LASSO regression of citalopram

#Libraries needed

library(glmnet)

library(caret)

library(nnet)

library(epiR)

library(ROCR)

#load data

data <- read.csv("/Users/krti1/Documents/823//STARDUNIQUE.csv")

View(data)

str(data)

ncol(data)

names(data)

#data preparation

#data preparation

data = subset(data, select = -c(1,2,4,5,6,7,8,9,10,11,13,14,15) )

#Data partition

set.seed(12345)

Citalopram<-sample(2, nrow(data), replace =T, prob=c(0.8,0.2))

train<- data[Citalopram==1,]

test<-data[Citalopram==2,]

str(train)

str(test)

# Custom Control Parameters

custom<-trainControl(method = "repeatedcv",

number=10,repeats =5,verboseIter=T)

#Lasso

set.seed(12345)

lasso<-train(Citalopram~.,

train,method='glmnet',

tuneGrid=expand.grid(alpha=1,lambda=seq(0.001,1,length=100)),

trControl=custom)

lasso$bestTune

best <- lasso$finalModel

coef(best, s = lasso$bestTune$lambda)

#save final model

saveRDS(lasso,"final\_model.rds")

fm <- readRDS("final\_model.rds")

print

summary(fm)

#prediction

p1 <- predict(fm, train)

a<-(sqrt(mean((train$Citalopram-p1)^2)))

b<-(sqrt(mean((train$Citalopram-mean(train$Citalopram))^2)))

#r2square-percent of variation

(1-(a/b))\*100

p2 <- predict(fm, test)

a<-(sqrt(mean((train$Citalopram-p2)^2)))

b<-(sqrt(mean((train$Citalopram-mean(train$Citalopram))^2)))

#r2square

(1-(a/b))\*100

sqrt(mean((test$remission-p2)^2))

#rmse

sqrt(mean((train$Citalopram-p1)^2))

sqrt(mean((test$Citalopram-p2)^2))