

# Rapid Miner

Tahar Kechadi

School of Computer Science & Informatics  
University College Dublin

## Outline

- Rapid Miner installation
- Rapid Miner key components
- Process view
- Parameter setting
- Process creation and execution
- Result views

# Rapid Miner Installation

- Installation Versions
  - Windows executable
    - Double click on “rapidminer-xxx.exe” and follow the instruction on the screen
  - Java version (for any other platform)
- Java Version
  - Unpack the downloaded archive file
  - Make sure you have **Java 6.0 or higher** installed on your machine
  - \$ cd <install location>/rapidminer
  - \$ ./scripts/RapidMinerGUI
  - *Follow the steps to create a new local repository (first time only)*

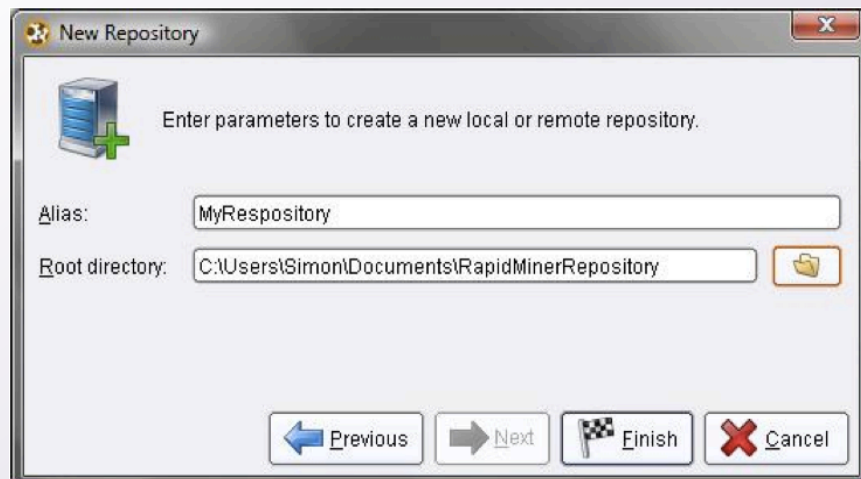
# Rapid Miner Installation

- Further details
  - **Wiki:** [http://rapid-i.com/wiki/index.php?title=Main\\_Page](http://rapid-i.com/wiki/index.php?title=Main_Page)
  - **User forum:** <http://forum.rapid-i.com/>
  - **Main website:** <http://rapid-i.com/>
- Tutorials
  - Available via the **Help -> rapidMiner Tutorial...** Menu option
  - Examples of processes are also available and many of them can be tested easily with the GUI version

