Week 1, video 4:

Classifying in RapidMiner 5.3

Hands-On Activity

- Running algorithm in RapidMiner 5.3

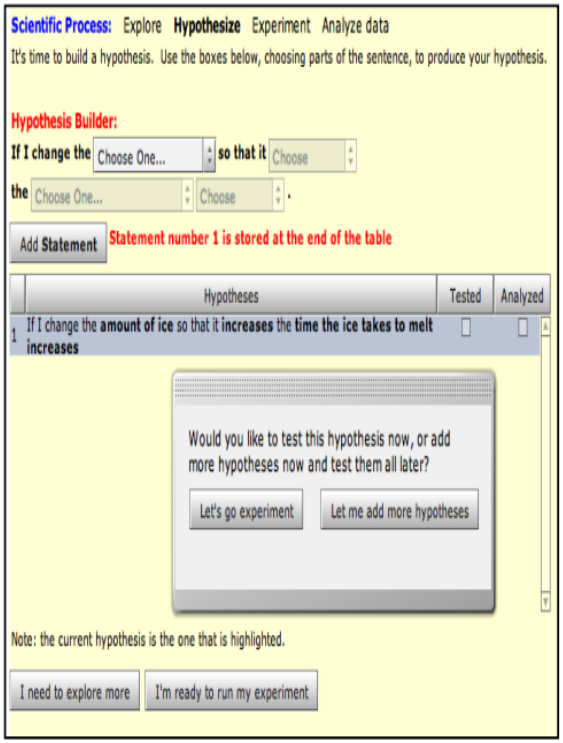
- Follow along on your own
  - Data set is on Coursera
  - SaoPedroetal(2013)_UMUAI_DesigningControlledExperiments_cummandlocalfeatures.csv

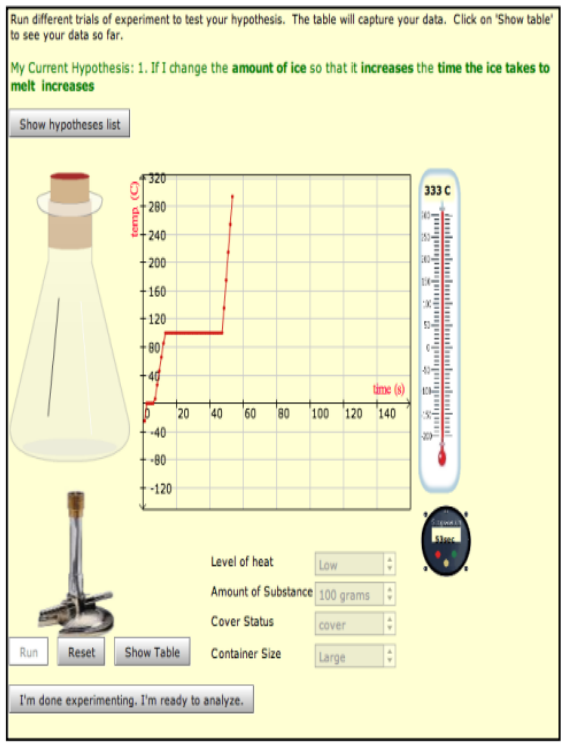
Data Comes From

- Sao Pedro, Baker, Gobert, Montalvo, & Nakama (2013) Leveraging Machine-Learned Detectors of Systematic Inquiry Behavior to Estimate and Predict Transfer of Inquiry Skill. *User Modeling and User-Adapted Interaction, 23* (1), 1-39.

Data Comes From

- Predicting whether students correctly design controlled experiments when learning in a science inquiry microworld, Inq-ITS

