PML Assignment

Ramsha Samreen

10/22/2020

For the week 4 Graded course project of practical machine learning using rstudio markdown and knitr

1) analysis

2) Introduction

Databases is collected from Nike band, Fitbit, jawbone, and we will be utilizing collected data for our analysis in this course assignment.

Now in this project, from the data of the accelerometer measure. of the individuals of their different-different class of physical activity

with the help of accessed data, we will be predict whether the individual is doing his exercises properly and the two files comprise testing and train data, we will basically predict the counting of exercise like the order of them as it should be predicted.

firstly of all load the data and proceed to process data and after we find exploratory analysis prediction for which model to select, finally for the predicting of the o/p of the testing set

```
library(data.table)
library(mlbench)
```

```
## Warning: package 'mlbench' was built under R version 4.0.3
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.0.3
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
library(klaR)
```

```
## Warning: package 'klaR' was built under R version 4.0.3
```

```
## Loading required package: MASS
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.0.3
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(rattle)
```

```
## Warning: package 'rattle' was built under R version 4.0.3
```

```
## Loading required package: tibble
```

```
## Warning: package 'tibble' was built under R version 4.0.3
```

```
## Loading required package: bitops
```

```
## Warning: package 'bitops' was built under R version 4.0.3
```

```
## Rattle: A free graphical interface for data science with R.
## Version 5.4.0 Copyright (c) 2006-2020 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```
##
## Attaching package: 'rattle'
```

```
## The following object is masked from 'package:randomForest':
##
##     importance

library(rpart)
library(rpart.plot)

## Warning: package 'rpart.plot' was built under R version 4.0.3

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.0.3

## corrplot 0.84 loaded
```

Now we will take data and clean to make it useful to exploring the data.

```
testUrl <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv"
traUrl  <- "https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv"

data_test <- read.csv(url(testUrl))
data_tra <- read.csv(url(traUrl))
```

Cleaning the Data Input

```
training_data <- data_tra[, colSums(is.na(data_tra)) == 0]
testing_data <- data_test[, colSums(is.na(data_test)) == 0]
```

now we will prepare the data for pred. in which we will consider 70% of the data for the training set and rest of which is thirty percentage of remaining data set to be tested and testing_data will be used furthaaaer again for the prediction of the 20% percentage of the cases

```
training_data <- training_data[, -c(1:7)]
testing_data <- testing_data[, -c(1:7)]
dim(training_data)

## [1] 19622    86

set.seed(1234)
datatraining <- createDataPartition(data_tra$classe, p = 0.7, list = FALSE)
training_data <- training_data[datatraining, ]
testing_data <- training_data[-datatraining, ]
```

```
dim(training_data)
```

```
## [1] 13737    86
```

```
dim(testing_data)
```

```
## [1] 4123   86
```

now we will remove the variables that are non zero from set of data which we have taken okayy

