1. Use the below given data set

Data Set

Ans:

> df <- read.csv("https://archive.ics.uci.edu/ml/machine- learning databases/00273/Example\_WearableComputing\_weight\_lifting\_exercises\_biceps\_curl\_variations.csv ,header = FALSE)

>data<-df[-1,]

>head(data)

>colnames(data) <- as.character(unlist(data[1,]))

>data1 = data[-1, ]

>str(data1)

>data1<-as.data.frame(lapply(data1, as.integer))

>table(data1$classe)

>data1$classe<-as.factor(data1$classe)

>replacement\_vec <- c("A", "B", "C", "D","E")

>levels(data1$classe) <- replacement\_vec

>head(data1)

>set.seed(1234)

>pd<-sample(2,nrow(data1),replace = TRUE,prob = c(0.7,0.3))

>train<-data1[pd==1,]

>test<-data1[pd==2,]

2. Perform the below given activities:

a. Create classification model using different random forest models

Ans:

>library(randomForest)

>rf<-randomForest(classe~.,data =train)

>rf

>library(caret)

>p1<-predict(rf,train)

>confusionMatrix(p1,train$classe)

>p2<-predict(rf,test)

>confusionMatrix(p2,test$classe)

b. Verify model goodness of fit

Ans:

#goodness of fit

>actual<-table(test$classe)

>expected<-table(predict(rf,test))

>chisq.test(actual, p = expected/sum(expected))

c. Apply all the model validation techniques

Ans:

#####cross validation

>control <- trainControl(method = 'repeatedcv',

number = 5,

repeats = 3)

>set.seed(7)

>mtry <- sqrt(ncol(train))

>rf\_random <- train(classe~.,

data = train,

method = 'rf',

metric = 'Accuracy',

tuneLength = 7,

trControl = control)

>print(rf\_random)

>plot(rf\_random)

>predictions<- predict(rf\_random,test)

>pred<- cbind(test,predictions)

>confusionMatrix<- confusionMatrix(pred$predictions,pred$classe)

d. Make conclusions

Ans:

> varImp(rf\_random)

rf variable importance

only 20 most important variables shown (out of 158)

Overall

raw\_timestamp\_part\_1 100.000

num\_window 98.965

magnet\_dumbbell\_y 40.247

roll\_forearm 35.908

user\_name 34.010

magnet\_dumbbell\_z 22.303

magnet\_dumbbell\_x 18.639

cvtd\_timestamp 14.568

roll\_arm 13.708

accel\_belt\_y 10.535

pitch\_belt 10.279

yaw\_belt 9.859

roll\_belt 8.849

gyros\_dumbbell\_y 8.049

gyros\_belt\_z 7.481

pitch\_forearm 7.194

gyros\_arm\_y 6.421

yaw\_dumbbell 5.877

accel\_arm\_z 5.280

magnet\_belt\_x 3.586

e. Plot importance of variables

>varImpPlot(rf)

