

University of Belgrade
Faculty of Organizational Sciences
Center for Business Decision Making

WHIBO GDT Evolutionary Search Example Process

Belgrade, November 2013

WHIBO generic decision tree (GDT) evolutionary search operator implements genetic algorithm which selects reusable components defined in .ass (algorithm search space) file.

Parameters:

- Algorithm search space file location – location of .ass file
- Parameters – list of parameters of genetic algorithm
- Wba file path macro name – macro pointing to .wba file
- Log file path – path where log file will be saved

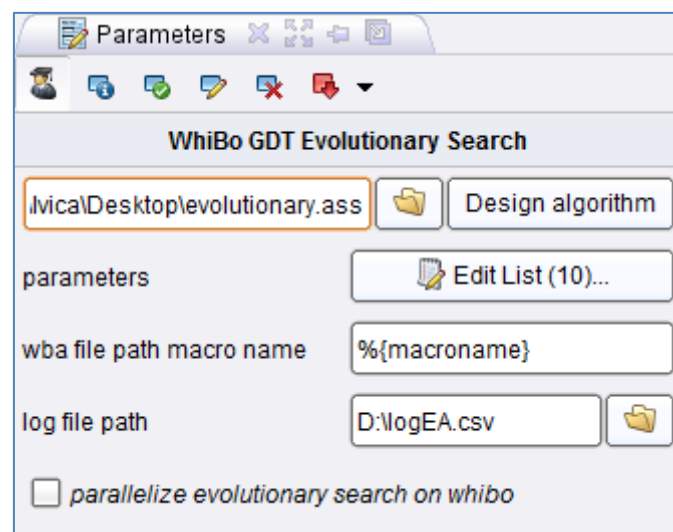


Figure 1 - Parameters for *WhiBo GDT Evolutionary Search*

User interface contains four panels:

Left panel contains an array of buttons. Every button represents a concrete sub-problem for a decision-tree algorithm design.

Central panel contains:

- Available RCs of selected sub-problem from the left panel.
- Available parameters (if available) for selected RCs.
- Buttons for including or disabling a RC from the current decision tree structure.

Right panel shows current state of user designed algorithm (saved sub-problems, RCs and parameters).

Top panel contains options for creating new, saving current or opening existing generic decision tree algorithm.