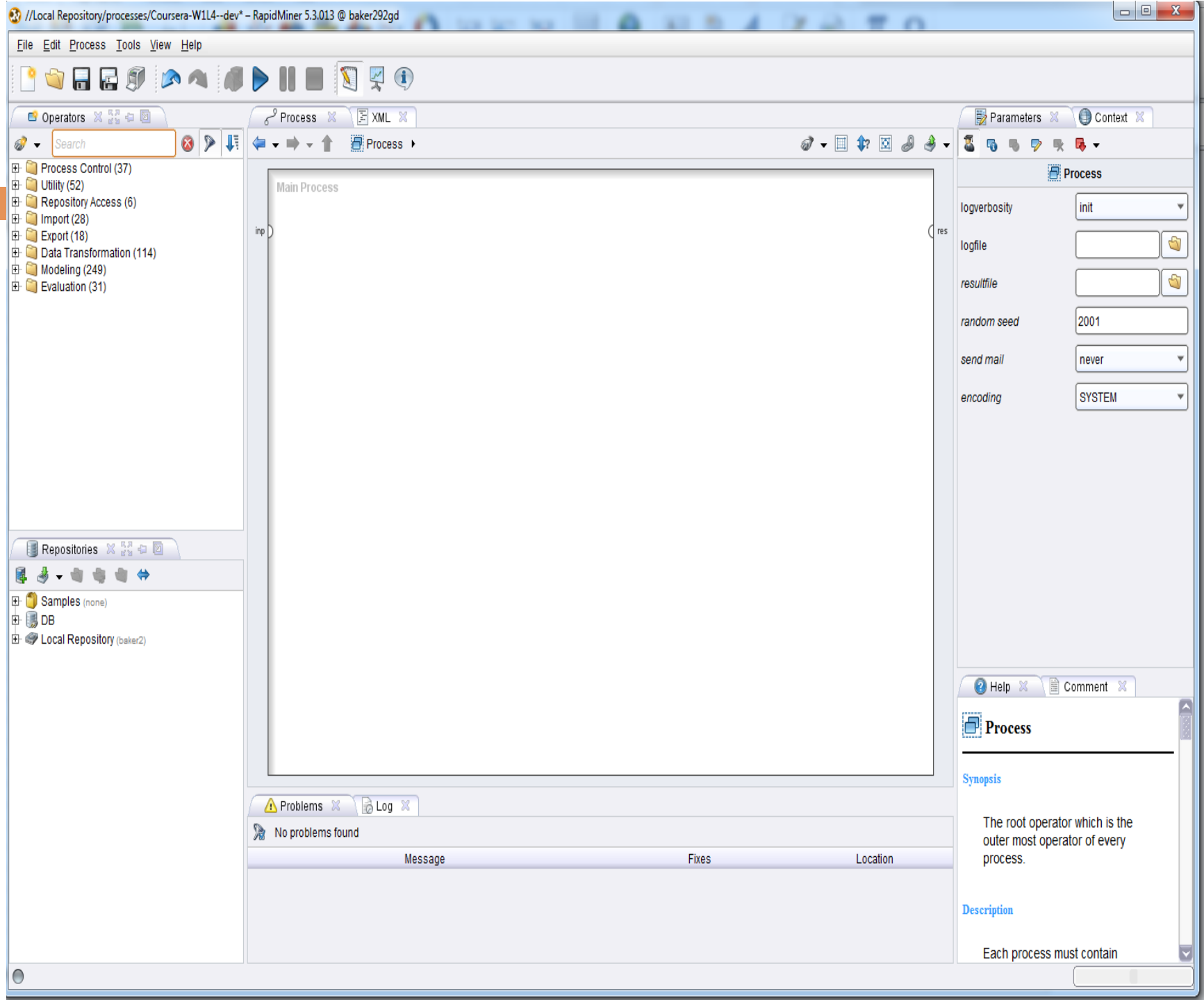


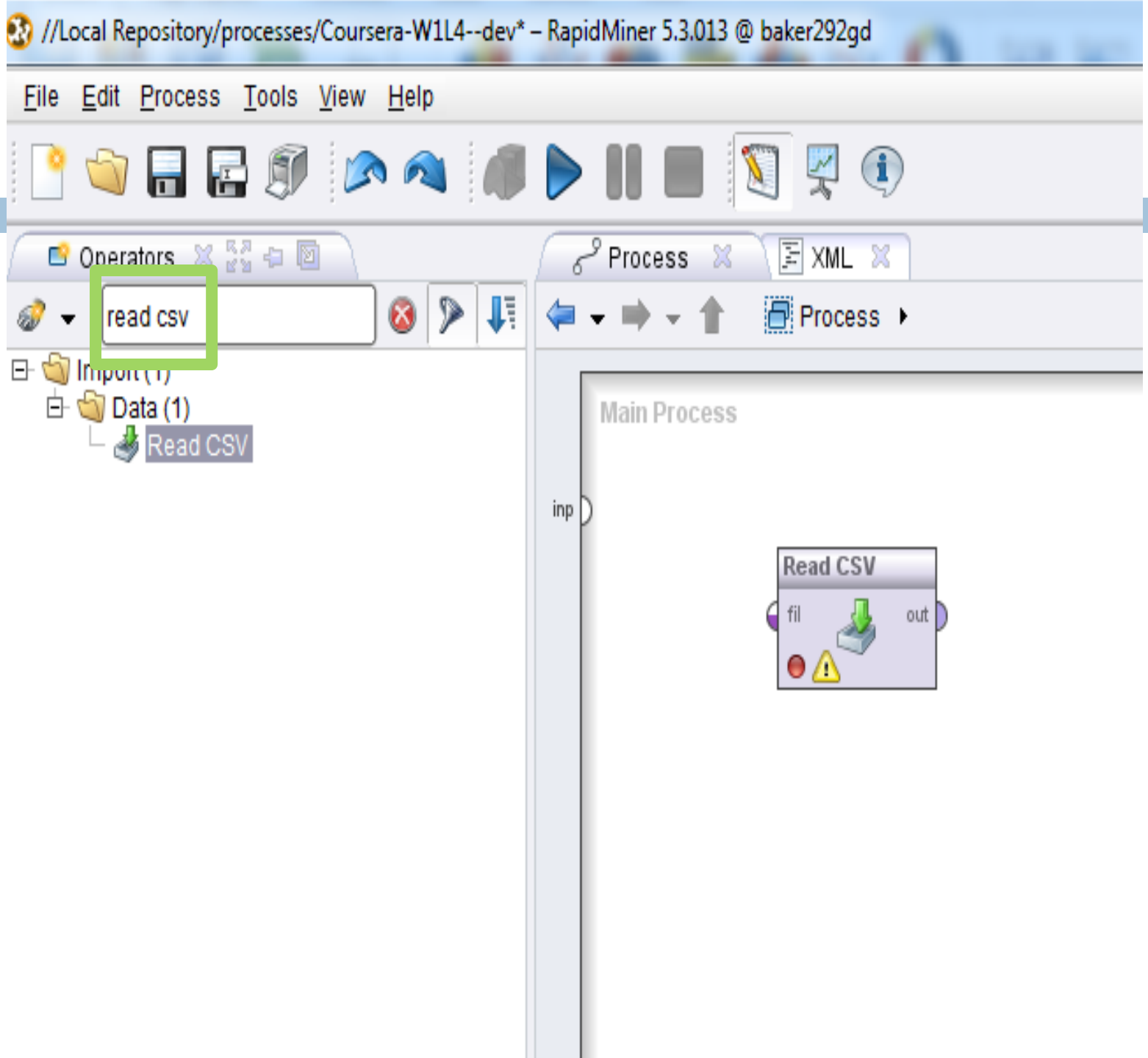


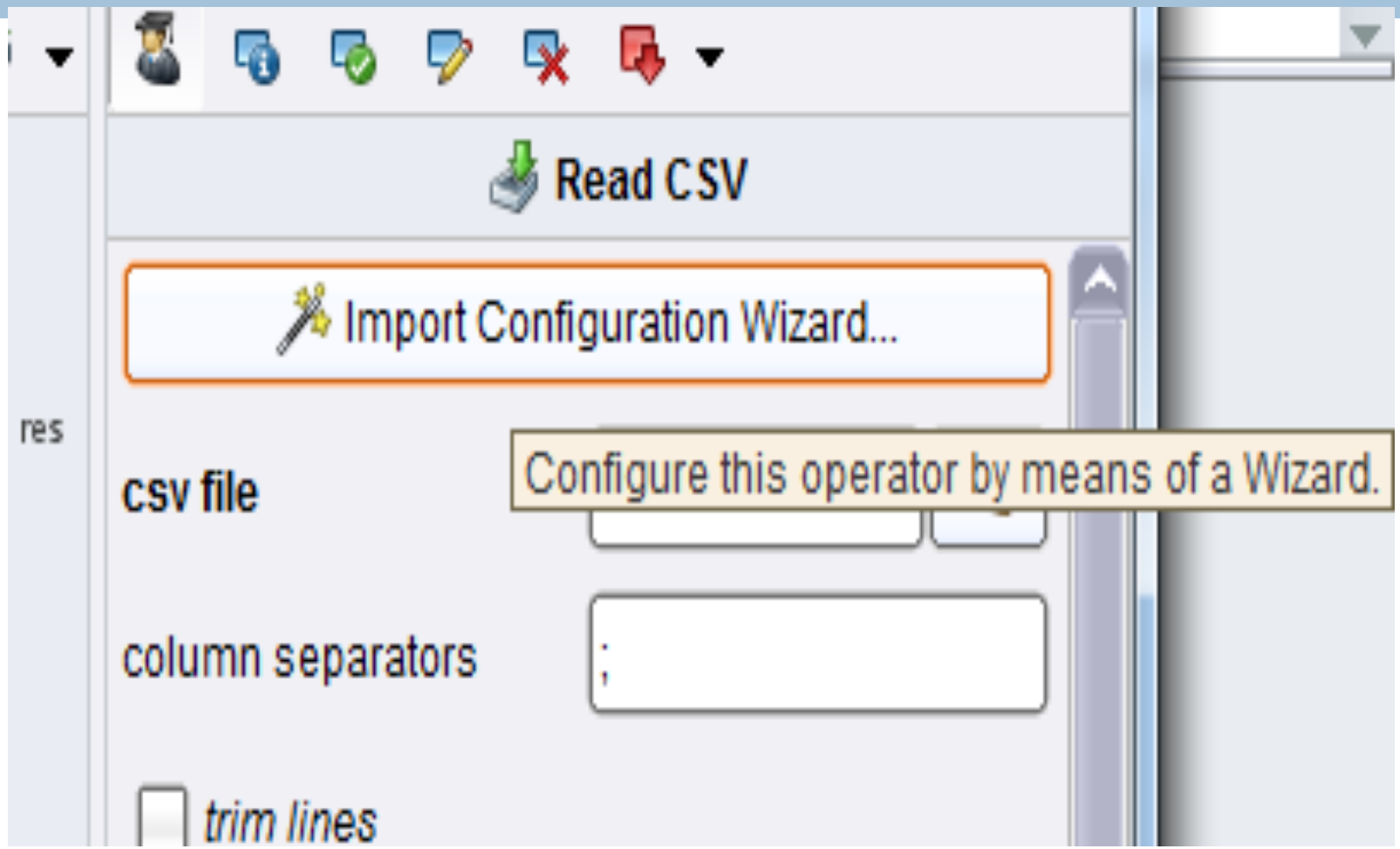
Let's Build Some Models

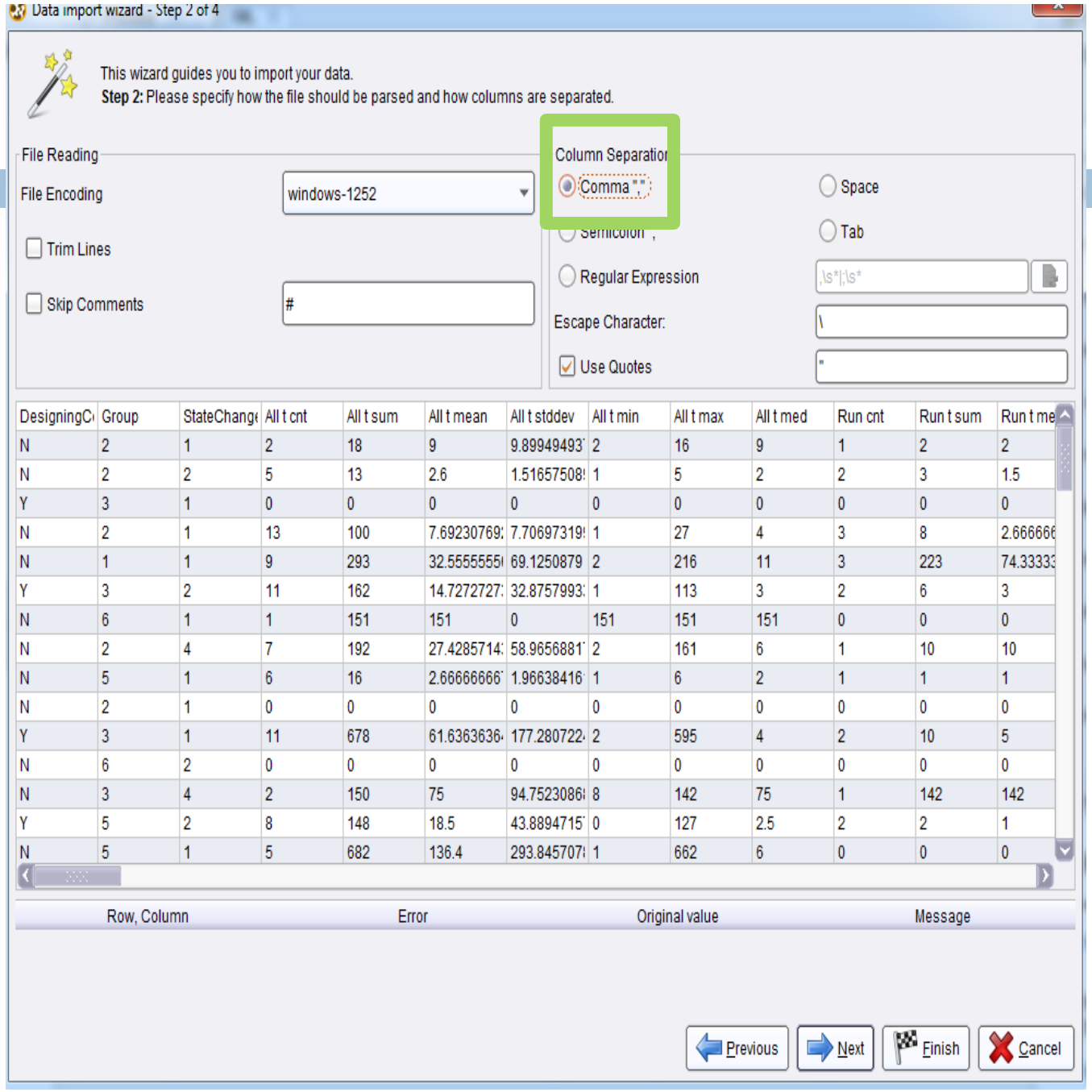
Open RapidMiner 5.3

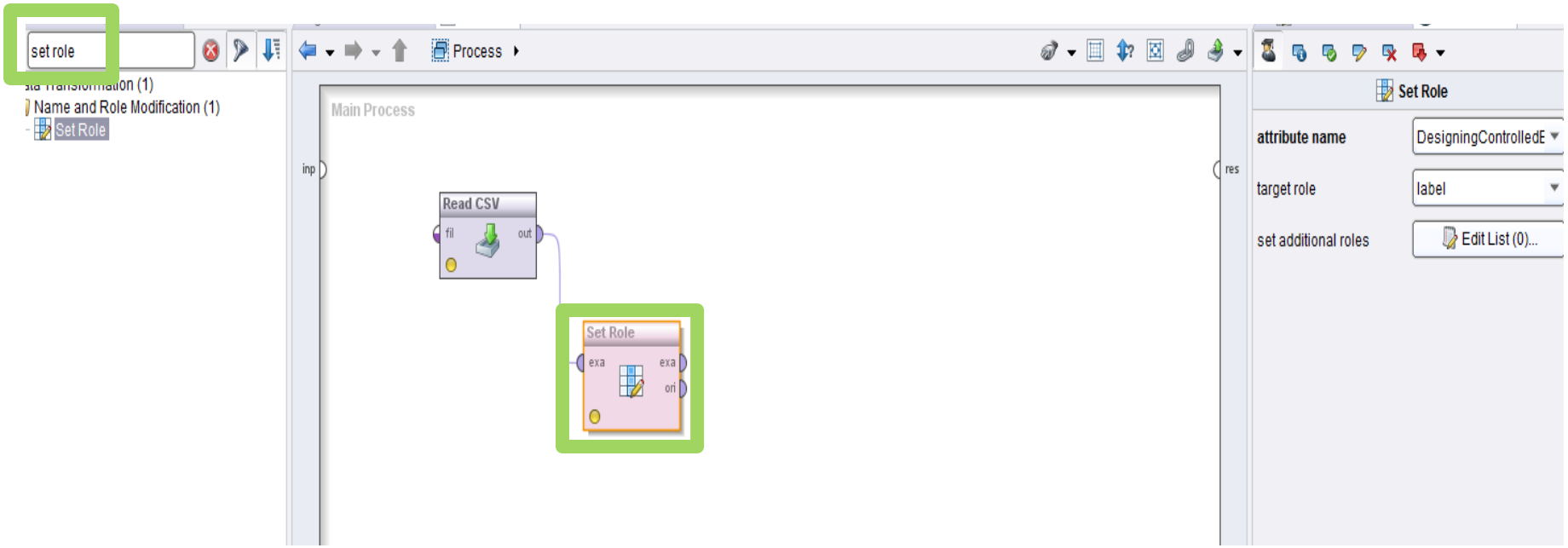
- And open a new process

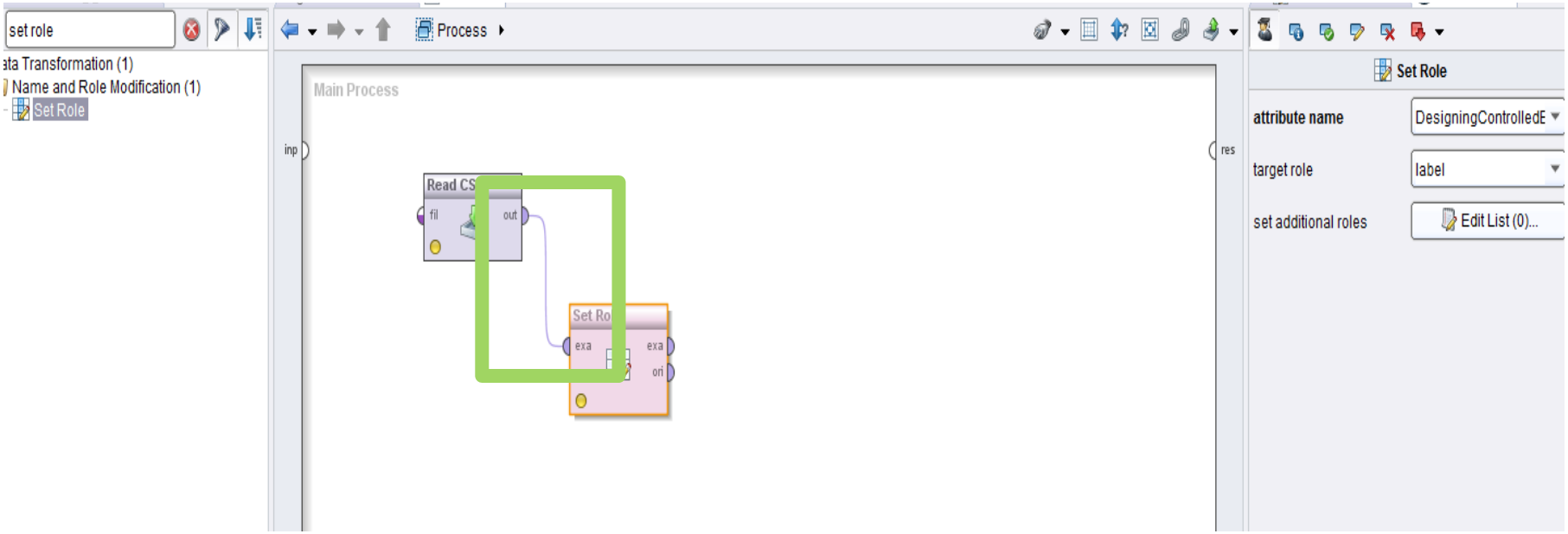

//Local Repository/processes/Coursera-W1L4--dev* – RapidMiner 5.3.013 @ baker292gd
File Edit Process Tools View Help
Operators
Search
Process Control (37)
Utility (52)
Repository Access (6)
Import (28)
Export (18)
Data Transformation (114)
Modeling (249)
Evaluation (31)
Repositories
Samples (none)
DB
Local Repository (baker2)
Process
XML
Process
Main Process
inp
res
Problems
Log
No problems found
Message
Fixes
Location
Parameters
Context
Process
logverbosity
init
logfile
resultfile
random seed
2001
send mail
never
encoding
SYSTEM
Help
Comment
Process
Synopsis
The root operator which is the outer most operator of every process.
Description
Each process must contain

