

Rapid Miner Studio

- แนะนำเมนูหลัก
- การจัดการข้อมูล (Data management)
- การนำข้อมูลเข้า
- Model
 - Euclidean Distance/k-NN Algorithm
 - Linear Regression Analysis
 - *k*-Means Clustering Algorithm
 - Artificial Neural Networks and Deep Learning

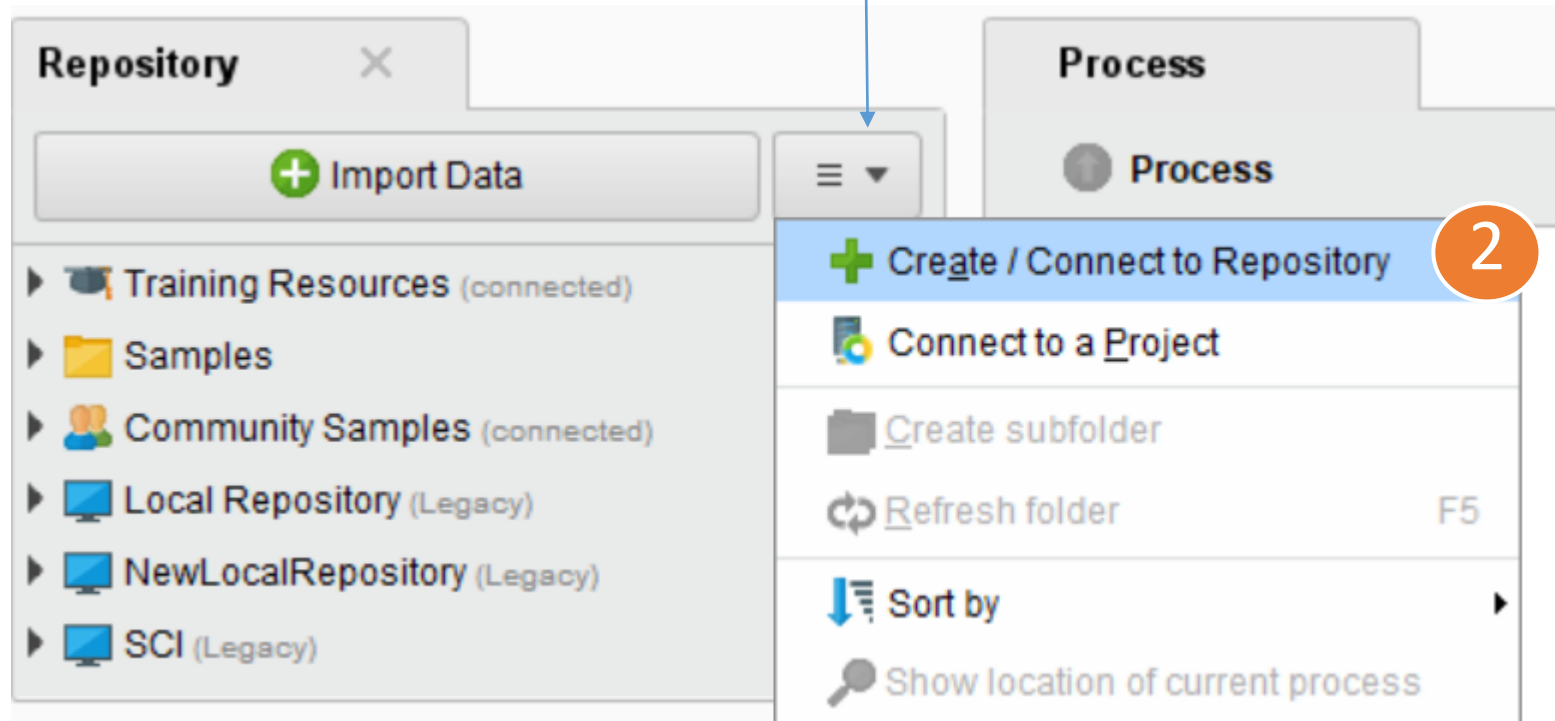
แนะนำเมนูหลัก 5 ฟังก์ชันการทำงานหลัก

The screenshot shows the RapidMiner Studio Free 9.1.000 interface. The title bar reads "<new process> - RapidMiner Studio Free 9.1.000 @ DESKTOP-EDIOK83". The menu bar includes File, Edit, Process, View, Connections, Cloud, Settings, Extensions, and Help. The toolbar contains icons for file operations and a 'Views' dropdown set to 'Design'. The main workspace is divided into several panels:

- Repository:** Contains a tree view of data sources (Sample, DB, Local, New, Cloud Repository) and an 'Import Data' button. A callout box labeled '1. เก็บข้อมูลและเก็บ Process ที่ได้สร้างไว้' (1. Collect data and collect the created process) points to this panel.
- Operators:** A list of available operators categorized by function (Data Access, Blending, Cleansing, Modeling, Scoring). A callout box labeled '2. ตัวดำเนินการต่าง ๆ' (2. Various operators) points to this panel.
- Process:** The central workspace for building a process flow. It shows a 'Process' node and a 'Drag data or operators here.' prompt. A callout box labeled '3. Process' points to this central area.
- Parameters:** A panel for configuring the selected process. It shows settings for 'logverbosity' (init) and 'logfile'. A callout box labeled '4. Parameter' points to this panel.
- Help:** A panel providing documentation for the selected process. It shows the 'Process' node and a 'Synopsis' section. A callout box labeled '5. Help' points to this panel.

การจัดการข้อมูล (Data management)

- Repository คือ ที่เก็บข้อมูลและเก็บ Process ที่ได้สร้างไว้ เพื่อไม่ต้องโหลดไฟล์ใหม่ทุกครั้ง
- เริ่มจากสร้างที่เก็บข้อมูลใหม่โดยคลิก 1



การจัดการข้อมูล (Data management)

เลือก

3

Create / Connect to Repository

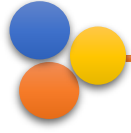
Enter parameters to create a new local or connect to a RapidMiner AI Hub repository.

☒ Create new local repository

☐ Connect to AI Hub repository

4

← Previous → Next Finish X Cancel



การจัดการข้อมูล (Data management)

- เปลี่ยนชื่อตามที่ต้องการ

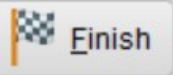
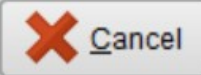
Create / Connect to Repository

Enter parameters to create a new local or connect to a RapidMiner AI Hub repository.

Alias: 5

Root directory: ☒ Use default location

✓

← Previous → Next  Finish  Cancel

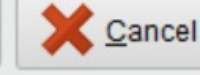
Create / Connect to Repository

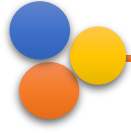
Enter parameters to create a new local or connect to a RapidMiner AI Hub repository.

