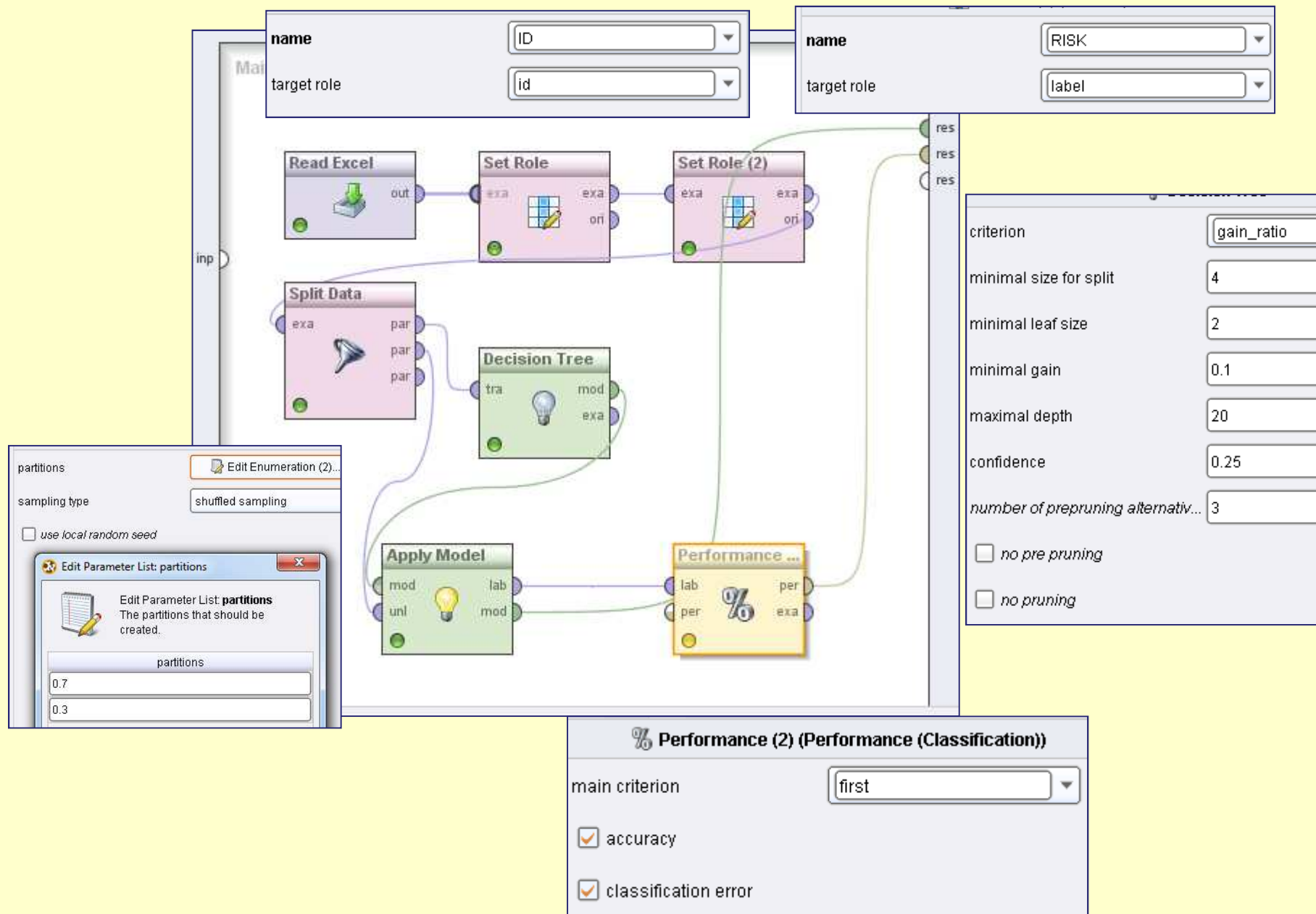


Week 09

Classification evaluation in Rapidminer

- **Dataset:** risk
week 09 – datasets.xls
- **Algorithms:** classification

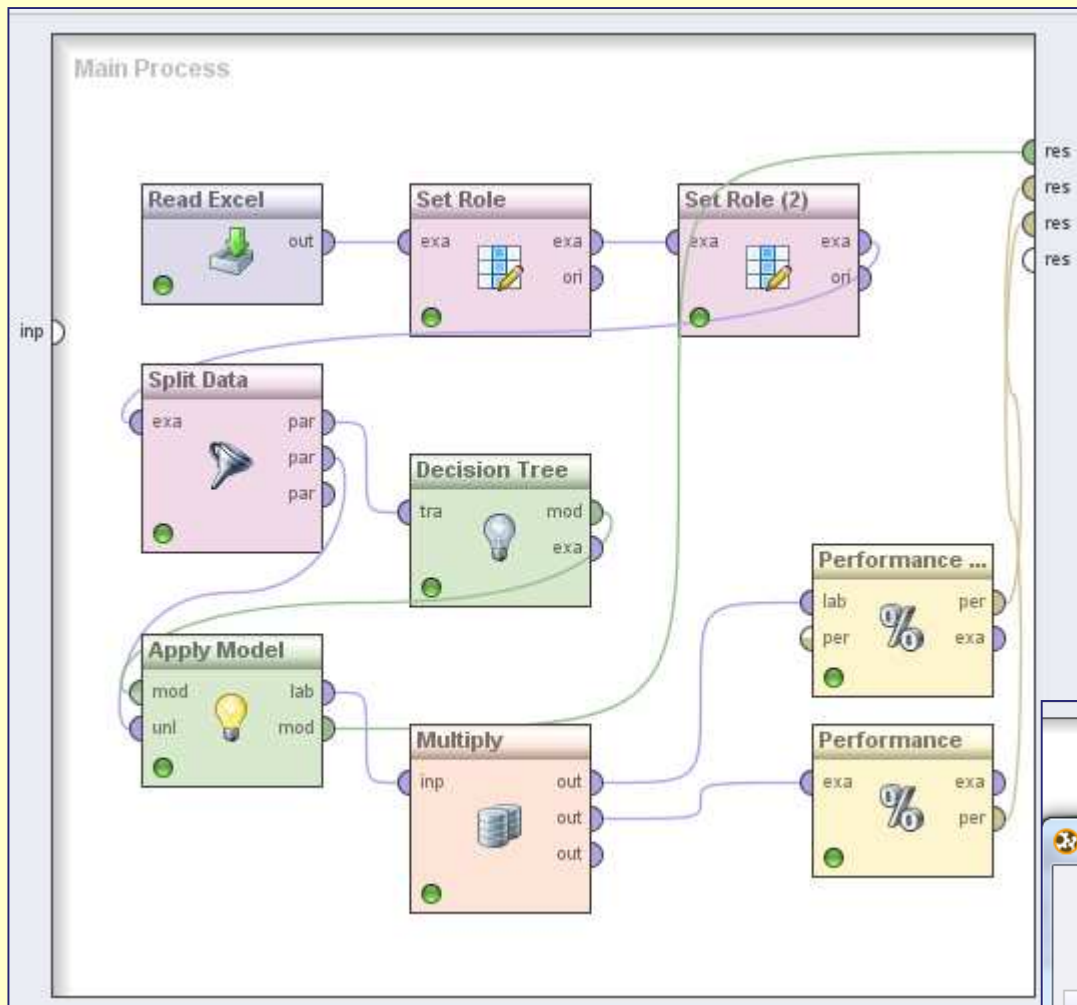
Decision tree for the risk dataset



accuracy: 75.79%

	true good risk	true bad loss	true bad profit	class precision
pred. good risk	171	27	43	70.95%
pred. bad loss	5	101	12	85.59%
pred. bad profit	77	135	664	75.80%
class recall	67.59%	38.40%	92.35%	

Computing the cost of the solution.



% Performance (Performance (Costs))

cost matrix Edit Matrix (3 x 3)...

Edit Parameter Matrix: cost matrix

The matrix of misclassification costs. Columns and Rows in order of internal mapping.

Cost Matrix	True Class 1	True Class 2	True Class 3
Predicted Class 1	0.0	1.0	1.0
Predicted Class 2	1.0	0.0	1.0
Predicted Class 3	1.0	1.0	0.0

accuracy: 75.79%				
	true good risk	true bad loss	true bad profit	class precision
pred. good risk	171	27	43	70.95%
pred. bad loss	5	101	12	85.59%
pred. bad profit	77	135	664	75.80%
class recall	67.59%	38.40%	92.35%	

