getwd()

library(dplyr)

install.packages("caret",dependencies=TRUE)

install.packages("randomForest")

library(caret)

library(randomForest)

data\_train\_person=read.csv("../input/cis660-2020fall/train\_participant\_data(trainset).csv")

glimpse(data\_train\_person)

head(data\_train\_person)

dim(data\_train\_person)

data\_train\_person$Obese = as.factor(data\_train\_person$Obese)

drop <- c("id","date","weight","height","start\_time","stop\_time","BMI\_level")

df\_person = data\_train\_person[,!(names(data\_train\_person) %in% drop)]

df\_person$Obese = as.factor(df\_person$Obese)

library(randomForest)

set.seed(71)

rf <-randomForest(Obese~.,data=df\_person, ntree=500)

print(rf)

floor(sqrt(ncol(df\_person) - 1))

mtry <- tuneRF(df\_person[-1],df\_person$Obese, ntreeTry=500,

stepFactor=1.5,improve=0.02, trace=TRUE, plot=TRUE)

best.m <- mtry[mtry[, 2] == min(mtry[, 2]), 1]

print(mtry)

print(best.m)

set.seed(71)

rf <-randomForest(Obese~.,data=df\_person, mtry=4, importance=TRUE,ntree=500)

print(rf)

importance(rf)

varImpPlot(rf)

library(randomForest)

library(mlbench)

library(caret)

control <- trainControl(method="repeatedcv", number=10, repeats=3)

seed <- 71

metric <- "Accuracy"

set.seed(seed)

mtry <- 4

tunegrid <- expand.grid(.mtry=8)

rf\_default <- train(Obese~., data=df\_person, method="rf", metric=metric, tuneGrid=tunegrid, trControl=control)

print(rf\_default)