Alias: 6

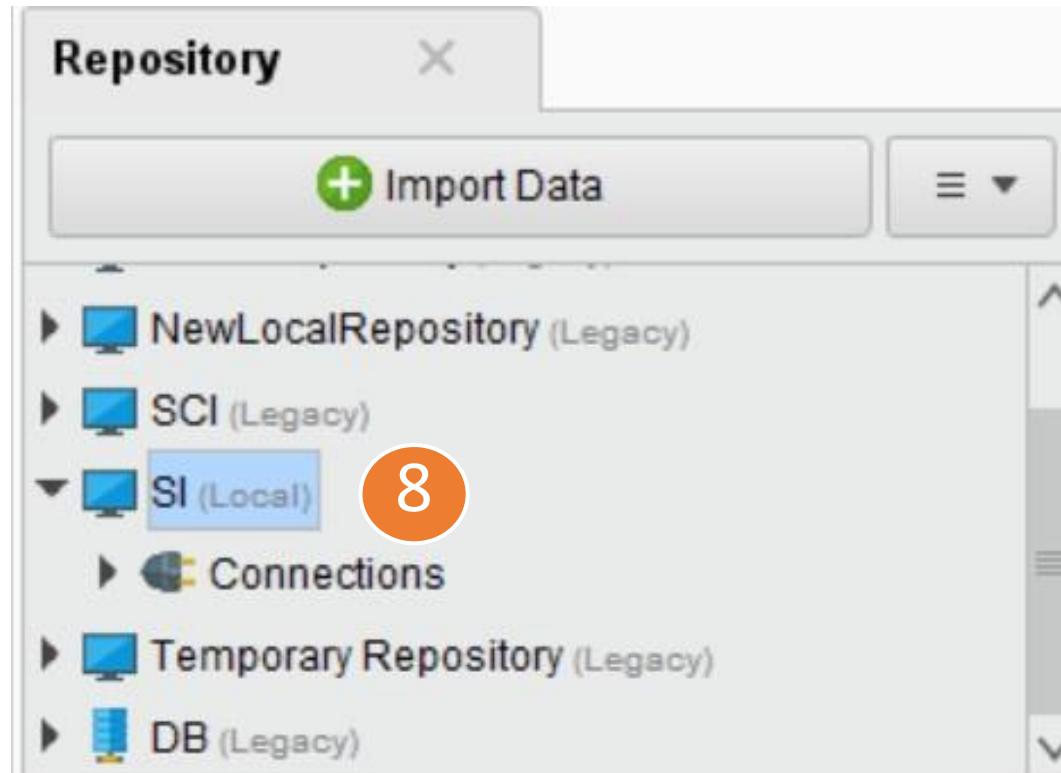
Root directory: ☒ Use default location

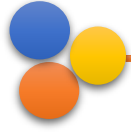
✓

← Previous → Next  7 Finish  Cancel



การจัดการข้อมูล (Data management)





นำข้อมูลต่อไปนี้ เข้า Excel และวิเคราะห์ข้อมูลด้วย RapidMiner

- ตารางที่ 1 ข้อมูลประวัติ (ข้อมูลฝึกฝน)

ID	Name	Weight	Height	Cholesterol	Blood sugar level	Result
1	One	70	175	175	80	Good
2	Two	81	159	250	250	Bad
3	Three	78	180	180	95	Good
4	Four	85	174	300	300	Bad
5	Five	76	163	250	245	Bad



การนำข้อมูลเข้า

1

<new process> – RapidMiner Studio Free 9.1.000 @ DESKTOP-EDIOK83

File Edit Process View Connections Cloud Settings Extensions Help

Design Results Turbo Prep

Repository

Process

100%

Import Data - Where is your data?

Where is your data?

2 My Computer Database

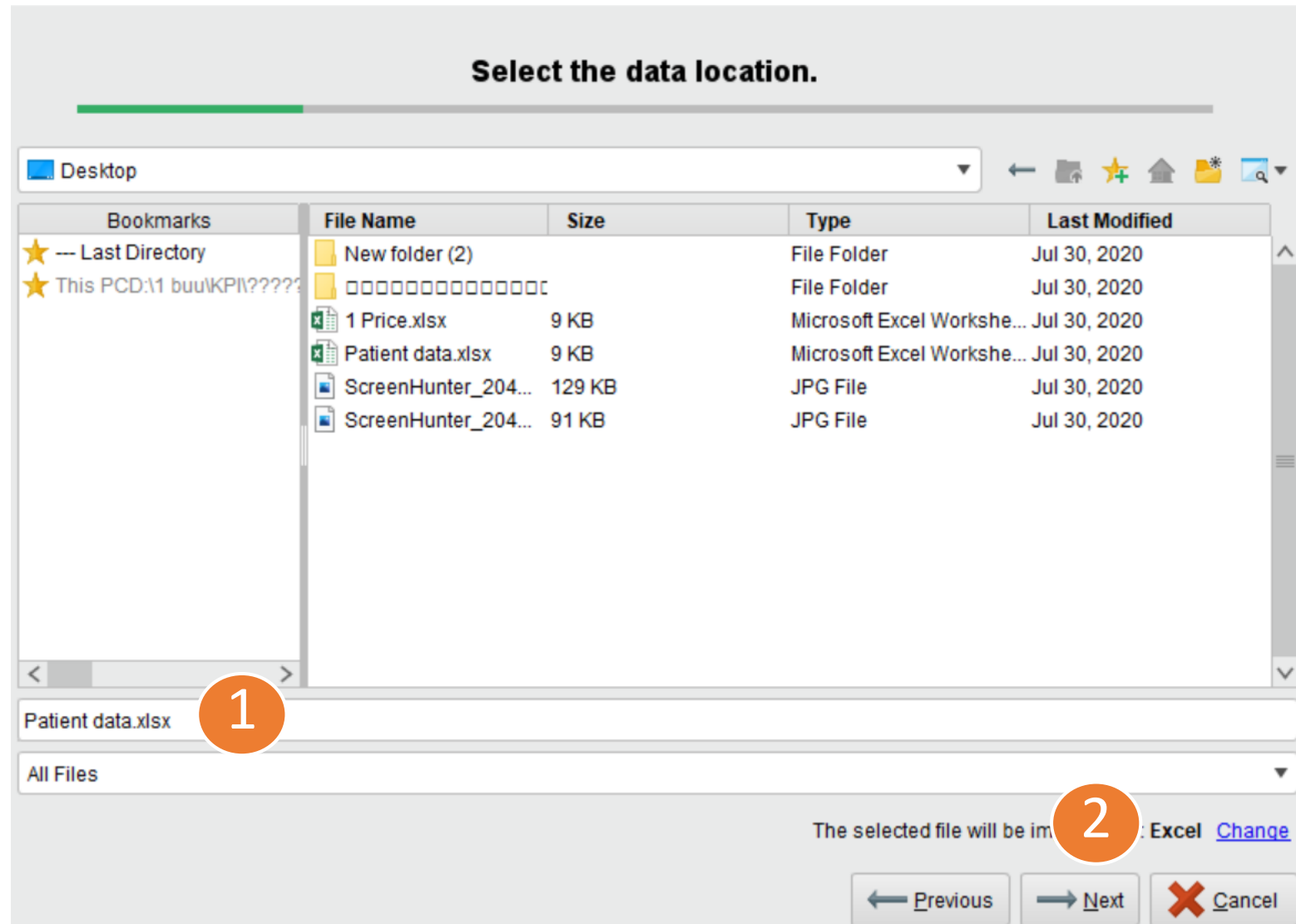
[Get support for more data sources from the RapidMiner Marketplace!](#)



การนำเข้าข้อมูลเข้า

- เลือกไฟล์ที่ต้องการวิเคราะห์

Import Data - Select the data location.





การนำเข้าข้อมูลเข้า

- 3 คือ เลือกข้อมูลว่าอยู่ใน Sheet ไດ
- 4 คือ เลือกช่วงของ column
- 5 คือ กำหนดว่าแถวใดเป็น header

Select the cells to import.

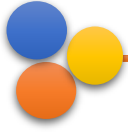
3

Sheet: **Sheet1** Cell range: **A:G** 4 **Select All** ☒ Define header row: **1** 5

	A	B	C	D	E	F	G
1	ID	Name	Weight	Height	Cholesterol	Blood sugar le...	Result
2	1.000	One	70.000	175.000	175.000	80.000	Good
3	2.000	Two	81.000	159.000	250.000	250.000	Bad
4	3.000	Three	78.000	180.000	180.000	95.000	Good
5	4.000	Four	85.000	174.000	300.000	300.000	Bad
6	5.000	Five	76.000	163.000	250.000	245.000	Bad

6

← Previous → Next ✕ Cancel



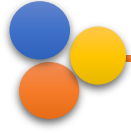
การนำข้อมูลเข้า

- กำหนดชนิดของข้อมูลให้เหมาะสำหรับการวิเคราะห์

Format your columns.

