MIS772

Predictive Analytics

Workshop: Association Rule Mining and sequential pattern analysis

(a.k.a. Market Basket Analysis)
Frequent itemset discovery and association rule analysis.

DEAKIN BUSINESS SCHOOL

AACSB ACCREDITED

EFMD EQUIS ACCREDITED

Workshop Plan

Objectives:
The task is to learn how to use RapidMiner to pre-process a data set for association rule analysis and then carry out the main tasks of rule analysis, i.e., frequent itemset discovery and association rule analysis.

Data Set:
Market Basket Analysis (from RapidMiner)

Method:
Attend the workshop, follow the tutor's demo and instructions, take notes. Note that the lecture and on-cloud lab session will be recorded and their videos linked to the CloudDeakin topic for later access and study.

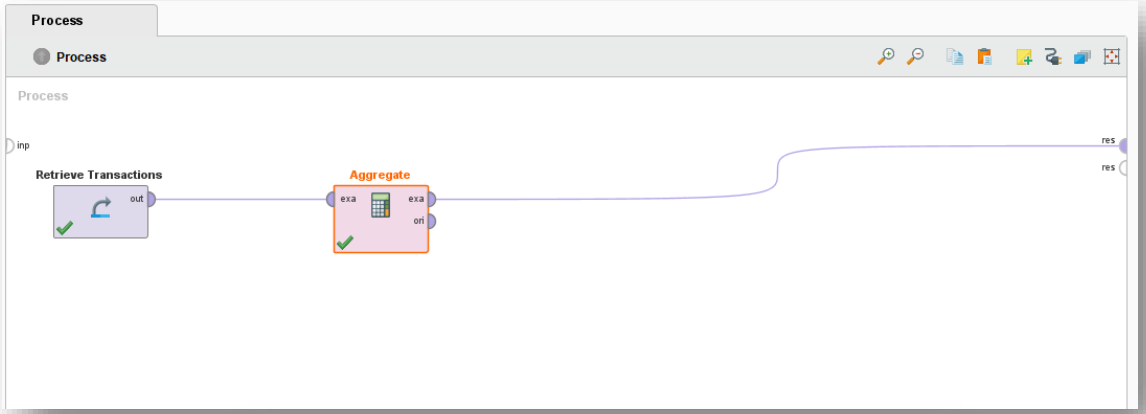
Overview of tasks

- **Task 1 – Acquire and aggregate data**
- **Task 2 – Pre-process transactions for rule analysis**
- **Task 3 – Extract frequent itemsets and association rules**

DEAKIN BUSINESS SCHOOL

Task 1 – Acquire and aggregate data

- Create a new process.
- In the “Repository” tab of RapidMiner, navigate to Samples>>Templates>>Market Basket Analysis.
- Drag the “Transactions” data store to the blank process and connect it to the “res” port of the process.
- Run the process and explore data in the “Results” perspective.
- Explore the data, e.g.,:
  - How many invoices and products are there in this data set?
  - What is the average Sales value in this data set?
- Add the “Aggregate” operator to the process. Select aggregation attributes, select Sales value with the aggregation function set to Sum. Set the “group by attributes” to Invoice.
- Run the process and explore data in the “Results” perspective.
  - Which invoices have the highest Sales value?