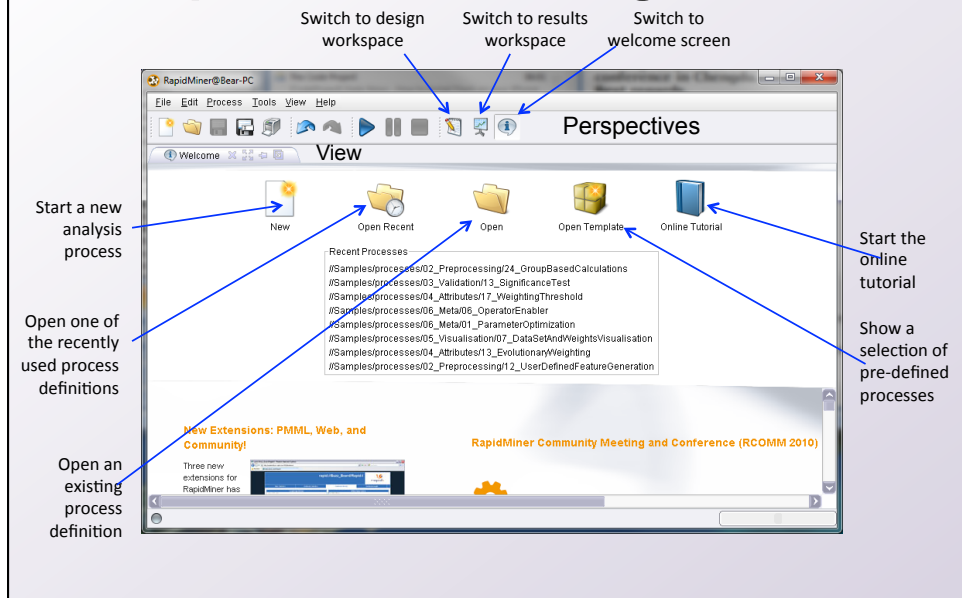
# RapidMiner Installation



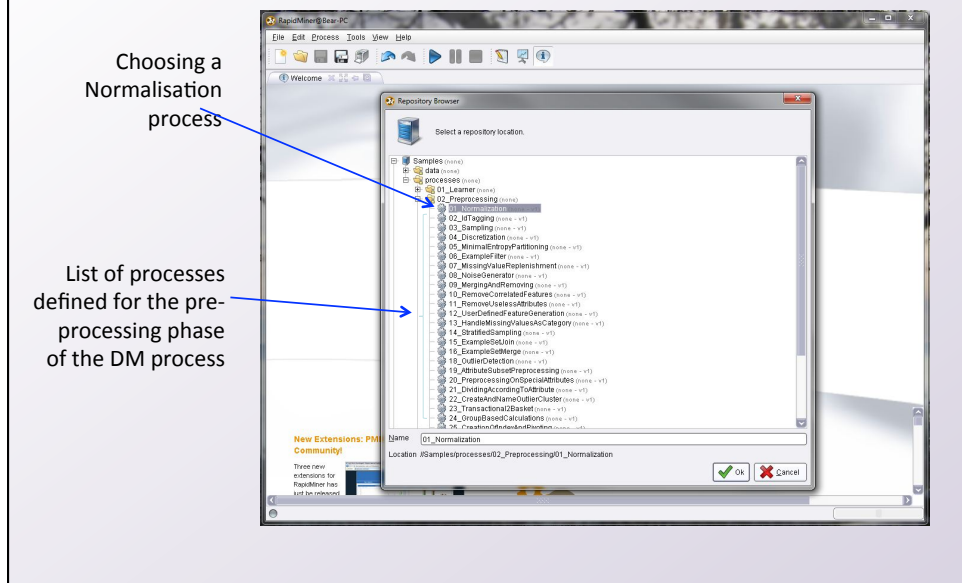
# RapidMiner Installation



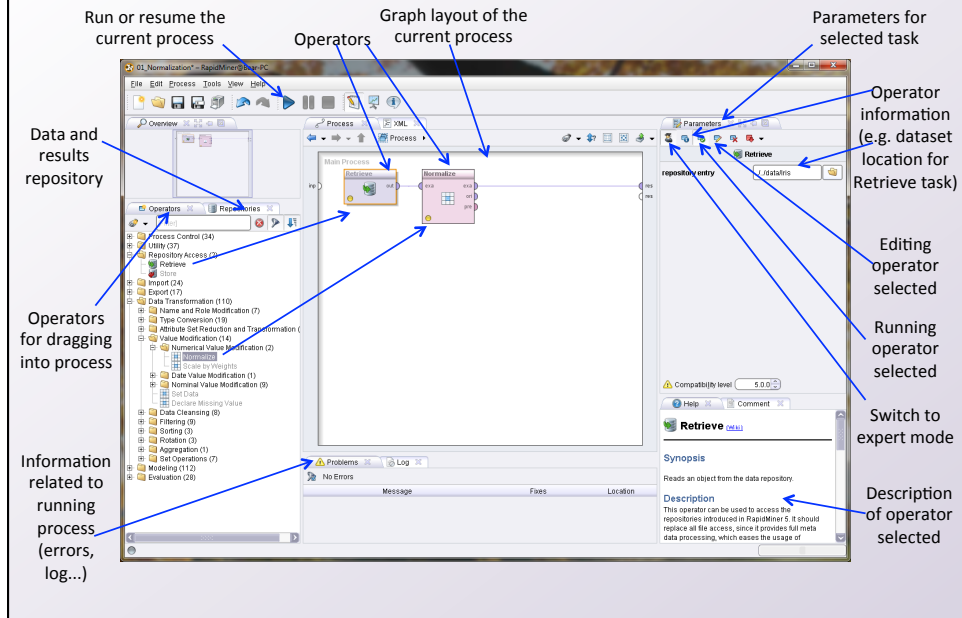
# Rapid Miner - Getting Started



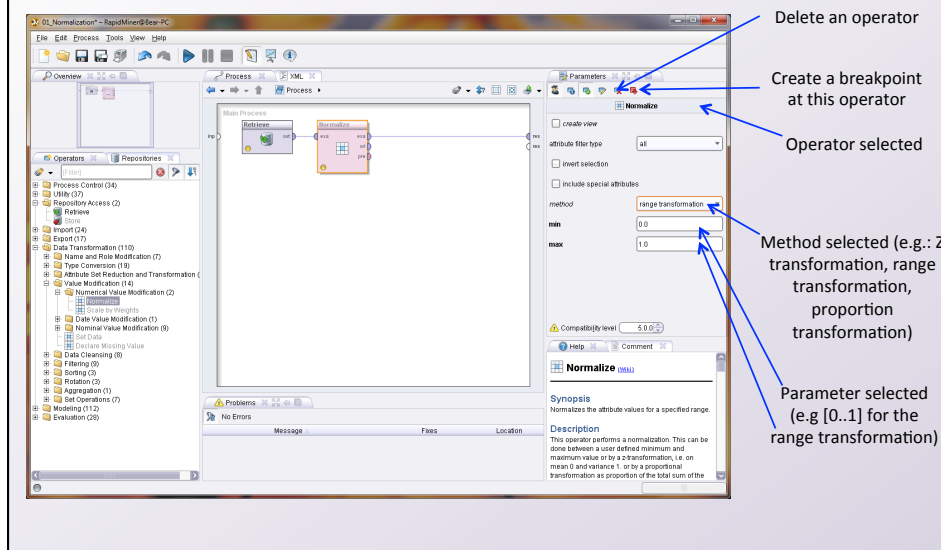
# Open Existing Process



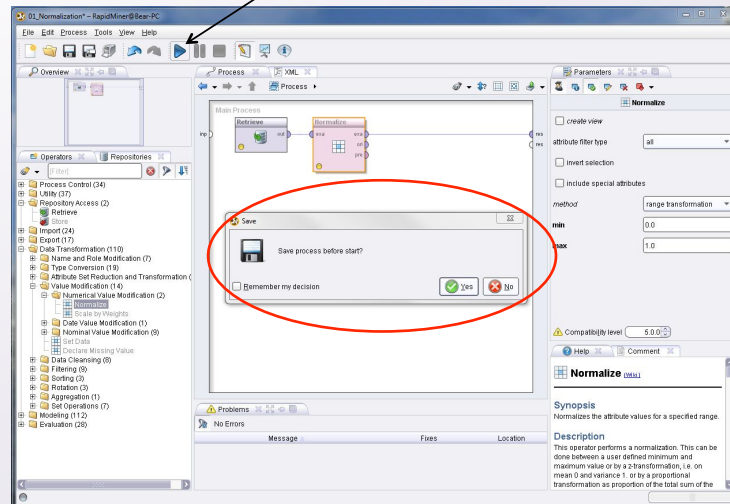
# Process View



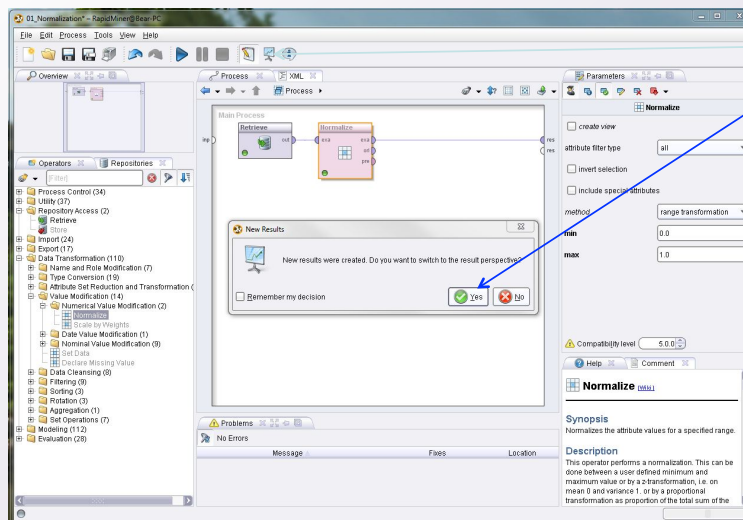
# Parameter Setting (Expert mode)



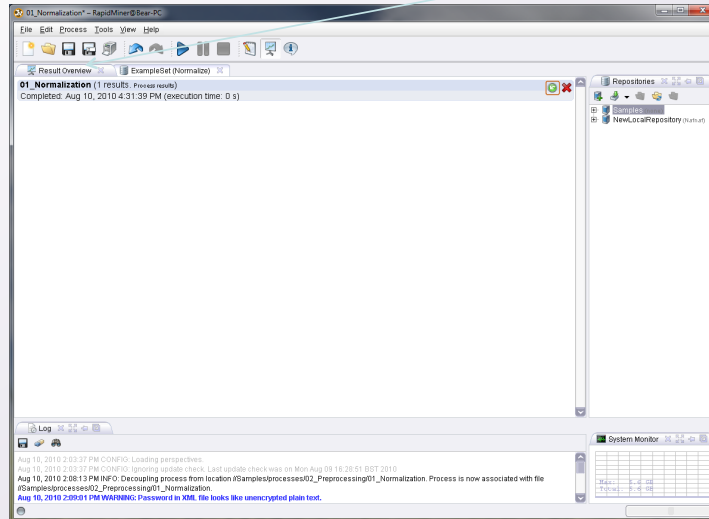
# Process Execution



# Process Execution



# Process Execution – Results Overview



# Process Execution – Results

Different views of results

Values standardised

Different options of view filter

Using Repository to open original dataset (e.g. to verify...)

ExampleSet (150 examples, 2 special attributes, 4 regular attributes)

View Filter (150 / 150): all

Repositories:

- data (source)
- Golf (source - v1)
- GolfTestset (source - v1)
- iris (source - v1)
- LoanNegotiation (source - v1)
- Market Data (source - v1)
- Polynomial (source - v1)
- RipleySet (source - v1)
- Sonar (source - v1)
- Transactions (source - v1)
- Weighting (source - v1)
- processes (source)
- NewLocalRepository (RAM)
- Test (RAM - v1, 8/10/10 11:10 AM - 718 bytes)

## Results vs. Original Dataset

ExampleSet (150 examples, 2 special attributes, 4 regular attributes)

