

Lecture 1

Hello! Rapidminer

Kim, Yang Sok
Dept. of MIS, Keimyung University

- **What is Rapidminer?**
- **User Interface**
- **Miscellaneous**
 - Exercise1: Use Local Repository
 - Exercise 2: Use Operators
 - Exercise 3: Set Breakpoint
- **Conclusion**

What is Rapidminer?

What is Rapidminer?

Development History

- **RapidMiner, formerly known as YALE (Yet Another Learning Environment), was developed in 2001 by Ralf Klinkenberg, Ingo Mierswa, and Simon Fischer at the Technical University of Dortmund.**
- **Starting in 2006, its development was driven by Rapid-I, a company founded by Ingo Mierswa and Ralf Klinkenberg in the same year.**
- **In 2007, the name of the software was changed from YALE to RapidMiner.**
- **In 2013, the company rebranded from Rapid-I to RapidMiner.**

What is Rapidminer?

Best Tool for Data Science & Machine Learning

[WHY RAPIDMINER](#)[INDUSTRIES](#)[PRODUCTS](#)[LEARN](#)[RESOURCES](#)[PARTNERS](#)[COMPANY](#)[GET STARTED](#)

RapidMiner
Exceeds Your
Expectations

Gartner

RapidMiner a Leader in the 2019 Gartner Magic Quadrant for Data Science and Machine Learning Platforms for the sixth year in a row

[Read the Report](#)

FORRESTER®

RapidMiner a Leader in the 2018 Forrester Wave Multimodel Predictive Analytics & Machine Learning Solutions for the second year in a row

[Read the Report](#)

What is Rapidminer?

Best Tool for Data Science & Machine Learning

- Rapidminer named as a leader in the Gartner's 2019 Magic Quadrant for Data Science and Machine Learning Platforms for six consecutive years
- Gartner recognizes leaders as having the highest combined scores **for ability to execute and completeness of vision.**
- Rapidminer signed or expanded relationships with over 10% of the companies currently on the 2017 Fortune 100 Global List, and nearly 40% of Fortune 500 are active users of RapidMiner.

Figure 1. Magic Quadrant for Data Science and Machine Learning Platforms



What is Rapidminer?

Studio, Server and Radoop

Rapidminer Server



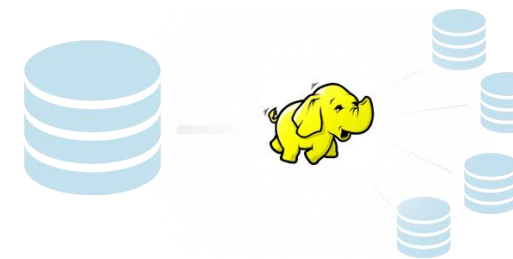
RapidMiner Server is an enterprise-grade environment for powerful predictive analytics, designed with central repository & management, dedicated computation power, and broad extensibility to operationalize analytic results. The three RapidMiner Server tiers encompass Collaboration, Computation and Deployment.

Rapidminer Studio



Forget sifting through code! RapidMiner is the most powerful, easy to use and intuitive graphical user interface for the design of analytic processes. Let the Wisdom of Crowds and recommendations from the RapidMiner community guide your way.

Rapidminer Radoop



RapidMiner Radoop, a core component of the RapidMiner Predictive Analytics Platform, extends predictive analytics to Hadoop and strongly supports Hadoop security implementations, all while delivering a seamlessly experience.

What is Rapidminer?

Characteristics

- **RapidMiner is very simple and easy to learn data mining.**
- **With RapidMiner, user can design an analytic process using operators and connections between operators. User can manage operators and connections through user friendly graphical user interface.**
- **While designing the analytics process, RapidMiner continually check whether the current analytics process is correct by using meta data transformation technology.**
- **Free for educational purpose, but need to pay for research.**

What is Rapidminer?

Download Rapidminer Studio

Go to <https://rapidminer.com/get-started/>



RapidMiner Studio: Start your free 30-day trial

Includes unlimited data rows, fastest performance, and premium features including Turbo Prep and Auto Model
After 30 days, you'll automatically revert to the free version of Rapidminer Studio

Submit work
email address

Enter your work email

DOWNLOAD

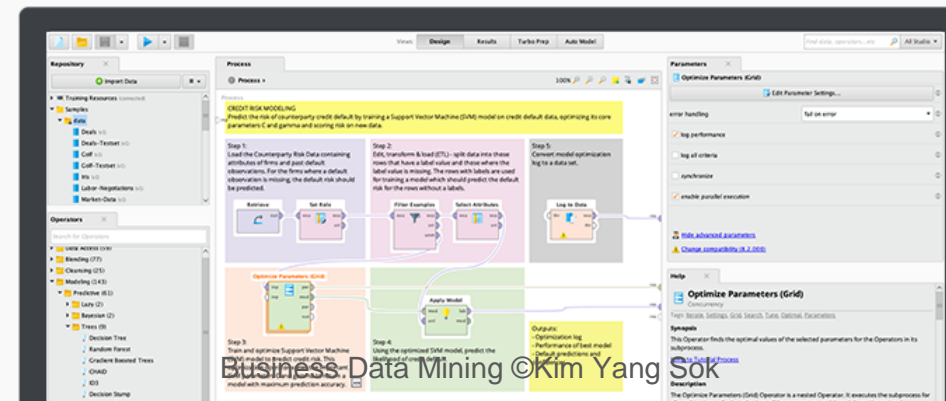
* Choose your role...

* Choose your indust...

Select the role that best describes you and the industry you're in. This will help us deliver more relevant resources.

[View our Privacy Policy](#) | [Need an Academic License?](#)

Click here to
download
Rapidminer



- To install Rapidminer Studio, refer to the installation site at Rapidminer website(<http://docs.rapidminer.com/studio/installation/>).
- RapidMiner Studio 9.4 will be used in this class.

User Interface

- **View is top level user interface and contains other user interface components**
- **Design and results view are basic views**
- **Change from one view to another view by using menu or icon in the menu bar.**
- **Each view consists of various panels.**
 - Design: Canvas and tools for building and managing processes.
 - Results: Visualization, in many varied formats, of design process results.

User Interface

Design View

The screenshot shows the RapidMiner Studio Educational 9.3.001 interface in Design View. The main workspace is empty, with a central area for designing a process. The interface includes several panels and a menu bar.

