

### Learning Objectives:

- To understand the applications of association rule mining
- To understand the basic flow of rule induction in Apriori algorithm
- To understand the concepts of support, lift and confidence of rules
- To implement association rules using Apriori algorithm in R

### Activity:

- a. Using the purchase data of stationery store across three days given below:

Trans Id	Item Purchased	Trans Id	Item Purchased
1	Pencils	5	Erasers
1	Markers	6	Envelop
1	Highlighters	7	Markers
1	Papers	7	Erasers
2	Markers	8	Pencils
2	Erasers	8	Markers
3	Stapler Pins	8	Stapler Pins
3	Papers	8	Post-It
3	Erasers	8	Highlighter
3	Card holders	8	Papers
3	Highlighters	8	Erasers
4	Papers	9	Stapler Pins
4	Erasers	9	Post-It
4	Card holders	9	Markers
5	Markers	9	Erasers
5	Post-It	10	Envelop

1. Manually compute rules using minimum support = 50%
2. Compute lift for the above rules



Activity Sheet  
Pb#1.xlsx

- b. Find the support confidence and lift for the rule, {Milk, Diaper} → Beer

TID	Items
1	Chips, Milk
2	Chips, Diaper, Beer, Cornflakes
3	Milk, Diaper, Beer, Pepsi
4	Chips, Milk, Diaper, Beer
5	Chips, Milk, Diaper, pepsi

## c. Association Rules for transaction data :

Steps to follow:

- Install and load 'arules' package  
`install.packages("arules")`
- Read 'Transactions.csv' data into R such that the arules package treats the input csv file as "transaction" data.  
`trans = read.transactions(file="Transactions.csv", rm.duplicates= FALSE,  
format="single",sep="," ,cols =c(1,2))`
- Explore and understand the data and items of transaction data  
`inspect(trans)`  
`trans`  
`image(trans)`  
  
`itemFrequency(trans)`  
`itemFrequencyPlot(trans)`
- Implementing association mining using 'Apriori' algorithm to extract rules  
`rules <- apriori(trans,parameter = list(sup = 0.5, conf = 0.6,target="rules"))`
- Understanding the rules  
`summary(rules)`  
`inspect(rules)`

**Assignment**

## d. Association Rules for "Flight Delay" dataset: Generate the rules and identify the patterns.

1. Read the data into R
2. Look at the summary of all the variables and convert the following variables as factors
  - Weather
  - DAY\_WEEK
  - Flight Status
3. Bin the numeric variable 'CRS\_DEP\_TIME' into 4 bins as follows: If time is between 6 AM to 12 noon code as 1 and If the time is between 2 , etc.

4. Convert the data frame in a transactions object. Look at the first 6 transactions to understand how 'arules' library requires data to be transformed. Use the below R code to view the transactions data
5. Apply 'arules' algorithm and play with various support, lift and confidence values.
6. Inspect all the rules.
7. Filter the rules with specific LHS and RHS conditions  
E.g.; Filter the rules with Flighstatus=0
8. Write the Rules into CSV