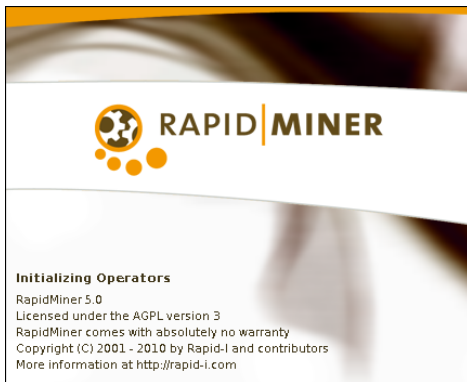


Rapid Miner

- RapidMiner is a tool for experimenting with machine learning and data mining algorithms
- An experiment is a set of operators that perform different tasks in the data
 - data input/output, data transformation, preprocessing, attribute selection, learning, evaluation, ...
- The experiments can be described visually as a process

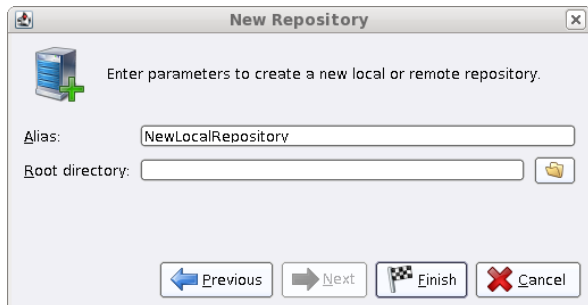
Rapid Miner - First contact



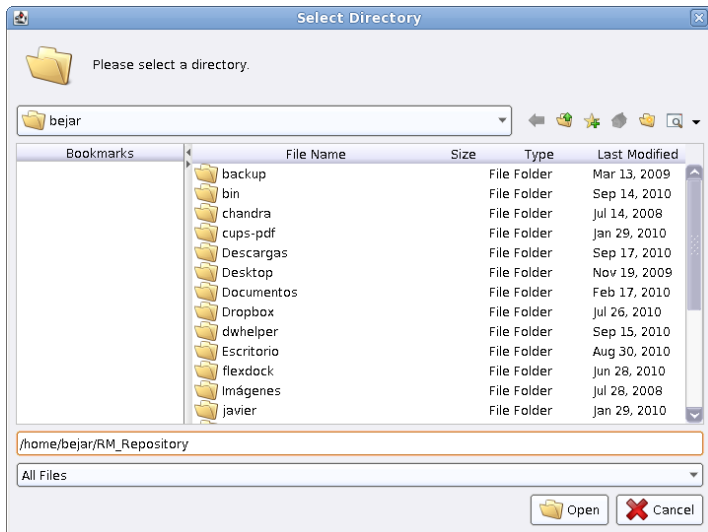
Rapid Miner - Setting a repository



Rapid Miner - Setting a repository



Rapid Miner - Setting a repository



Rapid Miner - Perspectives

The screenshot shows the RapidMiner application window with the title bar "RapidMiner@chandra.lsi.upc.edu". The menu bar includes "File", "Edit", "Process", "Tools", "View", and "Help". Below the menu is a toolbar with icons for file operations and help. The main area displays a "Welcome" message and several icons: "New", "Open Recent", "Open", "Open Template", and "Online Tutorial".

Sneak Peek Video of R-Extension

At the RapidMiner Community Meeting and Conference (RCMW 2010), the new RapidMiner extension for the integration of R will be presented for the first time

However, data analysts do not have to wait until the RapidMiner conference in order to get a first idea of what can be done with the new extension. The following video shows how easy R models and scripts can be integrated into the RapidMiner analysis processes.

"We make free software affordable."

According to the founder John Gilmore, this slogan was written on the support T-shirt of Cygnus, the first company offering professional services around open source products. Following this motto, the RapidMiner Enterprise Edition now allows you to deploy the most widely used open source data mining solution with confidence, security, and far lower total cost of ownership (TCO) than any other alternative.