Menu Bar: File, Edit, Process, View, Connections, Settings, Extensions, Help.

Views: Design (selected), Results, Turbo Prep, Auto Model.

Repository Panel: Contains a tree view of data sources. The "Local Repository (admin)" is selected. It includes options like "Import Data", "Training Resources (connected)", "Samples", "Community Samples (connected)", "Keras Samples", "DB (Legacy)", and "Connections (admin)".

Operators Panel: Displays a list of operators. The "Read Excel" operator is highlighted. It shows a tree view of operators under "Data Access (4)", "Files (4)", and "Read (3)".

Parameters Panel: Shows the parameters for the selected operator. Parameters include "logverbosity" (set to "init"), "logfile", "resultfile", "random seed" (set to "2001"), and "send mail" (set to "never"). There are also links for "Hide advanced parameters" and "Change compatibility (9.3.001)".

Help Panel: Provides information about the selected operator, showing the "Process" and "Synopsis" sections.

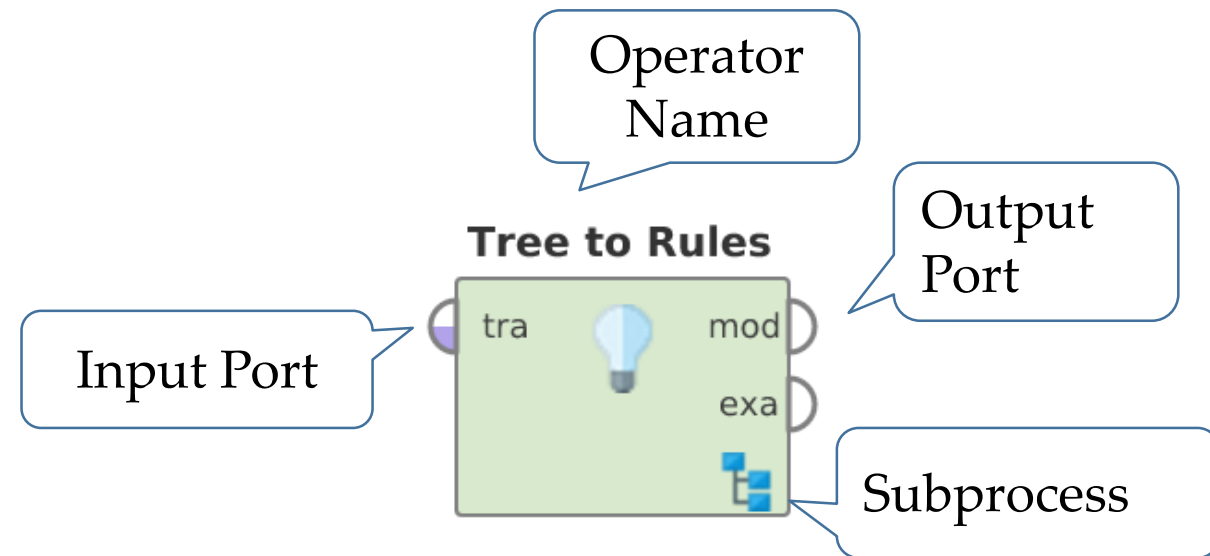
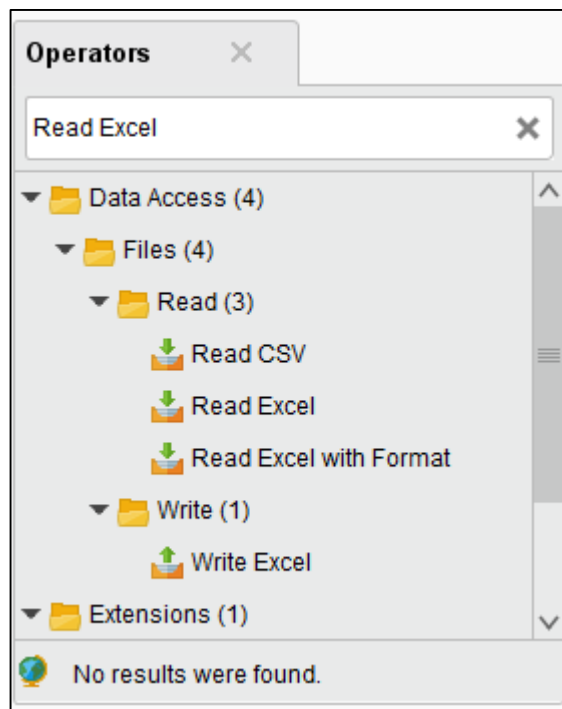
Callouts:

- Click "Design" button to activate Design View
- You can save data and process model "Local Repository" in Repository Panel
- You can find operators in Operators Panel for defining analysis process.
- You can design process in Process Panel
- You can define behaviors of an operator by setting parameters in Parameters Panel
- You can find information about an operator.

User Interface

Operator Panel

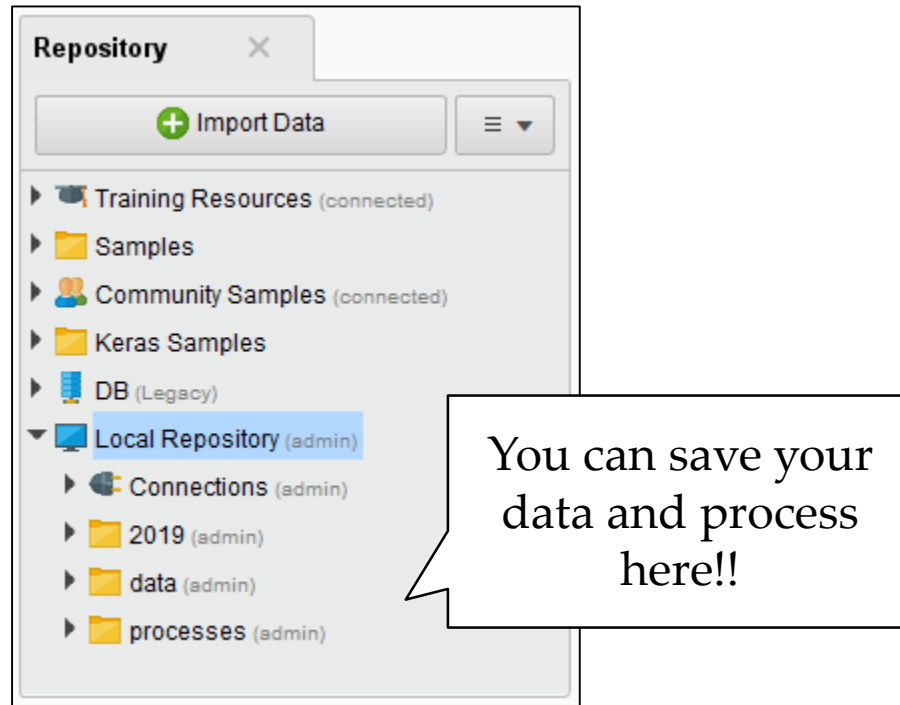
- Operator panel provides operators with in a tree, where related operators are grouped in one node, called package.
- User can find an operator by browsing the tree or by searching with keyword.



