Unit - III

Association Rule mining.

→ Association rule mining is a procedure which aims to observe frequently occuring patterns, correlations (or)

Association Analysis :-

It discovers the probability & occurence & items in a collection. Helps in discovering some interesting relationships in large data set.

A dataset contains data objects and each data object contains a set & attributes. An attribute is also called as dimension (or) feature (or) variable which represents the characteristics features & a data object

eg;- height, qualification, colour etc.

Association Rule mining ;-

It finds the interesting associations and relationships among large set & data items. This rule shows how frequently a itemset occurs in a transaction.

Eg:- Market Basket Data.

Transaction ID	Items.
1.	{Milk, Bread, Rice, Book}
2.	{Bread, Jam, Book, Pen}
3.	{Jam, Milk, Bread, Rice, Eggs}.
4.	{Rice, Eggs, Pen, Book}
5.	{Eggs, Pen, Milk, Bread, Jam}
6.	{Eggs, Rice, Bread, Jam}.

Let us consider one transaction like.

{Milk, Bread, Rice, Book} {Milk} → {Bread}

{Bread, Jam, Book, Pen} {Book} → {Pen}, {Bread} → {Jam}

Some similar associations.

{Dish wash liquid} → {Scrubber}

{Laptop} → {Mouse}

{Floor Stick} → {Floor cleaner}.

① Itemset Generation

② Frequent Itemset

③ Support findout

④ confidence

⑤ Rule mining

Itemset : {Milk, Bread, Jam, Rice, Eggs, Book, Pen}

TID	Items						
	Milk	Bread	Jam	Rice	Eggs	Book	Pen
1	1	1		1		1	
2		1	1			1	1
3	1	1	1	1	1		
4				1	1	1	1
5	1	1	1		1		1
6		1	1	1	1		

Frequent Itemsets :-

Two Itemsets : {Milk, Bread}, {Bread, Jam}, {Rice, eggs}, {Book, Pen}

three " : {Milk, Bread, Jam} {Rice, Eggs, Bread}, {Book, Pen, Eggs}

Four " : {Milk, Bread, Rice, eggs}. etc.

Support : It is a measure of how frequently a set of items occur in total no. of transactions.

{Milk, Bread} → {x, y} {x: milk); (y: Bread)

Therefore the frequency of occurance of x and y togeather in total no. of transactions is support.

{Milk, Bread, Jam} → {x, y} {(x: milk), (y: Bread, Jam)}

Here, the frequency of occuance of (Bread, Jam) with {Milk} in whole transactions is Support.

$$\text{Support } (s) = \frac{\sigma(X \cup Y)}{N}$$

Confidence :- It is a measure of how often items in Y appear in transactions that contain X.

{Milk, Bread, Jam} → {X, Y} :: (X: Milk), (Y: Bread, Jam).

Therefore the frequency of occurence of X and Y in all the transactions where X exists.

$$\text{Confidence } (c) = \frac{\sigma(X \cup Y)}{\sigma X}$$

Association Rule Mining :- Given a set of transactions T, the goal of association rule mining is to find all rules having.

Support ≥ minsup threshold

Confidence ≥ minconf threshold.

TID	Items.
1.	{Milk, Bread, Rice, Book}
2.	{Bread, Jam, Book, Pen}
3.	{Jam, Milk, Bread, Rice, Eggs}.
4.	{Rice, Eggs, Pen, Book}
5	{Eggs, Pen, Milk, Bread, Jam}.
6.	{Eggs, Rice, Bread, Jam}.

Eg:- Suppose, minsup = 0.3
min conf = 0.6.

Consider, {Rice, Eggs} → {X, Y}.

then, Support (s) = $\frac{\sigma(X \cup Y)}{N}$

Support (s) = $\frac{1+1+1}{6} = \frac{3}{6} = 0.5$

And Confidence (c) = $\frac{\sigma(X \cup Y)}{\sigma X}$

confidence (c) = $\frac{3}{4} = 0.75$.

Association Rule Mining,

here, Support: 0.5 ≥ minsup (0.3)
confidence: 0.75 ≥ minconf (0.6)

Therefore, we can mine {Rice, eggs} as a Rule.

Eg 2:

Suppose, minsup = 0.3

mincont = 0.6.

Consides, {Milk, Bread, Jam} → {X, Y}.

then, Support (s) = $\frac{\sigma(X \cup Y)}{N}$

$= \frac{1+1}{6} = \frac{2}{6} = 0.333.$

Confidence (c) = $\frac{\sigma(X \cup Y)}{\sigma X}$

$= \frac{2}{3} = 0.667.$

Association Rule mining,

here, Support: 0.333 ≥ minsup (0.3)

Confidence: 0.667 ≥ mincont (0.6)

Therefore, we can mine {Milk, Bread, Jam} as a Rule.