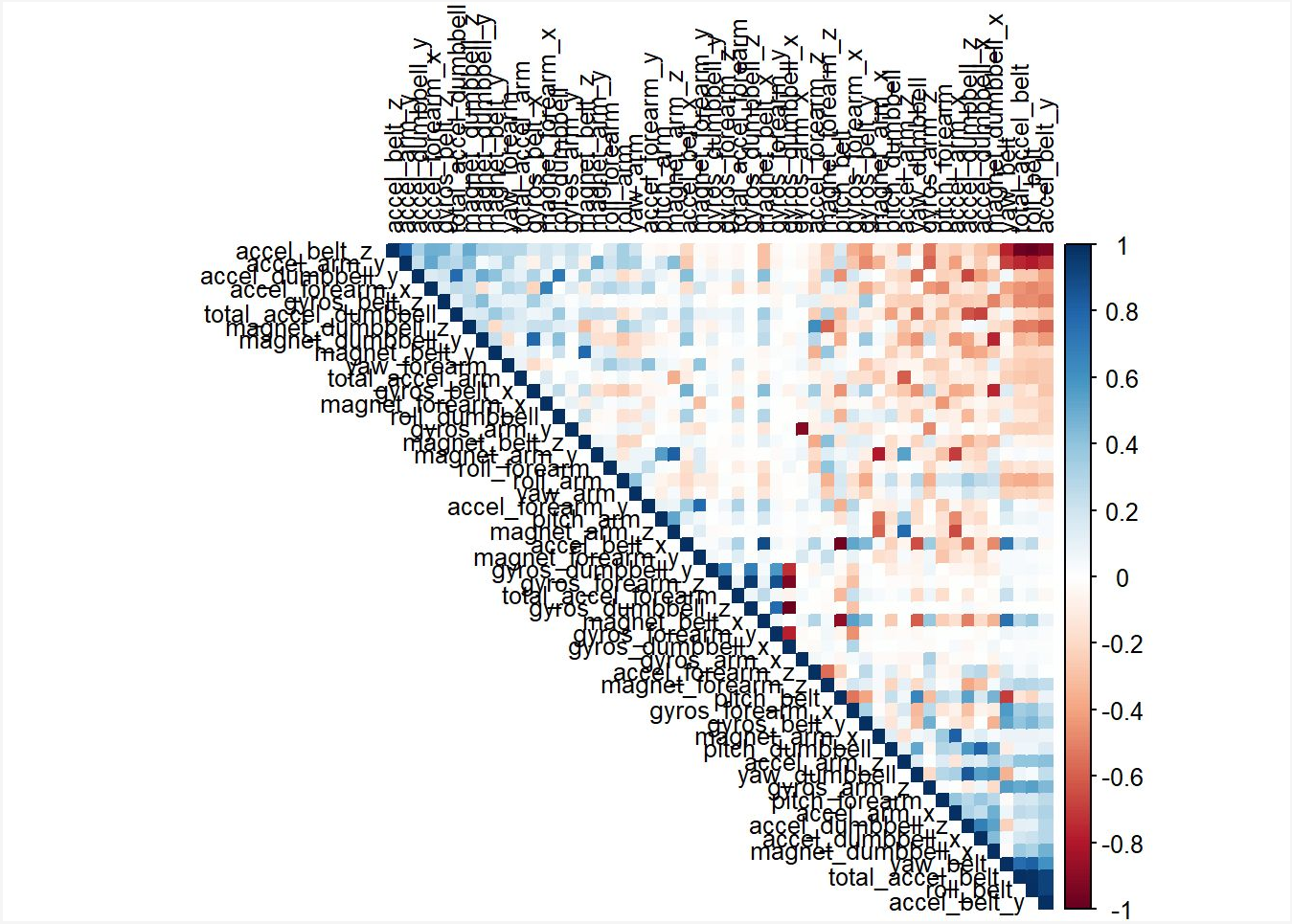
```
noneZero <- nearZeroVar(training_data)
training_data <- training_data[, -noneZero]
testing_data <- testing_data[, -noneZero]
dim(training_data)
```

```
## [1] 13737    53
```

```
dim(testing_data)
```

```
## [1] 4123   53
```

```
plot_cor <- cor(training_data[, -53])
corrplot(plot_cor, order = "FPC", method = "color", type = "upper", tl.cex = 0.8, tl.col =
rgb(0, 0, 0))
```