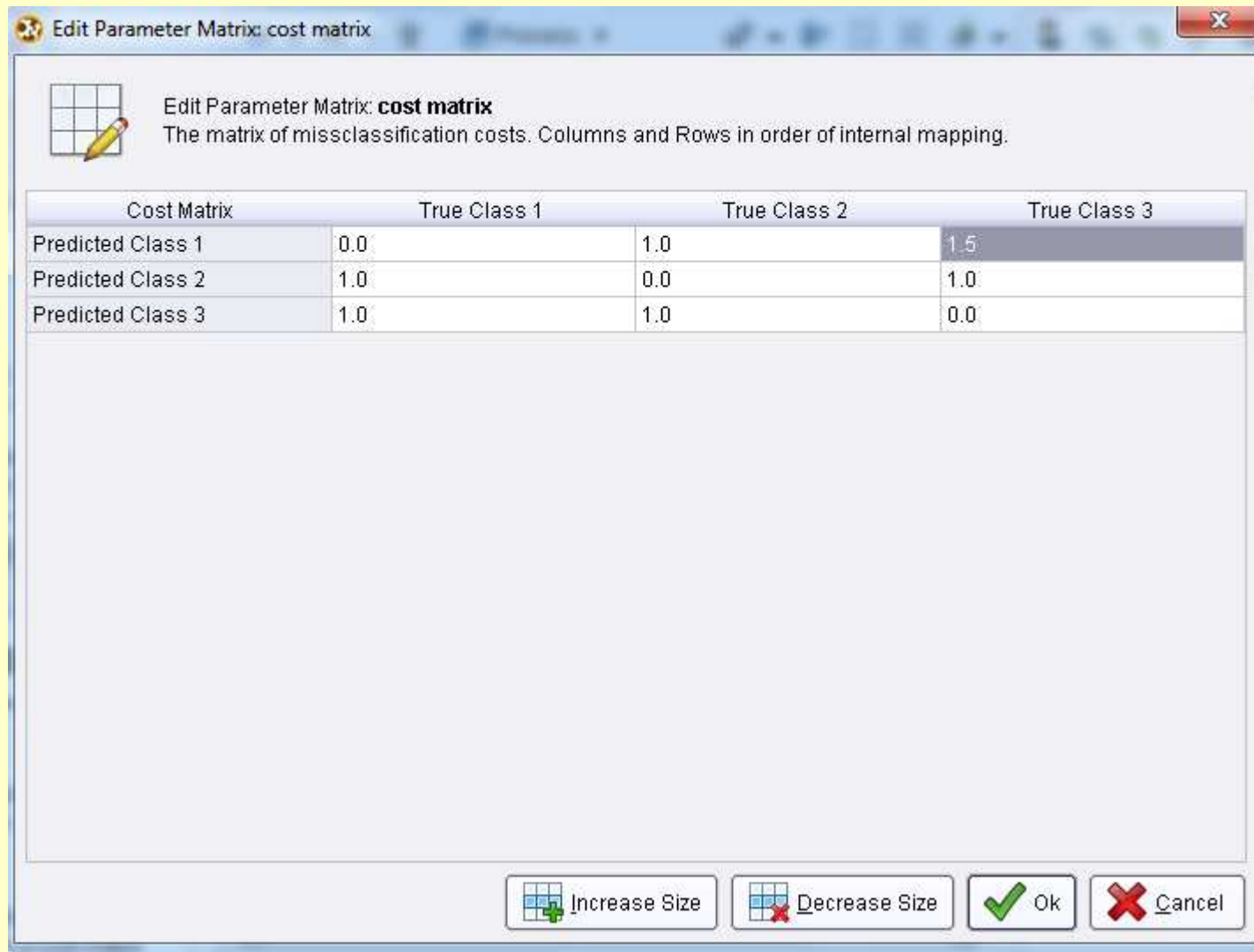
% PerformanceVector (Performance) ✕
% PerformanceVector (Performance (2)) ✕
💡 Tree (Decision Tree) ✕

Text View Annotations

Misclassificationcosts

Misclassificationcosts: 0.242 +/- 0.000

Changing the cost matrix values.



Edit Parameter Matrix: **cost matrix**
The matrix of misclassification costs. Columns and Rows in order of internal mapping.

Cost Matrix	True Class 1	True Class 2	True Class 3
Predicted Class 1	0.0	1.0	1.5
Predicted Class 2	1.0	0.0	1.0
Predicted Class 3	1.0	1.0	0.0

Increase Size Decrease Size Ok Cancel

Computing the cost of the solution while giving different costs to different errors.

Table View Plot View

accuracy: 75.79%

	true good risk	true bad loss	true bad profit	class precision
pred. good risk	171	27	43	70.95%
pred. bad loss	5	101	12	85.59%
pred. bad profit	77	135	664	75.80%
class recall	67.59%	38.40%	92.35%	

Confusion matrix is the same...