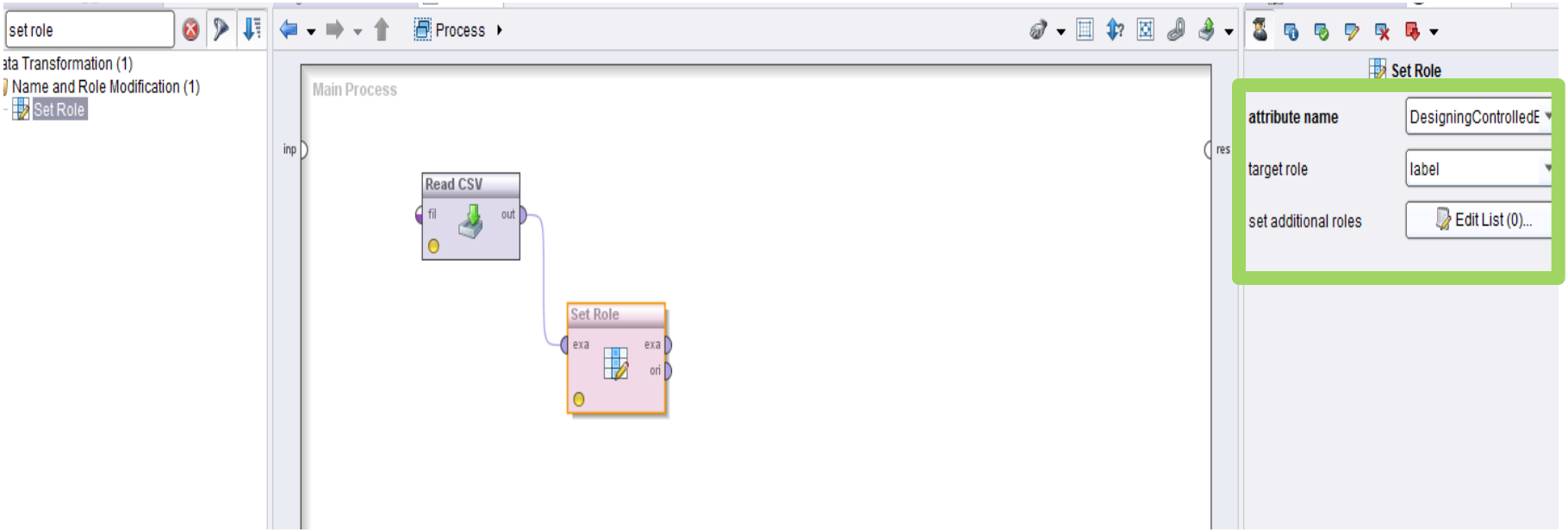

//Local Repository/processes/Coursera-W1L4--dev* – RapidMiner 5.3.013 @ baker292gd
File Edit Process Tools View Help
Operators
read csv
Data (1)
Read CSV
Process
XML
Process
Main Process
inp
Read CSV
fil
out

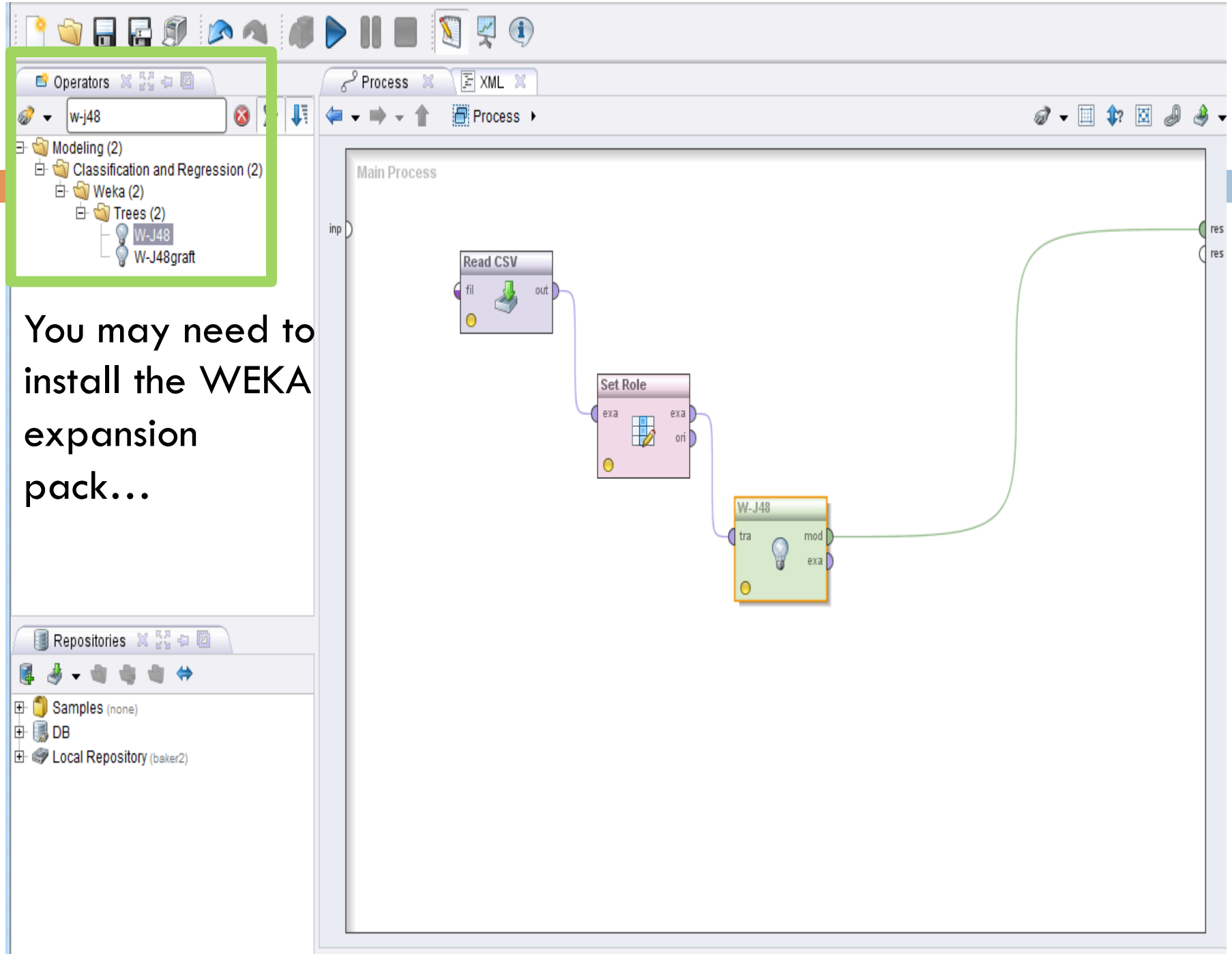

Read CSV
Import Configuration Wizard...
Configure this operator by means of a Wizard.
csv file
column separators
;
trim lines
res


Data import wizard - Step 2 of 4
This wizard guides you to import your data.
Step 2: Please specify how the file should be parsed and how columns are separated.
File Reading
File Encoding
windows-1252
Trim Lines
Skip Comments
#
Column Separation
Comma ","
Space
Tab
Regular Expression
,\s*|;\s*
Escape Character:
\
Use Quotes
"
DesigningC
Group
StateChange
All t cnt
All t sum
All t mean
All t stddev
All t min
All t max
All t med
Run cnt
Run t sum
Run t me
N 2 1 2 18 9 9.89949493 2 16 9 1 2 2
N 2 2 5 13 2.6 1.51657508 1 5 2 2 3 1.5
Y 3 1 0 0 0 0 0 0 0 0 0 0
N 2 1 13 100 7.69230769 7.70697319 1 27 4 3 8 2.666666
N 1 1 9 293 32.5555555 69.1250879 2 216 11 3 223 74.33333
Y 3 2 11 162 14.7272727 32.8757993 1 113 3 2 6 3
N 6 1 1 151 151 0 151 151 151 0 0 0
N 2 4 7 192 27.4285714 58.9656881 2 161 6 1 10 10
N 5 1 6 16 2.66666666 1.96638416 1 6 2 1 1 1
N 2 1 0 0 0 0 0 0 0 0 0 0
Y 3 1 11 678 61.6363636 177.280722 2 595 4 2 10 5
N 6 2 0 0 0 0 0 0 0 0 0 0
N 3 4 2 150 75 94.7523086 8 142 75 1 142 142
Y 5 2 8 148 18.5 43.8894715 0 127 2.5 2 2 1
N 5 1 5 682 136.4 293.845707 1 662 6 0 0 0
Row, Column
Error
Original value
Message
Previous
Next
Finish
Cancel

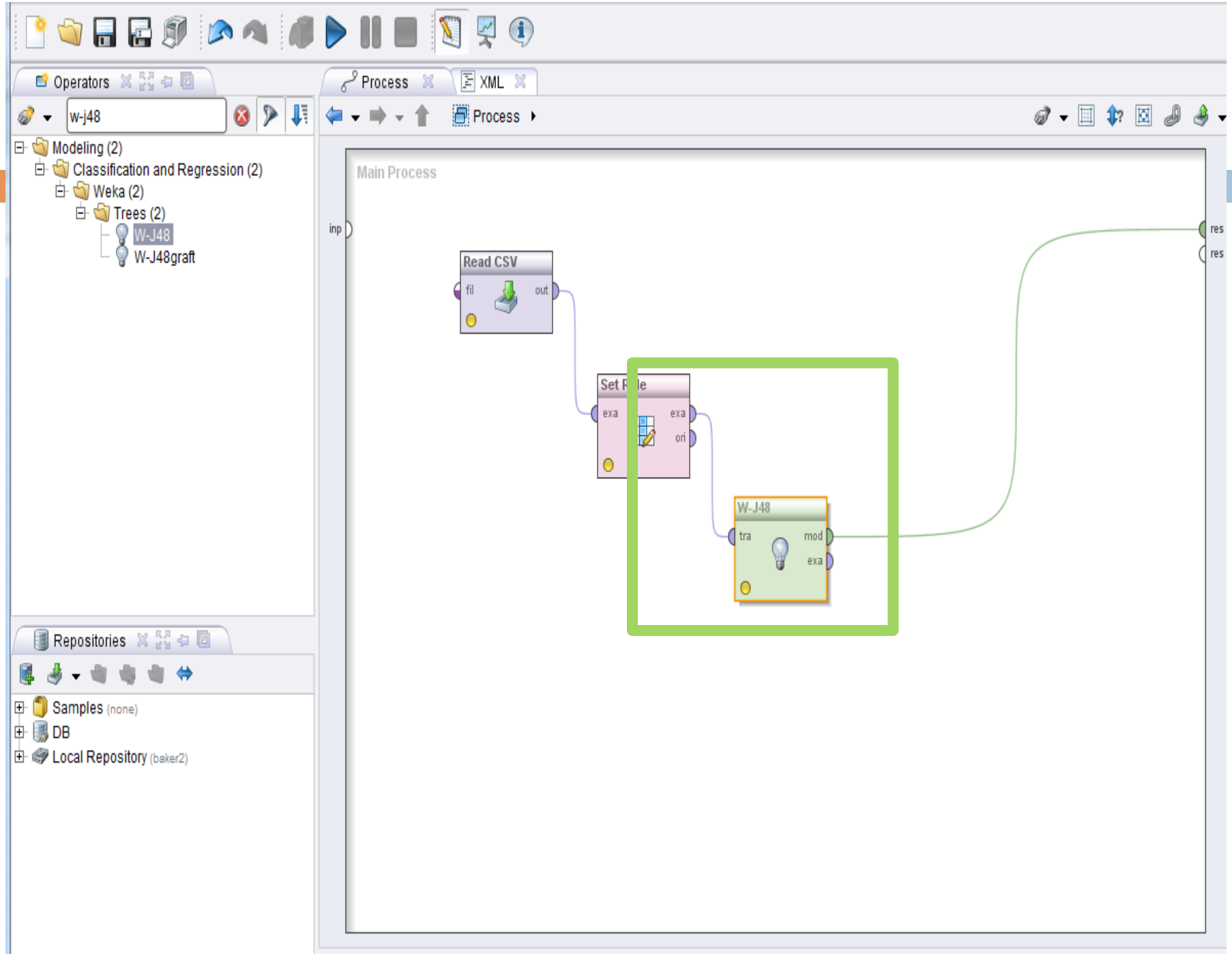

set role
Name and Role Modification (1)
Set Role
Process
Main Process
inp
res
Read CSV
fil
out
Set Role
exa
exa
ori
Set Role
attribute name
DesigningControlledE
target role
label
set additional roles
Edit List (0)...

