Stationery case study

The stationery database consists of 2 tables:

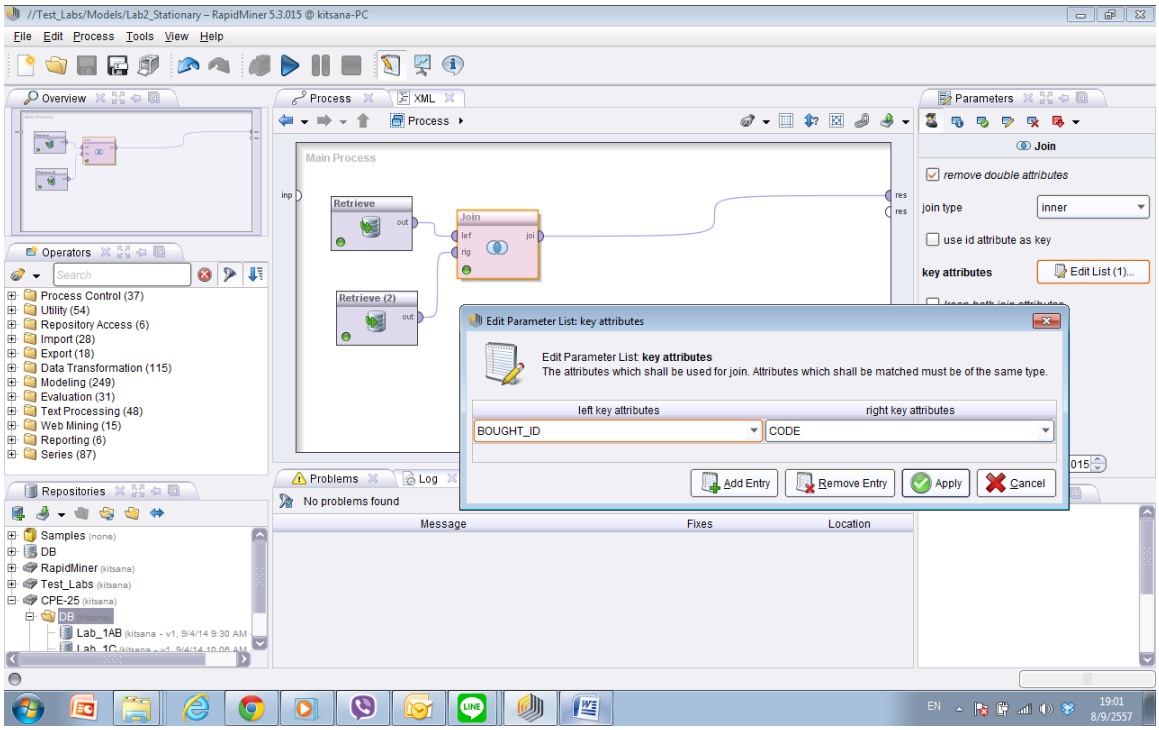
- "PRODUCT_SALE" table contains information about items bought by each customer. Each bought item is stored using its code in the column "BOUGHT_ID".
- "STOCK_NAME" table contains description of each item.

You have to:

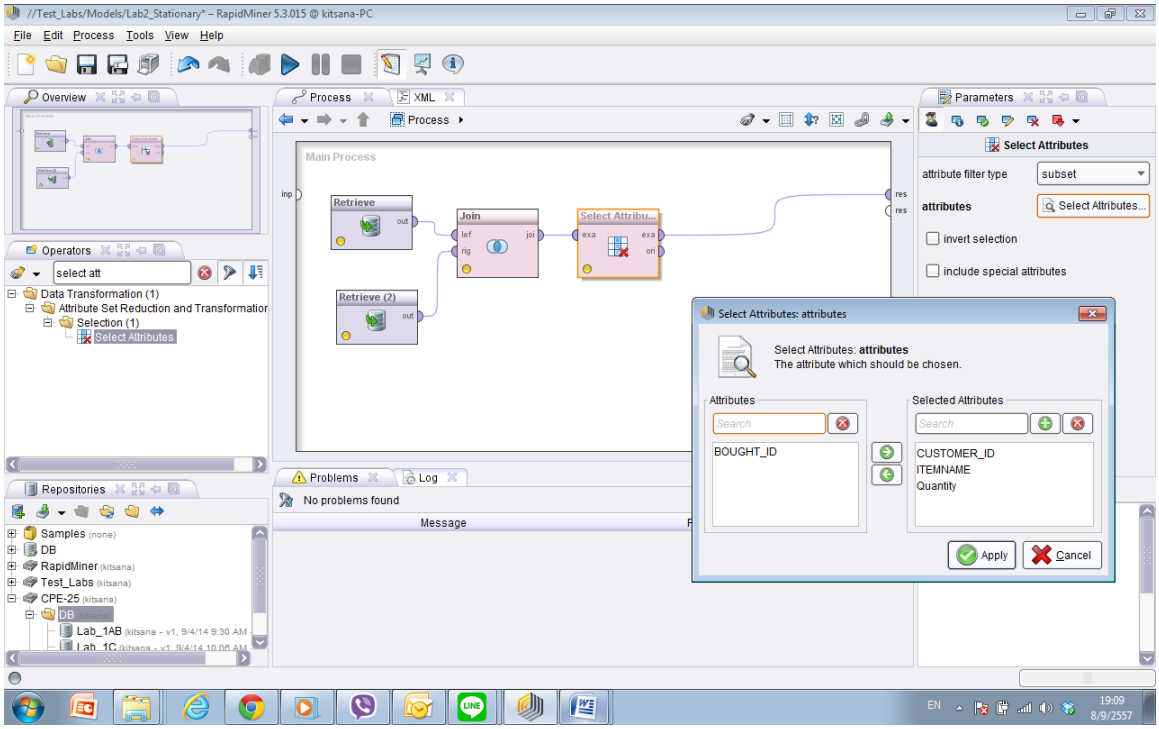
a) Retrieve the input data

b) Pre-process data using the following operators:

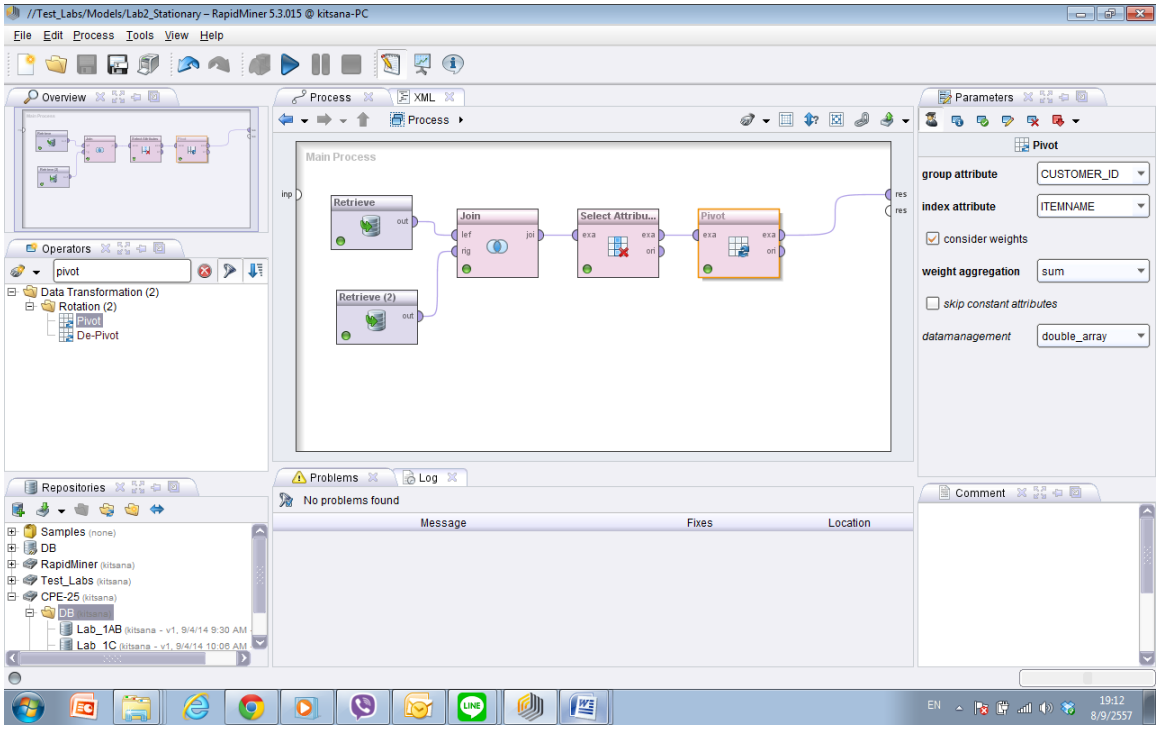
- Join



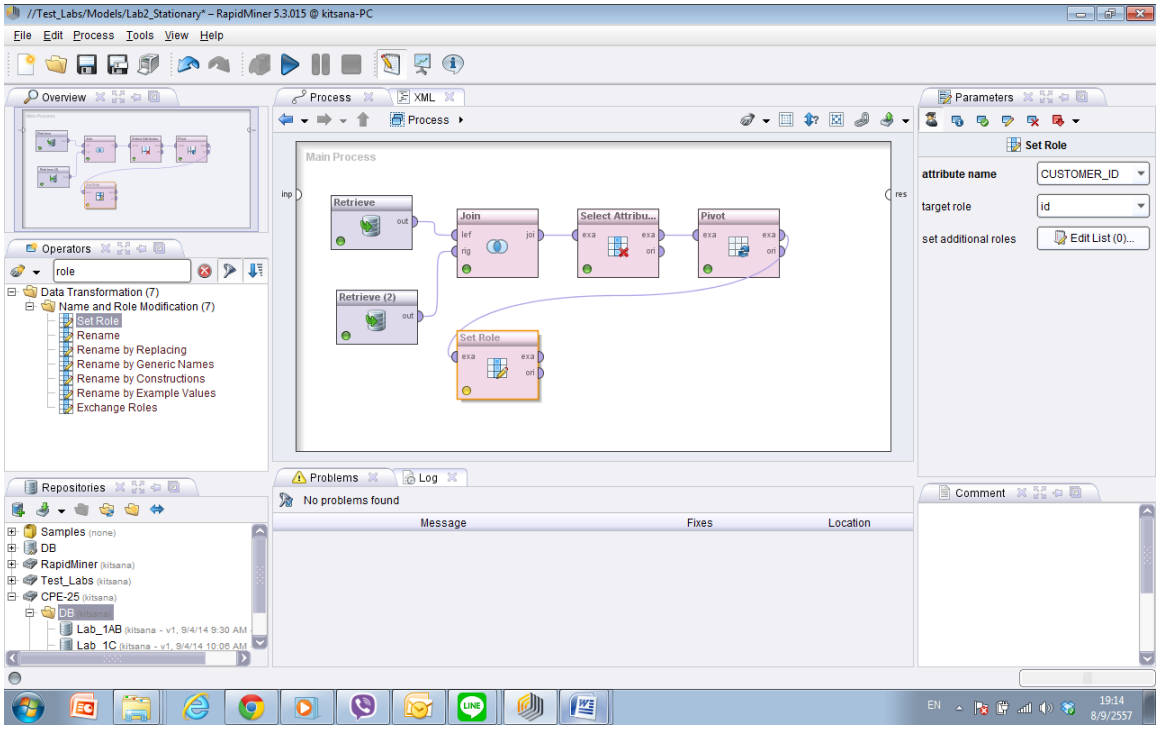
- Select attributes