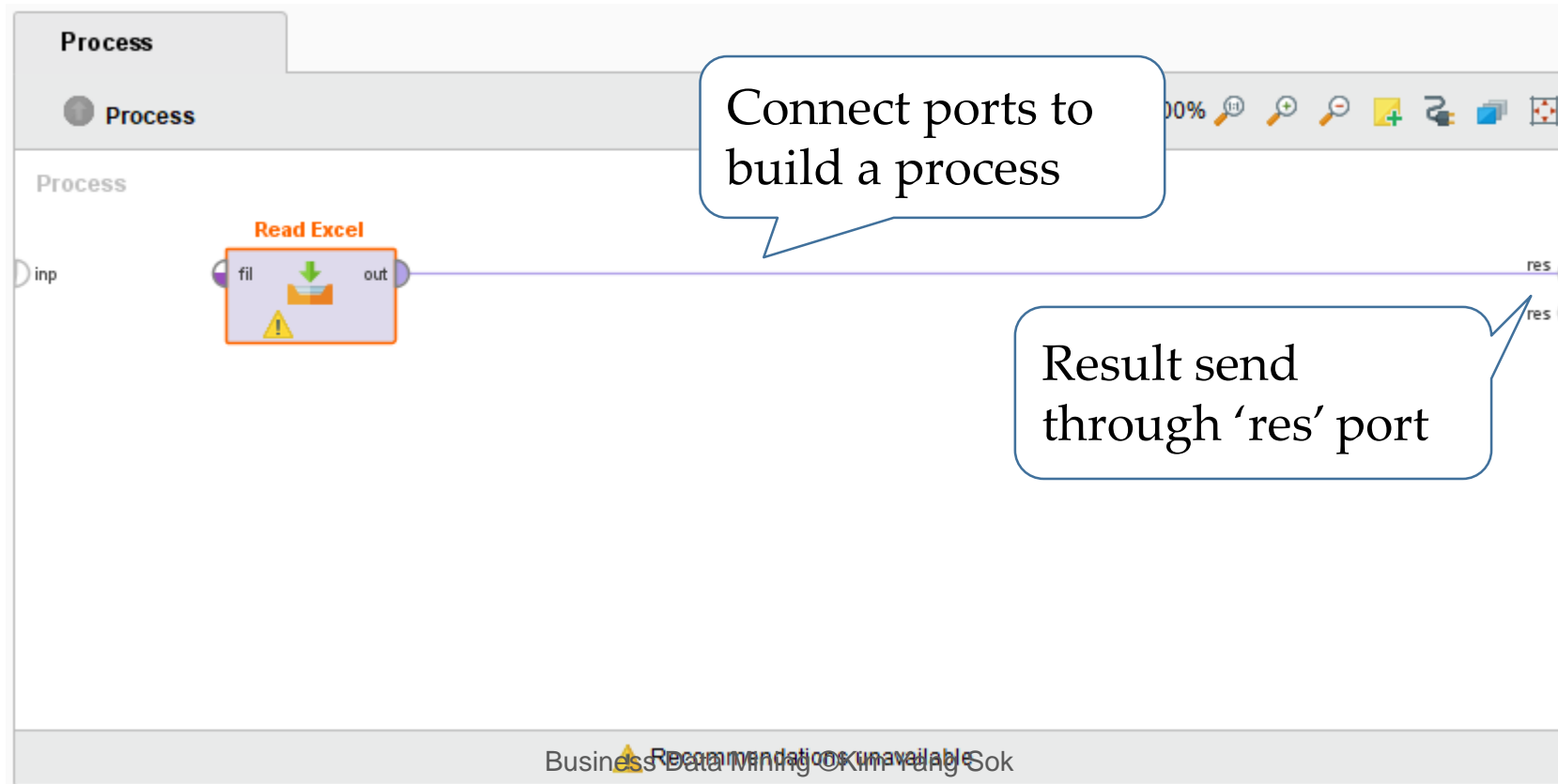
User Interface

Repository Panel

- Repository panel is used to save data and analysis process



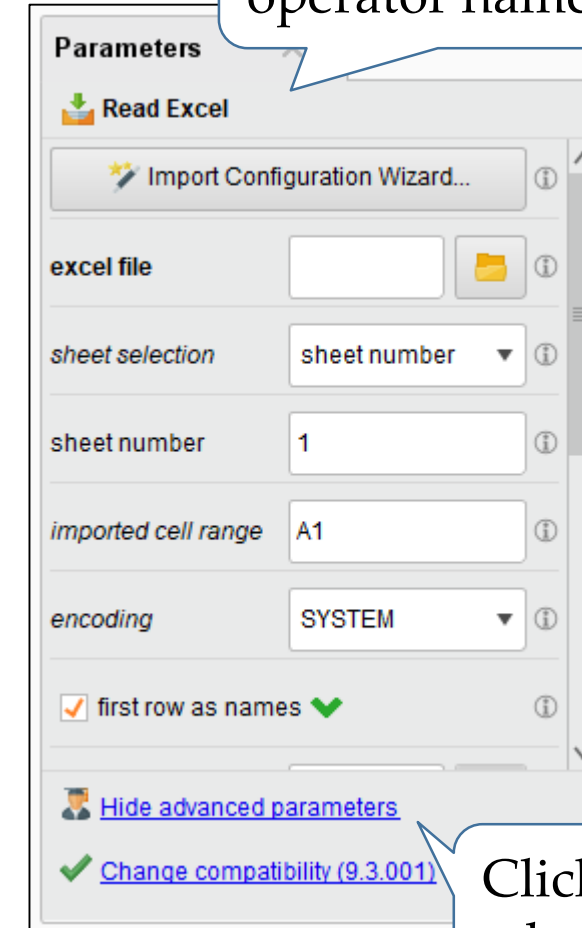
- **Process panel is a work space used to created analysis process.**
- **Analysis process design consists of two steps:**
 - Select and place operators from Operator Panel;
 - Connect two ports of operators and Process Panel (e.g., res port)



