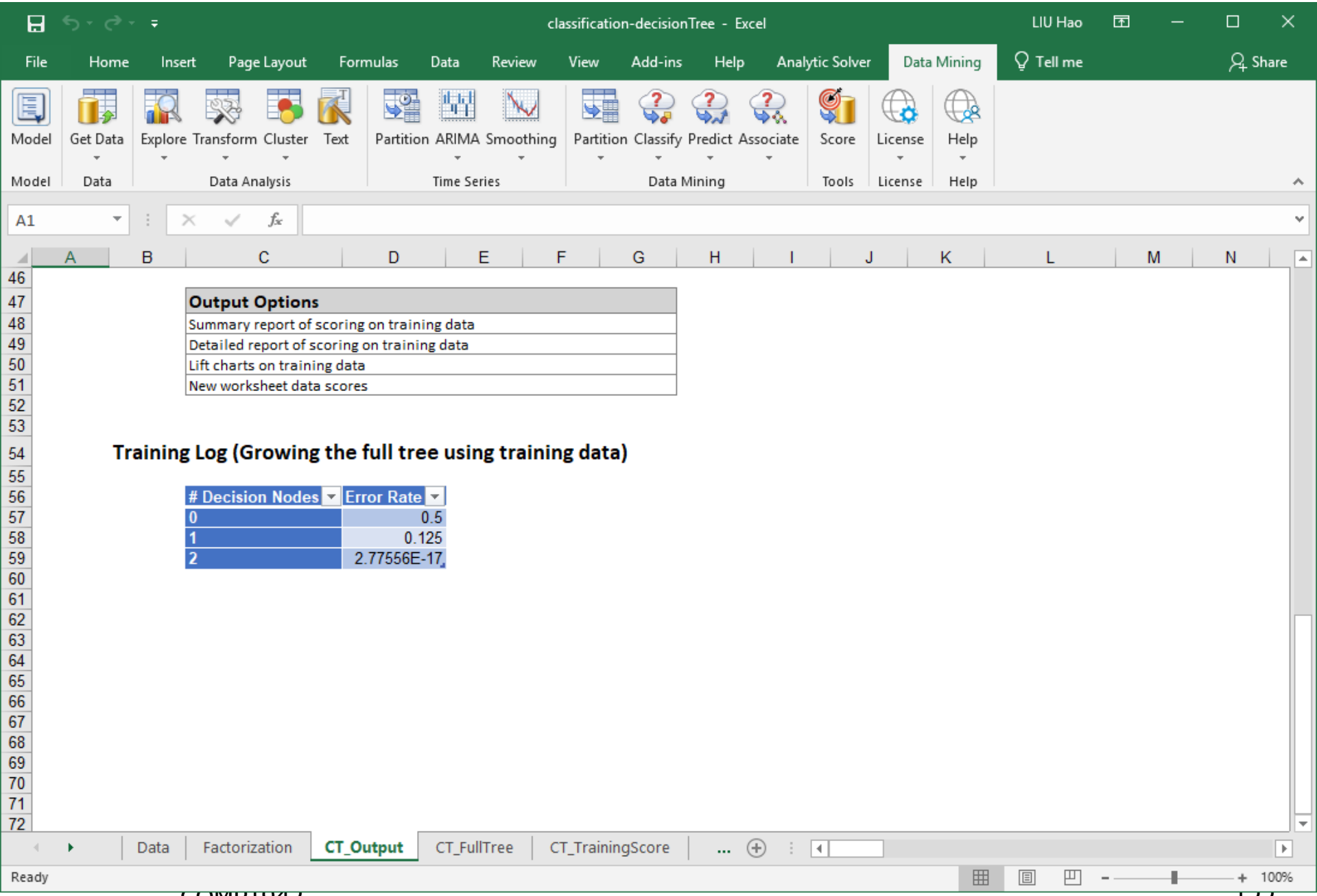

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
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
Decision Tree Classification: Model Parameters
# Classes 2
Success Class yes
Success Probability 0.5
Decision Tree: Reporting Parameters
Trees to draw Fully grown
# Max levels to display 5
Show feature importance? FALSE
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
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

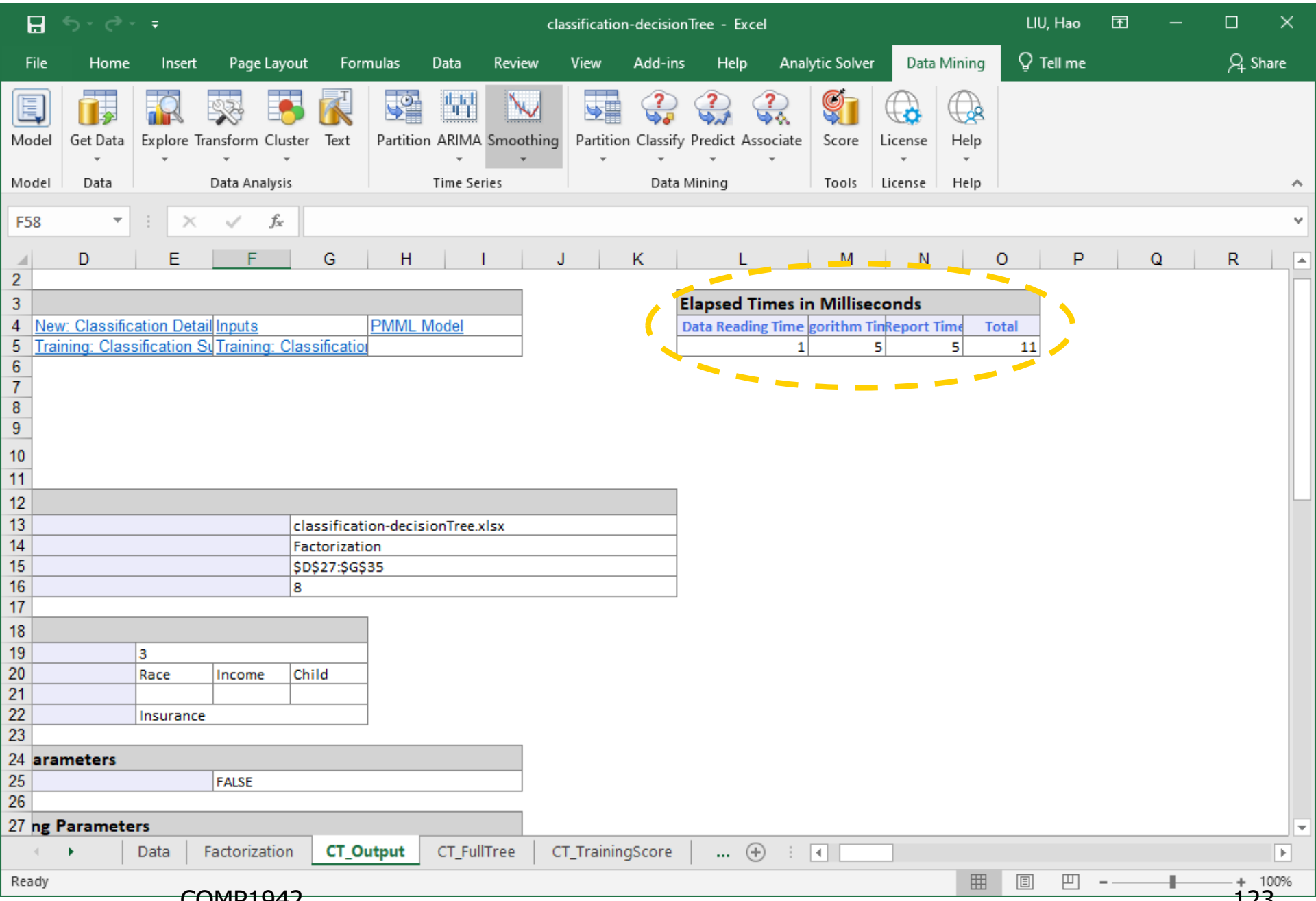

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
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
Decision Tree Classification: Model Parameters
# Classes 2
Success Class yes
Success Probability 0.5
Decision Tree: Reporting Parameters
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
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

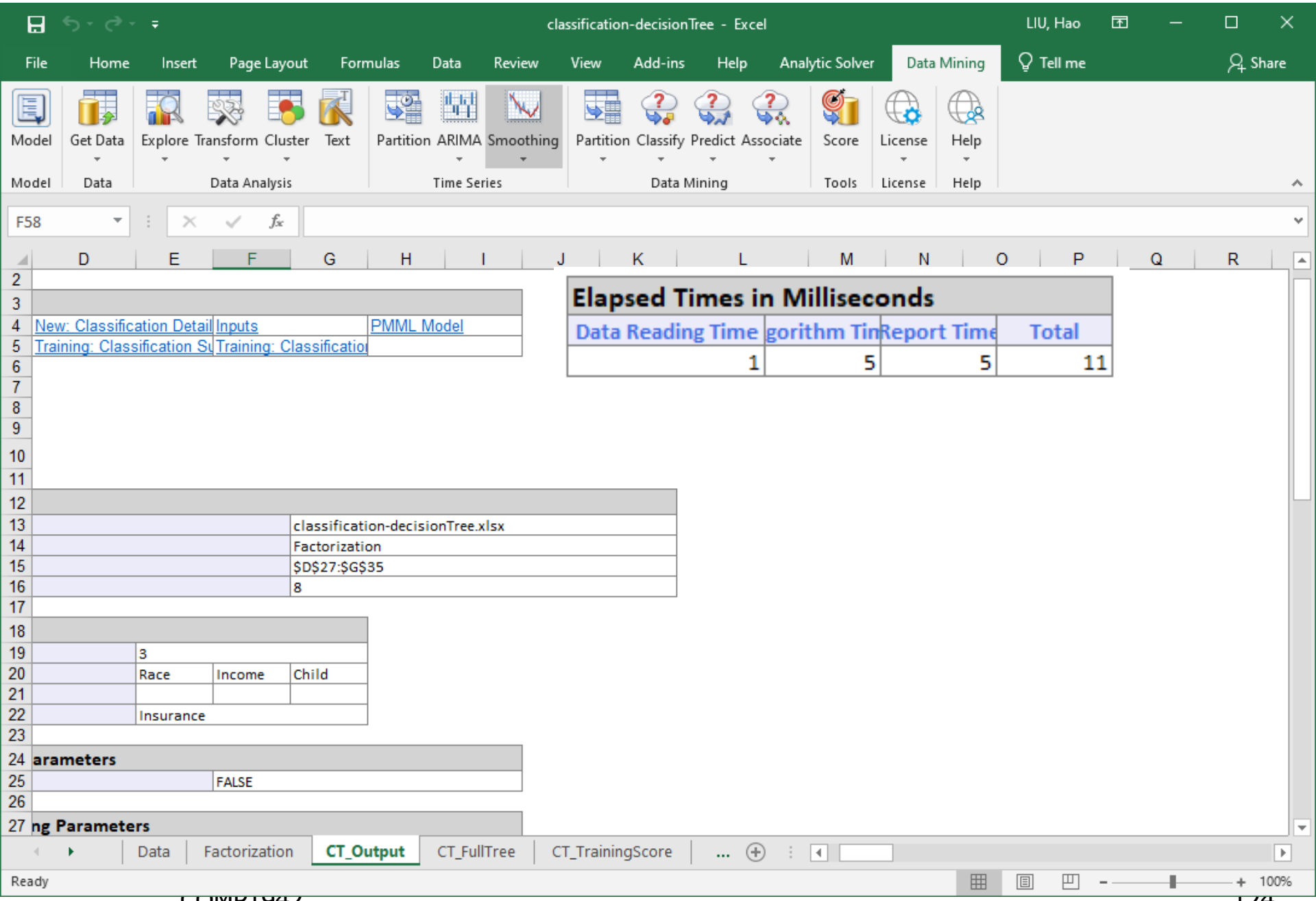

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
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
Decision Tree Classification: Model Parameters
# Classes 2
Success Class yes
Success Probability 0.5
Decision Tree: Reporting Parameters
Trees to draw Fully grown
# Max levels to display 5
Show feature importance? FALSE
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
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

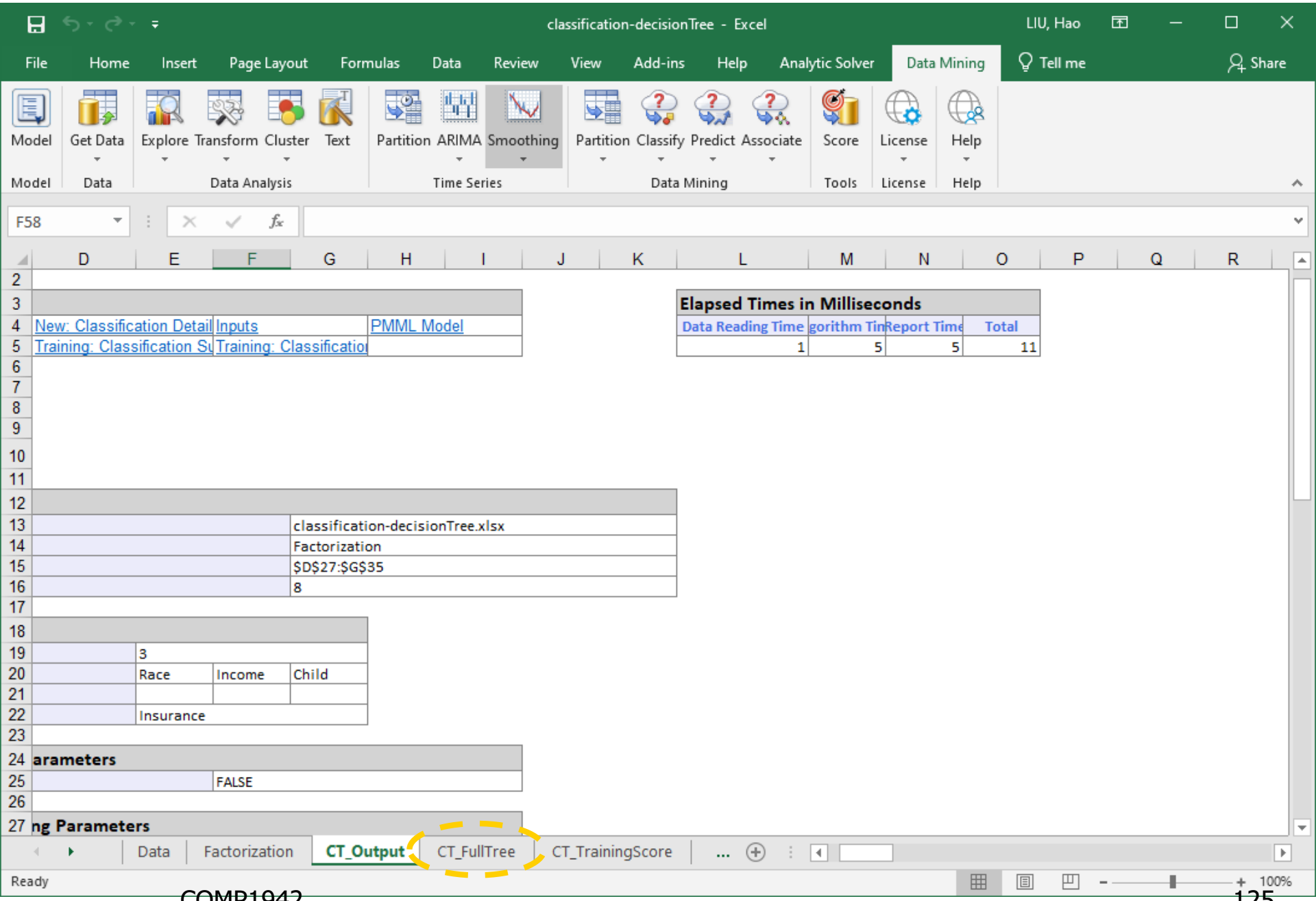

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
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
Training Log (Growing the full tree using training data)
# Decision Nodes
Error Rate
0
0.5
1
0.125
2
2.77556E-17
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Ready
100%

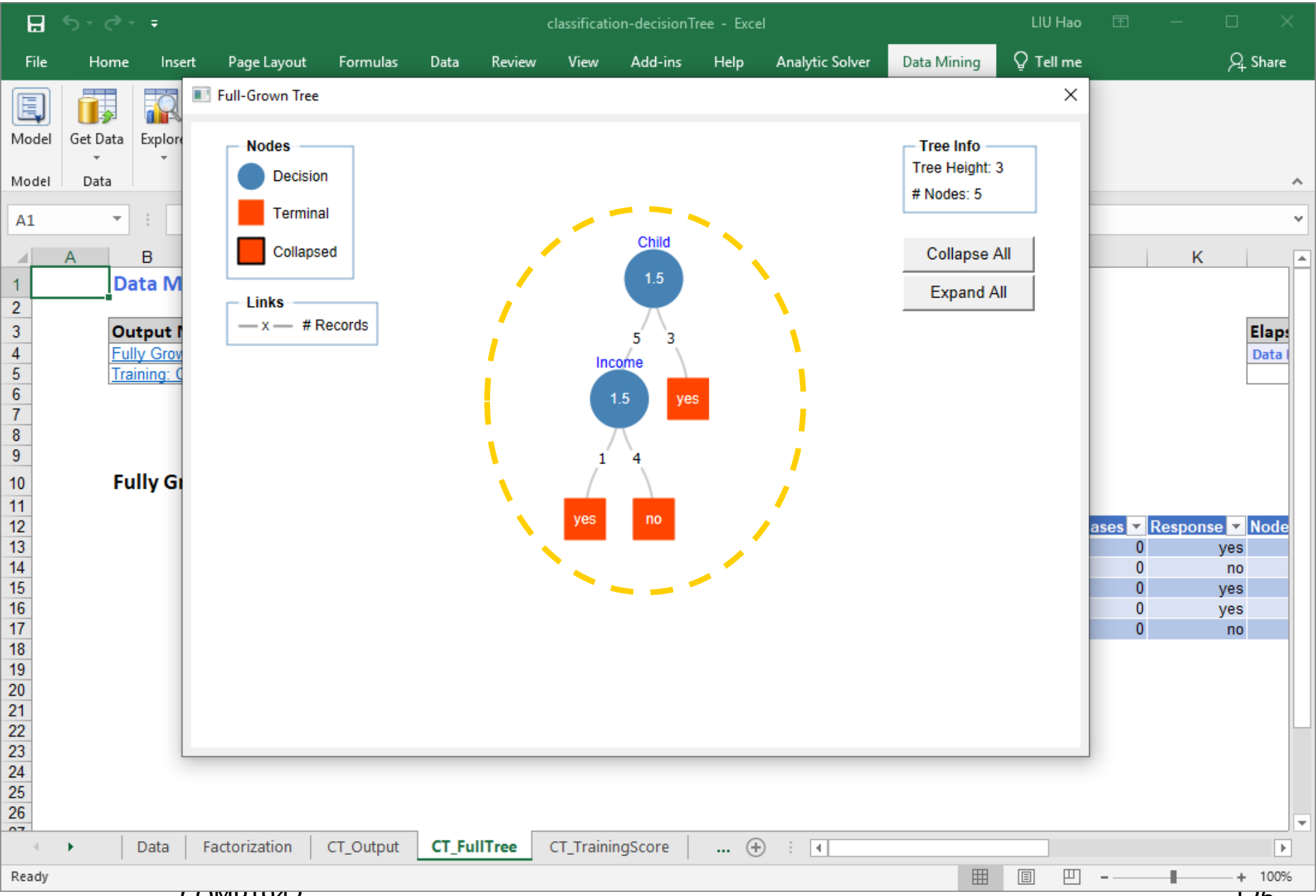

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
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
# Decision Nodes
Error Rate
0
0.5
1
0.125
2
2.77556E-17
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Ready
100%

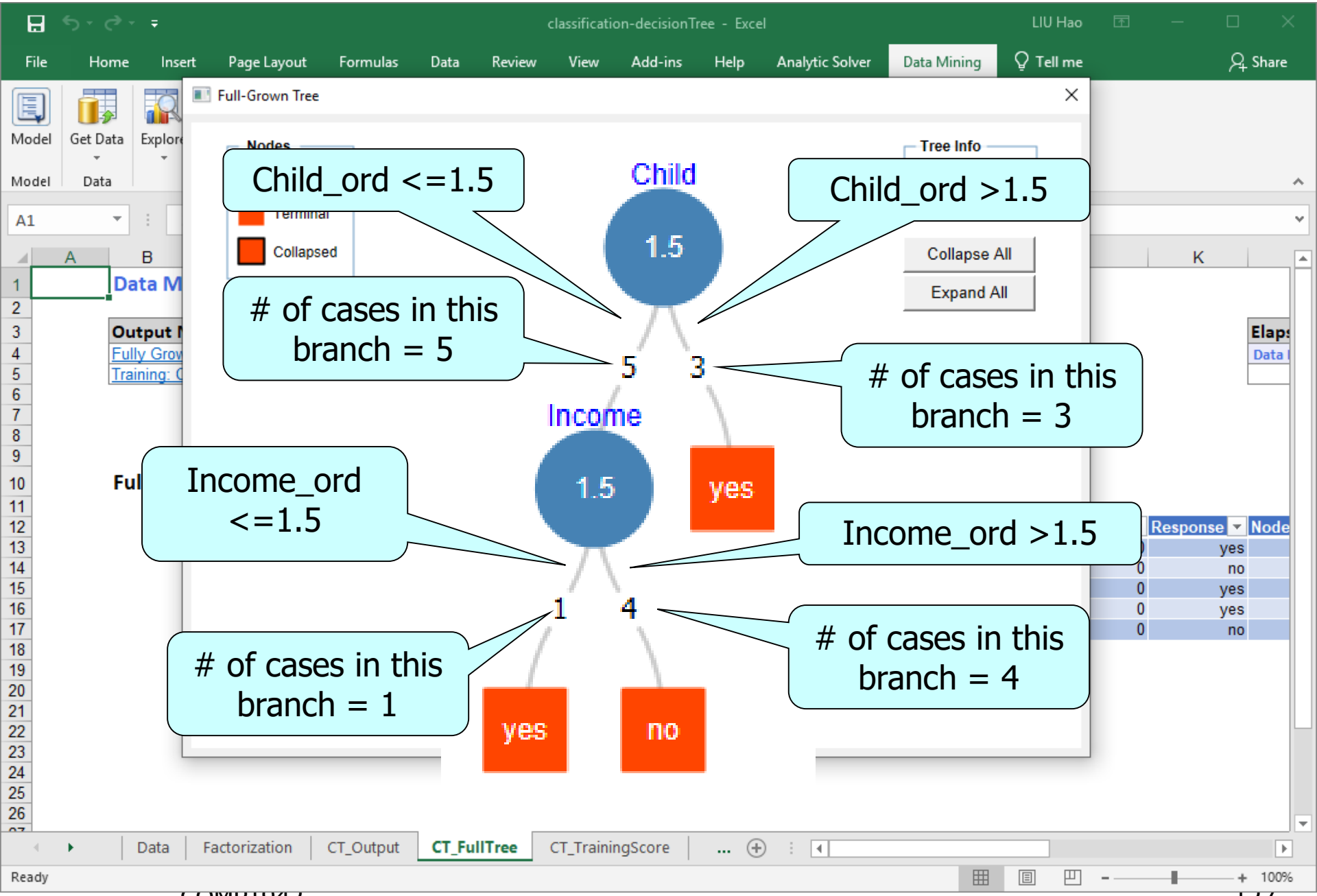

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
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
Training Log (Growing the full tree using training data)
# Decision Nodes
Error Rate
0
0.5
1
0.125
2
2.77556E-17
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Ready
100%

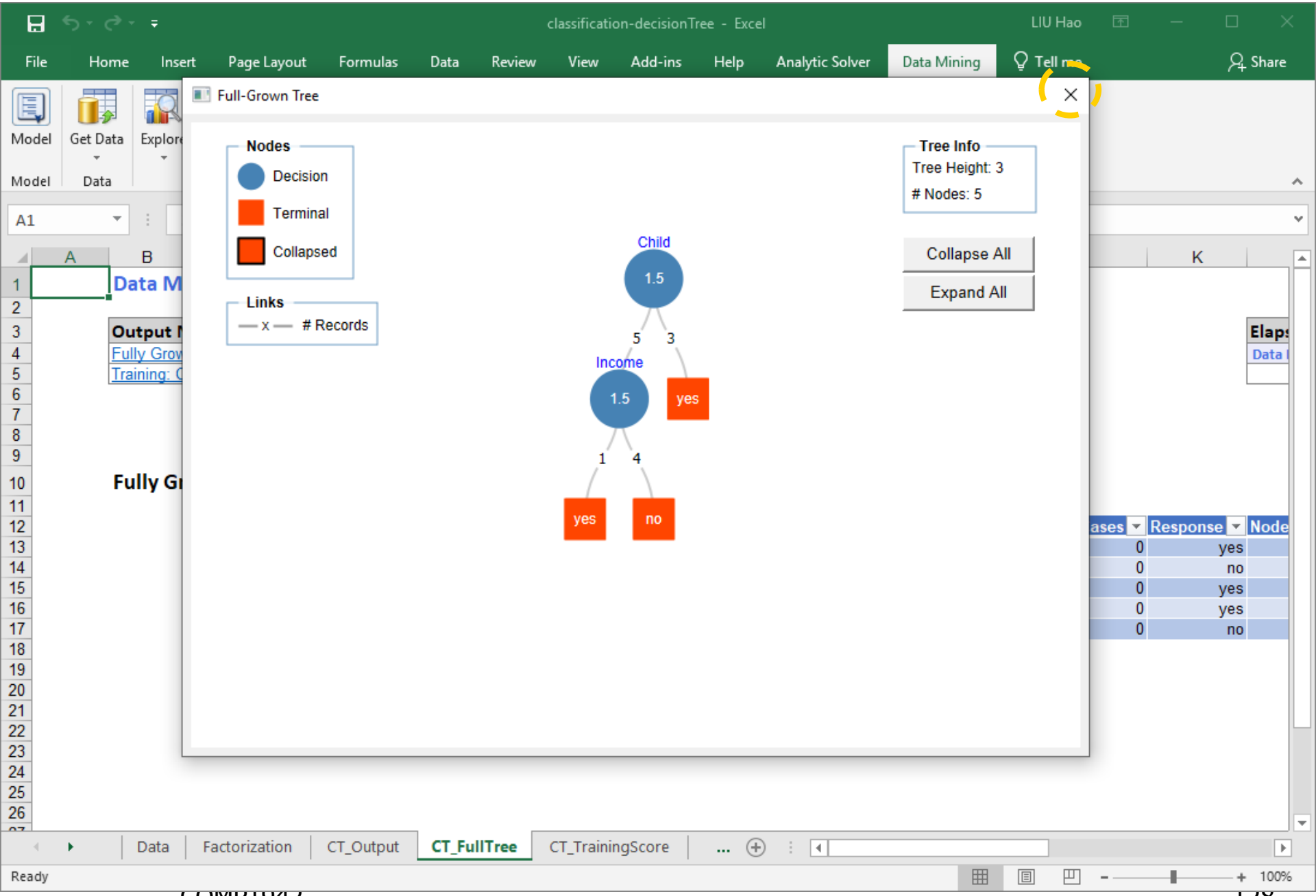

classification-decisionTree - Excel
LIU, Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
F58
Elapsed Times in Milliseconds
Data Reading Time gorithm Tin Report Time Total
1 5 5 11
New: Classification Detail Inputs PMML Model
Training: Classification Su Training: Classificatio
classification-decisionTree.xlsx
Factorization
$D$27:$G$35
8
3
Race Income Child
Insurance
arameters
FALSE
ng Parameters
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

