Lab 10

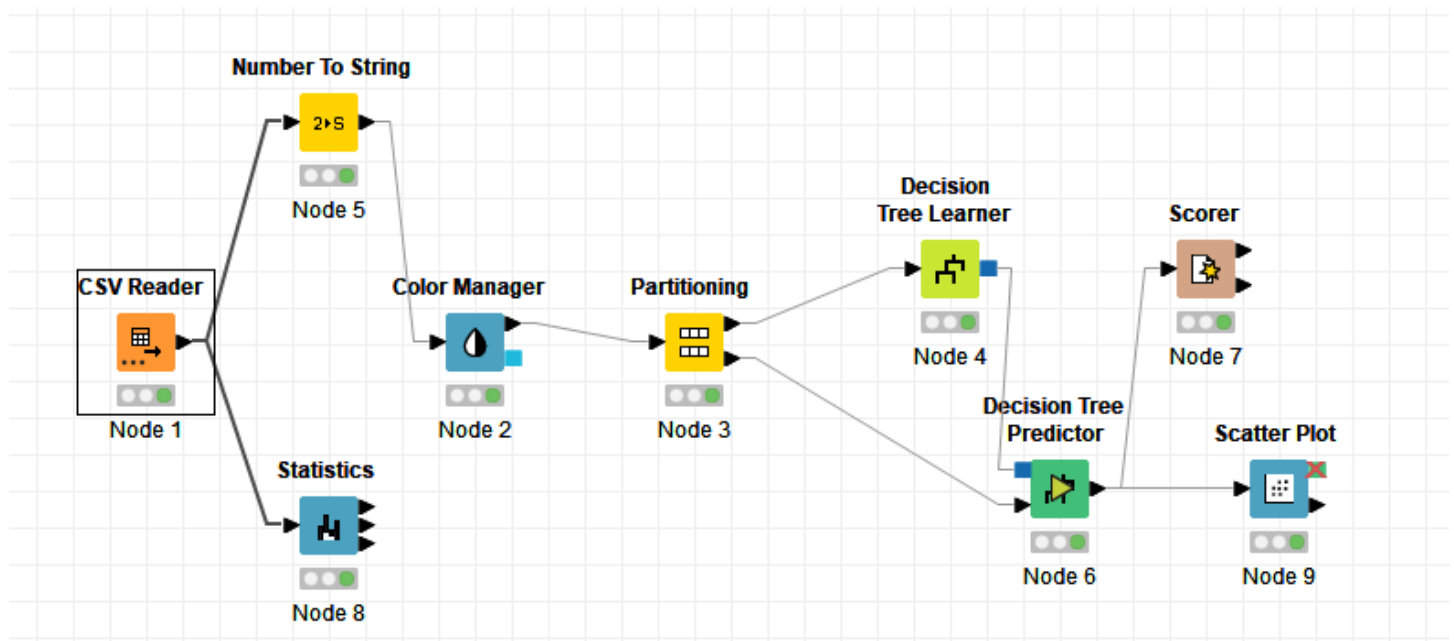## Task:

Health Care Machine Learning Supervised Learning in Knime Analytics Platform.

I'll be using heart dataset from UCI:

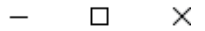https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data

Flow:

CSV Reader:

Statistics Node:

File

Options Histogram Flow Variables Memory Policy

☐ Calculate median values (computationally expensive)

Nominal values

⦿ Manual Selection ◯ Wildcard/Regex Selection ◯ Type Selection

Exclude

🔽 Filter

| D | oldpeak |

◯ Enforce exclusion

Include

🔽 Filter

I	ï»¿age
I	sex
I	cp
I	trestbps
I	chol
I	fbs
I	restecg
I	thalach
I	exang
I	slope

⦿ Enforce inclusion

Max no. of most frequent and infrequent values (in view): 20 ⬍

Max no. of possible values per column (in output table): 1,000 ⬍

☐ Enable HiLite

Number To String Node:

Color Manager Node:

File

Color Settings   Flow Variables   Memory Policy

Select one Column

S target ⌄

⦿ Nominal   ○ Range

■ 0
■ 1

Preview

Palettes   Swatches   HSV   HSL   RGB   CMYK   Alpha

○ Set 1

○ Set 2

○ Set 3 (colorblind safe)

⦿ Custom
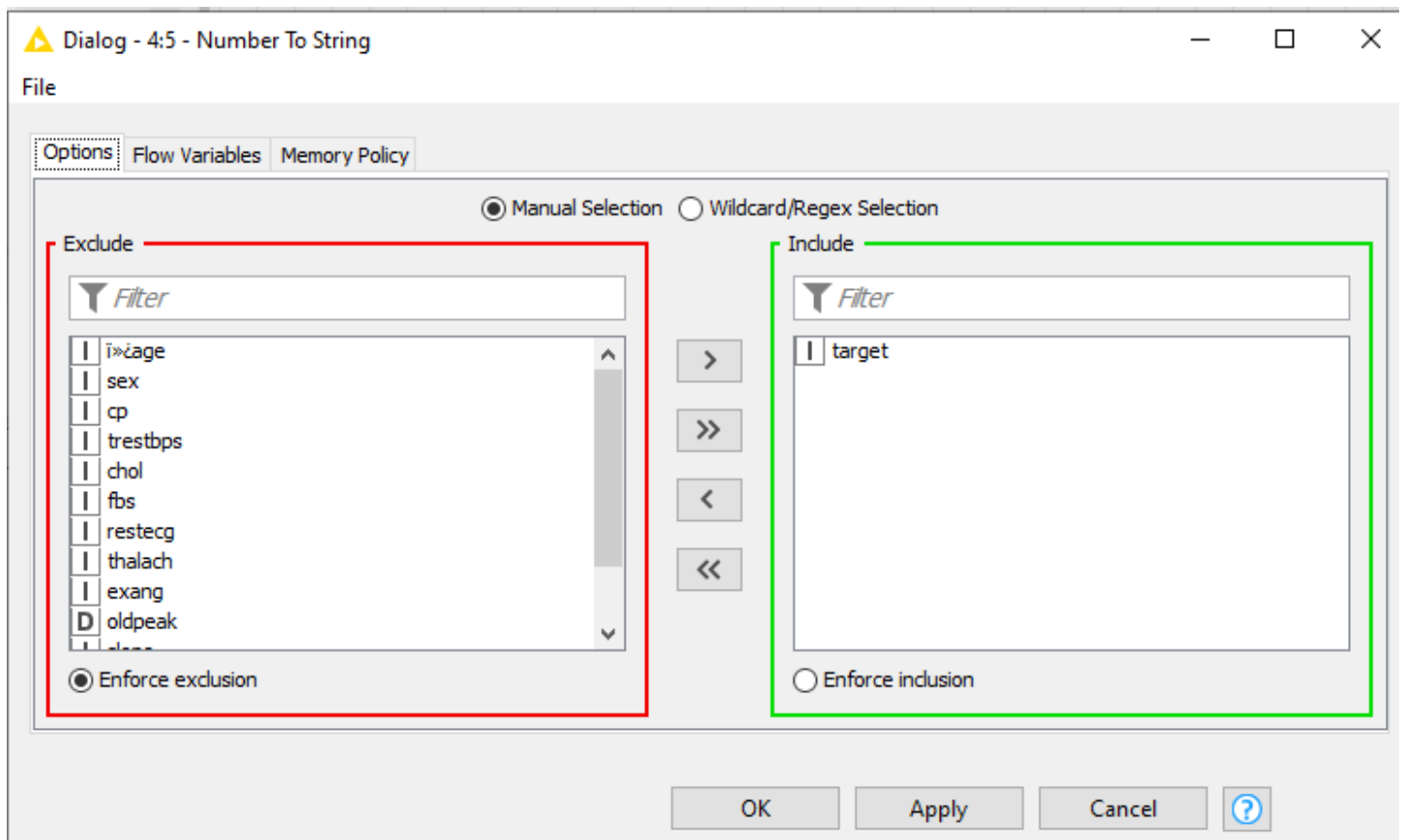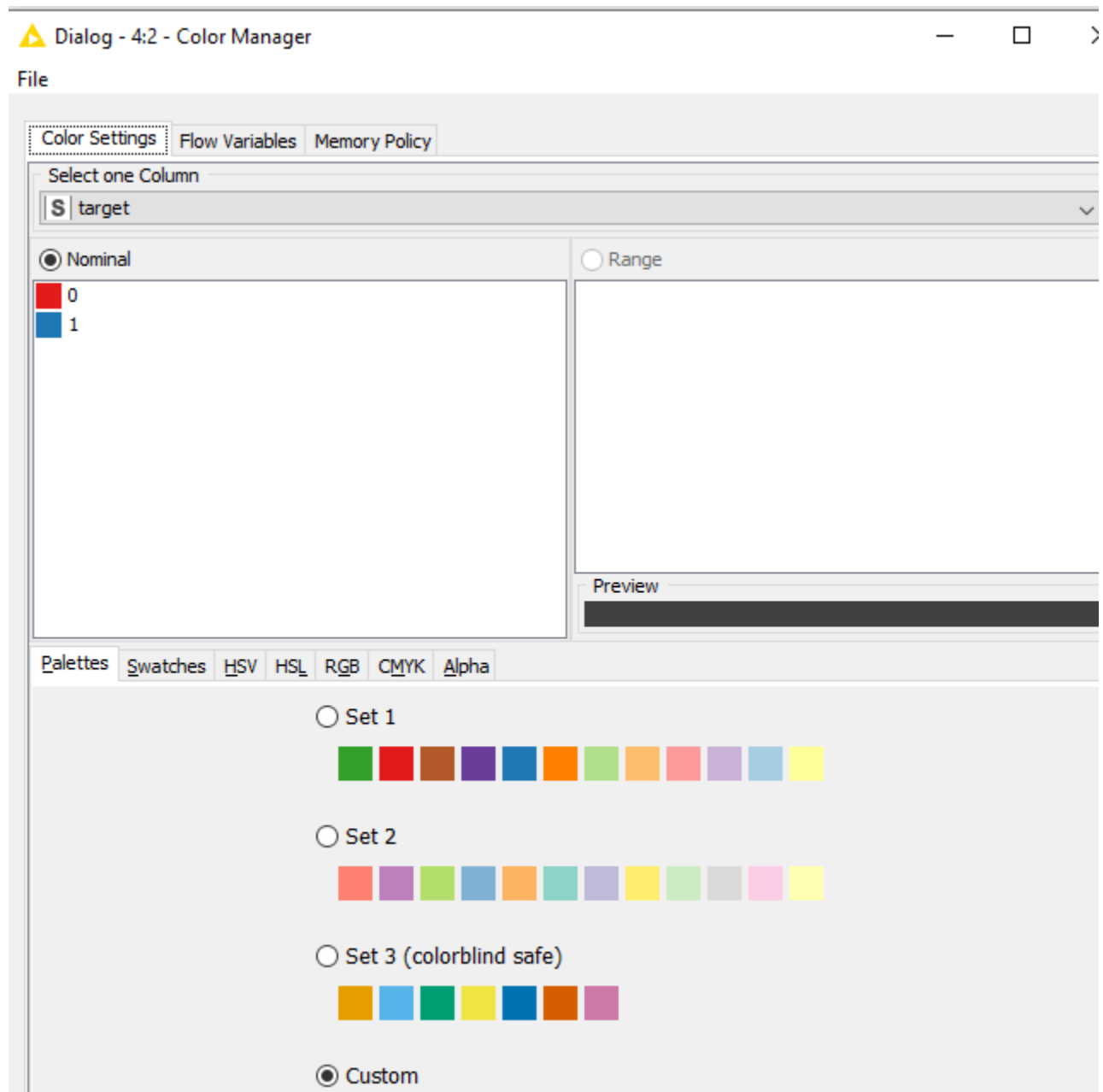
Color Manager Node Output:

Table with Colors - 4:2 - Color Manager

File Edit Hilite Navigation View

Table "default" - Rows: 303 | Spec - Columns: 14 | Properties | Flow Variables

Row ID	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	
Row153	66	0	2	146	278	0	0	152	0	0	1
Row154	39	0	2	138	220	0	1	152	0	0	1
Row155	58	0	0	130	197	0	1	131	0	0.6	1
Row156	47	1	2	130	253	0	1	179	0	0	2
Row157	35	1	1	122	192	0	1	174	0	0	2
Row158	58	1	1	125	220	0	1	144	0	0.4	1
Row159	56	1	1	130	221	0	0	163	0	0	2
Row160	56	1	1	120	240	0	1	169	0	0	0
Row161	55	0	1	132	342	0	1	166	0	1.2	2
Row162	41	1	1	120	157	0	1	182	0	0	2
Row163	38	1	2	138	175	0	1	173	0	0	2
Row164	38	1	2	138	175	0	1	173	0	0	2
Row165	67	1	0	160	286	0	0	108	1	1.5	1
Row166	67	1	0	120	229	0	0	129	1	2.6	1
Row167	62	0	0	140	268	0	0	160	0	3.6	0
Row168	63	1	0	130	254	0	0	147	0	1.4	1
Row169	53	1	0	140	203	1	0	155	1	3.1	0
Row170	56	1	2	130	256	1	0	142	1	0.6	1
Row171	48	1	1	110	229	0	1	168	0	1	0
Row172	58	1	1	120	284	0	0	160	0	1.8	1
Row173	58	1	2	132	224	0	0	173	0	3.2	2
Row174	60	1	0	130	206	0	0	132	1	2.4	1
Row175	40	1	0	110	167	0	0	114	1	2	1
Row176	60	1	0	117	230	1	1	160	1	1.4	2

Partitioning Node:

Dialog - 4:3 - Partitioning

File

First partition | Flow Variables | Memory Policy

Choose size of first partition

- ○ Absolute   100
- ● Relative[%]   80

- ○ Take from top
- ○ Linear sampling
- ○ Draw randomly
- ● Stratified sampling   S target ⌄

- ☐ Use random seed   1,619,987,652,4

OK   Apply   Cancel   ?

Decision Tree Learner Node:

Dialog - 4:4 - Decision Tree Learner

File

Options | PMMLSettings | Flow Variables

General

Class column | S | target ⌄

Quality measure | Gini index ⌄

Pruning method | No pruning ⌄

☑ Reduced Error Pruning

Min number records per node | 6 ⏶⏷

Number records to store for view | 10,000 ⏶⏷

☑ Average split point

Number threads | 4 ⏶⏷

☑ Skip nominal columns without domain information

Root split

☐ Force root split column
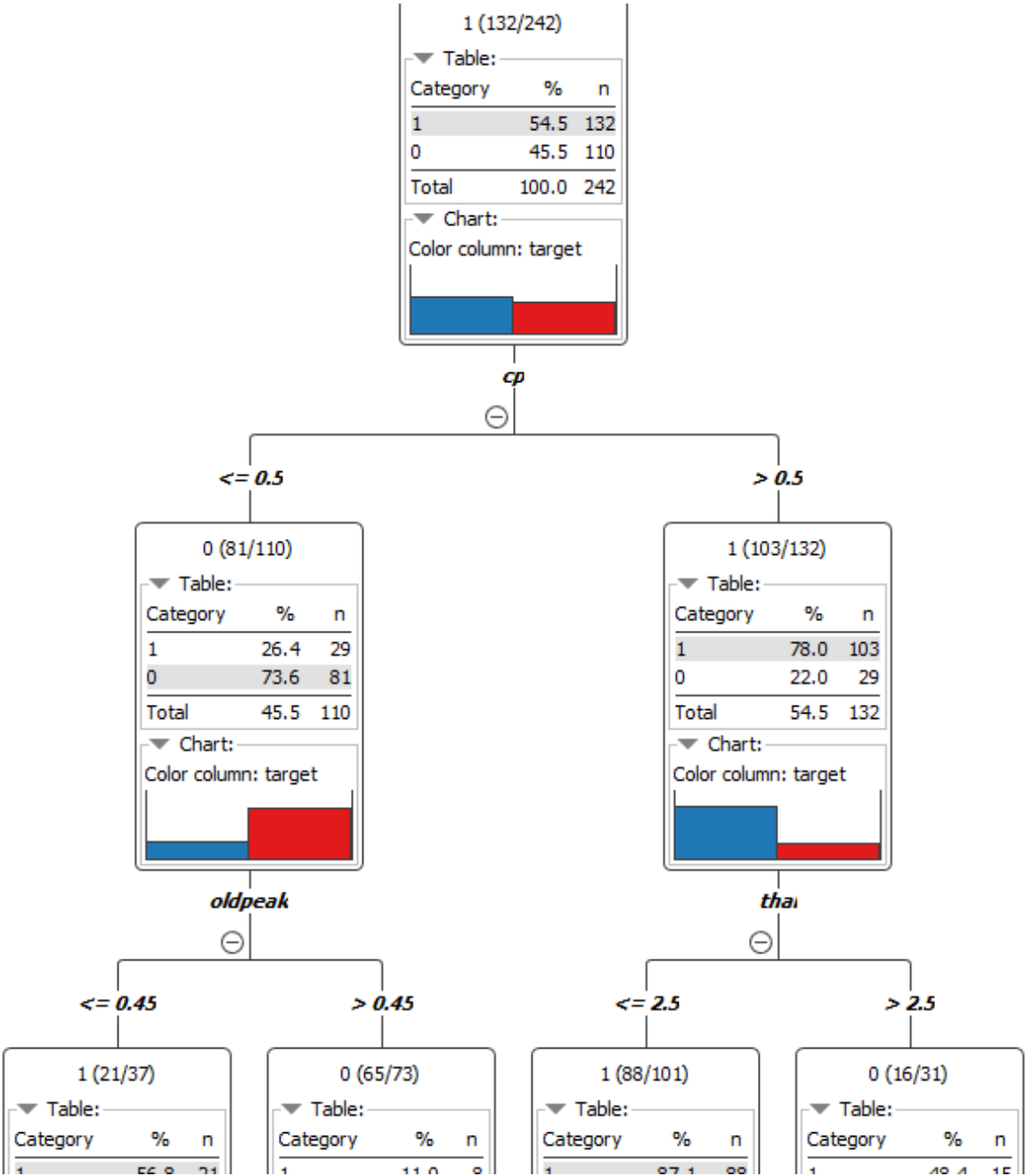
Root split column | I | thal ⌄

Binary nominal splits

☑ Binary nominal splits

Max #nominal | 10 ⏶⏷

☐ Filter invalid attribute values in child nodes

Decision Tree Learner Node Output:

1 (132/242)

▼ Table:

Category	%	n
1	54.5	132
0	45.5	110
Total	100.0	242

▼ Chart:
Color column: target

cp

⊖

<= 0.5

> 0.5

0 (81/110)

▼ Table:

Category	%	n
1	26.4	29
0	73.6	81
Total	45.5	110

▼ Chart:
Color column: target

1 (103/132)