User Interface

Parameters Panel

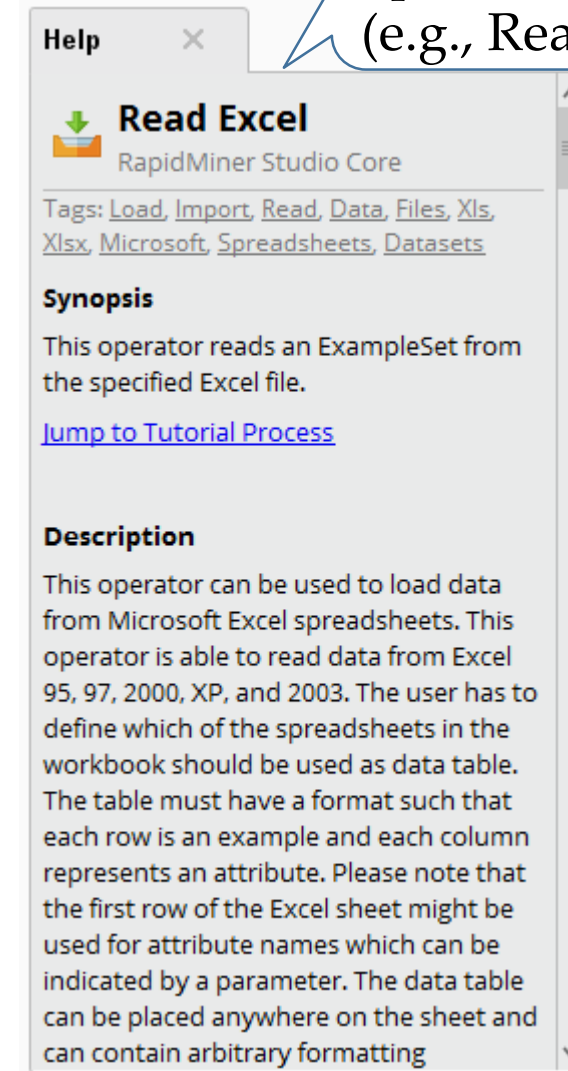
- If you choose an operator, it activates its Parameters Panel that shows parameters that should be set for its appropriate operation.
- There is general parameters and expert parameters.
 - While the former can be used by general users, the latter can be used by the expert who knows algorithms in depth
 - In order to activate the expert parameters, click “Show advanced parameters” link.



- **Parameter Types**

Types	Explanation	Example (Read Excel)
Check Box	User check an item	first row as name
Text Box	User type in value	sheet number
List	User select one item from the list.	date format
Button	User click this to open new dialog for further processing	Import Configuration Wizard ...

- Provide explanation on the selected operator.
- If you wish to see more details of the operator, please refer to operator manual (<http://docs.rapidminer.com/studio/operators/>)



Shows the selected operator name (e.g., Read Excel)

