library(caret)

library(dplyr)

library(readr)

#Combine 4 data files into one dataframe

Leedata2 <- list.files(path="D:/Work/TaylorsU/Lee Master Thesis/Dataset/Traning and Testing Tets (5% of the entier dataset)/All features", full.names = TRUE) %>%

 lapply(read\_csv) %>%

 bind\_rows

#Extract 10 features

Leedata3 <- select(Leedata2, c(flgs\_number,srate,drate,rate,max,state\_number,mean,min,stddev,seq,category))

#Category change to factor for classification

Leedata3$category <- as.factor(Leedata3$category)

#Static summary of dataframe

summary(Leedata3)

#Get data type of dataframe

apply(Leedata3, class)

#Extract 100 DDoS, 100 Dos, 100 Reconnaissance, and 76 Theft data

LeeDDos <- subset(Leedata3,category=="DDoS")

LeeDDos <- sample\_n(LeeDDos,100)

LeeDos <- subset(Leedata3,category=="DoS")

LeeDos <- sample\_n(LeeDos,100)

LeeRecon <- subset(Leedata3,category=="Reconnaissance")

LeeRecon <- sample\_n(LeeRecon,100)

LeeTheft <- subset(Leedata3,category=="Theft")

Leedata4 <- rbind(LeeDDos,LeeDos,LeeRecon,LeeTheft)

#Remove normal factor level

Leedata5 <- droplevels(Leedata4)

#export out dataset

write.csv(Leedata5,"D:/Work/TaylorsU/Lee Master Thesis/Dataset/Leedata5.csv",row.names=TRUE)

#split data to 80:20 ratio

LeeTrain <- createDataPartition(Leedata5$category, p=0.8, list=FALSE, times = 1)

Leetraining <- Leedata5[ LeeTrain, ]

Leetesting <- Leedata5[ -LeeTrain, ]

#Random Forest

library("randomForest")

Leemod\_RF <- train(category ~ ., data=Leetraining, method="rf")

Leepred\_RF = predict(Leemod\_RF, newdata=Leetesting)

confusionMatrix(Leepred\_RF,Leetesting$category)

# Random committee

library("kernlab")

Leemod\_ Random committee <- train(category ~ ., data=Leetraining, method=" Random committee Linear")

Leepred\_ Random committee = predict(Leemod\_ Random committee, newdata=Leetesting)

confusionMatrix(Leepred\_ Random committee Leetesting$category)