

## Stationery case study

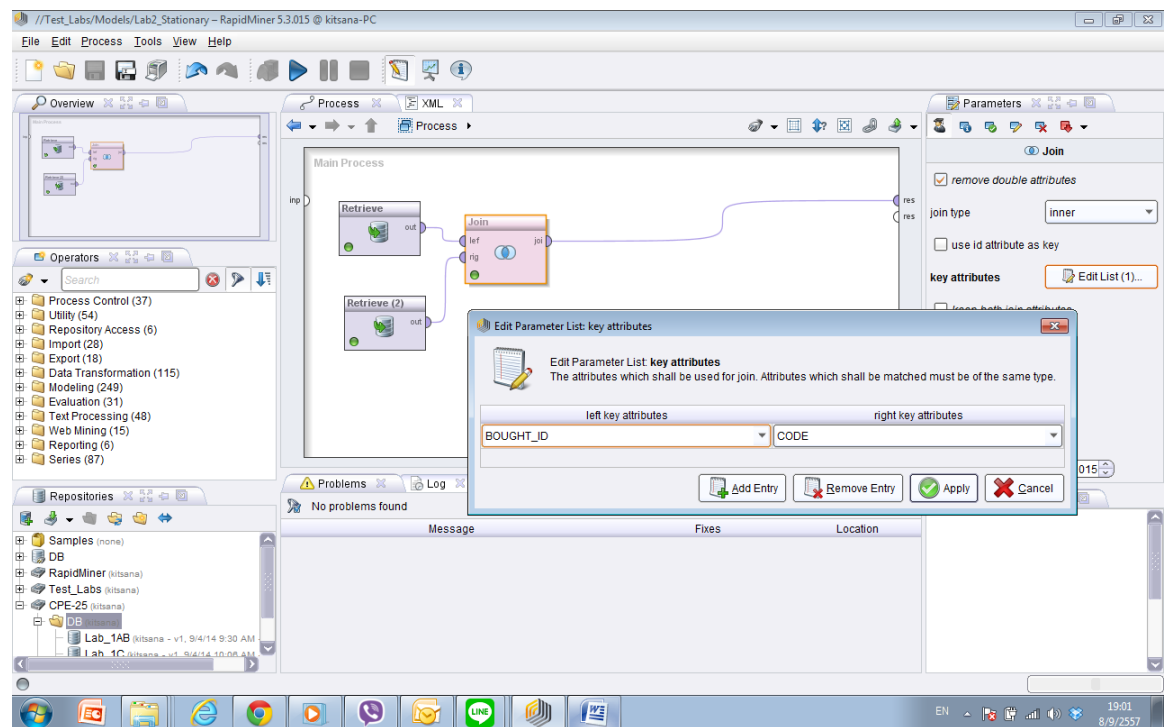
The stationery database consists of 2 tables:

- “PRODUCT\_SALE” table contains information about items bought by each customer. Each bought item is stored using its code in the column “BOUGHT\_ID”.
- “STOCK\_NAME” table contains description of each item.

You have to:

- Retrieve the input data
- Pre-process data using the following operators:

- **Join**



## - Select attributes

The screenshot displays the RapidMiner 5.3.015 interface. The main process canvas shows a workflow: 'Retrieve' (input) -> 'Join' -> 'Select Attributes' (output). The 'Select Attributes' process is highlighted, and its configuration panel is open. The 'attribute filter type' is set to 'subset'. The 'attributes' list is empty, and the 'Selected Attributes' list contains 'CUSTOMER\_ID', 'ITEMNAME', and 'Quantity'. The 'Apply' button is visible.

Process Overview:

- Retrieve (input)
- Join
- Select Attributes (output)

Select Attributes: attributes

Select Attributes: attributes  
The attribute which should be chosen.

Attributes: BOUGHT\_ID

Selected Attributes: CUSTOMER\_ID, ITEMNAME, Quantity

Apply

## - Pivot

The screenshot displays the RapidMiner 5.3.015 interface. The main process canvas shows a workflow: 'Retrieve' (input) -> 'Join' -> 'Select Attributes' -> 'Pivot' (output). The 'Pivot' process is highlighted, and its configuration panel is open. The 'group attribute' is set to 'CUSTOMER\_ID', the 'index attribute' is set to 'ITEMNAME', and the 'weight aggregation' is set to 'sum'. The 'Apply' button is visible.

