Apriori Algorithm

- Goal of any organization is to increase the revenue.

- Organizations began mining data relating to frequently bought items. Market basket analysis is one of the key techniques used by large retailers to uncover associations between items.

Example * customers who purchase bread have a 60% likelihood to also purchase Jam

* customers who buy laptops are more likely to purchase laptop bags as well.

** To find association between different items & products that can be sold together which gives in the right product placement

Ex People who buy bread also tend to buy butter & the marketing team at retail store should target customers ~~who~~ to make them buy a third item suppose eggs.

If a customer buys bread & butter & sees a discount offer on eggs, he will be encouraged to spend more & buy the eggs.

Association Rule Mining

The ARM rule can be thought of as an if-then relationship just to elaborate on it

$A \Longrightarrow B$

If then

Two ~~sub~~ elements to this rule first is if & second is then.

If is also known as antecedent this is an item or a group of items present in the itemset.

The later is known as consequent

Association Rule Mining algorithm helps the business to make profit.

ARM is all about building the rules

- If you buy A there is a slight possibility or chance that you may buy B also
- This type of relationship in which we can find relationship b/w these two items is known as single cardinality

Measures Association — Support
— Confidence
— lift

$$\text{Support} = \frac{\text{freq}(A,B)}{N}$$

Support is the frequency of or the combination of item A or B. Its basically the frequency of the items we have bought & combination of the items that we have bought

- To filter the items that have been bought less frequently

Confidence

$$\text{Confidence} = \frac{\text{freq}(A,B)}{\text{freq}(A)}$$

It gives us how often the items A & B occur together given the number of the times A occur.

$$\text{Lift} = \frac{\text{Support}(A \cup B)}{\text{Supp}(A) \times \text{Supp}(B)} \quad (A \to B)$$

Lift is basically the strength of any rule

Denomination gives the independent support value of A & B. This gives us the independent occurrence probability of A & B.

Denominator of lift is more, the occurrence of randomness is more.

Ex Transaction at a local market

TID	Items
T_1	A B C
T_2	A C D
T_3	B C D
T_4	A D E
T_5	B C E

Rules

1. $A \Rightarrow D$) A gives D if a person buys A he is most likely to buy D.
2. $C \Rightarrow A$
3. $A \Rightarrow C$
4. $B \& C \Rightarrow A$

Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow$	2/5	2/3	5/6
$B, C \Rightarrow A$	1/5	1/3	5/9

$$\text{Lift}_{A \to D} = \frac{\text{Support}(A \cup D)}{\text{Support}(A) \times \text{Support}(D)}$$

$$= \frac{\frac{2}{5}}{\frac{3}{5} \times \frac{3}{5}} = \frac{2}{5} \times \frac{25}{9} = \frac{5}{9}$$

Apriori algorithm uses frequent itemsets to generate association rules. It is based on the concept that a subset of a frequent itemset must also be a frequent itemset.

* A frequent itemset is an itemset whose support value is greater than a threshold value.

Example: If A & B is a frequent itemsets then A & B should also be frequent itemsets individually.

min_support = 2

TID	Items
T_1	1 3 4
T_2	2 3 5
T_3	1 2 3 5
T_4	2 5
T_5	1 3 5

→

C1

Itemset	Support
{1}	3
{2}	3
{3}	4
{4}	1
{5}	4

L1

Itemset	Support
{1}	3
{2}	3
{3}	4
{5}	4

→

Itemset of size 2

C2

Itemset	Support
{1, 2}	1
{1, 3}	3
{1, 5}	2
{2, 3}	2
{2, 5}	3
{3, 5}	3

L_2

Itemset	Support
{1,3}	3
{1,5}	2
{2,3}	2
{2,5}	3
{3,5}	3

C_3

Itemset	Support
{1,2,3}	✗
{1,2,5}	✗
{1,3,5}	2
{2,3,5}	2

L_3

Itemset	Support
{1,3,5}	2
{2,3,5}	2

C_3

Itemset	In L_2?
{1,2,3}, {1,2} {1,3} {2,3}	NO
{1,2,5}, {1,2} {1,5} {2,5}	NO
{1,3,5}, {1,5} {1,3} {3,5}	Yes
{2,3,5}, {2,3} {2,5} {3,5}	Yes

C_4

Itemset	Support
{1,2,3,5}	1

for $I = \{1,3,5\}$, subsets are $\{1,3\}, \{1,5\}, \{3,5\}, \{1\}, \{3\}, \{5\}$

for $I = \{2,3,5\}$ subsets are $\{2,3\}, \{2,5\}, \{3,5\}, \{2\}, \{3\}, \{5\}$

For every subset S of I,

$S \rightarrow (I - S)$ (S recommends $I - S$)

If support(I) / support(S) >= min_conf value

Applying Rules to Itemset 13

Rule 1: $\{1,3,5\}$

Rule 1: $\{1,3\} \rightarrow \{1,3,5\}$

confidence = support $(1,3,5)$ / support $(1,3)$

$= \frac{2}{3} = 66.66\% > 60\%$. Rule selected.

Rule 2 $\{1,5\} \rightarrow \{1,3,5\}$

confidence = support $(1,3,5)$ / support $(1,5)$

$= \frac{2}{2} = 100\% > 60\%$.

Rule 3 $\{3,5\} \rightarrow \{1,3,5\}$

confidence = support $(1,3,5)$ / support $\{3,5\}$

$= \frac{2}{3} = 66.66\% > 60\%$

Rule 4 $\{1\} \rightarrow \{1,3,5\}$

confidence = support $(1,3,5)$ / support (1)

$= \frac{2}{3} = 66.66\% > 60\%$.

Rule 5 $\{3\} \rightarrow \{1,3,5\}$

Confidence = support $(1,3,5)$ / support $\{3\}$

$= \frac{2}{4} = 50\% < 60\%$. Rule rejected

Rule 6 $\{5\} \rightarrow \{1,3,5\}$

Confidence = support $(1,3,5)$ / support (5)

$= \frac{2}{4} = 50\% < 60\%$. Rule Rejected.