Structure of Help

Synopsys

Description

Differentiation

Input

Output

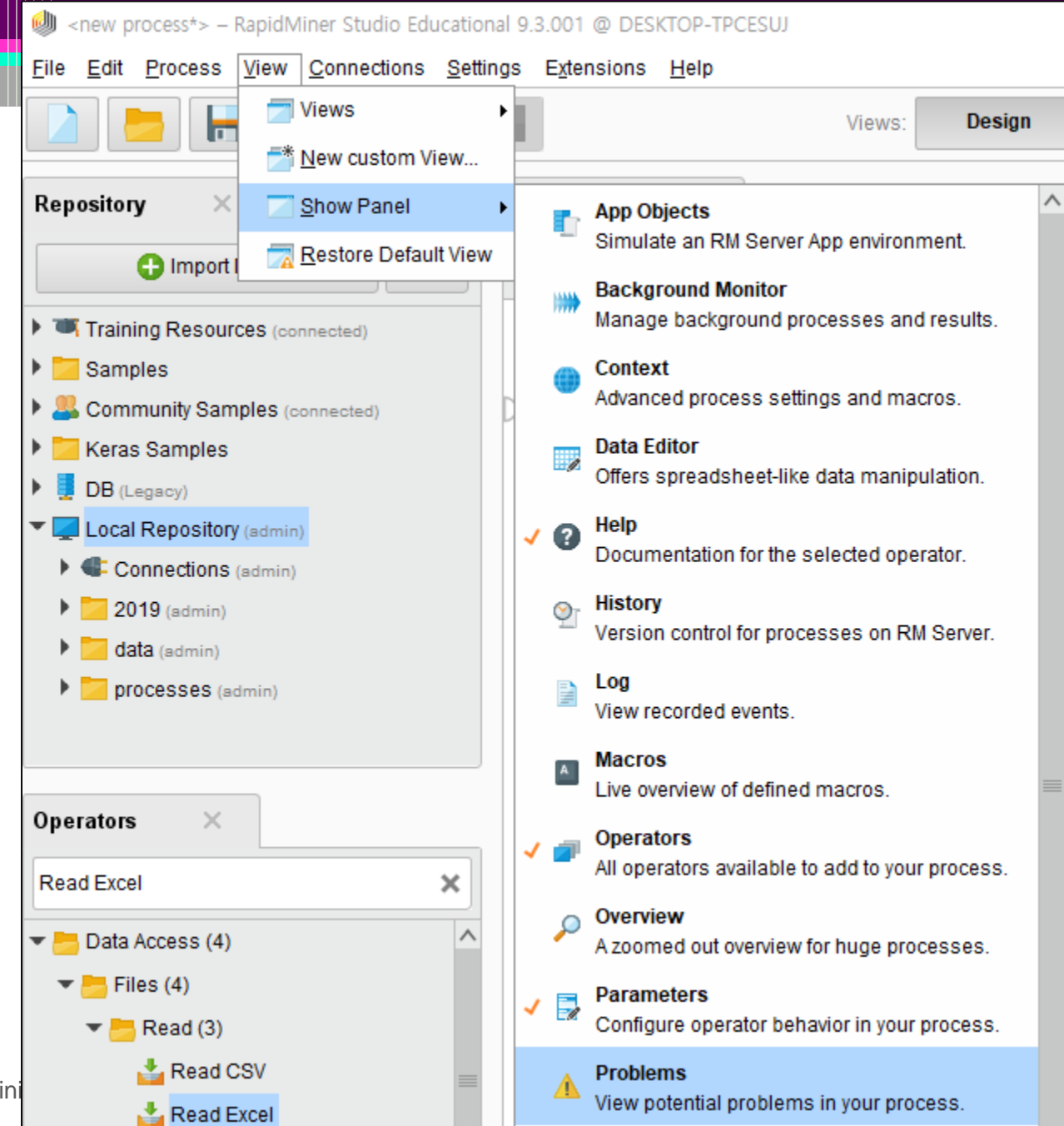
Parameters

Tutorial Process

User Interface

Add Other Panels

- When RapidMiner is installed, basic panels are activated.
- Add new panels (e.g., Problems) using “View>Show Panel” menu.
- You can change the place of each panels by using drag & drag a panel.



Miscellaneous

File Edit Process View Connections Settings Extensions Help

Views: Design Results

Repository

+ Import Data

Training Resources (connected)

Samples

Community Samples (connected)

Keras Samples

DB (Legacy)

Local Repository (connected)

Conn

2019

Busin

Process

Process

inp

Read CSV

fil out

You can execute various shortcut menu for operator

Operators

Read CSV

Data Access (1)

Files (1)

Read (1)

Read CSV

No results were found.

Recommended Operators

Process

Show Operator Info... F1

Enable Operator Ctrl+E

Rename operator F2

Replace Operator

Save as Building Block...

Cut Ctrl+X

Copy Ctrl+C

Paste Ctrl+V

Delete Delete

Add note

Breakpoint Before Shift+F7

Breakpoint After F7

Remove all Breakpoints

Show ExampleSet Result

Menu	Direct Key	Function
Show Operator Info...	F1	Show operator information.
Enable Operator	Ctrl+E	Make the operator enable or disable. This menu can be used for multiple operators.
Rename operator	F2	Change operator name.
Replace operator		Replace the selected with the other operator
Save as Building Block		Save selected operators (analytic process) as one unit.
Cut	Ctrl+X	Cut the selected operator/s.
Copy	Ctrl+C	Copy the selected operator/s.
Paste	Ctrl+P	Paste the copied operator/s.
Delete	Ctrl+D	Delete the selected operator.
Add note		Add a note.
Breakpoint Before	Shift+F7	Show project processing result before executing this operator.
Breakpoint After	F7	Show project processing result after executing this operator.

Local Repository/Business Data Mining/Wine Data Loading* - RapidMiner Studio Educational 9.3.001 @ DESKTOP-TPCESUJ

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

Repository

- Import Data
- Training Resources (connected)
- Samples
- Community Samples (connected)
- Keras Samples
- DB (Legacy)
- Local Repository (admin)
 - Connections (admin)
 - 2019 (admin)
 - Business Data Mining (admin)

Process

Process

inp

Read CSV

fil out

res

res

Parameters

Read CSV

Import Configuration Wizard...

csv file: alality-red.csv

column separators: ;

☐ trim lines

☒ use quotes

quotes character: "

escape character: \

[Hide advanced parameters](#)

[Change compatibility \(9.3.001\)](#)

Operators

Read CSV

- Data Access (1)
 - Files (1)
 - Read (1)
 - Read CSV

Show Operator Info... F1

Insert Operator

Adds this

No results were found.

Help

Read CSV

RapidMiner Studio Core

Tags: Load, Import, Read, Data, Files, Text, Commas, Spreadsheet, Excel, Datasets, Tsv

Synopsis

This Operator reads an ExampleSet from the specified CSV file.

23

1 Select an operator and place it on the process panel using drag & drop function.

2 Select an operator and double click it, and then it is placed in the process panel.

3 Click an operator and click the right button of your mouse and click "Insert Operator" menu

Exercise1: Use Local Repository

Exercise1: Use Local Repository

Task & Steps

- **Task**
 - Create a custom repository

- **Steps**
 1. Select “Local Repository” in the Repository Panel
 2. Execute “Create subfolder ” menu.
 3. Type “Business Data Mining” into Name text field of “New Folder” dialog



Repository

+ Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- Keras Samples
- DB (Legacy)
- Local Repository

Operators

Read Excel

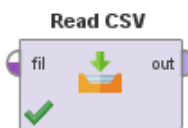
- Data Access (4)
- Files (4)
 - Read (3)
 - Read CSV
 - Read Excel
 - Read Excel with Format
 - Write (1)
 - Write Excel
- Extensions (1)

- Configure Repository
- Store Process Here
- Create Connection
- Create subfolder
- Paste
- Copy Location to Clipboard
- Delete
- Refresh folder
- Open in file browser

Process

Process

Process



Select Local Repository and activate shortcut menu. Click "Create subfolder" menu.

1

Parameters

Process

logverbosity	init
logfile	
resultfile	
random seed	2001
send mail	never
encoding	SYSTEM

[Hide advanced parameters](#)

[Change compatibility \(9.3.001\)](#)

Help

Process

RapidMiner Studio Core

Synopsis

The root operator which is the outermost operator of every process.



Views:

Design

Results

Turbo Prep

Auto Model

Find data, operators...etc

All Studio

Repository

+ Import Data

- Training Resources (connected)
- Samples
- Community Samples (connected)
- Keras Samples
- DB (Legacy)
- Local Repository (admin)
 - Connections (admin)
 - 2019 (admin)
 - data (admin)
 - processes (admin)

Operators

Read Excel

- Data Access (4)
 - Files (4)
 - Read (3)
 - Read CSV
 - Read Excel
 - Read Excel with Format
 - Write (1)
 - Write Excel
 - Extensions (1)

Process

Process

100%

Read CSV

inp

fil

out

New Folder

Please insert a name for the new folder.