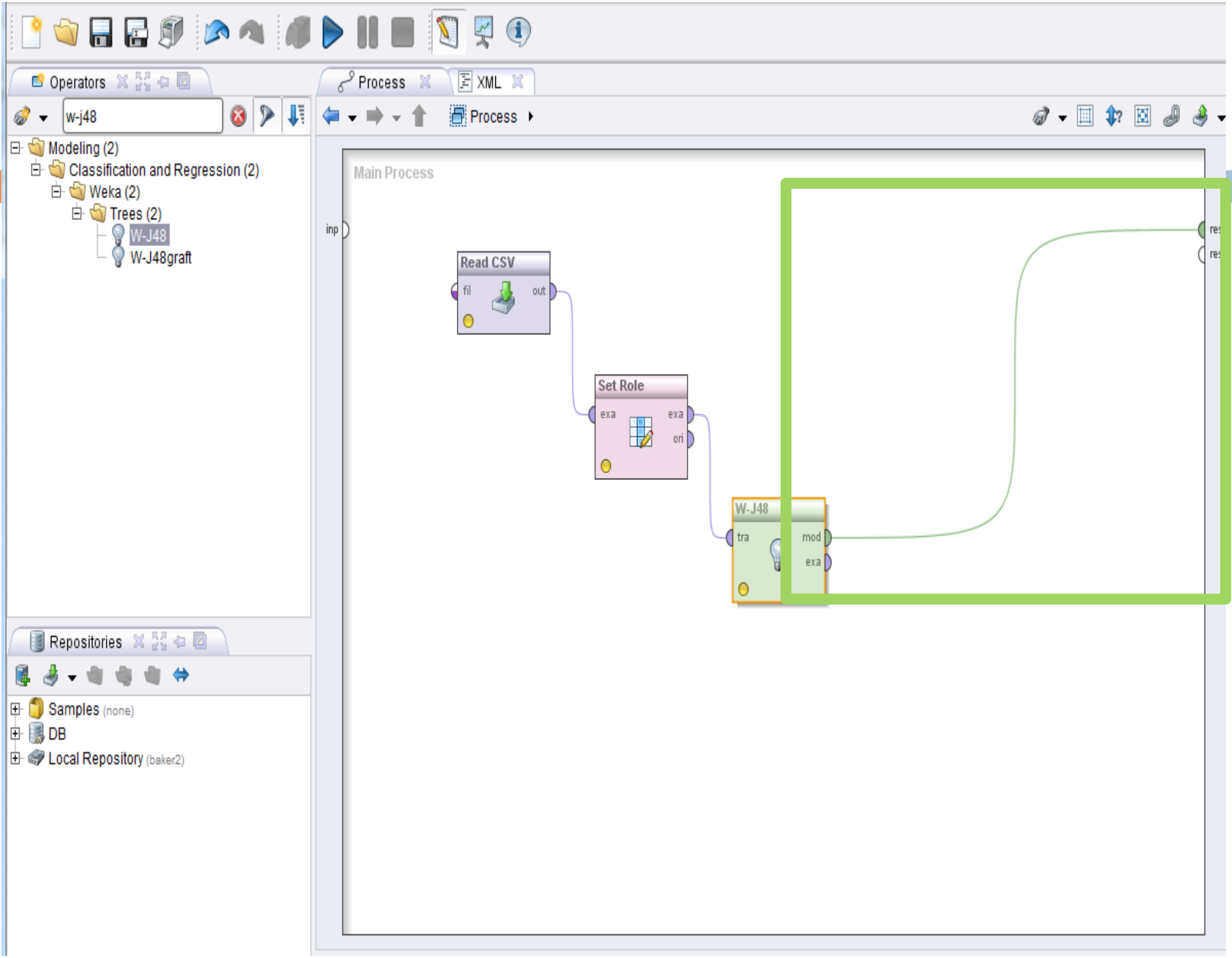

set role
ata Transformation (1)
Name and Role Modification (1)
Set Role
Process
Main Process
inp
res
Read CS
fil
out
Set Ro
exa
exa
ori
Set Role
attribute name
DesignigControlledE
target role
label
set additional roles
Edit List (0)...

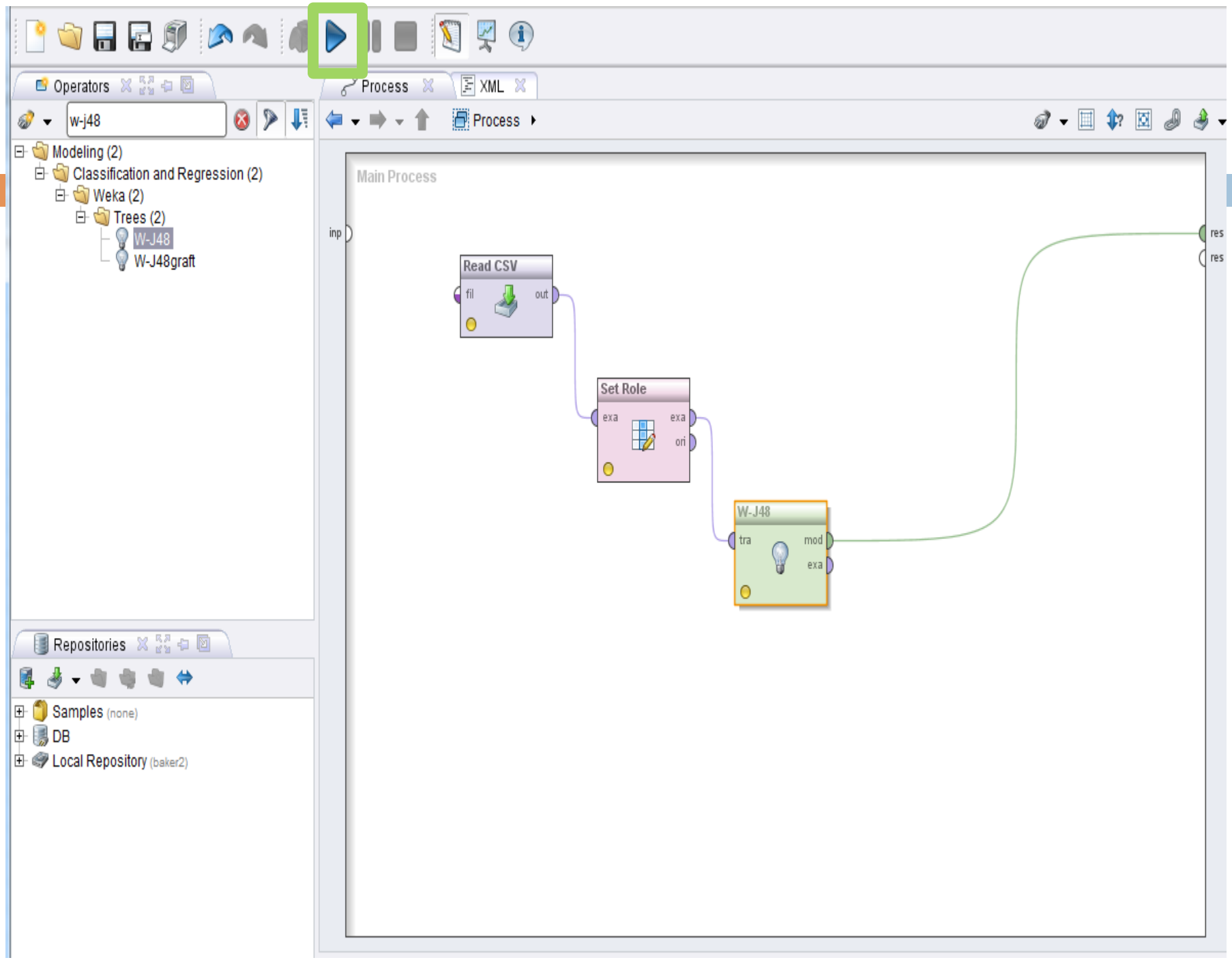

set role
ata Transformation (1)
Name and Role Modification (1)
Set Role
Process
Main Process
inp
res
Read CSV
fil
out
Set Role
exa
exa
ori
Set Role
attribute name
DesigningControlledE
target role
label
set additional roles
Edit List (0)...



You may need to install the WEKA expansion pack…


Operators
w-j48
Modeling (2)
Classification and Regression (2)
Weka (2)
Trees (2)
W-J48
W-J48graft
Repositories
Samples (none)
DB
Local Repository (baker2)
Process
XML
Process
Main Process
inp
Read CSV
fil
out
exa
exa
ori
W-J48
tra
mod
exa
res
res


Operators
w-j48
Modeling (2)
Classification and Regression (2)
Weka (2)
Trees (2)
W-J48
W-J48graft
Repositories
Samples (none)
DB
Local Repository (baker2)
Process
XML
Process
Main Process
inp
Read CSV
fil
out
Set Role
exa
exa
ori
W-J48
tra
mod
exa


Operators
w-j48
Modeling (2)
Classification and Regression (2)
Weka (2)
Trees (2)
W-J48
W-J48graft
Repositories
Samples (none)
DB
Local Repository (baker2)
Process
XML
Process
Main Process
inp
Read CSV
fil
out
Set Role
exa
exa
ori
W-J48
tra
mod
exa
res
res

Result Overview | W-J48 (W-J48)

Text View | Weka Result

W-J48

```
J48 pruned tree
------------------

Cm CVS cnt <= 0: N (271.0/2.0)
Cm CVS cnt > 0
|   CVS cnt <= 0
|   |   Run t sum <= 11
|   |   |   Hyp table show t sum <= 1
|   |   |   |   Cm Pause cnt <= 2
|   |   |   |   |   Data table show cnt <= 0
|   |   |   |   |   |   Hyp var change cnt <= 4
|   |   |   |   |   |   |   Cm Hyp var change cnt <= 12
|   |   |   |   |   |   |   |   Mw iv change t med <= 6.5
|   |   |   |   |   |   |   |   |   Cm Run cnt <= 4
|   |   |   |   |   |   |   |   |   |   All t min <= 1
|   |   |   |   |   |   |   |   |   |   |   Cm Incmplt run t min <= 2
|   |   |   |   |   |   |   |   |   |   |   |   Cm Cmplt run t sum <= 2: N (10.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   Cm Cmplt run t sum > 2
|   |   |   |   |   |   |   |   |   |   |   |   |   Hyp var change t min <= 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   Hyp make t sum <= 112
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Mw iv change t max <= 17
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Rept cnt <= 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Hyp make t stddev <= 33.234019
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Data table show cnt <= 2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm All t min <= 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm All t cnt <= 12: N (4.0/1.0)
```

Log

Result Overview | W-J48 (W-J48)

Text View | Weka Result

W-J48

```
J48 pruned tree
------------------

Cm CVS cnt <= 0: N (271.0/2.0)
Cm CVS cnt > 0
|   CVS cnt <= 0
|   |   Run t sum <= 11
|   |   |   Hyp table show t sum <= 1
|   |   |   |   Cm Pause cnt <= 2
|   |   |   |   |   Data table show cnt <= 0
|   |   |   |   |   |   Hyp var change cnt <= 4
|   |   |   |   |   |   |   Cm Hyp var change cnt <= 12
|   |   |   |   |   |   |   |   Mw iv change t med <= 6.5
|   |   |   |   |   |   |   |   |   Cm Run cnt <= 4
|   |   |   |   |   |   |   |   |   |   All t min <= 1
|   |   |   |   |   |   |   |   |   |   |   Cm Incmplt run t min <= 2
|   |   |   |   |   |   |   |   |   |   |   |   Cm Cmplt run t sum <= 2: N (10.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   Cm Cmplt run t sum > 2
|   |   |   |   |   |   |   |   |   |   |   |   |   Hyp var change t min <= 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   Hyp make t sum <= 112
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Mw iv change t max <= 17
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Rept cnt <= 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Hyp make t stddev <= 33.234019
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Data table show cnt <= 2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm All t min <= 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm All t cnt <= 12: N (4.0/1.0)
```