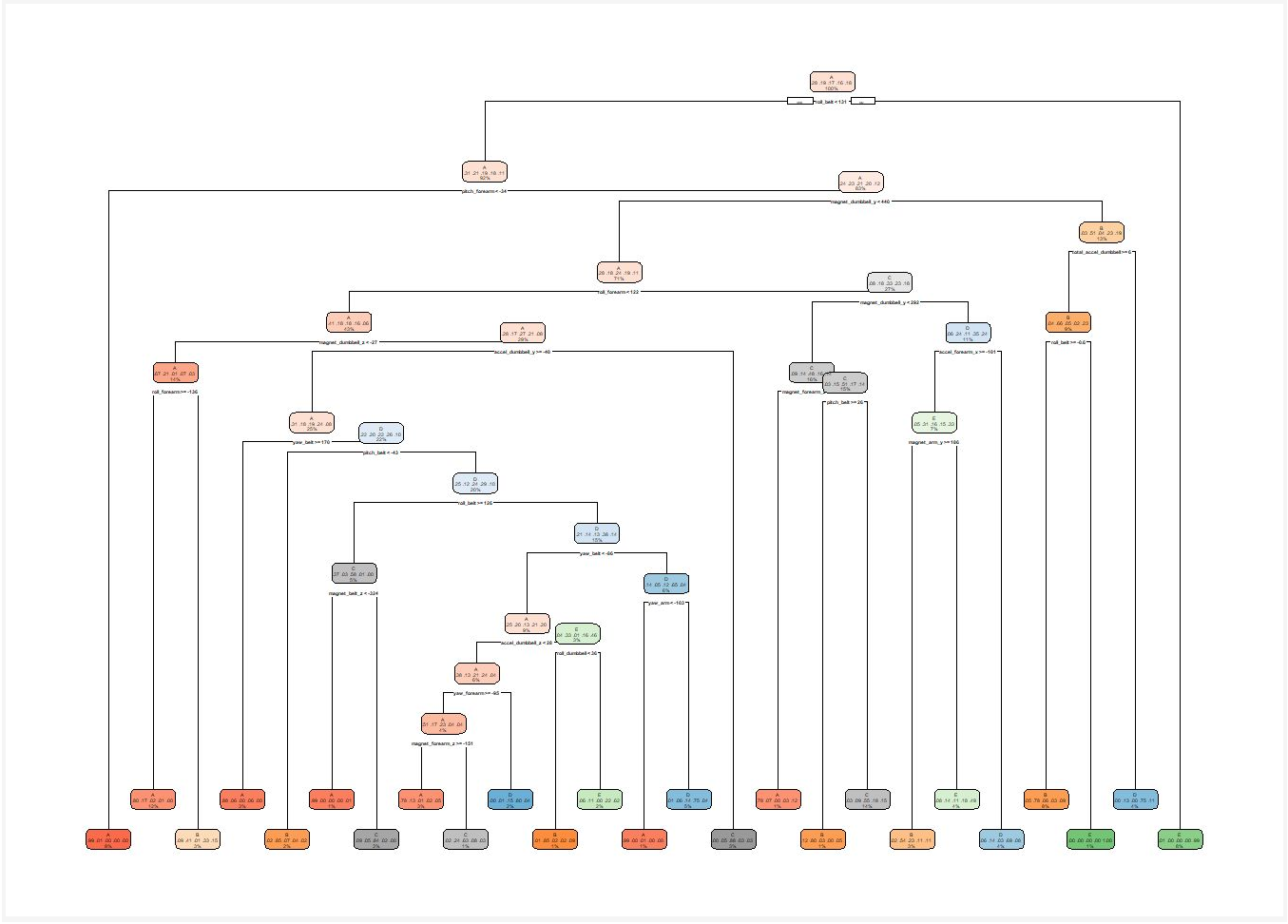


@@ now in this as we can see that the correcttt predictionsss are the ones with the dark colour intersectionnnnn @@

proceeding for the model building and for this we will use 2 different types of algorithms , trees and random forests for the prediction part oKayyy

```
set.seed(20000)
tredec <- rpart(classe ~ ., data=training_data, method = "class")
rpart.plot(tredec)
```

