Machine Learning

20 June 2015

Level-1 Reading Raw data from the directory

```
## Loading required package: lattice
## Loading required package: ggplot2
## Rattle: A free graphical interface for data mining with R.
## Version 3.4.1 Copyright (c) 2006-2014 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
## randomForest 4.6-10
## Type rfNews() to see new features/changes/bug fixes.
```

Level-2 Preprocessing: Removing columns with NA's and removing the first 6 coloumns which are not much useful.

```
training<-training[,colSums(is.na(training)) == 0]
training<-training[,-c(1:7)]
testing<-testing[,-c(1:7)]
```

Level-3 Creating training and testing set from the training data set, we are partitio in the ration of 75:25.

```
intrain<-createDataPartition(y=training$classe,p=0.75,list=FALSE)
training_data<-training[intrain,]
testing_data<-training[-intrain,]
```

Level-4 Training the model using the training_data set.

```
model2<-randomForest(classe~.,data=training_data,method="class")
```

Using the model we are doing prediction on the testing_data and comparing the actual output and the predicted output

```
##    predicted actual
## 1          A      A
## 13         A      A
## 15         A      A
## 17         A      A
## 18         A      A
## 21         A      A
```

Cross Validation

```
## Loading required namespace: e1071
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    A    B    C    D    E
##          A 1394    3    0    0    0
##          B    1  945    2    0    0
##          C    0    1  851    5    0
##          D    0    0    2  798    1
##          E    0    0    0    1  900
##
## Overall Statistics
##
##                Accuracy : 0.9967
##                  95% CI : (0.9947, 0.9981)
##     No Information Rate : 0.2845
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9959
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E
## Sensitivity            0.9993   0.9958   0.9953   0.9925   0.9989
## Specificity            0.9991   0.9992   0.9985   0.9993   0.9998
## Pos Pred Value         0.9979   0.9968   0.9930   0.9963   0.9989
## Neg Pred Value         0.9997   0.9990   0.9990   0.9985   0.9998
## Prevalence             0.2845   0.1935   0.1743   0.1639   0.1837
## Detection Rate         0.2843   0.1927   0.1735   0.1627   0.1835
## Detection Prevalence   0.2849   0.1933   0.1748   0.1633   0.1837
## Balanced Accuracy      0.9992   0.9975   0.9969   0.9959   0.9993
```

Submission Part

```
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##  B  A  B  A  A  E  D  B  A  A  B  C  B  A  E  E  A  B  B  B
## Levels: A B C D E
```