Log

Result Overview | W-J48 (W-J48)

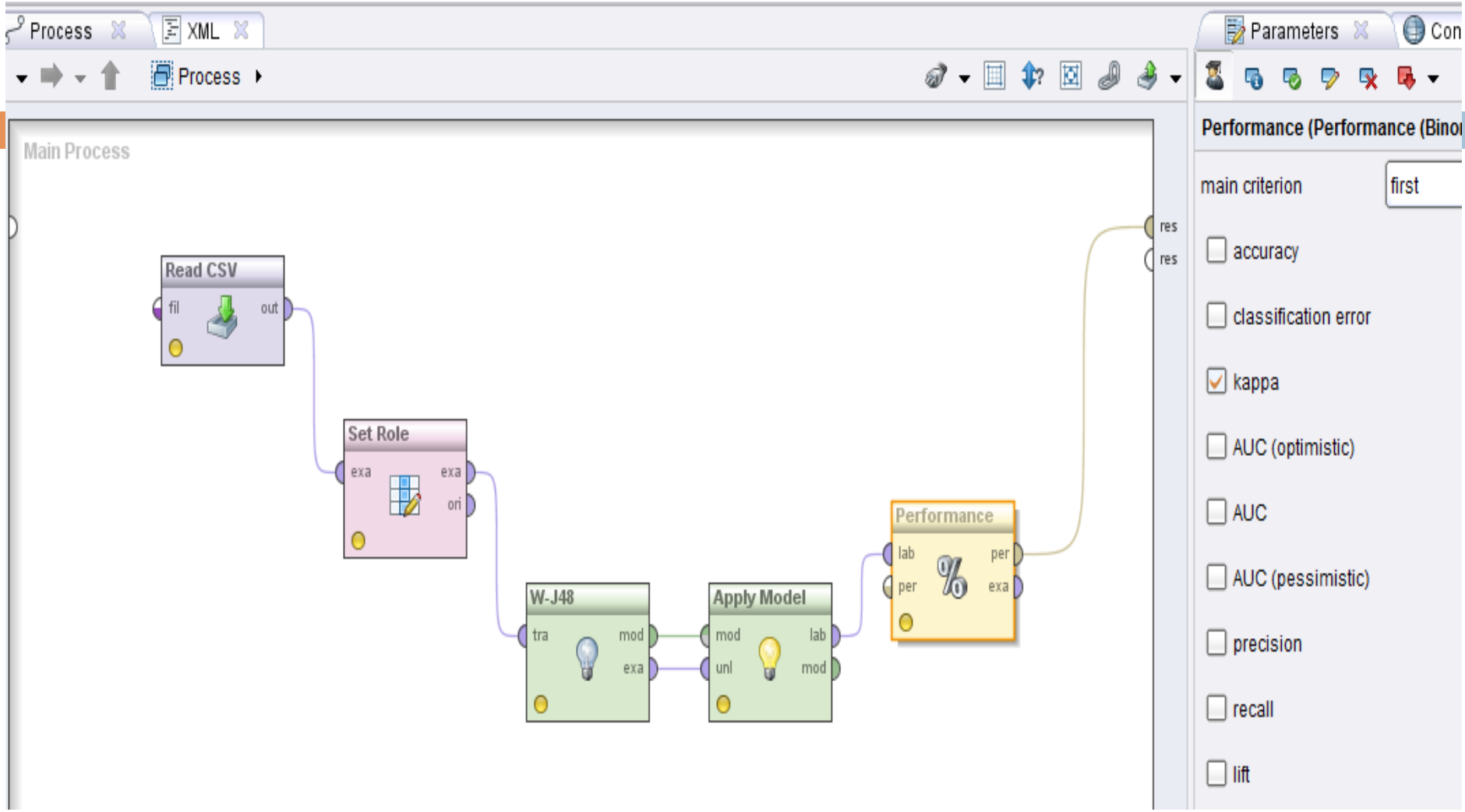
Text View | Weka Result

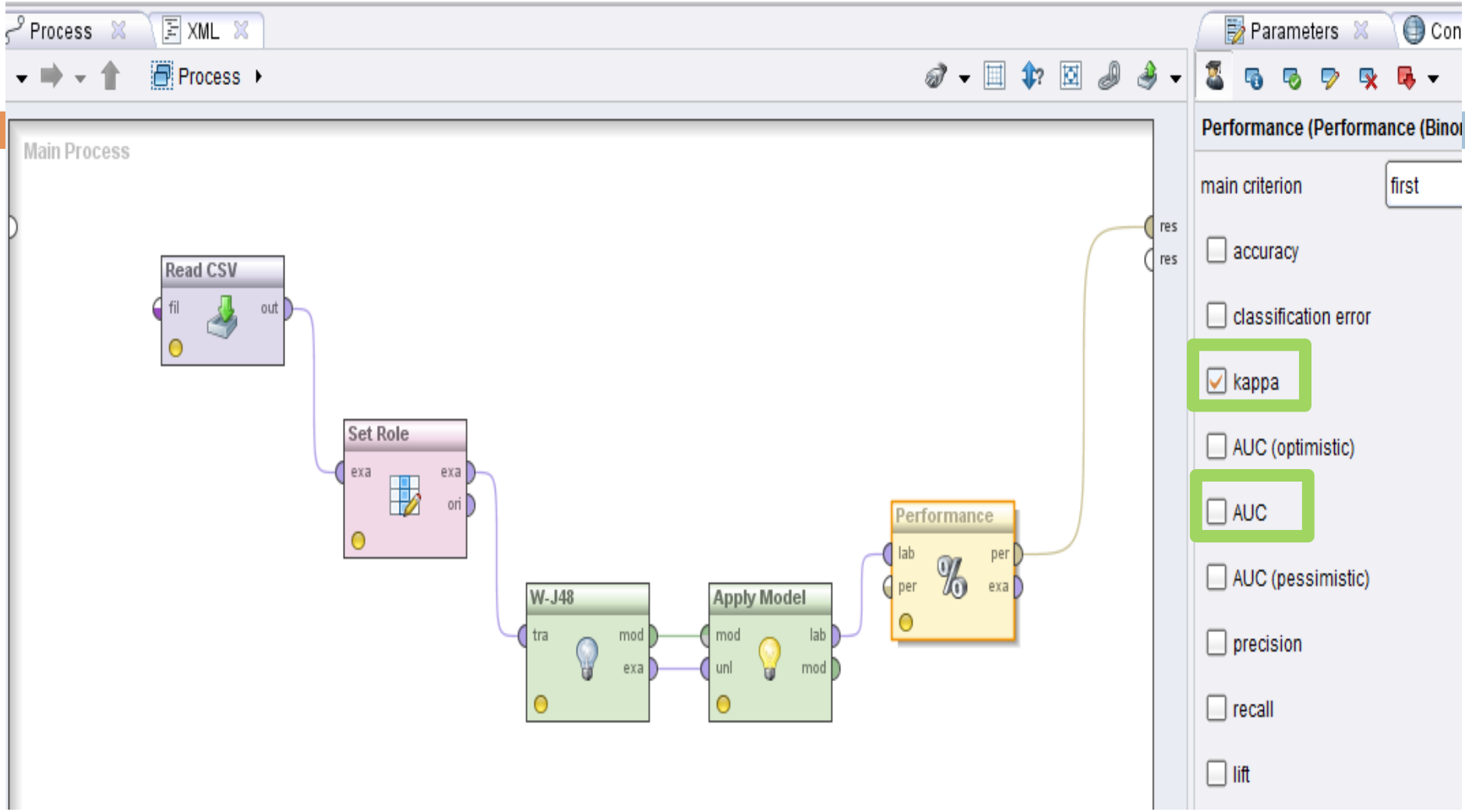
W-J48

```
J48 pruned tree
------------------

Cm CVS cnt <= 0: N (271.0/2.0)
Cm CVS cnt > 0
|   CVS cnt <= 0
|   |   Run t sum <= 11
|   |   |   Hyp table show t sum <= 1
|   |   |   |   Cm Pause cnt <= 2
|   |   |   |   |   Data table show cnt <= 0
|   |   |   |   |   |   Hyp var change cnt <= 4
|   |   |   |   |   |   |   Cm Hyp var change cnt <= 12
|   |   |   |   |   |   |   |   Mw iv change t med <= 6.5
|   |   |   |   |   |   |   |   |   Cm Run cnt <= 4
|   |   |   |   |   |   |   |   |   |   All t min <= 1
|   |   |   |   |   |   |   |   |   |   |   Cm Incmplt run t min <= 2
|   |   |   |   |   |   |   |   |   |   |   |   Cm Cmplt run t sum <= 2: N (10.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   Cm Cmplt run t sum > 2
|   |   |   |   |   |   |   |   |   |   |   |   |   Hyp var change t min <= 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   Hyp make t sum <= 112
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Mw iv change t max <= 17
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Rept cnt <= 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Hyp make t stddev <= 33.234019
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm Data table show cnt <= 2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm All t min <= 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Cm All t cnt <= 12: N (4.0/1.0)
```