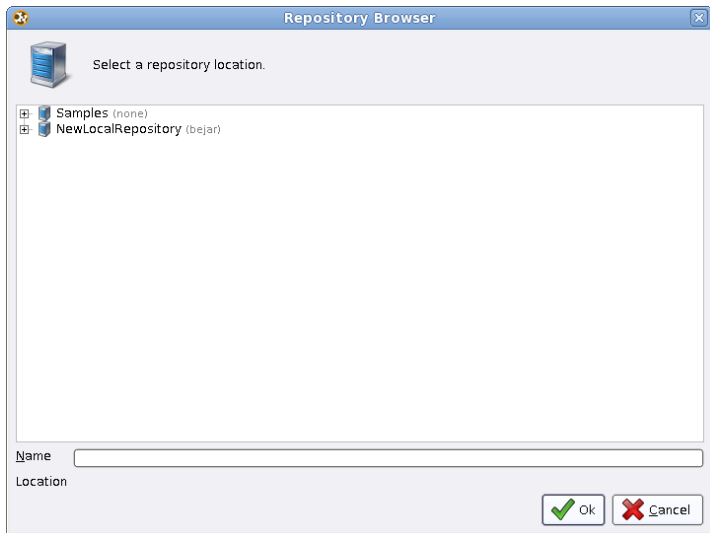
Using the free Community Edition of RapidMiner not only increases your risk of failure, it turns out to be more expensive. The RapidMiner Enterprise Edition delivers critical benefits like stabilized software, direct access to product expertise, and committed response times to help you save time and money. In the long short run, paying RapidMiner to maintain the software is cheaper than employing a specialist to do the work for

Promotion Offer!
RapidMiner Enterprise Edition Small, Standard, or Developer
Get Info Here!

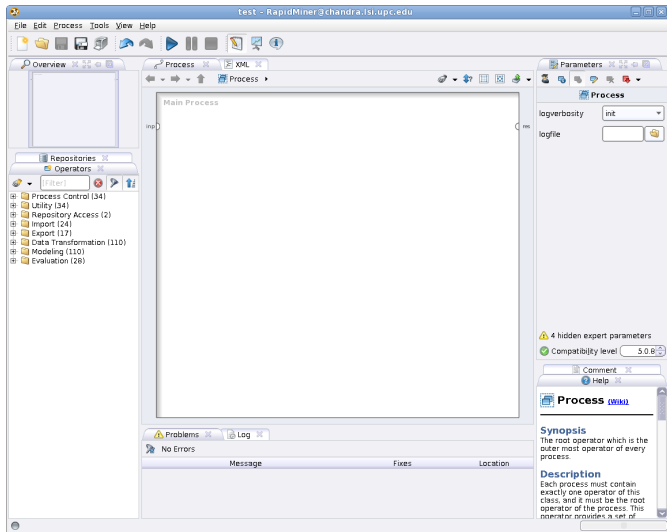
Courses & Webinars

- **Fast Track:** Introduction into Data Analysis with RapidMiner (Webinar)
- **Text Mining with RapidMiner**
- **Data Management & Manipulation:** Load and Transform your data (Webinar)
- **Sentiment Analysis, Opinion Mining, and Automatic Market Research**
- **Time Series Analysis with**

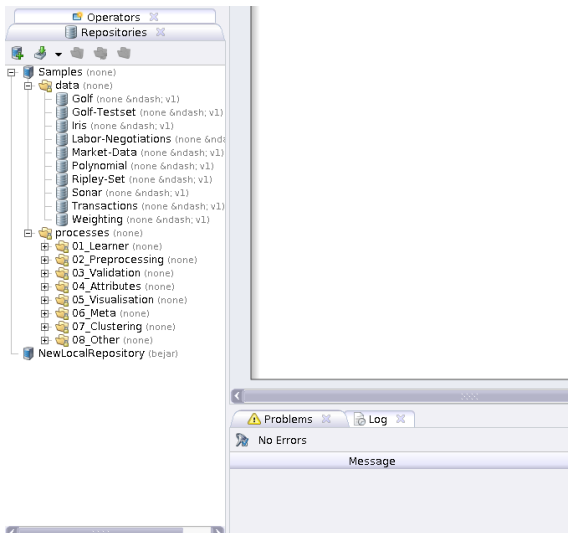
Rapid Miner - New project



Rapid Miner - Process perspective



Rapid Miner - Repository



Rapid Miner - Operators

The screenshot displays the RapidMiner application window titled "test - RapidMiner@chandra.lsi.upc.edu". The interface is divided into several panes:

- Overview:** A large empty central workspace.
- Process:** A central pane labeled "Main Process" with a large empty area for building a workflow.
- Parameters:** A pane on the right showing settings for the selected "Process" operator, including "logverbosity" set to "init" and an empty "logfile" field.
- Repositories:** A tree view on the left showing various operator categories:
 - Process Control (134)
 - Utility (14)
 - Repository Access (2)
 - Retrieve
 - Store
 - Import (25)
 - Data (18)
 - Read CSV
 - Read Excel
 - Read Access
 - Read AML
 - Read ARFF
 - Read ARFF
 - Read Database
 - Stream Database
 - Read SPSS
 - Read Stata
 - Read Sparse
 - Read DBase
 - Read C4.5
 - Read BibTeX
 - Read DasyLab
 - Read URL
 - Models (3)
 - Attributes (2)
 - Results (1)
 - Other (2)
 - Read
 - Export (17)
 - Data Transformation (110)
 - Modeling (243)
 - Evaluation (130)
 - Reporting (5)
- Problems & Log:** A pane at the bottom showing "No Errors" with columns for Message, Fixes, and Location.
- Help:** A pane on the right showing the "Process" operator's documentation, including a "Synopsis" and "Description".

The "Process" operator documentation includes:

- Synopsis:** The root operator which is the outer most operator of every process.
- Description:** Each process must contain exactly one operator of this class, and it must be the root operator of the process. This operator provides a set of

Rapid Miner - Adding operators

The screenshot displays the RapidMiner software interface. The main window is titled 'Process' and contains a 'Main Process' canvas. Inside the canvas, a 'Retrieve' operator is placed, connected to an 'inp' port on the left and an 'out' port on the right. The operator has a warning icon (yellow triangle with an exclamation mark) in its top-left corner.

On the right side, the 'Parameters' panel is open for the 'Retrieve' operator. It shows a 'repository entry' parameter with a text input field and a folder icon. Below this, the 'Compatibility level' is set to '5.0.8'. There are also 'Help' and 'Comment' tabs.

At the bottom of the interface, the 'Problems' panel shows an error message: '1 Error'. The message text is: 'The mandatory parameter "repository entry" is und... Set mandatory parame... Retrieve'. The error is located in the 'Retrieve' operator.

The 'Description' panel for the 'Retrieve' operator is also visible, providing a synopsis and description of the operator's function.

Synopsis
Reads an object from the data repository

Description
This operator can be used to access the repositories introduced in RapidMiner 5. It should replace all file access, since it provides full meta data processing, which eases the usage of RapidMiner a lot. In contrast

Rapid Miner - Connecting operators

The screenshot displays the RapidMiner software interface. At the top, there are tabs for 'Process' and 'XML'. Below the tabs is a toolbar with navigation and execution icons. The main workspace is titled 'Main Process' and contains a 'Retrieve' operator (represented by a database icon) with an 'out' port. A blue line connects the 'out' port of the 'Retrieve' operator to the 'res' port of a process flow on the right side of the canvas. The bottom of the interface features a 'Problems' and 'Log' section. The log shows the following messages:

```
11:53:11 CSDP 2010  
Sep 20, 2010 12:21:21 PM INFO: Decoupling process from location //NewLocalRepository/test.  
Process is now associated with file //NewLocalRepository/test.  
Sep 20, 2010 12:30:48 PM CONFIG: Saving property  
rapidminer.gui.resolve_relative_repository_locations=true
```

Rapid Miner - Metadata

The screenshot shows the RapidMiner interface with the 'Retrieve' operator selected. The main window displays the 'Retrieve.output (output)' metadata table, which provides details about the data source and its attributes.

Retrieve.output (output)
 Meta data: Data Table
 Number of examples = 150
 6 attributes:
 Generated by: [Retrieve output](#)

Role	Name	Type	Range	Missings	Comment
	a1	real	= 4.300 ...	= 0	
	a2	real	= 2 - 4 ...	= 0	
	a3	real	= 1 - 6 ...	= 0	
	a4	real	= 0.100 ...	= 0	
	id	id	nominal	= id_1..i...	= 0
	label	label	nominal	= Iris-se...	= 0

Press "F3" for focus.

The right-hand pane shows the 'Parameters' section for the 'Process' operator, including settings for log verbosity, logfile, resultfile, random seed (2001), send mail (never), and encoding (SYSTEM). Below this, there is a 'Compatibility level' set to 5.0.8 and a 'Process (wiki)' section with a 'Synopsis' and 'Description'.

Synopsis
 The root operator which is the outer most operator of every process.

