**Eric Diaz**

**03.12.2019**

**HW 1: hrules**

**Script:**

##Eric Diaz

#HW1 02/25/2019

setwd("~/Documents/Senior/Spring 2019/INFS 348/HW1")

install.packages("arules")

library(arules)

## Some suggestions: clean the name of the columns in excel to make it easier to work with.

## Do some data cleaning in excel to make life easier.

heartdata <- read.csv("dataclean.csv", header=TRUE)

# convert the dataframe to transactions

#Remove those columns that are 100% correlated. Hint: Look at buff in data.

# how to do it? test1, we talked about subsetting dataframe. select only those columns that make sense.

heartdata = heartdata[1:13]

htrans = as(heartdata, "transactions")

# set better support and confidence levels to learn more rules. If you get a lot of rules increase the support or/and confidence.

# similarly if you get very few rules, decrease support or/confidence.

#change the minlen to limit the number of "items". So, 2 in this case means arules with 2 items

hrules <- apriori(htrans, parameter = list(support =0.20, confidence = 0.75, minlen = 2))

#subset for buff or healthy

hrules.sub <- subset(hrules, subset= rhs %in% "class=buff"

& lift > 1.1)

#subset for sick

hrules.sub2 <-subset(hrules, subset= rhs %in% "class=sick"

& lift>1.1)

# subset those rules that pertain to rhs = buff or sick etc.

hrules\_df<- as(hrules, "data.frame")

hrules.sub<- as (hrules.sub, "data.frame")

hrules.sub2<- as (hrules.sub2, "data.frame")

str(hrules\_df)

str(hrules.sub)

str(hrules.sub2)

# write this data frame to a text file that we will analyze usin g Tableau.

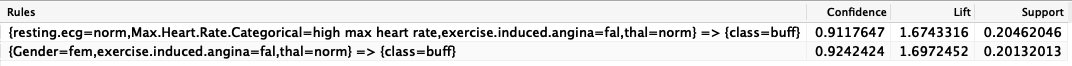
write.table(hrules\_df,file ="hrule.txt",sep="|",quote = FALSE,row.names=FALSE)

write.table(hrules.sub, file = "hruleBuff.txt", sep="|", quote = FALSE, row.names=FALSE)

write.table(hrules.sub2, file = "hruleSick.txt", sep="|", quote = FALSE, row.names=FALSE)

**Interesting Rules:**

**Healthy:**

****

After learning about the formulas for support, confidence and lift I decided to focus on the rules that only had low support and high confidence as well as a higher lift. Rules that have all of these attributes have the highest support, confidence and lift.

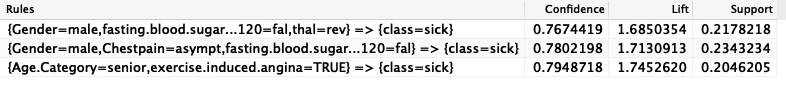
1. **{resting.ecg=norm,Max.Heart.Rate.Categorical=high max heart rate,exercise.induced.angina=fal,thal=norm} => {class=buff}**

This rule suggest that people with normal electrocardiographic outcomes, a high max heart rate, and no exercise induced angina, and a normal heart status (thal =norm), which 90% of the time is characteristic of healthy (buff) people.

1. **{Gender=fem,exercise.induced.angina=fal,thal=norm} => {class=buff}**

This rule suggests that people that identify as the femal gender, with no exercise induced angina, and a normal heart status (thal =norm), which 91% of the time is characteristic of healthy (buff) people.

**Sick:**



1. **{Gender=male,fasting.blood.sugar...120=fal,thal=rev} => {class=sick}**

This rule suggests that people who identify with the male gender, do not have a fasting blood sugar, with a heart defect that is reversible are 76.7% characteristic of sick people.

1. **{Gender=male,Chestpain=asympt,fasting.blood.sugar...120=fal} => {class=sick}**