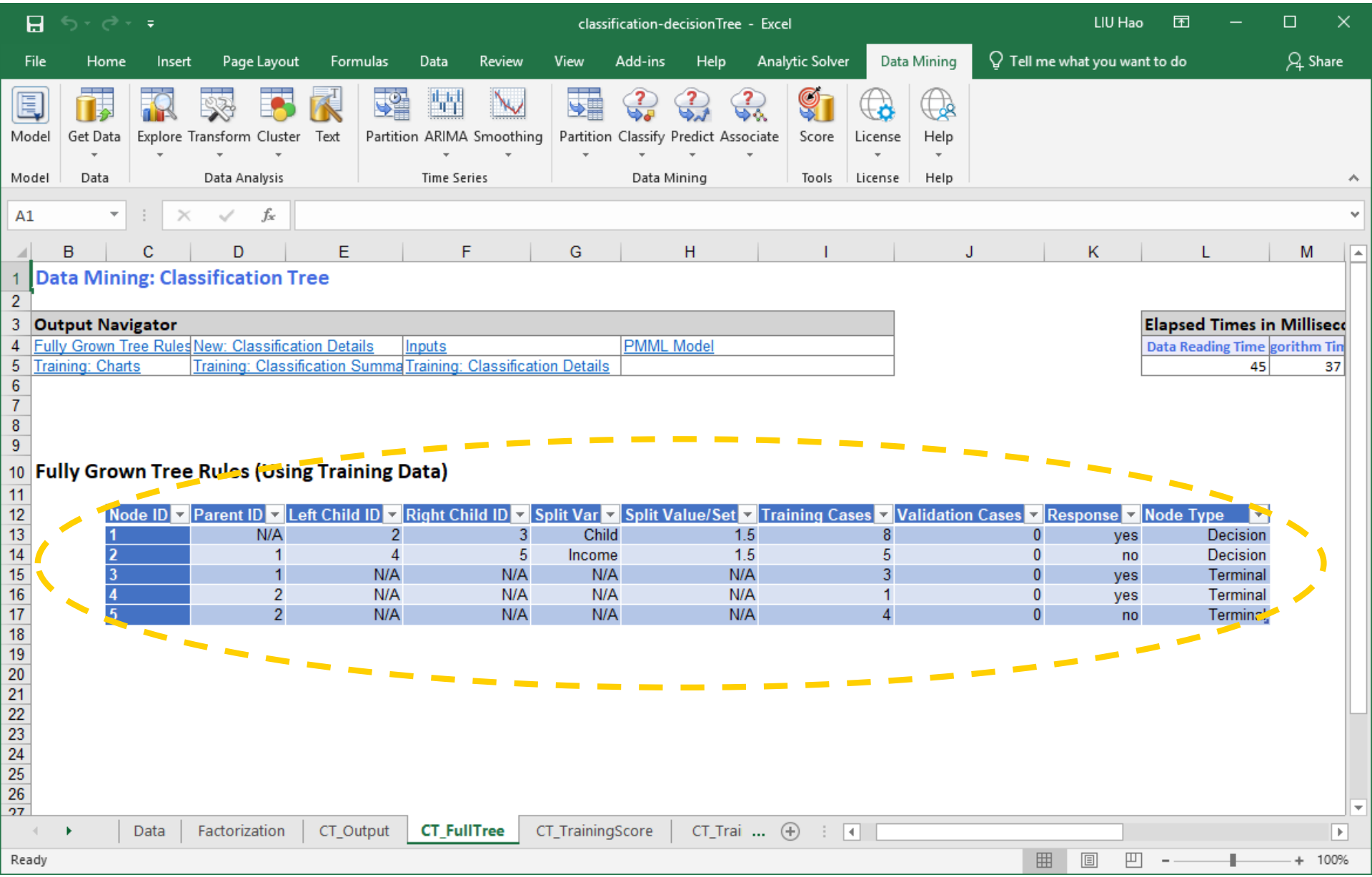

classification-decisionTree - Excel
LIU, Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
F58
Elapsed Times in Milliseconds
Data Reading Time gorithm Tin Report Time Total
1 5 5 11
New: Classification Detail Inputs PMML Model
Training: Classification Su Training: Classificatio
classification-decisionTree.xlsx
Factorization
$D$27:$G$35
8
3
Race Income Child
Insurance
arameters
FALSE
ng Parameters
Data Factorization CT_Output CT_FullTree CT_TrainingScore
Ready
100%

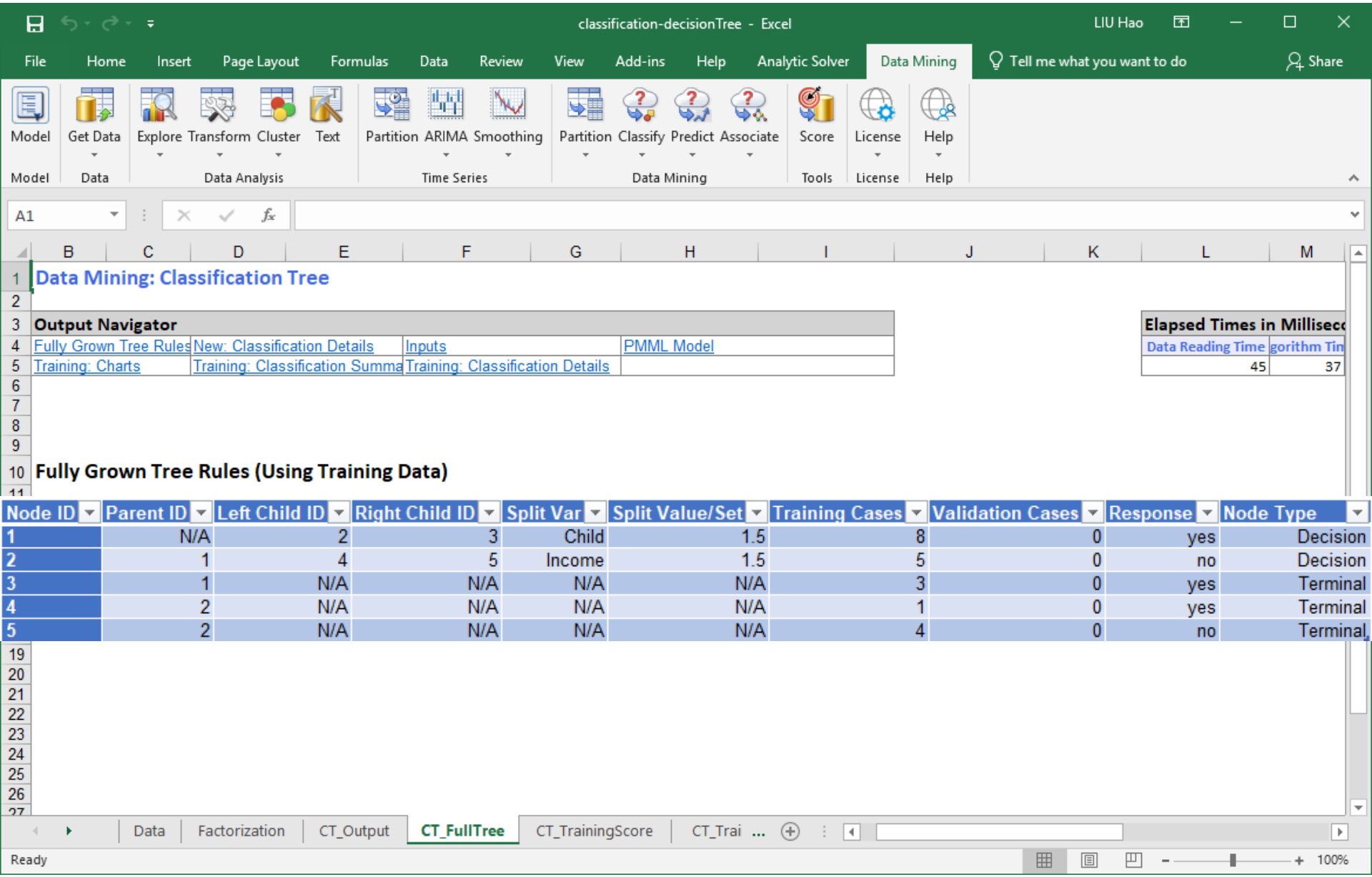

classification-decisionTree - Excel
LIU, Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
F58
New: Classification Detail
Inputs
PMML Model
Training: Classification Su
Training: Classificatio
Elapsed Times in Milliseconds
Data Reading Time
gorithm Tin
Report Time
Total
1
5
5
11
classification-decisionTree.xlsx
Factorization
$D$27:$G$35
8
3
Race
Income
Child
Insurance
arameters
FALSE
ng Parameters
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Ready
100%

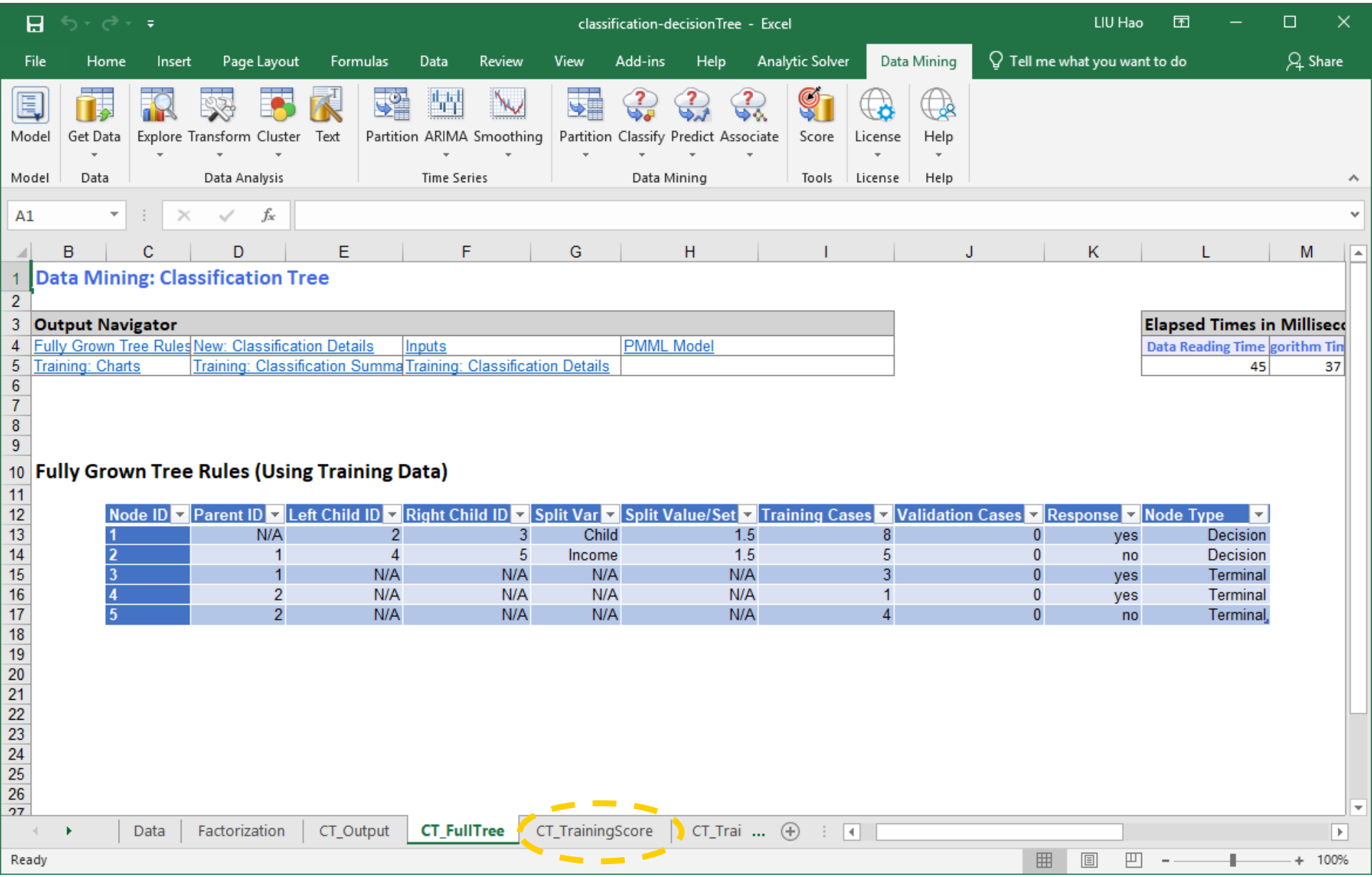

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Data
Full-Grown Tree
Nodes
Decision
Terminal
Collapsed
Links
x
# Records
Tree Info
Tree Height: 3
# Nodes: 5
Collapse All
Expand All
Child
1.5
5
3
Income
1.5
yes
1
4
yes
no
Response
yes
no
yes
yes
no
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Ready
100%

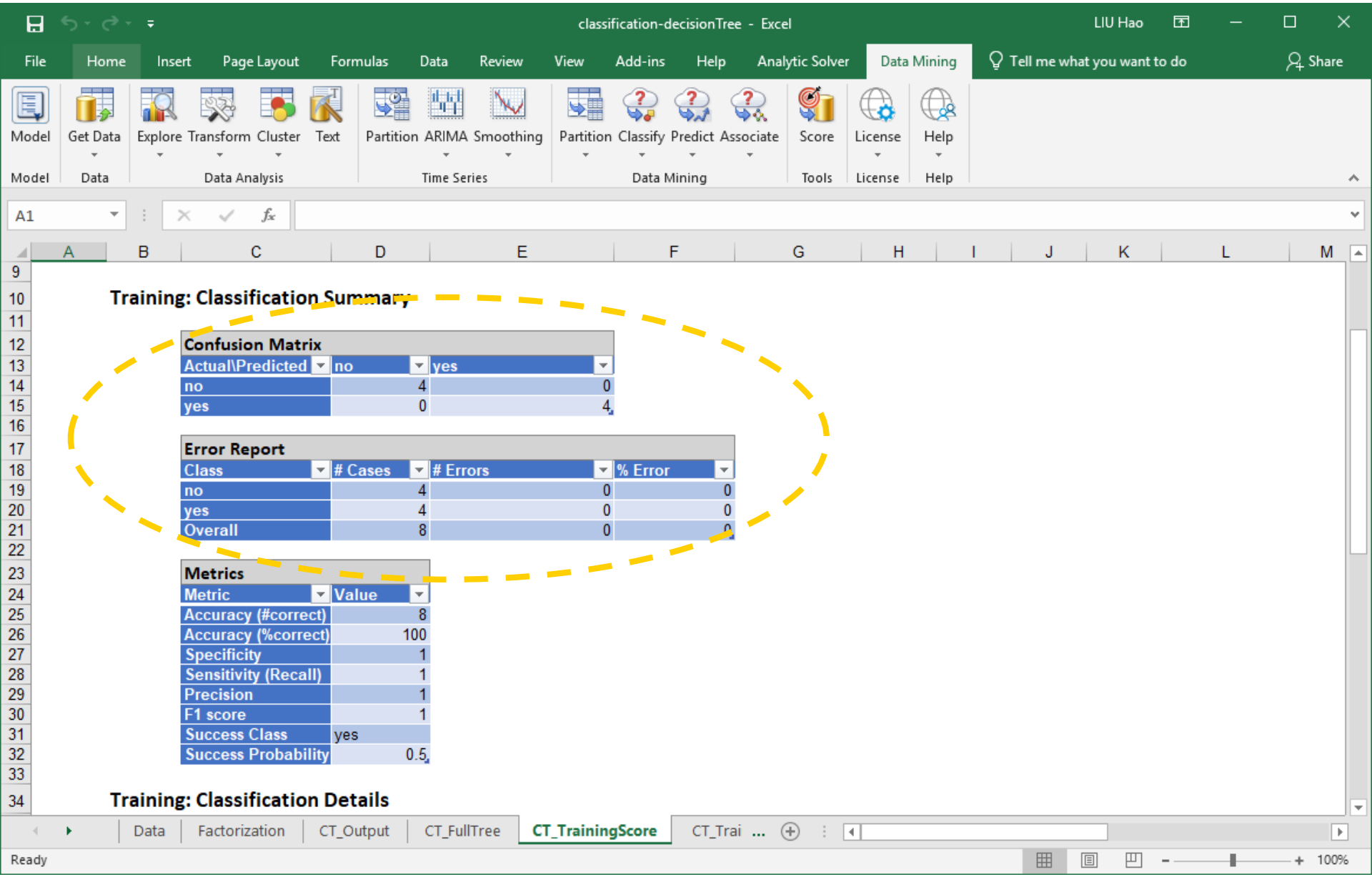

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Model
Data
Full-Grown Tree
Nodes
Terminal
Collapsed
Tree Info
Collapse All
Expand All
Child
1.5
Child_ord <=1.5
Child_ord >1.5
# of cases in this branch = 5
# of cases in this branch = 3
5
3
Income
1.5
yes
Income_ord <=1.5
Income_ord >1.5
1
4
# of cases in this branch = 1
# of cases in this branch = 4
yes
no
Response
Node
yes
no
yes
yes
no
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Ready
100%

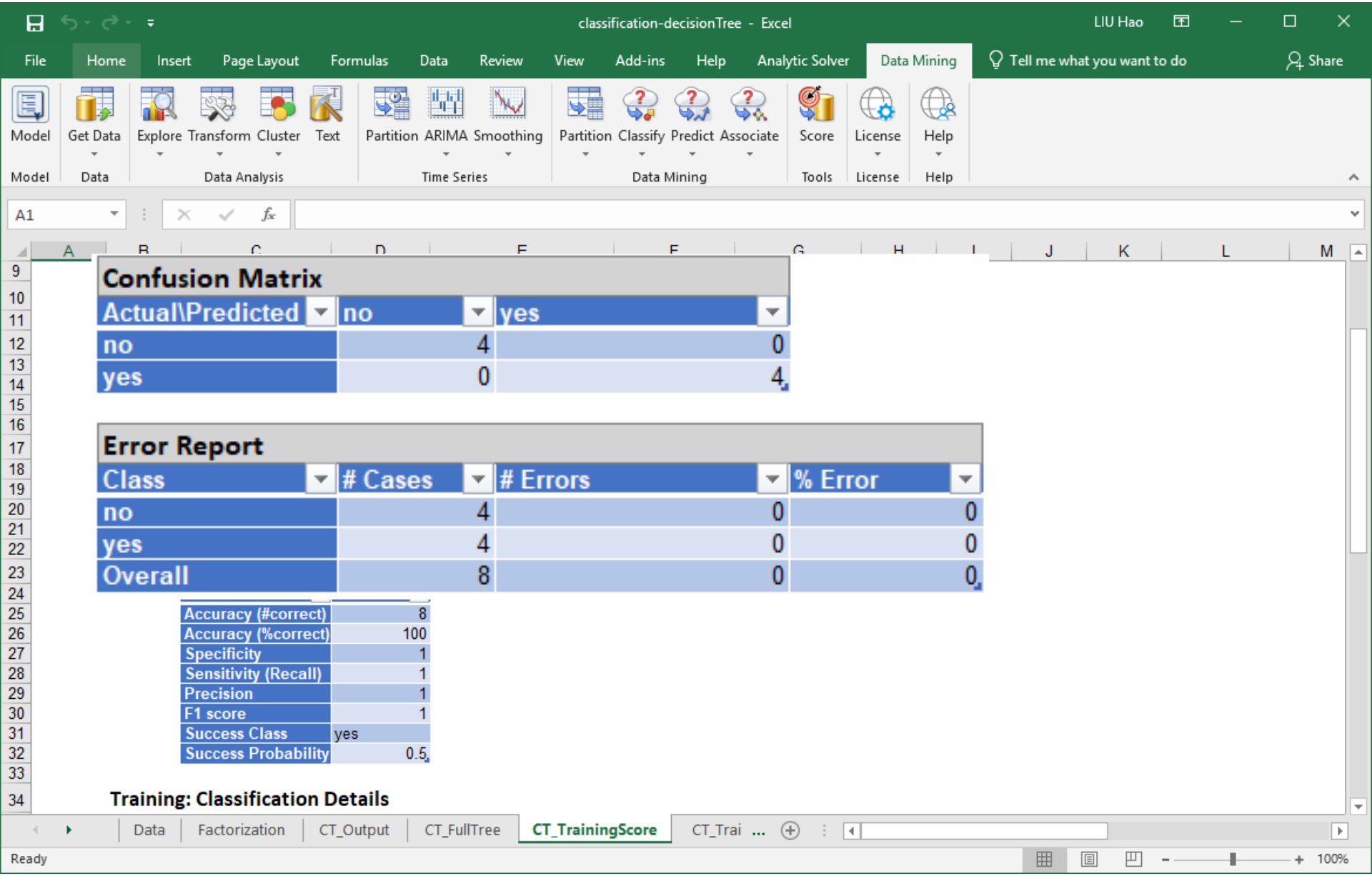

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me
Share
Model
Get Data
Explore
Model
Data
A1
Full-Grown Tree
Nodes
Decision
Terminal
Collapsed
Links
x
# Records
Tree Info
Tree Height: 3
# Nodes: 5
Collapse All
Expand All
Child
1.5
5
3
Income
1.5
yes
1
4
yes
no
Data M
Output
Fully Grow
Training: C
Fully Gr
Elaps
Data
ases
Response
Node
0
yes
0
no
0
yes
0
yes
0
no
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Ready
100%

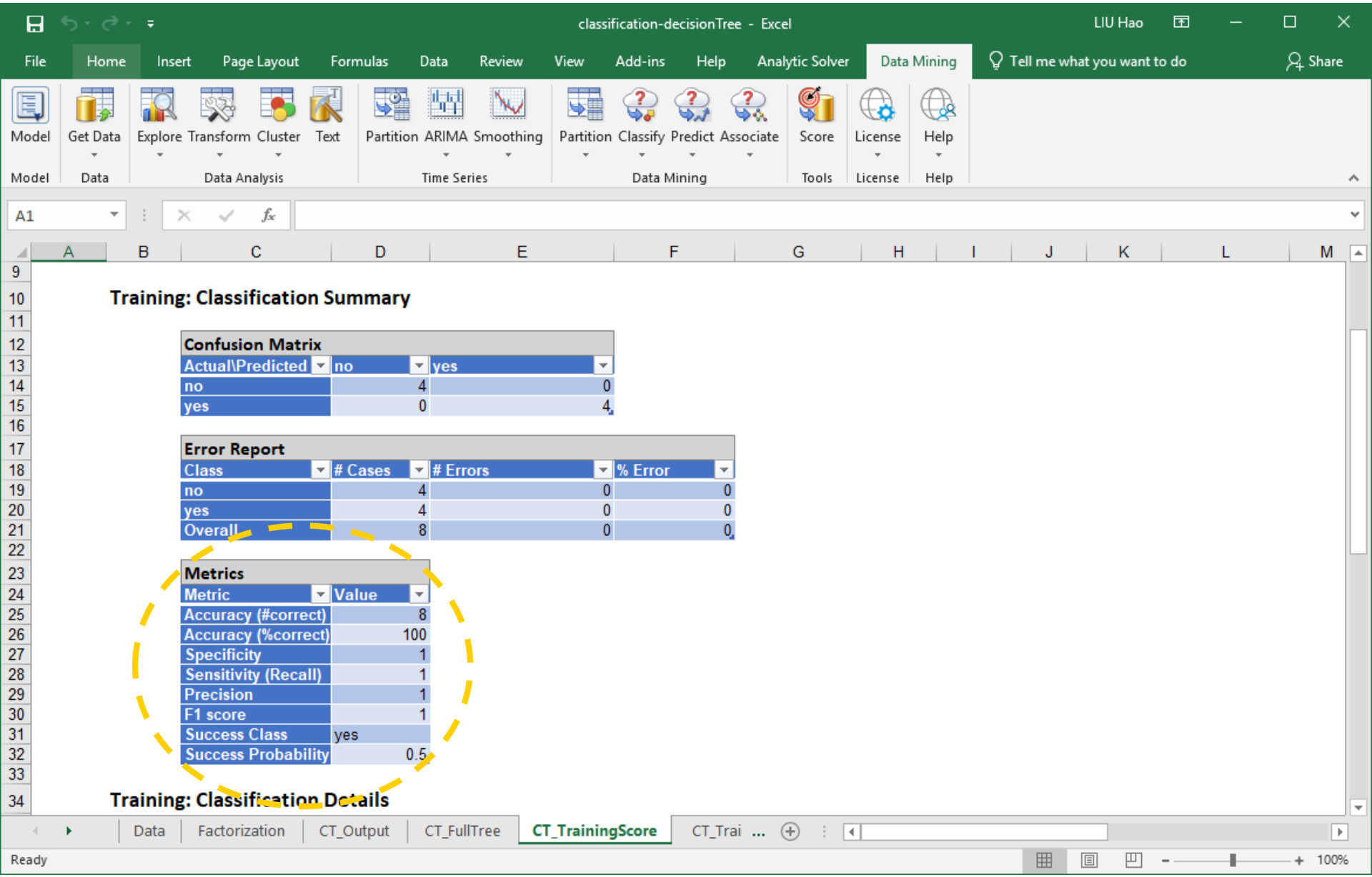

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Data Mining: Classification Tree
Output Navigator
Fully Grown Tree Rules
New: Classification Details
Inputs
PMML Model
Training: Charts
Training: Classification Summa
Training: Classification Details
Elapsed Times in Millisec
Data Reading Time
gorithm Tin
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_Trai ...
Ready
100%

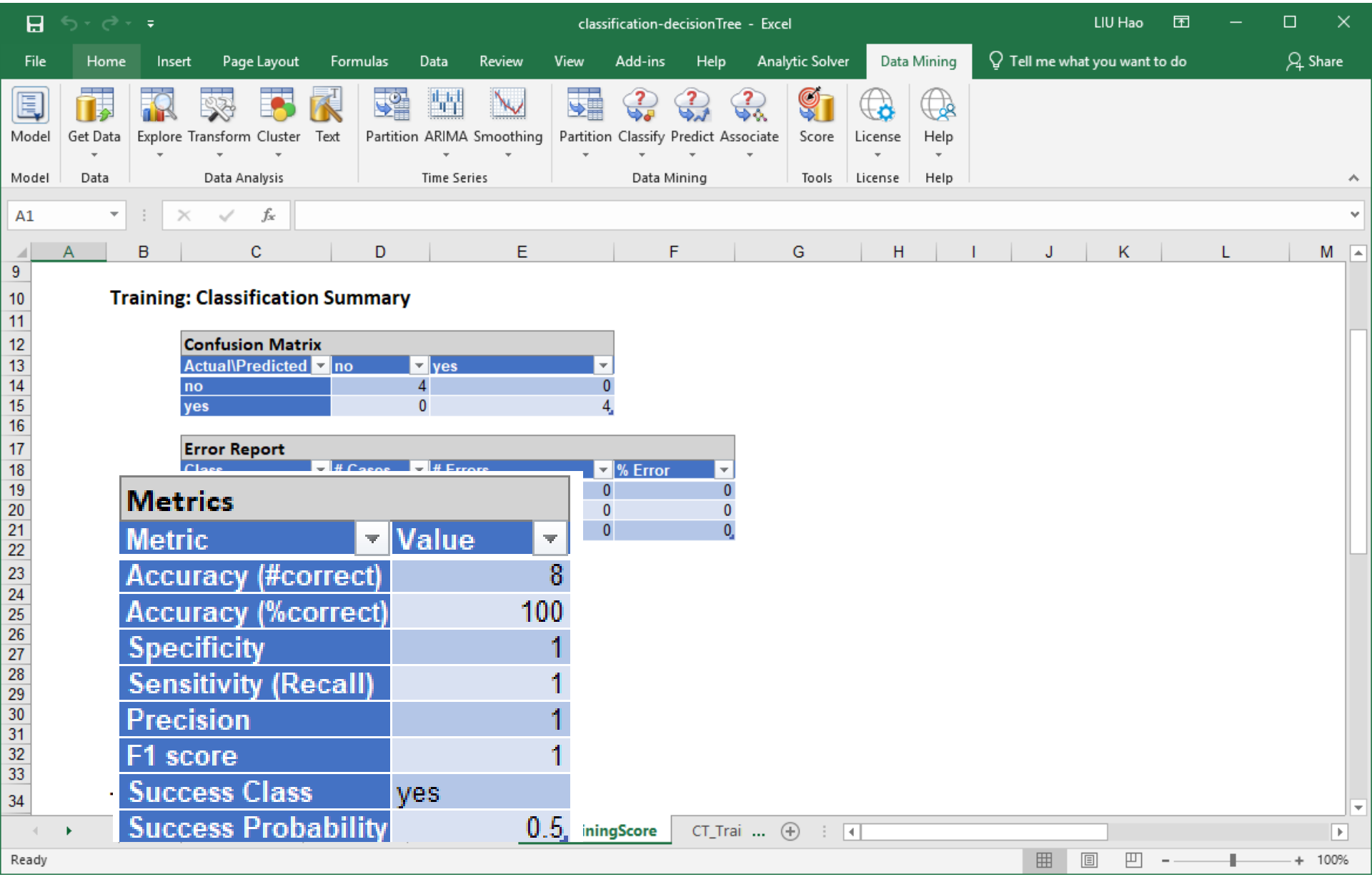

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Data Mining: Classification Tree
Output Navigator
Fully Grown Tree Rules
New: Classification Details
Inputs
PMML Model
Training: Charts
Training: Classification Summa
Training: Classification Details
Elapsed Times in Millisec
Data Reading Time
gorithm Tin
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_Trai ...
Ready
100%