Name Business Data Mining

OK Cancel

Recommended Operators

Business Data Mining ©Kim Yang Sok

Type in "Business Data Mining" in to Name text field.

Click "OK" button

Parameters

Process

logverbosity init

logfile

resultfile

random seed 2001

send mail never

encoding SYSTEM

[Hide advanced parameters](#)

[Change compatibility \(9.3.001\)](#)

Synopsis

The root operator which is the outermost operator of every process

27

Exercise 2:

Use Operators

Exercise 2: Use Operators

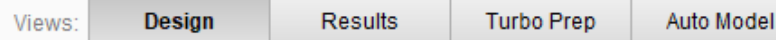
Task & Steps

- **Task**

- Find operators and add them to Process Panel

- **Steps**

1. Find <<Generate Direct Mailing Data>> by 'Search' function
2. Add <<Generate Direct Mailing Data>> into Process Panel
3. Link [output] of <<Generate Direct Mailing Data>> to [res] of Process Panel
4. Execute analysis process
5. Find <<Split Data>> by 'Browse' function
6. Split data into subsets by using <<Split Data>>
7. Link [par]s of <<Split Data>> to [res] of Process Panel
8. Execute analysis process



All Studio ▼

 Import Data

1

Process

Generate Direct Mailing Data

1

2

3

number examples	100
-----------------	-----

100

☐ use local random seed ✓ [Hide advanced parameters](#)

Generate Direct Mailing Data

RapidMiner Studio Core

Tags: [Random Data Generation](#)