Text View Annotations

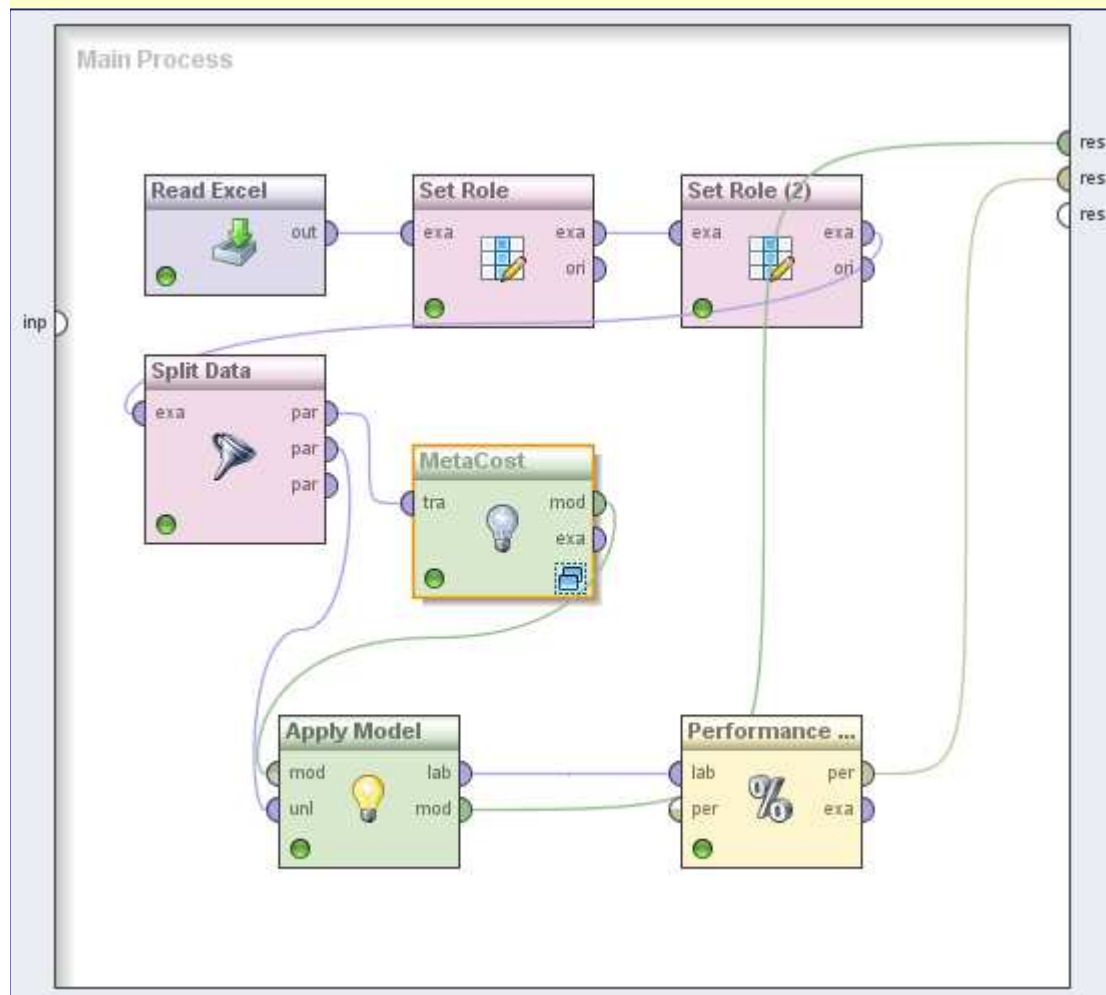
Misclassification costs

Misclassification costs: 0.260 +/- 0.000

but the overall cost is not.

<i>Confusion matrix</i>				<i>Cost matrix</i>			
	true good risk	true bad loss	true bad profit	class precision			
pred. good risk	171	27	43	70.95%	0	1	1.5
pred. bad loss	5	101	12	85.59%	1	0	1
pred. bad profit	77	135	664	75.80%	1	1	0
class recall	67.59%	38.40%	92.35%				
					<i>product of each pair of cells</i>		
			n	1235	0	27	64.5
			errors	299	5	0	12
			% error	0.24211	77	135	0
					sum		0.260

Using the cost matrix when building the model



The screenshot shows the 'MetaCost' interface with the following settings:

- cost matrix:** [Grid icon]
- use subset for training:** 0.8
- iterations:** 1
- sampling with replacement*
- use local random seed*
- parallelize learning process*

The 'Edit Parameter Matrix: cost matrix' dialog box is open, showing the following table:

Cost Matrix	True Class 1	True Class 2	True Class 3
Predicted Class 1	0.0	1.0	1.0
Predicted Class 2	1.0	0.0	1.0
Predicted Class 3	1.0	1.0	0.0