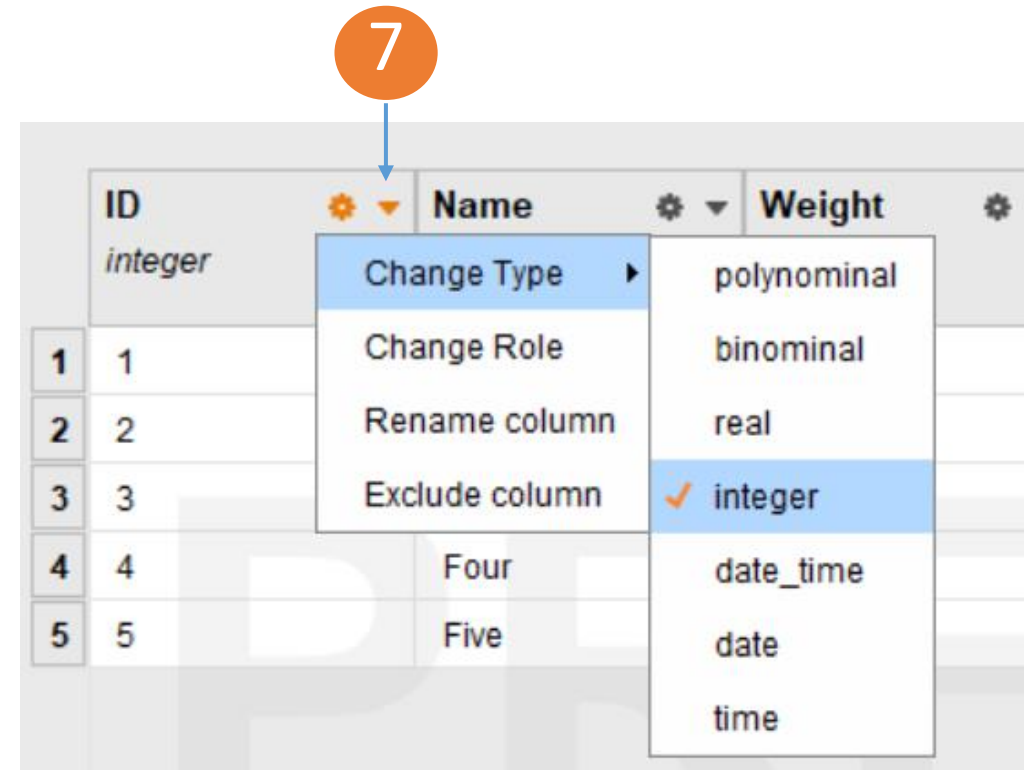
☐ Replace errors with missing values ⓘ

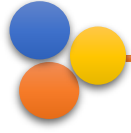
	ID <i>integer</i>	Name <i>polynomial</i>	Weight <i>integer</i>	Height <i>integer</i>	Cholesterol <i>integer</i>	Blood suga... <i>integer</i>	Resi <i>poly</i>
1	1	One	70	175	175	80	Goc
2	2	Two	81	159	250	250	Bad
3	3	Three	78	180	180	95	Goc
4	4	Four	85	174	300	300	Bad
5	5	Five	76	163	250	245	Bad



ชนิดของข้อมูลใน RapidMiner

- ชนิดของข้อมูล (Value type) คือ ค่าของข้อมูลที่เก็บในแต่ละแอตทริบิวต์ (Attribute)
 - Polynominal ข้อมูลประเภท Category (ข้อมูลที่ไม่ใช่ตัวเลข) มีค่ามากกว่า 2 ค่าขึ้นไป เช่น อาชีพ และศาสนา เป็นต้น
 - Binominal ข้อมูลประเภท Category (ข้อมูลที่ไม่ใช่ตัวเลข) มีค่าเพียง 2 ค่าเท่านั้น เช่น เพศ เป็นโรคหรือไม่เป็นโรค เป็นต้น
 - Real ทศนิยม
 - Integer ข้อมูลประเภทตัวเลขจำนวนเต็มบวก ลบ ศูนย์





ชนิดของข้อมูล

- คลิก 7 กำหนดให้ ID เป็นชนิด ID เนื่องจากเป็น Attribute ที่ไม่นำมาวิเคราะห์

☐ Replace errors with missing values

	ID	Name
1	1	
2	2	
3	3	
4	4	Four
5	5	Five

7

Change Type
Change Role
Rename column
Exclude column

Format your columns.

☐ Replace errors with missing values ⓘ

	ID	Name	Weight	Height	Cholesterol	Blood
1	1					
2	2					
3	3					
4	4					
5	5					

Change role

Please enter the new role:

id
label
id
weight

8

9 สีฟ้า

☐ Replace errors with

	ID	Name
1	1	One
2	2	Two
3	3	Three
4	4	Four
5	5	Five

9



การนำข้อมูลเข้า

Format your columns.

☐ Replace errors with missing values ⓘ

	Name <i>polynomial</i>	Weight <i>integer</i>	Height <i>integer</i>	Cholesterol <i>integer</i>	Blood suga... <i>integer</i>	Result <i>polynomial</i>
1	One	70	175	175	80	Good
2	Two	81	159	250	250	Bad
3	Three	78	180	180	95	Good
4	Four	85	174	300	300	Bad
5	Five	76	163	250	245	Bad

Change role

Please enter the new role:

label

OK Cancel

10



การนำข้อมูลเข้า

Format your columns.

☐ Replace errors with missing values ⓘ

	Name <i>polynomial</i>	Weight <i>integer</i>	Height <i>integer</i>	Cholesterol <i>integer</i>	Blood suga... <i>integer</i>	Result <i>polynomial label</i>
1	One	70	175	175	80	Good
2	Two	81	159	250	250	Bad
3	Three	78	180	180	95	Good
4	Four	85	174	300	300	Bad
5	Five	76	163	250	245	Bad

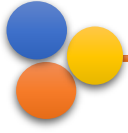
< ||| >

✓ no problems.

← Previous → Next ✗ Cancel

11

สี่เขียว



การนำข้อมูลเข้า

Where to store the data?

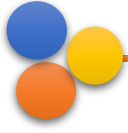
- ▶ Local Repository (Legacy)
- ▶ NewLocalRepository (Legacy)
- ▶ SCI (Legacy)
- ▶ SI (Local)
- ▶ Temporary Repository (Legacy)

Name

Location //SI/Patient data

12

Previous Finish Cancel



การนำข้อมูลเข้า

- เมื่อนำข้อมูลเข้าเรียบร้อยแล้ว

File Edit Process View Connections Settings Extensions Help

Design Results Turbo Prep All Studio

Result History ExampleSet (//SI/Patient data) ExampleSet (//SI/Patient data)

Open in Turbo Prep Auto Model Filter (5 / 5 examples): all

Row No.	ID	Result	Name	Weight	Height	Cholesterol	Blood sugar ...
1	1	Good	One	70	175	175	80
2	2	Bad	Two	81	159	250	250
3	3	Good	Three	78	180	180	95
4	4	Bad	Four	85	174	300	300
5	5	Bad	Five	76	163	250	245

Data

Statistics

Visualizations

เมื่อนำข้อมูลเข้าแล้ว สามารถดูสถิติได้

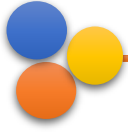
เช่น ค่า Max, Min, Average

1

Result History		ExampleSet (//SI/Patient data)		ExampleSet (//SI/Patient data)	
Name	Type	Missing	Statistics		Filter (7 / 7 attributes): <input type="text" value="Search for Attributes"/>
✓ ID	Integer	0	Min 1	Max 5	Average 3
✓ Result	Polynomial	0	Least Good (2)	Most Bad (3)	Values Bad (3), Good (2)
✓ Name	Polynomial	0	Least Two (1)	Most Five (1)	Values Five (1), Four (1), ...[3 more]
✓ Weight	Integer	0	Min 70	Max 85	Average 78
✓ Height	Integer	0	Min 159	Max 180	Average 170.200
✓ Cholesterol	Integer	0	Min 175	Max 300	Average 231
✓ Blood sugar level	Integer	0	Min 80	Max 300	Average 194