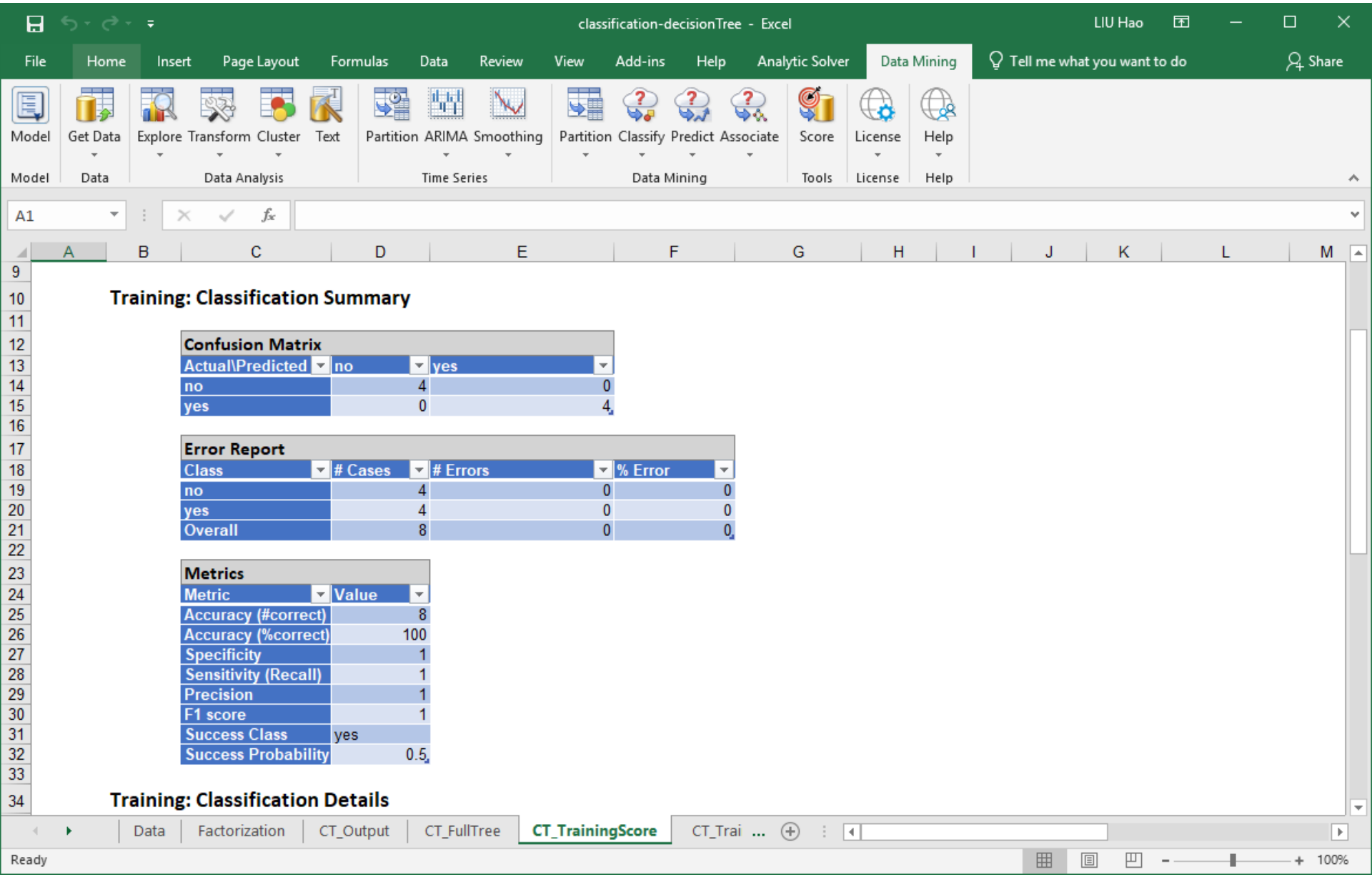

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Data Mining: Classification Tree
Output Navigator
Fully Grown Tree Rules
New: Classification Details
Inputs
PMML Model
Training: Charts
Training: Classification Summa
Training: Classification Details
Elapsed Times in Millisec
Data Reading Time
gorithm Tin
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_Trai ...
Ready
100%

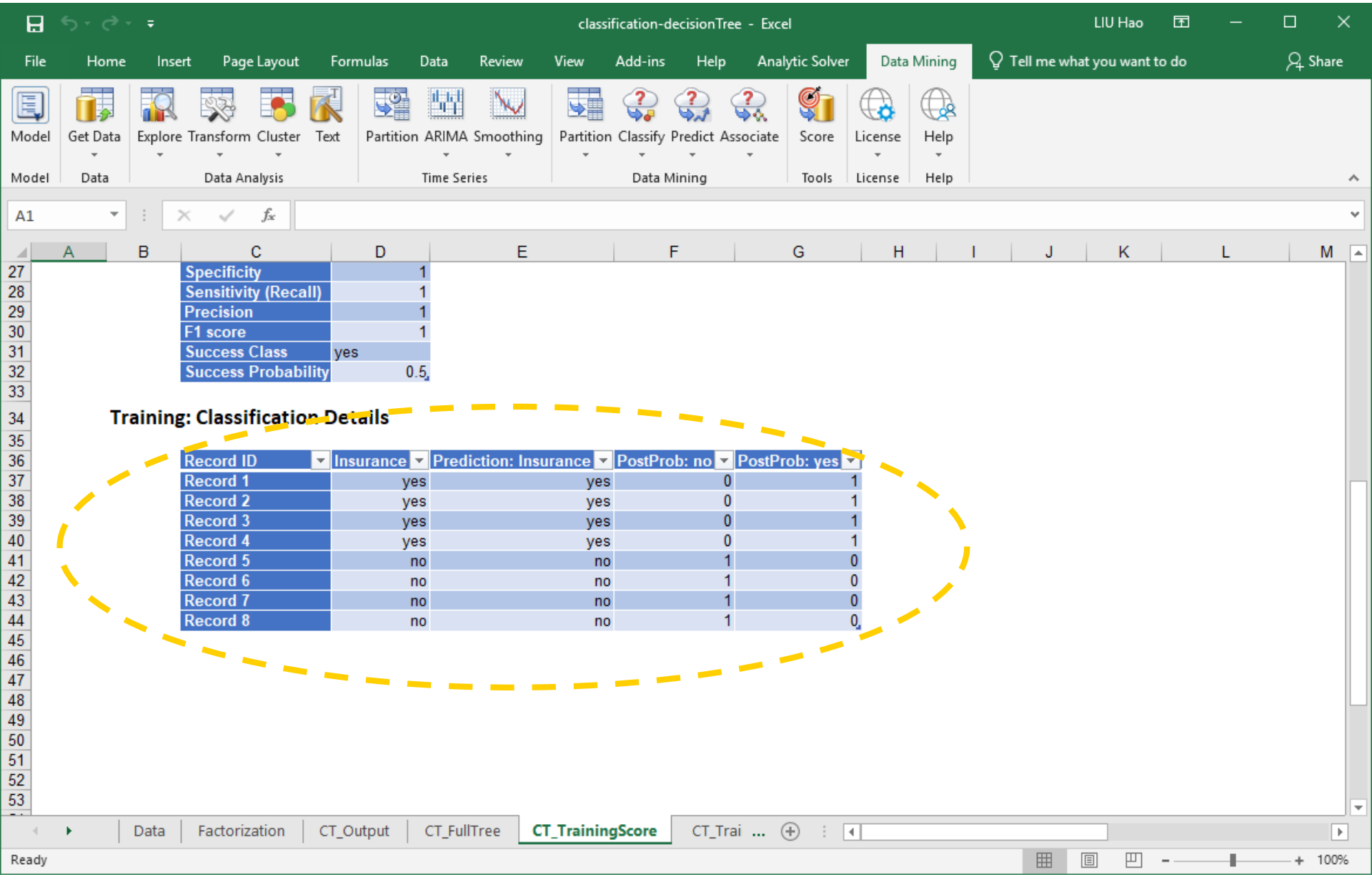

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
Training: Classification Summary
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
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
Training: Classification Details
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
CT_Trai ...
Ready
100%

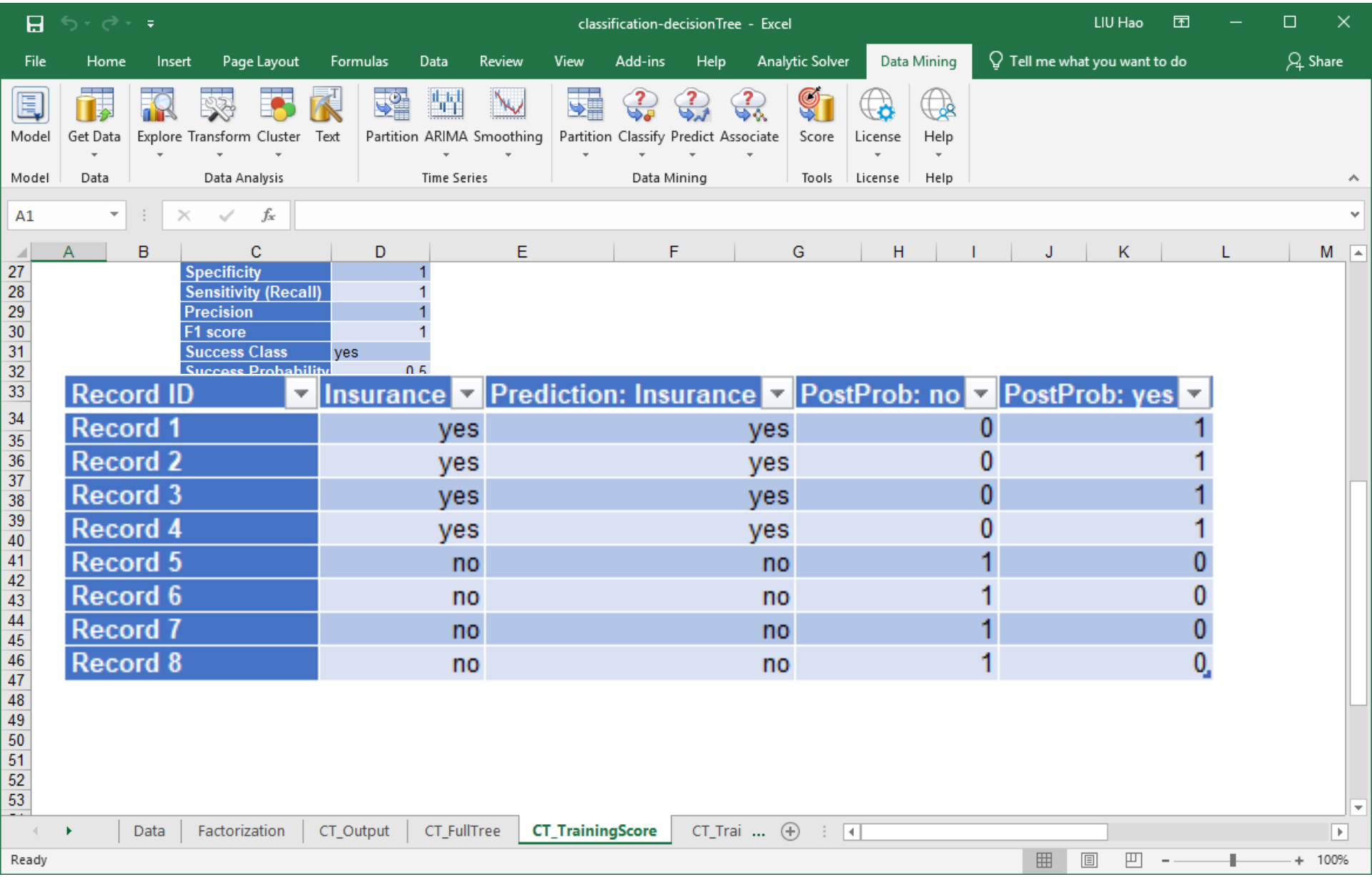

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
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
100%

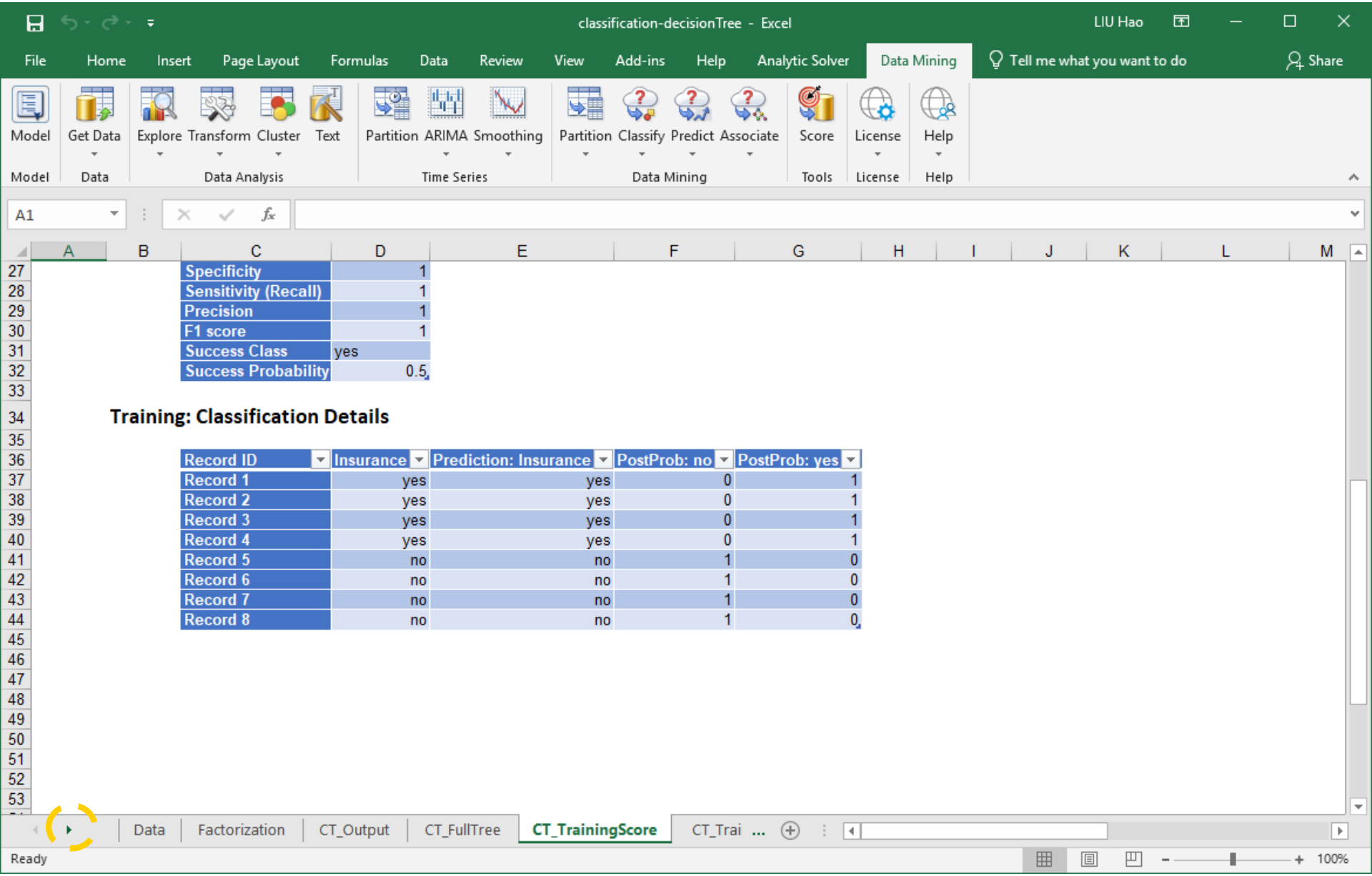

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
Training: Classification Summary
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
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
Training: Classification Details
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
CT_Trai ...
Ready
100%

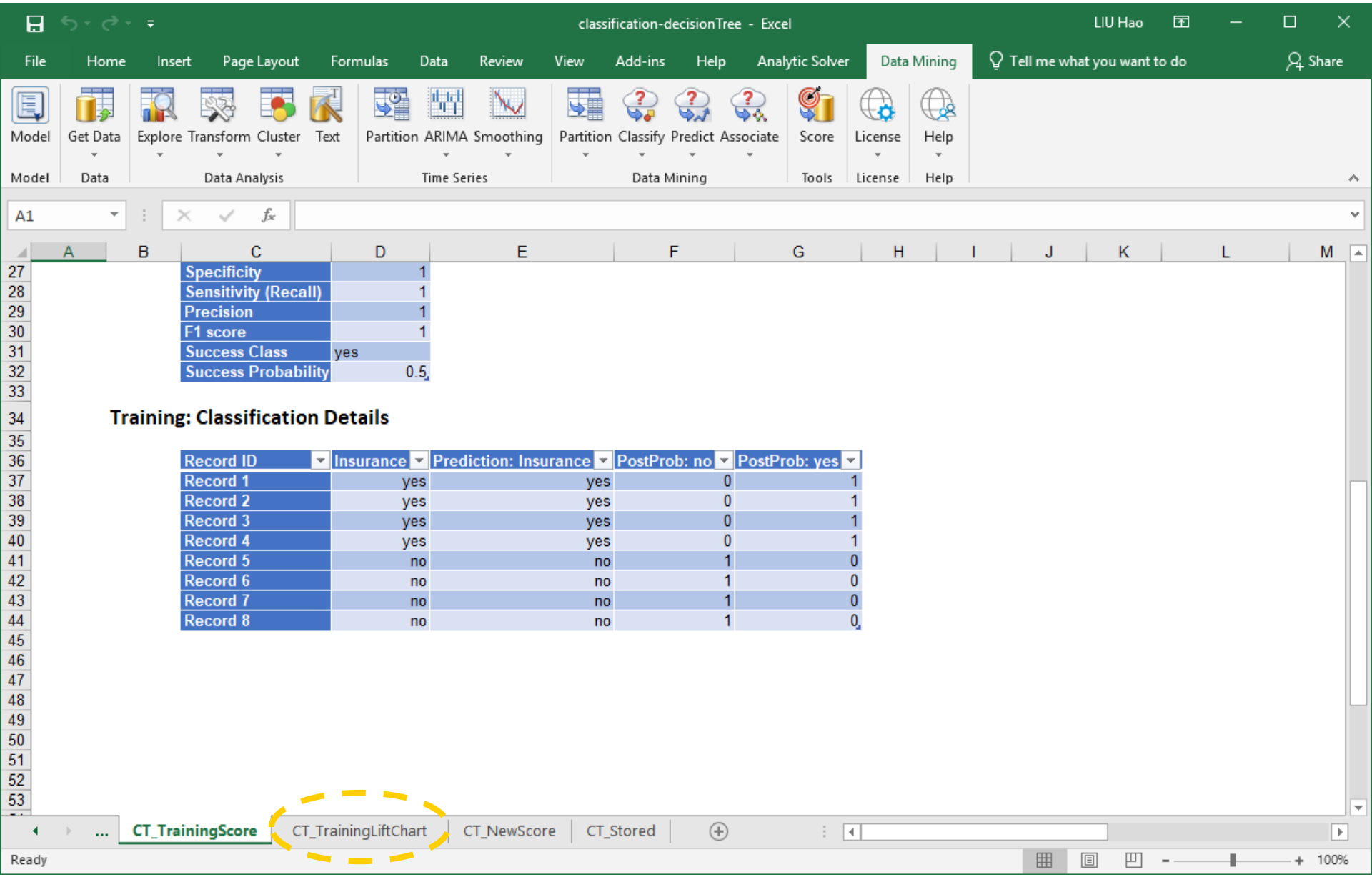

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
Training: Classification Summary
Confusion Matrix
Actual\Predicted
no
yes
no
4
0
yes
0
4
Error Report
% Error
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
iningScore
CT_Trai ...
Ready
100%

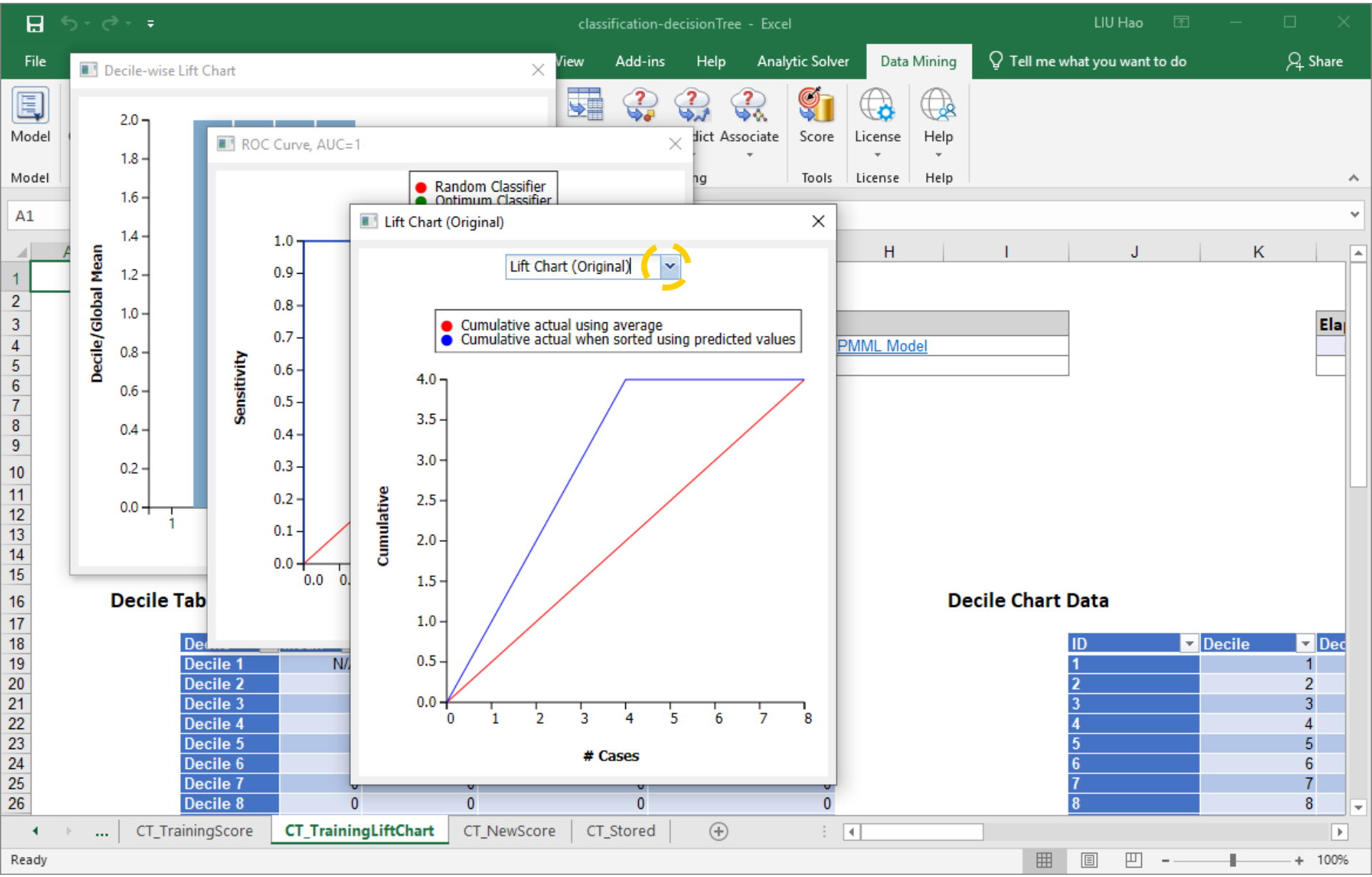

classification-decisionTree - Excel
LIU Hao
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Get Data
Explore
Transform
Cluster
Text
Partition
ARIMA
Smoothing
Partition
Classify
Predict
Associate
Score
License
Help
Model
Data
Data Analysis
Time Series
Data Mining
Tools
License
Help
A1
Training: Classification Summary
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
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
Training: Classification Details
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
CT_Trai ...
Ready
100%

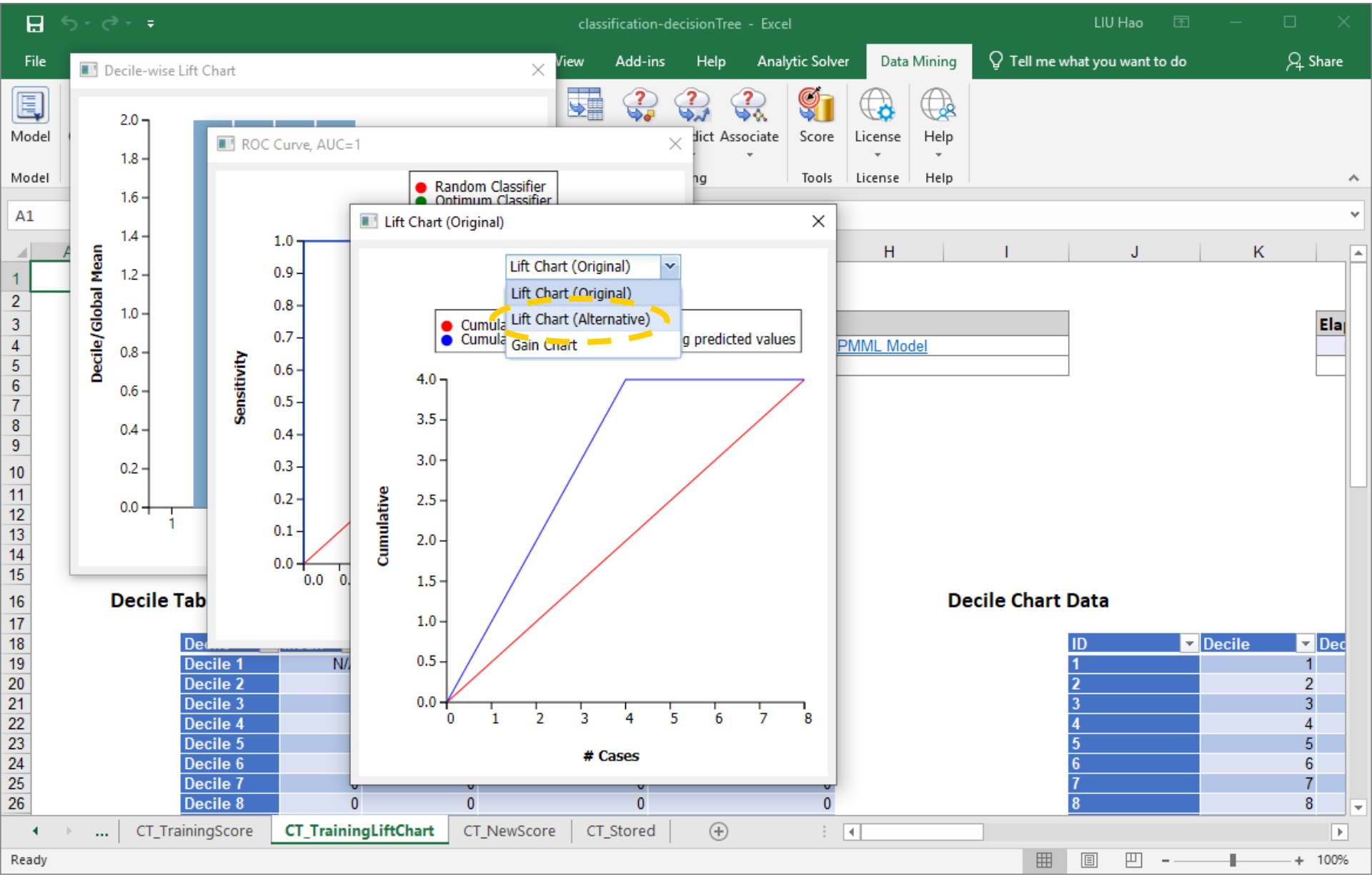

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Specificity 1
Sensitivity (Recall) 1
Precision 1
F1 score 1
Success Class yes
Success Probability 0.5
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%

