DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Experiment - 1.3

Student Name: Praduman Kumar **UID: 20BCS9446**
Branch: BE-CSE **Section/Group: 714/A**
Semester: 6th **Date of Performance: 03/03/2023**
Subject Name: Data Mining Lab **Subject Code: 20CSP-376**

1. Aim:

Demonstration of association rule mining using Apriori algorithm on supermarket data .

2. Objective:

To demonstration of association rule mining using Apriori algorithm on supermarket data .

3. Summary :

ARM : ARM (Association Rule Mining) is a popular technique in data mining that involves discovering patterns or relationships among data items in large databases. Specifically, it is used to identify frequent itemsets and association rules, which are often used for making recommendations or predicting customer behavior.

In ARM, the data is represented as transactions or sets of items, and the algorithm searches for frequent itemsets, which are subsets of items that appear together in a significant number of transactions. Once frequent itemsets are identified, association rules are generated by examining the relationships between the items in the itemsets.

APRIORI ALGORITHM : The Apriori algorithm is a classic algorithm used in association rule mining, a technique used in data mining to discover interesting relationships or patterns between items in a dataset. The algorithm is named after the fact that it employs an "a priori" knowledge of frequent itemsets to efficiently search for all association rules.

The Apriori algorithm is widely used in market basket analysis, customer behavior analysis, and other areas where discovering relationships between items is important. Its efficiency and effectiveness in identifying frequent itemsets and association rules have made it a popular and widely used algorithm in the field of data mining.

In data mining, support and confidence are two measures used in association rule mining to identify patterns or relationships between different items in a dataset.

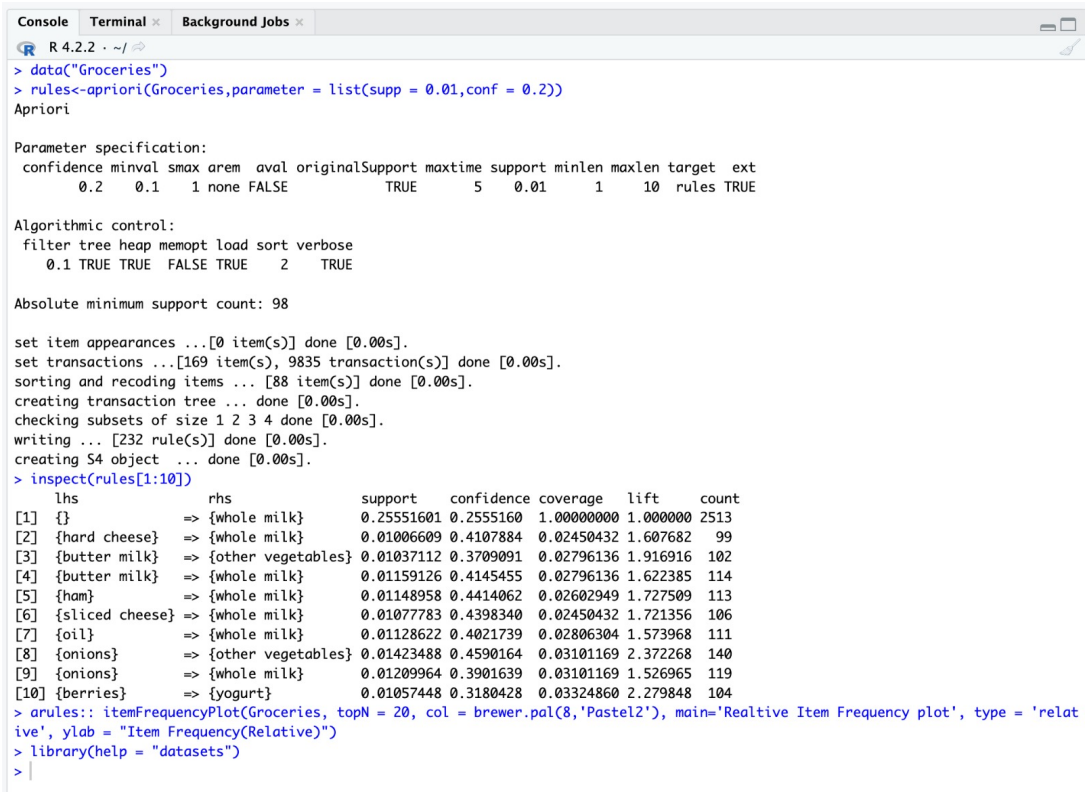
Support refers to the frequency with which a particular itemset appears in a dataset. It is calculated as the number of transactions that contain the itemset divided by the total number of transactions in the dataset. The support value ranges from 0 to 1, where 0 means the itemset never appears in the dataset, and 1 means it appears in all transactions.

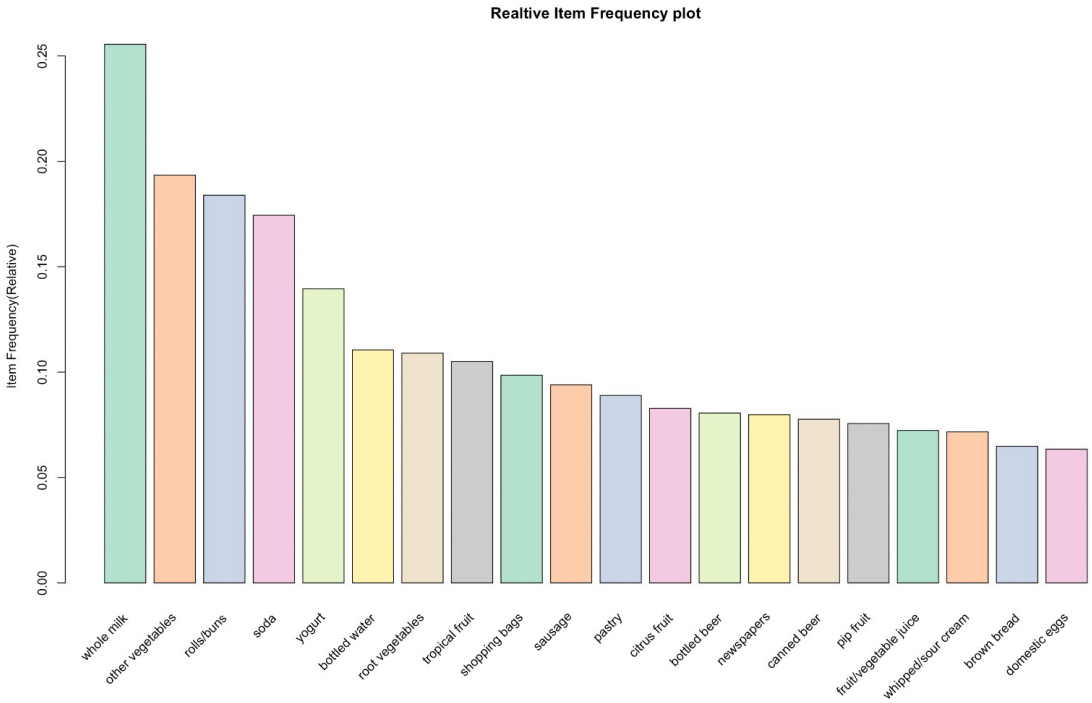
Confidence, on the other hand, measures the strength of the association between two items in an itemset. It is calculated as the number of transactions that contain both items divided by the number of transactions that contain the first item. The confidence value also ranges from 0 to 1, where 0 means no association, and 1 means a perfect association between the items.

4. Script/Code/Steps:

```
install.packages("arules")

install.packages("arulesViz")

install.packages("RColorBrewer")

#import dataset

data("Groceries")

#using apriori() function

rules<-apriori(Groceries,parameter = list(supp = 0.01,conf = 0.2))

#using inspect() function

inspect(rules[1:10])

#using itemFrequency() function

arules:: itemFrequencyPlot(Groceries, topN = 20, col = brewer.pal(8,'Pastel2'), main='Realtive Item Frequency plot', type = 'relative', ylab = "Item Frequency(Relative)")

library(help = "datasets")
```

4. Output Screenshots :




Realtive Item Frequency plot
Item Frequency(Relative)
0.00
0.05
0.10
0.15
0.20
0.25
whole milk
other vegetables
rolls/buns
soda
yogurt
bottled water
root vegetables
tropical fruit
shopping bags
sausage
pastry
citrus fruit
bottled beer
newspapers
canned beer
pip fruit
fruit/vegetable juice
whipped/sour cream
brown bread
domestic eggs