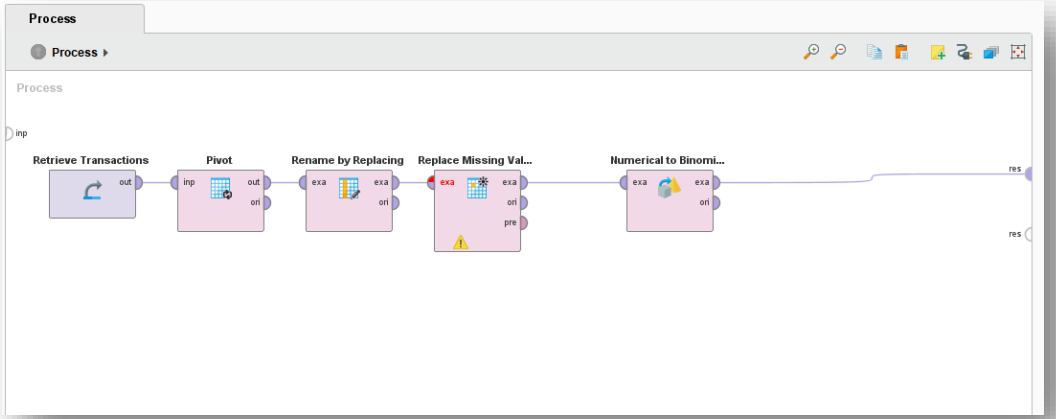


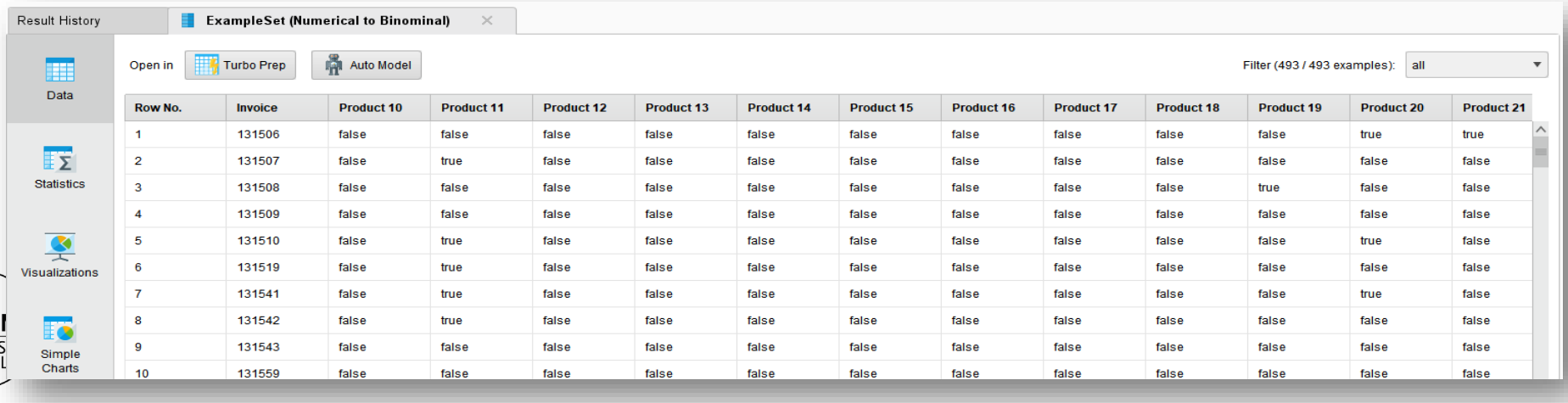
Row No.	Invoice	sum(Sale...
176	646220	345
202	646324	265
480	647978	253
20	1306954	245
196	646318	242
347	647181	240
243	646552	224
368	647285	221
305	646982	214
390	647381	212
326	647096	210
416	647562	210

DEAKIN BUSINESS SCHOOL

Task 2 – Pre-process transactions for rule analysis

- Continue with the process from Task 1.
- Add “Pivot” to the process and select Invoice for “group by attributes”, product 1 for “column grouping attribute”, and Orders as “aggregation attributes” with Sum as aggregation function.
- Add “Rename by Replacing” to the process. Set “attribute filter type” to all, and set “replace what” to **sum\(Orders\)_**
- Add “Replace Missing values” to the process. Set “attribute filter type” to all, and default to zero.
- Add “Numerical to Binomial” to the process. Set “attribute filter type” to all.
- Run the process and explore the results in the “Results” perspective.
  - What does this data set show?
  - If you wanted to see the original Transactions data in the “Results” perspective too, what would you add to this process and where?





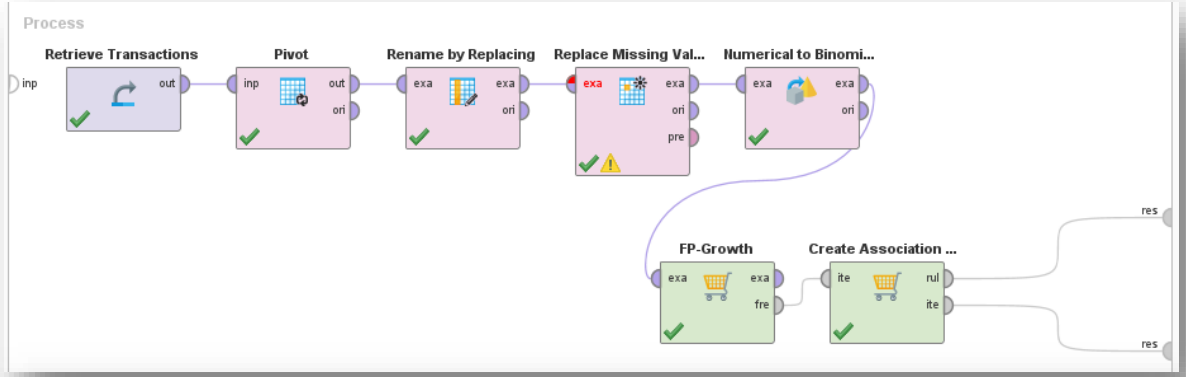
Row No.	Invoice	Product 10	Product 11	Product 12	Product 13	Product 14	Product 15	Product 16	Product 17	Product 18	Product 19	Product 20	Product 21
1	131506	false	false	false	false	false	false	false	false	false	false	true	true
2	131507	false	true	false	false	false	false	false	false	false	false	false	false
3	131508	false	false	false	false	false	false	false	false	false	true	false	false
4	131509	false	false	false	false	false	false	false	false	false	false	false	false
5	131510	false	true	false	false	false	false	false	false	false	false	true	false
6	131519	false	true	false	false	false	false	false	false	false	false	false	false
7	131541	false	true	false	false	false	false	false	false	false	false	true	false
8	131542	false	true	false	false	false	false	false	false	false	false	false	false
9	131543	false	false	false	false	false	false	false	false	false	false	false	false
10	131559	false	false	false	false	false	false	false	false	false	false	false	false

