**a. Predict the no of comments in next H hrs**

**Ans:**

**predstrain<-predict(reducedmodel,train,interval = 'confidence')**

**b. Use regression technique**

Ans:

setwd("E:/Dataset/Dataset/Training")

data<-read.csv("Features\_Variant\_1.csv")

head(data)

names(data)<-c(1:54)

names(data) <- paste("I",colnames(data), sep="")

data1<-data[,-which(names(data) == "I54")]

correlation<-cor(data$I54,data1)

library("dplyr")

set.seed(123)

ind <- sample(2, nrow(data), replace = TRUE, prob = c(0.7, 0.3))

train <- data[ind==1,]

test <- data[ind==2,]

head(train)

mymodel<-lm(I54~I3+I5+I6+I7+I8+I9+I10+I11+I12+I13+I14+I16+I17+I18+I19+I20+I21+I22+

I23+I24+I25+I26+I27+I29+I30+I31+I32+I33+I34+I35+I37,data = train )

summary(mymodel) #by seeing the summary we can remove some more

varibles which are not significantt

reducedmodel<-lm(I54~I3+I5+I6+I7+I8+I9+I13+I20+I21+I22+

I23+I24+I30+I31+I32+I33+I35+I37,data = train )

anova(reducedmodel,mymodel)

**c. Report the training accuracy and test accuracy**

Ans:

predstrain<-predict(reducedmodel,train,interval = 'confidence')

train\_preds <- data.frame(cbind(actuals=train$I54, predicteds=predstrain))

correlation\_accuracy <- cor(train\_preds)

RESULT:

actuals fit lwr upr

actuals 1.0000000 0.5939358 0.5931398 0.5936519

fit 0.5939358 1.0000000 0.9990512 0.9991518

lwr 0.5931398 0.9990512 1.0000000 0.9964105

upr 0.5936519 0.9991518 0.9964105 1.0000000

preds<-predict(reducedmodel,test,interval = 'confidence')

test\_preds <- data.frame(cbind(actuals=test$I54, predicteds=preds))

correlation\_accuracy <- cor(test\_preds)

correlation\_accuracy

actuals fit lwr upr

actuals 1.0000000 0.5154665 0.5140881 0.5160187

fit 0.5154665 1.0000000 0.9992071 0.9992863

lwr 0.5140881 0.9992071 1.0000000 0.9969901

upr 0.5160187 0.9992863 0.9969901 1.0000000