Synopsis

This operator generates an `ExampleSet` that represents direct m

30

30



Click to execute the analysis process

1

Show examples generated

2

Result History ExampleSet (Generate Direct Mailing Data) x

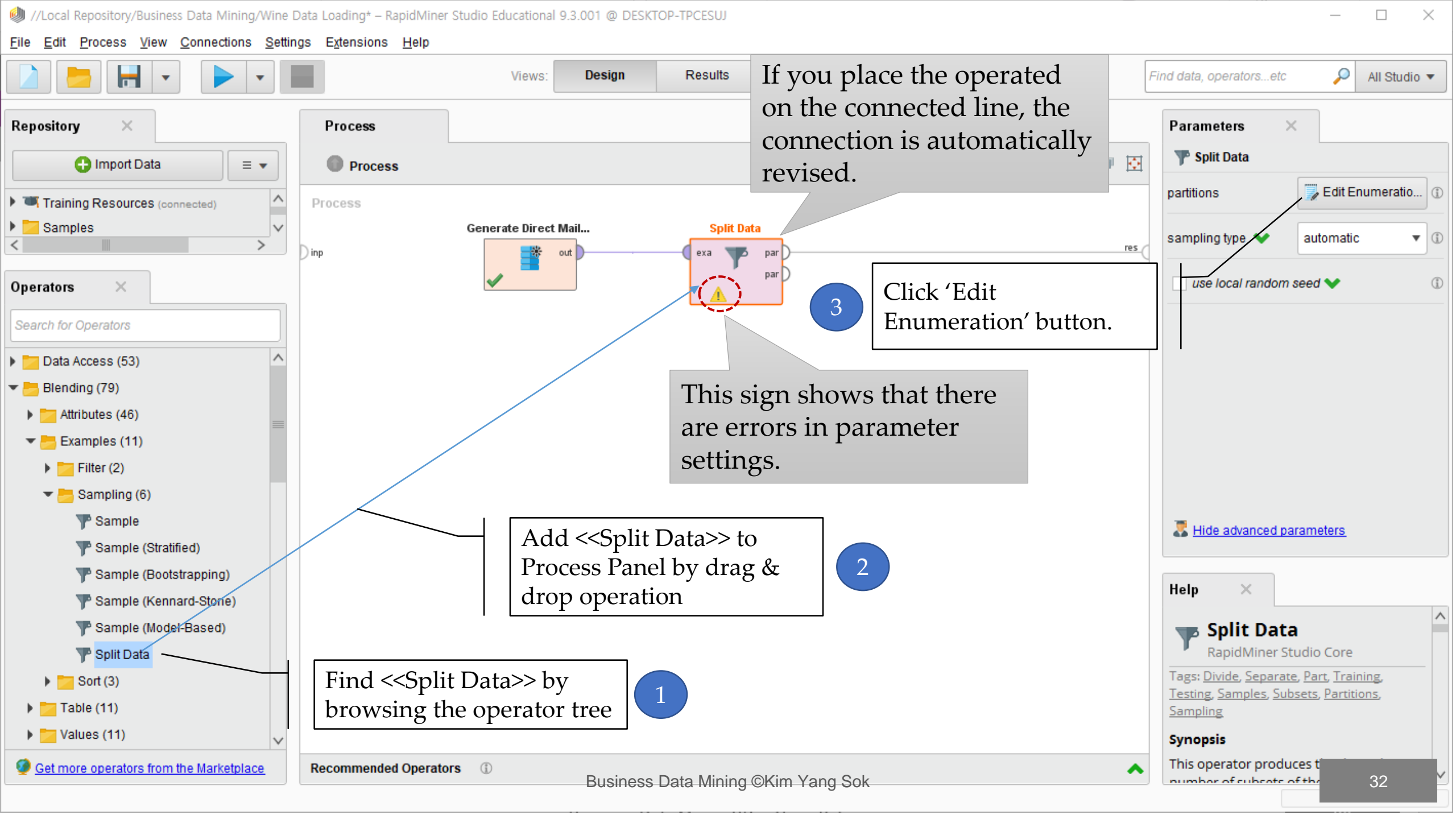
Open in Turbo Prep Auto Model

Filter (100 / 100 examples): all

Data
Statistics
Visualizations
Annotations

Row No.	label	name	age	lifestyle	zip code	family status	car	sports	earnings
1	no response	VnSEFOuL	62	cozily	50168	married	practical	athletics	102526
2	no response	8Tv0hcce	34	active	66479	married	expensive	soccer	33006
3	response	Zny9ysbk	69	healthy	16592	single	expensive	badminton	118760
4	response	HV3xCamM	57	cozily	50068	married	practical	soccer	131429
5	response	sflRsQ6v	66	cozily	35988	single	practical	badminton	96003
6	no response	w9voDHj0	21	healthy	60039	single	expensive	badminton	83376
7	no response	VUKT58V5	22	healthy	69662	married	practical	badminton	49826
8	response	HzMvC0c1	59	cozily	40432	single	practical	soccer	41166
9	response	B9MsYPOR	63	active	55994	single	practical	badminton	134340
10	no response	mWUY9Cjs	31	active	37460	married	practical	athletics	108695
11	response	Lavs4Ms	60	cozily	36782	single	practical	badminton	49723
12	response	kNAihUVY	38	healthy	12447	single	expensive	soccer	105801
13	no response	FHL6d3WH	36	active	56504	married	expensive	athletics	79703
14	no response	taO86r6P	30	cozily	84343	single	expensive	soccer	68693
15	no response	qQSz8WWj	64	healthy	66428	single	expensive	badminton	111755
16	no response	waFEJp2V	37	cozily	18272	married	practical	badminton	36430
17	response	CGFIIPRt	61	cozily	24254	married	practical	soccer	129420

ExampleSet (100 examples, 1 special attribute, 8 regular attributes)



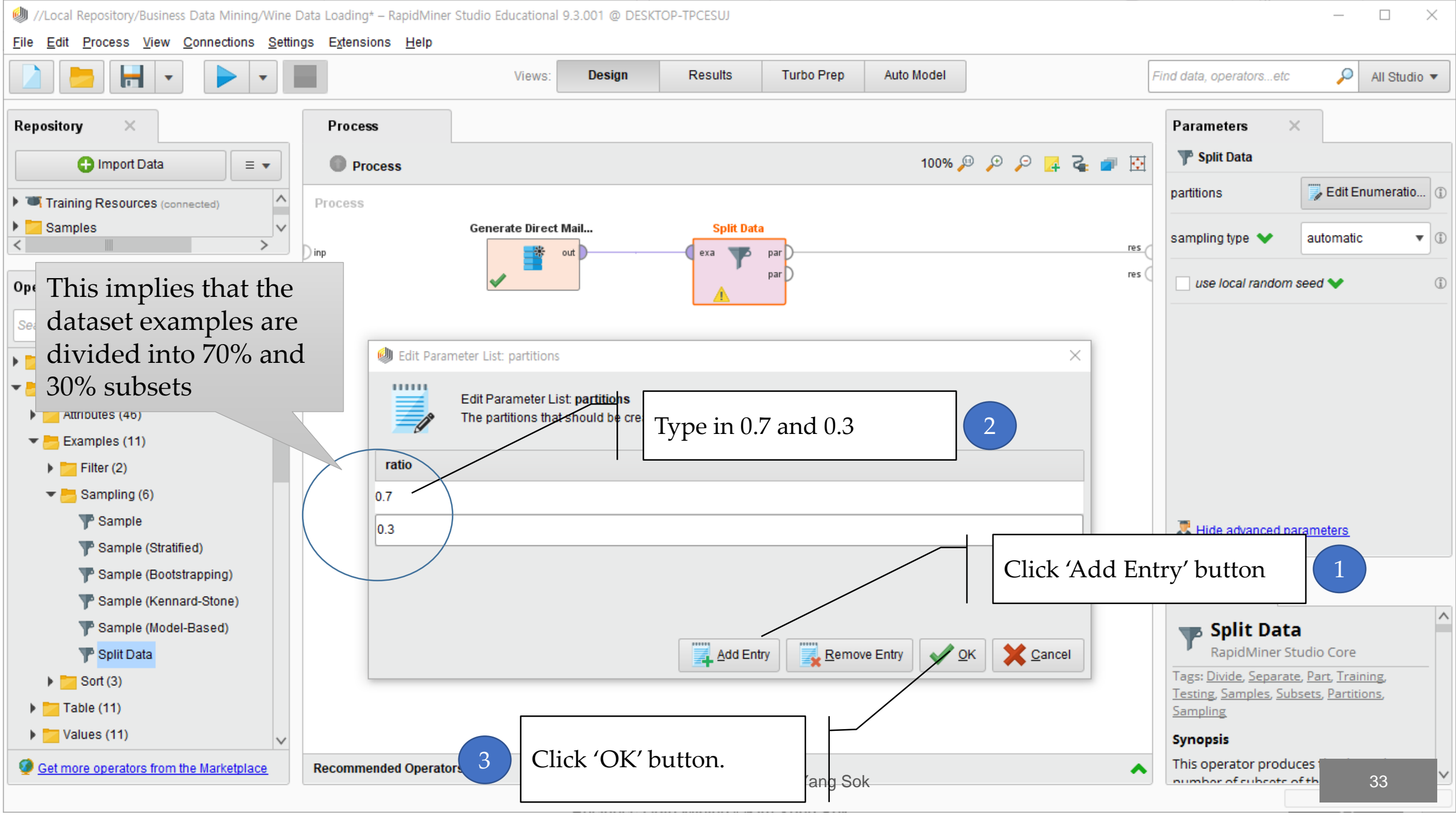
If you place the operated on the connected line, the connection is automatically revised.

Click 'Edit Enumeration' button.

This sign shows that there are errors in parameter settings.

Add <<Split Data>> to Process Panel by drag & drop operation

Find <<Split Data>> by browsing the operator tree



File

Edit

Process

View

Connections

Settings

Extensions

Help

Repository

Import Data

Training Resources (connected)

Samples

Operators

Search for Operators

Data Access (53)

Blending (79)

Attributes (46)

Examples (11)

Filter (2)

Sampling (6)

Sample

Sample (Stratified)

Sample (Bootstrapping)

Sample (Kennard-Stone)

Sample (Model-Based)

Split Data

Sort (3)

Table (11)

Values (11)

Get more operators from the Marketplace

Views: DesignResultsTurbo PrepAuto Model

1

2

Process

Generate Direct Mail...

Split Data

res

res

res

Execute the analysis process

Connect [par] and [res] for each [par] output.

Each [par] output 70% and 30% examples

Parameters

Split Data

partitions

sampling type automatic

use local random seed

Hide advanced parameters

Help

Split Data

RapidMiner Studio Core

Tags: Divide, Separate, Part, Training, Testing, Samples, Subsets, Partitions, Sampling

Synopsis

This operator produces t

number of subsets of the

Recommended Operators

Business Data Mining ©Kim Yang Sok

ALT + click to disconnect.