Row No.	id	label	a1	a2	a3	a4
1	id_1	lrs-setosa	5.100	3.500	1.400	0.200
2	id_2	lrs-setosa	4.900	3	1.400	0.200
3	id_3	lrs-setosa	4.700	3.200	1.300	0.200
4	id_4	lrs-setosa	4.600	3.100	1.500	0.200
5	id_5	lrs-setosa	5	3.600	1.400	0.200
6	id_6	lrs-setosa	5.400	3.900	1.700	0.400
7	id_7	lrs-setosa	4.600	3.400	1.400	0.300
8	id_8	lrs-setosa	5	3.400	1.500	0.200
9	id_9	lrs-setosa	4.400	2.900	1.400	0.200
10	id_10	lrs-setosa	4.900	3.100	1.500	0.100
11	id_11	lrs-setosa	5.400	3.700	1.500	0.200
12	id_12	lrs-setosa	4.800	2.400	1.600	0.200
13	id_13	lrs-setosa	4.800	3	1.400	0.100
14	id_14	lrs-setosa	4.300	3	1.100	0.100
15	id_15	lrs-setosa	5.800	4	1.200	0.200
16	id_16	lrs-setosa	5.700	4.400	1.500	0.400
17	id_17	lrs-setosa	5.400	3.900	1.300	0.400
18	id_18	lrs-setosa	5.100	3.500	1.400	0.300
19	id_19	lrs-setosa	5.700	3.800	1.700	0.300
20	id_20	lrs-setosa	5.100	3.600	1.500	0.300
21	id_21	lrs-setosa	5.400	3.400	1.700	0.200
22	id_22	lrs-setosa	1.00	3.700	1.500	0.400
23	id_23	lrs-setosa	4.900	2.600	1	0.600
24	id_24	lrs-setosa	5.100	3.300	1.700	0.200
25	id_25	lrs-setosa	4.800	4.000	1.800	0.200

ExampleSet (150 examples, 2 special attributes, 4 regular attributes)

Row No.	id	label	a1	a2	a3	a4
1	id_1	lrs-setosa	0.222	0.625	0.068	0.147
2	id_2	lrs-setosa	0.191	0.417	0.068	0.042
3	id_3	lrs-setosa	0.177	0.300	0.051	0.042
4	id_4	lrs-setosa	0.083	0.458	0.085	0.042
5	id_5	lrs-setosa	0.194	0.667	0.068	0.042
6	id_6	lrs-setosa	0.306	0.792	0.119	0.125
7	id_7	lrs-setosa	0.083	0.583	0.068	0.083
8	id_8	lrs-setosa	0.194	0.583	0.085	0.042
9	id_9	lrs-setosa	0.028	0.375	0.068	0.042
10	id_10	lrs-setosa	0.167	0.458	0.085	0
11	id_11	lrs-setosa	0.306	0.708	0.085	0.042
12	id_12	lrs-setosa	0.139	0.583	0.102	0.042
13	id_13	lrs-setosa	0.139	0.417	0.068	0
14	id_14	lrs-setosa	0	0.417	0.017	0
15	id_15	lrs-setosa	0.417	0.833	0.034	0.042
16	id_16	lrs-setosa	0.389	1	0.085	0.125
17	id_17	lrs-setosa	0.306	0.792	0.051	0.125
18	id_18	lrs-setosa	0.222	0.625	0.068	0.083
19	id_19	lrs-setosa	0.389	0.750	0.119	0.083
20	id_20	lrs-setosa	0.222	0.750	0.085	0.083
21	id_21	lrs-setosa	0.306	0.583	0.119	0.042
22	id_22	lrs-setosa	0.132	0.708	0.085	0.125
23	id_23	lrs-setosa	0.083	0.667	0	0.042
24	id_24	lrs-setosa	0.222	0.542	0.119	0.167
25	id_25	lrs-setosa	0.139	0.667	0.119	0.083

Original Dataset

## Results: Metadata view

ExampleSet (150 examples, 2 special attributes, 4 regular attributes)

Role	Name	Type	Statistics	Range	Missings
id	id	nominal	mode = id_1 (1); head = [id_1 (1), id_10 (1), id_100 (0]		
label	label	nominal	mode = lrs-setosa (50); lrs-setosa (50); lrs-versi: 0		
regular	a1	real	avg = 5.043 +/- 0.628	(4.300; 7.900]	0
regular	a2	real	avg = 3.054 +/- 0.436	(2.000; 4.400]	0
regular	a3	real	avg = 3.759 +/- 1.764	(1.000; 6.900]	0
regular	a4	real	avg = 1.198 +/- 0.763	(0.100; 2.500]	0