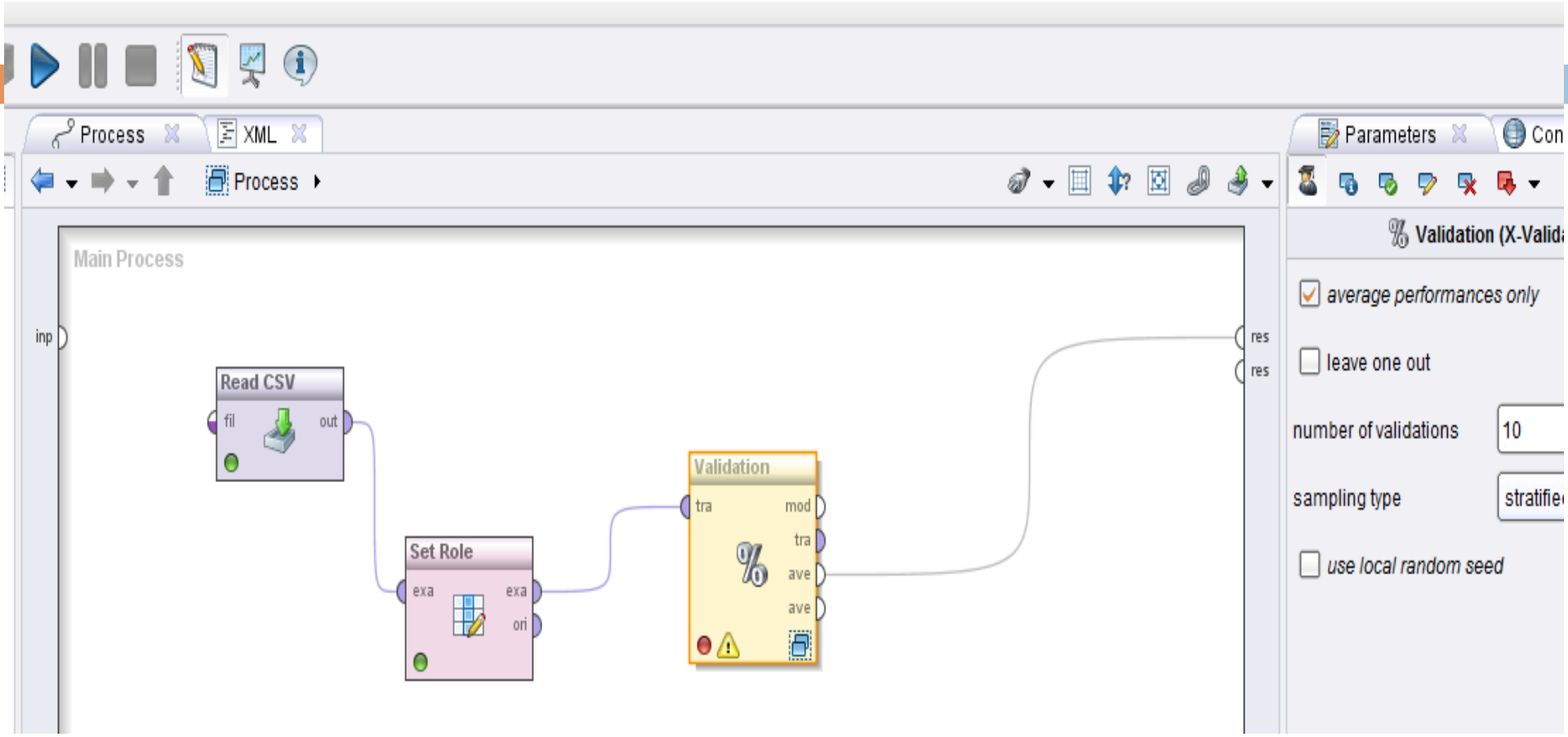
Log

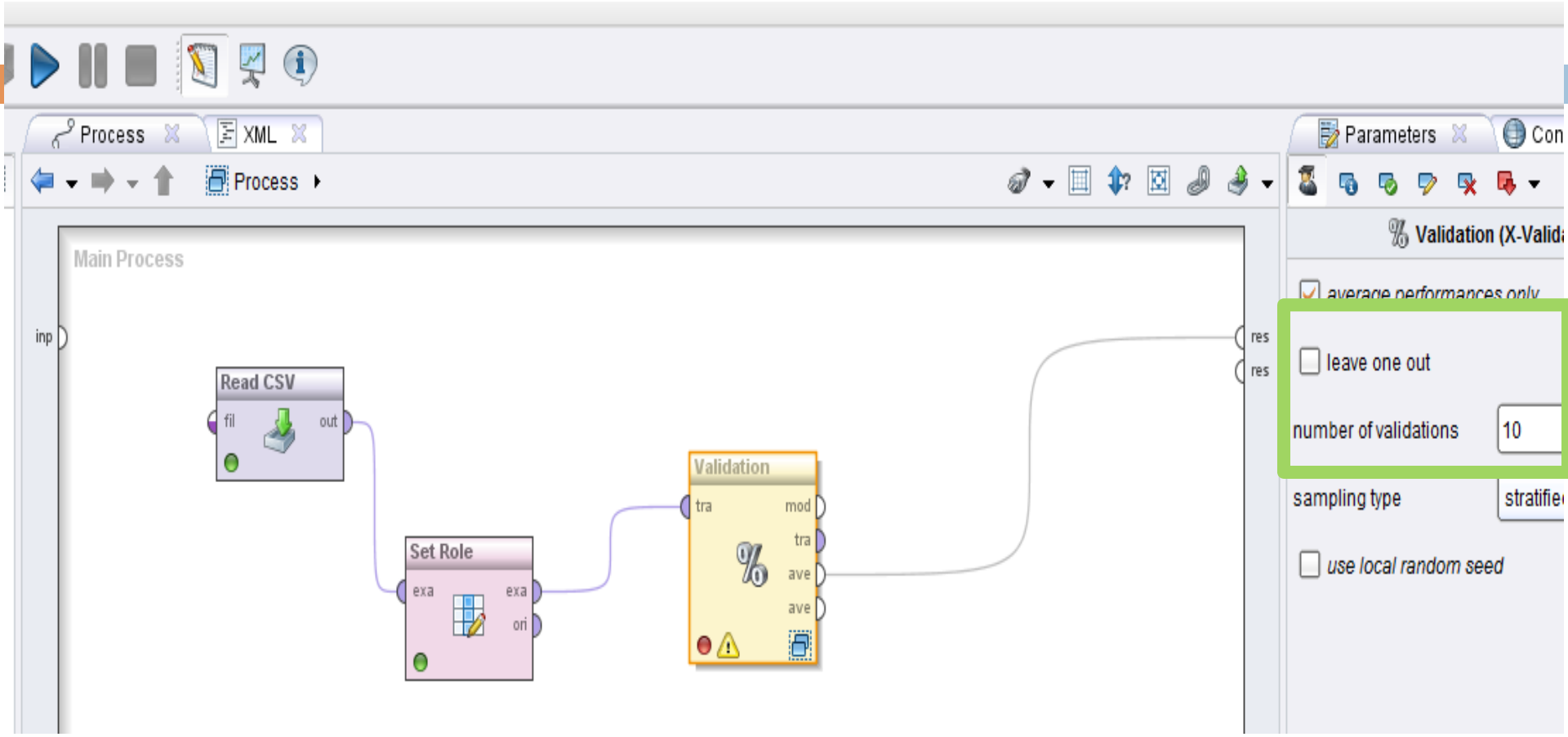

Process
XML
Process
Parameters
Con
Main Process
Read CSV
fil
out
Set Role
exa
exa
ori
W-J48
tra
mod
exa
Apply Model
mod
lab
unl
mod
Performance
lab
per
per
exa
res
res
Performance (Performance (Bino
main criterion
first
accuracy
classification error
kappa
AUC (optimistic)
AUC
AUC (pessimistic)
precision
recall
lift

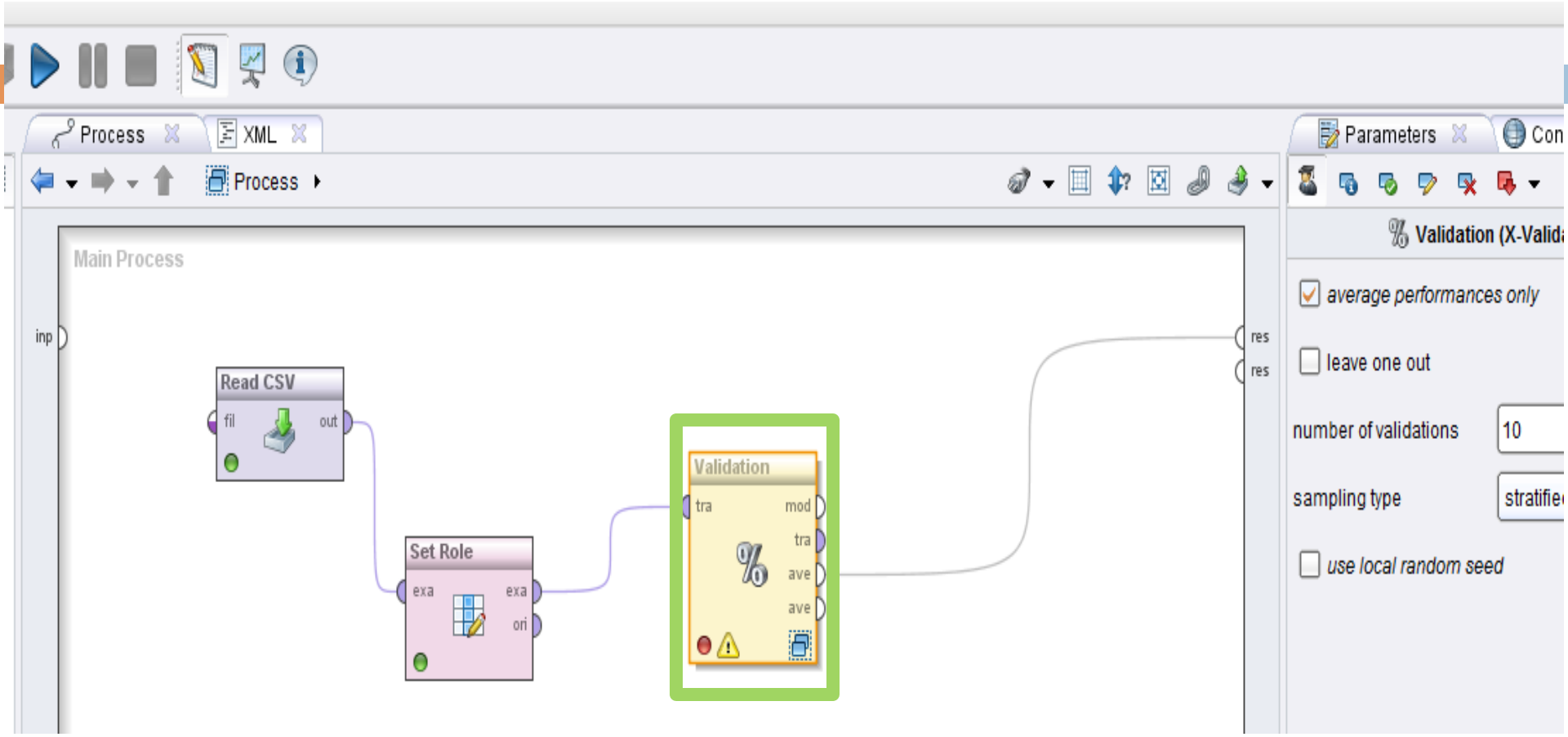

Process
XML
Process
Main Process
Read CSV
fil
out
Set Role
exa
exa
ori
W-J48
tra
mod
exa
Apply Model
mod
lab
unl
mod
Performance
lab
per
per
exa
res
res
Parameters
Performance (Performance (Bino
main criterion
first
accuracy
classification error
kappa
AUC (optimistic)
AUC
AUC (pessimistic)
precision
recall
lift