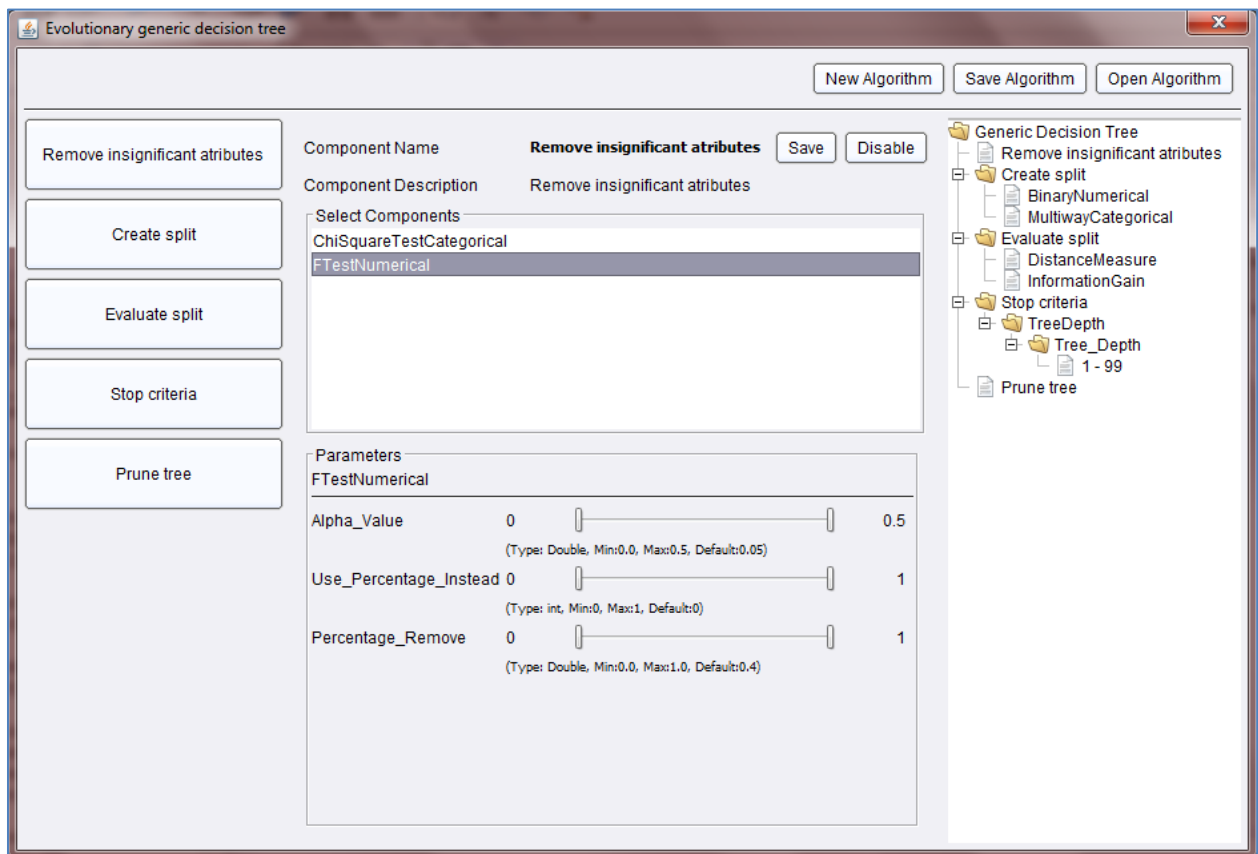


Figure 2 - WhiBo GDT evolutionary search user interface for design of algorithm search space

General procedure for designing algorithm search space:

- Select sub-problem from left panel. When sub-problem is selected, possible solutions (RCs) are shown in central panel.
- Select RC (or components if multiple) for sub-problem from central panel. If parameters for component(s) are available, they will be shown in bottom part of central panel with lower and upper values selected. User can modify these values.
- Click on save component button. Components and defined parameters for selected sub-problem will be shown in the right panel as part of current GDT algorithm.
- This procedure should be repeated for every sub-problem (Create split and Evaluate split sub-problem are basic for decision tree growth and they must be defined. Definition of other sub-problems is optional).

When all sub-problems, components and parameters are defined algorithm should be saved on file system (click on save button from upper panel). By default algorithms are saved with *.ass* (algorithm search space) extension.

After definition of algorithm search space parameters for genetic algorithm should be defined.

Parameters:

- MAX_ALLOWED_EVOLUTIONS – maximal numbers of generations of genetic algorithms (default value - 50).
- POPULATION_SIZE – number of units (decision trees) in one generation (default value - 30).
- MUTATION_RATE – percentage of genes (components) will be changed (default value - 6).
- CROSSOVER_RATE – rate of crossover of chromosomes in genetic algorithm (default value – 0.35)
- SWITCH_FROM_SURROGATE_PERCENTAGE_EVOLUTIONS – defines how many units should be removed from previous generation (default value – 0.4)
- SURROGATE_PERCENTAGE – defines how many units should be selected from previous generation (default value – 0.4)
- mutateComponents – boolean value indicating weather reusable components should be mutated (default value - true).
- mutateParameters – boolean value indicating weather parameters should be mutated (default value - true).
- componentsMutationRate – mutation rate of components (default value - 1).
- parametersMutationRate – mutation rate of parameters (default value - 1).

Edit Parameter List: parameters

Edit Parameter List: **parameters**
The parameters.

parameter name	values
MAX_ALLOWED_EVOLUTIONS	50
POPULATION_SIZE	30
MUTATION_RATE	6
CROSSOVER_RATE	0.35
SWITCH_FROM_SURROGATE_PERCENTAGE_EVOLUTION:	0.4
SURROGATE_PERCENTAGE	0.3
mutateComponents	true
mutateParameters	false
componentMutationRate	1
parametersMutationRate	1

Add Entry Remove Entry Apply Cancel

Figure 3 - Parameters of genetic algorithm

WHIBO example process

1. Import dataset into main RapidMiner process. For this purpose Iris dataset will be imported. It can be found in Samples RapidMiner repository in data folder.

