Association Rule

* Association rule learning can be divided into 3 kinds of algorithm
  * Apriori
  * Eclat
  * T-P Growth

* Association rule learning works on the concept of if and else statement

  eg: If A then B.

  where if A is an incident

  then B is consequent.

Let us discuss about Apriori algorithm

* A priori algorithm is generally considered as unsupervised learning algorithm

~~problem~~ formula's

~~For~~

$$\text{support} = \frac{\text{No of items x \& Y bought together}}{\text{Total no of transcations}}$$

$$\boxed{\text{Support} = \frac{fre(x,y)}{N.}}$$

$$\text{confidence} = \frac{\text{Total no. of items x \& Y bought together}}{\text{Total no. of items 'x' bought}}$$

$$\text{confidence} = \frac{\text{Freq}(x,y)}{\text{Freq}(x)}.$$

$$\text{lift} = \frac{\text{support}(x,y)}{\text{support}(x) + \text{support}(y)}.$$

Problem

Construct the following transcations apply the association rule mining to get the association rule with minimum support two and confidence of 50%.

⇒ Association rule mining is explained using Apriori Algorithm

Transcational data for All electronics branch.

T_{ID}	List of item IDs
T_{100}	I_1, I_2, I_5
T_{200}	I_2, I_4
T_{300}	I_2, I_3
T_{400}	I_1, I_2, I_4
T_{500}	I_1, I_3
T_{600}	I_2, I_3
T_{700}	I_1, I_3
T_{800}	I_1, I_2, I_3, I_5
T_{900}	I_1, I_2, I_3

Step 1 :

Now scan each itemset and count how many candidate each has.

1-itemset	supcount
$\{I_1\}$	6
$\{I_2\}$	7
$\{I_3\}$	6
$\{I_4\}$	2
$\{I_5\}$	2

Step 2

Check whether the candidate support count is with minimum of "two".

1-itemset	supcount
$\{I_1\}$	6
$\{I_2\}$	7
$\{I_3\}$	6
$\{I_4\}$	2
$\{I_5\}$	2

step3

Now Generate two candidate keys.

2-itemset

{I1,I2}
{I1,I3}
{I1,I4}
{I1,I5}
{I2,I3}
{I2,I4}
{I2,I5}
{I3,I4}
{I3,I5}
{I4,I5}.

step4:

Now Scan the support count of candidate keys.

2-itemset	Supcount.
{I1,I2}	4
{I1,I3}	4
{I1,I4}	1
{I1,I5}	2
{I2,I3}	4
{I2,I4}	2
{I2,I5}	2
{I3,I4}	0
{I3,I5}	1
{I4,I5}	0

5:

Check whether the candidate support count is with minimum of 'two', if not remove the candidate key. (from step 4)

2-itemset	supcount
$\{I_1, I_2\}$	4
$\{I_1, I_3\}$	4
$\{I_1, I_5\}$	2
$\{I_2, I_3\}$	4
$\{I_2, I_4\}$	2
$\{I_2, I_5\}$	2

Step 6:

Now Generate three candidate keys. with sup count

3 itemset	supcount
$\{I_1, I_2, I_3\}$	2
$\{I_1, I_2, I_5\}$	2
$\{I_1, I_2, I_4\}$	1
$\{I_1, I_3, I_4\}$	0
I_1, I_3, I_5	1
I_1, I_4, I_5	0
I_2, I_3, I_4	0
I_2, I_3, I_5	1
I_2, I_4, I_5	0

Step 7:

Check whether the candidate support count is with minimum of 'two'

3-itemset	supcount
$\{I_1, I_2, I_3\}$	2
$\{I_1, I_2, I_5\}$	2

Step 8:

Generate four candidate keys.

itemset	supCount
$\{I_1\ I_2\ I_3 I_5\}$	1

So not possible, because it has only 1 supCount

* Non-empty subsets of frequency sets are three item sets

$\{I_1\ I_2\ I_3\}$ $\{I_1\ I_2\ I_5\}$.

→ $\{\{I_1),(I_2),(I_3),(I_1,I_2),(I_1,I_3),(I_2,I_3)\}$

→ $\{(I_1),(I_2),(I_5),(I_1,I_2),(I_1,I_5),(I_2,I_5)\}$.

* Association rule for every non-empty subsets

$$S \Rightarrow (I - S)$$

* Let us consider, Association rule mining subset creation between three Items sets.

$I \Rightarrow \{I_1\ I_2\ I_3\}$.

Non-empty subsets.

$\{(I_1)\ (I_2)\ (I_3)\ (I_1\ I_2)\ (I_1, I_3)\ (I_2 I_5)\}$.

le 1 :

$\{I_1\} \rightarrow \{I_2, I_3\}$.

$\downarrow$ S $\downarrow$ I-S

support $= \frac{2}{9} = 22.22\%$

Confidence $=$ support $\frac{(1,2,3)}{\text{support}(1)} = \frac{2/9}{6/9} = 1/3 = 33.33\%$.

R ∴ The condition is invalid ($< 50\%$).

Rule 2

$\{2\} = \{1,3\}$

Support $= 2/9 = 22.22\%$.

Confidence $= \frac{\text{support } (1,2,3)}{\text{support } (2)} = \frac{2/9}{7/9} = 2/7 = 28.57\%$.

∴ The condition is invalid ($< 50\%$)

Rule 3

$\{3\} = \{1,2\}$

support $= 2/9 = 22.22\%$.

Confidence $= \frac{\text{support } (1,2,3)}{\text{support } (3)} = \frac{2/9}{6/9} = 2/6 = 1/3$ 33.34

∴ The condition is invalid ($< 50\%$

Rule 4

$\{1,2\} = \{3\}$

support $= 2/9 = 22.22\%$.

Confidence = $\frac{\text{support}(1,2,3)}{\text{support}(1,1)} = \frac{2/9}{4/9} = 2/4 = 50\%$

∴ The above condition is valid (>=50%).

le 5

$\{1,3\} = \{2,3\}$

support = $2/9 = 22.22\%$

confidence = $\frac{\text{support}(1,2,3)}{\text{support}(1,3)} = \frac{2/9}{4/9} = 2/4 = 50\%$

∴ The above condition is valid (>=50%)

le 6

$\{2,3\} = \{1,3\}$

support = $2/9 = 22.22\%$

confidence = $\frac{\text{support}(1,2,3)}{\text{support}(2,3)} = \frac{2/9}{7/9} = 2/7 = 28.57$

∴ The above condition is invalid (< 50%)