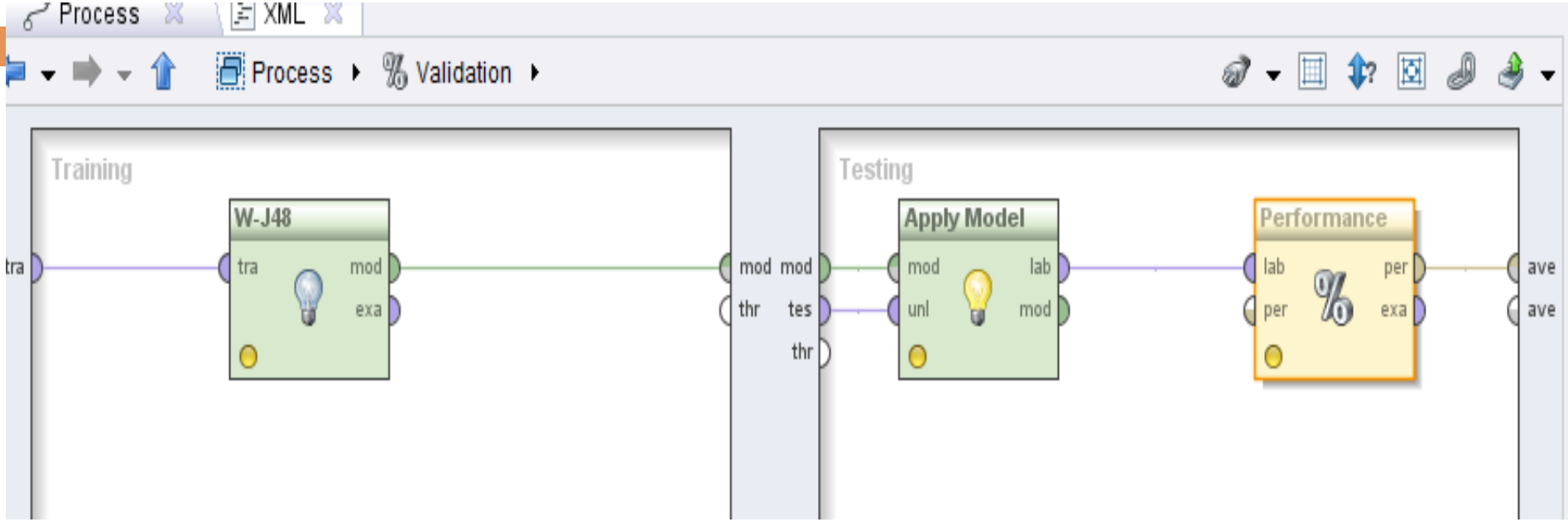
kappa: 0.933

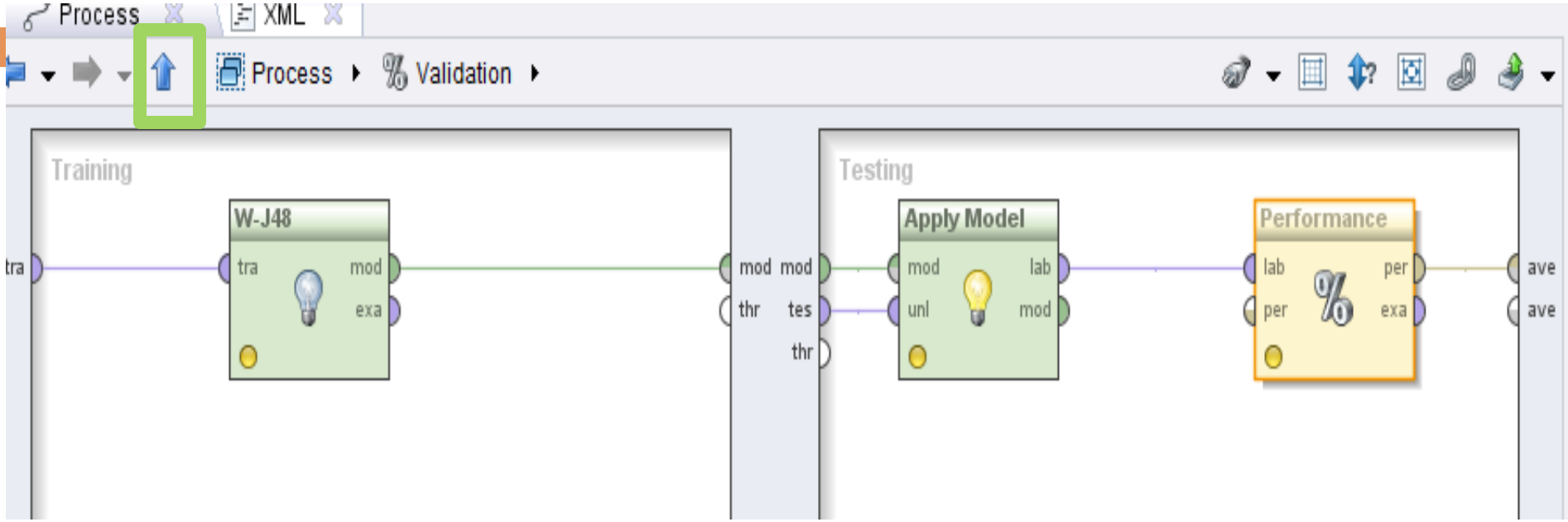
	true N	true Y	class precision
pred. N	383	11	97.21%
pred. Y	5	165	97.06%
class recall	98.71%	93.75%	

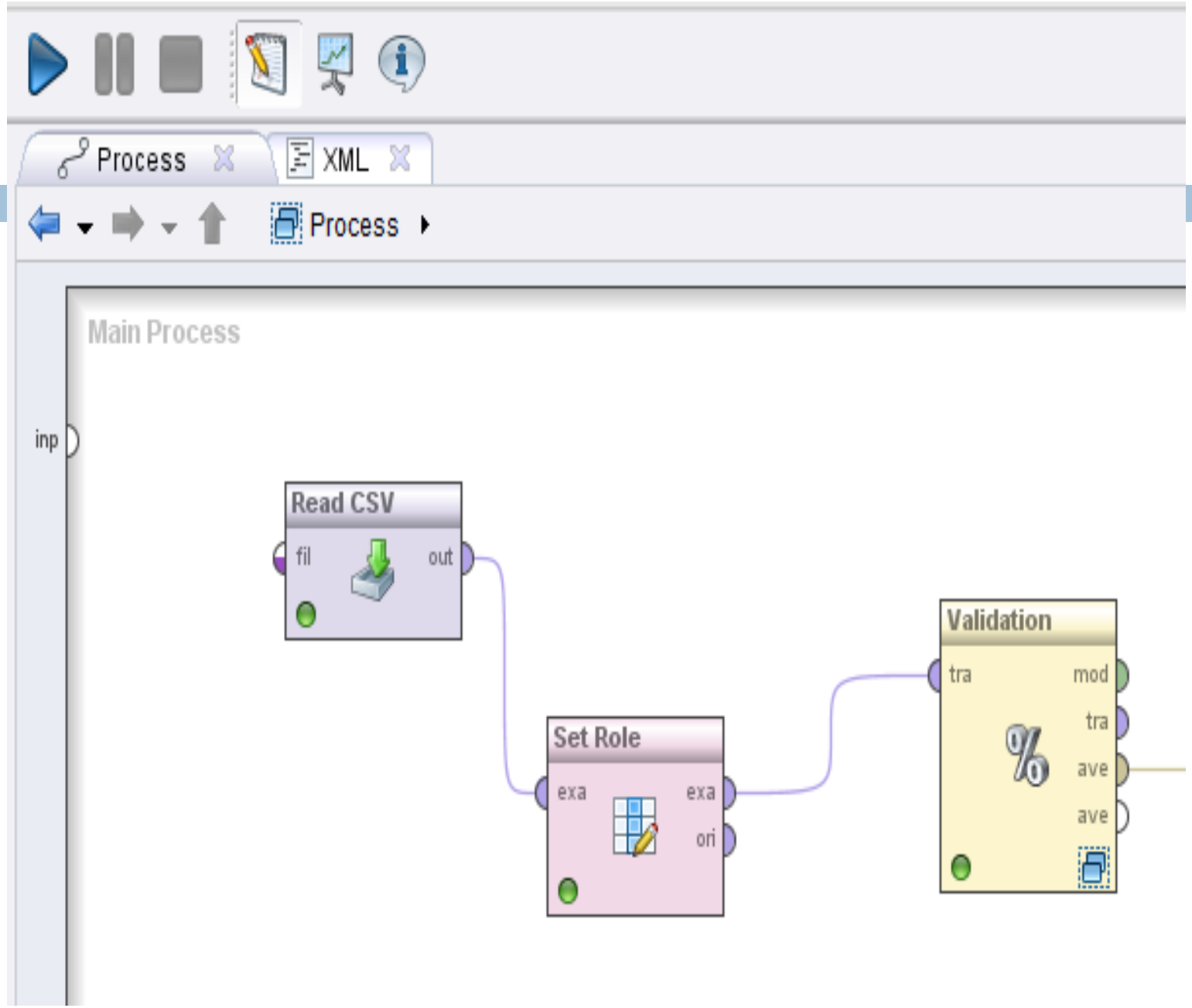

Process
XML
Process
Main Process
inp
Read CSV
fil
out
Set Role
exa
exa
ori
Validation
tra
mod
tra
ave
ave
res
res
Parameters
Validation (X-Valida
average performances only
leave one out
number of validations
10
sampling type
use local random seed

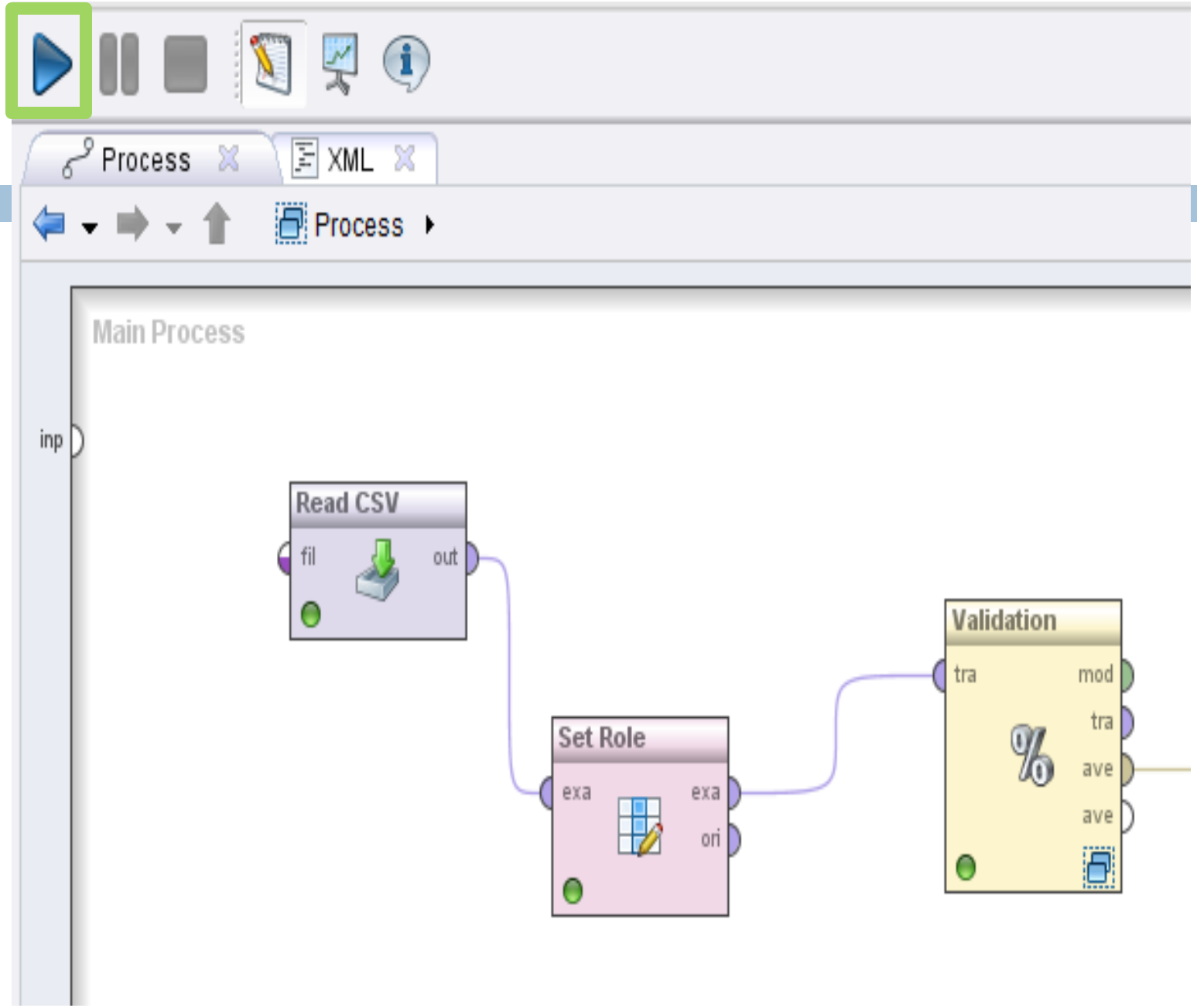

Process
XML
Process
Main Process
inp
Read CSV
fil
out
Set Role
exa
exa
ori
Validation
tra
mod
tra
ave
ave
res
res
Parameters
Validation (X-Valida
average performances only
leave one out
number of validations
10
sampling type
stratifie
use local random seed


Process
XML
Process
Main Process
inp
Read CSV
fil
out
Set Role
exa
exa
ori
Validation
tra
mod
tra
ave
ave
res
res
Parameters
Validation (X-Valida
average performances only
leave one out
number of validations
10
sampling type
stratifie
use local random seed


Process
XML
Process
Validation
Training
W-J48
tra
mod
exa
mod
thr
Testing
Apply Model
mod
unl
lab
mod
Performance
lab
per
per
exa
ave
ave