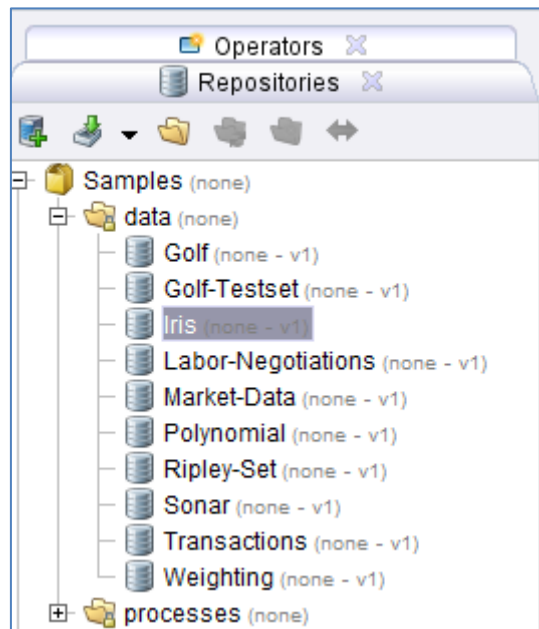


Figure 4 - Sample Repository

Drag Iris dataset into main RapidMiner process.

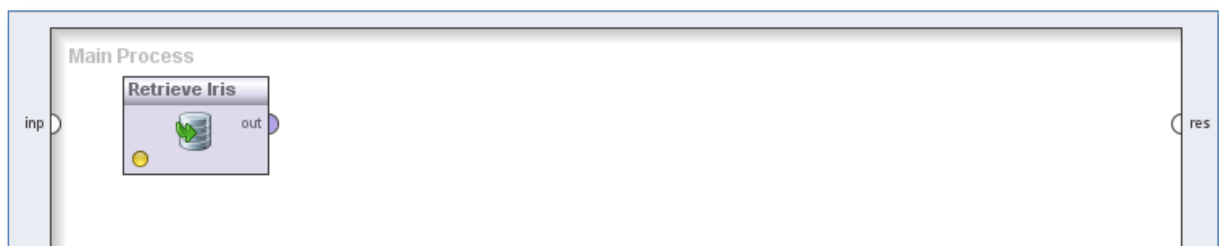


Figure 5 - Main RapidMiner process (1)

2. Import **WhiBo GDT Evolutionary Search** operator.

Generic decision tree operator can be found inside *Operators* panel, in *Whibo* folder.

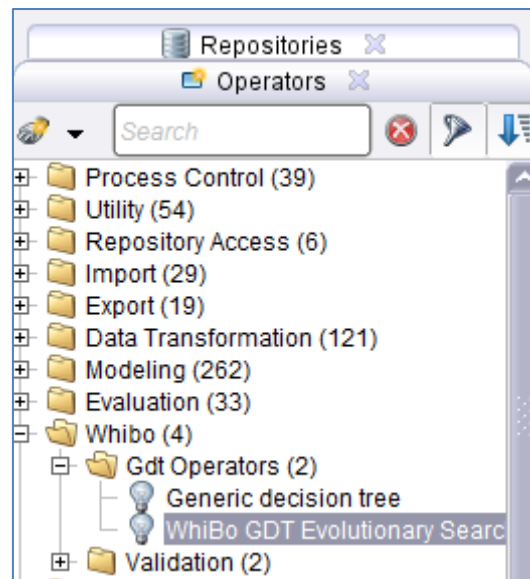


Figure 6 – GDT Evolutionary search operator

Click on this operator and drag it into main RapidMiner process.