▼ Table:

Category	%	n
1	78.0	103
0	22.0	29
Total	54.5	132

▼ Chart:
Color column: target

oldpeak

⊖

thal

⊖

<= 0.45

> 0.45

<= 2.5

> 2.5

1 (21/37)

▼ Table:

Category	%	n
1	56.8	21

0 (65/73)

▼ Table:

Category	%	n
1	11.0	8

1 (88/101)

▼ Table:

Category	%	n
1	87.1	88

0 (16/31)

▼ Table:

Category	%	n
1	48.4	15

Decision Tree Predictor Node:

Dialog - 4:6 - Decision Tree Predictor — □ ✕

File

Options | Flow Variables | Memory Policy

Maximum number of stored patterns for HiLite-ing: [10,000 ▲▼]

☐ Change prediction column name

[Prediction (target)]

☑ Append columns with normalized class distribution

Suffix for probability columns [ ]

Decision Tree Predictor Node Output:

Classified Data - 4:6 - Decision Tree Predictor — □ ✕

File Edit Hilite Navigation View

Table "default" - Rows: 61 | Spec - Columns: 17 | Properties | Flow Variables

Row ID	g	thalach	exang	oldpeak	slope	ca	thal	target	P (targ...	P (targ...	Predicti...
Row0		150	0	2.3	0	0	1	1	0.956	0.044	1
Row1		187	0	3.5	0	0	2	1	0.956	0.044	1
Row4		163	1	0.6	2	0	2	1	0.714	0.286	1
Row11		139	0	0.2	2	0	2	1	0.956	0.044	1
Row13		144	1	1.8	1	0	2	1	0.956	0.044	1
Row20		161	0	0.5	1	0	3	1	0	1	0
Row22		178	0	0	2	0	2	1	0.773	0.227	1
Row34		125	1	1.4	2	1	2	1	0.933	0.067	1
Row45		172	0	0.2	2	0	2	1	0.956	0.044	1
Row52		146	0	1.8	1	3	3	1	0.316	0.684	0
Row53		175	0	0.6	1	0	2	1	0.956	0.044	1
Row56		186	0	0	2	0	2	1	0.773	0.227	1
Row58		174	0	0	2	0	2	1	0.956	0.044	1
Row79		154	1	0.6	1	0	3	1	0.316	0.684	0
Row83		178	0	1.2	1	0	3	1	0.316	0.684	0
Row84		122	0	0.6	1	0	2	1	0.15	0.85	0
Row87		156	0	0	2	0	3	1	0.75	0.25	1
Row88		158	0	1.6	1	0	2	1	0.956	0.044	1
Row89		122	0	1	1	0	2	1	0.15	0.85	0
Row101		145	0	4.2	0	0	3	1	0.316	0.684	0
Row104		163	0	0	2	0	2	1	0.956	0.044	1
Row105		115	0	1.5	1	0	2	1	0.956	0.044	1
Row110		154	1	0	2	0	2	1	0.773	0.227	1
Row111		173	0	0.2	2	1	3	1	0.75	0.25	1

Scorer Node:

Scorer Node Output:

Row ID	ePo...	I FalsePo...	I TrueNe...	I FalseN...	D Recall	D Precision	D Sensitivity	D Specificity	D F-meas...	D Accuracy	D Cohen'...
1		6	22	10	0.697	0.793	0.697	0.786	0.742	?	?
0		10	23	6	0.786	0.688	0.786	0.697	0.733	?	?
Overall		?	?	?	?	?	?	?	?	0.738	0.478

Scatter Plot Node:

| Options | Axis Configuration | General Plot Options | View Controls | Flow Variables | Memory Policy |

Create image at outport

Maximum number of rows: 2,500

Selection column name: Selected (Scatter Plot)

Choose column for x axis

| I | ï»¿age

Choose column for y axis

| I | trestbps

Report on missing values

Scatter Plot Output: