# Load the libraries

library(arules)

library(arulesViz)

library(datasets)

# Load the data set

data(Groceries)

# Create an item frequency plot for the top 20 items

itemFrequencyPlot(Groceries,topN=20,type="absolute")

# Get the rules

rules <- apriori(Groceries, parameter = list(supp = 0.001, conf = 0.8))

# Show the top 5 rules, but only 2 digits

options(digits=2)

inspect(rules[1:5])

set of 410 rules

rule length distribution (lhs + rhs): sizes

3 4 5 6

29 229 140 12

summary of quality measures:

support conf. lift

Min. :0.00102 Min. :0.80 Min. : 3.1

1st Qu.:0.00102 1st Qu.:0.83 1st Qu.: 3.3

Median :0.00122 Median :0.85 Median : 3.6

Mean :0.00125 Mean :0.87 Mean : 4.0

3rd Qu.:0.00132 3rd Qu.:0.91 3rd Qu.: 4.3

Max. :0.00315 Max. :1.00 Max. :11.2

mining info:

data n support confidence

Groceries 9835 0.001 0.8

rules<-sort(rules, by="confidence", decreasing=TRUE)

subset.matrix <- is.subset(rules, rules)

subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA

redundant <- colSums(subset.matrix, na.rm=T) >= 1

rules.pruned <- rules[!redundant]

rules<-rules.pruned

rules<-apriori(data=Groceries, parameter=list(supp=0.001,conf = 0.08),

appearance = list(default="lhs",rhs="whole milk"),

control = list(verbose=F))

rules<-sort(rules, decreasing=TRUE,by="confidence")

inspect(rules[1:5])