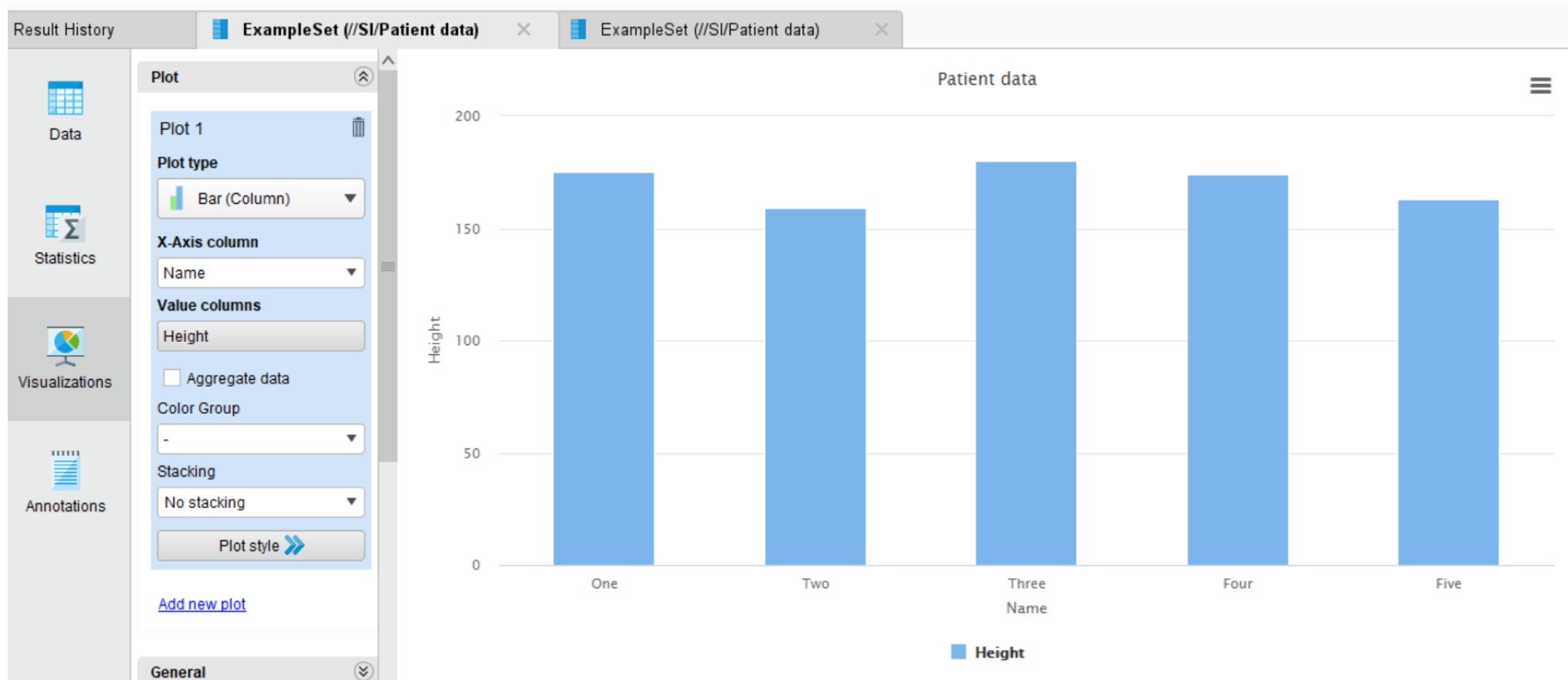
Showing attributes 1 - 7

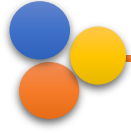
Examples: 5 Special Attributes: 2 Regular Attributes:



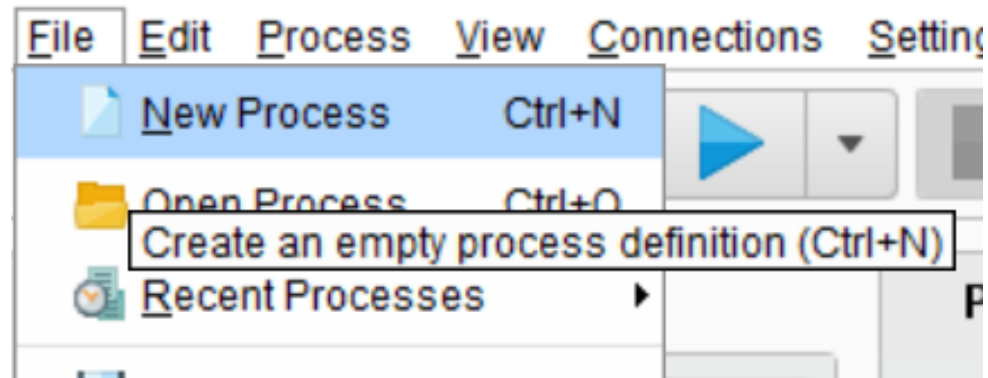
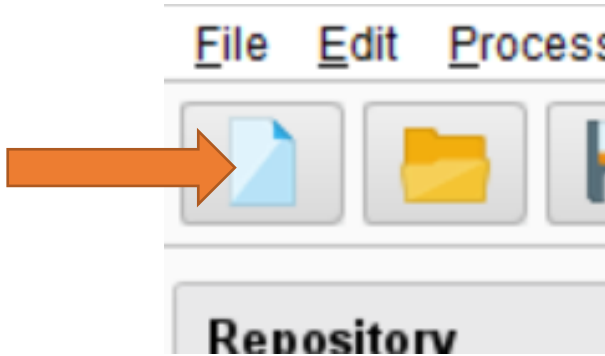
เมื่อนำข้อมูลเข้าแล้ว สามารถ Visualization ได้

- Visualization คือ นำข้อมูลดิบมาเปลี่ยนเป็นกราฟ แผนภูมิ หรือแม้กระทั่งวิดีโอที่ช่วยอธิบายปริมาณ หรือตัวเลข เพื่อให้เข้าใจได้ง่ายขึ้น

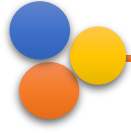




สร้าง process ใหม่



Euclidean Distance/k-NN Algorithm



Euclidean Distance/k-NN Algorithm

การนำข้อมูลทดสอบ ชื่อ Patient data test

ID	Name	Weight	Height	Cholesterol	Blood sugar level
1	A	45	150	199	72
2	B	56	175	284	141
3	C	68	162	158	124
4	D	81	179	248	210



k-NN

- นำข้อมูล Patient data test เข้า

Format your columns.

☐ Replace errors with missing values ⓘ

	ID <i>integer</i> <i>id</i>	Name <i>polynomial</i>	Weight <i>integer</i>	Height <i>integer</i>	Cholesterol <i>integer</i>	Blood sugar... <i>integer</i>
1	1	A	45	150	199	72
2	2	B	56	175	284	141
3	3	C	68	162	158	124
4	4	D	81	179	248	210

PREVIEW...

✓ no problems.

← Previous → Next ✗ Cancel



k-NN

File Edit Process view Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

Result History ExampleSet (/SI/Patient data test1) ExampleSet (Apply Model)

Open in Turbo Prep Auto Model

Row No.	ID	Name	Weight	Height	Cholesterol	Blood sugar ...
1	1	A	45	150	199	72
2	2	B	56	175	284	141
3	3	C	68	162	158	124
4	4	D	81	179	248	210