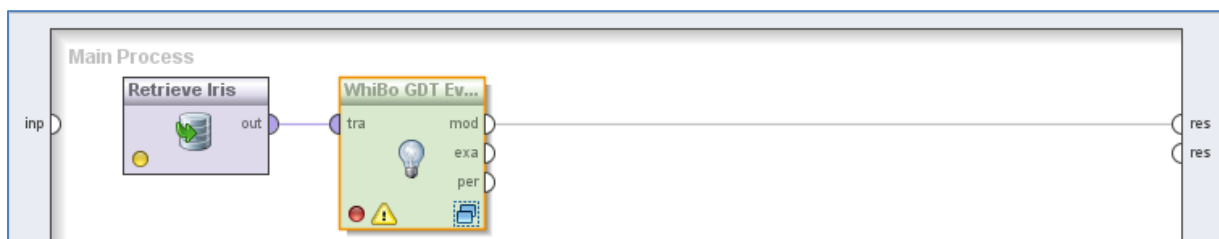


Figure 7 – Main RapidMiner process (2)

Since *GDT Evolutionary search* is complex operator it is needed to create subprocess for him.

3. Create subprocess for **Whibo GDT Evolutionary Search** operator.

Double click on **Whibo GDT Evolutionary Search** operator and subprocess will show up.

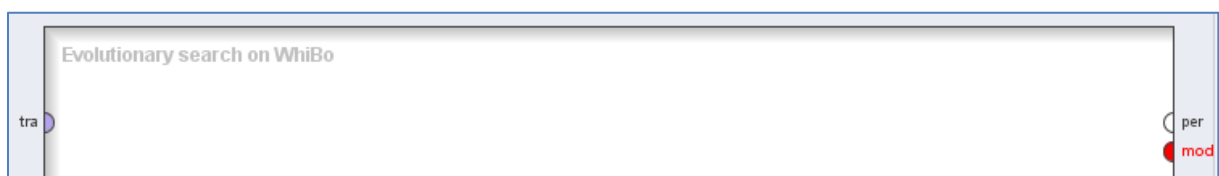


Figure 8 - Whibo GDT Evolutionary Search subprocess (1)

It is recommended to insert any *Validation* operator in this subprocess. In this example it will be **X-Validation** operator.

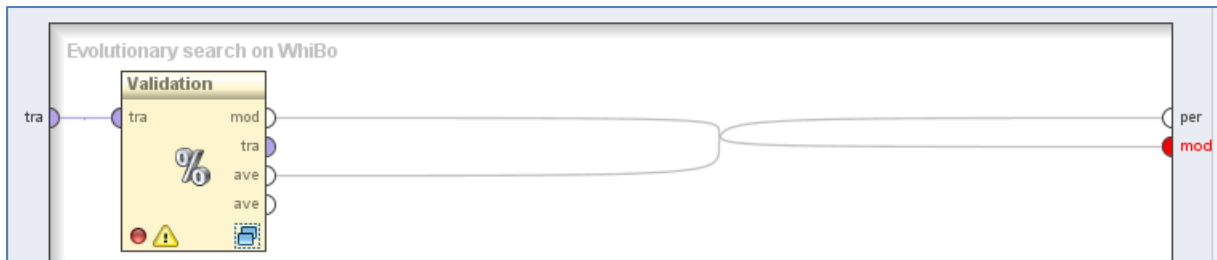


Figure 9 - Whibo GDT Evolutionary Search subprocess (2)

Validation is also complex operator and has its own subprocess, which contains two parts. First one is *Training*, where *Generic decision tree* operator **must be inserted**. Second part is *Test* panel, where *Apply model* and *Performance (Classification)* operators are inserted.

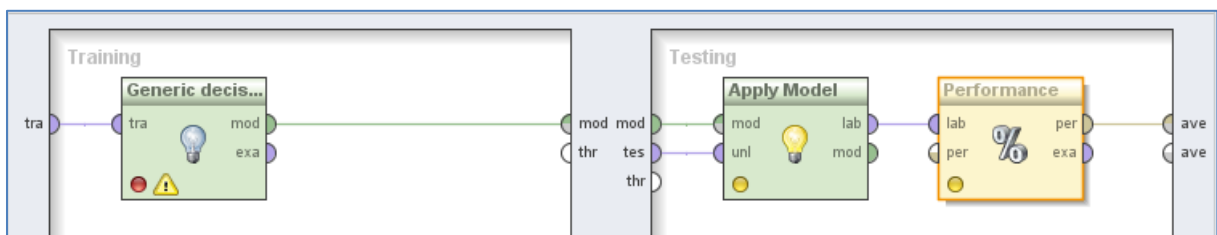


Figure 10 – X-Validation subprocess

Generic decision tree operator must have algorithm definition file. In this example CART is created (from Generic decision tree tutorial).