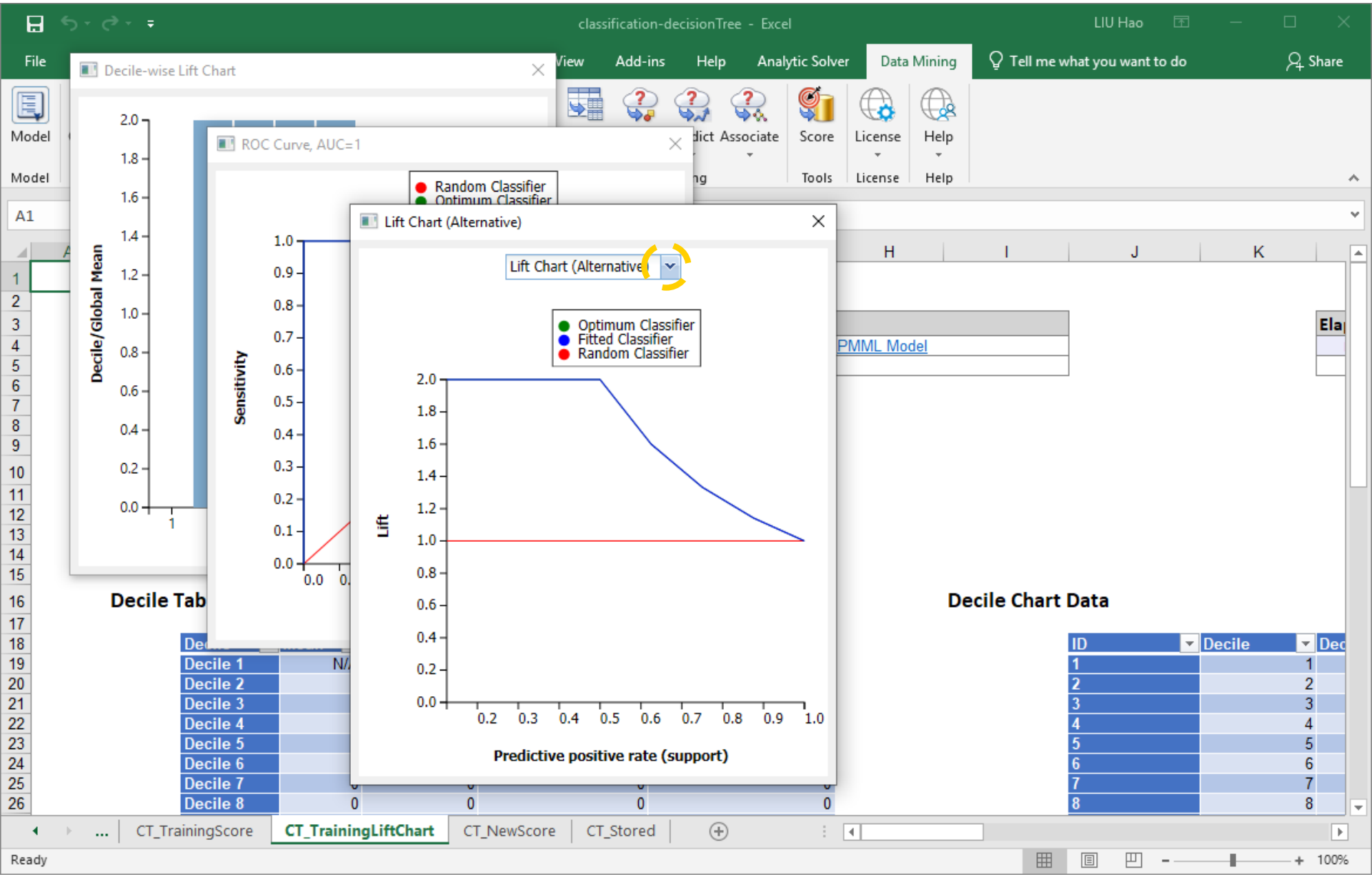

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Specificity 1
Sensitivity (Recall) 1
Precision 1
F1 score 1
Success Class yes
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%

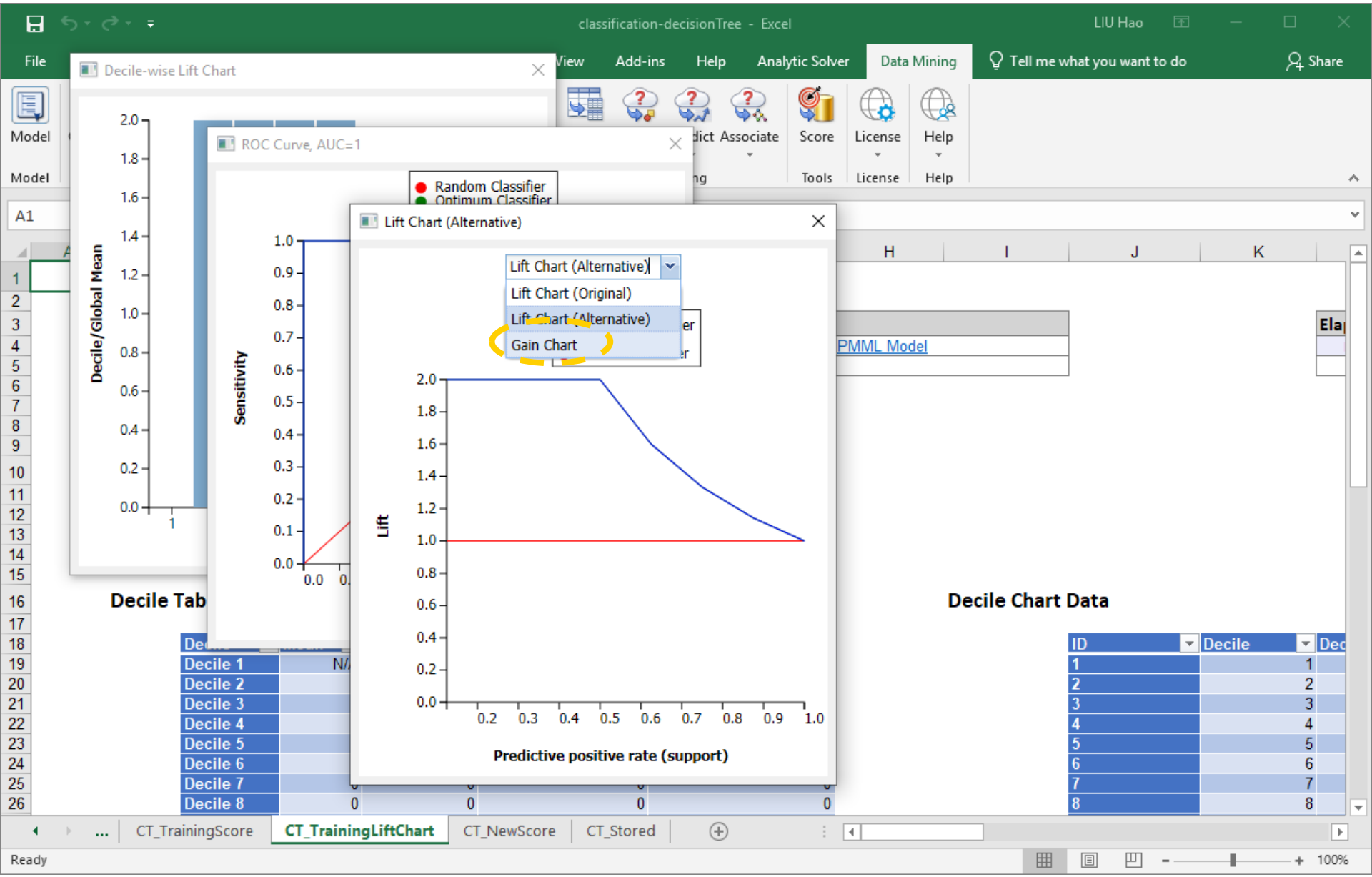

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Specificity 1
Sensitivity (Recall) 1
Precision 1
F1 score 1
Success Class yes
Success Probability 0.5
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%

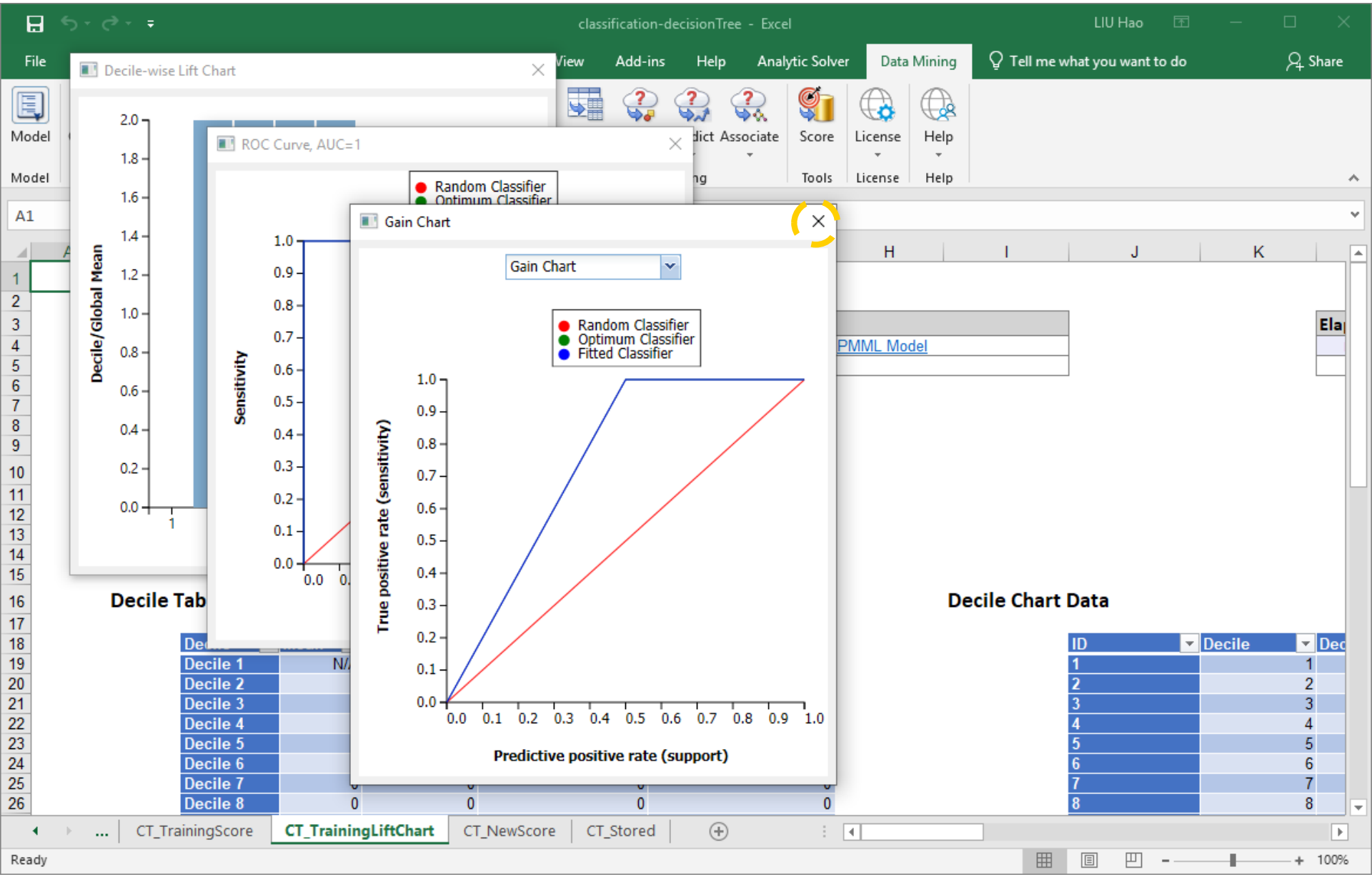

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Specificity 1
Sensitivity (Recall) 1
Precision 1
F1 score 1
Success Class yes
Success Probability 0.5
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%

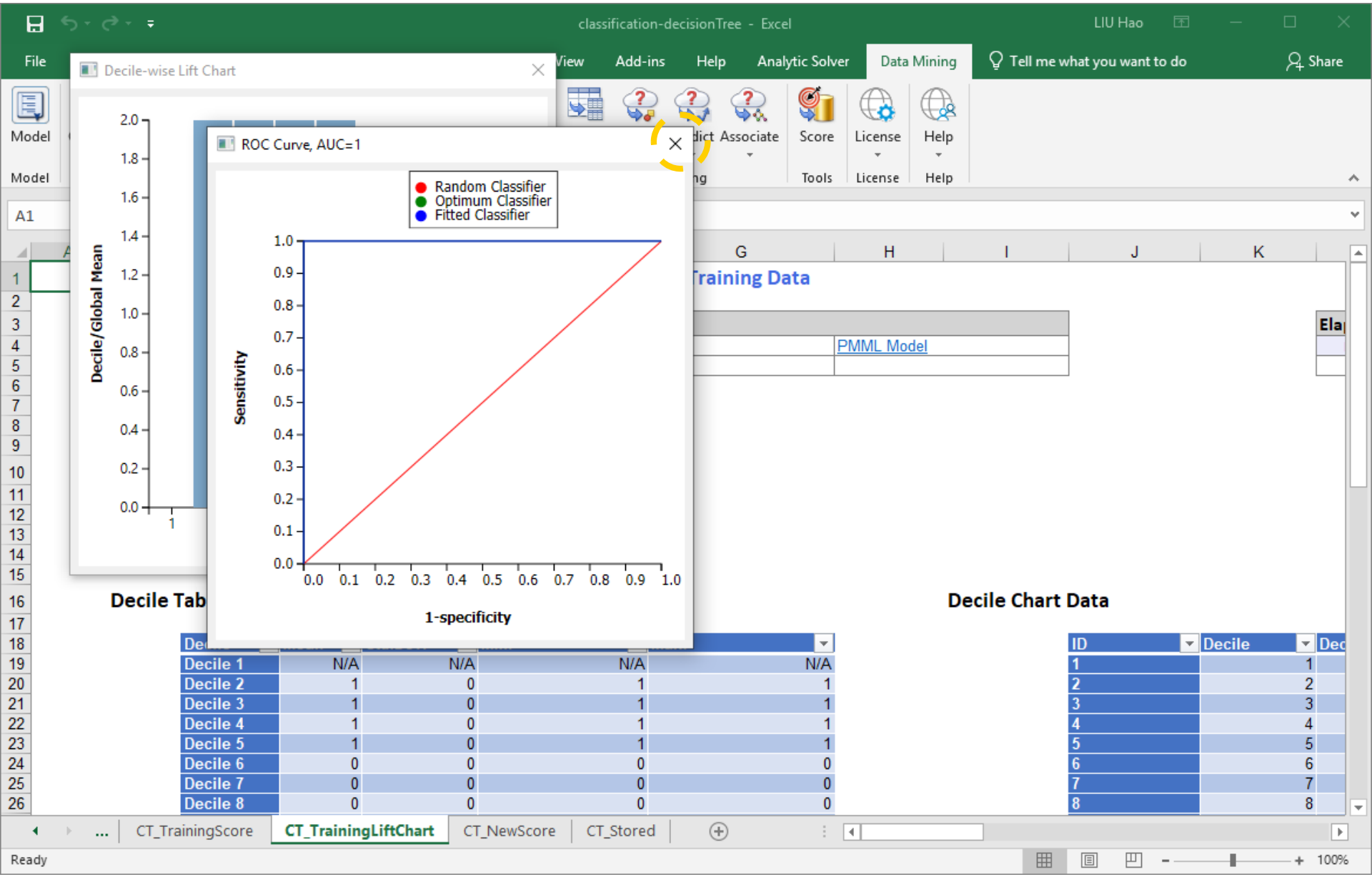

classification-decisionTree - Excel
LIU Hao
File
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Associate
Score
License
Help
Tools
Decile-wise Lift Chart
Decile/Global Mean
ROC Curve, AUC=1
Random Classifier
Optimum Classifier
Sensitivity
Lift Chart (Original)
Lift Chart (Original)
Cumulative actual using average
Cumulative actual when sorted using predicted values
Cumulative
# Cases
PMML Model
Decile Tab
Decile Chart Data
Decile 1
Decile 2
Decile 3
Decile 4
Decile 5
Decile 6
Decile 7
Decile 8
ID
Decile
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Ready
100%

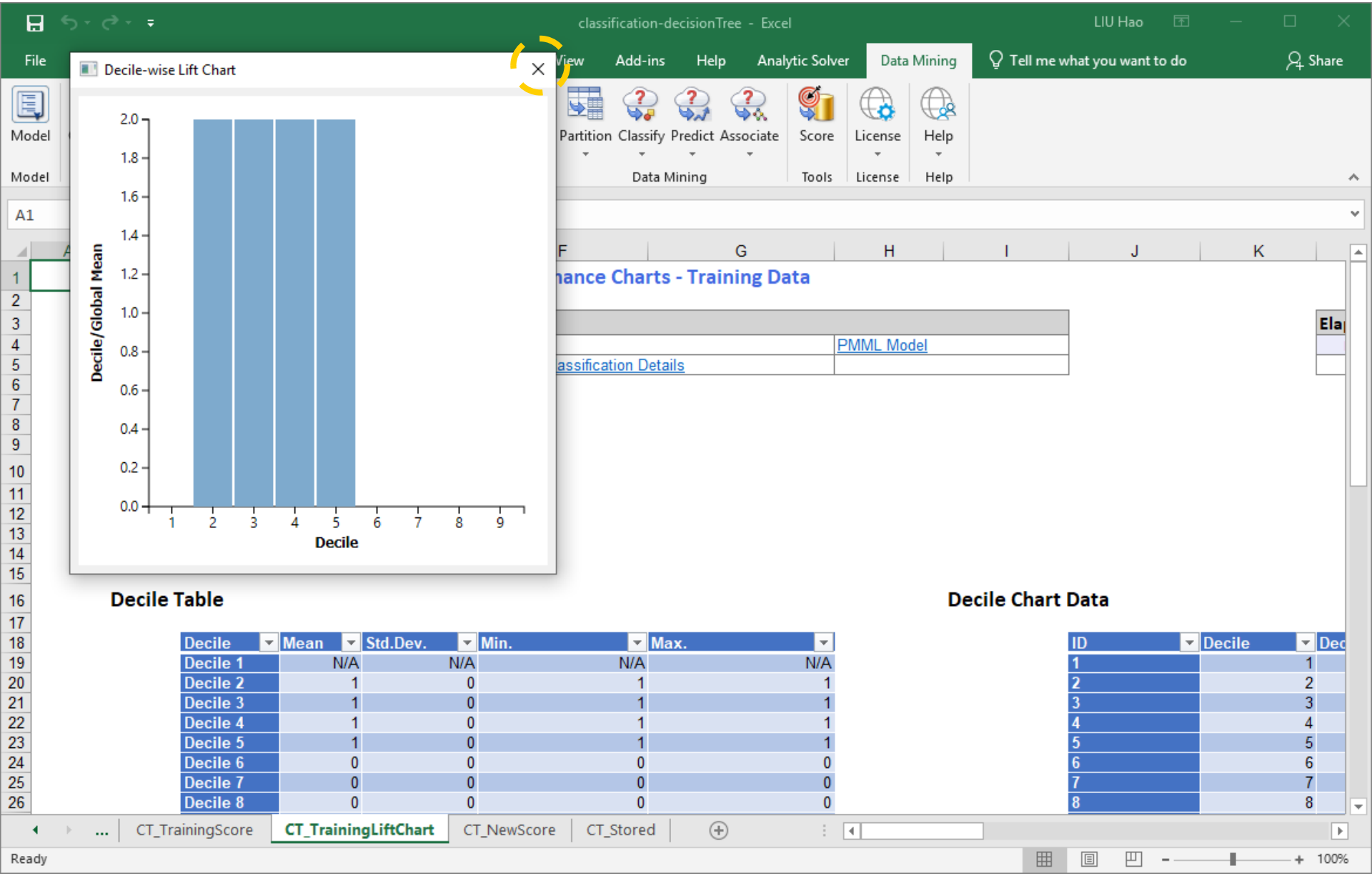

classification-decisionTree - Excel
LIU Hao
File
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Associate
Score
License
Help
Tools
Decile-wise Lift Chart
Decile/Global Mean
ROC Curve, AUC=1
Random Classifier
Optimum Classifier
Sensitivity
Lift Chart (Original)
Lift Chart (Original)
Lift Chart (Alternative)
Gain Chart
Cumulative
# Cases
PMML Model
Decile Tab
Decile Chart Data
Decile 1
Decile 2
Decile 3
Decile 4
Decile 5
Decile 6
Decile 7
Decile 8
ID
Decile
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Ready
100%

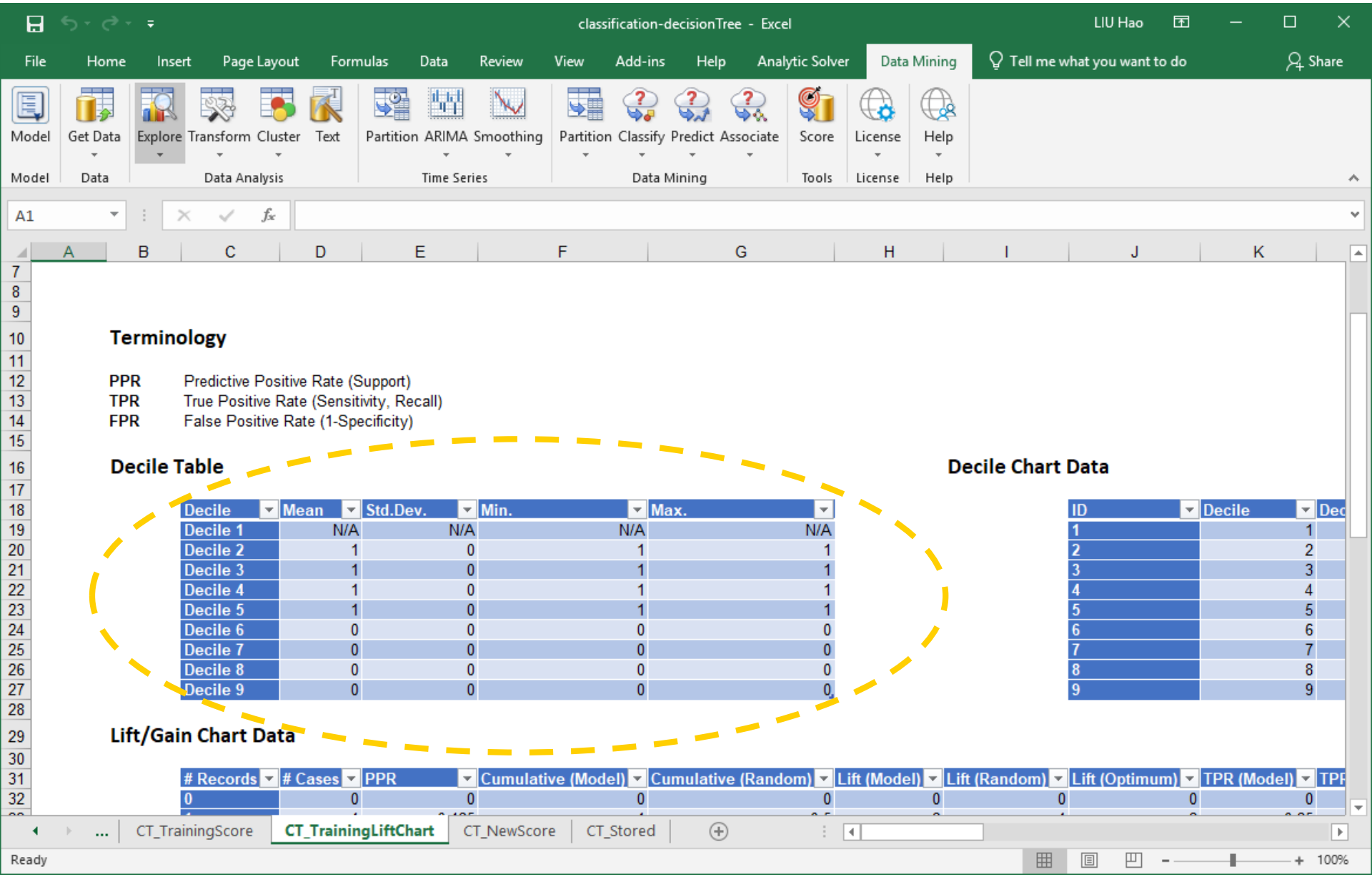

classification-decisionTree - Excel
LIU Hao
File
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Associate
Score
License
Help
Tools
Decile-wise Lift Chart
Decile/Global Mean
ROC Curve, AUC=1
Random Classifier
Optimum Classifier
Sensitivity
Lift Chart (Alternative)
Lift Chart (Alternative)
Optimum Classifier
Fitted Classifier
Random Classifier
Lift
Predictive positive rate (support)
PMML Model
Decile Tab
Decile Chart Data
Decile 1
Decile 2
Decile 3
Decile 4
Decile 5
Decile 6
Decile 7
Decile 8
ID
Decile
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Ready
100%

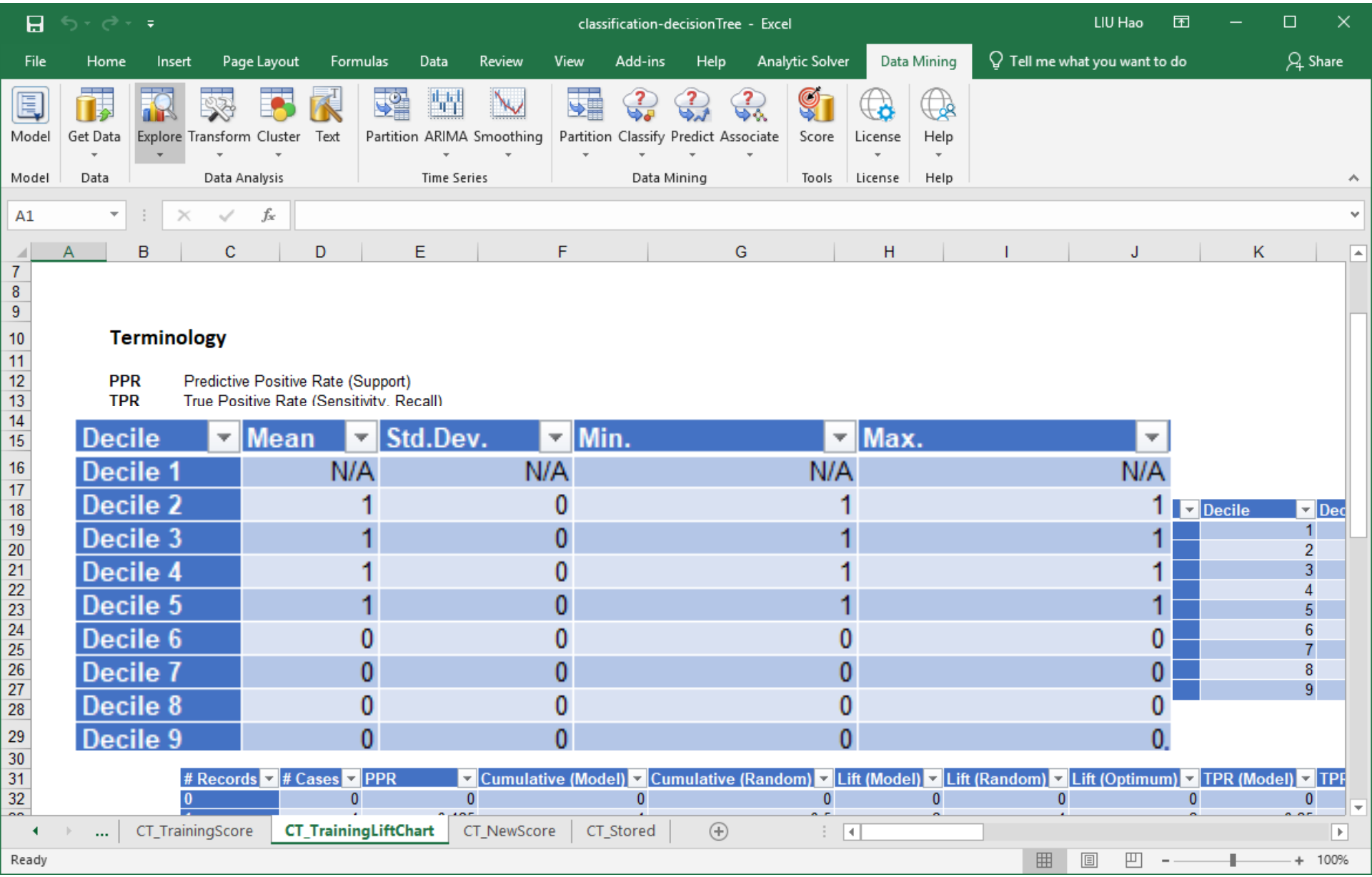

classification-decisionTree - Excel
LIU Hao
File
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Associate
Score
License
Help
Tools
A1
Decile-wise Lift Chart
Decile/Global Mean
ROC Curve, AUC=1
Random Classifier
Optimum Classifier
Sensitivity
Lift Chart (Alternative)
Lift Chart (Alternative)
Lift Chart (Original)
Lift Chart (Alternative)
Gain Chart
Lift
Predictive positive rate (support)
PMML Model
Decile Tab
Decile Chart Data
Decile 1
Decile 2
Decile 3
Decile 4
Decile 5
Decile 6
Decile 7
Decile 8
ID
Decile
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Ready
100%