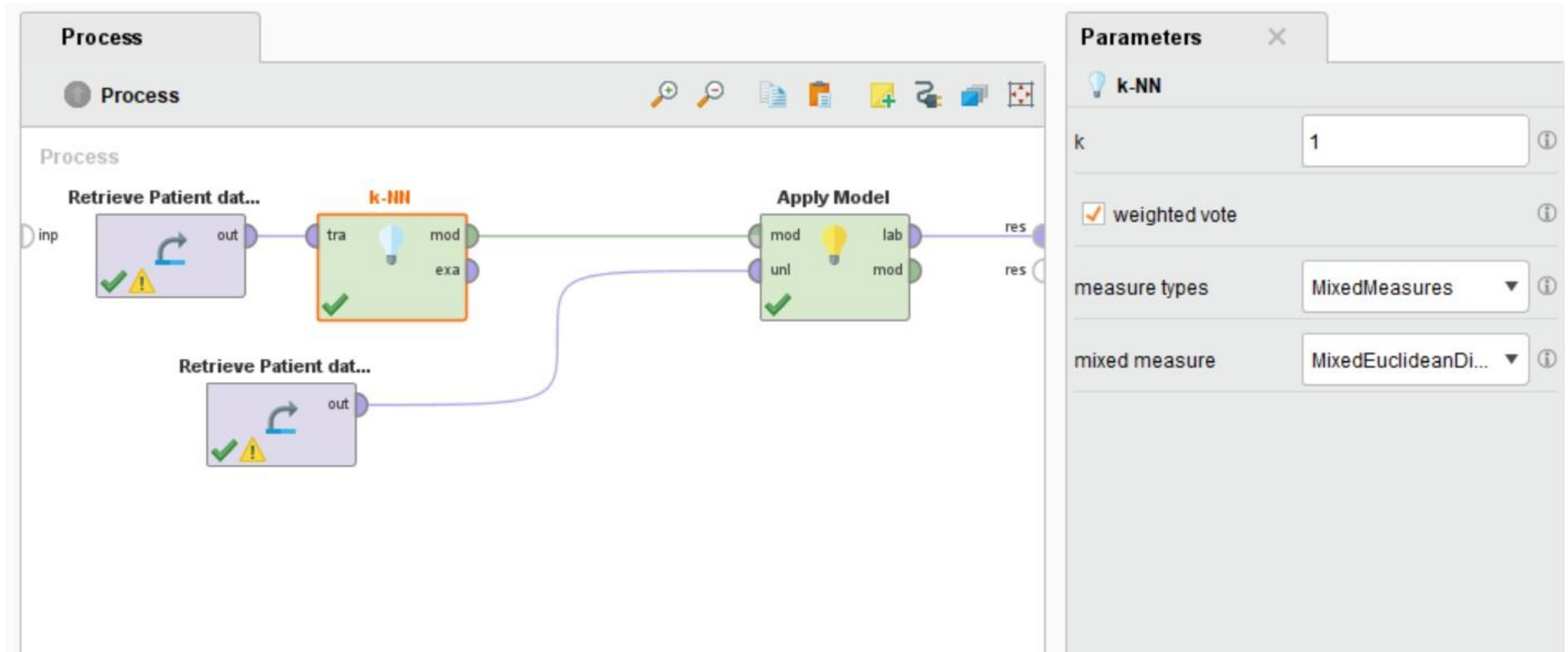
Data

Statistics

Visualizations



Annotations

ถ้าเลือก $k=1$



Result History

ExampleSet (Apply Model) X

Open in  Turbo Prep  Auto Model

Row No.	ID	prediction(R...
1	1	Good
2	2	Bad
3	3	Good
4	4	Bad

Data

Statistics

Visualizations



งาน

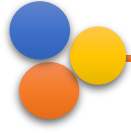
1. จากนั้นใช้ k -NN Algorithm, $k = 3$
2. จากนั้นใช้ k -NN Algorithm, $k = 5$

Linear Regression Analysis

คำถาม จงทำนายว่าเมื่อ Temperature = 10 แล้วจะมีความดันเท่าไร

ตาราง ข้อมูล Temperature Pressure อุณหภูมิและความดัน

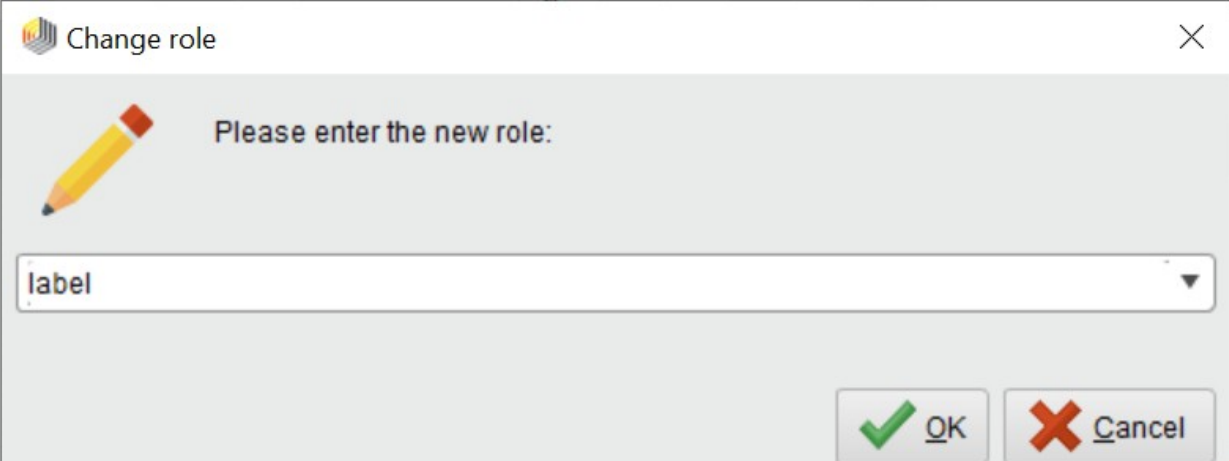
Temperature (x)	Pressure (y)
1	2
2	4
3	5
4	4
5	5
10	?



ข้อที่ 1 ข้อมูล Temperature Pressure

- นำข้อมูล train เข้าและเปลี่ยนให้ pressure เป็น label
- จากนั้น นำข้อมูล test เข้า

Temperature (x) <i>integer</i>		Pressure (y) <i>integer</i>
1	1	2
2	2	4
3	3	5
4	4	4
5	5	5



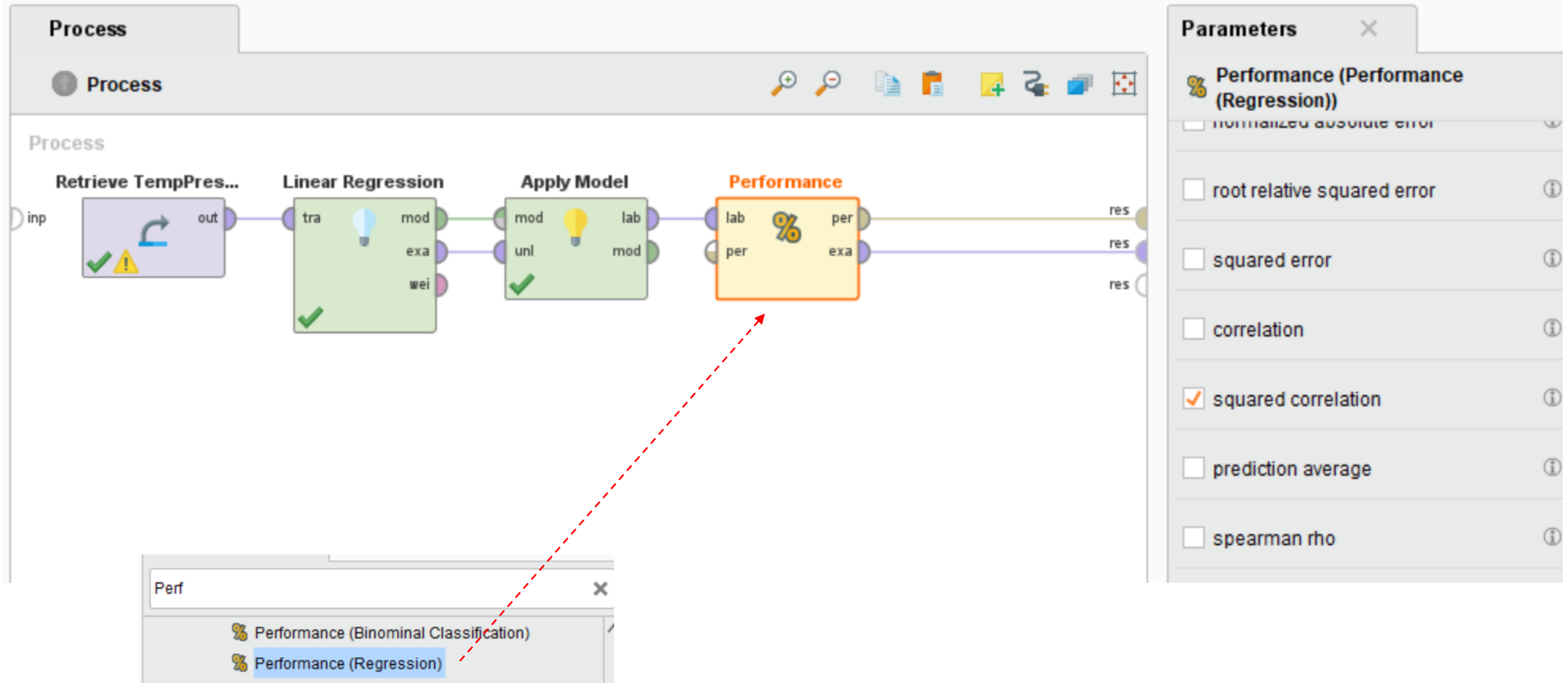
Change role

Please enter the new role:

label

OK Cancel

ข้อที่ 1 ข้อมูล Temperature Pressure



ข้อที่ 1 ข้อมูล Temperature Pressure



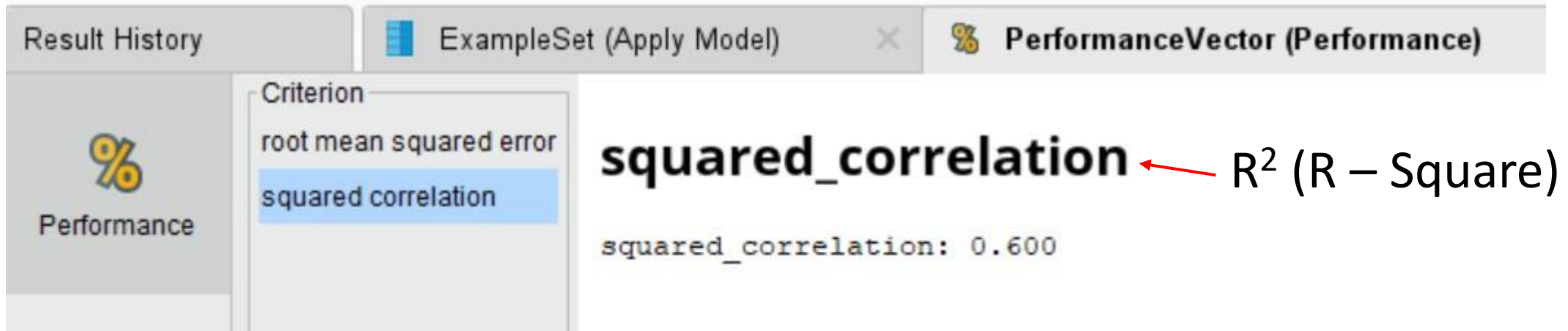
Result History ExampleSet (Apply Model) PerformanceVector (Performance)

Criterion

- root mean squared error
- squared correlation

root_mean_squared_error

root_mean_squared_error: 0.693 +/- 0.000



Result History ExampleSet (Apply Model) PerformanceVector (Performance)

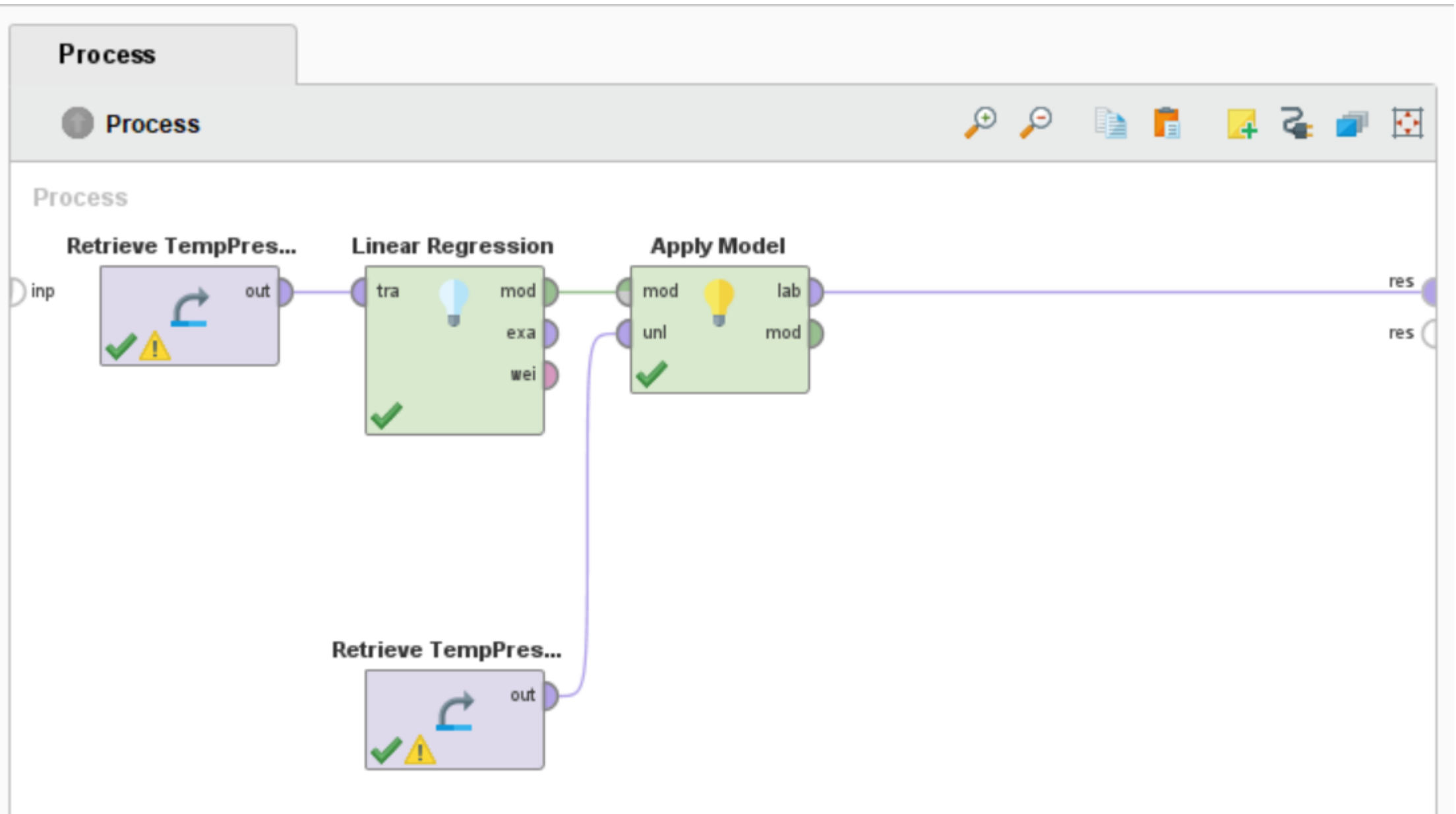
Criterion

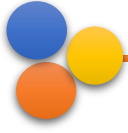
- root mean squared error
- squared correlation

squared_correlation ← R^2 (R – Square)



squared_correlation: 0.600

Testing ข้อมูล Temperature Pressure

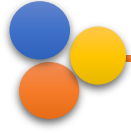




ผลลัพธ์ของค่าความดัน

Result History		
Open in		
 Turbo Prep		
 Auto Model		
Row No.	prediction(P...	Temperatur...
1	8.200	10