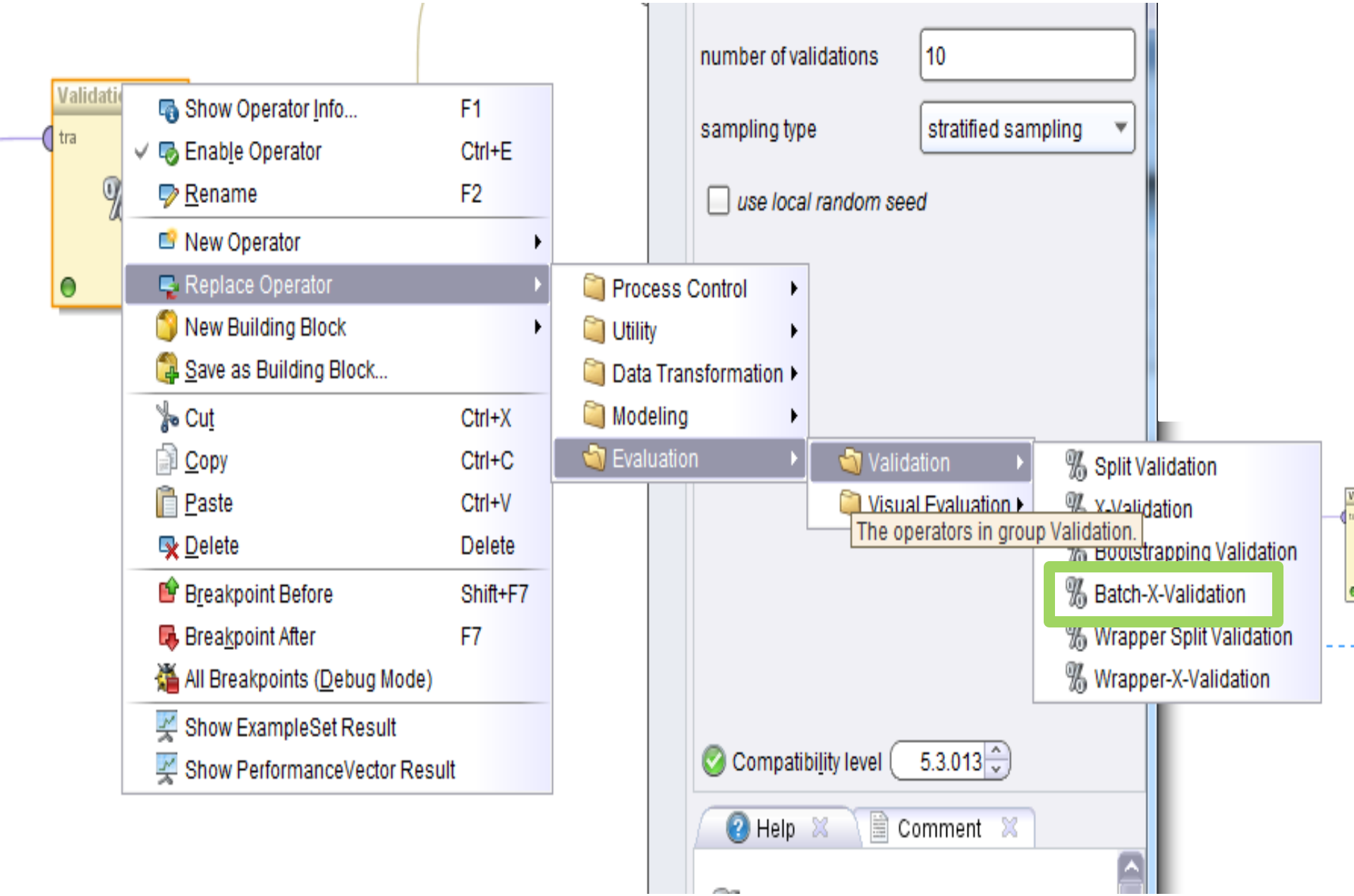

Process
XML
Process
Validation
Training
W-J48
tra
mod
exa
Testing
Apply Model
mod
unl
lab
mod
Performance
lab
per
per
exa
mod
thr
mod
thr
tes
ave
ave

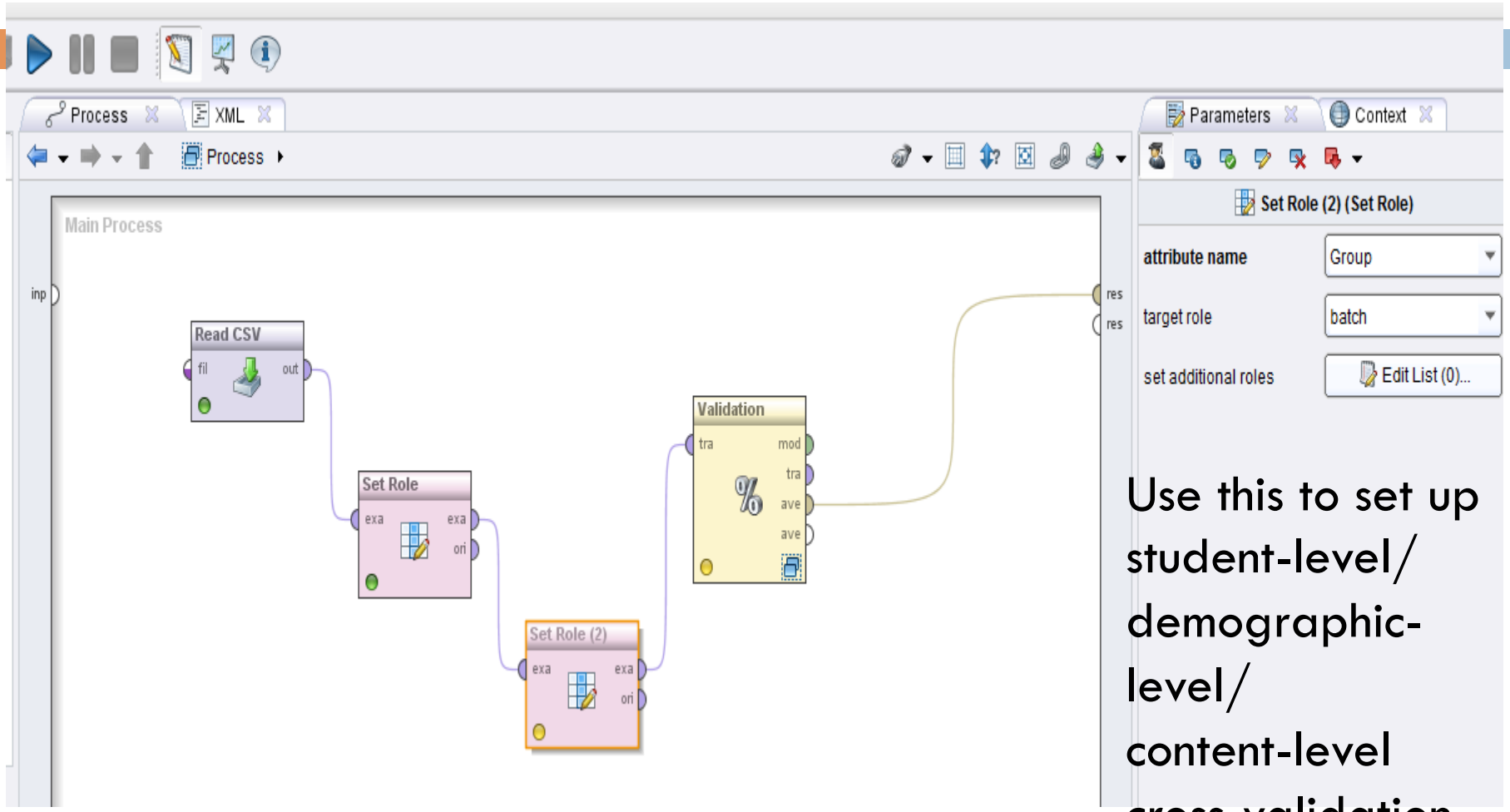

Process
XML
Process
Main Process
inp
Read CSV
fil
out
Set Role
exa
exa
ori
Validation
tra
mod
tra
ave
ave


Process
XML
Process
Main Process
inp
Read CSV
fil
out
Set Role
exa
exa
ori
Validation
tra
mod
tra
ave
ave

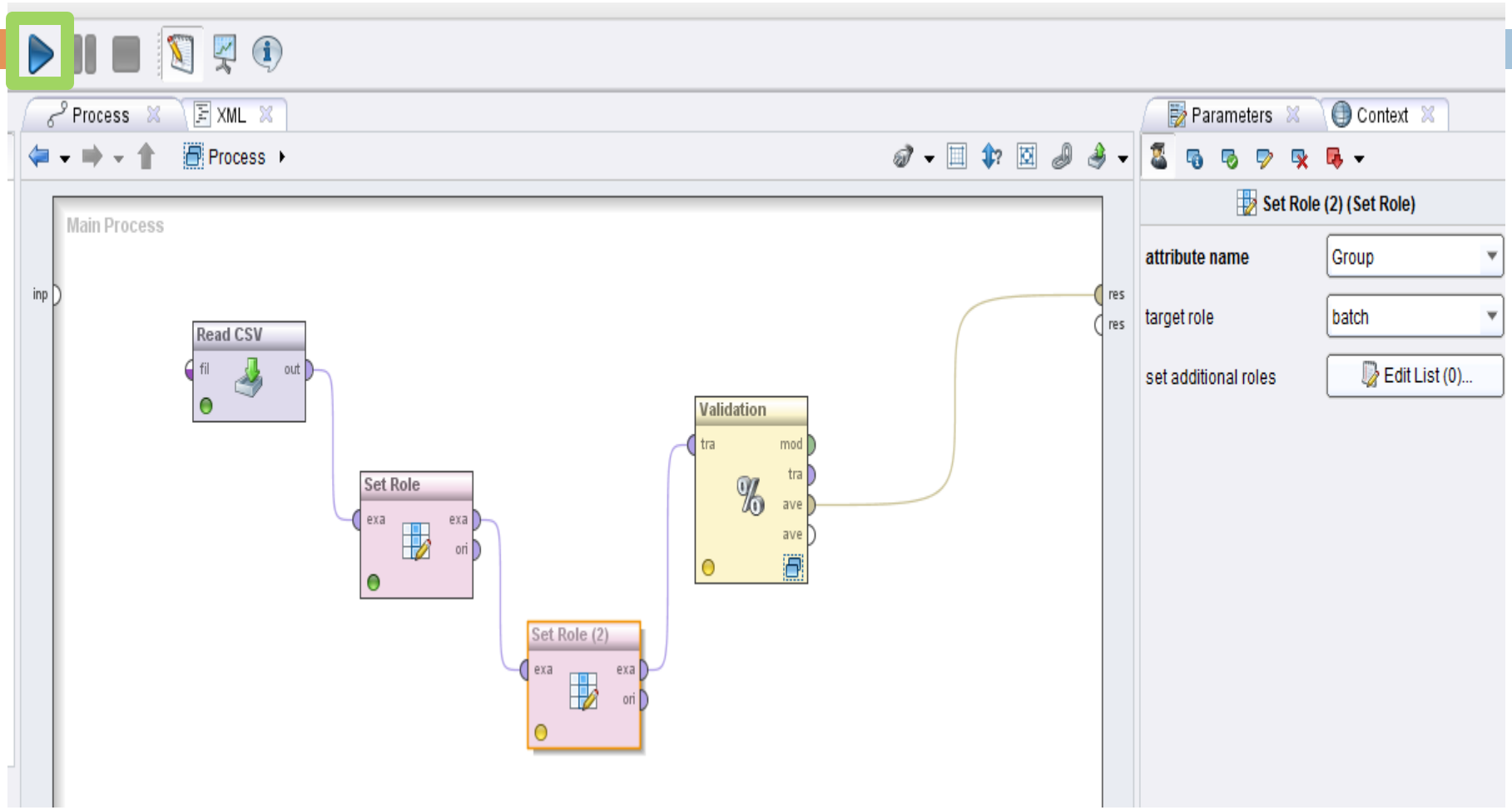
kappa: 0.442 - 0.153 (mikro: 0.445)

	true N	true Y	class precision
ored. N	325	70	82.28%
ored. Y	63	106	62.72%
class recall	83.76%	60.23%	


number of validations
10
sampling type
stratified sampling
use local random seed
Show Operator Info... F1
Enable Operator Ctrl+E
Rename F2
New Operator
Replace Operator
New Building Block
Save as Building Block...
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Delete Delete
Breakpoint Before Shift+F7
Breakpoint After F7
All Breakpoints (Debug Mode)
Show ExampleSet Result
Show PerformanceVector Result
Process Control
Utility
Data Transformation
Modeling
Evaluation
Validation
Visual Evaluation
The operators in group Validation.
Split Validation
X-Validation
Bootstrapping Validation
Batch-X-Validation
Wrapper Split Validation
Wrapper-X-Validation
Compatibility level
5.3.013
Help
Comment



Use this to set up student-level/ demographic-level/ content-level cross-validation


Process
XML
Process
Main Process
inp
Read CSV
fil
out
Set Role
exa
exa
ori
Set Role (2)
exa
exa
ori
Validation
tra
mod
tra
ave
ave
res
res
Parameters
Context
Set Role (2) (Set Role)
attribute name
Group
target role
batch
set additional roles
Edit List (0)...

kappa: 0.445 +/- 0.154 (mikro: 0.448)

	true N	true Y	class precision
pred. N	326	70	82.32%
pred. Y	62	106	63.10%
class recall	84.02%	60.23%	

Try it yourself with other algorithms!

- W-JRip
- W-KStar
- Linear Regression (implements Step Regression)