IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Yicun Deng
Class: M003
Date: 10/29/20
Chapter Number: #17
Title of Chapter: Hi Ho, Hi Ho—Data Mining We Go

Module 9 - Association Rules Mining

Data mining is using R or other tools for the computer to generate interesting and novel pattern within data
Different dataset embedded different relationship between data points and this needs to be accounted. Just like the supermarket dataset
Support value:0.67
Diapers and beer occurred in two out of three carts
Confidence value:0.5
Beer occurs 50% of the time diapers were in the cart
Lift value:2.5(probability of beer in cart:20%)
[0,5/0.2(0.5 from confidence/0.2 from beer itself)]
Real world constraints of rules should be taken into consideration too
Install.packages("arules")
Library("arules")
data(Groceries)
summary(Groceries)
itemFrequencyPlot(Groceries,support=0.1)
itemFrequencyPlot(Groceries,support=0.05,cex.names=0.5)
apriori()(LHS left hand, predict right side, multiple items, RH right hand, predict left side, one item)
apriori(Groceries,parameter=list(support=0.005,confidence=0.5))
use summary()
inspect()

Module 9 - R Code for Association Rules Mining

The data itself has bias
Install.packages("arulesViz")
Library(arulesViz)
plot(ruleset)
goodrules <- ruleset[quality(ruleset)$lift>3.5]
inspect(goodrules)
data<- g@data
means<-rowMeans(data)
labels<-g@itemInfo[,1]
labels[1:10]
levels<-g@itemInfo$level2
levels[1:10]

IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

```
df<-data.frame(means,labels,levels)
df1<-df[order(-df$means),]
df1[1:10,]
```

Exercise Review
Load the dataset from Epub data from Vienna University of Economics. Use apriori()
Of support=0.005, confidence= 0.5. And see what the rules are using the plot()

Question for Class
How are you? Please have a good day!