ข้อมูลสำหรับวิเคราะห์ Multiple linear regression



ตารางที่ 2 ข้อมูลสำหรับวิเคราะห์ Multiple linear regression

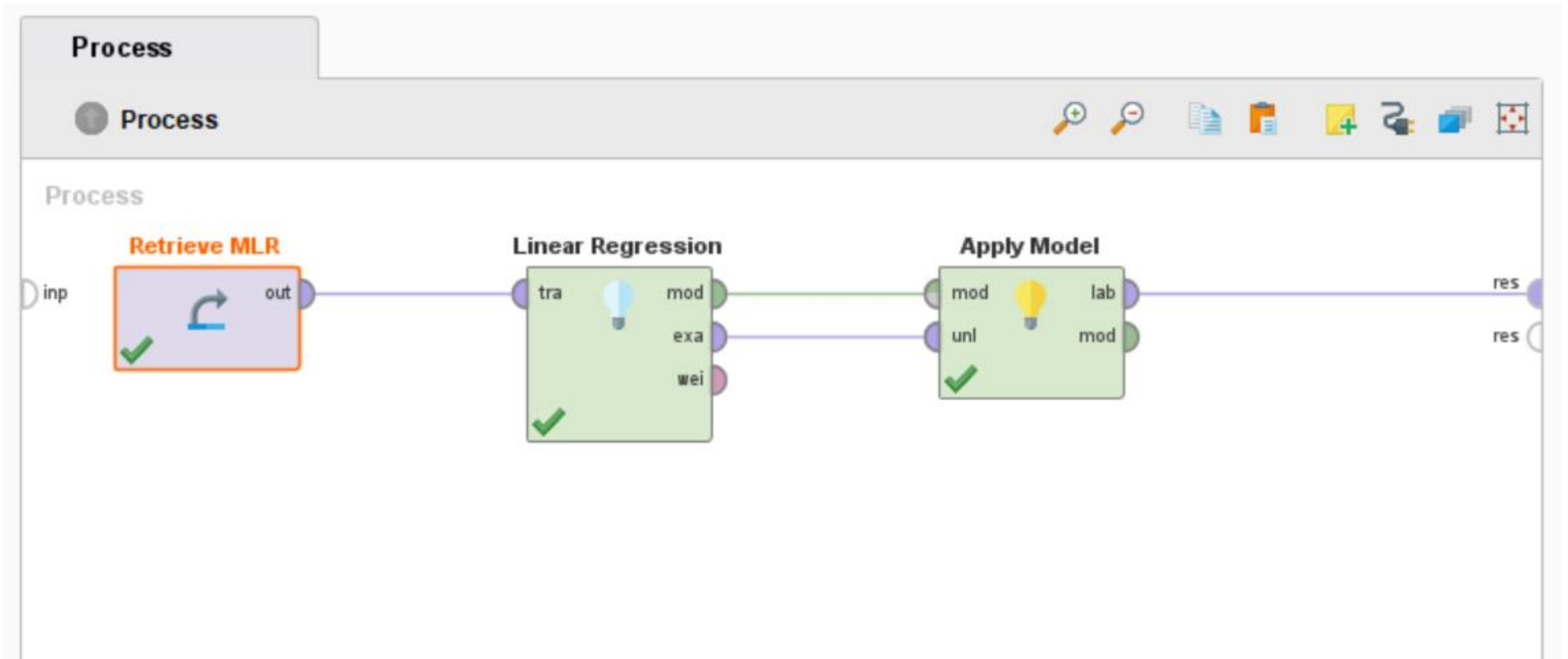
- Training

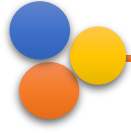
Weight	Height	Blood sugar level	Cholesterol
70.2	175.9	80.36	175.2
81.3	159.3	250.3	250.1
78.3	180.3	95.6	180.3
85.2	174.1	300.4	300.0
76.1	163.3	245.65	250.7

- Test

Weight	Height	Blood sugar level	Cholesterol
70.2	175.9	80.36	?
81.3	159.3	250.3	?
78.3	180.3	95.6	?
85.2	174.1	300.4	?
76.1	163.3	245.65	?


การวิเคราะห์ Multiple linear regression







วิเคราะห์ Multiple linear regression


ExampleSet (Apply Model) ✕



Data


Statistics


Visualizations

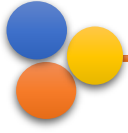
Open in

 Turbo Prep

 Auto Model

Row No.	Cholesterol	prediction(C...	Weight	Height	Blood sugar ...
1	175.200	170.430	70.200	175.900	80.360
2	250.100	250.194	81.300	159.300	250.300
3	180.300	184.937	78.300	180.300	95.600
4	300	298.301	85.200	174.100	300.400
5	250.700	252.439	76.100	163.300	245.650

k-Means Clustering Algorithm



กำหนดให้มีข้อมูล Customer ดังตารางข้างล่างนี้ ให้แบ่งกลุ่มข้อมูลออกเป็น 3 กลุ่ม

Name	รายได้ (พันบาทต่อเดือน)	จำนวนหนี้ (พันบาทต่อเดือน)	จำนวนสมาชิก ที่ต้องดูแล
A	25	1	2
V	80	15	8
X	6	1	5
T	8	2	6
G	30	0	9
R	27	3	2
Y	50	12	10

กลุ่มที่ 1 คือใครบ้าง.....

กลุ่มที่ 2 คือใครบ้าง.....

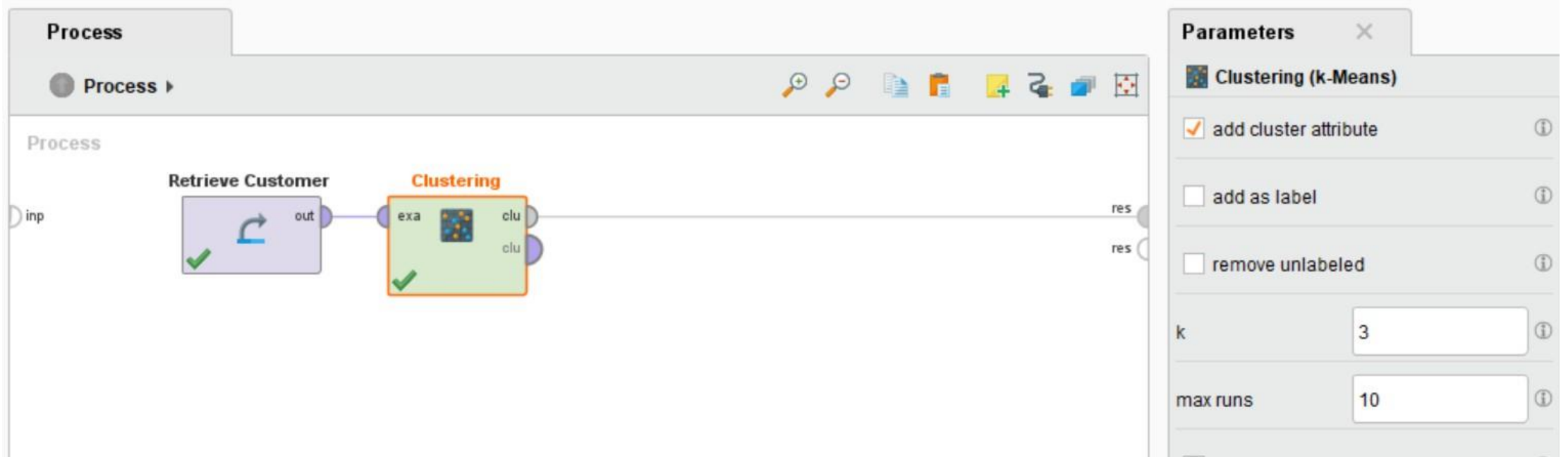
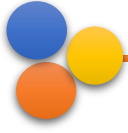
กลุ่มที่ 3 คือใครบ้าง.....

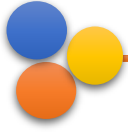


Format your columns.

☐ Replace errors with missing values ⓘ

	Name <i>polynomial id</i>	▼ in <i>integer</i>	▼ out <i>integer</i>	▼ care <i>integer</i>
1	A	25	1	2
2	V	80	15	8
3	X	6	1	5
4	T	8	2	6
5	G	30	0	9
6	R	27	3	2
7	Y	50	12	10





Result History

Cluster Model (Clustering)

Cluster Model

Cluster 0: 5 items
Cluster 1: 1 items
Cluster 2: 1 items
Total number of items: 7

Description

Folder View

Graph

Result History

Cluster Model (Clustering)

