**CHAID in R**

### R version should more than 3.4.X and R-Studio can be any version or you see this error

“package ‘CHAID’ is not available (for R version 3.1.3)”

### **Install the CHAID library in R terminal**

install.packages("CHAID", repos="[http://R-Forge.R-project.org](http://R-Forge.R-project.org/)")

### **Load the library**

library(CHAID)

**# Data Sets(Play)**

# Data sets Features – Outlook,Temperature,Humidity,Wind,Play\_tennis

# Dependent Variable- Play\_tennis

# Independent variable - Outlook,Temperature,Humidity,Wind

# Some of the datasets

"Outlook","Temperature","Humidity","Wind","Play\_tennis"

1,"Sunny","Hot","High","Weak","No"

2,"Sunny","Hot","High","Strong","No"

3,"Overcast","Hot","High","Weak","Yes"

4,"Rain","Mild","High","Weak","Yes"

5,"Rain","Cool","Normal","Weak","Yes"

**### Fit tree to data**

#CHAID class works on three arguments

# Creating the object of CHAID class

**chaid\_object <- chaid(formula=,control=,data=)**

**1. Formula**

**2. Controling Parameter of spliting of tree**

**3. Data set**

**1. Formula**

formula = Dependent\_Variable ~ All\_Independent\_Variable\_You\_Need

Example on above dataset(Play)

formula = Play\_tennis ~ Outlook+Temperature+Humidity+Wind

or(use “.” To select all Ivs or “+” to select individual)

formula = Play\_tennis ~.

**2.Controling Parameter of spliting of tree will use in CHAID object**

# synatx of chaid\_control (Default values)

control\_vector\_name <- chaid\_control(alpha2 = 0.05, alpha3 = -1, alpha4 = 0.05,

minsplit = 20, minbucket = 7, minprob = 0.01,

stump = FALSE, maxheight = -1)

Arguments of chaid\_control()

Note: -1 represent max

1. alpha2 = Level of significance used for merging of predictor categories

2. alpha3 = level of significance used for the the splitting of former merged categories of the predictor.

3. alpha4 = Level of significance used for splitting of a node in the most significant predictor.

4. minsplit = Number of observations in splitted response at which no further split is desired.

5. minbucket = Minimum number of observations in terminal nodes.

6. minprob = Mininimum frequency of observations in terminal nodes.

7. stump = only root node splits are performed.

8. maxheight = Maximum height for the tree.

Example

vector\_name <- chaid\_control(minsplit = 200, minprob = 0.1)

**3. Data set**

data = name\_of\_DataFrame

**All at one place**

**control\_vector\_name <- chaid\_control(minsplit = 200, minprob = 0.1)**

**dt.chaid <- chaid(formula = Play\_tennis ~., control = control\_vector\_name, data = dataFrame\_Name)**

# Summary of Tree

**summary(dt.chaid)**

# Plot the tree

**plot(dt.chaid)**

# Show the text in plot

**text(dt.chaid)**