Name : Sakshi Aher
Class: BE (A)

Roll No: 26

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. For Example: Market Basket Analysis.

Objective:

1. Model associations between products by determining sets of items frequently purchased together and building association rules to derive recommendations.
2. Learn Market Basket Analysis using A-priori & FP growth Algorithm using RapidMiner

Hardware Requirement:

Any CPU with Pentium Processor or similar, 256 MB RAM or more,1 GB Hard Disk or more

Software Requirements:

32/64-bit Linux/Windows Operating System, latest RapidMiner Tool

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 A-priori 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 suggests Colgate and tongue cleaner

- MILK > BREAD
- MILK > EGGS
- MILK > BREAD > EGGS

Support :

Market Basket Optimisation: $$\text{support}(\boldsymbol{I}) = \frac{\#\ \text{transactions containing } \boldsymbol{I}}{\#\ \text{transactions}}$$

Movie Recommendation: $$\text{support}(\boldsymbol{M}) = \frac{\#\ \text{user watchlists containing } \boldsymbol{M}}{\#\ \text{user watchlists}}$$

- **How many people have seen X-Machina**
- A: 10 / 100
- **Support = 10% Confidence:**

Market Basket Optimisation: $$\text{confidence}(\boldsymbol{I_1} \rightarrow \boldsymbol{I_2}) = \frac{\#\ \text{transactions containing } \boldsymbol{I_1} \text{ and } \boldsymbol{I_2}}{\#\ \text{transactions containing } \boldsymbol{I_1}}$$

Movie Recommendation: $$\text{confidence}(\boldsymbol{M_1} \rightarrow \boldsymbol{M_2}) = \frac{\#\ \text{user watchlists containing } \boldsymbol{M_1} \text{ and } \boldsymbol{M_2}}{\#\ \text{user watchlists containing } \boldsymbol{M_1}}$$

- People who have Watched Interstellar, are likely to like Ex-Machine as well
- A: 40 watched Interstellar
- out of 40, only 7 watched Ex-Machina
- **Confidence = 7 / 40 = 17.5%**

Lift :

Market Basket Optimisation: $$\text{lift}(\boldsymbol{I_1} \rightarrow \boldsymbol{I_2}) = \frac{\text{confidence}(\boldsymbol{I_1} \rightarrow \boldsymbol{I_2})}{\text{support}(\boldsymbol{I_2})}$$

Movie Recommendation: $$\text{lift}(M_1 \rightarrow M_2) = \frac{\text{confidence}(M_1 \rightarrow M_2)}{\text{support}(M_2)}$$

- People who watched
- Interstellar 40 / 100
- Ex-Machina 07 / 40
- What is the Likely hood if we recommend Ex-Machina to person who has watched Interstellar?
- LIFT = Confidence / Support

=17.5% / 10%

= 1.75

Algorithm:

1. Set a min support & confidence
2. Take all the Subsets in transactions
3. Take all the rules these subsets having higher confidence than minimum confidence
4. Sort the rules by decreasing lift

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 1: Data Set

GivenM: inimum **Suppor=t**60% Minimum **Confidenc=e**80%

Candidate Tabl:e NCo**1**w 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 2 : 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 3: 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 4: 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 5: 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 6: Candidate Table C3

Support Count

Now again compare the item set with min support 3

Item Set	Support Count
O,K,E	3

Table 7: L3 Support CountNow

**creaatsseociation 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 8: Association Rule

- Compare this with Minimum Confidence=80%

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

Table 9: 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.

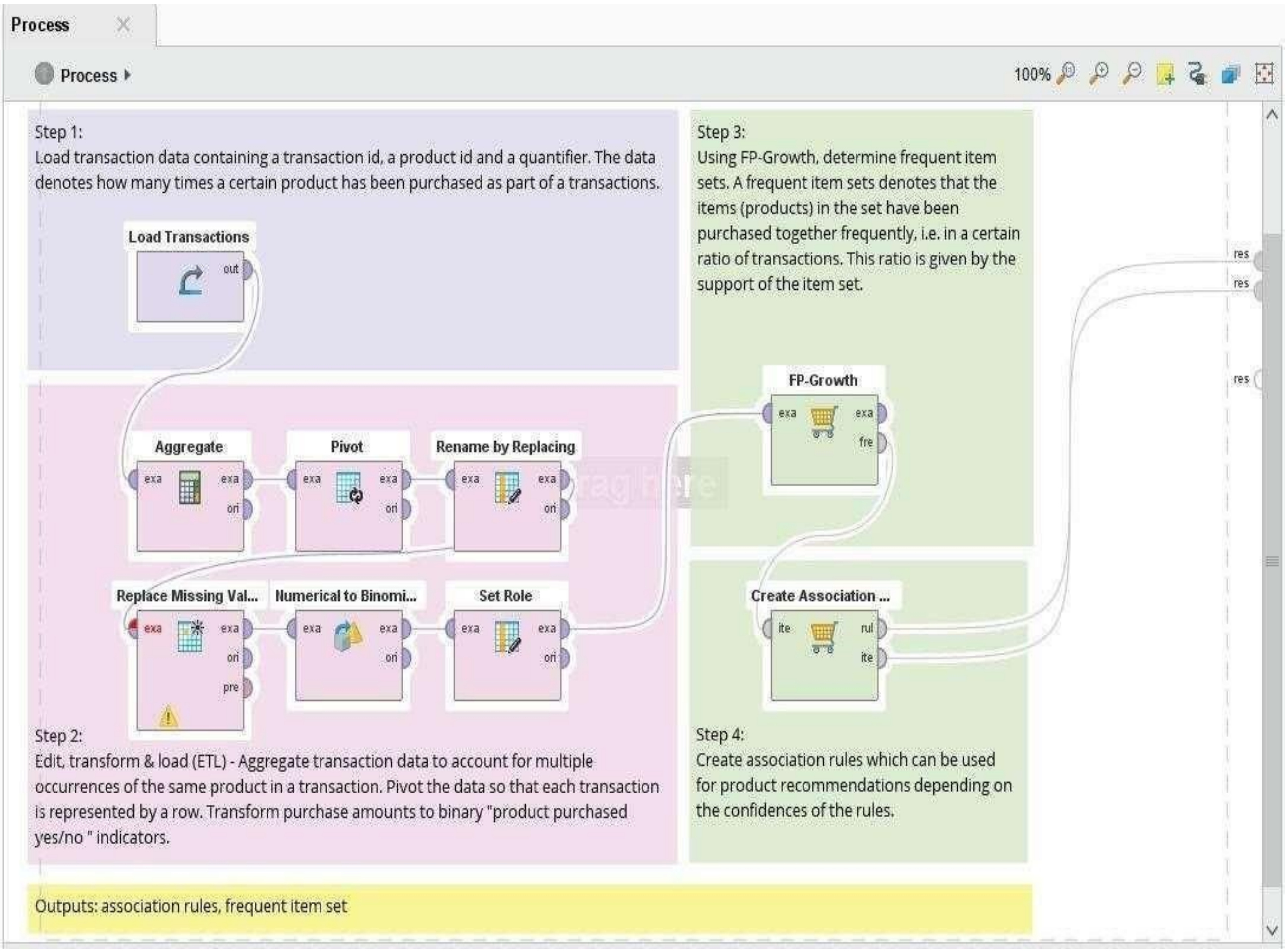

Figure 1: MARKET BASKET ANALYSIS

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 2: 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