Parameters of X-Validation are not important for this example process and will be default values.

4. Set parameters of **Whibo GDT Evolutionary Search** operator.

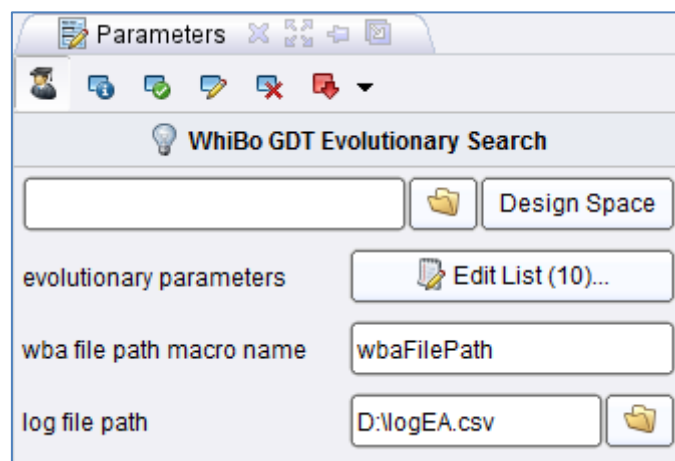


Figure 11 – Parameters panel of *Whibo GDT Evolutionary Search* operator

First, algorithm search space (ass) must be defined. Click on **Design Space** button.

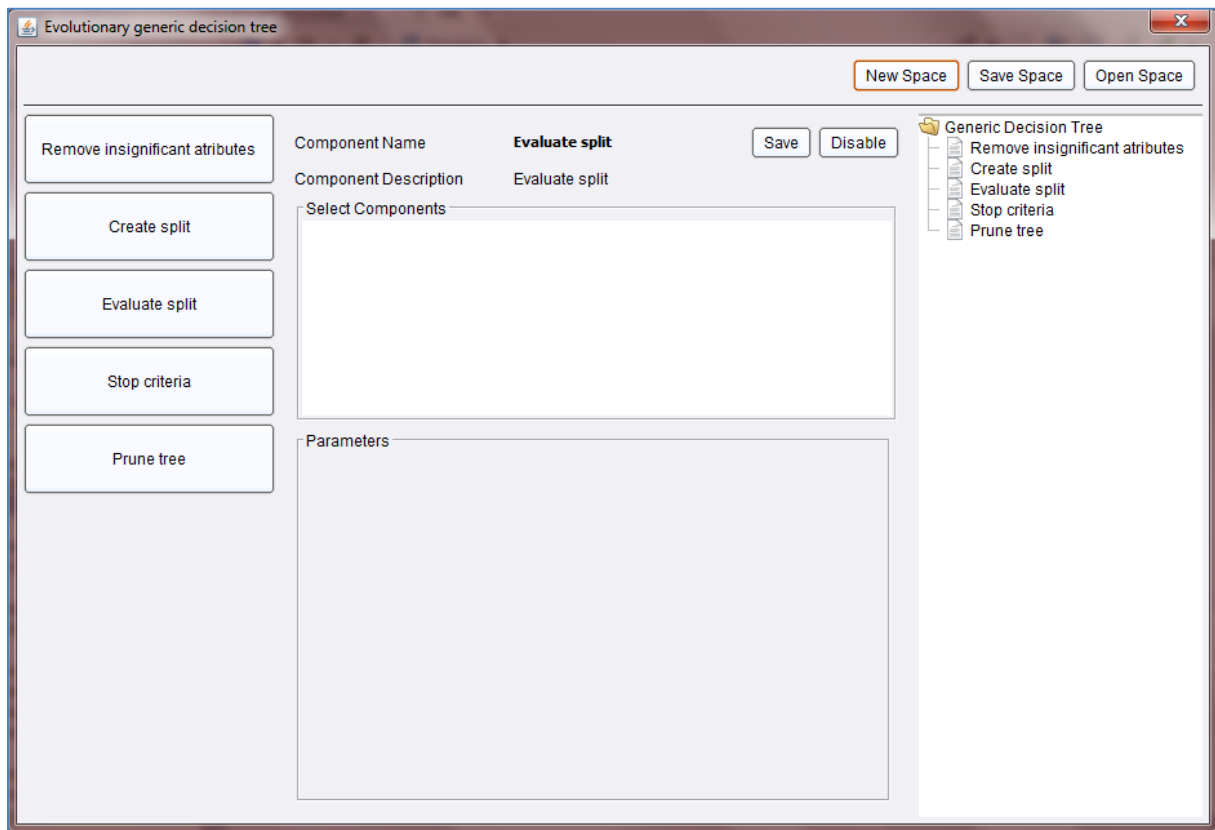


Figure 12 – Algorithm search space dialog (1)

Select components for each subproblem. In this example following components will be chosen:

- Remove insignificant attributes – none.
- Create split – *BinaryCategorical*, *BinaryNumerical*, *MultiwayCategorical* and *SignificantCategorical* with default parameters.
- Evaluate split – *ChiSquare*, *GainRatio* and *GiniIndex*.
- Stop criteria – none.
- Prune tree – none.

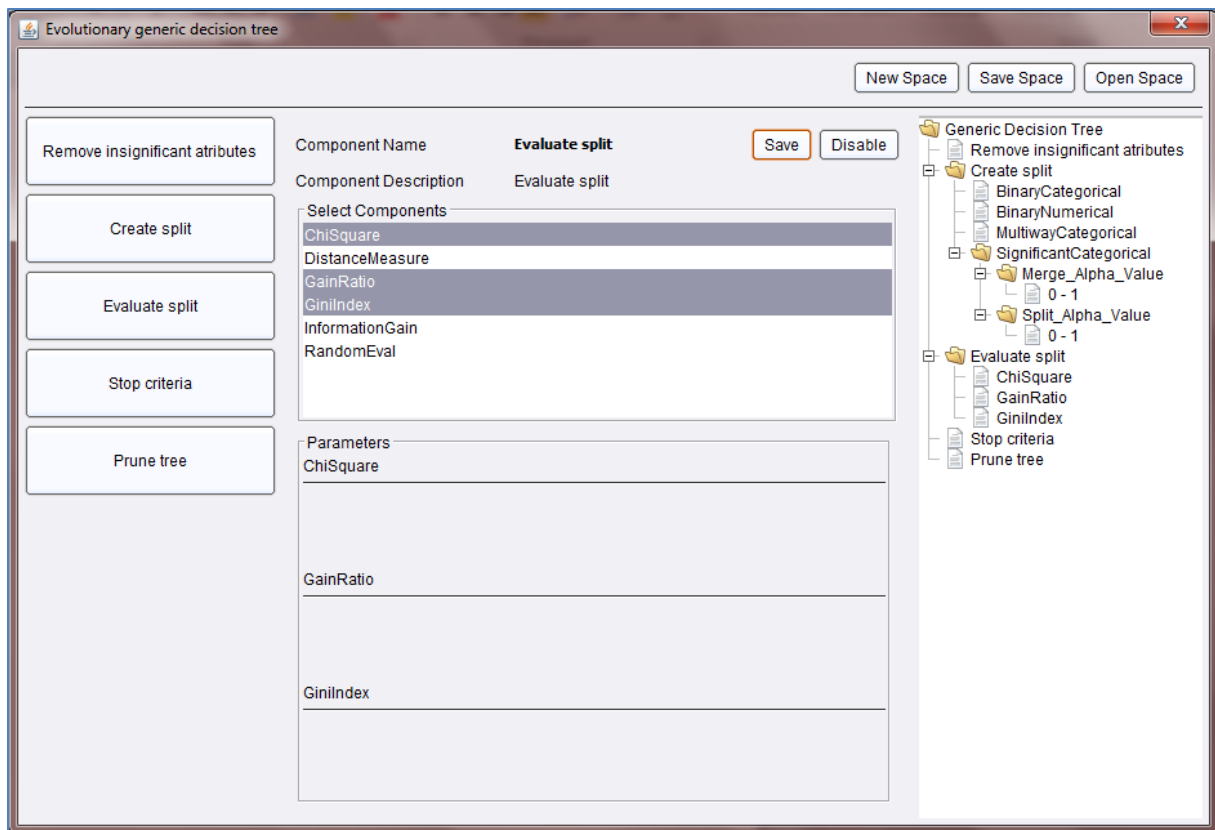


Figure 13 – Algorithm search space dialog (2)

Click **Save Space** button on top right part of the screen and save algorithm search space on local hard drive. After that, load .ass file by clicking **Folder** button in *Parameters* panel.