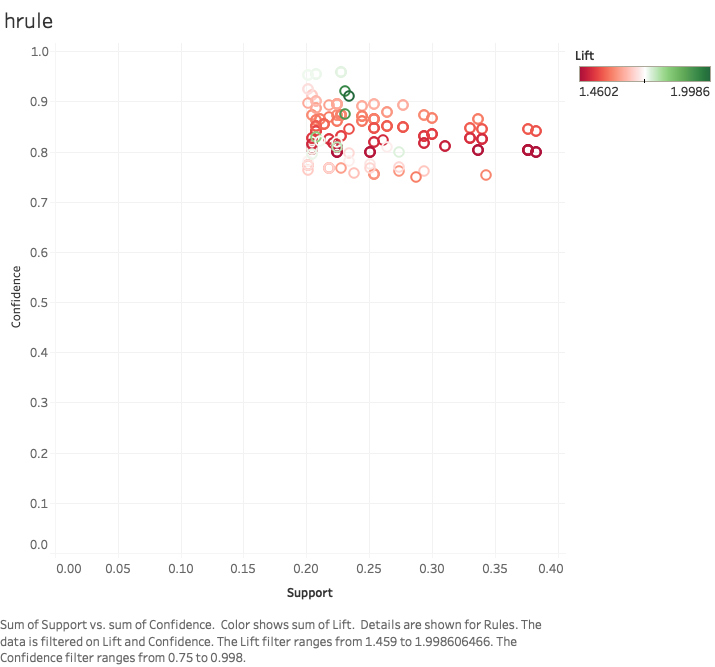
This rule suggests that people who identify with the male gender, are having no chest pain symptoms, and do not have a fasting blood sugar level, which 78% are characteristic of sick people.

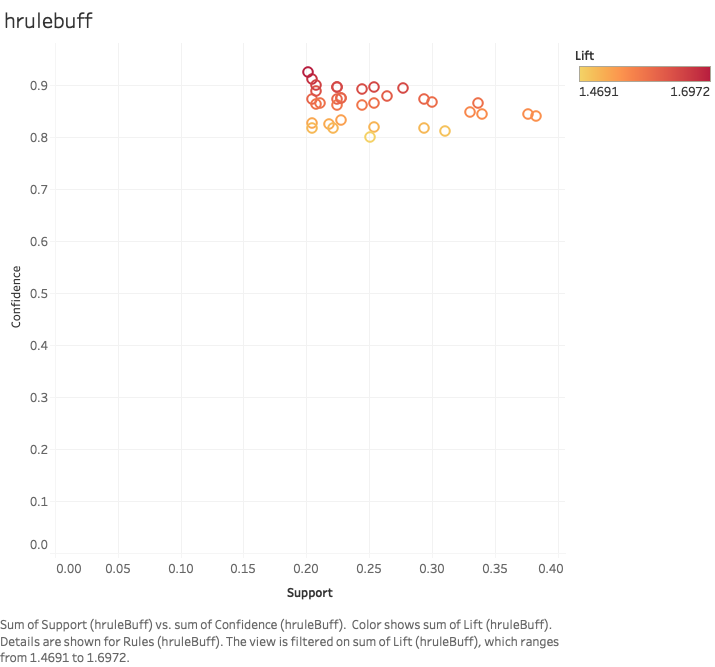
1. **{Age.Category=senior,exercise.induced.angina=TRUE} => {class=sick}**

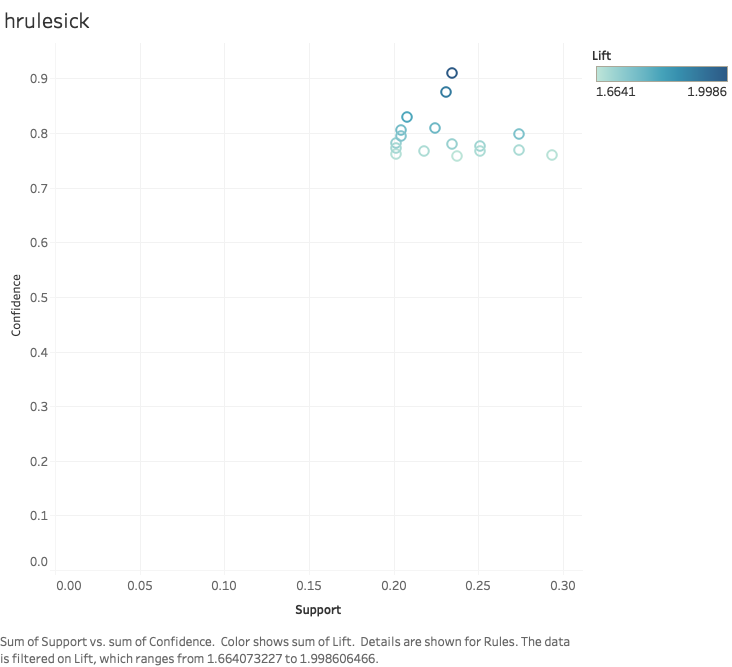
This rule suggests that people who are considered in the age category of seniors who had exercise induced angina are 79.4% characteristic of sick people.

**Images of Rules (supp vs. confidence):**

*General image of all of the “arules” with the respective lift.*

****

****

******

*Image of Rules with Highest Lift, Confidence*