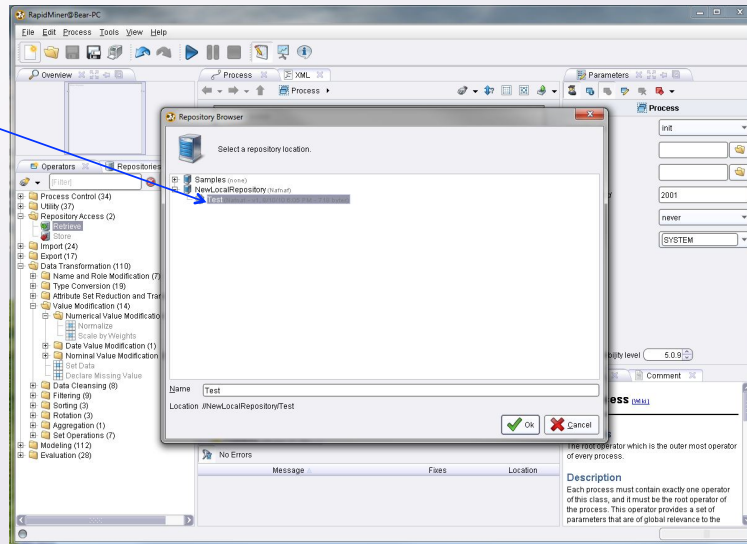
ExampleSet (150 examples, 2 special attributes, 4 regular attributes)

Role	Name	Type	Statistics	Range	Missings
id	id	nominal	mode = id_1 (1); lrs-setosa (50); lrs-versi: 0		
label	label	nominal	mode = lrs-setosa (50); lrs-versi: 0		
regular	a1	real	avg = 0.429 +/- 0.100	(0.000; 1.000]	0
regular	a2	real	avg = 0.429 +/- 0.100	(0.000; 1.000]	0
regular	a3	real	avg = 0.468 +/- 0.100	(0.000; 1.000]	0
regular	a4	real	avg = 0.458 +/- 0.100	(0.000; 1.000]	0



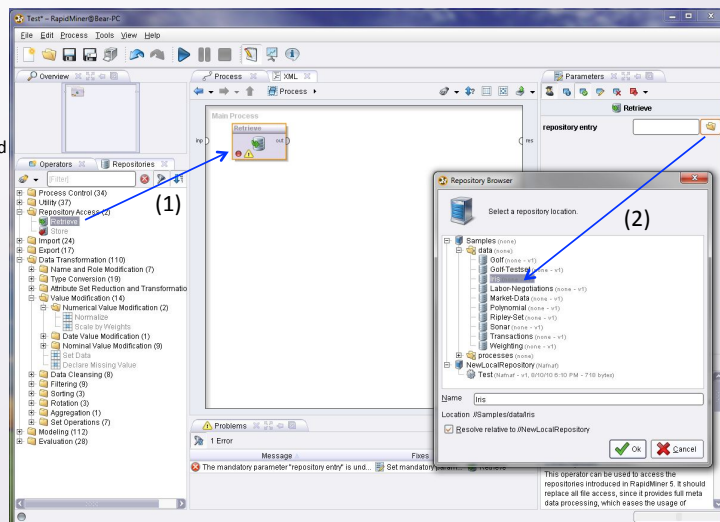
# Create a new Process

Select a location to store this new process



## Create a new Process (cont.)

1. Click and Drag the Retrieve Operator to the Graph Layout



2. Select dataset for this Retrieve Operator

## Create a new Process (cont.)

3. Click and Drag the Normalise Operator to the Graph Layout

4. Click and Drag to create links between Retrieve Operator and Normalise Operator as well as between Normalise Operator and res (result)

5. Click on Run icon to execute this process. We will obtain the same results as the previous example