Process Overview:

- Retrieve (input)
- Join
- Select Attributes
- Pivot (output)

Pivot

group attribute: CUSTOMER\_ID

index attribute: ITEMNAME

consider weights: ☒

weight aggregation: sum

skip constant attributes: ☐

datamanagement: double\_array

Apply

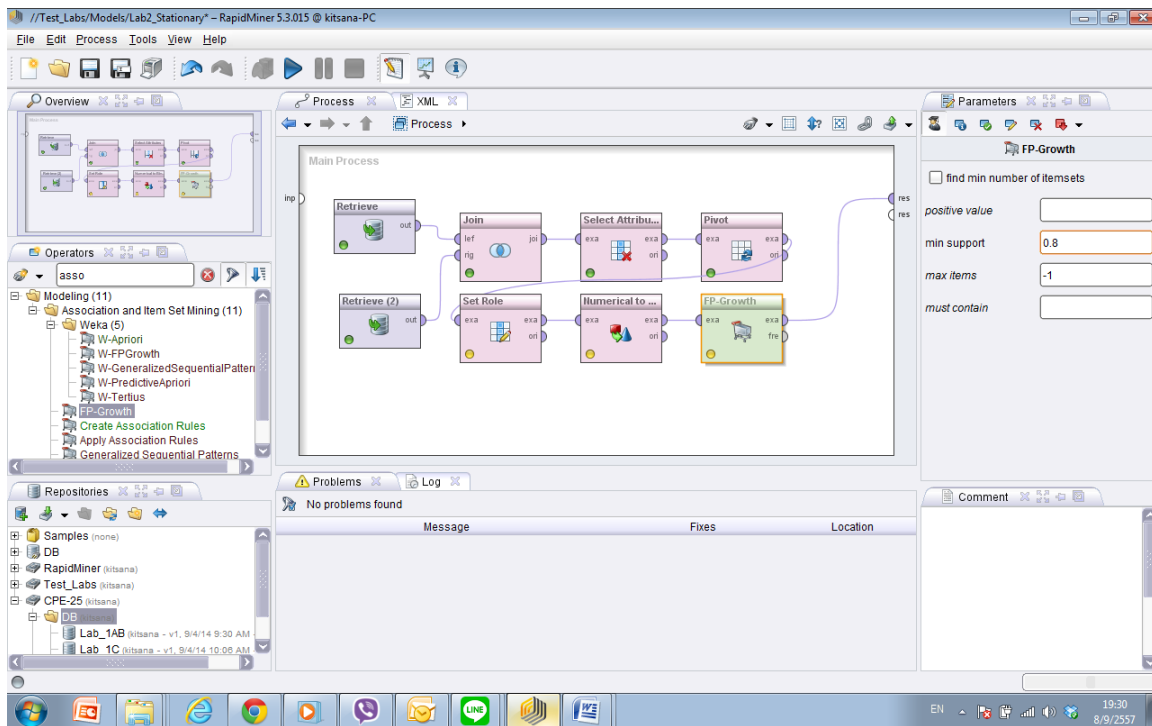
## – Set role

The screenshot shows the RapidMiner 5.3.015 interface. The main process canvas displays a workflow: 'Retrieve' → 'Join' → 'Select Attribute...' → 'Pivot'. A second 'Retrieve (2)' operator is connected to the 'Set Role' operator, which is placed after the 'Join' operator. The 'Set Role' operator is configured with 'attribute name' set to 'CUSTOMER\_ID' and 'target role' set to 'id'. The left sidebar shows the 'Operators' panel with 'Data Transformation (7)' expanded, and 'Set Role' is selected under 'Name and Role Modification (7)'. The bottom status bar shows the time as 19:14 on 8/9/2557.