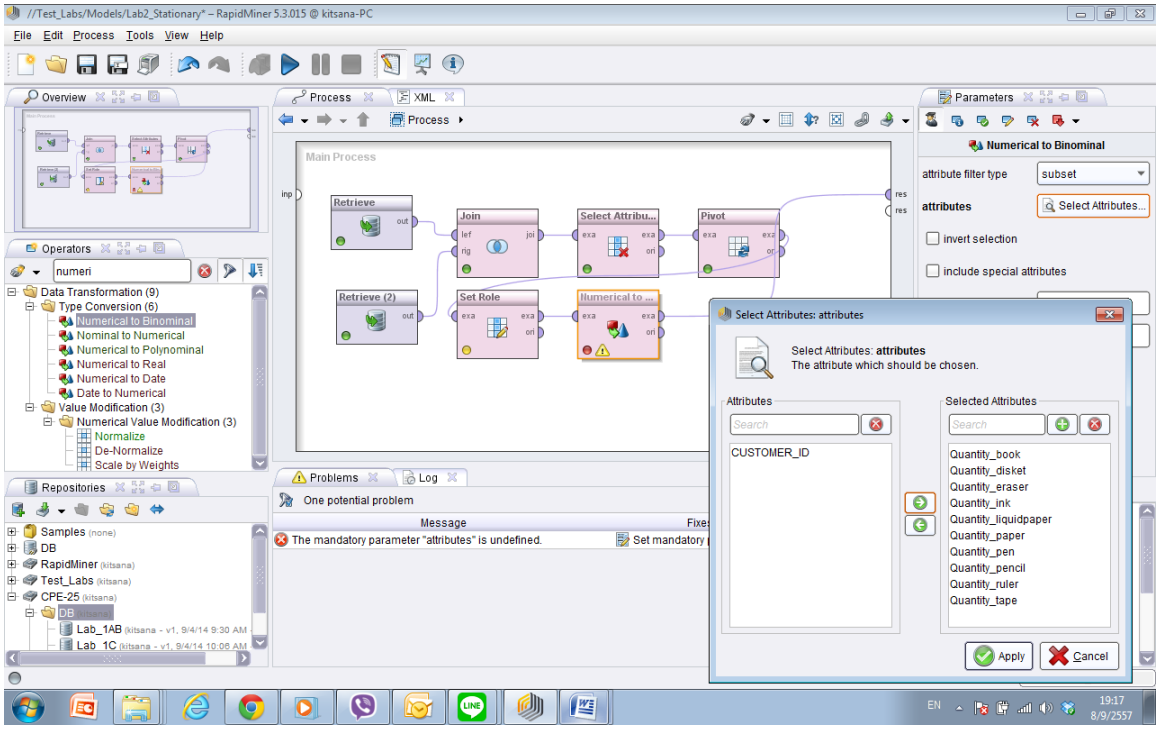
- Pivot



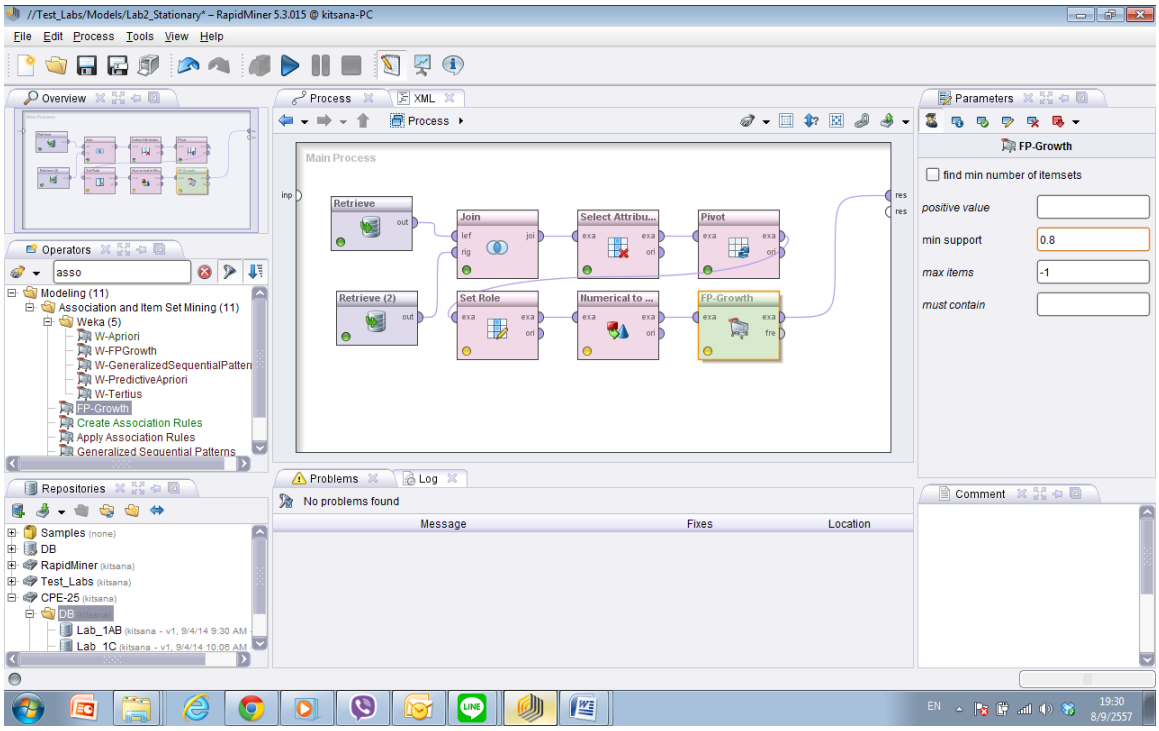
- Set role