Description
 Each process must contain exactly one operator of this class, and it must be the root operator of the process. This operator provides a set of parameters that are of global

Rapid Miner - Complex Process

The screenshot displays the RapidMiner interface with a workflow process titled "Main Process". The process consists of several interconnected operators:

- Read ARFF**: The starting point of the process.
- Set Role**: Configures the roles of the data attributes.
- Split Data**: Divides the data into training and testing sets.
- Decision Tr...**: A decision tree model operator.
- Apply Model**: Applies the trained model to the data.
- Performan...**: Evaluates the performance of the model.

The interface includes several panels:

- Overview**: A high-level view of the process.
- Process**: The main workspace for building the workflow.
- Parameters**: A panel for configuring the selected operator, currently showing "Apply Model" with options like "application param..." and "create view".
- Operators**: A library of available operators, including "Rule Induction (5)", "Neural Net Training (2)", "Function Fitting (7)", "Logistic Regression (2)", "Support Vector Modeling (6)", "Discriminant Analysis (3)", "Meta Modeling (13)", "Data Transformation (3)", "Weka (110)", "Attribute Weighting (33)", "Clustering and Segmentation (18)", "Association and Item Set Mining (10)", "Correlation and Dependency Computation", "Simulty Computation (4)", "Model Application (11)", "Thresholds (3)", "Confidences (2)", "Apply Model", "Update Model", "Group Models", "Ungroup Models", "Generate Top-k Predictions", "Create Formula", "Evaluation (31)", "Validation (7)", "Performance Measurement (19)", "Classification and Regression (6)", "Performance (Classification)", "Performance (Binominal Classification)", "Performance (Regression)", "Performance (Costs)", "Performance (Ranking)", "Performance (Support Vector Classification)", and "Attributes (3)".
- Problems**: A panel showing "No Errors".
- Compatibility level**: Set to "5.0.8".
- Help**: A panel for the "Apply Model" operator, including a "Synopsis" and "Description".

Rapid Miner - Results Perspective

The screenshot displays the Rapid Miner Results Perspective window. The main area shows two results:

- test (1 results, Process results)**: Completed: Oct 13, 2010 5:15:08 PM (execution time: 3 s)
- test (2 results, Process results)**: Completed: Oct 13, 2010 5:24:06 PM (execution time: 3 s)

Two panels are visible:

- Decision Tree (Decision Tree)**: Shows a complex decision tree structure.
- Performance Vector (Performance)**: Displays the following metrics:

PerformanceVector:	
accuracy:	66.13%
ConfusionMatrix:	
True:	Rock Mine
Rock:	18 9
Mine:	12 23

At the bottom, there is a **Log** window and a **System Monitor** window.

Log window content:

```

Oct 13, 2010 5:13:19 PM CONFIG: Loading perspectives.
Oct 13, 2010 5:13:19 PM WARNING: Plugin initializer com.rapidminer.PluginInitR.initGui of Plugin R Extension
causes an error: null
Oct 13, 2010 5:13:21 PM CONFIG: ignoring update check. Last update check: was on Wed Oct 13 16:19:52 CEST 2010
Oct 13, 2010 5:13:21 PM INFO: Connecting to: http://www.myexperiment.org/workflows.xml?num=100
Oct 13, 2010 5:13:40 PM INFO: Decoupling process from location //NewLocalRepository/test. Process is now associated with file
//NewLocalRepository/test.
Oct 13, 2010 5:15:05 PM INFO: Saved process definition at: //NewLocalRepository/test.
Oct 13, 2010 5:15:05 PM INFO: No filename given for result file, using stdout for logging results!
Oct 13, 2010 5:15:05 PM INFO: Loading initial data.
Oct 13, 2010 5:15:05 PM INFO: Process starts
Oct 13, 2010 5:15:05 PM WARNING: Using deprecated example set stream version 1
Oct 13, 2010 5:15:08 PM INFO: Saving results.
Oct 13, 2010 5:15:08 PM INFO: Process finished successfully after 3 s.
  