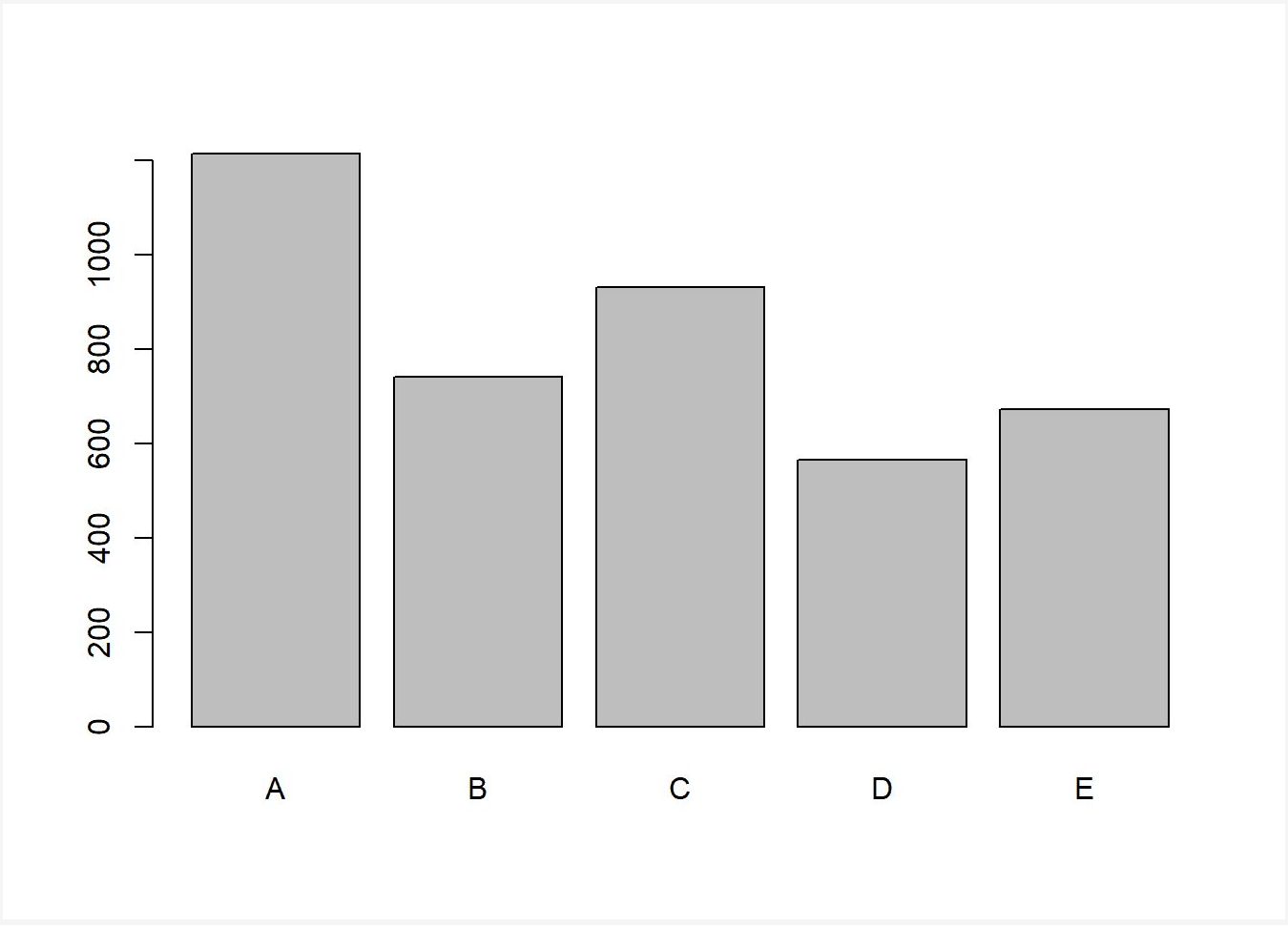
```
## Warning: labs do not fit even at cex 0.15, there may be some overplotting
```



now we will be validating our model and there we goo yess.. so good

```
modelpre <- predict(tredec, testing_data, type = "class")
#ab <- confusionMatrix(modelpre, testing_data$classe)
#ab

plot(modelpre)
```



we are going apply two models one by one lastly will apply the first is general boosted model and then the second oneis be gbm model for this

```
set.seed(10000)
ctr_gbm <- trainControl(method = "repeatedcv", number = 5, repeats = 1)
valid_gbm <- train(classe ~ .,data=training_data, method = "gbm", trControl = ctr_gbm,
verbose = FALSE)
valid_gbm$finalModel

## A gradient boosted model with multinomial loss function.
## 150 iterations were performed.
## There were 52 predictors of which 52 had non-zero influence.
```

So in this project, we tried to predict the order wise the someone did the exercise, and then we created the analysis in which we did some cross-validation and why I chose this specific way towards approaching and then predicted for 20. and i have attached the link to GitHub, which contained the HTML and rmd file. Still, due to some unprecedented reason, I could not attach the

file, which consisted of the output, so I have attached the pdf file and the rmd file. Please consider the request, and thank you it was an amazing experience overall thank you, mentors and university