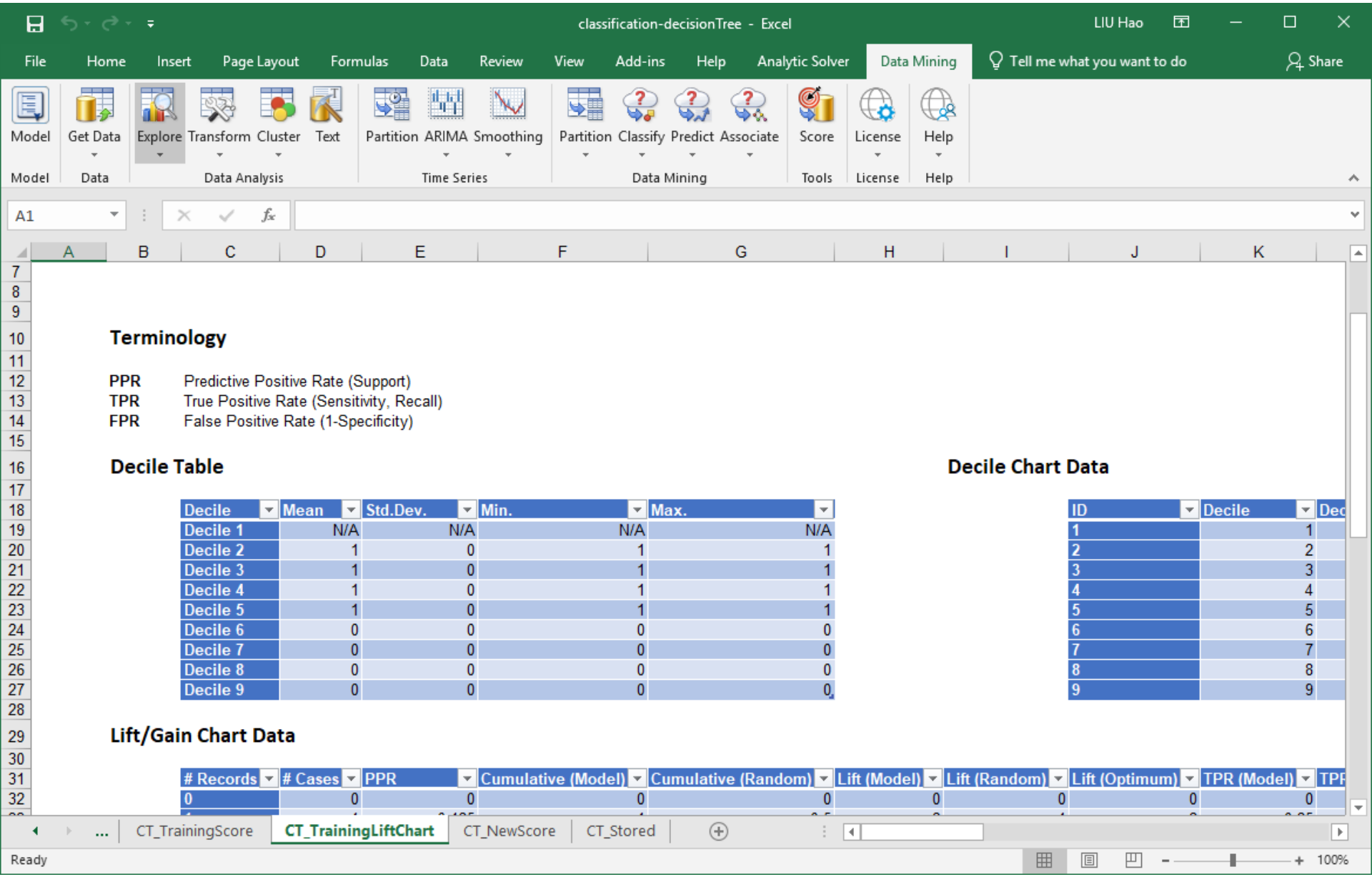

classification-decisionTree - Excel
LIU Hao
File
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Associate
Score
License
Help
Tools
Decile-wise Lift Chart
Decile/Global Mean
ROC Curve, AUC=1
Random Classifier
Optimum Classifier
Sensitivity
Gain Chart
Gain Chart
Random Classifier
Optimum Classifier
Fitted Classifier
True positive rate (sensitivity)
Predictive positive rate (support)
PMML Model
Decile Tab
Decile Chart Data
Decile 1
Decile 2
Decile 3
Decile 4
Decile 5
Decile 6
Decile 7
Decile 8
ID
Decile
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Ready
100%

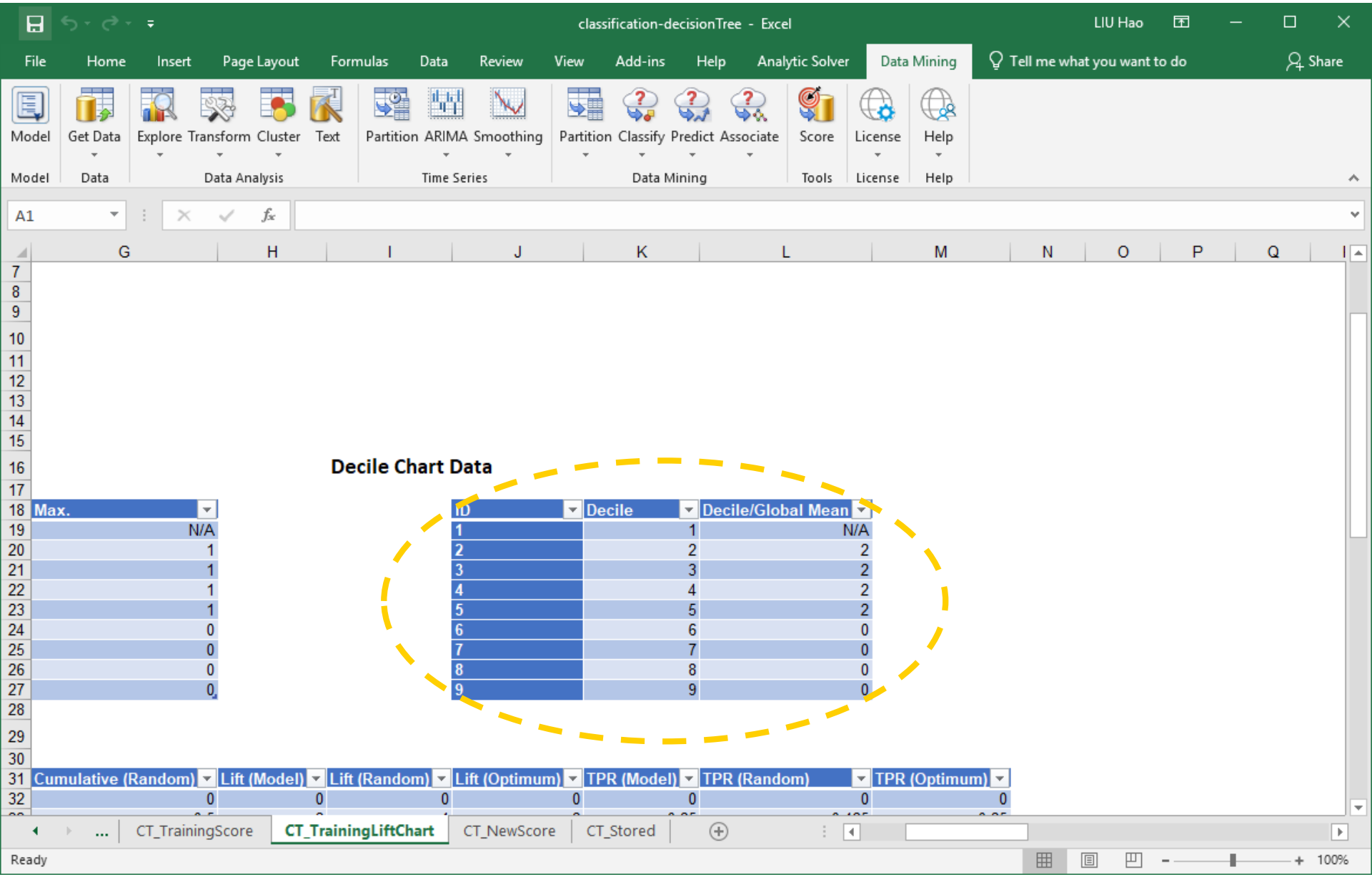

classification-decisionTree - Excel
LIU Hao
File
View
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Associate
Score
License
Help
Tools
License
Help
A1
Decile-wise Lift Chart
Decile/Global Mean
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
ROC Curve, AUC=1
Random Classifier
Optimum Classifier
Fitted Classifier
Sensitivity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
1-specificity
Training Data
PMML Model
Decile Tab
Decile Chart Data
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
ID Decile
1 1
2 2
3 3
4 4
5 5
6 6
7 7
8 8
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Ready
100%

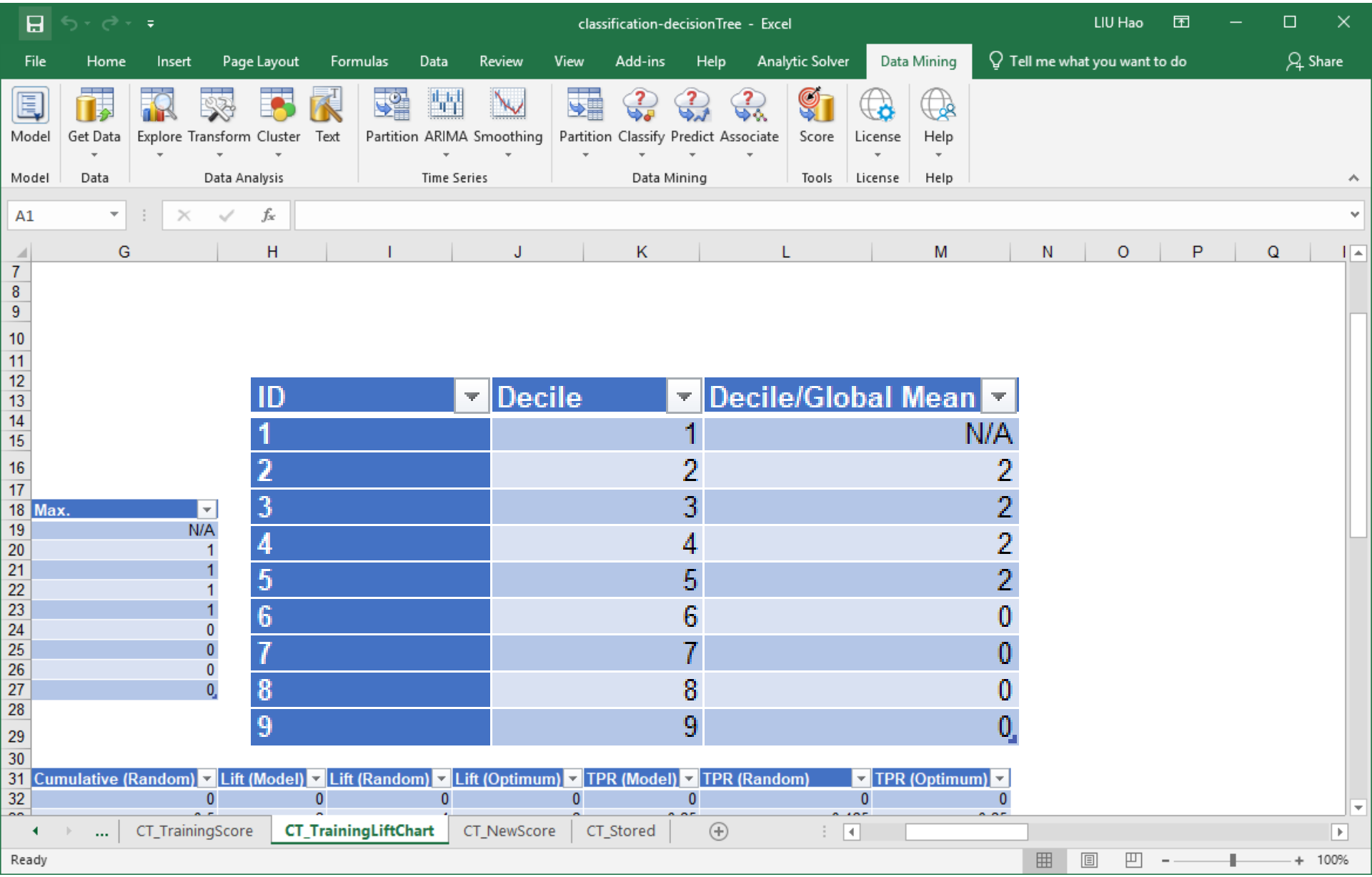

classification-decisionTree - Excel
LIU Hao
File
Add-ins
Help
Analytic Solver
Data Mining
Tell me what you want to do
Share
Model
Partition
Classify
Predict
Associate
Score
License
Help
Data Mining
Tools
License
Help
A1
Decile-wise Lift Chart
Decile/Global Mean
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
Decile
nance Charts - Training Data
PMML Model
assification Details
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 Chart Data
ID
Decile
1 1
2 2
3 3
4 4
5 5
6 6
7 7
8 8
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Ready
100%

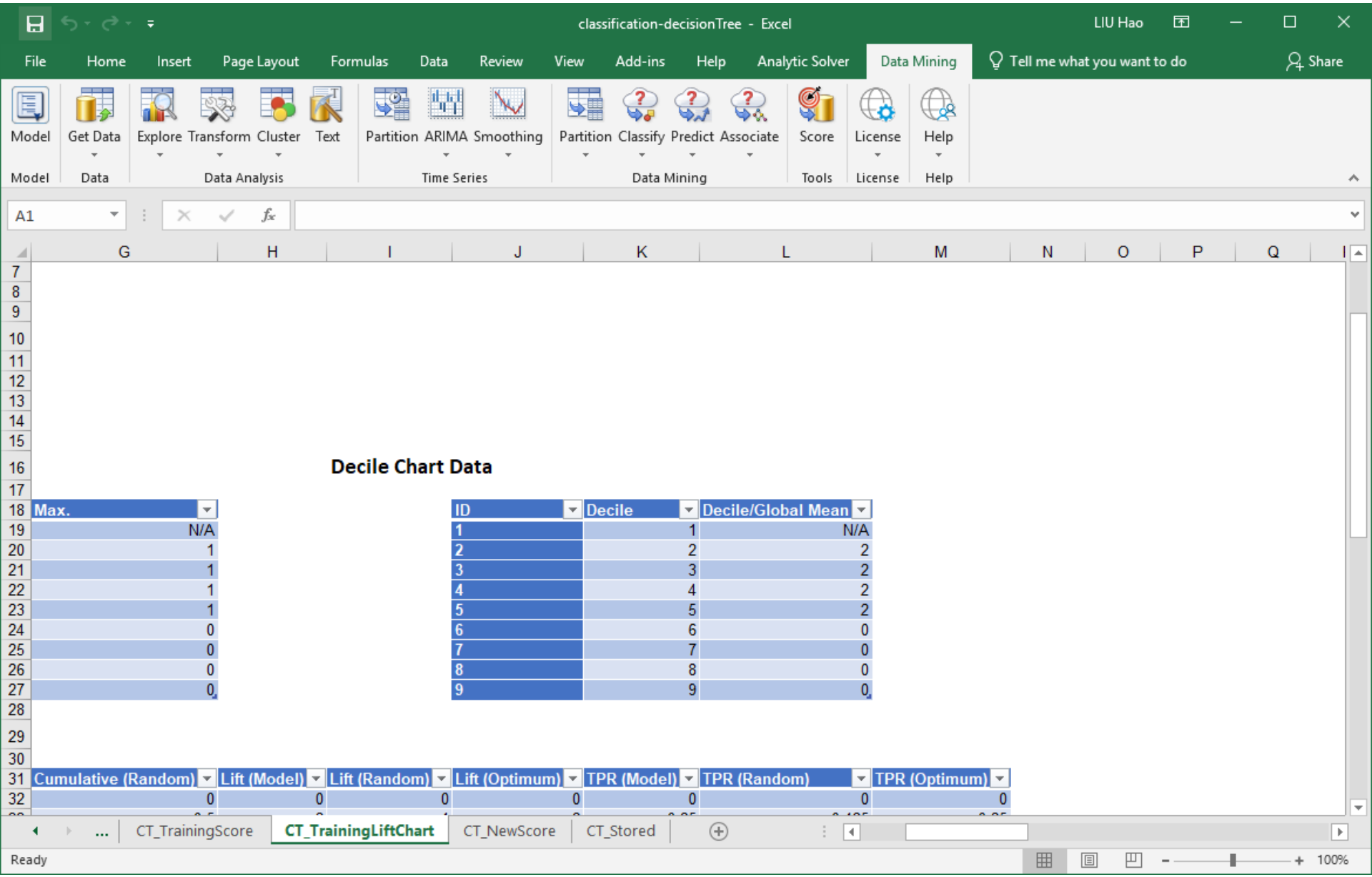

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining
Tell me what you want to do
Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Terminology
PPR Predictive Positive Rate (Support)
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
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)
0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0
CT_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored
Ready
100%

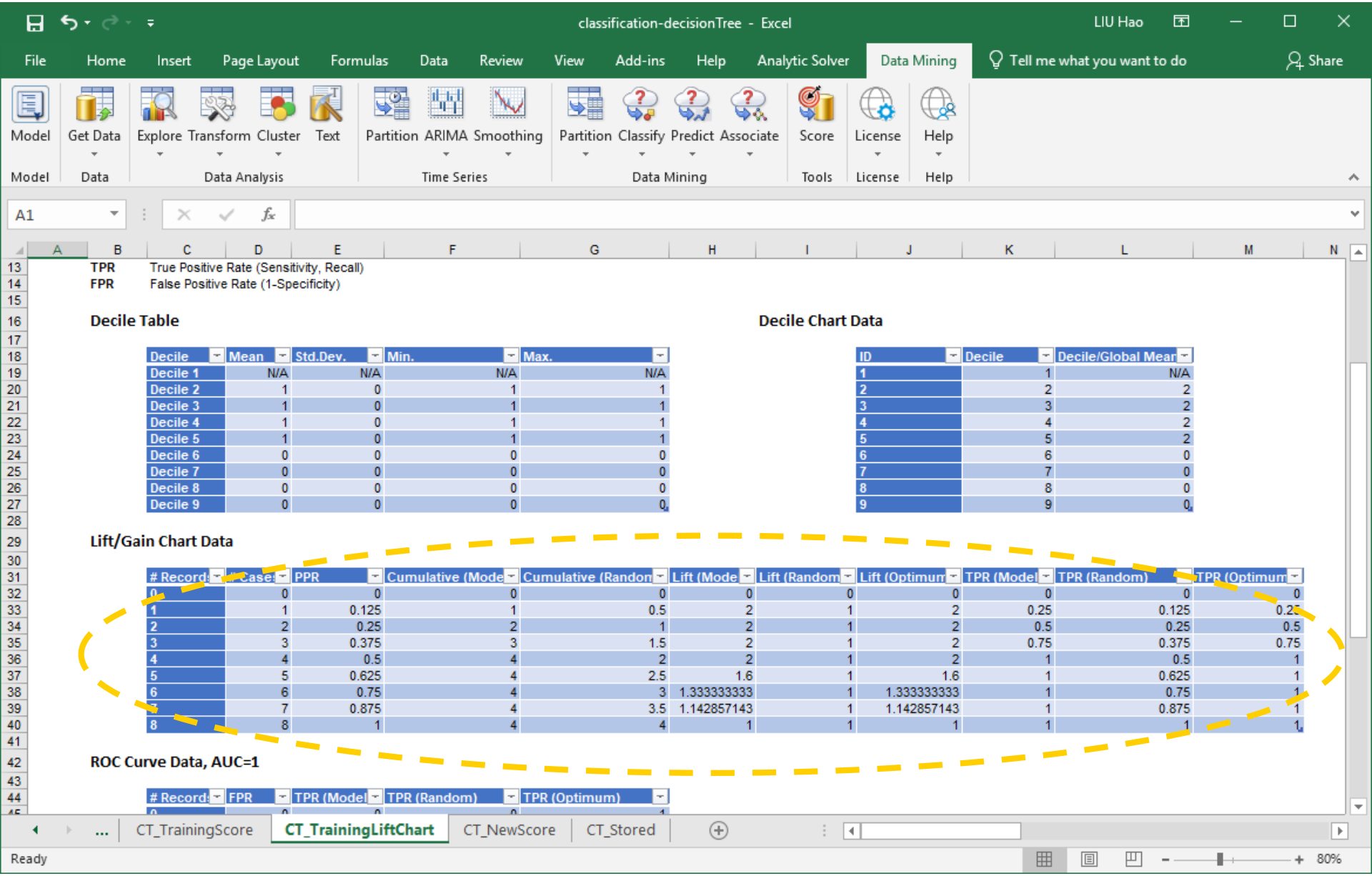

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Terminology
PPR Predictive Positive Rate (Support)
TPR True Positive Rate (Sensitivity, Recall)
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
1
2
3
4
5
6
7
8
9
# Records # Cases PPR Cumulative (Model) Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model)
0 0 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
Terminology
PPR Predictive Positive Rate (Support)
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
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)
0 | 0 | 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
Decile Chart Data
Max.
N/A
1
1
1
1
0
0
0
0
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
Cumulative (Random)
Lift (Model)
Lift (Random)
Lift (Optimum)
TPR (Model)
TPR (Random)
TPR (Optimum)
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
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
Max.
N/A
1
1
1
0
0
0
0
Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model) TPR (Random) TPR (Optimum)
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
Decile Chart Data
Max.
N/A
1
1
1
1
0
0
0
0
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
Cumulative (Random)
Lift (Model)
Lift (Random)
Lift (Optimum)
TPR (Model)
TPR (Random)
TPR (Optimum)
0 0 0 0 0 0 0
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Ready
100%




**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**

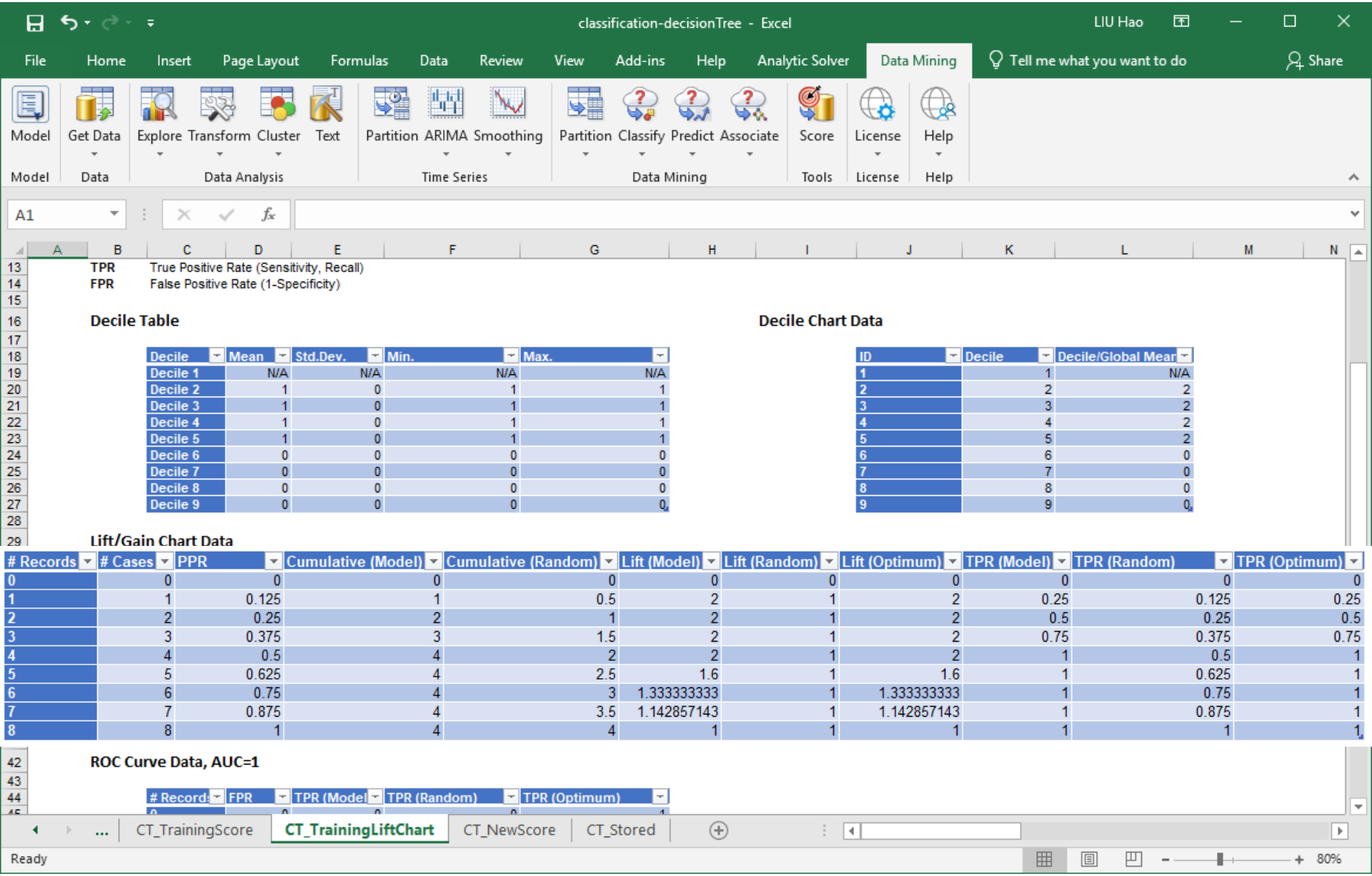
| ID | Decile | Decile/Global Mean |
|---|---|---|
| 1 | 1 | N/A |
| 2 | 2 | 2 |
| 3 | 3 | 2 |
| 4 | 4 | 2 |
| 5 | 5 | 2 |
| 6 | 6 | 0 |
| 7 | 7 | 0 |
| 8 | 8 | 0 |
| 9 | 9 | 0 |

**Lift/Gain Chart Data**

| # 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 |

**ROC Curve Data, AUC=1**

| # Records | FPR | TPR (Model) | TPR (Random) | TPR (Optimum) |
|---|---|---|---|---|




**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**

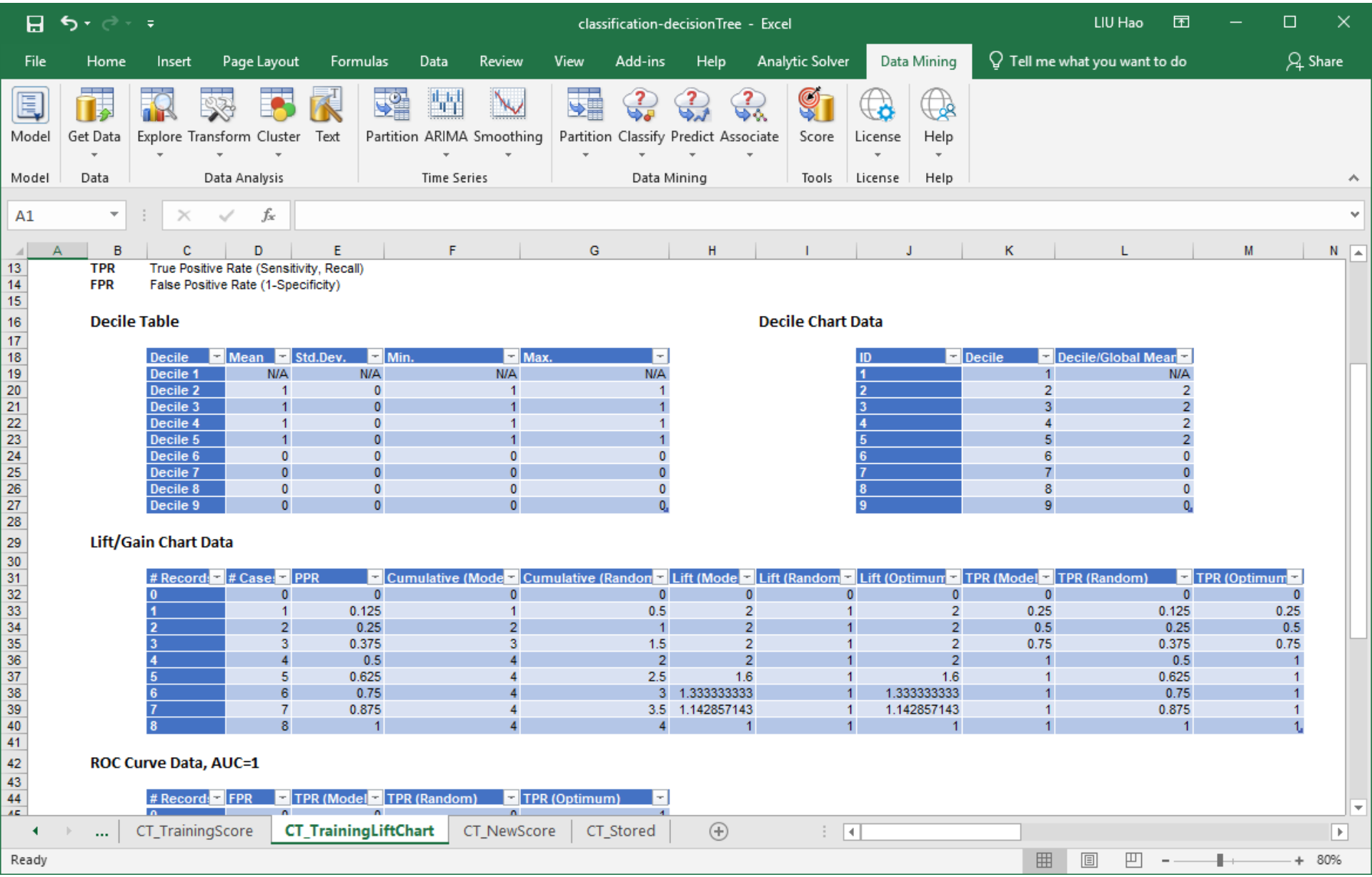
| ID | Decile | Decile/Global Mean |
|---|---|---|
| 1 | 1 | N/A |
| 2 | 2 | 2 |
| 3 | 3 | 2 |
| 4 | 4 | 2 |
| 5 | 5 | 2 |
| 6 | 6 | 0 |
| 7 | 7 | 0 |
| 8 | 8 | 0 |
| 9 | 9 | 0 |

