**1. Use the below given data set**

**Data Set**

**2. Perform the below given activities:**

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

ANS:

x=data.matrix((subset(train, select = -I54)), rownames.force = NA)

y=data.matrix((train$I54), rownames.force = NA)

library(lars)

fit <- lars(x=x,y=y,type="lasso")

# summarize the fit

summary(fit)

#result

fit

# select a step with a minimum error

best\_step <- fit$df[which.min(fit$RSS)]

best\_step

# make predictions

predictions <- predict(fit, x, s=best\_step, type="fit")$fit

Note:-

**1. Use LASSO, Elastic Net and Ridge and other regression techniques that are covered in the**

**module**

ANS:

library(mlbench)

library(glmnet)

library(caret)

library(psych)

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="")

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)

custom<-trainControl(method = 'repeatedcv',

number=10,

repeats=5,

verboseIter = T)

#linear model

lm<-train(I54~.,

train,

method='lm',

trControl=custom)

#ridge regression

ridge<-train(I54~.,

train,

method='glmnet',

tuneGrid=expand.grid(alpha=0,

lambda=seq(0.0001,1,length=5)),

trControl=custom)

#lasso regression

set.seed(1234)

lasso<-train(I54~.,

train,

method='glmnet',

tuneGrid=expand.grid(alpha=1,

lambda=seq(0.0001,1,length=5)),

trControl=custom)

#Elastic net regression

set.seed(1234)

en<-train(I54~.,

train,

method='glmnet',

tuneGrid=expand.grid(alpha=seq(0,1,length=10),

lambda=seq(0.0001,1,length=5)),

trControl=custom)

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

ANS:

#linear model

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

predtest<-predict(lm,test,interval='confidence')

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

correlation\_accuracy\_lm <- cor(train\_preds)

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

correlation\_accuracy\_lm<- cor(test\_preds)

#ridge regression

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

predtest<-predict(ridge,test,interval='confidence')

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

correlation\_accuracy\_ridge <- cor(train\_preds)

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

correlation\_accuracy\_ridge<- cor(test\_preds)

#lasso regression

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

predtest<-predict(lasso,test,interval='confidence')

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

correlation\_accuracy\_lasso<- cor(train\_preds)

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

correlation\_accuracy\_lasso<- cor(test\_preds)

#Elastic net regression

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

predtest<-predict(en,test,interval='confidence')

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

correlation\_accuracy\_en<- cor(train\_preds)

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

correlation\_accuracy\_en<- cor(test\_preds)

**3. compare with linear models and report the accuracy**

Ans:

modle\_list<-list(Linearmodle=lm,Ridge=ridge,Lasso=lasso,ElasticNet=en)

res<-resamples(modle\_list)

summary(res)

Call:

summary.resamples(object = res)

Models: Linearmodle, Ridge, Lasso, ElasticNet

Number of resamples: 50

MAE

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

Linearmodle 7.625293 8.098122 8.281878 8.354208 8.578254 9.255513 0

Ridge 7.599021 7.983570 8.141161 8.205179 8.347624 9.428709 0

Lasso 6.668703 7.350526 7.528830 7.554320 7.856077 8.457384 0

ElasticNet 6.668703 7.350526 7.528830 7.554320 7.856077 8.457384 0

RMSE

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

Linearmodle 19.41145 24.14653 27.44947 28.07549 31.66233 42.33065 0

Ridge 20.05419 24.43864 27.73843 27.98851 30.78163 39.00985 0

Lasso 17.80462 25.17770 27.31164 27.66513 30.83183 35.33892 0

ElasticNet 17.80462 25.17770 27.31164 27.66513 30.83183 35.33892 0

Rsquared

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

Linearmodle 0.1338622 0.2721218 0.3437854 0.3383928 0.4147493 0.4854419 0

Ridge 0.1124264 0.2868360 0.3391347 0.3336796 0.3855723 0.5049407 0

Lasso 0.1916371 0.2855412 0.3394958 0.3382394 0.4074597 0.4989840 0

ElasticNet 0.1916371 0.2855412 0.3394958 0.3382394 0.4074597 0.4989840 0

**4. create a graph displaying the accuracy of all models**

Ans:

plot(correlation\_accuracy\_lm,correlation\_accuracy\_en,correlation\_accuracy\_lasso)