Name: Varun Ahire

Roll No.: A-20

Subject: Data Mining and Warehousing

Experiment No. : 3

Title :

Apply a-priori algorithm to find frequently occurring items from given data and generate strong association rules using support and confidence thresholds.

Objectives :

Model associations between products by determining sets of items frequently purchased together and building association rules to derive recommendations.

Hardware Requirement :

Pentium or higher processor, 2GB RAM and 500 GB HDD.

Software Requirement :

Rapid Miner

Outcomes :

Create association rules which can be used for product recommendations depending on the confidences of the rules

Theory :

- **Association rule for mining:**
    - Proposed by R Agrawal and R Srikant in 1994.
    - It is an important data mining model studied extensively by the database and data mining community.
    - Assume all data are categorical.

- Initially used for Market Basket Analysis to find how items purchased by customers are related.

- **The Apriori algorithm:**

  - The best known algorithm
  - Two steps:
    - Find all item sets that have minimum support (frequent item sets, also called large item sets).
    - It Create Association rule with support and Confidence.
    - E.g. if we buy tooth brush : it suggest Colgate and tongue cleaner

- **Data Set**

T-Id	Item Set
T-1000	M,O,N,K,E,Y
T-1001	D,O,N,K,E,Y
T-1002	M,A,K,E
T-1003	M,U,C,K,Y
T-1004	**C,O,O,K,E**

Table : Data Set

Given: Minimum **Support** = 60%
Minimum **Confidence** = 80%

- **Candidate Table C1**: Now find support count of each item set

Item Set	Support Count
M	3
O	4
N	2
E	4
Y	3
D	1
A	1
U	1
C	2
K	5

Table: Candidate Table C1

- Now find out minimum Support
- Support = 60/100*5
  =3
- Where 5 is Number of entry
- Compare Min Support with each item set

- **L1 Support Count**

Item Set	Support Count
M	3
O	4
K	5
E	4
Y	3

Table: L1 Support Count

- **Candidate Table C2:**

Item Set	Support Count
MO	1
MK	3
ME	2
MY	2
OK	3
OE	3
OY	2
KE	4
KY	3
EY	2

Table: Candidate Table C2

- Now again Compare C2 with Min Support 3

- **L2 Support Count**

Item Set	Support Count
MK	3
OK	3
OE	3
KE	4
KY	3

Table: L2 Support Count

- After satisfied minimum support criteria
- Make Pair to generate C3

- **Candidate Table C3**

Item Set	Support count
M,K,O	1
M,K,E	2
M,K,Y	2
O,K,E	3
O,K,Y	2

Table: Candidate Table C3

- **L3 Support Count**

  Now again compare the item set with min support 3

Item Set	Support Count
O,K,E	3

Table: L3 Support Count

- **Now create** association rule with support and Confidence for {O,K,E}
  - Confidence =Support/No. of time it Occurs

Association Rule	Support	Confidence	Confidence (%)
O ^ K ⇒ E	3	3/3 = 1	1*100=100
O ^ E ⇒ K	3	3/3 = 1	1*100=100
K ^ E ⇒ O	3	3/4 = 0.75	0.75*100=75
E⇒ O ^ K	3	3/4 = 0.75	0.75*100=75
K⇒ O ^ E	3	3/5 = 0.6	0.6*100=60
O⇒ K ^ E	3	3/4 = 0.75	0.75*100=75

Table: Association Rule

- Compare this with Minimum Confidence=80%

Rule	Support	Confidence
O ^ K ⇒ E	3	100
O ^ E ⇒ K	3	100

Table: Support and Confidence

Hence final Association rule are
{O ^ K ⇒ E}
{O ^ E ⇒ K}

- From first observation we predict that if the customer buy item O and item K then defiantly he will by item E
- From Second observation we predict that the customer buy item O and item E then defiantly he will by item K

- **Market Basket Analysis using Rapid Miner**

Rapid Miner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization Rapid Miner is developed on an open core model. The Rapid Miner Studio Free Edition, which is limited to 1 logical processor and 10,000 data rows, is available under the AGPL license.

Commercial pricing starts at $2,500 and is available from the developer.

- **MARKET BASKET ANALYSIS**

  Model associations between products by determining sets of items frequently purchased together and building association rules to derive recommendations.

