ICS

ASSIGNMENT-7

AIM: BASKET ANALYSIS USING APRIORI ALGORITHM IN R

THEORY:

- **MARKET BASKET ANALYSIS:**
- Technique to identify the strength of association b/w pairs of products together & identify patterns of co-occurrence.

- **ASSOCIATION RULE MINING:**
- Used when you want to find an association b/w different objects in a set, find frequent patterns in a transaction database, relational database & any other info repository.

- It can tell you what items customers frequently buy together by generating a set of rules called association rules.

Ex:

ID.	ITEMS
1.	{ bread, milk }
2.	{ bread, diapers, beer, eggs }
3.	{ milk, diaper, beer, cola }
4.	{ bread, milk, diaper, beer }
5.	{ bread, milk, diaper, cola }

- From the above, association rule that diaper & beer are always bought together is figured.

- Different factors considered to arrive at these association rules are :

   i. Itemset : collection of one or more items

   ii. Support count : Frequency of occurrence of item set

   iii. Support (x) : Fraction of transactions containing item set x.

$$Support\ (x) = \frac{freq.\ (x)}{N}$$

For a rule $A \to B$, support $= \dfrac{freq.\ (A, B)}{N}$

   iv. Confidence : For $A \to B$ confidence shows %age of B bought with A.

$$confidence\ (A \to B) = P(A|B) = \frac{P(A \cdot B)}{P(B)} = \frac{freq.\ (A \cdot B)}{freq\ (B)}$$

   v. Frequent itemsets : support $\geq$ min. (support). Set on user choice.

   vi. Strong Rules : If $A \to B$ (support, confidence) satisfies min. support & min. confidence, then strong rule.

vii. lift : correlation b/w A & B in A→B

$$lift(A→B) = \frac{support}{support(A) \cdot support(B)}$$

- APRIORI ALGORITHM :
- Divided into 2 phases :
  - i. Frequent item-set generation
  - ii. Rule generation.

- Frequent itemset generation :
- Most computationally expensive step. Requires scan of complete database.

- By considering support values, we keep grouping items together & remove those whose support lies below min. support. Finally, only those itemsets are considered whose support is greater than min. support.

- Rule generation :
- Finds confidence for all the frequent items sets & all those sets whose confidence is equal to greater than min. confidence are chosen as association rules.

ex:

transactions	items	
1	I_1, I_2	min. supp = 2
2	I_2, I_3, I_4, I_5	min. conf. : 60%
3	I_2, I_3, I_4, I_5	min. conf. : 60%
4	I_1	
5	I_1, I_2, I_3	

Item	Support count
I_1	3
I_2	4
I_3	3
I_4	1
I_5	1

remove items whose support is less than 2 →

Item	Support
I_1	3
I_2	4
I_3	3

Generating 2 - itemset combinations:

Itemset	Support
$\{I_1, I_2\}$	2
$\{I_1, I_3\}$	1
$\{I_2, I_3\}$	3

→

Itemset	Support
$\{I_1, I_2\}$	2
$\{I_2, I_3\}$	3

$\{I_1, I_3\}$ was removed

Finally

Itemset	Support
$\{I_1, I_2, I_3\}$	1

$\therefore$ support $(\{I_1, I_2, I_3\}) \leq 2$, not considered

Frequent itemsets : $\{I_1, I_2\}$, $\{I_3, I_4\}$

Association Rules	Confidence	
$I_2 \to I_1$	$2/4 = 50\%$	✗
$I_1 \to I_2$	$2/3 = 66\%$	✓
$I_2 \to I_3$	$3/4 = 75\%$	✓
$I_3 \to I_2$	$3/3 = 100\%$	✓

$I_2 \to I_1$ not association rule as its confidence is less than min. confidence.

Final rules,
$I_1 \to I_2$, $I_2 \to I_3$, $I_3 \to I_2$

- PACKAGES IN R:
- arules : represent, manipulate & analyze txn data & patterns
- arulesViz : Extends arules for visualisation of association rules & itemset
- tidyverse : data science
- knitr : dynamic report generation

- lubridate : date & time

- plyr : splitting, applying, combining data.

CONCLUSION:
Hence Market Basket Analysis using Apriori
algorithm has been understood & Implemented.