**Lift/Gain Chart Data**

| # 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 |

**ROC Curve Data, AUC=1**

| # Record | FPR | TPR (Model | TPR (Random) | TPR (Optimum) |
|---|---|---|---|---|




**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**

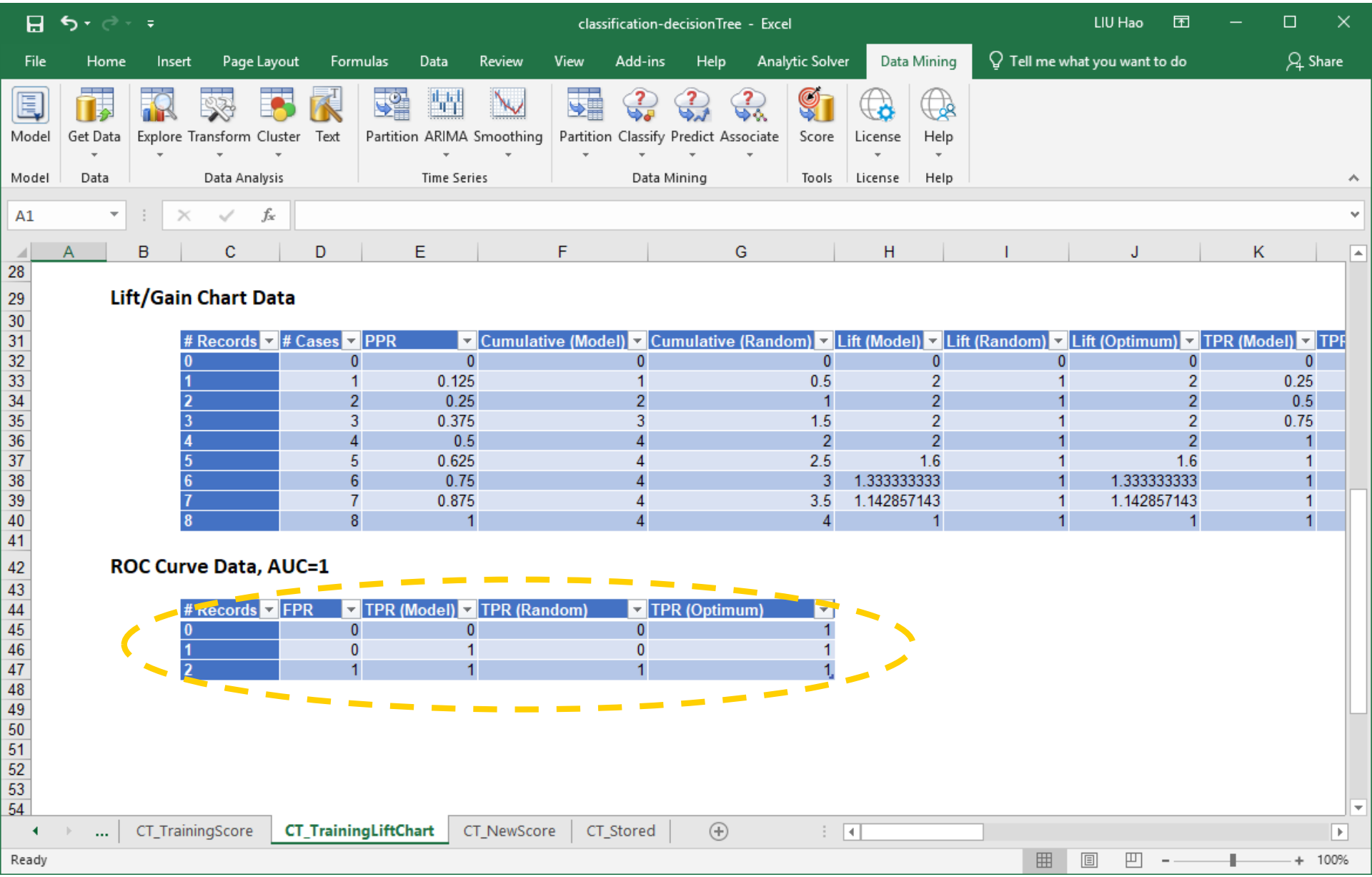
| ID | Decile | Decile/Global Mean |
|---|---|---|
| 1 | 1 | N/A |
| 2 | 2 | 2 |
| 3 | 3 | 2 |
| 4 | 4 | 2 |
| 5 | 5 | 2 |
| 6 | 6 | 0 |
| 7 | 7 | 0 |
| 8 | 8 | 0 |
| 9 | 9 | 0 |

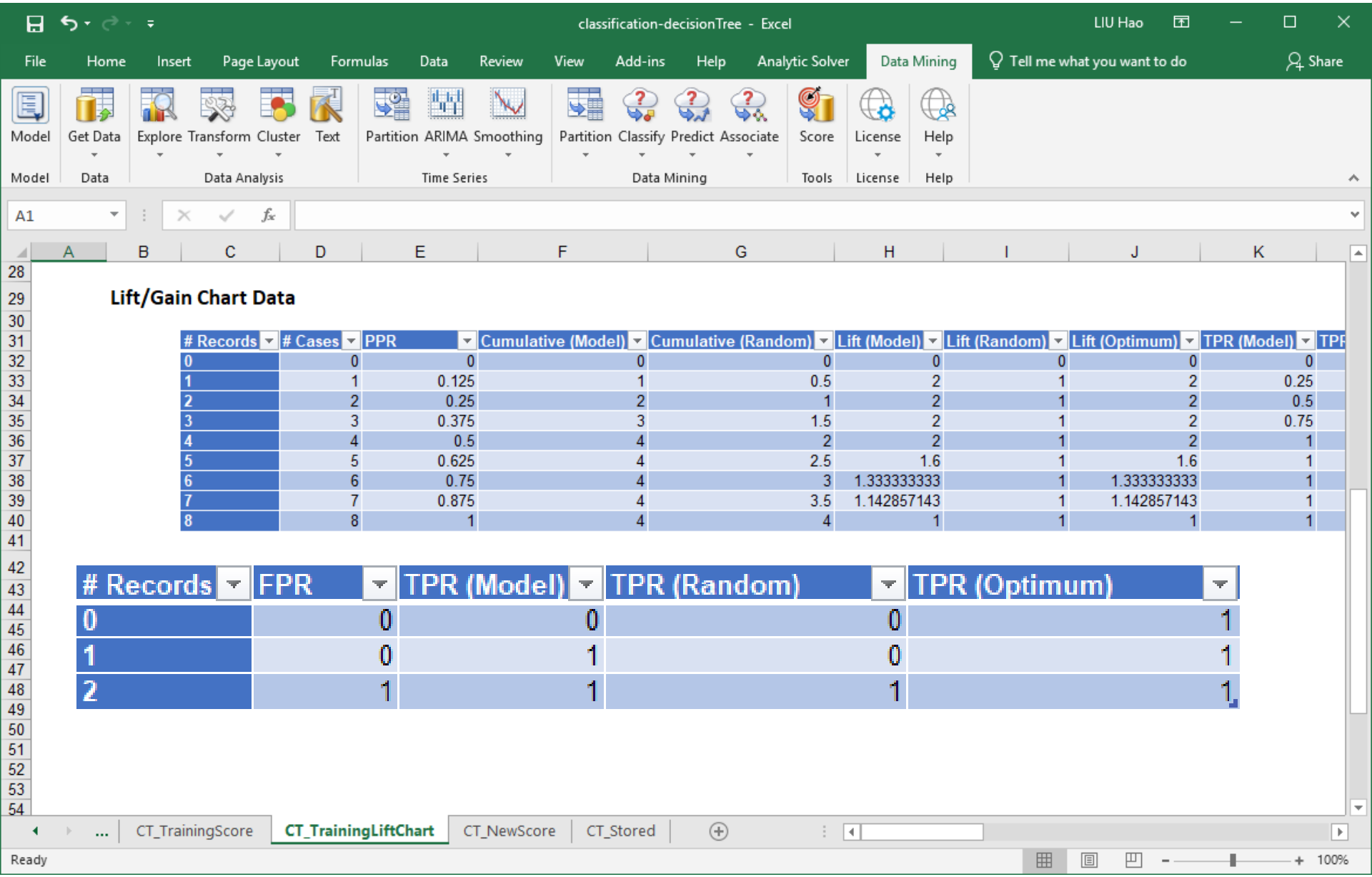
**Lift/Gain Chart Data**

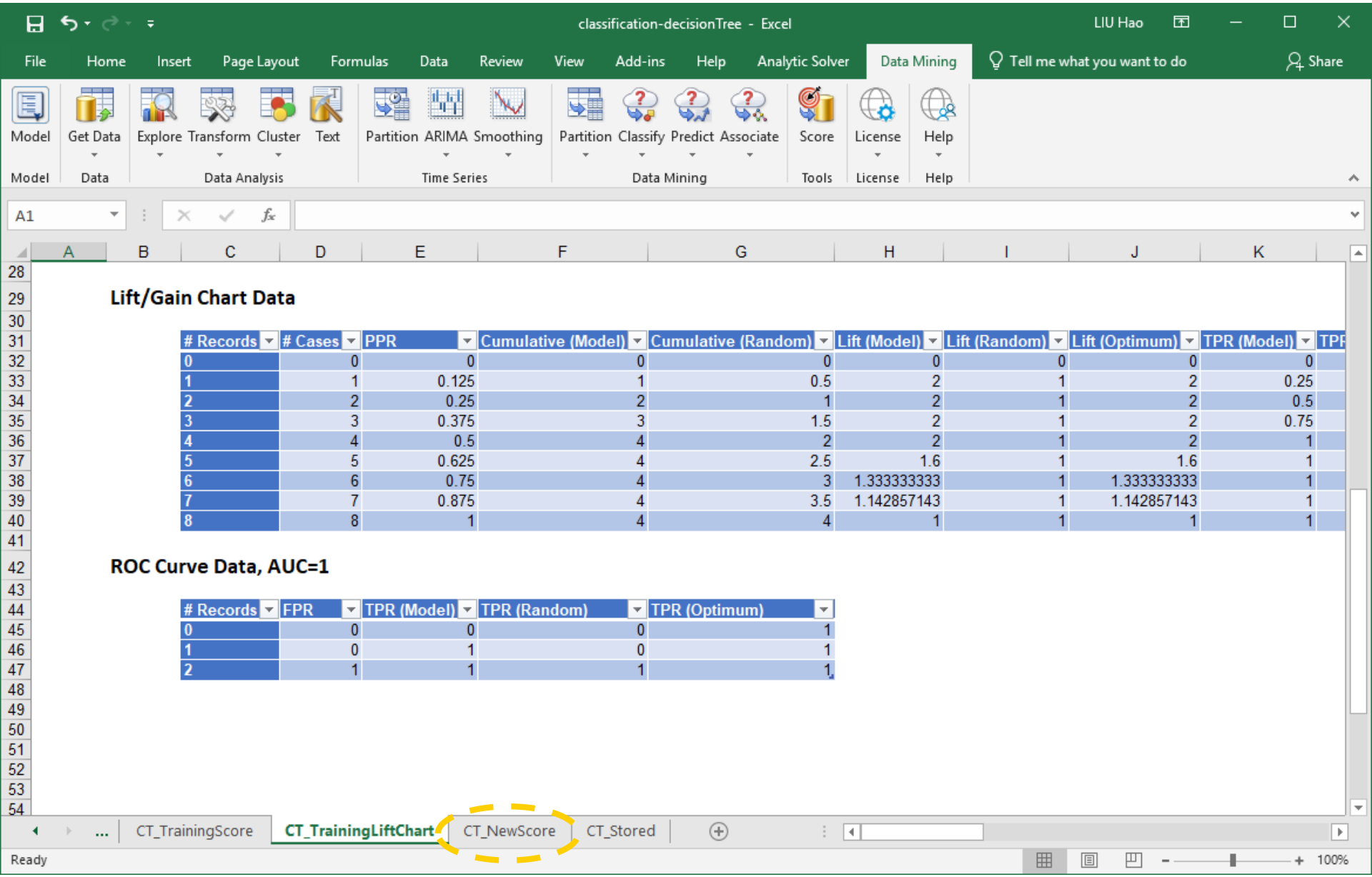
| # Records | # Cases | PPR | Cumulative (Model) | Cumulative (Random) | Lift (Model) | Lift (Random) | Lift (Optimum) | TPR (Model) | TPR (Random) | TPR (Optimum) |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0.125 | 1 | 0.5 | 2 | 1 | 2 | 0.25 | 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 |

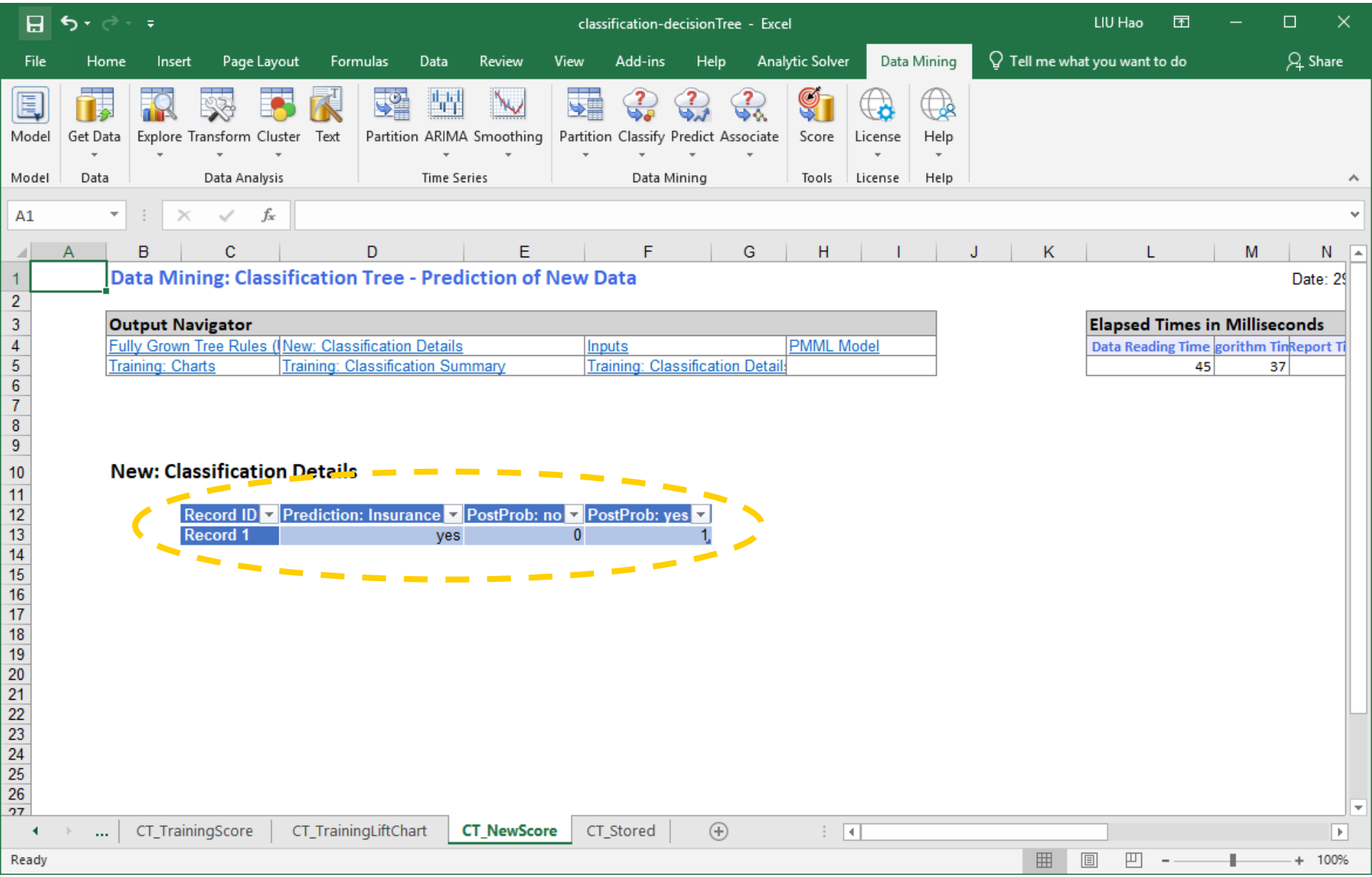
**ROC Curve Data, AUC=1**

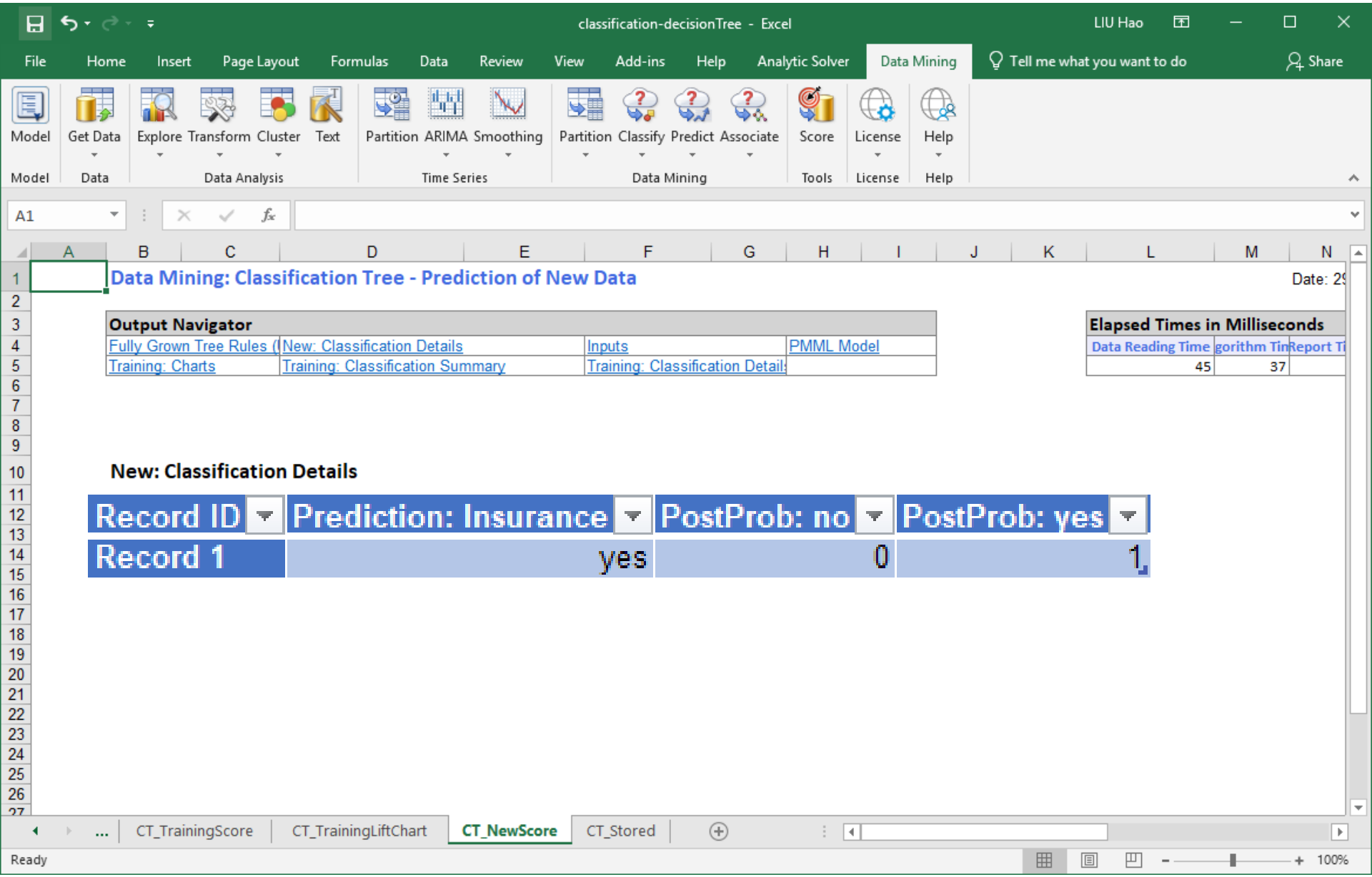
| # Records | FPR | TPR (Model) | TPR (Random) | TPR (Optimum) |
|---|---|---|---|---|


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
Lift/Gain Chart Data
# Records # Cases PPR Cumulative (Model) Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model)
0 0 0 0 0 0 0 0 0
1 1 0.125 1 0.5 2 1 2 0.25
2 2 0.25 2 1 2 1 2 0.5
3 3 0.375 3 1.5 2 1 2 0.75
4 4 0.5 4 2 2 1 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
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_TrainingScore CT_TrainingLiftChart CT_NewScore CT_Stored
Ready
100%

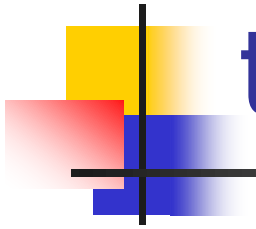

classification-decisionTree - Excel
LIU Hao
File Home Insert Page Layout Formulas Data Review View Add-ins Help Analytic Solver Data Mining Tell me what you want to do Share
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Associate Score License Help
Model Data Data Analysis Time Series Data Mining Tools License Help
A1
Lift/Gain Chart Data
# Records # Cases PPR Cumulative (Model) Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model)
0 0 0 0 0 0 0 0 0
1 1 0.125 1 0.5 2 1 2 0.25
2 2 0.25 2 1 2 1 2 0.5
3 3 0.375 3 1.5 2 1 2 0.75
4 4 0.5 4 2 2 1 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
# Records FPR TPR (Model) TPR (Random) TPR (Optimum)
0 0 0 0 1
1 0 1 0 1
2 1 1 1 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
Lift/Gain Chart Data
# Records # Cases PPR Cumulative (Model) Cumulative (Random) Lift (Model) Lift (Random) Lift (Optimum) TPR (Model)
0 0 0 0 0 0 0 0 0
1 1 0.125 1 0.5 2 1 2 0.25
2 2 0.25 2 1 2 1 2 0.5
3 3 0.375 3 1.5 2 1 2 0.75
4 4 0.5 4 2 2 1 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
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_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
Data Mining: Classification Tree - Prediction of New Data
Date: 2
Output Navigator
Fully Grown Tree Rules (
New: Classification Details
Inputs
PMML Model
Training: Charts
Training: Classification Summary
Training: Classification Details
Elapsed Times in Milliseconds
Data Reading Time
gorithm Tin
Report Ti
45
37
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
Data Mining: Classification Tree - Prediction of New Data
Date: 2
Output Navigator
Fully Grown Tree Rules (
New: Classification Details
Inputs
PMML Model
Training: Charts
Training: Classification Summary
Training: Classification Details
Elapsed Times in Milliseconds
Data Reading Time
gorithm Tim
Report Ti
45
37
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%

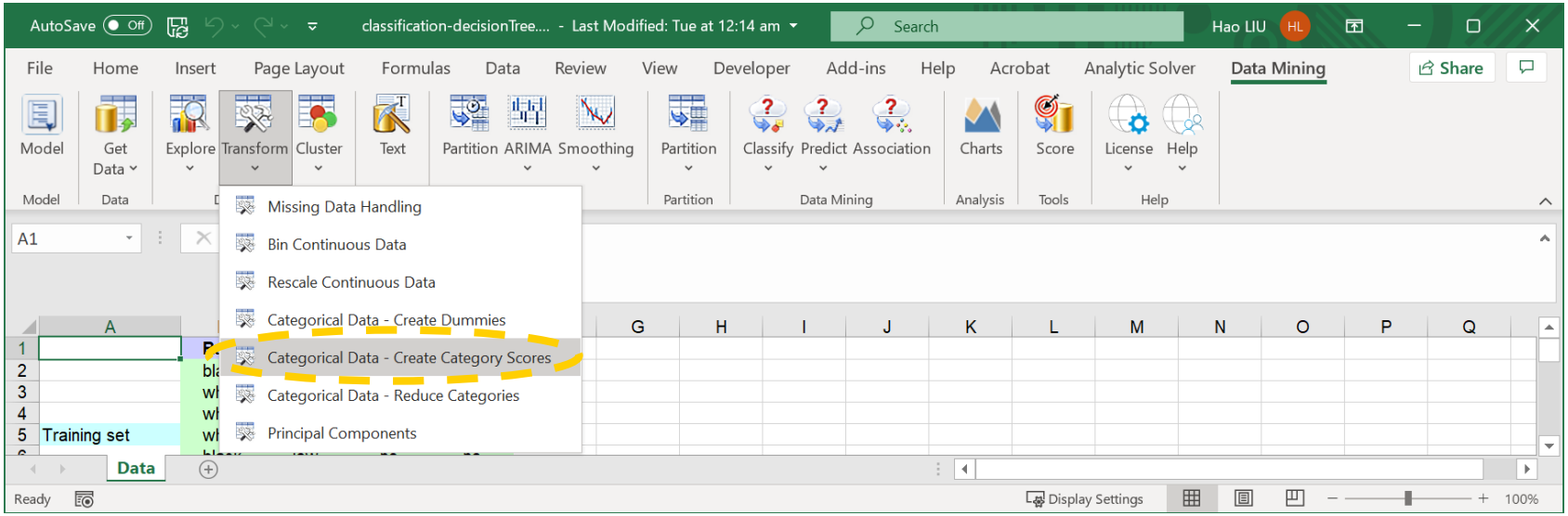
# How to use the data mining tool



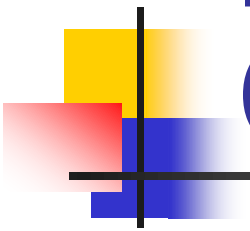
- We have the following 2 versions.
  - XLMiner Desktop (installed in either the CSE lab machine or your computer)
  - XLMiner Cloud (installed as a plugin in your Office 365 Excel)

# 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



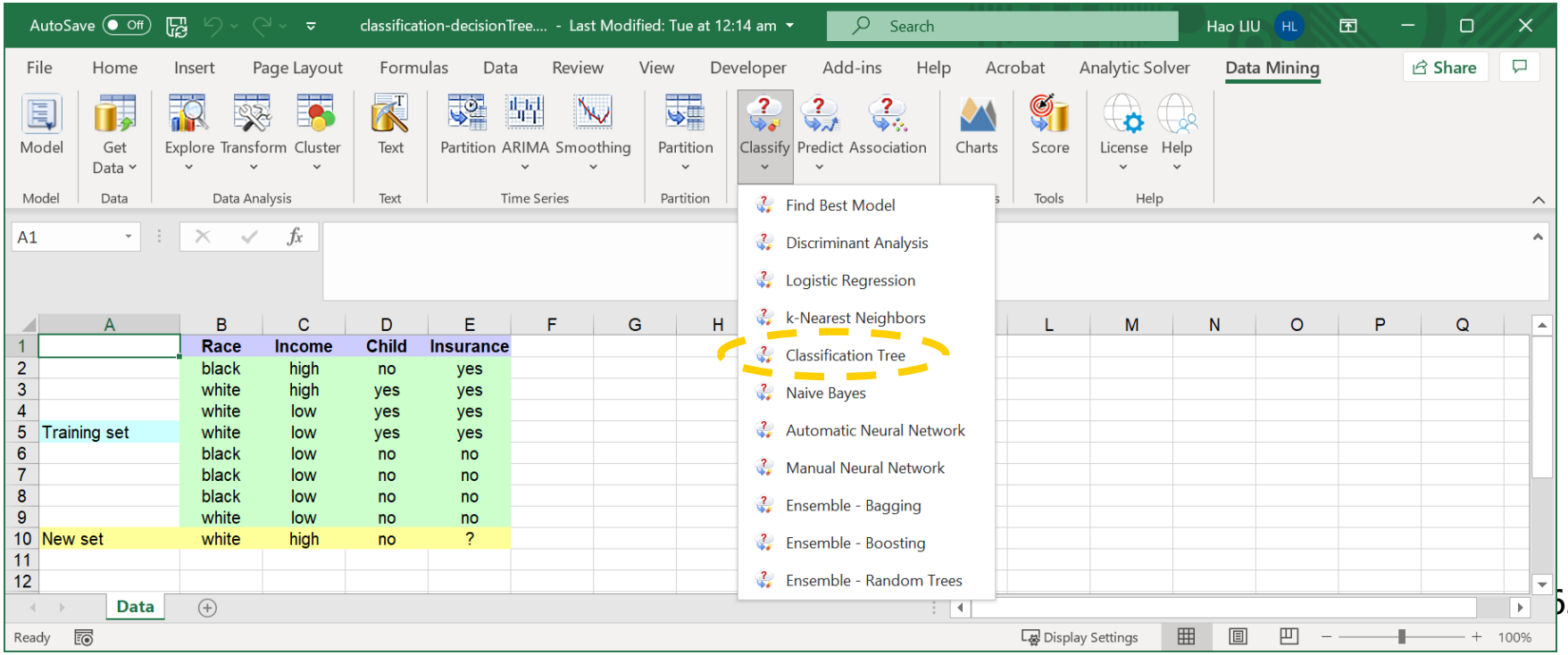
# How to use the data mining tool (XLMiner Cloud)



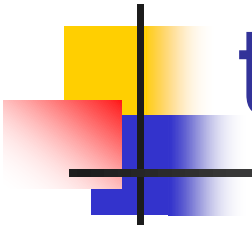
- The steps and output format of "create category scores" in XLMiner Cloud are similar to the steps in XLMiner Desktop.
- The transformation result of XLMiner Cloud Platform is the same as that from XLMiner Desktop.