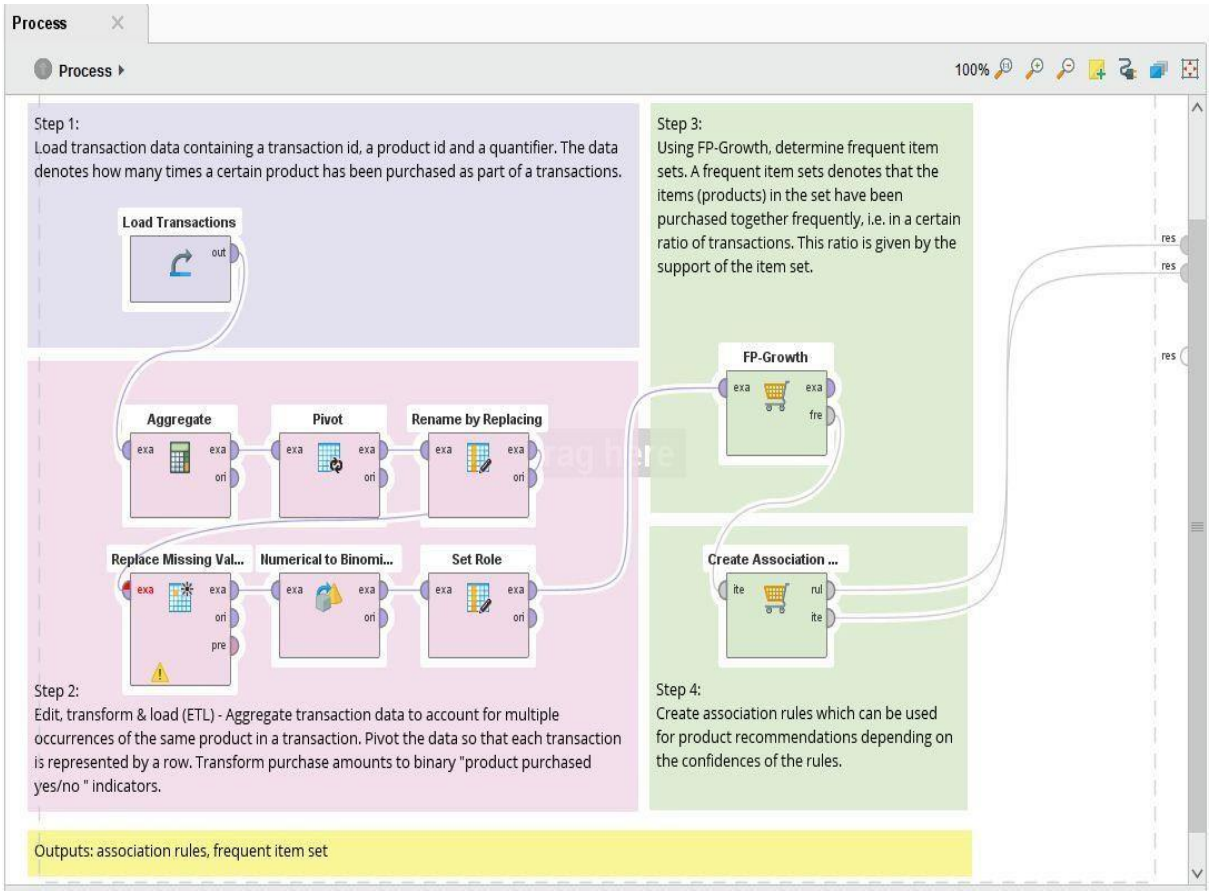

Figure: MARKET BASKET ANALYSIS

No. of Sets: 47
Total Max. Size: 3

Min. Size: 1
Max. Size: 3
Contains Item:
Update View

Size	Support	Item 1	Item 2	Item 3 ↓
3	0.006	Product 12	Product 20	Product 27
3	0.006	Product 11	Product 12	Product 20
3	0.006	Product 11	Product 20	Product 19
1	0.138	Product 11		
1	0.136	Product 12		
1	0.103	Product 20		
1	0.079	Product 10		
1	0.079	Product 18		
1	0.079	Product 23		
1	0.073	Product 15		
1	0.071	Product 26		
1	0.067	Product 13		
1	0.059	Product 21		

Figure: Frequent Item Sets (FP Growth)

Conclusion :

Thus we learn that to find frequently occurring items from given data and generate strong association rules using support and confidence thresholds using a-priori algorithm.