34

Open in Turbo Prep Auto Model

Filter (30 / 30 examples): all

Row No.	label	name	age	family status	car	sports	earnings
1	no response	8TV		married	expensive	soccer	
2	response	Zny		single	expensive	badmi	
3	response	Lavst4MS	60	single	practical	badmi	
4	response	kNAihUVY	38	single	expensive	soccer	
5	no response	waFEJp2V	37	married	practical	badminton	36430
6	no response	39OUyKYK	41	married	practical	athletics	137658
7	no response	BQg9yrw	24	married	practical	soccer	67545
8	response	3yYxpFte	27	married	expensive	soccer	149929
9	no response	FnmvD5Rp	27	married	expensive	badminton	77466
10	no response	66MA2Tqh	34	married	expensive	athletics	70684
11	no response	afMFwPFx	26	married	practical	soccer	104627
12	response	cAmqKr3E	58	single	practical	badminton	125512
13	no response	13a8vQ3d	28	single	practical	soccer	87285
14	response	CATLXO8O	56	single	practical	badminton	63512
15	response	zamLhLJM	57	single	expensive	soccer	79263
16	no response	KiSILJtj	36	married	practical	athletics	97491
17	response	rhkSk76X	59	single	practical	badminton	141220

ExampleSet (30 examples, 1 special attribute, 8 regular attributes)

If you click this tab, you can see 70% subset

This shows the number of 30% subset

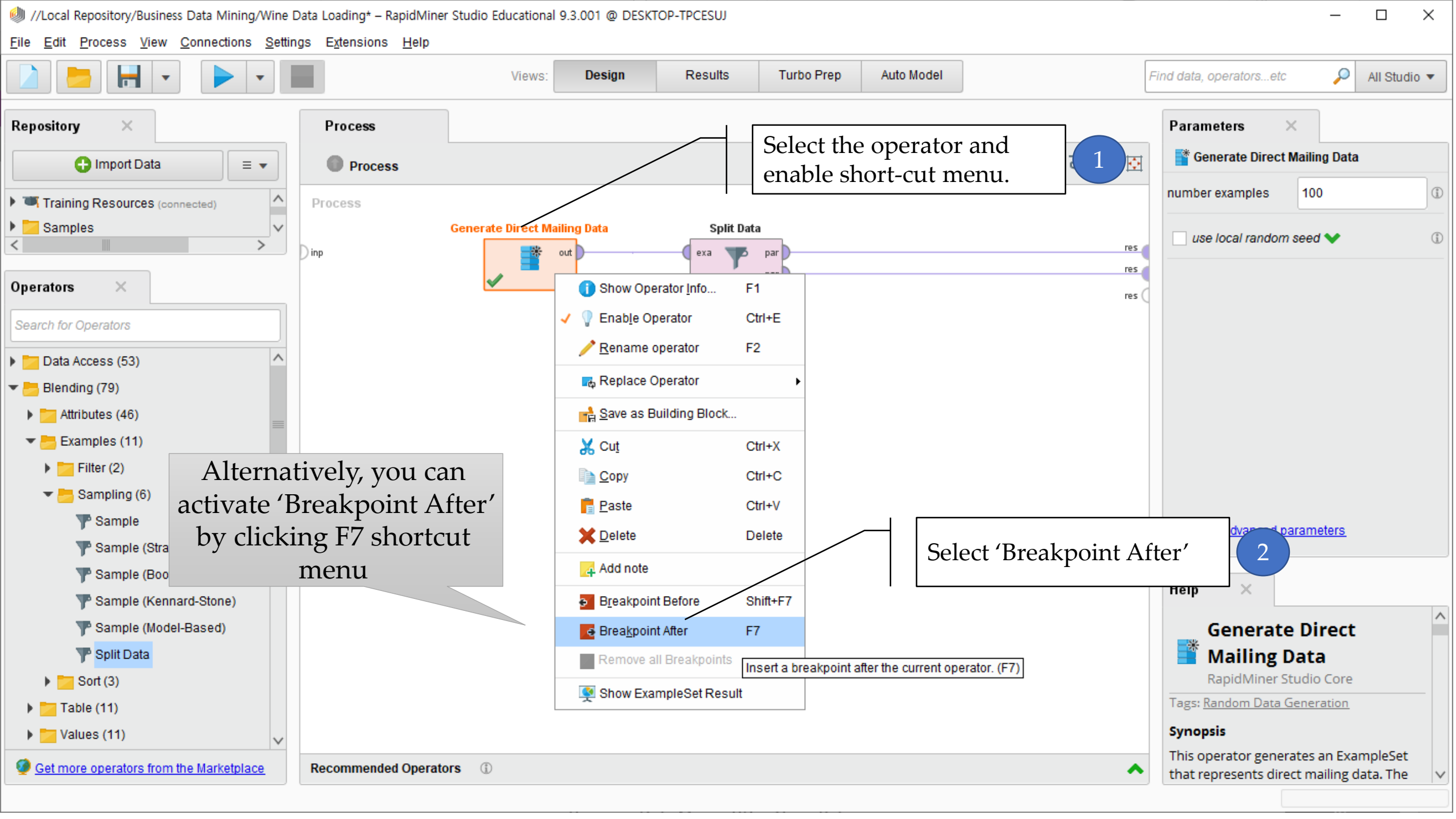
Exercise 3: Set Breakpoint

Exercise 3: Set Breakpoint

Task & Steps

- **Task**
 - Set breakpoint on the specific operator

- **Steps**
 1. Select <<Generate Direct Mailing Data>>
 2. Click right mouse button to display short-cut menu
 3. Select 'Breakpoint After'
 4. Execute analysis process



Select the operator and enable short-cut menu.

Alternatively, you can activate 'Breakpoint After' by clicking F7 shortcut menu

Select 'Breakpoint After'

Insert a breakpoint after the current operator. (F7)

Conclusion

- This lecture covers basics of Rapidminer
- You will use this software during the class, so it is necessary to install this software and you should be familiar with the user interface of this software.
- In the next lecture, you will learn how to perform data mining. In particular, we will focus on CRISP-DM.



QUESTIONS?