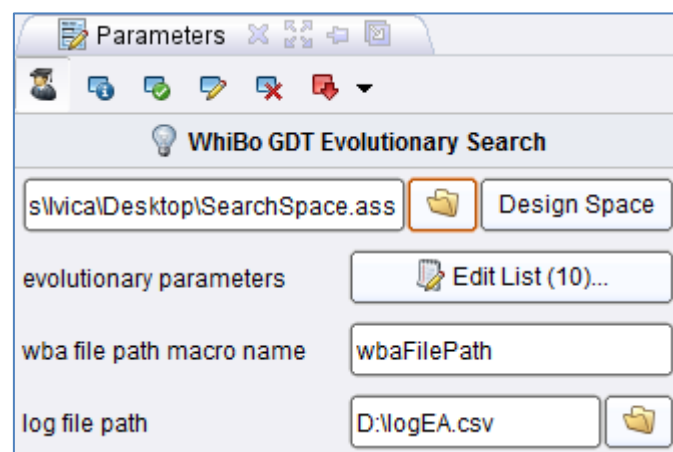


Figure 14 – Parameters panel after selecting .ass file

Since this operator use evolutionary algorithm to search across algorithm search pace for best combination of components it has to have parameters for evolutionary algorithm. These parameters are listed next to *evolutionary parameters* label.

Edit Parameter List: evolutionary parameters

The parameters.

parameter name	values
MAX_ALLOWED_EVOLUTIONS	50
POPULATION_SIZE	30
MUTATION_RATE	6
CROSSOVER_RATE	0.35
SWITCH_FROM_SURROGATE_PERCENTAGE_EVOLUTION:	0.4
SURROGATE_PERCENTAGE	0.3
mutateComponents	true
mutateParameters	false
componentMutationRate	1
parametersMutationRate	1

Buttons: Add Entry, Remove Entry, Apply, Cancel

Figure 15 – Evolutionary algorithm parameters

Evolutionary parameters are already explained.

wba file path macro name must be declared. Because of that *Set Macro* operator must be defined in main RapidMiner process, which will show file path to .wba file.

Parameters

Set Macro

macro: wbaFile

value: Tutorija/wba files\CART.wba

Figure 16 – Set Macro operator Parameters panel

This macro point to .wba file and its name is wbaFile. In **Whibo GDT Evolutionary Search** operator *Parameters* panel in field for *wba file path macro name* macro must be inserted like `%{macroName}`, where *macroName* is name of defined Macro.

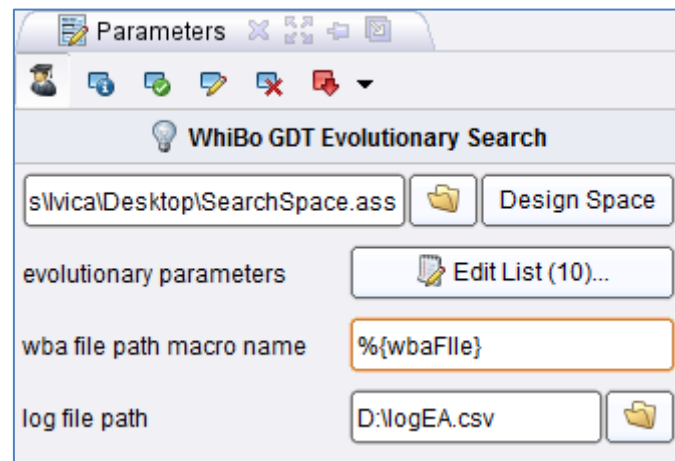


Figure 17 – Setting macro name

Also path to log file must be defined. In log file every iteration is saved, with used components and its parameters (if component have it), selected performance and time needed to finish that iteration.