Description

Folder View

Graph

root

cluster_0

- A
- X
- T
- G
- R

cluster_1

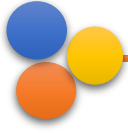
- V

cluster_2

- Y

Artificial Neural Networks

โครงข่ายประสาทเทียม



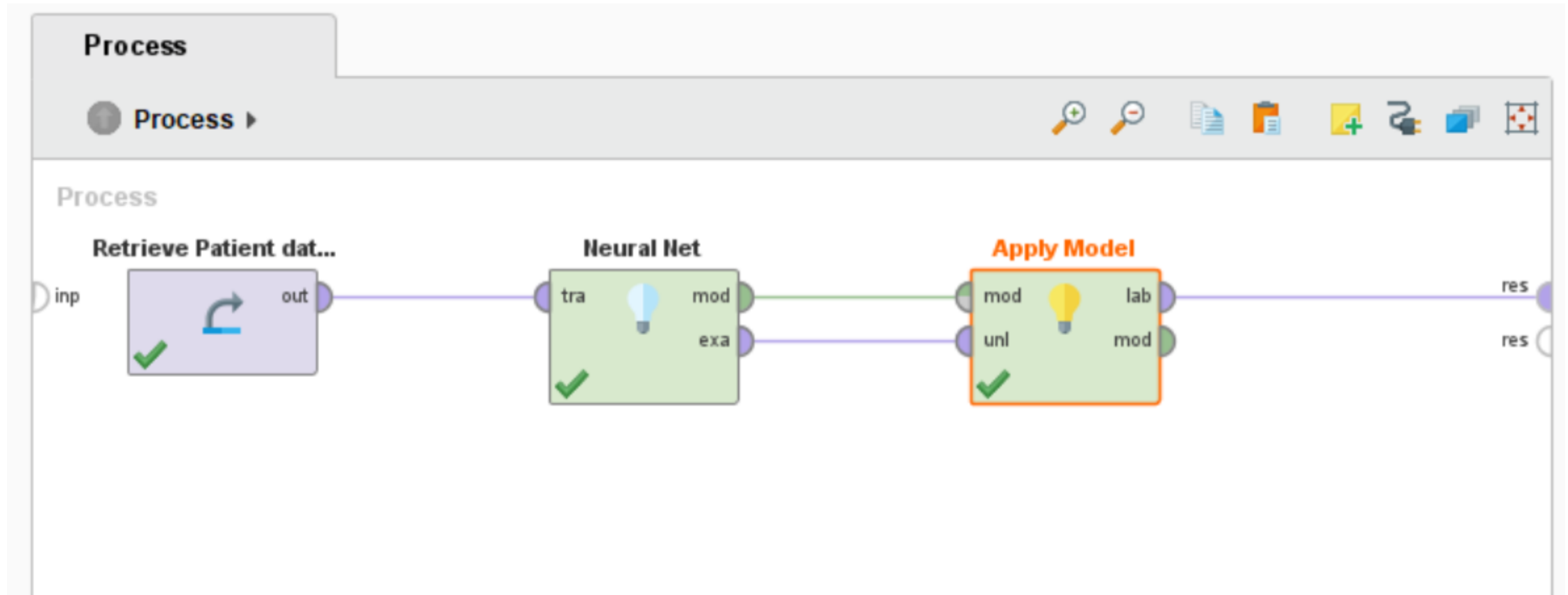
ตารางที่ 1 ประวัติคนไข้ (Patient data)

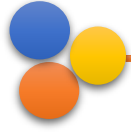
ID	Name	Weight	Height	Cholesterol	Blood sugar level	Result
1	One	70	175	175	80	Good
2	Two	81	159	250	250	Bad
3	Three	78	180	180	95	Good
4	Four	85	174	300	300	Bad
5	Five	76	163	250	245	Bad

ตารางที่ 2 ประวัติคนไข้



ID	Name	Weight	Height	Cholesterol	Blood sugar level
1	A	45	150	199	72
2	B	56	175	284	141
3	C	68	162	158	124
4	D	81	179	248	210

เมื่อใช้ข้อมูลตารางที่ 1 และ 2

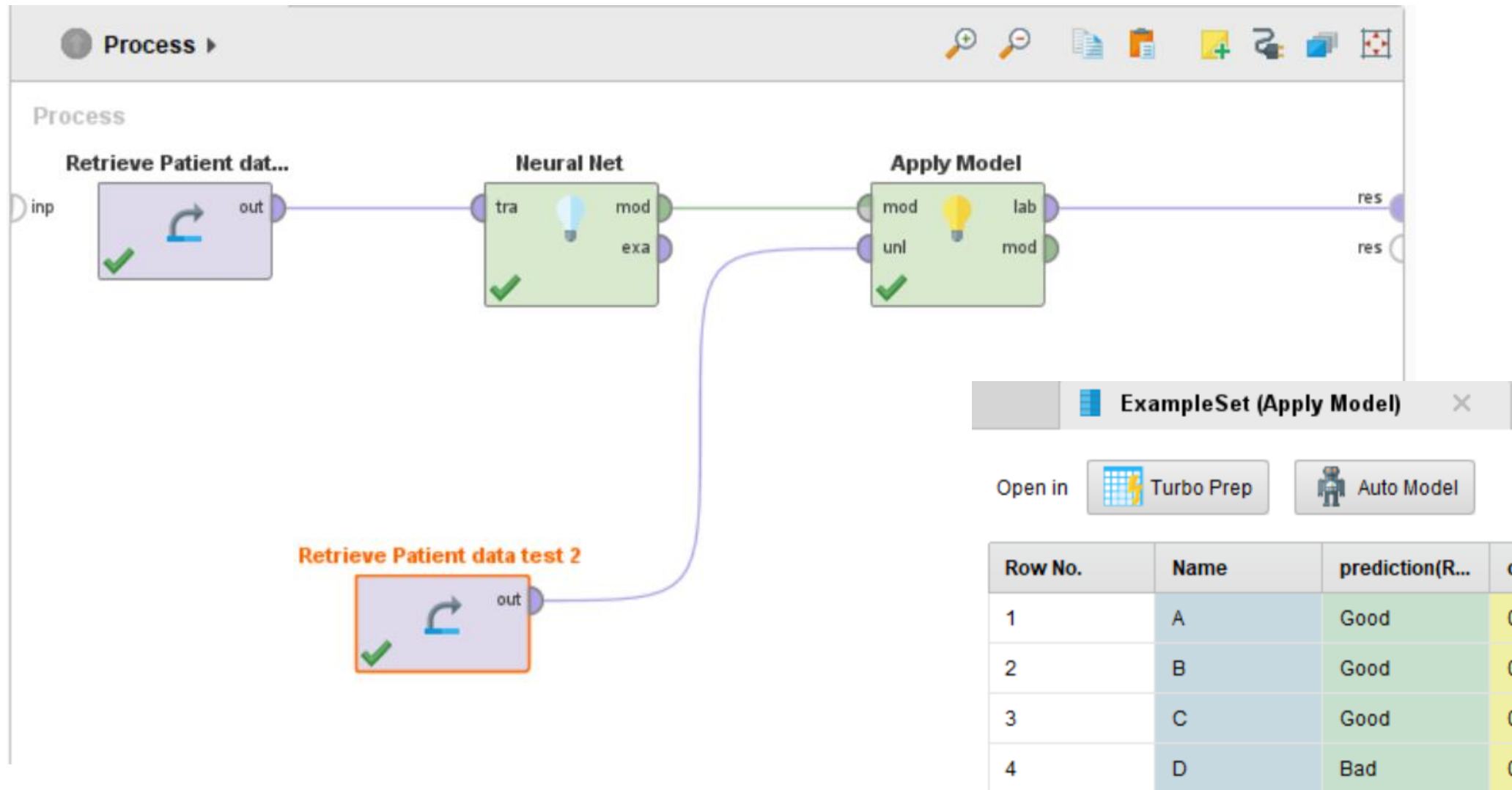




เมื่อใช้ข้อมูลตารางที่ 1 และ 2

ExampleSet (Apply Model) X			
Open in  Turbo Prep  Auto Model			
Row No.	Name	Result	prediction(R...
1	One	Good	Good
2	Two	Bad	Bad
3	Three	Good	Good
4	Four	Bad	Bad
5	Five	Bad	Bad

เมื่อใช้ข้อมูลตารางที่ 1 และ 2 กรณี test data



เมื่อใช้ Neural Net กับข้อมูล TempPressure

The screenshot displays a software interface for a neural network workflow. The main workspace shows a sequence of four blocks: 'Retrieve TempPres...', 'Neural Net', 'Apply Model', and 'Performance'. The 'Neural Net' block is highlighted with an orange border. To the right, a 'Parameters' panel for the 'Neural Net' block shows settings: 'hidden layers' (with an 'Edit List (0)...' button), 'training cycles' (1000), 'learning rate' (0.01), 'momentum' (0.9), and a 'decay' checkbox. Below the main workspace, there are two 'PerformanceVector (Performance)' panels. The left panel shows the 'root_mean_squared_error' criterion with a value of 0.381 +/- 0.000. The right panel shows the 'squared_correlation' criterion with a value of 0.879. A 'Result History' panel on the far left shows a 'Performance' icon.

Process

Process

Retrieve TempPres... Neural Net Apply Model Performance

Parameters

Neural Net

hidden layers Edit List (0)...

training cycles 1000

learning rate 0.01

momentum 0.9

☐ decay

Result History

Performance

PerformanceVector (Performance)

Criterion

root mean squared error

squared correlation

root_mean_squared_error

root_mean_squared_error: 0.381 +/- 0.000

Performance

PerformanceVector (Performance)

Criterion

root mean squared error

squared correlation

squared_correlation

squared_correlation: 0.879