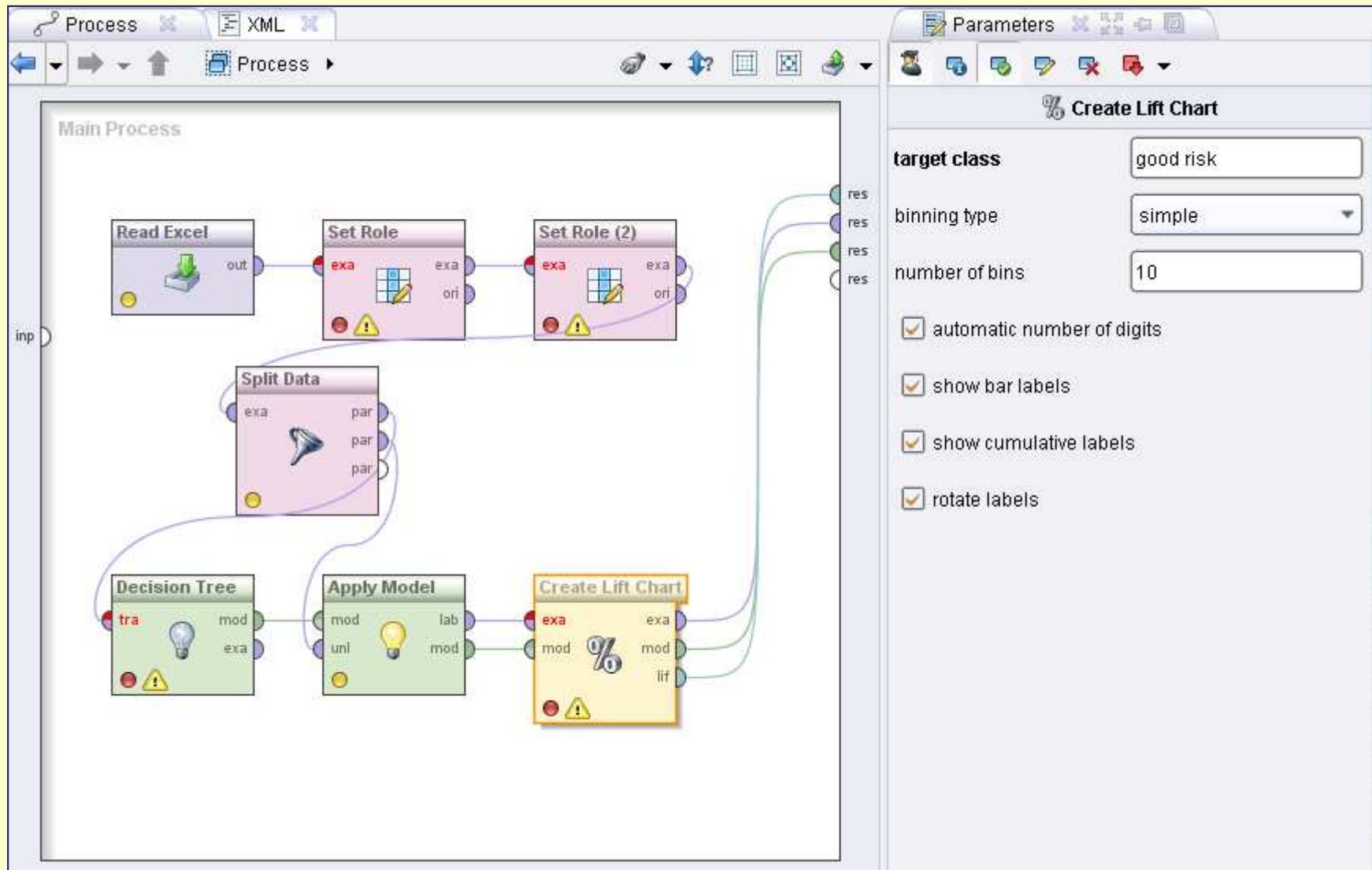
Multiclass Classification Performance Annotations