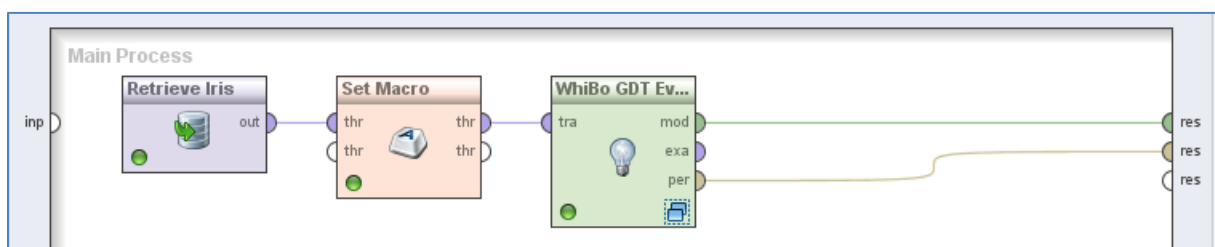


Figure 18 - Main RapidMiner process (3)

Final process looks like this. **Whibo GDT Evolutionary Search** operator generates two outputs. First one is model, which generate decision tree model. Second generates performance of operator.

5. Click **Run** button.

When RapidMiner finish executing process it outputs of **Whibo GDT Evolutionary Search** operator are shown.

accuracy: 96.00% +/- 3.27% (mikro: 96.00%)				
	true Iris-setosa	true Iris-versicolor	true Iris-virginica	class precision
pred. Iris-setosa	50	0	0	100.00%
pred. Iris-versicolor	0	47	3	94.00%
pred. Iris-virginica	0	3	47	94.00%
class recall	100.00%	94.00%	94.00%	

Figure 19 – Performance

Performance is shown as confusion matrix, while model is shown as decision tree.

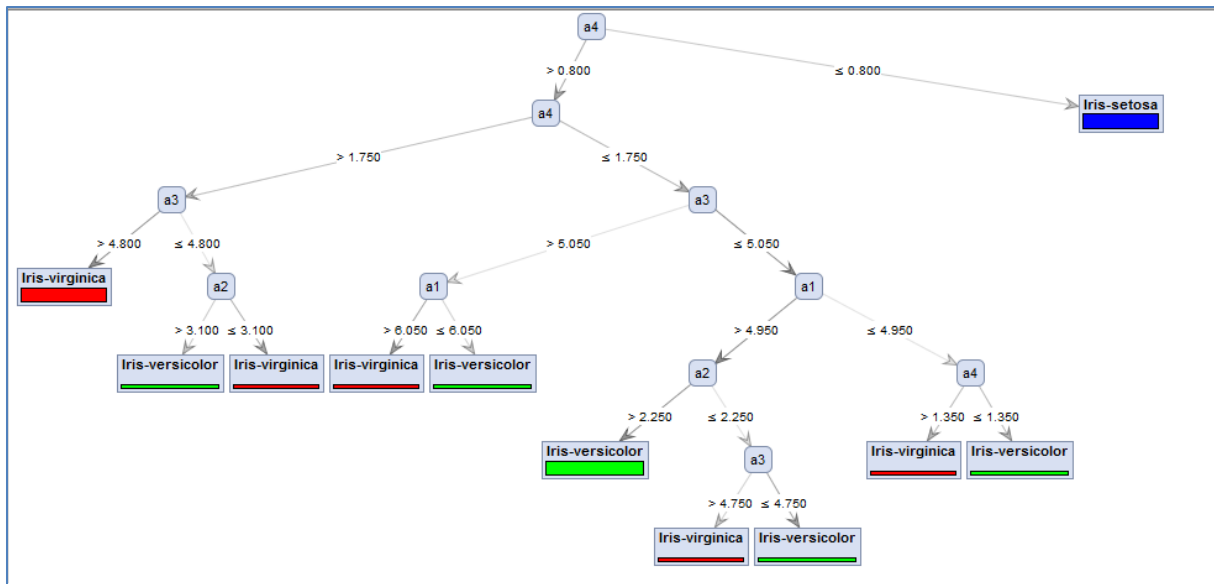


Figure 20 – Decision tree model

For any further information feel free to send e-mail ([contacts](#)) or post on [forum](#).