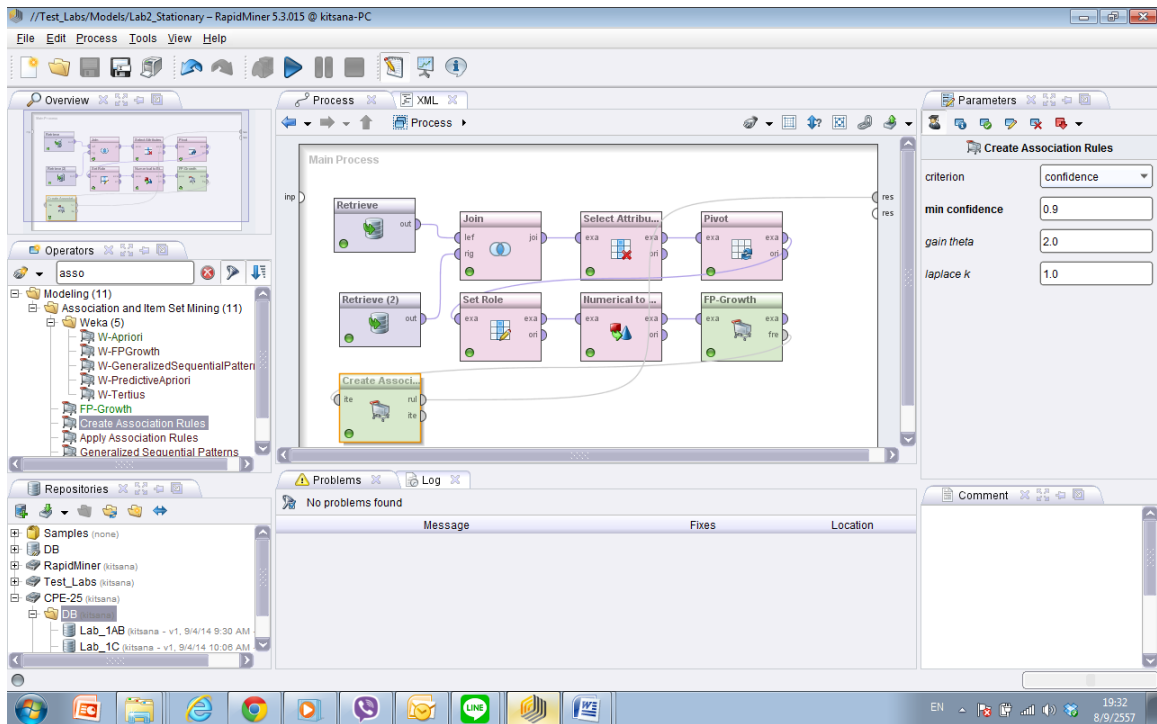
## – Numerical to Binomial

The screenshot shows the RapidMiner 5.3.015 interface with the 'Numerical to Binomial' operator selected in the 'Operators' panel. The main process canvas shows the workflow: 'Retrieve' → 'Join' → 'Select Attribute...' → 'Pivot'. A 'Set Role' operator is connected to the 'Join' operator, and a 'Numerical to Binomial' operator is connected to the 'Pivot' operator. The 'Numerical to Binomial' operator is configured with 'attribute filter type' set to 'subset' and 'attributes' set to 'CUSTOMER\_ID'. A warning message is displayed in the bottom status bar: 'One potential problem: The mandatory parameter "attributes" is undefined.' The right sidebar shows the 'Parameters' panel for the 'Numerical to Binomial' operator. The bottom status bar shows the time as 19:17 on 8/9/2557.

c) Apply FP-growth to determine all the frequent itemsets. Set minimum support value to 80%



d) Apply "Create Association Rules" operators. Set confidence value to 90%



e) Answer the following questions

- List all the frequent itemsets that have been found (with their support)?
- What is the maximal frequent itemset found? Give its support? List all the strong association rules that can be generated from it.
- According to the rule {liquidpaper} => {book,paper}, explain its meaning
- What items are expected to be bought by customers who buy liquidpaper?

No.	Premises	Conclusion	Support	Confid...	LaPl...	Gain	p-s	Lift
1	Quantity_liquidpaper	Quantity_book, Quantity_paper	0.834	0.903	0.953	-1.01	0.006	1.007
2	Quantity_paper	Quantity_liquidpaper	0.872	0.923	0.963	-1.01	-0.00	0.999
3	Quantity_book, Quantity_paper	Quantity_liquidpaper	0.834	0.931	0.967	-0.95	0.006	1.007
4	Quantity_book	Quantity_liquidpaper	0.883	0.931	0.966	-1.01	0.006	1.007
5	Quantity_liquidpaper	Quantity_paper	0.872	0.944	0.973	-0.97	-0.00	0.999
6	Quantity_book, Quantity_liquidpaper	Quantity_paper	0.834	0.945	0.974	-0.93	0.000	1.001
7	Quantity_book	Quantity_paper	0.897	0.945	0.973	-1	0.001	1.001
8	Quantity_paper	Quantity_book	0.897	0.949	0.975	-0.99	0.001	1.001
9	Quantity_liquidpaper	Quantity_book	0.883	0.955	0.978	-0.96	0.006	1.007
10	Quantity_paper, Quantity_liquidpaper	Quantity_book	0.834	0.957	0.980	-0.91	0.007	1.009