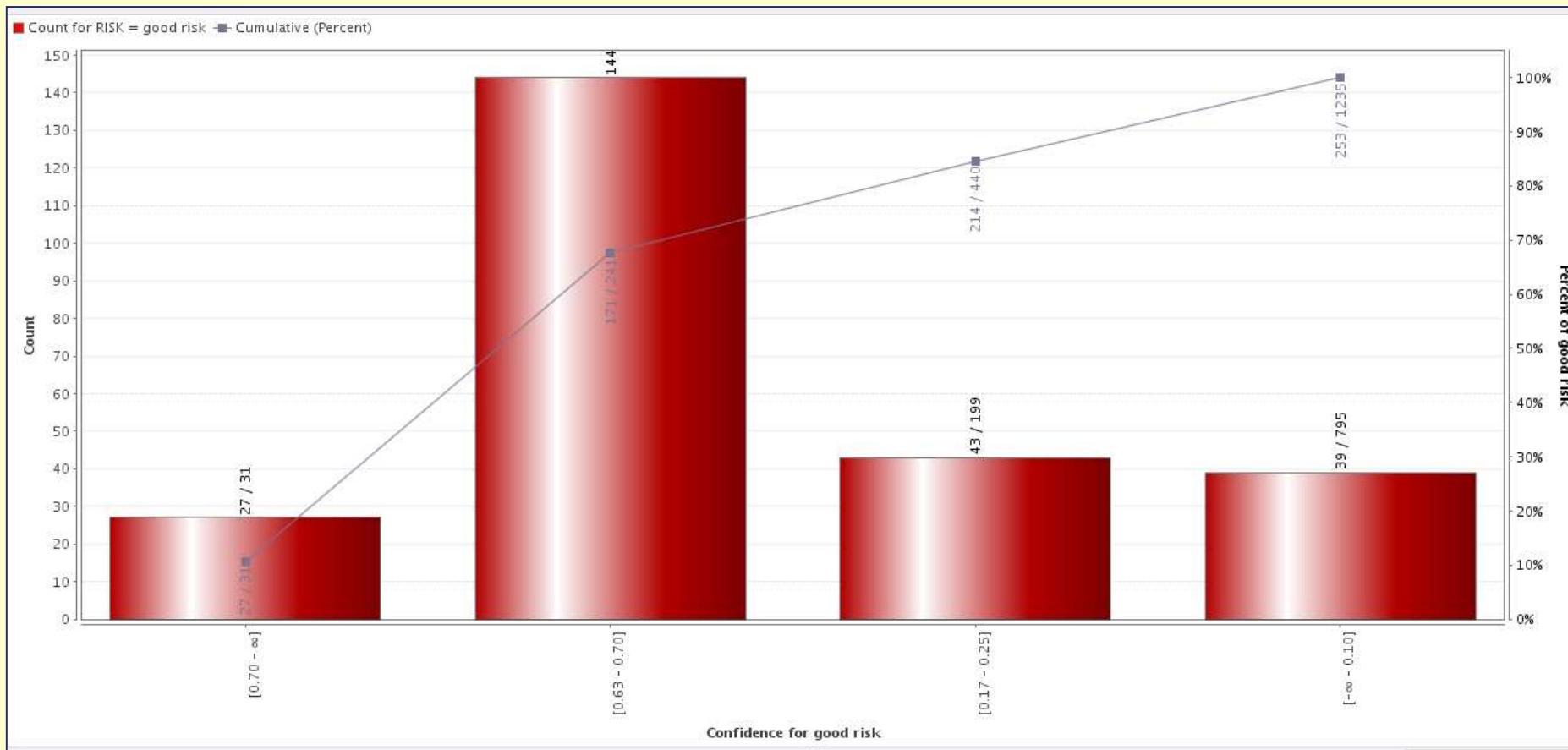
Table View Plot View

accuracy: 73.93%

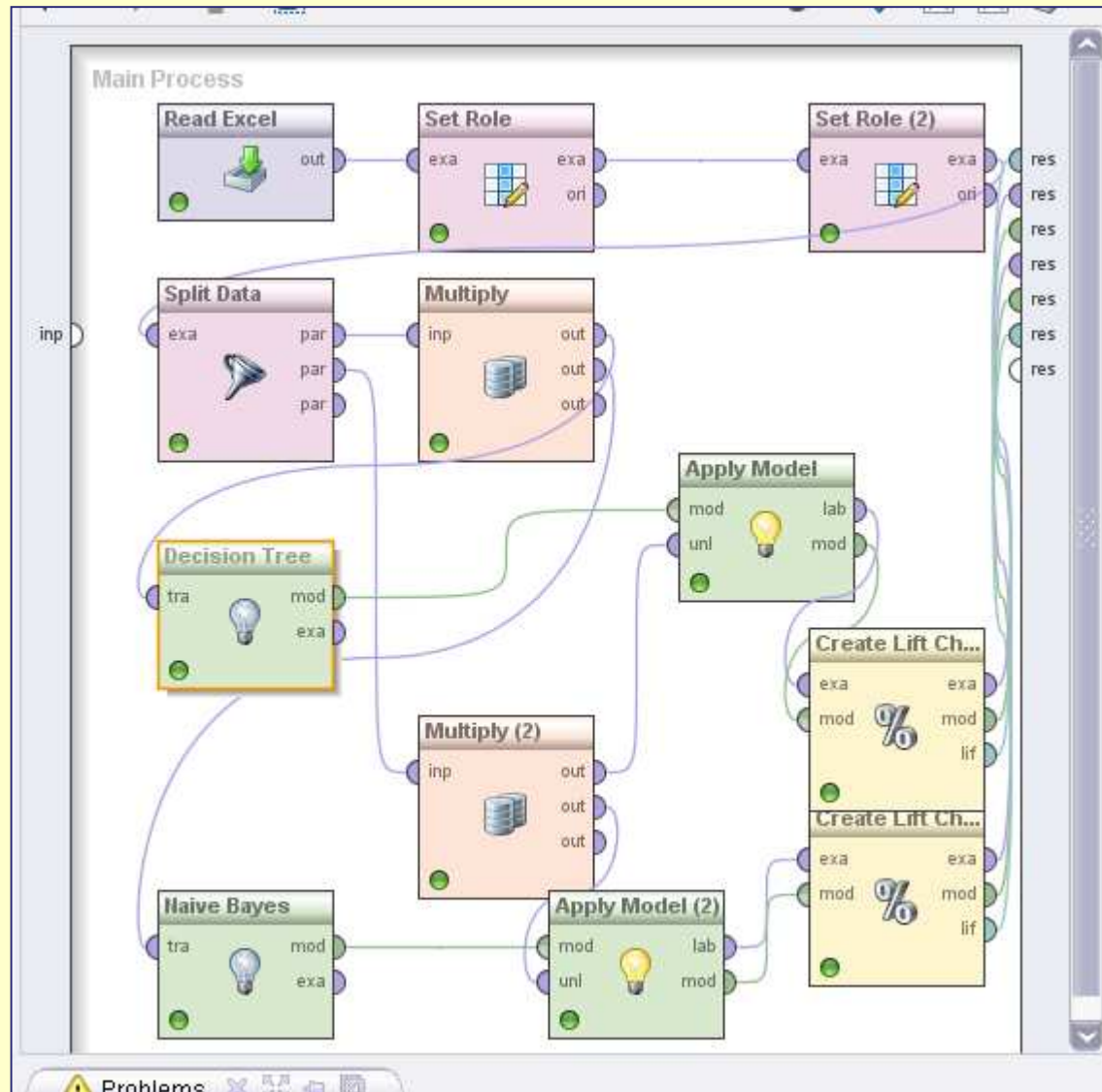
	true good risk	true bad loss	true bad profit	class precision
pred. good risk	144	27	39	68.57%
pred. bad loss	5	101	12	85.59%
pred. bad profit	104	135	668	73.65%
class recall	56.92%	38.40%	92.91%	