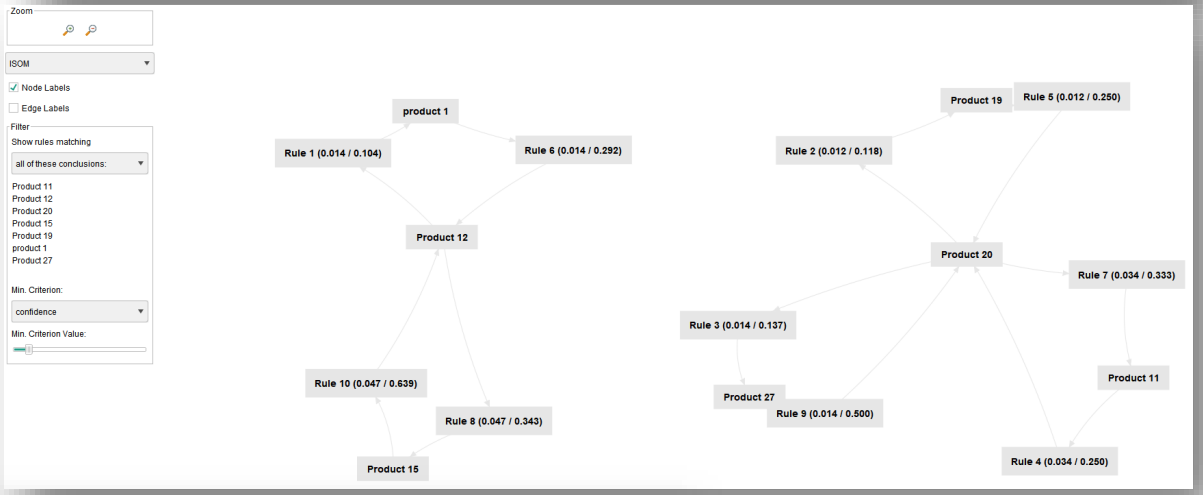
Task 3 – Extract frequent itemsets and association rules

- Continue with the process from Task 2.
- Add “FP-Growth” to the process. Connect the “Numerical to Binomial” operator to FP-Growth.
- In the FP-Growth parameters list, set “positive value” to true, “min support” to 0.05, and “min number of itemsets” to 30.
- Add “Create Association Rules” to the process and connect the “fre” port of “FP-Growth” to the “ite” port of “Create Association Rules”, and then, connect both output ports of the latter to the result ports of the process.
- Run the process and explore the results in the “Results” perspective.

  What is the most frequent two-item itemset found using “FP-Growth”?

- Save your work in the folder for this workshop, under your local repository.





Show rules matching

all of these conclusions:

Product 11
Product 12
Product 20
Product 15
Product 19
product 1
Product 27

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift ↓	Convicti...
9	Product 27	Product 20	0.014	0.500	0.986	-0.043	0.011	4.833	1.793
8	Product 12	Product 15	0.047	0.343	0.921	-0.225	0.037	4.701	1.412
10	Product 15	Product 12	0.047	0.639	0.975	-0.099	0.037	4.701	2.393
4	Product 11	Product 20	0.034	0.250	0.909	-0.241	0.020	2.417	1.195
5	Product 19	Product 20	0.012	0.250	0.965	-0.085	0.007	2.417	1.195
7	Product 20	Product 11	0.034	0.333	0.938	-0.172	0.020	2.417	1.293
6	product 1	Product 12	0.014	0.292	0.967	-0.083	0.008	2.146	1.220

DEAKIN BUSINESS SCHOOL