# 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

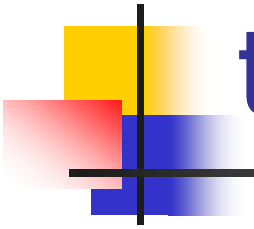


# How to use the data mining tool (XLMiner Cloud)

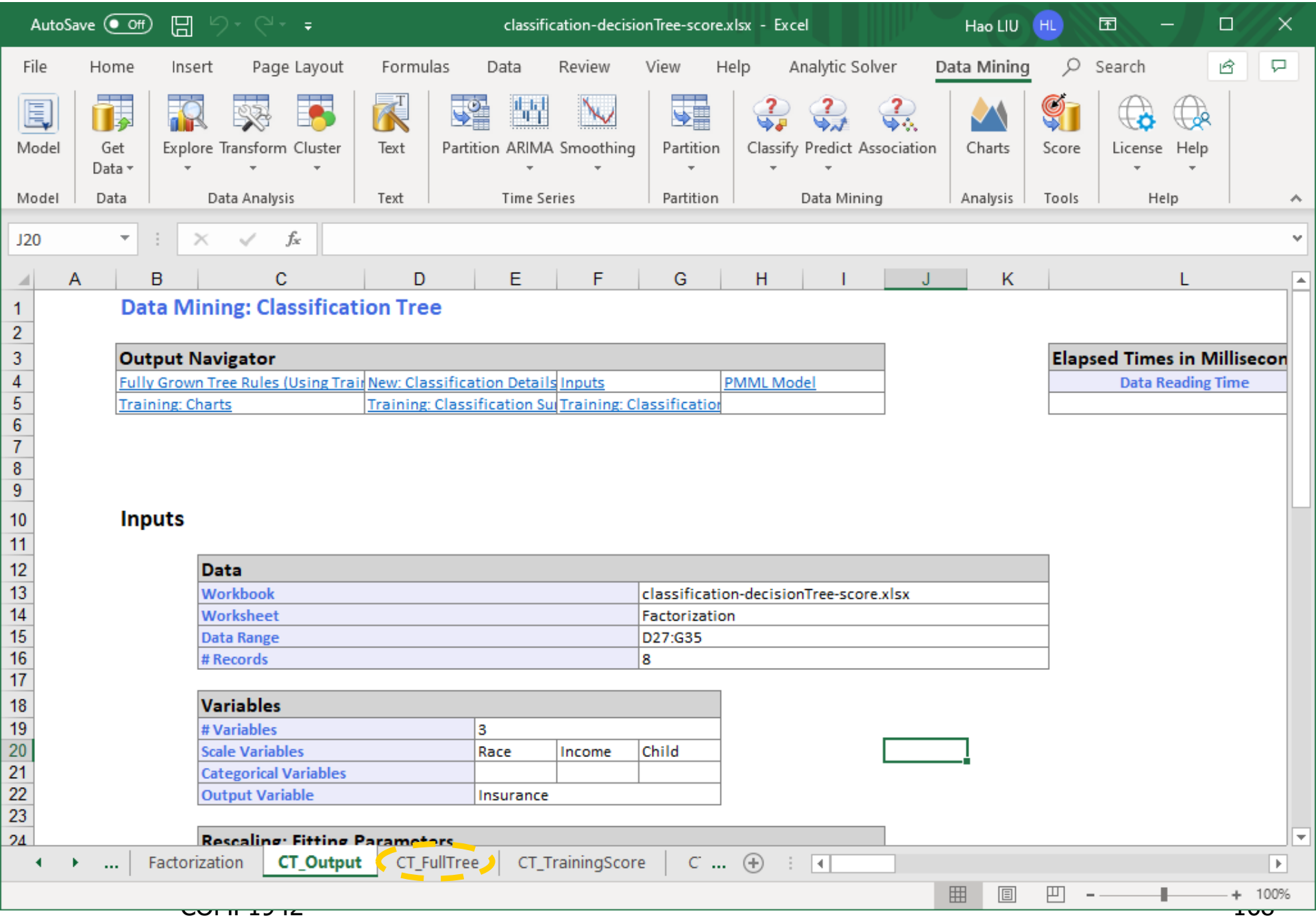


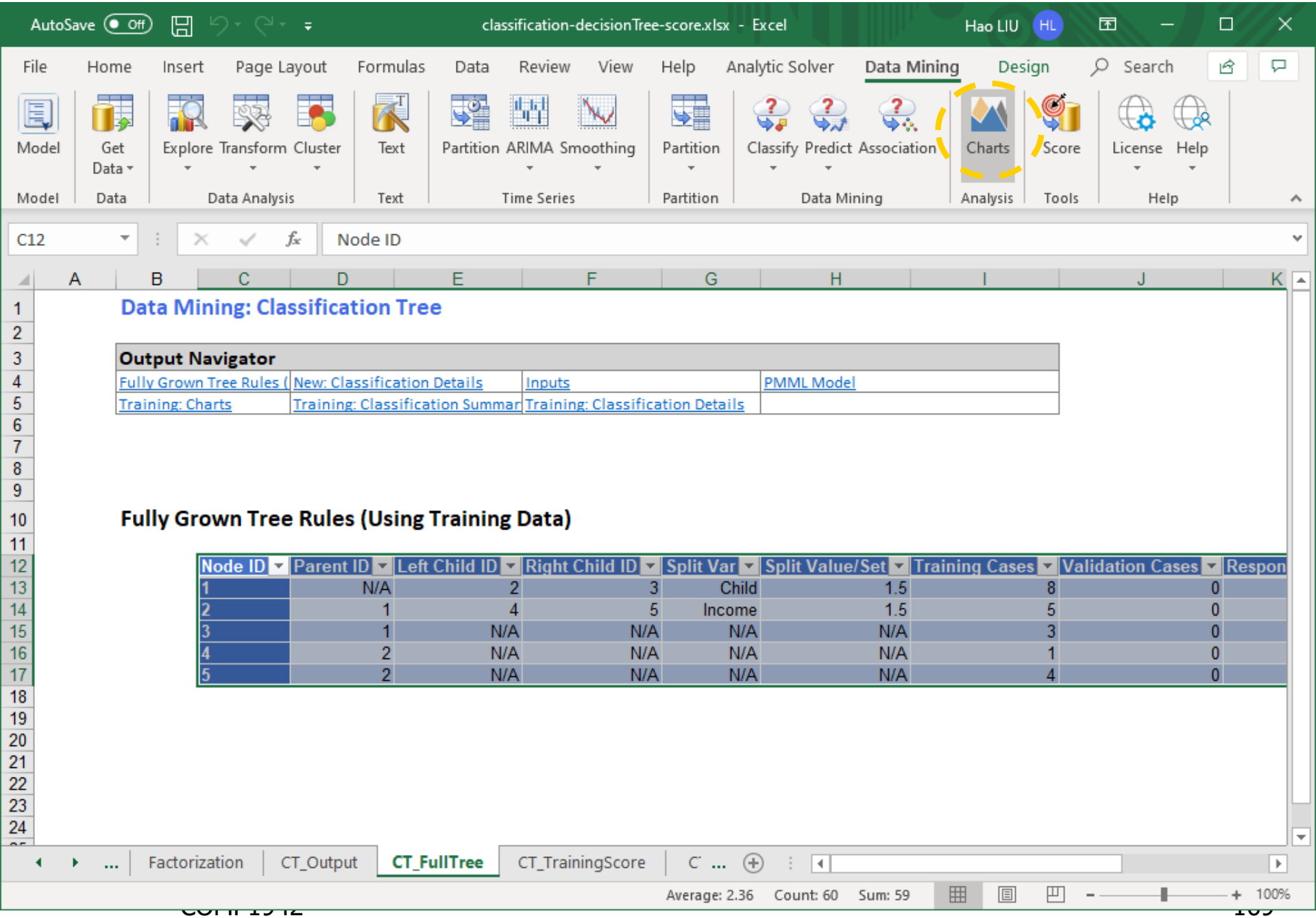
- The steps of performing “classification tree” in XLMiner Cloud is similar to the steps in XLMiner Desktop.
- The decision tree and classification result of XLMiner Cloud is the same as that from XLMiner Desktop.

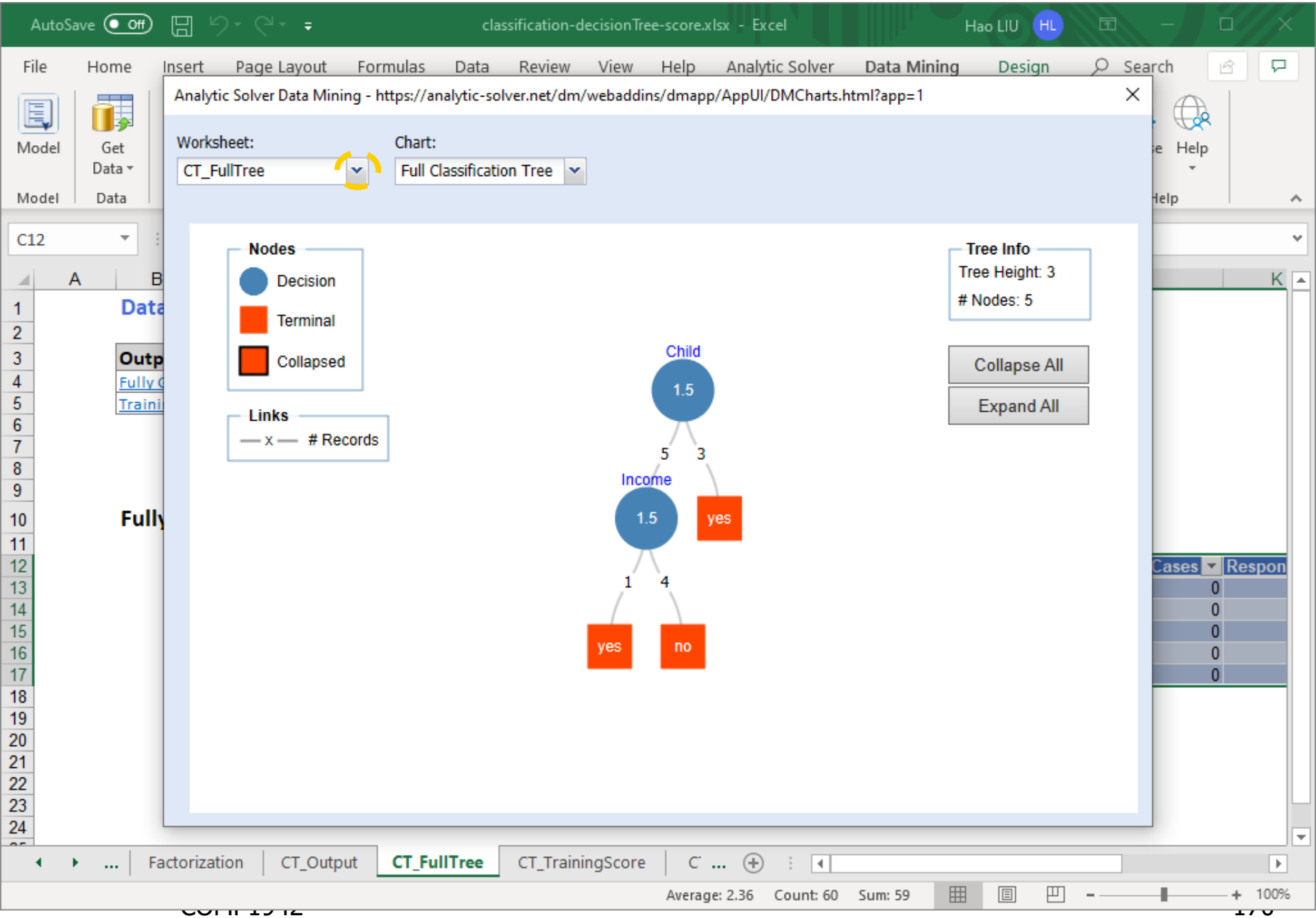
# How to use the data mining tool (XLMiner Cloud)

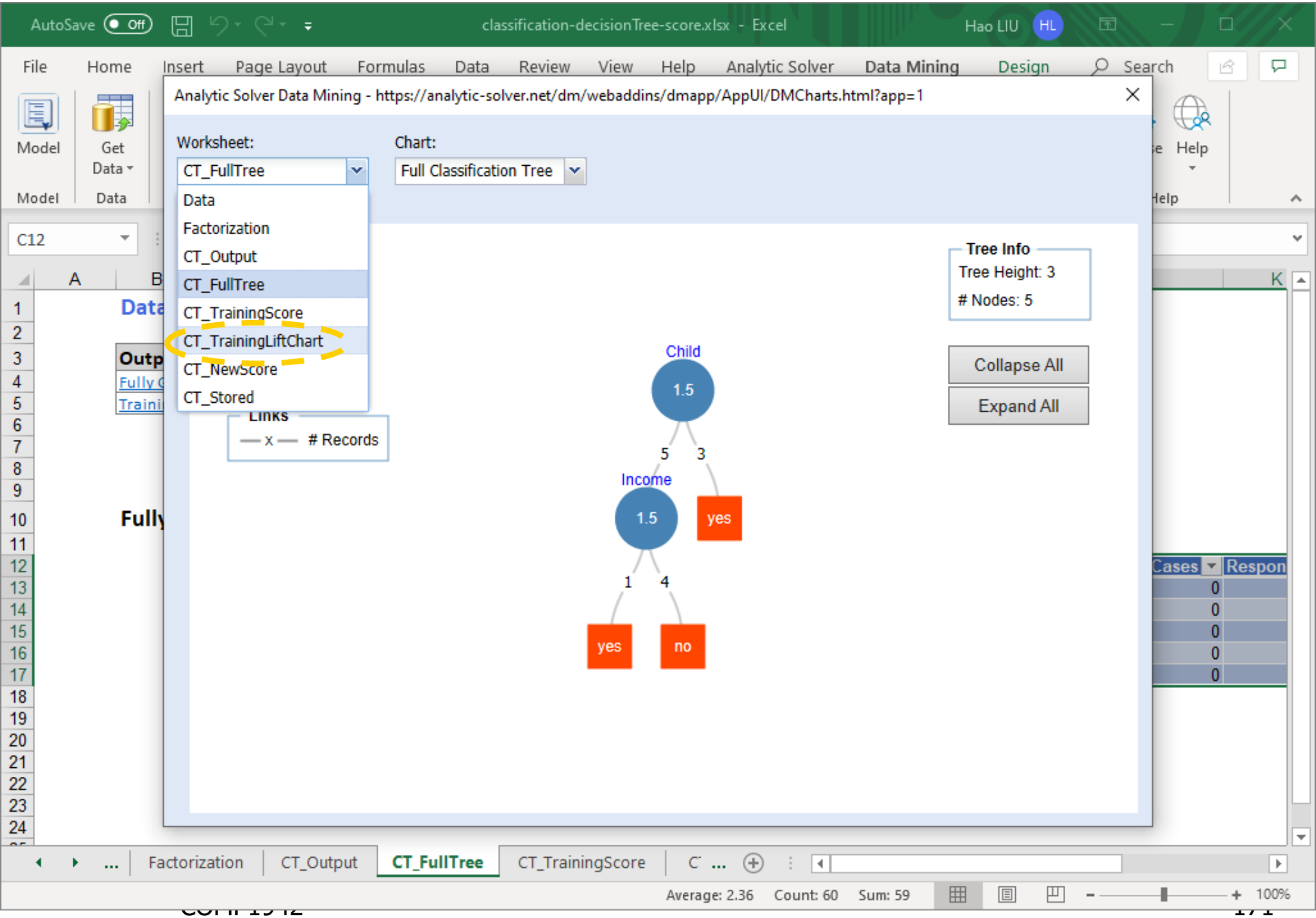


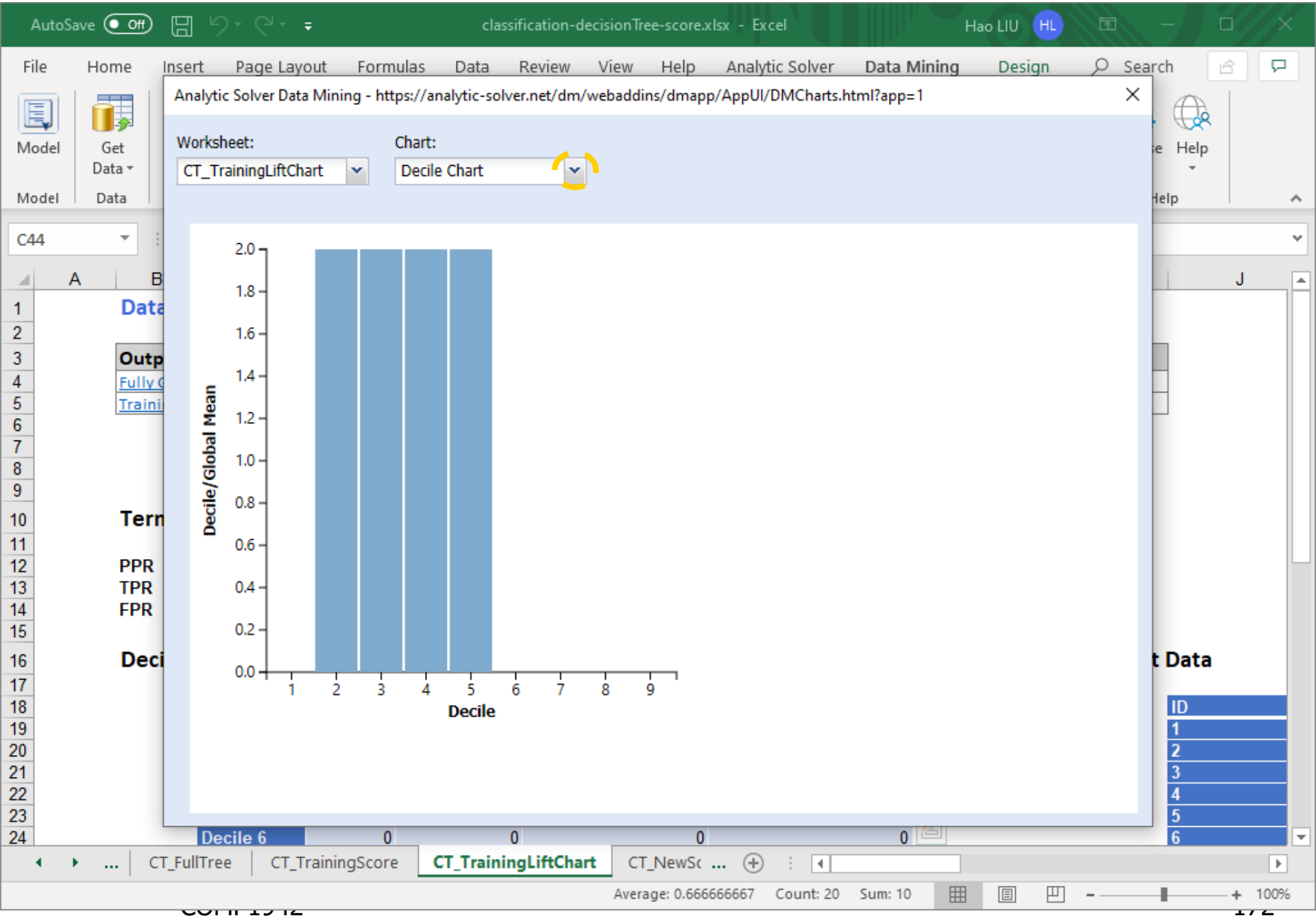
- The output format of XLMiner Cloud is similar to the output in XLMiner Desktop.
- However, to display the classification tree and the lift charts, you need to call out the “Charts” window.

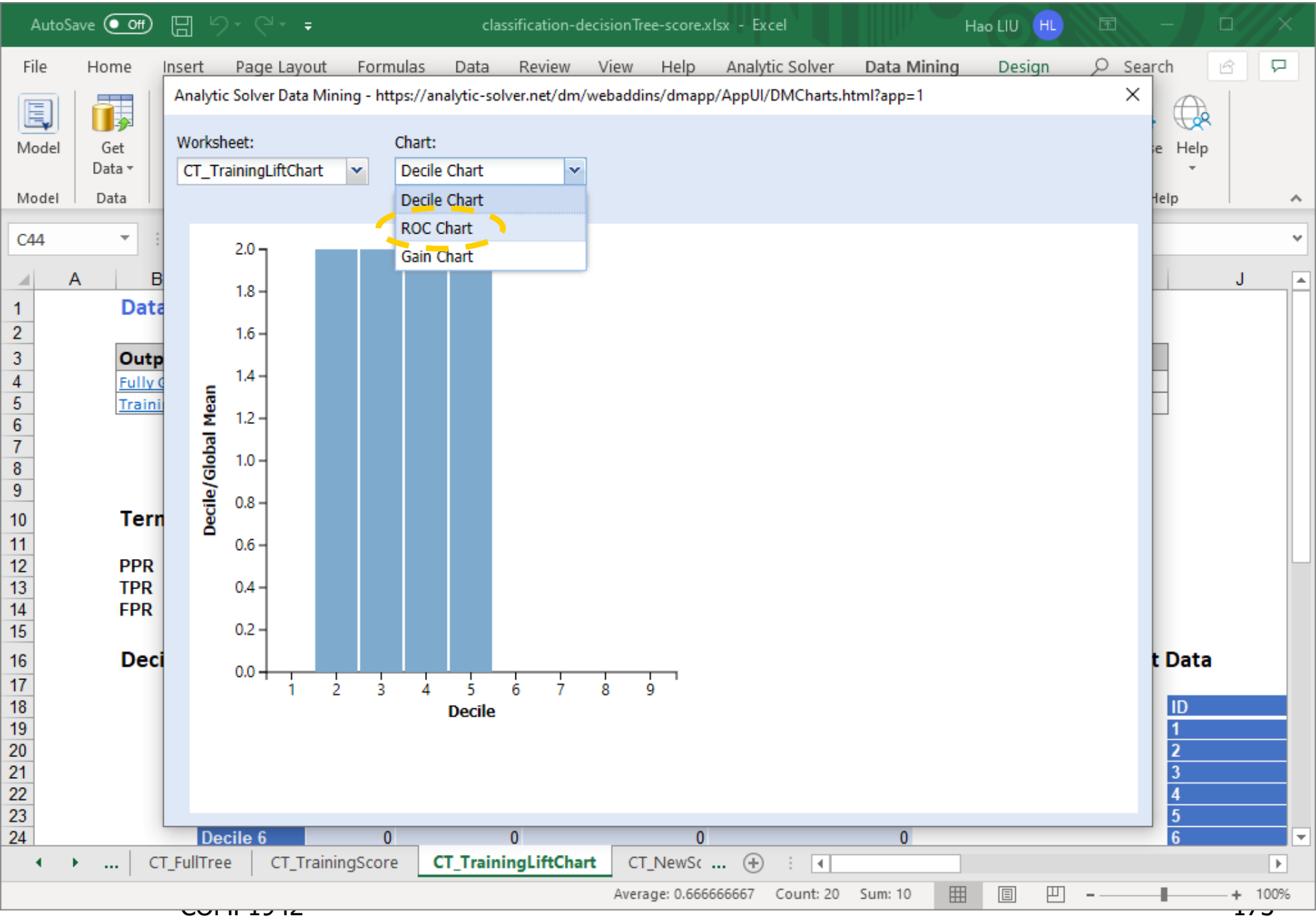

AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Search
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Association Charts Score License Help
Model Data Data Analysis Text Time Series Partition Data Mining Analysis Tools Help
J20
Data Mining: Classification Tree
Output Navigator
Fully Grown Tree Rules (Using Trai
New: Classification Details
Inputs
PMML Model
Training: Charts
Training: Classification Su
Training: Classificatior
Elapsed Times in Millisecon
Data Reading Time
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
100%

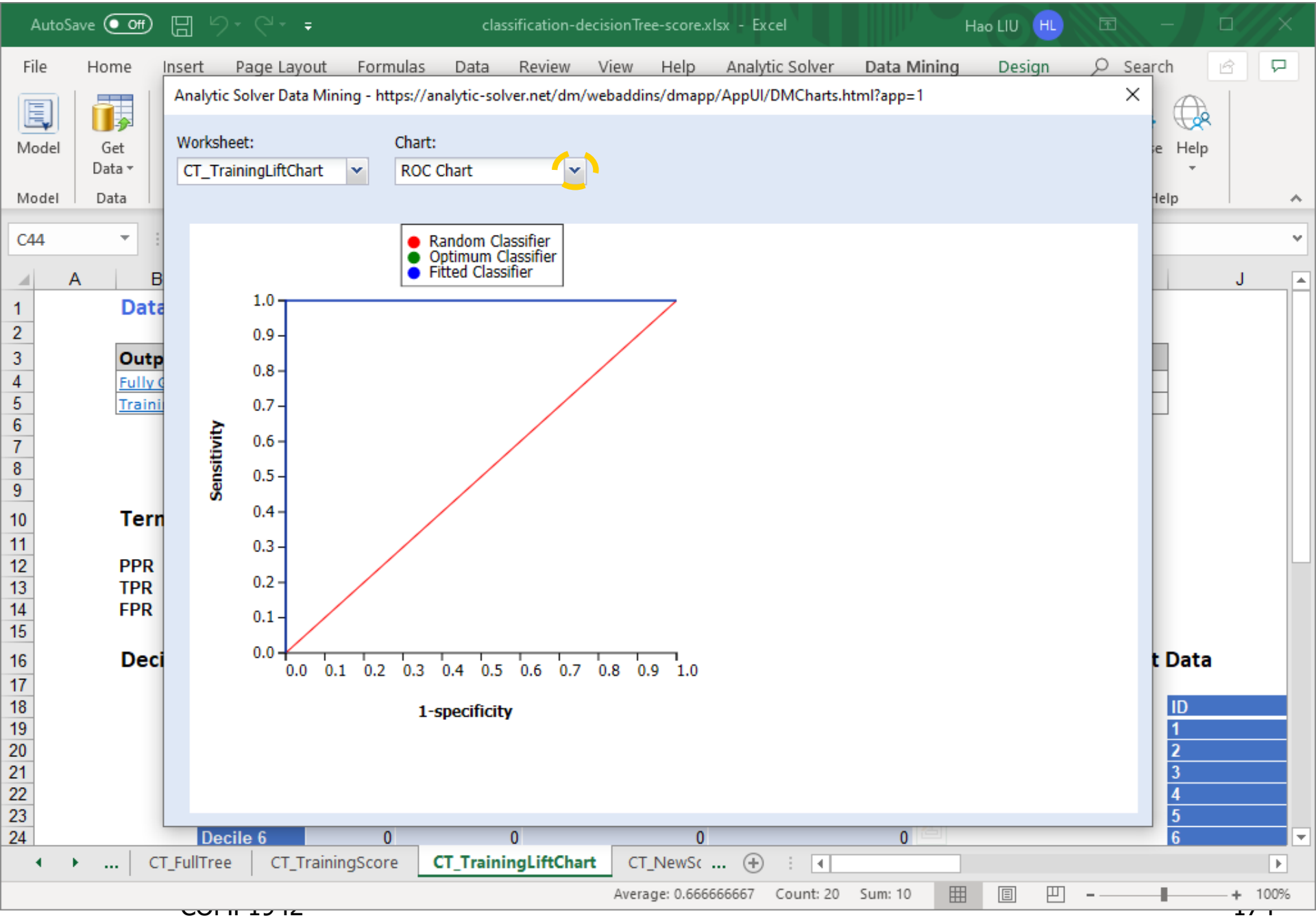

AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model Get Data Explore Transform Cluster Text Partition ARIMA Smoothing Partition Classify Predict Association Charts Score License Help
Model Data Data Analysis Text Time Series Partition Data Mining Analysis Tools Help
C12
Node ID
Data Mining: Classification Tree
Output Navigator
Fully Grown Tree Rules ( New: Classification Details Inputs PMML Model
Training: Charts Training: Classification Summar Training: Classification Details
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 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
Factorization CT_Output CT_FullTree CT_TrainingScore
Average: 2.36 Count: 60 Sum: 59
100%