creating a lift chart

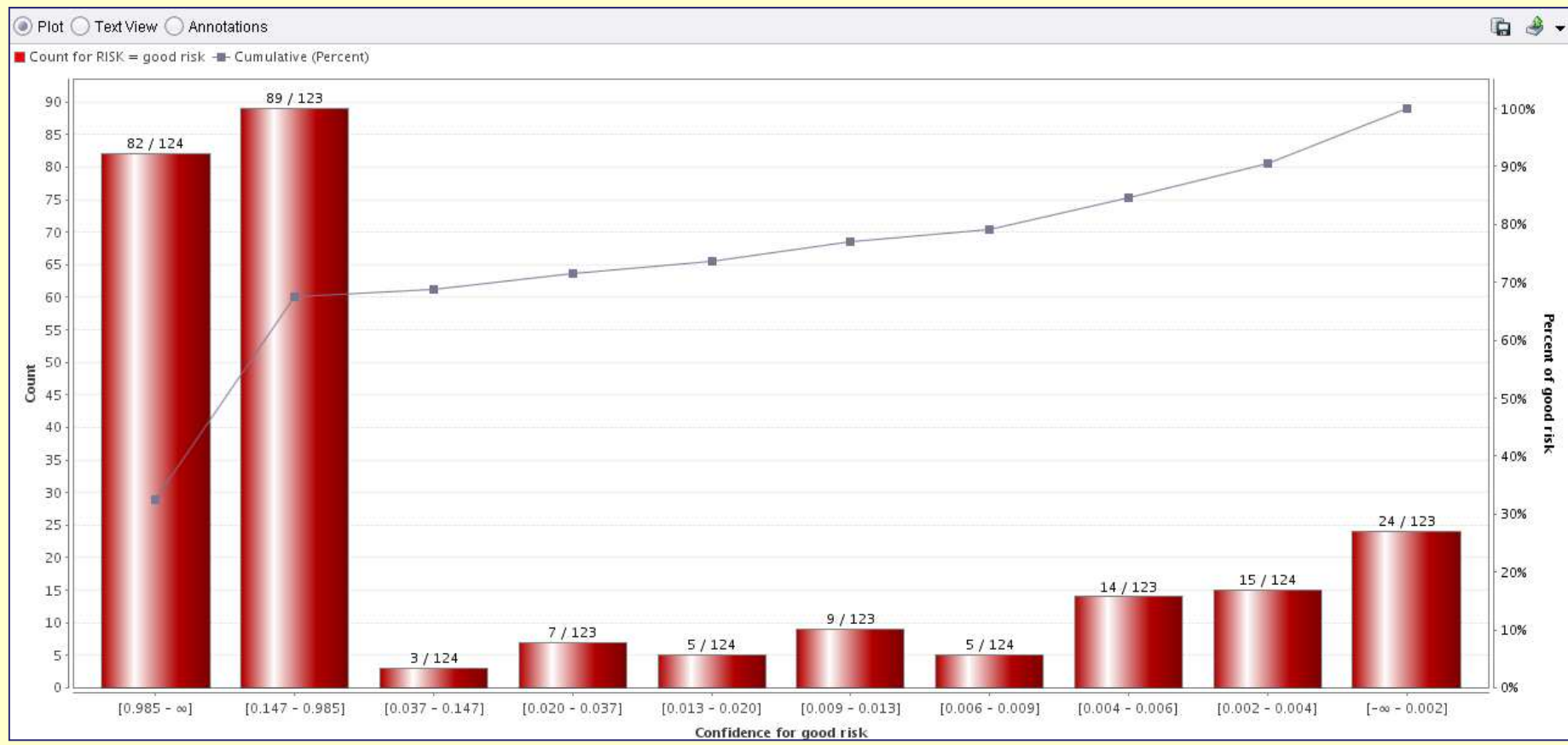




lift chart for two models



Lift chart for the naive Bayes model



Nominal to Binominal

create view

attribute filter type:

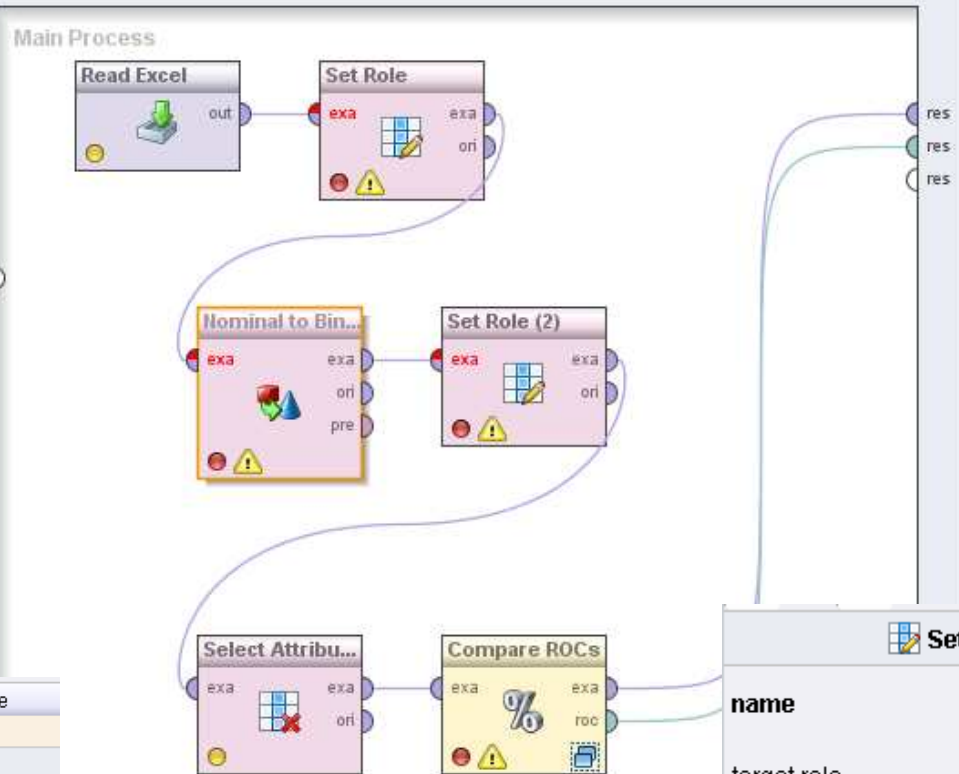
attribute:

invert selection

include special attributes

transform binominal

use underscore in name



Set Role (2) (Set Role)

name:

target role:

Role	Name	Type
id	ID	integer
regular	RISK = good risk	binominal
regular	RISK = bad loss	binominal
regular	RISK = bad profit	binominal
regular	AGE	integer
regular	INCOME	integer
regular	GENDER	binominal
regular	MARITAL	nominal
regular	NUMKIDS	integer
regular	NUMCARDS	integer
regular	HOWPAID	binominal
regular	MORTGAGE	binominal
regular	STORECAR	integer
regular	LOANS	integer

Select Attributes: attributes

Select Attributes: **attributes**
The attribute which should be chosen.

Attributes:

- ID
- RISK = bad loss
- RISK = bad profit

Selected Attributes:

- AGE
- GENDER
- HOWPAID
- INCOME
- LOANS
- MARITAL
- MORTGAGE
- NUMCARDS
- NUMKIDS
- RISK = good risk
- STORECAR

Ok Cancel

