###########################################################################

Feature Selections

########################################################################

install.packages("FSelector")

library("FSelector")

NOTE:- If we are using FSelector package for feature selection then in case of Window-7 32 bit it can be install and included, however in case of window7 64bit it can install but not include through library and error of rJava package come, so you have disable the JAVA\_HOME system variable.

1. R command to check JAVA\_HOME

Sys.getenv("JAVA\_HOME")

1. R Command to set the JAVA\_HOME

Sys.setenv(JAVA\_HOME="")

#########################################################################

CFS Filter

#########################################################################

setwd("D:/Experimentjournal/catalogs/GoFFeatureSelection/Original")

Dataset.file<- "D:/Experimentjournal/catalogs/GoFFeatureSelection/80Sparse/80sparse\_GoF\_TFIDF.csv"

Dataset <- read.csv(Dataset.file,header=TRUE)

Cate <- Dataset[,1]

dat<- Dataset[,-1]

subset<- cfs(Cate~., dat)

f <- as.simple.formula(subset, "Category")

print(f)

#########################################################################

Chi-square

#########################################################################

setwd("D:/Experimentjournal/catalogs/GoFFeatureSelection/Original")

Dataset.file<- "D:/Experimentjournal/catalogs/GoFFeatureSelection/80Sparse/80sparse\_GoF\_TFIDF.csv"

Dataset <- read.csv(Dataset.file,header=TRUE)

Cate <- Dataset[,1]

dat<- Dataset[,-1]

weights<- chi.squared(Cate~., dat)

subset<- cutoff.k(weights, 10)

f <- as.simple.formula(subset, "Category")

print(f)

#########################################################################

Correlation

#########################################################################

setwd("D:/Experimentjournal/catalogs/GoFFeatureSelection/Original")

Dataset.file<- "D:/Experimentjournal/catalogs/GoFFeatureSelection/80Sparse/80sparse\_GoF\_TFIDF.csv"

Dataset <- read.csv(Dataset.file,header=TRUE)

Cate <- Dataset[,1]

dat<- Dataset[,-1]

weights<- linear.correlation(Cate~., dat)

subset<- cutoff.k(weights, 10)

f <- as.simple.formula(subset, "Category")

print(f)

#########################################################################

Information Gain

#########################################################################

setwd("D:/Experimentjournal/catalogs/GoFFeatureSelection/Original")

Dataset.file<- "D:/Experimentjournal/catalogs/GoFFeatureSelection/Original/Orig\_GoF\_Enteropy.csv"

Dataset <- read.csv(Dataset.file,header=TRUE)

Cate <- Dataset[,1]

dat<- Dataset[,-1]

weights<- information.gain (Cate~., dat)

subset<- cutoff.k(weights, 20)

f <- as.simple.formula(subset, "Category")

print(f)

#########################################################################

Gain Ratio

#########################################################################

setwd("D:/Experimentjournal/catalogs/GoFFeatureSelection/Original")

Dataset.file<- "D:/Experimentjournal/catalogs/GoFFeatureSelection/80Sparse/80sparse\_GoF\_TFIDF.csv"

Dataset <- read.csv(Dataset.file,header=TRUE)

Cate <- Dataset[,1]

dat<- Dataset[,-1]

weights<- gain.ratio (Cate~., dat)

subset<- cutoff.k(weights, 10)

f <- as.simple.formula(subset, "Category")

print(f)

#########################################################################

RandomForest

#########################################################################

setwd("D:/Experimentjournal/catalogs/GoFFeatureSelection/Original")

Dataset.file<- "D:/Experimentjournal/catalogs/GoFFeatureSelection/80Sparse/80sparse\_GoF\_TFIDF.csv"

Dataset <- read.csv(Dataset.file,header=TRUE)

Cate <- Dataset[,1]

dat<- Dataset[,-1]

weights<- random.forest.importance (Cate~., dat)

subset<- cutoff.k(weights, 10)

f <- as.simple.formula(subset, "Category")

print(f)

#########################################################################

oneR

#########################################################################

setwd("D:/Experimentjournal/catalogs/GoFFeatureSelection/Original")

Dataset.file<- "D:/Experimentjournal/catalogs/GoFFeatureSelection/80Sparse/80sparse\_GoF\_TFIDF.csv"

Dataset <- read.csv(Dataset.file,header=TRUE)

Cate <- Dataset[,1]

dat<- Dataset[,-1]

weights<- oneR (Cate~., dat)

subset<- cutoff.k(weights, 10)

f <- as.simple.formula(subset, "Category")

print(f)