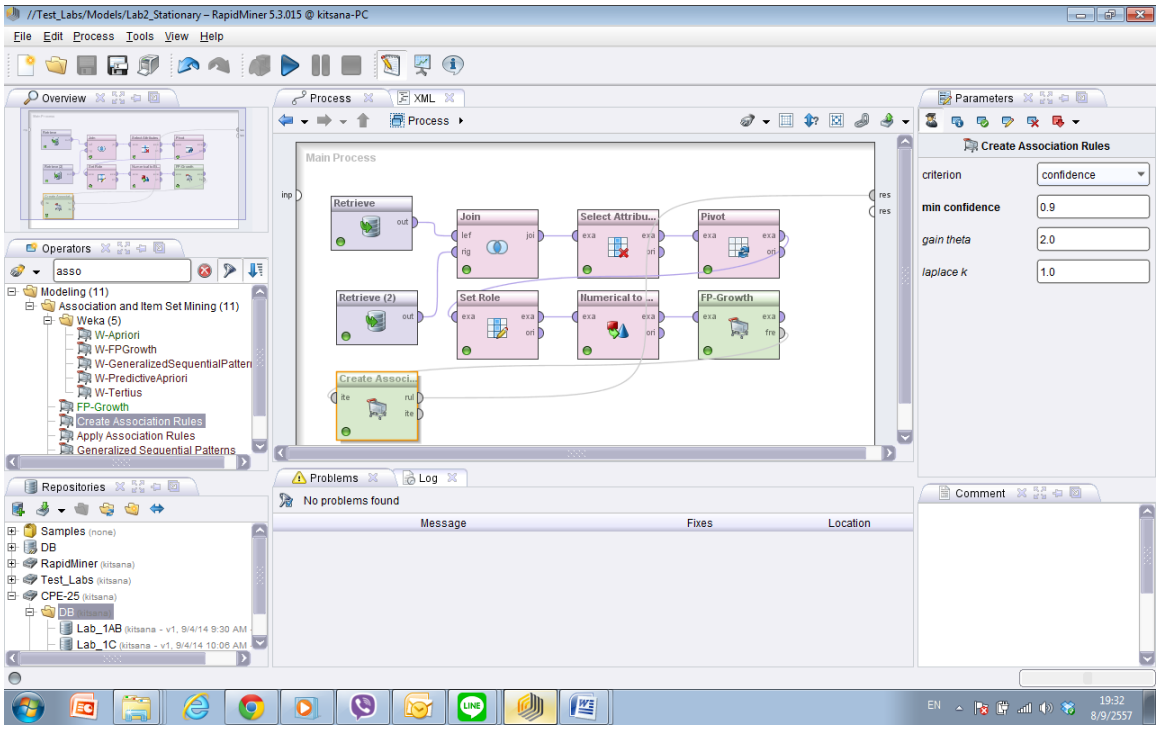
- Numerical to Binomial



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