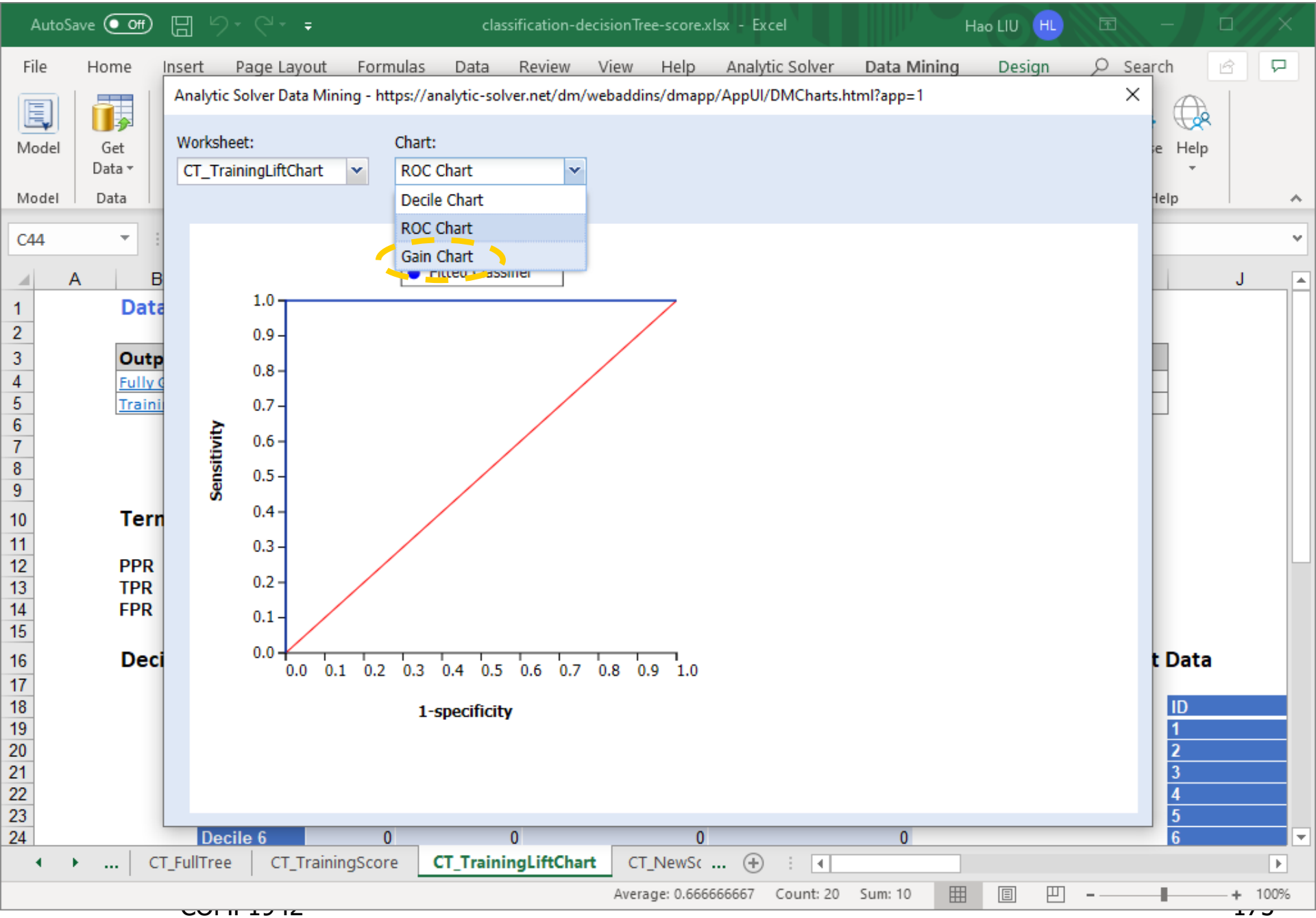

AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model
Get Data
Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_FullTree
Chart:
Full Classification Tree
Nodes
Decision
Terminal
Collapsed
Links
x # Records
Tree Info
Tree Height: 3
# Nodes: 5
Collapse All
Expand All
Child
1.5
5
3
Income
1.5
yes
1
4
yes
no
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Average: 2.36
Count: 60
Sum: 59
100%

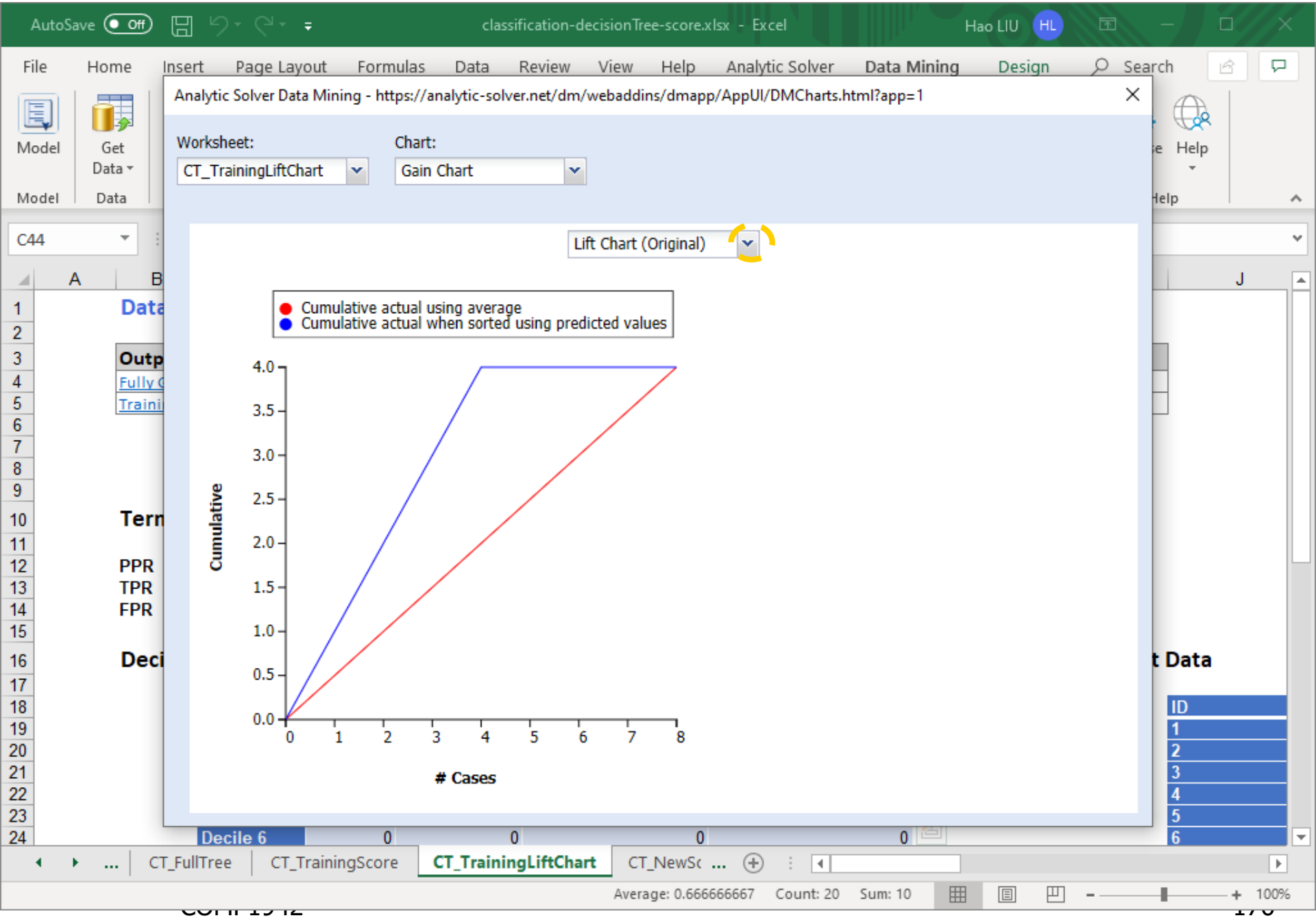

AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model
Get Data
Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_FullTree
Data
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
CT_TrainingLiftChart
CT_NewScore
CT_Stored
Chart:
Full Classification Tree
Links
x # Records
Tree Info
Tree Height: 3
# Nodes: 5
Collapse All
Expand All
Child
1.5
5
3
Income
1.5
yes
1
4
yes
no
Factorization
CT_Output
CT_FullTree
CT_TrainingScore
Average: 2.36
Count: 60
Sum: 59
100%

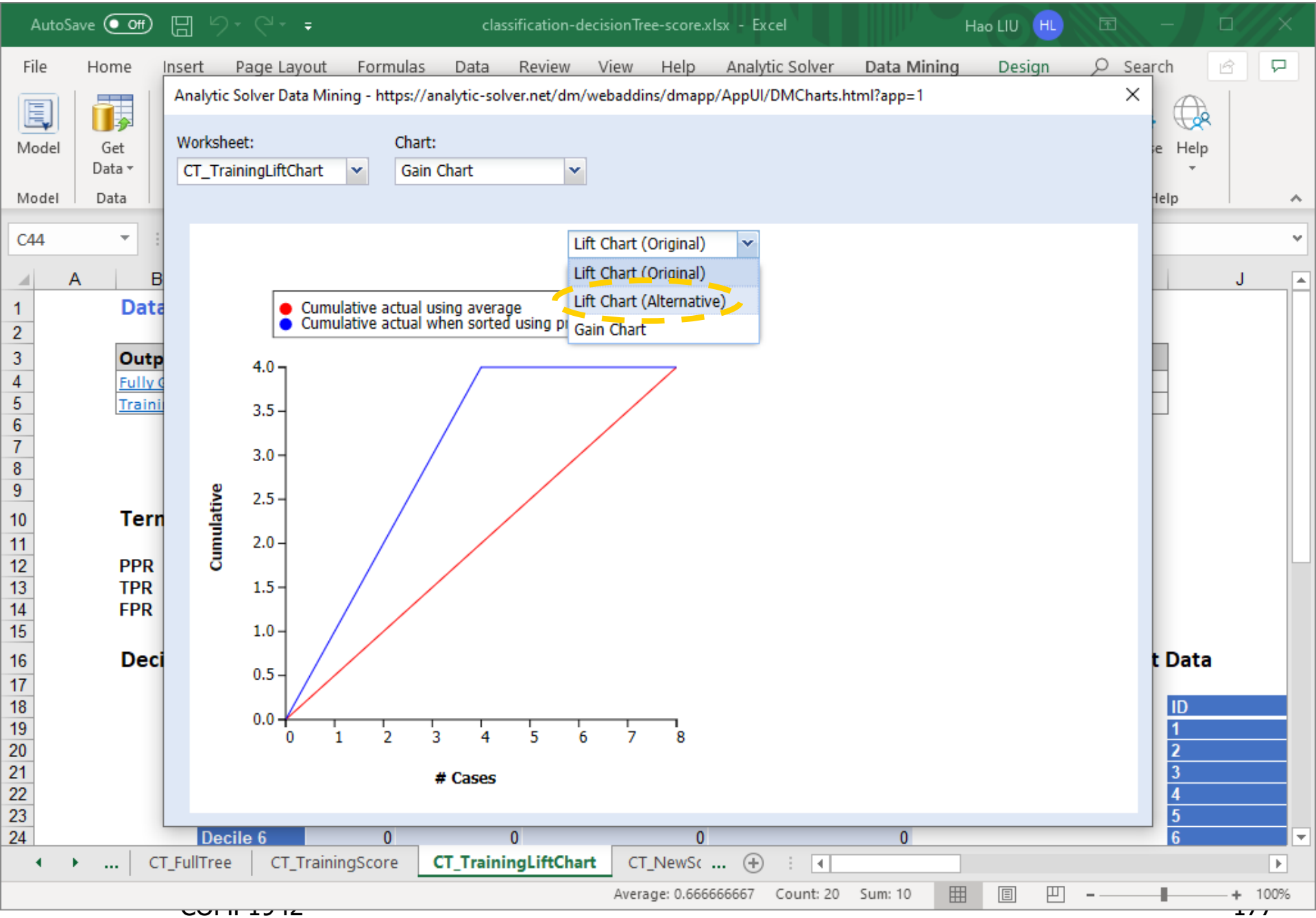

AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model Get Data Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
Decile Chart
Decile/Global Mean
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 6 7 8 9
Decile
Data
Outp
Fully
Traini
Tern
PPR
TPR
FPR
Deci
t Data
ID
1
2
3
4
5
6
Decile 6 0 0 0 0
CT_FullTree CT_TrainingScore CT_TrainingLiftChart CT_NewSc ...
Average: 0.666666667 Count: 20 Sum: 10
100%

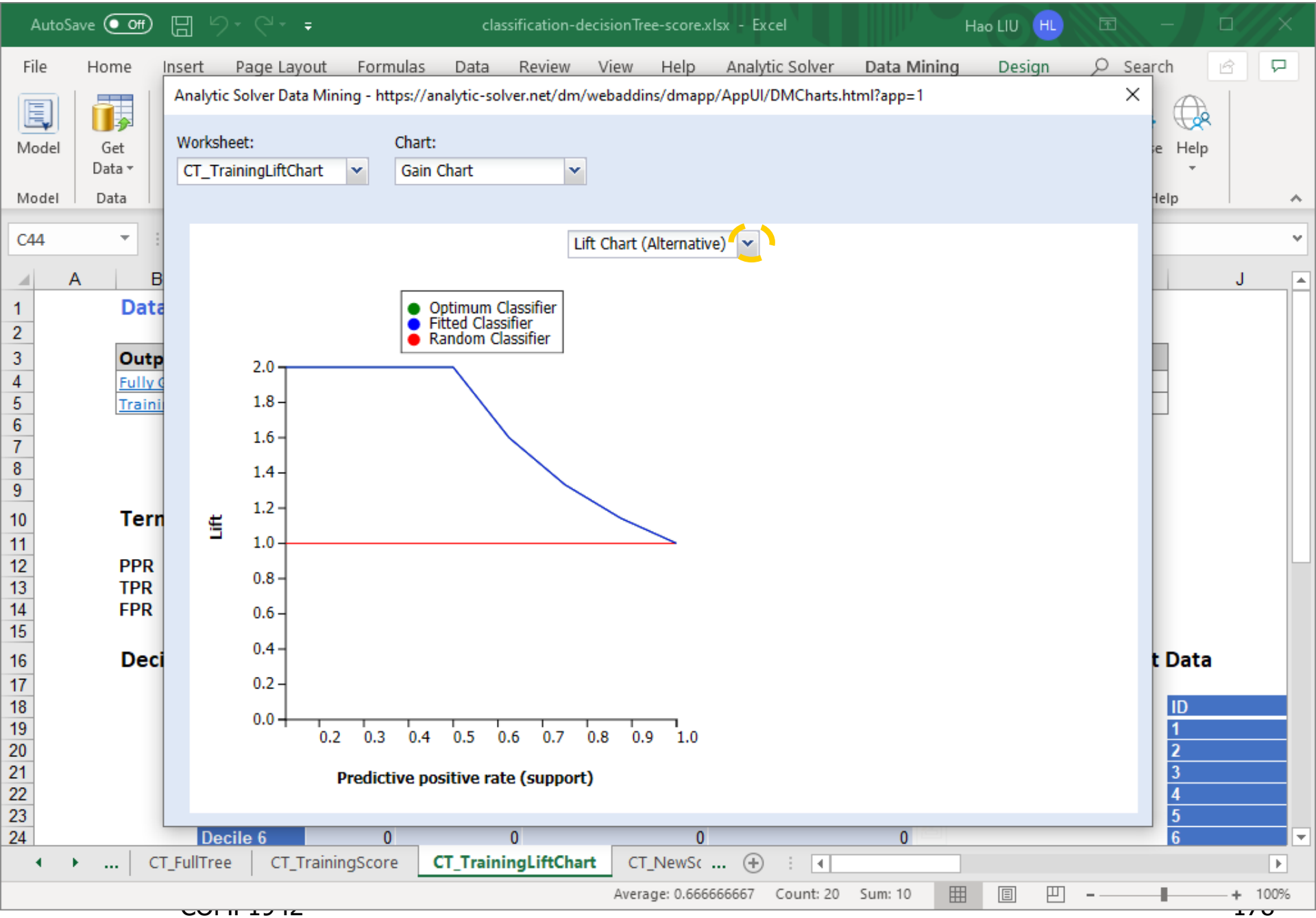

AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model
Get Data
Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
Decile Chart
Decile Chart
ROC Chart
Gain Chart
Decile/Global Mean
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 6 7 8 9
Decile
Data
Outp
Fully
Traini
Tern
PPR
TPR
FPR
Deci
t Data
ID
1
2
3
4
5
6
Decile 6 0 0 0 0
CT_FullTree
CT_TrainingScore
CT_TrainingLiftChart
CT_NewSc ...
Average: 0.666666667 Count: 20 Sum: 10
100%

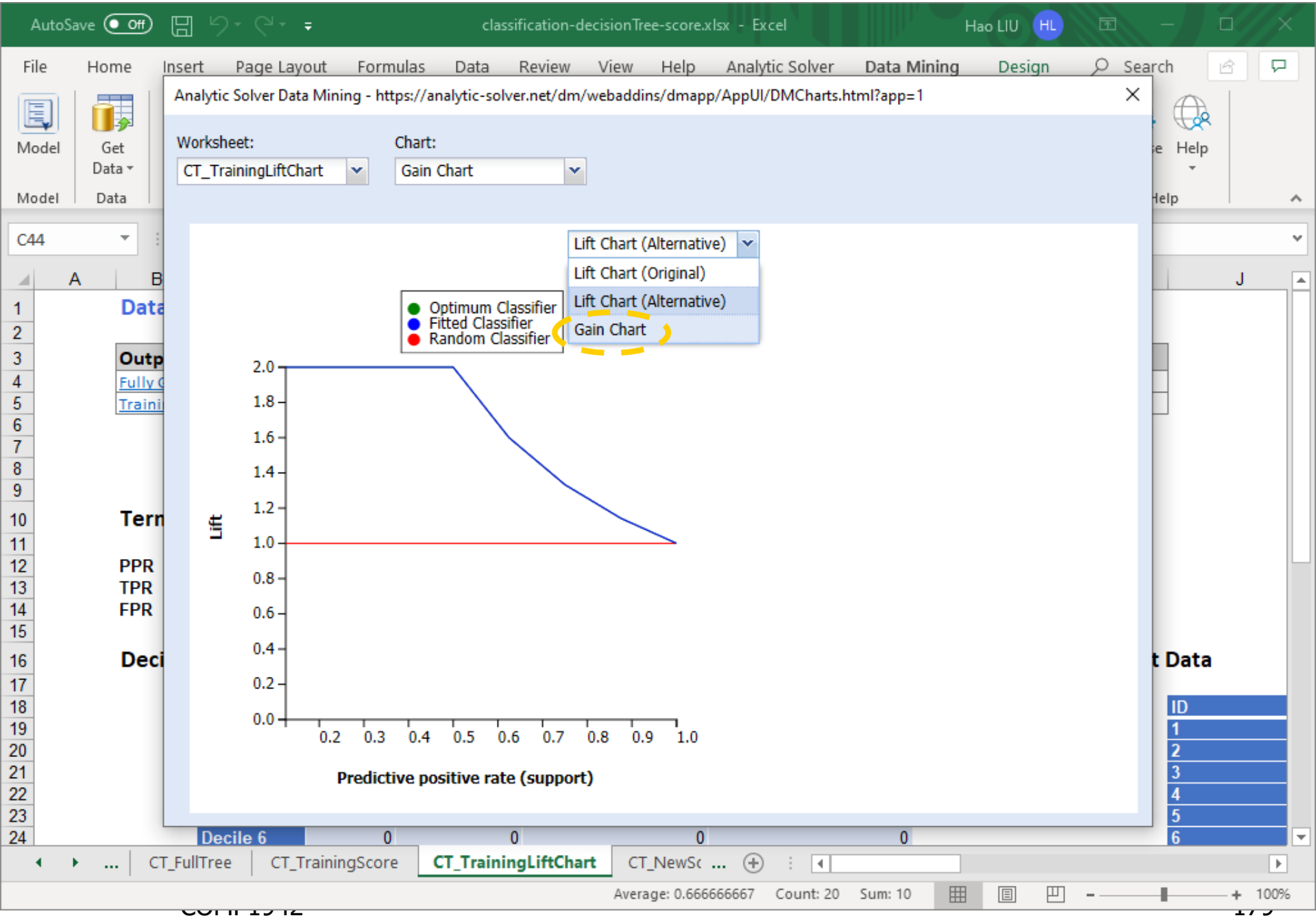

AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model Get Data Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
ROC Chart
Random Classifier
Optimum Classifier
Fitted Classifier
Sensitivity
1-specificity
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Data
Outp
Fully
Traini
Tern
PPR
TPR
FPR
Deci
t Data
ID
1
2
3
4
5
6
Decile 6 0 0 0 0
CT_FullTree CT_TrainingScore CT_TrainingLiftChart CT_NewSc ...
Average: 0.666666667 Count: 20 Sum: 10
100%

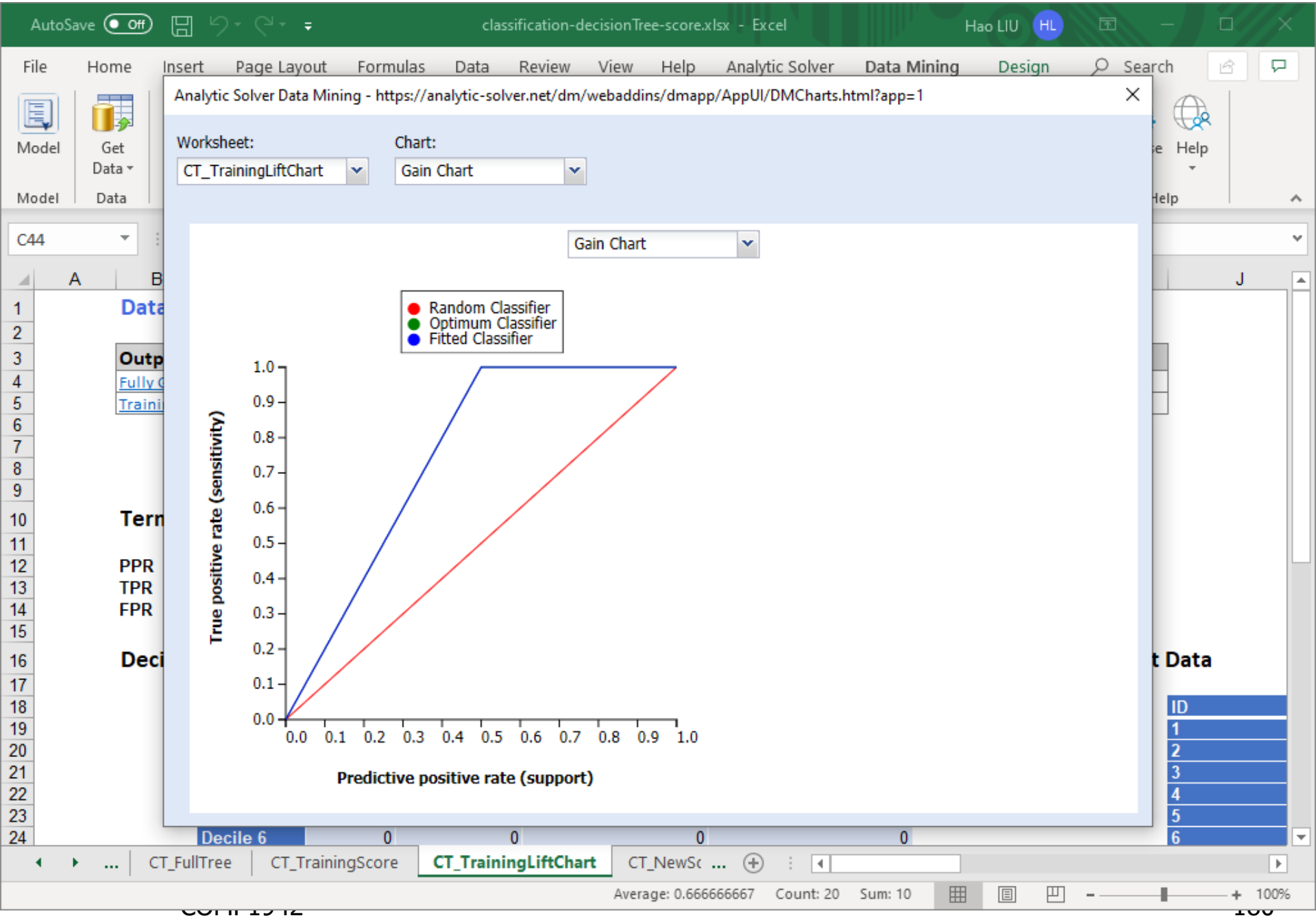

AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model Get Data
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
ROC Chart
Decile Chart
ROC Chart
Gain Chart
Sensitivity
1-specificity
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Data
Outp
Fully
Traini
Tern
PPR
TPR
FPR
Deci
t Data
ID
1
2
3
4
5
6
Decile 6 0 0 0 0
CT_FullTree CT_TrainingScore CT_TrainingLiftChart CT_NewSc ...
Average: 0.666666667 Count: 20 Sum: 10
100%


AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model Get Data Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
Gain Chart
Lift Chart (Original)
Cumulative actual using average
Cumulative actual when sorted using predicted values
Cumulative
# Cases
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0 1 2 3 4 5 6 7 8
Data
Outp
Tern
PPR
TPR
FPR
Deci
t Data
ID
1
2
3
4
5
6
Decile 6 0 0 0 0
CT_FullTree CT_TrainingScore CT_TrainingLiftChart CT_NewSc ...
Average: 0.666666667 Count: 20 Sum: 10
100%


AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model Get Data Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
Gain Chart
Lift Chart (Original)
Lift Chart (Original)
Lift Chart (Alternative)
Gain Chart
Cumulative actual using average
Cumulative actual when sorted using p
Cumulative
# Cases
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
1
2
3
4
5
6
7
8
Data
Outp
Tern
PPR
TPR
FPR
Deci
t Data
ID
1
2
3
4
5
6
Decile 6
CT_FullTree
CT_TrainingScore
CT_TrainingLiftChart
CT_NewSc
Average: 0.666666667
Count: 20
Sum: 10
100%


AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model
Get Data
Model Data
Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
Gain Chart
Lift Chart (Alternative)
Optimum Classifier
Fitted Classifier
Random Classifier
Lift
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predictive positive rate (support)
C44
Data
Outp
Fully
Traini
Tern
PPR
TPR
FPR
Deci
t Data
ID
1
2
3
4
5
6
Decile 6 0 0 0 0
CT_FullTree
CT_TrainingScore
CT_TrainingLiftChart
CT_NewSc ...
Average: 0.666666667 Count: 20 Sum: 10
100%


AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model
Get Data
Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
Gain Chart
Lift Chart (Alternative)
Lift Chart (Original)
Lift Chart (Alternative)
Gain Chart
Optimum Classifier
Fitted Classifier
Random Classifier
Lift
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predictive positive rate (support)
Data
Outp
PPR
TPR
FPR
Deci
t Data
ID
1
2
3
4
5
6
Decile 6 0 0 0 0
CT_FullTree
CT_TrainingScore
CT_TrainingLiftChart
CT_NewSc ...
Average: 0.666666667 Count: 20 Sum: 10
100%


AutoSave Off
classification-decisionTree-score.xlsx - Excel
Hao LIU HL
File Home Insert Page Layout Formulas Data Review View Help Analytic Solver Data Mining Design Search
Model Get Data Help
Analytic Solver Data Mining - https://analytic-solver.net/dm/webaddins/dmapp/AppUI/DMCharts.html?app=1
Worksheet:
CT_TrainingLiftChart
Chart:
Gain Chart
Gain Chart
Random Classifier
Optimum Classifier
Fitted Classifier
True positive rate (sensitivity)
Predictive positive rate (support)
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Decile 6 0 0 0 0
ID 1 2 3 4 5 6
CT_FullTree CT_TrainingScore CT_TrainingLiftChart
Average: 0.666666667 Count: 20 Sum: 10
100%