```

System Monitor window content:

Max:	253 MB
Total:	62 MB

Rapid Miner - Results Perspective

test - RapidMiner@chandra.lsi.upc.edu

File Edit Process Tools View Help

Result Overview PerformanceVector (Performance) Tree (Decision Tree)

Table / Plot View Text View Annotations

Criterion Selector

Multiclass Classification Performance Annotations

Accuracy

Table View Plot View

accuracy: 66.13%

	true Rock	true Mine	class precision
pred. Rock	18	9	66.67%
pred. Mine	12	23	65.71%
class recall	60.00%	71.88%	

Log

System Monitor

Oct 13, 2010 5:13:19 PM CONFIG: Loading perspectives.
Oct 13, 2010 5:13:19 PM WARNING: Plugin initializer com.rapidminer.PluginIntr.initGui of Plugin R Extension causes an error: null
 Oct 13, 2010 5:13:21 PM CONFIG: ignoring update check. Last update check: was on Wed Oct 13 16:19:52 CEST 2010
 Oct 13, 2010 5:13:21 PM INFO: Connecting to: http://www.myexperiment.org/workflows.xml?num=100
 Oct 13, 2010 5:13:40 PM INFO: Decoupling process from location //NewLocalRepository/test. Process is now associated with file //NewLocalRepository/test.
 Oct 13, 2010 5:15:05 PM INFO: Saved process definition at: //NewLocalRepository/test.
 Oct 13, 2010 5:15:05 PM INFO: No filename given for result file, using stdout for logging results!
 Oct 13, 2010 5:15:05 PM INFO: Loading initial data.
 Oct 13, 2010 5:15:05 PM INFO: Process starts
Oct 13, 2010 5:15:05 PM WARNING: Using deprecated example set stream version 1
 Oct 13, 2010 5:15:08 PM INFO: Saving results
 Oct 13, 2010 5:15:08 PM INFO: Process finished successfully after 3 s.

Max.: 253 MB
 Total: 62 MB

Rapid Miner - Results Perspective

The screenshot displays the Rapid Miner interface with a decision tree and a system monitor. The decision tree is a hierarchical structure of nodes and edges. The root node is 'attribute_45', which splits into two branches: > 0.386 and ≤ 0.386 . The > 0.386 branch leads to a leaf node labeled 'Mine'. The ≤ 0.386 branch leads to 'attribute_11', which splits into > 0.171 and ≤ 0.171 . The > 0.171 branch leads to 'attribute_40', which splits into > 0.477 (leaf 'Rock') and ≤ 0.477 (leaf 'Rock'). The ≤ 0.171 branch leads to 'attribute_23', which splits into $\geq 0.5\alpha$ (leaf 'Mine') and $< 0.5\alpha$ (leaf 'Rock'). The ≤ 0.477 branch of 'attribute_40' leads to 'attribute_39', which splits into > 0.085 (leaf 'Rock') and ≤ 0.085 (leaf 'Rock'). The ≤ 0.477 branch of 'attribute_39' leads to 'attribute_8', which splits into > 0.347 (leaf 'Rock') and ≤ 0.347 (leaf 'Rock'). The ≤ 0.347 branch of 'attribute_8' leads to 'attribute_39', which splits into > 0.044 (leaf 'Rock') and ≤ 0.044 (leaf 'Rock'). The ≤ 0.044 branch of 'attribute_39' leads to 'attribute_32', which splits into $> 0.6\alpha$ (leaf 'Mine') and $\leq 0.6\alpha$ (leaf 'Rock').

The System Monitor window shows a graph of memory usage over time. The graph has a grid background and a blue line representing memory usage. The y-axis is labeled 'Max: 253 MB' and 'Total: 62 MB'. The x-axis represents time.

The Log window shows the following messages:

```

Oct 13, 2010 5:13:19 PM CONFIG: Loading perspectives.
Oct 13, 2010 5:13:19 PM WARNING: Plugin initializer com.rapidminer.PluginIntr.initGui of Plugin R Extension
causes an error: null
Oct 13, 2010 5:13:21 PM CONFIG: ignoring update check. Last update check: was on Wed Oct 13 16:19:52 CEST 2010
Oct 13, 2010 5:13:21 PM INFO: Connecting to http://www.myexperiment.org/workflows.xml?num=100
Oct 13, 2010 5:13:40 PM INFO: Decoupling process from location //NewLocalRepository/test. Process is now associated with file
//NewLocalRepository/test.
Oct 13, 2010 5:15:05 PM INFO: Saved process definition at //NewLocalRepository/test.
Oct 13, 2010 5:15:05 PM INFO: No filename given for result file, using stdout for logging results!
Oct 13, 2010 5:15:05 PM INFO: Loading initial data.
Oct 13, 2010 5:15:05 PM INFO: Process starts
Oct 13, 2010 5:15:05 PM WARNING: Using deprecated example set stream version 1
Oct 13, 2010 5:15:06 PM INFO: Saving results.
Oct 13, 2010 5:15:06 PM INFO: Process finished successfully after 3 s.
  
```